

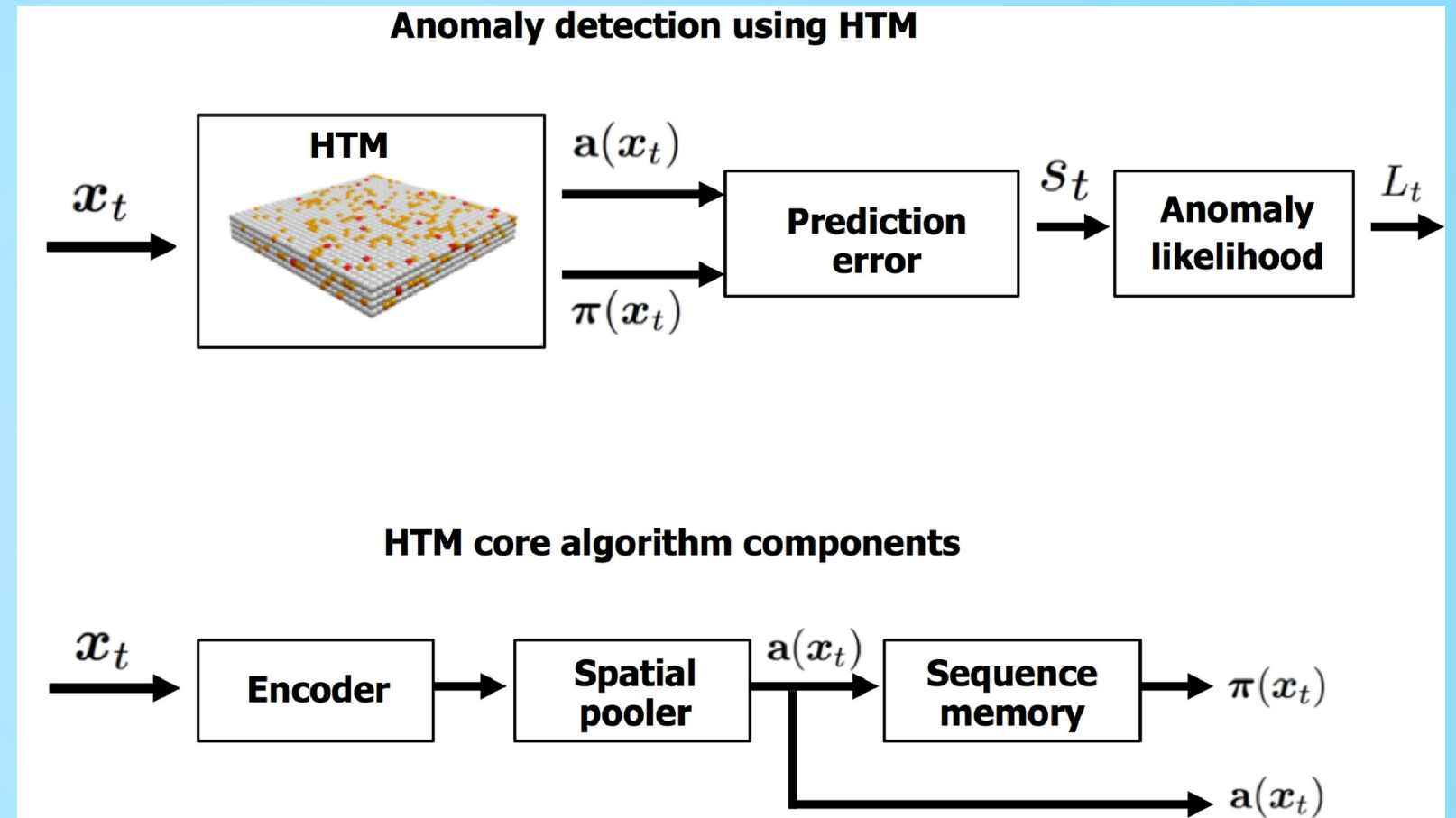
Experience Using NuPIC to Detect Anomalies in Controls Data

Ted D'Ottavio, Philip Dyer, Joseph Piacentino, Jr., Matthew Tomko

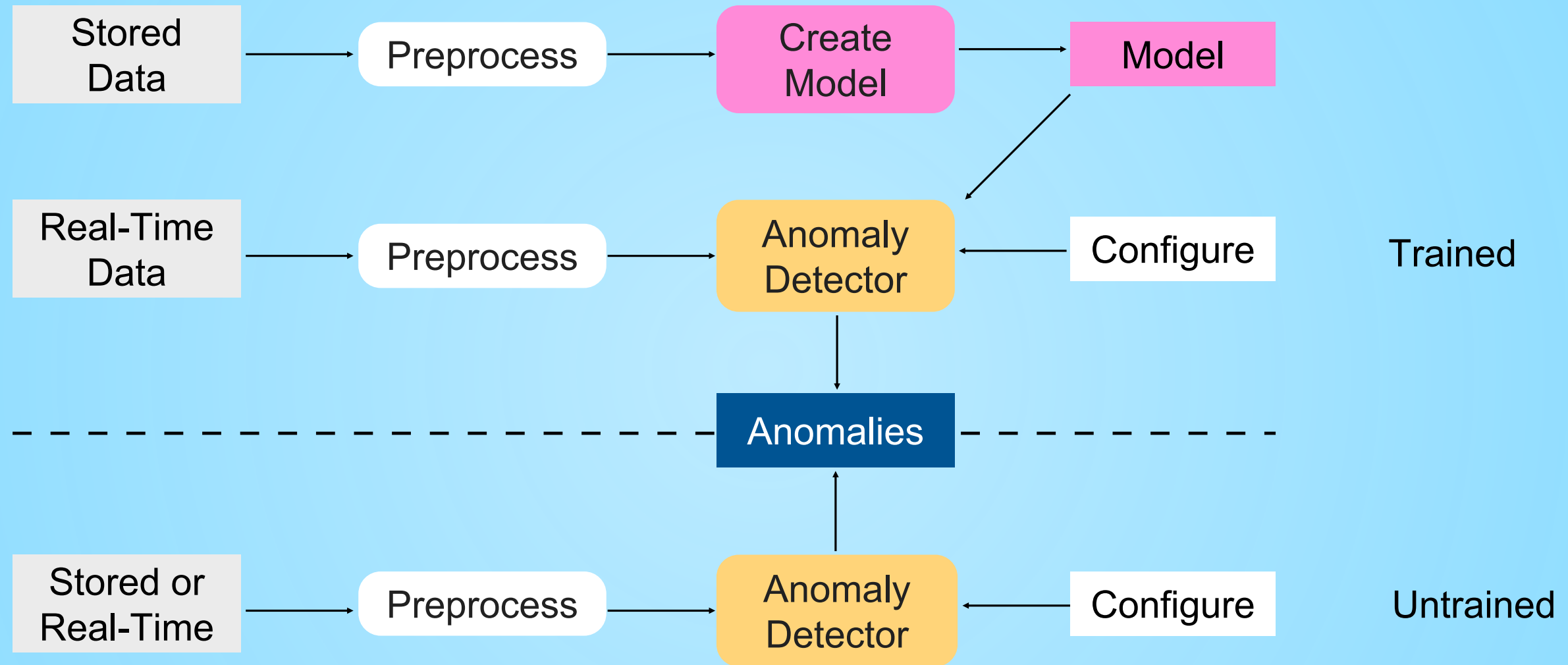
Collider-Accelerator Department
Brookhaven National Laboratory

