advanced network science initiative (ansi)

Statistical learning based online prediction, detection, and classification of anomalies in power grids

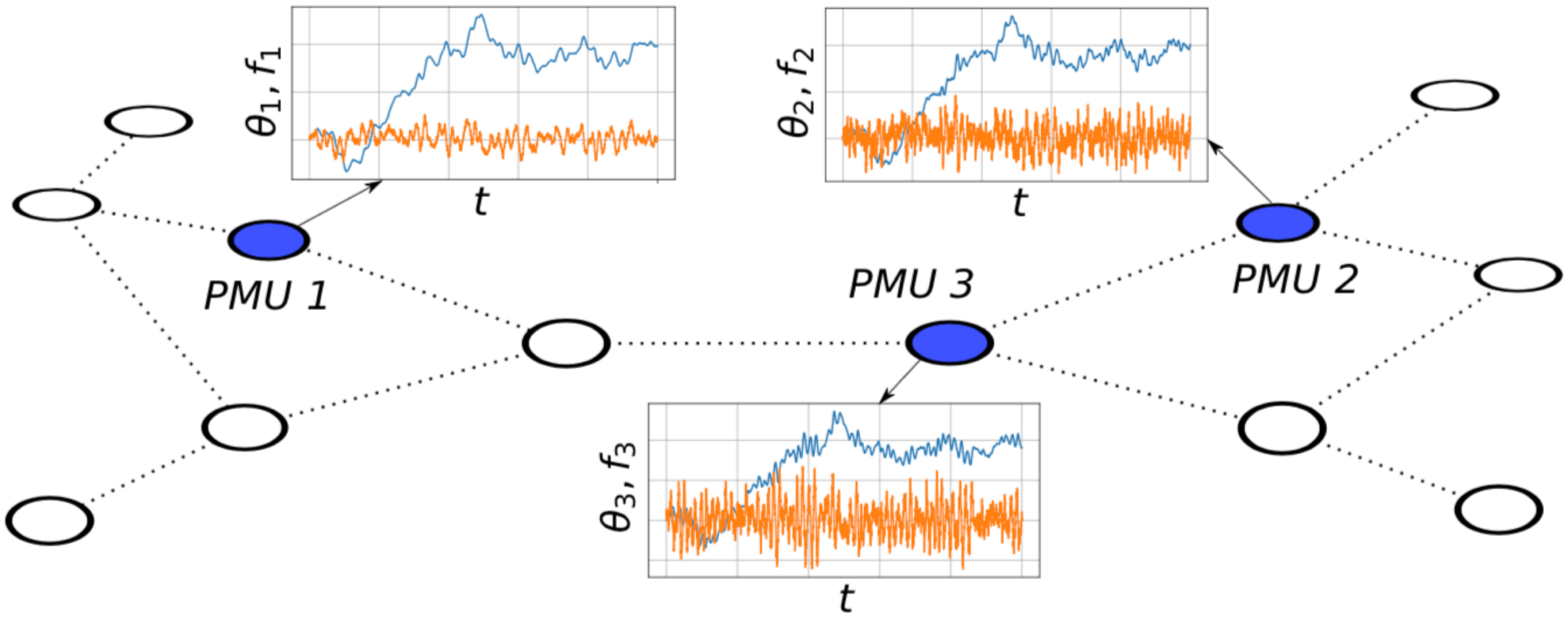
Christopher Hannon, Deepjyoti Deka, Marc Vuffray, Andrey Lokhov

Los Alamos National Laboratory
Advanced Network Science Initiative

NASPI meeting, San Diego CA, April 2019

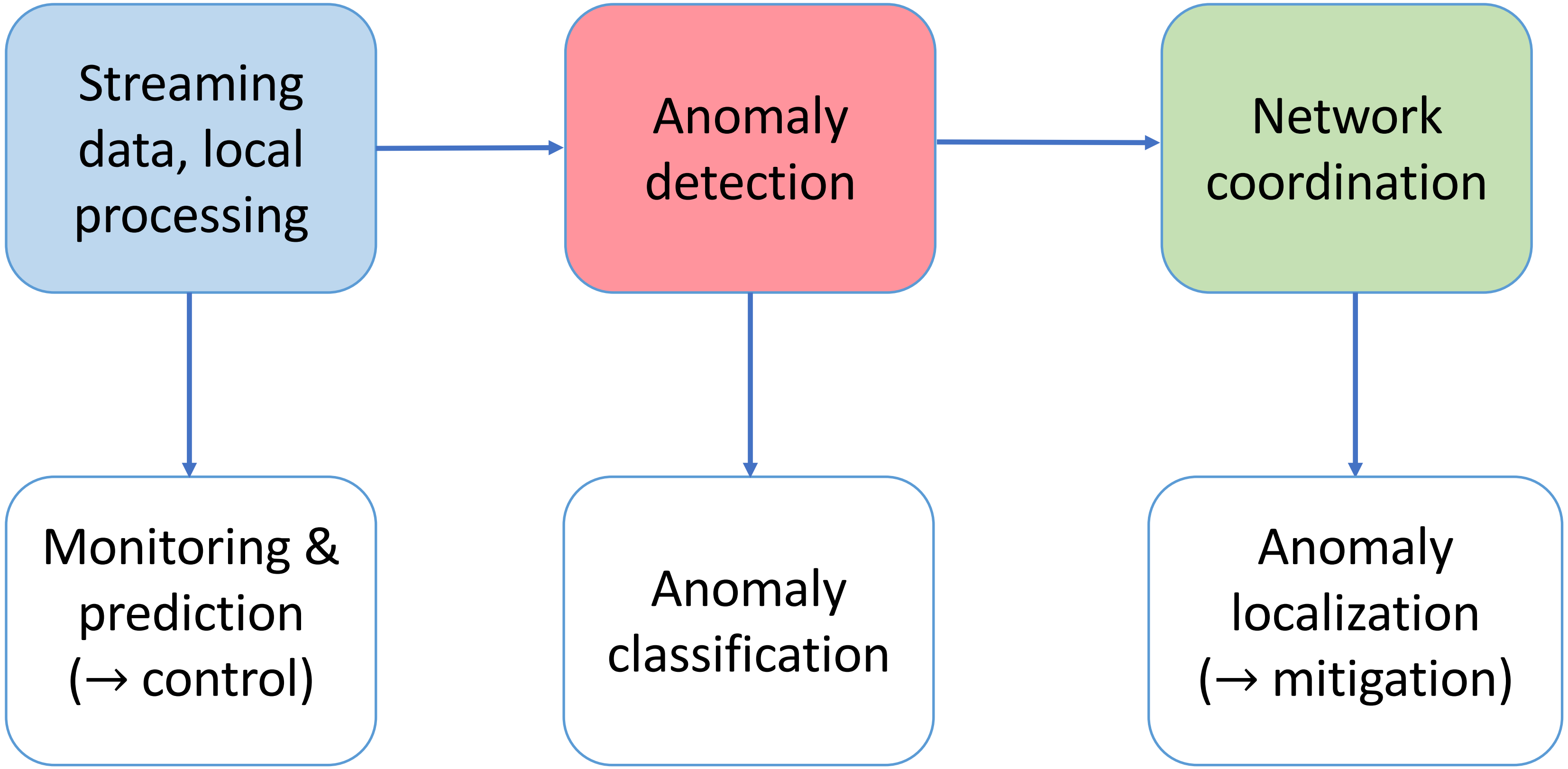
PMU data analytics in power grids

- ✓ State estimation
- ✓ Model validation
- ✓ Forced oscillations
- ✓ System stability
- ✓ Control Parameter calibration
- ✓ Generation measures
- ✓ re-dispatch
- ✓ …



This talk: dealing with anomalies (ideally, methods coordinated with other tasks above)

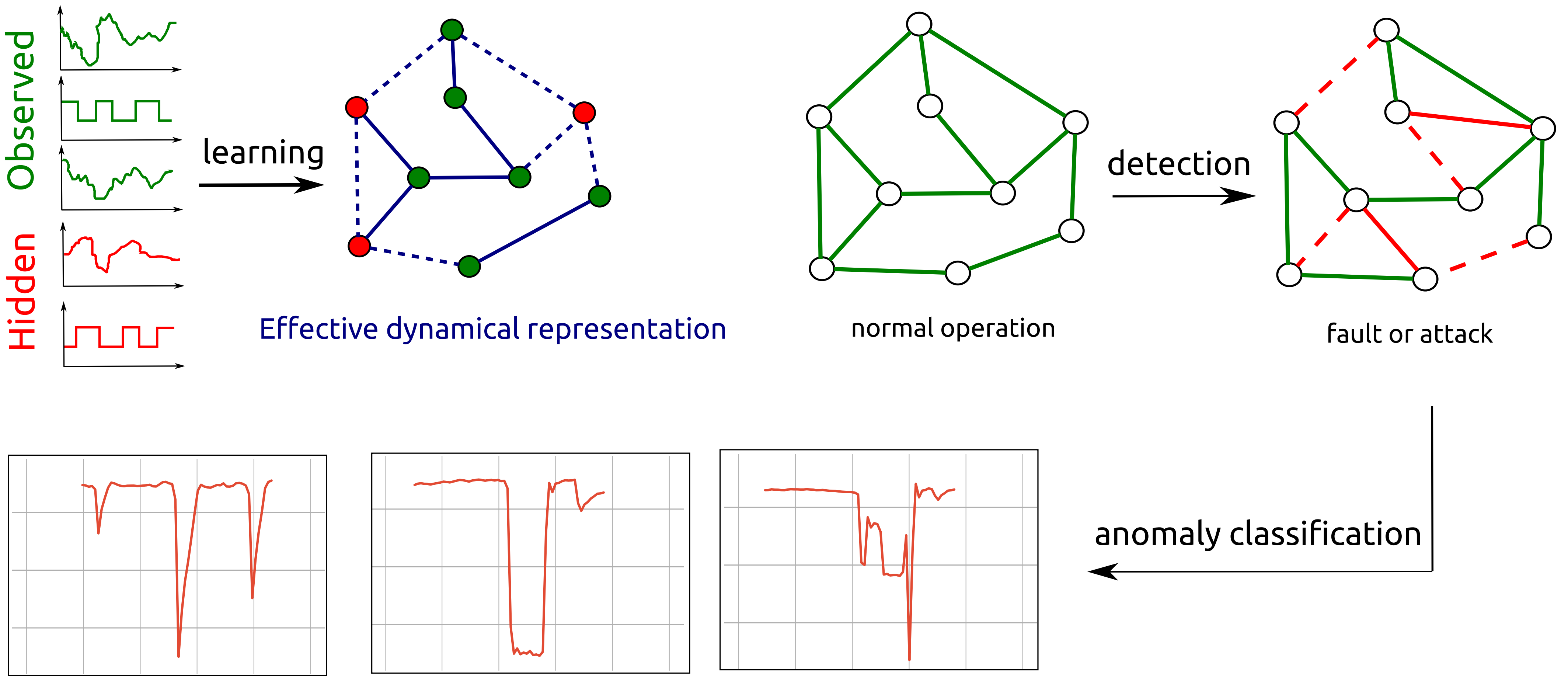
Monitoring of anomalies: tasks



End goal: assisting operators through a software with automatic algorithmic processing

- ✓ Computational efficiency
- ✓ Rigorous guarantees
- ✓ Interpretability (explanations)
- ✓ Simplicity & vizualization

Statistical learning framework: unified approach with guarantees



After local processing: network-wide communication, anomaly localization and implementation of control measures

Principle challenges in anomalies for power grid

- Precise model often unknown or parameters changing in time
- Rigorous guarantees on quality of predictions often absent
- Identification of predefined anomaly types can be vulnerable to previously unseen anomalies
- Existing approaches (e.g. with tunable parameters) are often not transferrable
- Logic behind decisions through some algorithms are not always readily interpretable
- The need for algorithms to run faster than real time and make predictions

Solutions offered by the statistical learning framework

- Precise model often unknown or parameters changing in time
  Learning a stochastic model at every time point

- Rigorous guarantees on quality of predictions often absent
  Theoretical and empirical guarantees

- Identification of predefined anomaly types can be vulnerable to previously unseen anomalies
  Agnostic to predefined faults or attack vectors (CPS perspective)

- Existing approaches (e.g. with tunable parameters) are often not transferrable
  No tunable hyperparameters, transferability of methods

- Logic behind decisions through some algorithms are not always readily interpretable
  Probabilistic scoring, explanation of decisions

- The need for algorithms to run faster than real time and make predictions
  Fast learning process generating predictive model

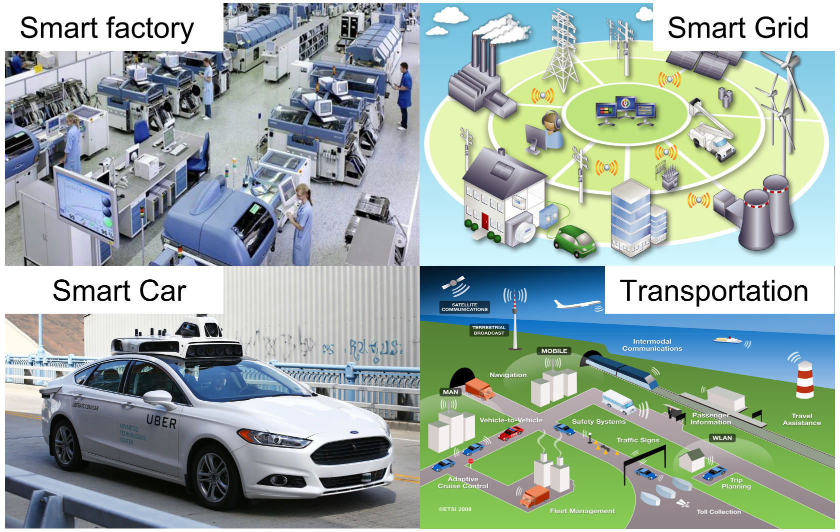
Diving into details: learning of an effective model

Sometimes, we have a good idea of the model form…

$$M_i \ddot{\delta}_i + D_i \dot{\delta}_i = - \sum_{(ij) \in E} \beta_{ij} \left(\delta_i - \delta_j \right) + \delta P_i$$

But often, we don't even have that!

