mapDDX:
Descriptive and
Diagnostic plots, and
Xcorrelation analysis of activity maps

Nov. 17

Jungsik Noh

UT Southwestern Medical Center

Introduction

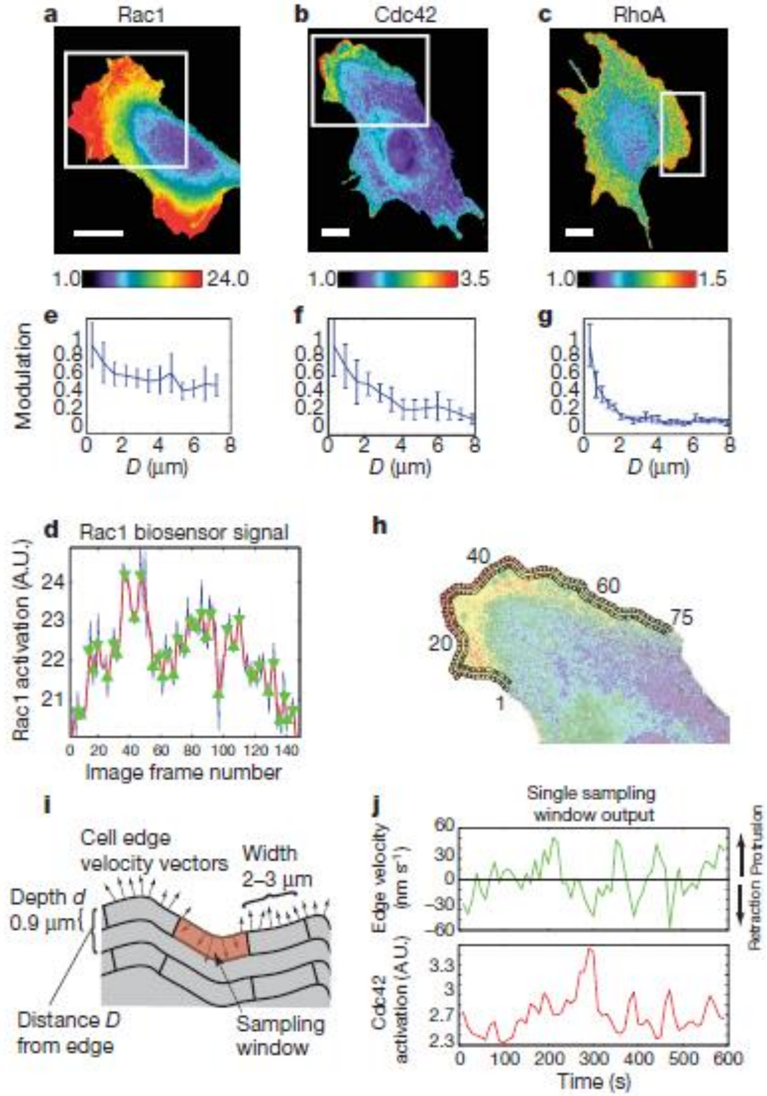


Figure 1 | Activation of Rho GTPases in migrating MEFs. a–c, Rac1 (a), Cdc42 (b) and RhoA (c) activation reported by biosensors. White box indicates region of interest selected for analysis. Scale bar, 20 µm. Colour

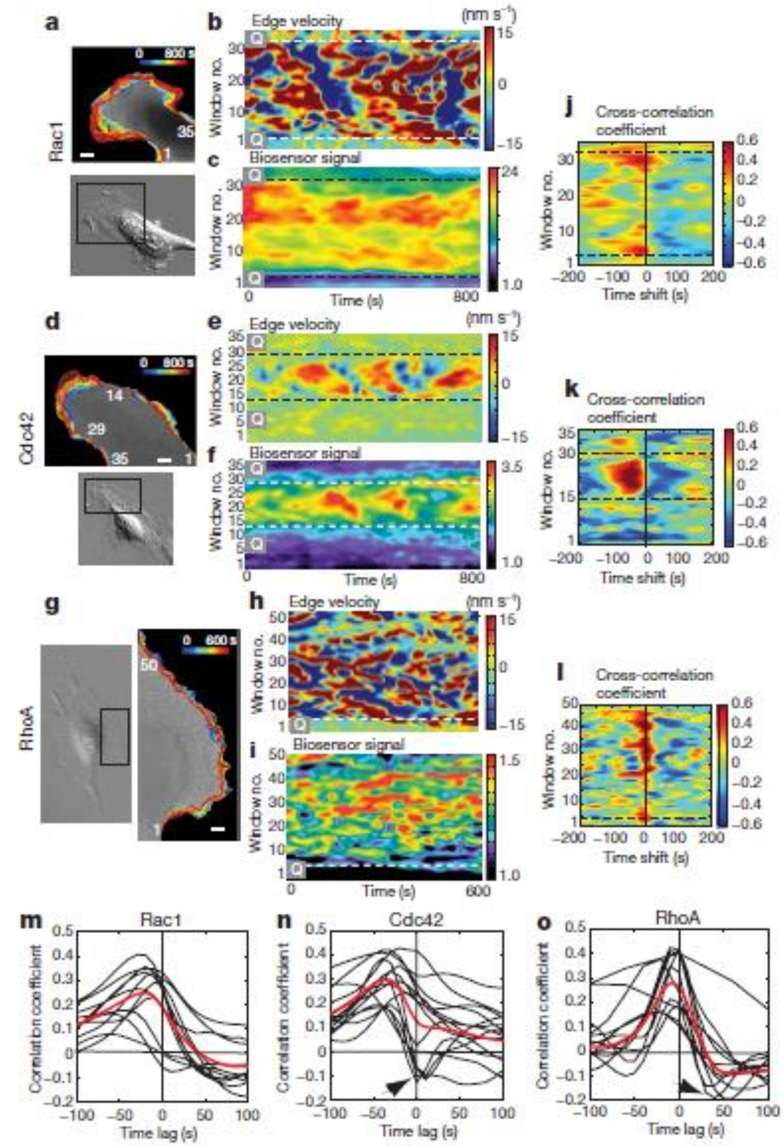


Figure 2 | Dynamics of cell edge morphology and GTPase activation.

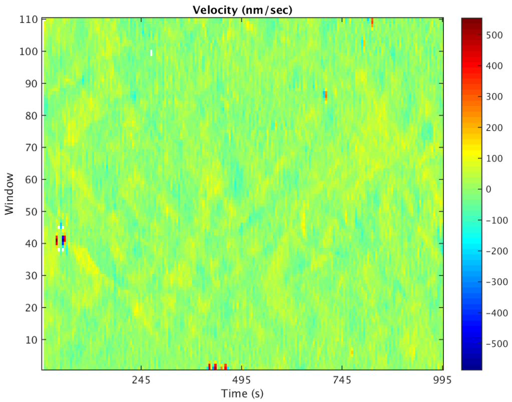
*Machacek et al. (2009, Nature)

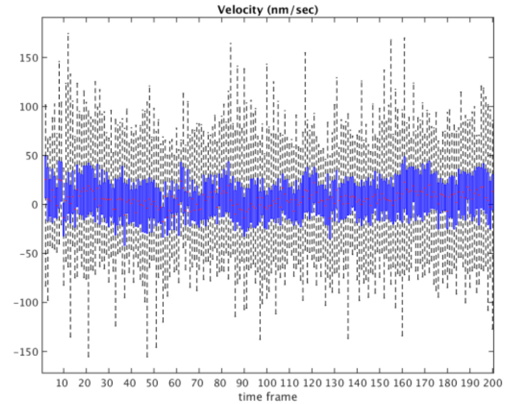
mapDDX

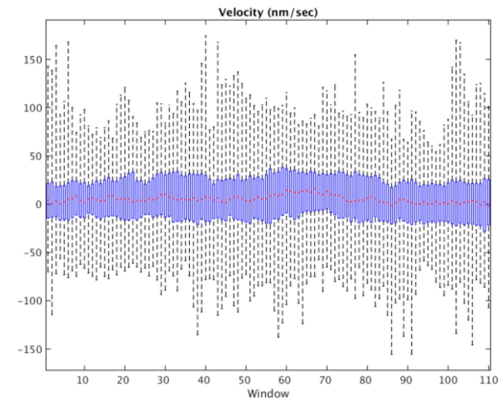
```
function mapDescriptives_OneChan(MD, iChan, maxLayer, chanName, chanTitle, figuresDir, varargin)
% mapDescriptives_OneChan Draw descriptive plots of an activity map of
% the specified channel in movieData.
%
% Usage:
%       mapDescriptives_OneChan(MD, 1, 3, 'Actin', 'Actin', ...
%       fullfile(MD.outputDirectory_, 'mapDescriptives'), 'impute', 0, ...
%       'parpoolNum', 4)
%
% Input:
%       MD          - movieData object
%       iChan       - channel index
%       maxLayer    - maximum layer to which activity maps are drawn
%       chanName    - a short name for the channel. eg. 'Actin'
%       chanTitle   - a more detailed name for the channel
%                     eg. 'Velocity (nm/sec)'
%       figuresDir  - a directory where plots are saved as png files
%
% Output: png files are saved in the figuresDir
```

