Real-time Tracking of Neuronal Network Structure using Data Assimilation

Franz Hamilton,[1] Tyrus Berry,[2] Nathalia Peixoto,[1] and Timothy Sauer[2]

[1]*Electrical and Computer Engineering, George Mason University, Fairfax, VA 22030*
[2]*Mathematical Sciences, George Mason University, Fairfax, VA 22030*

(Dated: May 17, 2013)

A nonlinear data assimilation technique is applied to determine and track effective connections between ensembles of cultured spinal cord neurons measured with multielectrode arrays. The method is statistical, depending only on confidence intervals, and requiring no form of arbitrary thresholding. In addition, the method updates connection strengths sequentially, enabling real-time tracking of nonstationary networks. The ensemble Kalman filter is used with a generic spiking neuron model to estimate connection strengths as well as other system parameters to deal with model mismatch. The method is validated on noisy synthetic data from Hodgkin-Huxley model neurons before being used to find network connections in the neural culture recordings.

PACS numbers: 87.19.lj, 07.05.Mh

A dynamic of particular interest in neuronal networks is the connectivity structure. Connectivity changes as a function of network development and in response to applied electrical or pharmacological stimuli. Tracking these changes is essential for understanding the underlying dynamical evolution of the network, and serves as a valuable tool for experimental interventions.

A method for tracking connectivity should ideally meet several criteria. First, it should be statistical in nature. A strictly statistically-based method will require only a pre-determined confidence level, removing the need to provide an arbitrary decision threshold or decide how to make conclusions from a ROC curve. Second, the method should be robust to error. For our purposes, we are interested in two main types: error caused by system noise, and error caused by a mismatch in dynamics between acquired data and the model used to represent the data. Third, the method should have sequential, or real-time, implementation capability. Real-time analysis of a network's structure is particularly important in experimental scenarios where feedback is critical for determining or controlling the direction of the experiment. This letter describes a new method for tracking connectivity in neuronal networks that meets these three criteria.

Methods for detecting network links from complex time series have been the focus of several recent studies. The analysis of interactions between nonlinear processes was pursued in [1, 2], and ideas from compressed sensing were introduced in [3]. The concept of Granger causality [4] has been exploited by calculating partial directed coherence in [5].

Parallel to these developments are a wide range of connectivity methods for spike train measurements. Higher moment methods such as coherence, joint densities and cumulant spectra were pursued in [6–8]. An information theory approach was proposed in [9]. Maximum likelihood methods for spike train neuronal interaction were investigated in [10–13]. These methods do not meet the first criteria of being statistically based.

The Cox method, introduced for neural networks in [14] and refined in [15], was the first statistical test for connectivity in networks analyzed by spike trains. The method was expanded in [16] to include a test for changes in connectivity. The Cox method [17] is based on a proportional hazard model and uses multiple spike trains to compute a maximal likelihood connectivity parameter β_{ik} along with a confidence interval. The confidence interval is then used to test for a connection against the null hypothesis $H_0 : \beta_{ik} = 0$. It was shown in [15, 16] that to be empirically useful, a test for connectivity in large networks should produce confidence intervals that can be analyzed via multiple hypothesis testing. Additionally, the Cox method is designed to be independent of the individual neuron models, making the test very robust.

While the Cox method satifies the first two criteria of being statistical and robust to error, it fails to meet the third. Data requirements were found in [16] to be at least 10^3 spikes per neuron, and prior to the conclusion of processing, the method provides no estimate of the network connectivity structure. Furthermore, if new data is acquired then the entire data set must be re-processed before it can be included in the estimate, making use of the Cox method more suitable for offline analysis.

In the following we report on use of a parameter-augmented version of the ensemble Kalman filter (EnKF) to estimate connection strengths of a network of cultured neurons sampled by a multielectrode array (MEA). The Kalman filter [18] is a natural choice for this problem, because in addition to estimating the connection parameters it will estimate their uncertainties, which will allow a statistical test for connectivity to be developed. We will show that the method exhibits a robustness to error much like the Cox method, but offers a significant advantage in terms of application for real-time analysis.

