Electrical Power and Energy Systems 121 (2020) 106088

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Clustering methods assessment for investment in zero emission neighborhoods' energy system

Dimitri Pinel

Department of Electrical Power Engineering, NTNU, Elektrobygget, O. S. Bragstads plass 2E, E, 3rd floor, 7034 Trondheim, Norway

ARTICLE INFO

Keywords:
Clustering
Design
Optimization
Distributed energy resources
Zero emission

ABSTRACT

This paper investigates the use of clustering in the context of designing the energy system of Zero Emission Neighborhoods (ZEN). ZENs are neighborhoods who aim to have net zero emissions during their lifetime. While previous work has used and studied clustering for designing the energy system of neighborhoods, no article dealt with neighborhoods such as ZEN, which have high requirements for the solar irradiance time series, include a CO_2 factor time series and have a zero emission balance limiting the possibilities. To this end several methods are used and their results compared. The results are on the one hand the performances of the clustering itself and on the other hand, the performances of each method in the optimization model where the data is used. Various aspects related to the clustering methods are tested. The different aspects studied are: the goal (clustering to obtain days or hours), the algorithm (k-means or k-medoids), the normalization method (based on the standard deviation or range of values) and the use of heuristic. The results highlight that k-means offers better results than k-medoids and that k-means was systematically underestimating the objective value while k-medoids was constantly overestimating it. When the choice between clustering days and hours is possible, it appears that clustering days offers the best precision and solving time. The choice depends on the formulation used for the optimization model and the need to model seasonal storage. The choice of the normalization method has the least impact, but the range of values method show some advantages in terms of solving time. When a good representation of the solar irradiance time series is needed, a higher number of days or using hours is necessary. The choice depends on what solving time is acceptable.

1. Introduction

More than just accuracy, solving time and complexity are key elements that needs to be taken into account when designing optimization models. Indeed, certain applications require certain solving speeds. The unit commitment problem or the control of processes are good examples of applications that need a solution in time. In general, a shorter solving time increases the practicality of using the model. To keep the solving time within acceptable bounds, which needs to be defined on a case-by-case basis, different approaches are available. Some applications can accept sub-optimal solutions within an optimality gap and can simply stop the optimization after a given amount of time. In other cases, the complexity of the model can be reduced, by reducing the number of binary variables or changing the formulation of certain constraints. Finally, it is also possible to reduce the dimensionality of the problem. The time is one of the dimensions that can most commonly be reduced, by reducing the granularity (modelling hourly instead of every 15-min for instance) or by using clustering algorithms to group hours by features. Clustering algorithms can gather similar points from a dataset of any dimensions into groups called clusters. Each clusters is then represented by one point. Several methods can be used to assess how similar the points of the datasets are and how the representative of each clusters should be created.

In this paper, we investigate the use of clustering in a mixed integer linear program (MILP) called ZENIT. The goal is to identify which technique performs best for this application regarding the time necessary to solve the model, the optimality gap, and the representation of some timeseries of particular importance.

ZENIT (Zero Emission Neighborhood (ZEN) Investment Tool) is a program based on optimization that helps design the energy system of neighborhoods in a cost-optimal way and with a goal of having achieved net zero emissions of CO_2 in the neighborhood's lifetime. It is developed as a part of the research center on Zero Emission Neighborhoods in Smart Cities in Norway. The goal of this center is to research solutions to reduce the emission of neighborhoods in various fields such as architecture, urban planning and materials.

In this paper the focus of the clustering is a reduction of the time dimensionality, i.e. using less timesteps. The dimension of the dataset to

E-mail address: dimitri.q.a.pinel@ntnu.no.

https://doi.org/10.1016/j.ijepes.2020.106088
Received 29 July 2019; Received in revised form 25 March 2020; Accepted 6 April 2020
Available online 21 April 2020

Nomenclature

Nomenclature

$t\,(\mathcal{T})$	timestep in hour within year
$\kappa\,(\mathcal{K})$	cluster representative (centroid)
$t_\kappa\,(\mathscr{T}_\kappa)$	timestep within cluster κ
$b\,(\mathcal{B})$	building or building type
$i\,(\mathcal{I})$	energy technology, $\mathcal{I} = \mathcal{F} \cup \mathcal{E} \cup \mathcal{HST} \cup \mathcal{EST};\mathcal{I} = \mathcal{Q} \cup \mathcal{G}$
$f\,(\mathcal{F})$	technology consuming fuel (gas, biomass, …)
$e\,(\mathcal{E})$	technology consuming electricity
$hst\,(\mathcal{HST})$	heat storage technology
$est\,(\mathcal{EST})$	electricity storage technology
$q\,(\mathcal{Q})$	technologies producing heat
$g\,(\mathcal{G})$	technologies producing electricity
$b\,(\mathcal{B})$	building or building type
$C_{i,b}^{var,disc}$, $C_{i,b}^{fix,disc}$	variable/fix investment cost of i in b discounted to the beginning of the study including potential re-investments and salvage value [€/kWh]/[€]
$\varepsilon_{r,D}^{tot}$	discount factor for the duration of the study D with discount rate r
$C_{i,b}^{maint}$	annual maintenance cost of i in b [€/kWh]
P_f^{fuel}	price of fuel of g [€/kWh]
P_t^{spot}	spot price of electricity at t [€/kWh]
P^{grid}	electricity grid tariff [€/kWh]
P^{ret}	retailer tariff on electricity [€/kWh]
η_{est}, η_{hst}	efficiency of charge and discharge
η_i	efficiency of i
η_{inv}	efficiency of the inverter
$\phi_t^{CO_2,e}$	CO_2 factor of electricity at t [gCO_2/kWh]
$\phi^{CO_2,f}$	CO_2 factor of fuel type f [gCO_2/kWh]
α_{CHP}	heat to electricity ratio of the CHP
α_i	part load limit as ratio of installed capacity
GC	size of the neighborhood grid connection [kW]
X_i^{max}	maximum investment in i [kW]
X_i^{min}	minimum investment in i [kW]
$E_{b,t}$	electric load of b at t [kWh]
$H_{b,t}^{SH}$, $H_{b,t}^{DHW}$	heat (space heating/domestic hot water) load of b at t [kWh]
$COP_{hp,b,t}$	coefficient of performance of heat pump hp
$\dot{Q}_{st}^{max}$	maximum charge/discharge rate of est/hst [kWh/h]
IRR_t^{tilt}	total irradiance on a tilted plane [W/m^2]
G^{stc}	irradiance in standard test conditions: 1000 W/m^2
T^{coef}	temperature coefficient
T_t	ambient temperature at t [°C]
T^{noct}	normal operating cell temperature [°C]
T^{stc}	ambient temperature in standard test conditions [°C]
σ_κ	number of occurrences of cluster κ in the year
C^{HG}	cost of investing in the heating grid [€]
M	"Big M", taking a large value
B_q^{DHW}	binary parameter stating whether q can produce DHW
Q_{b_1,b_2}^{HGloss}	heat loss in the heating grid in the pipe going from b_2 to b_1
$\dot{Q}_{b_1,b_2}^{MaxPipe}$	maximum heat flow in the heating grid pipe going from b_2 to b_1 [kWh]
$P_{hp,b,t}^{input,max}$	maximum power consumption of hp at t based on manufacturer data and output temperature
b^{HG}	binary for the investment in the Heating Grid
$b_{i,b}$	binary for the investment in i in b
$x_{i,b}$	capacity of i in b
$f_{f,t,b}$	fuel consumed by f in b at t [kWh]
$d_{e,t,b}$	electricity consumed by e in b at t [kWh]
y_t^{imp}, y_t^{exp}	electricity imported from the grid to the neighborhood/exported at t [kWh]
$y_{t,g,b}^{exp}$	electricity exported by g to the grid at t [kWh]
$g_{t,g,b}^{selfc}$	electricity generated by g self consumed in the neighborhood at t [kWh]
$g_{t,g,b}^{ch}$	electricity generated by g used to charge the batteries at t [kWh]
$y_{t,est,b}^{imp}$	electricity imported from the grid to est at t [kWh]
$y_{t,est,b}^{exp}$	electricity exported from the est to the grid at t [kWh]
$g_{g,t,b}$	electricity generated by g at t [kWh]
$q_{q,t,b}$	heat generated by q in b at t [kWh]
$y_{t,est,b}^{dch}$	electricity discharged from est to the neighborhood at t [kWh]
$y_{t,est,b}^{ch}$	electricity charged from on-site production to est at t [kWh]
$q_{t,st,b}^{ch}$, $q_{t,st,b}^{dch}$	energy charged/discharged from the neighborhood to the storage at t [kWh]
$v_{t,st,b}^{stor}$	level of the storage st in building b at t [kWh]
$g_{t,b}^{curt}$	solar energy production curtailed [kWh]
$g_{g,t,b}^{dump}$	electricity generated but dumped by g at t [kWh]
$q_{t,b}^{dump}$	heat dumped at t in b [kWh]
$q_{b_1,b_2,t}^{HGtrans}$	heat transferred via the heating grid from b_1 to b_2 at t [kWh]
$q_{b,t}^{HGused}$	heat taken from the heating grid by b at t [kWh]
$o_{i,t,b}$	binary controlling if i in b is on or off at t
$\overline{x_{i,b,t}}$	maximum production from i [kWh]

cluster depends on the length of the time series used (usually a year: 8760 h) and the number of buildings in the neighborhood. The objective is to contain the solving time as well as keep a good representation of the original timeseries, with a particular focus on the solar irradiance.

Section 2 presents relevant existing literature regarding clustering in power systems applications and in particular for the design of neighborhoods energy systems and present the contribution of this paper. Section 3 presents the clustering methods investigated in this paper and Section 4, the results of those methods with regard to certain metrics. in Section 5 the optimization models are presented and the results of the clustering methods in the optimization are analyzed in Section 6.

2. State of the art and contribution

Clustering algorithms have been studied extensively since the 1930s [1] and improved since then. They are used in various applications across many fields. The principle of those algorithms is to gather similar observations of a dataset into clusters based on a given metric. The outputs of such algorithms are a list of all original data points and the cluster to which they belong as well as a representative vector for each cluster. Many algorithms exist but, in this paper the focus is on the k-means and the k-medoids algorithms because they are the most commonly use for power systems applications. These algorithms differ in the way the representative vector of each cluster are chosen. The k-means algorithm uses a centroid as the representative vector, i.e. the vector with the smallest squared distance to every member of the cluster [2]. The k-medoids algorithm chooses the representatives of the clusters by choosing the vector in the original data with the smallest distance to every other members of the cluster [3]. In the power systems field, it has been for example used in the context of grid expansion planning in [4], national energy system planning [5,6] and unit commitment models [7].

[5] suggests that the best clustering technique depends on the data to process and the model in which they are going to be used. It is thus important to compare different methods in order to find the best choice for our particular needs. It also gives insights in the choice of the

number of clusters to use. Several articles compare, with different approaches, the possible clustering techniques. Among them, [5] compares the performance of downsampling, k-means and hierarchical clustering as well as different heuristics and combinations of previously mentioned methods. It finds that for their energy system planning model and in the context of pluri-annual time series, some heuristics appear promising. The clustering is performed on days, with 4 different time series and multiple locations giving a rather large number of dimensions.

