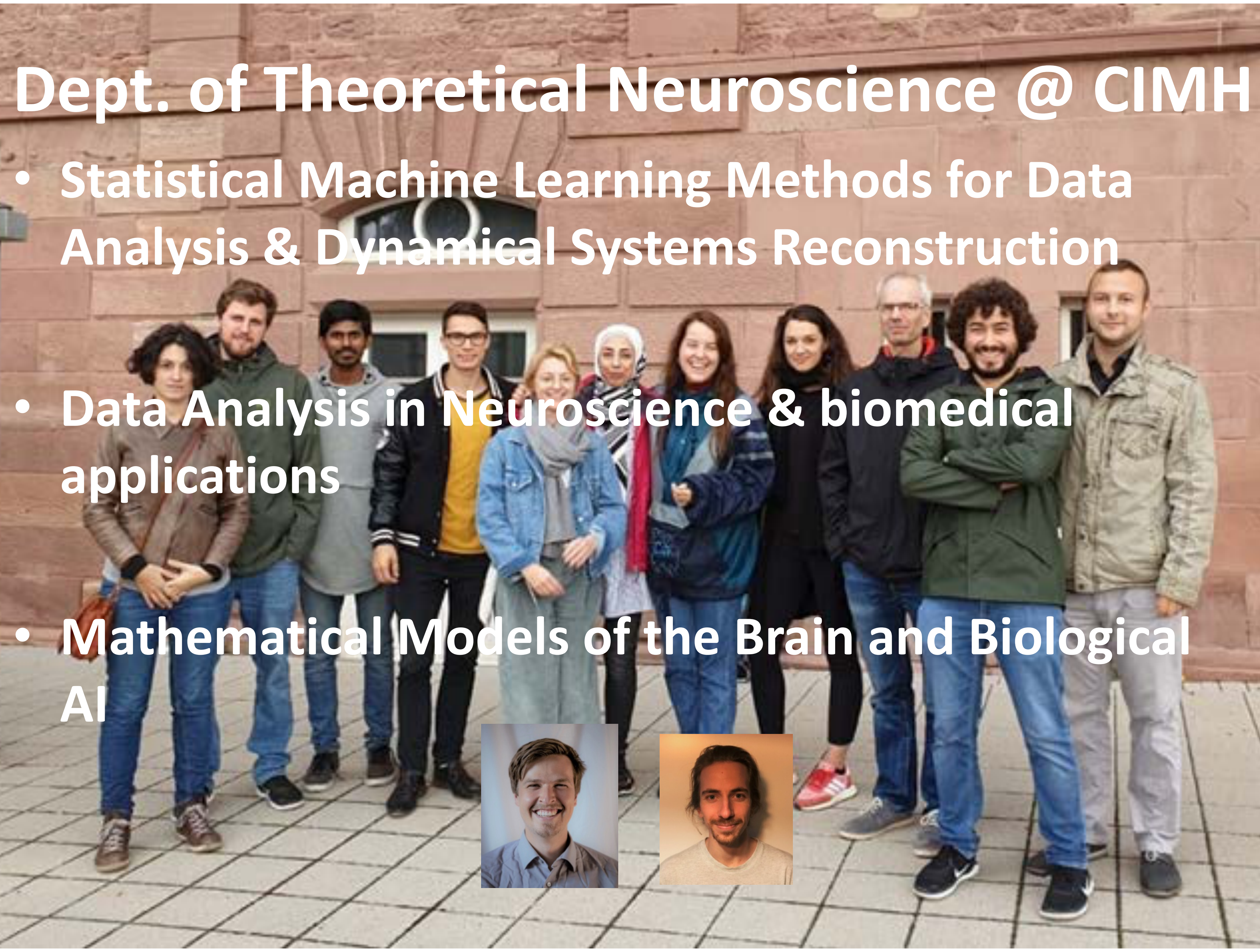
TIME SERIES ANALYSIS & RECURRENT NEURAL NETWORKS #1

Introduction & basic terms

Main lecture: Daniel Durstewitz

Exercises: Leonard Bereska, Manuel Brenner, Daniel Kramer, Georgia Koppe

Heidelberg University


Dept. of Theoretical Neuroscience @ CIMH
• Statistical Machine Learning Methods for Data Analysis & Dynamical Systems Reconstruction
• Data Analysis in Neuroscience & biomedical applications
• Mathematical Models of the Brain and Biological AI

$X = \{x_1, \dots x_T\} \equiv \{x_{1:T}\} \equiv \{x_t\}, t = 1 \dots T$

$(x_1, \dots, x_T)$

x_t , $x(t)$

Goals of TSA?

- prediction

  $x_1 \dots x_T \longrightarrow x_{T+1}, x_{T+2}, \dots$

- hypothesis testing

- Identify generating system $\dots x_t \dots$

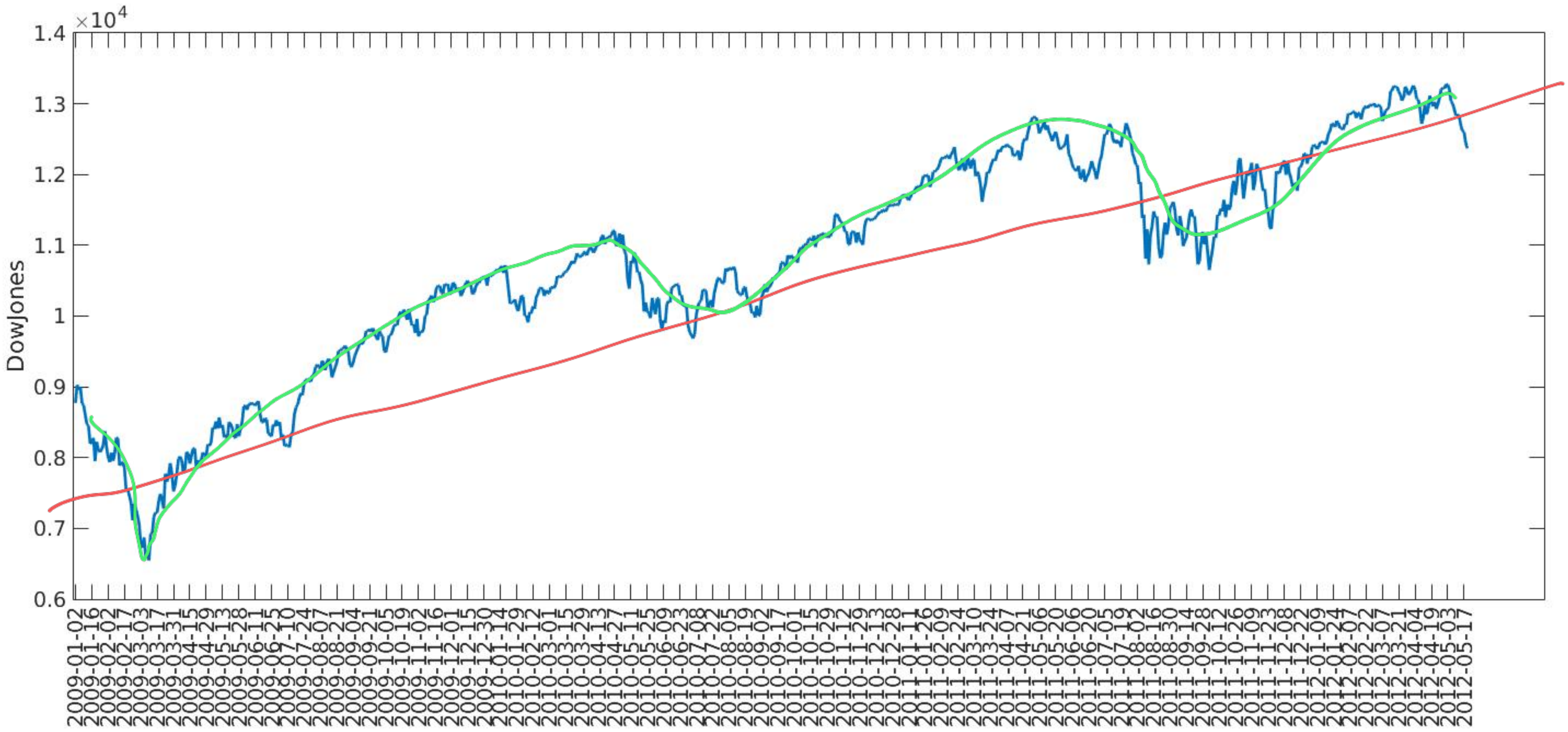
  $\uparrow$

  DS

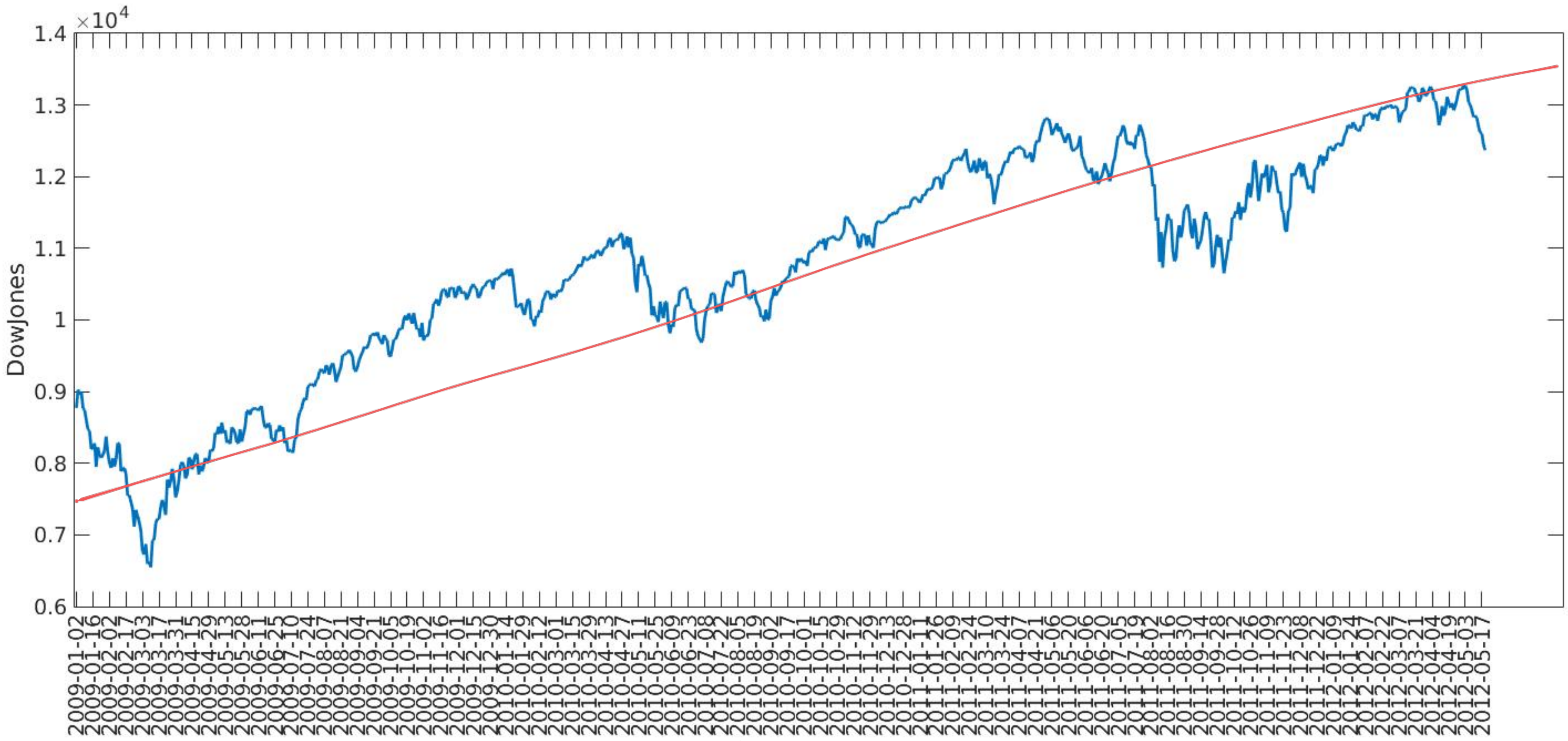
- Model-based TSA

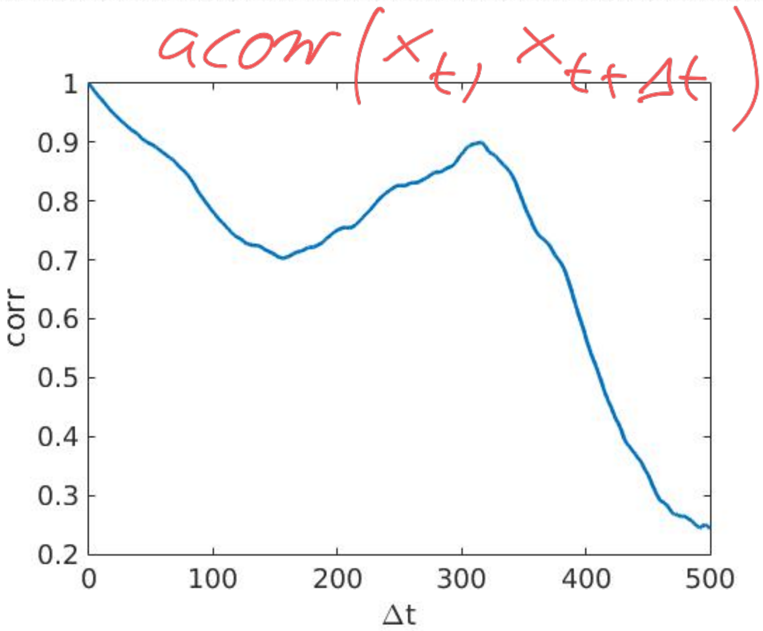
  $$x_t = F_\theta(x_{t-1} \dots x_{t-k})$$

- Statistical perspective
  - Dynamical systems persp.
    - → generating DS → TS
    - → model analysis

