Statistics and Its Interface Volume 16 (2023) 227–257

Smooth online parameter estimation for time varying VAR models with application to rat local field potential activity data

Anass El Yaagoubi Bourakna*, Marco Pinto, Norbert Fortin, and Hernando Ombao*

Multivariate time series data appear often as realizations of non-stationary processes where the covariance matrix or spectral matrix smoothly evolve over time. Most of the current approaches estimate the time-varying spectral properties only retrospectively – that is, after the entire data has been observed. Retrospective estimation is a major limitation in many adaptive control applications where it is important to estimate these properties and detect changes in the system as they happen in real-time. To overcome this limitation, we develop an online estimation procedure that gives a real-time update of the time-varying parameters as new observations arrive. One approach to modeling non-stationary time series is to fit time-varying vector autoregressive models (tv-VAR). However, one major obstacle in online estimation of such models is the computational cost due to the high-dimensionality of the parameters. Existing methods such as the Kalman filter or local least squares are feasible. However, they are not always suitable because they provide noisy estimates and can become prohibitively costly as the dimension of the time series increases. In our brain signal application, it is critical to develop a robust method that can estimate, in real-time, the properties of the underlying stochastic process, in particular, the spectral brain connectivity measures. For these reasons we propose a new smooth online parameter estimation approach (SOPE) that has the ability to control for the smoothness of the estimates with a reasonable computational complexity. Consequently, the models are fit in real-time even for high dimensional time series. We demonstrate that our proposed SOPE approach is as good as the Kalman filter in terms of mean-squared error for small dimensions. However, unlike the Kalman filter, the SOPE has lower computational cost and hence scalable for higher dimensions. Finally, we apply the SOPE method to local field potential activity data from the hippocampus of a rat performing an odor sequence memory task. As demonstrated in the video, the proposed SOPE method is able to capture the dynamics of the connectivity as the rat samples the different odor stimuli.

*Corresponding author.

Keywords and phrases: Online parameter estimation, Locally stationary processes, Dynamic brain connectivity, Dynamic spectral connectivity.

1. INTRODUCTION

Many time series data recorded from various research areas including finance, econometrics, biology, atmospheric sciences and neuroscience, exhibit nonstationarity. However, these time series data may have second moment structures (variance, cross-covariance, spectral matrix) that may evolve over the duration of the observation period. In this paper, we examine local field potential (LFP) signals recorded from the hippocampus of a rat (see Figure 1 and 2), which display changes in the second moment as the rat is presented with sequences of odors during the experiment. Our goal here is to develop a computationally efficient approach for estimating rat brain functional connectivity that can track these changes in real-time.

There are a number of available methods for analyzing non-stationary time series data. One approach is to model

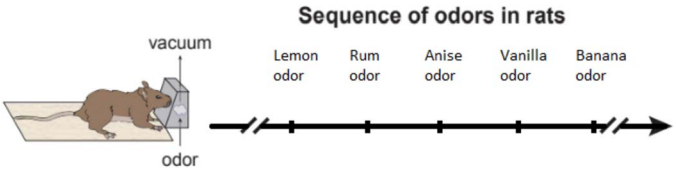

Figure 1. Odor sequence memory task.

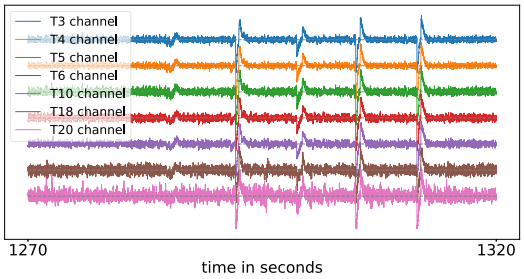

Figure 2. LFP recordings from seven tetrodes of the rat.

the data as a realization of some "locally stationary" process which essentially assumes that the second moment structure (i.e., the spectral matrix) is approximately constant within a narrow time interval. This idea was introduced in Priestley (1965) and then reformulated in Dahlhaus (1997) with a framework that admits a sequence of consistent estimators for the time-varying spectral matrix. These two approaches use the Fourier complex exponentials as the stochastic building blocks for representing non-stationary signals. However, there are other building blocks for signal representation In Ombao, Sachs and Guo (2005) and Nason, Von Sachs and Kroisandt (2000) the representations use non-decimated wavelets and SLEX (smooth localized complex exponentials) as building blocks. Thus, under these stochastic representations, the time-varying auto-spectra and time-varying cross-coherence are rigorously defined.

