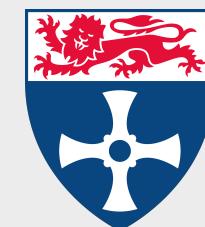


Investigating the effect of down-sampling and interpolating time series data using similarity metrics.

Tolu Awosanya; Matthew Forshaw;
Joe Matthews

Newcastle University,
t.awosanya1@ncl.ac.uk

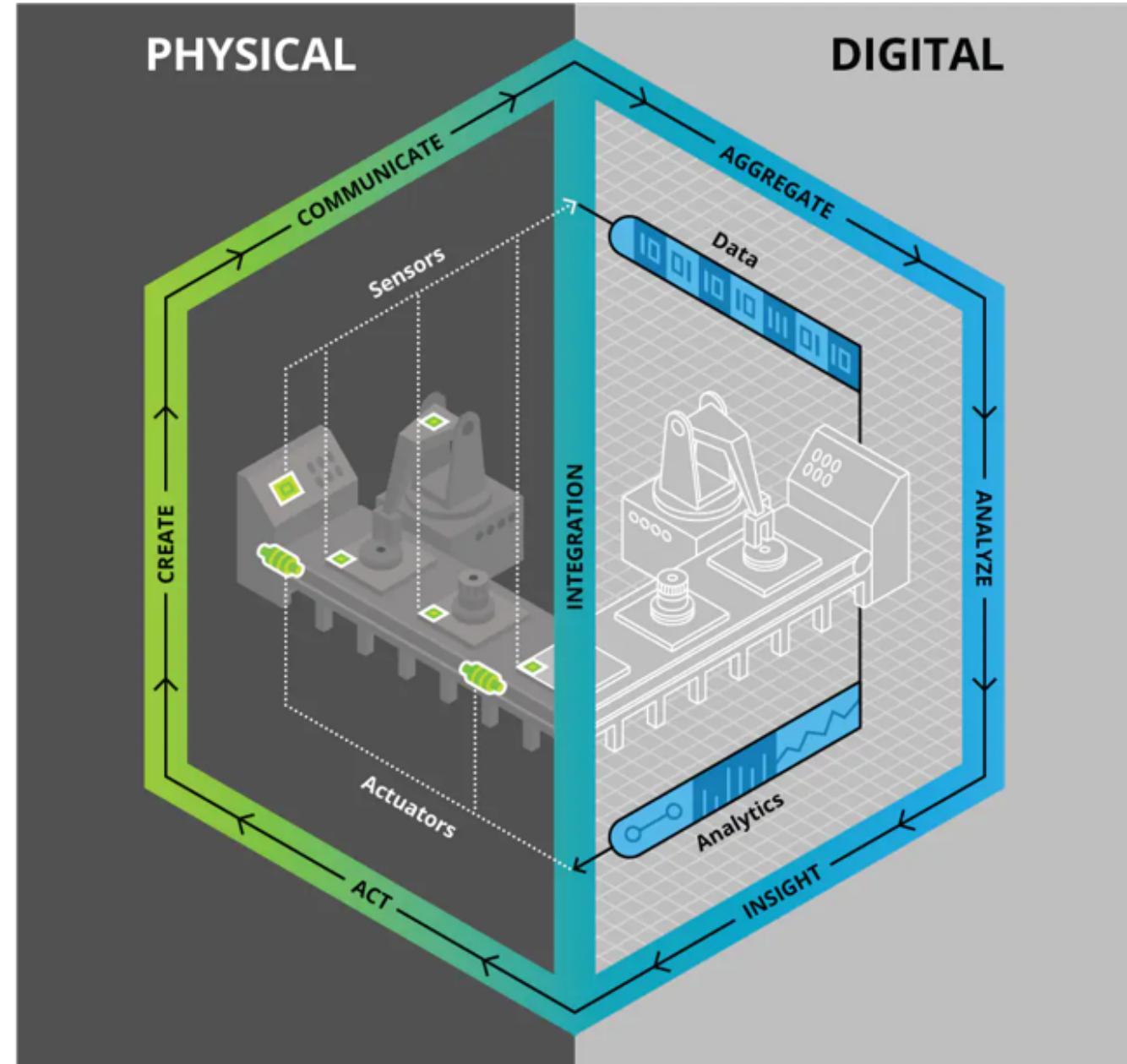
Sixth UK Systems Research Challenges
Workshop, 30th Nov - 2st Dec 2021



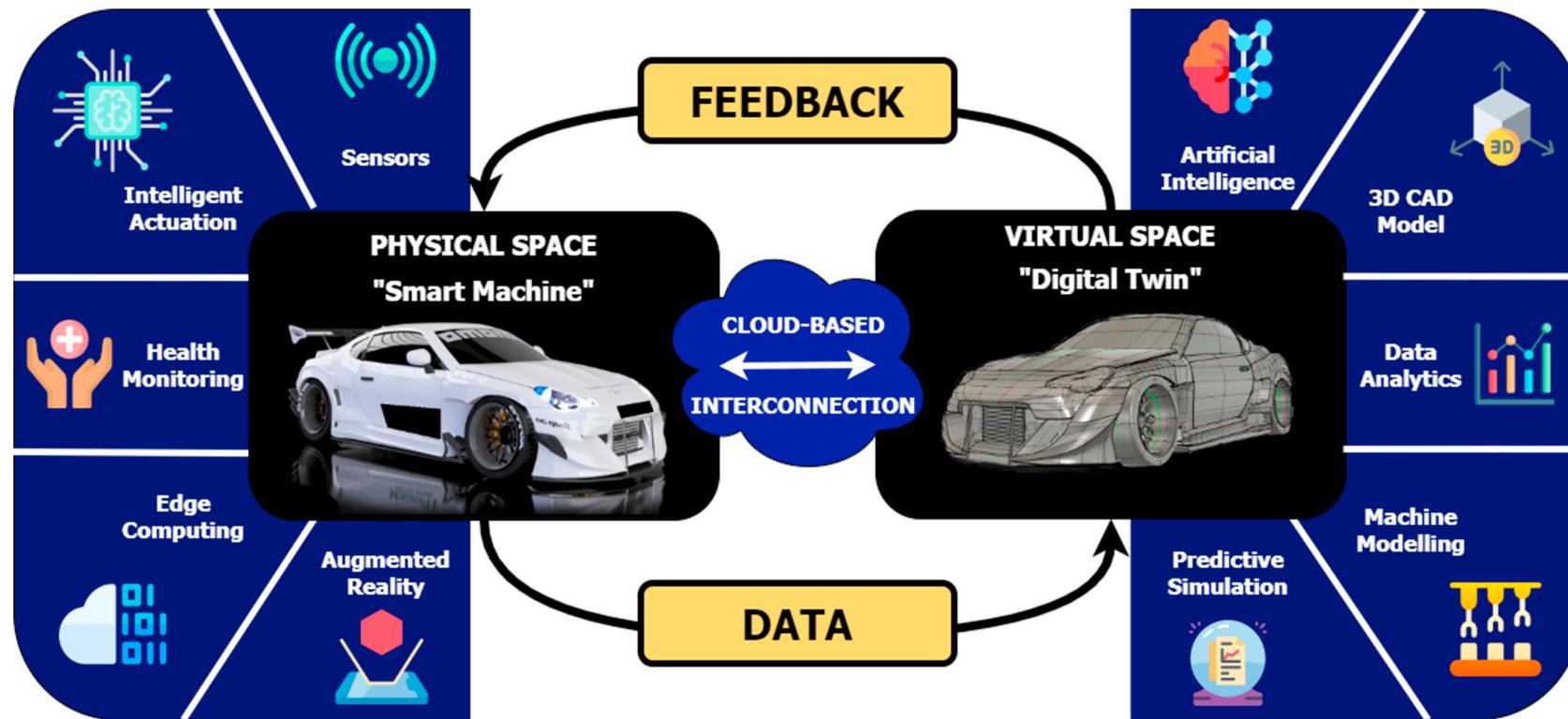
Newcastle
University

A dynamic virtual representation of a physical object or system across its lifecycle, using real-time data to enable understanding, learning, and reasoning.

