IDEAL 2019

Detection of Abnormal Load Consumption in the Power Grid Using Clustering and Statistical Analysis

Matúš Cuper Marek Lóderer Viera Rozinajová

Slovak University of Technology

14.11.2019, Manchester UK

Problem statement

- Identification of abnormal behavior
  - data evolution, unlabeled dataset, noise, anomaly definition etc.
  - classification, clustering, statistical methods etc.
  - local and global anomalies
- Prevention of illegal consumption, identification of malfunctioning devices, optimization of energy distribution
- Behavior that differs from behavior of neighboring instances

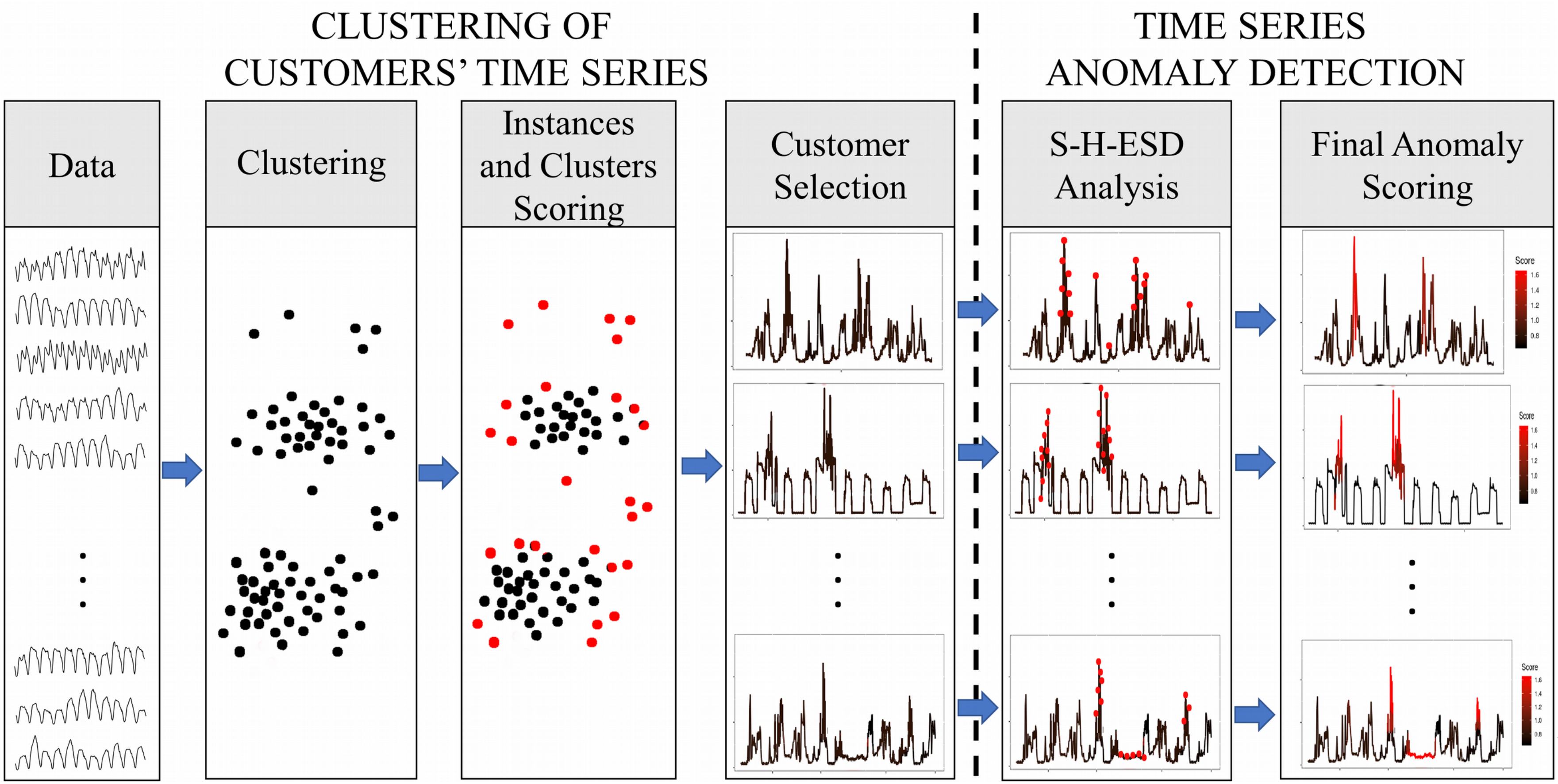
Existing solutions

- Clustering based
  - sliding windows to keep relevant information
  - suitable for quasi-periodic time series
- Statistical based
  - robust to anomalous fluctuations
  - reduction of false-alarms
- Other
  - distance based
  - density based