Another approach to modeling non-stationary signals is to use time-domain representations with time-varying coefficients. Vector autoregressive (VAR) models are popular because they are a natural extension of the autoregressive model to the multivariate setting and thus can capture the cross-dependence between different components. Under the VAR model, the conditional mean is explicitly expressed as a linear combination of its own past as well as the past values of the other time series. Thus, the VAR model provides a direct mechanism for forecasting future values of one time series based on its own past values as well as the past value of others, see Sims (1980) and Lütkepohl (1991). In general, VAR models are extensively used in numerous fields such as economics, weather forecasting and brain imaging, see Brockwell and Davis (1986), Shumway and Stoffer (2005), Gorrostieta et al. (2012), Yu et al. (2016), Samdin et al. (2016), Hu et al. (2020).

Some of the main qualities that make VAR models attractive to practitioners are mainly their ability to capture contemporaneous and lagged linear relationships between different components of a time series and also their ability to provide analytical definitions of connectivity based on the spectral density information, see Granger (1969), Baccala and Sameshima (2001), Wang, Ting and Ombao (2016), Ting et al. (2018). For a technical review of the VAR models see Lütkepohl (1991). There are some approaches that allow offline estimation of the parameters of high-dimensional time series under the VAR model. For example the use of regularization techniques, such as the l_1 regularization (LASSO-type) can ensure sparsity in high-dimensional settings (see Tibshirani (1996), Basu and Michailidis (2015) and Hu, Fortin and Ombao (2019)). Other examples leverage the low-rank property of the transition matrix of the VAR model to gain insight on the network of Granger causal interactions between time series components (see Basu, Li and Michailidis (2019) and Ting et al. (2018)).

Over the last two decades, several methods have been proposed for analyzing non-stationary signals including, Nason, Von Sachs and Kroisandt (2000), Ombao, Sachs and Guo (2005), Dahlhaus and Subba Rao (2006), Dahlhaus and Subba Rao (2007), Cole et al. (2010), Park, Eckley and Ombao (2014), Samdin et al. (2016), Bruce et al. (2018), Ting et al. (2018), Shappell et al. (2019), Li et al. (2019), Zhao and Prado (2020) and Li et al. (2021). However, none of these methods can track changes in local field potentials in real-time. This is a significant limitation since these methods are not able to provide immediate feedback to the experimenter about, for instance, whether or not the stimulus type and intensity are eliciting the anticipated brain response. Hence, the adequacy of the experimental settings can be verified only retrospectively, that is, after the data acquisition process has been completed. Here, we will develop a procedure that has the ability to provide immediate feedback regarding the connectivity. Thus, the experimenter would be able to adapt the stimulus type or intensity in real-time during the experiment, rather than after the fact. This opportunity for real-time manipulation is critical to investigate the dynamic nature of information processing in the brain. Therefore, we envision two main benefits to the development of robust tools for online estimation of spectral power and coherence in electrode arrays. First, these tools could have an immediate impact on neuroscience by improving the data collection process. This is particularly important for technically challenging and time-consuming projects, like the rat LFP experiment described here, for which data collection on a given day represents the culmination of months of work (i.e., training the animal, building and surgically implanting the microdrive, and slowly driving the electrodes to the target location). In this case, accurate online estimation of LFP power and coherence could help improve electrode positioning and determine whether the neural response to specific stimuli or conditions is of sufficient effect size to warrant beginning data collection. Second, these tools are also critical to the ongoing development of the next generation of neuroscience experiments, including closed-loop approaches in which experimental manipulations are performed based on the detection of specific activity patterns in real-time. In this case, rapid and robust online estimation of LFP power and coherence is critical to quickly identify key neural states and trigger the corresponding experimental intervention within the necessary time window. These interventions can range from to the manipulation of experimental parameters, such as the precise timing and intensity of stimuli, to the manipulation of neural circuits through electrical or optogenetic stimulation, to measure the effect of disrupting upstream regions on local information processing.

