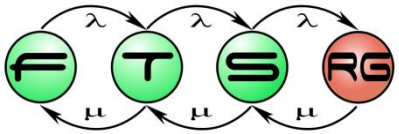
Analysis of Multi-dimensional data

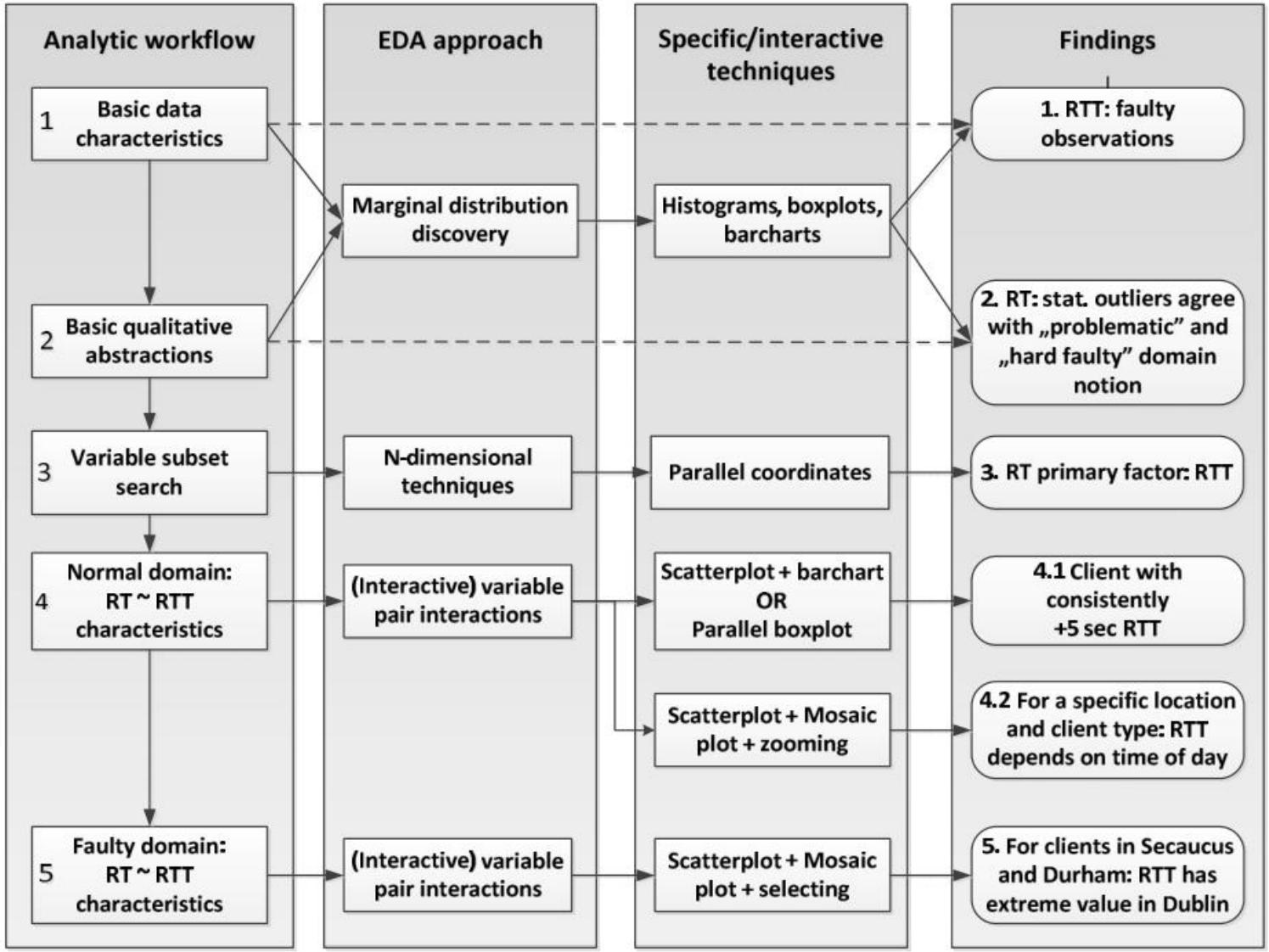
Visual analysis of measurement data

10/17/2019

Budapest University of Technology and Economics
Fault Tolerant Systems Research Group