$$X_{t+1} = F(X_t, X_{t-1}, \dots, X_{t-k}, \xi)$$



Diving into details: learning of an effective model

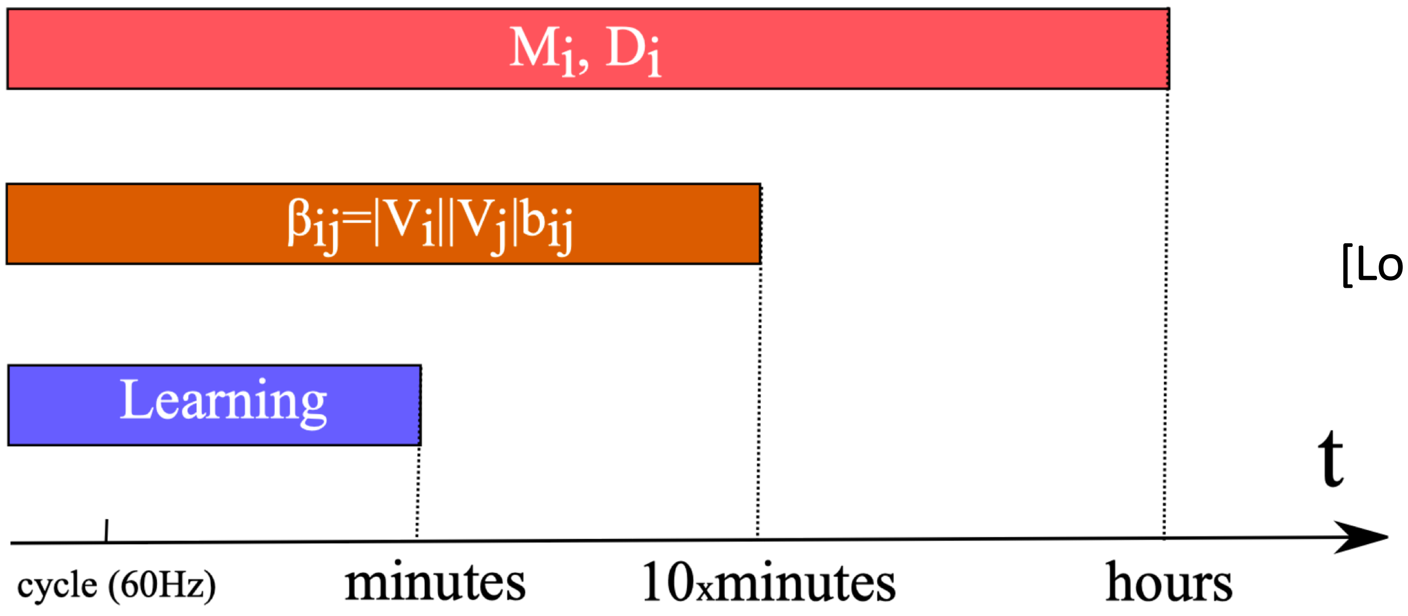
By default, the “true” model can be a very complicated non-linear object with memory:

$$X_{t+1} = F(X_t, X_{t-1}, \dots, X_{t-k}, \xi)$$

Linearization at short time scales and steady systems:

$$X_{t+1} = AX_t + BX_{t-1} + \cdots + \Sigma\xi_t$$

$$X_{t+1} = AX_t + \Sigma\xi_t$$



[Lokhov *et al.* PSCC 2018]

Method: maximum likelihood estimator

Proposition: using T observations $\{X_t\}_{t=1,\dots,T}$

$$\boxed{\widehat{A} = \operatorname*{argmin}_{A} \sum_{t=1}^{T-1} \|X_{t+1} - AX_t\|_2^2}$$

$$\Sigma_1 = \frac{1}{T-1}\sum_{t=1}^{T-1} X_{t+1}X_t^\top, \qquad \Sigma_0 = \frac{1}{T-1}\sum_{t=1}^{T-1} X_tX_t^\top$$

Unconstrained Maximum Likelihood (UML) estimator:

$$\widehat{A} = \Sigma_1\Sigma_0^{-1}$$

UML estimator: theoretical analysis

Theorem: bound on the expected estimation error

$$\|\widehat{A} - A\|_{\mathsf{F}} \leq \frac{\|B\|_2}{\epsilon\sqrt{T-1}} \sqrt{\mathbb{E}\left[\mathrm{Tr}(\Sigma_0)\right] \mathbb{E}\left[\|\Sigma_0^{-1}\|_{\mathsf{F}}^2\right]}$$

Remark: translates into an estimate on the amount of data that guarantees reconstruction to a given accuracy

[Lokhov *et al.* PSCC 2018], see also [Simchowitz *et al.* COLT 2018]

UML estimator: theoretical analysis

Theorem: bound on the expected estimation error

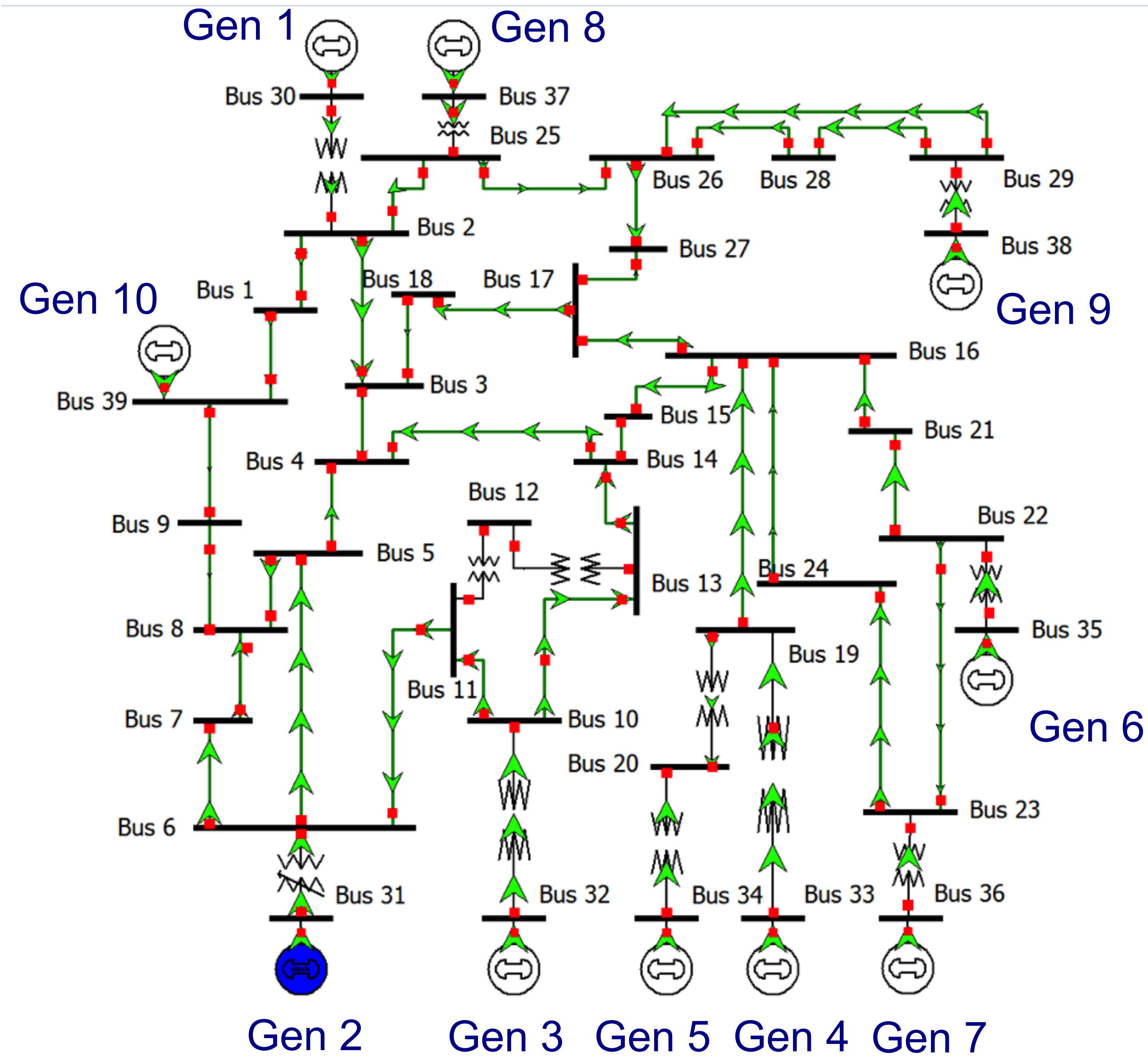
$$\widehat{A}_d = \frac{\widehat{A} - \mathbb{I}_{2N\times 2N}}{\Delta t}$$

$$\|\widehat{A}_d - A_d\|_{\mathrm{F}} \leq \epsilon^{-1} \sqrt{\frac{\sum_{i=1}^{N} M_i^{-2}\sigma_{P_i}^2}{\Delta t(T-1)} \mathbb{E}\left[\mathrm{Tr}(\Sigma_0)\right] \mathbb{E}\left[\|\Sigma_0^{-1}\|_{\mathrm{F}}^2\right]}$$

Remark: translates into an estimate on the amount of data that guarantees reconstruction to a given accuracy   Dependence only on $t_{\mathrm{obs}} = T\Delta t$

[Lokhov *et al.* PSCC 2018], see also [Simchowitz *et al.* COLT 2018]

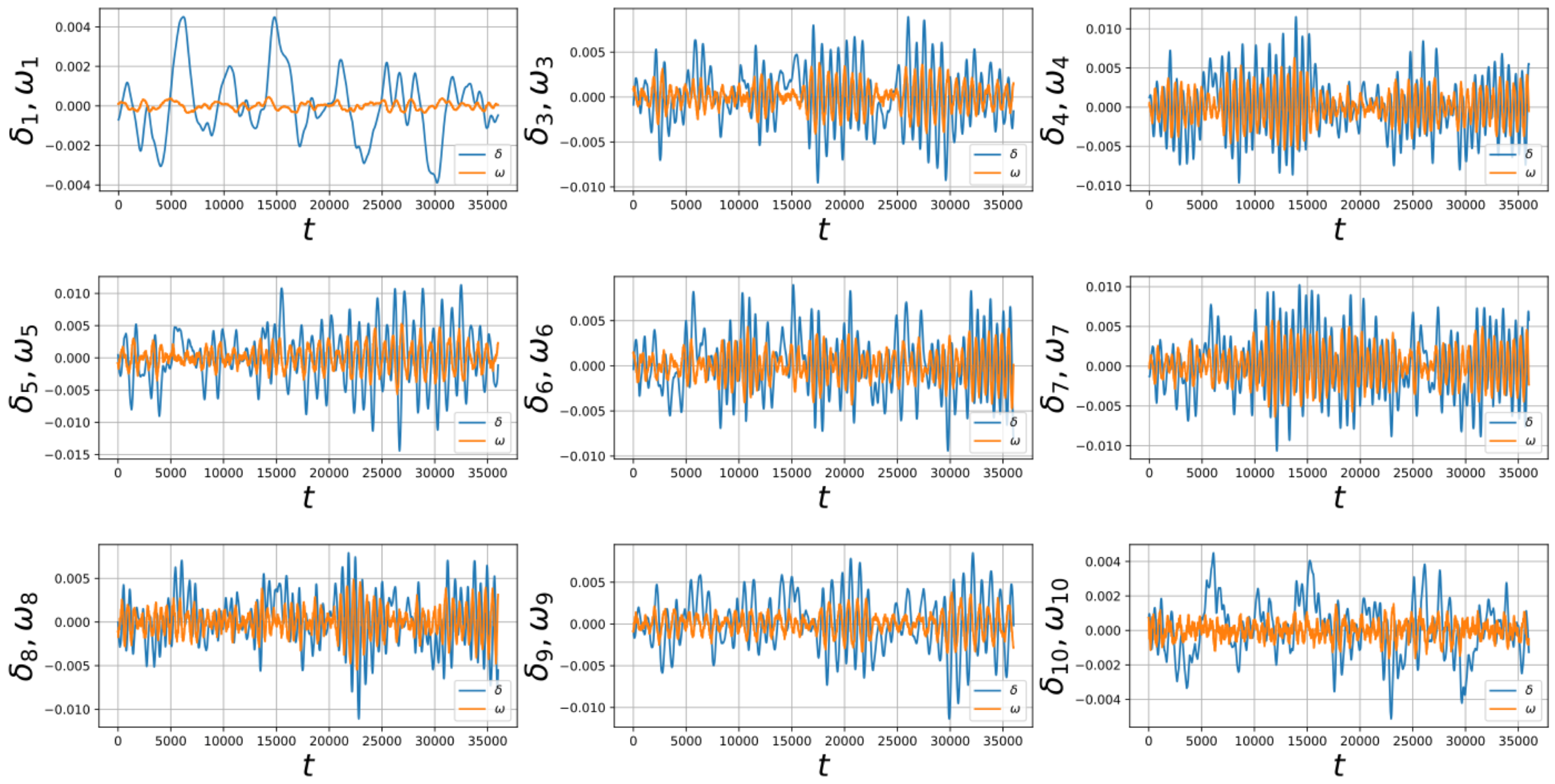
Numerical illustration: 39-bus system with 10 generators



[Lokhov *et al.* PSCC 2018]

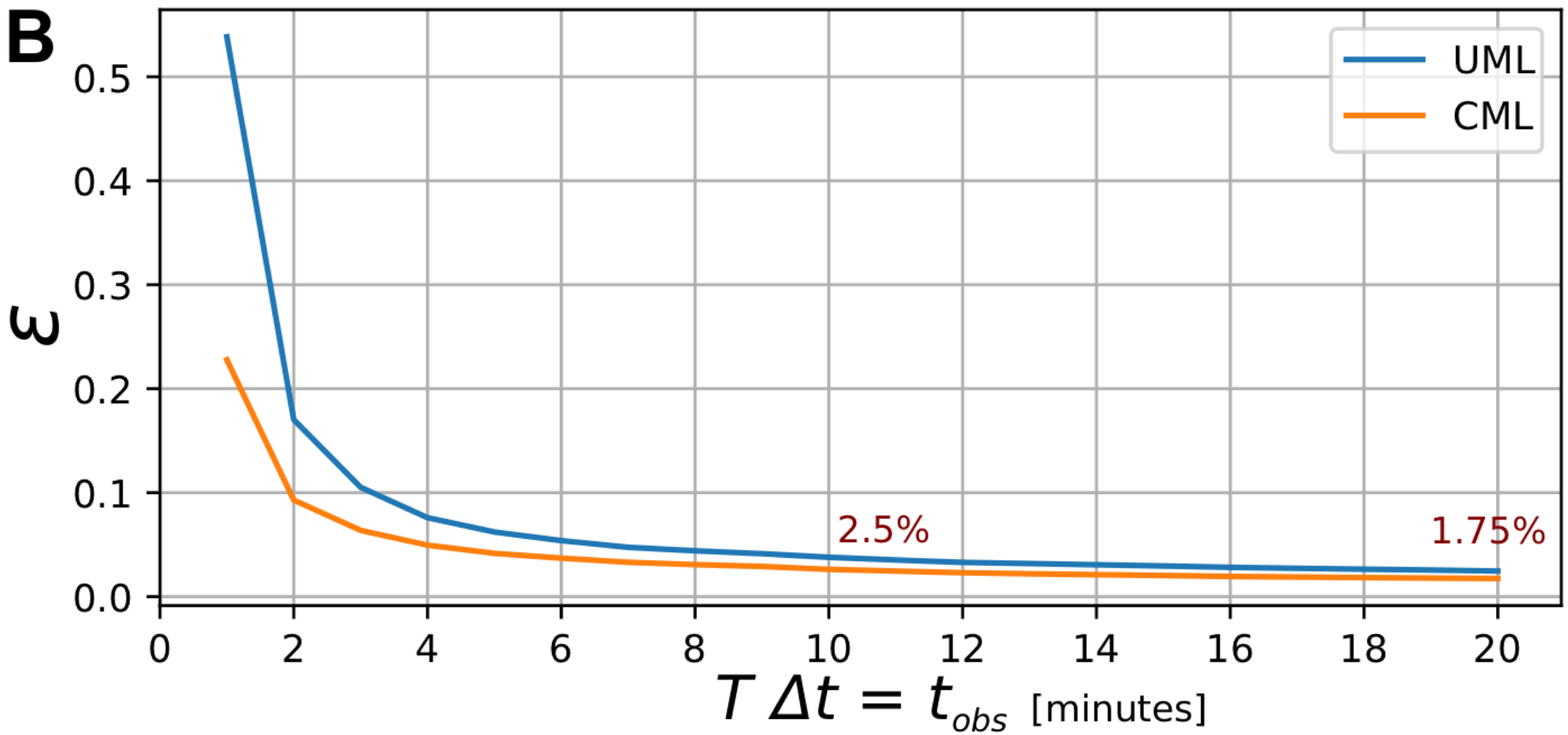
Numerical illustration: 39-bus system with 10 generators

Sampled at the smallest resolution, $\Delta t = 1/60$ sec (1 cycle), $\sigma_{P_i} = 0.01$ p.u.