Our goal in this paper is to develop a procedure that is able to describe these changes of connectivity in local field potentials in real-time. These non-stationary local field potentials will be modeled using time-varying vector autoregressive (tv-VAR) processes. Under the tv-VAR framework,

we will develop an online procedure for estimating the parameters. Consequently, the procedure will also be able to estimate any functional of these parameters including cross-coherence, partial cross-coherence and partial directed coherence. In Pagnotta and Plomp (2018), it have been shown that the tv-VAR model has the ability to capture the dynamics in the multivariate brain signals. However, this comes at a heavy computational cost due to the high dimensionality of the parameter space (which is quadratic in the dimension of the multivariate time series). In this paper we present a novel smooth online parameter estimation (SOPE) procedure using the penalized least squares criterion. Compared to the Kalman filter, which is the industry standard, the proposed SOPE method has the advantage of being computationally very efficient and that it has a Bayesian interpretation. In Section 2 we describe the model, introduce the idea behind smooth online parameter estimation (SOPE), and discuss the infill behavior of our estimator as a function of the penalization parameters. In Section 3 we present a brief overview of some classical brain connectivity measures that are derived from the tv-VAR model. In Section 4 we show that SOPE provides similar performance as the Kalman filter, in terms of mean square error (MSE) using simulated data. However, the main advantage of SOPE is the significantly shorter computational time. In Section 5 we apply SOPE to LFP data taken from the hippocampus of a rat performing a hippocampus-dependent sequence memory task and the proposed SOPE method shows results consistent with the literature. Finally, in Section 6 we present a brief discussion of our results.

2. SMOOTH ONLINE PARAMETER ESTIMATION (SOPE)

2.1 The tv-VAR model

Let $X(t) = [X_1(t), ..., X_P(t)]'$ be a P dimensional time series collected from a network with P nodes. Here, $X(t)$ could represent the observed brain signal (e.g., electroencephalogram or local field potential) over P locations (electrodes) on the scalp, over time points $t = 1, 2, \ldots, T$. Thus, the tv-VAR model of order K, denoted tv-VAR(K), is defined as follows:

$$X(t) = \sum_{\ell=1}^{K} \Phi_{t,\ell} X(t-\ell) + E(t), \tag{1}$$

where the set $\{\Phi_{t,\ell}\}_{\ell=1}^{K}$ represents the tv-VAR parameters at time t (K, $P \times P$ matrices); $E(t)$ is Gaussian white noise term with zero mean and variance Σ_E. In the following we will denote the concatenated unknown time varying parameter matrices by $\Phi(t)$, and we will denote the concatenated previous K observations at time t by $U(t)$:

$$\Phi(t) = [\Phi_{t,1}, \ldots, \Phi_{t,K}], \tag{2}$$

$$U(t) = [X(t-1)', \ldots, X(t-K)']', \tag{3}$$

Now the model definition simplifies to

$$X(t) = \Phi(t)U(t) + E(t) \tag{4}$$

and the parameter $\Phi(t)$ will be estimated in real-time.

2.2 Recursive least squares estimators

The least squares method is a very powerful and fundamental estimation procedure in statistics which dates back to the work of Carl Friedrich Gauss and Adrien-Marie Legendre, see Stigler (1981). For almost 150 years since its inception, the least squares estimation procedure did not attract much attention until the early 1940s when Kolmogorov (1939) and Wiener (1949) independently introduced the field of linear filtering from the stochastic processes point of view. It was only near a decade later that the efficient matrix formulation of the recursive least squares (RLS) was proposed in Plackett (1950). In Kalman (1960), the following major development to the field of linear filtering was proposed using the state space formulation of the problem. More details on the historical development of the RLS/linear filtering can be found in Sorenson (1970) and Young (2011). Dahlhaus and Subba Rao investigated online and recursive inference for time varying ARCH models in Dahlhaus and Subba Rao (2006) and Dahlhaus and Subba Rao (2007). However, their model can only handle a one dimensional time series, which is not suitable for multivariate time series, as it is often the case for local field potentials.