Proposed method

1. Clustering of time series
2. Scoring of instances and clusters
3. Customer selection
4. S-H-ESD analysis
5. Combining of computed scores

Proposed method



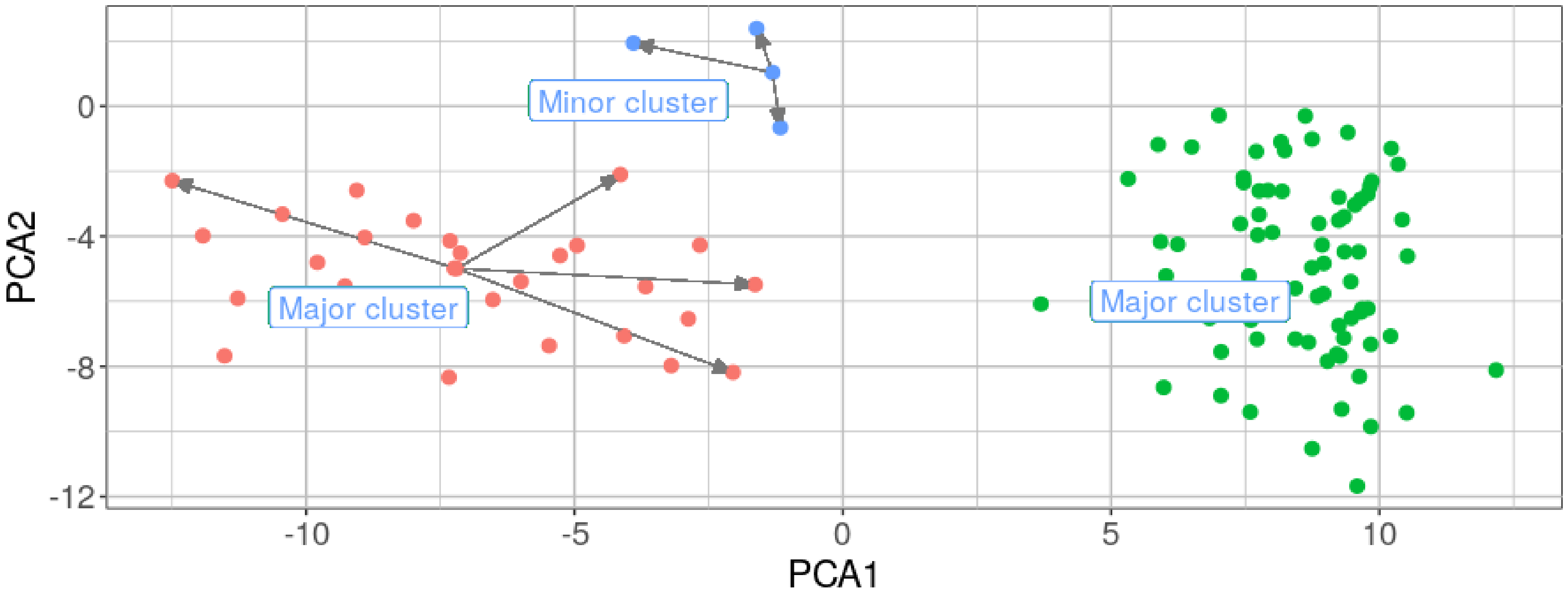
1. Clustering of time series

- Data preprocessing
  - normalization (z-score)
  - window alignment
  - division (workdays and non-workdays)
- Dataset aggregation by sliding windows
  - windows size (2 weeks)
  - window shift (1 week)
- Clustering using k-medoids
  - distance metric (GAK)

2. Scoring of instances and clusters

- Instance scoring
  - different values for different instances
  - relative distance of instance within given cluster
- Cluster scoring
  - different values for different clusters
  - small cluster penalization

