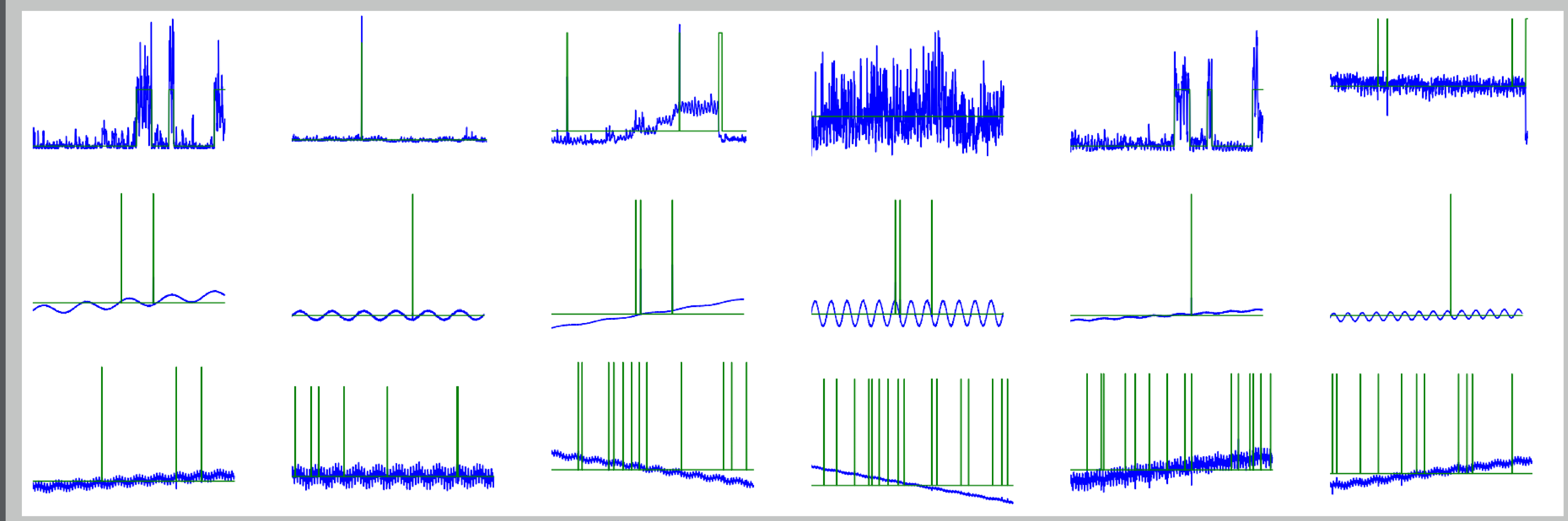
Multi-scale streaming anomalies detection for time series

