

# Assessment of Unsupervised Standard Pattern Recognition Methods for Industrial Energy Time Series

Nicole Ludwig, Simon Waczowicz, Ralf Mikut, Veit Hagenmeyer

## Motivation

**Unsupervised Learning:** How do we measure success?

Two frequently used methods are fuzzy-c-means clustering and motif discovery. We want to see how their results differ and whether it is important to consider which method is used to find patterns in energy time series.

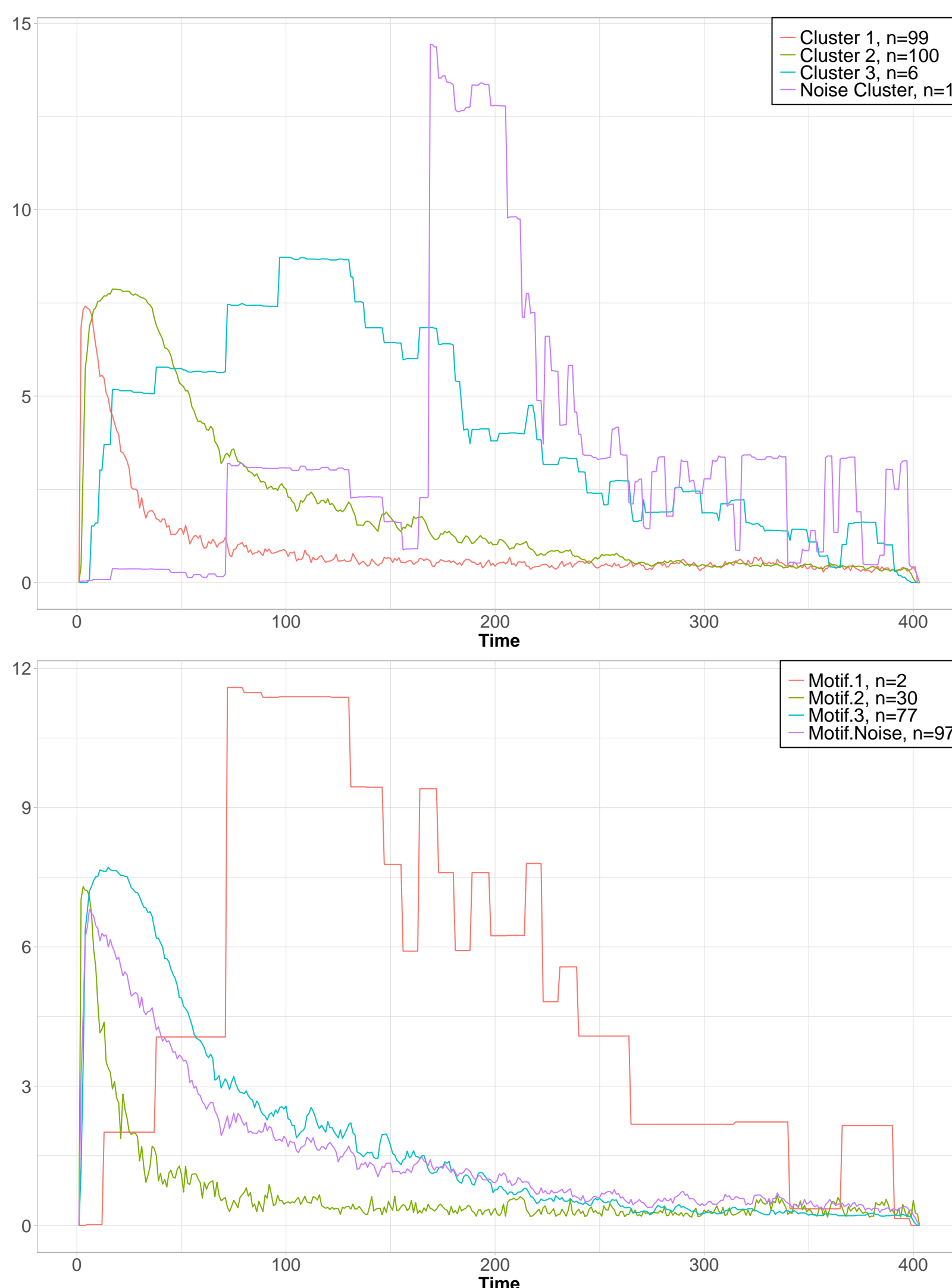
## Methodology

- Fuzzy-C-Means Clustering [1]
- Motif Discovery (SAX & Random Projection) [2]
- Industrial energy time series data set [3]

We divide our evaluation into two steps:

- Compare the results on a mean level
- Investigate the subsequence level of the results

