

On the Tradeoff between Energy Data Aggregation and Clustering Quality

Holger Trittenbach (KIT), Jakob Bach (KIT), Klemens Böhm (KIT)

Problem Statement

Volume of energy data challenges the scalability of data processing and analysis

- High frequency: energy data collected in second intervals, from many devices
- Multiple Measurements: e.g., voltages, currents, harmonic distortion

Unclear how to assess effects of temporal aggregation for energy data clustering

- Clustering to discover consumer groups or recurring patterns
- Data reduction by downsampling or aggregation, e.g., average over 15 min
- Tradeoff between data volume and information content

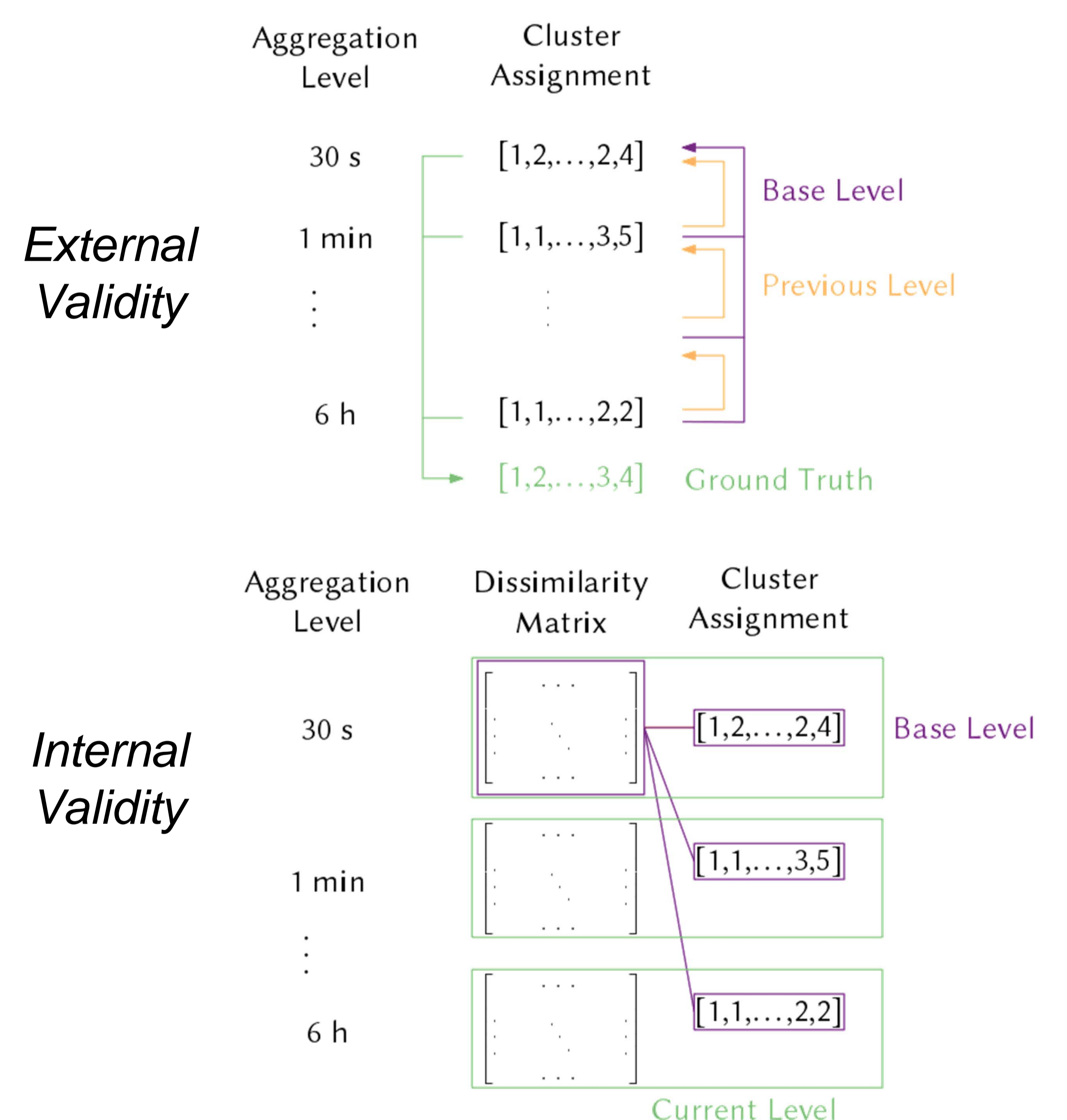
Experimental Design

Design Space

- Data Set:** machine-dependent and grid-dependent electrical quantities
- Clustering Algorithm:** representative-based, hierarchical, density-based
- Dissimilarity:** lock-step, elastic, complexity-based
- Aggregation Function:** location (max, mean), dispersion (standard deviation), shape (skewness)
- Aggregation Level:** intervals from 1 min up to 6 h
- Over 43,000 experiment settings