B Ravi Kiran
Université Lille 3, CRIStAL, Lille, France

Université de Lille 3 Sciences humaines et sociales

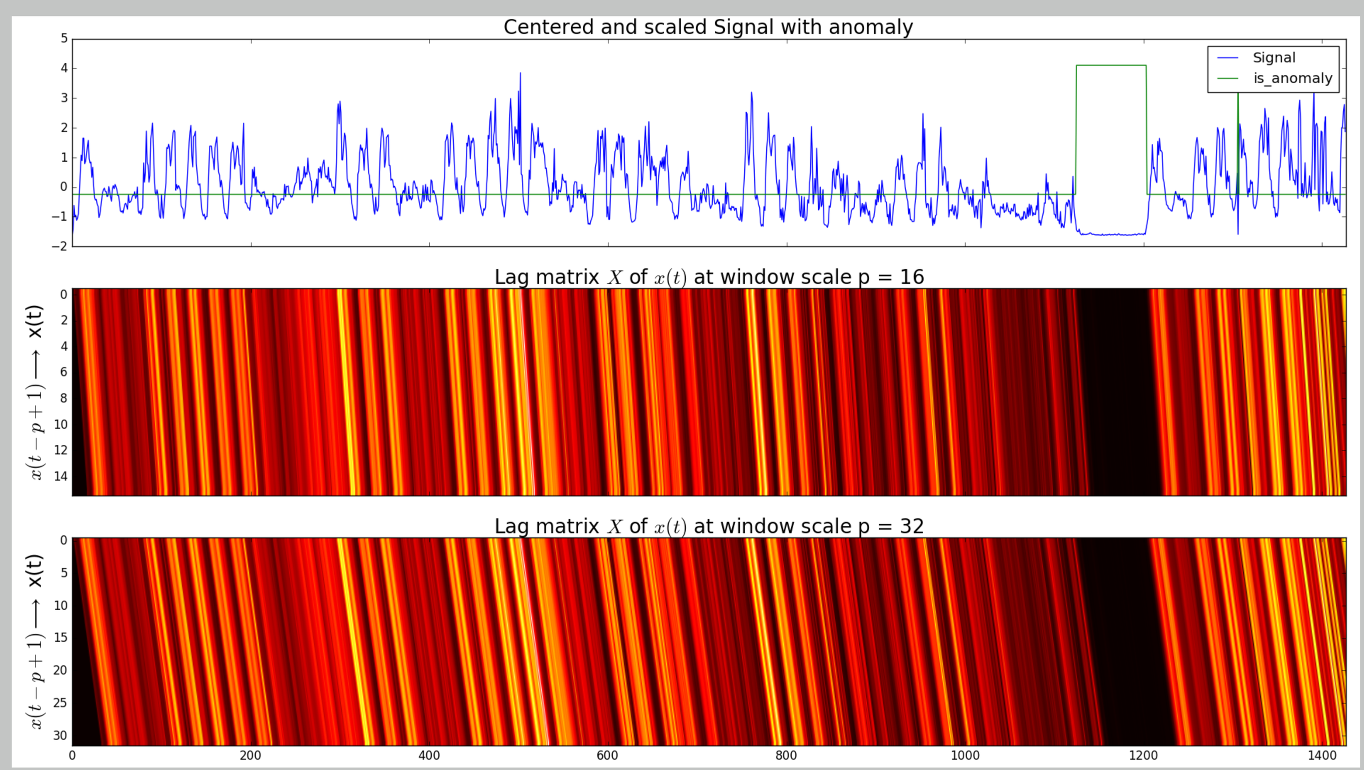
Streaming anomaly detection

- $x(t)$ are observations over time t where new data arrives over time t.
- **Problem** : Window $x(t), [t : t-p+1]$ which "deviates" from normal behavior.
- **Motivation** : How to become invariant to the window-size/scale of pseudo-periodic structure in $x(t)$?



Streaming multi-scale Lag-matrix construction

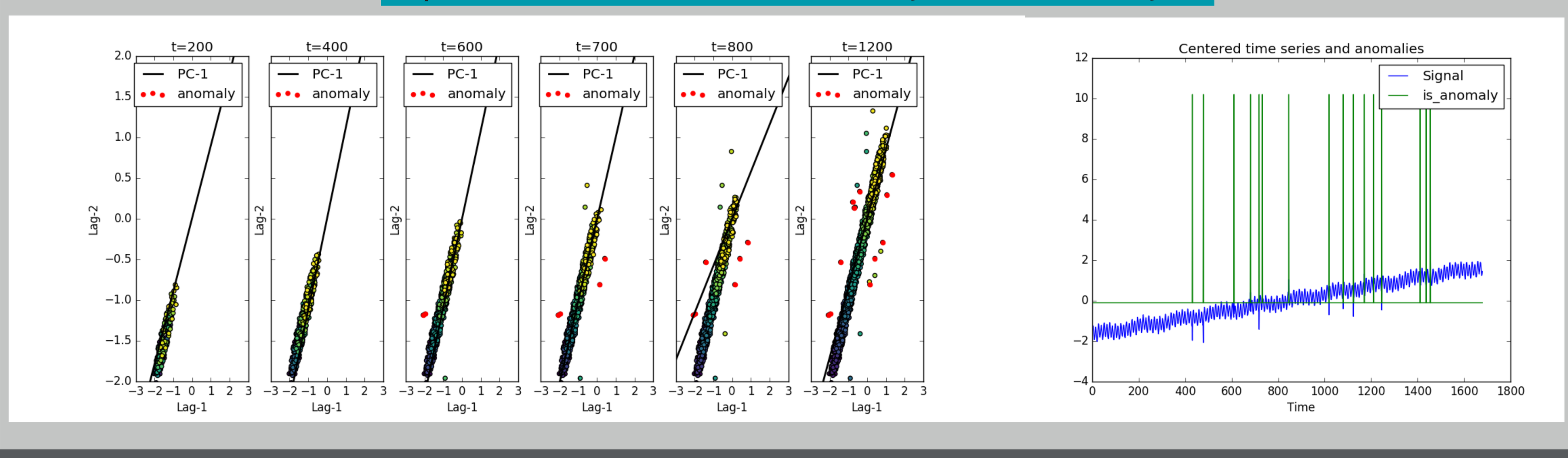
$$X_t^p = [x_t, x_{t-1}, \dots, x_{t-p+1}]^T \in \mathbb{R}^p$$