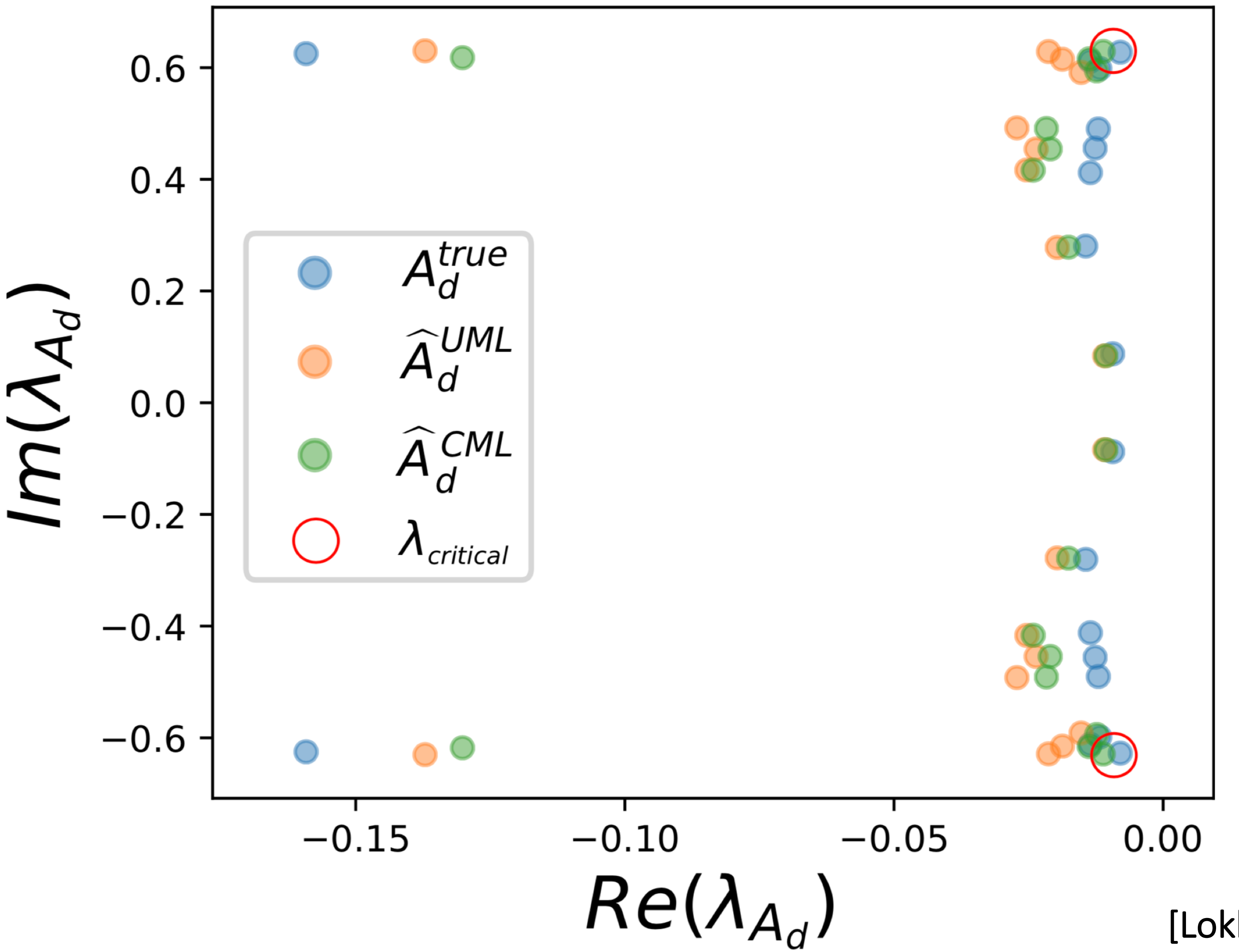
[Lokhov *et al.* PSCC 2018]

Dependence on total observation time t_{obs}



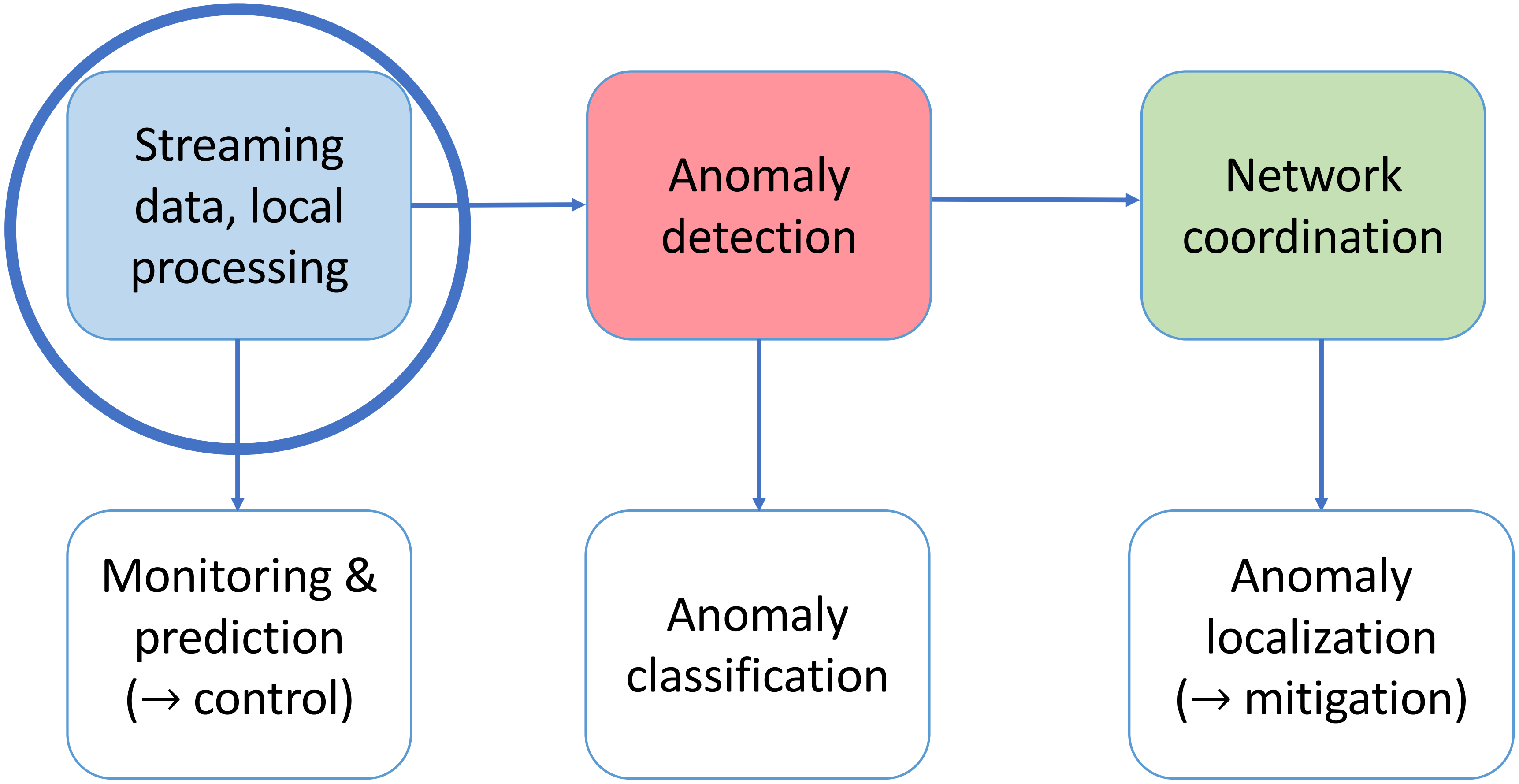
[Lokhov *et al.* PSCC 2018]

Prediction of critical eigenvalues

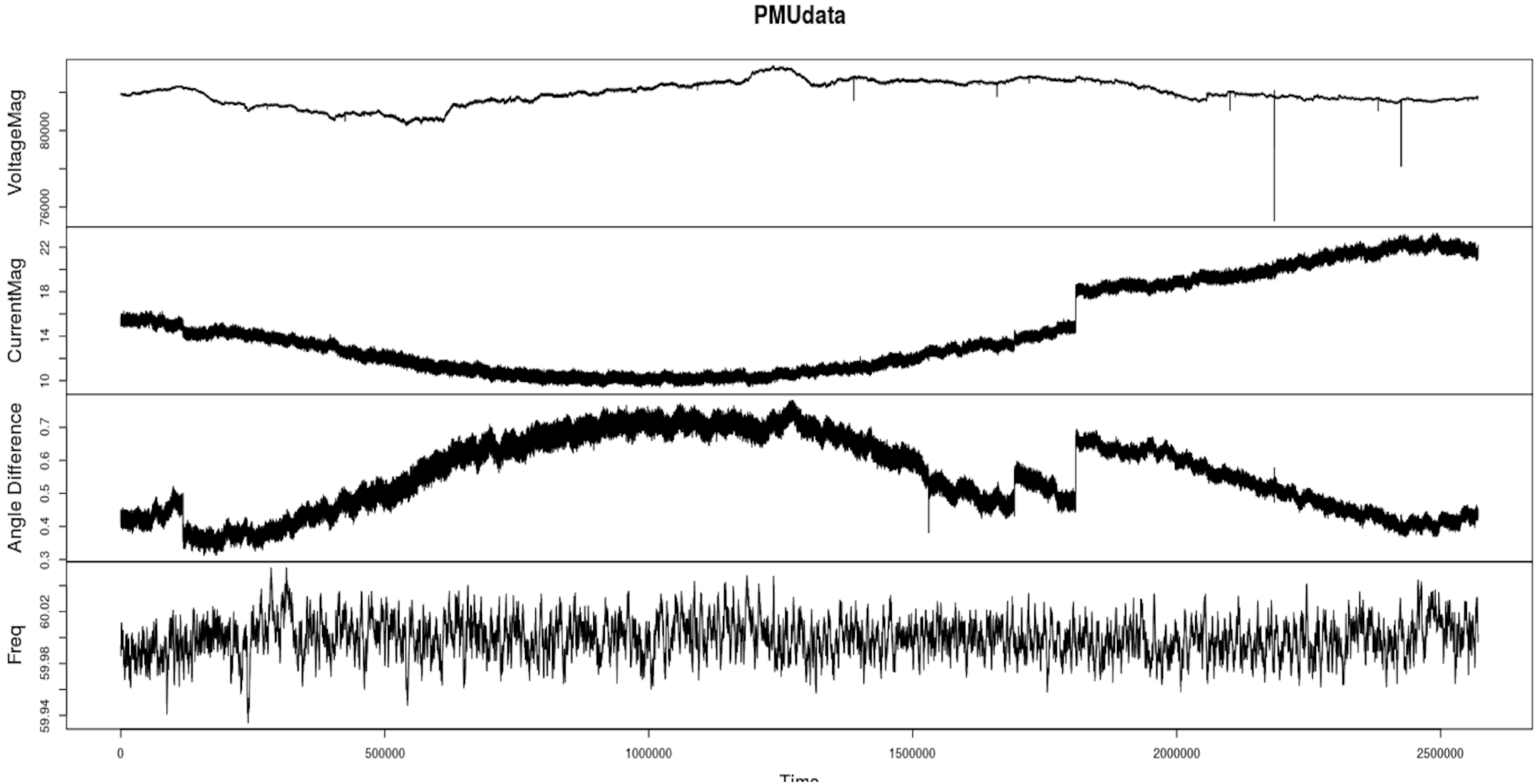


[Lokhov *et al.* PSCC 2018]

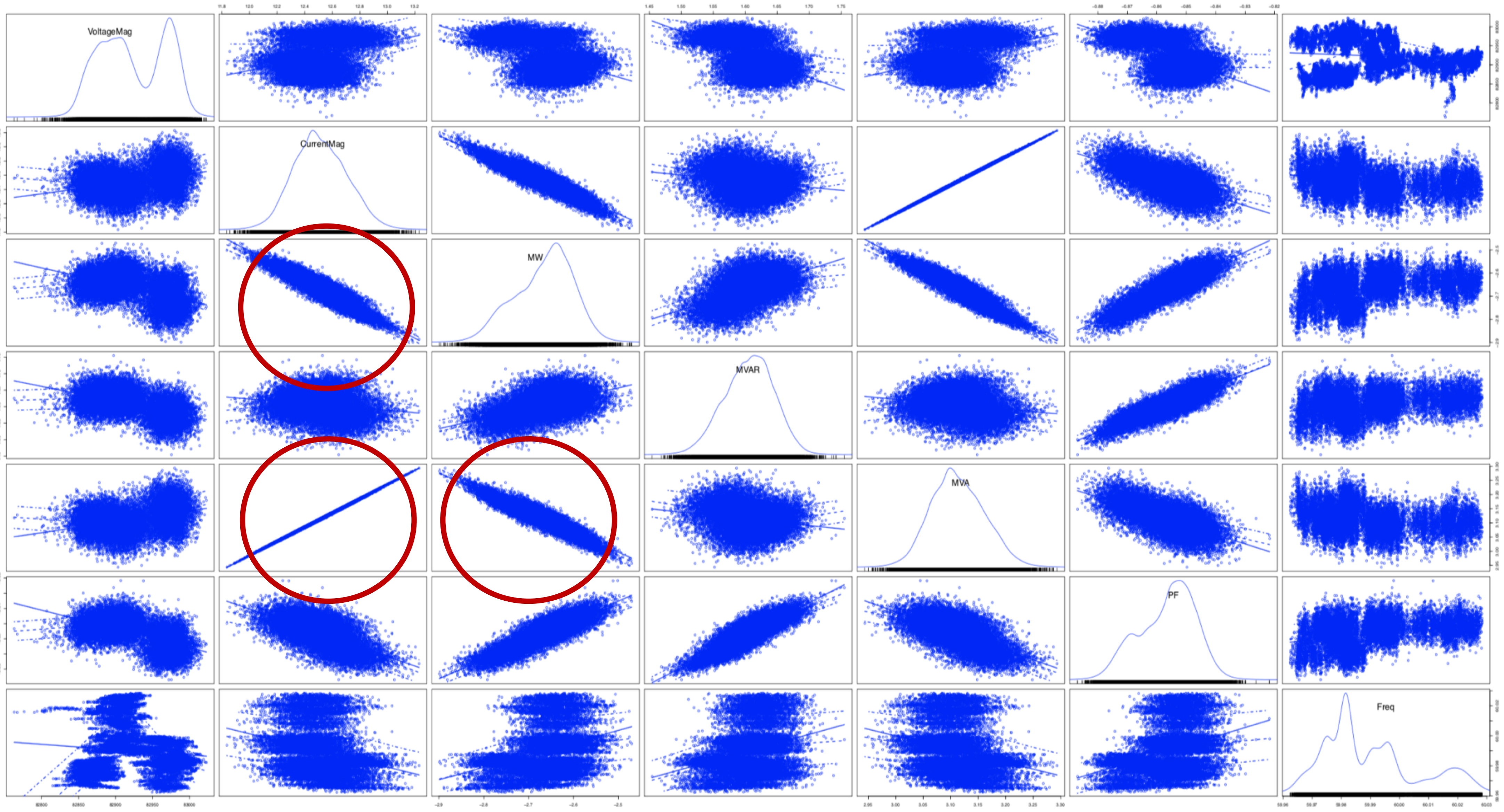
Statistical learning framework: local streaming processing



