Data Analysis Techniques for Monitoring District Heating Substations

Shahrooz Abghari
shahrooz.abghari@bth.se

Co-authors: *Veselka Boeva, Christian Johansson, Jens Brage Håkan Grahn , and Niklas Lavesson*

September 10, 2019

District heating domain

- Faults in substations [1]:
  - Comfort problem and physical faults
  - Known faults but unsolved
  - Faults which require new fault detection methods

[1]. Gadd, H., Werner, S.: Fault detection in district heating substations. Applied Energy 157 (2015) 51–59

Hybrid method

- Higher Order Mining [2]: mining from *derived data rather than raw data*
- Combination of methods:
  - Sequential Pattern Mining - > to find ***frequently occurring patterns*** with respect to a user-specified threshold.
  - Clustering Analysis,
    - Weekly clustering - > to group extracted patterns based on their similarities
    - Consensus clustering, -> to group patterns that are coming from two weeks explaining the same phenomenon.
  - Minimum Spanning Tree -> to detect ***outlying*** patterns

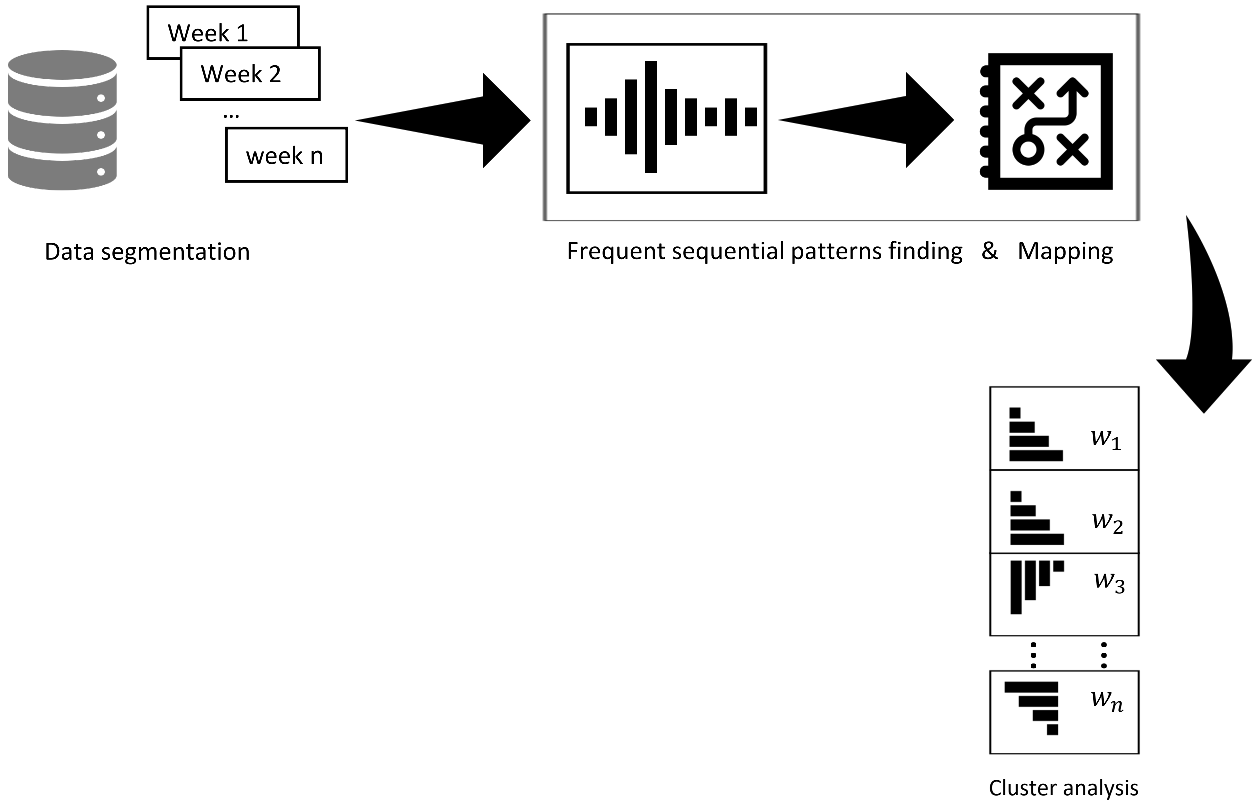
[2] F. Roddick, M. Spiliopoulou, D. Lister, and A. Ceglar, “Higher order mining,” ACM SIGKDD Explorations Newsletter, vol. 10, no. 1, pp. 5–17, 2008.