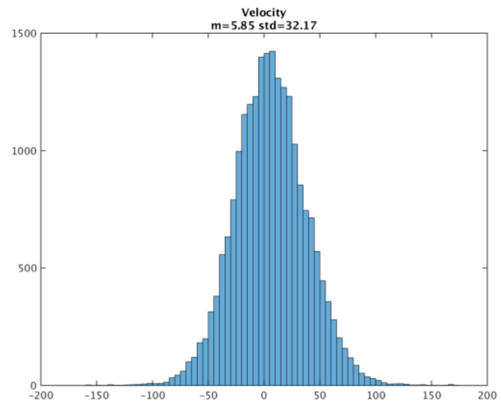
```
function mapXcorrCurvePermutation(MD, iChan1, iChan2, chan1Name, chan2Name, layerMax, figuresDir, varargin)
% mapXcorrCurvePermutation Perform cross correlation analysis between two
% channels. It plots cross correlation maps, their mean curves
% at each layer together with confidence bounds based on permutation, and a
% topograph of the cross correlations at lag 0. The cross correlations at
% lag h are Corr(chan1_{t+h}, chan2_t).
% It computes cross correlations in a fashion that can handle many NaN's
% by utilizing nanXcorrMaps.m function.
%
% Usage:
%       mapXcorrCurvePermutation(MD, 2, 1, 'mDia1', 'Actin', 3, ...
%           fullfile(MD.outputDirectory_, 'mapCrossCorr'), 'impute', 1, 'parpoolNum', 4)
%
% Input:
%       MD          - a movieData object
%       iChan1      - the 1st channel index
%       chan1Name   - a short name for channel1.
%       iChan2      - the 2nd channel index
%       chan2Name   - a short name for channel2.
%       layerMax    - maximum layer to be analyzed
%       figuresDir  - a directory where plots are saved as png files
%
% Output: png files are saved in the figuresDir.
```

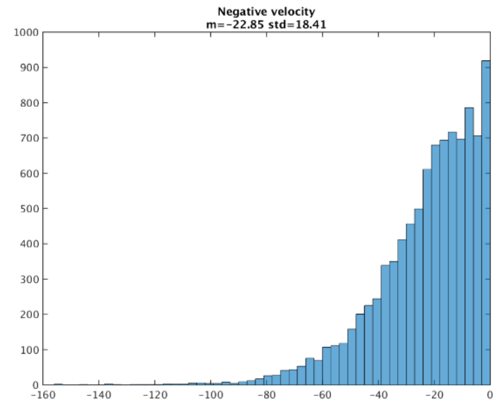
Descriptives: edge velocity

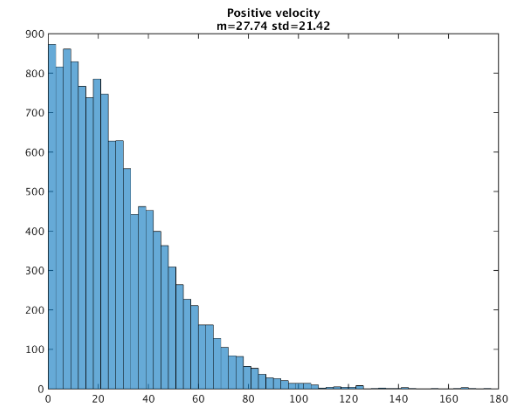


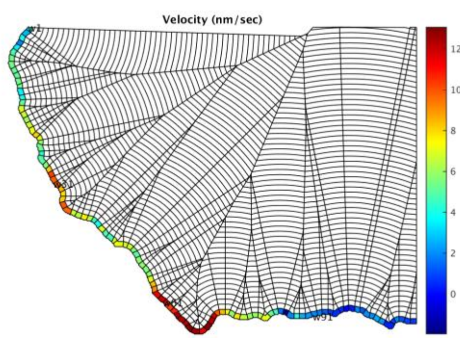


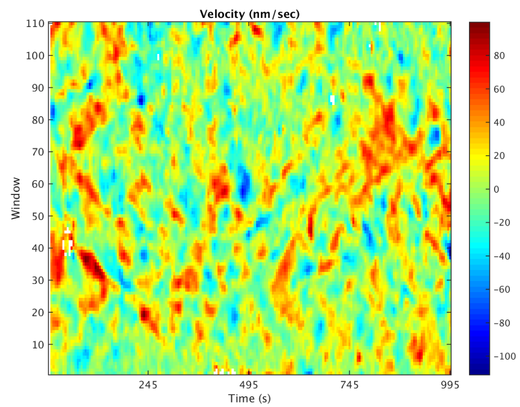


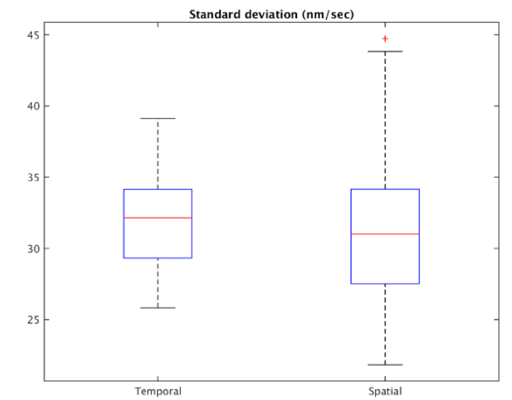






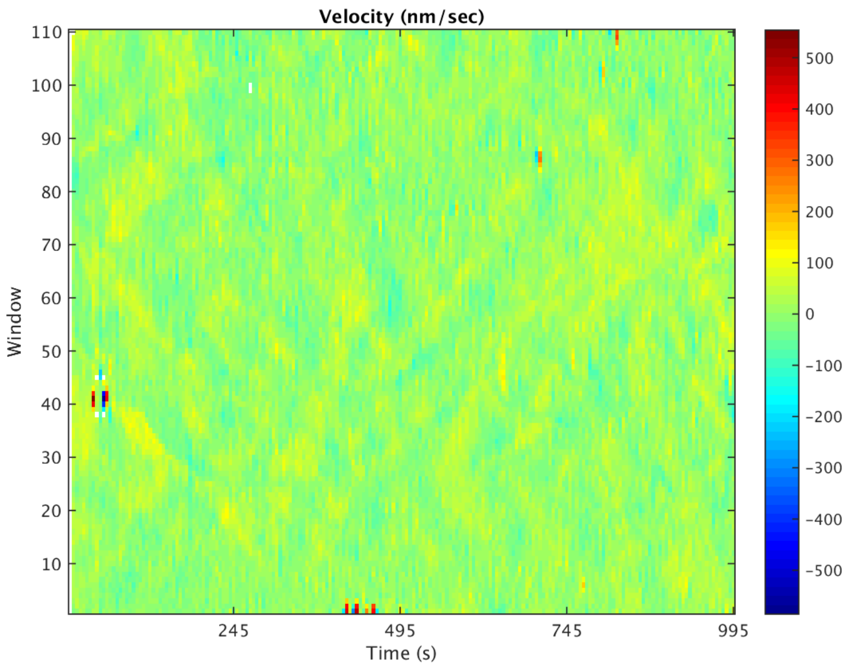




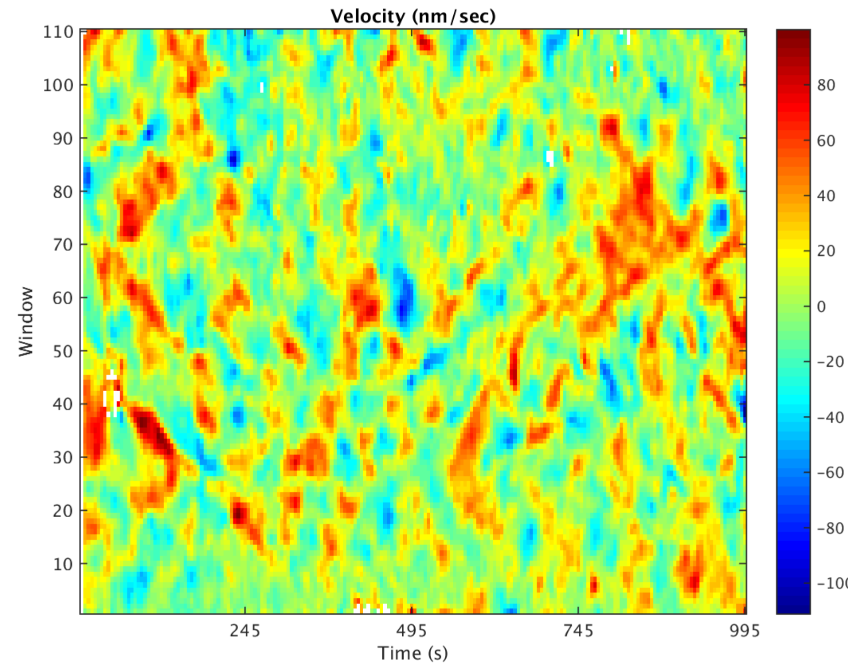


Ptk1 Dec.-1 (cell#5)

