Machine Learning Assisted Analysis of Gas Consumption Usage

User Consumption Profile Analysis focusing on Uilenstede Campus

BUILDING TECHNOLOGY MASTER TRACK

Faculty of Architecture and the Built Environment

First Mentor: Dr. Laure Itard

Second Mentor: Dr. Charalampos Andriotis

Student: Yu Hsiu Tung

Machine Learning Assisted Analysis of Gas Consumption Usage

User Consumption Profile Analysis focusing on Uilenstede Campus

BUILDING TECHNOLOGY MASTER TRACK

Faculty of Architecture and the Built Environment

First Mentor: Dr. Laure Itard

Second Mentor: Dr. Charalampos Andriotis

Student: Yu Hsiu Tung

Content

Problem background

Energy Crisis and Rising Gas Price

- Energy Efficient Building
- Cost Savings

€

Moving towards an energy efficient building

Building

Energy usage Gas €

DUWO's Uilenstede project

The site

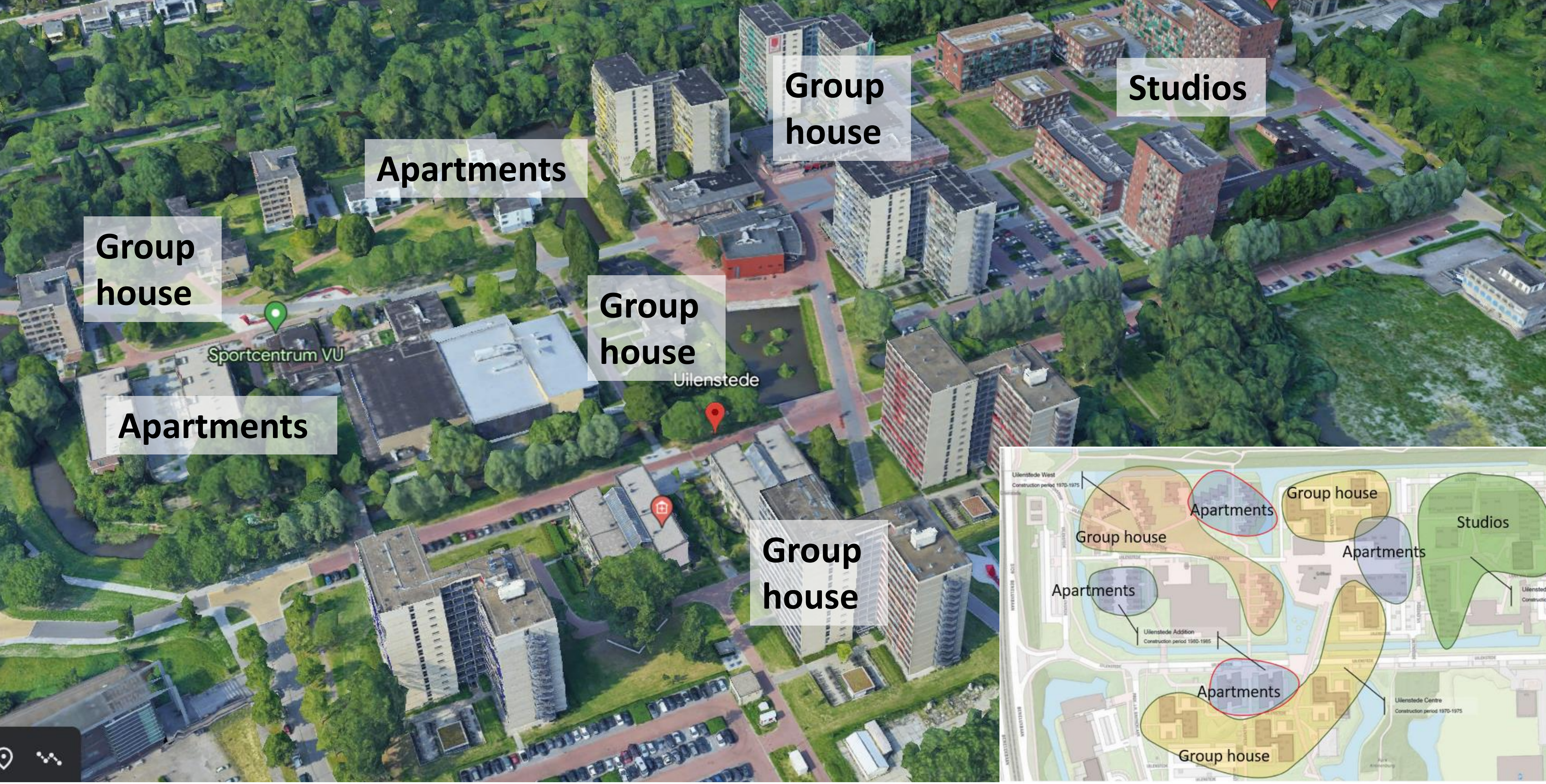
South of Amsterdam Uilenstede

Site location

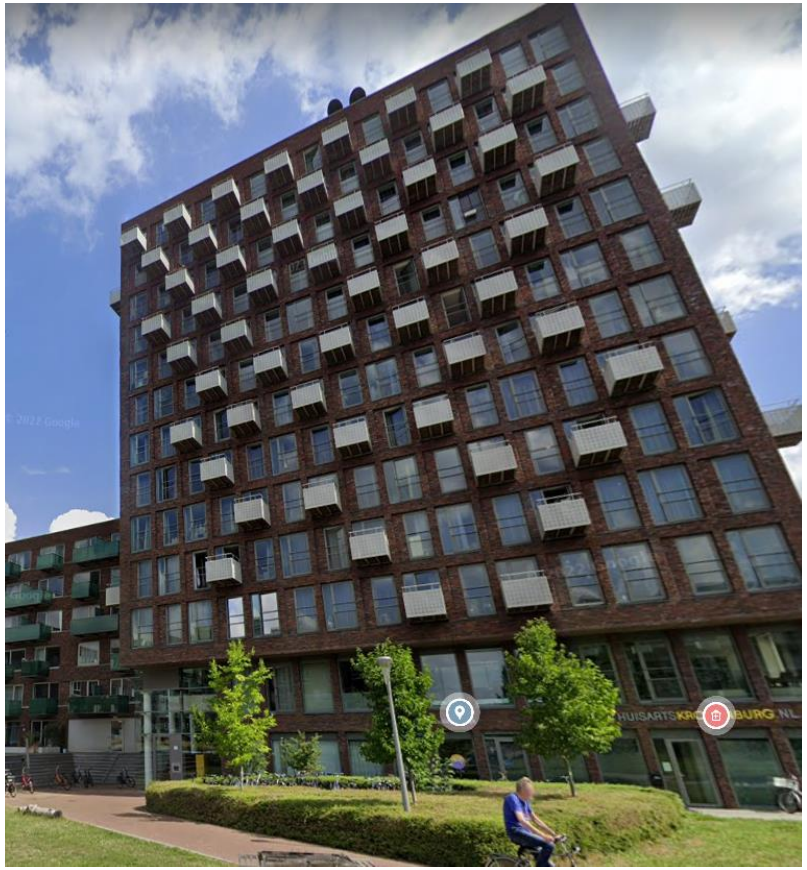
Source: google earth

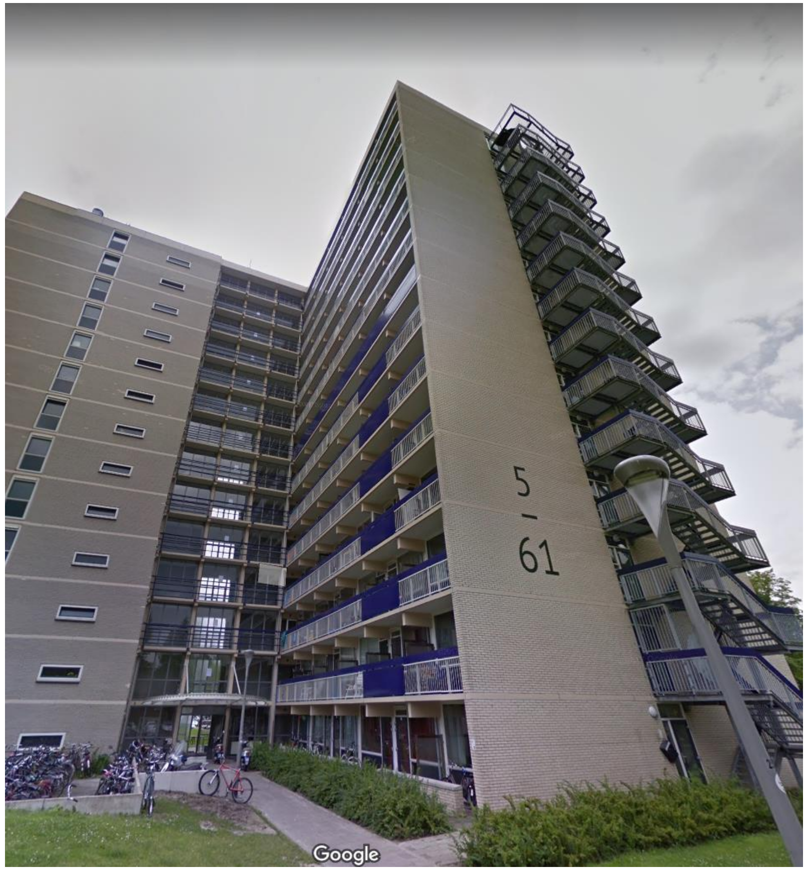

2010s
1970s
1990s
Sportcentrum VU
Uilenstede
1980s
1970s
2D

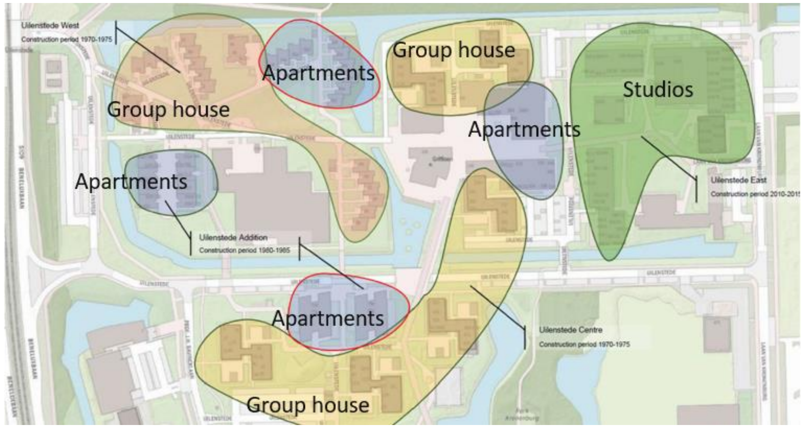
Built year



Housing types

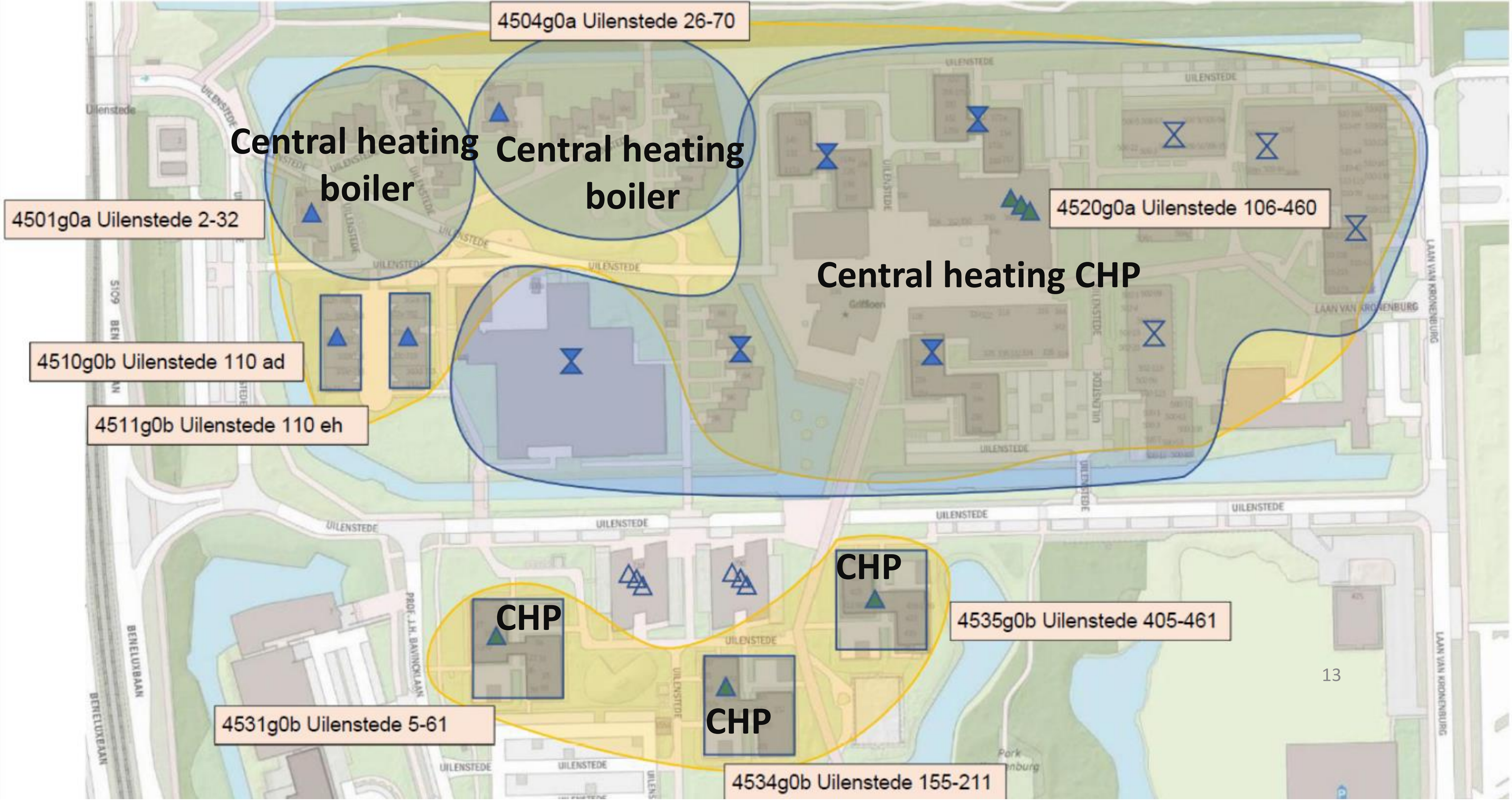






Housing types

Central heating system with Combined Heat and Power (CHP) systems and boiler



Heating system of the building

Information

Building Type & Energy Usage

Uilenstede project

Research Question

How can energy consumption profiles guide energy conservation strategies?

Understanding Patterns of Gas Consumption

Influence of Building Features on Energy Efficiency

Research framework

Consumption data

Usage profiles

Building features

Built year, insulation values, heating system

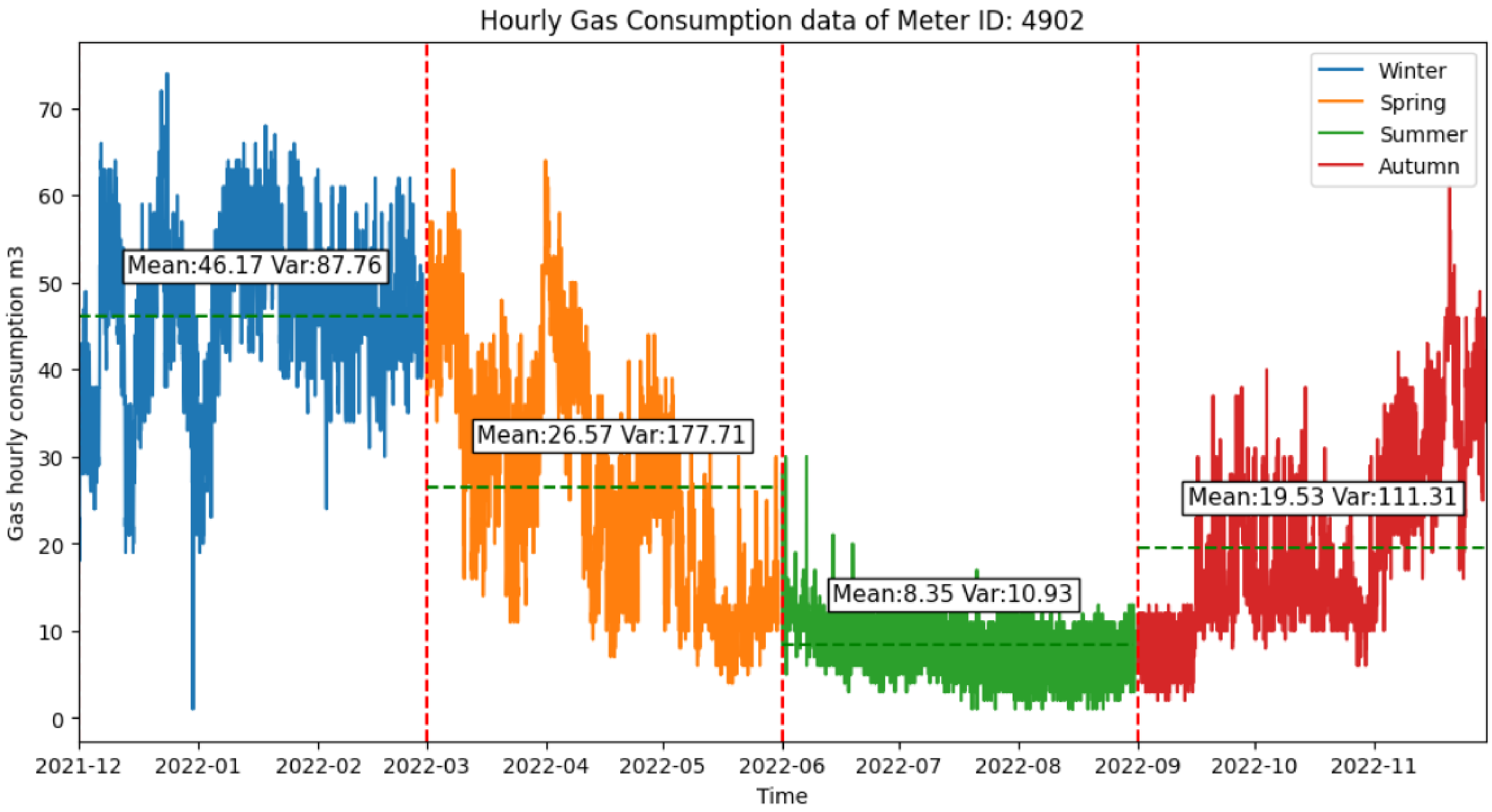
Relation between

Summarize data

Features of the buildings

Type	Value
Roof	Enhanced with insulation of approximately 5 – 10 cm Rc = 1.5 (m2K)/W.
Walls	Contain uninsulated cavities, resulting in an insulation value of Rc = 0.36 (m2K)/W.
Windows, Glass + Frame, and Doors	Insulation value of U=1.7W/(m2K).

Hourly gas consumption

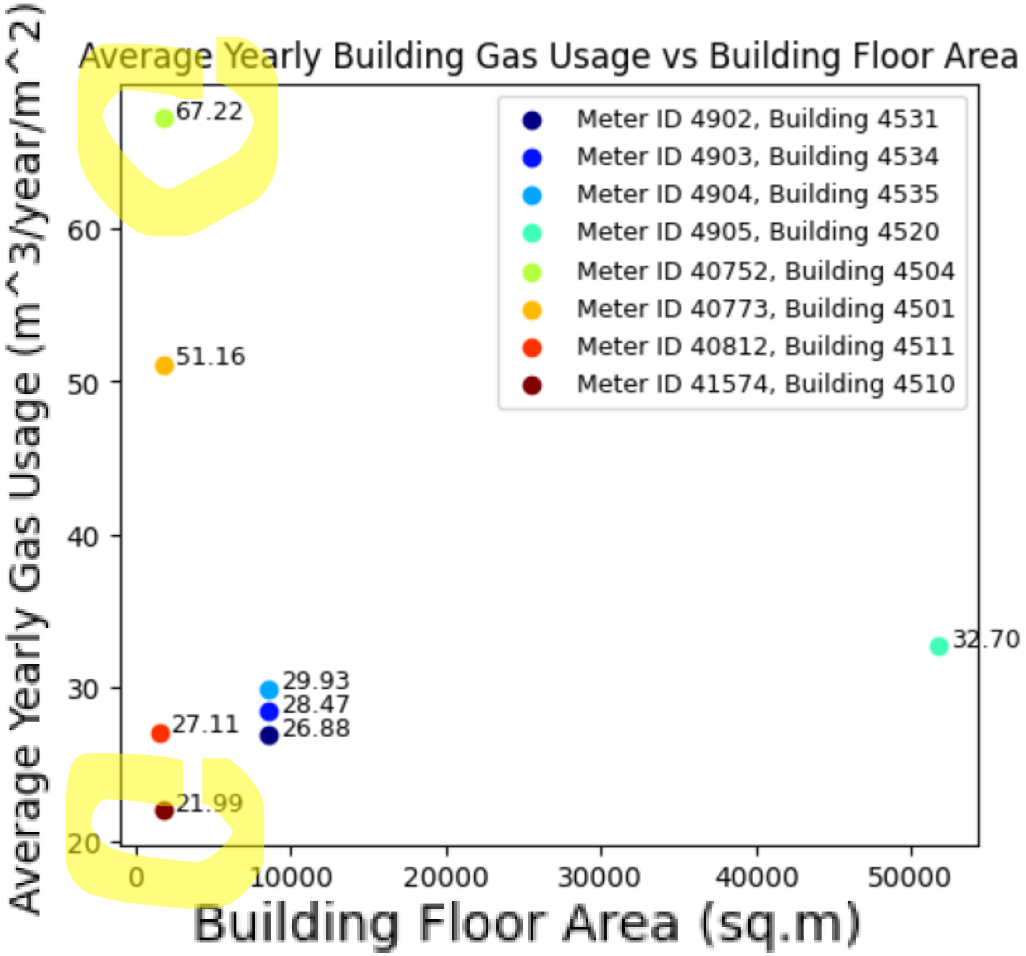


Analyze data

Average gas usage per floor area

Hot tap water percentage

Annual gas consumption normalized by floor area



- Building 4504 is nearly **three times difference** with Building 4510 despite both serving **similar floor areas**
- Varying energy efficiency of these buildings.

Table 2.: Building Numbers, ID Descriptions, and Floor Areas

Building no.	Complex Number	Meter ID	Floor Area (sq.m)
4531	4902	Uilenstede 5-61 (Toren 1)	8555
4534	4903	Uilenstede 155-211 (Toren 2)	8553
4535	4904	Uilenstede 405-461 (Toren 3)	8552
4520	4905	Uilenstede 106-460	51766
4504	40752	Uilenstede 36-70	1801
4501	40773	Uilenstede 2-32	1843
4511	40812	Uilenstede 102 e-h (trad)	1534
4510	41574	Uilenstede 102 a-d (H-inst)	1798

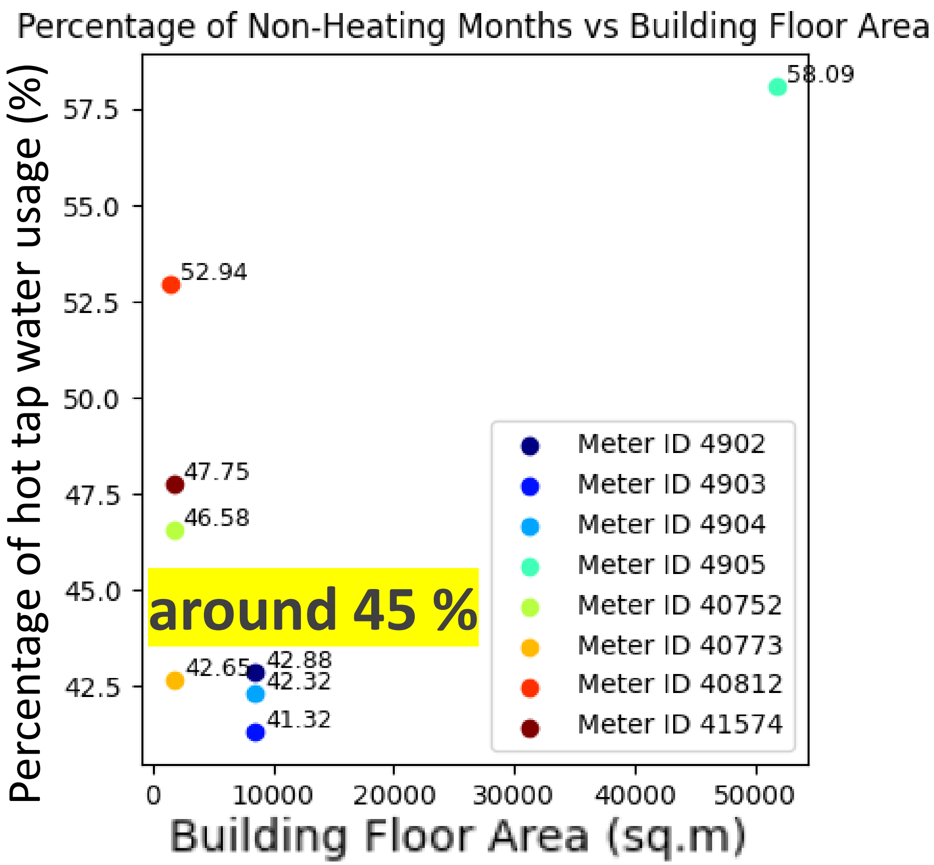
Hot tap water gas consumption percentage

Shower, cooking

summer months, heating is not required

$$\text{Hot tap water percentage} = \frac{\text{Averaged summer months consumption (May-Sep)}}{\text{Averaged total consumption}}$$

Hot tap water percentage vs floor area

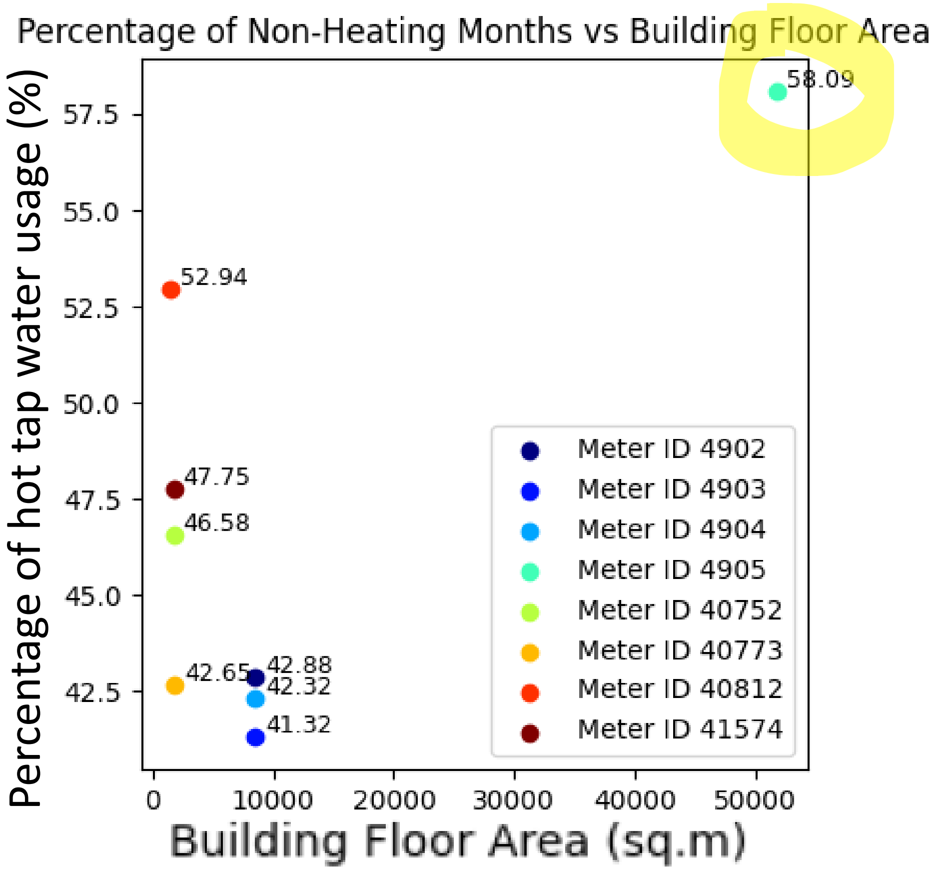


Typical residential home, it's often estimated that about **15-25%**

Table 2.: Building Numbers, ID Descriptions, and Floor Areas

Building no.	Complex Number	Meter ID	Floor Area (sq.m)
4531	4902	Uilenstede 5-61 (Toren 1)	8555
4534	4903	Uilenstede 155-211 (Toren 2)	8553
4535	4904	Uilenstede 405-461 (Toren 3)	8552
4520	4905	Uilenstede 106-460	51766
4504	40752	Uilenstede 36-70	1801
4501	40773	Uilenstede 2-32	1843
4511	40812	Uilenstede 102 e-h (trad)	1534
4510	41574	Uilenstede 102 a-d (H-inst)	1798

Hot tap water percentage vs floor area



•Meter #4905: Has the highest hot tap water usage percentage of 58.09%, it has a **central heating** system, attributed to heat losses that occur during the distribution process.