Data & preprocessing

- Hourly numerical measurements, 2-year data
- Feature selection
  - *Outdoor temperature* and ***five*** features:
    - *‘sec_delta_t’, ‘prim_delta_t’, ‘consumed_energy’, volume_flow’, ‘station_effectivness’*
    - $station\_effectivness = \dfrac{prim\_delta\_t}{primary\ supply\ temperature - secondaryreturn\ temperature}$
- Data categorization
  - Transform the continues data into ***four*** categories
  - ***Four*** seasons = *low, low_medium, medium_high, high*

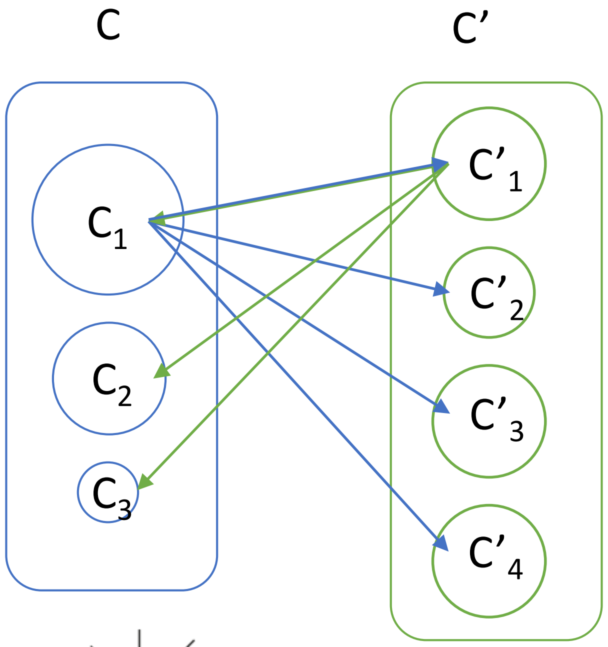
sec_delta_t	prim_delta_t	consumed_energy	volume_flow	station_effectivness
high	*high*	*medium_high*	*medium_high*	*high*