Raw activity map

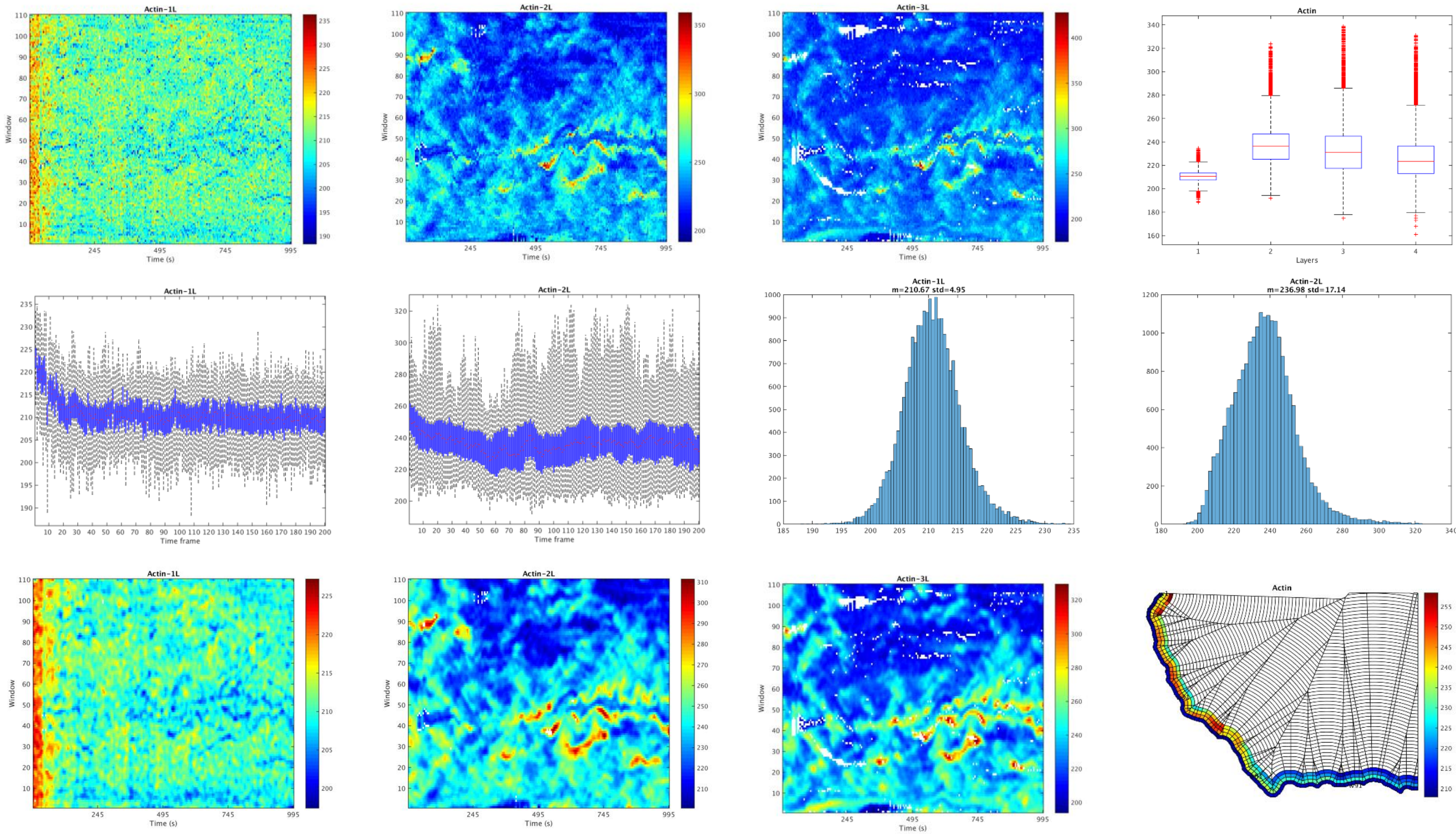


Outlier adj./smoothed activity map



- Velocity map = $\{V(t,w): t \geq 2\ (frames), w \geq 1\ (windows)\}$
- $V(t,w)$: displacement (nm) between the frames at (t-1) and t divided by time interval (sec)
- Z-scores of all the $\{V(t,w)\}$ are computed
- |Z-scores| > 5 => detected outliers
- Outliers and missing windows can be imputed by using the most similar windows (`knnimpute.m` function)

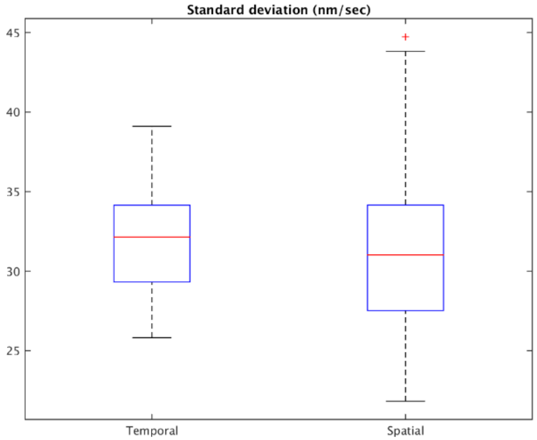
Descriptives: Actin 1-3 layers



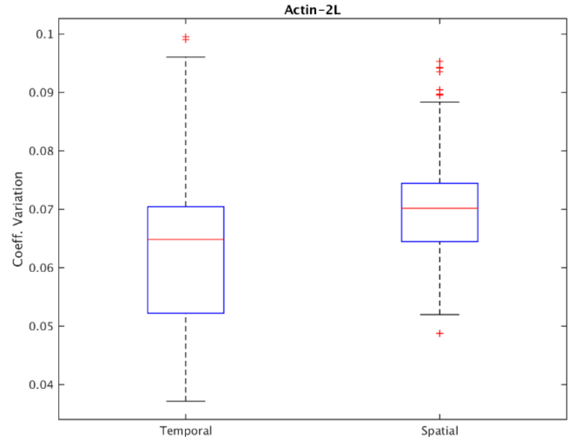
- Check layer-wise heterogeneity.
- Use the boxplot over time to check if there is a trend over time which can mislead temporal correlation analysis.
- A topograph shows spatial distribution of activities averaged over time.

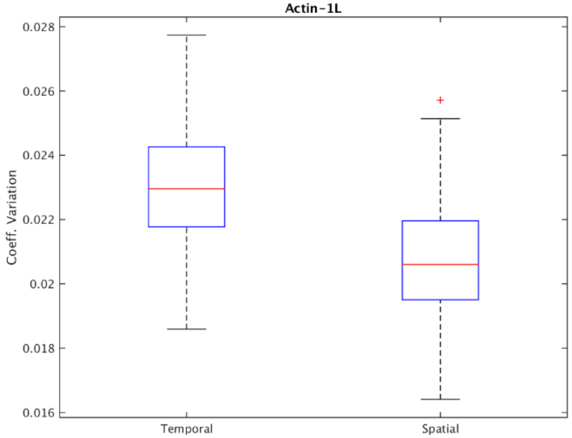
Diagnostics: coefficient of variation (CV)

- $CV(X) = SD(X)/E(X)$
- CV is a statistical measure of dispersion of a probability distribution.
- CV has no unit and more suitable for non-negative observations.

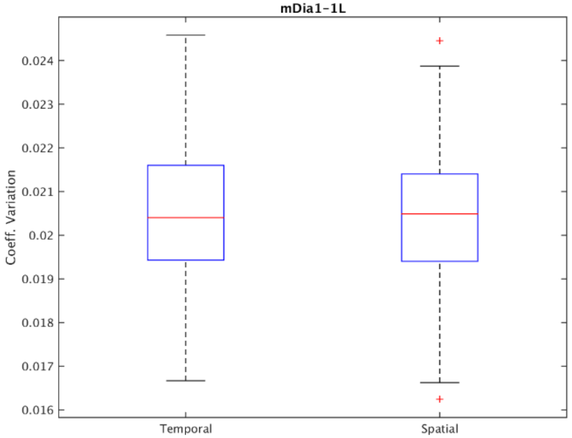


- Temporal/spatial variations of velocity are similar for the above cell.
- Common case

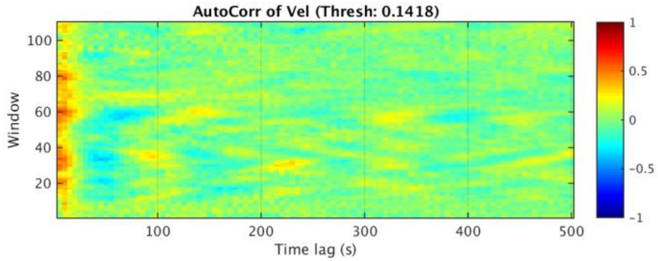




- Temporal CV is bigger than spatial CV, which suggests temporal heterogeneity.
- Such heterogeneity may not be relevant to underlying biological questions.