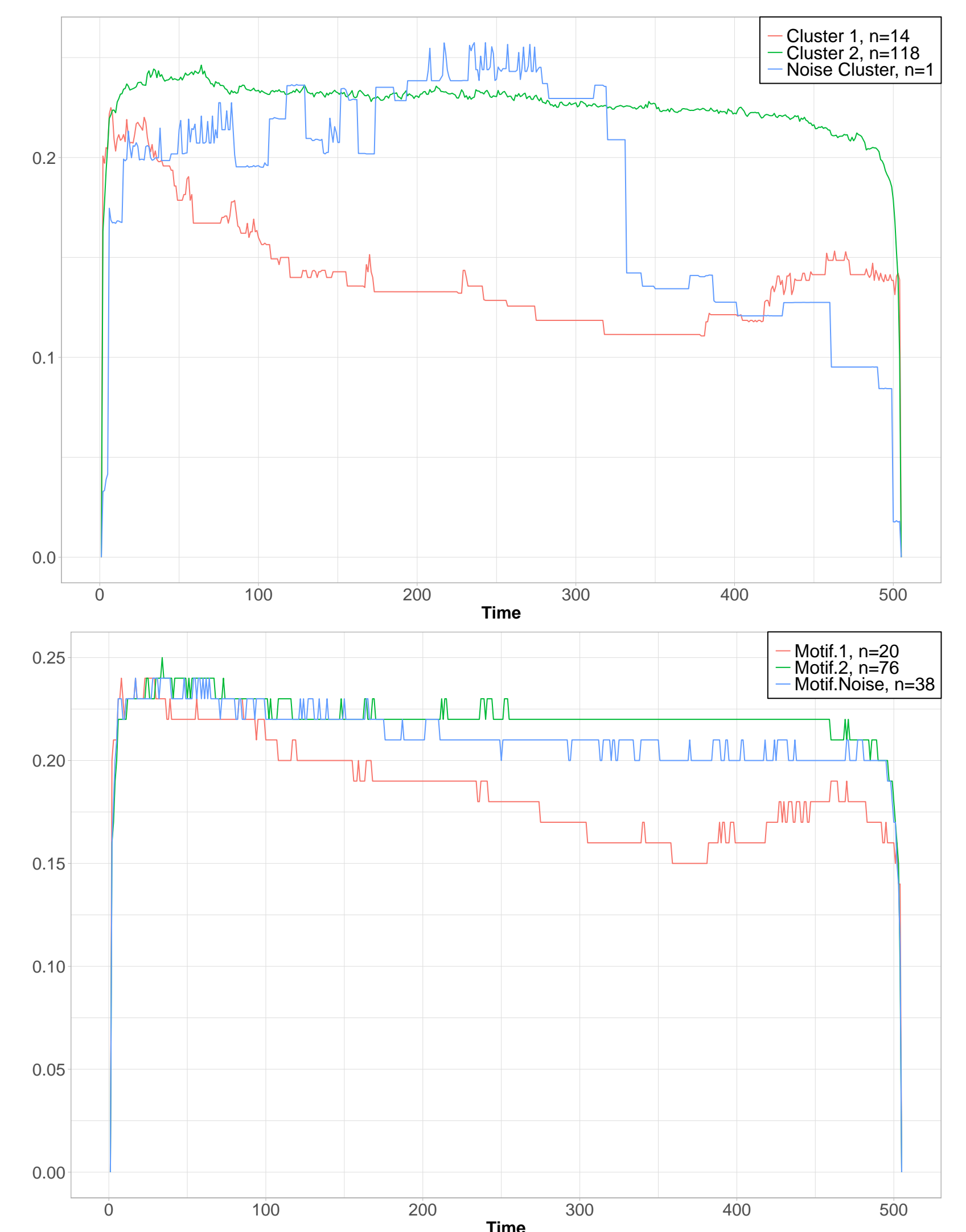
## Mean level comparison



Two prominent differences:

- The magnitude of the value of the mean curves varies substantially
- The clustering finds information in patterns where the motif does not (Noise)

Overall, however, the shape of the mean curves are similar for both methods, indicating that on a coarse level the choice of methodology seems unimportant.



## Subsequence level comparison

The number of subsequences assigned to each cluster in comparison to their motif assignment.

	Motif 1	Motif 2	Motif 3	Noise Motif
Cluster 1.1	3	4	NA	7
Cluster 1.2	17	70	NA	31
Noise Cluster 1	-	1	NA	-
Cluster 2.1	1	17	35	46
Cluster 2.2	1	11	40	48
Cluster 2.3	-	2	2	2
Noise Cluster 2	-	-	-	1
Cluster 3.1	3	19	17	8
Cluster 3.2	1	-	6	1
Cluster 3.3	-	1	6	3
Noise Cluster 3	-	-	1	-
Cluster 4.1	3	29	NA	15
Cluster 4.2	1	6	NA	1
Cluster 4.3	-	5	NA	5
Noise Cluster 4	-	-	NA	1

## Conclusion

- Mean shapes are comparable
- The sequences assigned to each group differ greatly
- There are no exhaustive studies on the performance of the two algorithms
- If this result is true for other data, an uneducated choice of methodology distorts the result

**Future research:** Are there specific patterns, which one algorithm can undoubtedly find while the other cannot and which algorithm is preferable for which specific application.

## References

- [1] Waczowicz et al. (2015): *Demand response clustering - How do dynamic prices affect household electricity consumption?* In: 2015 IEEE Eindhoven PowerTech. Eindhoven, Netherlands: IEEE, S. 1- 6.
- [2] Chiu et al. (2003): *Probabilistic discovery of time series motifs*. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. The Ninth ACM SIGKDD International Conference. Washington, D.C. ACM Special Interest Group on Knowledge Discovery in Data; ACM Special Interest Group on Management of Data. New York, NY: ACM, S. 493 - 498.
- [3] Bischof et al. (2018): *HIPE - An Energy-Status-Data Set from Industrial Production*. In: Proceedings of the Ninth International Conference on Future Energy Systems - e-Energy '18. The Ninth International Conference. Karlsruhe, Germany. New York, New York, USA: ACM Press, S. 599 - 603.