2. Scoring of instances and clusters



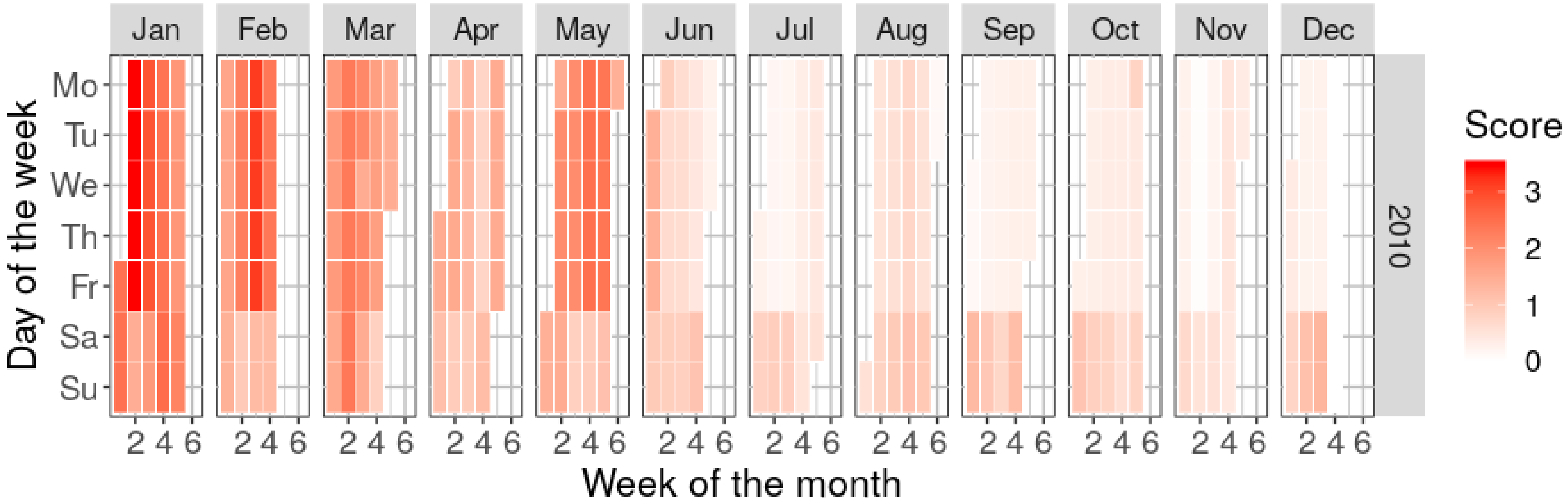
3. Customer selection

- Interval of interquartile rule for anomaly detection

<Q1 – 1.5*IQR, Q3 + 1.5*IQR>

- FeaClip method
- Based on suspicion
- Visualization

3. Customer selection



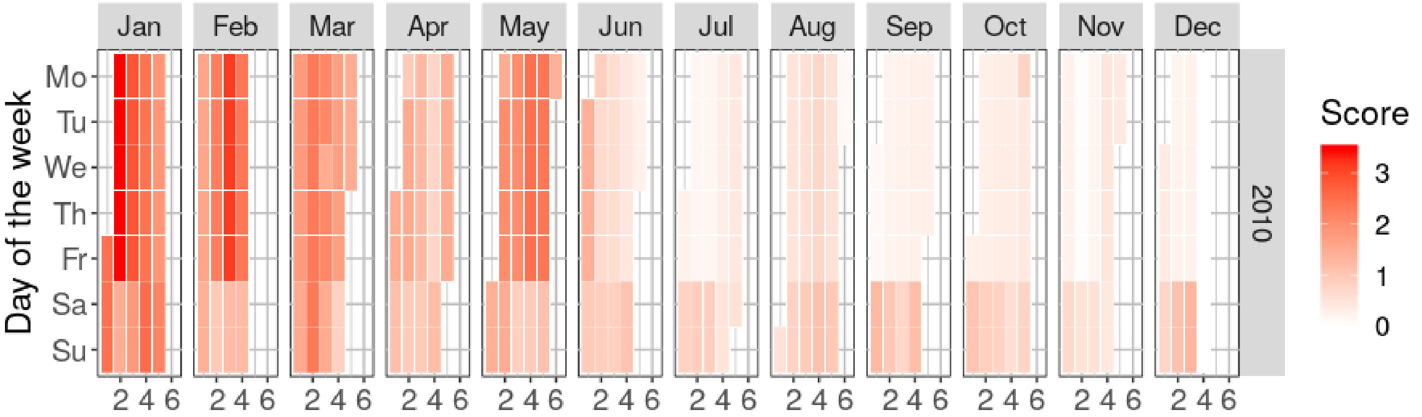
4. S-H-ESD analysis

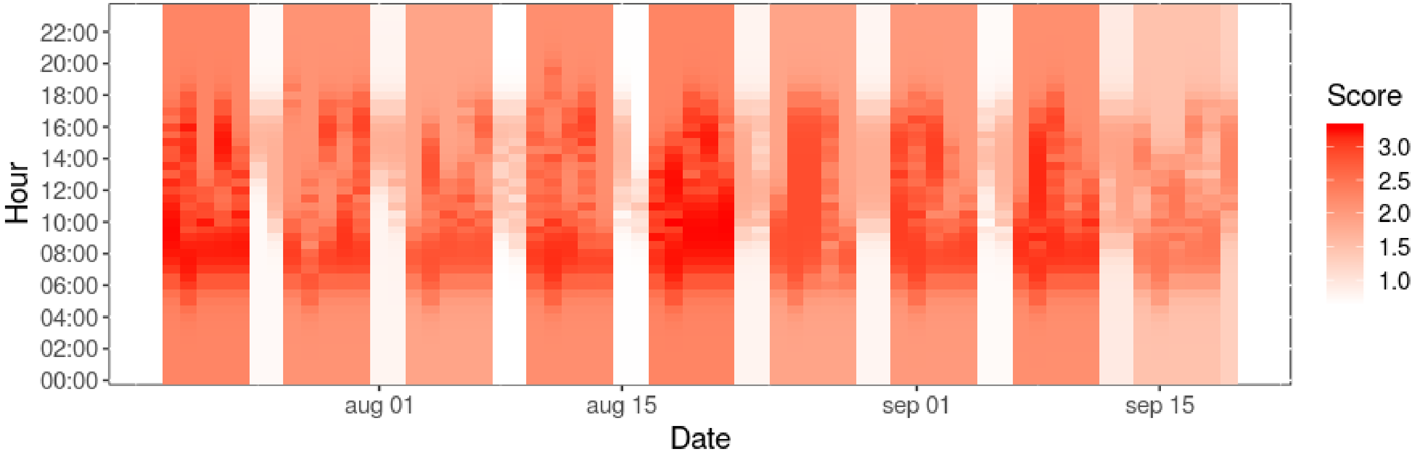
- Smoothing in order to eliminate numerous local anomalies
- S-H-ESD analysis
  - based on Grubb’s test
  - using median absolute deviation
  - capable of identifying up to 50% anomalies
- Output of S-H-ESD analysis is a flag
  - Intervals with dense occurrence of flags are grouped
  - Intervals with sparse occurrence of flags are smoothed

5. Combining of computed scores

- Combining of smoothed flags from S-H-ESD analysis and computed score from k-medoids
- Lower granularity
- Result is not flag indicating anomalousness, but number representing degree of anomalousness of given measurement