There are two main paradigms under the general RLS framework. The first is the classical RLS which corresponds to the stationary case where the parameters are static. The second is the weighted RLS which corresponds to the non-stationary case where parameters are evolving over time. Naturally, our interest lies in the second case since the brain functional signals are dynamic. There are two general estimation approaches to weighted RLS: sliding window and exponentially weighted RLS. To estimate the unknown parameters at a fixed time point, the sliding window recursive least squares (SWRLS) uses a fixed number of observations around this particular time point. By sliding the window across time, this results in time varying parameter estimates where the estimator only borrows information from time-localized observations. However, this approach suffers from high sensitivity to noise in the observations. A solution to this problem is to take a larger window (that is, include more observations), but these results in a more biased estimator due to the bias-variance trade-off. Similarly, the exponentially weighted least squares (EWRLS) uses exponentially decaying weights to estimate the time varying parameters. Therefore, this approach uses all past observations but gives a lower weight to temporally distant observations in order to limit their contribution to the present estimate. In order to control the bias-variance trade-off, a "forgetting factor" $\kappa \in (0, 1)$ is selected where values of κ are close to 0 will give more importance to recent observations and therefore

will produce estimators with lower bias but these will have higher variance. Conversely, values of κ that are close 1 will result in a more biased estimator but with a smaller variance. There is an extensive literature concerning the choice of the decaying weights for the RLS algorithms depending on the problem at hand. However, the flexibility of such algorithms is limited due to the linear nature of the estimator. The above mentioned approaches try to reduce the variance by increasing the window size but this leads to a more biased estimator. Furthermore, those approaches do not leverage potential prior information concerning the smoothness of the estimated curves. In contrast, the proposed SOPE method has the ability to produce estimates with lower variance through the roughness penalty and at the same time preserve the local structure so that the bias is controlled.

2.3 Kalman filter

The Kalman filter (KF) is also a recursive algorithm. In fact, under simplistic conditions, the RLS is a special case of the KF. To estimate $\{\Phi(t)\}$ using the KF, we need to formulate the problem under the framework of state-space models. First, vectorize the parameters (5) and define observation matrix (6):

$$a(t) = vec(\Phi(t)'), \tag{5}$$

$$C(t) = I_P \otimes vec([X'(t-1), ..., X'(t-K)]). \tag{6}$$

Using the above notation, the formal model is defined in terms of the state transition and observation model, respectively, as follows:

$$a(t) = a(t-1) + w(t), \tag{7}$$

$$X(t) = C(t)a(t) + v(t), \tag{8}$$

where $w(t)$ and $v(t)$ are the transition noise and observation noise, respectively. The dynamics of the parameters are governed by Equation 7. Under appropriate assumptions (see Kalman (1960)), the KF can deliver optimal results in terms of minimal variance of the error. However, the problem at hand is more complicated because the optimal parameters of the filter are unknown and will have to be estimated along with the tv-VAR parameters. Hence, the KF will not necessarily be optimal in the sense of minimal error variance. Indeed, the automatic selection of the filter variance matrices is possible, but that is another problem on its own. For more details on adaptive filtering see Mehra (1972), Lippuner and Moschytz (2004), Emami and Taban (2018).

There are two main limitations of the Kalman filter as a tool for practical online estimation. The primary problem is the high level of uncertainty (variance) of the parameter estimators. Consequently, since connectivity measures (e.g., coherence, partial directed coherence) are highly non-linear functions of the tv-VAR parameters, small perturbations in the tv-VAR estimates can lead to substantial changes in the connectivity estimates and hence an even higher level of uncertainty in the connectivity estimators. The other problem of the KF is its prohibitively high computational cost which requires manipulation of a covariance matrix of the state vector. Since the state vector lives in a KP^2-dimensional space, its covariance matrix will be very expensive to compute and to store in memory when the dimension P of the problem is high, which is the case with EEG and LFP data where P is in the range $20 \sim 256$.

2.4 Proposed method: SOPE

We now propose the SOPE method which leverages the local stationarity assumption to provide real-time estimates of the tv-VAR model parameters. Under local stationarity, there is some smoothness assumption on the time-varying parameters. Thus, in our approach we consider the parameters to be differentiable to some given order with respect to rescaled time.

Under Gaussianity ($E(t) \overset{iid}{\sim} \mathcal{N}(0, \Sigma_E)$), the generalized least squares approach (GLS) is equivalent to maximizing the conditional likelihood. Using the notation in Equations 2 and 3 we obtain the the following:

$$\begin{aligned} &X(t)\big|\Phi(t), U(t) \sim \mathcal{N}(\Phi(t)U(t), \Sigma_E) \\ \implies &\widehat{\Phi}(t) = \underset{b \in \mathbb{R}^{P \times KP}}{\arg\min} \big|\big|X(t) - bU(t)\big|\big|^2_{\Sigma_E^{-1}}. \end{aligned} \tag{9}$$

In this particular problem the MLE will not be an optimal choice for obvious reasons which we state here for completeness. First, the problem is ill posed since $U(t)U(t)'$ is singular, there is only have one observation $X(t)$ at time t to estimate the tv-VAR parameters $\Phi(t)$. Second, even with regularization the estimator will be biased (shrinkage to zero) and will have a high variance due to the fact that only data from a single time point is used for estimation. Therefore, it will be necessary to borrow information from neighboring time points. However, since our goal is to develop an online estimation method, this neighborhood will consist only of *past* observations.