Quick Overview of NuPIC

- NuPIC = Numenta Platform for Intelligent Computing open source on GitHub since 2013 - python, C++, java
- Numenta's goal is to reverse-engineer the neocortex and apply that knowledge to the creation of machine intelligence (2005)

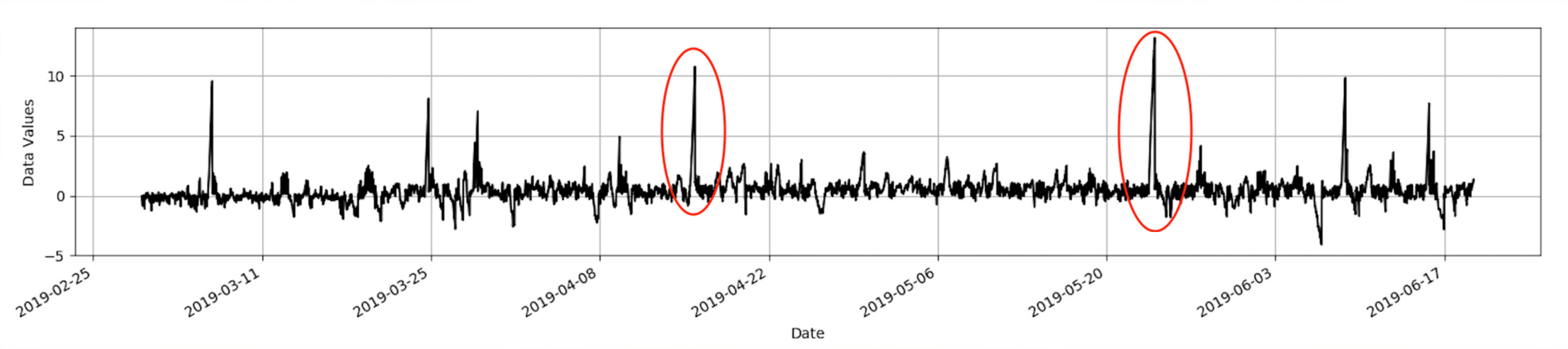


Using NuPIC for Anomaly Detection

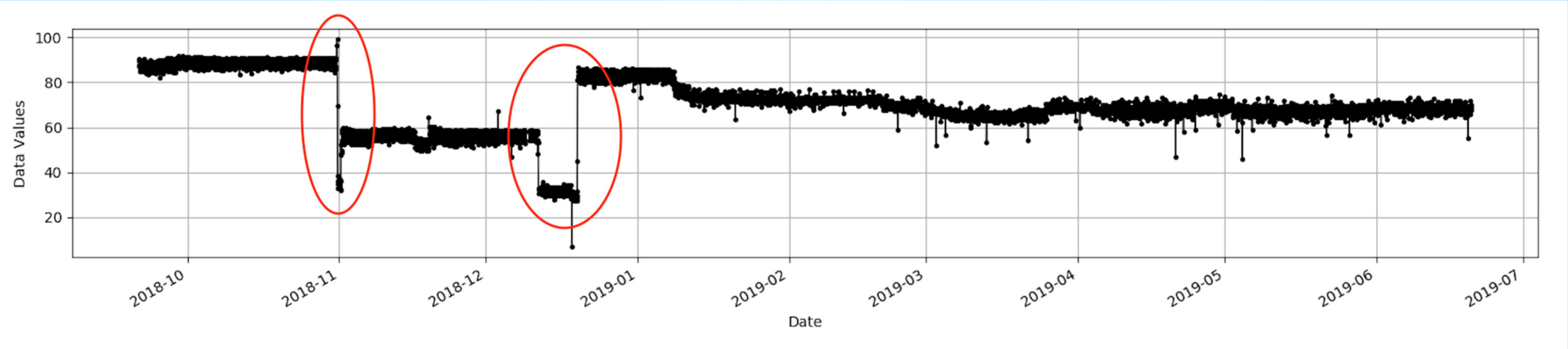


Preprocess - average noisy data, filter unwanted data
Configure - set resolution, use timestamp (y/n), learning on (y/n)