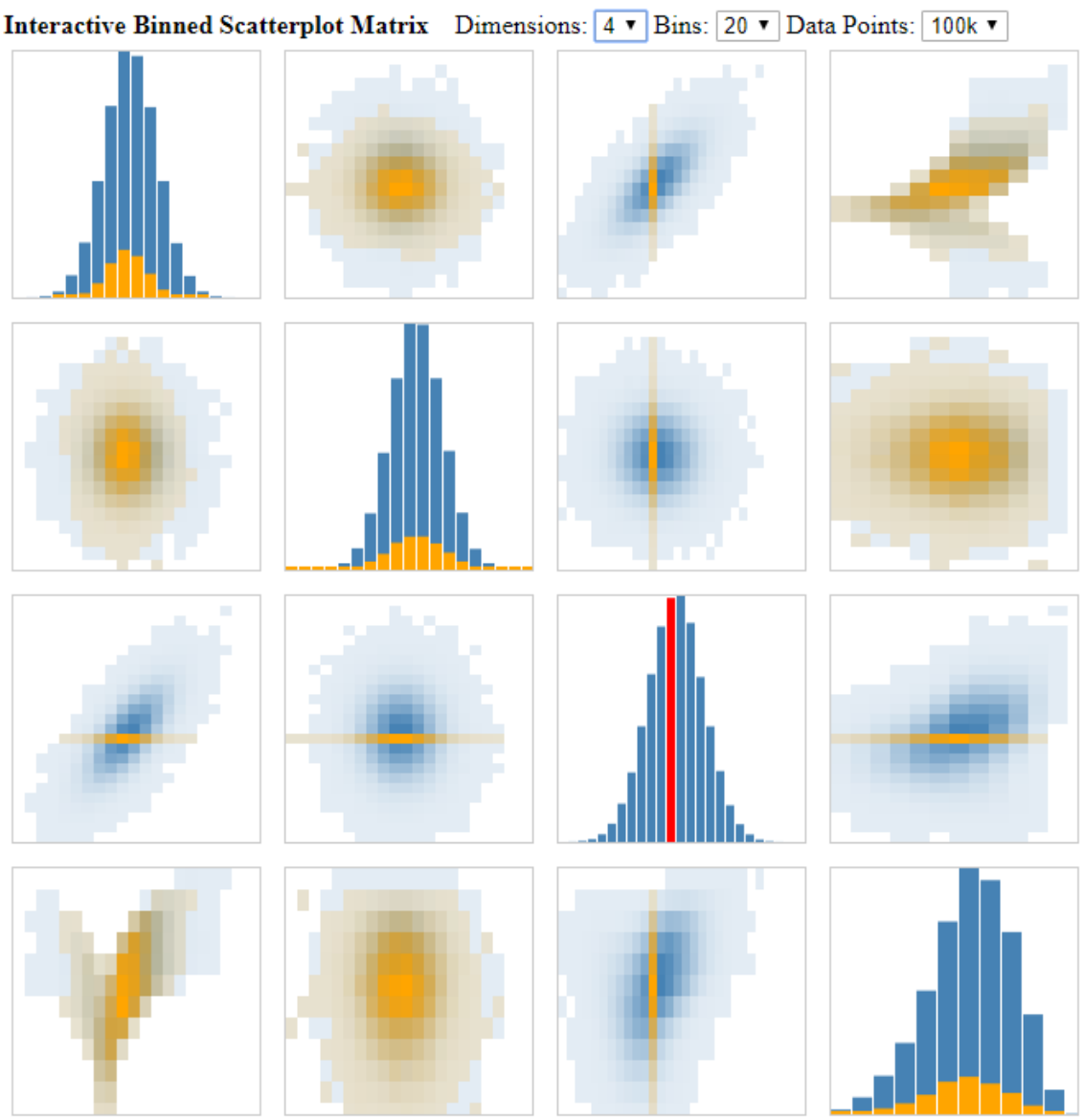


Generalized „workflow”



Patariczа, András, Imre Kocsis, Ágnes Salánki, and László Gönczy. "Empirical Assessment of Resilience." In *International Workshop on Software Engineering for Resilient Systems*, pp. 1-16. Springer, Berlin, Heidelberg, 2013.