Our proposed method is based on the framework of a locally stationary process, which implicitly assumes some smoothness of the time-varying parameters, and hence the "immediate" past observations must contain relevant information for the current observations. Moreover, one important assumption here is that, for each time t, the true physiological signal-generating process can be approximated by a VAR model with parameters $\Phi(t)$ that change smoothly with respect to (rescaled) time. Therefore, a meaningful approach is to obtain an estimator that is reasonably smooth and at the same time provides a good fit based on least squares or some objective criterion (such as the likelihood function). This can be formalized in terms of the penalized least squares criterion below:

$$\widehat{\Phi}(t) = \underset{b \in \mathbb{R}^{P \times KP}}{\arg\min} \big|\big|X(t) - bU(t)\big|\big|^2_{\Sigma_E^{-1}} + \lambda P(b), \tag{10}$$

where $\lambda > 0$ is a regularization parameter and $P(b)$ denotes a penalization term that controls the smoothness of

the estimated function $\widehat{\Phi}(t)$. In this paper, we will investigate different forms of the penalty term $P(b)$ including the Frobenius norm (note that the Frobenius norm of a matrix A is the square root of the sum of its elements squared, i.e., $||A||_F = \sqrt{tr(A'A)}$) of the first order difference with previous estimates or the second order difference or a combination of the two:

$$(11)\quad P_1(b) = \left|\left|b - \widehat{\Phi}(t-1)\right|\right|_F^2,$$

$$(12)\quad P_2(b) = \left|\left|b - 2\widehat{\Phi}(t-1) + \widehat{\Phi}(t-2)\right|\right|_F^2,$$

$$(13)\quad P_3(b) = \left|\left|b - \left[\widehat{\Phi}(t-1) + \beta\big(\widehat{\Phi}(t-1) - \widehat{\Phi}(t-2)\big)\right]\right|\right|_F^2.$$

The penalty function $P_1(b)$ in Equation 11 corresponds to the first order difference penalization; $P_2(b)$ in Equation 12 corresponds to the second order difference penalization. Clearly, $P_3(b)$ in Equation 13 is a combination of the previous two penalty functions with penalization parameters λ_1 and λ_2, where $\lambda = \lambda_1 + \lambda_2$ and $\beta = \frac{\lambda_2}{\lambda_1+\lambda_2}$. For the following derivations, we will assume that $\Sigma_E = I$. This assumption will greatly simplify computations without unnecessarily limiting the flexibility of the tv-VAR model for capturing the connectivity structure of the brain network. Note that $\Sigma_E = I$ does not imply that the components of the multivariate time series are independent. Instead, this means that the connectivity in the brain network is fully captured by $\Phi(t)$. Moreover, this assumption also allows the computation to be carried out in an online fashion that is computationally robust. The previous problems are all quadratic in b. Therefore, minimizing with respect to b leads to the following recursive formulas:

$$(14)\quad \widehat{\Phi}(t) = \big(X(t)U(t)' + \lambda\widehat{\Phi}(t-1)\big) \cdot \Big(U(t)U(t)' + \lambda I\Big)^{-1},$$

$$(15)\quad \widehat{\Phi}(t) = \Big(X(t)U(t)' + \lambda\big[\widehat{\Phi}(t-1) + \big(\widehat{\Phi}(t-1) - \widehat{\Phi}(t-2)\big)\big]\Big) \cdot \Big(U(t)U(t)' + \lambda I\Big)^{-1},$$

$$(16)\quad \widehat{\Phi}(t) = \Big(X(t)U(t)' + \lambda\big[\widehat{\Phi}(t-1) + \beta\big(\widehat{\Phi}(t-1) - \widehat{\Phi}(t-2)\big)\big]\Big) \cdot \Big(U(t)U(t)' + \lambda I\Big)^{-1}.$$

In order to invert $U(t)U(t)' + \lambda I$ efficiently, it is necessary to use the Sherman-Morrison-Woodbury inversion formula in Hager (1989).