Figure 1. Manufacturing process digital twin model

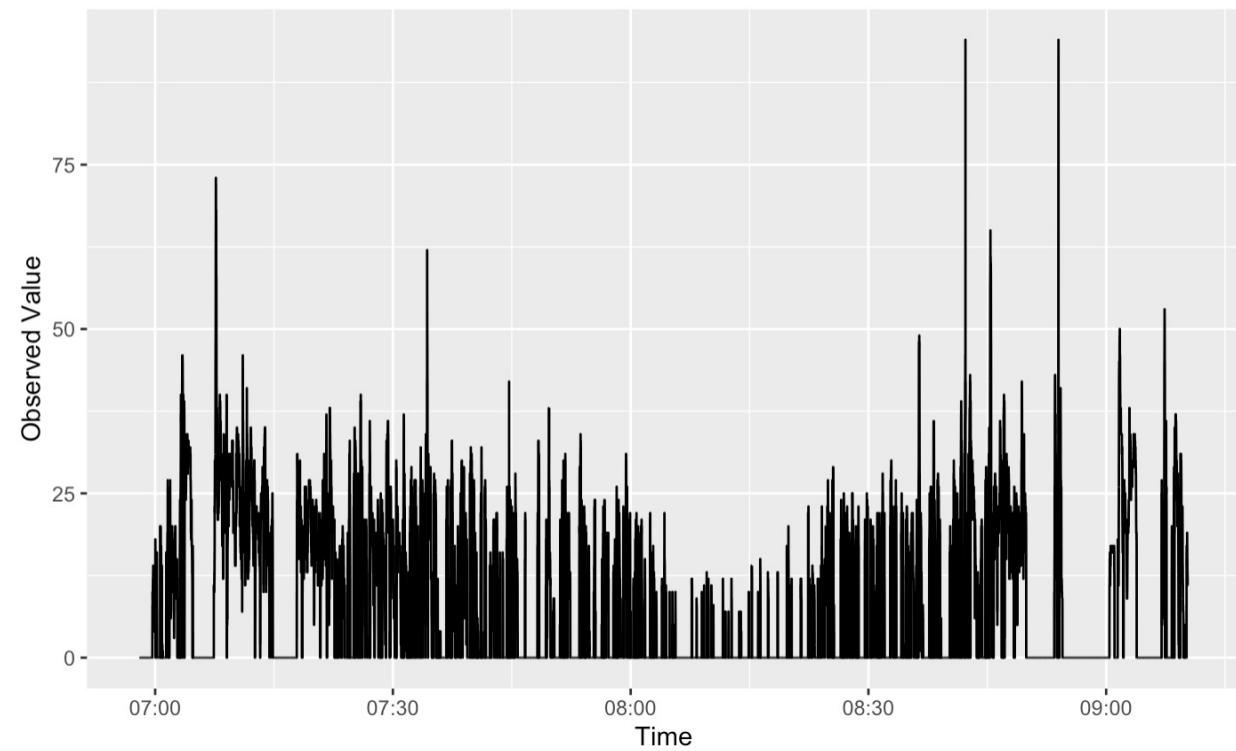
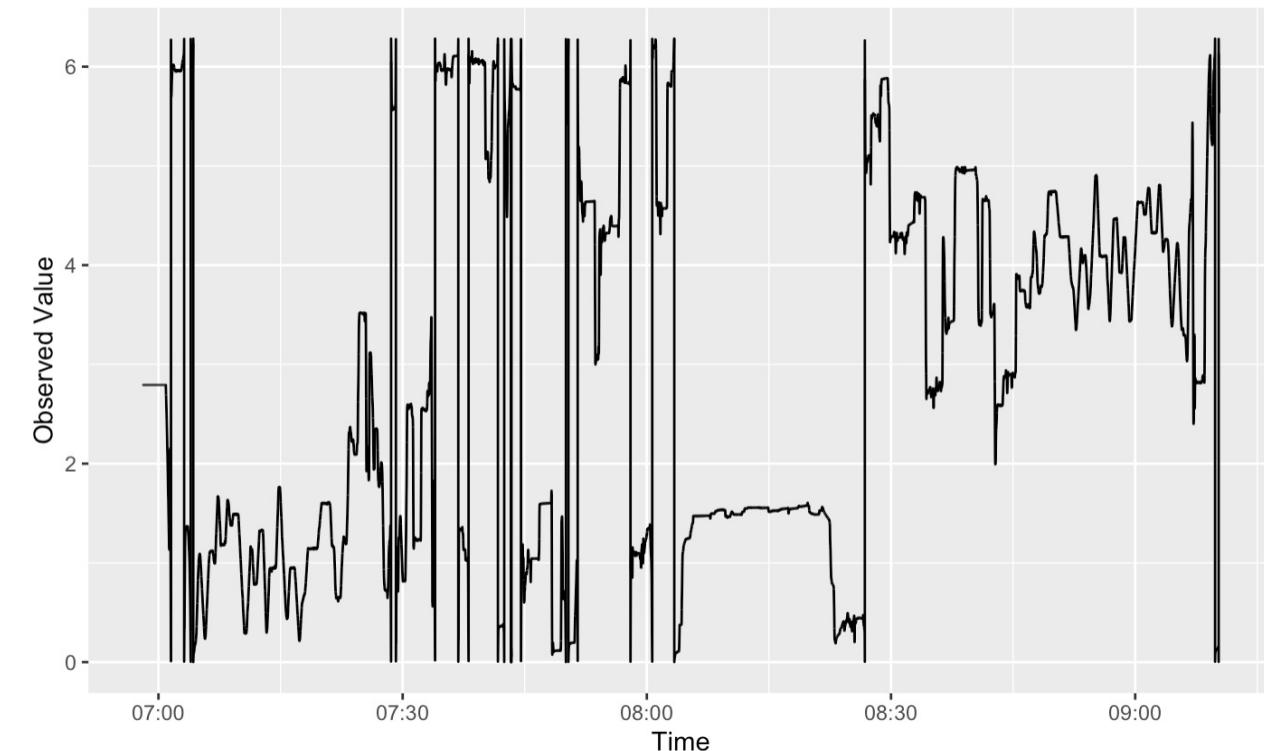