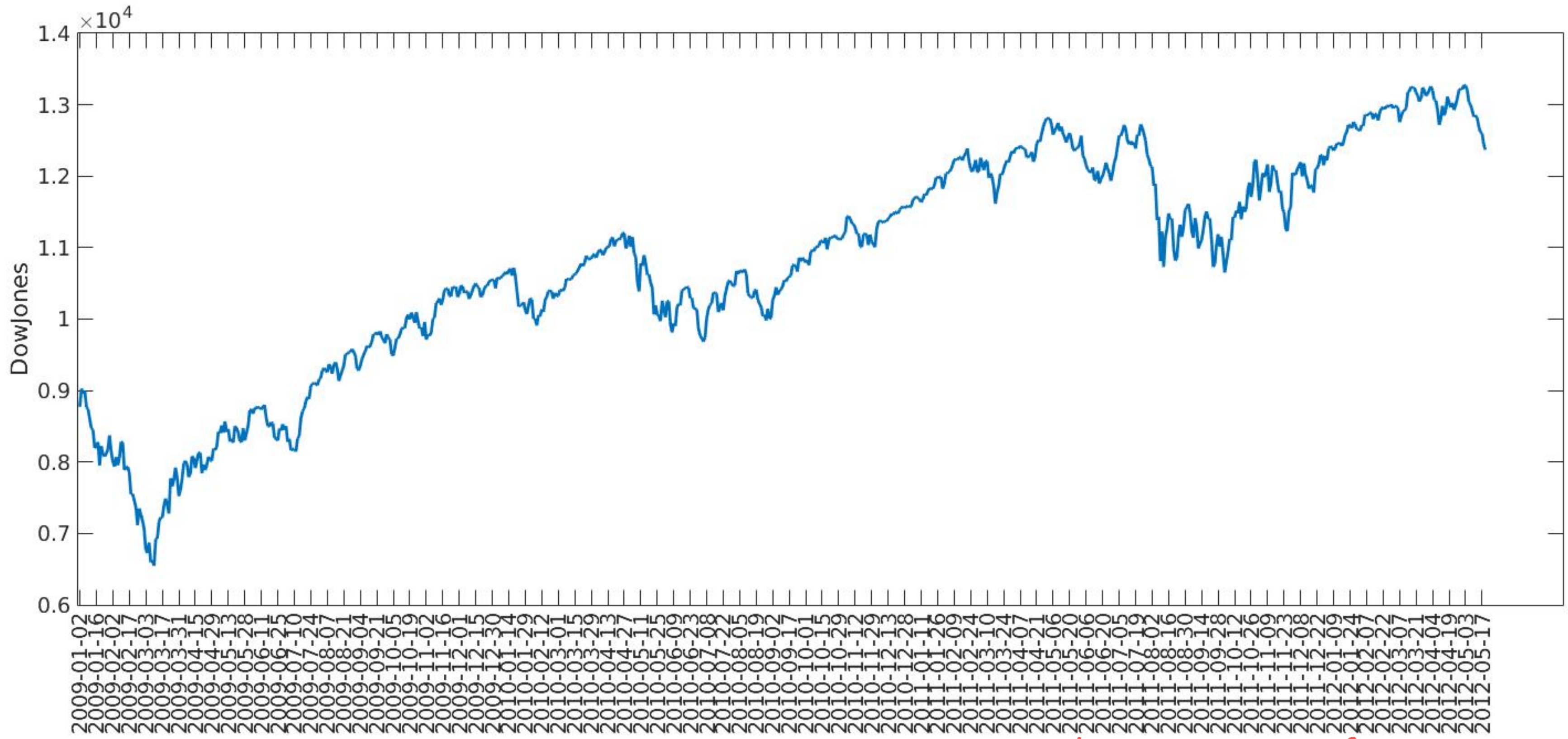


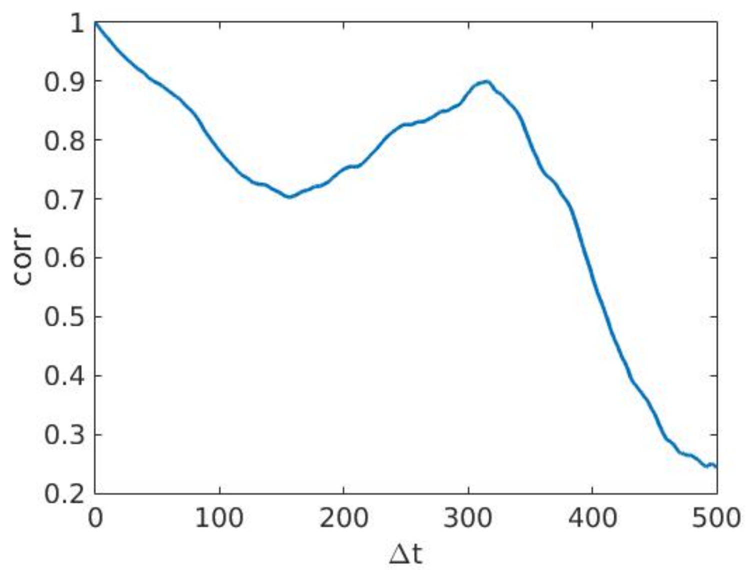
Source: http://www.wessa.net/finmardata.wasp



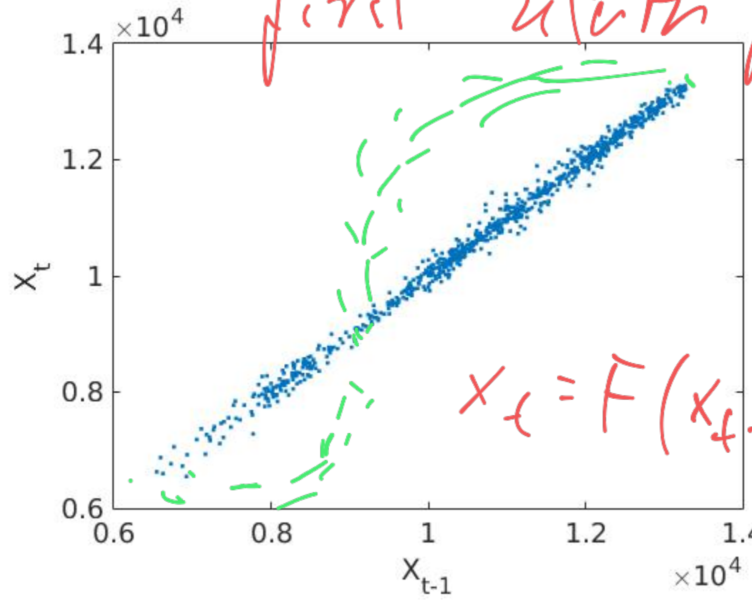


Source: http://www.wessa.net/finmardata.wasp



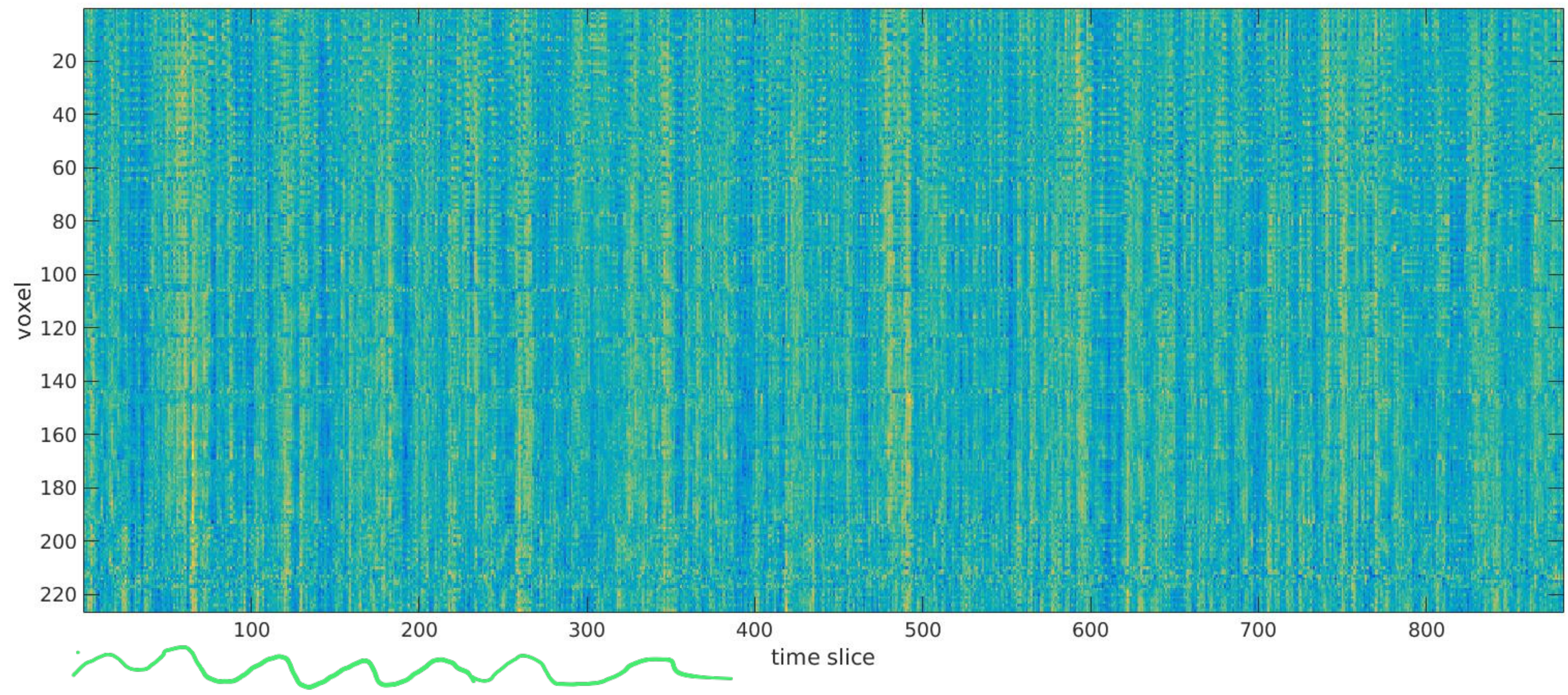


first-return plot

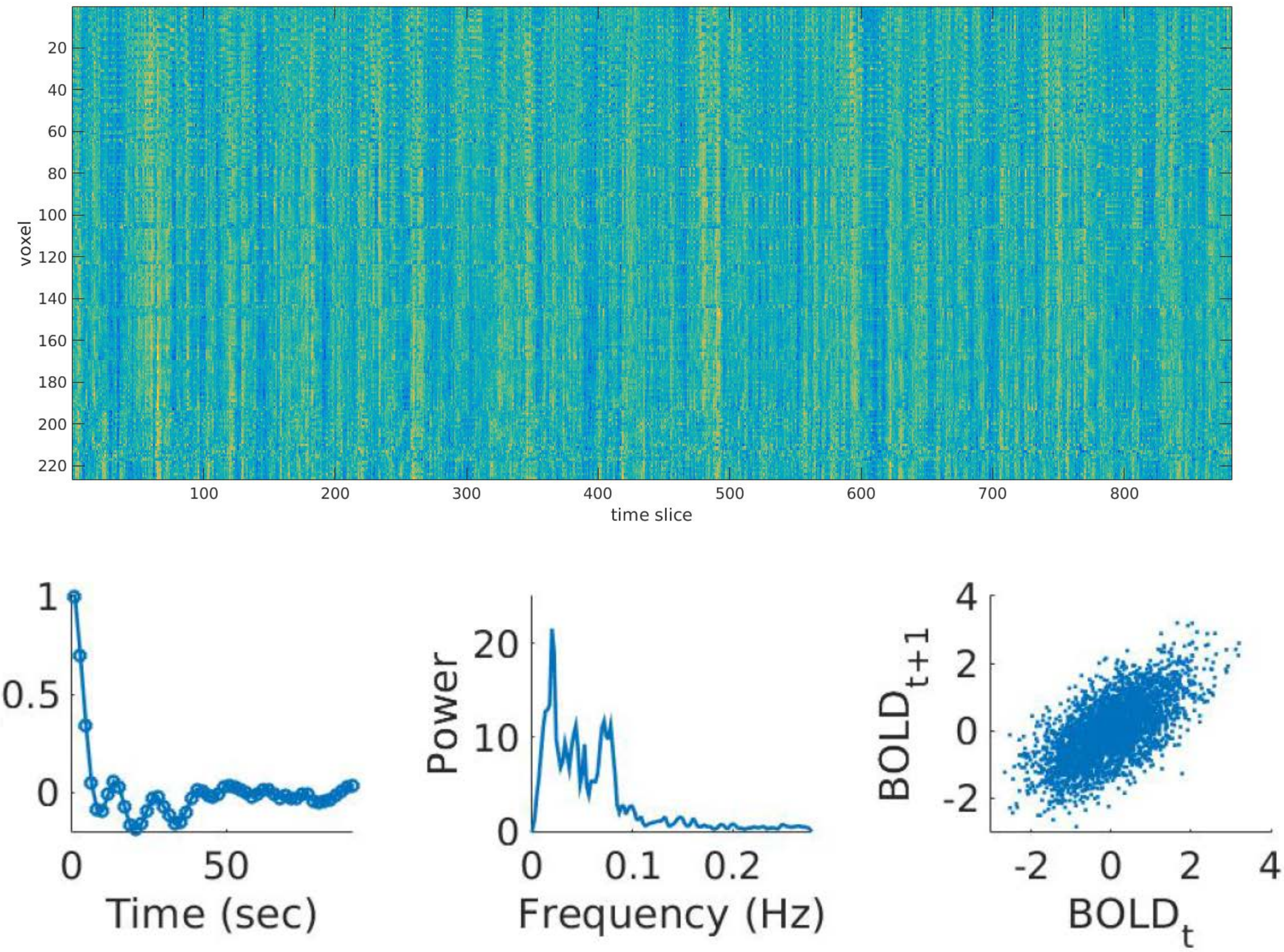


$x_t = F(x_{t-1})$

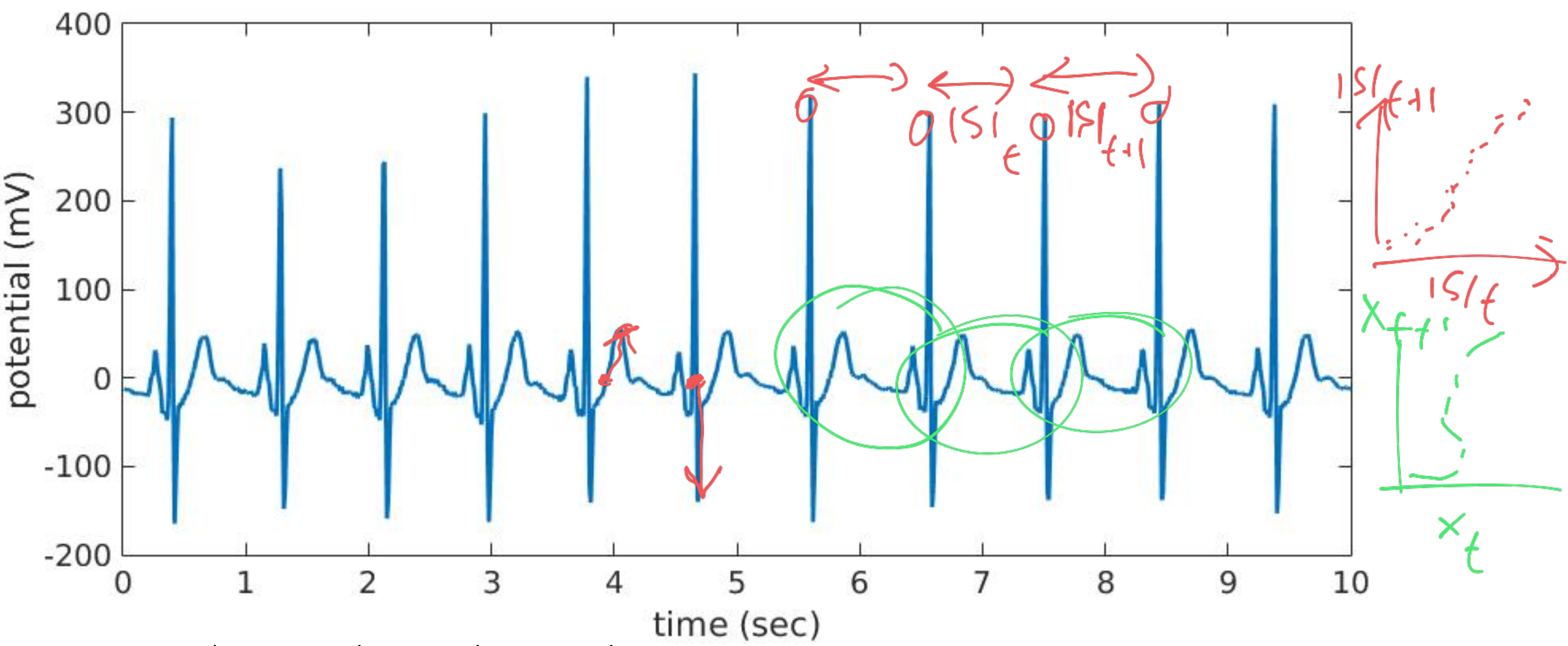
Source: http://www.wessa.net/finmardata.wasp