2.5 Bayesian interpretation

The tv-VAR parameters are assumed to vary smoothly over time and therefore they must have a Taylor expansion at every time point t. Hence, $\Phi(t) \sim \Phi(t-1) + \dot{\Phi}(t-1)$, where the derivative $\dot{\Phi}(t-1)$ can be approximated by $\Phi(t-1) - \Phi(t-2)$ which leads to $\Phi(t) \sim \Phi(t-1) + \big(\Phi(t-1) - \Phi(t-2)\big)$. Depending on the information available regarding the smoothness of the estimated function and the density of the observations, i.e., the Δt between consecutive observations $X(t)$ and $X(t-1)$, this will lead to different levels of confidence in the previous approximation, which translates to $\Phi(t) \sim \Phi(t-1) + \beta\big(\Phi(t-1) - \Phi(t-2)\big)$ with β close to one when the function is smooth and when the observations density is very high and close to zero otherwise.

Thus, given estimates $\widehat{\Phi}$ of Φ at times $t-1$ and $t-2$, a potential prior for the parameters at time t would be $\mathcal{N}(\mu, \Sigma)$, with $\mu = \widehat{\Phi}(t-1) + \beta\big(\widehat{\Phi}(t-1) - \widehat{\Phi}(t-2)\big)$ and $\Sigma = \lambda^{-1}I$, here λ is the precision parameter of the prior and it is supposed to represent our confidence on the "location" of the next iterate, and I is a $KP^2 \times KP^2$ identity matrix, the smoother the function the higher the λ. Of course, moving forward, these parameters need to be vectorized for the interpretation to make sense. Under Gaussianity of the noise, this Bayesian formulation implies the following conditional posterior:

$$(17)\quad f\big(b\big|X(t), U(t)\big) \propto f\big(X(t)\big|b, U(t)\big) f\big(b\big)$$

Therefore, the point estimate $\widehat{\Phi}(t)$ of $\Phi(t)$ at time t is derived to be:

$$(18)\quad \begin{aligned}\widehat{\Phi}(t) &= \underset{b\in\mathbb{R}^{P\times KP}}{\arg\max}\ f\big(X(t)\big|b, U(t)\big) f\big(b\big|\mathcal{I}_{t-1}\big),\\ &= \underset{b\in\mathbb{R}^{P\times KP}}{\arg\min}\ \left|\left|X(t) - bU(t)\right|\right|^2_{\Sigma_E^{-1}} + \lambda P(b).\end{aligned}$$

Note that the maximization problem of the (log-)posterior in Equation 18 is equivalent to the minimization problem stated in Equation 10, with penalties defined in Equations 11, 12 and 13 corresponding to different prior choices. A choice of $\beta = 0$ is equivalent to the choice of a Gaussian prior that assumes a constant function. This is consistent with penalizing the gradient since this penalization will have as a result the vanishing of the estimate's gradient. Moreover, taking $\beta = 1$ is equivalent to having a Gaussian prior that assumes a constant gradient/linear function of time. This is consistent with penalizing the curvature since this penalization produces a vanishing second derivative. Note that it is possible to pursue higher order penalties, one could extend this reasoning indefinitely to chose the corresponding prior that assumes some degree of smoothness.

Given the penalization parameters α, β and the time series $X(t)$ the following recursive algorithm (SOPE) provides online estimates of the parameters:

Algorithm 1 Smooth Online Parameter Estimation for tv-VAR models

procedure GetSmoothEstimates($X(1), \ldots, X(T)$)
  Initialize:
  $\widehat{\Phi}(K-1) =$ Least Squares
  $\widehat{\Phi}(K) \quad = \widehat{\Phi}(K-1)$
  $\alpha \in (0, \infty)$, $\beta \in [0, \ldots, 1)$
  for $t = K+1, \ldots, T$ **do**
    $U(t) = \left[X(t-1)', \ldots, X(t-K)'\right]'$
    $A = X(t)U(t)' + \lambda\Big[\widehat{\Phi}(t-1) + \beta\big(\widehat{\Phi}(t-1) - \widehat{\Phi}(t-2)\big)\Big]$
    $B = \big(U(t)U(t)' + \lambda I\big)^{-1}$
    $\widehat{\Phi}(t) = AB$
  end for
end procedure

2.6 Infill asymptotics

Leaving the world of stationarity comes with many complications. Classical asymptotic results do not necessarily apply here, since the distant future or the distant past might not provide relevant information about the present due to the nonstationarity. Therefore, by considering locally stationary processes, one should look into infill asymptotics on the parameter $\Phi(t)$. Following Dahlhaus (2012), we analyze the behavior of our estimator using infill asymptotics by rescaling the time index of the tv-VAR parameter to the unit interval $[0, 1]$. Define $u = \frac{t}{T}$, then as $T \to \infty$, the range of u will become dense in $[0, 1]$.