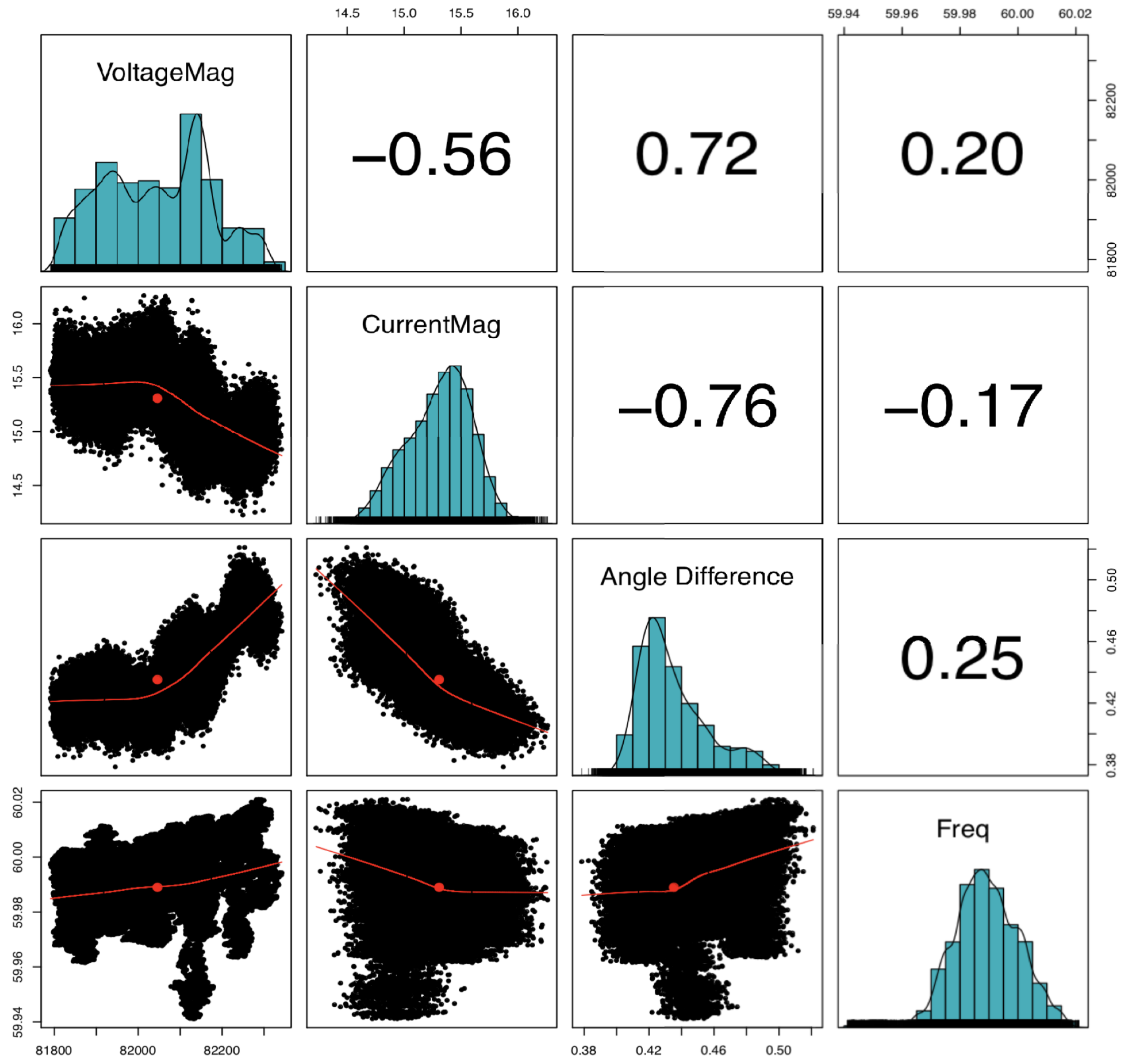
Example of raw PMU data



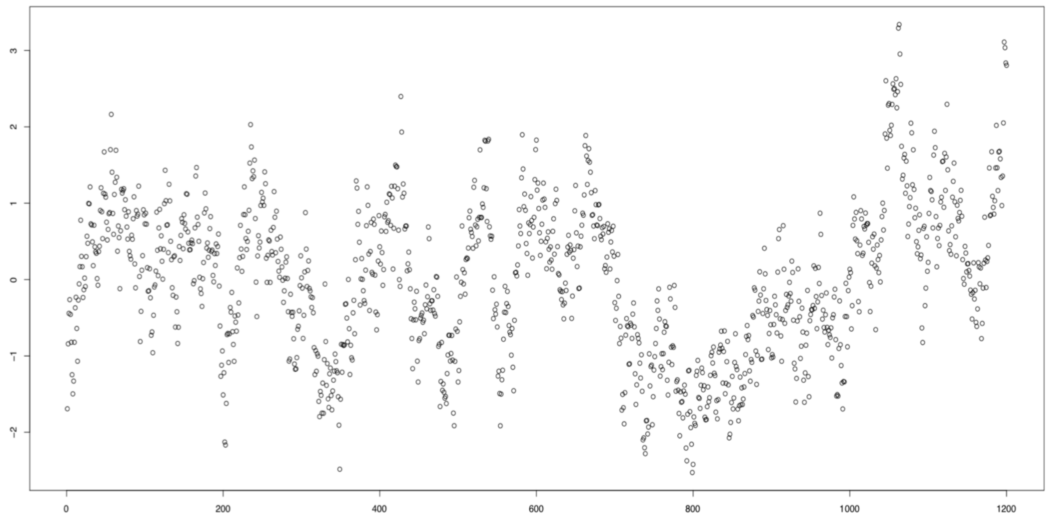
Eliminating dependent quantities that can be derived

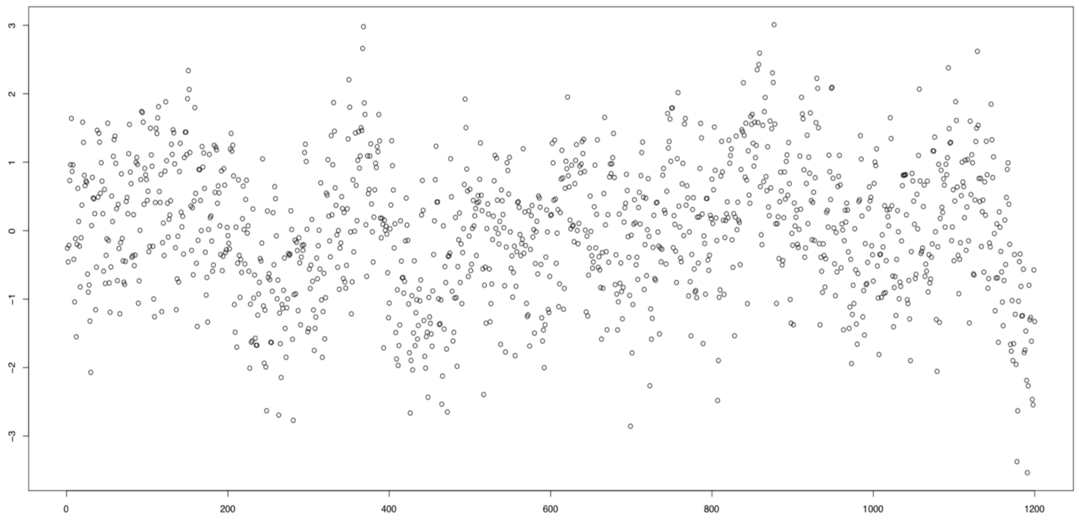


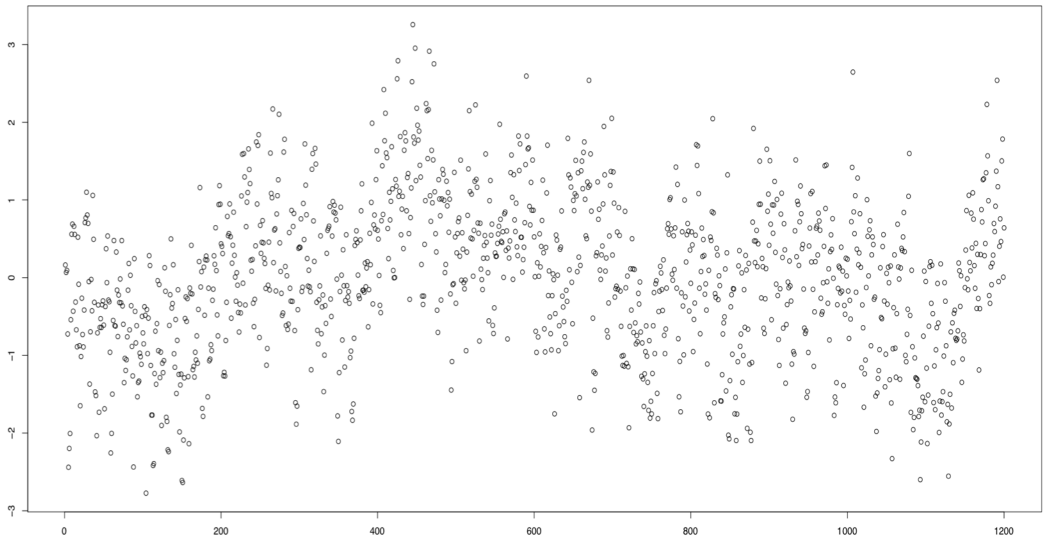
Data streams per PMU we will be working with

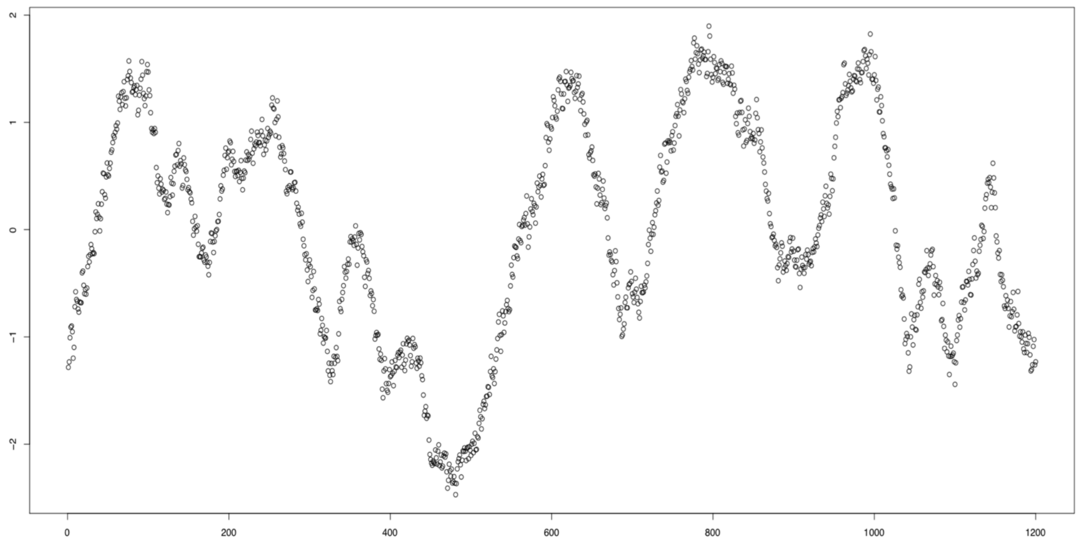


Data standardization (trend, variance, prediction time scale)