For a grid expansion planning problem, [4] compares systematic sampling, k-means, k-medoids, hierarchical clustering with Ward's linkage and moment matching. It clusters on hours and 5 dimensions. In this case, hierarchical and k-medoids appear to perform equally well.

Closer to the ZENIT model needs, Ref. [8] compared clustering algorithms (k-means, k-centers, k-medoids, k-medians, monthly averaged days, and seasonal days) to find representative days for a model investing and operating the energy system of a building. It finds k-medoids as the best suited method for this application.

Reference [9] also compares different techniques in the context of different local energy systems (averaging, k-means, k-medoids, hierarchical) for obtaining representative days, 3-days or weeks. It finds that medoids perform better than centroids but recommends overall the use of hierarchical clustering due to the reproducibility of the results.

It is also interesting to look at the choices made for other models similar to ZENIT, i.e. model for investment and operation in the energy system of buildings or neighborhoods. Those choices are naturally dependant on factors such as the scale of the neighborhood, the level of detail of the model, the target run time, the machine used to solve the model or its goal: investment and/or operation and in some cases grid layout, but it remains a good indication nonetheless.

Many authors choose to use season based clustering (SBC), where they choose or average the time series to form one representative day for each season [10] or only for the summer, the winter and the mid-season [11–13]. They also have varying choices in terms of number of periods for the chosen days: from hourly (i.e. 24 periods) [12], to twelve [11,12], or six periods [12,13]. Similarly, some choose to use one average day per month [14–16], or several days per month, such as [17] with a week day, a week end day and a peak day per month or [18] with 2 days of 12 periods each per month.

The exact method used to determine the days is not always clear [19]; points this out and suggests a graphical method using the load duration curves. Another method relying on k-means clustering is proposed in [20].

Reference [21] uses weekly downsampling to allow the model to run faster and checks the scheduling with a 24 h rolling horizon model with hourly resolution. Complete years with hourly resolution are also used in some models [22].

Other studies rely on clustering [8]. Reference [23] suggest a way to keep seasonal storage operation while using design days found with k-means clustering. Similarly [24] relies on k-medoids clustering to find design days. However, only outside temperature and global irradiance are used, assuming that the other time series are correlated to either of those two. The other time series are reconstructed from the clusters after the clustering. K-means is also used in [25], where two models are coupled, for providing representative weeks and for providing representative hours. The hours clustering is preceded by the removal of peaks from the time series and followed by their re-introduction.

In this paper different methods of clustering, normalizing and treating peaks are compared in the specific case of ZENIT. In addition, design days and representative hours are compared to find the strengths and weaknesses of each approach. This study stands out from other comparative studies by limiting the number of algorithm used but also considering the choices for normalizing and handling peaks. The Zero Emission context also brings specific problems to overcome. For example, the zero emission balance constraint in the optimization model limits the way one can reduce the number of timesteps. Another example is the strong requirements on the solar irradiance time series due to the importance of PV in the results. To the best of the author's knowledge, no paper tackles clustering in the context of ZEN or in a similar context.

This paper contributes to the existing literature regarding clustering in the context of power systems and in the context of the design of the energy system of neighborhoods by addressing the optimal clustering methods for designing the energy system of ZENs. This is important as the best method is specific to each application ([5]). In particular, it investigates the impact of the zero emission balance and other ZEN's specific requirements on the performance of clustering techniques. It also addresses two aspects that are little discussed in the existing literature: the choice of days or hours for the clusters and the impact of the normalization method.

3. Reduction of the number of timesteps

Many possibilities exist in order to reduce the number of timesteps in the optimization. However some are not adequate for the model. Downsampling for instance is not well suited. With the downsampling method, the time series are reduced by averaging the values on a certain period of time. A six hours downsampling would average the values of the time series on intervals of six hours, dividing by six the total number of timesteps. This method reduces the precision of the data and is not well suited for applications with renewable energies, which vary rapidly. The use of heuristic is often considered, and there are different approaches depending on the application. The heuristic could be reducing the time series to a collection of extreme events found in the time series, such as the hours with the maximum load or the lowest temperature or any combination of such criteria. In the case of ZENIT, this is not an acceptable solution on its own. Despite the reduction of the level of details induced, which could be somewhat overcome by tuning the heuristic chosen, the biggest reason that contraindicates its use is the Zero Emission balance constraint. Indeed, using this constraint requires to take into account every hour in the year, which is difficult with heuristics. On the contrary, clustering allows the use of the Zero Emission balance. In clustering, an algorithm is used to gather similar timesteps into clusters. Each original timestep is then represented by a cluster. We choose this approach over downsampling and heuristic in order to keep the original time granularity and the use of the emission constraint.

Several clustering algorithm exists and we limit this study to k-means and k-medoids clustering. In addition we consider the use of heuristics in combination to the clustering. This approach is recommended in this kind of optimization application because the clustering alone would likely 'dilute' the extreme events' timesteps, such as the hour with the maximum load, into a cluster represented by a lower value, which would lead to an under-dimensioned solution. A simple heuristic in addition to the clustering allows to correct this. In this paper, the heuristic chosen is the time (day or hour) with the highest total load, defined as the sum of the domestic hot water load (DHW), space heating (SH) and electric load, and the time with the lowest irradiance. In addition, normalizing the data before clustering can be beneficial [26]. Several ways to normalize the data before the clustering algorithm exists and we also consider two options: a normalization based on the range of each time series (1) and one based on the standard deviation (2).

$$X' = \frac{X - min(X)}{max(X) - min(X)} \tag{1}$$

$$X' = \frac{X}{std(X)} \tag{2}$$

Lastly, as mentioned in the literature review, mainly two approaches exist for clustering, one clusters directly the hours, the other focuses on design days. The design day approach uses clustering for selecting representative days in the year and then use the hourly values for each representative day. This approach is often favored when storages are modelled. Indeed, because the relation between timesteps inside a day are kept, it allows for daily operation of storages contrary to hours clustering.

The clustering is performed in Python using PyClustering [27] for the k-medoids algorithm and Scipy for the k-means [26,28]. The practical handling of the clustering is described in the flowchart in Fig. 1.

The data entering the clustering process consists of several hourly time series covering one year. The data is composed of the following time series: one domestic hot water load (DHW), one space heating load (SH) and one electric load for each building (or building type) in the neighborhood; outside air temperature, total irradiance and CO_2 factor of electricity.

4. Clustering results

The different clustering approaches presented in the previous section were performed for various number of clusters: for the clustering of design days, up to 100, and for the clustering of hours, up to 2400 (with 6 h steps). This allows to determine which number of clusters to use in the optimization model. The representatives of clusters and their sequence are combined to rebuild a complete year and then compared to the original data to compute errors. In this section, the errors are presented as Root Mean Square Deviations (RMSD) and as Normalized RMSD (NRMSD) when comparing the errors of different time series. All figures below share the same legend presented on Fig. 2.

Fig. 3 presents the NRMSD across all timeseries which can be interpreted as an indicator of the overall performance of the tested methods. It does not however give insight about the errors for individual timeseries.

Considering all figures in this section, it is clear that the k-means algorithm offers a better representation of the original timeseries than its k-medoids counterpart. Indeed in all of the following figures, the green line representing k-means reaches lower levels of RMSD and converges to it faster than the k-medoids ones. This indicates a closer match between the clusters obtained with k-means and the original timeseries than what is obtained with k-medoids for a given number of clusters. This is what we could expect. Indeed, the k-medoids uses vectors from the original datasets instead of creating centroids, which are better representatives. However, this ensures that the chosen representatives of clusters in the case of k-medoids are meaningful and realistic.

Another thing one could expect is that the performance of the clusterings monotonically improves. However, this is not the case of our results, especially in the case of design days. For the performance regarding individual time series, this could be explained because of a better performance of other time series for this particular number of clusters. However, this lack of monotony can still be found in the aggregated result of Fig. 3. One possible explanation for the lack of monotony could be that k-means and k-medoids algorithms do not always find the global optimums but can provide solutions that are only local optimums. Hierarchical clustering or running the clustering algorithms several time with different initial conditions could provide more consistent results.

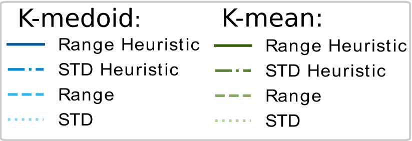
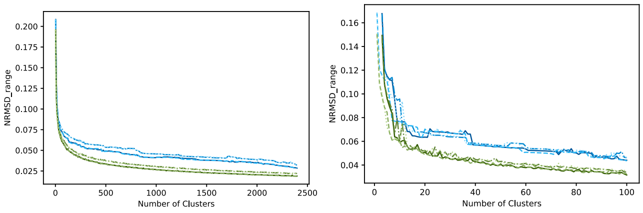


Fig. 2. Legend of the Results.



(a) Hours

(b) Days

Fig. 3. Average of the NRMSD of All Clustered Time Series, Normalized with Range.

Looking at Fig. 3, the use of heuristic results in a tiny advantage for the heuristic versions on the overall error of the clustering. This is especially true in the case of clustering on hours. For design days clustering, the difference between clustering with and without heuristic disappears after around 8 design days. The lower the amount of design days, the higher the impact of forcing two days to be extreme events is, while for hours, the forced hours are "diluted" faster.

From all figures, considering an equivalent resulting number of timesteps (translating to the complexity to solve the model) clustering on hours gives much better results than clustering on days. For the overall error, Fig. 3, the error for the hours clustering is about 50% lower than for the design days.

The performance for individual time series is discussed in the following.

For the CO_2 factor of electricity in Fig. 4, the convergence rate is much lower in the case of days than of hours. The decrease is almost linear, compared to exponential. In addition, there are high variations for days that are not present for hours. For 100 days, the RMSD is about 4.5 gCO_2/kWh against 2 for an equivalent number of hours.

In the case of spot price in Fig. 5, the difference between the

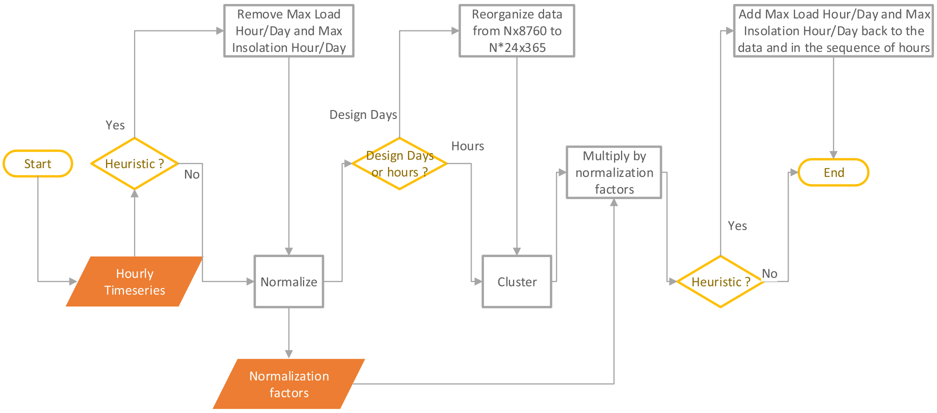
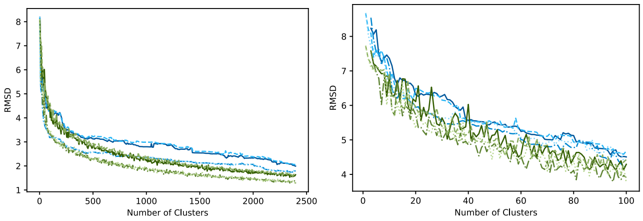
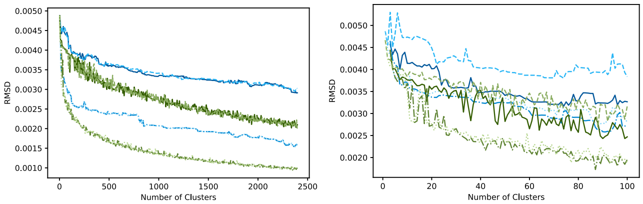

Start
Hourly Timeseries
Heuristic ?
Yes
No
Remove Max Load Hour/Day and Max Insolation Hour/Day
Normalize
Normalization factors
Design Days or hours ?
Design Days
Hours
Reorganize data from Nx8760 to N*24x365
Cluster
Multiply by normalization factors
Heuristic ?
Yes
No
Add Max Load Hour/Day and Max Insolation Hour/Day back to the data and in the sequence of hours
End

Fig. 1. Flowchart of the Clustering Process.