Types of Anomalies

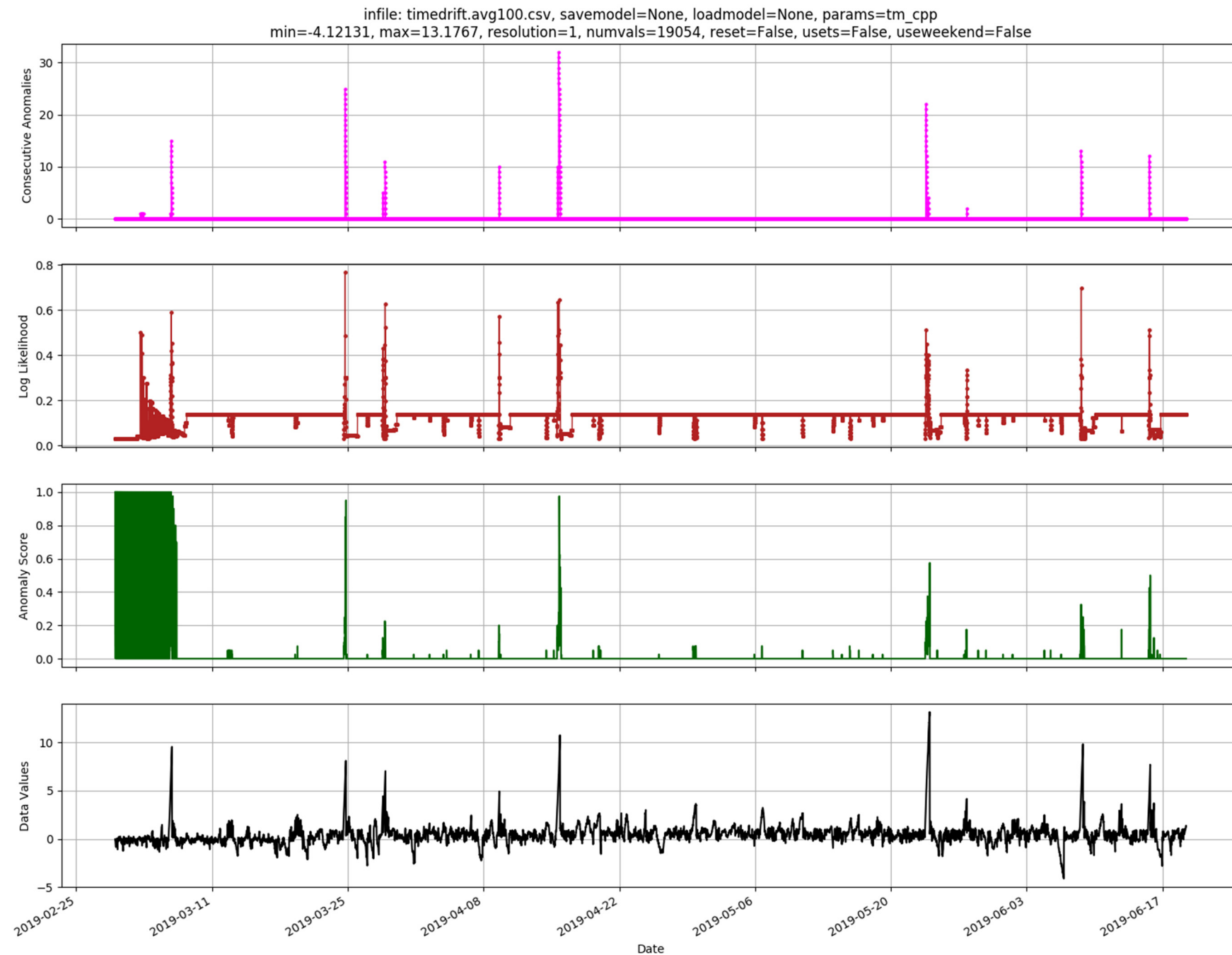


Global Outlier Anomaly

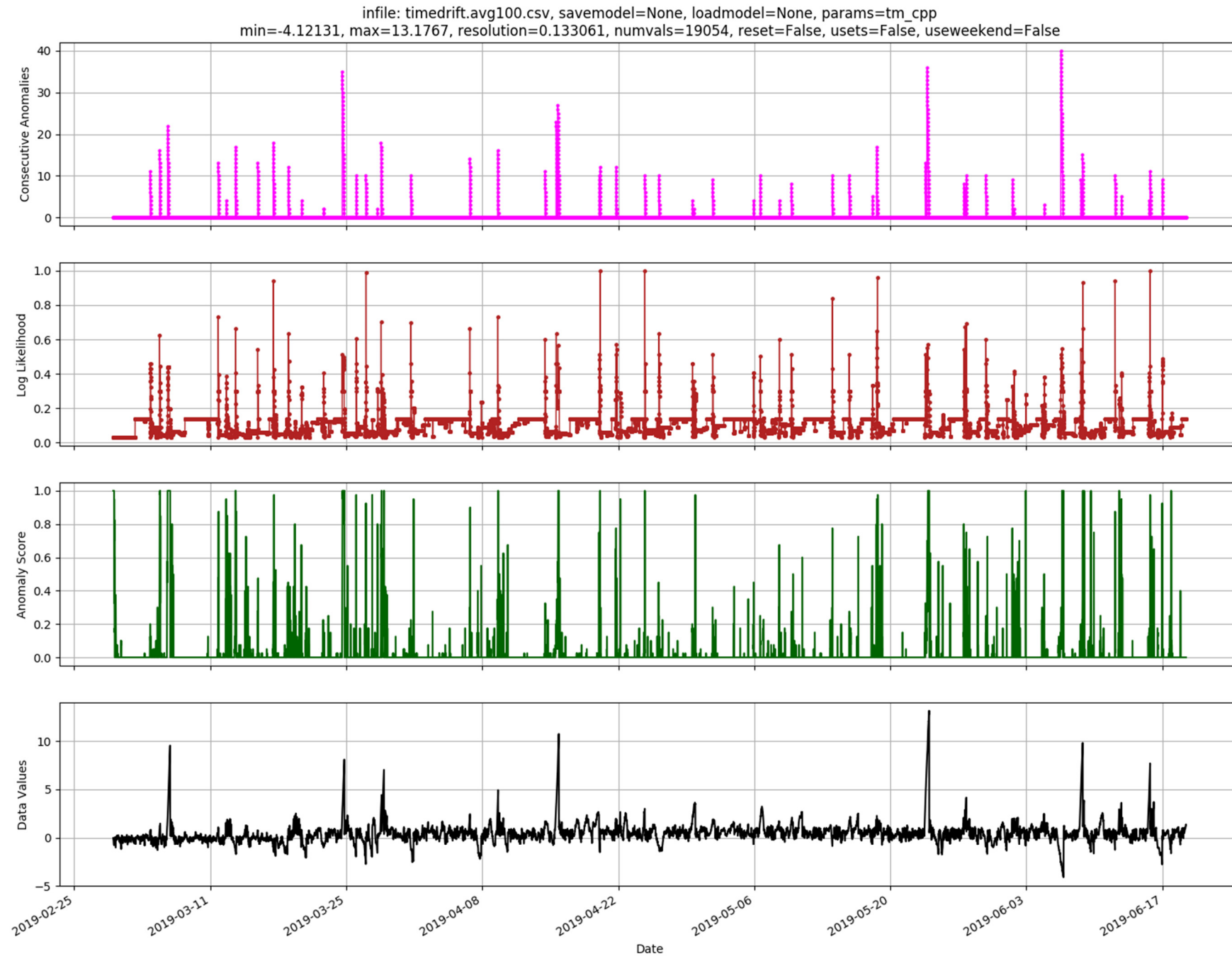


Level Change Anomaly