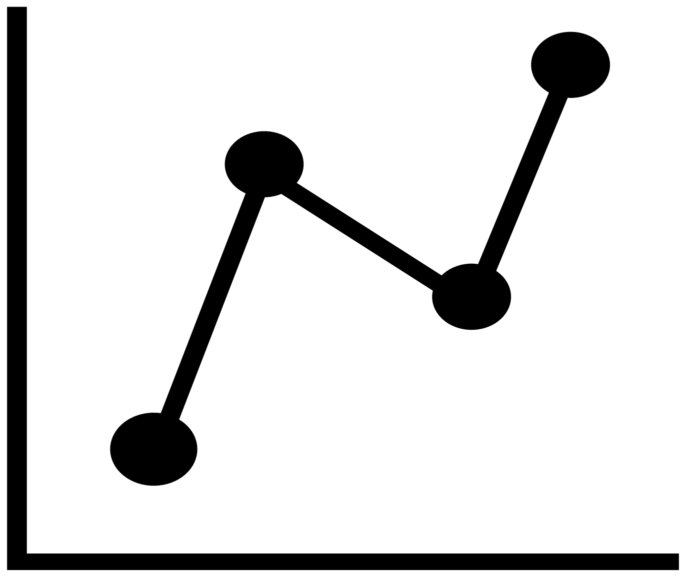
Table 2.: Building Numbers, ID Descriptions, and Floor Areas

Building no.	Complex Number	Meter ID	Floor Area (sq.m)
4531	4902	Uilenstede 5-61 (Toren 1)	8555
4534	4903	Uilenstede 155-211 (Toren 2)	8553
4535	4904	Uilenstede 405-461 (Toren 3)	8552
4520	4905	Uilenstede 106-460	51766
4504	40752	Uilenstede 36-70	1801
4501	40773	Uilenstede 2-32	1843
4511	40812	Uilenstede 102 e-h (trad)	1534
4510	41574	Uilenstede 102 a-d (H-inst)	1798

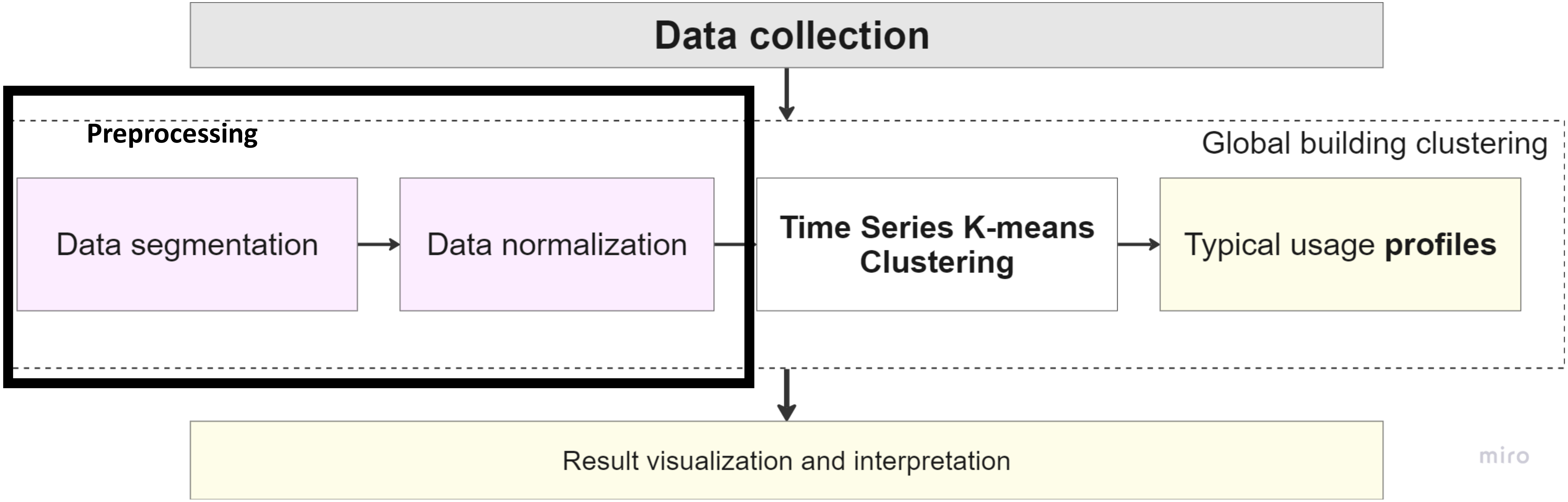
Consumption profiles

Consumption profiles

By identifying when and how energy is used, it is possible to find opportunities to reduce consumption.



Framework



Data collection
Preprocessing
Global building clustering
Data segmentation
Data normalization
Time Series K-means Clustering
Typical usage profiles
Result visualization and interpretation
miro

Find the periodicity

- Hourly
- Daily
- Weekly?

Autocorrelation function

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T} \underbrace{(r_t - \bar{r})}_{\text{original}} \underbrace{(r_{t-k} - \bar{r})}_{\text{shifted}}}{\sum_{t=1}^{T} (r_t - \bar{r})^2}$$

- **Variables Explained**:
- r(t) = original data
- r(t-k) = data shifted by k units.
- r_bar = The average of original data

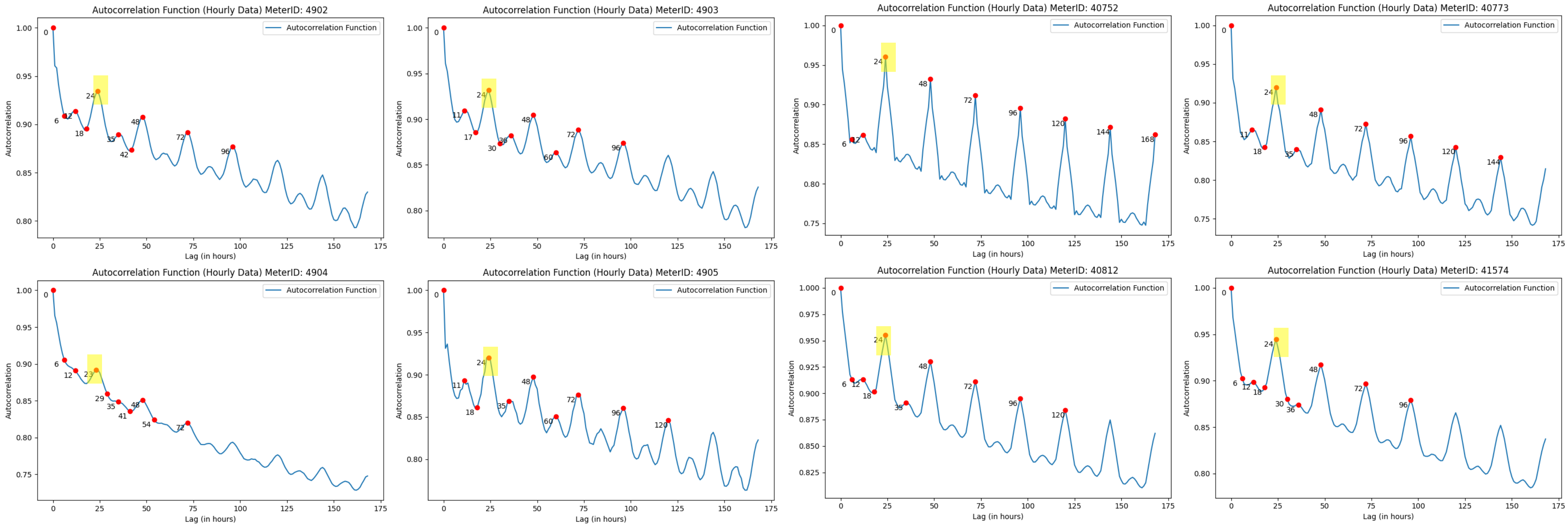
Autocorrelation function

K = k-unit shifted data

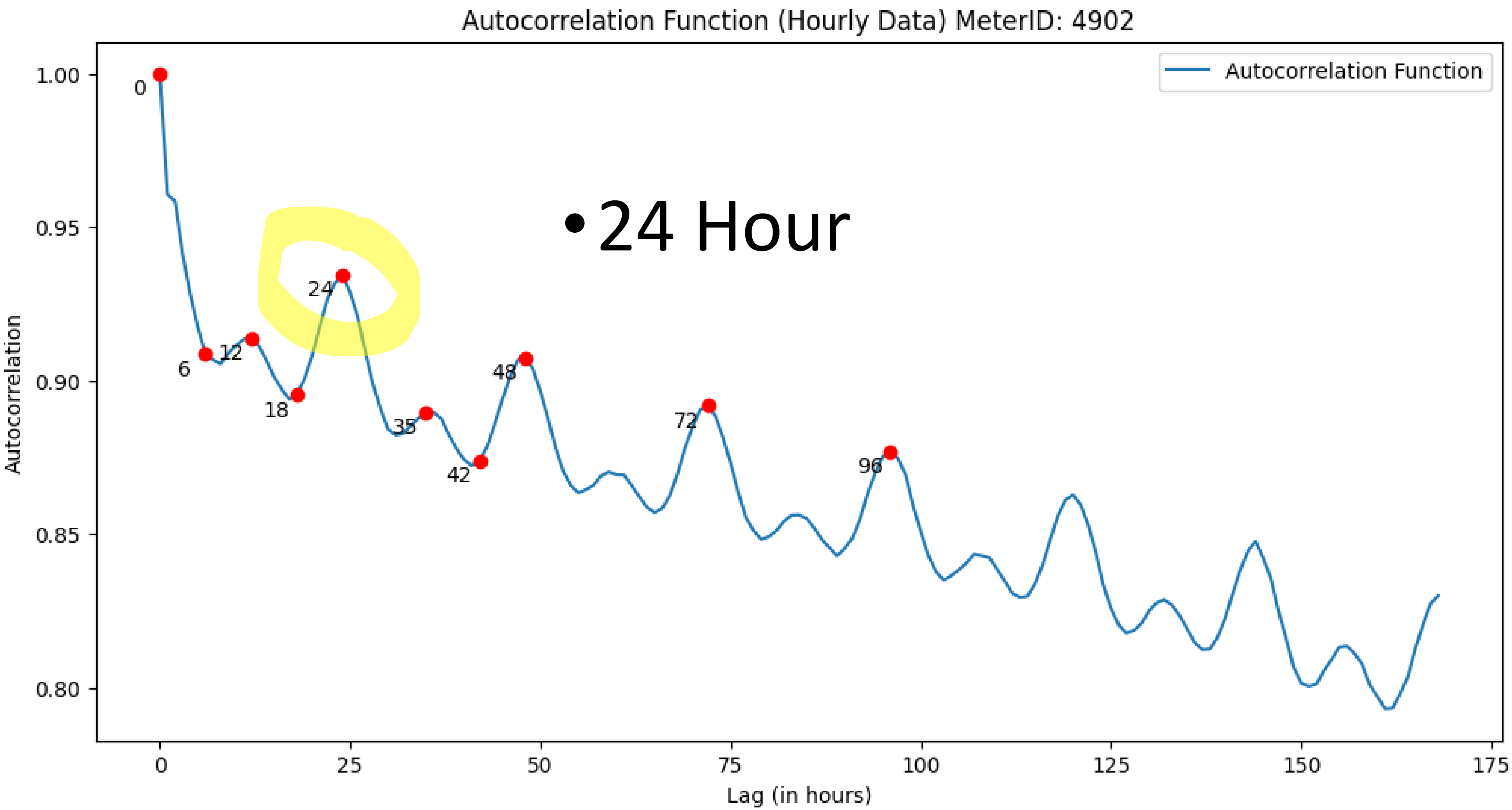
$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T} (r_t - \bar{r})(r_{t\boxed{-k}} - \bar{r})}{\sum_{t=1}^{T} (r_t - \bar{r})^2}$$

- **Variables Explained**:
- r(t) = original data
- r(t-k) = data shifted by k units.
- r_bar = The average of original data

Autocorrelation across all meters

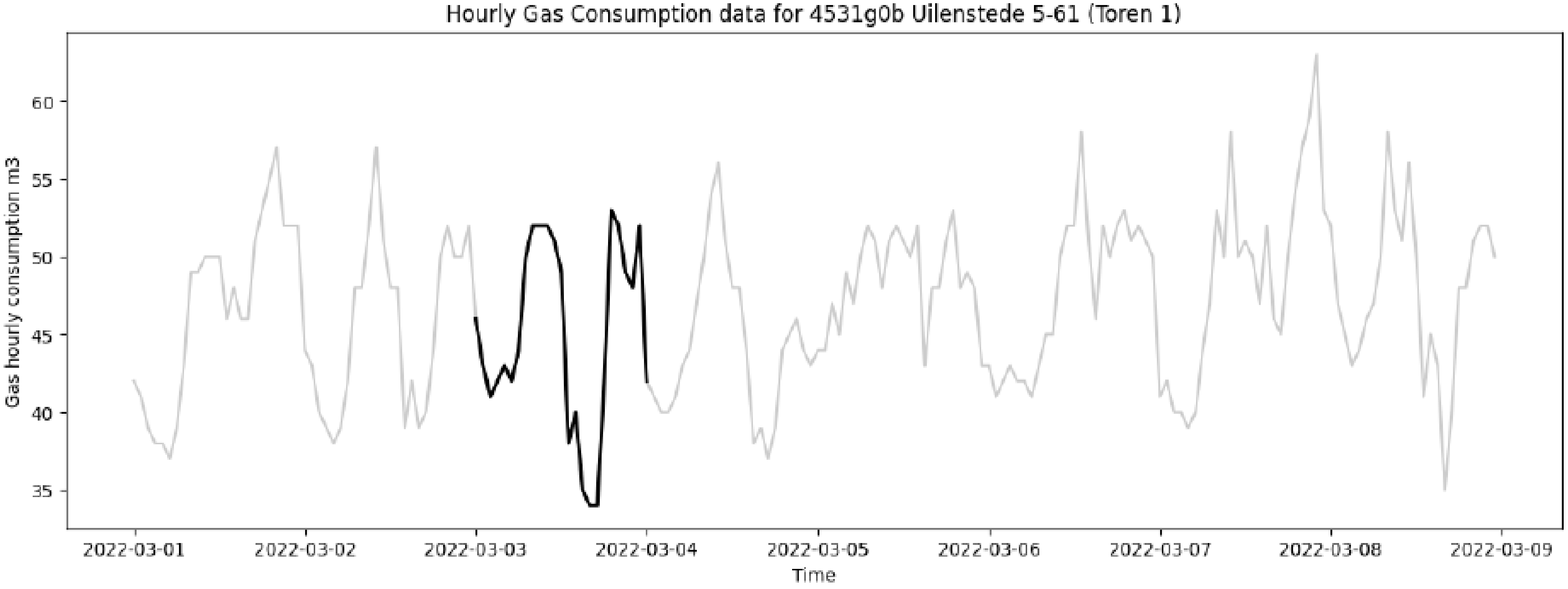


Autocorrelation result - hourly



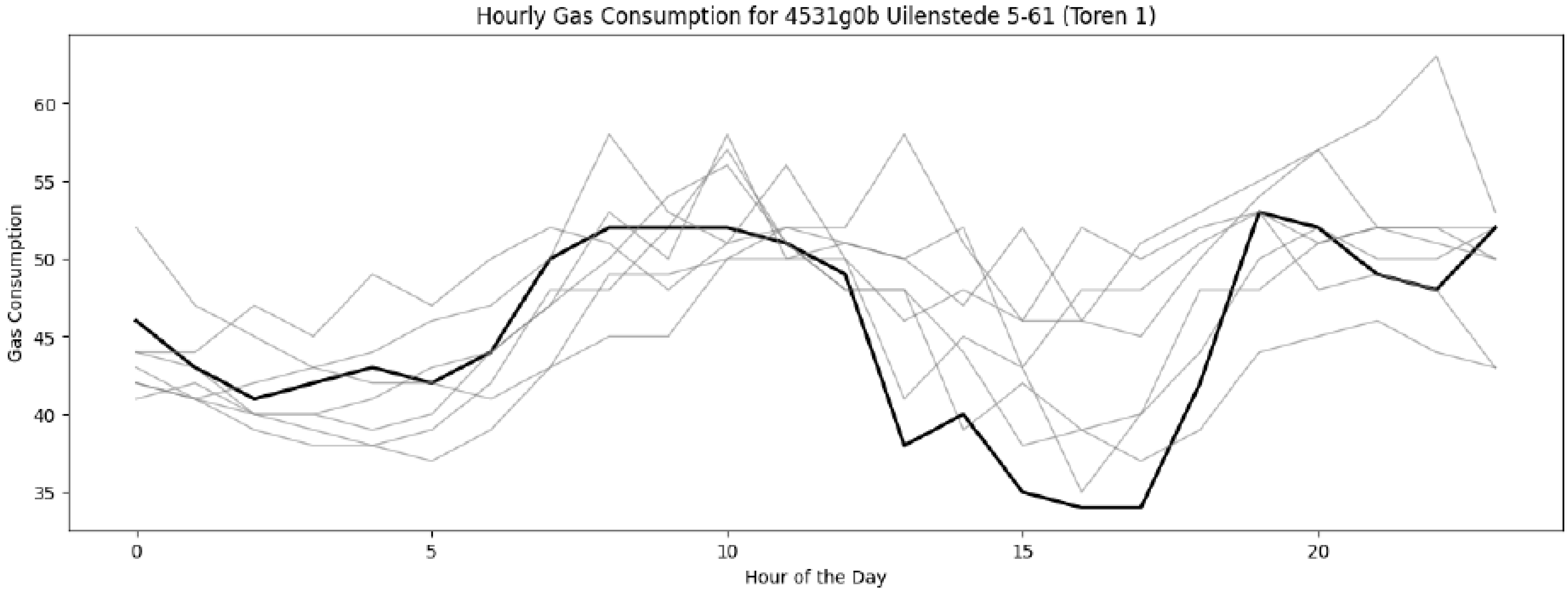
- 24 Hour

Segmentation – yearly data



(a) Usage profiles before the data segmentation of building 4531

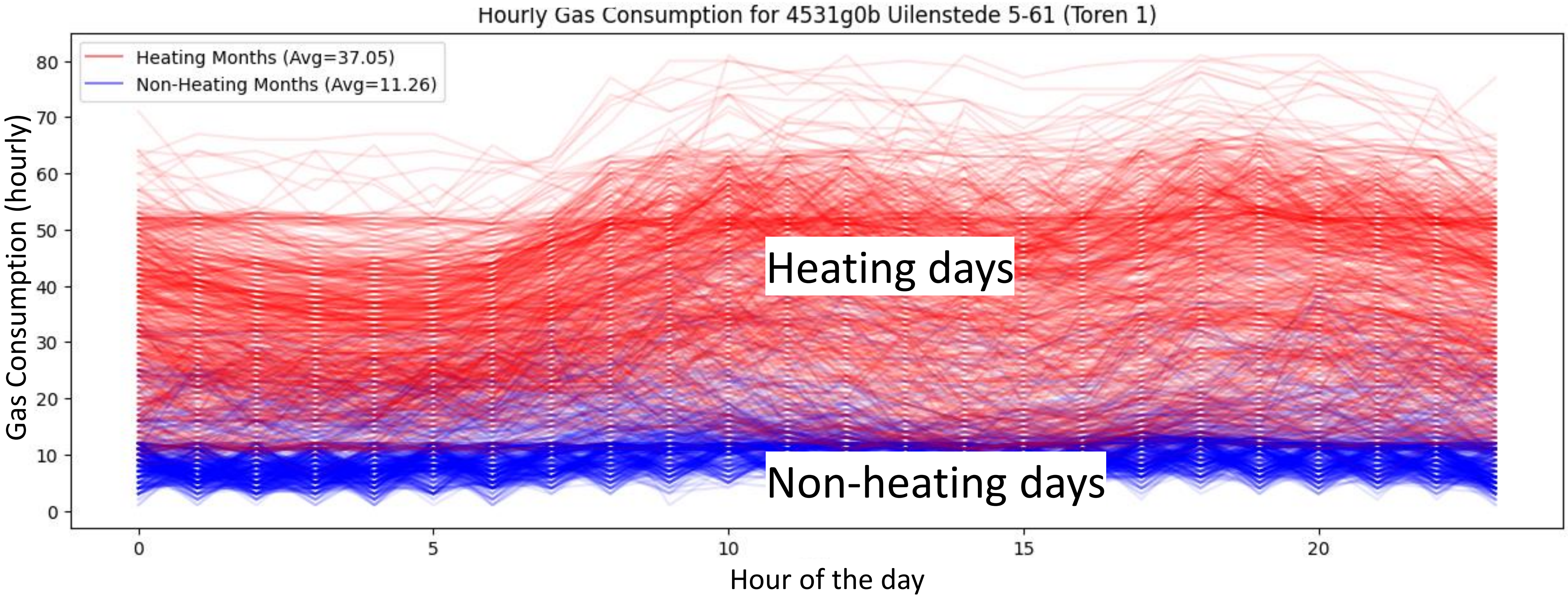
Segmentation – daily data



(b) Usage profiles after the data segmentation of building 4531

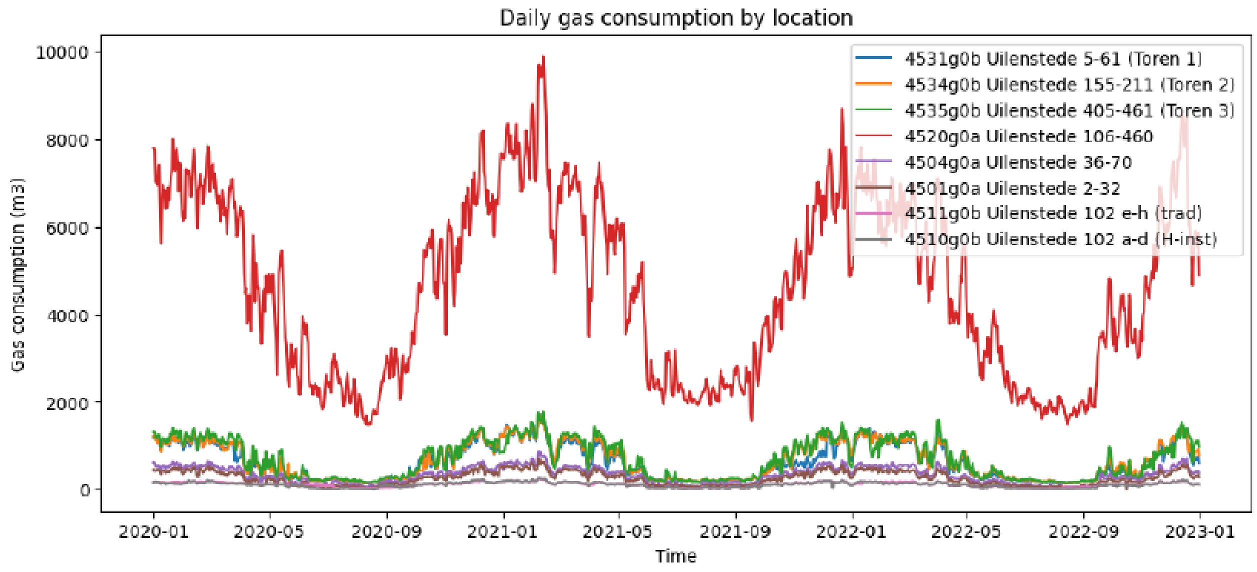
Consumption on heating and non heating months