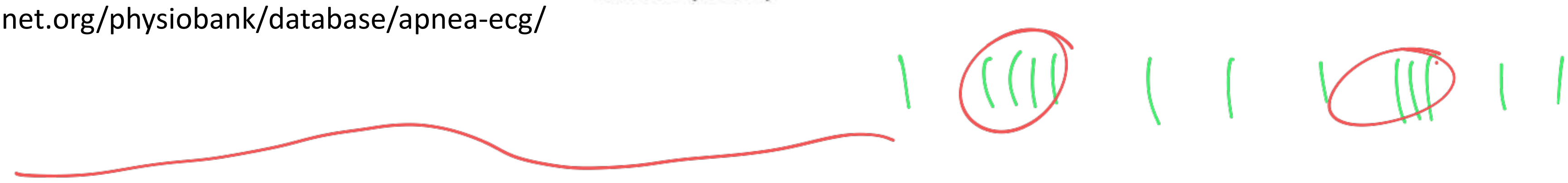
Source: data from Bähner et al. (2015) Neuropsychopharmacology

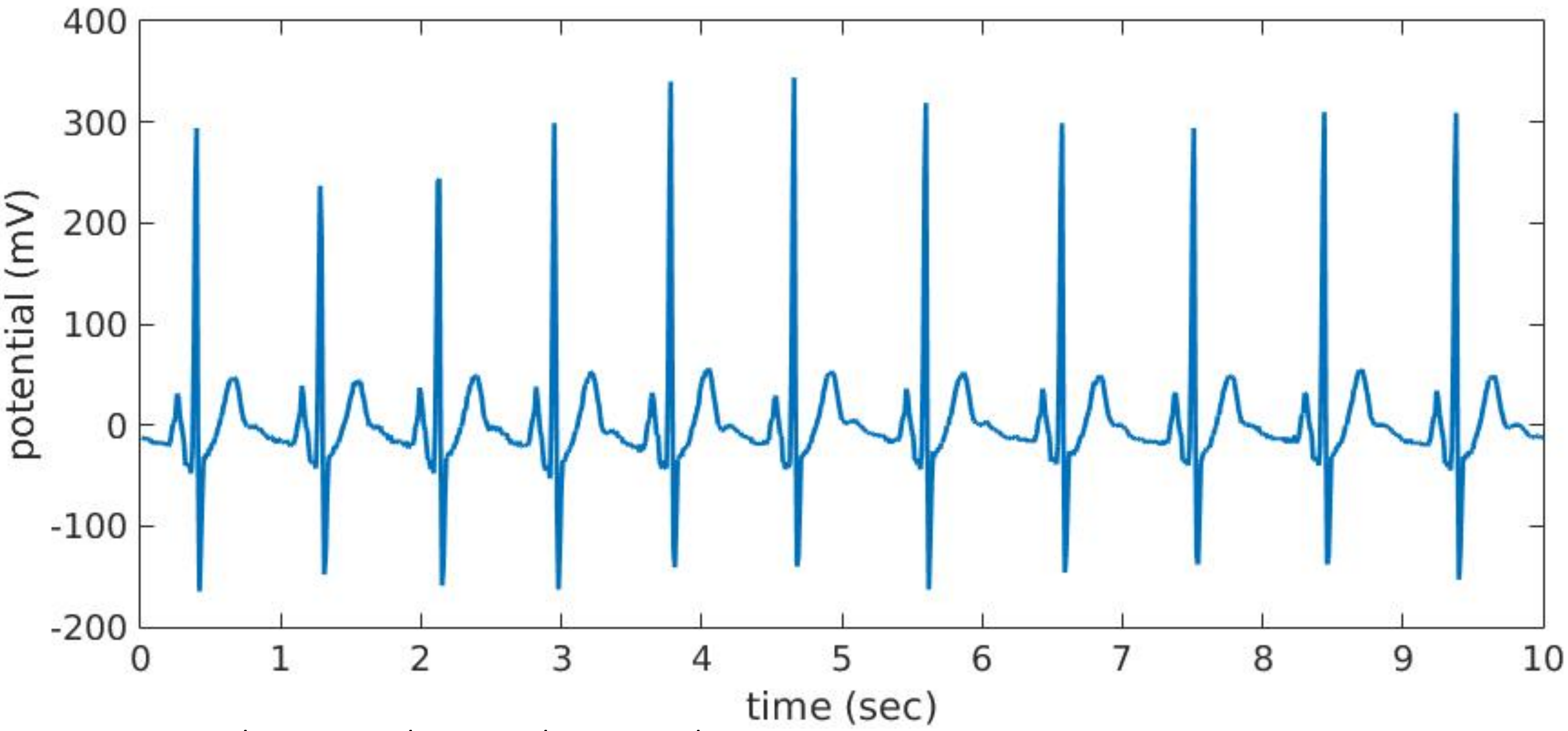


Source: data from Bähner et al. (2015) Neuropsychopharmacology

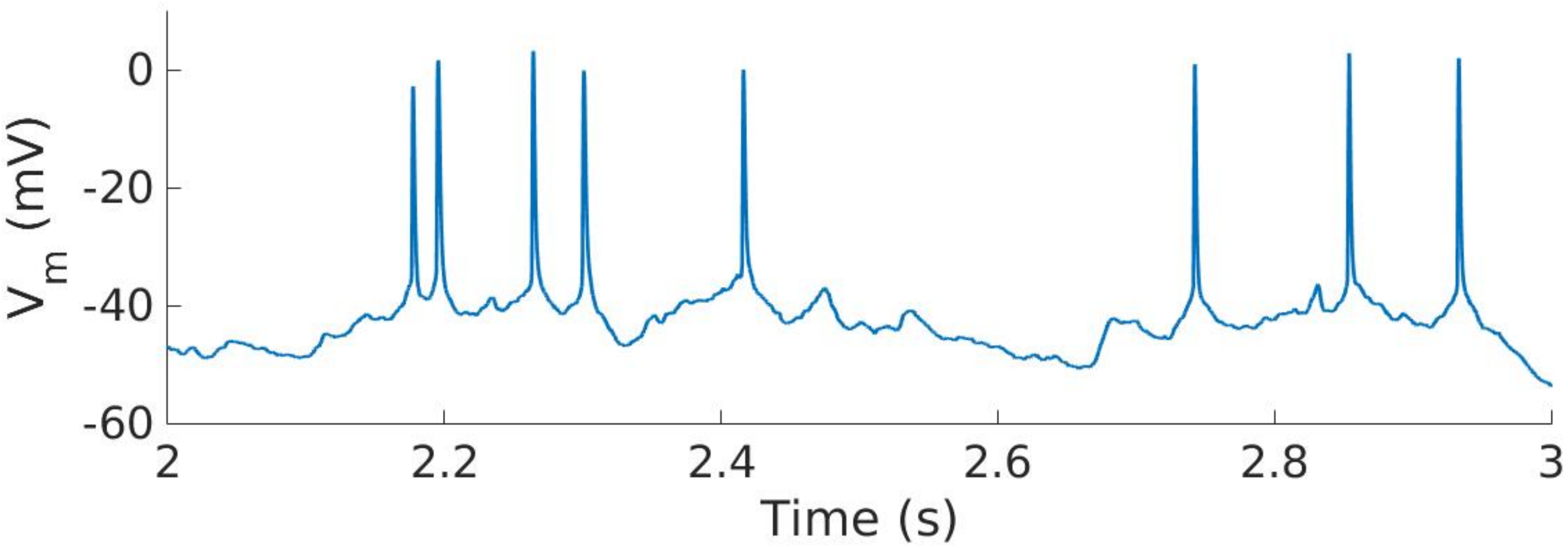


Source: www.physionet.org/physiobank/database/apnea-ecg/

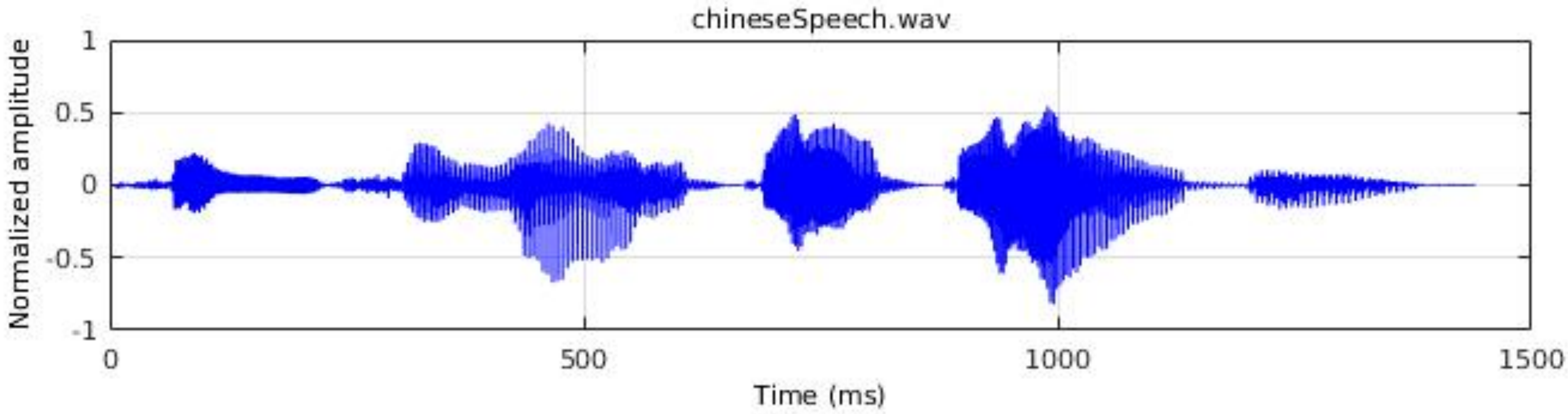




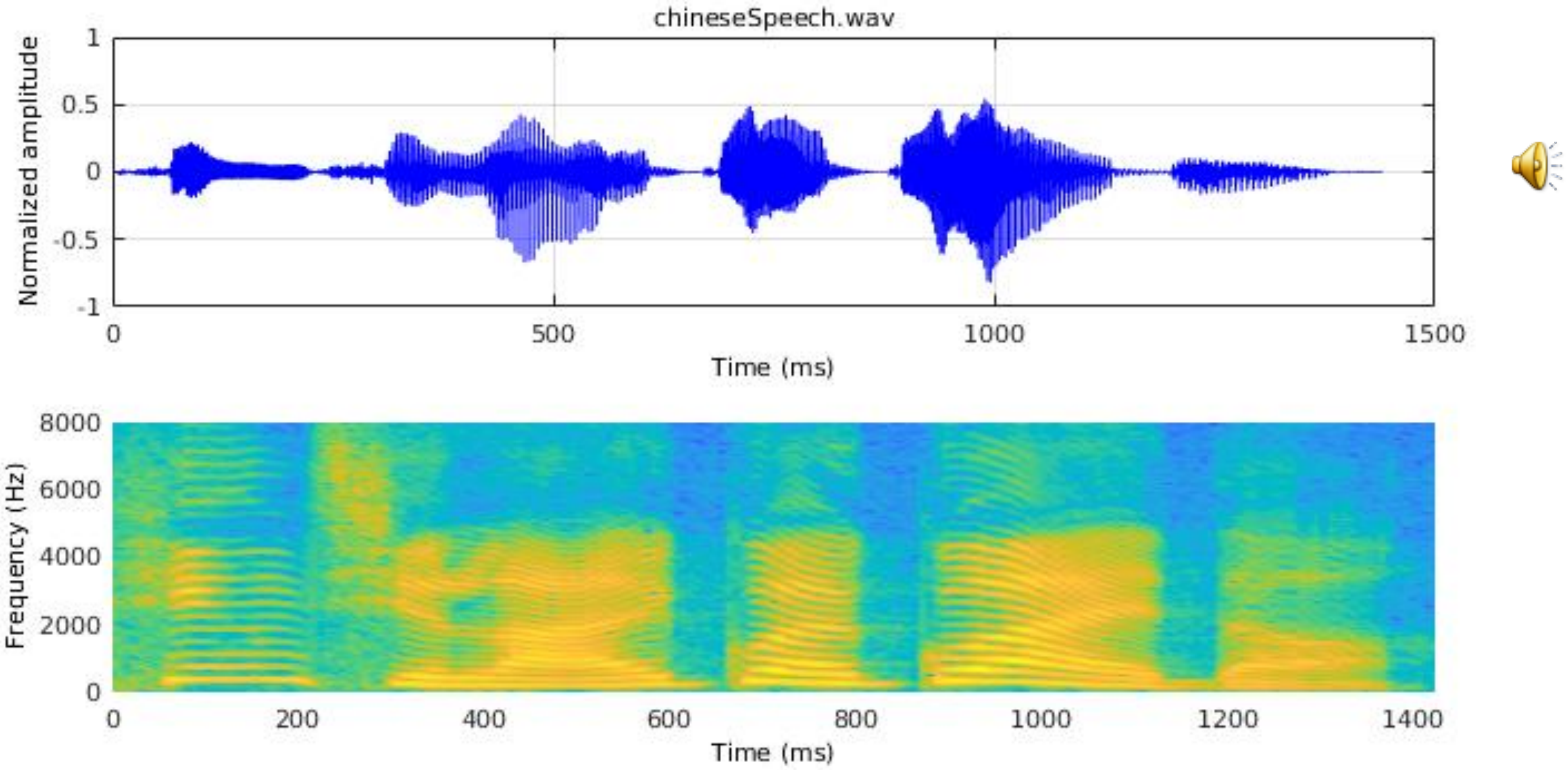
Source: www.physionet.org/physiobank/database/apnea-ecg/