Pairwise analysis



http://vis.stanford.edu/projects/datavore/splom/

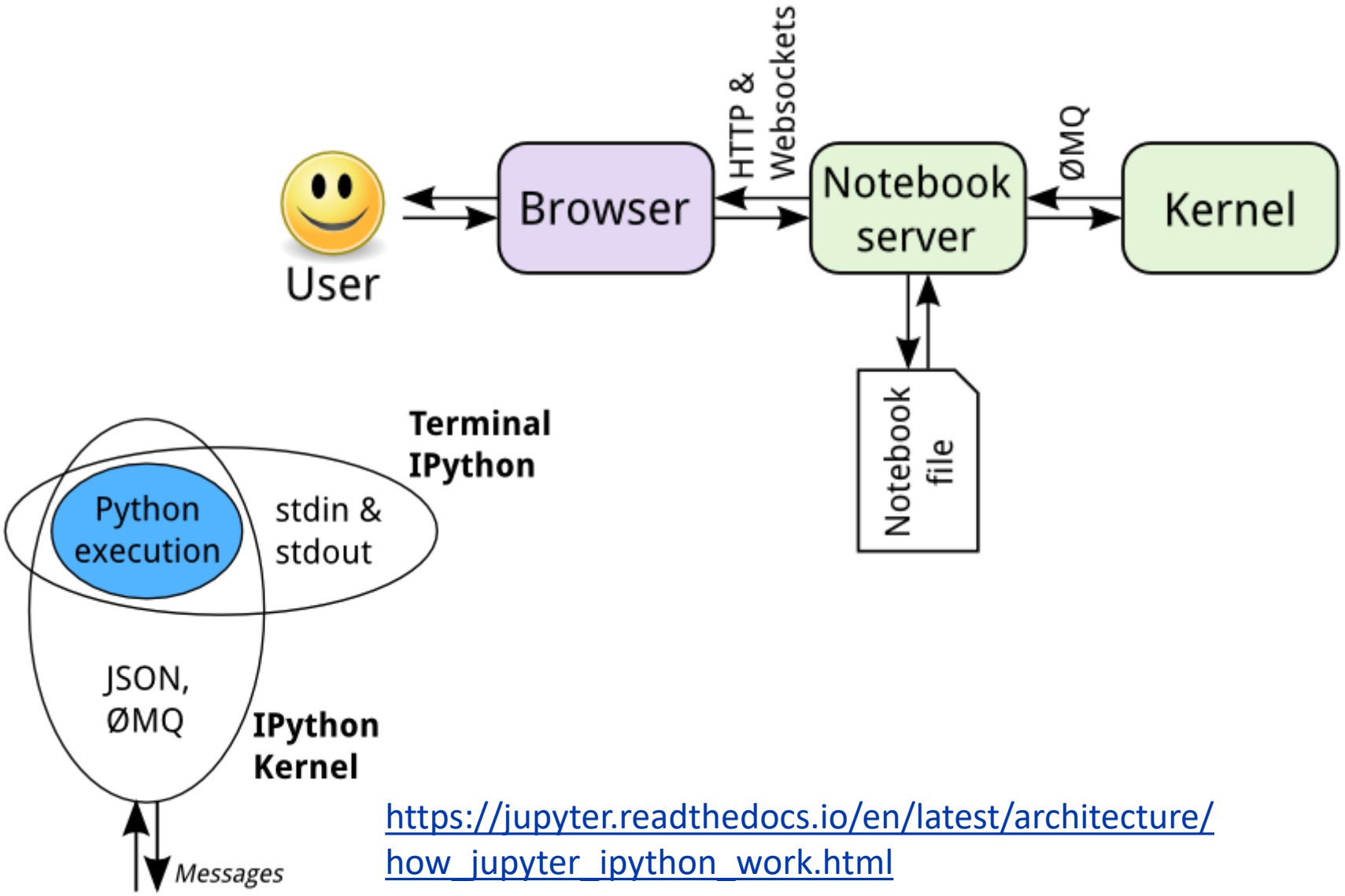
JUPYTER NOTEBOOK EXAMPLE

Effective multi-dimensional data visualization

https://github.com/dipanjanS/practical-machine-learning-with-python/tree/master/bonus%20content/effective%20data%20visualization

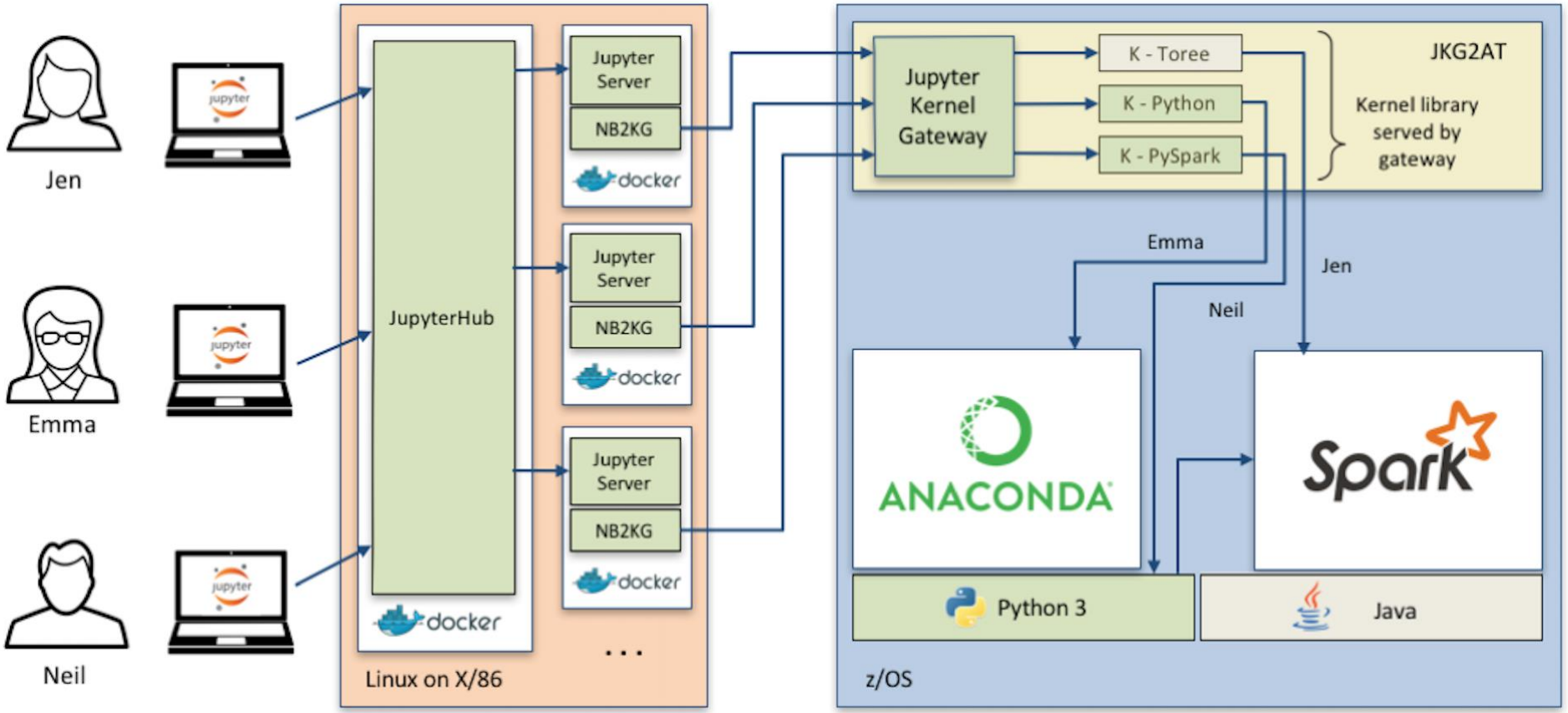
https://towardsdatascience.com/the-art-of-effective-visualization-of-multi-dimensional-data-6c7202990c57

