
Temporal Data Mining for Sustainable Data Centers

Naren Ramakrishnan
Department of Computer Science
Virginia Tech
naren@cs.vt.edu

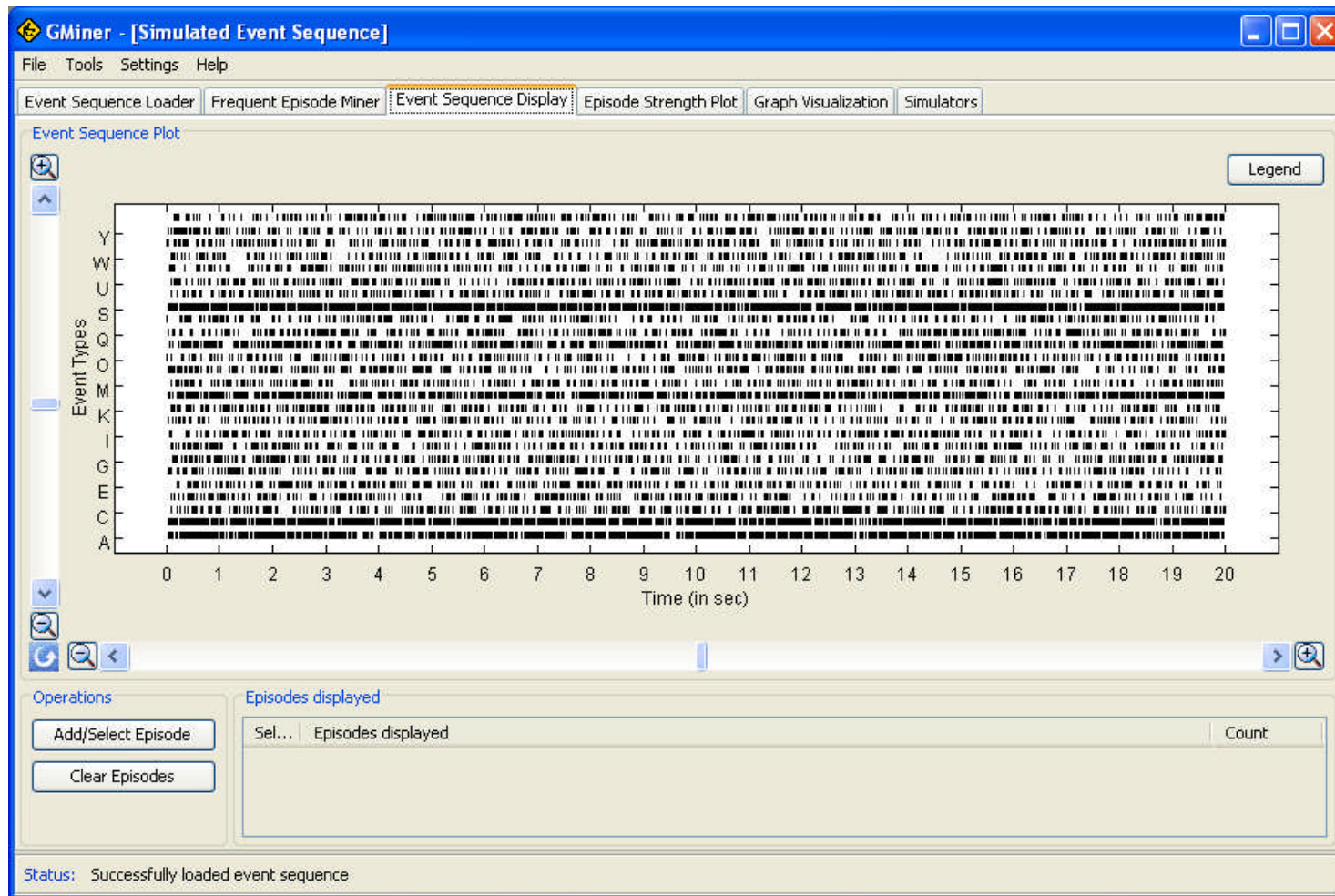
Central Themes

- Temporal data mining
 - Detecting relationships between physical variables in a time-series dataset
 - Network reconstruction
 - Summarizing the above relationships in the form of a graphical model
-