(a) Hours (b) Days

Fig. 4. RMSD of CO_2 Factor of Electricity.



(a) Hours (b) Days

Fig. 5. RMSD of Spot Price.

clustering performed using the standard deviation method and the range method seems very significant: for hours clustering between 0.001 and 0.0015 €/kWh or a factor of 2, the standard deviation is performing better. The difference between k-medoids and k-means is also considerably to the advantage of k-means: for hours clustering between 0.0005 and 0.001 €/kWh. For design days clustering the overall difference between methods is similar but there is more variability and some differences specific to this case. For instance, there are differences between the cases with and without heuristic, with the heuristic case performing better. Those differences are rather small for the standard deviation normalization and larger in the range case, especially in the k-medoids case.

The errors for the temperature time series are very similar to the overall ones commented before. The RMSD of temperature plateaus rather quickly to around 2 for the hours, and 2.8 for the days.

In the context of Zero Emission Neighborhoods, the irradiance has a very important role. Indeed, solar power is the main source of local (on the site) energy for neighborhoods. This means that solar irradiance and the production from the solar technologies will be crucial in compensating the emissions in the Zero Emission balance. Thus, in order to obtain designs that actually are Zero Emission, the precision of the clustering of the irradiance is essential. The behaviour for the hours clustering, Fig. 6, is similar to the overall behaviour. The RMSD for 100 clusters is around 35 W/m^2 for k-means and 55 W/m^2 for k-medoids. For days clustering, the convergence rate is slow and after 100 cluster, the RMSD is around 80 for k-means and 110 for k-medoids. The slow convergence rate means that for small numbers of clusters the difference between days and hours clustering is even worse. For 10 days, the RMSD is 140 for k-means and 170 for k-medoids. For 240 h, the RMSD is about 60 for k-means and 100 for k-medoids. Those values are high in comparison to the standard test condition (STC) of solar panels of 1000 W/m^2.

Only the performance for one of the three buildings is shown in this section. The other buildings can be found in the appendix.

For the electric load, Fig. 7, in the case of days, the convergence has a steep rate but it happens slightly later around 10 days. After the convergence, the difference between all methods is close to zero. For the clustering on hours, the convergence is fast. The main difference from the behavior in the mean RMSD is that the cases with k-medoids and standard deviation normalization have a higher RMSD. The plateau is around 0.0013 Wh m^{-2} h^{-1} versus 0.0005 Wh m^{-2} h^{-1} for k-means range and 0.0008 Wh m^{-2} h^{-1} for the others.

The RMSD for the SH and DHW time series behave as the mean of the RMSDs. The mean of the RMSD is influenced greatly by the loads because they behave similarly and because of the presence of 3 time series for each building.

Another metric of interest is the Yearly Average Error (YAE), this metric allows us to have information about the distribution of the error. With RMSD, there is no information on the sign of the errors. YAE allows to know if the errors are, on average, compensated or rather accumulate from timestep to timestep.

The results for k-means are not in Table 1 because the YAE stays at 0 for all number of clusters and all cases. Instead the values of the RMSD are presented in Table 2.

Comparing the RMSD and YAE from Table 1 and Table 2 gives us insights in how much the errors in irradiance cancel each other, at least in terms of annual values. In the case of irradiance, the negative signs first informs us that it is under-represented. The difference between the RMSD and the YAE values also suggest that the errors tends to be compensated by one another and they compensate completely in the k-means cases. In general, the hours clustering performs better than the daily one. k-means is better than k-medoids in terms of YAE for the same reasons that it is better for RMSD. The performance of STD or range on their own or in addition to heuristic is not consistent but the gains here are less big than between days and hours clustering.

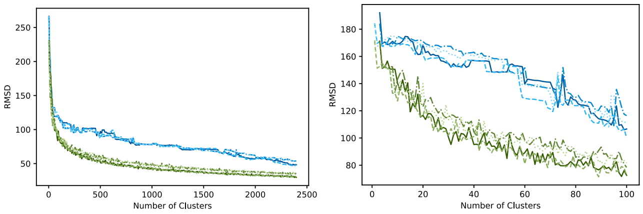
Two other metrics, the covered variance and the correlation error, can also be used in order to assess the clustering methods such as in [6]. They are defined in the same way as in [6]:

$$VC = \frac{var(X^*)}{var(X)} \tag{3}$$

$$CE = |corr(X_1^*, X_2^*) - corr(X_1, X_2)| \tag{4}$$

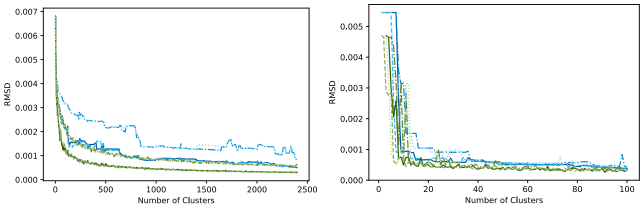
The covered variation (VC) of one timeseries is calculated as quotient of the variance of the timeseries reconstituted from the clusters (X^*) and the variance of the original timeseries (X). The correlation error (CE) between two timeseries is calculated as the absolute difference between the Pearson correlation coefficients calculated using the reconstructed timeseries and using the original timeseries.

These metrics are calculated for different numbers of clusters and the results are presented in Figs. 8 and 9. From both figures, k-medoids performs slightly better for really low number of clusters (less than 10 days/240 h) and the performances even out after that. The normalization based on standard deviation has a little edge over the range-based one but the difference is not large enough to be significant. It takes more day-clusters than hour-clusters to achieve similar performances. For instance, a covered variance of 0.9 is achieved with 250 h versus around 45 days. The combination of k-means clustering with a range normalization is significantly worse (about twice) at representing the correlation between the timeseries for a number of clusters between 20 days (240 h) and 60 days (1440 h). Overall, the results for those metrics are quite good for all methods from about 240 h or 20 days. If we look a bit more into the details, the variability covered is best for the loads and for the temperature. the covered variability of the irradiance is a bit worse and the variability covered for the spot prices and the CO_2 factors are the lowest. The spot price timeserie also has the highest



(a) Hours (b) Days

Fig. 6. RMSD of Irradiance on a Tilted Surface.



(a) Hours (b) Days

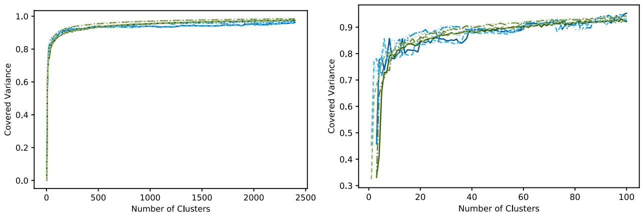
Fig. 7. RMSD of Electric Load in the Normal Offices.

Table 1
Yearly Average Error (YAE) and RMSD for 10 and 100 days and equivalent number of hours for the irradiance with k-medoids (STD.:Standard Devation, H: with heuristic, $\cancel{H}$: without heuristic).

	Days				Hours			
	STD.		Range		STD.		Range	
	H	$\cancel{H}$	H	$\cancel{H}$	H	$\cancel{H}$	H	$\cancel{H}$
YAE 10	−34	−24	−37	−38	−18	−18	−13	−15
YAE 100	−3.5	−0.67	−0.26	−2.3	−3.1	−2.8	−3.0	−2.1
RMSD 10	175	173	169	170	95.3	95.6	107	107
RMSD 100	116	112	107	100	53.9	53.8	48.3	47.8

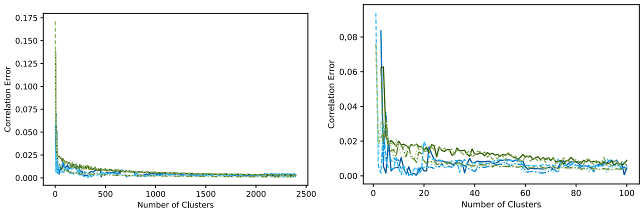
Table 2
RMSD for 10 and 100 days and equivalent number of hours for the irradiance with k-means (STD.:Standard Deviation, H: with heuristic, $\cancel{H}$: without heuristic).

	Days				Hours			
	STD.		Range		STD.		Range	
	H	$\cancel{H}$	H	$\cancel{H}$	H	$\cancel{H}$	H	$\cancel{H}$
RMSD 10	143	133	140	127	71.0	72.2	63.7	66.6
RMSD 100	78.4	81.7	72.6	72.5	35.5	35.1	29.7	30.5



(a) Hours (b) Days

Fig. 8. Mean Covered Variance for all Timeseries Depending on the Number of Clusters.



(a) Hours (b) Days

Fig. 9. Mean of all Correlation Error Between Each Timeseries Depending on the Number of Clusters.

correlation error to the other timeseries. The CO_2 factors, spot price and, to a smaller extent, the irradiance timeseries benefit the most from increases in the number of clusters.

From the results presented in this section, k-means and hours clustering are the best choices. For instance, with a focus on the irradiance, the choice would be range and heuristic. Overall the biggest impact can be made by choosing the correct clustering algorithm and the correct resolution. When it comes to the normalization method and the use of heuristic, the choice has less importance and varies depending on the goal. However there appears to be better results with the range normalization and without the heuristic. These results are however not enough. They only display some metrics for how close the clusters come to the original data. This does not guarantee that the one performing best in this section would also perform best in the optimization.

5. Models and implementation

In this section, the main equations of the ZENIT model are presented along with two variations for using either representative days or hours then the implementation and data used is briefly presented. The variations will be called M0 and M1 and are based on [23].