— Heating months: Oct - Apr

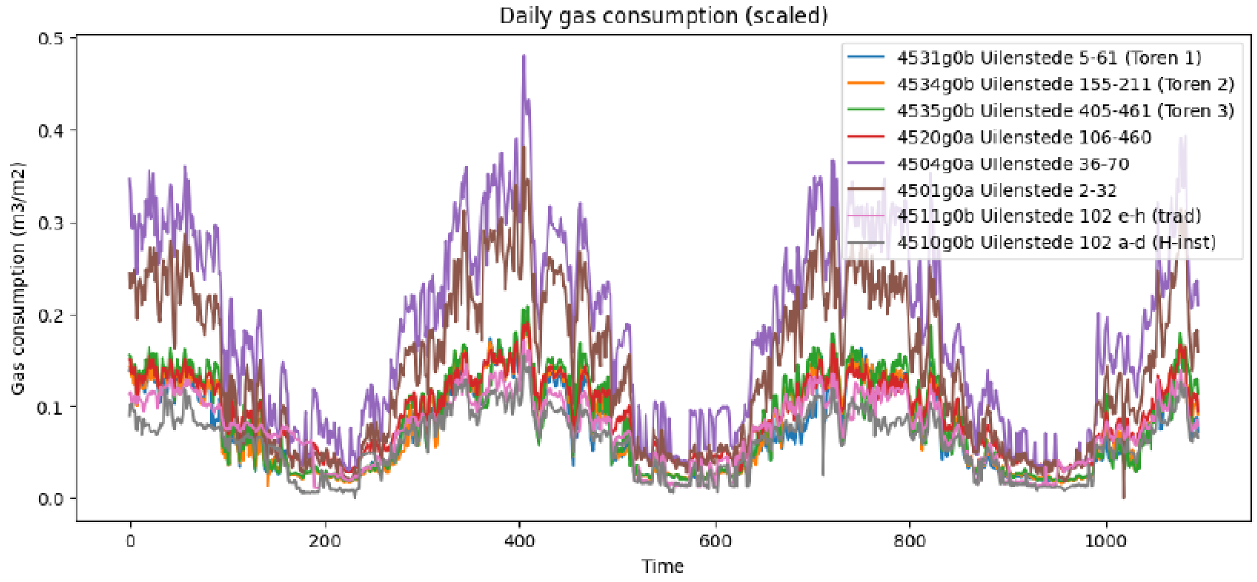


Scaling

- By floor area



((a)) Daily consumption data 2020-2023 across locations

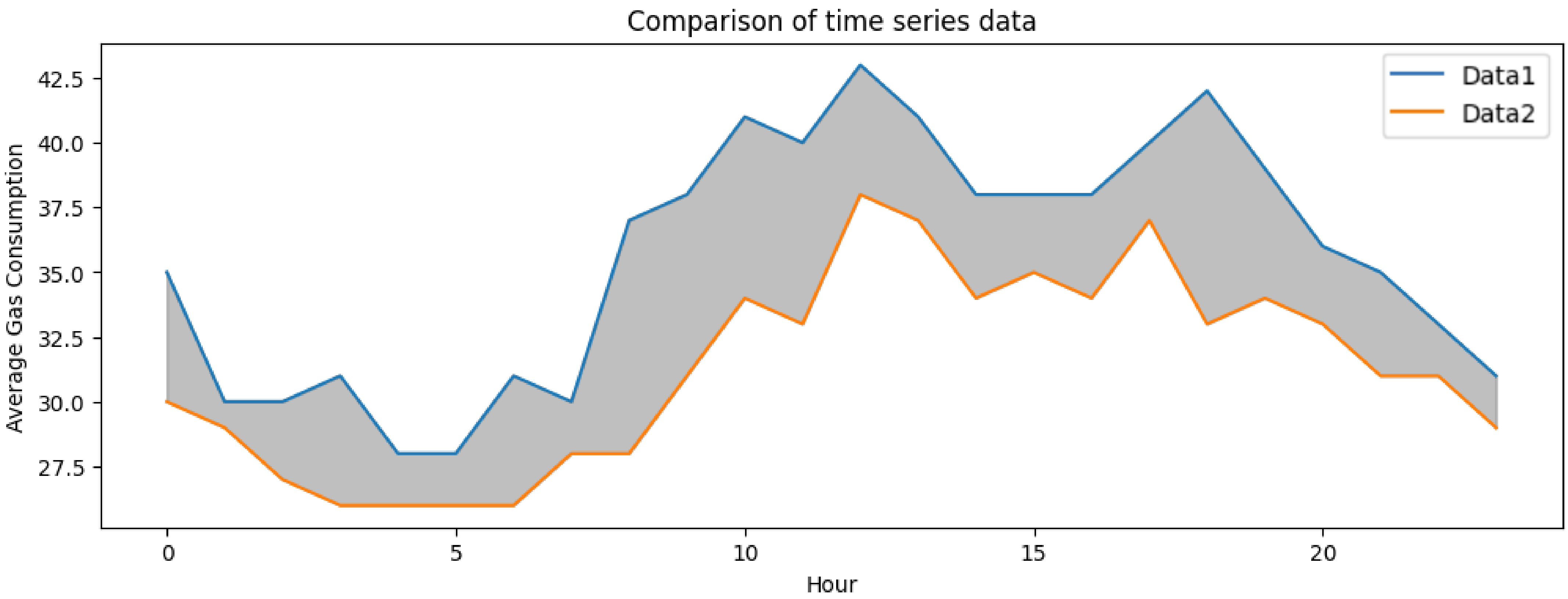


((b)) Scaled Daily consumption data 2020-2023 across locations

Figure 4.13.: Usage profiles for Building 4531

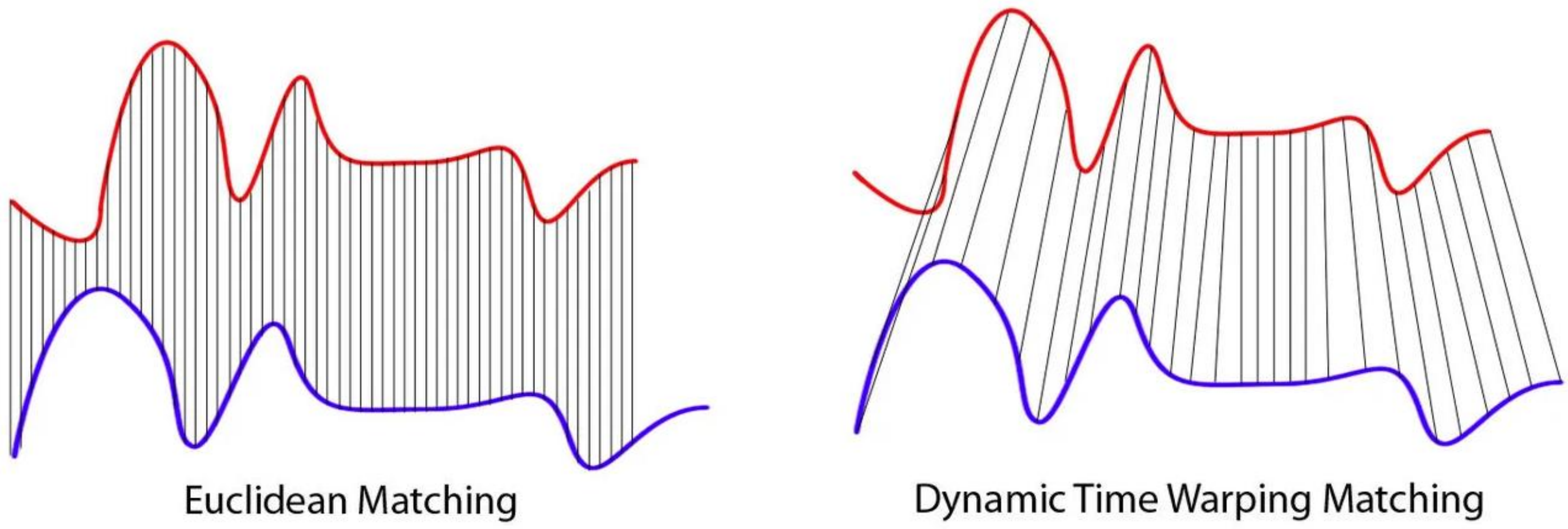
How to find similarity

Finding similar time-series in shape



Distance measurements

- Euclidean distance vs Dynamic Time Warping



Source: https://medium.com/mlearning-ai/what-is-dynamic-time-warping-253a6880ad12

Dynamic time warping

- Calculating distance between time series data

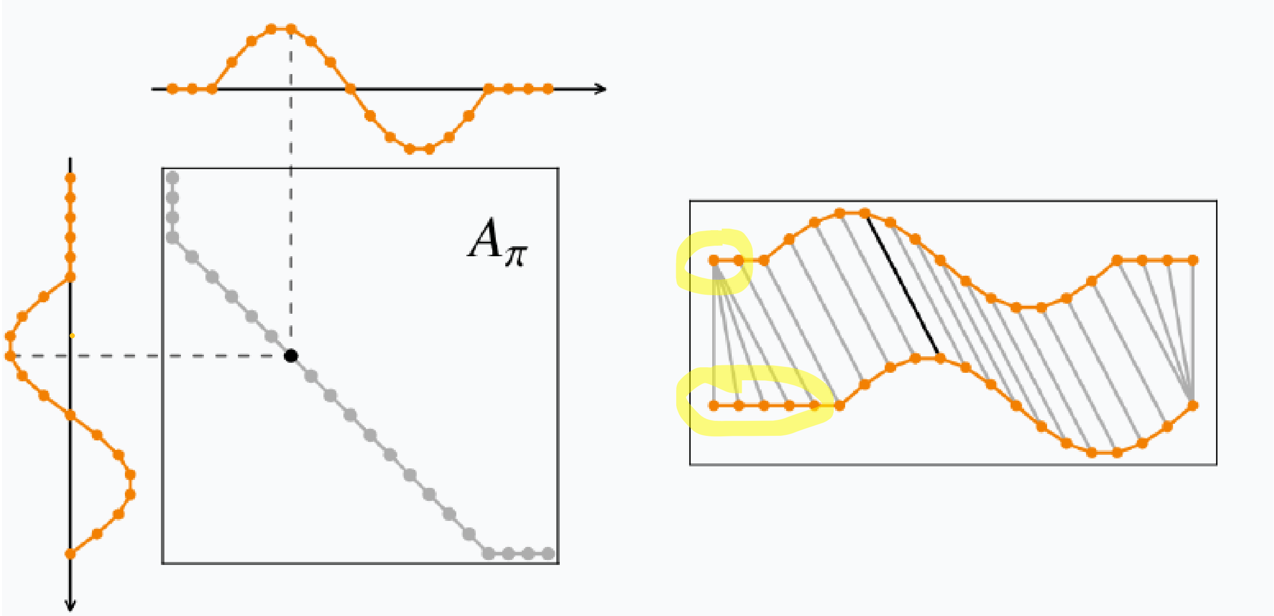


Figure 4.9.: Visual representation of Dynamic Time Warping

Tavenard, R. (n.d.). *An introduction to Dynamic Time Warping*. Github.Io. Retrieved June 1, 2023, from https://rtavenar.github.io/blog/dtw.html

How to group similar time-series data

Identify similar patterns

How to group similar time-series data

Distance Matrix (pairwise distance)

-> computationally intensive process **time complexity O(n^2)**

Principal component analysis (PCA)

-> suitable for high-dimensional data, data is only two dimension

Support Vector Machines (SVM)

-> require a lot of tunning for kernel and the regularization parameter

1D Convolutional Neural Networks

->black box approach, difficult to explain why certain classifications were made

How to group similar time-series data

~~**Distance Matrix (pairwise distance)**~~

~~-> computationally intensive process **time complexity O(n^2)**~~

~~Principal component analysis (**PCA**)~~

~~-> suitable for high-dimensional data, data is only two dimension~~

~~**Support Vector Machines (SVM)**~~

~~-> require a lot of tunning for kernel and the regularization parameter~~

~~**1D Convolutional Neural Networks**~~

~~->black box approach, difficult to explain why certain classifications were made~~

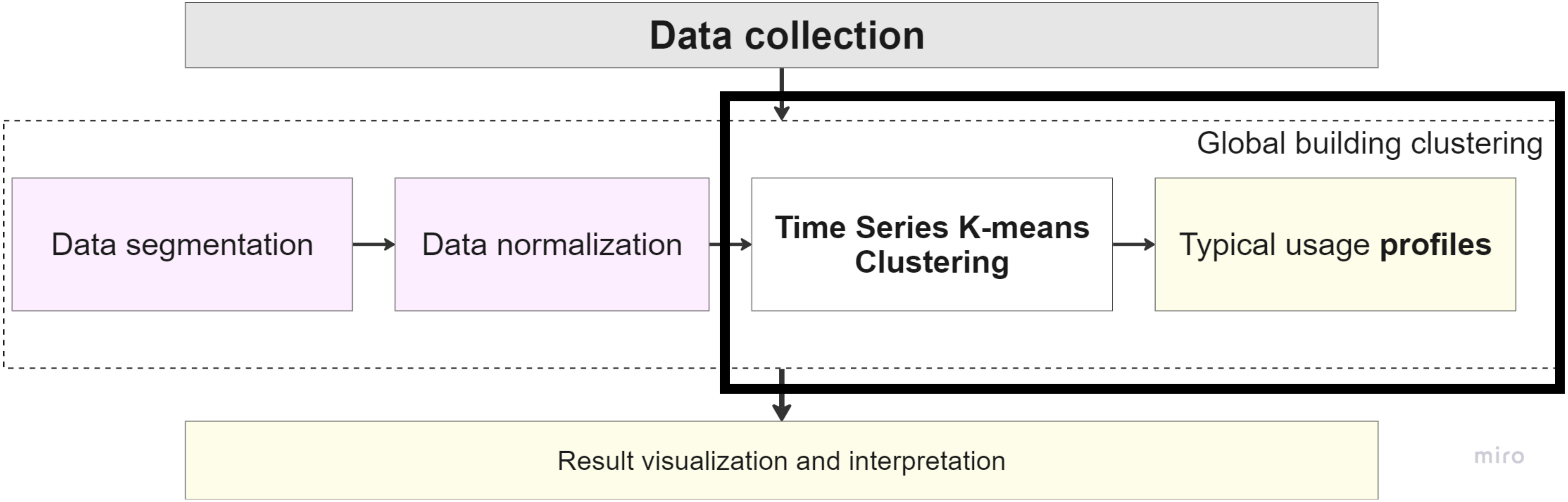
Clustering

Clustering

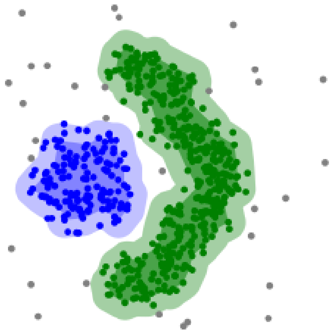
For exploratory analysis

- Straightforward
- Interpretable

Clustering

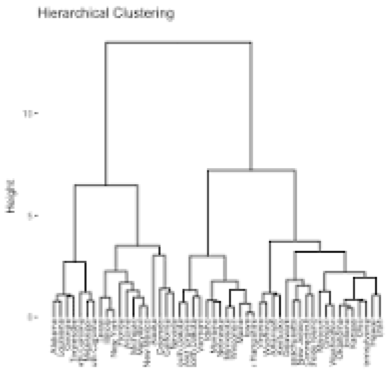

Data collection
Global building clustering
Data segmentation
Data normalization
Time Series K-means Clustering
Typical usage profiles
Result visualization and interpretation
miro

Different clustering algorithms



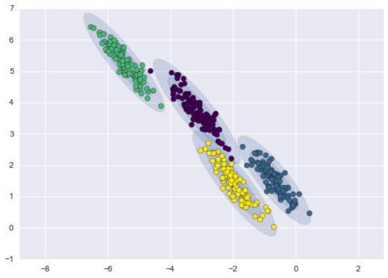
Density-based Clustering (DBSCAN):

Density don't always hold for time-series data.



Hierarchical Clustering:

Computationally intensive, time complexity of O(n^2 log(n))



Gaussian Mixture (GM) Model:

Each data point is drawn independently from its Gaussian distributions, does not consider temporal dependency

Different clustering methods

~~***Density-based Clustering (DBSCAN)***:~~

~~Density don't always hold for time-series data.~~

~~***Hierarchical Clustering***~~:

~~Computationally intensive, time complexity of O(n^2 log(n))~~

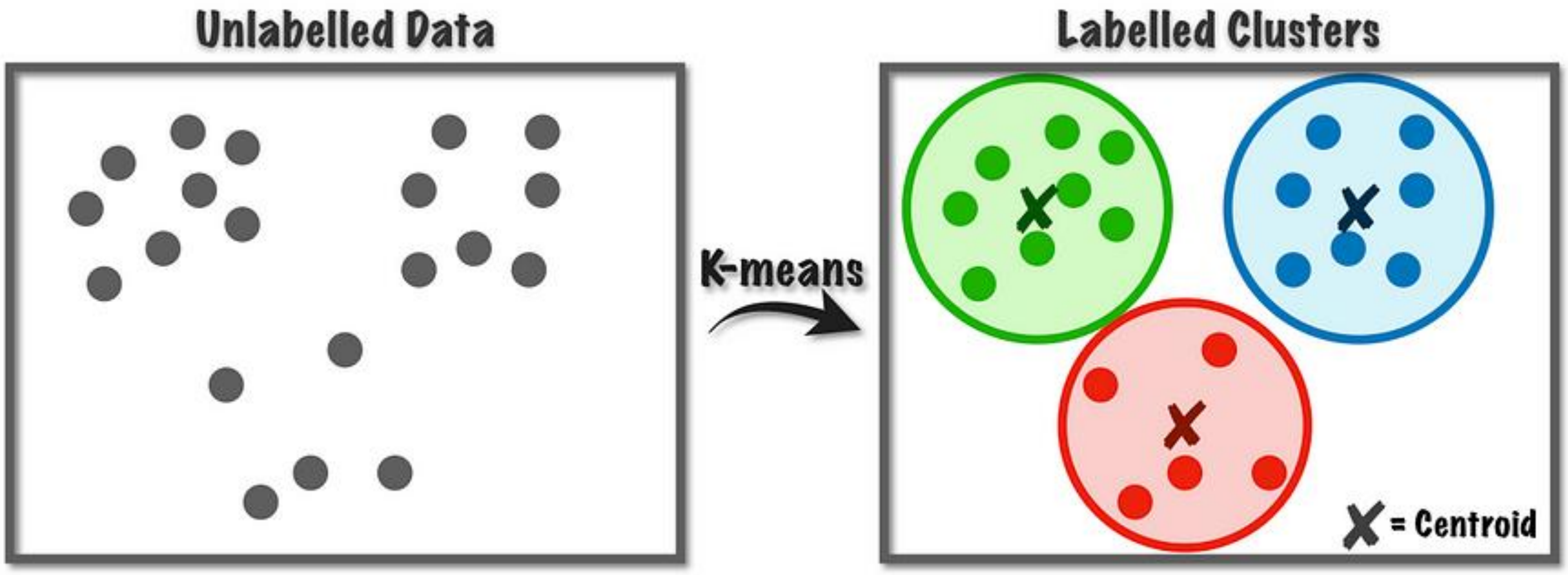
~~***Gaussian Mixture (GM) Model***~~:

~~Each data point is drawn independently from its Gaussian distributions, does not consider temporal dependency~~

K-Means clustering is faster and easier to interpret for initial analysis.

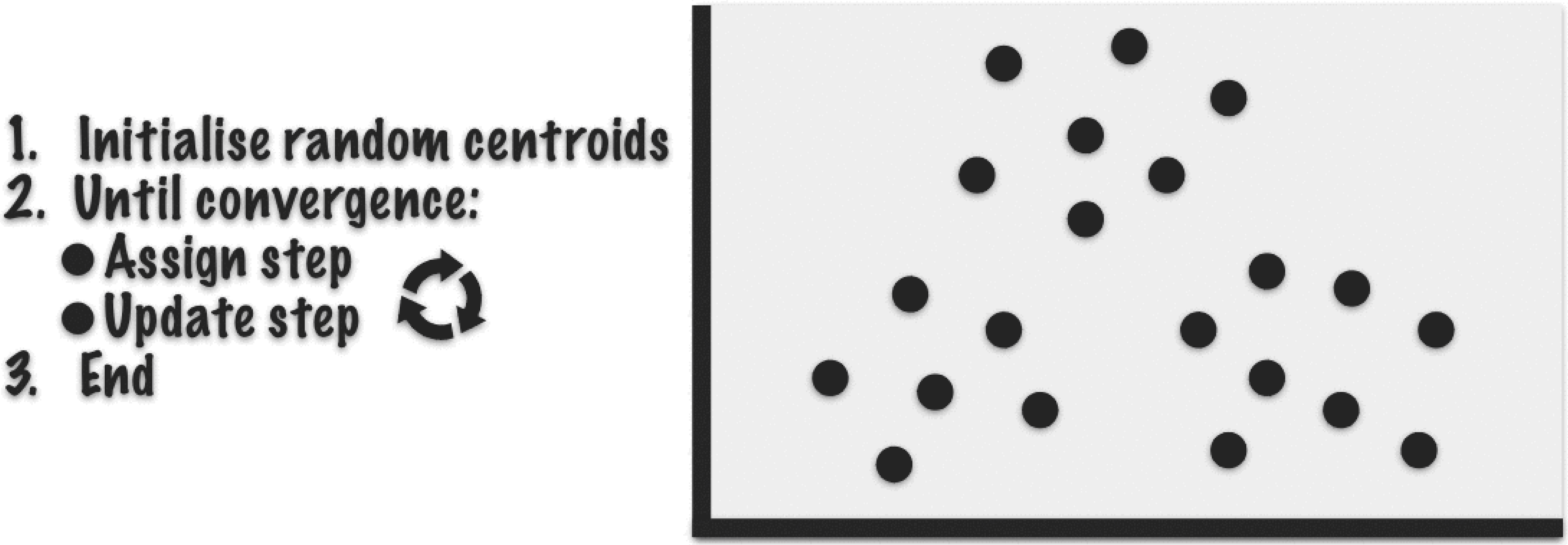
K-Means Clustering

Group similar data together



Source: https://towardsdatascience.com/k-means-a-complete-introduction-1702af9cd8c

K-Means Clustering



- Source: https://towardsdatascience.com/k-means-a-complete-introduction-1702af9cd8c