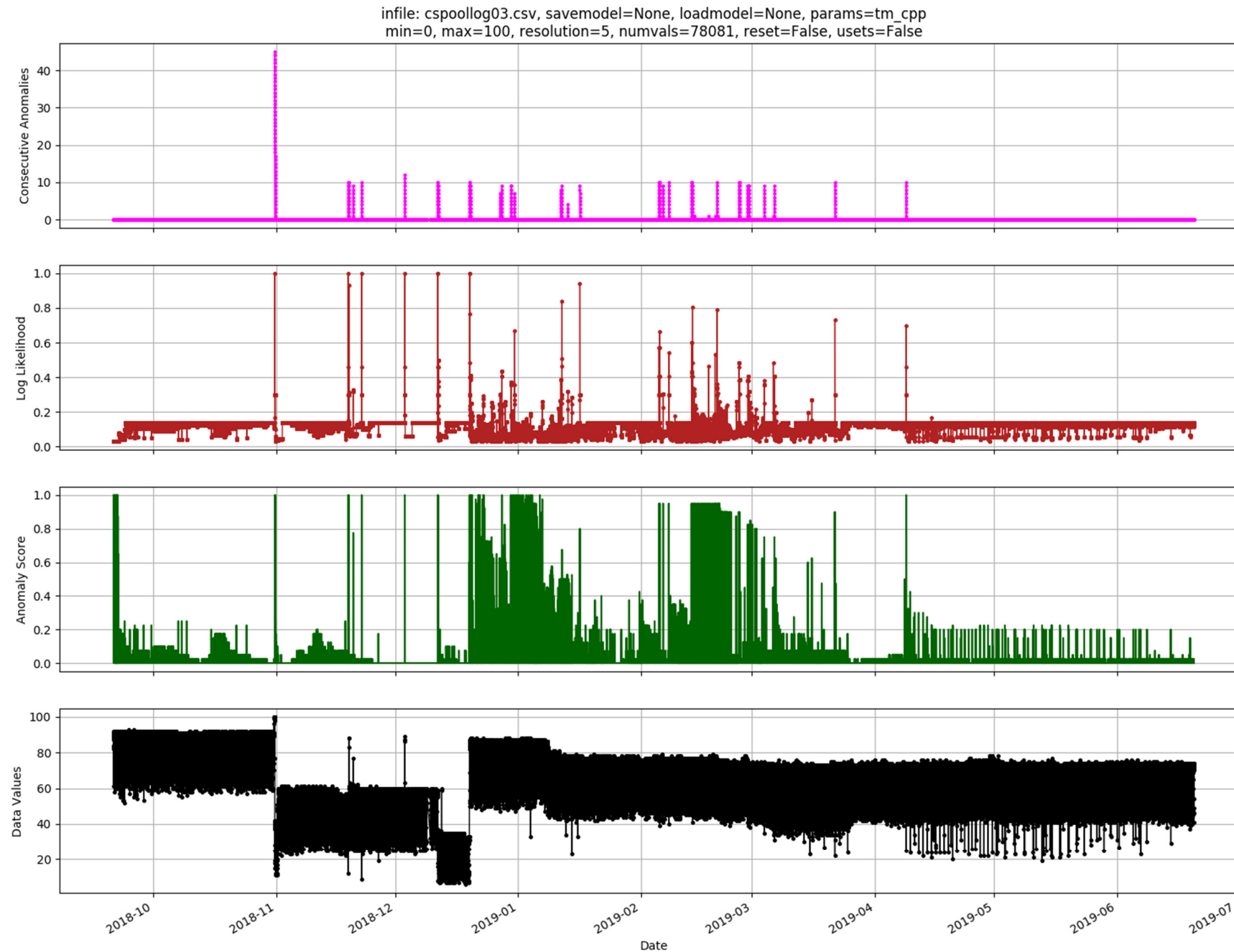
Global Outlier Anomalies



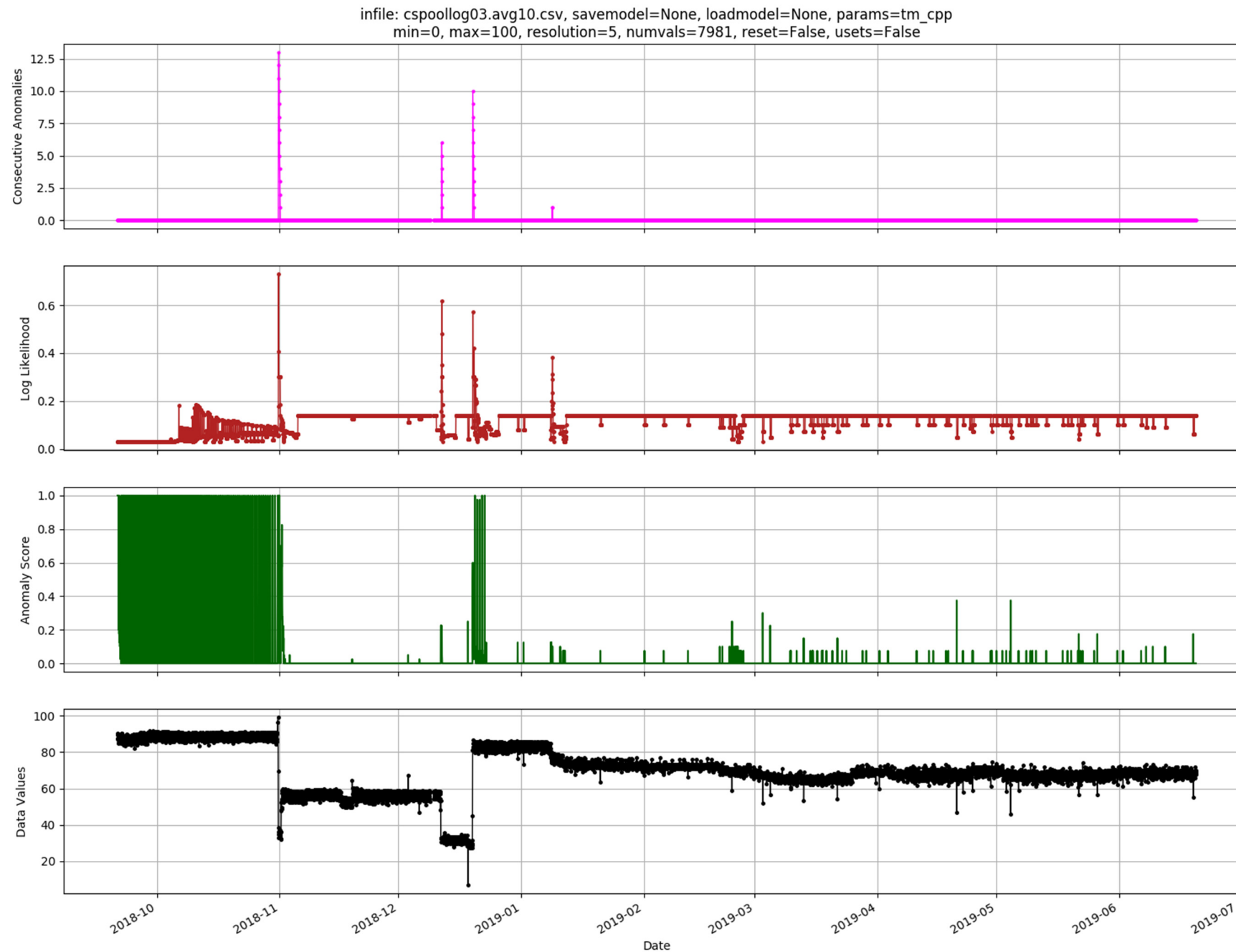
Global Outlier Anomalies cont.