Proposed method

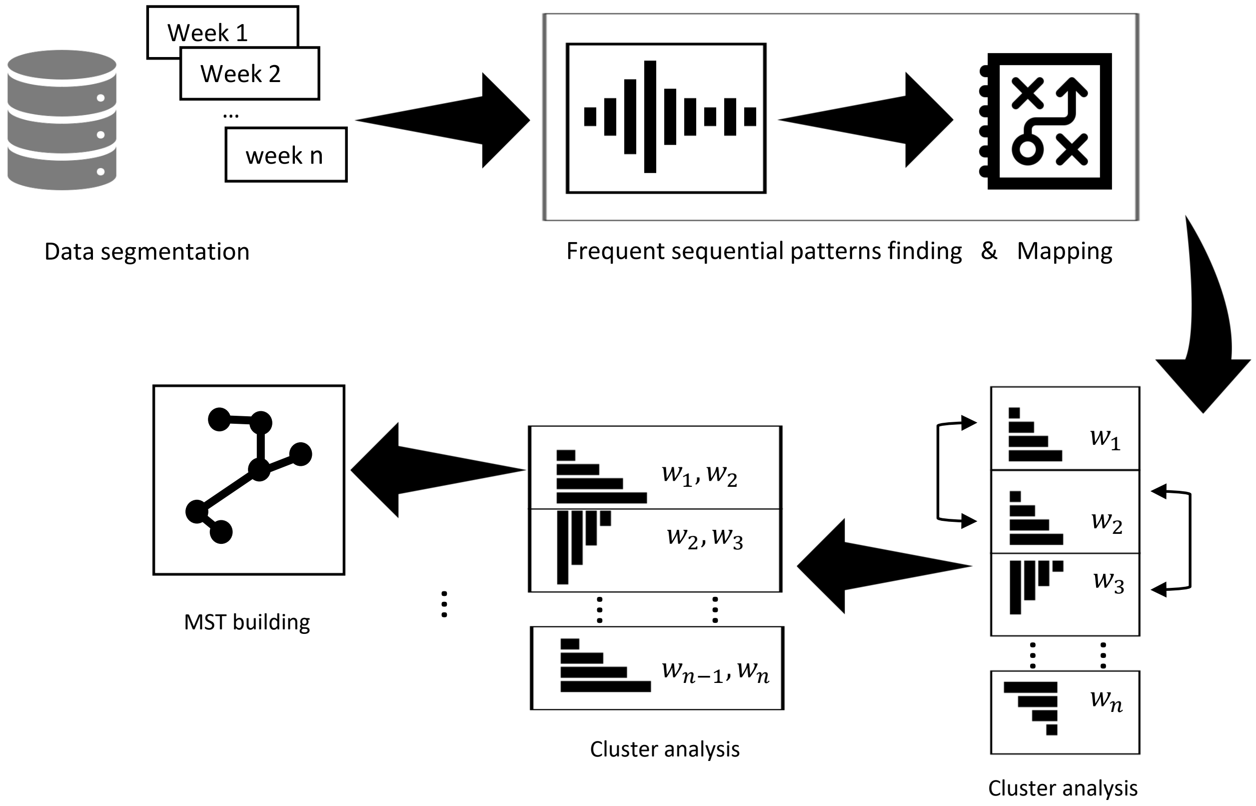


Similarities between two weeks

- Given two clustering solutions of datasets X and X':
  - $C \;=\; \{C_1, C_2, \ldots, C_n\}$
  - $C' \;=\{C'_1, C'_2, \ldots, C'_m\}$
- $S_w(C, C') = \dfrac{\sum_{i=1}^{n}(min_{j=1}^{m} w_i.d(c_i,c\prime_j))}{2} + \dfrac{\sum_{j=1}^{m}(min_{j=1}^{n} w\prime_j.d(c_i,c\prime_j))}{2}$
  - $w_i = \dfrac{|c_i|}{X}$
  - d, distance measure



Proposed method

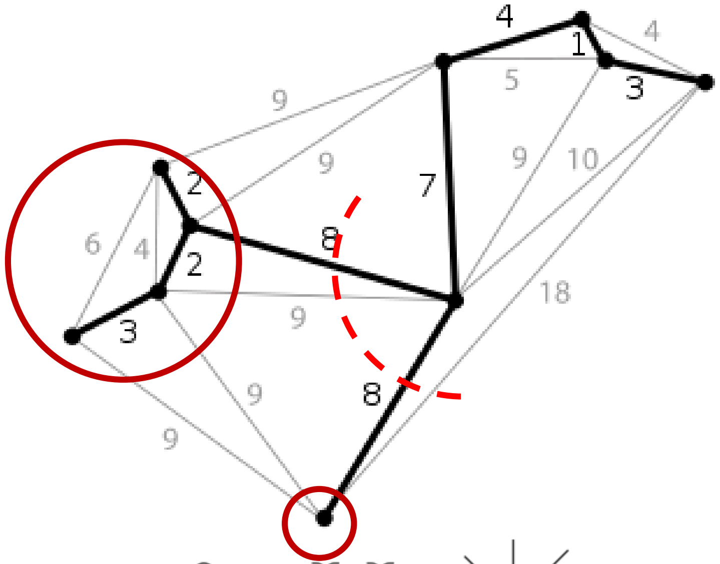


Minimum spanning tree (MST)