Clustering Result

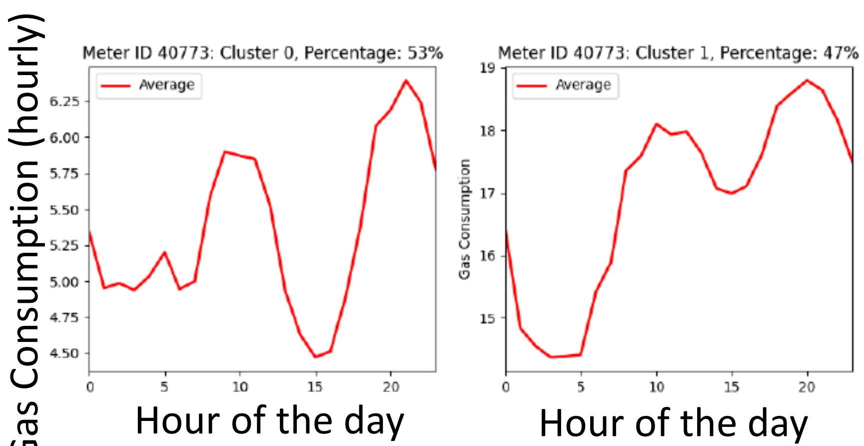
Single building

Globally across all meters

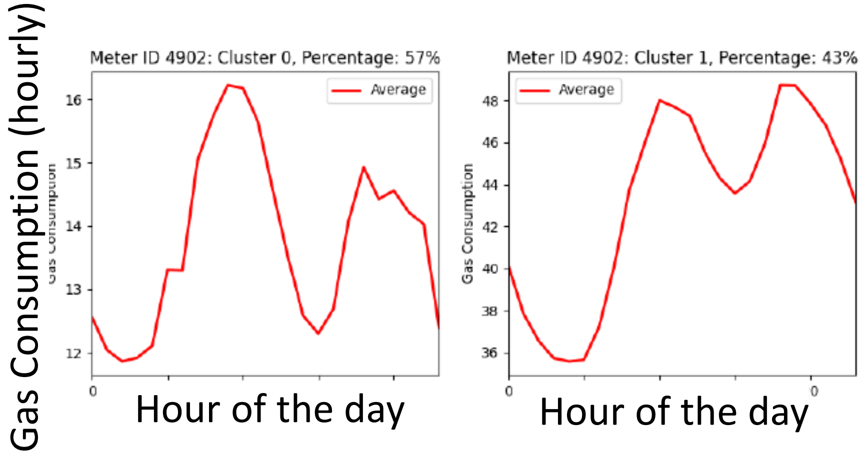
Clustering result for single building

Consumption pattern for each building

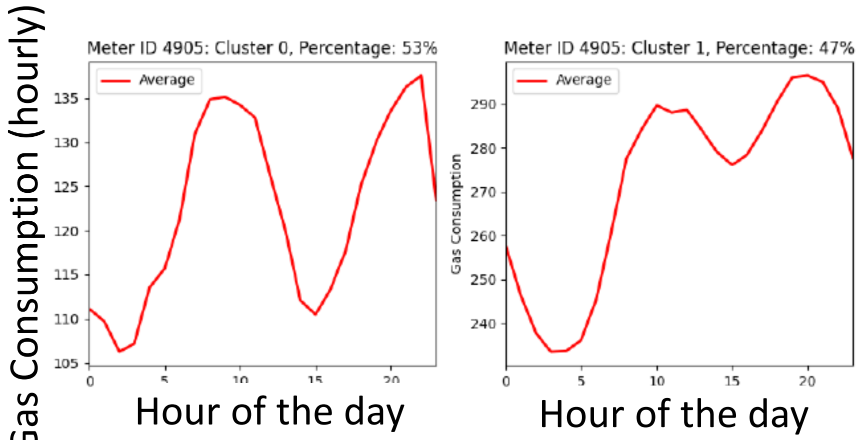
Typical daily profiles of 4 building



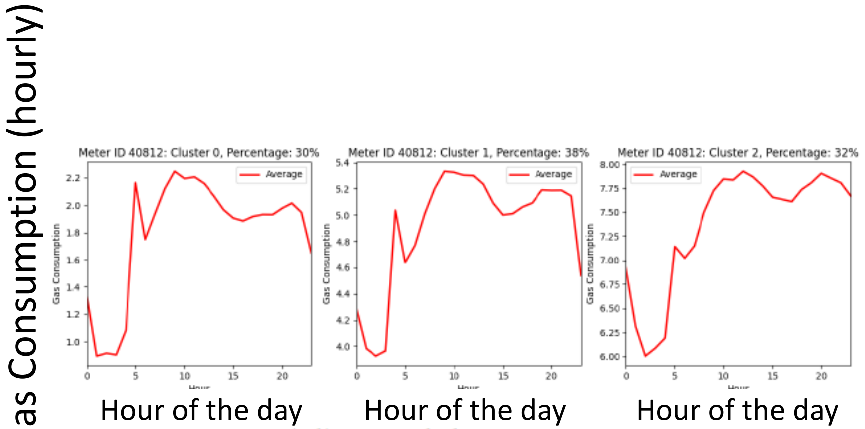
(a) Building 4501



(b) Building 4531

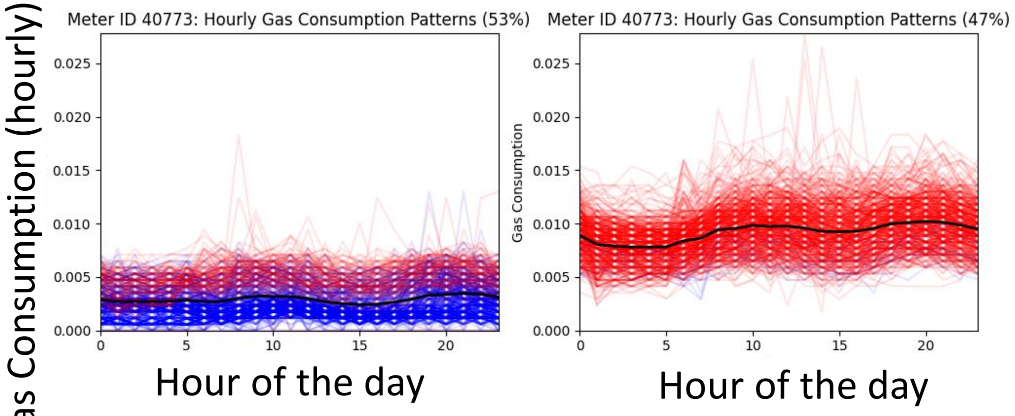


(c) Building 4520

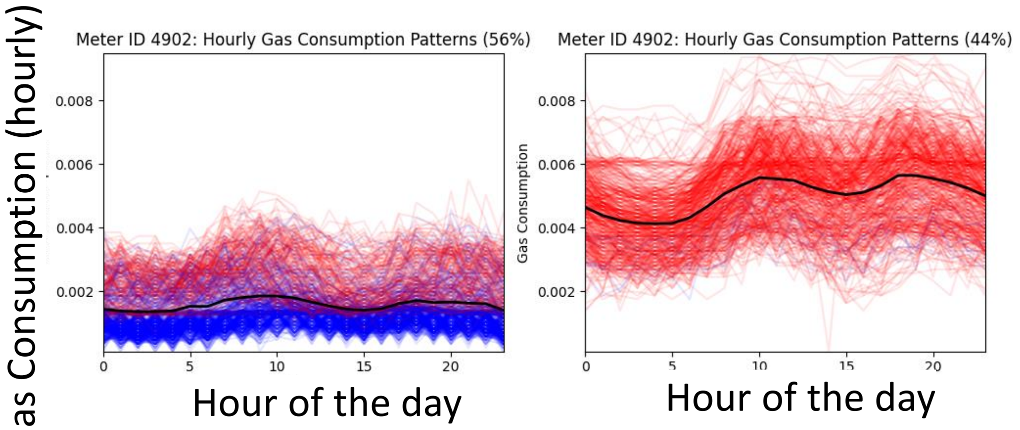


(d) Building 4511

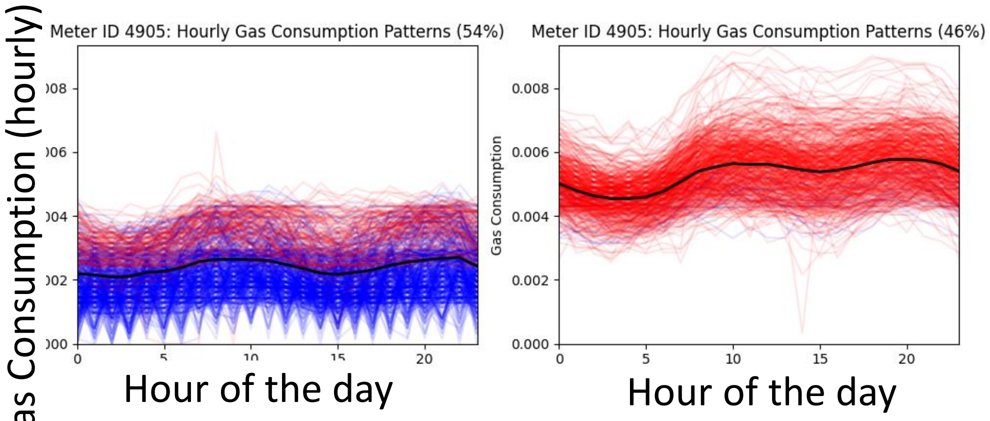
Typical daily profiles of 4 building



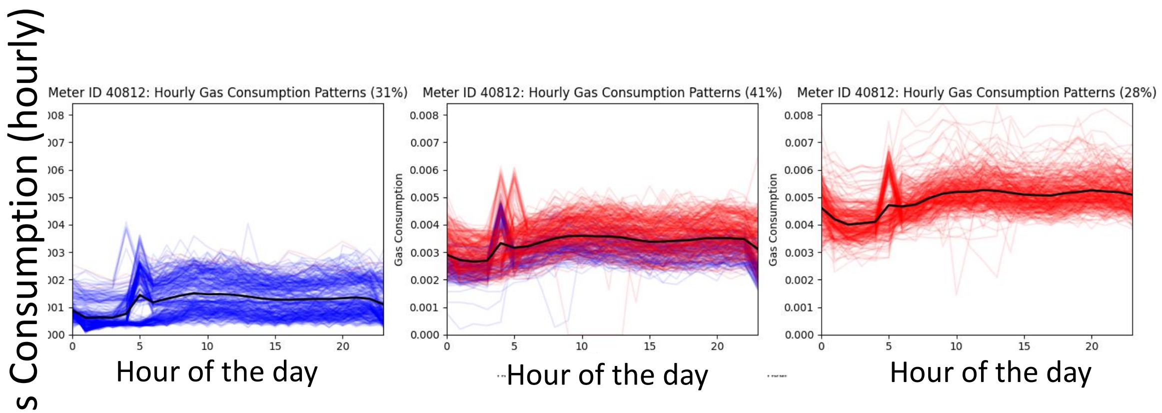
(a) Building 4501



(b) Building 4531



(c) Building 4520



(d) Building 4511

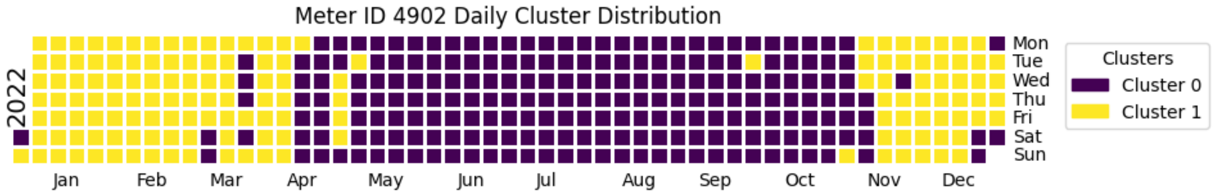
Typical Daily Usage Profiles of Building 4531

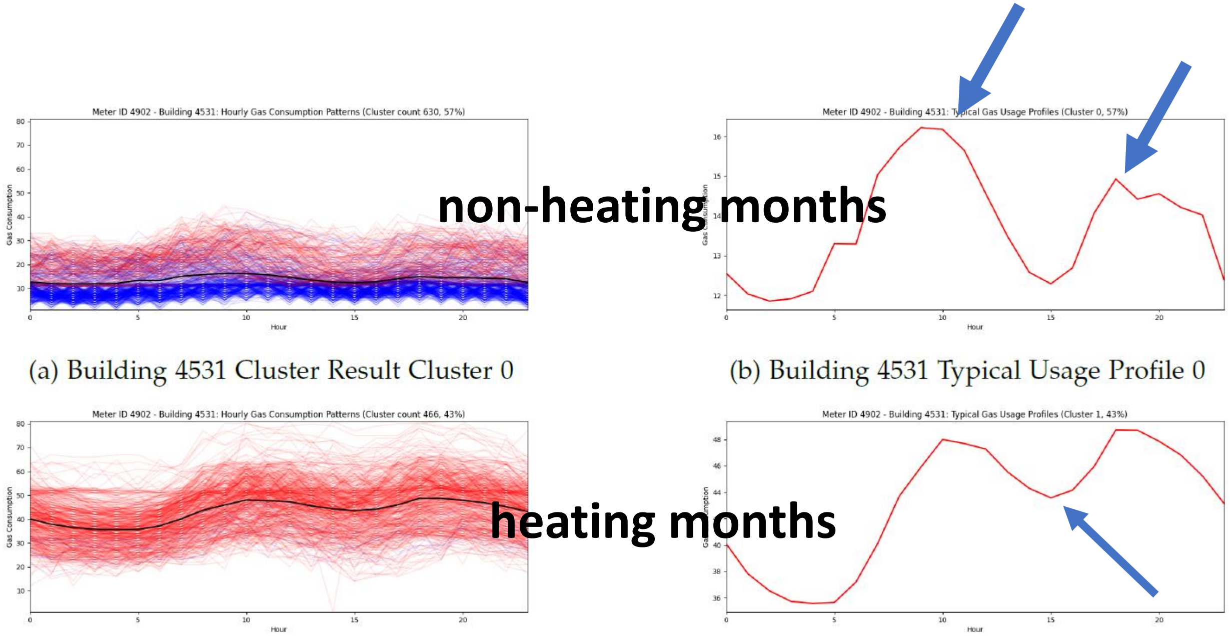
- High rise
- Group house
- Individual CHP

Typical Daily Usage Profiles of Building 4531

- The longer daylight hours, leading to a wider interval between peaks in **non-heating months**.
- **Heating months**, consumption relative higher at 3pm.

Even distribution





(a) Building 4531 Cluster Result Cluster 0

(b) Building 4531 Typical Usage Profile 0

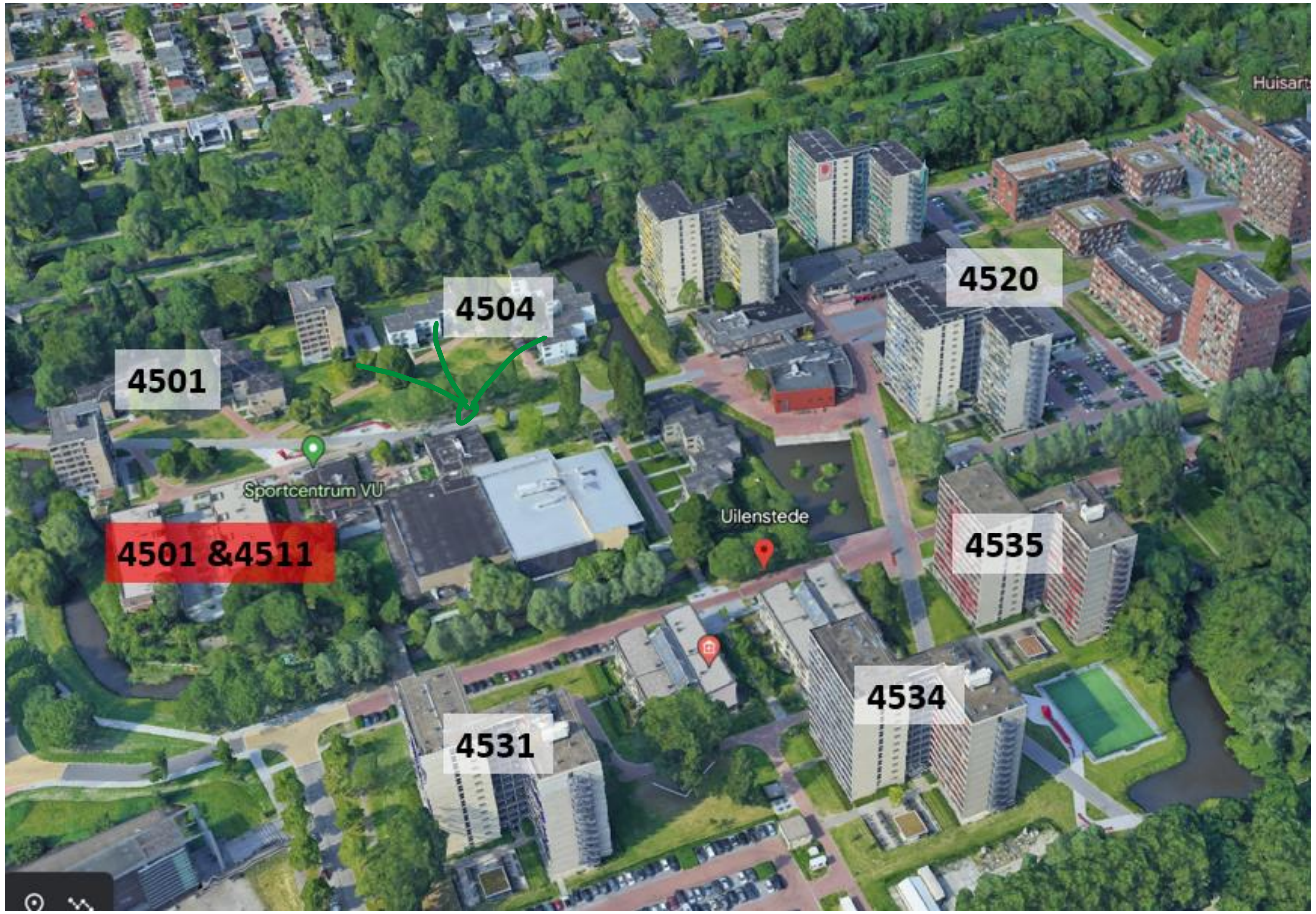
(c) Building 4531 Cluster Result Cluster 1

(d) Building 4531 Typical Usage Profile 1

Figure 4.15.: Visualisation of Cluster Results for Building 4531

Typical Daily Usage Profiles of Building 4511

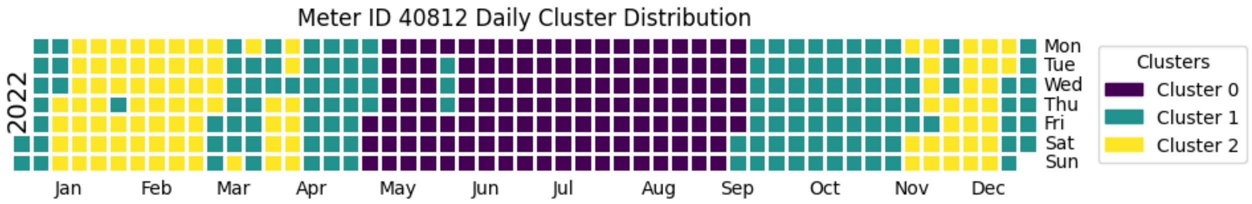
- Low rise
- Apartments
- Shared central heating system

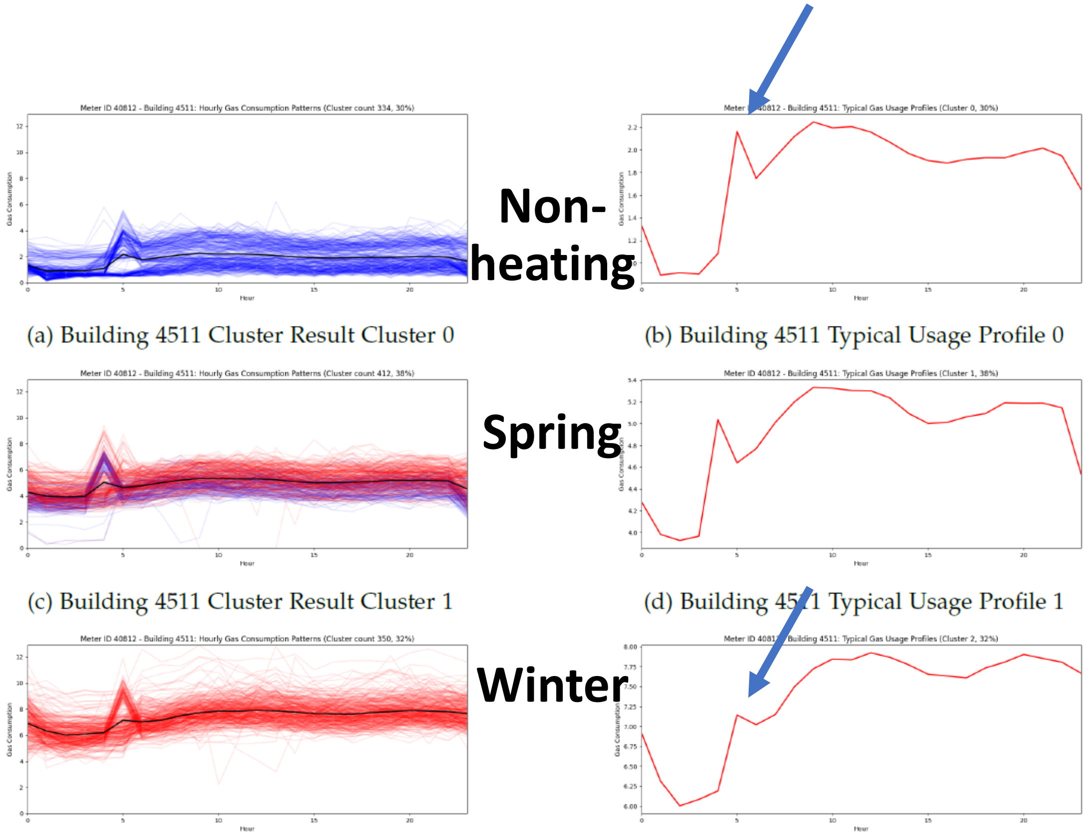


Typical Daily Usage Profiles of Building 4511

- Water heating scheduling at 6 am -> **heating up hot water storage** to set temperature
- Winter cluster has a less pronounced spike due to overall high gas consumption during heating season

Different seasons





(a) Building 4511 Cluster Result Cluster 0

(b) Building 4511 Typical Usage Profile 0

(c) Building 4511 Cluster Result Cluster 1

(d) Building 4511 Typical Usage Profile 1

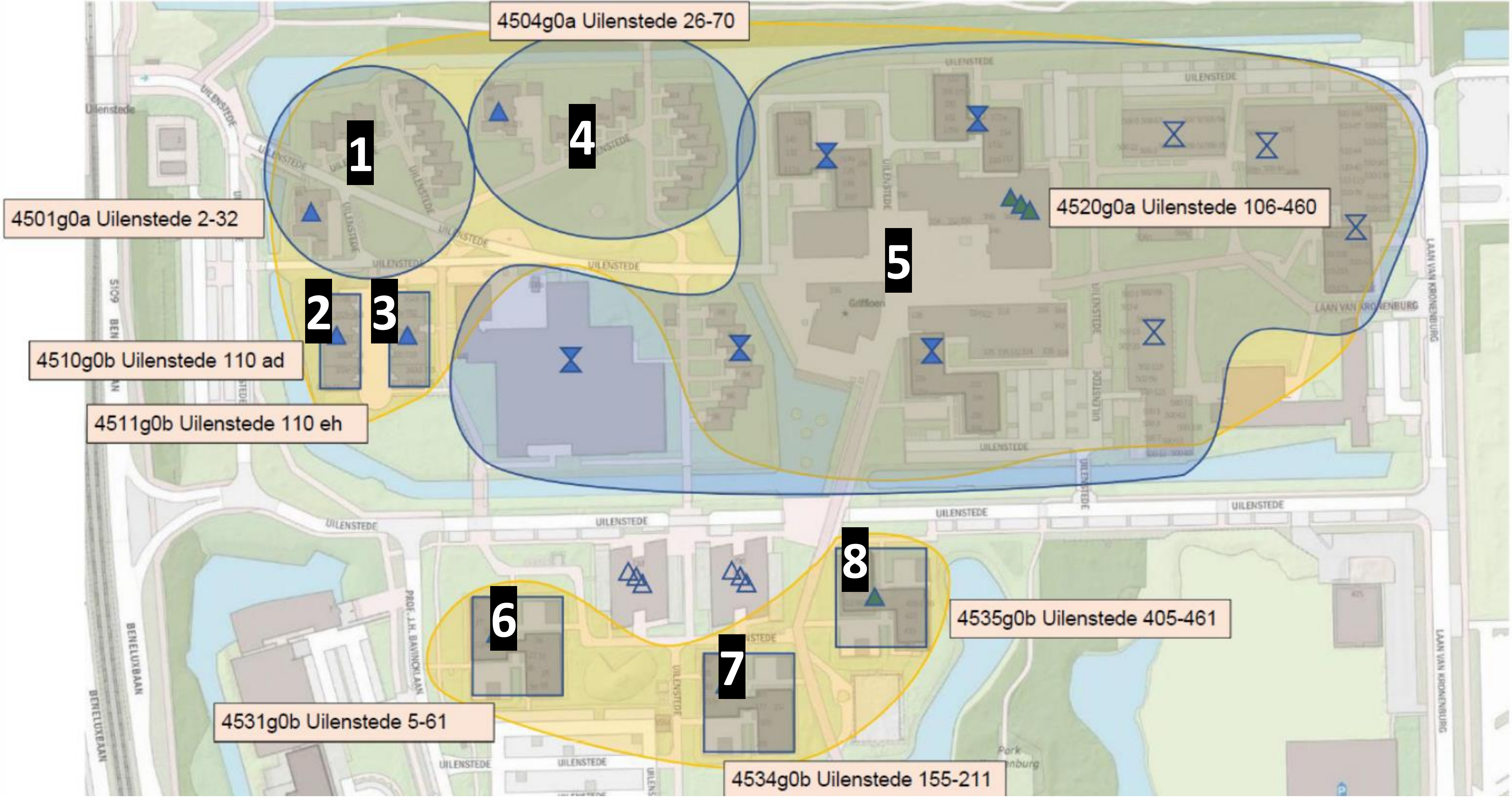
(e) Building 4511 Cluster Result Cluster 2

(f) Building 4511 Typical Usage Profile 2

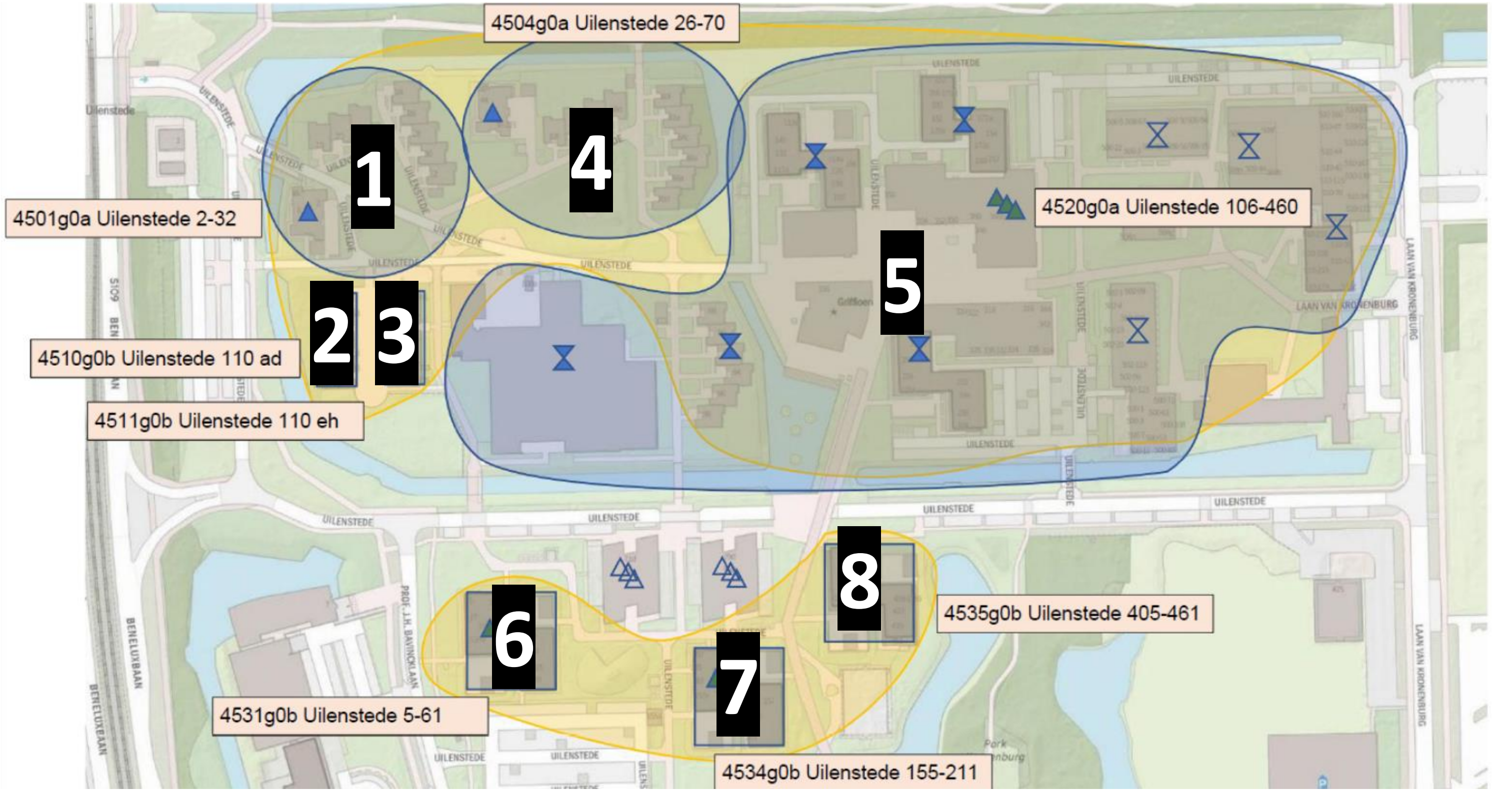
Figure 4.17.: Visualisation of Cluster Results for Building 4511

Clustering Result for all building

Similar patterns across all buildings **globally**


4504g0a Uilenstede 26-70
4501g0a Uilenstede 2-32
4520g0a Uilenstede 106-460
4510g0b Uilenstede 110 ad
4511g0b Uilenstede 110 eh
4535g0b Uilenstede 405-461
4531g0b Uilenstede 5-61
4534g0b Uilenstede 155-211
1
2
3
4
5
6
7
8