Jupyter Notebooks



https://jupyter.readthedocs.io/en/latest/architecture/how_jupyter_ipython_work.html

Scaling and enterprise usage



https://izoda.github.io/site/ecosystem/

Analysis of wine characteristics

- Cortez P, Cerdeira A, Almeida F, Matos T, Reis J. Modeling wine preferences by data mining from physicochemical properties. Decision Support Systems. 2009 Nov 1;47(4):547-53.

EXAMPLE: VISUAL ANALYSIS OF ENERGY CONSUMPTION

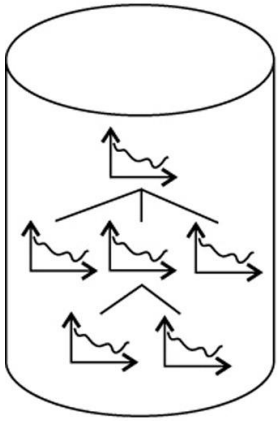
Janetzko, H., Stoffel, F., Mittelstädt, S. and Keim, D.A., 2014. Anomaly detection for visual analytics of power consumption data. *Computers & Graphics*, *38*, pp.27-37.

https://www.sciencedirect.com/science/article/pii/S0097849313001477

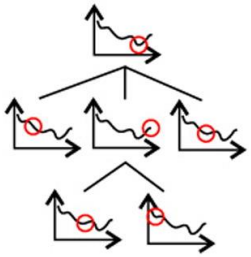
Visual identificaton of anomalies

- Goals
  - Examine behavior
  - Find anomalies
- Approach
  - Hierarchical consumption model
  - Predict values
  - Compare with measured data