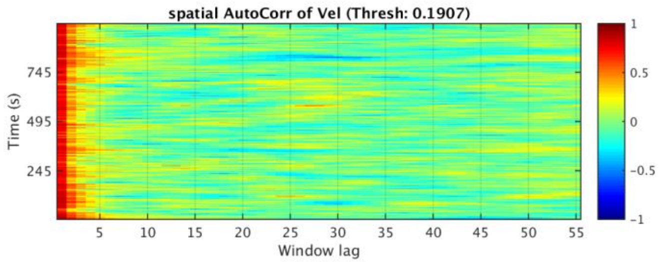


Diagnostics: temporal/spatial autocorrelations



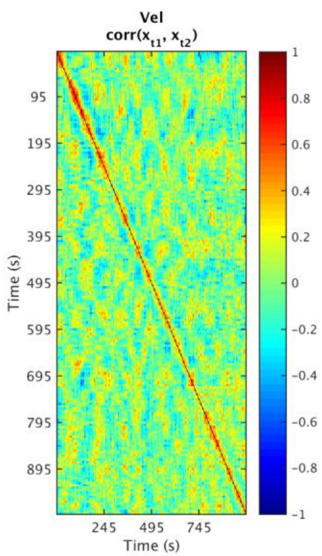
Temporal autocorr.

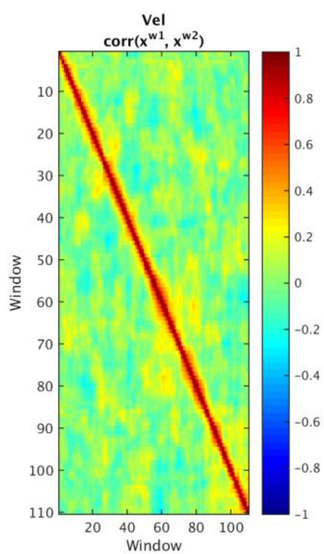
$$\rho_w(h) = \frac{1}{t_m} \sum_{t=1}^{t_{max}-h} \dot{X}(t+h,w)\ \dot{X}(t,w)$$



Spatial autocorr.

$$\rho_t(h) = \frac{1}{w_{max}} \sum_{w=1}^{w_{max}} \dot{X}(t,w+h)\ \dot{X}(t,w)$$





Corr. between two time frames (two columns)
- Left

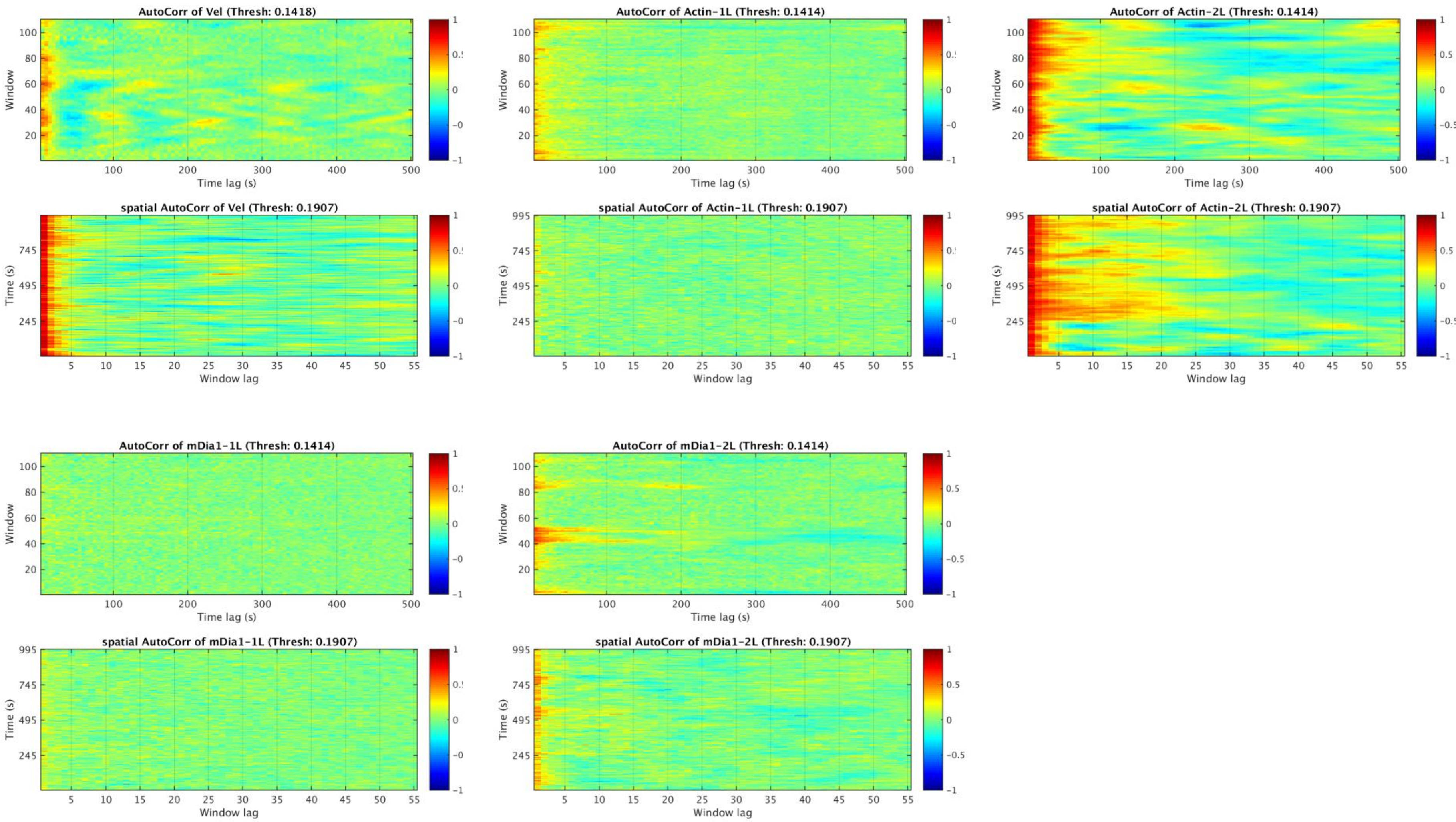
$$\rho(t_1,t_2) = \frac{1}{w_{max}} \sum_{w=1}^{w_{max}} \dot{X}(t_1,w)\ \dot{X}(t_2,w)$$

Corr. between two windows (two rows)
- Right

$$\rho(w_1,w_2) = \frac{1}{t_{max}} \sum_{t=1}^{t_{max}} \dot{X}(t,w_1)\ \dot{X}(t,w_2)$$

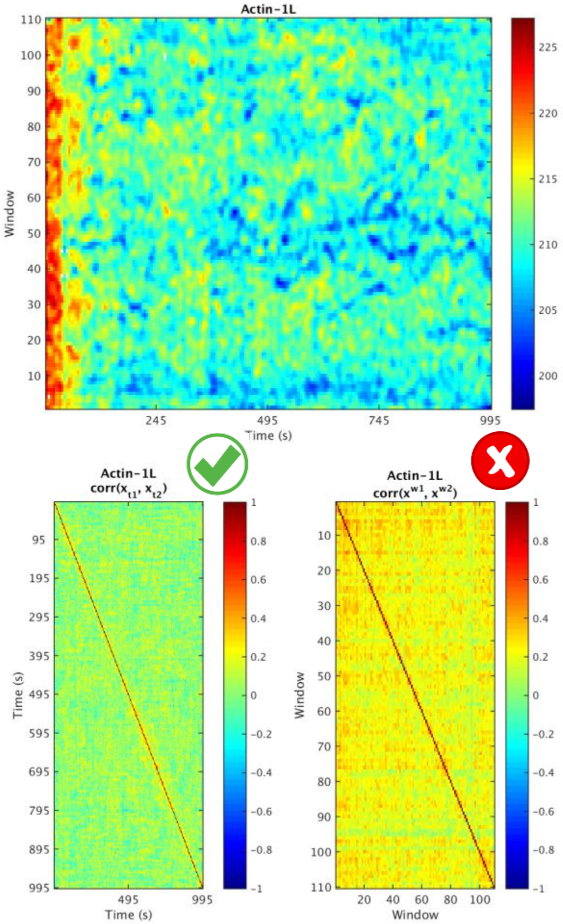
- Here, $\dot{X}(t,w)$ denotes a standardized variable.
- The standardization can be different in different context.