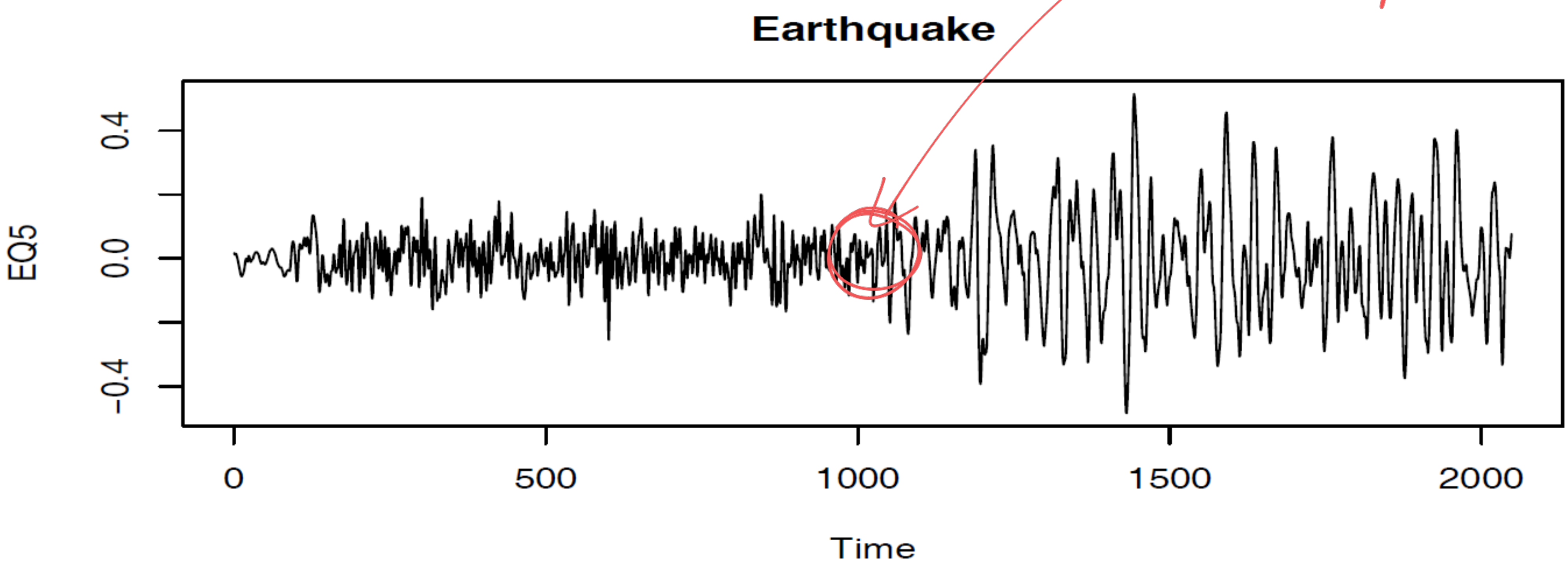
Source: in-vivo patch-clamp recording by Thomas Hahn, ZI MA

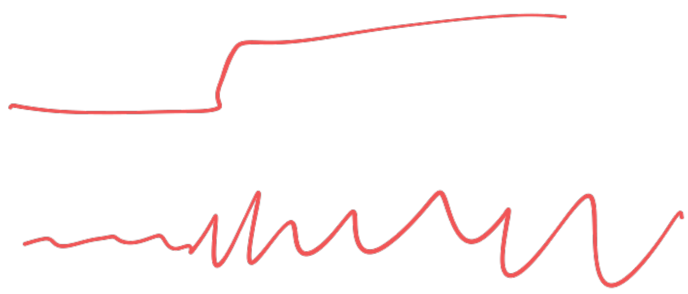


Provided by Emili Balaguer-Ballester (Bournemouth)