8 meters



Set the cluster num to **8 groups**

Objective : find **different patterns** for each meter

Global clustering result across all meters

Heating months more variant in terms of consumption

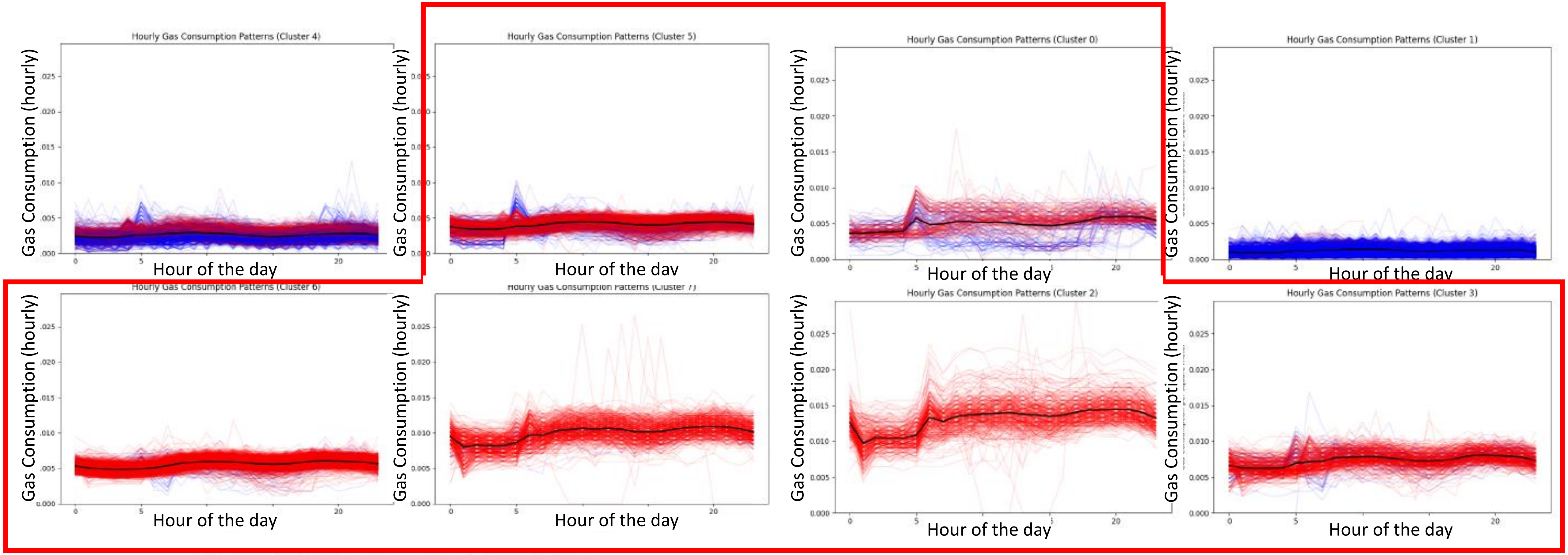
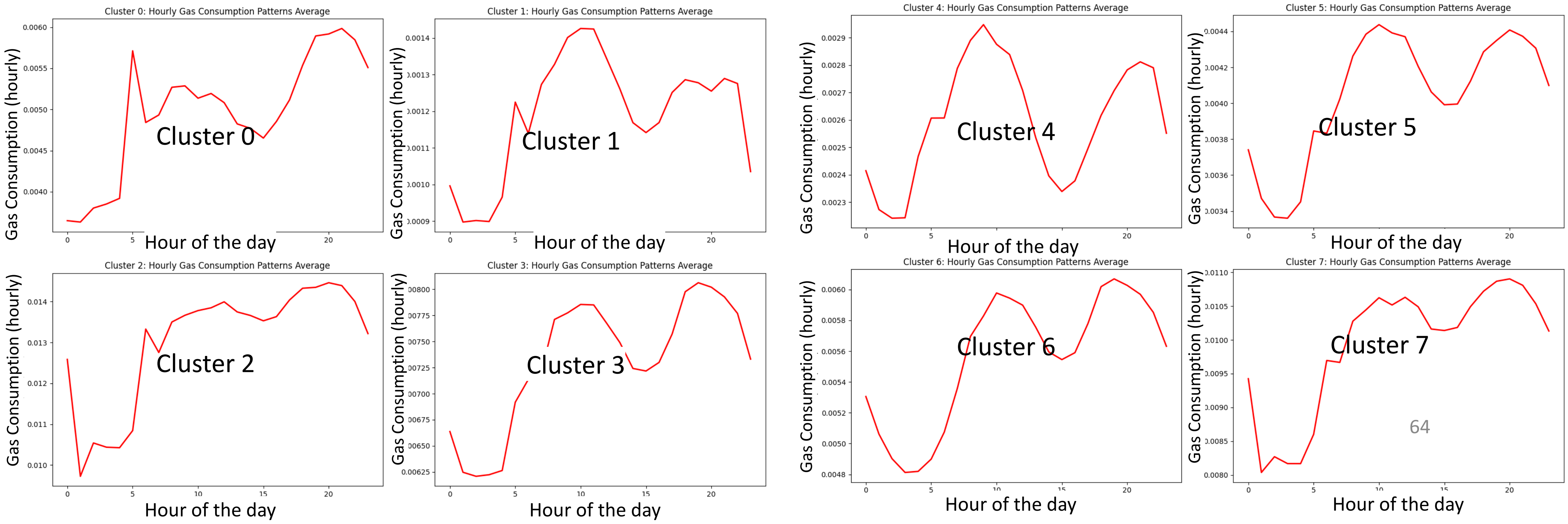


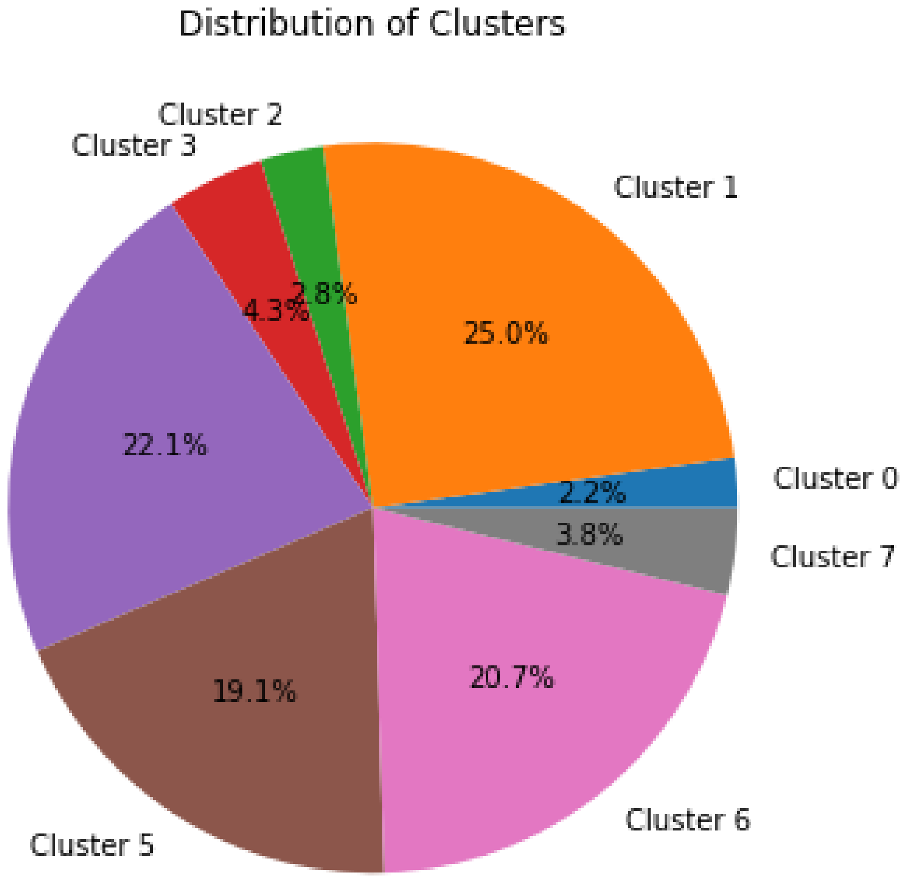
Figure 4.21.: Cluster result of all buildings

Global clustering result across all meters

Similar to each other



Amount of data in each cluster



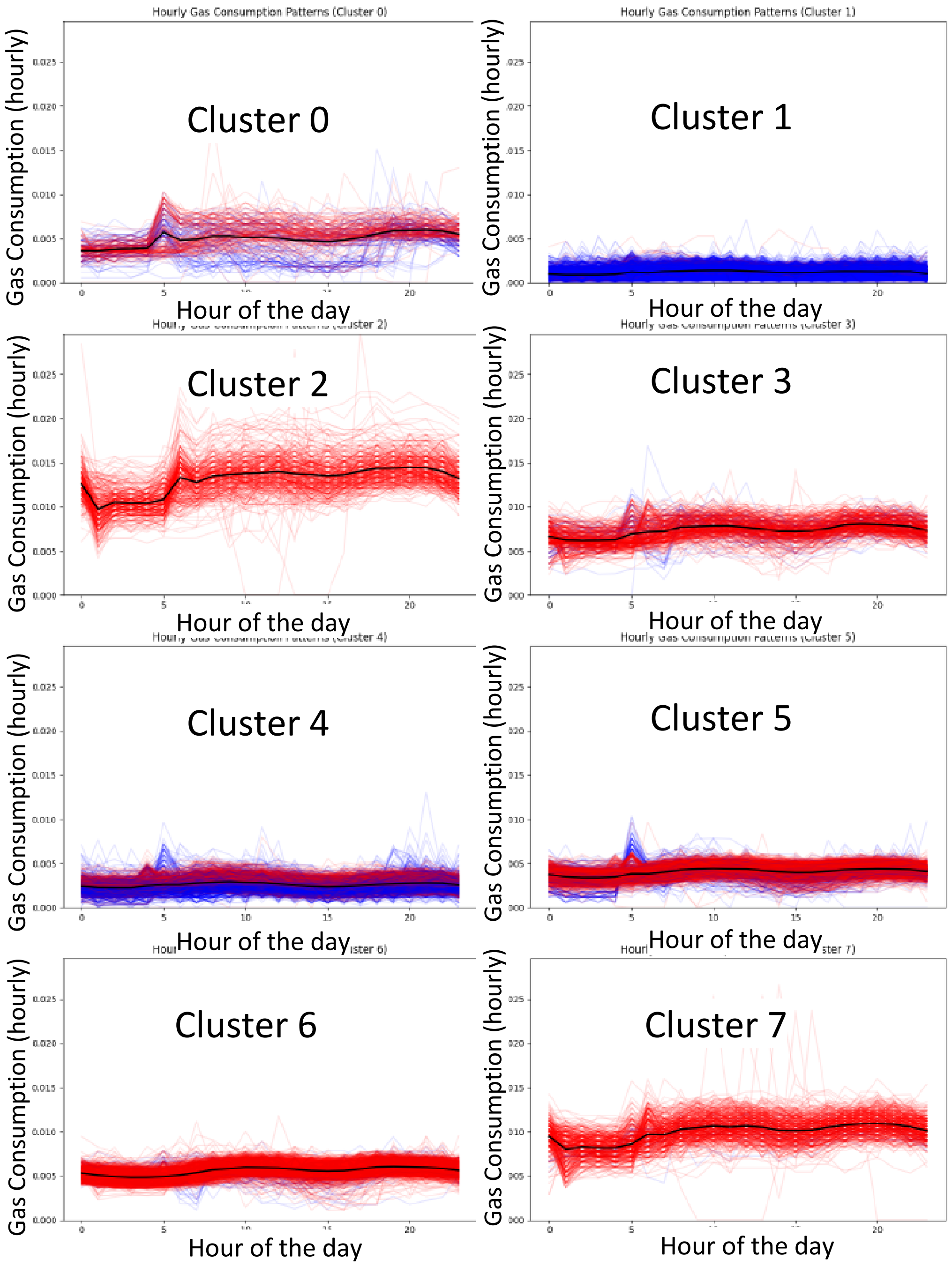
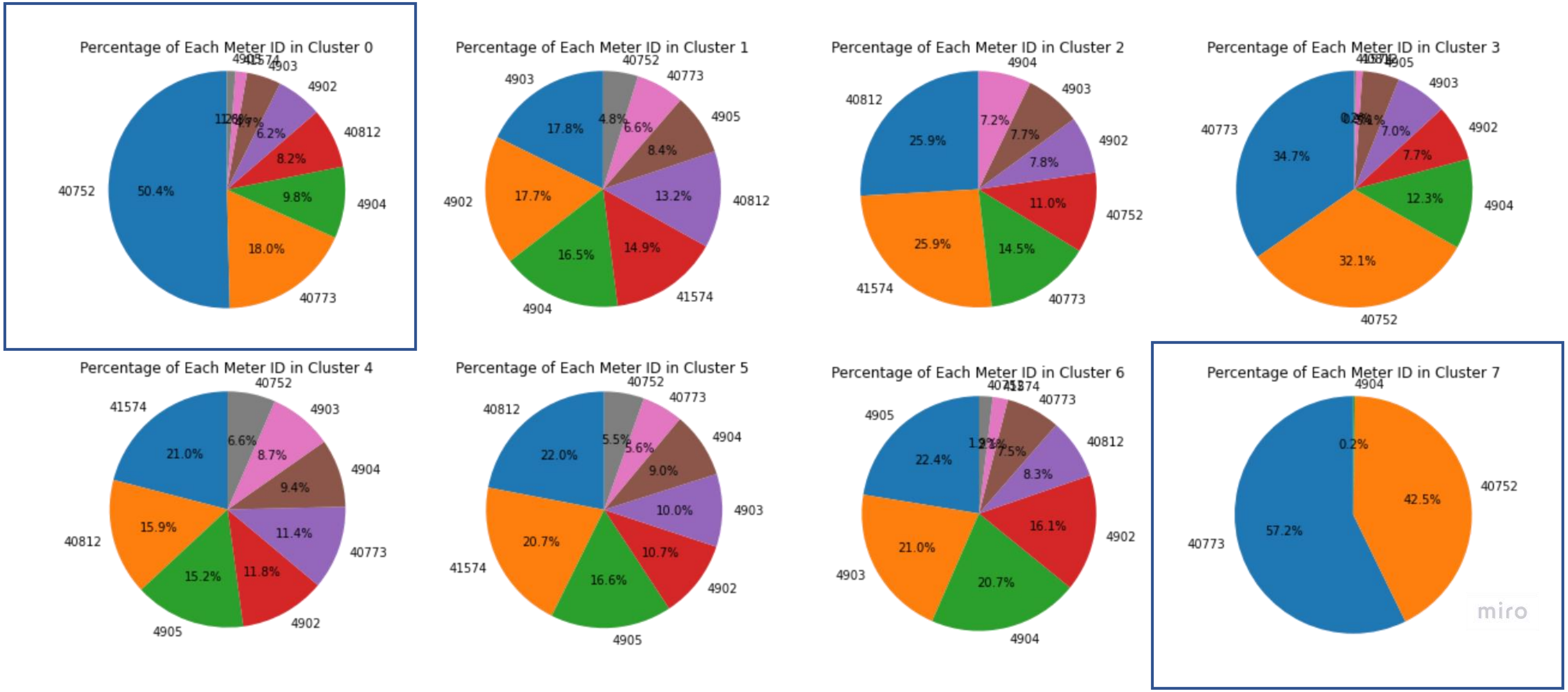


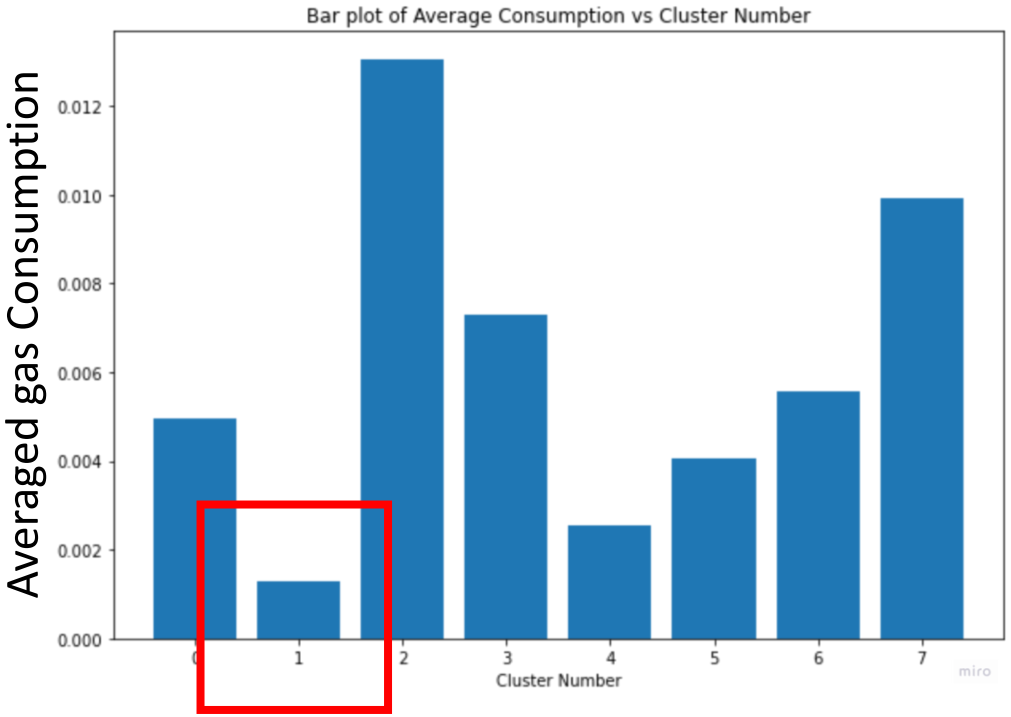
Figure 4.21.: Cluster result of all buildings

Meter distribution in each group

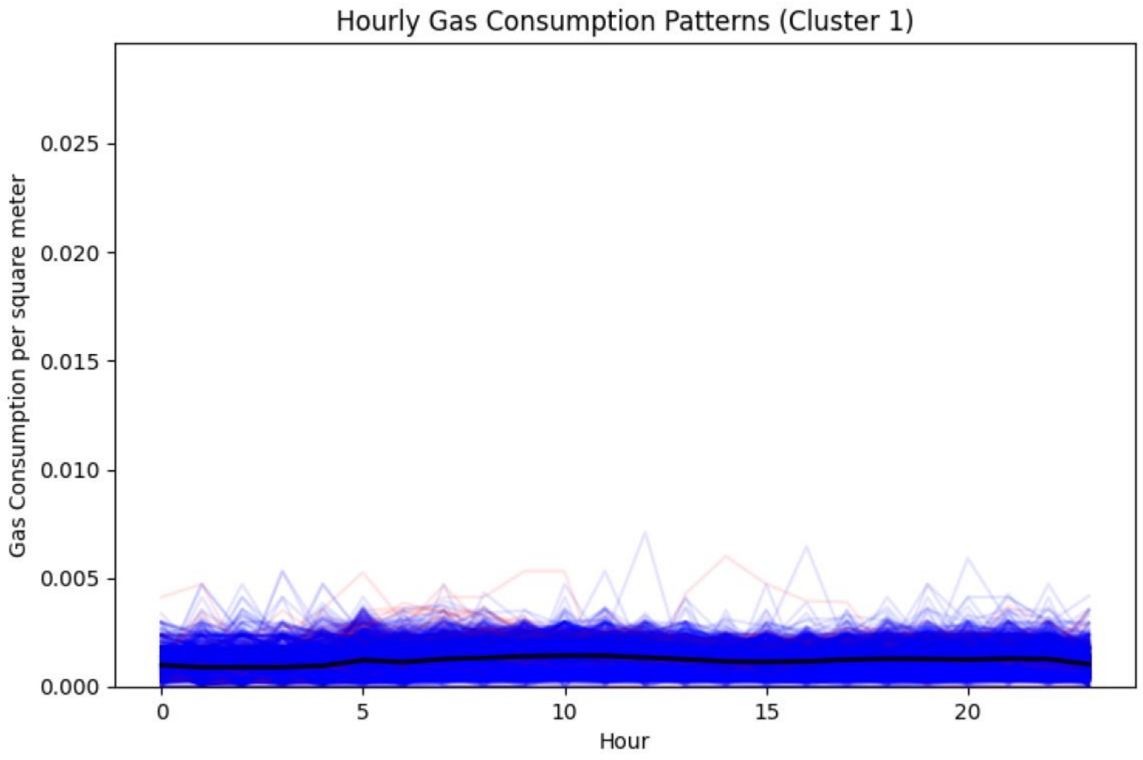


Gas consumption of each cluster

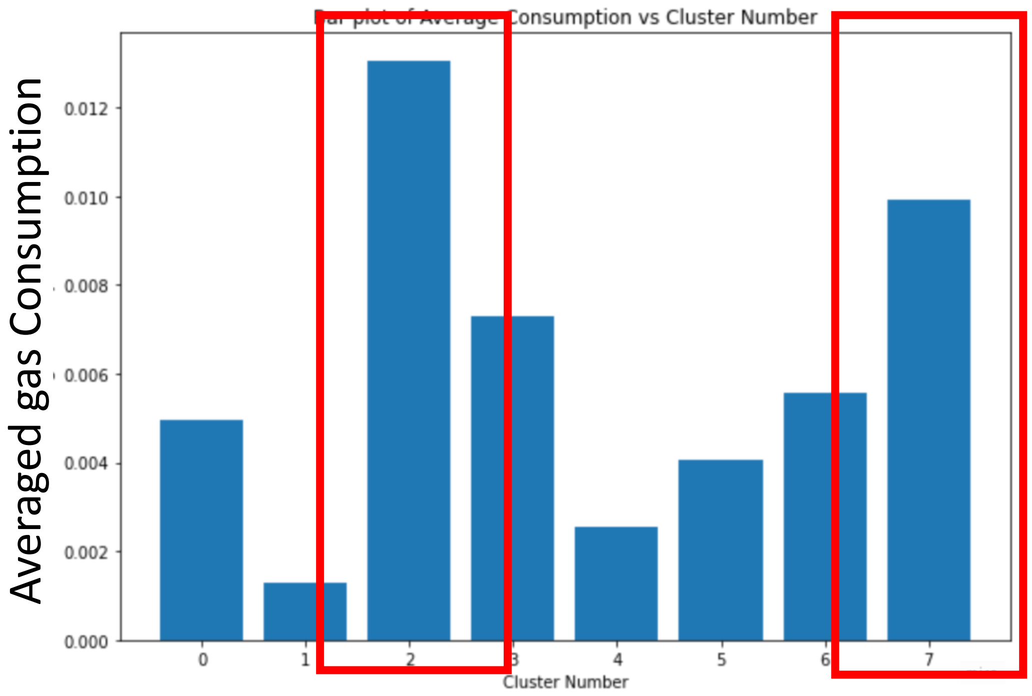
Cluster 1 has low consumption due to it’s composed of non-heating month



Here show the percentage of winter summer



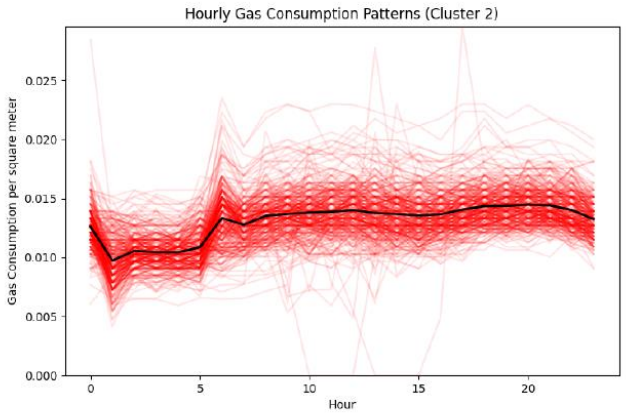
Gas consumption of each cluster

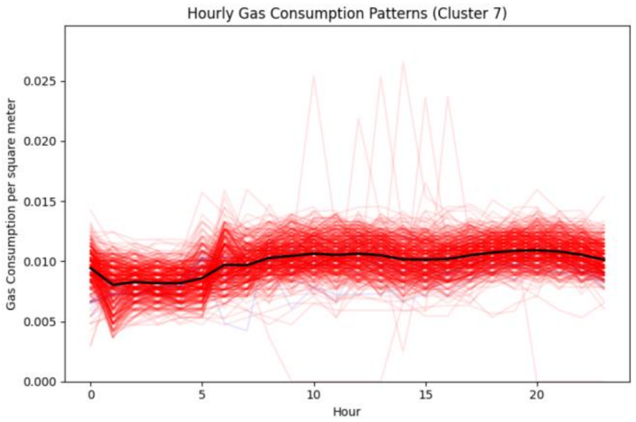


Cluster 2 and Cluster 7:

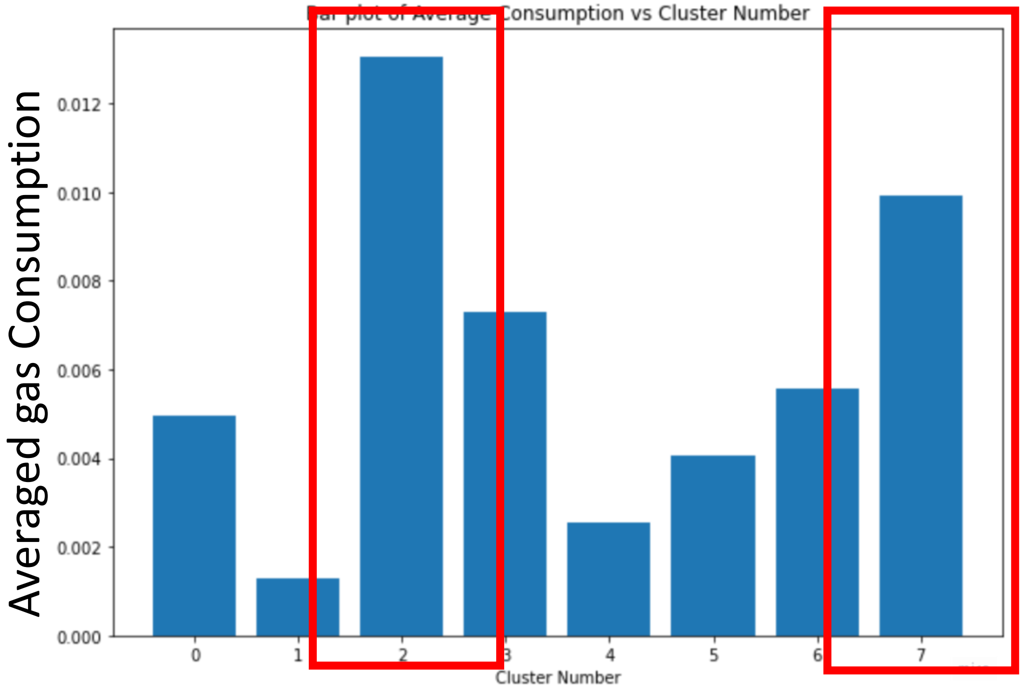
(both built in 1991)

lower insulation values



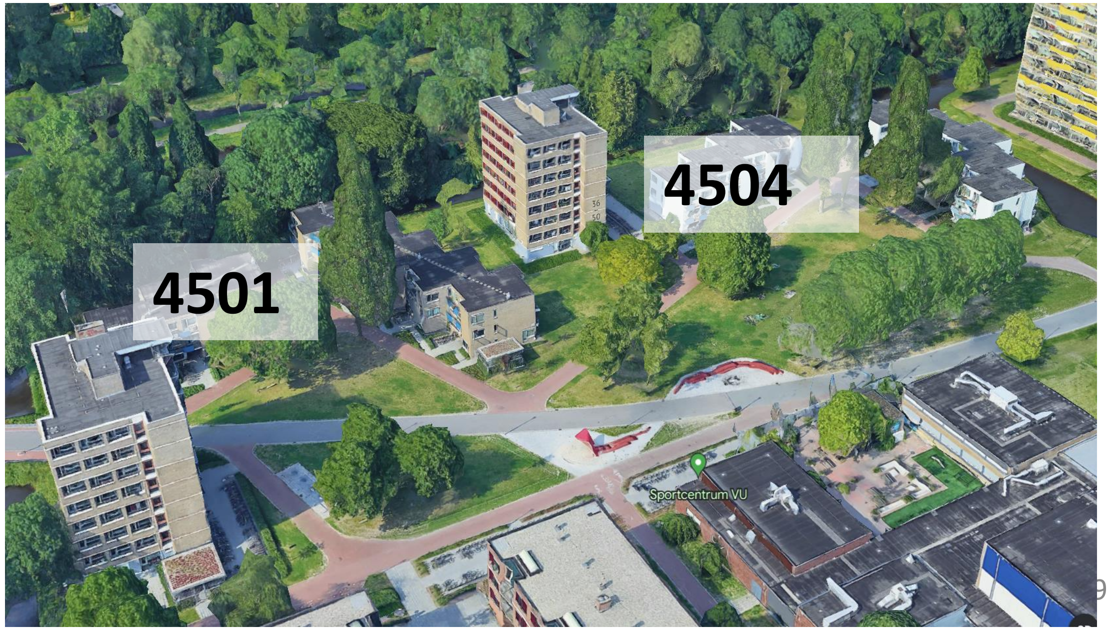


Clusters consumption

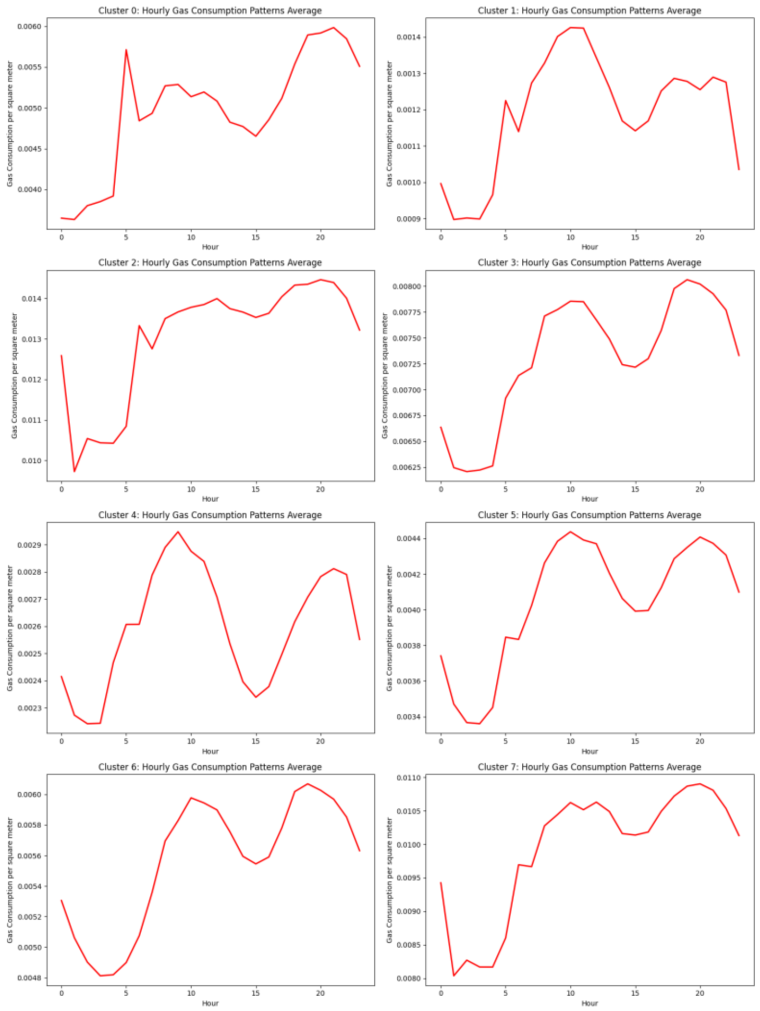


Cluster 2 and Cluster 7:
(both built in 1991)
having lower insulation values

- Low rise
- Apartments
- Shared central heating system



8 profiles are all similar



Too many number of groups

Can we reduce?