ZENIT aims is to design the energy system of a neighborhood so that it can be Zero Emission during its lifetime. Thus, it considers the investment as well as the operation of the neighborhood to find the cost optimal solution. The objective function is: *Minimize*:

$$b^{HG}\cdot C^{HG} + \sum_{b}\sum_{i}\left(\left(C_{i,b}^{var,disc} + \frac{C_{i,b}^{maint}}{\varepsilon_{r,D}^{tot}}\right)\cdot x_{i,b} + C_{i,b}^{fix,disc}\cdot b_{i,b}\right) + \sum_{t_\kappa}\frac{\sigma_\kappa}{\varepsilon_{r,D}^{tot}}\left(\sum_{b}\sum_{f} f_{f,t,b}\cdot P_f^{fuel} + \left(P_t^{spot} + P^{grid} + P^{ret}\right)\cdot\left(y_t^{imp} + \sum_{b}\sum_{est} y_{t,est,b}^{imp}\right) - P_t^{spot}\cdot y_t^{exp}\right) \tag{5}$$

It considers the investment cost in technologies ($C_{i,b}^{var,disc}$, $C_{i,b}^{fix,disc}$) and the heating grid (C^{HG}), as well as operation and maintenance related costs ($C_{i,b}^{maint}$). A binary variable controls the investment in the heating grid (b^{HG}). The subscript used in the equations are b for the buildings, i for the technologies, t for the timesteps, f for fuels and est for batteries. ε are the discount factors with interest rate r for the duration of the study D. $x_{i,b}$ is the capacity of the technologies and $b_{i,b}$ the binary related to whether it is invested in or not. σ_κ is the number of occurrences of cluster κ in the full year and t_κ is the timestep in the cluster. P are the prices of fuel, electricity on the spot market, grid tariff or retailer tariff. f is the consumption of fuel and y are the imports or exports of electricity.

In order to fulfill the Zero Emission requirement, the following constraint, called the Zero Emission Balance is necessary:

$$\phi_t^{CO_2,e}\sum_{t_\kappa}\sigma_\kappa\left(y_t^{imp} + \sum_{b}\sum_{est} y_{t,est,b}^{imp}\right) + \sum_{t_\kappa}\sigma_\kappa\sum_{b}\sum_{f}\phi^{CO_2,f}\cdot f_{f,t,b} \leqslant \phi_t^{CO_2,e}\cdot\sum_{t_\kappa}\sigma_\kappa\left(\sum_{b}\sum_{est}\eta_{est}\cdot y_{t,est,b}^{exp} + \sum_{b}\sum_{g} y_{t,g,b}^{exp}\right) \tag{6}$$

It forces the emissions of CO_2 to be at least equal to the compensations. The principle of the compensation is that the energy produced in the neighborhood, by renewable sources, that is exported to the national grid reduces the global production. The corresponding amount of saved CO_2 is counted as compensation for the neighborhood. The CO_2 factors are represented by $\phi_{e,t}^{CO_2}$ for electricity and $\phi_f^{CO_2}$ for other fuels. η_{est} is the charging efficiency of the battery.

Eqs. (7a), (7b) and (7c) represent respectively the equations for the electric load, the DHW load and the SH load. ∀ t:

$$y_t^{imp} + \sum_b \left(\sum_{est} y_{t,est,b}^{dch} \cdot \eta_{est} + \sum_g g_{g,t,b}^{selfc} \right) = \sum_b \left(\sum_e d_{e,t,b} + E_{b,t} \right) \tag{7a}$$

$\forall\ t,\ b$:

$$\sum_q q_{q,t,b}^{DHW} + \sum_{hst} \left(\eta_{hst} \cdot q_{t,hst,b}^{DHWdch} - q_{t,hst,b}^{DHWch} \right) + q_{t,b}^{HGusedDHW} = H_{b,t}^{DHW} + q_{t,b}^{dump} \tag{7b}$$

$$\sum_q q_{q,t,b}^{SH} + \sum_{hst} \left(\eta_{hst} \cdot q_{t,hst,b}^{SHdch} - q_{t,hst,b}^{SHch} \right) + q_{t,b}^{HGusedSH} = H_{b,t}^{SH} \tag{7c}$$

The electricity consumed in the neighborhood (the load and the use of some heating technologies) need to be balanced by the imports, discharges from the batteries or consumption of on-site production. The principles are the same for the heat but separately for each building. At the production plant, the heat produced is either stored, dumped or fed to the heating grid (Eq. (8a)). The heat flow through the pipes is limited (Eq. (8b)). We model the grid in a way that the buildings cannot feed heat into the heating grid (Eq. (8c)). In addition, the larger technologies of the central plant are only available if the optimization invests in the heating grid (Eq. (8f)).

$$\sum_q q_{q,t,'PP'} + \sum_{hst} \left(\eta_{hst} \cdot q_{t,hst,'PP'}^{dch} - q_{t,hst,'PP'}^{ch} \right) = \sum_{b \setminus 'PP'} q_{t,'PP',b}^{HGtrans} + q_{t,'PP'}^{dump} \tag{8a}$$

$\forall\ b,\ b',\ t$

$$q_{t,b',b}^{HGtrans} \leqslant \dot{Q}_{b',b}^{MaxPipe} \tag{8b}$$

$\forall\ b,\ t$

$$\sum_{b'} q_{t,b,b'}^{HGtrans} \leqslant \sum_{b''} (q_{t,b'',b}^{HGtrans} - Q_{b'',b}^{HGloss}) \tag{8c}$$

$$q_{t,b}^{HGused} = q_{t,b}^{HGusedSH} + q_{t,b}^{HGusedDHW} \tag{8d}$$

$$q_{t,b}^{HGused} = \sum_{b''} (q_{t,b'',b}^{HGtrans} - Q_{b'',b}^{HGloss}) - \sum_{b'} q_{t,b,b'}^{HGtrans} \tag{8e}$$

$\forall\ i$

$$x_{i,'ProductionPlant'} \leqslant X_i^{max} \cdot b^{HG} \tag{8f}$$

The import and export are limited by the size of the grid connection:

$$y_t^{imp} + \sum_b \sum_{est} y_{t,est,b}^{imp} + \sum_b \sum_g y_{t,g,b}^{exp} \leqslant GC \tag{9}$$

The fuel or electricity consumption depends on the heat produced and the efficiency (Eq. 10) and in the case of CHPs, the Heat to power ratio regulates how much electricity is produced as a by product (10b). In the implementation, α_{CHP} has a fixed value.

$\forall\ \gamma \in \mathcal{F} \cap Q,\ t,\ b$:

$$f_{\gamma,t,b} = \frac{q_{\gamma,t,b}}{\eta_\gamma} \tag{10a}$$

$\forall\ \gamma \in \mathcal{E} \cap Q,\ t,\ b$:

$$d_{\gamma,t,b} = \frac{q_{\gamma,t,b}}{\eta_\gamma} \tag{10b}$$

$\forall\ t',\ CHP',\ b$:

$$g_{CHP,t,b} = \frac{q_{CHP,t,b}}{\alpha_{CHP}} \tag{12}$$

Some heating technologies can only supply the SH. Eq. (14) controls which technology can produce DHW. $\forall\ q,\ t,\ b$:

$$q_{q,t,b} = q_{q,t,b}^{DHW} + q_{q,t,b}^{SH} \tag{13}$$

$$q_{q,t,b}^{DHW} < = M \cdot B_q^{DHW} \tag{14}$$

The solar technologies output depends on the solar irradiance and the module efficiency. In the case of PV, the efficiency is defined as in [29].

$$g_{PV,t} + g_t^{curt} = \eta_{PV,t} \cdot x_{PV} \cdot IRR_t^{tilt} \tag{14a}$$

$$q_{ST,t} = \eta_{ST} \cdot x_{ST} \cdot IRR_t^{tilt} \tag{14b}$$

$$\eta_{PV,t} = \frac{\eta^{inv}}{G^{stc}} \cdot (1 - T^{coef} \cdot (T^c - T^{stc})) \tag{14c}$$

$$T^c = T_t + \left(T^{noct} - 20 \right) \cdot \frac{IRR_t^{tilt}}{800} \tag{14d}$$

Eqs. (15a) and (15b) link the heat produced to the COP and the electrical consumption. The COPs are different for SH and DHW due to different temperature set points. They also depend on the outside temperature and are calculated before the optimization. Eq. (15c) regulates how the heat pump can be used for both SH and DHW and enforces that the capacity invested is not exceeded. $P^{input,max}$ represents the maximum power input to the heat pump at the timestep based on the temperature set point for a 1 kW unit. $d_{hp,b,t}^{SH}$ and $d_{hp,b,t}^{SH}$ represent the electric consumption of the heat pump for SH and DHW while $q_{hp,b,t}^{DHW}$ and $q_{hp,b,t}^{DHW}$ are the heat production.

$$d_{hp,b,t}^{SH} = \frac{q_{hp,b,t}^{SH}}{COP_{hp,b,t}^{SH}} \tag{15a}$$

$$d_{hp,b,t}^{DHW} = \frac{q_{hp,b,t}^{DHW}}{COP_{hp,b,t}^{DHW}} \tag{15b}$$

$$\frac{d_{hp,b,t}^{DHW}}{P_{hp,b,t}^{input,max,DHW}} + \frac{d_{hp,b,t}^{SH}}{P_{hp,b,t}^{input,max,SH}} \leqslant x_{hp,b} \tag{15c}$$

Some technologies have part-load limitations, they cannot be operated from 0 to 100%. This leads to a large number of binary variables in the model:

$$\overline{x_{i,b,t}} \leqslant X_{i,b}^{max} \cdot o_{i,t,b} \tag{16a}$$

$$\overline{x_{i,b,t}} \leqslant x_{i,b} \tag{16b}$$

$$\overline{x_{i,b,t}} \geqslant x_{i,b} - X_{i,b}^{max} \cdot (1 - o_{i,t,b}) \tag{16c}$$

$$q_{i,b,t} \leqslant \overline{x_{i,b,t}} \tag{16d}$$

$$q_{i,b,t} \geqslant \alpha_{i,b} \cdot \overline{x_{i,b,t}} \tag{16e}$$

Some technologies have a minimum investment capacity and are modelled as semi-continuous variables:

$$x_{i,b} \leqslant X_{i,b}^{max} \cdot b_{i,b} \tag{17a}$$

$$x_{i,b} \geqslant X_{i,b}^{min} \cdot b_{i,b} \tag{17b}$$

The electricity production from on-site technologies can be exported, consumed directly, stored or dumped:

$$g_{g,t,b} = y_{t,g,b}^{exp} + g_{g,t,b}^{selfc} + g_{t,g,b}^{ch} + g_{t,g,b}^{dump} \tag{18}$$

To distribute the production to the batteries, we have $\forall\ t,\ b$:

$$\sum_g g_{t,g,b}^{ch} = \sum_{est} y_{t,est,b}^{ch} \tag{19}$$

The handling of the storages is what differentiates model M0 and M1. Model M0 is not able to handle seasonal storage while model M1 can be used for that. In ZENIT, each battery is modelled as 2 separate virtual batteries, with one connecting the neighborhood to the grid: allowing import and export between the grid and the battery, and import from the battery to the neighborhood's loads, and the other connecting the technologies producing electricity in the neighborhood and the neighborhood's loads: allowing exports to the grid and to the neighborhood. This distinction provides traceability of the electricity in the batteries. The origin of the electricity is important because of the different CO_2 factors.

Both models are presented in the following subsections.

5.1. Model M0

Model M0 uses a classical formulation for storages in models using clustering and that do not need seasonal storage. The equation for electric and heat storages are similar, so only a generic equation is presented, fitting both cases.