The *in vitro* neuronal network we study consists of mammalian spinal cord neurons plated onto microelectrode arrays. An MEA simultaneously records the extracellular potential of spiking neurons plated near the array's electrodes. The neurons were extracted from embryonic day 18 (E18) mice, mechanically dissociated, and plated onto 64-channel MEAs at a density of about 200,000 cells per array. The connectivity strength between neural assemblies will be represented by a parameter vector $\boldsymbol{\beta}$, where β_{ik} indicates the connection parameter from neuron k to neuron i. Our goal is to continuously estimate these β_{ik} parameters and determine which of the estimates are statistically significant, resulting in an identified network link.

For assimilation purposes, we posit a general nonlinear system with an n-dimensional state vector x and m-dimensional observation vector y evolving according to

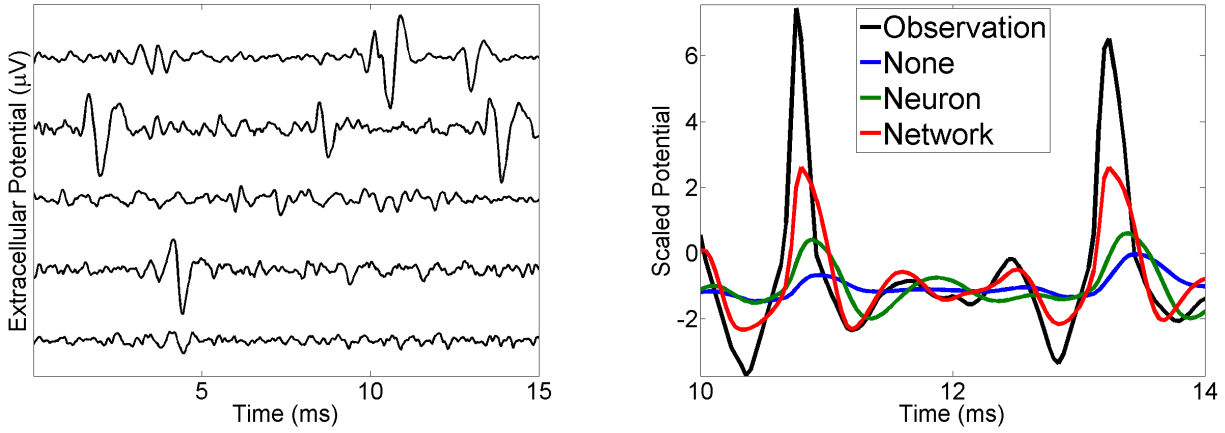
$$\begin{aligned} x_{k+1} &= f(x_k, t_k) + w_k \\ y_k &= h(x_k, t_k) + v_k \end{aligned} \tag{1}$$

where w_k and v_k are Gaussian noise terms with covariance matrices Q and R respectively.

An ensemble Kalman filter (EnKF) approximates a nonlinear system using a finite weighted ensemble [19]. We initialize the filter with state vector $x_0^+ = 0_{n\times 1}$ and covariance matrix $P_0^+ = I_{n\times n}$. At the k-th step of the filter we assume we are given an estimate of the state x_{k-1}^+ and a covariance matrix P_{k-1}^+. We use the singular value decomposition to find the unique symmetric positive definite square root S_{k-1}^+ of the matrix P_{k-1}^+. Using the algorithm of scaled unscented transformation (see for example [20]) a weighted ensemble of state vectors is generated. This ensemble is then propagated forward a single time step using the model f and observed with h. The mean of the resulting state ensemble is the a priori estimate x_k^- and the mean of the observed ensemble is the predicted observation y_k^-. Denoting the covariance matrices P_k^- and P_k^y of the resulting state and observed ensemble, and the cross-covariance matrix P_k^{xy} between the state and observed ensembles, the equations

$$\begin{aligned} K_k &= P_k^{xy}(P_k^y)^{-1} \\ P_k^+ &= P_k^- - P_k^{xy}(P_k^y)^{-1}P_k^{yx} \\ x_k^+ &= x_k^- + K_k\left(y_k - y_k^-\right). \end{aligned} \tag{2}$$

update the state and covariance estimates using the observation y_k.



(a) (b)

FIG. 1: (a) Recorded extracellular potential *in vitro* (five traces shown). Sample neural spike waveforms from the MEA differ in shape from generic neuron models, which simulate intracellular currents. (b) Three assimilation results. Observation consists of preprocessed (see text) extracellular waveform compared with its intracellular model counterpart. When no parameters of the model are fit (marked None) the observation cannot track. By fitting individual neuron parameters (marked Neuron) more of the observation is tracked, with a short delay. When the full network parameters are fit in addition to the neuron parameters (marked Network), the observation is fairly effectively tracked, showing the importance of including the network structure in assimilation.