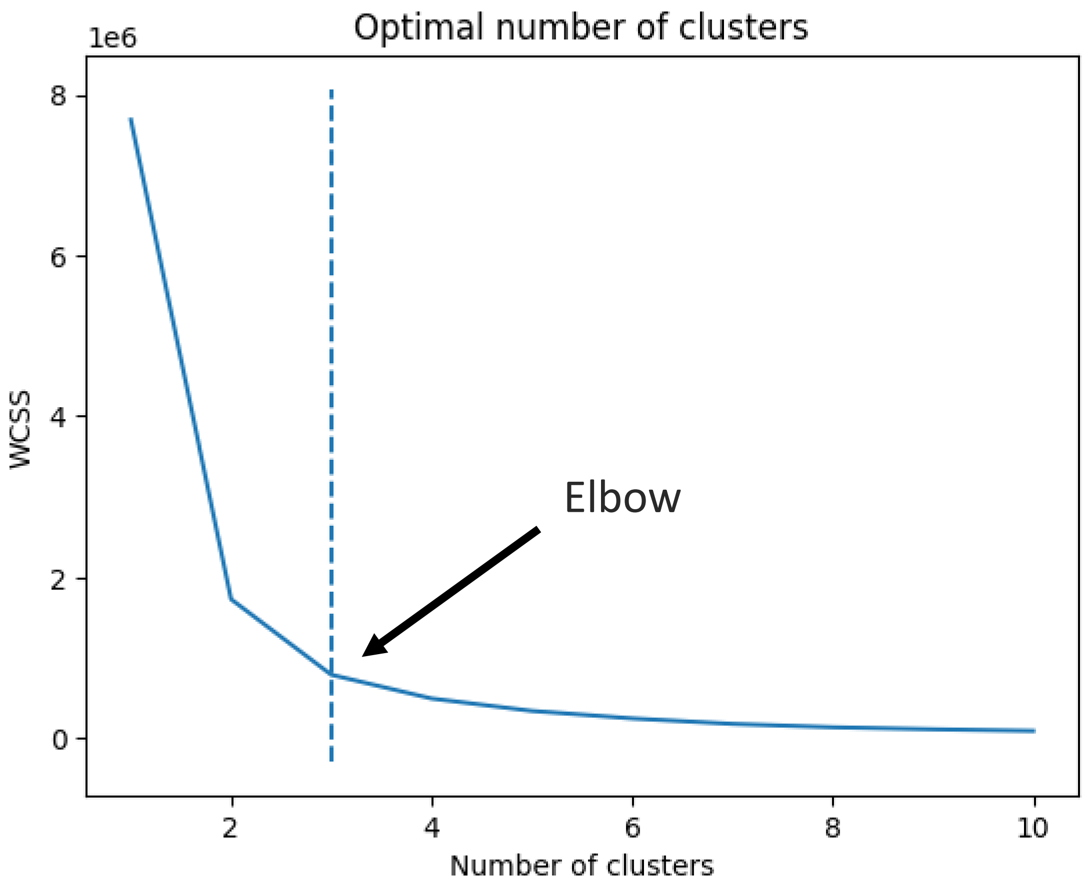
What is the optimal numbers of groups?

Elbow Method

- Within-Cluster Sum of Square

$$WCSS = \sum_i (x_i - c_j)^2$$

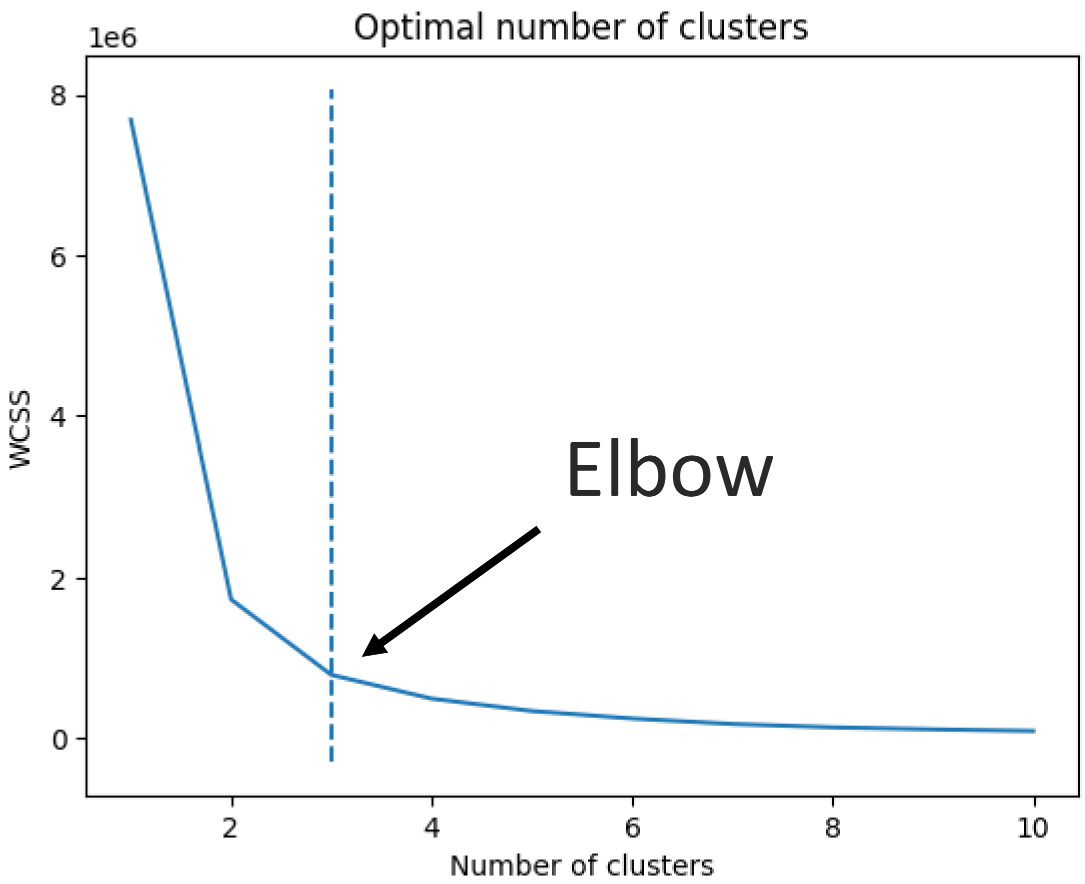
Minimize sum of the squared distances



xi data point,
cj centroid of the cluster to which xi is assigned.

Elbow Method

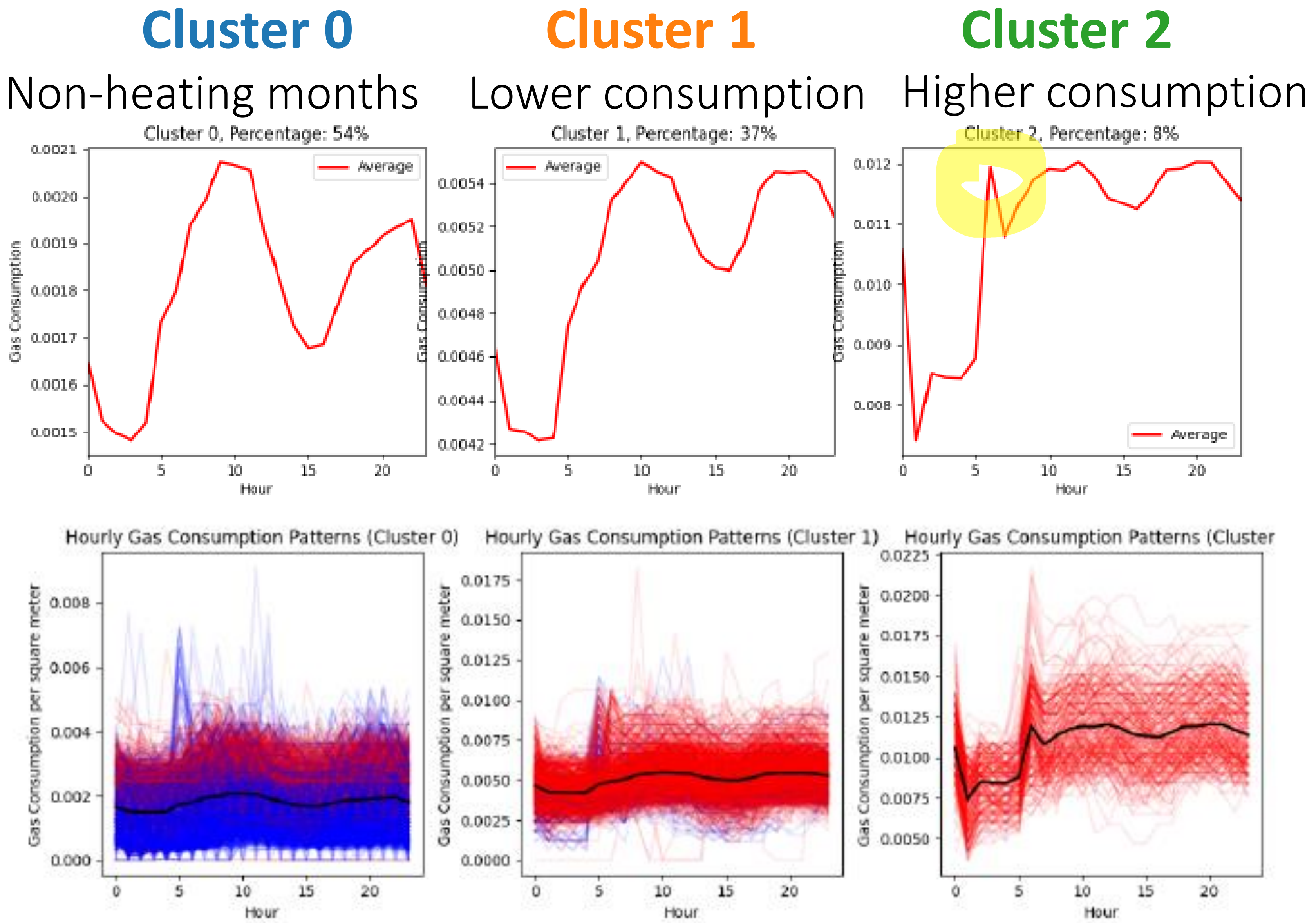
- Within-Cluster Sum of Square



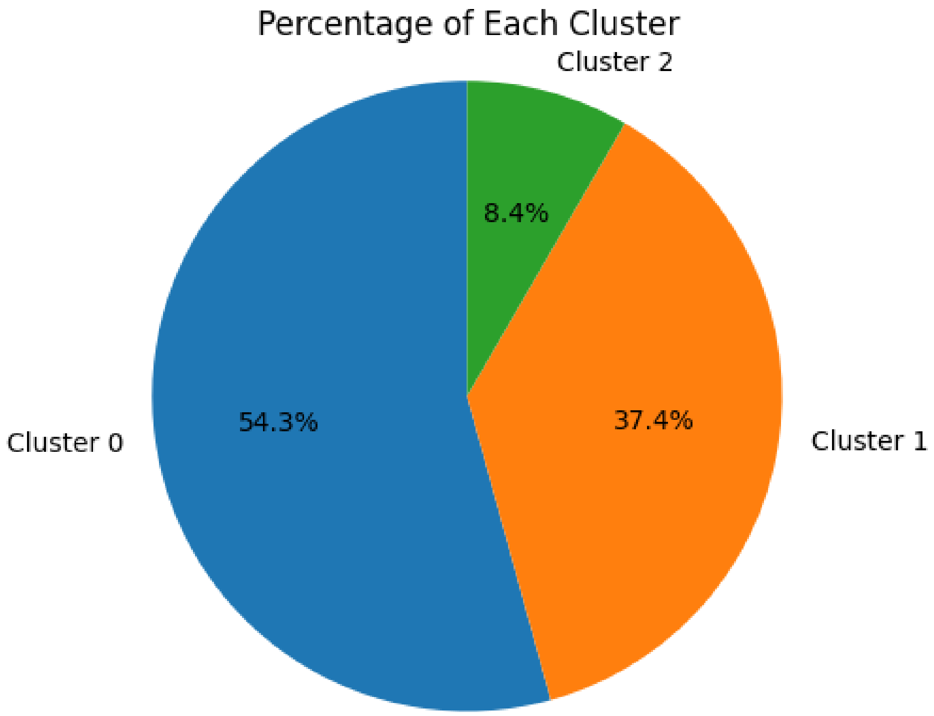
3 clusters

Divide all metered data into 3 clusters

Global clustering result



Amount of data in each cluster



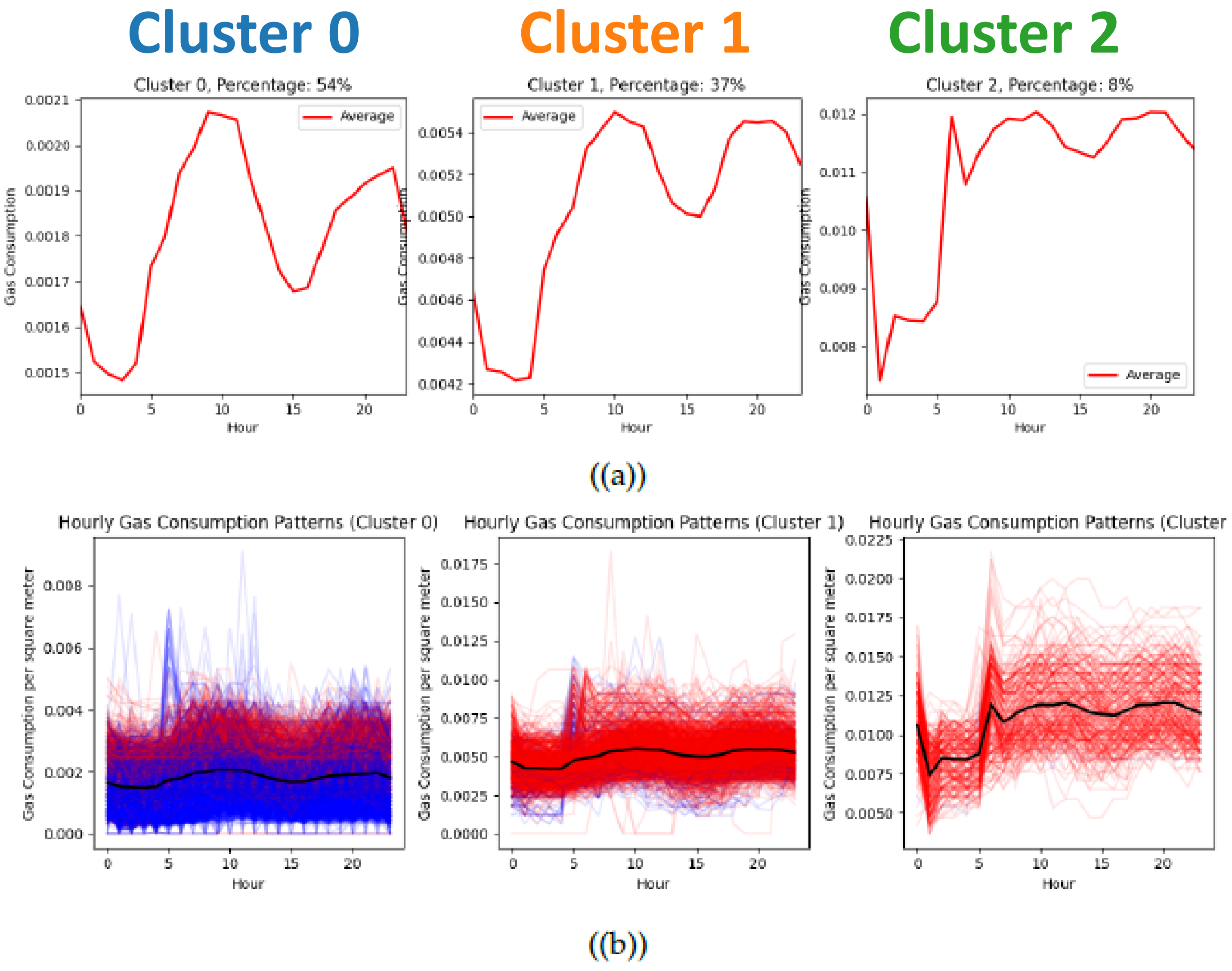
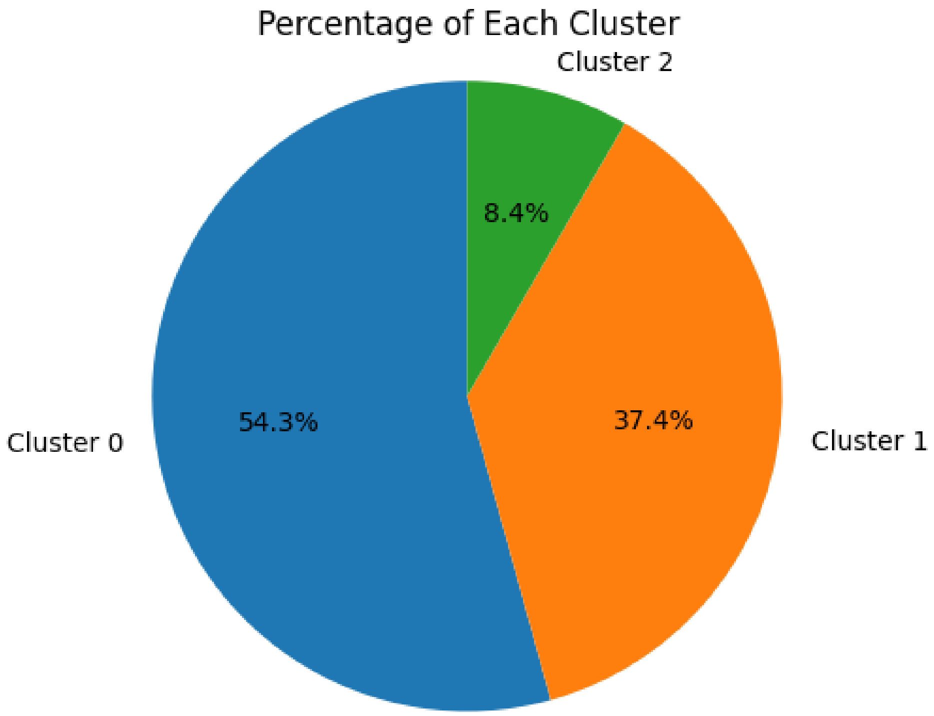