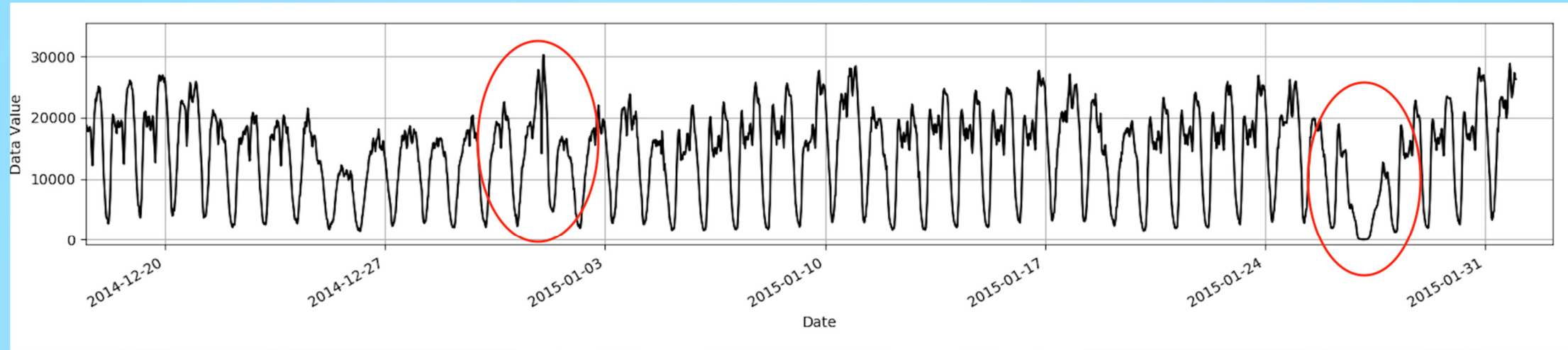
Level Change Anomalies



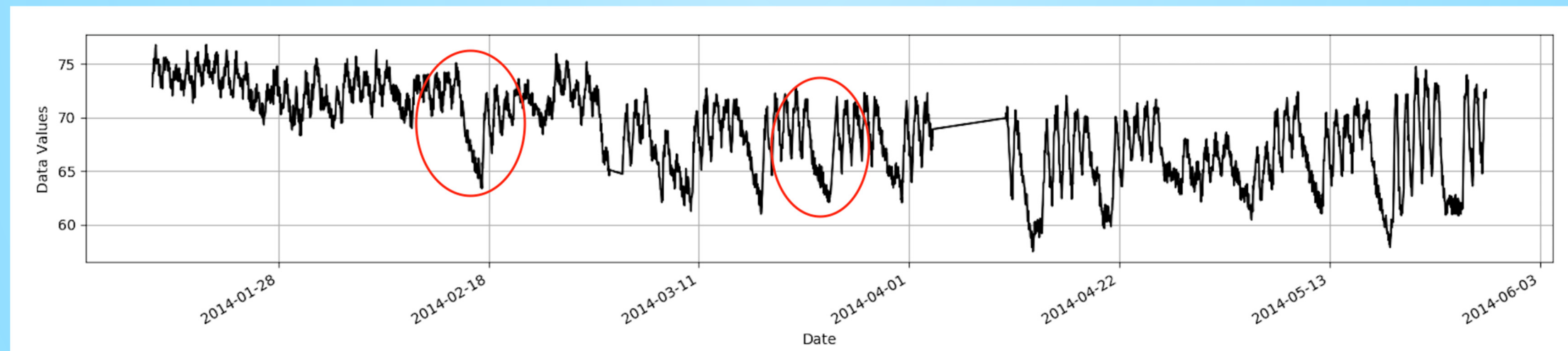
Level Change Anomalies cont.



Types of Anomalies cont.