Problem

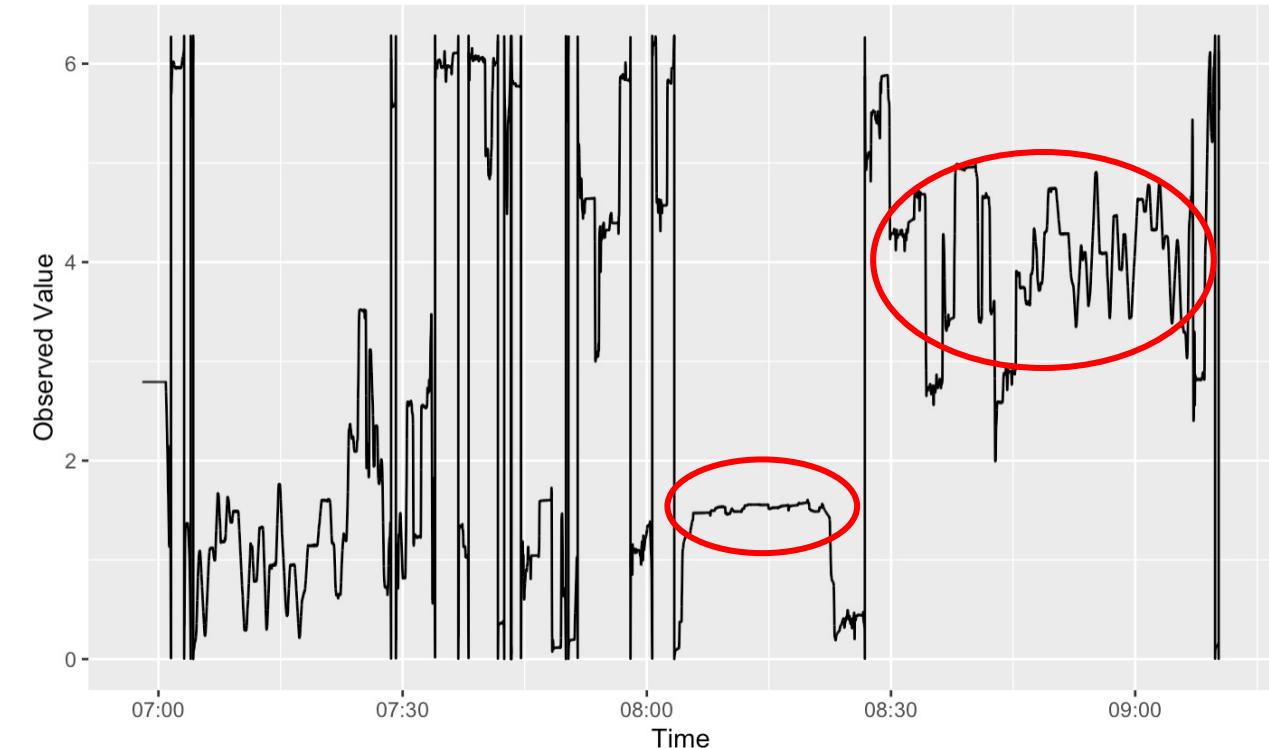


- Increased Data Growth
- Currently reducing sampling, throwing information away for current assets.
- Numerical understanding of down-sampling rather than visual.
- Higher rate of sampling and more informed down sampling