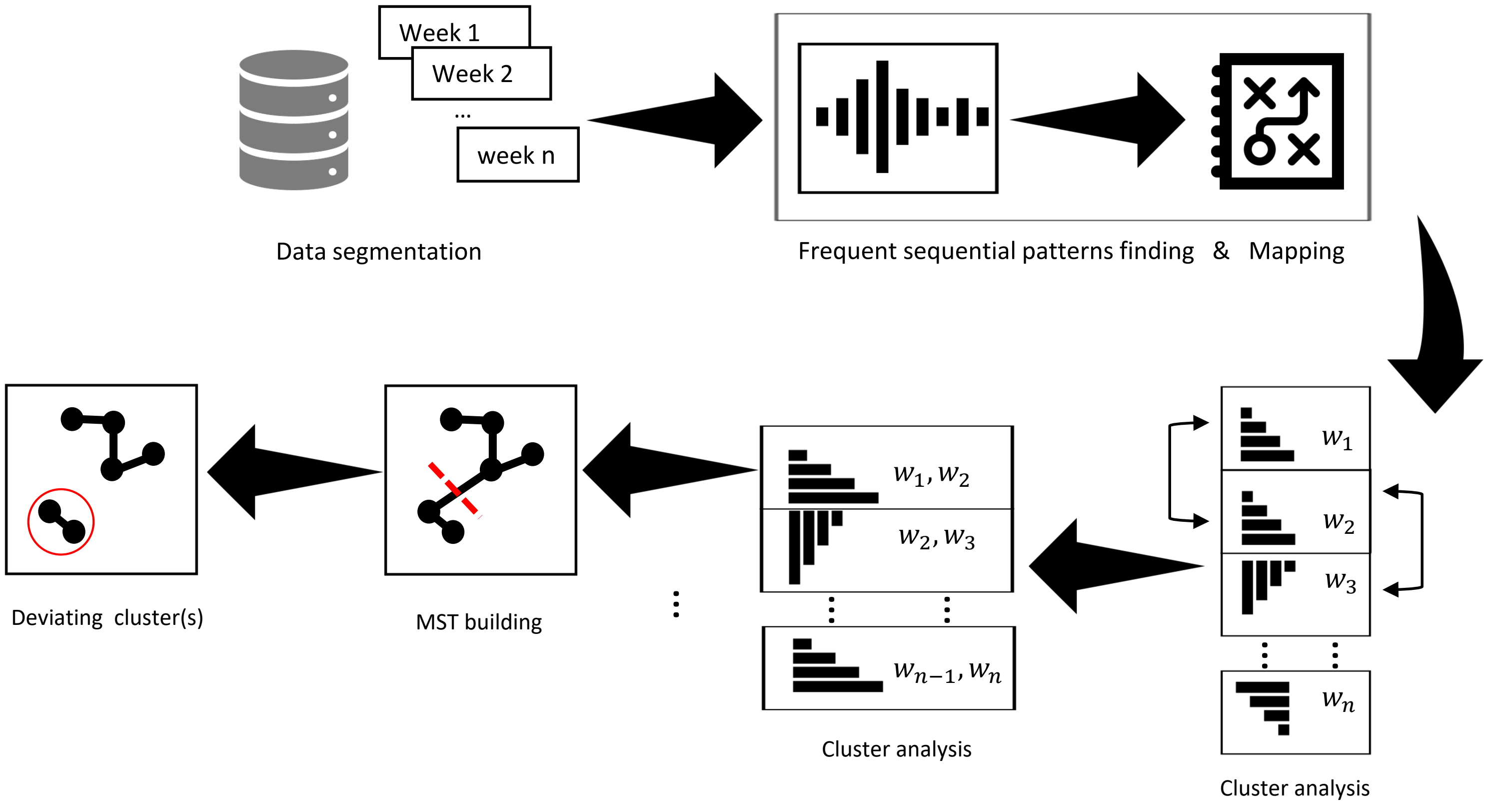
- Given an undirected, connected, and weighted graph $G=(V,E)$
  - a spanning tree of the graph G is a tree that covers all the nodes of G
  - a tree with the minimum cost of traversing is called minimum spanning tree

- MST is applied on top the clustering solutions
  - V = exemplars of the clusters
  - E = dissimilarity between V