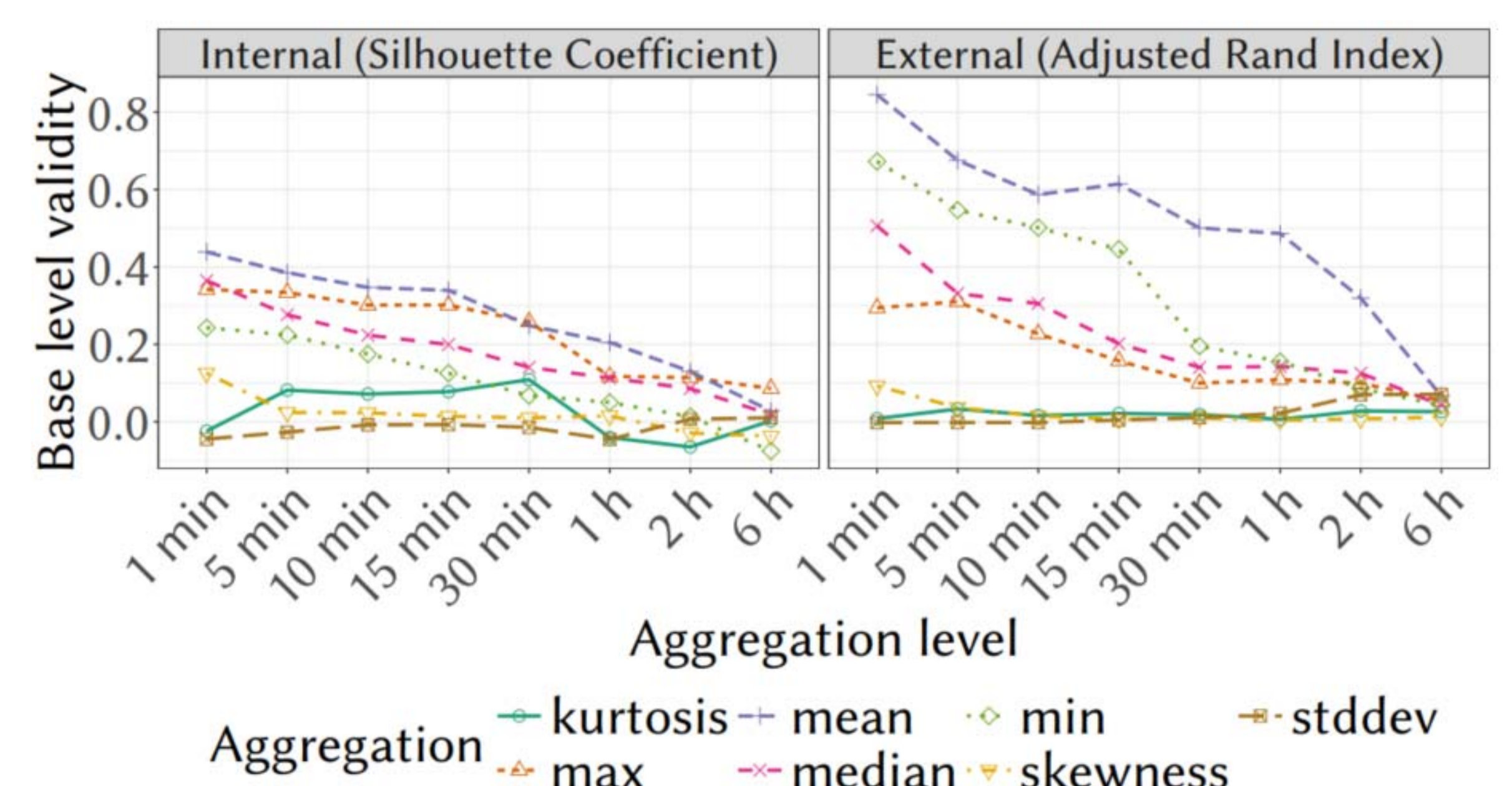
Evaluation

- Clustering Structure:** distribution of cluster sizes
- Internal Validity:** quality of clustering
- External Validity:** comparison to target assignment
- Forecasting:** error with clustering as pre-processing



Preliminary Results

- Silhouette Coefficient:** higher values indicate higher clustering quality
- Adjusted Rand Index:** high value if cluster assignment for aggregated and unaggregated data are similar
- Result:** location statistics yield best quality, decreasing trend with increasing aggregation



Future Work

- In-depth analysis of various evaluation metrics
- Study of several electrical quantities
- Guidelines how to evaluate aggregation for domain experts

Acknowledgments: This work was supported by the German Research Foundation (DFG) as part of the Research Training Group GRK 2153: Energy Status Data – Informatics Methods for its Collection, Analysis and Exploitation.