For our purposes, f in (1) is chosen to be a generic spiking model. We used the Hindmarsh-Rose intracellular model

[21] parametrized as in [22]:

$$\begin{aligned}\dot{V}_i &= a_i V_i^2 - V_i^3 - y_i - z_i + \sum_{k\neq i}^{n} \Gamma_{\mathrm{HR}}(V_k) V_k + \sigma \dot{W}_{Vi} \\ \dot{y}_i &= (a_i + \alpha_i) V_i^2 - y_i + \sigma \dot{W}_{yi} \\ \dot{z}_i &= \mu_i (b_i V_i + c_i - z_i) + \sigma \dot{W}_{zi}\end{aligned} \tag{3}$$

for $i = 1 \ldots n$, and where $\Gamma_{\mathrm{HR}}(V_k) = \beta_{ik}/(1 + 9e^{-10V_k})$ is a gating function that regulates the amount of input between neurons. Here V is the membrane variable, y (resp., z) represents the fast (resp., slow) channel dynamics, I is the input current and a, b, c, α, μ are parameters whose values determine the different spiking behaviors of the neuron. The parameters β_{ik} represent the connectivity parameter from neuron k to neuron i, and $\dot{W}$ represents standard white noise. The observable is the membrane voltage $h(V, y, z) = V$.

Since the recorded extracellular potential waveform (see Fig. 1) differs from the Hindmarsh-Rose waveform, preprocessing is required. We connected the recorded potential to its intracellular counterpart through the relationship $dI/dt \approx E$ where E is the recorded extracellular potential and I is the intracellular potential [23]. The integration required was done using a simple moving average algorithm. The resulting approximation was then scaled as necessary to ensure stability when assimilated to our general model.

As discussed in [24–27], we can treat the l-dimensional parameter vector p as extra states of the model, with trivial dynamics. The augmented state vector consisting of the original n model states and l parameters is estimated by the EnKF. Fig. 1(b) compares the reconstruction of the observable by the EnKF in three senarios: when (1) no parameters, (2) only the model parameters, and (3) the model plus network parameters $\boldsymbol{\beta}$ are used in the assimilation.

Using the ensemble Kalman filter produces an estimate of the error of the parameter estimate, given by the diagonal entry of the covariance matrix P_k^+ corresponding to the parameter in question. Let $\beta = (x_k^+)_i$ be an estimated parameter in the state vector and $\sigma^2 = (P_k^+)_{ii}$ its variance. The standard confidence interval for β with confidence level γ_0 is $\left[\beta - \kappa_{(1-\gamma_0)/2}\sigma, \beta + \kappa_{(1-\gamma_0)/2}\sigma\right]$ where $\kappa_{(1-\gamma_0)/2}$ is the $(1-\gamma_0)/2$ quantile of the normal distribution. Thus, if β is a connectivity parameter we can apply a Wald test of $\beta = 0$ by asking whether 0 lies in the above confidence interval.