- Removing the longest edges
  - Smaller sub-trees are labeled as outlier

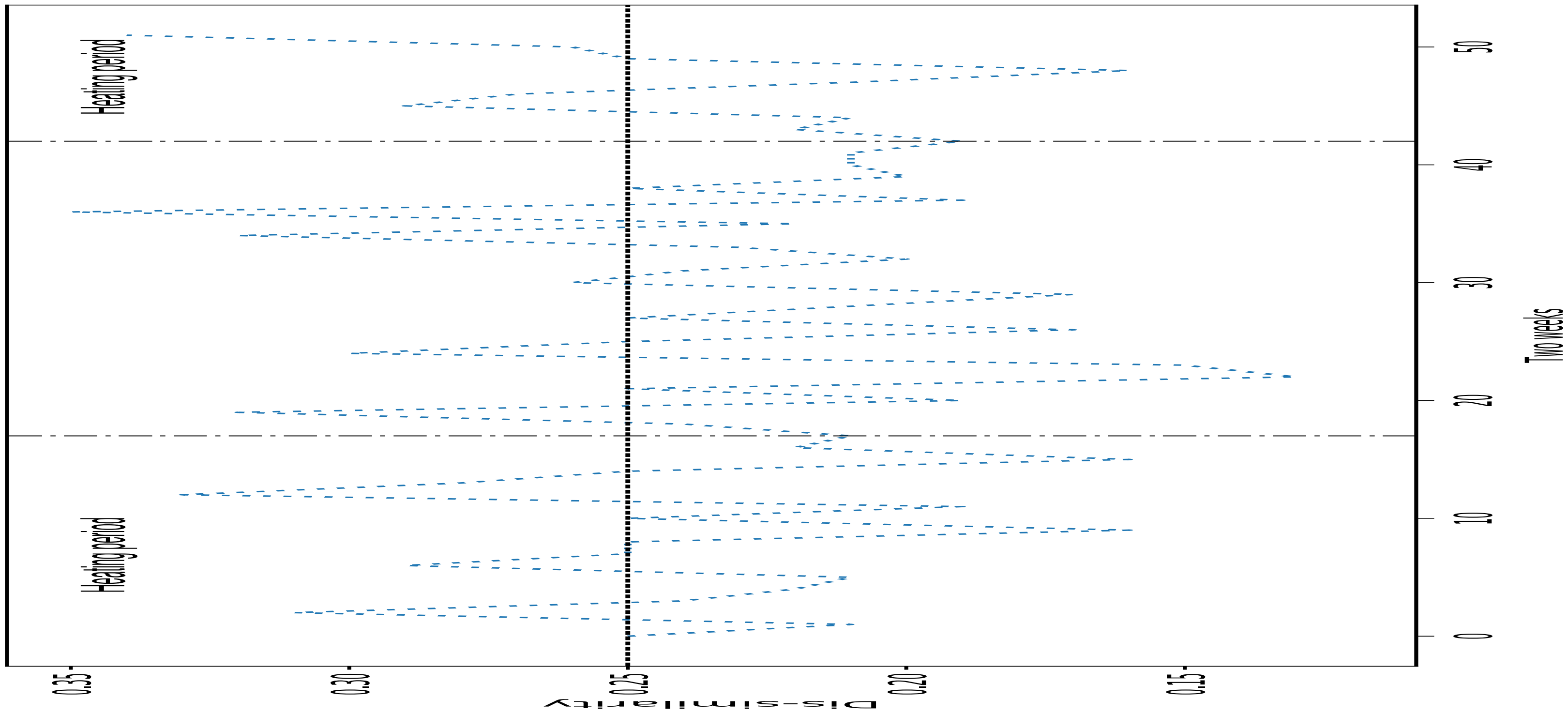


Proposed method (cont.)



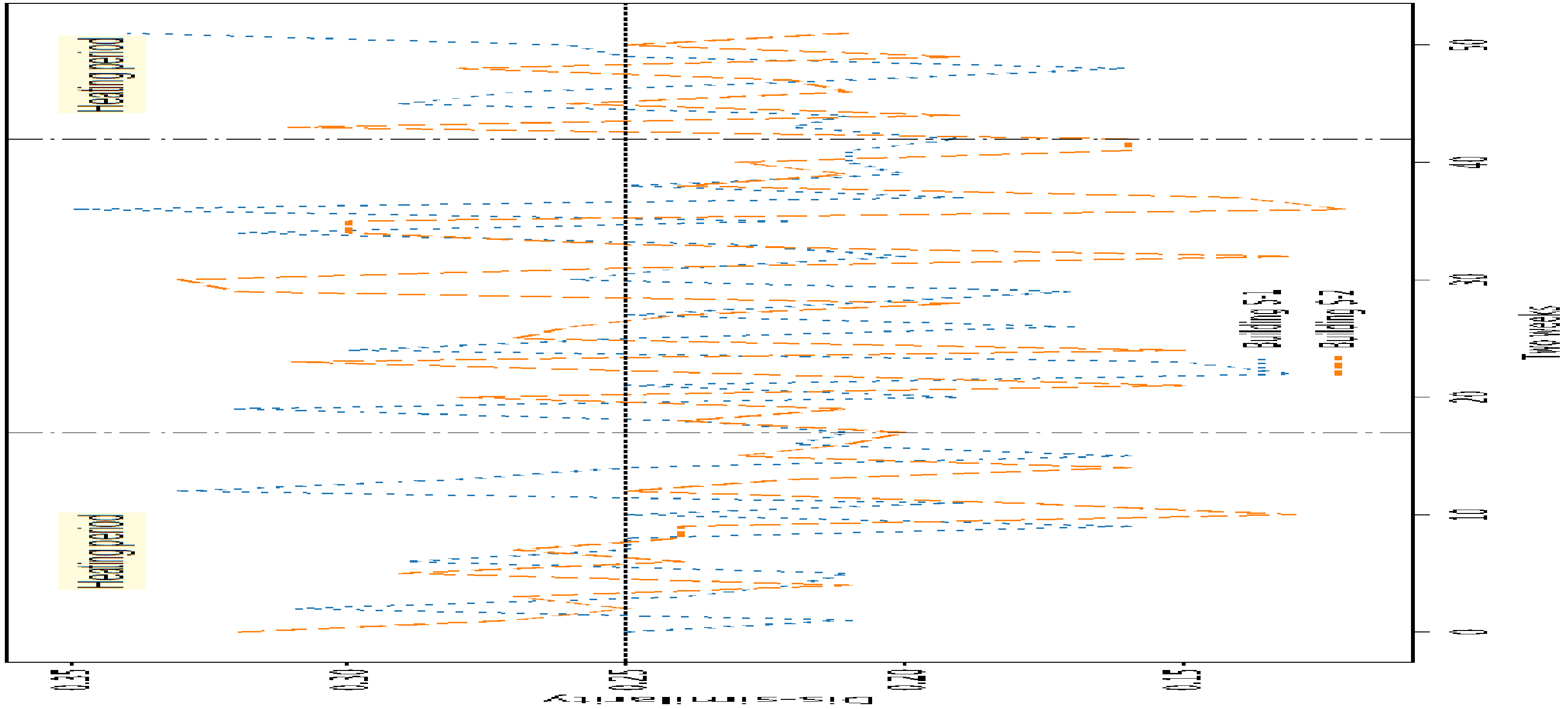
Results:

Substation’s bi-weekly profile, 2017



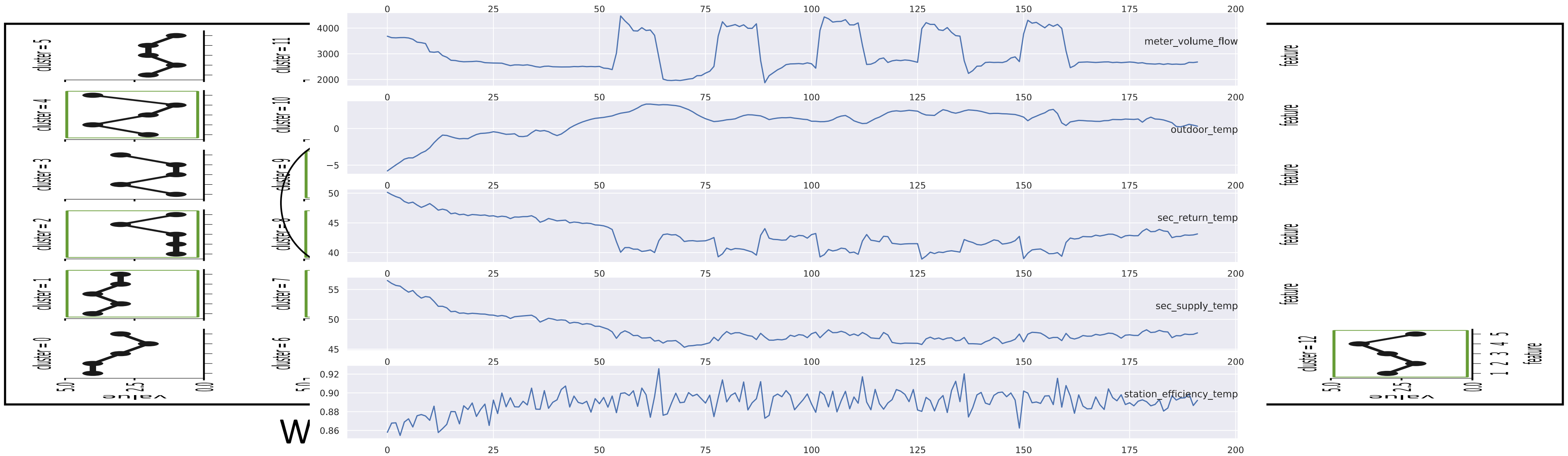
Results:

Comparison of two profiles, 2017



Results:

Weekly clustering models and consensus model



Future work

- We aim to pursue further analysis and evaluation of the proposed approach
- In the long-term perspective,
  - Label weekly patterns with some *performance indicators*

Thank You for listening!

Shahrooz Abghari
shahrooz.Abghari@bth.se