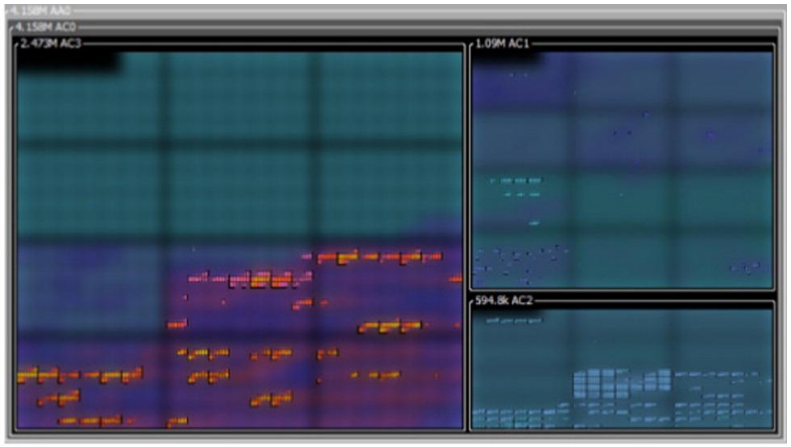
Main steps



Time series



Anomaly detection



Pixel-based time series & anomaly visualization

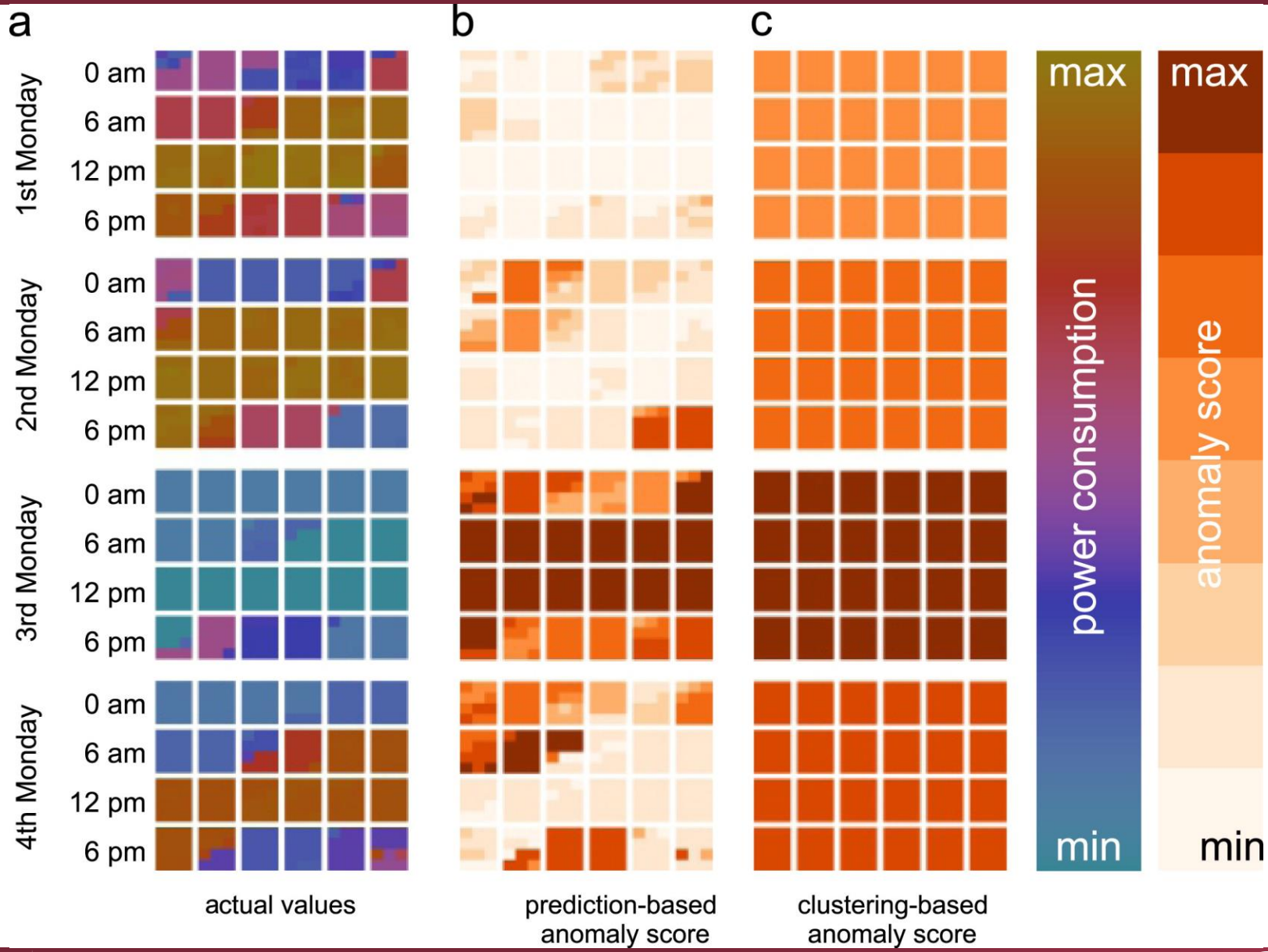
Anomalies

- Prediction-based approach
  - Weighted prediction for values

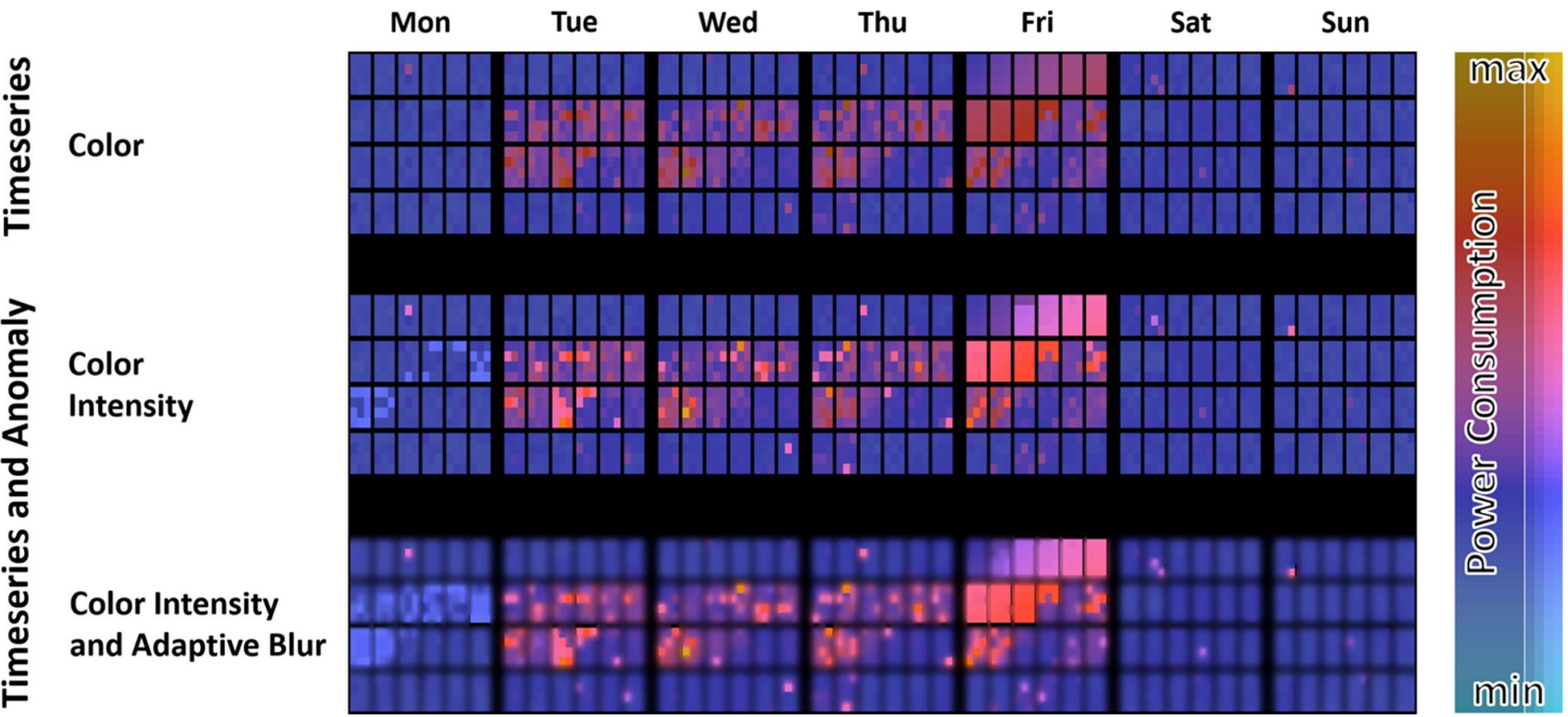
$$anomaly[time] = \frac{|predVal[time] - obsVal[time]|}{avg_{t \in Time}(|predVal[t] - obsVal[t]|)}$$

  - Is the consumption close to predicted?
- Clustering-based methods
  - Daily data transformed to a multi-dimensional representation
  - „Density” of points is considered
  - Are there similar days?