$\forall\ \kappa,\ t \in [1,\ T^{clu}],\ st,\ b$

$$v^{stor}_{\kappa,t,st} = v^{stor}_{t-1,st} + \eta^{stor}_{st}\cdot q^{ch}_{t,st} - q^{dch}_{t,st} \tag{20}$$

$\forall\ t \in [0,\ T^{clu}],\ st,\ b$

$$v^{stor}_{\kappa,t,st,b} \leqslant x_{st,b} \tag{21}$$

$$q^{ch}_{\kappa,t,st,b} \leqslant \dot{Q}^{max}_{st} \tag{22}$$

$$q^{dch}_{\kappa,t,st,b} \leqslant \dot{Q}^{max}_{st} \tag{23}$$

$\forall\ p,\ st,\ b,\ \kappa$

$$v^{stor}_{\kappa,0,st,b} = v^{stor}_{\kappa,T^{clu},st,b} \tag{24}$$

The state of charge of the storage *st* (either heat or electric storage) is represented by v^{stor} while *qch* and *qdch* are the energy charged and discharged. The maximum charge and discharge rate is Q^{max}_{st}. The differences between this model and a model with full year data is that the starting value of the storage is "free" at the beginning of each cluster instead of only at the beginning of the year. This model is valid for different T^{clu} even though in our case it is 24 for days clustering. It allows for a daily operation of the storage. Different values of T^{clu} could be used for allowing different ranges of operation of the storage. A bigger value allows longer operation but probably increases the number of timesteps to get the same clustering precision and reducing it reduces the possible range. The daily range makes sense because of the daily cycle of the loads, that allows us to make the assumption of Eq. (24). This model does not make sense with $T^{clu} = 1$, i.e. the hours clustering, because the resolution of the data used is also one hour; hourly storage operation does not make sense.

5.2. Model M1

In model M1, the main difference with model M0 is that the storage level equation becomes: $\forall\ \kappa,\ t \in [1,\ 8760],\ st,\ b$

$$v^{stor}_{\kappa,t,st} = v^{stor}_{t-1,st} + \eta^{stor}_{st}\cdot q^{ch}_{t_\kappa,st} - q^{dch}_{t_\kappa,st} \tag{25}$$

The end value of the storage constraint is also replaced by: $\forall\ st,\ b$

$$v^{stor}_{0,st,b} = v^{stor}_{8760,st,b} \tag{26}$$

Where t_κ is the time corresponding to t in the cluster. It is found by using the sequence of cluster's representatives (ξ) either directly for the hourly case or with the day number corresponding to t and the hour in the day:

Hours:

$$t_\kappa = \xi(t) \tag{27}$$

$$t_\kappa = \xi\left(\left\lfloor \frac{t}{24} \right\rfloor\right) + t - \left\lfloor \frac{t}{24} \right\rfloor\cdot 24 \tag{28}$$

This means that the storage level is not decoupled between the different clusters. The charging and discharging is defined for each timestep in each cluster but the storage level is defined for every hour in the year. This model comes from the assumption that days or hours with similar conditions in terms of the time series (loads, spot price, temperature, …), i.e. belonging to one cluster will behave in the same way in terms of charging and discharging of the storage. This formulation however comes at the expense of longer computation time. Both hourly and days clustering can be used with M1.

In [23], another variation is presented to improve further model M1 by defining only the variables related to operations binary variables (on/off status) for each cluster while other variables are defined for each hour of the year. That model has not been implemented because it increases the computation time even more.

5.3. Implementation

The model is implemented on a test case based on a small neighborhood, a campus at Evenstad in Norway. The buildings are gathered in three categories to only have three buildings in the optimization. We assume every building has a hydronic system.

The economical and technical data of the technologies are taken from the Danish Energy Agency.[1] In total, 22 technologies are implemented with, at the building level: solar panel, solar thermal, air-air heat pump, air–water heat-pump, ground source heat pump, bio boiler with wood logs or pellets, electric heater and electric boiler, biomethane boiler, biogas and biomethane CHP; and at the neighborhood level: biogas boiler, wood chips and pellets boiler and CHPs, ground source heat pump and electric boiler. When it comes to the storage technologies, lithium-ion is used for electrical storage and hot water tanks for the heat storage. The storage technologies at the neighborhood level are not included to make it easier to compare the objective values obtained from the runs with M0 and M1.

The spot price of electricity is obtained from Nordpool's website.[2] The temperature data comes from Agrometeorology Norway,[3]. The solar irradiance (diffuse horizontal (DHI) and direct normal(DNI)) are obtained from Solcast.[4] The irradiance on a tilted surface IRR^{Tilt} which is an input of the clustering is derived from the DHI and DNI with:

$$IRR^{Tilt}_t = DHI_t\frac{1+cos(\phi_1)}{2} + \alpha\cdot(DNI_t + DHI_t)\frac{1-cos(\phi_1)}{2} + DNI_t\left(\frac{cos(\varphi_t)\cdot sin(\phi_1)\cdot cos(\phi_2-\psi_t)}{sin(\varphi_t)} + \frac{sin(\varphi_t)\cdot cos(\phi_1)}{sin(\varphi_t)}\right) \tag{29}$$

We assume that for some sun positions (sun elevations (φ) below 1 degree and sun azimuths (ψ) between -90 and 90 degrees), no direct beam reaches the panels. This means that the last term of Eq. (29) is removed at such times. We use a constant albedo factor (α) of 0.3 for the whole year. Hourly albedo values could also be used to reflect the impact of snow in the winter better. The tilt angle of the solar panel is ϕ_1; the orientation of the solar panel regarding the azimuth is ϕ_2.

The hourly CO_2 factors of electricity are obtained with the methodology presented in [30] while the other CO_2 factors come from [31].

The prices of wood pellets comes from [32], the price of wood logs from [33], the price of wood chips from [34] and the price of biogas from [35].

The electric and heat load profiles for the campus are derived from [36]. The domestic hot water (DHW) and Space Heating (SH) are then based on the time series from a passive building in Finland [37].

The model is implemented in Python and is solved using Gurobi. It is run on a laptop with an Intel Core i7-7600U dual core processor at 2.8Ghz and 16 GB of RAM.

[1] https://ens.dk/en/our-services/projections-and-models/technology-data.

[2] https://www.nordpoolgroup.com/Market-data1/#/nordic/table.

[3] https://lmt.nibio.no. Fåvang station

[4] https://solcast.com.au.

6. Model results and discussion

In this section we present the results obtained with the different clustering methods and variations from the earlier sections. We always use the heuristic in order to guarantee that the peak load is covered.

6.1. Simplified model

In order to get a reference objective value to base our analysis on, a simplified version of the model is run. This simplified model leaves out several of the constraints using binaries, namely the part load constraints, the minimum investment capacity (turning the semi-continuous variables into continuous variables) and changing the cost function from $a{\cdot}x + b$ to $c{\cdot}x$. Without simplifying the model, solving the model with 365 days or 8760 h would take too long. It is important to note that this simplified model is not directly obtained by removing constraints but by setting the input associated to the binary to zero. For example, the fixed investment costs and the minimum capacity are set to 0 but the constraints are still there. In the case of the minimum load during operation, the minimum loads are set to 0 but the related constraints are not written when the model is generated in Gurobi. The results for the non-simplified model are presented after without a reference value.

Because M1 allows for seasonal storage modelling while M0 does not and in order to obtain results that can be compared more easily between M0 and M1, the storages at the neighborhood level were not included in the technological option input in this study.

We chose the number of days and hours in this section graphically at the elbow of the curves in Fig. 3. The number of clusters is chosen so that adding clusters does not bring considerable improvements. For the case of hours, this corresponds to around 120 h; 96 and 144 h are also studied as a 20% variation. We also consider the corresponding number of days, i.e. 4,5,6. Indeed this gives an equal number of timesteps in the optimization but the performance of the clustering on days for such low numbers of days should give poor result considering Fig. 3. In addition we choose a number of design days with similar graphical elbow considerations. However, we consider Fig. 6 instead of the NRMSD figure because in the case of clustering days the performances for the irradiance were converging slower. This leads us to choose 30 days. We also take the 20% variations, which corresponds to 24 and 36 days.

From Table 3, k-means range seems to be the overall best choice, but it underestimates the objective value. k-medoids constantly overestimates the objective value, with significant errors for low numbers of days. On the other hand, k-means gives good results even for a low number of days.

From Tables 4 and 5 it appears that the hours clustering performs the best on problem M1, especially with the range normalization and k-medoids. For approaching the reference value from below, the best approach is k-means with hours clustering. Here the range method seems slightly better than STD. k-medoids constantly overestimates the objective value while k-means constantly underestimates it. However, in general and for around 30 days and 120 h, the k-means seems to be the appropriate choice. Indeed, even though k-medoids with STD also has good results, it appears less consistent. With this algorithm, the performance does not always improve with an increasing number of clusters; choosing the correct amount of clusters would become harder. k-means, while not completely exempt from this flaw, appears more robust in this regard.

For M1, the average of the run time for days clustering for 24, 30 and 36 days is 3500 s with extreme values of 2 289 and 5628 s. For the hours clustering, the average runtime is 5 973 s with extremes of 2 421 and 12 500 s. Days clustering is on average almost twice as fast as hours clustering on this simplified model despite having more timesteps overall. As a reference, to solve this simplified problem without any clustering (using a complete year) takes around 30 000 s.

For M0 the runtimes are low with all values below 360 s.

It is also interesting to look at the actual systems resulting of each investment run. Fig. 10 shows these investments for the runs with the simplified model.

From Fig. 10, it is noteworthy that there tends to be an investment in the heating grid when using k-medoids while it is not often the case with k-means. In general, the element with the biggest impact on the investment appears to be the clustering algorithm chosen. Indeed, there is are quite distinct groups of investments with k-medoids on the one side and k-means on the other emerging from the figure. Both reference runs have a very similar system, with the exception of the amount of space heating storage. This suggests that only the amount of storage invested will be affected by the choice of M0 or M1, leading to possible over-investment in storage if using M0. The investments resulting from runs with k-medoids seem to be closer to the references in general than the runs with k-means. One important exception is regarding the amount of PV invested where it tends to over-invest more than k-means (which is also over-investing). This over-investment stems from the representation of the solar irradiance in the clustered data; k-means offering a better representation as seen in Fig. 6.

6.2. Complete model

For the complete model, no reference value is presented because running the models with a complete year of data takes too long and it is the reason clustering is explored in the first place.

Fig. 11 presents the objective values resulting from the optimization in the case of M0. Without a reference value, it is impossible to reach a conclusion regarding the performances of each approach. However we can make some remarks. The objective values follow the same patterns as in the case of the simplified model and from the results we can expect that in this case as well k-means underestimates and k-medoids overestimates the objective value. It also appears that even a few days are enough to get satisfying results when using k-means.

Regarding runtime for M0 (Table 6), k-means with STD is clearly the fastest while k-medoids with STD is the slowest being about half as fast. k-medoids range and k-means range have comparable runtimes except for the case of 36 days where the k-medoids version is about 20% slower. k-means range is itself 25% slower than k-means STD.

For M1, the same remarks hold. K-medoids and k-means seem to respectively over- and underestimate the objective value. Fig. 12 confirms that for k-medoids, the range method performs better than STD as in Table 4 and 5. For k-means we also find that the results are similar.