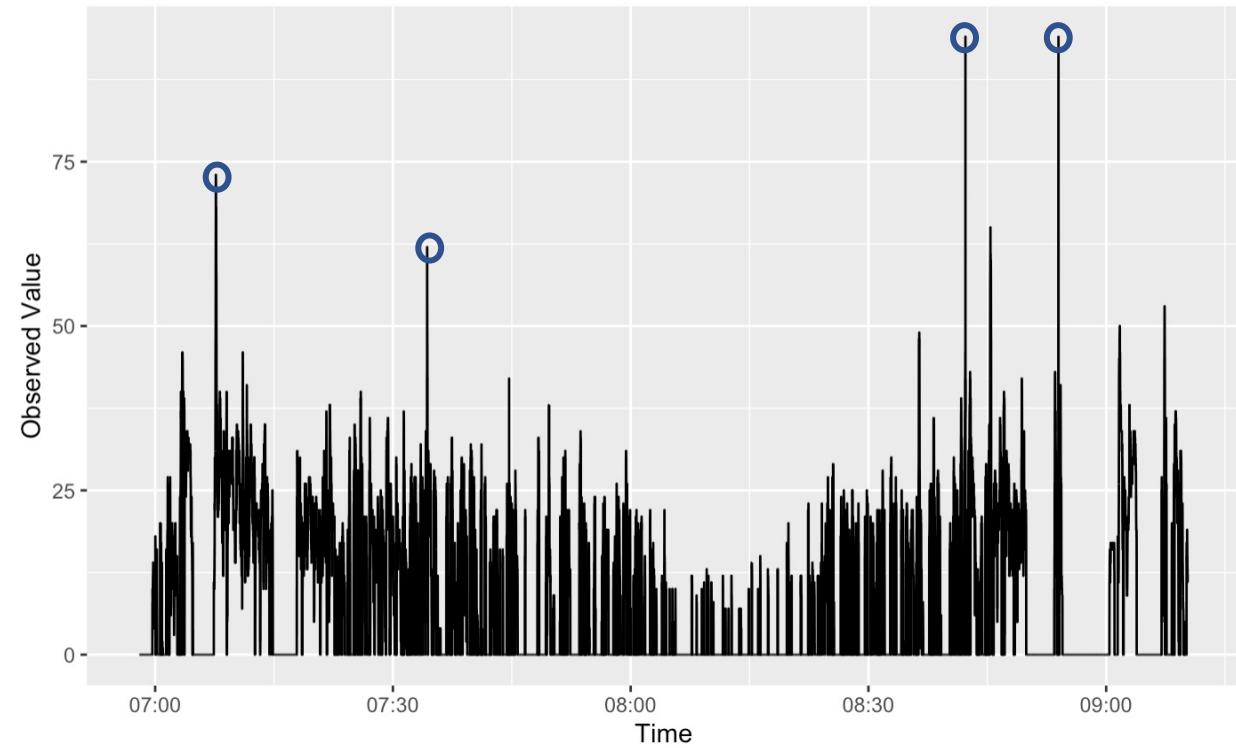
Time series



Time series

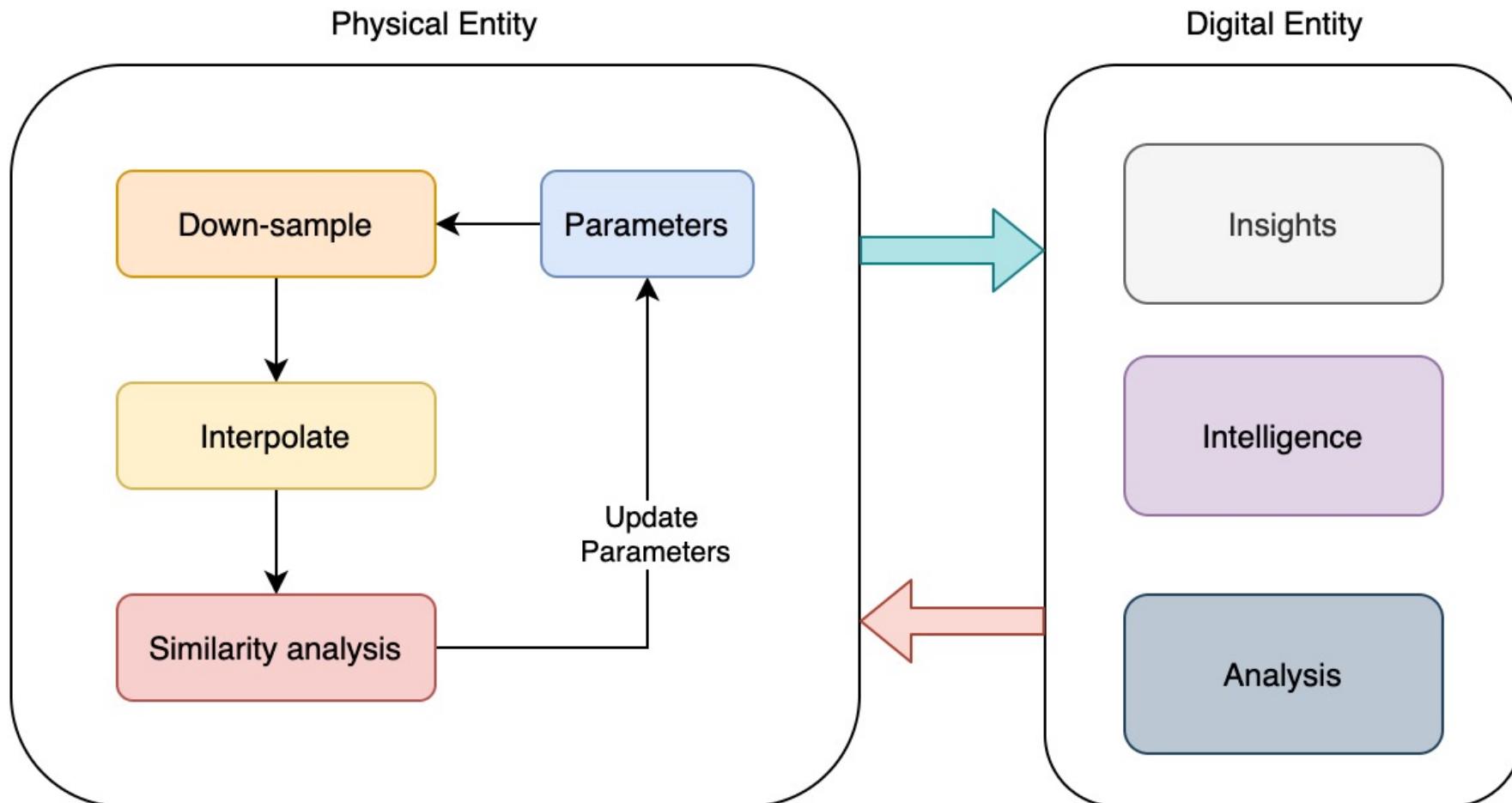


- Change Point
- Adaptability for variability in operating conditions



- Trends
- Outliers/ anomalies

System Process



Down-sampling

In signal processing, down-sampling is the process of reducing the sampling rate of a signal. This is usually done to reduce the data rate or the size of the data. This can be performed by compression, or simply removal of values.

Good attributes of down-sampling approaches

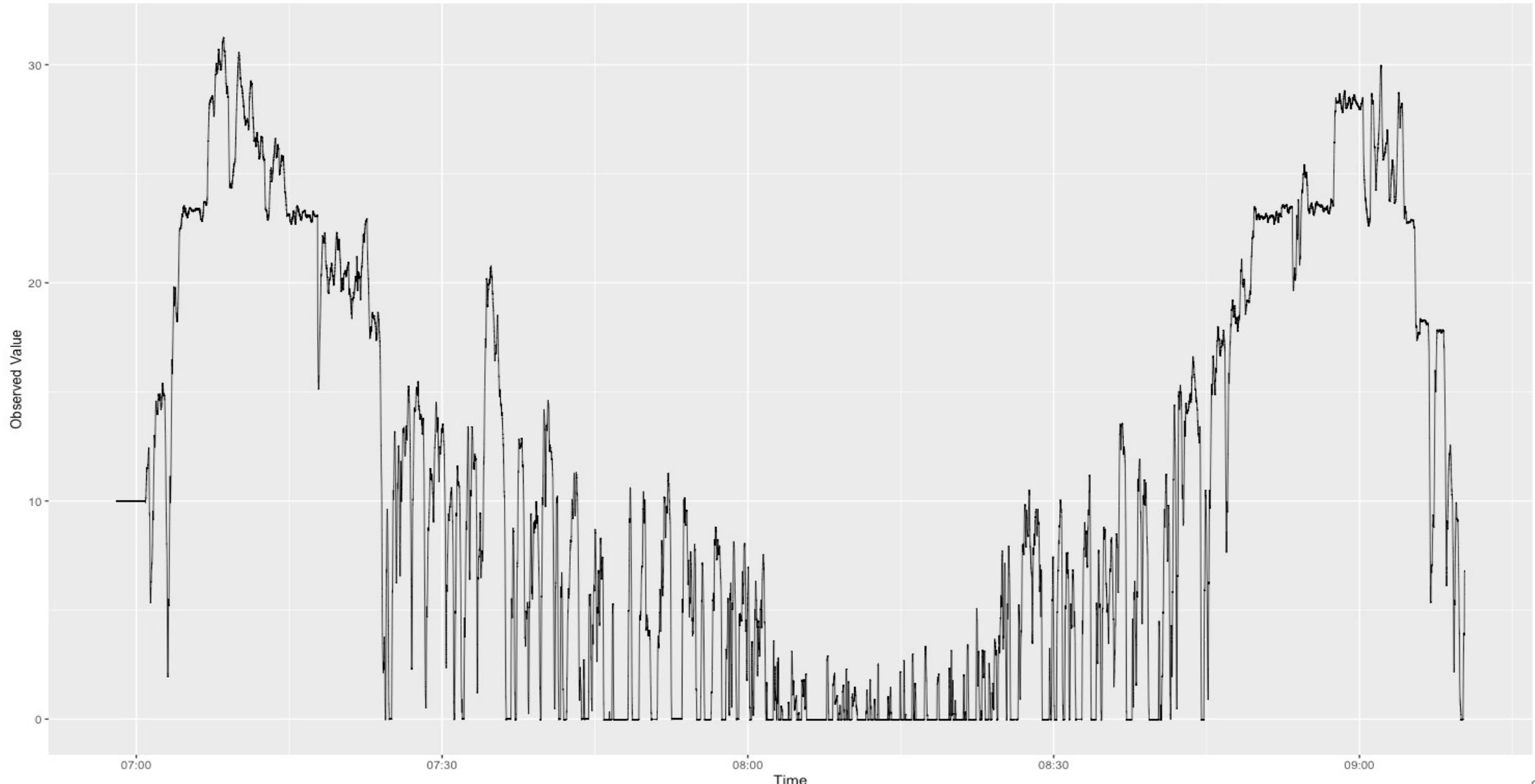
- It should be adaptive to considering various operating conditions that might exist.
- It should take outliers into considerations as they are important, and imperative to a good understanding of the physical entity.
- It should be able to function in a real time environment, for better latency.