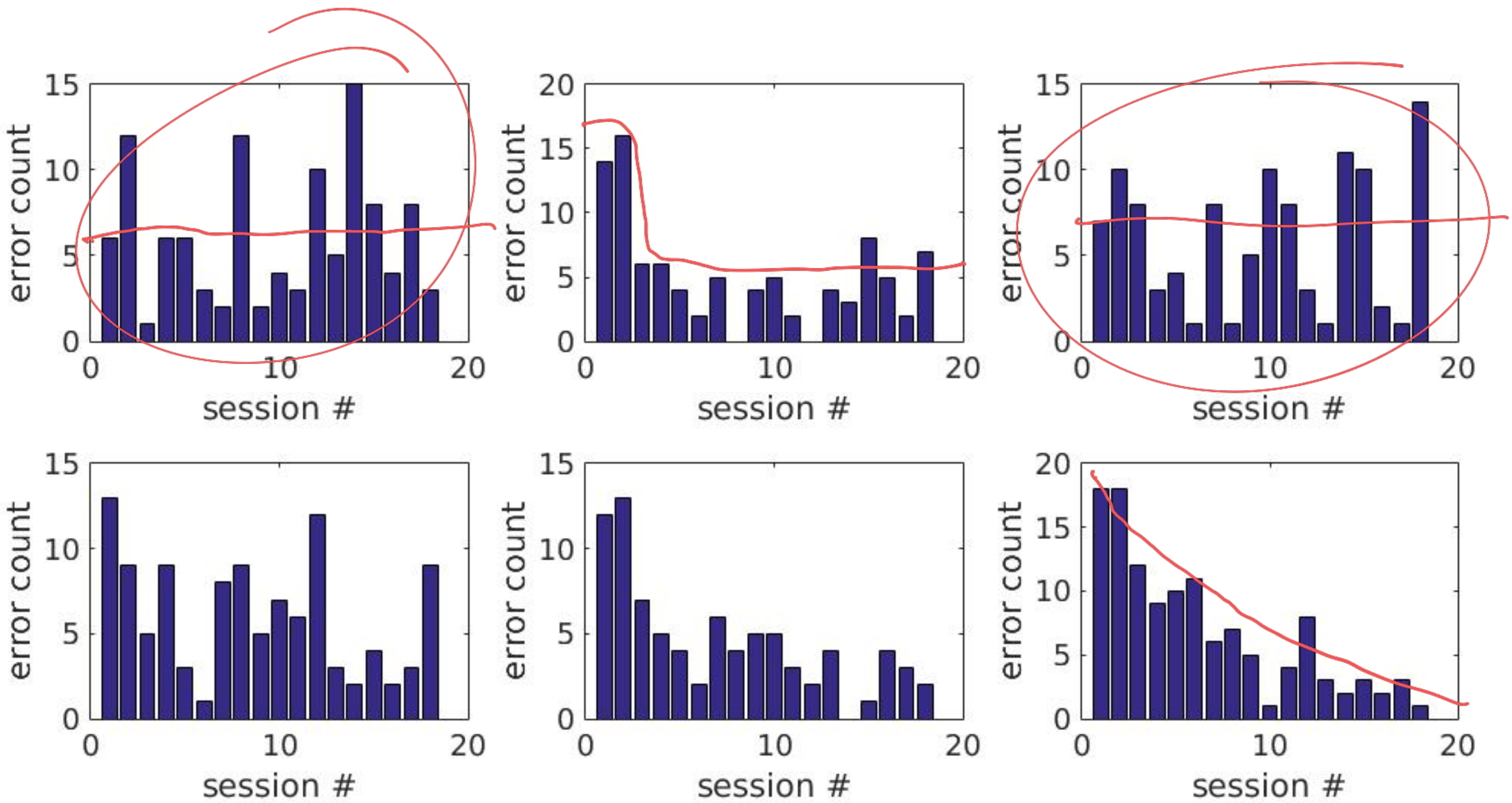


Provided by Emili Balaguer-Ballester (Bournemouth) change point



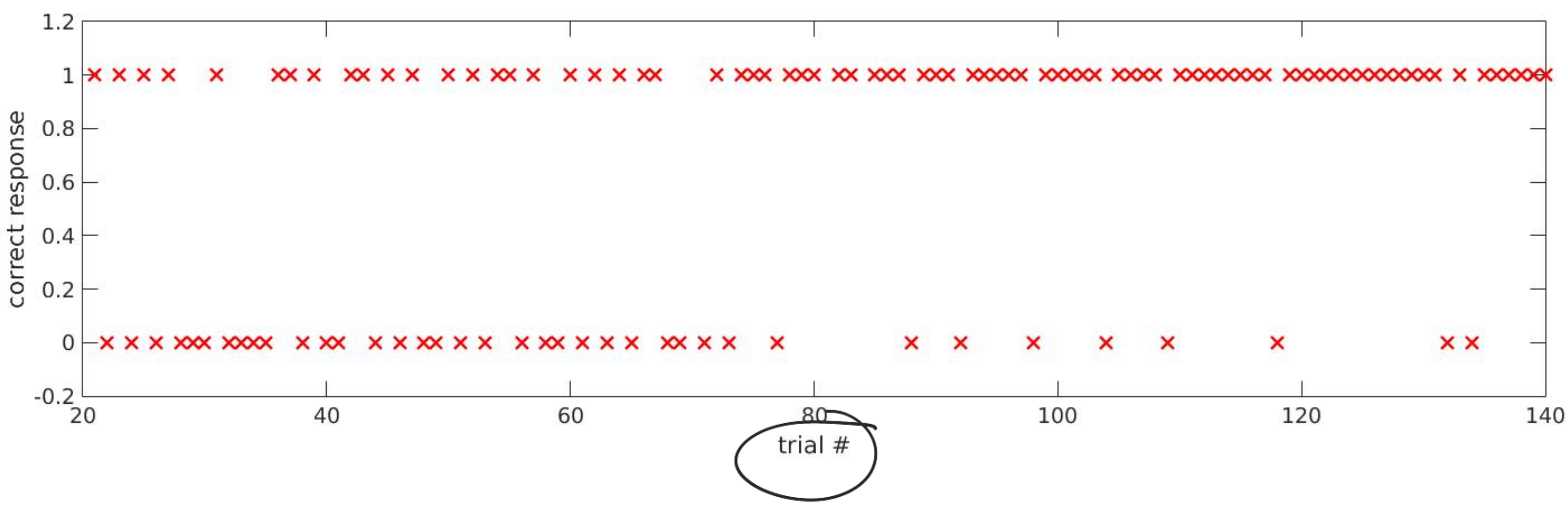


From Shumway & Stoffer (2011), Time Series Analysis

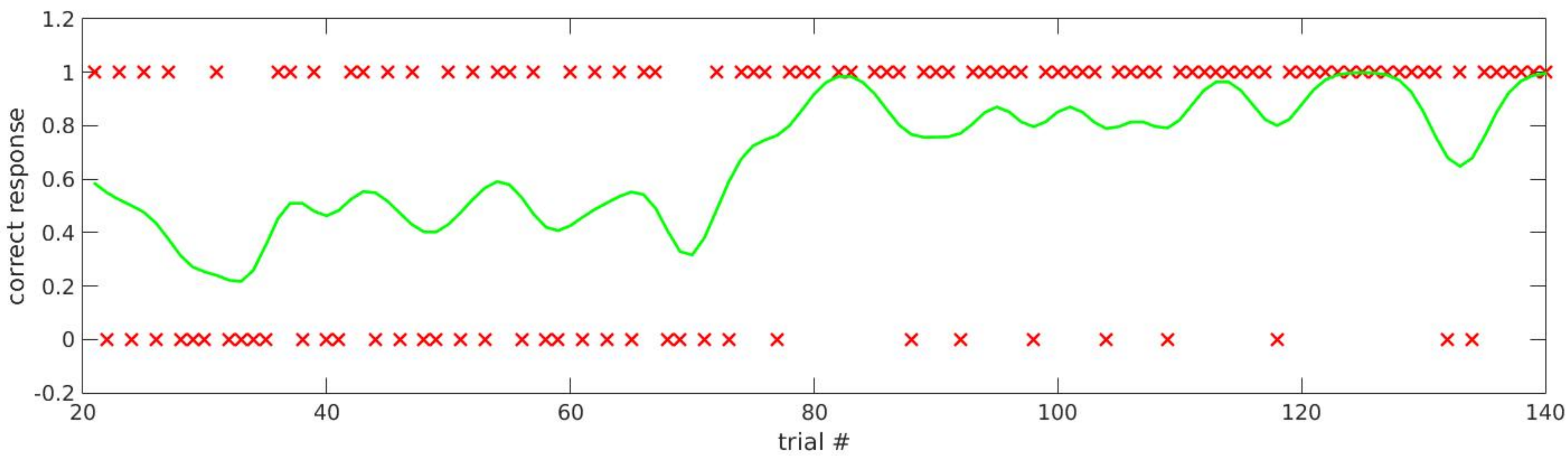


Source: data from Richter et al. (2013) PLoS One

$\{0, 1\}$

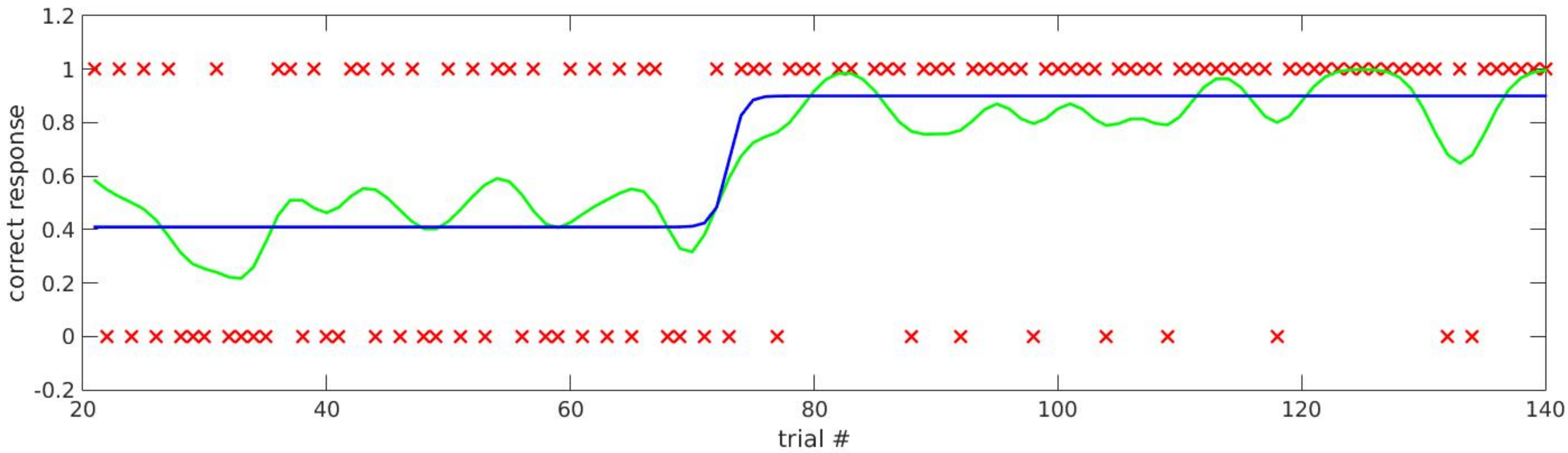


Source: data from Durstewitz et al. (2010) Neuron

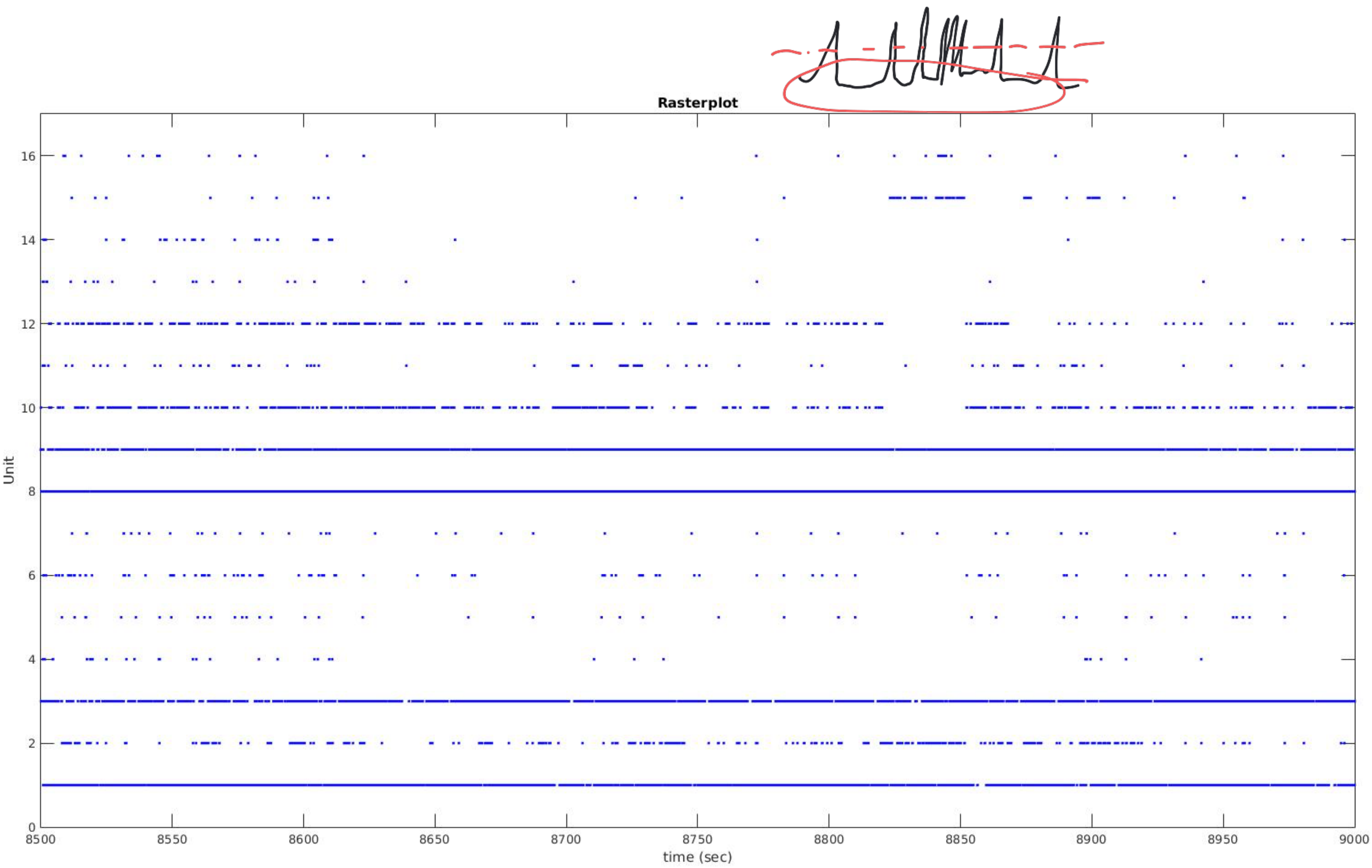


Source: data from Durstewitz et al. (2010) Neuron

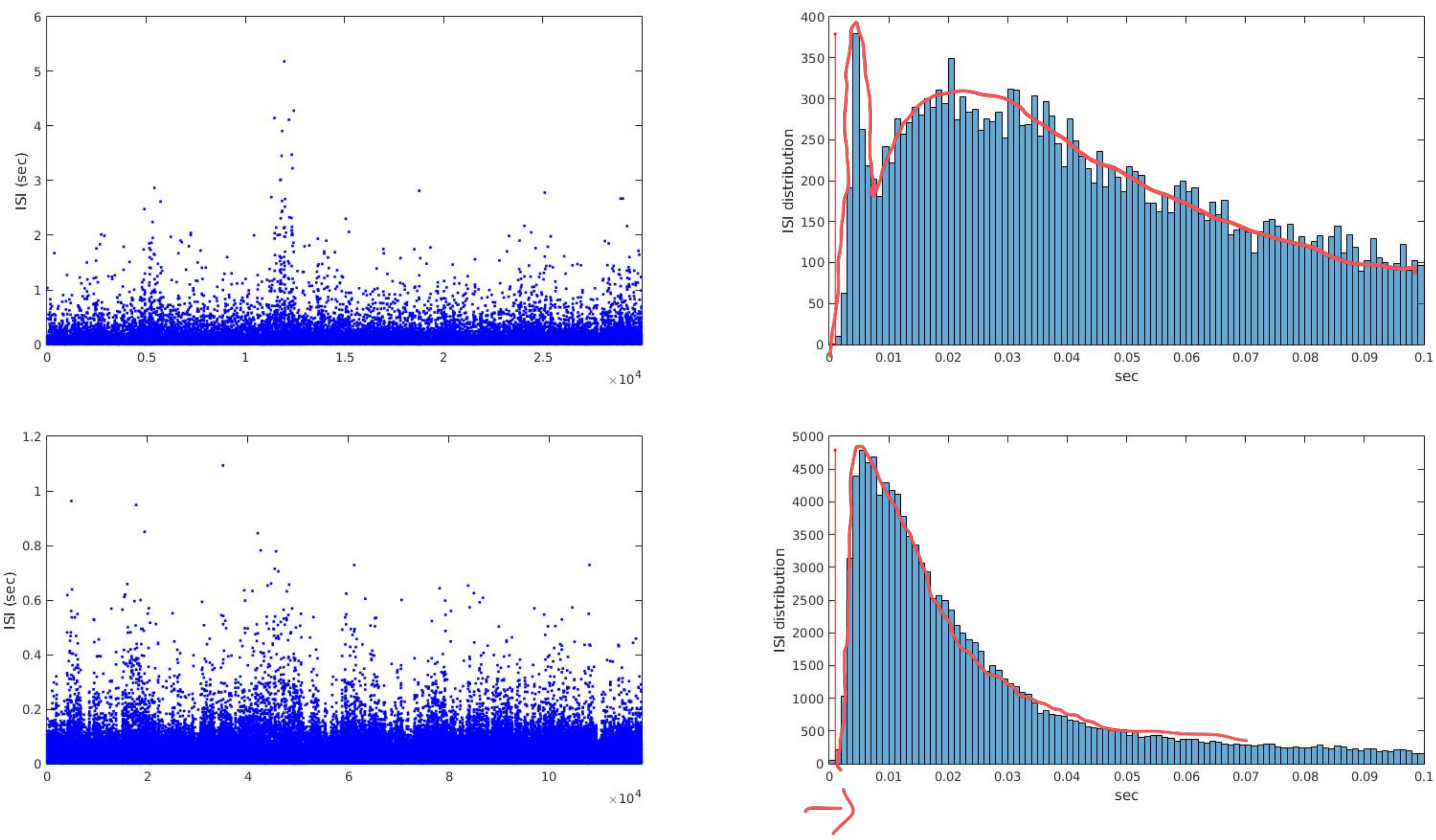
$$\mu + \frac{\theta}{1 + e^{-\alpha(x-\beta)}}$$



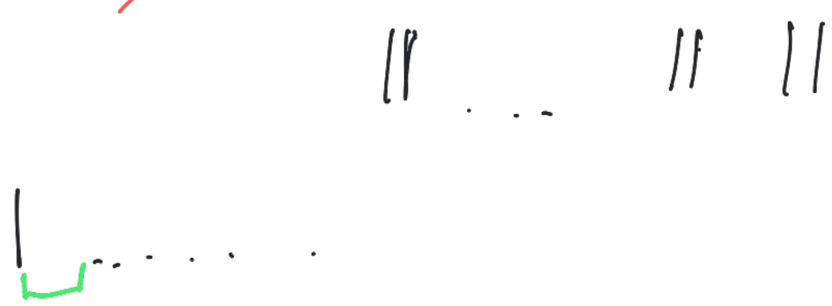
Source: data from Durstewitz et al. (2010) Neuron

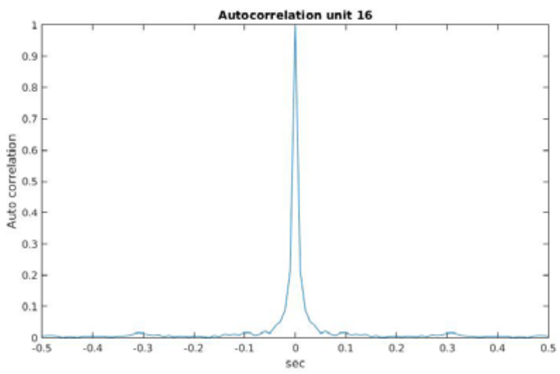


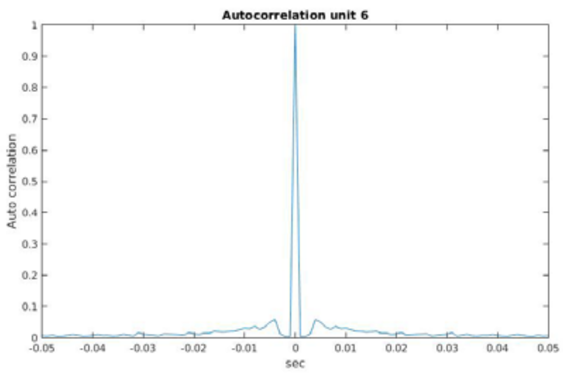
$\{t_i\}$ point process

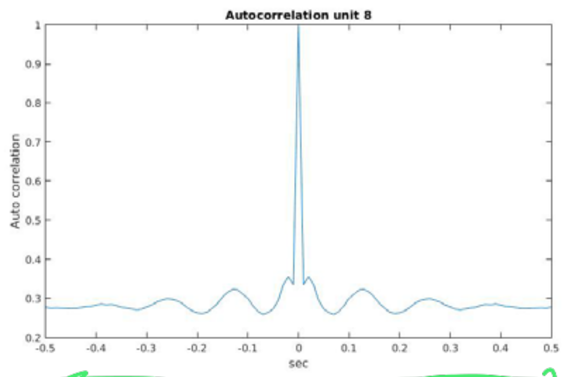


$\{ t_{i+1} - t_i \}$

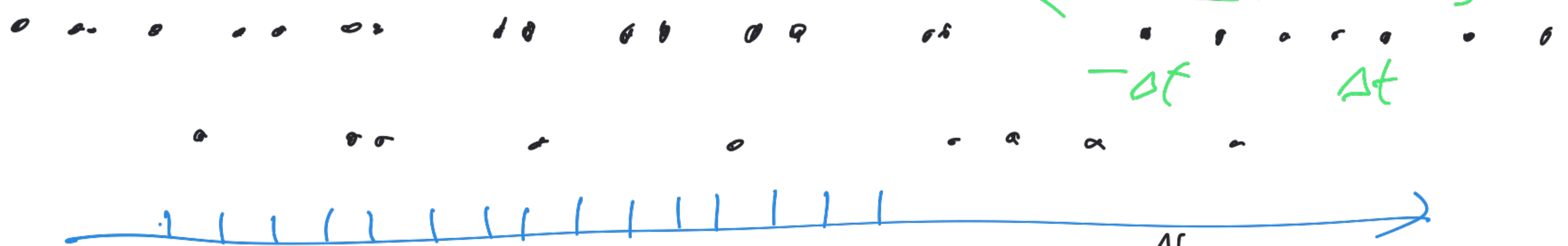








$-\Delta t$ Δt

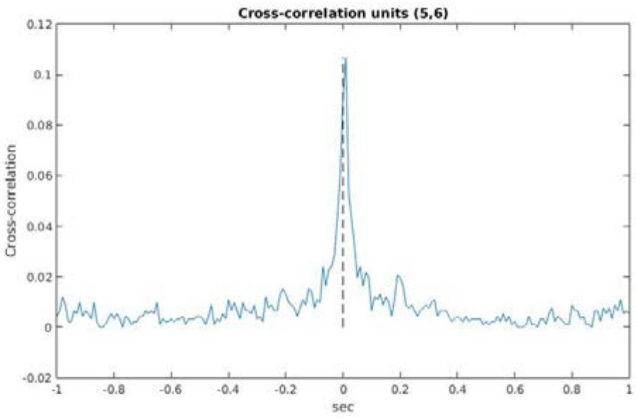


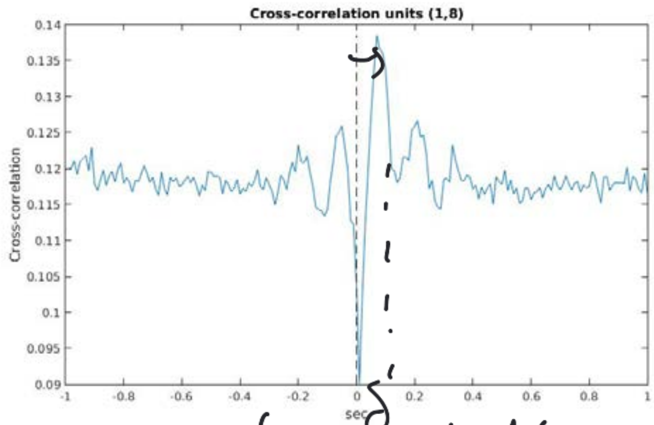
0 1 0 0 1 1 0 0 1 0 0 0 1

$x_t \in \{0, 1\}$

$$\mathrm{acov}[x_t, x_{t+\Delta t}] = \underbrace{E[x_t \cdot x_{t+\Delta t}]}_{p(x_t = 1 \wedge x_{t+\Delta t} = 1)} - \underbrace{E[x_t]}_{} E[x_{t+\Delta t}]$$

$$\underbrace{p(x_t = 1 \wedge x_{t+\Delta t} = 1)}_{p(x_t = 1)\, p(x_{t+\Delta t} = 1)} - p(x_t = 1)\, p(x_{t+\Delta t} = 1)$$