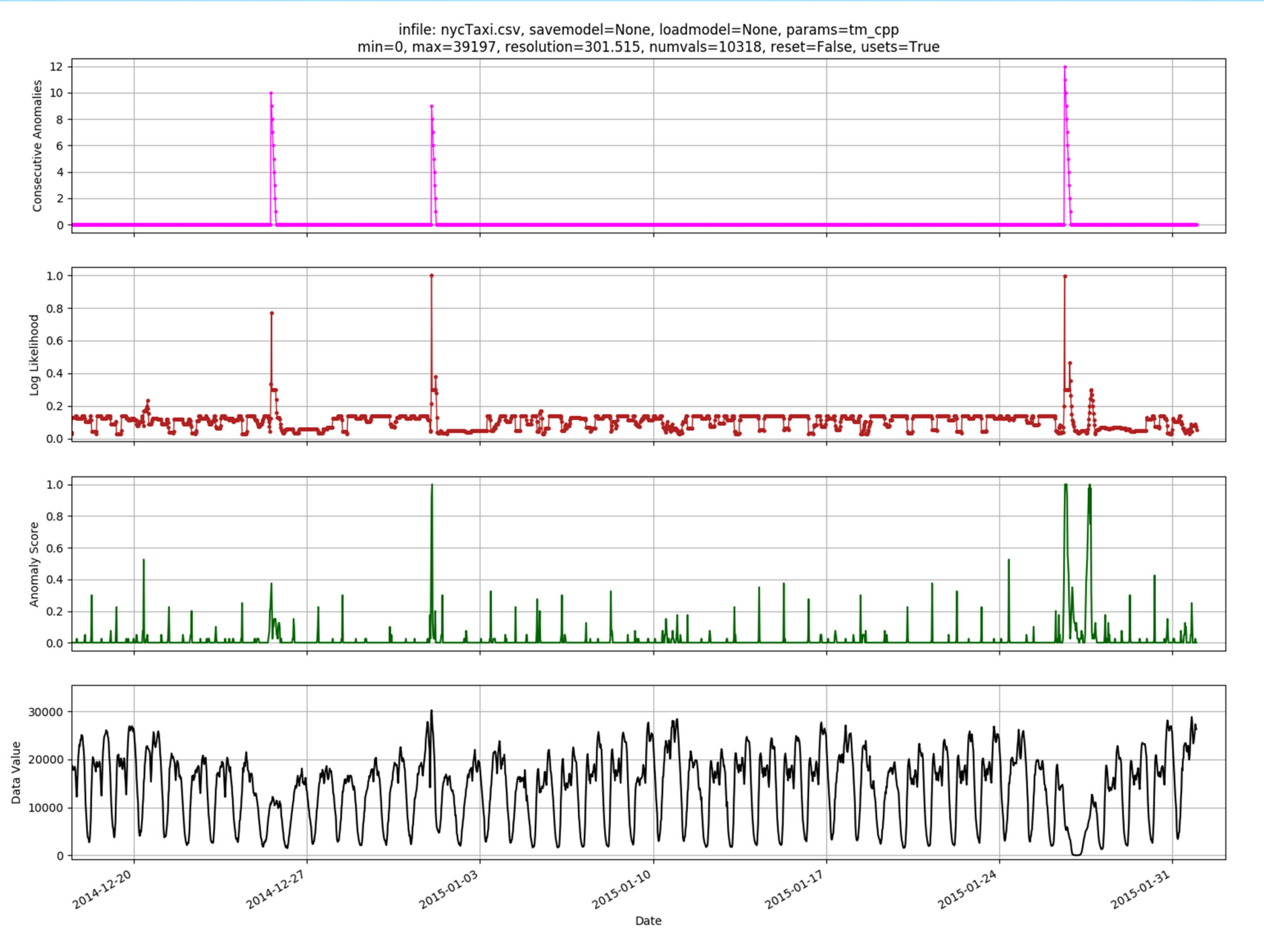


Pattern Deviation Anomaly

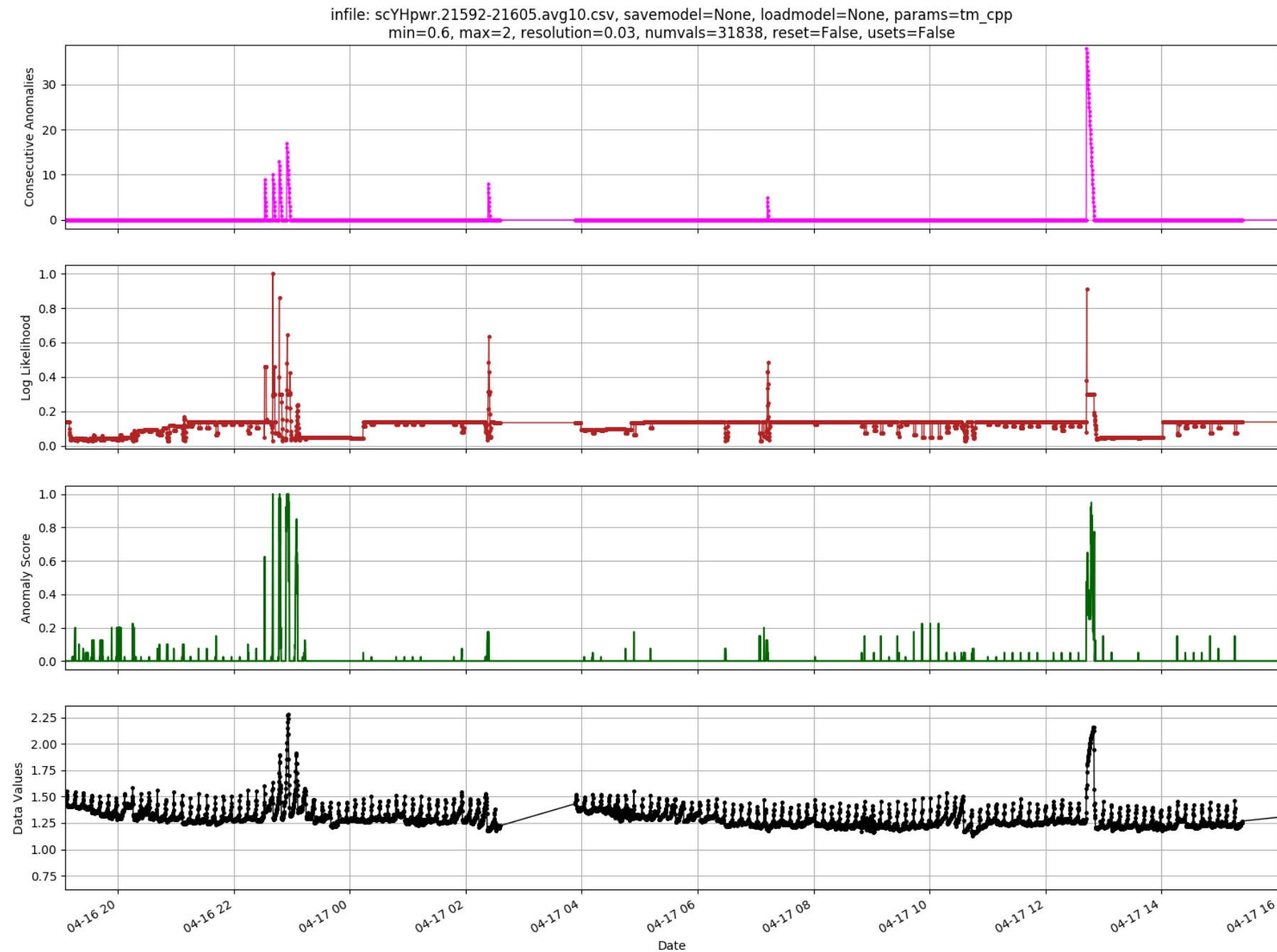


Pattern Change Anomaly

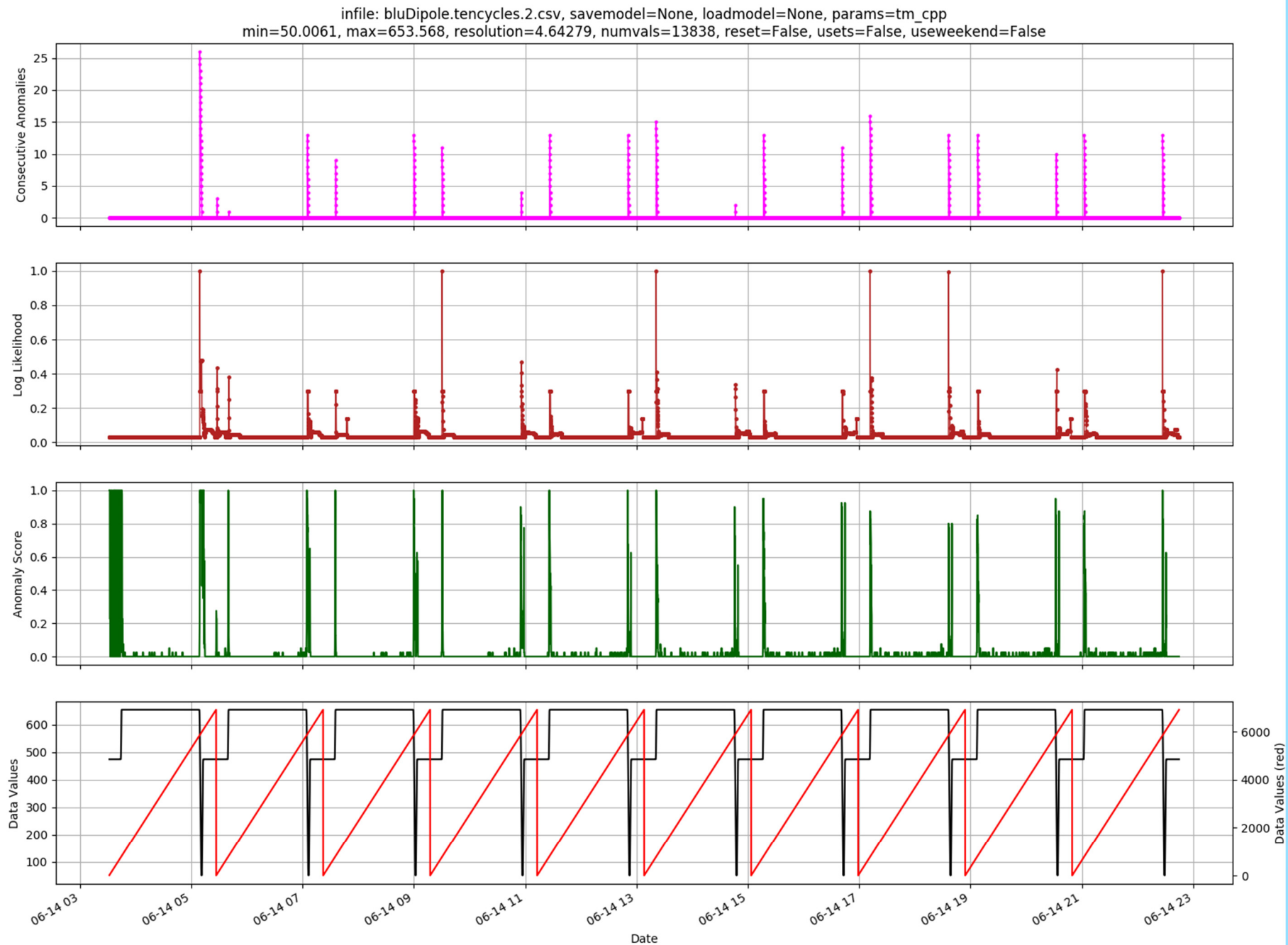