Largest Triangle Three buckets concept

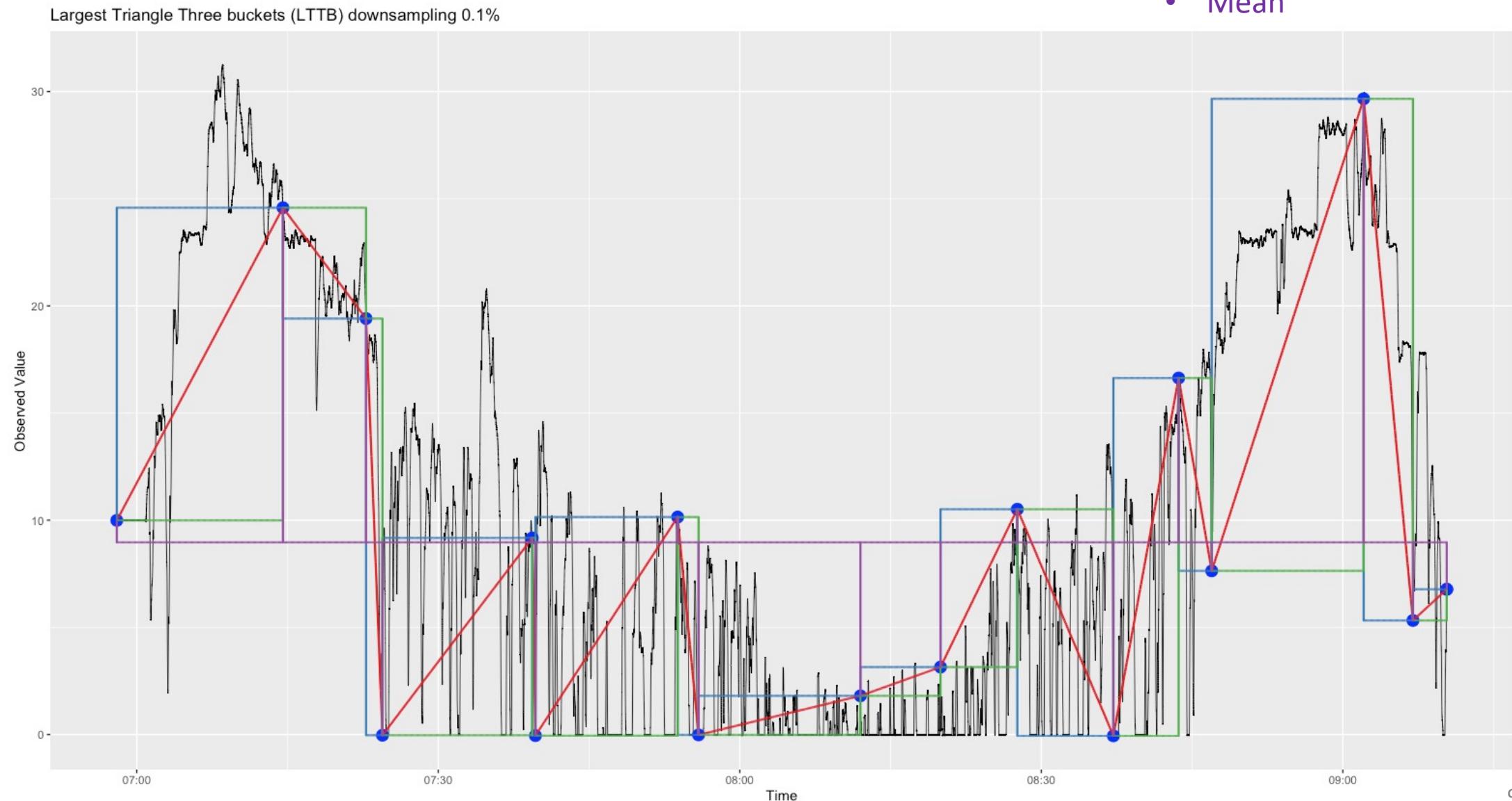
Down-sampling

Largest Triangle Three buckets (LTTB) downsampling 0.1%



Down-sampling

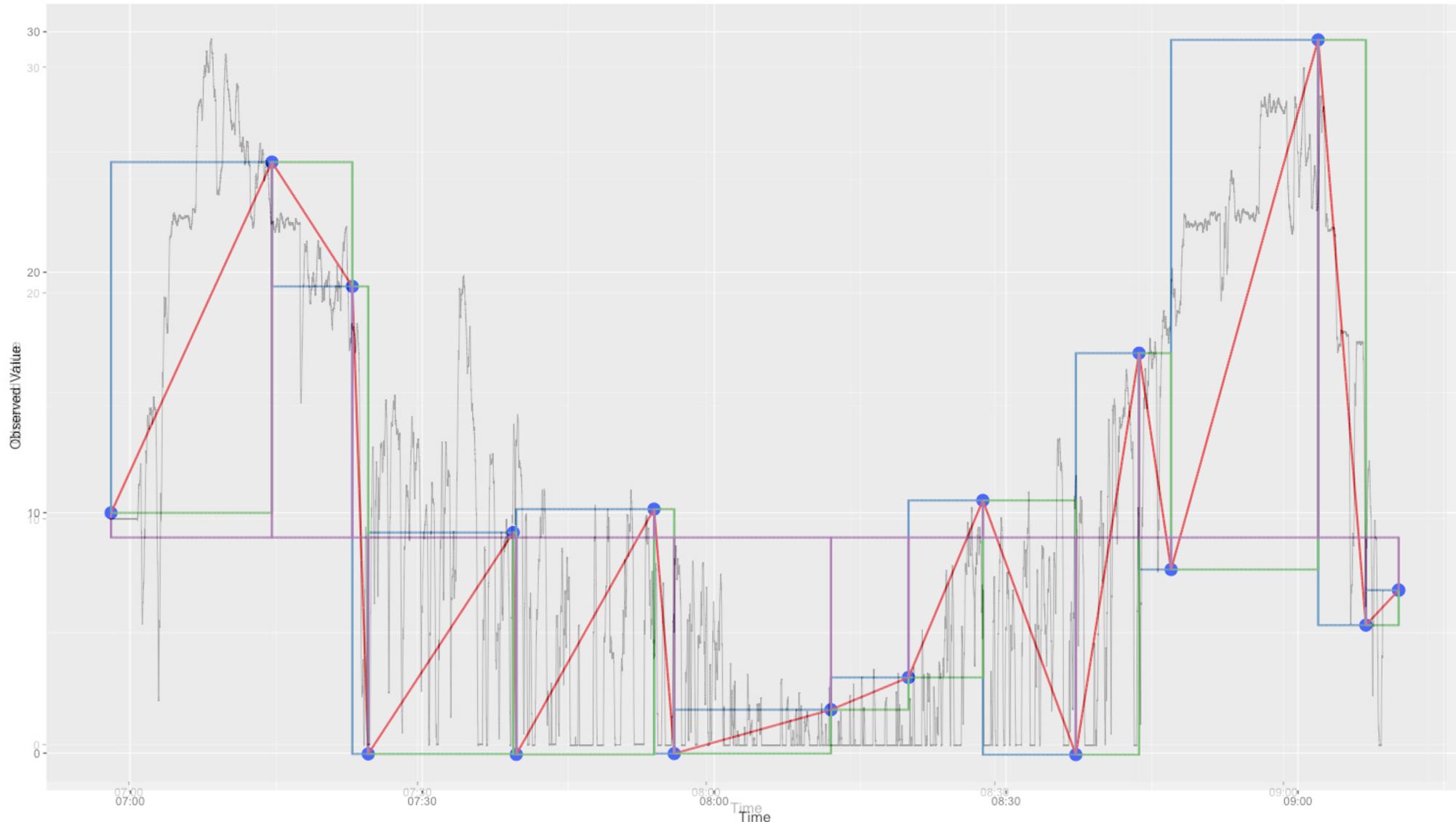
- Linear
- Last object carried forward
- Next object carried backwards
- Mean



Down-sampling

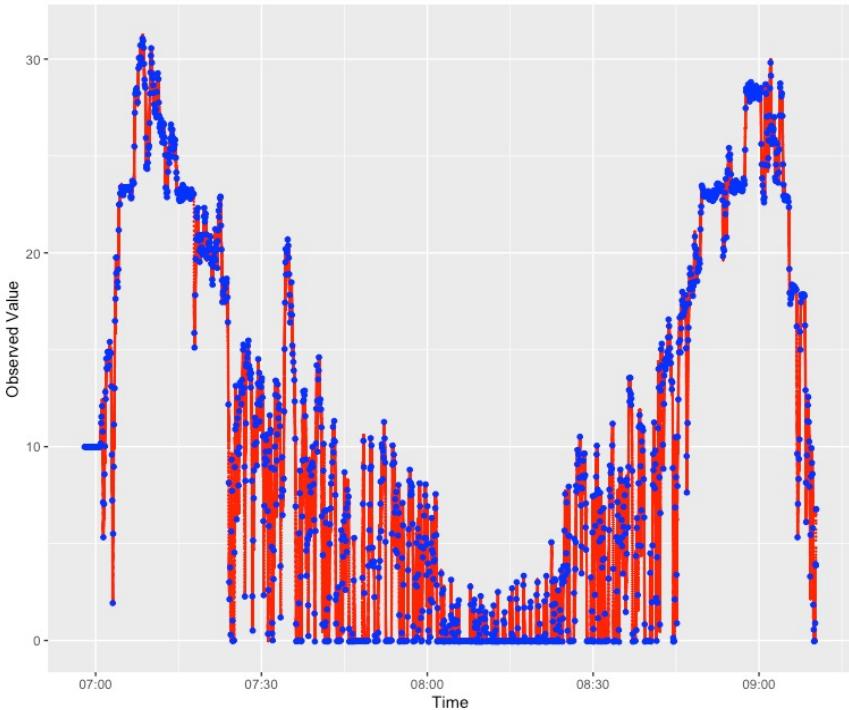
Largest Triangle Three buckets (LTTB) downsampling 0.1%