M1 is between 15 and 40 times longer to solve than M0 for the days (Tables 6 and 7). When it comes to the difference between the days and the hours, even though the number of timesteps are the same, the hourly model takes at least 10 times longer to solve than the daily model. This difference is hard to explain. Indeed both models get the same number of timesteps and are identical with the exception of what is presented in Eqs. (27) and (28).

Fig. 13 shows the investment resulting from the runs using the full model. The systems obtained are similar to the ones visible in Fig. 10 but there is a lower diversity of technologies. The systems are comprised of a different amount of biomethane boilers, air–water heat

Table 3
Variations in objective value from the reference for different numbers of representative days for M0 with simplified model (**STD**: Standard Devation, **R**:Range), Reference Value for 365 days: **2,056,849** €.

		Days					
		4	**5**	**6**	**24**	**30**	**36**
k-means	STD	−10.29	−9.50	−9.42	−6.14	−5.21	−4.82
	R	−10.29	−10.64	−9.68	−4.80	−4.74	−3.82
k-medoids	STD	28.27	22.71	33.53	9.61	10.04	7.49
	R	11.57	8.78	23.16	5.36	4.84	8.27

Table 4
Variations in objective value from the reference for different number of representative days for M1 with simplified model (**STD**: Standard Devation, **R**:Range), Reference Value for a complete year: **2,060,612** €.

		4	5	6	24	30	36
k-means	STD	−10.16	−9.18	−8.78	−6.07	−5.38	−6.14
	R	−10.16	−10.38	−9.09	−5.03	−5.38	−4.44
k-medoids	STD	28.42	22.80	33.55	9.64	10.08	7.54
	R	11.78	8.91	23.19	5.40	4.90	8.31

Table 5
Variations in objective value from the reference for different number of representative hours for M1 with simplified model (**STD**: Standard Devation, **R**:Range), Reference Value for a complete year: **2,060,612** €.

		96	120	144
k-means	STD	−5.58	−5.54	−4.95
	R	−4.66	−5.51	−4.60
k-medoids	STD	8.45	11.06	9.73
	R	3.07	4.59	3.62

pumps, PV and heat storages. The heating grid is never chosen. A different system is appearing only in one of the cases of M0 with k-medoids and a low number of days, where solar thermal replaces partly the air–water heat pump. There is still a distinction between k-means and k-medoids as in Fig. 10 but it is less clear, especially in the case of the storage. Furthermore, the investments with model M1 with hours seems to be less sensitive to the number of clusters used, especially when it comes to the storage.

If the use of k-medoids is required for any reason, then using the hourly method can bring significant improvements to the precision over the daily method. These improvements needs to be considered in regard to the increased solving time to choose the method to use. Otherwise, k-means should be preferred. In that case, the improvements of the precision is insufficient to justify using the hourly method. One such possible reason is to have a good representation of the solar irradiance which is the case for ZENIT. By using the day method with low numbers of days, even though the solving time and objective values are good, the representation of the solar irradiance is problematic as seen in Fig. 6. In our case and to get a good solar irradiance representation, the use of k-means and hours clustering in M1 is preferable.

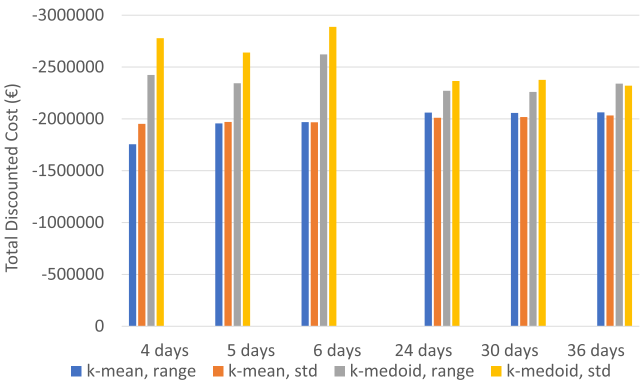


Fig. 11. Objective Values for M0 with Design Days with Complete Model.

Table 6
Runtime for M0 in seconds with days (**STD**: Standard Devation, **R**:Range).

		4	5	6	24	30	36
k-means	STD	70.98	96.47	108.8	1708	2846	3342
	R	116.3	115.2	155.4	1924	3504	4320
k-medoids	STD	63.42	62.98	288.4	3544	4157	6163
	R	57.52	115.0	127.5	2088	3442	5288

Overall, with regards to Zero Emission Neighborhood Energy System, the k-means performs better than the k-medoids algorithm. This is the opposite of what has been found in several studies in other energy system applications, such as in [8] or [6]. However, this is an illustration of the findings of [5] that the best clustering technique is dependant on the data to process and the application. In our particular case, the reason that k-means performs better than k-medoids could be that an averaging of all points inside each clusters leads to a better

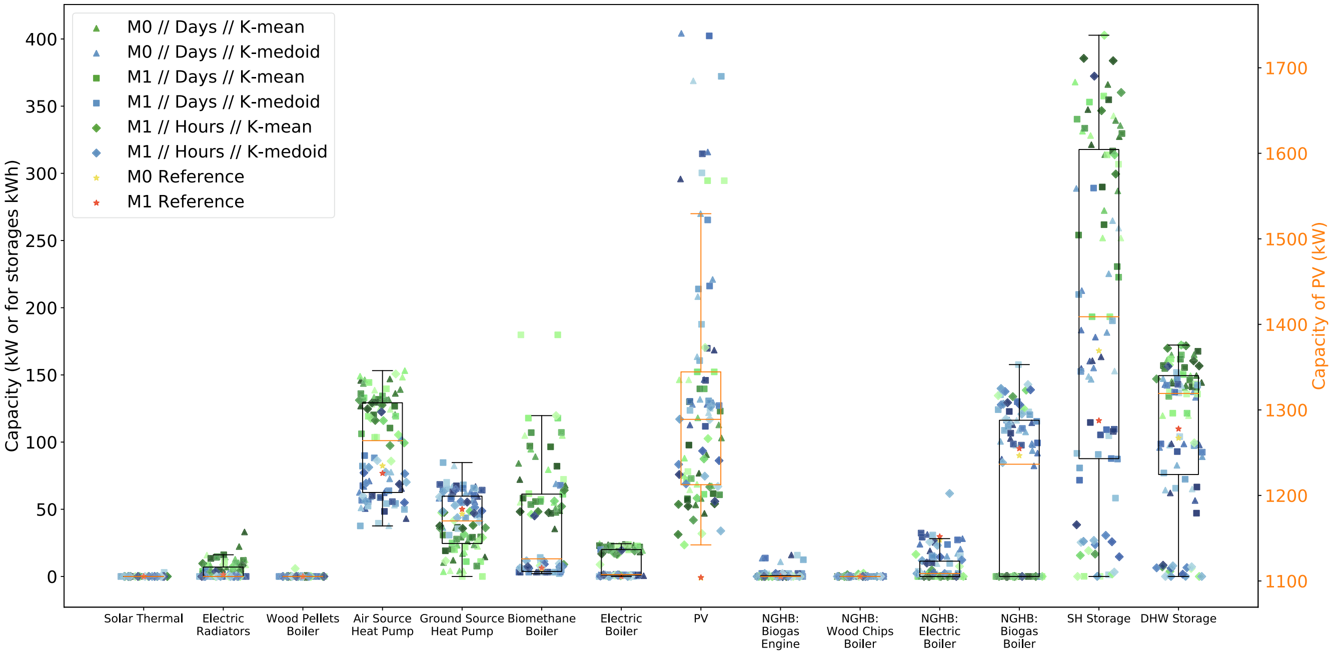


Fig. 10. Investments Resulting from the Runs with the Simplified Model. The color gradient represents the number of clusters, the clearer the least clusters and the darker the more clusters. “NGHB:” Before the technology name means that it is a technology at the neighborhood scale and also implies the presence of the heating grid. The technologies at the building level are aggregated for all the buildings.

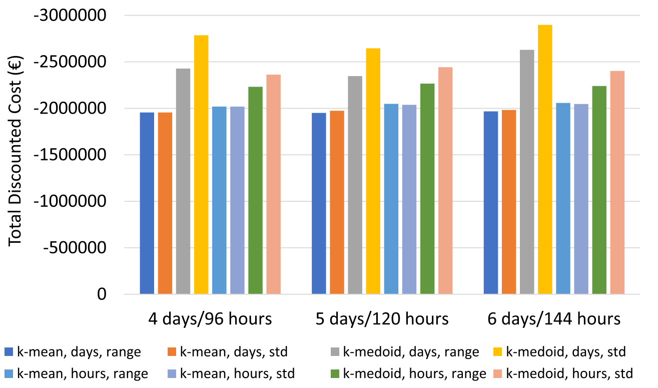


Fig. 12. Objective Values for M1 with Days and Hours with Complete Model.

Table 7
Runtime for M1 in seconds (**STD**: Standard Devation, **R**:Range).

	Days	**4**	**5**	**6**
k-means	STD	1386	2250	4340
	R	2059	3393	3519
k-medoids	STD	1723	1717	5838
	R	1139	2319	2626
	Hours	**96**	**120**	**144**
k-means	STD	14789	29159	62239
	R	13342	45048	60165
k-medoids	STD	20288	55509	105672
	R	18632	19860	59055

representation of the solar irradiance (as can be seen in Tables 2 and 1) while the points closest to the mean of the clusters may be pushed towards a better representation of the loads due to the number of load timeseries and their correlation. The better representation of solar then allows to reduce the investment in PV and to reach an energy system closer to the references cases.

7. Limitations

There are different limitations that should be mentioned regarding this paper. Regarding the studied methods, the fact that only clustering algorithms are studied have been explained; however other clustering algorithms could offer advantages. Many heuristics, either new or variations around the one used, could also be studied and finding the overall best heuristic presents a challenge. The clustering has been used on a specific case and we cannot guarantee that the same result holds true for larger cases or in other countries where the correlation between the different inputs are different. Unfortunately no reference value is shown for the complete model and a simplified model had to be used in order to compare the precision.