Figure 4.22.: Result of the Inter-Building Clustering using Elbow Method

Cluster distribution



Cluster 0 Cluster 1 Cluster 2

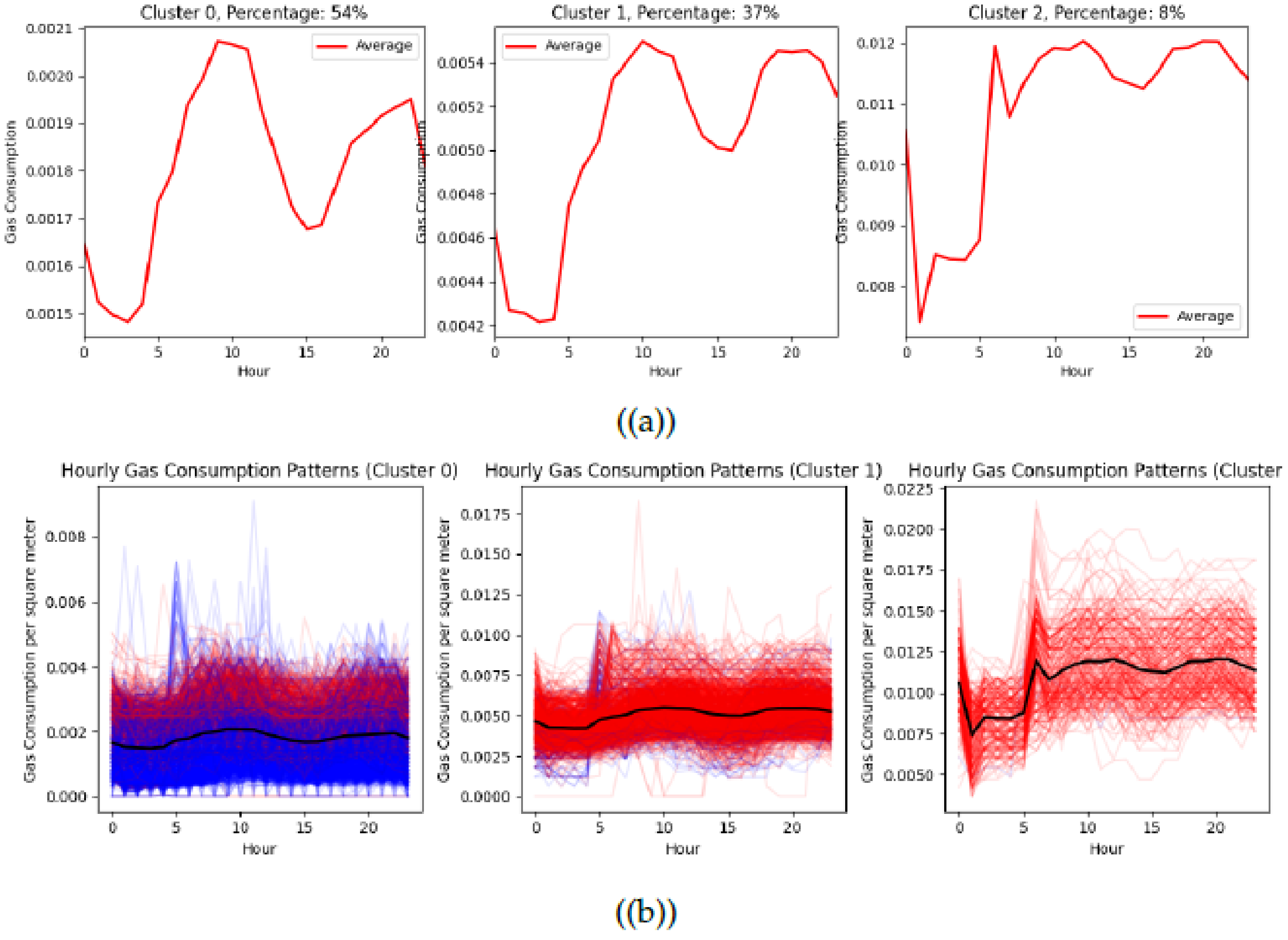


Figure 4.22.: Result of the Inter-Building Clustering using Elbow Method

Average gas consumption per floor area

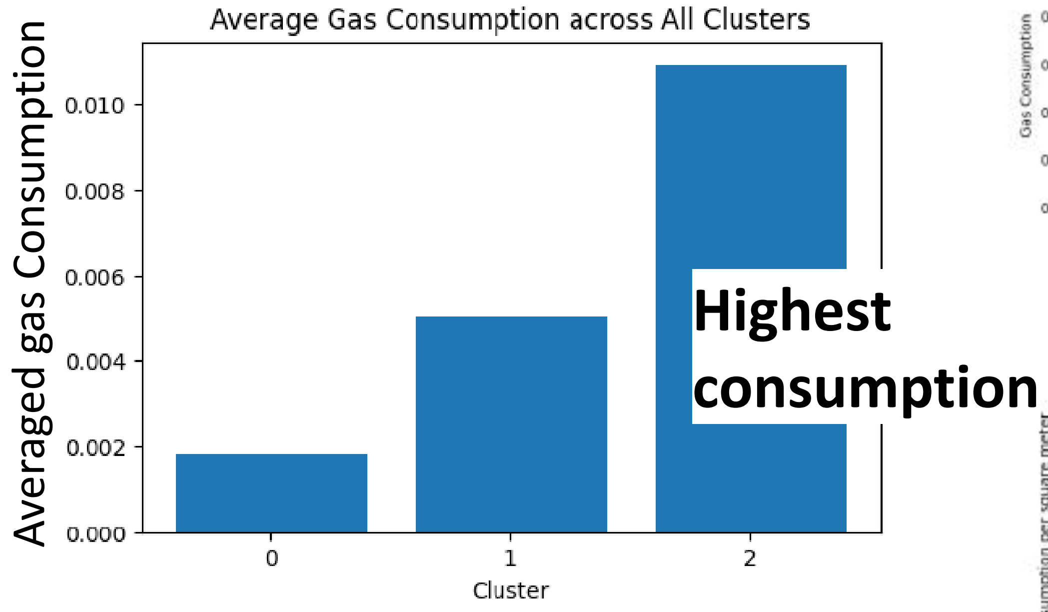


Figure 4.23.: Gas consumption per floor area Inter-Building Clusters

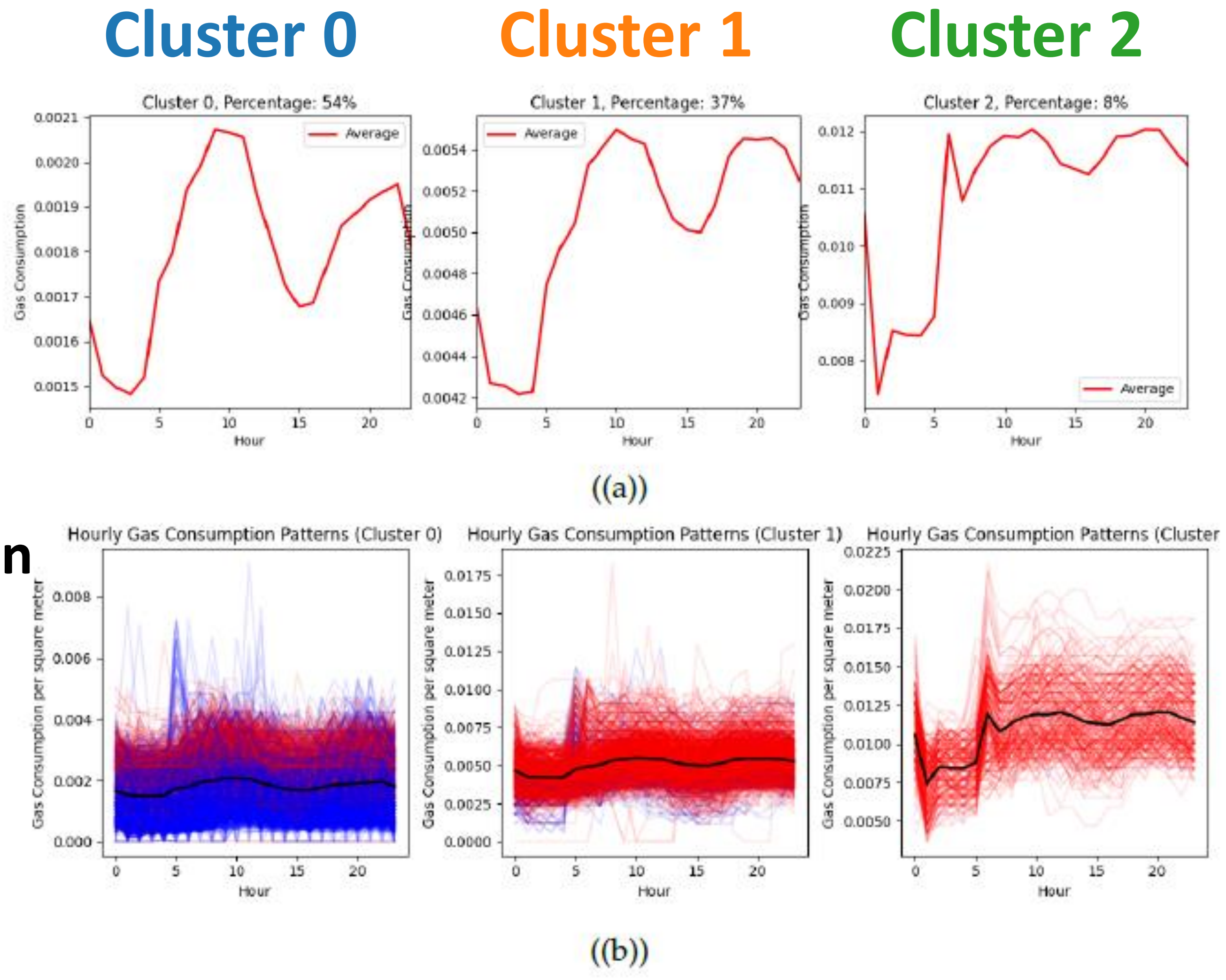
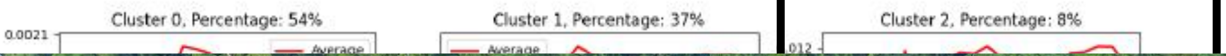


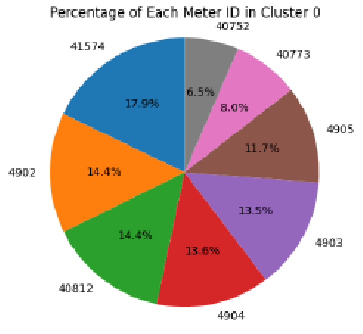
Figure 4.22.: Result of the Inter-Building Clustering using Elbow Method

Meter distribution in each group

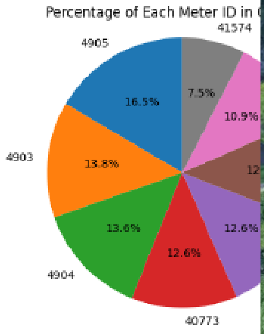
Cluster 0 Cluster 1 Cluster 2



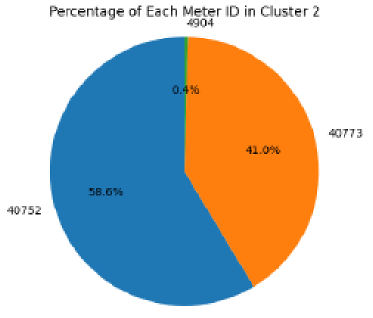
G.0.1. The percentage of meter ID in each cluster



((a)) Percentage of meter id in cluster 0

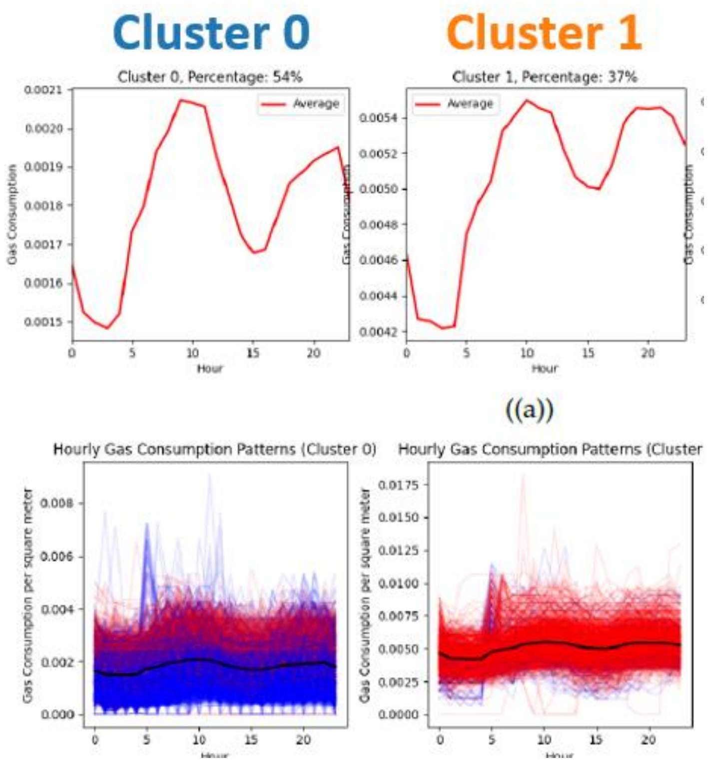


((b)) Percentage of meter id



((c)) Percentage of meter id in cluster 2

Figure 4.22.: Result of the Inter-Building Clustering using Elbow Method

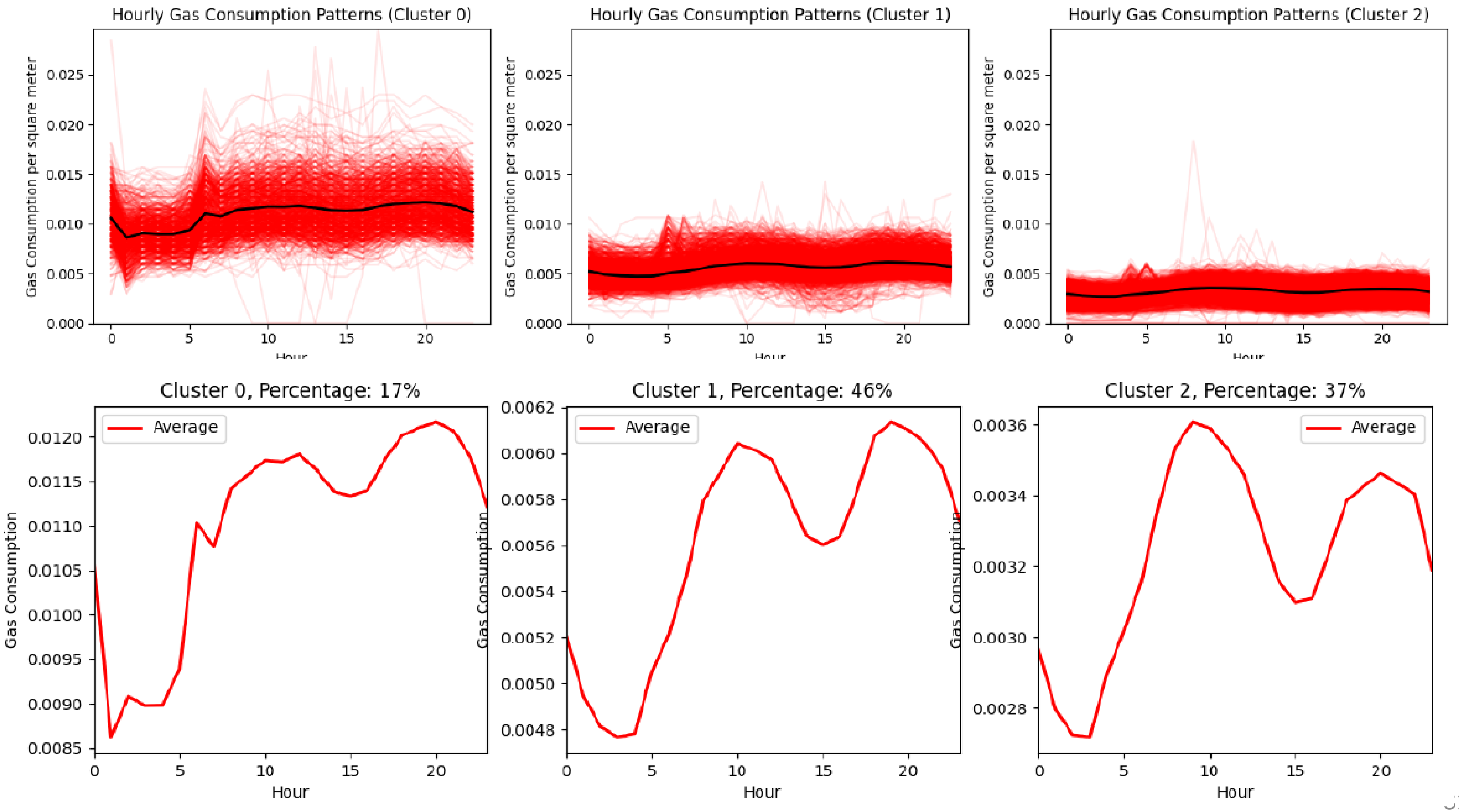


Only heating months

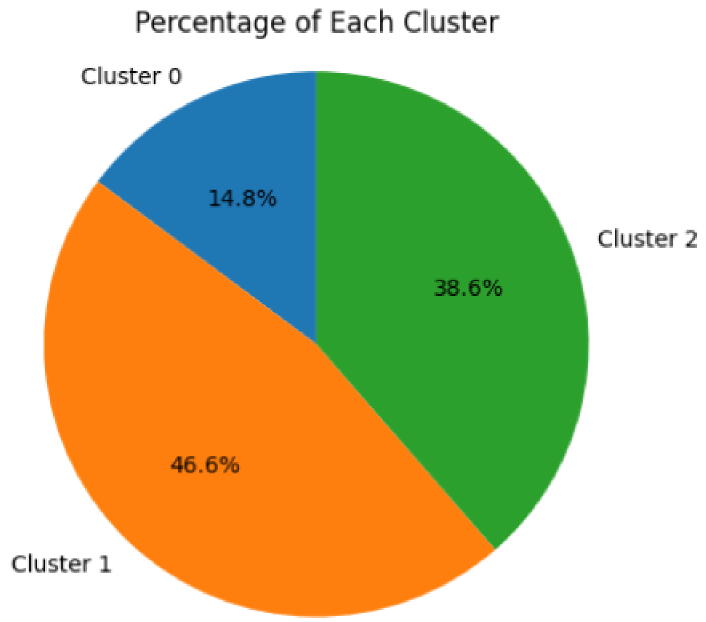
Heating months clustering results

Cluster 0 **Cluster 1** **Cluster 2**

3 clusters

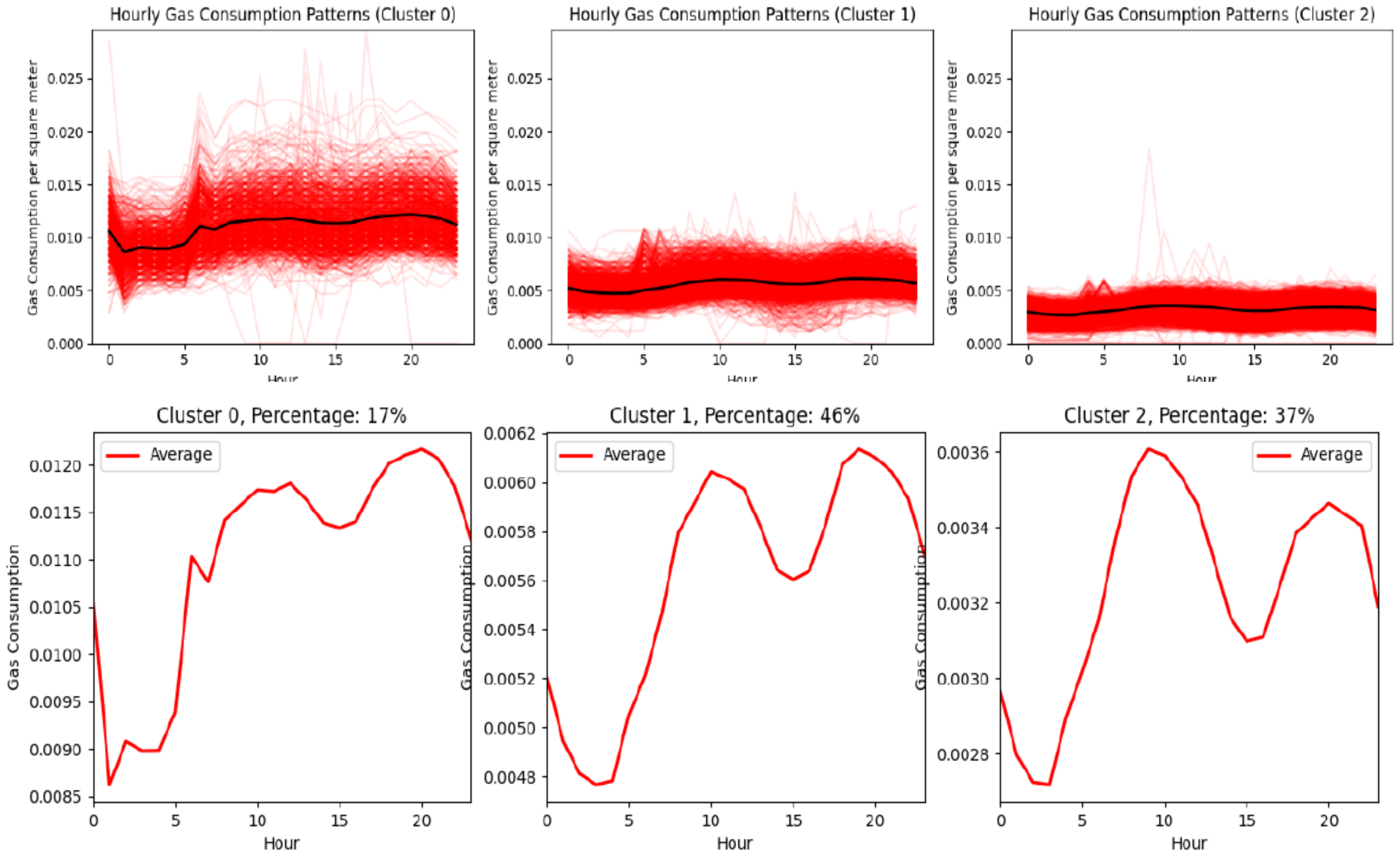


Amount of data in each cluster

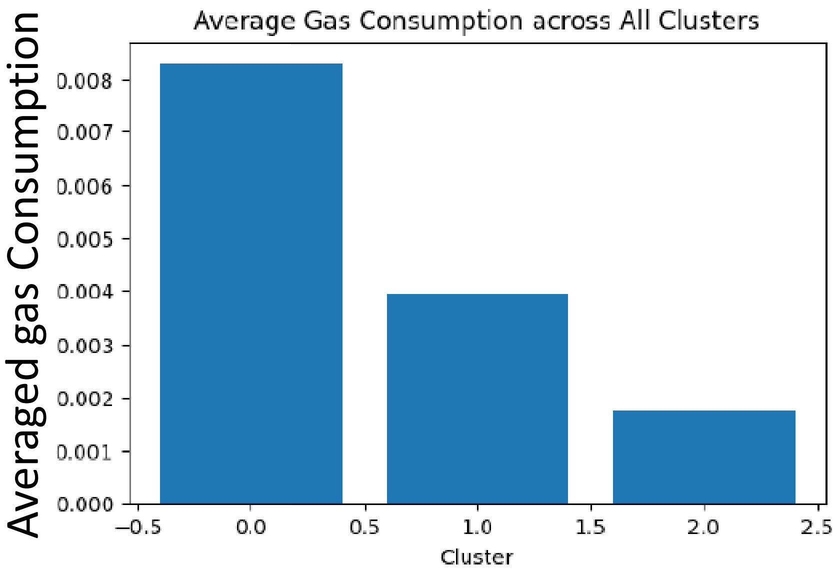


(a) Percentage of all clusters

Cluster 0 **Cluster 1** **Cluster 2**

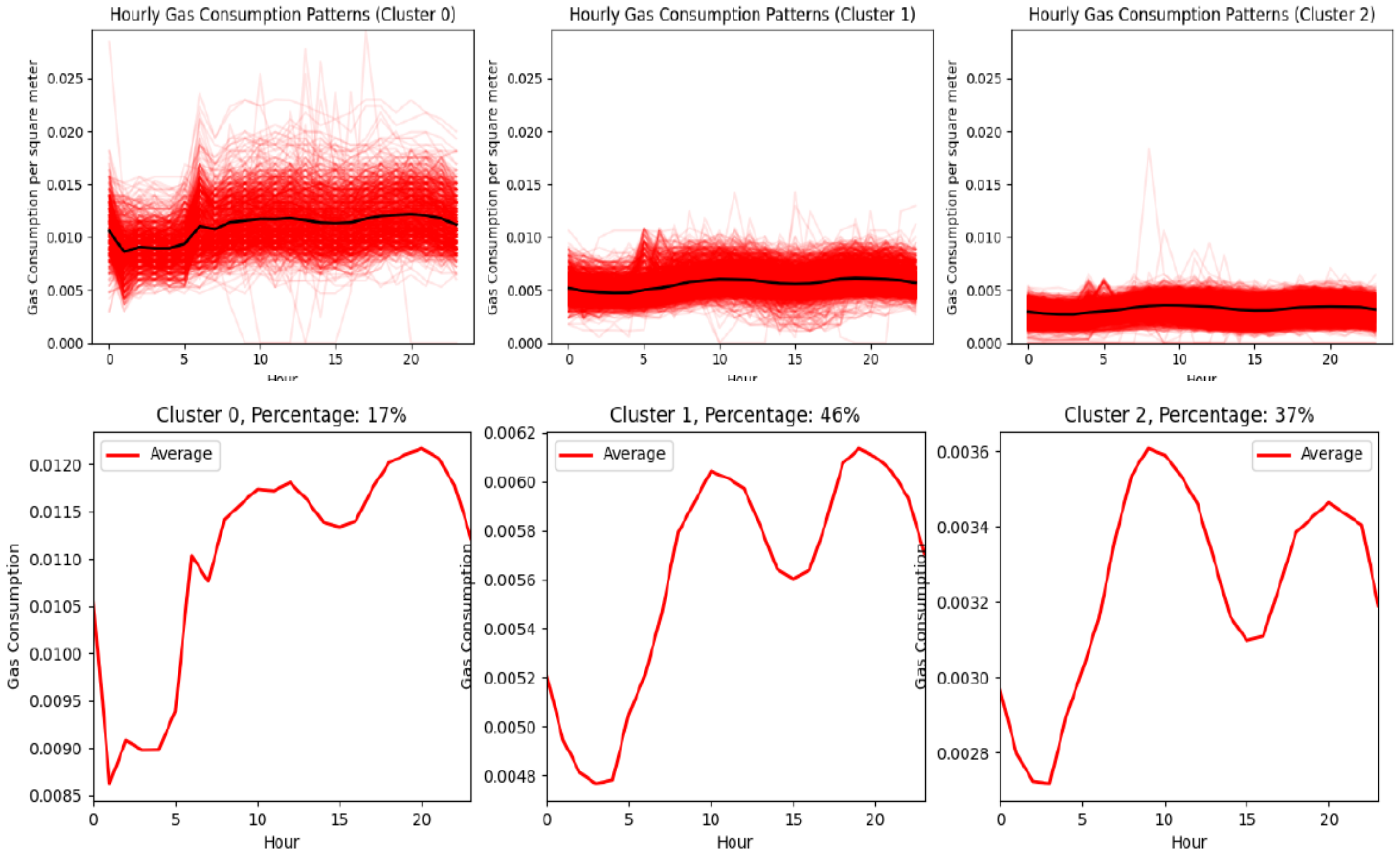


Average gas consumption



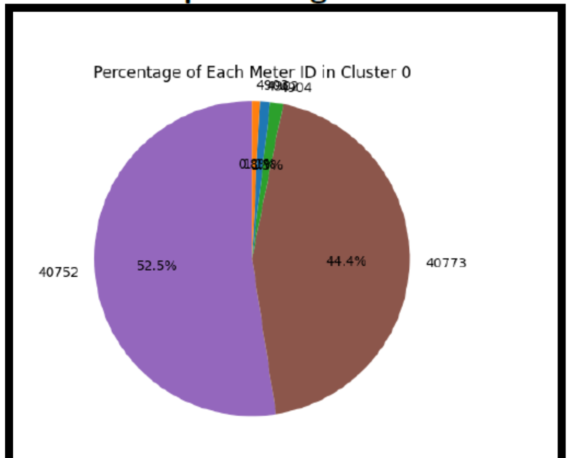
(b) Gas consumption per floor area Inter-Building Clusters

Cluster 0 **Cluster 1** **Cluster 2**

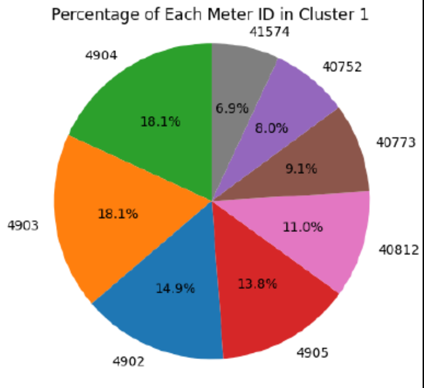


Meter distribution in each group

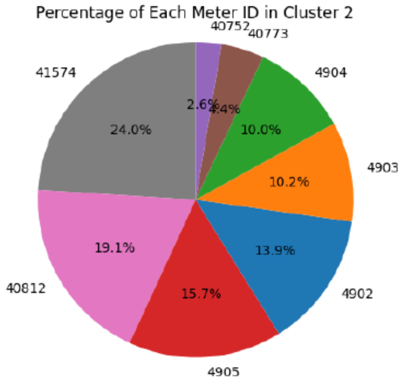
H.0.2. The percentage of meter ID in each cluster