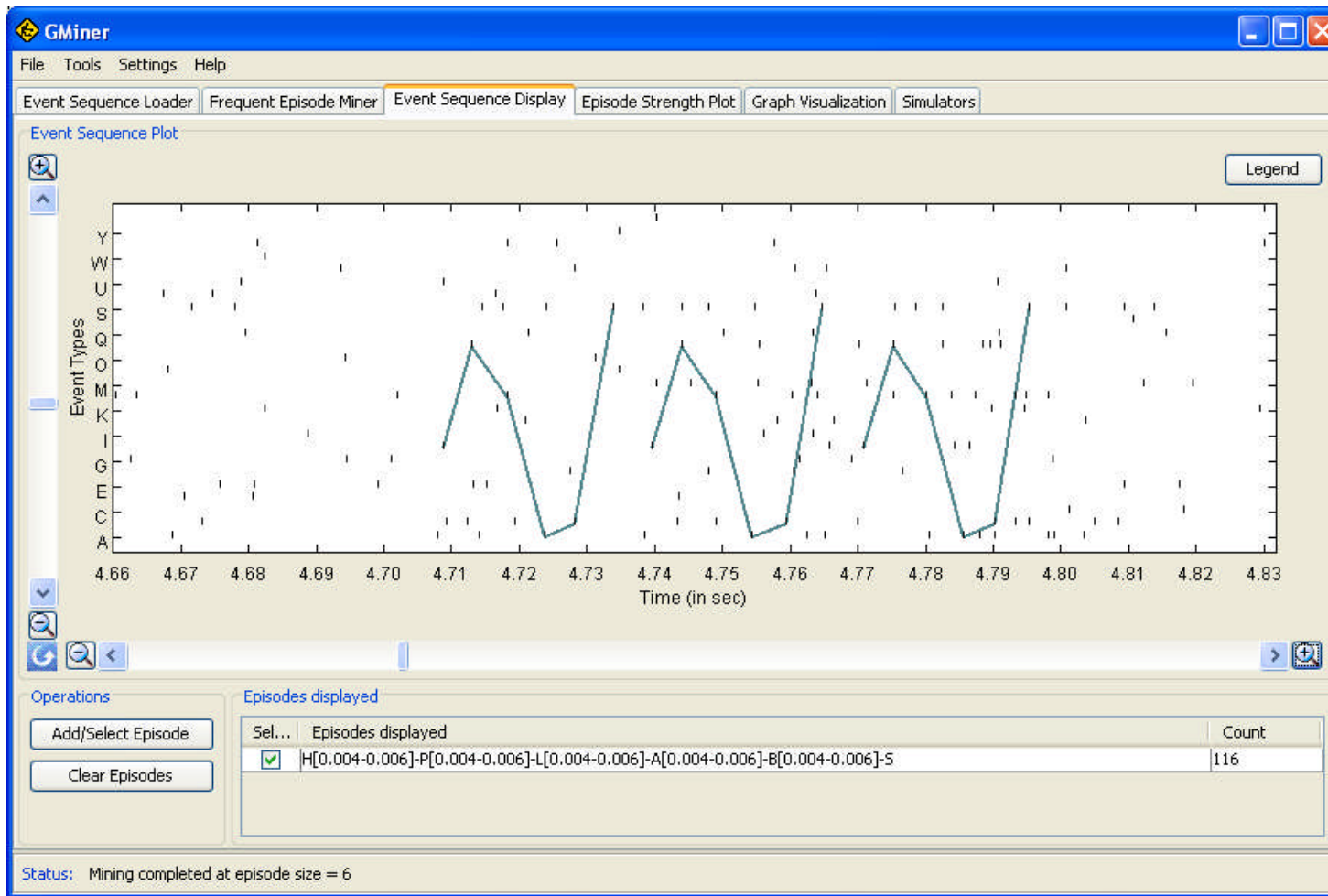
Why is this problem non-trivial?