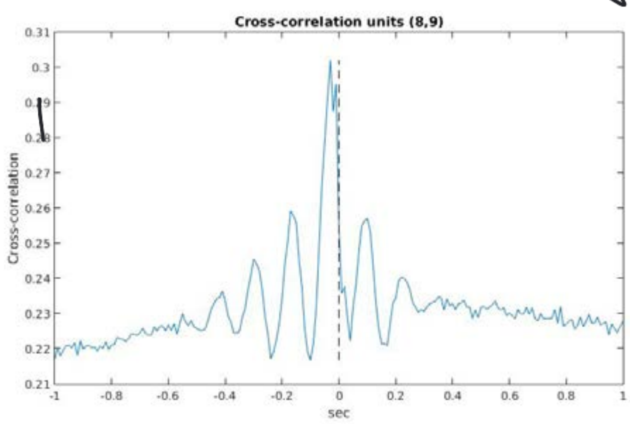
$-\Delta t \qquad \delta \qquad +\Delta t$

x

y

δ

0 1 0 1 1 0 1



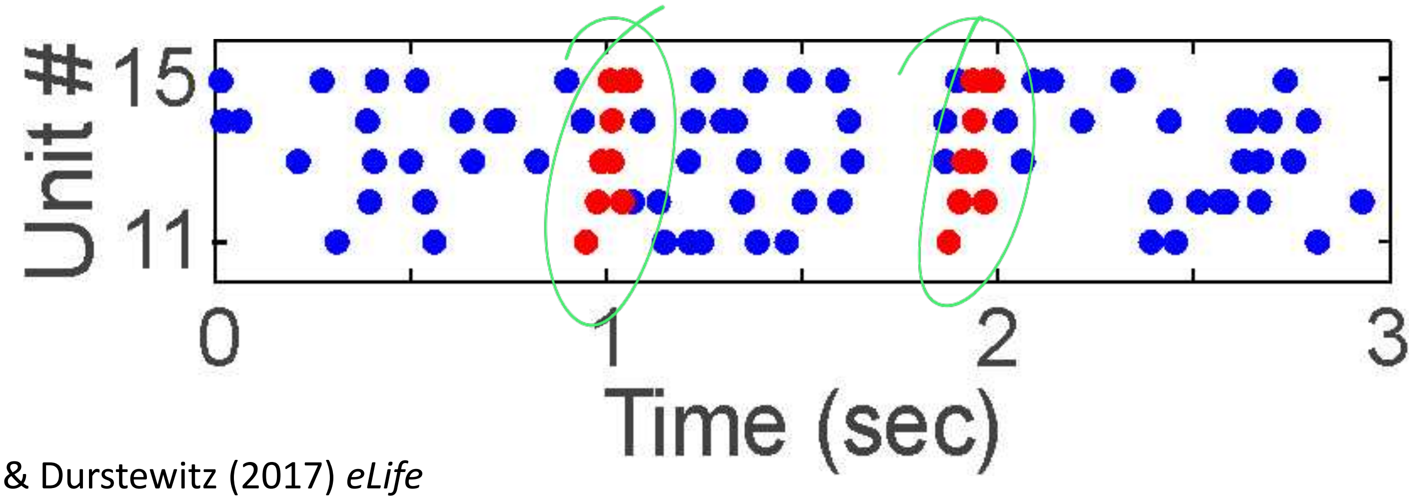
Δt

$x_t \in \{0,1\}$

$y_t \in \{0,1\}$

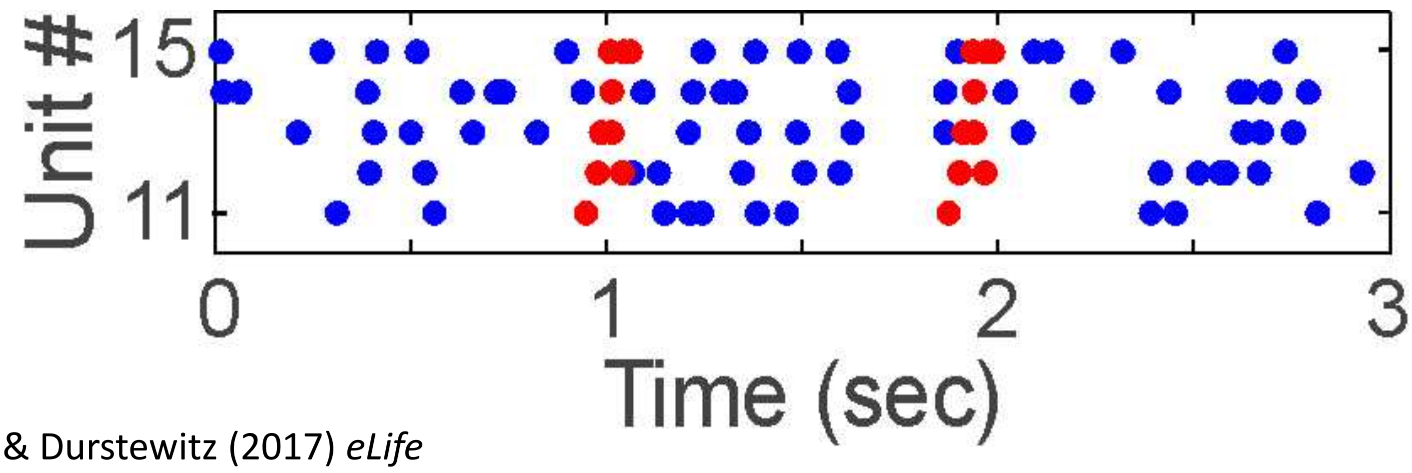
$\mathrm{xcorr}(x_t, y_{t+\Delta t})$

Precise non-sequential patterns

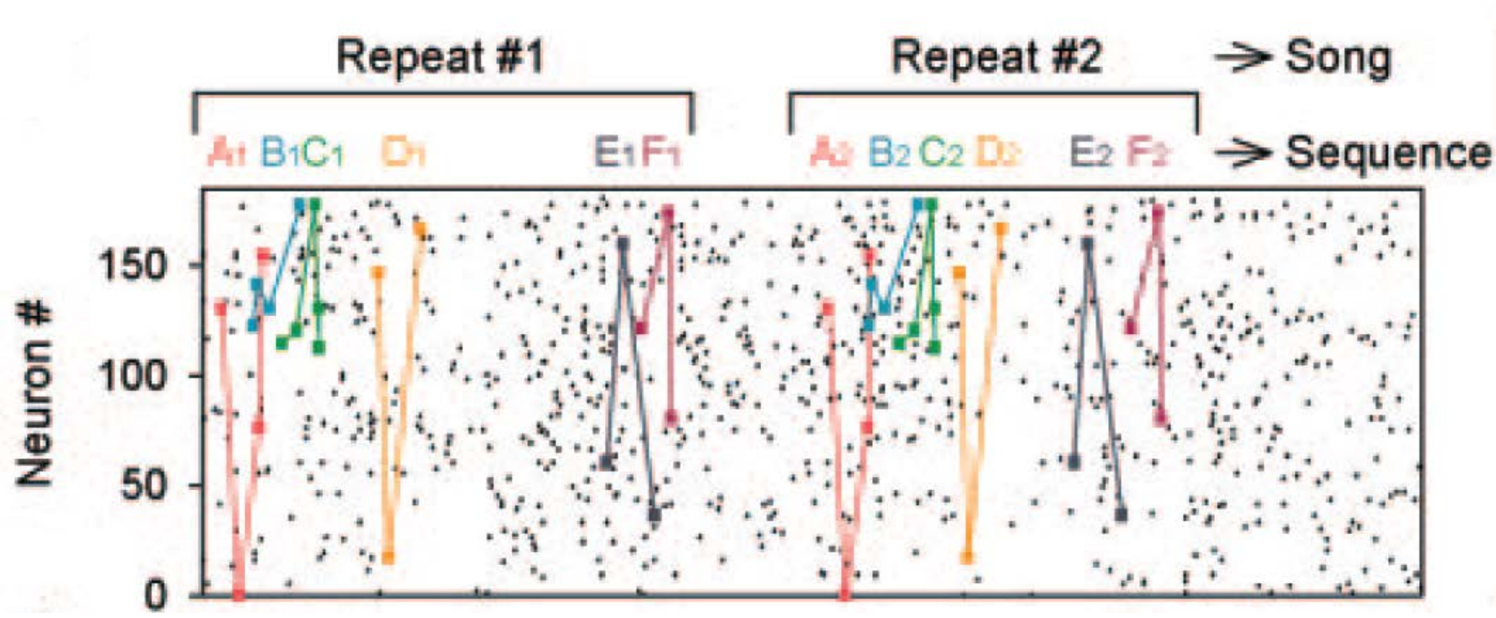


Source: Russo & Durstewitz (2017) *eLife*

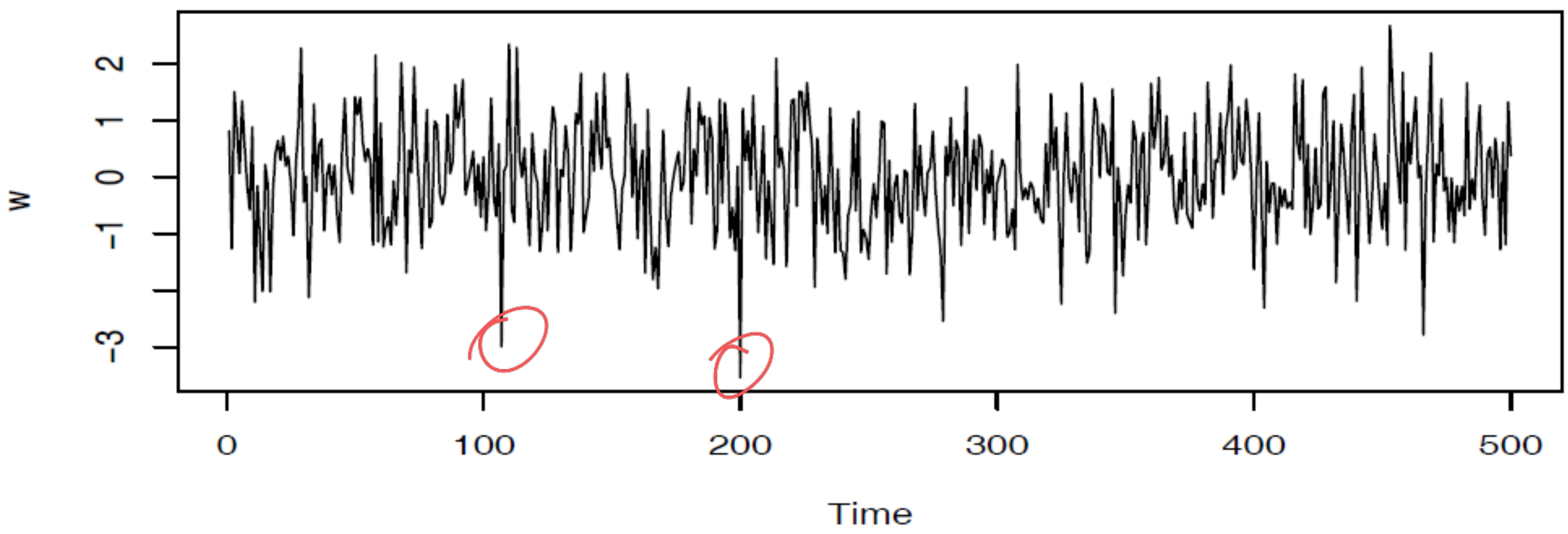
Precise non-sequential patterns



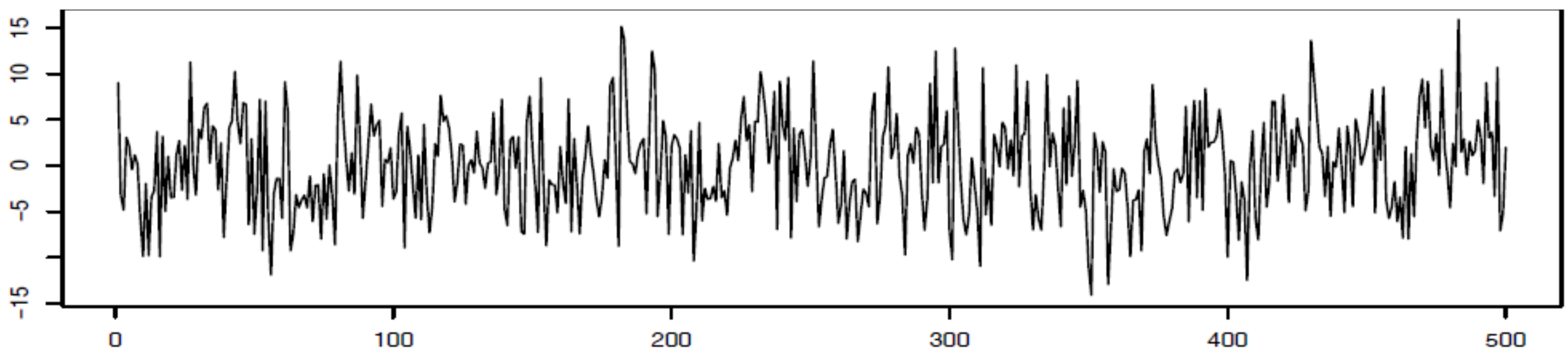
Source: Russo & Durstewitz (2017) *eLife*