Streaming PCA

- Dimensionality reduction for time series lag embedding
- Recursive update for principal subspace
- **Linear Principal Component Analysis criterion** : Given $X^p \in \mathbb{R}^{T \times p}$, $\mathbf{w}_p$ is defined as 1-D projection capturing most of the energy of samples :

$$\mathbf{w}_p = \arg\min_{\|\mathbf{w}\|=1} \sum_{t=1}^{T} \|X_t^p - (\mathbf{w}\mathbf{w}^T)X_t^p\|^2$$



Multiscale streaming PCA

Algorithm 1 Streaming PCA

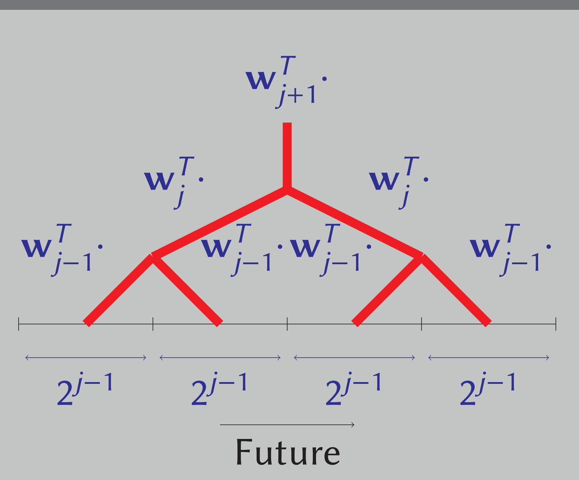
Initialization: $\mathbf{w}_j \leftarrow \mathbf{0}$, $\sigma_j^2 \leftarrow \epsilon$ with $\epsilon \ll 1$
for $t = 1, \dots, T$ **do**
  for $j = 1, \dots, J$ **do**
    $Z_t^j \leftarrow H_{2^j}^T X_t^j$
    $y_t^j \leftarrow \mathbf{w}_j^T Z_t^j$
    $\sigma_j^2 \leftarrow \sigma_j^2 + (y_t^j)^2$
    $\mathbf{e}_t^j \leftarrow Z_t^j - y_t^j \mathbf{w}_j$
    $\mathbf{w}_j \leftarrow \mathbf{w}_j + \sigma_j^{-2} y_t^j \mathbf{e}_t^j$
    $\widetilde{Z}_t^j \leftarrow \mathbf{w}_j^T Z_t^j \mathbf{w}_j$
    $\alpha_t^j \leftarrow \|\widetilde{Z}_t^j - Z_t^j\|^2$
  end for
end for
return $\boldsymbol{\alpha} \in \mathbb{R}^{T \times J}$

Given $x(t) \in \mathbb{R}$ for $t = 1 : T$

- We evaluate the lag-matrix $X \in \mathbb{R}^{T \times p}$ where $p = 2^j$.
- For each $X_t \in \mathbb{R}^p$ we perform a change of basis $Z_t := \Phi^T X_t$
- H refers to a unitary tx. that can
  - localize a deviation from the local mean and variations.
  - Preserve the variance.
- Haar transform $\Phi = H$:

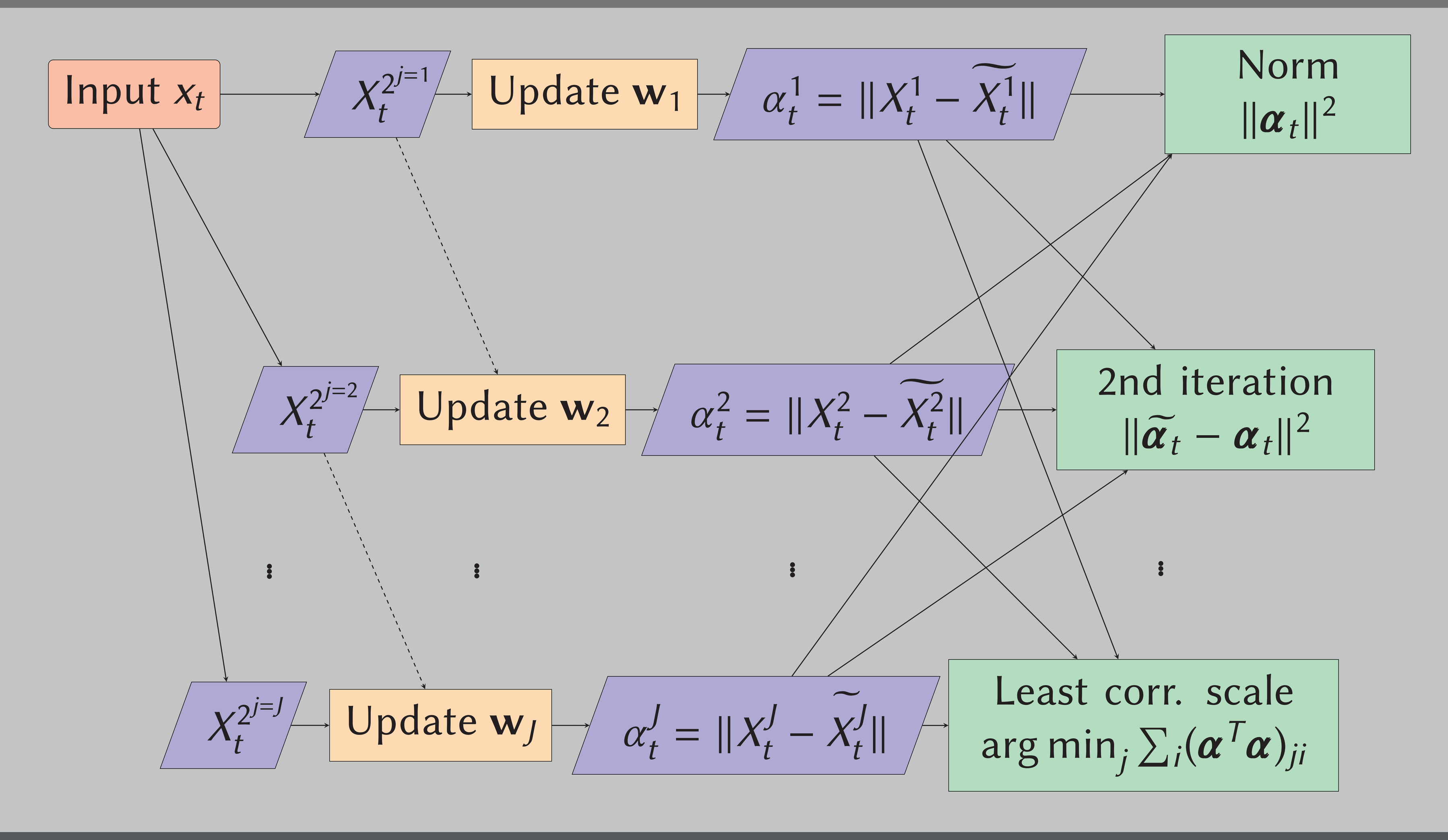
$$H_{2N} = \frac{1}{\sqrt{2}} \begin{bmatrix} H_N \otimes [1, 1] \\ I_N \otimes [1, -1] \end{bmatrix}$$