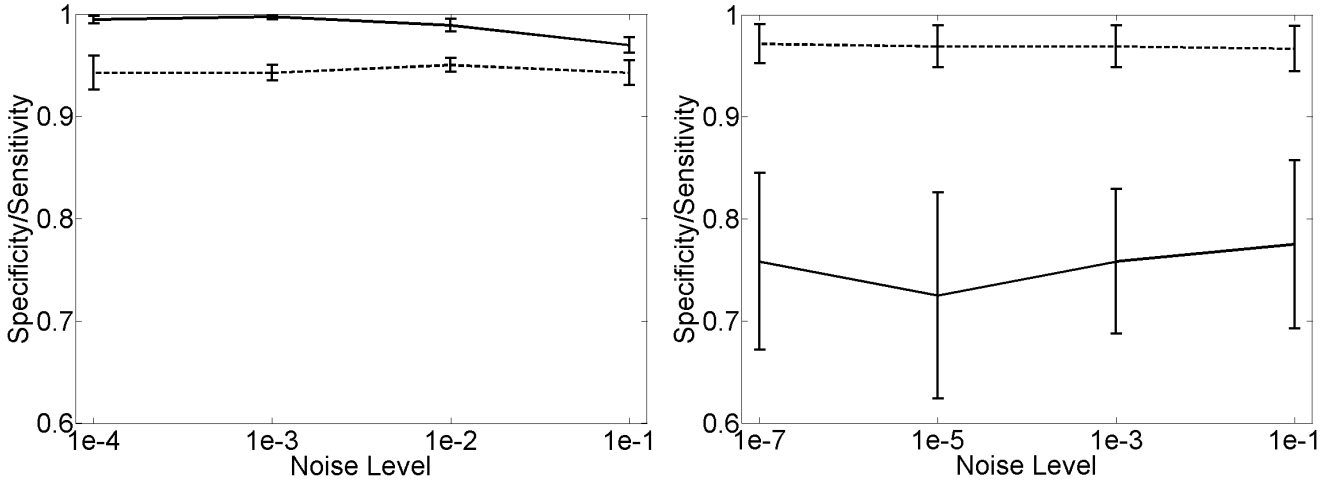


FIG. 2: Specificity (dotted line) and sensitivity (solid line) results of our method's identification of network links at the 95% confidence level when model error caused by noise is considered. Networks consist of (a) 10 Hindmarsh-Rose neurons (b) 5 Hodgkin-Huxley neurons with 40% network connectivity. In both cases, Hindmarsh-Rose was the EnKF network model. Error bars represent standard error over 10 network realizations of length 8 sec. Identified links are those whose estimated β connectivity parameter is greater than two times their filter-estimated standard deviation.

Figure 2(a) shows that when data is fit with the correct model, the specificity and sensitivity results are correct at the 95% confidence level, independent of system noise. Here we used the Hindmarsh-Rose equations (3) both to generate data and as f in the EnKF equations.

To validate the method for use with real data, we test significant model error, caused by a mismatch in dynamics. Consider the Hodgkin-Huxley model [28] with original parameters as implemented by [29], expanded to a network of

n equations, and with gating function $\Gamma_{\rm HH}(V_k) = \beta_{ik}/(1+e^{-10(V_k+40)})$ connecting neurons as in (3). The observables are the voltage variables V_i perturbed with noise, namely $V_i + \sigma\dot{W}_i$. We assimilate this data to model (3) and use the EnKF method to estimate the connectivity parameters.

The identification of network links using data from a significantly different neuron model, Hodgkin-Huxley, presents a new challenge. To handle the increased model error, we designate certain of the model parameters as free states while fixing others. We fix $b = 9, c = 5, \mu = .001$ and leave a, α, I as free states to be estimated along with the model variables. Additionally, we have to carefully tune Q in the EnKF algorithm to prevent the covariance matrix from becoming degenerate and causing filter divergence.

In an attempt to develop a fair test for our method, we simulate random networks of five noisy Hodgkin-Huxley neurons where each connected pair has a bi-directional relationship. That is, for every β_{ik} there is a β_{ki} in the network. Observational noise is used on the horizontal axis rather than system noise since the discrepancy in dynamics between the Hodgkin-Huxley data and our neuron model is already so large that the addition of system noise has no real effect.

Fig. 2(b) shows that the EnKF method is successful in determining the neurons that are connected to one another, identifying that a statistically significant connection exists between a connected pair. As the amount of observational noise increases, our method maintains specificity around 95% showing that our results are statistically significant. Sensitivity persists around 70% showing that our method is finding the majority of the network links.