- Linear
- Last object carried forward
- Next object carried backwards
- Mean

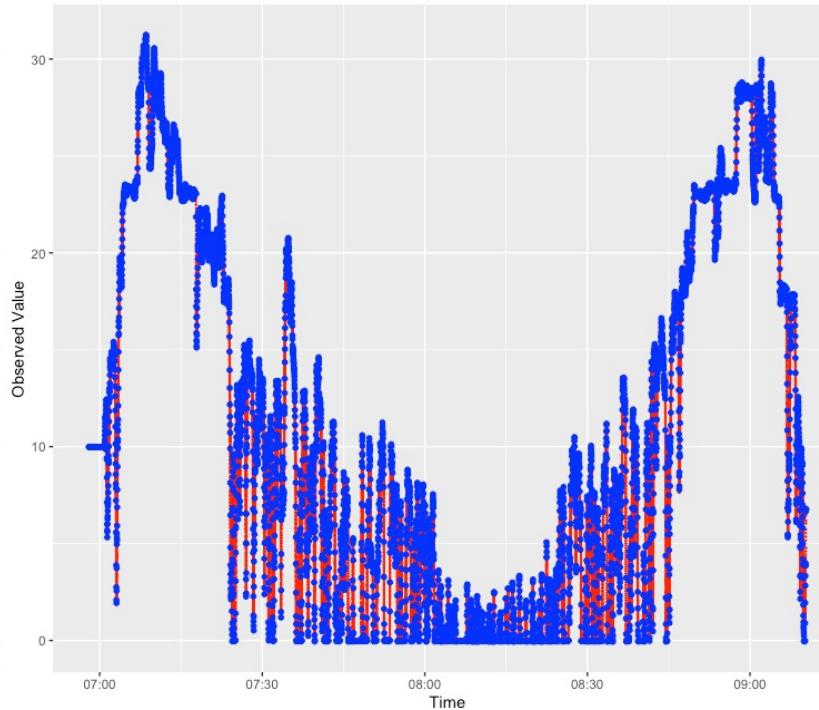


Down-sampling Example

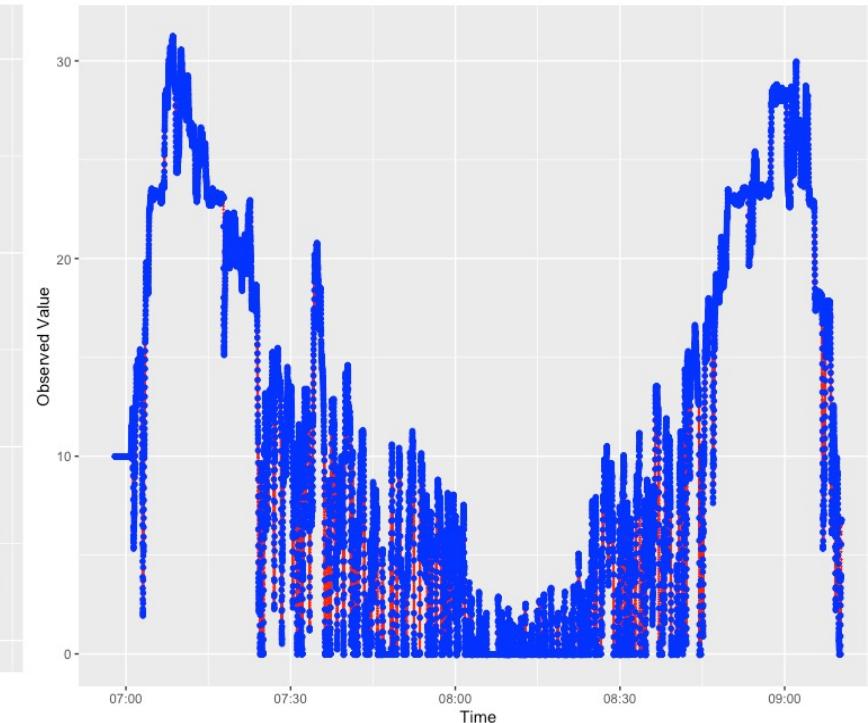
Largest Triangle Three buckets (LTTB) downsampling 1%



Largest Triangle Three buckets (LTTB) downsampling 5%



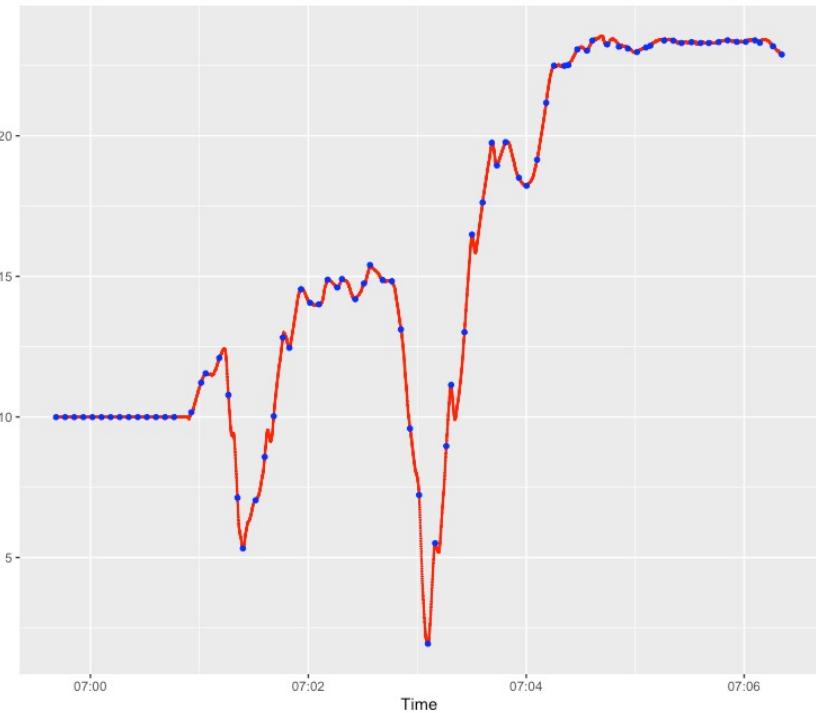
Largest Triangle Three buckets (LTTB) downsampling 10%



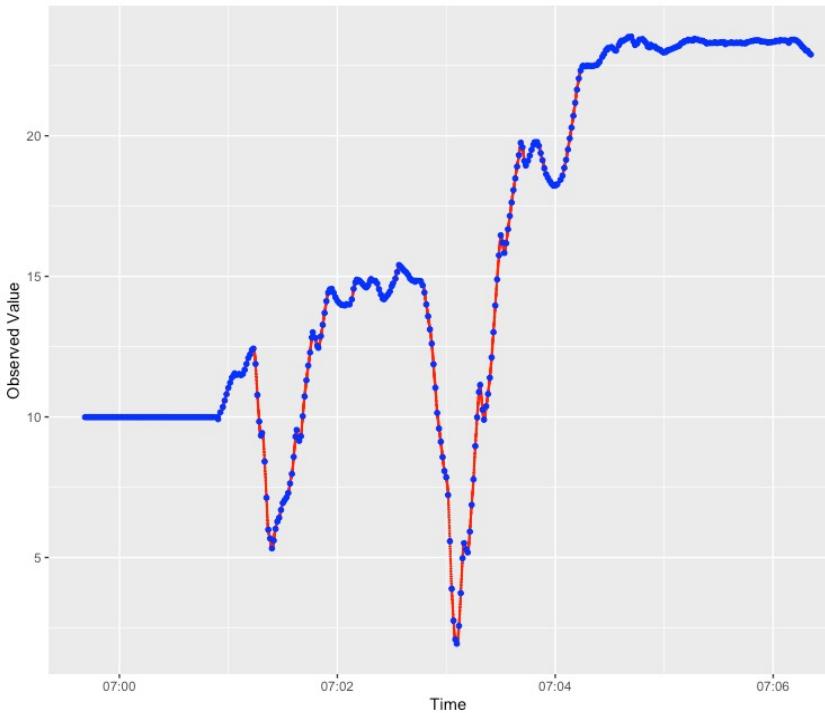
- The higher your down-sampling variable more points kept.
- No discernable pattern from the down-sampled points.