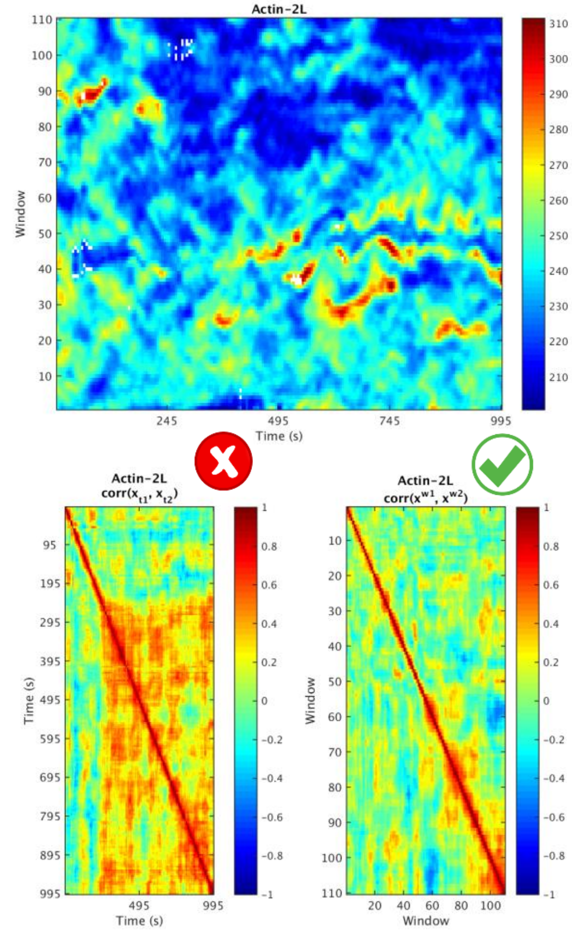
Autocorrelation maps for vel, actin, mDia1



- The quality of 1st layer is not good.
- Autocorrelations of actin are higher than mDia1.

Correlations between column or row vectors

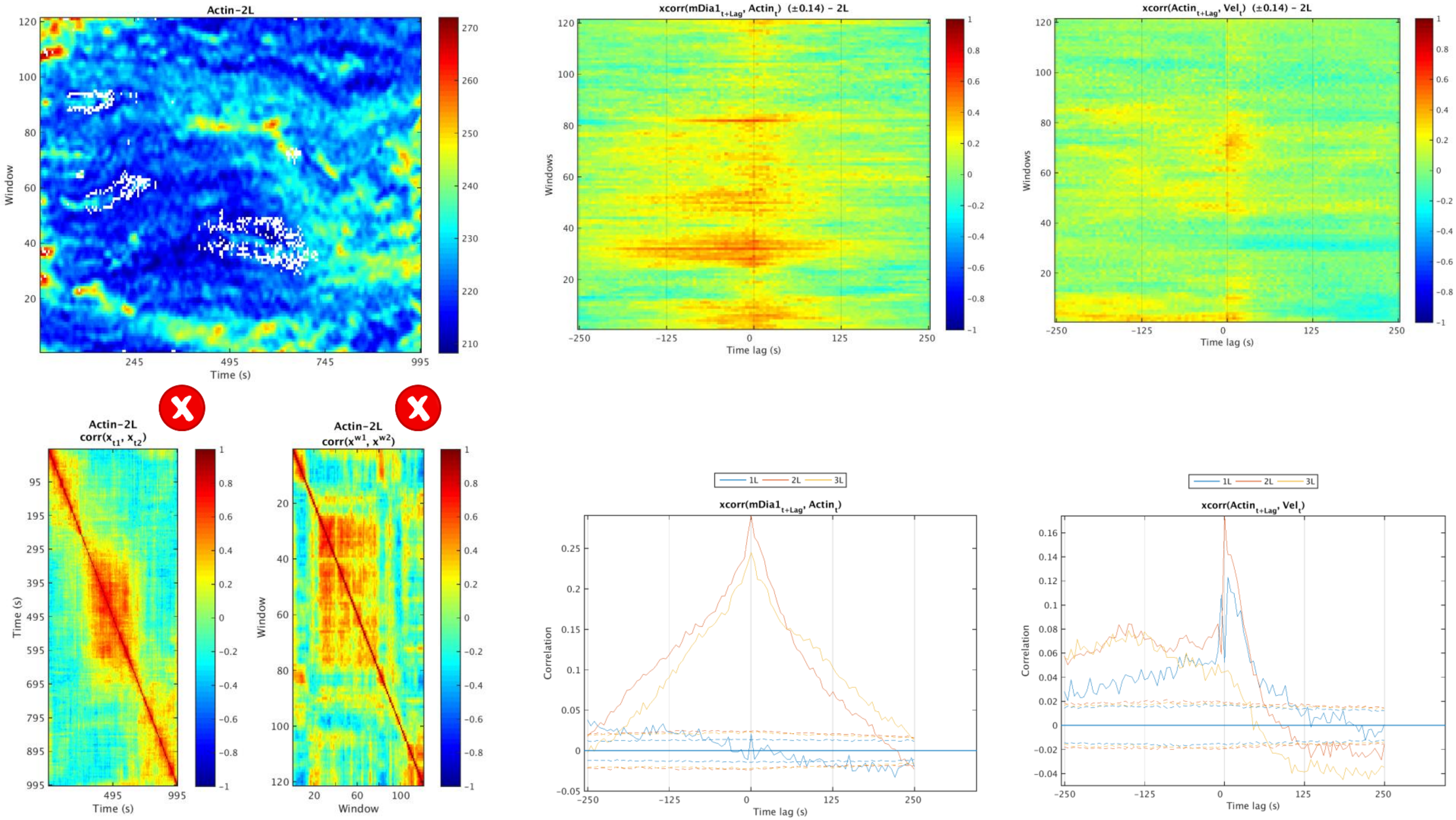




- High correlations between windows indicate that row vectors share common temporal pattern (Activities are high at the beginning).
- In such case, temporal analysis can be misleading.

- High correlations between time frames indicate that column vectors share common spatial pattern (Activities are low in win80~110 after 250 sec) .
- In such case, spatial analysis can be misleading.

Ptk1 Cell#1



- The above cross correlations are not typical.
- The common temporal pattern in win20~60 may result in spurious high correlations.

Augmented Dickey-Fuller (ADF) test

- `adftest.m` tests whether a TS is stationary or non-stationary.
- Matlab manual

Autoregressive model variant, which specifies a test of the null model

$$y_t = y_{t-1} + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \ldots + \beta_p \Delta y_{t-p} + \varepsilon_t$$

against the alternative model

$$y_t = \phi y_{t-1} + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \ldots + \beta_p \Delta y_{t-p} + \varepsilon_t,$$

with AR(1) coefficient, $\phi < 1$.

$(\Delta y_t = y_t - y_{t-1})$

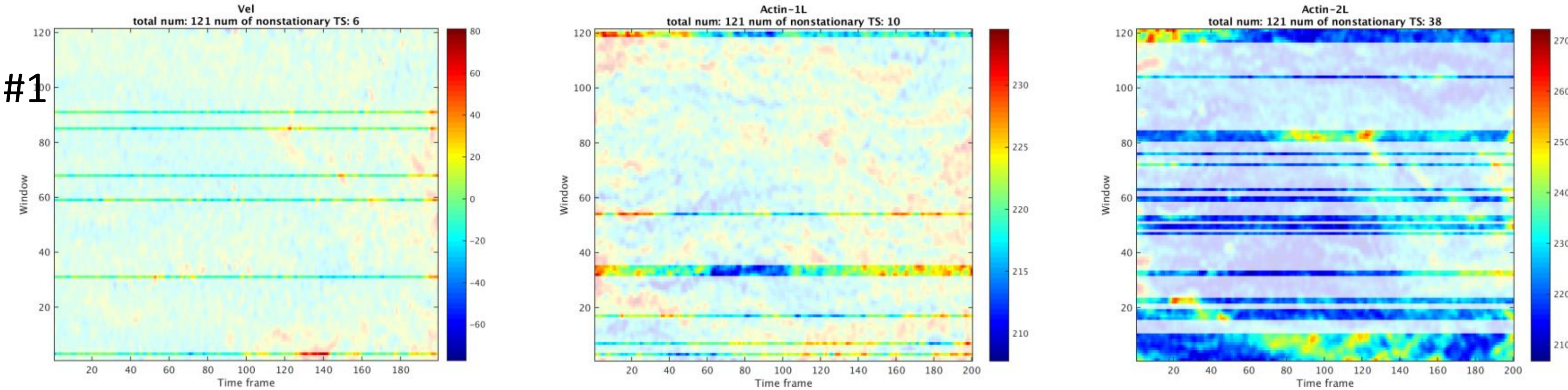
- Idea:
  - A stationary TS has the property that $E(Y_t) = constant$ which is a long-term equilibrium level.
  - Consider a TS regression model:

$$\Delta y_t = \alpha + \gamma y_{t-1} + \beta_1 \Delta y_{t-1} + \epsilon_t$$

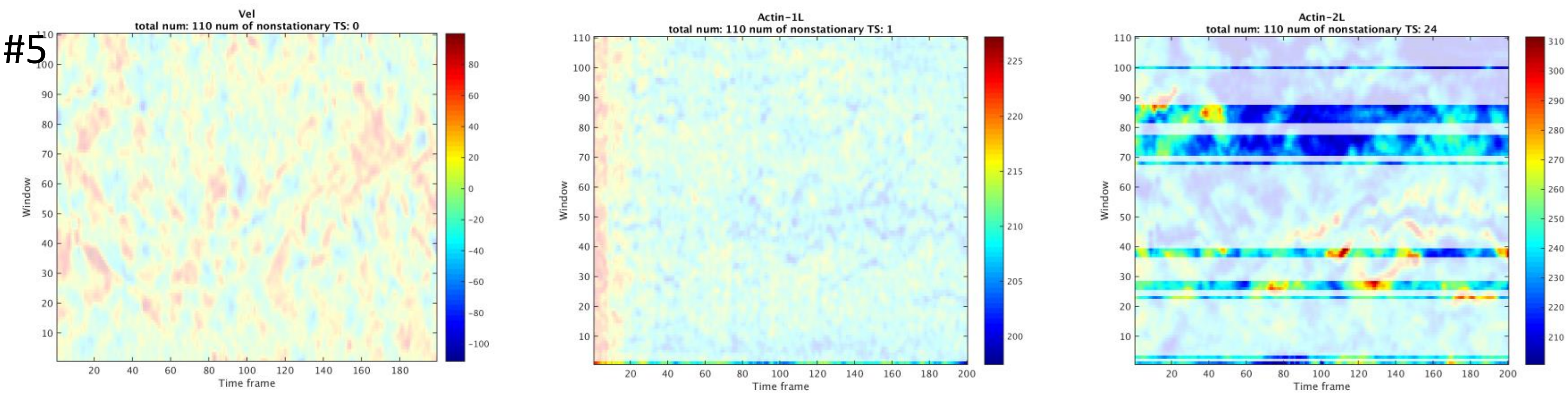
  - For the stationarity, y_{t-1} needs to be negatively correlated with $\Delta y_t = y_t - y_{t-1}$.
  - We can test the non-stationarity (H0: $\gamma = 0$) vs. stationarity (H1: $\gamma < 0$).
  - If H0 is true, then $\{\Delta y_t : t \geq 1\}$ is an AR(1) process, and $\{y_t : t \geq 1\}$ is an integrated process like a random walk process which does not have a constant long-term equilibrium level.

- Non-transparency indicates windows tested to be non-stationary (Transparent: stationary).