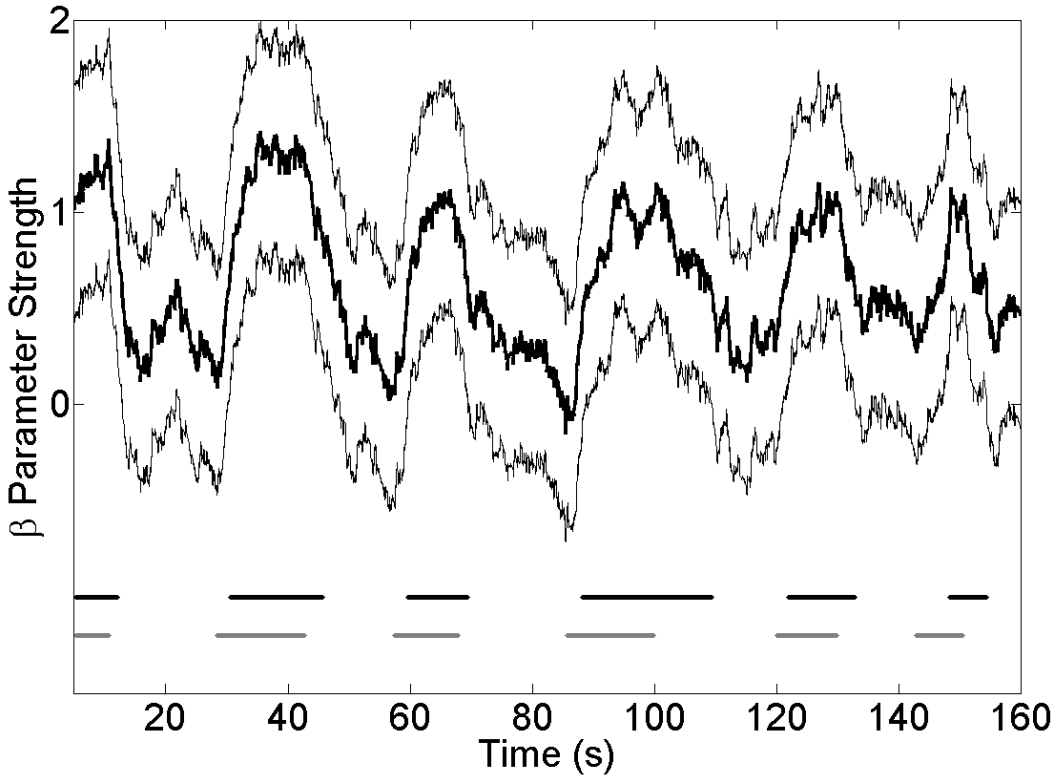


FIG. 3: Tracking results of a time-varying connection in a five neuron network of Hodgkin-Huxley neurons over 160 seconds of simulation data. Connection turned on/off randomly. Bottom part of figure shows the actual "on" time (grey bar) and the filter-estimated "on" time of the connection (black bar). Top part of figure shows the filter estimated connection strength and corresponding 99.5% confidence region.

A significant application of this method is to detect the appearance and disappearance of connections in non-stationary networks, such as neural cultures. To verify the methodology for this purpose, in the Hodgkin-Huxley model-mismatch scenario, we turned a particular β_{ij} on and off randomly, and asked the method to track the changes. We established the following protocol for determining the on/off transitions of the connection. When the connection is off, the protocol says to declare a transition from "off" to "on" when the mean estimated connection strength increases beyond the minimum (since the last "on" transition) of the confidence interval maximum; and vice versa for "off". Fig. 3 shows the ability of the method to detect these changes with minimal delay.

Finally, to track network structure, it is important to ensure that the network as a whole, rather than an individual

neuron, is correct at a given statistical level. To properly account for the repeated testing of different nodes in a network, we control the Familywise Error Rate (FWER) with the Bonferroni correction, which corrects the error rate by dividing by the number of tests. If the desired FWER for a set of n tests is γ the individual tests should be conducted at confidence level $\gamma_0 = 1 - (1 - \gamma)/n$. For example, in the case of a 10 neuron network, to be 95% confident that the entire network was correct, each connection is tested at a confidence level of 99.94%.