Pattern Deviation Anomalies



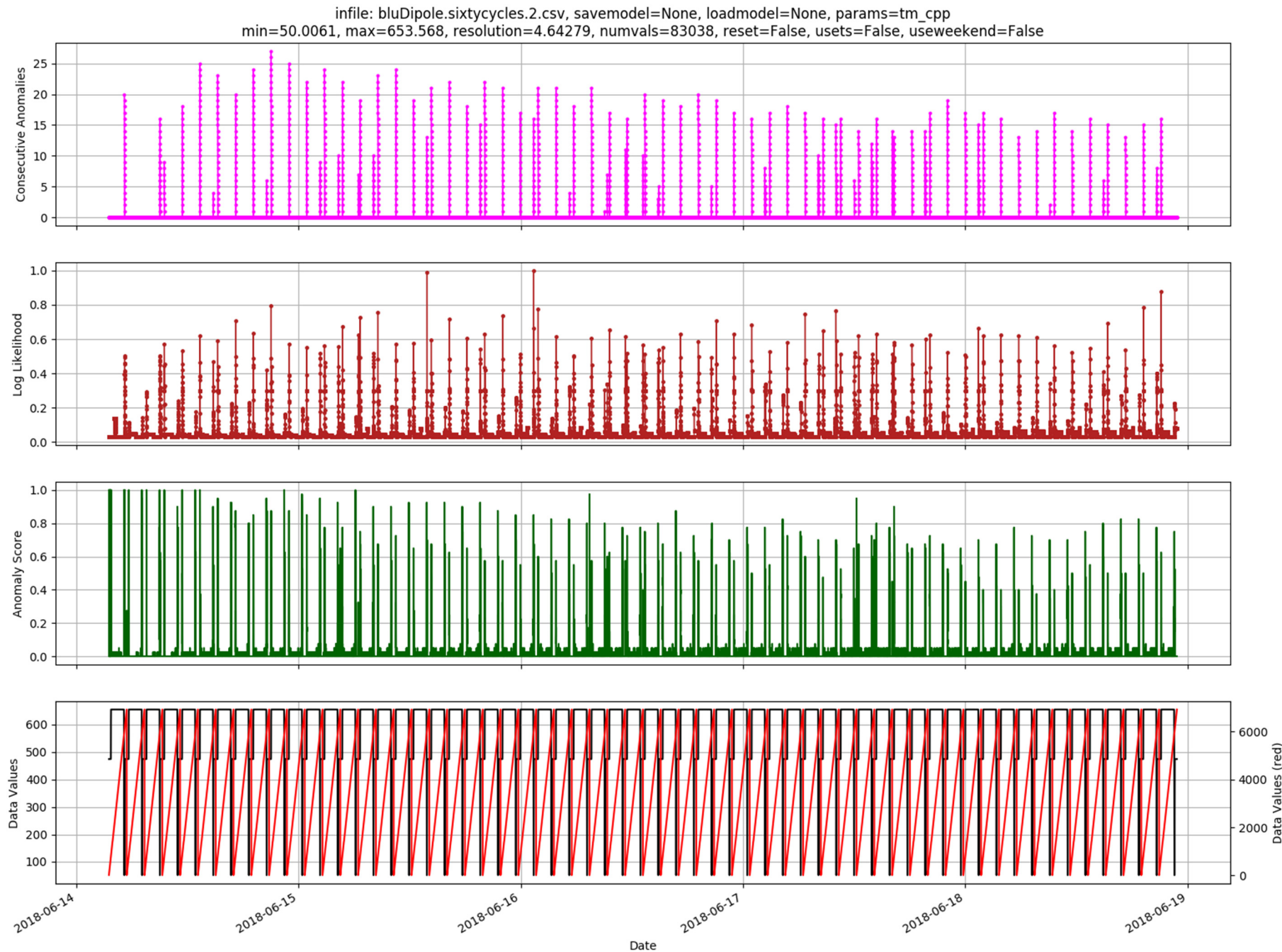
Pattern Deviation Anomalies cont.



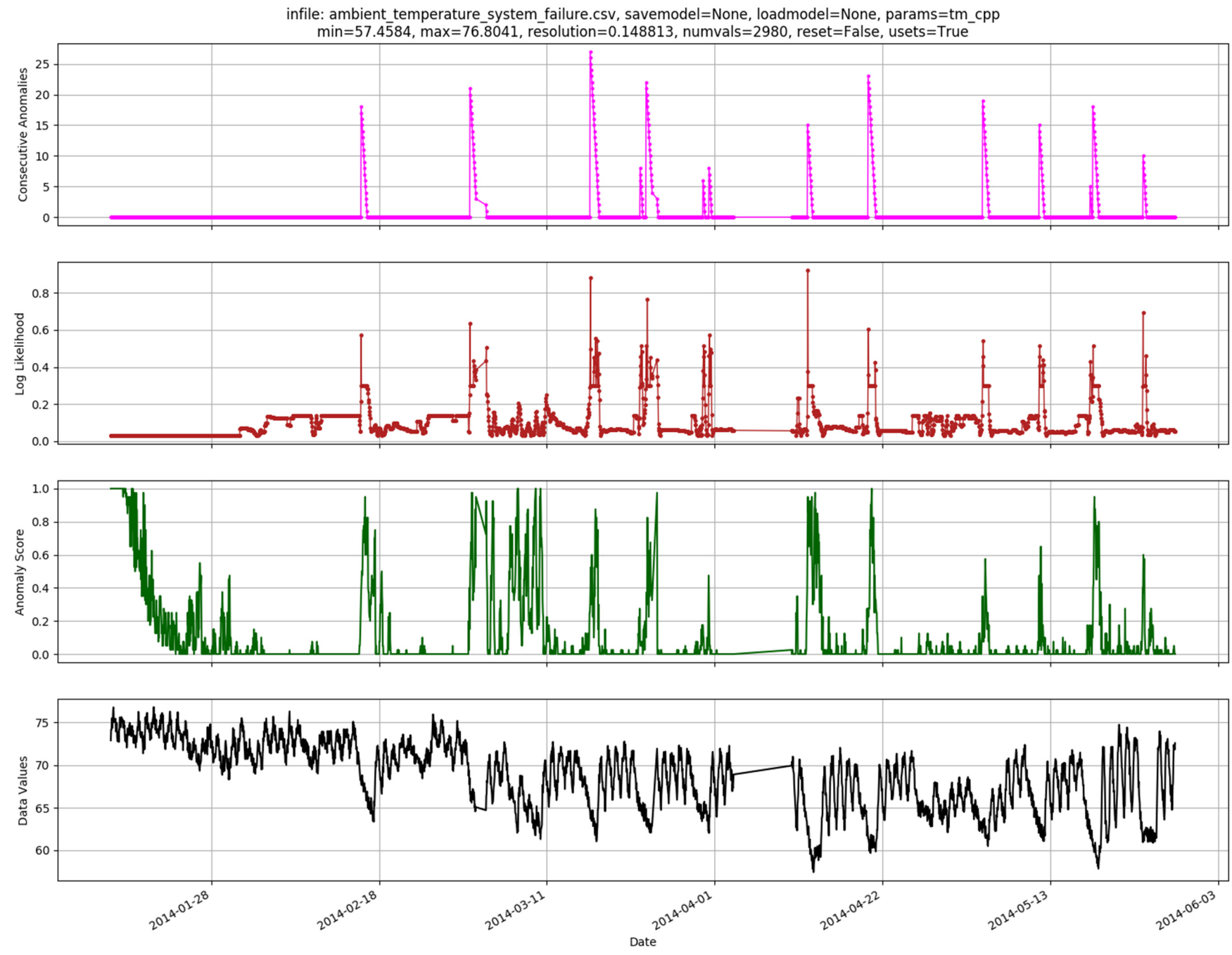
Pattern Deviation Anomalies cont.