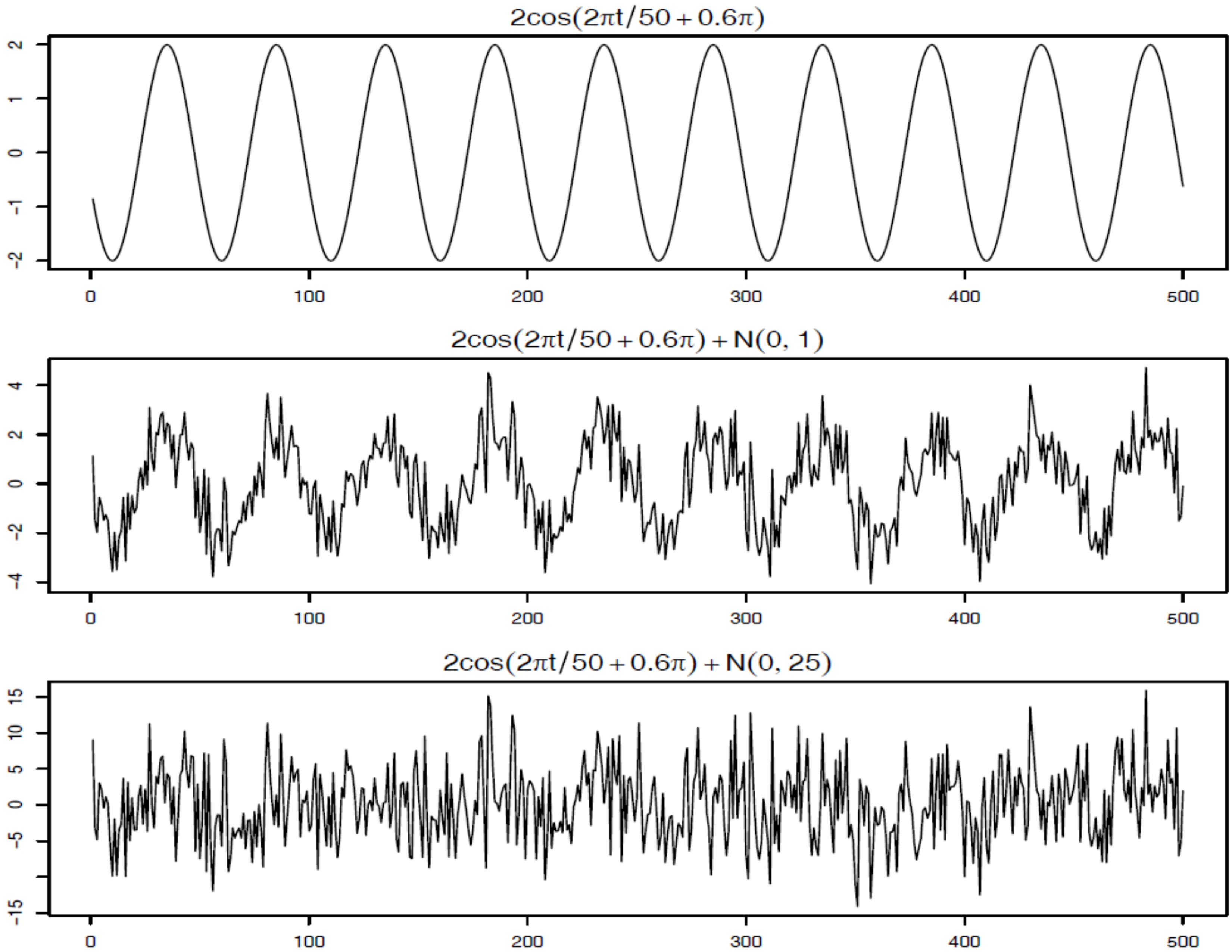
Ikegaya et al. (2004), *Nature*



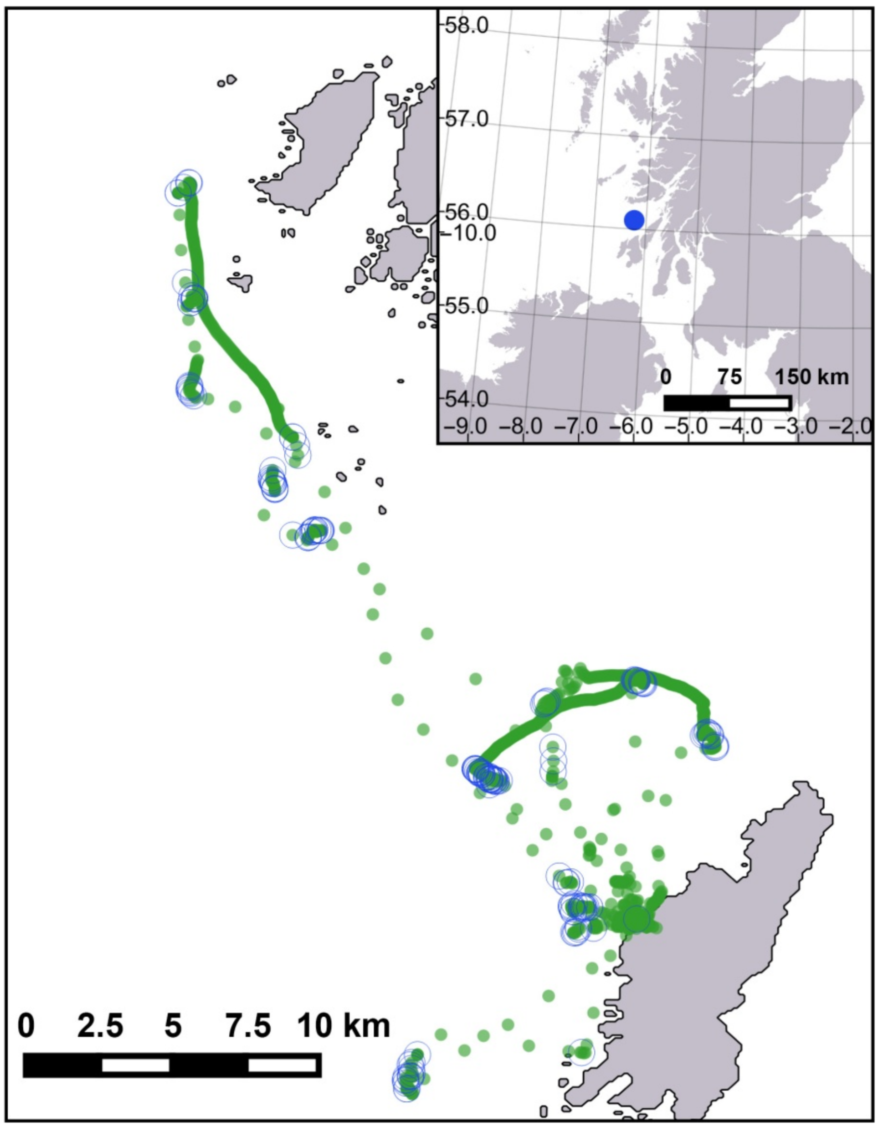
From Shumway & Stoffer (2011), Time Series Analysis



From Shumway & Stoffer (2011), Time Series Analysis



From Shumway & Stoffer (2011), Time Series Analysis

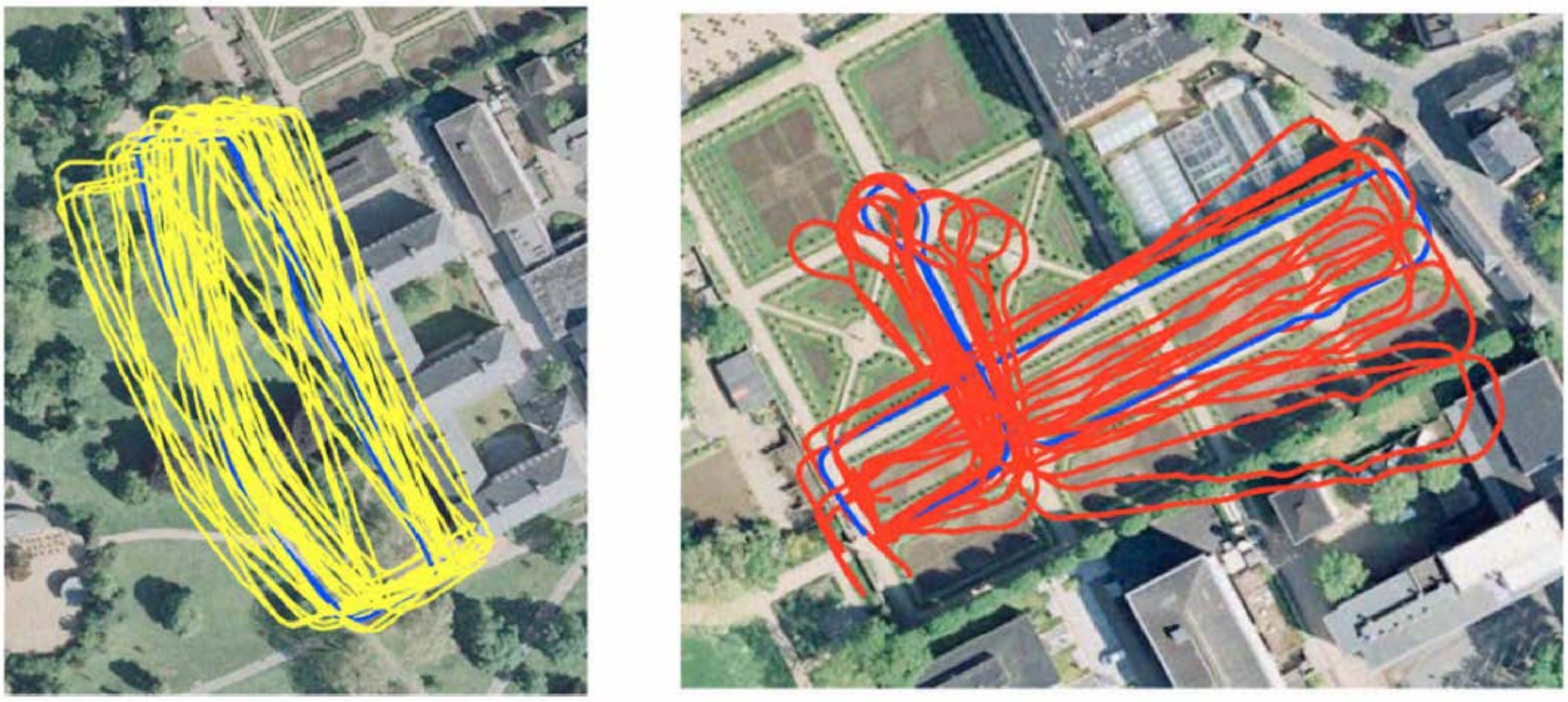


$$\begin{pmatrix} x_t \\ y_t \end{pmatrix}$$

<u>Source:</u> Browning et al. (2018) Predicting animal behaviour using deep learning: GPS data alone accurately predict diving in seabirds. Methods in Ecology and Evolution 9 (3): 681-692

Pedestrian Dead Reckoning

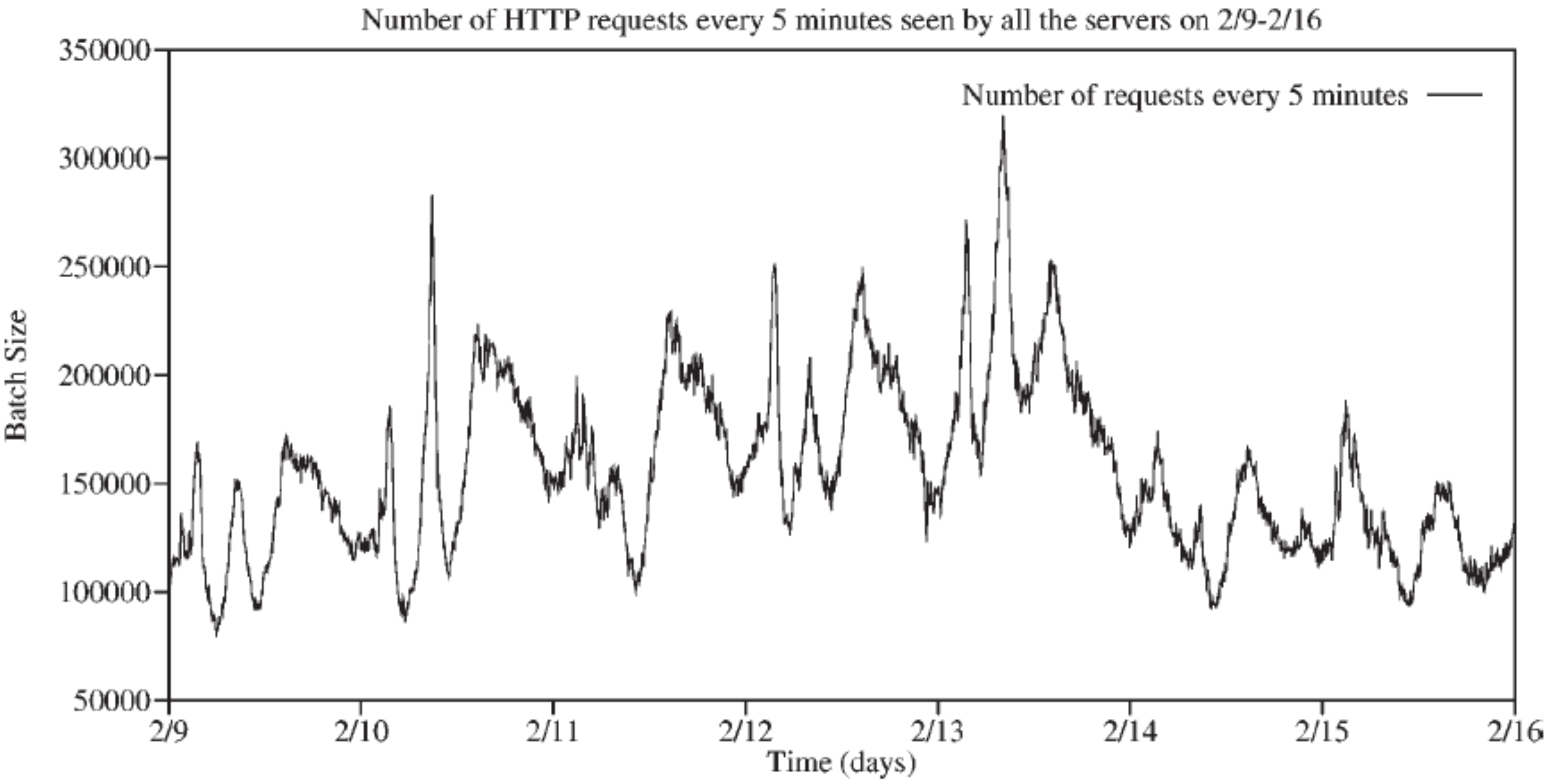
- Pedestrian tracking with step detection:



- Problem: drift → error accumulates over time

Source: Chen, C et al. (2018), *IONet: Learning to Cure the Curse of Drift in Inertial Odometry*, arXiv preprint, arXiv:1802.02209.