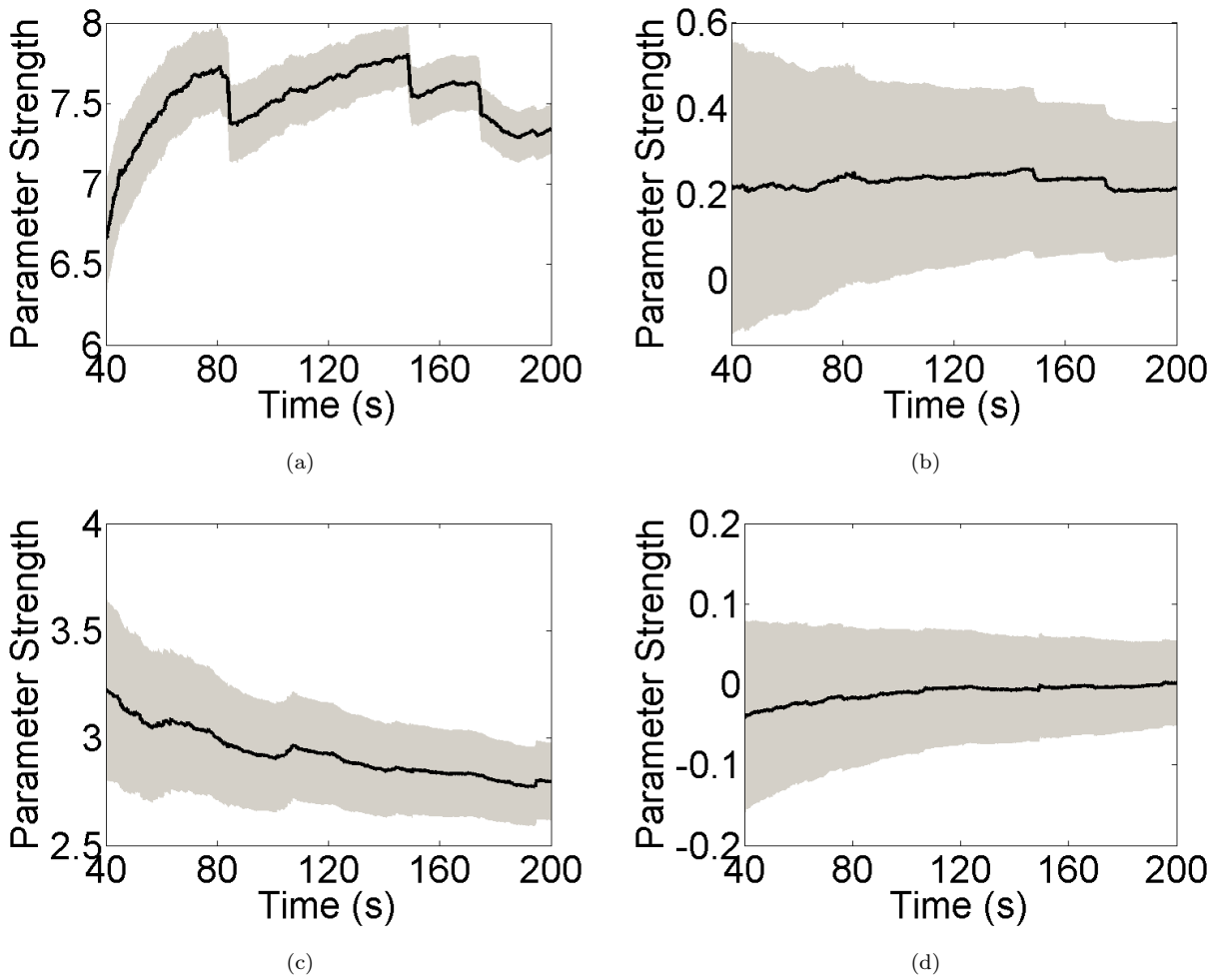


FIG. 4: In vitro β_{ik} parameters estimated over the last 160 seconds of data. Solid black line indicates the estimated connection strength and shaded grey area denotes 99.94% confidence region. A β_{ik} parameter is identified as a statistically significant connection when its confidence region does not include 0. (a) β_{41} (b) β_{43} (c) β_{53} (d) β_{15}. Panels (a) and (c) show statistically significant connections for the entire 160 seconds. Panel (b) is initially not significant, but at around 120 seconds it becomes a identified connection. Panel (d) shows a non-connection over the entire 160 seconds.

Fig. 4 shows the method's continuous tracking of β_{ik} connectivity parameters with the appropriate 99.75% shaded confidence region over 160 seconds of recorded data. Panels (a) and (c) show identified connections within the confidence level, while panel (d) shows a clear non-connection. Panel (b) shows a difficult case, As the filter gains more information about the network, it is stating that the connection exists, at least within the confidence level used.

A method that can carry out sequential tracking of neuronal network links has a major advantage for experimental purposes. Using an EnKF with a general neuron model is a way to provide a statistically driven estimate of network connectivity that is updated continuously as data arrives, unlike offline methods such as [14–16]. This allows application to network structures with moderate nonstationarity, a typical feature of neural ensembles and biological networks in general. As shown above, this algorithm for detecting connectivity meets our criteria of being statistically based, robust to error and capable of real-time implementation.