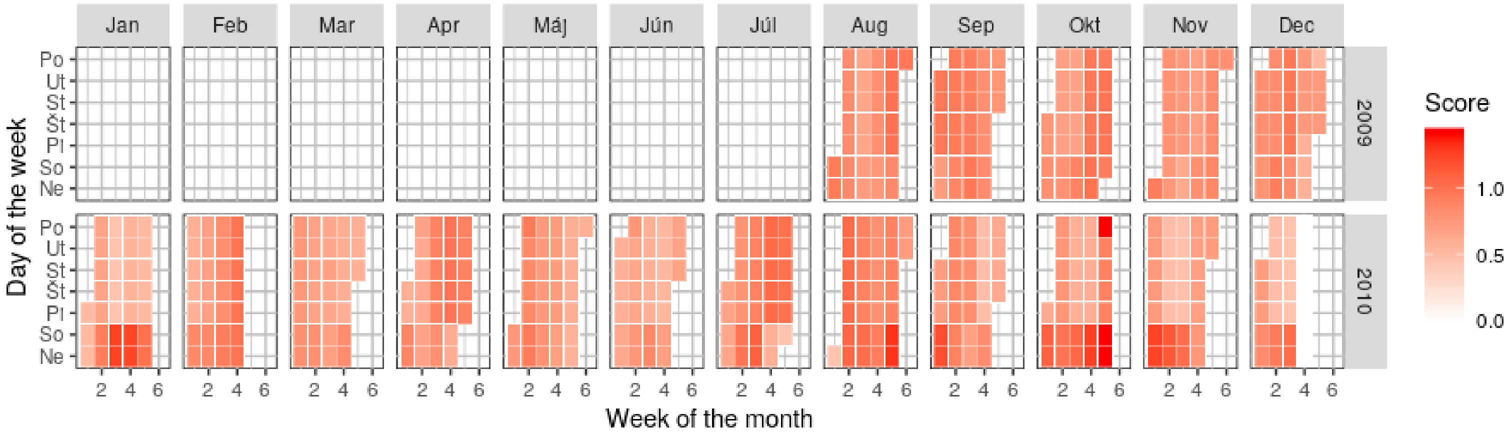
5. Combining of computed scores

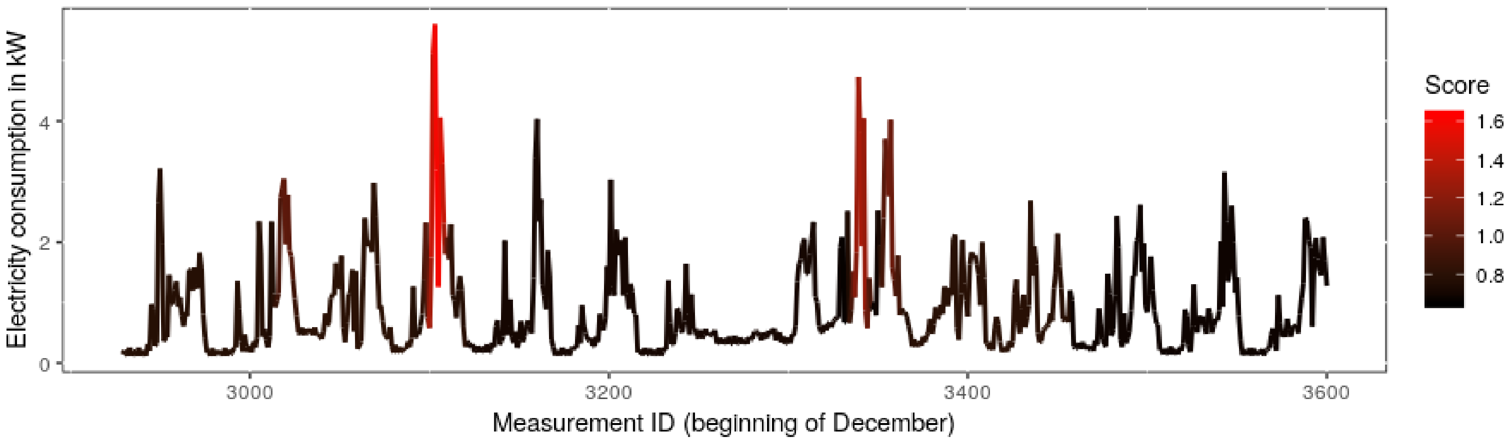




Anomaly score =
Consumer score + S-H-ESD score

5. Combining of computed scores



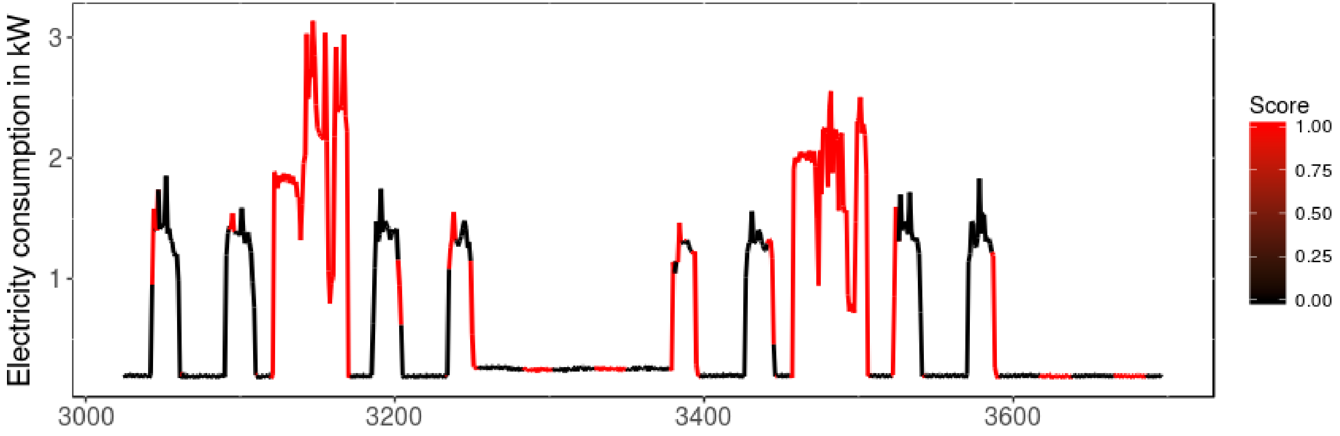


Evaluation

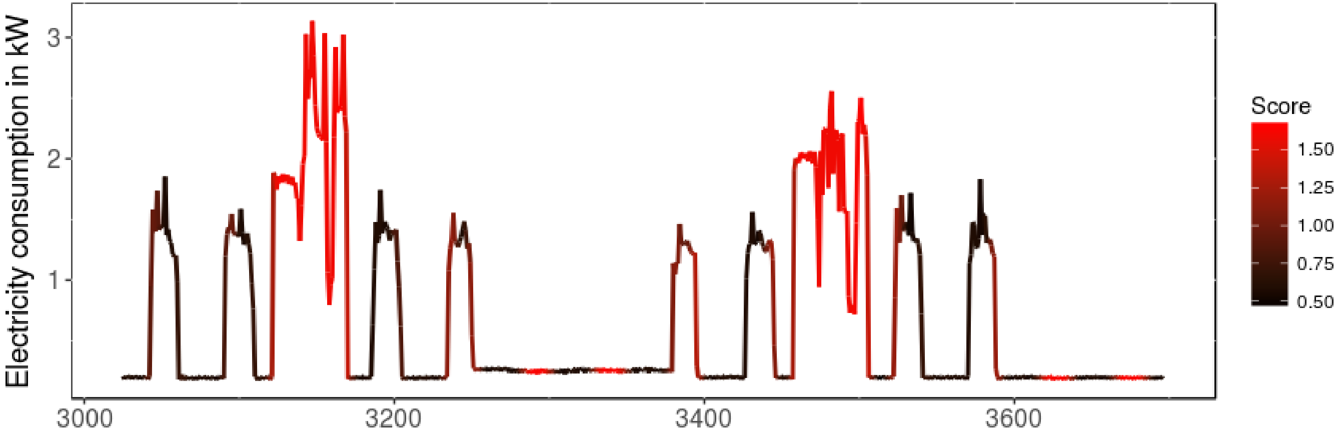
- Dataset from Irish CER Smart Metering Project
- Clustering
  - experiments
  - cluster validation indexes
- Proposed method compared to
  - PEWMA
  - SD-EWMA
  - KNN-LDCD
  - S-H-ESD

Evaluation

S-H-ESD



Proposed Method



Evaluation

Method Name	TP	TN	FP	FN	Precision	Recall	F1 score
PEWMA	0.6 %	82.5 %	3.3 %	13.6 %	17.67 %	4.90 %	7.67 %
SD-EWMA	0.7 %	83.1 %	2.6%	13.6 %	20.36 %	4.69 %	7.62 %
KNN-LDCD	0.04 %	85.6 %	0.03 %	14.2 %	60.00 %	0.31 %	0.62 %
S-H-ESD	6.4 %	58.1 %	27.6 %	7.9 %	18.82 %	44.90 %	26.52 %
Proposed Method	**4.4 %**	**80.6 %**	**5.1 %**	**9.9 %**	**42.84 %**	**39.27 %**	**40.98 %**
Reality	14.3 %	85.7 %	-	-	-	-	-

Summary

- Only consumers with high score are analyzed
- Flag replaced by degree of anomalousness
- Possible online processing

Thank you for your attention