- Patterns are spread over time
 - with possibly “junk” in between
 - Sensors dynamically cluster
 - around “interesting” time points
 - Raise the level of abstraction
 - to relate mined patterns to underlying system infrastructure
-

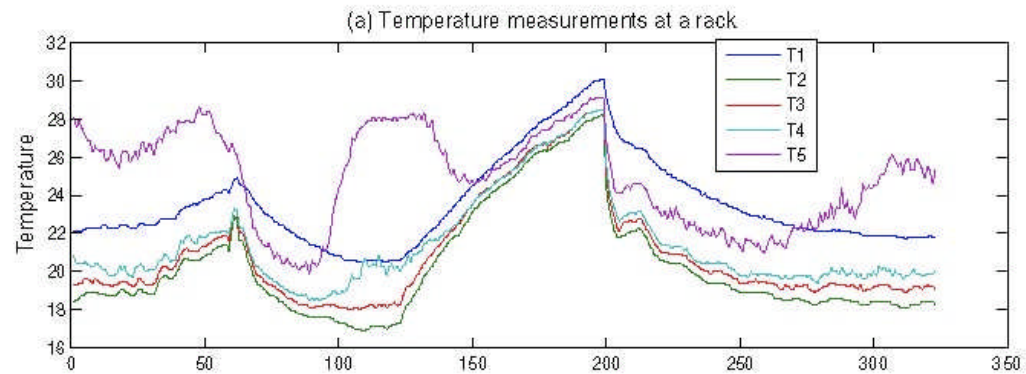
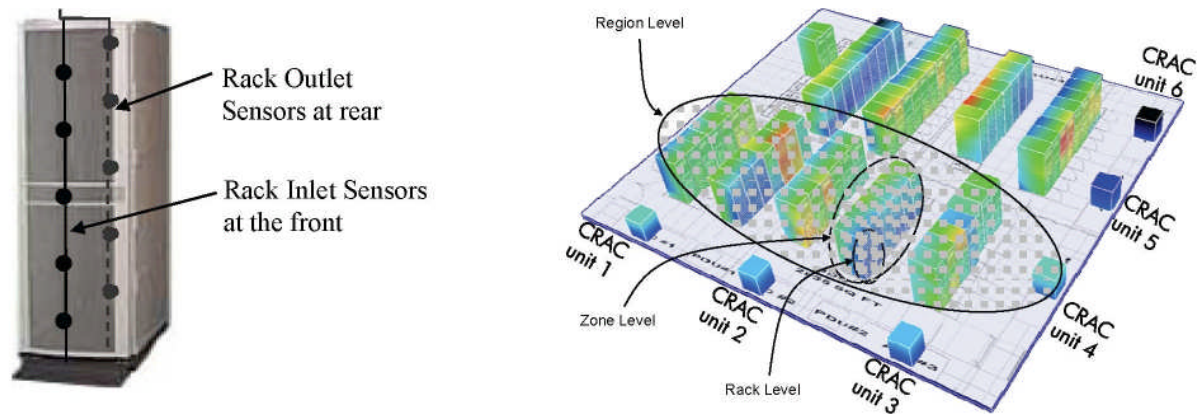
Planting and mining episodes