The method could serve as a powerful tool in pharmacological or stimulation studies where the effects of an applied drug or pulse pattern on connectivity was to be determined. The ability to continuously track connectivity is designed to provide real-time feedback in an experimental situation, allowing an researcher to adjust observation, control or

other interventions as necessary.

[1] R. Dahlhaus, M. Eichler, and J. Sandkhler, Journal of Neuroscience Methods **77**, 93 (1997), ISSN 0165-0270.
[2] M. G. Rosenblum and A. S. Pikovsky, Physical Review E **64**, 045202 (2001).
[3] D. Napoletani and T. D. Sauer, Physical Review E **77**, 026103 (2008), ISSN 1539-3755, 1550-2376.
[4] C. W. J. Granger, Econometrica **37**, 424 (1969), ISSN 0012-9682.
[5] L. Sommerlade, M. Thiel, B. Platt, A. Plano, G. Riedel, C. Grebogi, J. Timmer, and B. Schelter, Journal of Neuroscience Methods **203**, 173 (2012), ISSN 0165-0270.
[6] D. R. Brillinger, The Annals of Probability **3**, 909 (1975), ISSN 0091-1798.
[7] A. Aertsen and G. Gerstein, Brain Research **340**, 341 (1984).
[8] A. M. Aertsen, G. L. Gerstein, M. K. Habib, and G. Palm, Journal of Neurophysiology **61**, 900 (1989).
[9] M. Garofalo, T. Nieus, P. Massobrio, and S. Martinoia, PLoS ONE **4**, e6482 (2009).
[10] E. S. Chornoboy, L. P. Schramm, and A. F. Karr, Biological Cybernetics **59**, 265 (1988), ISSN 0340-1200, 1432-0770.
[11] M. Okatan, M. A. Wilson, and E. N. Brown, Neural Computation **17**, 1927 (2005), ISSN 0899-7667.
[12] I. Stevenson, J. Rebesco, L. Miller, and K. Krding, Current Opinion in Neurobiology **17** (2008).
[13] I. H. Stevenson, J. M. Rebesco, N. G. Hatsopoulos, Z. Haga, L. E. Miller, and K. P. Kording, IEEE Transactions on Neural Systems and Rehabilitation Engineering **17**, 203 (2009), ISSN 1534-4320.
[14] G. N. Borisyuk, R. M. Borisyuk, A. B. Kirillov, E. I. Kovalenko, and V. I. Kryukov, Biological Cybernetics **52**, 301 (1985), ISSN 0340-1200, 1432-0770.
[15] M. S. Masud and R. Borisyuk, Journal of Neuroscience Methods **196**, 201 (2011), ISSN 0165-0270.
[16] T. Berry, F. Hamilton, N. Peixoto, and T. Sauer, J. Neurosci. Meth. **209**, 388 (2012).
[17] D. R. Cox, Stochastic Point Processes pp. 55–56 (1972).
[18] R. Kalman, J. Basic Eng. **82**, 35 (1960).
[19] E. Kalnay, *Atmospheric modeling, data assimilation, and predictability* (Cambridge Univ. Press, 2003).
[20] D. Simon, *Optimal State Estimation: Kalman, H_∞, and Nonlinear Approaches* (John Wiley and Sons, 2006).
[21] J. Hindmarsh and R. Rose, Proc. Roy. Soc. **221**, 87 (1984).
[22] I. Belykh, E. d. Lange, and M. Hasler, Physical Review Letters **94**, 188101 (2005).
[23] D. Johnston and S. Miao-Sin Wu, *Foundations of cellular neurophysiology* (Bradford Books, 1995).
[24] H. Voss, J. Timmer, and J. Kurths, Int. J. Bif. Chaos **14**, 1905 (2002).
[25] C. Trudinger, M. Raupach, P. , Rayner, and I. Enting, Environmetrics **19**, 849 (2008).
[26] G. Ullah and S. Schiff, Phys. Rev. E **79**, 040901 (2009).
[27] G. Ullah and S. Schiff, PLoS Computational Biology **6**, e1000776 (2010).
[28] A. Hodgkin and A. Huxley, J. Physiology **117**, 500 (1952).
[29] H. Hasegawa, Physical Review E **61**, 718 (2000).