Asymptotically, when $T \to \infty$, the (rescaled) time between consecutive observations becomes infinitesimal, i.e., $h = \frac{1}{T} \to dt$. Furthermore, if the penalization coefficients are selected correctly, the finite difference penalization terms will converge to the derivatives, see Appendix 2.6. Moreover, minimizing the problem globally ($\forall t \in [1, \ldots, T]$) with penalty terms as defined in Equations 11 and 12 will asymptotically result into the following calculus of variation problems:

$$\begin{cases} \mathcal{L}(u, b, \dot{b}) & = \left|\left|X(\lfloor uT \rfloor) - b(u)U(\lfloor uT \rfloor)\right|\right|_2^2 + \\ & \quad c_1\left|\left|\dot{b}(u)\right|\right|_F^2 \\ \mathcal{J}[b] & = \int_0^1 \mathcal{L}(u, b, \dot{b})du \\ \widehat{\Phi} & = \arg\min_{b \in B} \mathcal{J}[b] \end{cases}, \tag{19}$$

$$\begin{cases} \mathcal{L}(u, b, \dot{b}, \ddot{b}) & = \left|\left|X(\lfloor uT \rfloor) - b(u)U(\lfloor uT \rfloor)\right|\right|_2^2 + \\ & \quad c_2\left|\left|\ddot{b}(u)\right|\right|_F^2 \\ \mathcal{J}[b] & = \int_0^1 \mathcal{L}(u, b, \dot{b}, \ddot{b})du \\ \widehat{\Phi} & = \arg\min_{b \in B} \mathcal{J}[b] \end{cases}, \tag{20}$$

where the first equations represent the Lagrangian which is the least squares term (or the likelihood of the observations) plus a penalty term for the roughness of the function; the second equations represent the functionals to be minimized over B (class of smooth enough functions). To solve such problems, we consider the necessary condition for optimality, also known as Euler-Lagrange equations (see Appendix 2.6 for more details and Gelfand and Fomin (1964) for more background on Euler-Lagrange equations and calculus of variation in general):

$$\ddot{\widehat{\Phi}}(u) = \frac{1}{c_1}\nabla_{\widehat{\Phi}}||X(\lfloor uT \rfloor) - \widehat{\Phi}U(\lfloor uT \rfloor)||_2^2, \tag{21}$$

$$\dddot{\widehat{\Phi}}(u) = -\frac{1}{c_2}\nabla_{\widehat{\Phi}}||X(\lfloor uT \rfloor) - \widehat{\Phi}U(\lfloor uT \rfloor)||_2^2. \tag{22}$$

If the derivative is penalized with c_1 being small (asymptotically "weak" penalization) then the solution boils down to the least squares estimator. However, if c_1 is large, the second derivative vanishes which implies that the solution is affine:

$$\begin{cases} c_1 \to 0 & \implies \forall u, \nabla_{\widehat{\Phi}}||X(\lfloor uT \rfloor) - \widehat{\Phi}(u)U(\lfloor uT \rfloor)||_2^2 \to 0 \\ c_1 \to \infty & \implies \forall u, \ddot{\widehat{\Phi}}(u) \to 0, \end{cases}$$

Similarly, if we penalize the second derivative with c_2 being small (asymptotically "weak" penalization) the solution boils down to the least squares estimator. However, if c_2 is large, the third derivative vanishes which means that the solution is quadratic in time:

$$\begin{cases} c_2 \to 0 & \implies \forall u, \nabla_{\widehat{\Phi}}||X(\lfloor uT \rfloor) - \widehat{\Phi}(u)U(\lfloor uT \rfloor)||_2^2 \to 0 \\ c_2 \to \infty & \implies \forall u, \dddot{\widehat{\Phi}}(u) \to 0. \end{cases}$$

Therefore, based on the choice of the penalty paramters c_1 and c_2, the SOPE method will produce an estimator that is an intermediate between the LSE (or MLE) on one side and the best linear (or parabolic) curve on the other side depending on the exact regularization that is used.