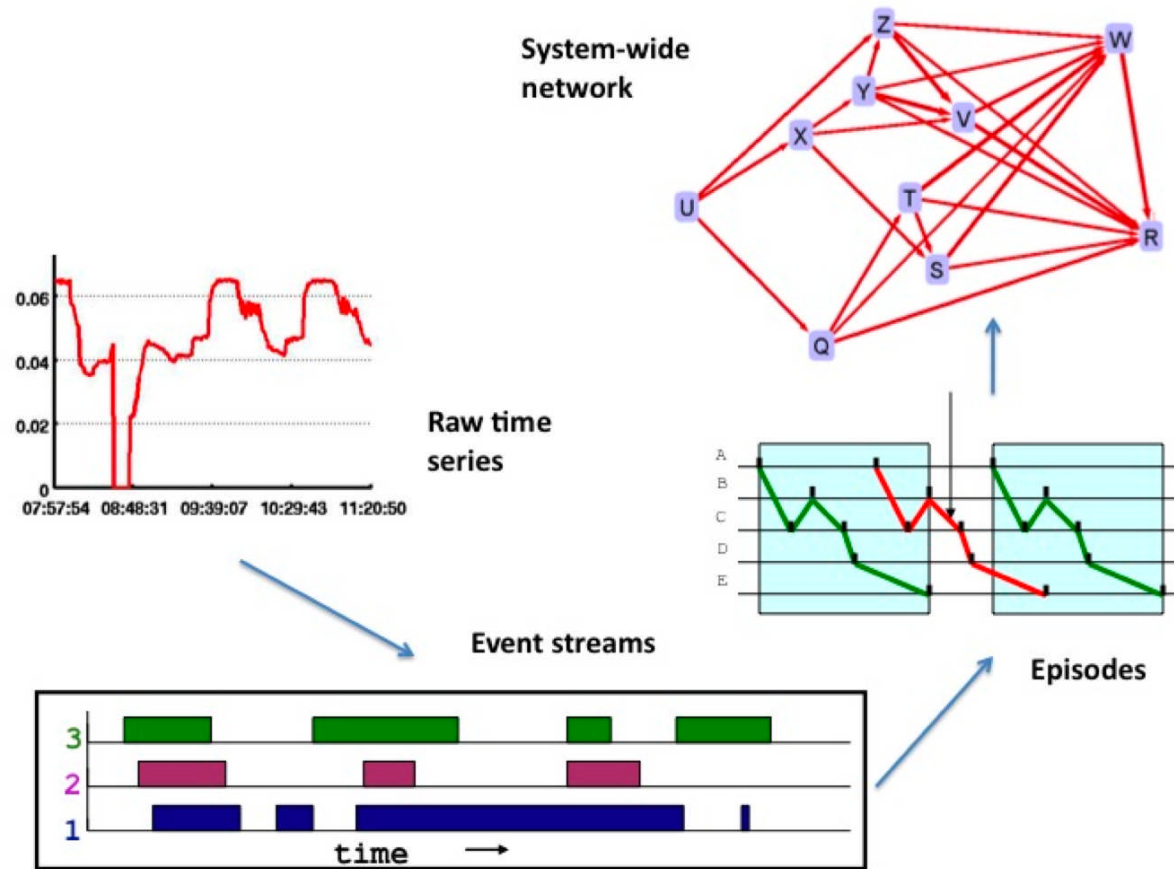
Planting and mining episodes



Data center architecture and event streams



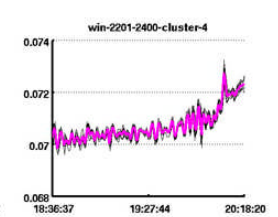
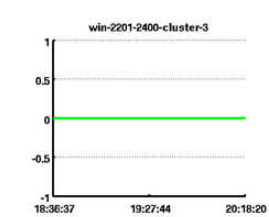
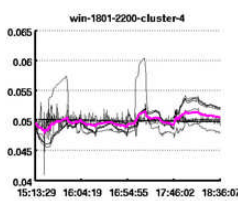
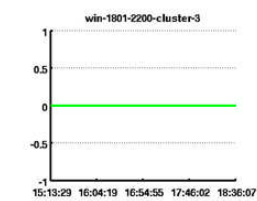
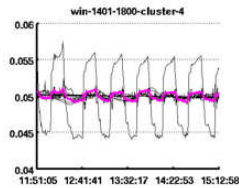
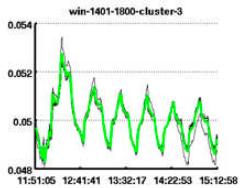
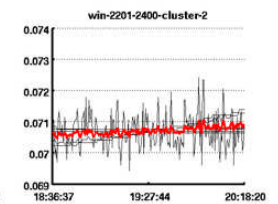
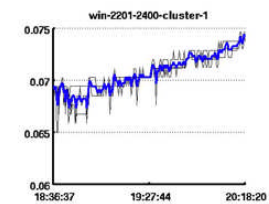