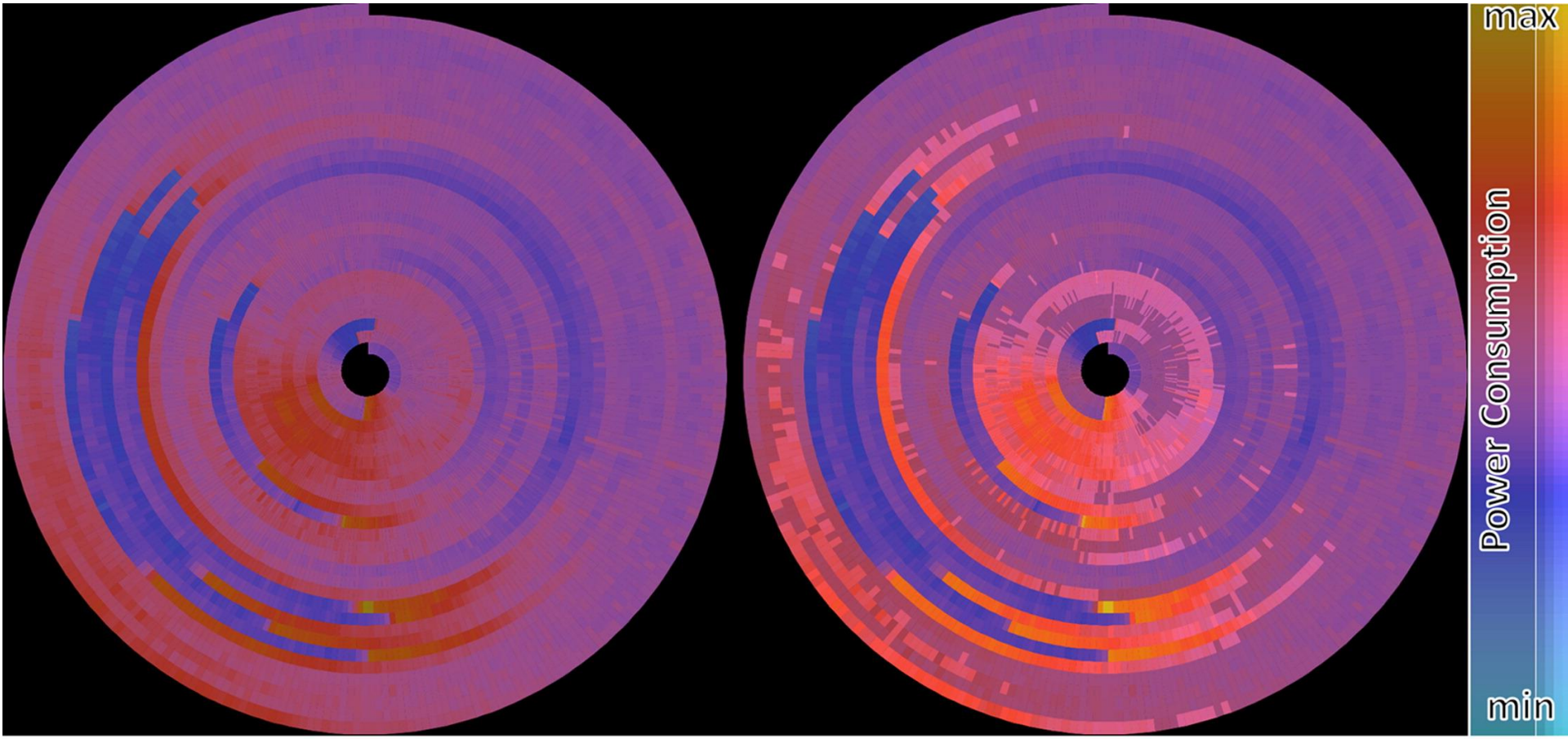
Visual comparison of methods



Recursive patterns



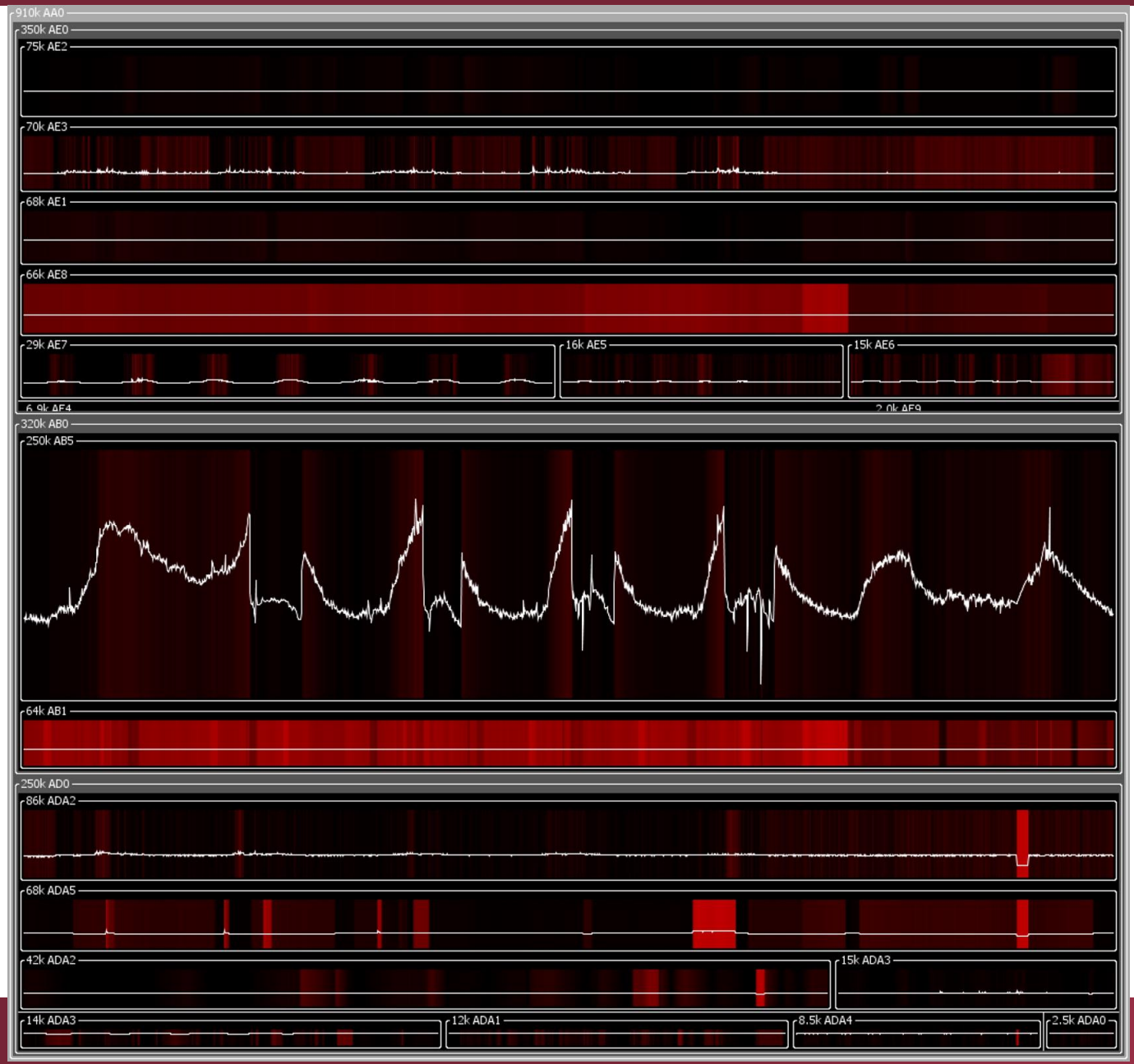
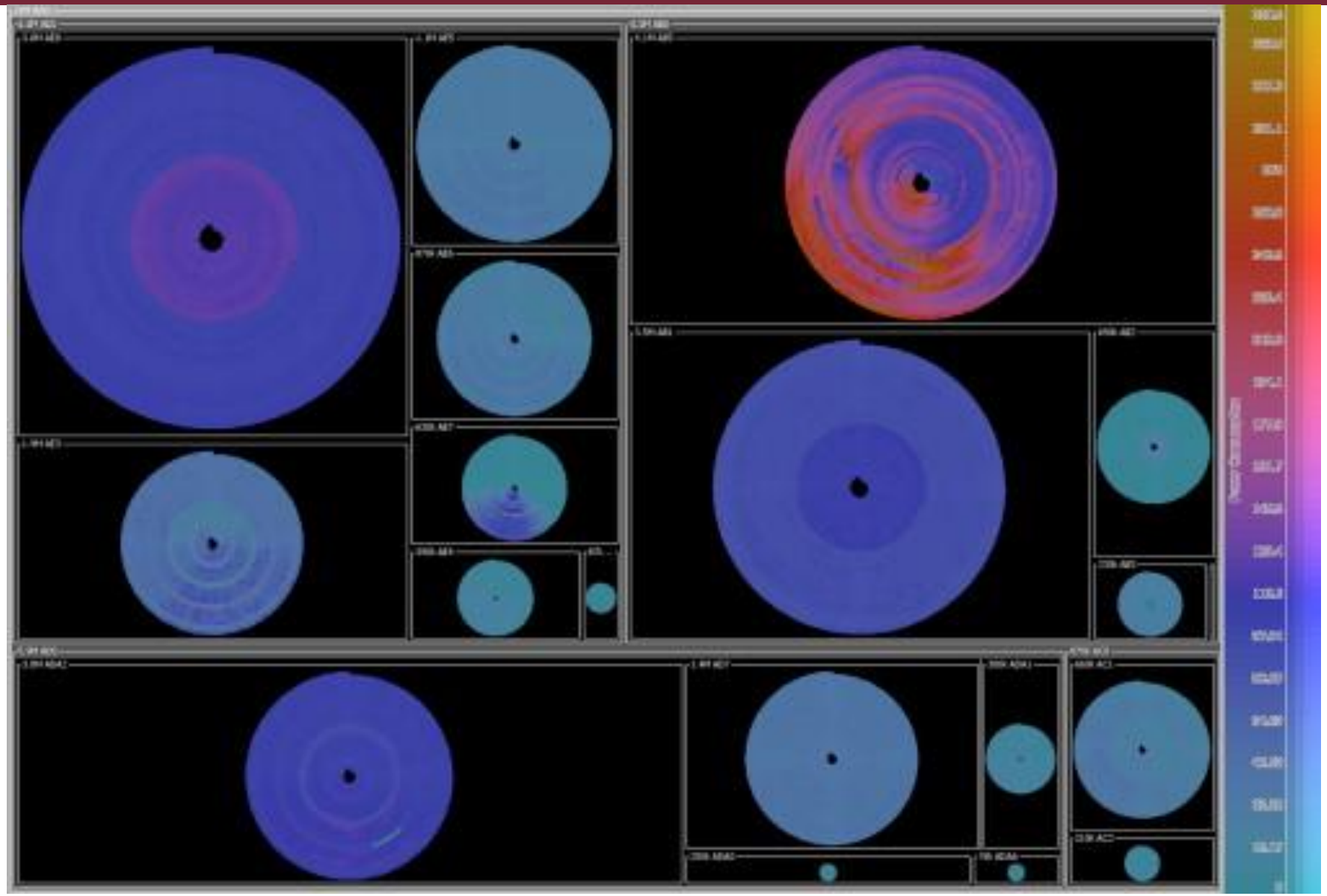
Spiral encoding



Timeseries Color Encoding

Timeseries and Anomaly Score Intensity Encoding

Value + anomaly: extended line chart



Hierarchy + visualization