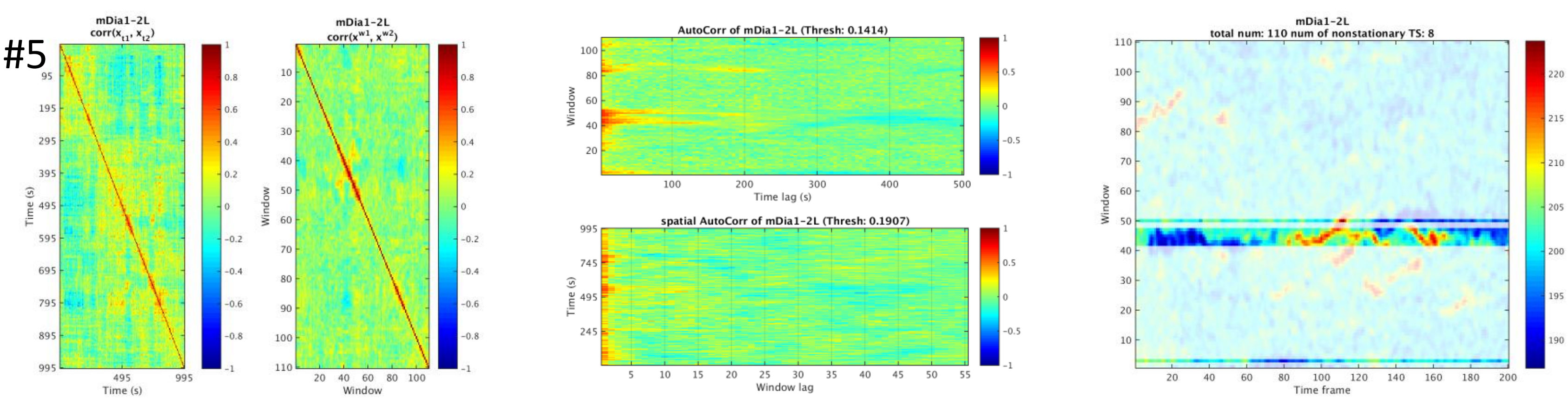
Ptk1 Cell#1



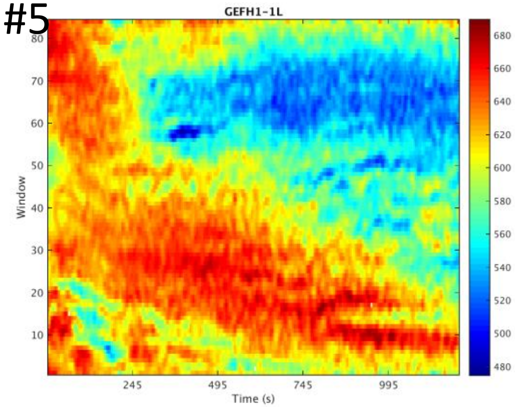
Ptk1 Cell#5

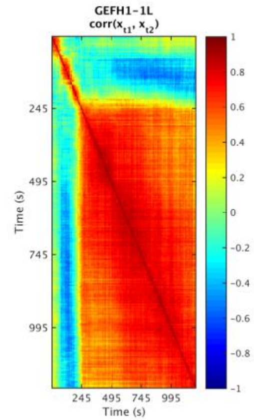


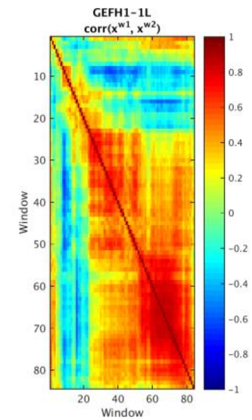
Ptk1 Cell#5
mDia1

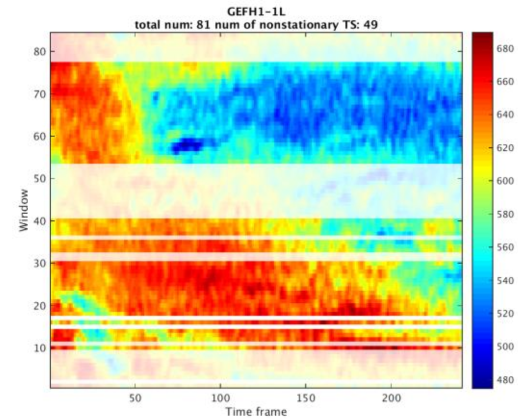


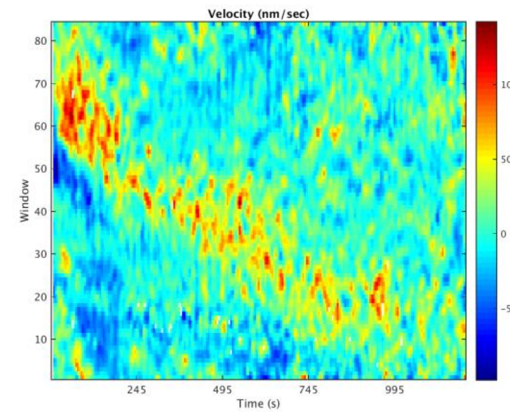
MDA Cell#5

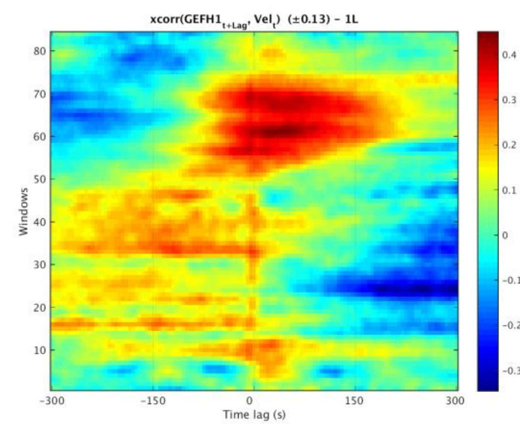








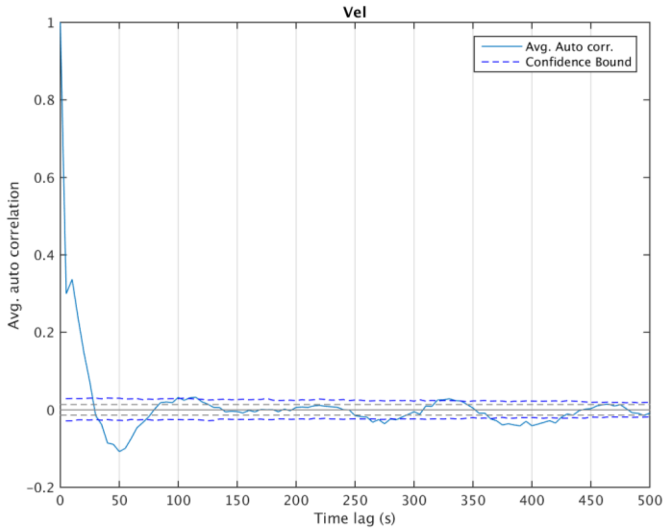


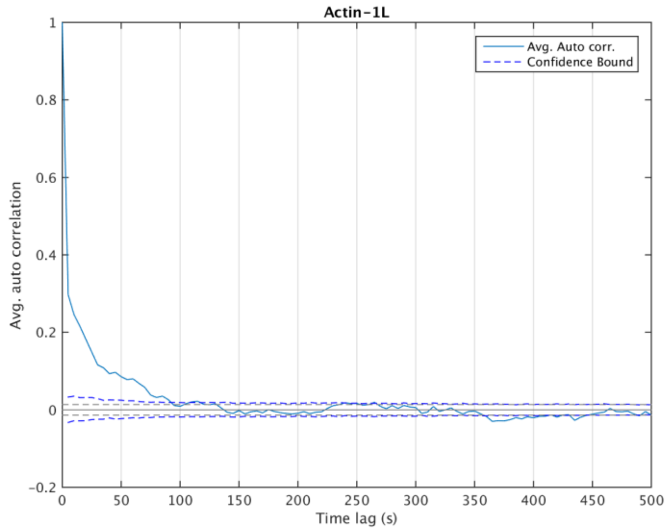


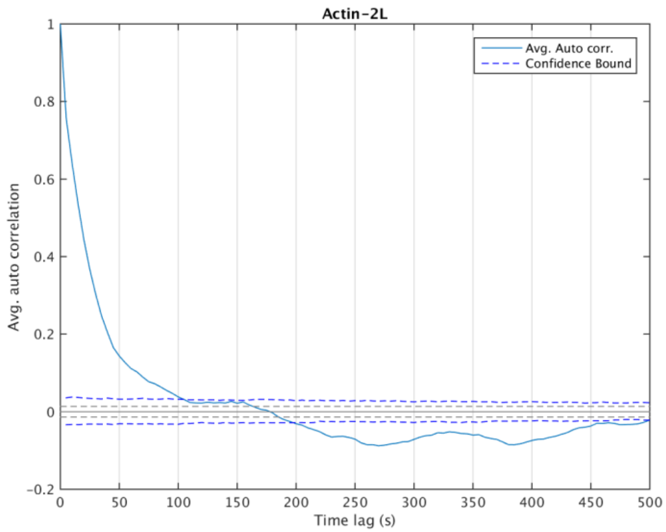
- Highly non-stationary case
- Correlations can be spurious.
- Methods for non-stationary TS are recommended.

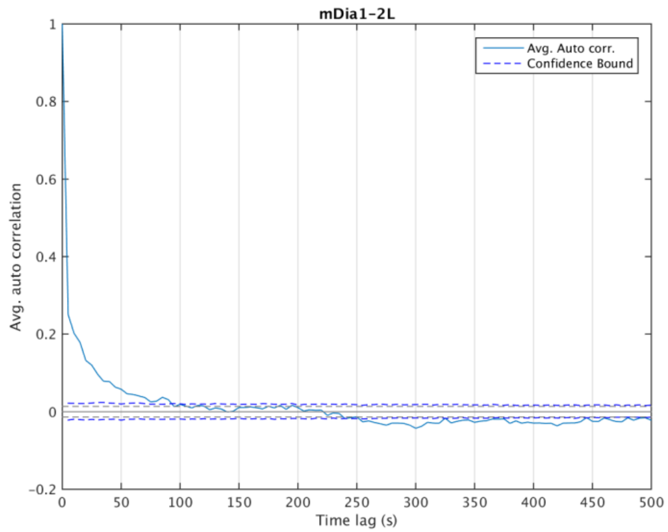
Autocorrelation curves (temporal)