8. Conclusion

After introducing the use of reduction techniques and clustering in energy systems and in particular in the design of the energy system of neighborhoods, this paper discussed why clustering is chosen over other solutions such as downsampling. Different clustering methods have then been evaluated, first directly on their ability to come close to the original dataset and then on the results they give when used in ZENIT. K-means and k-medoids have been compared and the study allowed to highlight that counter to what is found for many other energy system applications, k-means performs better than k-medoids. The study also highlights the role of the normalization method on the performances by comparing a method using the standard deviation and one using the range of values. We find occurrences of models using clustered days (or design days) and of instances using clustered hours in the literature but the reason for the choice are not always clear. In this study, both approaches are implemented and the relation between the performance, the solving time and the possible uses of each are reviewed. The impact of the use of a simple heuristic is also studied. Two versions of the optimization models were used with different capabilities when it comes to storage: M0 for daily storage operation and M1 for storage without time limitation. While the use of M0 or M1 should be considered on the basis of the necessity to include seasonal storage, the

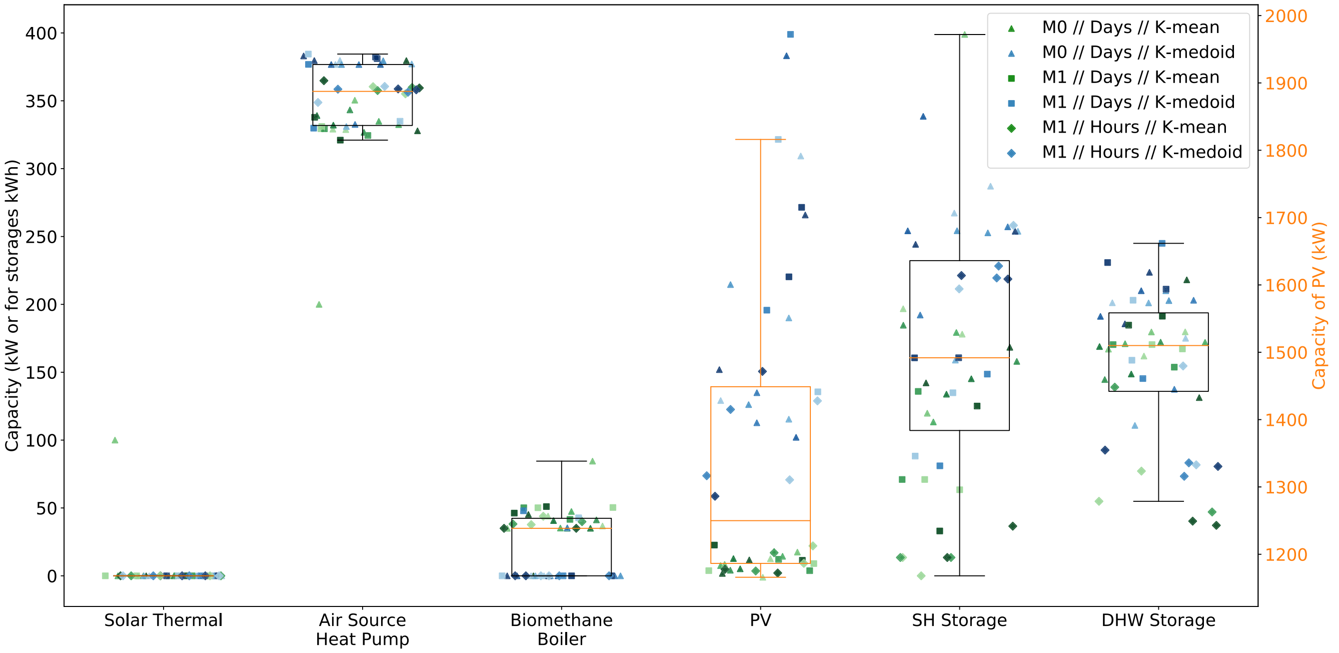
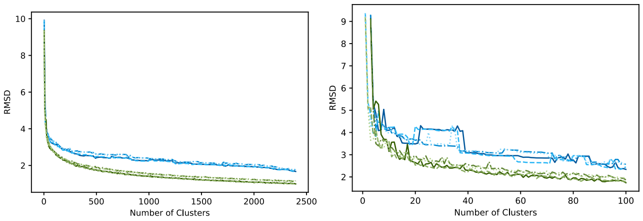
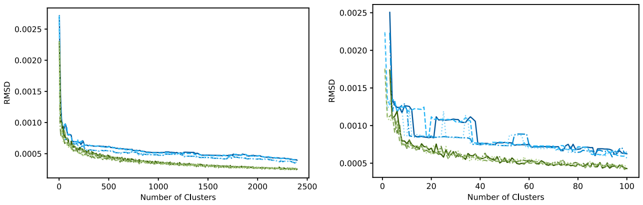


Fig. 13. Investments Resulting from the Runs with the Complete Model. The color gradient represents the number of clusters, the clearer, the least clusters and the darker, the more clusters. "NGHB:" Before the technology name means that it is a technology at the neighborhood scale and also implies the presence of the heating grid. The technologies at the building level are aggregated for all the buildings. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



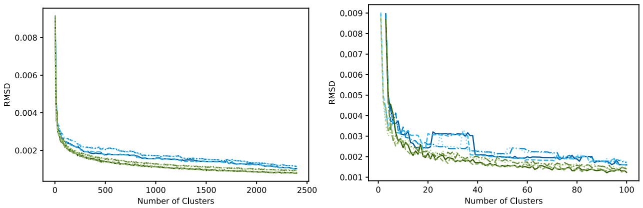
(a) Hours (b) Days

Fig. 14. RMSD of Temperature.



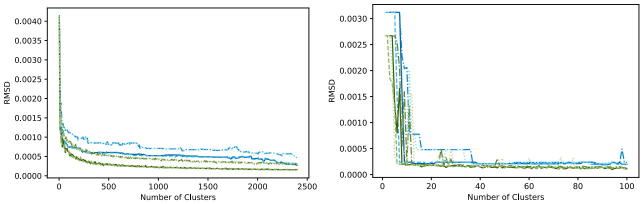
(a) Hours (b) Days

Fig. 15. RMSD of DHW Load in the Normal Offices.



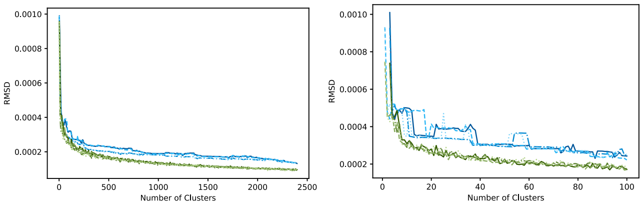
(a) Hours (b) Days

Fig. 16. RMSD of SH Load in the Normal Offices.



(a) Hours (b) Days

Fig. 17. RMSD of Electric Load in the Passive Offices.



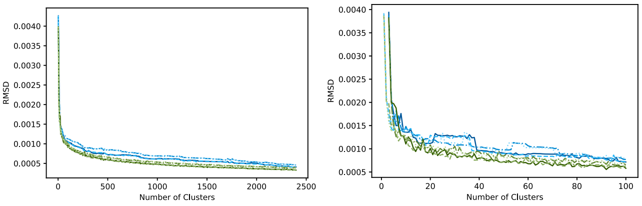
(a) Hours (b) Days

Fig. 18. RMSD of DHW Load in the Passive Offices.

choice of the clustering method (algorithm, cluster type and normalization method) can be made based on the results presented in this paper. For the particular application of designing the energy system of neighborhoods with an objective of zero emissions, the best method appears to be to use the k-means algorithm with the range normalization and days as cluster type. A low number of days is fine but it can be interesting to increase it to improve the representation of the solar irradiance for example. The trade-off between time and precision should then be considered. Further work could extend the result to other cases and study if the results presented in this paper scale to bigger neighborhoods. Other clustering algorithms or heuristics could also be investigated.

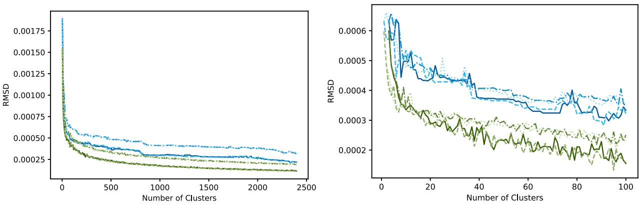
Declaration of Competing Interest

None.



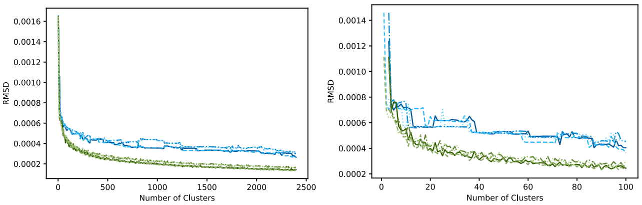
(a) Hours (b) Days

Fig. 19. RMSD of SH Load in the Passive Offices.



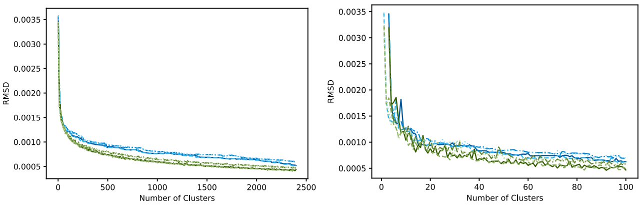
(a) Hours (b) Days

Fig. 20. RMSD of Electric Load in the Student Housing.



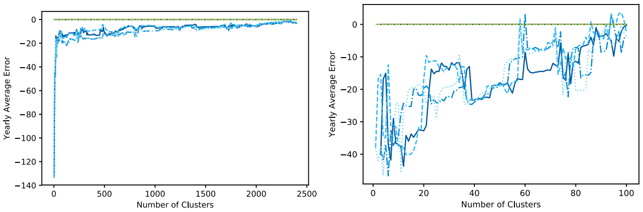
(a) Hours (b) Days

Fig. 21. RMSD of DHW Load in the Student Housing.



(a) Hours (b) Days

Fig. 22. RMSD of SH Load in the Student Housing.



(a) Hours (b) Days

Fig. 23. YAE of Irradiance.

Acknowledgment

This article has been written within the Research Center on Zero Emission Neighborhoods in Smart Cities (FME ZEN). The author gratefully acknowledges the support from the ZEN partners and the Research Council of Norway.

The author would also like to thank John Clauß for providing the hourly CO_2 factor for electricity data. .

Appendix A. Additional results of the clustering

Additional results are presented in this appendix. In particular, the RMSD for the time series that were not included in Section 4 are shown in this section.

Table 8
Data of technologies producing heat and/or electricity in the complete model.

Tech.	η_{th} (%)	Fix. Inv. Cost (€)	Var. Inv. Cost (€/kW)	α_i (% Inst. Cap.)	Min. Cap. (kW)	Annual O&M Costs (% of Var Inv. Cost)	Lifetime (year)	Fuel	α_{CHP}	El.	Heat
At building level											
PV[1]		0	730	0	50	1.42	35			1	0
ST[2]	70	28350	376	0	100	0.74	25			0	1
ASHP[3]	f(T_t)	42300	247	0	100	0.95	20	Elec.		0	1
GSHP[4]	f(T_t)	99600	373	0	100	0.63	20	Elec.		0	1
Boiler[5]	85	32200	176	30	100	2.22	20	Wood Pellets		0	1
Heater	100	15450	451	0	100	1.18	30	Elec.		0	1
Boiler	100	3936	52	20	35	2.99	25	Biomethane		0	1
At neighborhood level											
CHP[6]	47	0	1035	50	200	1.03	25	Biogas	1.09	1	1
CHP	98	0	894	20	1000	4.4	25	Wood Chips	7.27	1	1
CHP	83	0	1076	20	1000	4.45	25	Wood Pellets	5.76	1	1
Boiler[7]	115	0	680	20	1000	4.74	25	Wood Chips		0	1
Boiler[7]	100	0	720	40	1000	4.58	25	Wood Pellets		0	1
CHP[8]	66	0	1267	10	10	0.84	15	Wood Chips	3	1	1
Boiler[9]	58	0	3300	70	50	5	20	Biogas		0	1
GSHP[4]	f(T_t)	0	660	010	1000	0.3	25	Elec.		0	1
Boiler	99	0	150	5	60	0.71	20	Elec.		0	1
Boiler	100	0	60	15	500	3.25	25	Biogas		0	1

[1] Area Coefficient: 5.3 m^2/kW.
[2] Area Coefficient: 1.43 m^2/kW.
[3] Air Source Heat Pump.
[4] Ground Source Heat Pump.
[5] Automatic stoking of pellets.
[6] Gas Engine.
[7] HOP.
[8] Gasified Biomass Stirling Engine Plant.
[9] Solid Oxyde Fuel Cell (SOFC).