Down-sampling Example

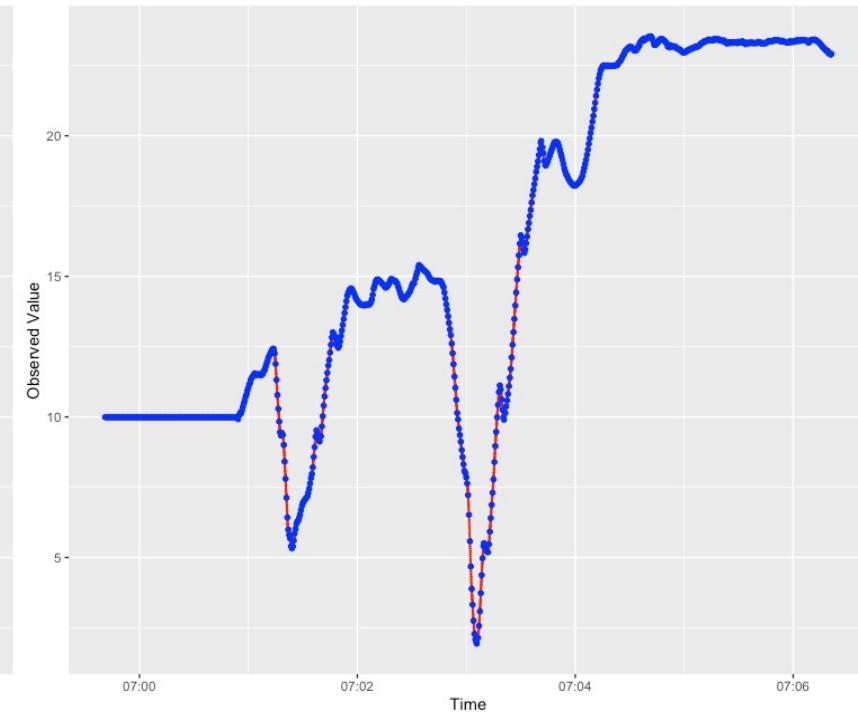
Largest Triangle Three buckets (LTTB) downsampling 1%



Largest Triangle Three buckets (LTTB) downsampling 5%



Largest Triangle Three buckets (LTTB) downsampling 10%

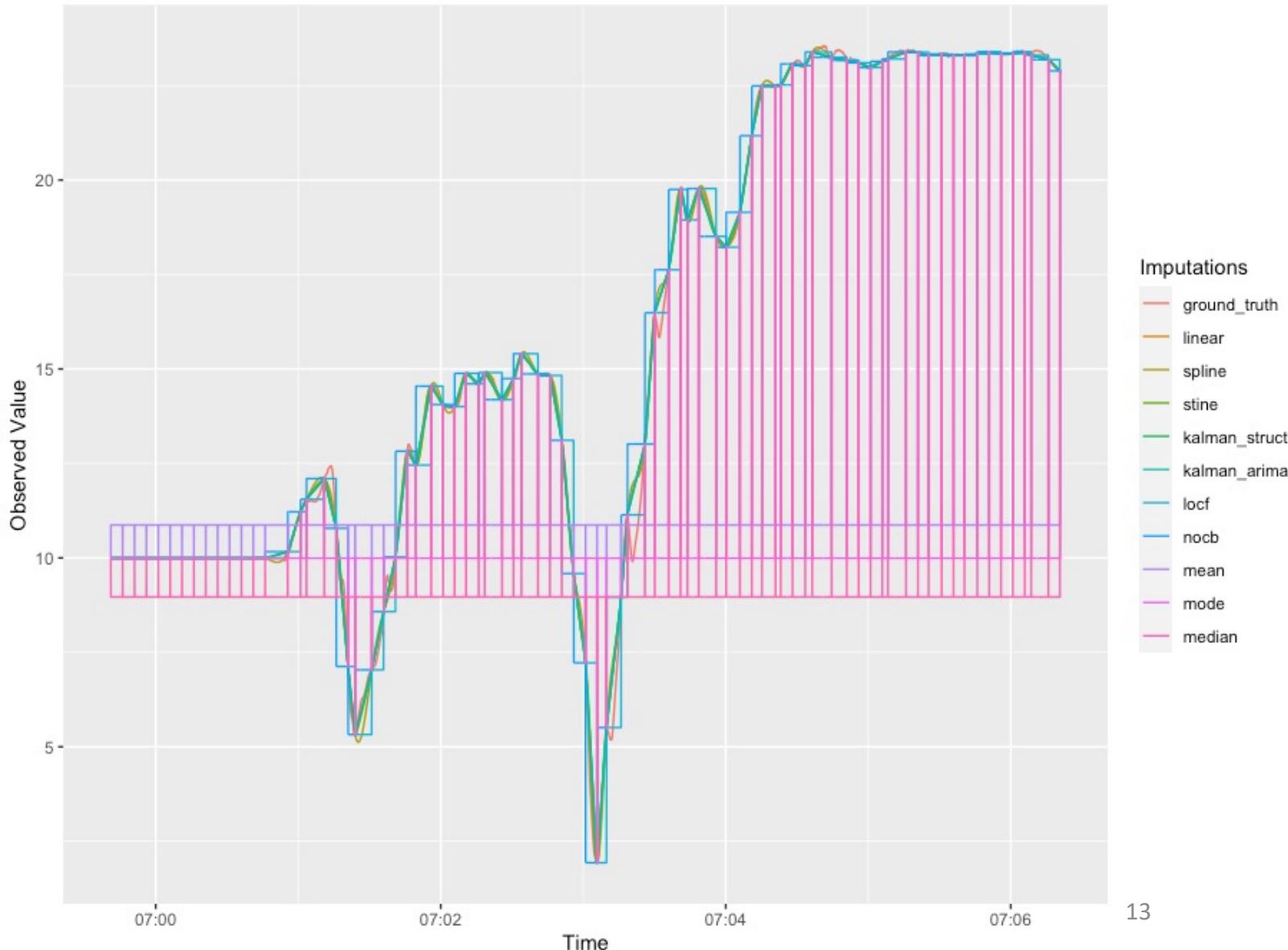


Observations are first added area with rapid movement.

Interpolation

Good attributes of interpolation approaches.