Ptk1 Cell#5

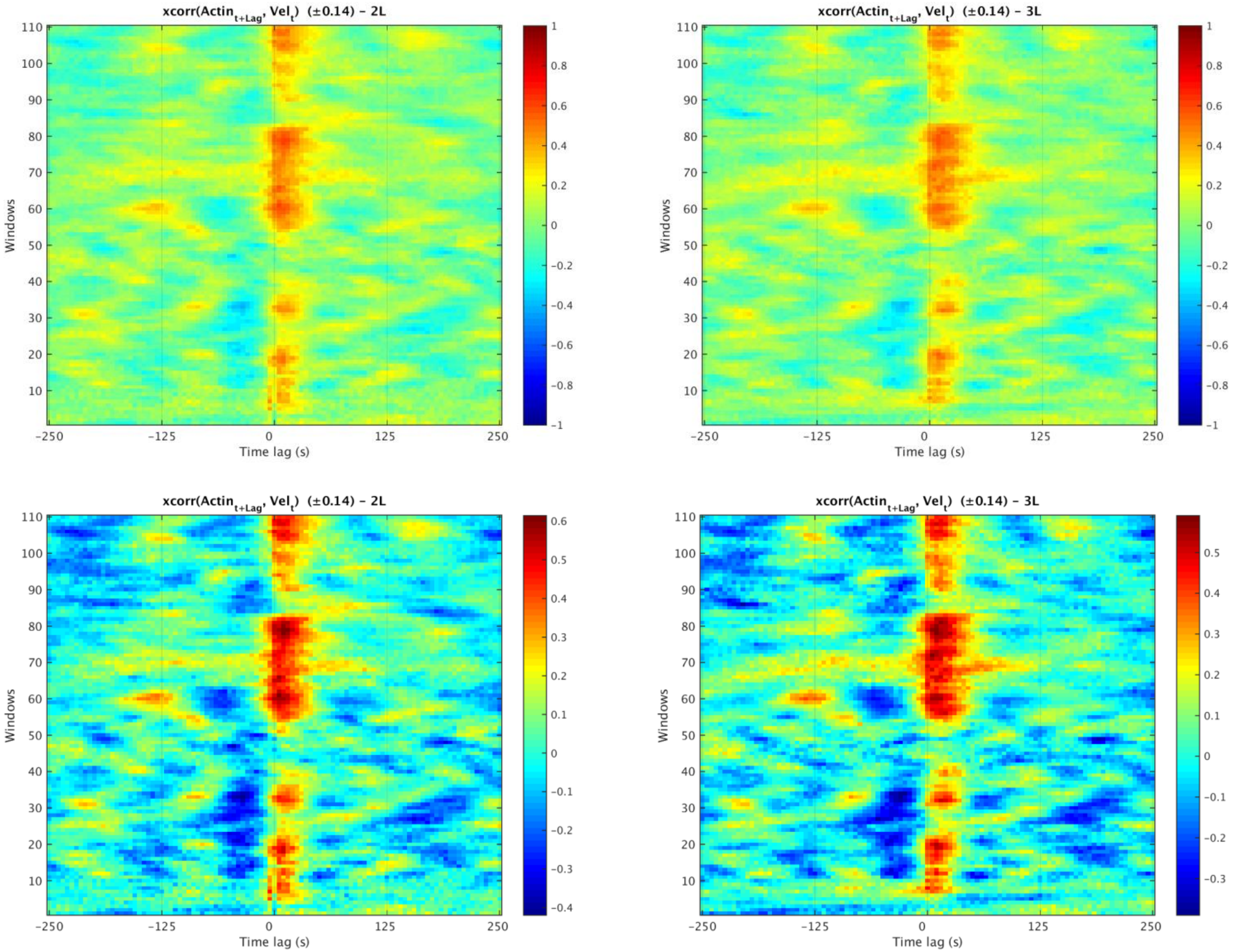








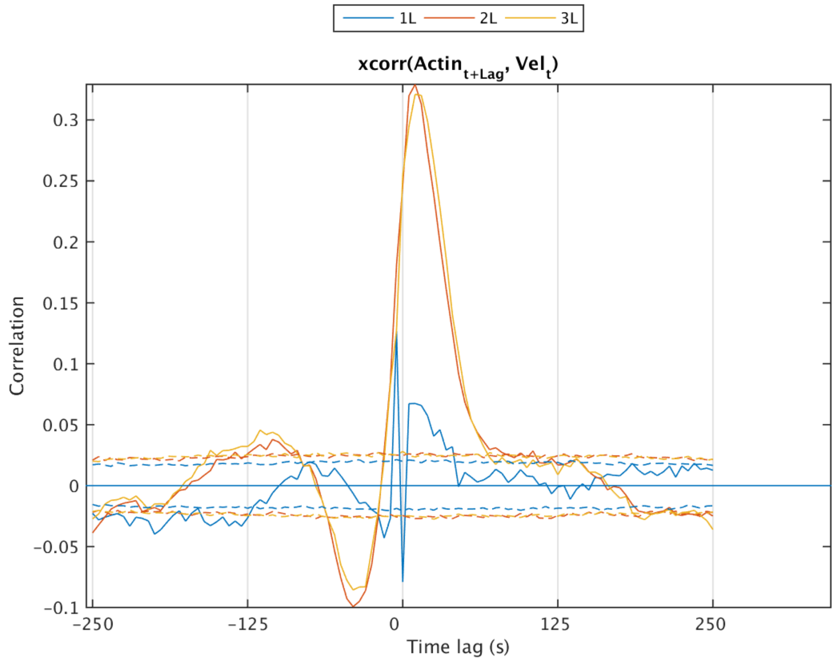
Cross correlation maps

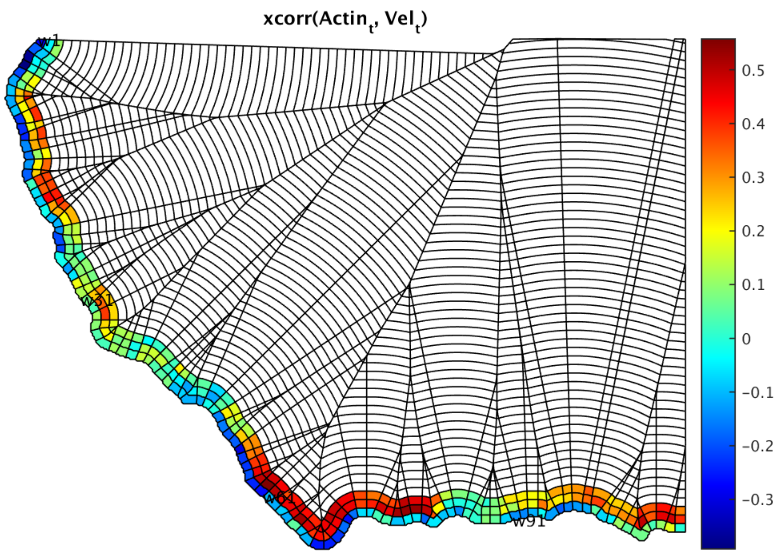


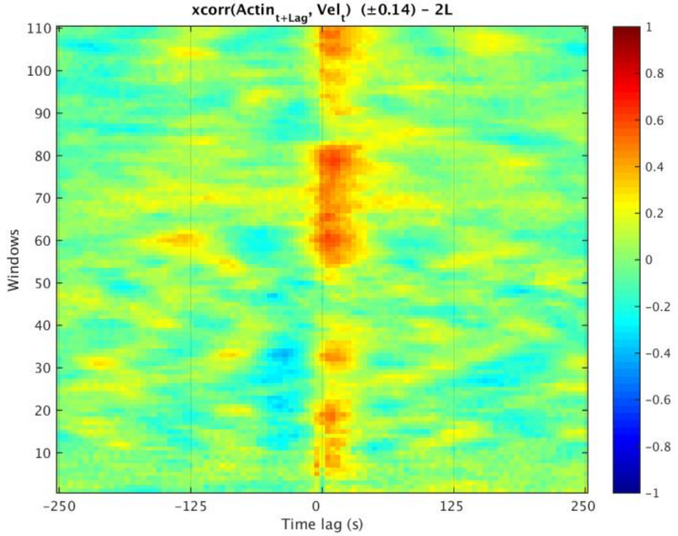
[-1, 1] range

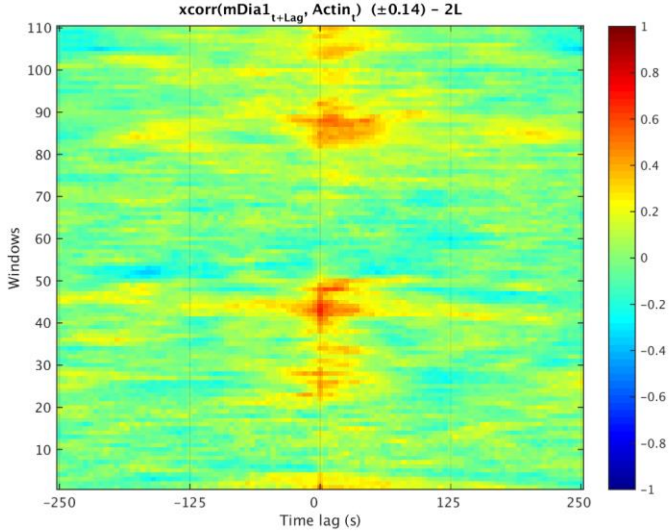
Range adjusted

Cross correlations

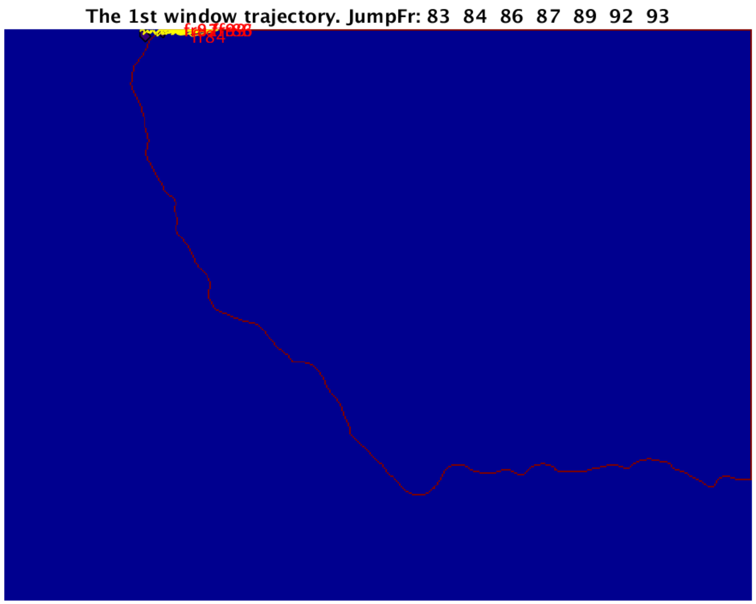


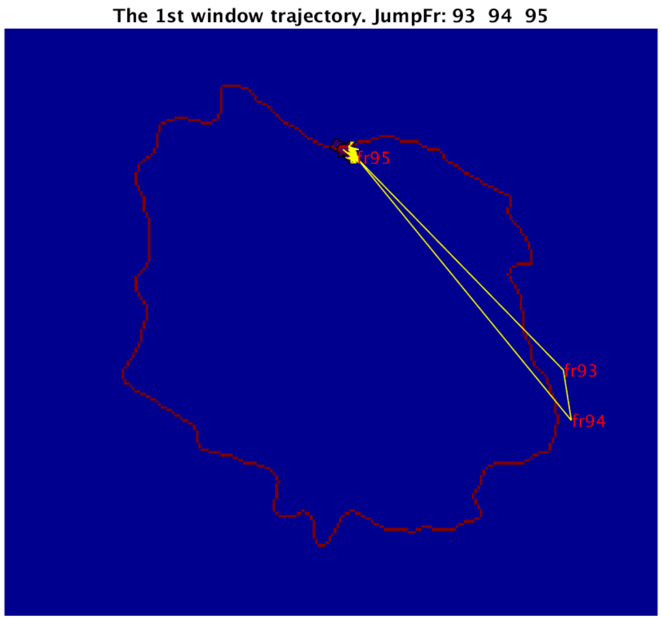






Diagnostics: Check windowing jumps





- It plots the trajectory of the 1st window in the 1st layer in order to find frames where window jumping happened.
- It plots the trajectory, the cell boundary at the last frame and estimated jump frames.

mapDiagnosticsXcorrCurves: .m functions

1. mapDescriptives_OneChan, mapDescriptives_Vel
    1. mapOutlierImputation > myknnimpute
    2. topographMD
    3. TimeSpaceAutoCorPlot
        1. autoCorrMap > nanXcorrelation
    4. autoCorrCurvePermTest > autoCorrMap
    5. nanAdfTestMap
    6. nanZscore

2. mapXcorrCurvePermutation, mapXcorrCurvePermutation_Vel
    1. mapOutlierImputation
    2. topographMD
    3. xcorrCurvePermutationTest
        1. xcorrMapPlot > nanXcorrMaps > nanXcorrelation
        2. timePermDistXcorrMean > nanXcorrMaps