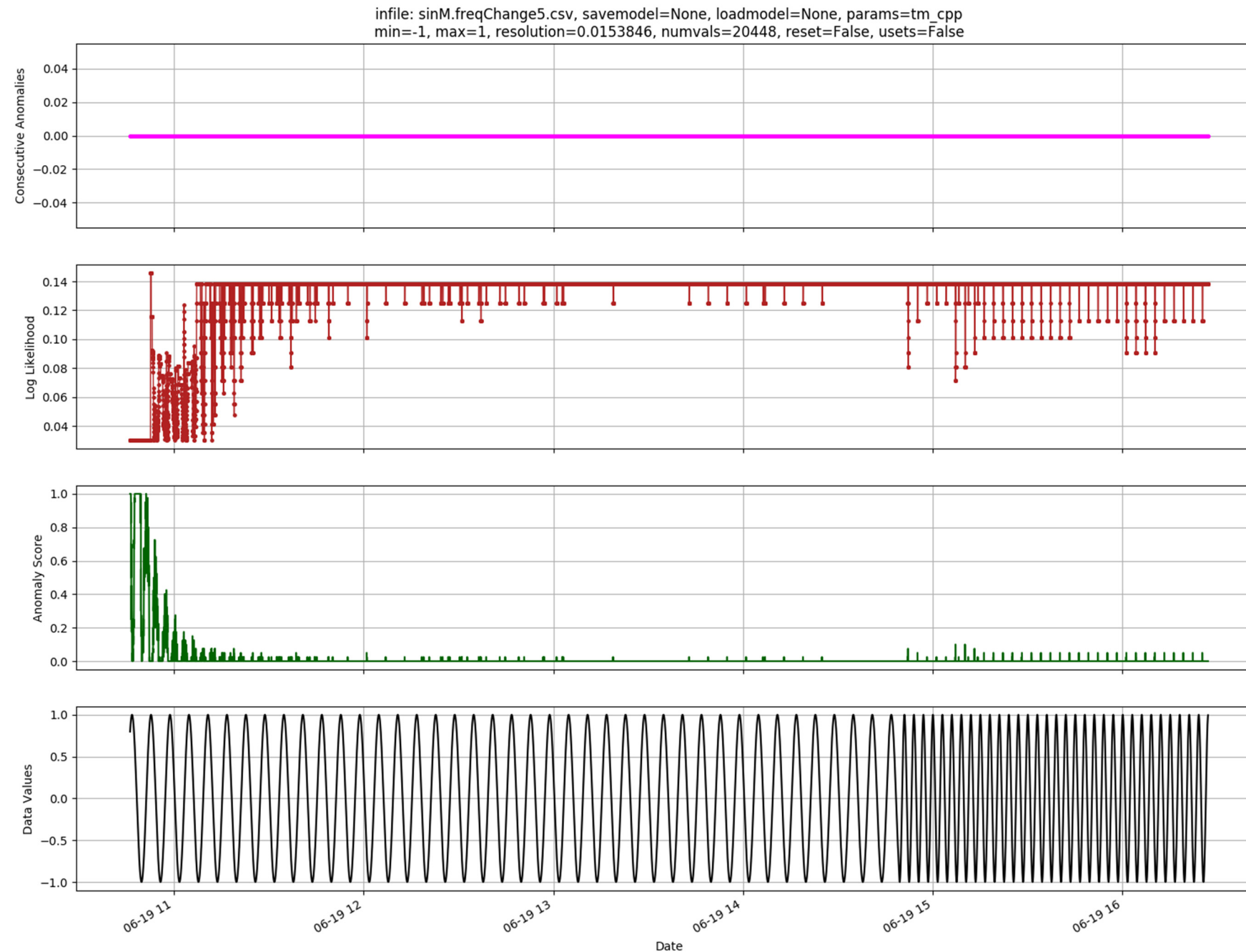
Pattern Deviation Anomalies cont.



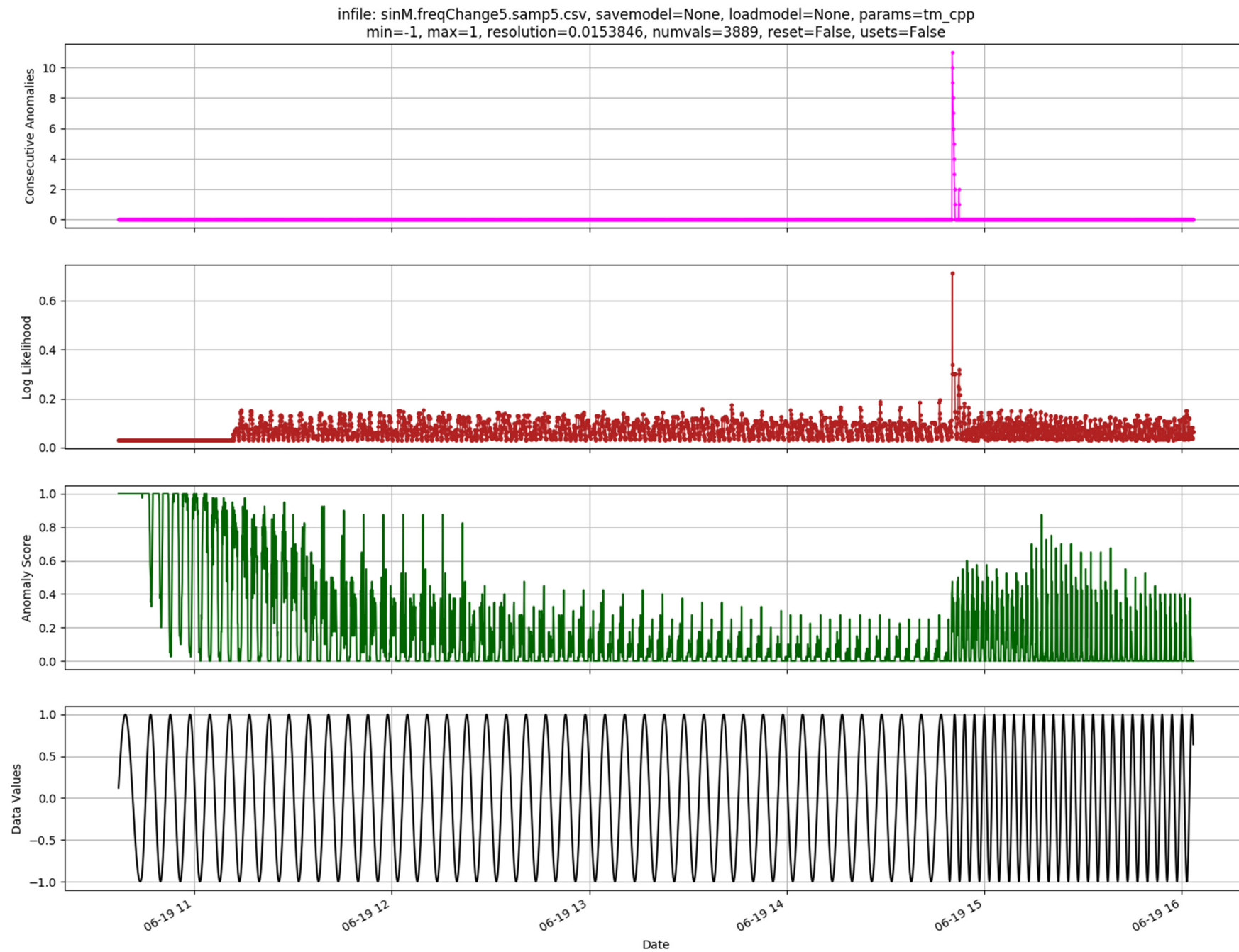
Pattern Change Anomalies



Pattern Change Anomalies cont.



Pattern Change Anomalies cont.



Getting the Most Out of NuPIC

- Preprocess the data as needed using averaging and/or filtering
Gather around 10k points (more is OK, min 1k)
- Adjust sensitivity using the resolution parameter
- If possible (and if necessary), train and save a model
- Run the anomaly detector, loading the model if saved
- An anomaly is found if you generate 10 consecutive Log Anomaly Likelihood values greater than 0.3

NuPIC Pros

- Can be set up to run with or without training
- General purpose - can be used for a wide range of data
- Can be used with both stored and real-time data.
Real-time monitoring requires 2.5% CPU to monitor a single control point returning data at once/second.
- Works well for Global Outlier and Level Change anomalies
- Works well when there are diurnal patterns in the data

NuPIC Cons

- May require preprocessing and/or configuration
- Mixed results with Pattern Deviation and Pattern Change anomalies
- Can be slow for batch processing - about 10k / minute
- Can detect anomalies for only one data set at a time
- Custom algorithms for a particular data set may work just as well or better

Questions?