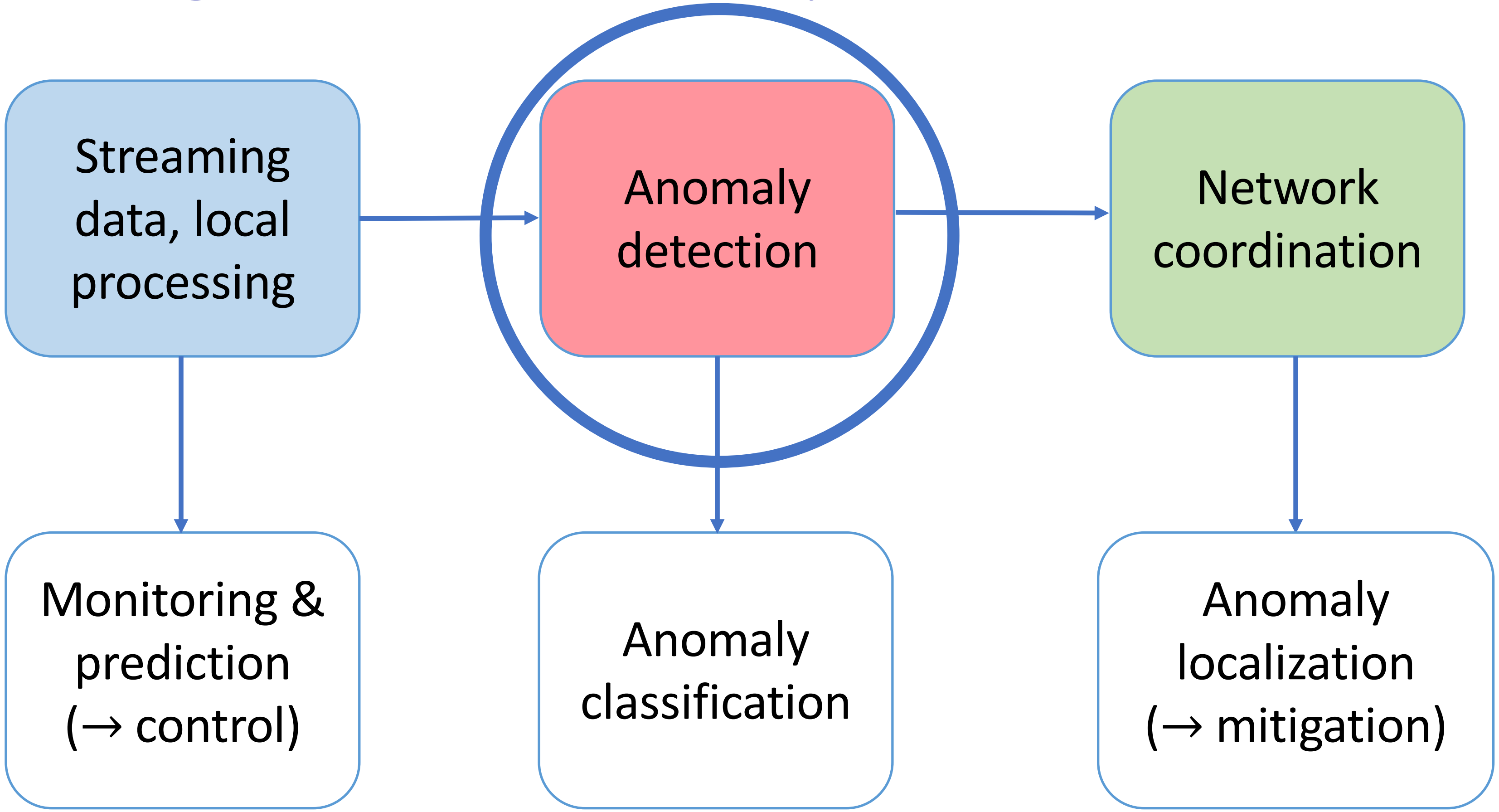








Statistical learning framework: anomaly detection



Learning a linear model over 10 minutes with prediction scale 0.5 sec

$$X_{t+1} = AX_t + \Sigma\xi_t$$

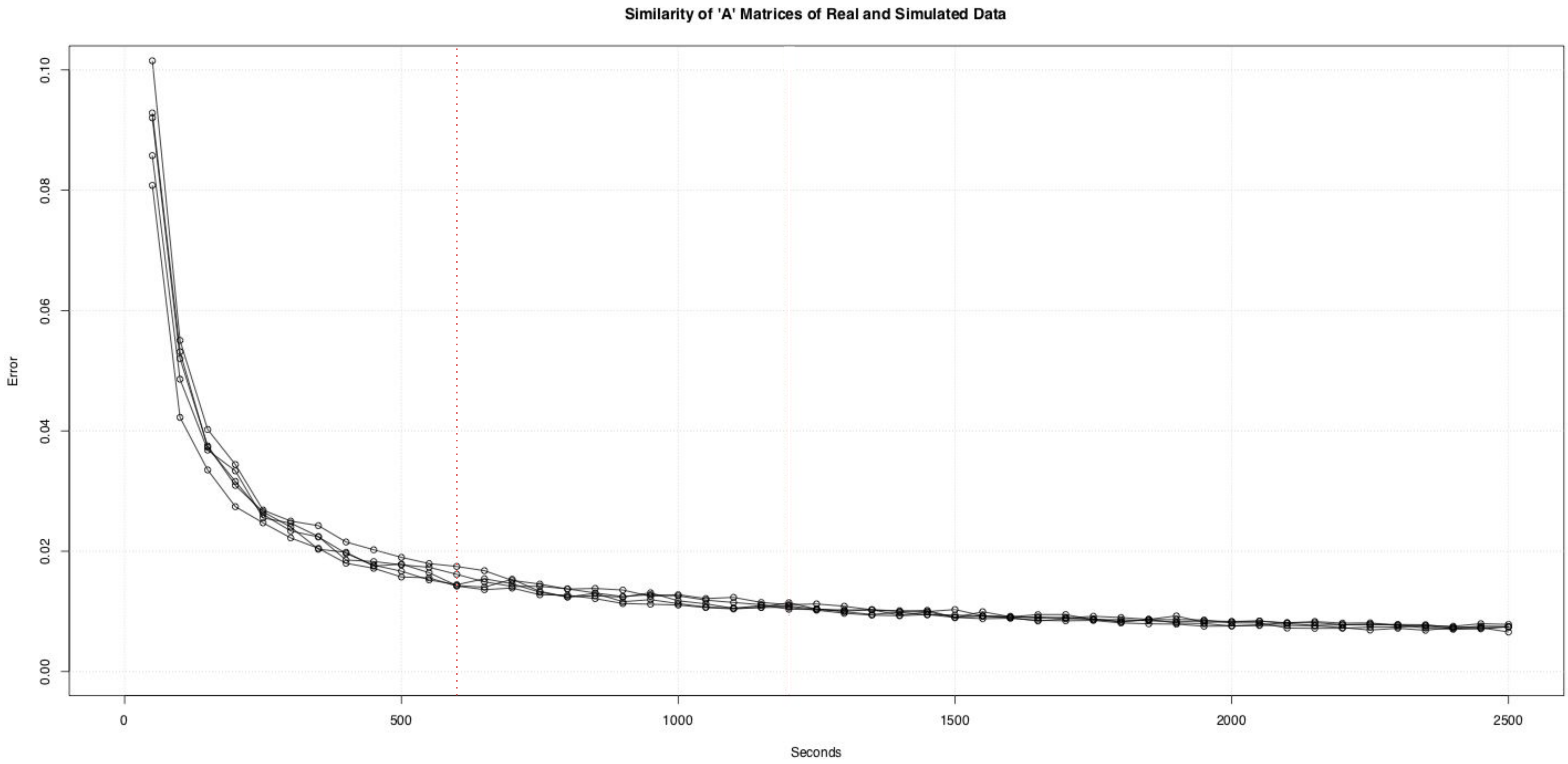
Example of an outcome (learning time is faster than real time 0.5 sec):

$$A = \begin{pmatrix} 0.882118137 & 0.012060263 & 0.006386051 & -0.029965801 \\ -0.014293795 & 0.478200526 & -0.192183794 & -0.017579340 \\ 0.022435590 & -0.080999570 & 0.598828527 & 0.001700428 \\ 0.002928399 & -0.002927597 & 0.001563269 & 0.996497662 \end{pmatrix}$$

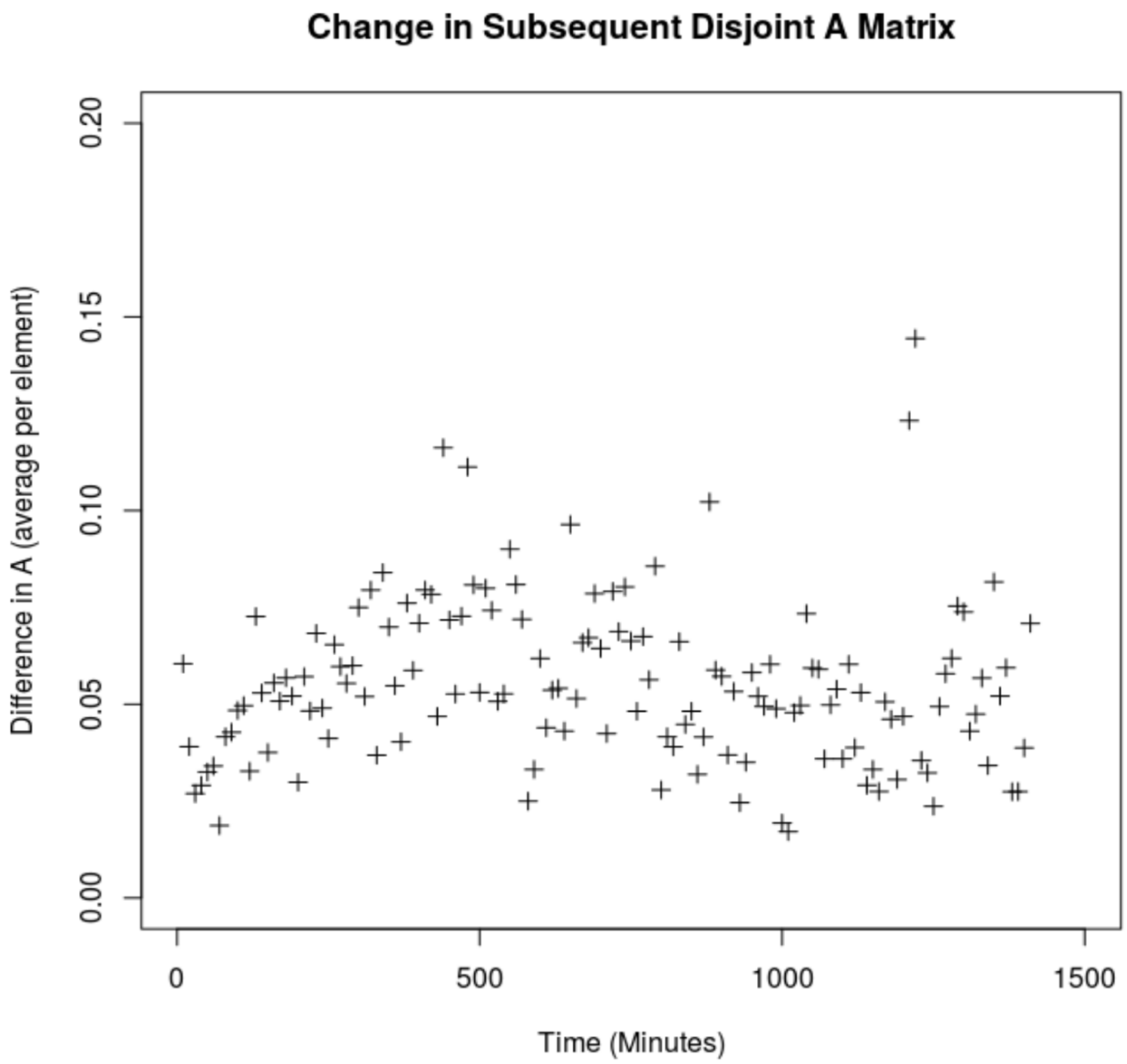
Residuals well described by a multivariate Normal distribution with non-trivial covariance matrix

Is 10 min enough to guarantee acceptable accuracy? Is linear model adequate?

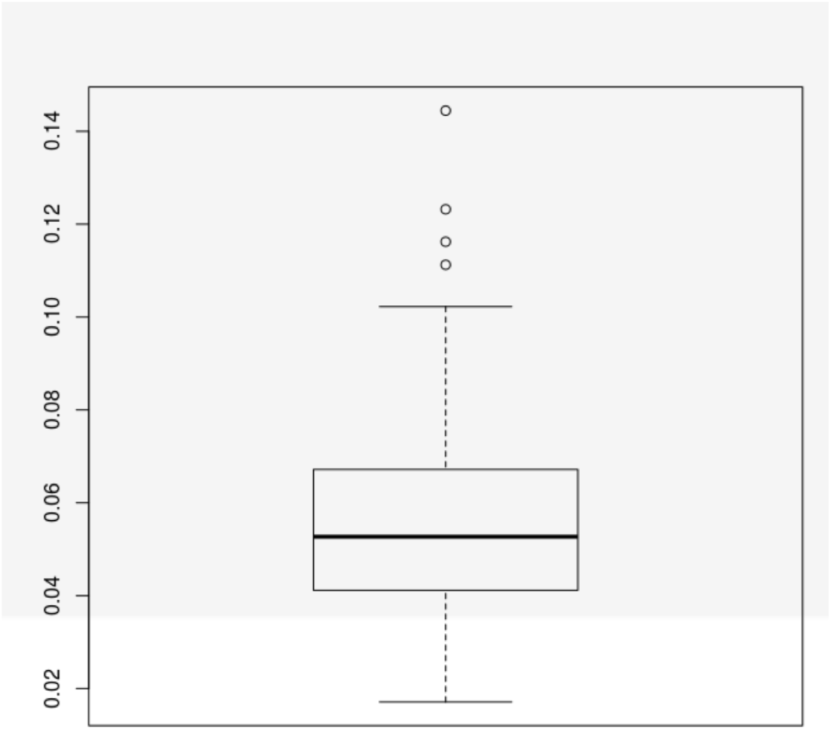
Empirical validation of T=10 minutes learning scale



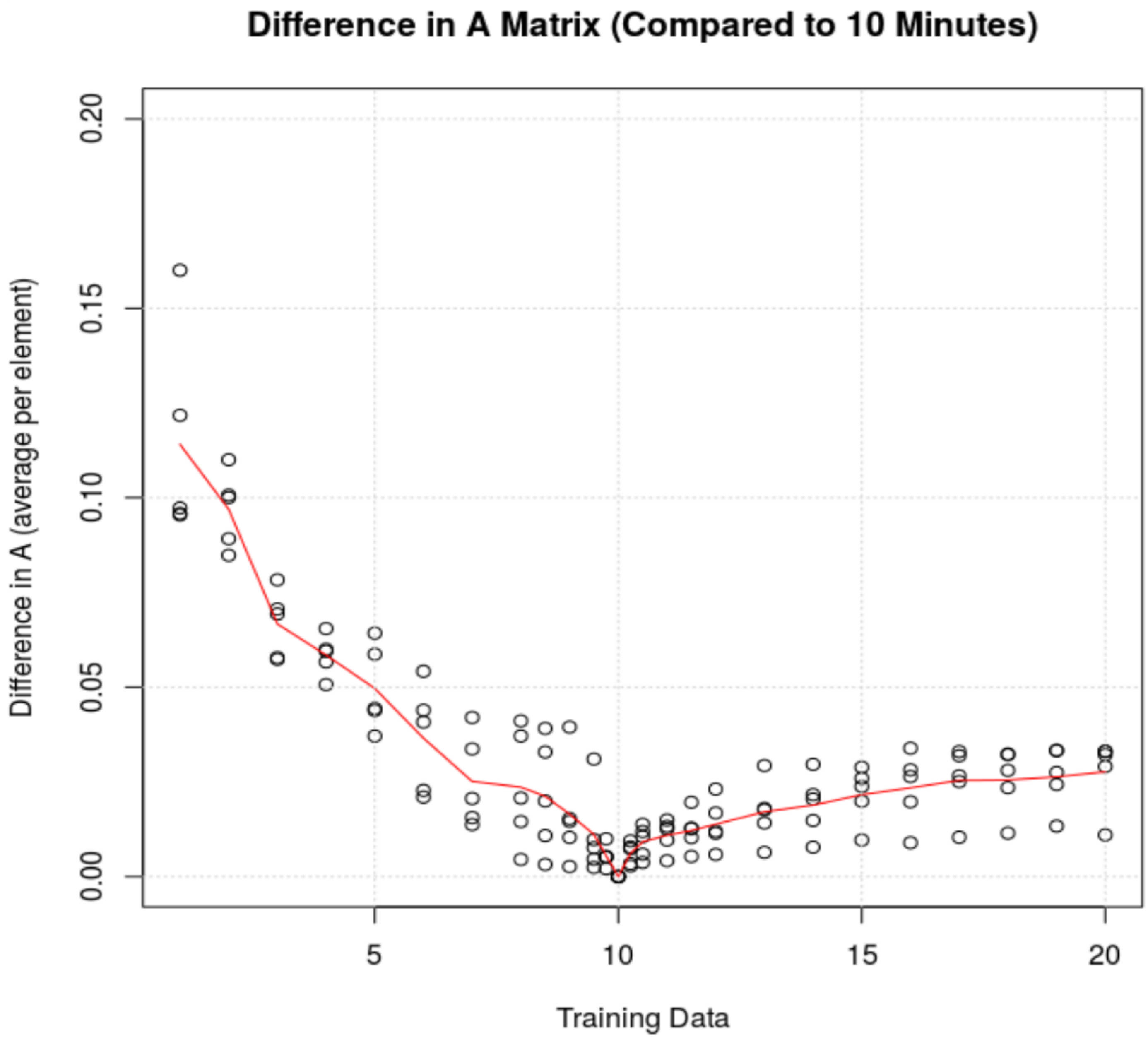
Why not taking larger T compared to 10 min?



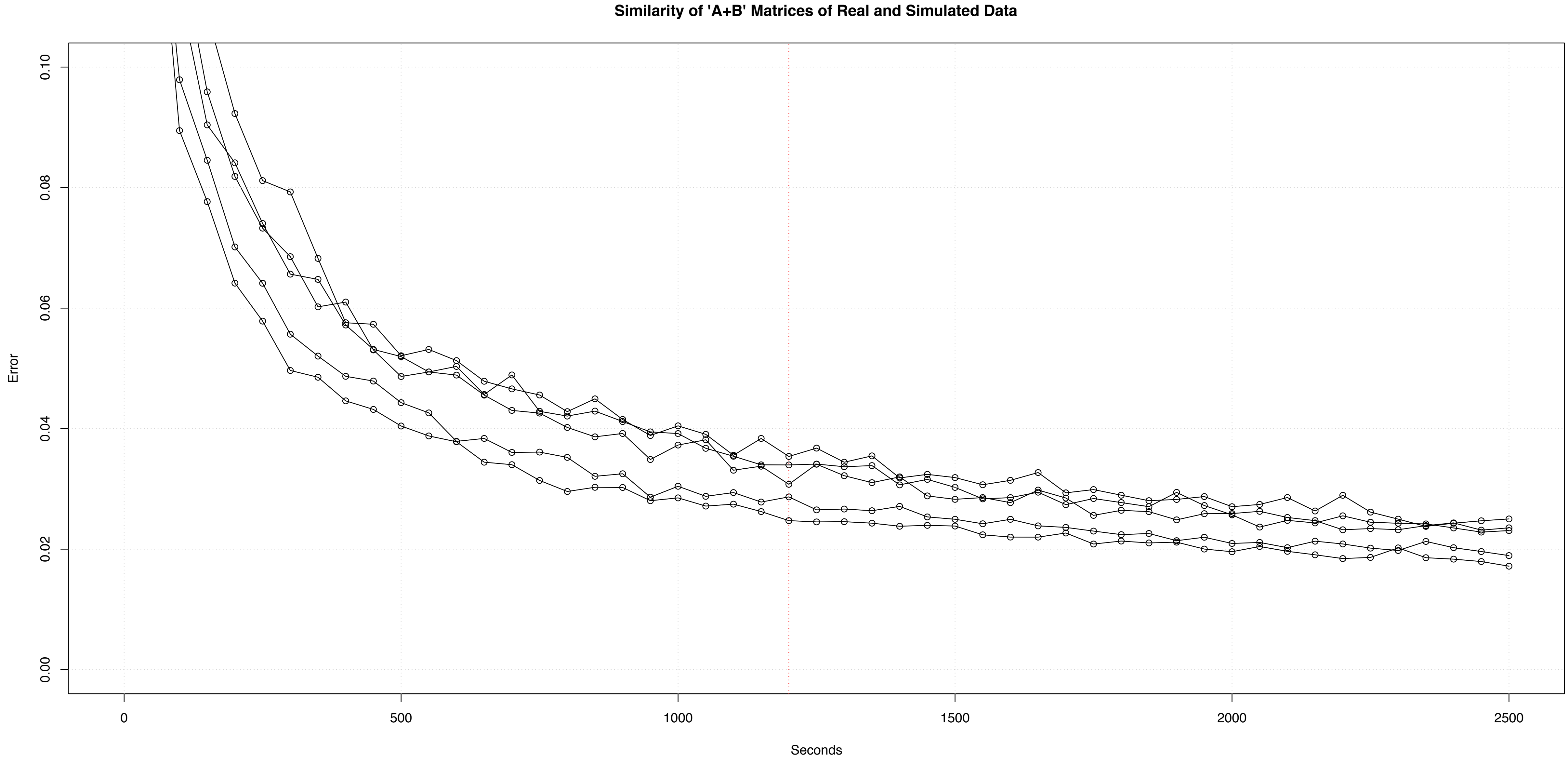
Model change is statistically significant beyong 10 min:



T=10 as a trade-off between accuracy and model consistency



Empirical validation of model adequacy



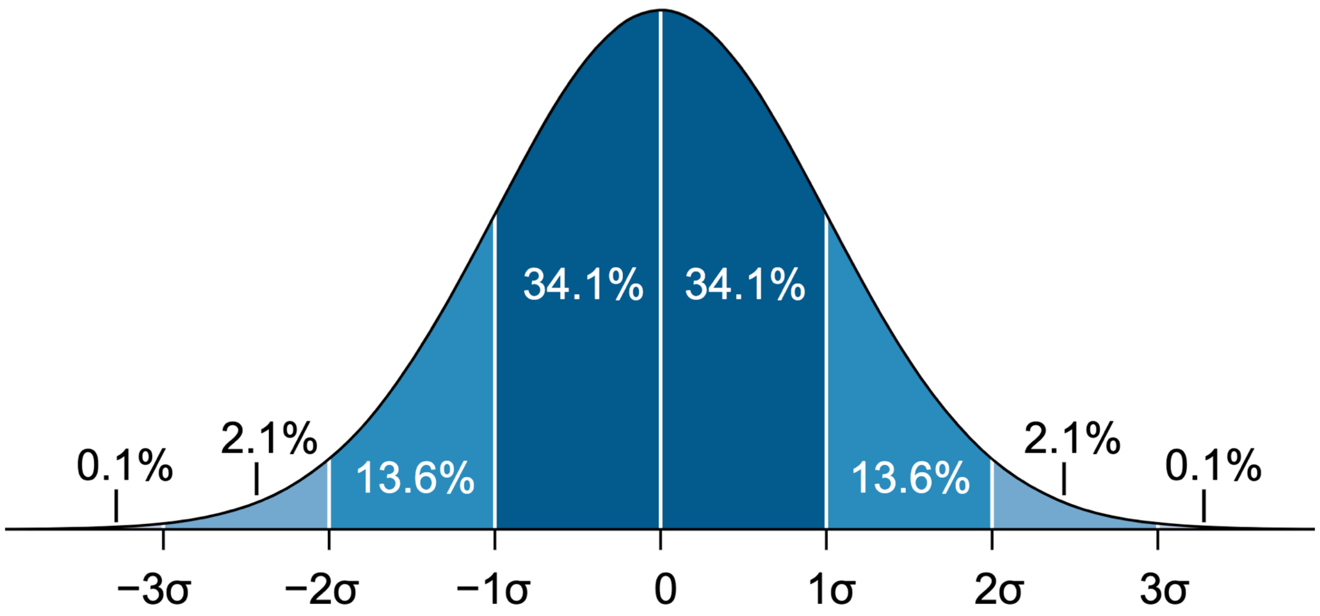
Anomaly detection and scoring

Multivariate case: use of Mahalanobis distance

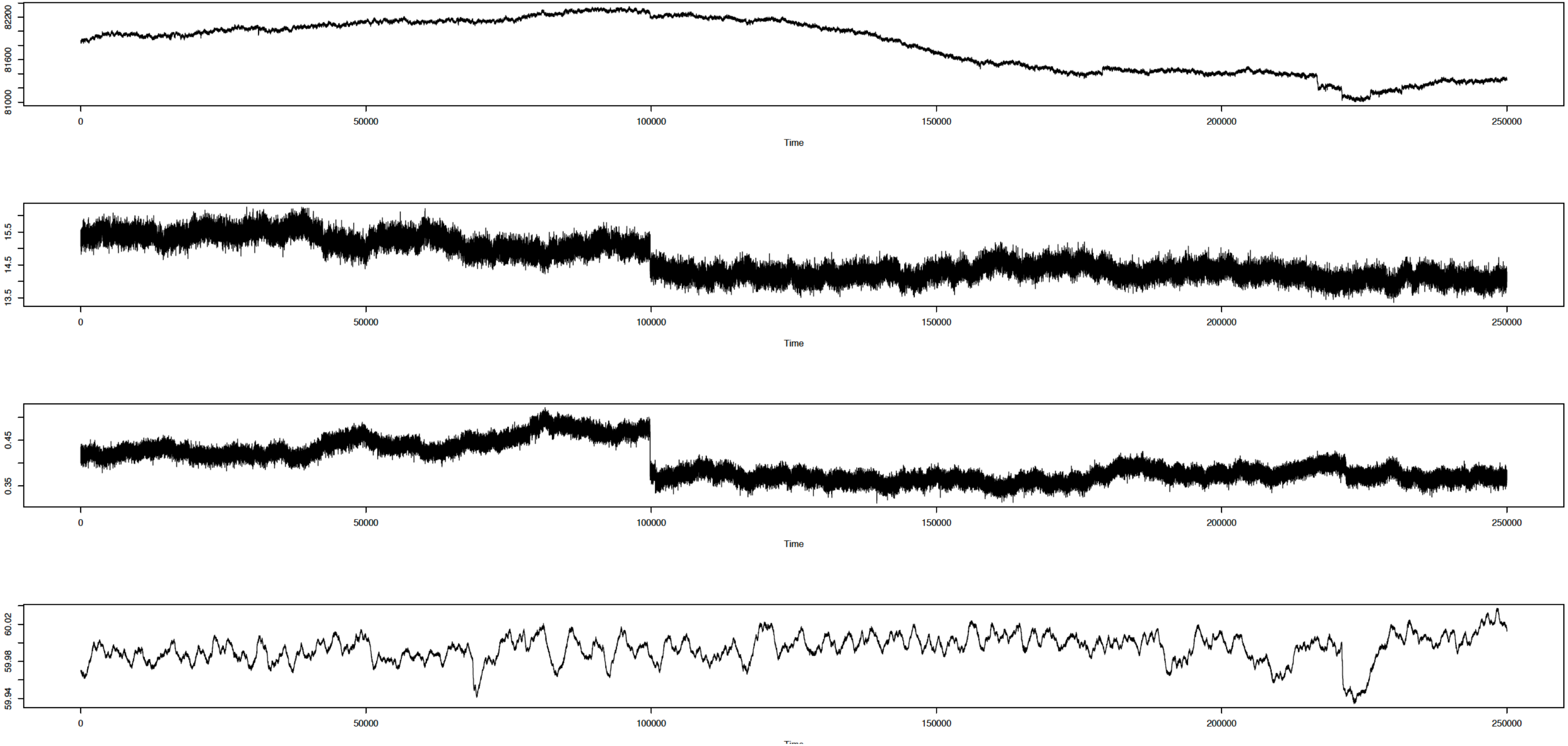
$$X_{t+1} - AX_t = \Sigma\xi_t \propto \mathcal{N}(0, \Sigma^2)$$

$$D_t = \sqrt{(X_{t+1} - AX_t)^T\, \Sigma^{-2}(X_{t+1} - AX_t)}$$

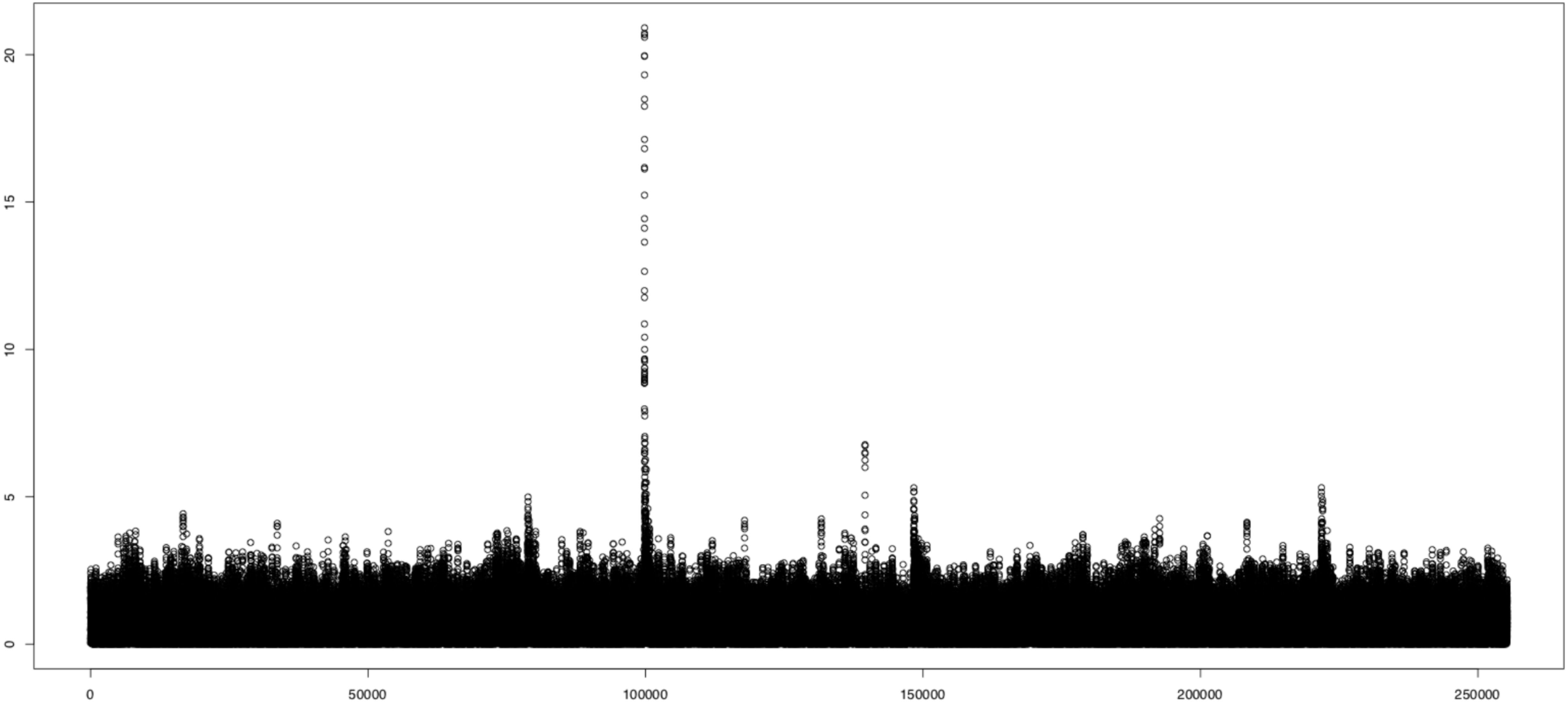
For single streams, use of the conditional distribution, reduction to a simple one-dimensional problem:



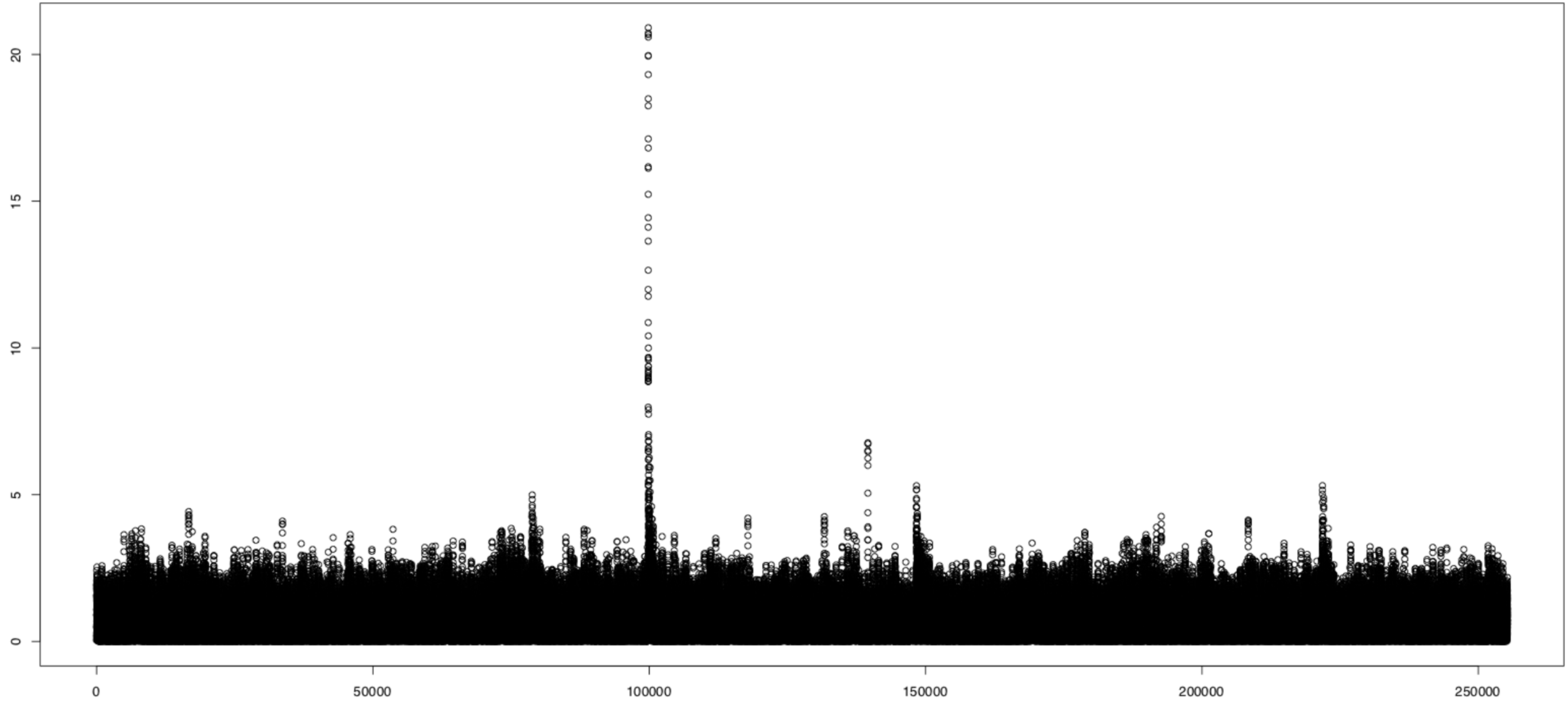
Anomaly scoring: one day example



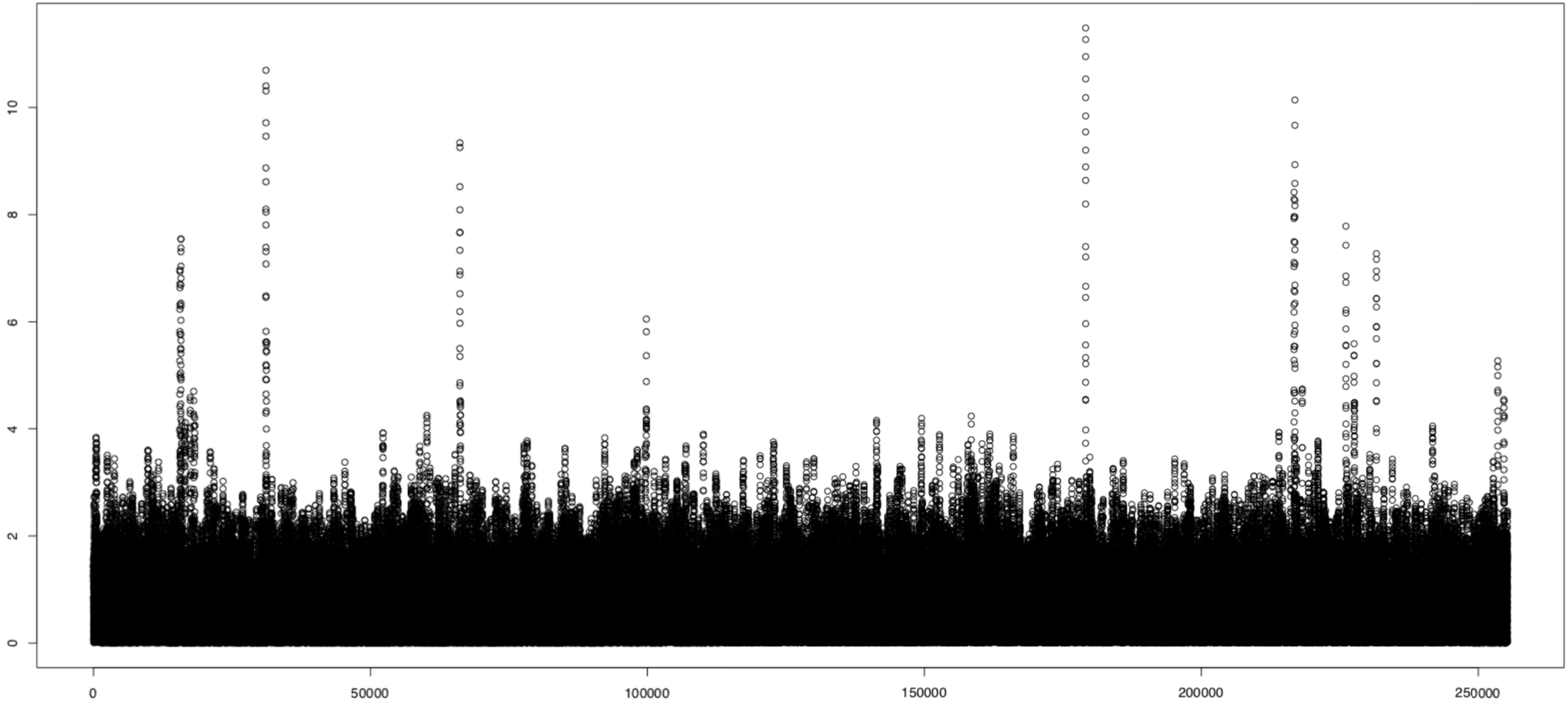
Anomaly scoring: current



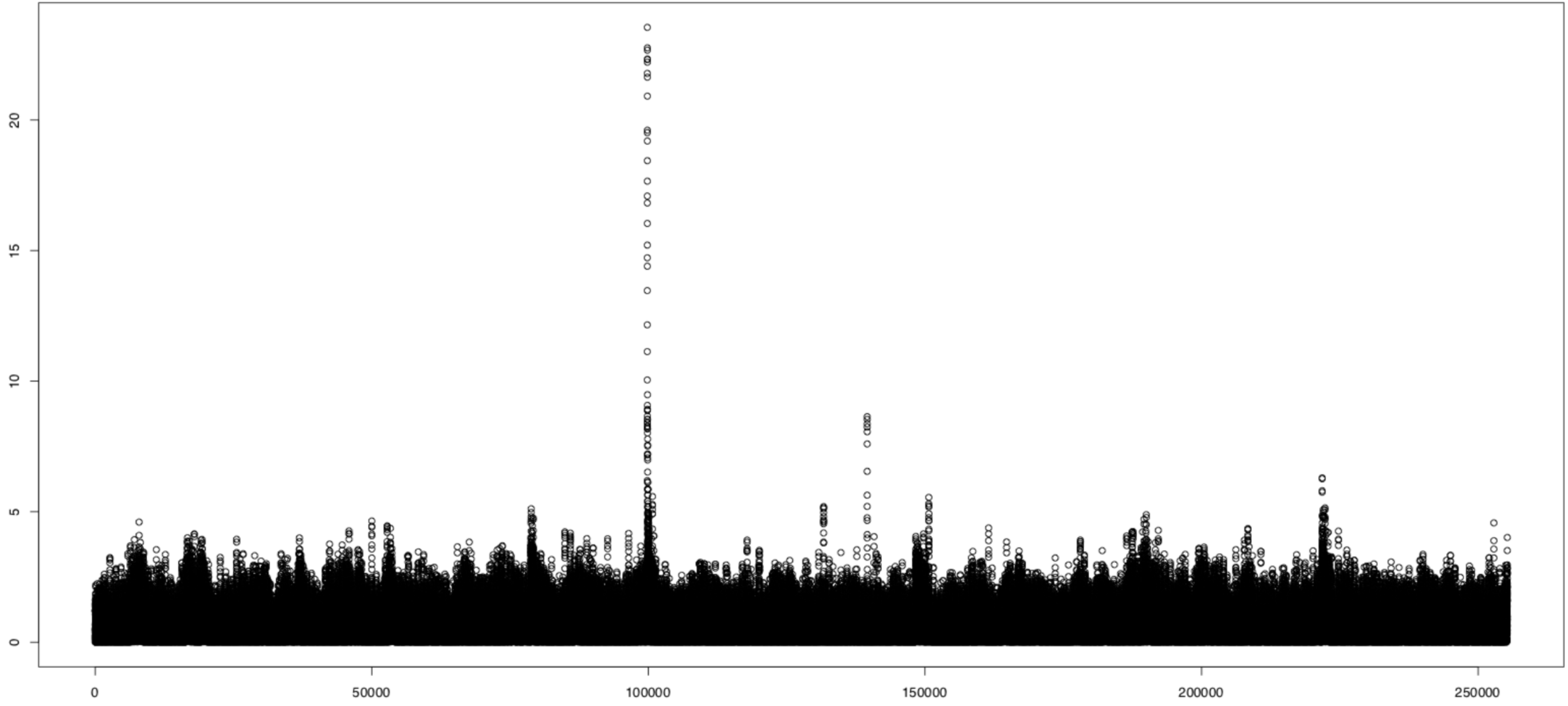
Anomaly scoring: current



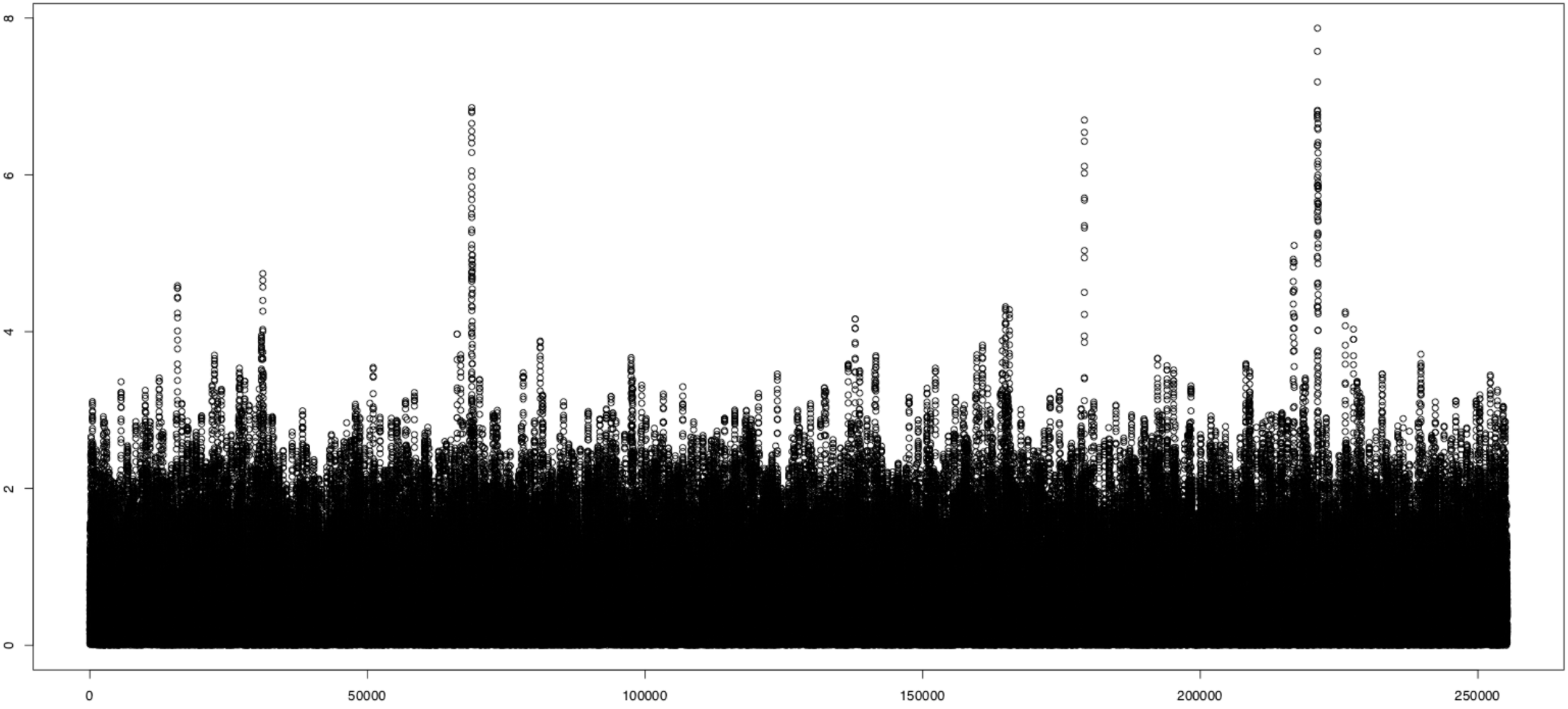
Anomaly scoring: voltage



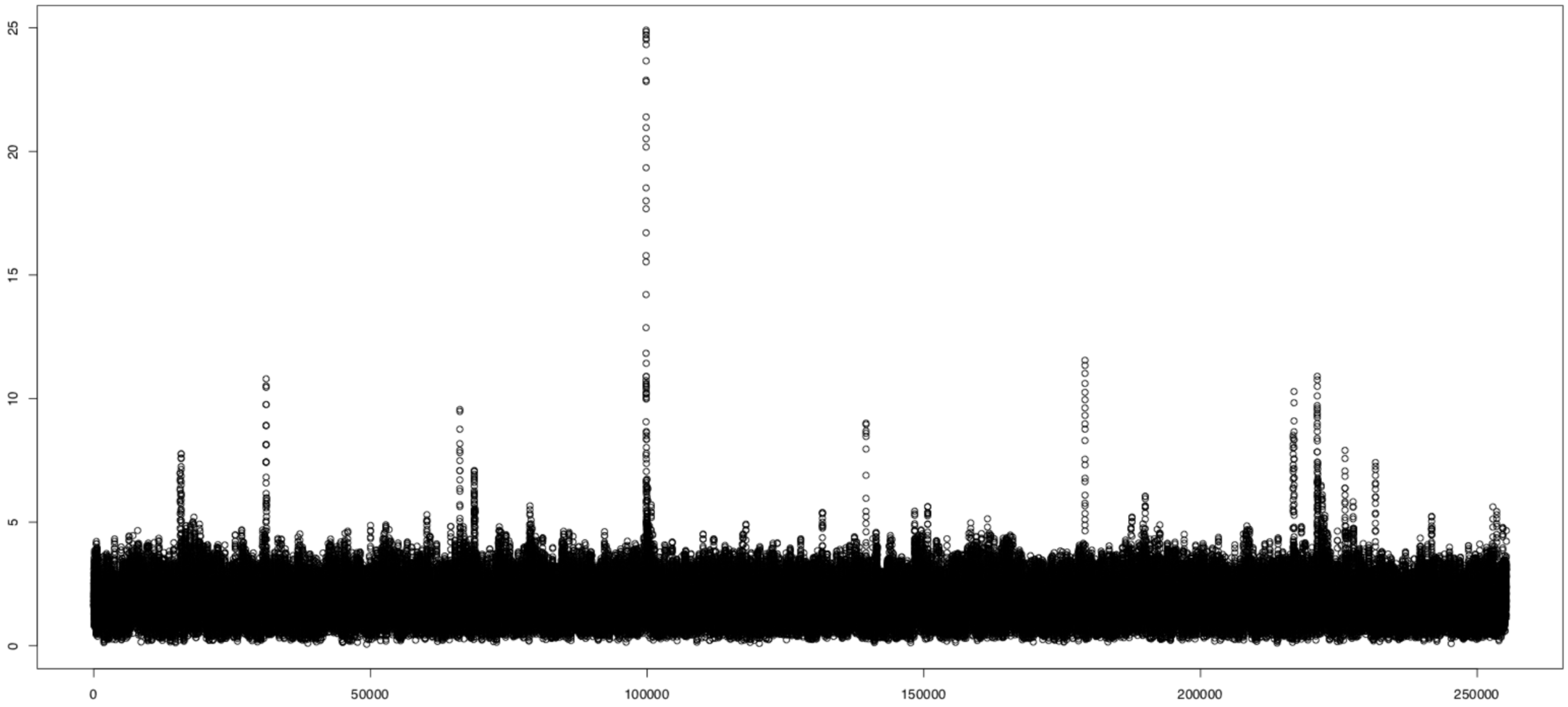
Anomaly scoring: angle



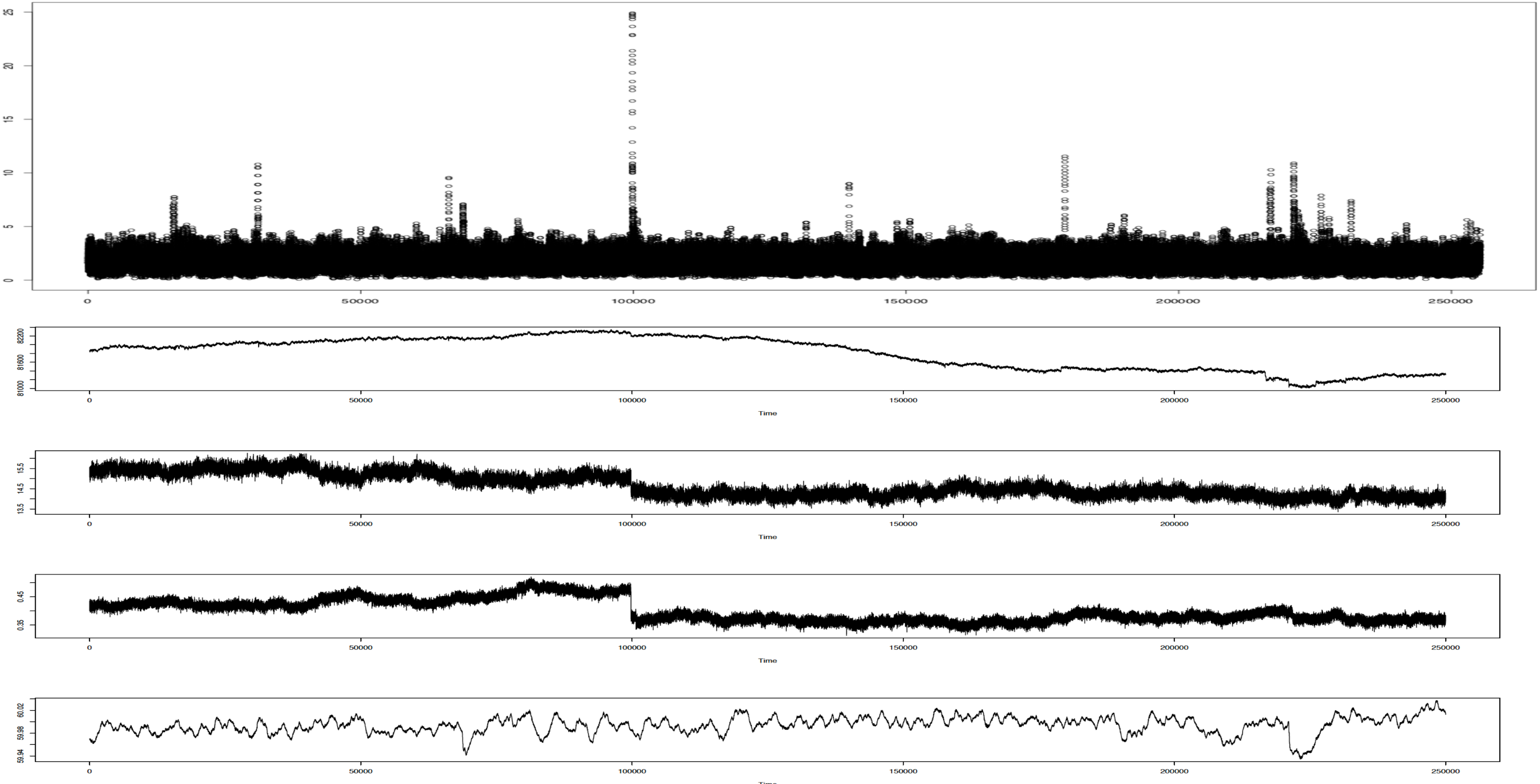
Anomaly scoring: frequency



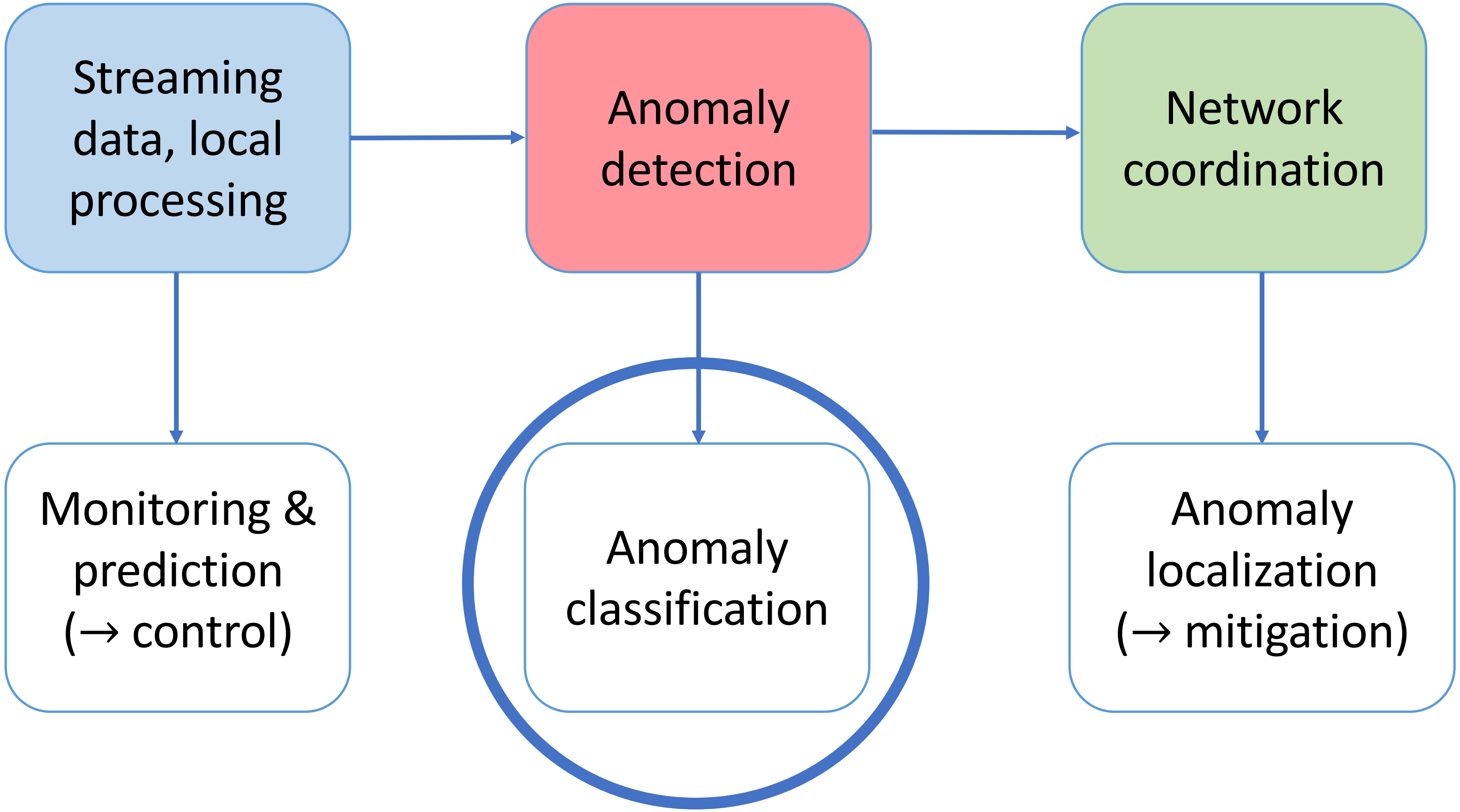
Anomaly scoring: multivariate



Anomaly scoring: comparison with actual time series



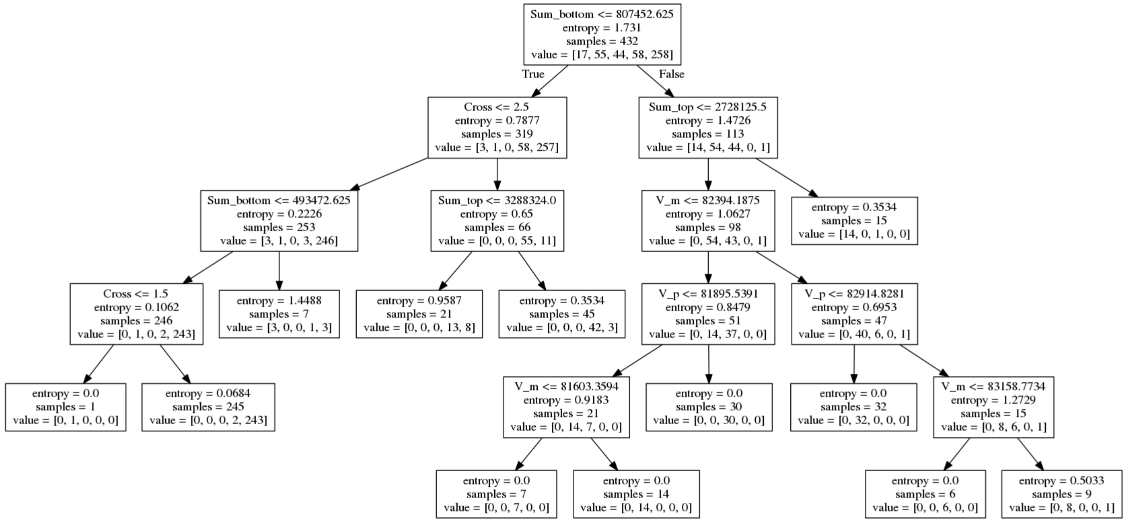
Statistical learning framework: anomaly classification (preliminary)