(a) Percentage of meter id in cluster 0



(b) Percentage of meter id in cluster 1

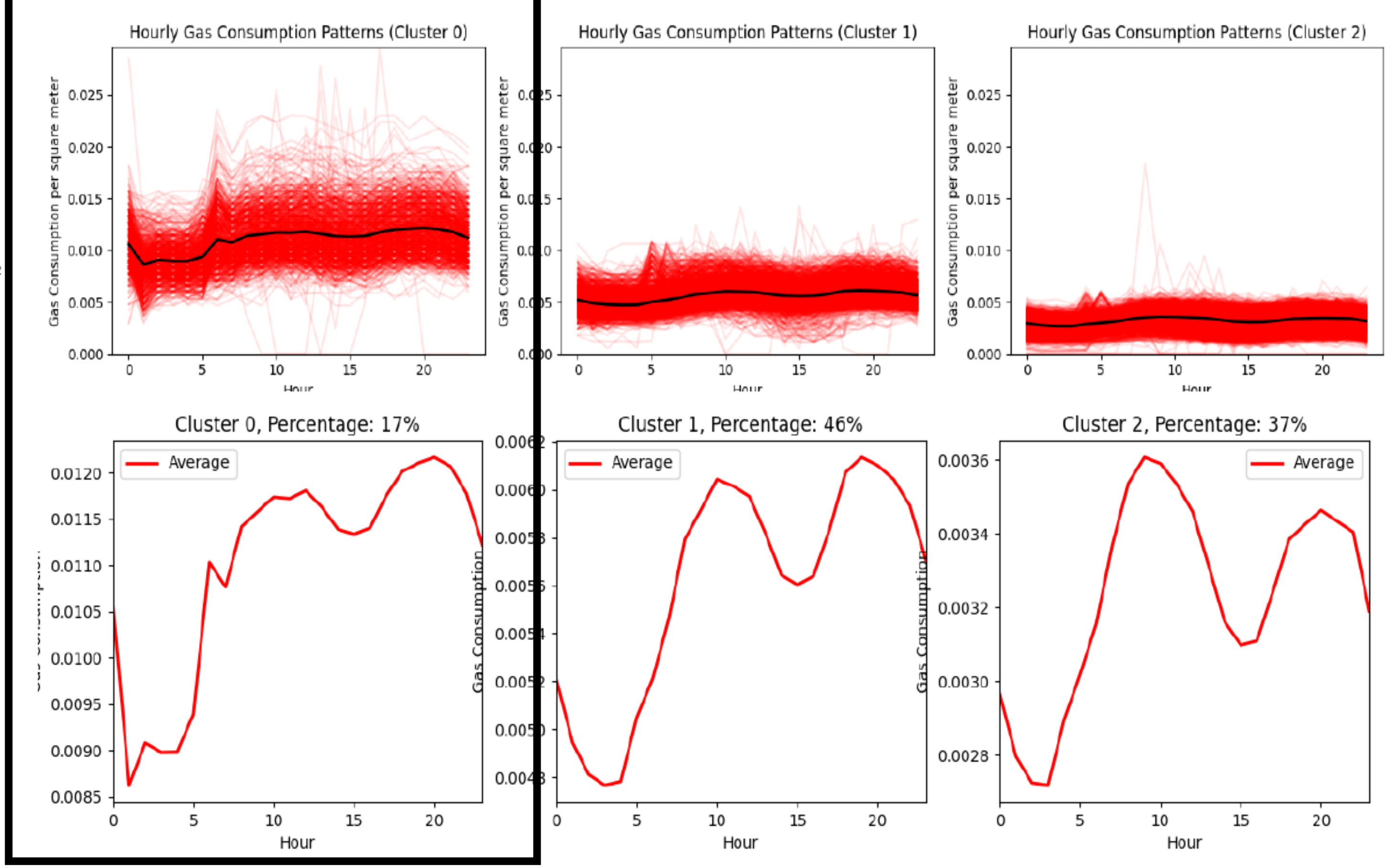


(c) Percentage of meter id in cluster 2

Cluster 0

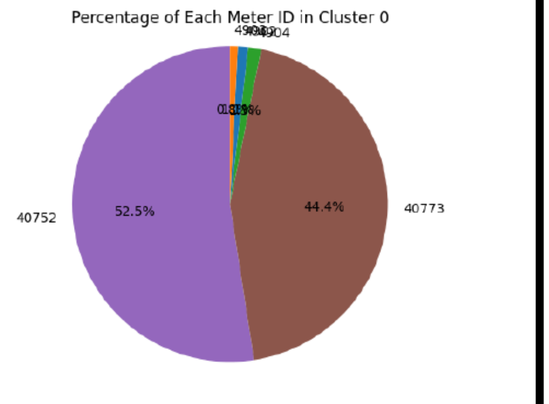
Cluster 1

Cluster 2

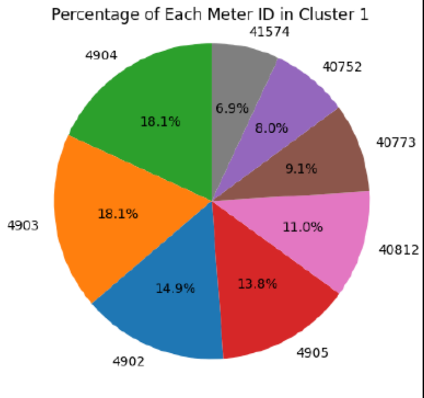


Meter distribution in each group

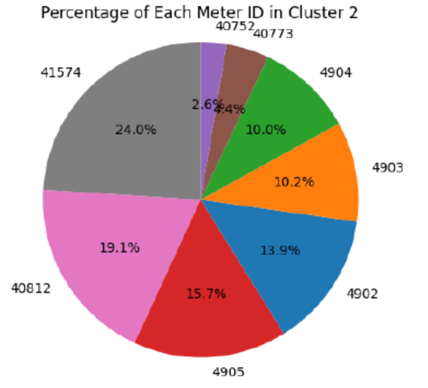
H.0.2. The percentage of meter ID in each cluster



(a) Percentage of meter id in cluster 0

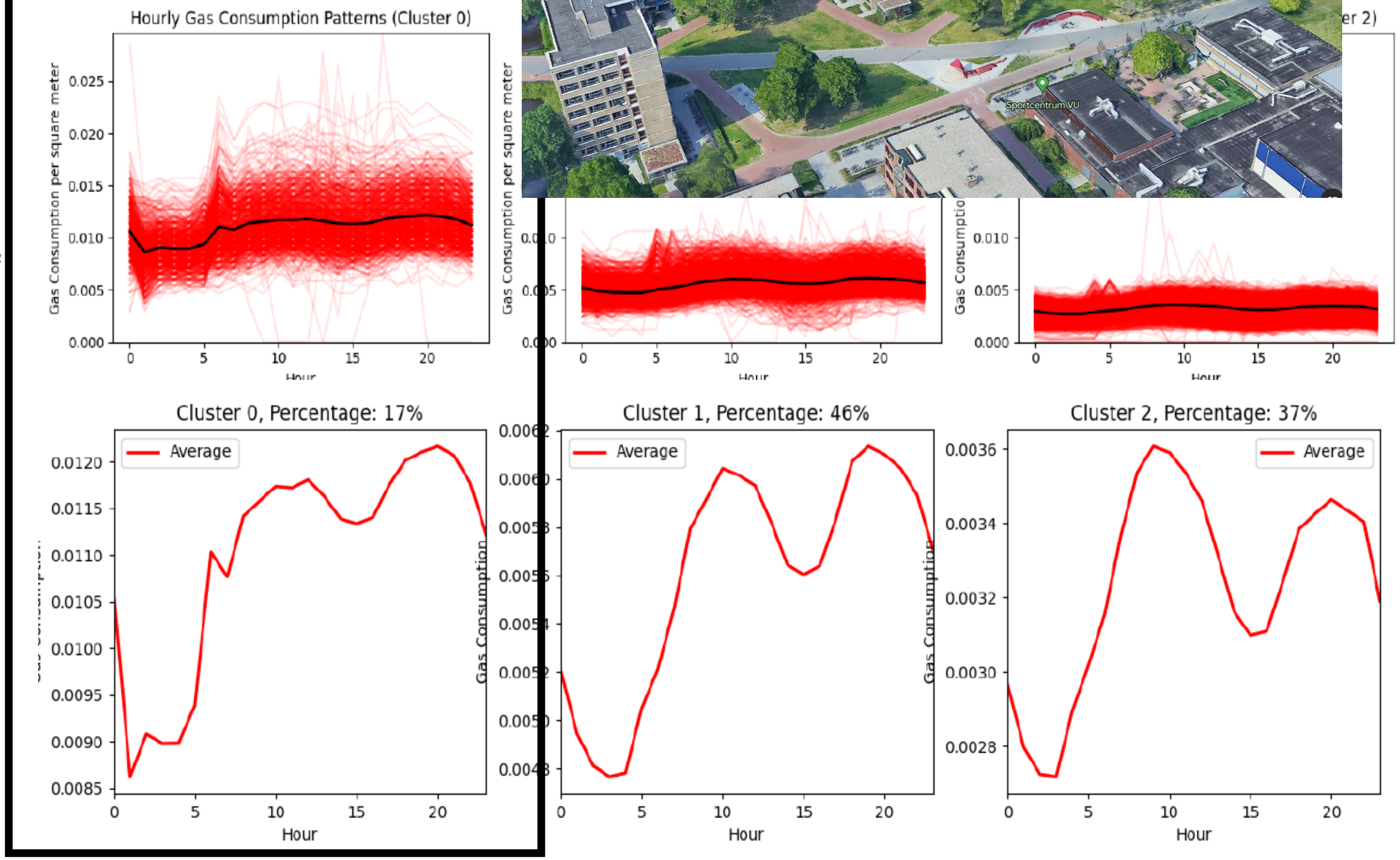


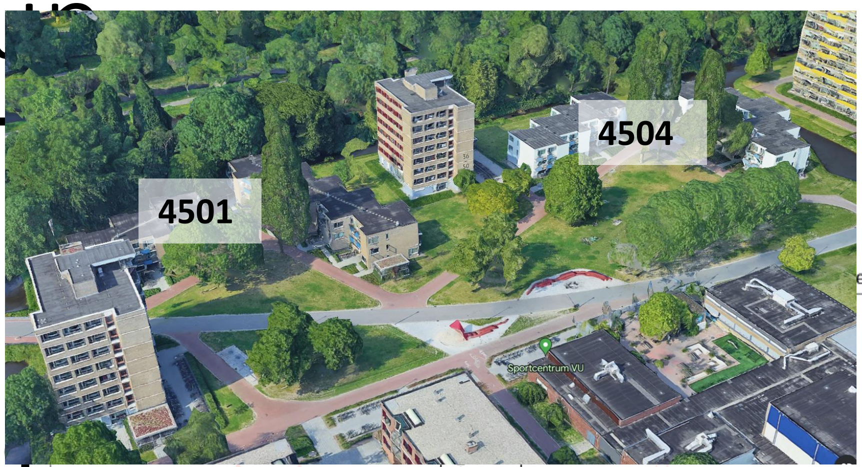
(b) Percentage of meter id in cluster 1



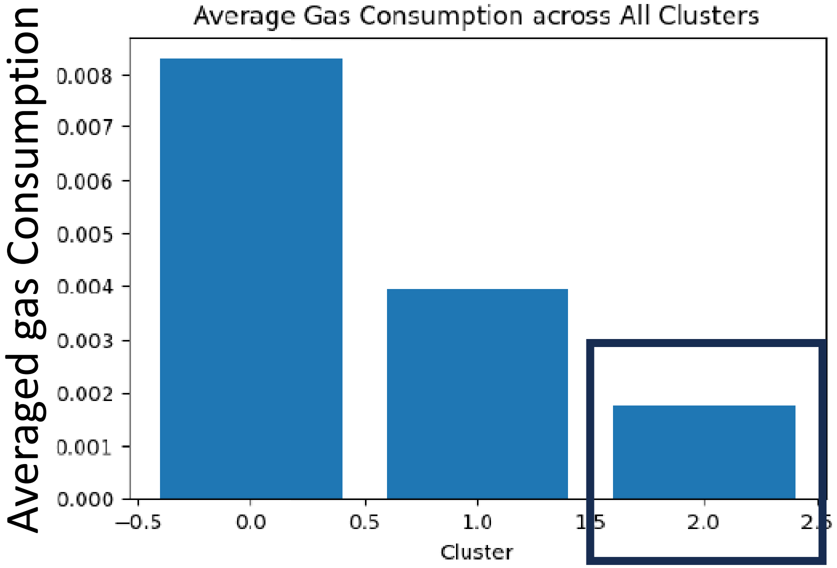
(c) Percentage of meter id in cluster 2

Cluster 0





Average gas consumption &Cluster information



(b) Gas consumption per floor area Inter-Building Clusters

Table H.4.: Cluster Statistics

Cluster	Year Highest Proportion	Avg. Roof Insulation (m^2 K/W)	Avg. EP1 (kWh/m^2)	Avg. EP2 (kWh/m^2)	Avg. EP2 EMG forf. (kWh/m^2)	Average Insulation Floor [m^2K/W]
0	1991	2.21	122.22	201.3	201.3	1.30
1	2014	2.98	91.17	174.96	174.96	2.41
2	1982	2.33	142.19	207.43	207.43	1.66

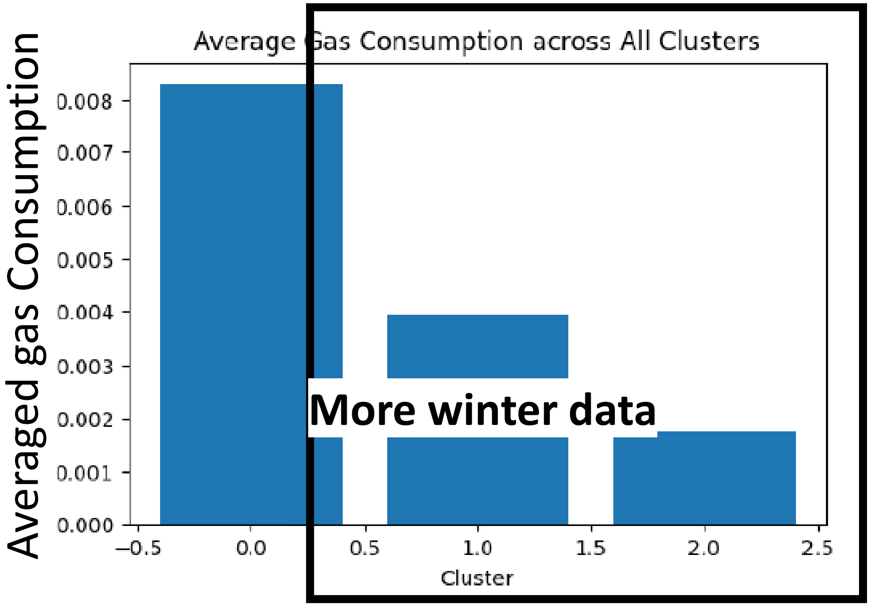
Cluster	Avg. Insulation Facades Excl. AOR (m^2 K/W)	Avg. Facades Ex AOR (m^2)	Avg. Insulation Windows (W/m^2 K)	Avg. Until. Window(m^2)
0	2.00	29.73	9.28	1.8
1	2.28	13.71	10.06	2.08
2	1.58	13.10	10.92	2.15

Lowest average insulation, oldest built year

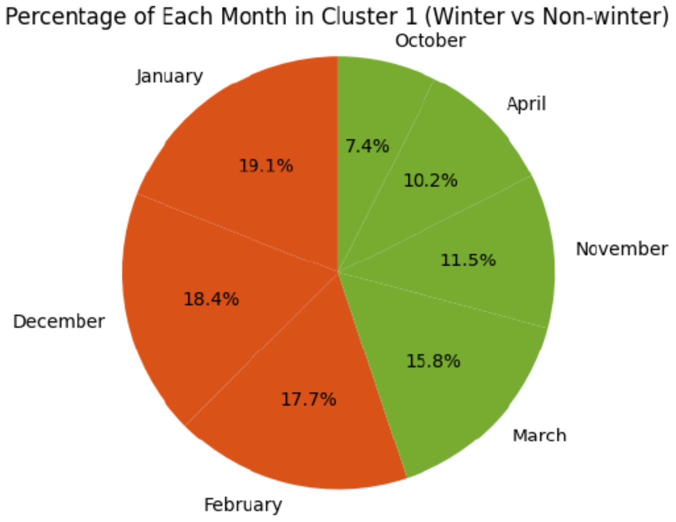
Average gas consumption

Winter months

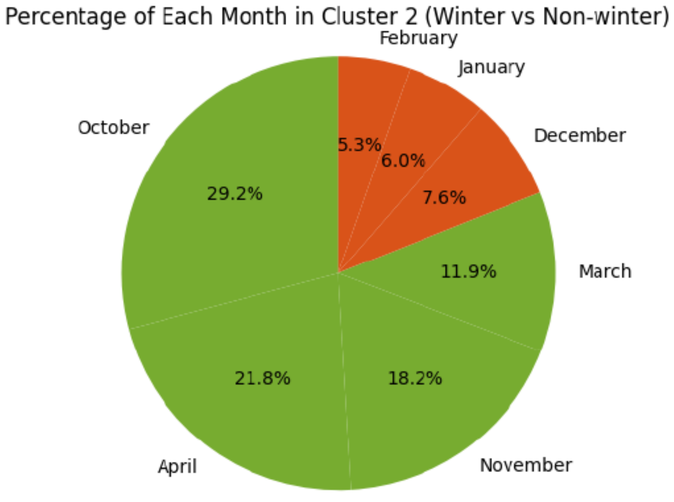
Non winter months



(b) Gas consumption per floor area Inter-Building Clusters



(a) Percentage of Each Month in cluster 1



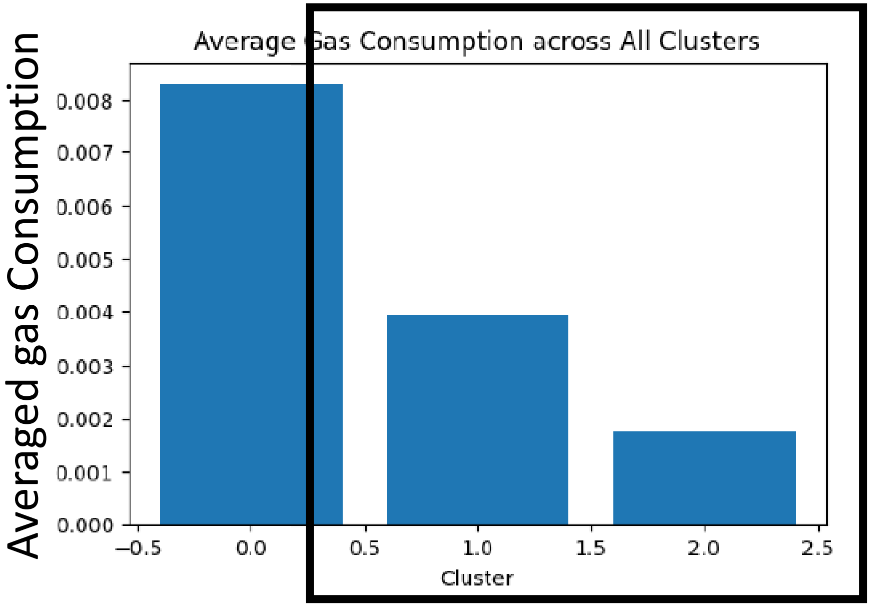
(b) Percentage of Each Month in cluster 2

Cluster 1

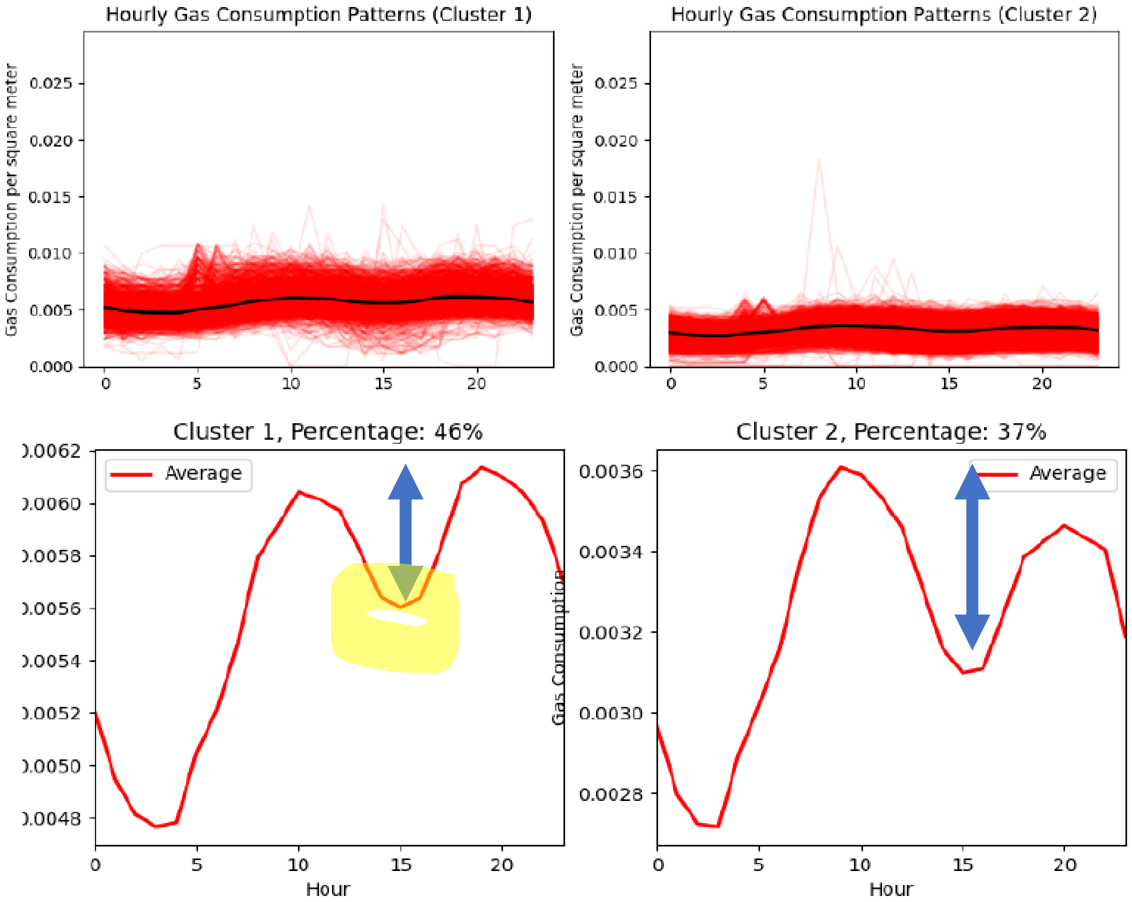
Cluster 2

Higer consumption
more data from winter

Average gas consumption



(b) Gas consumption per floor area Inter-Building Clusters



Cluster 1

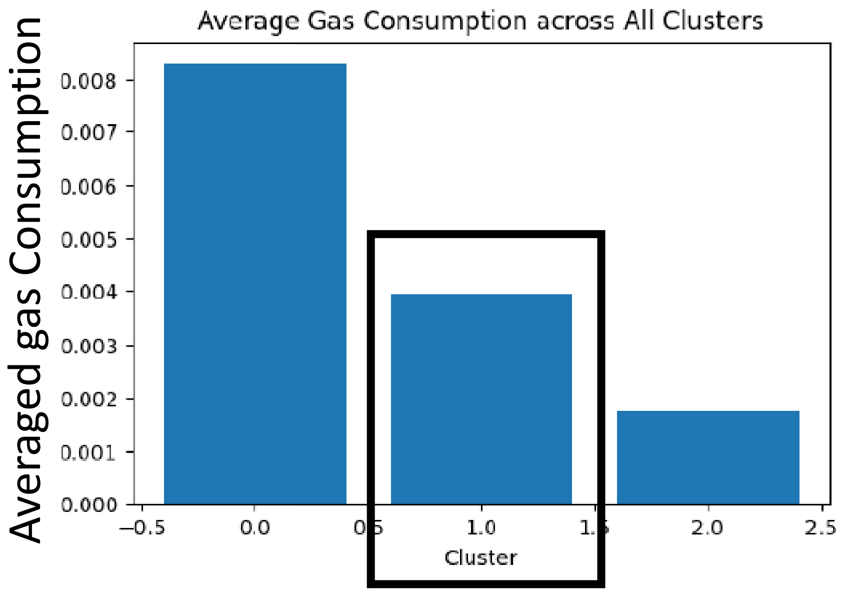
Cluster 2

Higer consumption

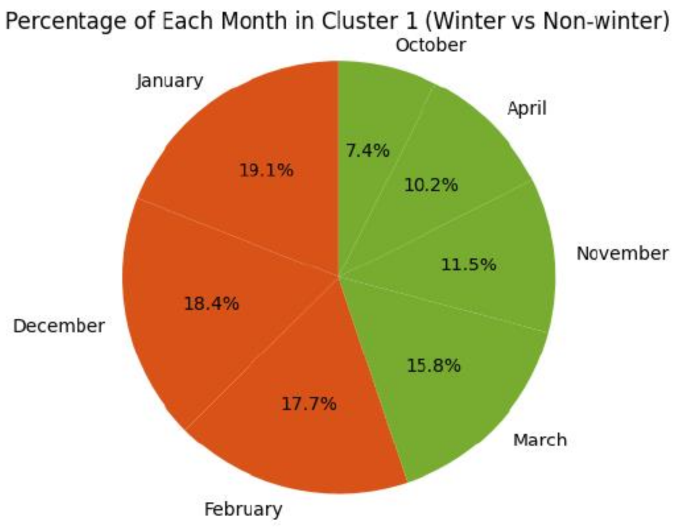
more data from winter

Cluster 1

Buildings **mostly** from 2014, **highest average insulation.**
Not the lowest consumption, **heat losses in the system**.
Percentage of winter months

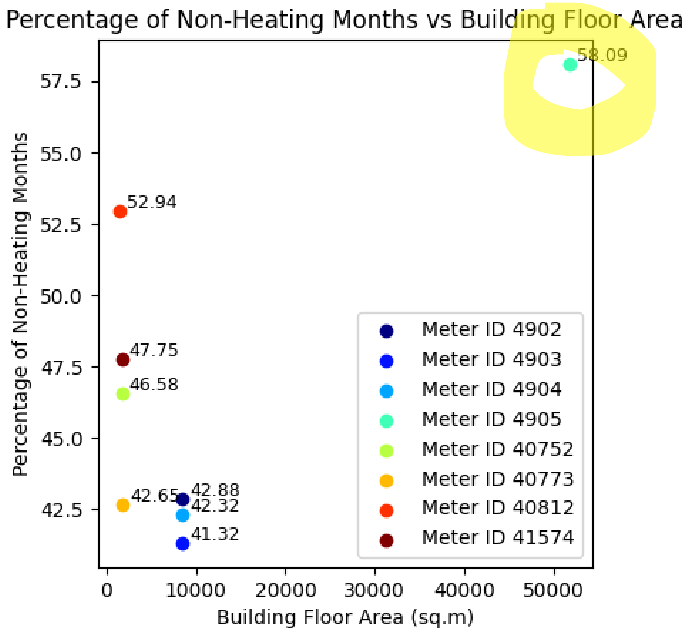


(b) Gas consumption per floor area Inter-Building Clusters



(a) Percentage of Each Month in cluster 1

Winter months



Large percentage of hot tap water

Conclusion & Limitations

Conclusion

- Similar gas usage patterns across all building **student accommodations**.
- Highlight the importance of heating schedules in energy usage.
- Heat losses in the heating system might reduce the energy efficiency more than level of insulation in the building.

Limitations

- Lack of direct end-user metering
- More building information
- Focus on student accommodations may limit the scope and generalizability of results.

How can energy consumption profiles guide energy conservation strategies?

By understanding patterns of gas consumption, we see what building factors indicate higher energy consumption.

Such as the higher influence of heating systems and distribution system on energy efficiency.

Consumption

Patterns & value

Introduce more building features?
Feature extraction?
What features to consider?

Discussion

More information?

Future research

- Explore diverse building types
- Alternative methods
- Building Features: year, building size (floor area, number of floors), occupancy, equipment (heating, cooling, lighting systems), and insulation levels.
- Feature engineering: Encoding, hashing, transformer, feature crossing, feature extraction, feature crossing, PCA

Thank you

Machine Learning Assisted Analysis of Gas Consumption Usage

User Consumption Profile Analysis focusing on Uilenstede Campus

BUILDING TECHNOLOGY MASTER TRACK

Faculty of Architecture and the Built Environment

First Mentor: Dr. Laure Itard

Second Mentor: Dr. Charalampos Andriotis

Student: Yu Hsiu Tung

Only winter months

December January February

Clustering only winte