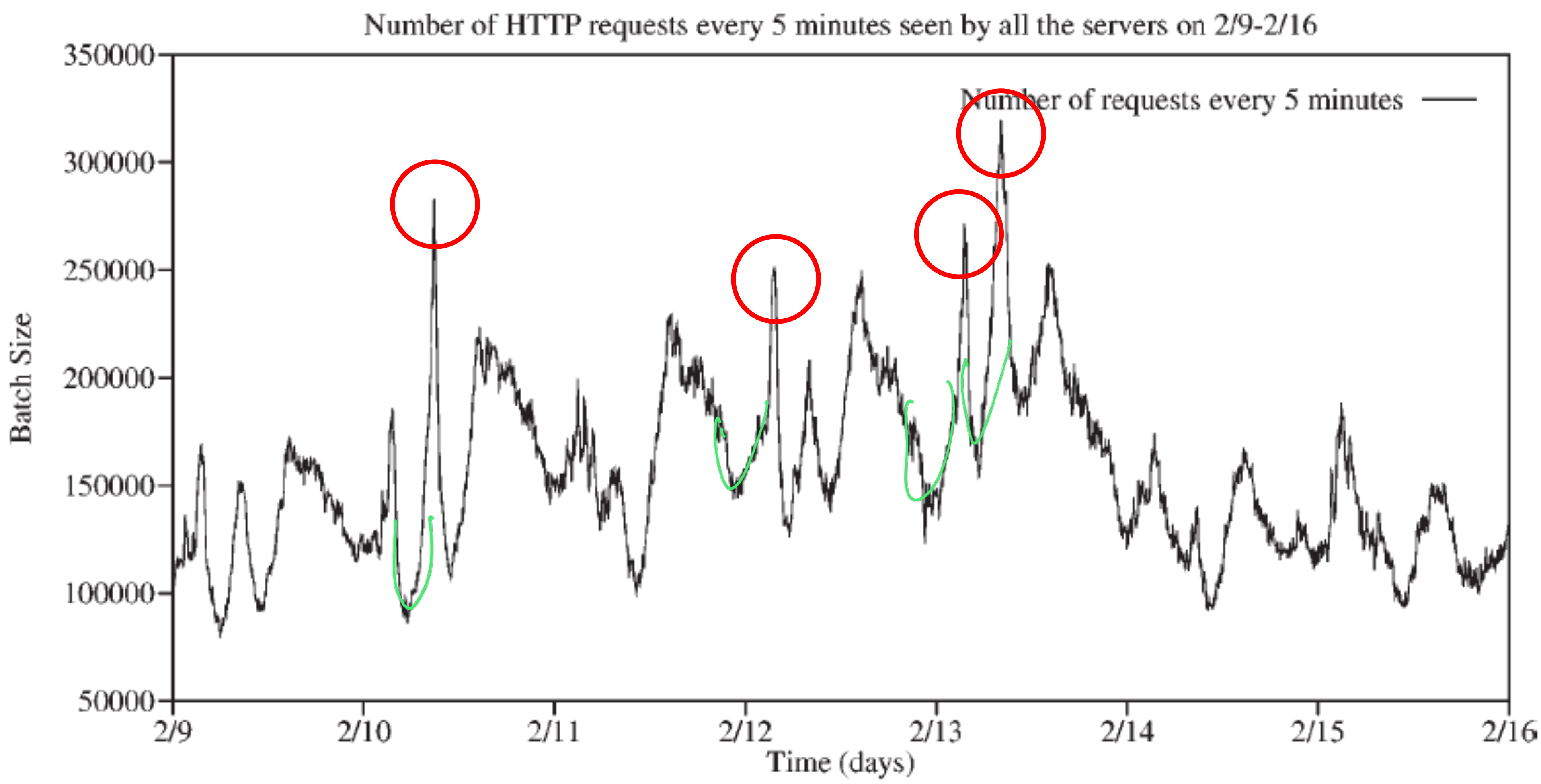
Website traffic prediction



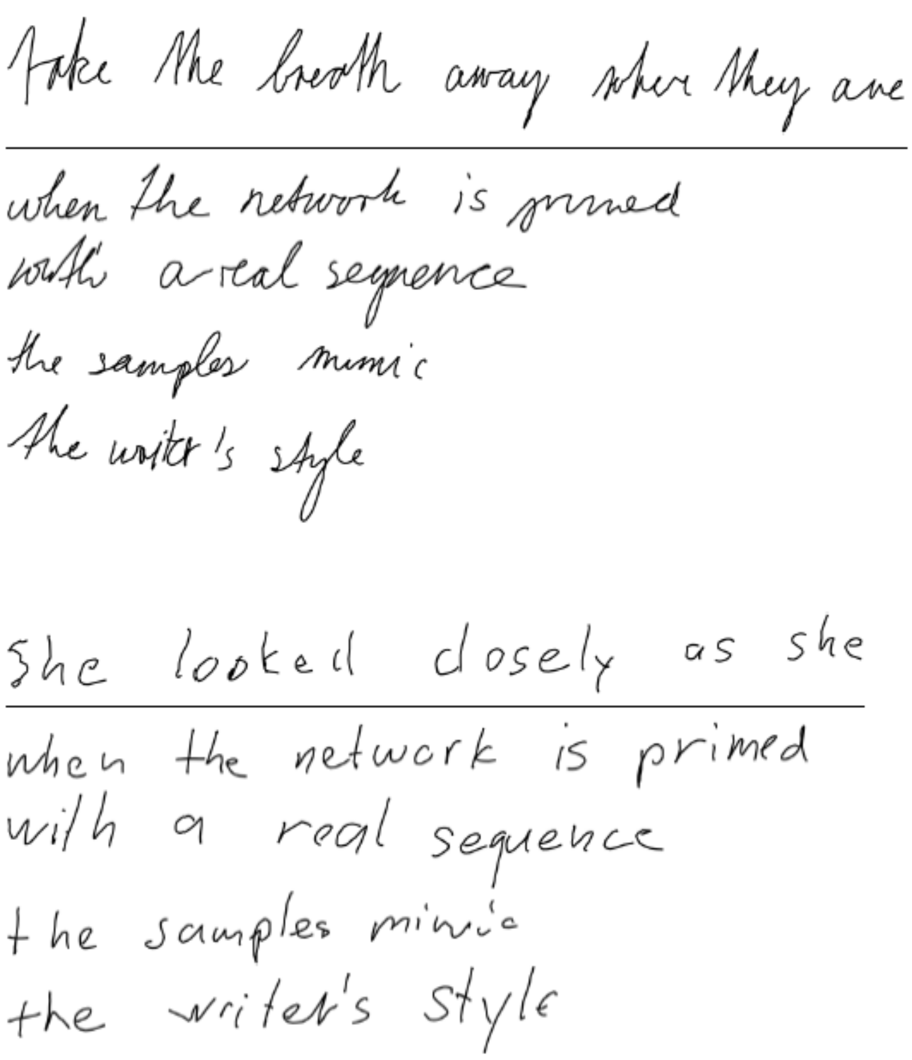
Source: Iyengar, A.K et al. L. *World Wide Web* (1999) 2: 85. https://doi.org/10.1023/A:1019244621570

Website traffic prediction

Goal: predict sharp maxima



Source: Iyengar, A.K et al. L. *World Wide Web* (1999) 2: 85. https://doi.org/10.1023/A:1019244621570



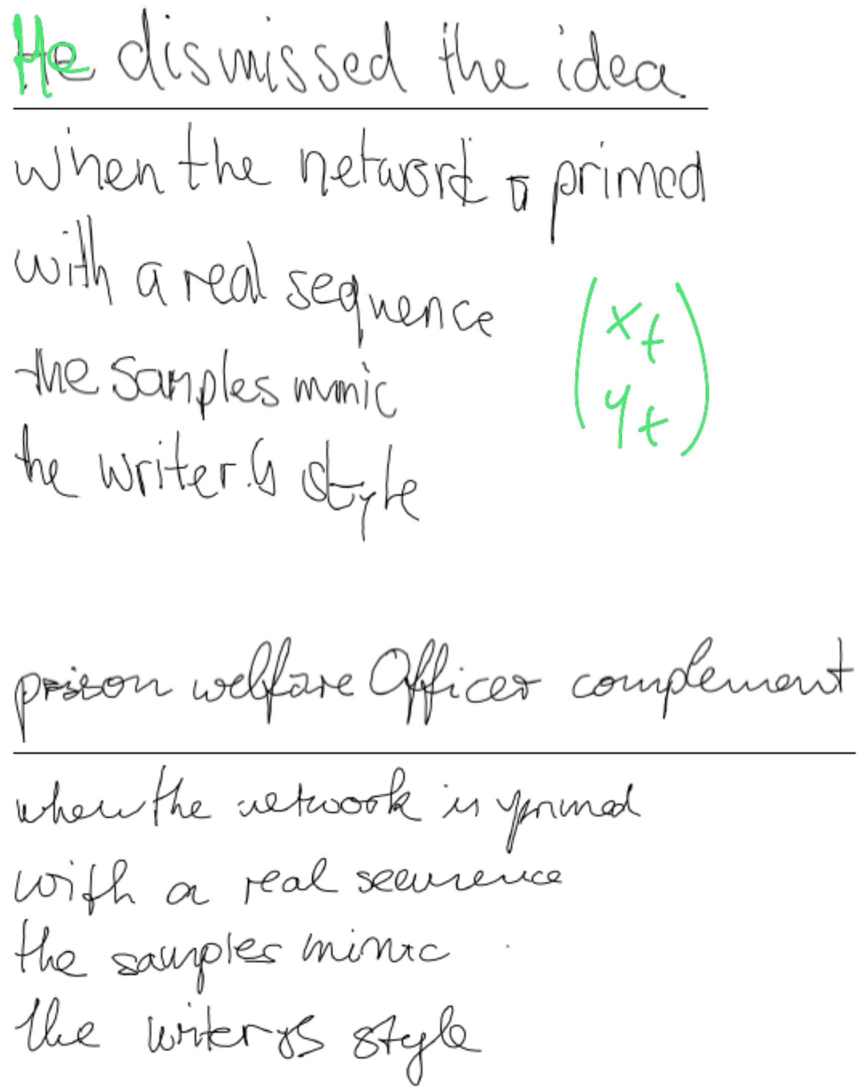


Figure 18: **Samples primed with real sequences.** The priming sequences (drawn from the training set) are shown at the top of each block. None of the lines in the sampled text exist in the training set. The samples were selected for legibility.

Graves (2014)
arXiv:1308.0850v5

With sweaty palms and heart racing, John drove to Sally's house for their first date. Sally, her pretty white dress flowing in the wind, carefully entered John's car. John and Sally drove to the movie theatre. John and Sally parked the car in the parking lot. Wanting to feel prepared, John had already bought tickets to the movie in advance. A pale-faced usher stood before the door; John showed the tickets and the couple entered. Sally was thirsty so John hurried to buy drinks before the movie started. John and Sally found two good seats near the back. John sat down and raised the arm rest so that he and Sally could snuggle. John paid more attention to Sally while the movie rolled and nervously sipped his drink. Finally working up the courage to do so, John extended his arm to embrace Sally. He was relieved and ecstatic to feel her move closer to him in response. Sally stood up to use the restroom during the movie, smiling coyly at John before that exit.

Source: Mark Riedl, Scheherazade, Entertainment Intelligence Lab (http://eilab.gatech.edu/)

http://www.bbc.com/future/story/20180829-the-worlds-most-prolific-writer-is-a-chinese-algorithm

$$a: \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad s: \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} ; \quad \dots \dots$$

$$\{x_t\},\ t = 1, \dots, T$$

$$x_t \in \mathbb{B}^{36}, \quad 36 = M, \quad x_{it} \in \{0, 1\}$$

$$\sum_{i=1}^{M} x_{it} = 1$$

Outline

Statistical foundations

$x_t = F_\theta(x_{t-1} \ldots x_{t-k})$

linear func.

- 4.11.: Examples & basic definitions/ terms
- 11.11.: Linear time series models (ARMA)
- 18.11.: Statistical inference in ARMA models/ multivariate AR models, Granger causality
- 25.11.: Non-Gaussian AR models for count & point processes

Outline

$x_t = \cancel{F_\theta(x_{t-1}, \ldots)}$ $g(z_t)$

$z_t = F_\theta(z_{t-1})$

linear

Statistical foundations

- 4.11.: Examples & basic definitions/ terms
- 11.11.: Linear time series models (ARMA)
- 18.11.: Statistical inference in ARMA models/ multivariate AR models, Granger causality
- 25.11.: Non-Gaussian AR models for count & point processes

Linear Dynamical Systems

- 2.12.: Linear dynamical systems & latent variable models
- 9.12.: Inferring unknown states in LDS (EM, Kalman filter-smoother)
- 16.12.: Inferring parameters in LDS; state space models for count & point processes

Outline

Recurrent Neural Networks

nonlinear

- 13.1.: Intro to discrete-time nonlinear dynamical systems
- 20.1.: Recurrent neural networks (RNN), training by grad descent (BPTT, RTRL) & 2nd order methods $x_t = F_\theta(x_{t-1}, h_t)$
- 27.1.: Capturing long-term temporal dependencies, vanishing/ exploding grad problem, deep RNN: LSTM/ GRU/ regularized PLRNN
- 3.2.: Generative (deep) RNNs: Statistical inference by extended & unscented KF, numerical sampling (particle filters), Laplace approx. x_t $z_t = F_\theta(z_{t-1})$
- 10.2.: Variational inference in generative RNN, sequential Variational Autoencoder RNN
- 17.2.: Stochastic Gradient Variational Bayes (SGVB), structured variational densities; Generative Adversarial Networks
- 24.2.: Attention, Self-Attention, & Transformer Networks

Literature

- Chatfield, C.: The Analysis of Time Series: An Introduction, Sixth Edition, Chapman and Hall/CRC (2003)
- Shumway, R.H., Stoffer, D.S.: Time Series Analysis and Its Applications: With R Examples, Springer, Heidelberg (2011) x_t
- Durbin, J., Koopman, S.J.: Time Series Analysis by State Space Methods (Oxford Statistical Science) (2012)
- Strogatz, S.H.: Nonlinear dynamics and chaos. Addison-Wesley Publ. (1994) z_t
- Bishop, C.: Pattern Recognition & Machine Learning (2006), [Ch. 10, 11, 13]
- Goodfellow, Bengio, Courville (2016) Deep Learning. MIT. [Ch. 10, 19, (20)]
- Durstewitz, D.: Advanced Data Analysis in Neuroscience – Integrating Statistical & Computational Models. Springer (2017) [Ch. 7-9]