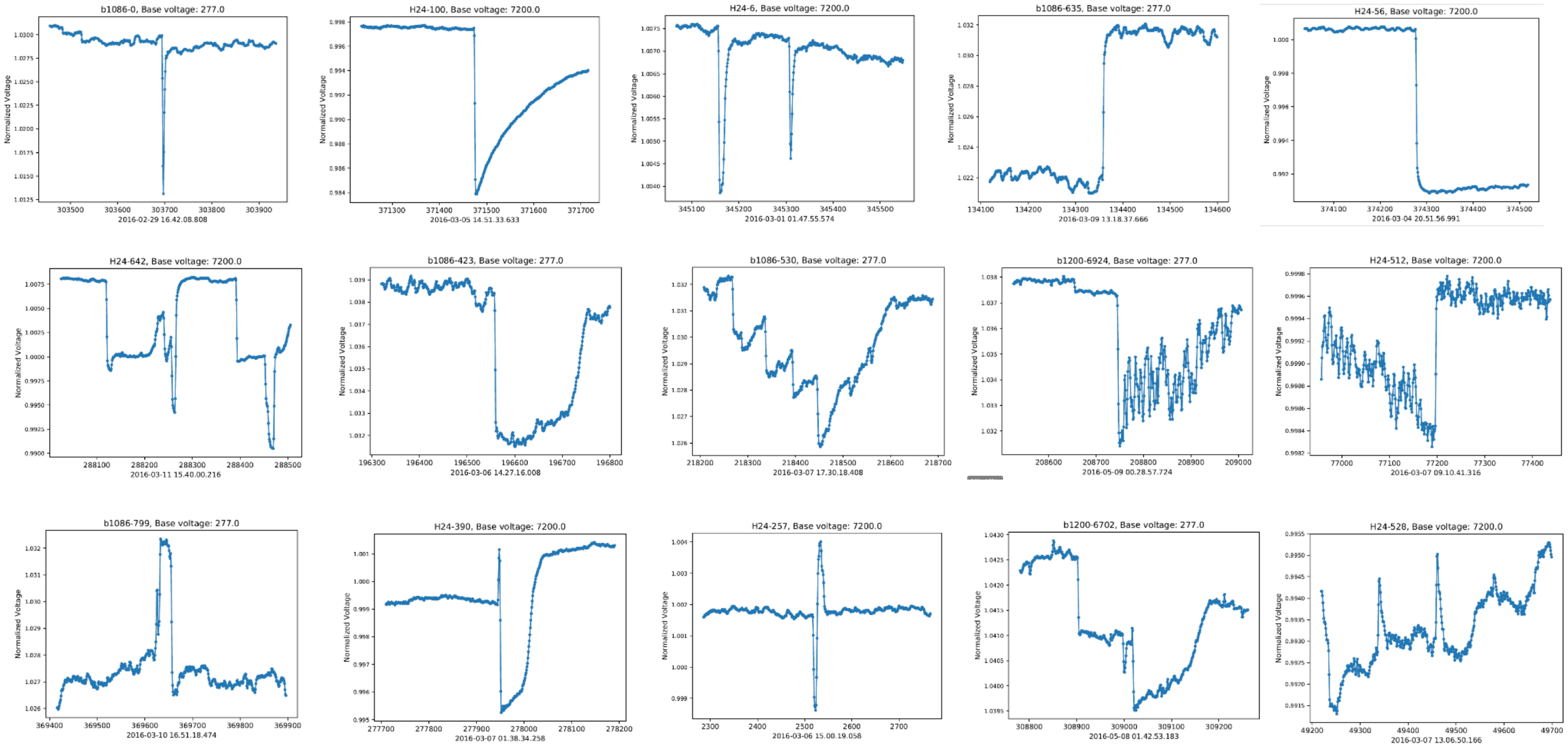


Anomaly classification

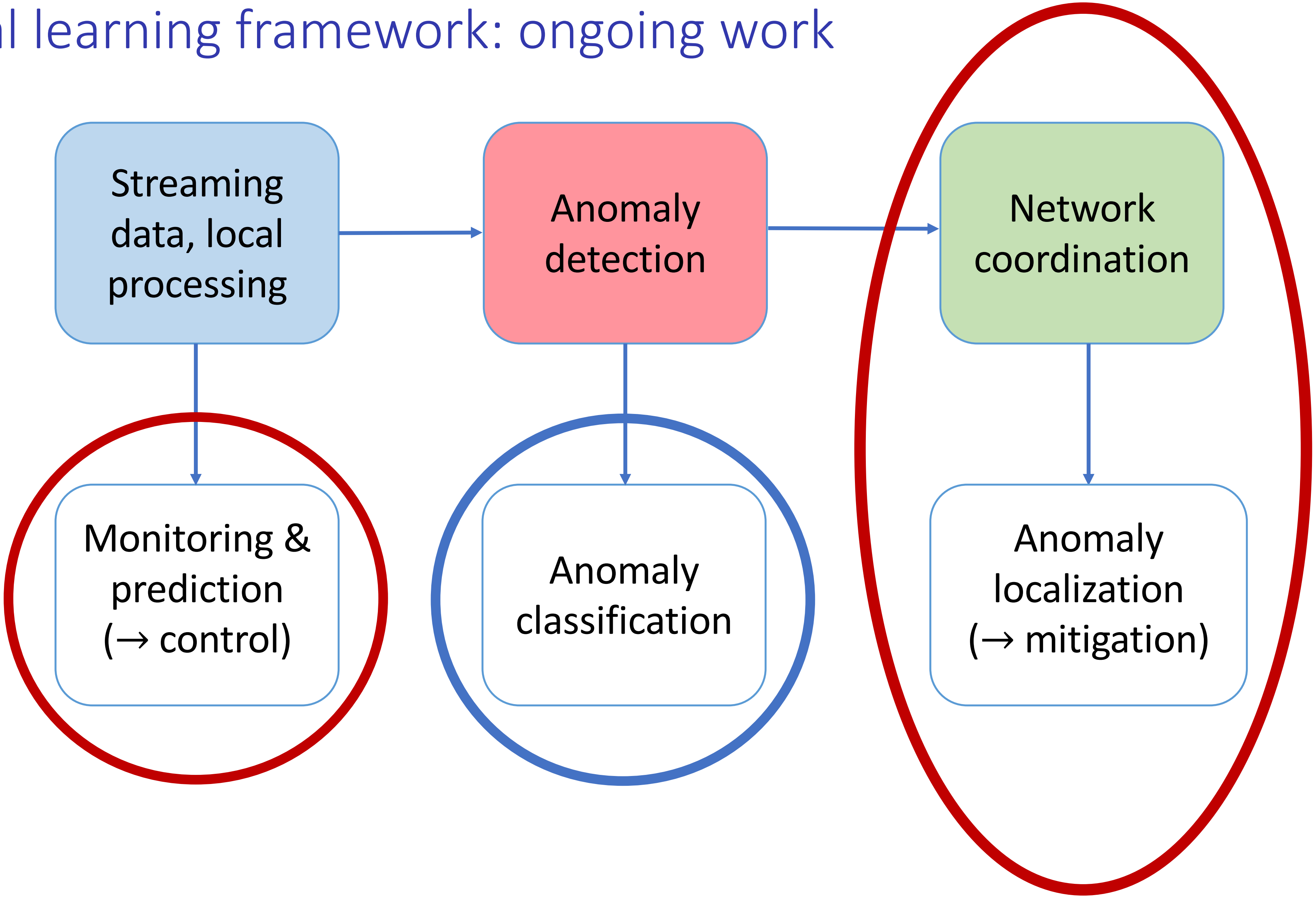
Focus on interpretability

Features extracted through detection





Statistical learning framework: ongoing work



Outlook

- Need **feedback** from NASPI community on the developed framework (constructive critique)

- **Finalize** the work: classification and prediction phases, tests on different data sets (Partner utilities, UC Riverside, Illinois Institute of Technology), software release, update with network-wide algorithms

- **Acknowledgement**: GMLC project on Machine Learning for Distribution Grids