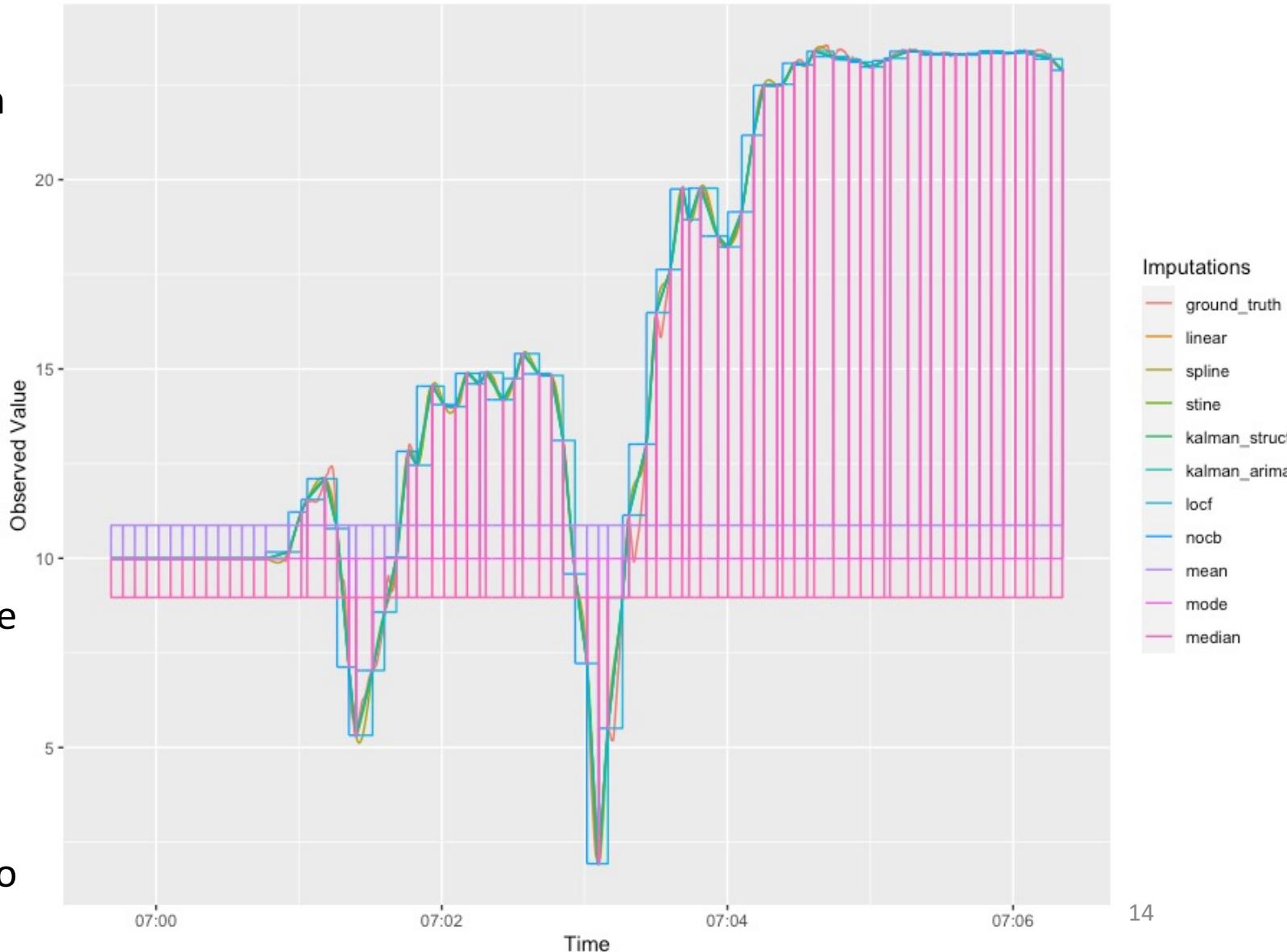
- Knowledge of the semantics of the down-sampling approaches, such as the thresholds, down-sampling configuration.
- Statistical accuracy
- Visual accuracy
- Quick execution



Interpolation

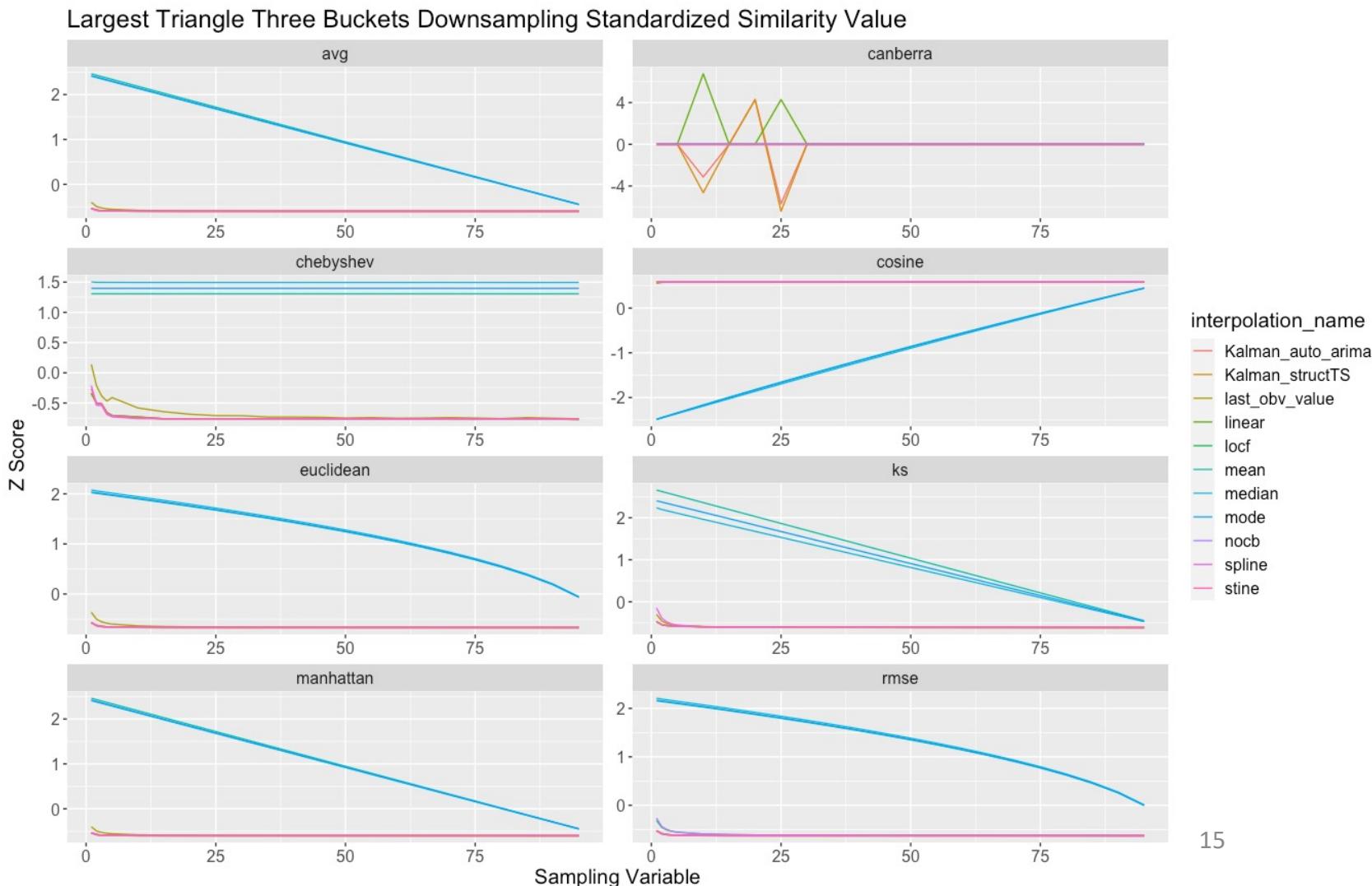
There are tradeoffs to be considered with it comes to selecting interpolation methods.

- The methods currently listed don't have an implicit knowledge into how the data was down-sampled.
- The mean mode and median does not provide a good visual and statistical accuracy.
- The Spline interpolation does not provide a good visual accuracy as the values interpolated could end up interpolating into the negative axis.
- The Kalman based interpolations are slower when compared to the rest due to the way it functions.



Similarity metrics

- The sampling variable of the largest triangle three buckets have an impact on the similarity metric value.
- The interpolation type has an impact on the final similarity metric value.
- The similarity variable matters when paired up with the specific interpolation type.



Future directions

- Aim to further complete the permutations of down-sampling algorithms, imputation methods and similarity metrics, as the more visited the more insights can be gleamed from this body of work.
- Expand into multivariate down-sampling.
- Develop a testable streaming implementation of the system on a physical hardware.