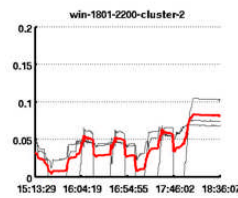
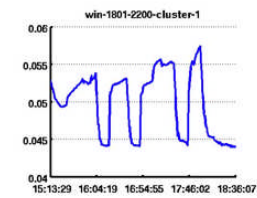
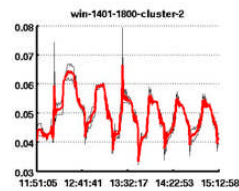
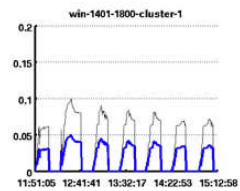
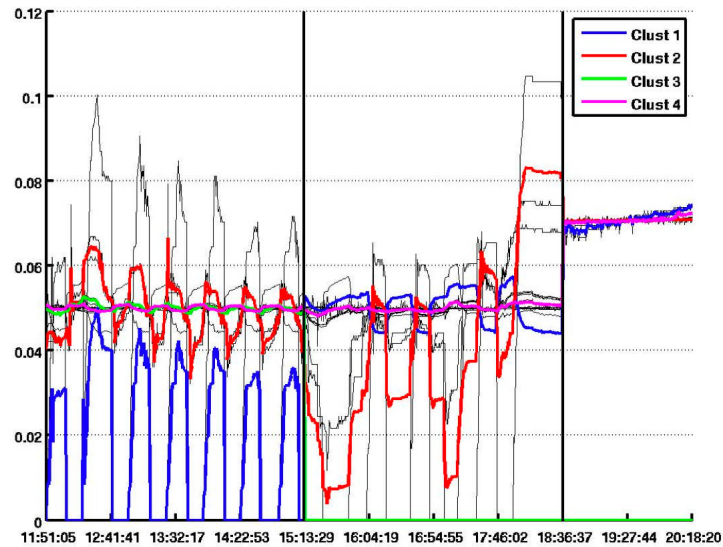
Multi-level modeling



But ...

- Sensor streams are not quite stationary
 - Workload changes
 - Dynamic steering and control
 - Faults
 - Need to identify segments around which qualitative behavior changes
-