Hierarchical (Approximation) PCA



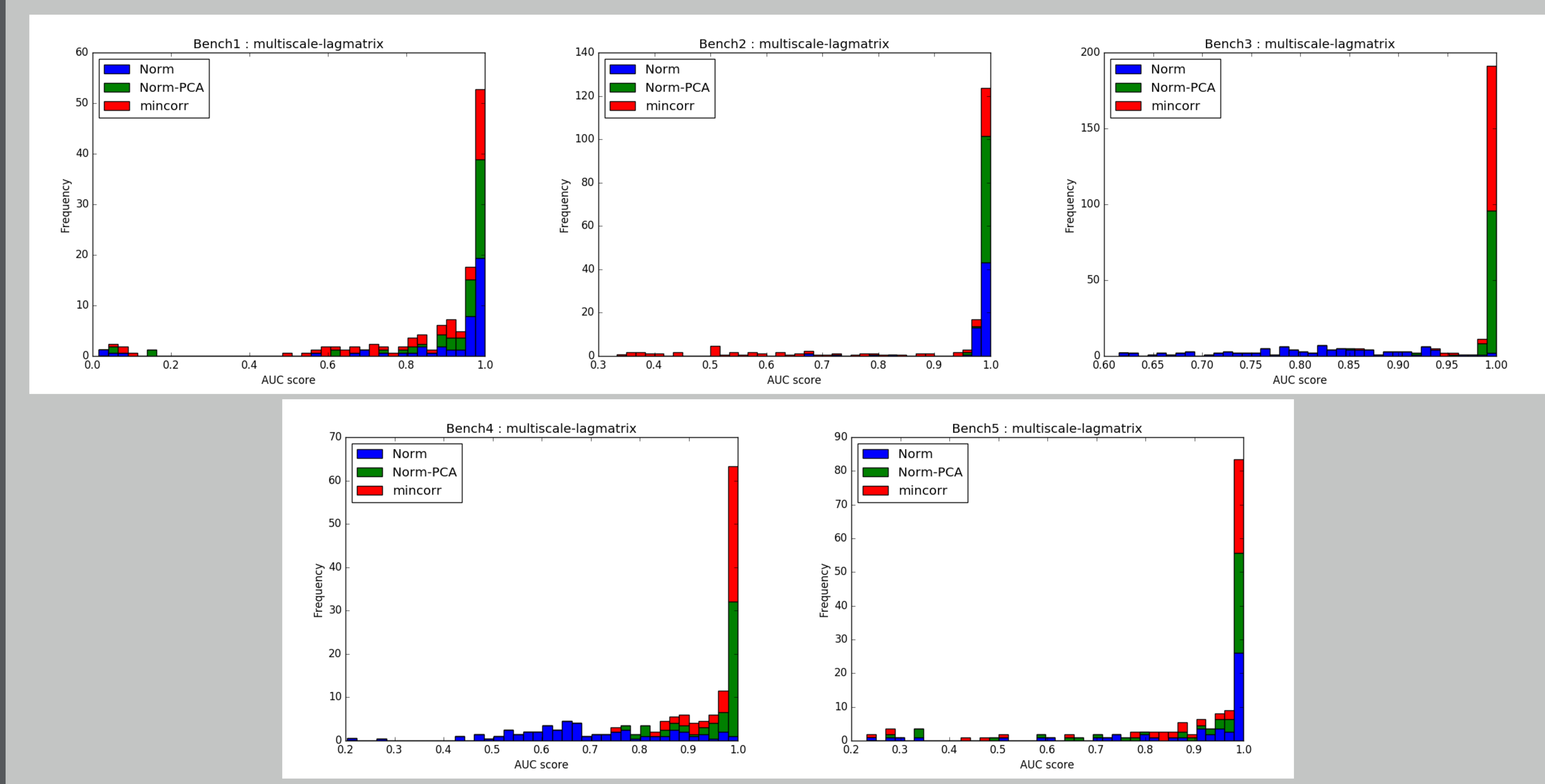
For scale $p_{j+1} = 2p_j$, a reduced lag matrix Z_t^{j+1} is obtained by projection of each component of size p_j on the principal direction obtained at this scale, *i.e.* $Z_t^{j+1} = [\mathbf{w}_j^T Z_t^j, \mathbf{w}_j^T Z_{t-2^j}^j]^T$ with $Z_t^1 = X_t^1$. The principal direction is updated after this step.