Table 9
Data of technologies producing heat and/or electricity in the simplified model. There is no fixed investment cost, no minimum size and no part load limitation. The other parameters are the same as in Table 8.

Technology	Var. Inv. Cost (€/kW)	Technology	Var. Inv. Cost
At building level		**At neighborhood level**	
PV	730	Biogas CHP	1035
ST	376	Wood Chips CHP	894
ASHP	670	Wood Pellets CHP	1076
GSHP	1369	Wood Chips Boiler	680
Wood Pellet Boiler	498	Wood Pellets Boiler	720
Elec. Heater	605	Wood Chips CHP	1267
Biomethane Boiler	91	Biogas Boiler	3300
		GSHP	660
		Elec. Boiler	150
		Biogas Boiler	60

Table 10
Data of Fuels.

Fuel	Fuel Cost (€/kWh)	CO_2 factor (gCO_2/kWh)
Electricity	$f(t)$	$f(t)$
Wood Pellets	0.03664	40
Wood Chips	0.02592	20
Biogas	0.07	0
Biomethane	0.07	100

The errors for the temperature time series, Fig. 14, are very similar to the overall ones. The RMSD of temperature plateaus rather quickly to around 2 for the hours, and 2.8 for the days.

The RMSD of the loads of the normal offices are presented in Figs. 15–17. For the offices already at the passivhus standard, the results are presented in Figs. 17–19.

For the student housings, the results are presented in Figs. 20–22.

The figures for the yearly average errors presented in Table 1 are presented in Fig. 23.

Appendix B. Technology Data

The data for technologies in Tables 8 and 9 come mainly from the Danish Energy Agency and Energinet.[5] The data for storages is presented in Table 11.

Table 11
Data of storage.

Index	One	Inv. Cost	O&M	Lifetime	Min. Cap.	Charge/
	way eff.		Cost			Discharge rate
	(%)	(€/kWh)	(% of Inv.	(year)	(kWh)	(% of Cap)
			Cost)			
Battery						
1[1]	95	577	0	10	13.5	37
2[2]	938	500	0	15	210	23
3[3]	95	432	0	20	1000	50
Heat Storage						
1[4]	95	75	0	20	0	20
2[3]	98	3	0.29	40	45 000	1.7

[1] Based on Tesla Powerwall.
[2] Based on Tesla Powerpack.
[3] Based on Danish energy agency data.
[4] Same data are used for the heat storage at the building or neighborhood level and for both SH and DHW.

The data for prices of fuels (Table 10) come from different sources. For the wood pellets and wood chips, they come from the Norwegian Bioenergy Association.[6] The data for the biogas and biomethane come from the European Biogas Association.[7].

The data for CO_2 factor of fuels come from a report from Cundall[8].

References

[1] Driver HE, Kroeber AL. Quantitative expression of cultural relationships. University of California Publ Am Archaeol Ethnol 1932:211–56.

[2] MacQueen J. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley symposium on mathematical statistics and probability, Volume 1: Statistics. Berkeley, Calif.: University of California Press; 1967. p. 281–97. [Online]. Available: <https://projecteuclid.org/euclid.bsmsp/1200512992>.

[3] Kaufmann L, Rousseeuw P. Clustering by means of medoids. Data Anal L1-Norm and Related Methods 1987:405–16.

[4] Härtel P, Kristiansen M, Korpås M. Assessing the impact of sampling and clustering techniques on offshore grid expansion planning. Energy Proc 2017;137:152–61 14th Deep Sea Offshore Wind RD Conference, EERA DeepWind'2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1876610217353043.

[5] Pfenninger S. Dealing with multiple decades of hourly wind and pv time series in energy models: a comparison of methods to reduce time resolution and the planning implications of inter-annual variability. Appl Energy 2017;197:1–13 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306261917302775.

[6] Nahmmacher P, Schmid E, Hirth L, Knopf B. Carpe diem: a novel approach to select representative days for long-term power system modeling. Energy 2016;112:430–42 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544216308556.

[7] Wogrin S, Galbally D, Reneses J. Optimizing storage operations in medium- and long-term power system models. IEEE Trans Power Syst 2016;31(4):3129–38.

[8] Schütz T, Schraven MH, Fuchs M, Remmen P, Müller D. Comparison of clustering algorithms for the selection of typical demand days for energy system synthesis. Renew Energy 2018;129:570–82 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148118306591.

[9] Kotzur L, Markewitz P, Robinius M, Stolten D. Impact of different time series aggregation methods on optimal energy system design. Renew Energy 2018;117:474–87 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148117309783.

[10] Capuder T, Mancarella P. Techno-economic and environmental modelling and optimization of flexible distributed multi-generation options. Energy 2014;71:516–33 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544214005283.

[11] Yang Y, Zhang S, Xiao Y. Optimal design of distributed energy resource systems coupled with energy distribution networks. Energy 2015;85:433–48 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544215004107.

[12] Yokoyama R, Hasegawa Y, Ito K. A milp decomposition approach to large scale optimization in structural design of energy supply systems. Energy Convers Manage 2002;43(6):771–90 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0196890401000759.

[13] Weber, Shah N. Optimisation based design of a district energy system for an eco-town in the united kingdom. Energy 2011;36(2):1292–308 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544210006407.

[14] Harb H, Reinhardt J, Streblow R, Müller D. Mip approach for designing heating systems in residential buildings and neighbourhoods. J Build Perform Simul 2016;9(3):316–30 [Online]. Available: doi: 10.1080/19401493.2015.1051113.

[15] Morvaj B, Evins R, Carmeliet J. Optimising urban energy systems: simultaneous system sizing operation and district heating network layout. Energy 2016;116:619–36 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544216314207.

[16] Schütz T, Schiffer L, Harb H, Fuchs M, Müller D. Optimal design of energy conversion units and envelopes for residential building retrofits using a comprehensive milp model. Appl Energy 2017;185:1–15 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306261916314933.

[17] Mashayekh S, Stadler M, Cardoso G, Heleno M. A mixed integer linear programming approach for optimal der portfolio sizing and placement in multi-energy microgrids. Appl Energy 2017;187:154–68 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306261916316051.

[18] Piacentino A, Barbaro C. A comprehensive tool for efficient design and operation of polygeneration-based energy μgrids serving a cluster of buildings part ii: analysis of the applicative potential. Appl Energy 2013;111:1222–38 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306261913000706.

[19] Ortiga J, Bruno J, Coronas A. Selection of typical days for the characterisation of energy demand in cogeneration and trigeneration optimisation models for buildings. Energy Convers Manage 2011;52(4):1934–42 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0196890410005315.

[20] Fazlollahi S, Bungener SL, Mandel P, Becker G, Maréchal F. Multi-objectives, multi-period optimization of district energy systems: I. selection of typical operating periods. Comput Chem Eng 2014;vol. 65: 54–66 [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0098135414000751>.

[21] Li B, Roche R, Miraoui A. Microgrid sizing with combined evolutionary algorithm and milp unit commitment. Appl Energy 2017;188:547–62 [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0306261916318013>.

[22] Ashouri A, Fux , Benz MJ, Guzzella L. Optimal design and operation of building services using mixed-integer linear programming techniques. Energy 2013;59: 365–76. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0360544213005525>.

[23] Gabrielli P, Gazzani M, Martelli E, Mazzotti M. Optimal design of multi-energy systems with seasonal storage. Appl Energy, 2018;219: 408–24, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/

[5] https://ens.dk/en/our-services/projections-and-models/technology-data
[6] http://nobio.no/wp-content/uploads/2018/01/Veien-til-biovarme.pdf.
[7] https://www.europeanbiogas.eu/wp-content/uploads/2019/07/Biomethane-in-transport.pdf

S0306261917310139>.

[24] Stadler P, Ashouri A, Maréchal F. Model-based optimization of distributed and renewable energy systems in buildings. Energy Build 2016;120: 103–13. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0378778816302079>.

[25] Fleischhacker A, Lettner G, Schwabeneder D, Auer H. Portfolio optimization of energy communities to meet reductions in costs and emissions. Energy 2019;173: 1092–105. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0360544219303032>.

[26] Jones E, Oliphant T, Peterson P, et al. SciPy: Open source scientific tools for Python; 2001–, [Online; accessed 22.05.2019]. [Online]. Available: <http://www.scipy.org/>.

[27] Novikov A. PyClustering: data mining library. J Open Source Software, 2019;4(36): 1230. [Online]. Available: https://doi.org/10.21105/joss.01230.

[28] Arthur D, Vassilvitskii S. k-means: the advantages of careful seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM symposium on discrete algorithms, SODA 2007, New Orleans, Louisiana, USA; January 2007, p. 9.

[29] Hellman HP, Koivisto M, Lehtonen M. Photovoltaic power generation hourly modelling. Proceedings of the 2014 15th international scientific conference on electric power engineering (EPE). 2014. p. 269–72.

[30] Clauß J, Stinner S, Solli C, Lindberg KB, Madsen H, Georges L. Evaluation Method for the Hourly Average CO2eq. Intensity of the electricity mix and its application to the demand response of residential heating. Energies 2019;12(7): 1345. [Online]. Available: <https://www.mdpi.com/1996-1073/12/7/1345>.

[31] Clark D. CO_2 Emissions from Biomass and Biofuels. Information paper: Cundall; 2013.

[32] Bioenergi i Norge: Markedsrapport for pellets 2017, Norsk Bioenergiforening, NOBIO. Tech. Rep.; 2017. <http://nobio.no/wp-content/uploads/2019/01/Pris-og-salgsstatistikk-for-pellets-i-Norge-2017.pdf>, Accessed June 19.

[33] Vinterbäck J, Porsö C. EUBIONET3 WP3 - Wood fuel price statistics in europe - d 3. 3, Swedish University of Agricultural Sciences, Uppsala, Tech. Rep.; 2011. <https://ec.europa.eu/energy/intelligent/projects/sites/iee-projects/files/projects/documents/eubionet_iii_wood_fuels_price_statistics_in_europe_en.pdf>, Accessed June 19.

[34] Trømborg E. IEA Bioenergy Task 40: Country report 2013 for Norway, Norwegian University of Life Sciences, Ås. Tech. Rep.; 2015, <http://task40.ieabioenergy.com/wp-content/uploads/2013/09/iea-task-40-country-report-2014-norway.pdf>, Accessed June 19.

[35] Biomethane in transport, European Biogas Association. Tech. Rep.; 2016. <http://european-biogas.eu/wp-content/uploads/2016/05/BiomethInTransport.pdf>, Accessed June 19.

[36] Lindberg KB. Impact of Zero Energy Buildings on the Power System: A study of load profiles, flexibility and system investments, Ph.D. dissertation, NTNU; 2017.

[37] Pal SK, Alanne K, Jokisalo J, Siren K. Energy performance and economic viability of advanced window technologies for a new Finnish townhouse concept. Appl Energy 2016;162:11–20.