Detecting shifts in clustering across time segments



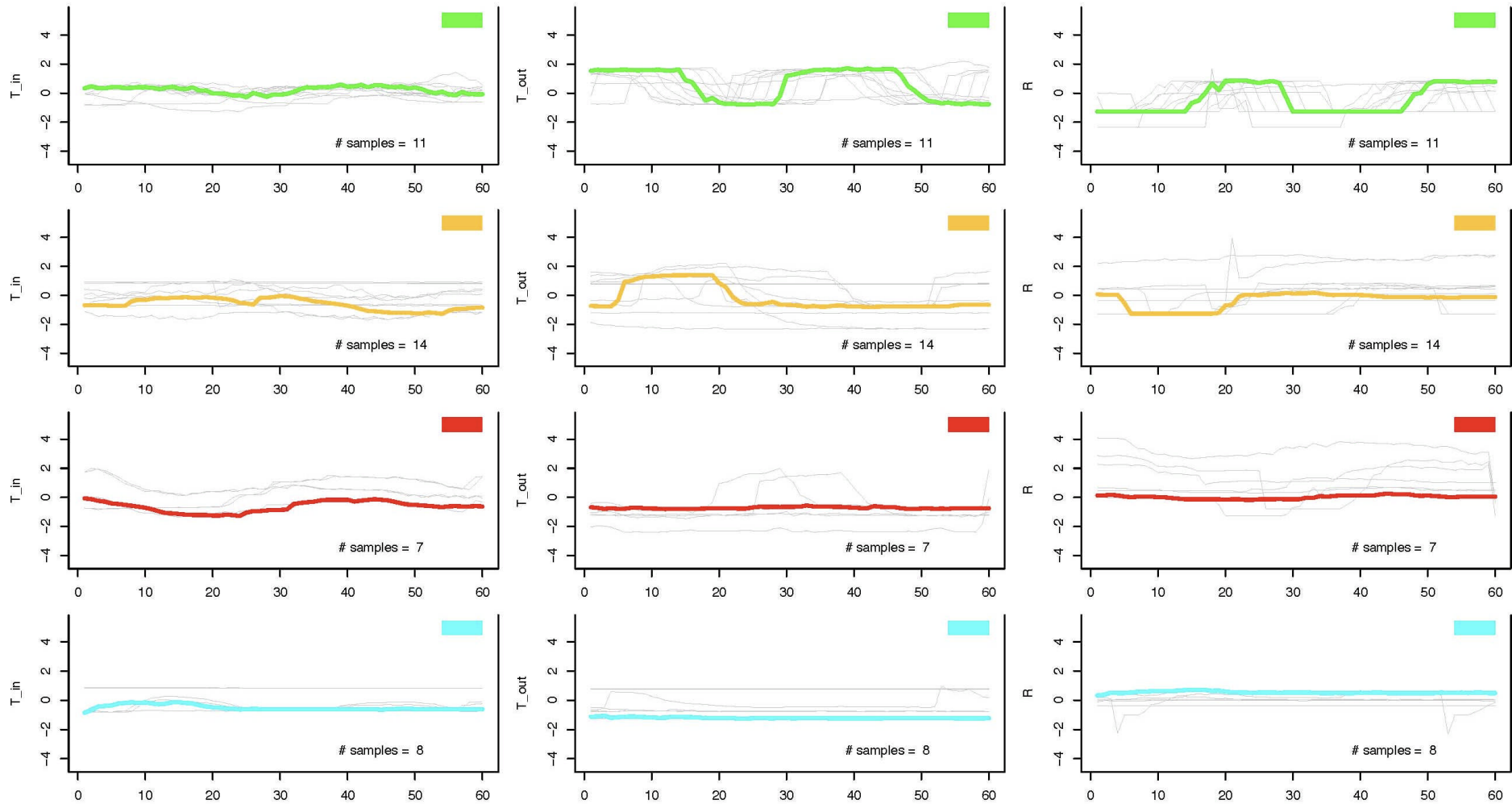
Types of patterns

- Clusterings and segmentations
 - Given: continuous time-series
 - Find: clusters that dynamically re-group
 - Frequent episodes
 - Given: event stream <event, time, dur>
 - Find: e1[0-5] -> e3[0-2] -> e7[0-10]
 - Dynamic temporal redescription
 - Integrated methodology for bridging multiple levels of representation
-