Global flow

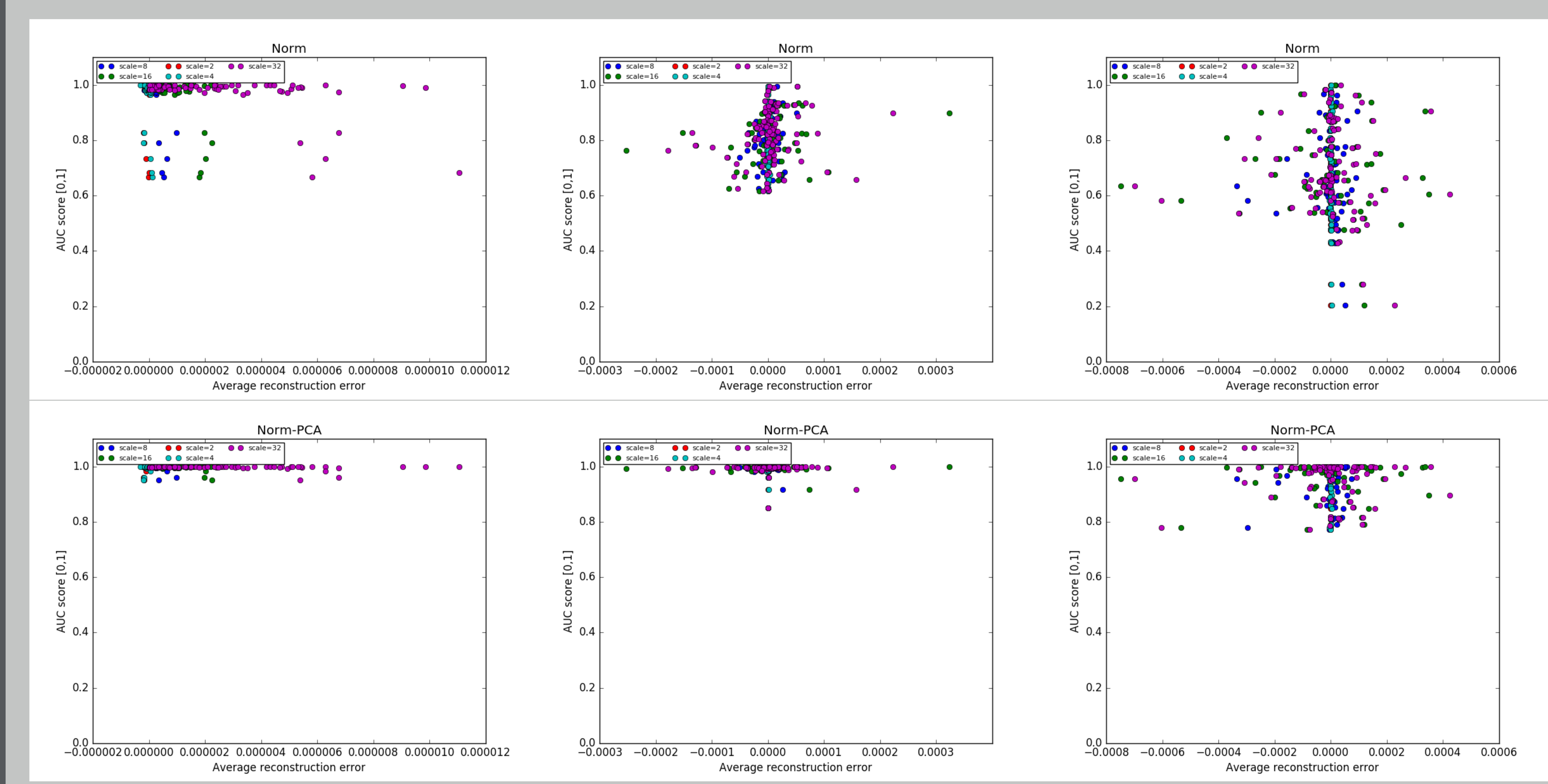


AUC performances across different aggregation methods

Performance Evaluation : Area under the receiver operators characteristics curve (AUC) (FPR *vs* TPR), 0 (worst) & 1 (perfect).



Effect of 2nd iteration of Streaming PCA



Least correlated scale Vs 2nd iteration of Streaming PCA :

- Least correlated scale and 2nd iteration of streaming PCA decorrelates the reconstruction error across scales.
- The least correlated scale performs better when there is a single scale structure across time series. Second iteration performs better when there are multiple scales.

Future work

- Understand bounds on reconstruction error $\boldsymbol{\alpha}(t)$ for Streaming PCA.
- Better base-line by comparing with streaming covariance estimation.
- Probabilistic anomaly score using gaussian neg-log score.
- Recursively calculable multi-scale time series representation $H^T X_t$ to model long range dependencies.

References

- Papadimitriou et al. Optimal multi-scale patterns in time series streams, 2006 ACM SIGMOD
- Bin Yang, Projection approximation subspace tracking. Trans. Sigal Processing, Jan. 1995
- Evaluating real-time anomaly detection algorithms, Numenta benchmark, ICMLA 2015