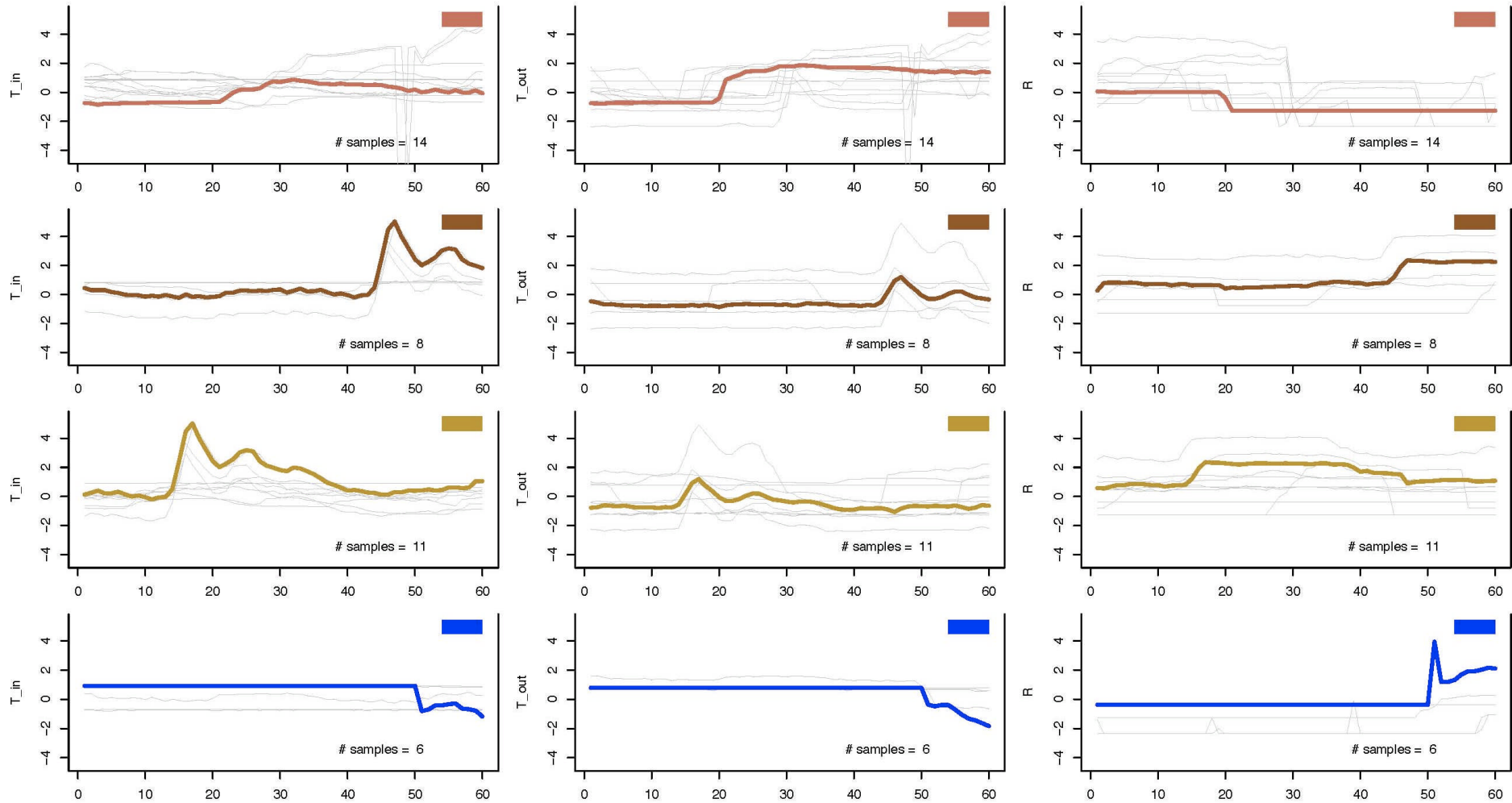
Integrated methodology

- Use domain knowledge to group sensor variables into “units”
 - {T1,T2,R1}, {T3,T4, R2}
 - Breakup event streams into units
 - Detect coordinated trends across units using k-medoid clustering
 - Redescribe original data using identified clusters
-

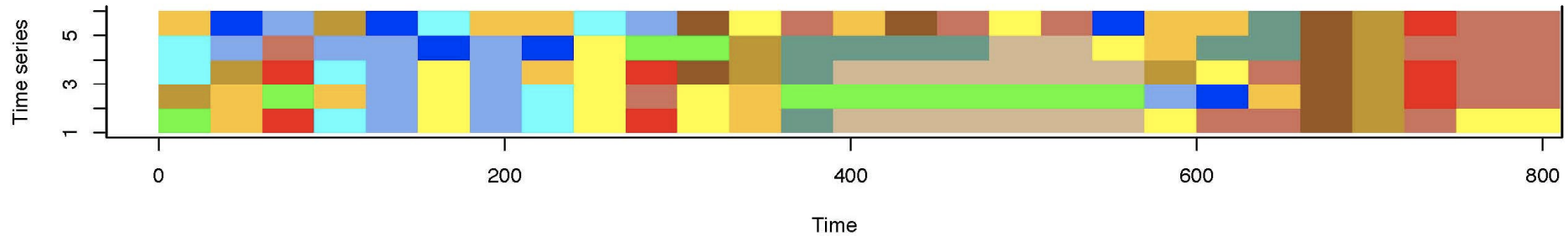
Preliminary results



Preliminary results (contd.)



Redescribing event streams



Use clusters as descriptors over time-series data

Identify trends across time

Identify trends across space

Spatio-temporal dynamic modeling

The Holy Grail

- Reconstruct the transfer function underlying key system variables
 - Operate on streaming data
 - Provide human-in-the-loop discovery capabilities
-

Thank you

- Acknowledgements
 - Manish Marwah and Ratnesh Sharma
 - Deb Patnaik
 - Contact
 - naren@cs.vt.edu
 - <http://www.cs.vt.edu/~naren>
-