Note that if Σ_E is unknown, it needs to be estimated online. Therefore, the previous results will not be valid anymore. However, if the covariance estimates improve as more data is observed, then previous results will hold after some burn in period that is necessary for getting a good estimate for Σ_E, see Appendix 7.1 for the generalized SOPE algorithm.

2.7 SOPE and the Kalman smoother

In addition to the previous two subsections, it is possible to make a connection between smooth online parameter estimation as described in Equation 13 and state space models that bridges the gap between the Bayesian interpretation and the asymptotic behaviour of the obtained estimator.

On the one hand the overall SOPE estimator could be summarized as the following minimization problem

$$\widehat{\Phi} = \arg\min_{b_K, \ldots, b_T} \sum_{t=K}^{T} \left|\left|X(t) - b(t)U(t)\right|\right|_2^2 + \lambda\left|\left|\ddot{b}(t)\right|\right|_F^2 \tag{23}$$

with respect to the time varying VAR parameters. Where λ controls the smoothness of the estimates, the larger the λ the

smoother the estimates. If $\lambda = 0$, the minimizer is the OLS estimator, which leads to non smooth estimates. If $\lambda = \infty$, then the only acceptable solution is one that vanishes the second derivative, which in turn leads to an estimator that is linear in time as presented in Section 2.6.

On the other hand assuming the following model

$$X(t) = \Phi(t)U(t) + v(t) \text{ and } \ddot{\Phi}(t) = w(t) \tag{24}$$

where $v(t)$ and $w(t)$ are independent white noise processes with $Var(v(t)) = \sigma_v^2 I$ and $Var(w(t)) = \sigma_w^2 I$. One can show that the model defined in 24 leads to a state space model, where minimizing the log likelihood with respect to the states is equivalent to maximizing the complete data likelihood which is equivalent to the minimization problem in Equation 23 with $\lambda = \frac{\sigma_v^2}{\sigma_w^2}$. Thus, the variance σ_w^2 controls the smoothness of the estimates, the smaller σ_w^2 the larger λ and the smoother $\widehat{\Phi}(t)$. This connection between the SOPE and the Kalman smoother is somehow very similar to the connection between smoothing splines and state space models. Instead of smoothing the data itself (i.e., obtaining $\widehat{\mu}(t)$ from $y_t = \mu(t) + \epsilon(t)$) we aim to smooth the time varying parameters. Refer to Shumway and Stoffer (2005) for more details on how to build the state space model from Equation 24 and to see more details on the connection between smoothing splines and the Kalman smoother in general.

3. ONLINE ESTIMATES OF BRAIN CONNECTIVITY MEASURES

For centuries, scientists have been interested in localizing brain functions (e.g., Temkin (1947)). In recent decades, considerable progress has been made bringing to light the anatomical and structural map of the brain. However, understanding how information is processed and integrated across brain regions is a much more difficult challenge because it involves dynamics and the notion of causality (see Friston (2011) and Lindquist (2008)).

Functional connectivity is often defined as the statistical dependence between distant populations/groups of neurons. Depending on the data modality (in particular, the sampling rate) it is usually assessed using cross-correlation (for functional magnetic resonance imaging data) or cross-coherence (for electroencephalograms or local field potentials). Over the last twenty years there have been an increasing trend in the literature related to the assessment of effective connectivity, which can be understood as the influence of a group of neurons over another, see Cribben et al. (2012), Xu and Lindquist (2015), Fiecas and Ombao (2016), Zhou et al. (2016), Ombao et al. (2018), Ting et al. (2020). Consequently, effective connectivity is a measure of causal influence/information flow between two regions. Baccala and Sameshima developed the concept of partial directed coherence (PDC) between groups of neurons which measures the direction of information flow inside the brain structure, see Saito et al. (1981) and Baccala and Sameshima (2001).

In this paper, we discuss three measures of connectivity, coherence, partial coherence and partial directed coherence. However, in our data analysis, we will only use coherence and partial directed coherence. Coherence is the relative synchrony between a pair of signals. Coherence at some pre-specified frequency band is the squared cross-correlation between a pair of filtered signals (whose power are each concentrated at the specific band). This intuitive exposition of coherence is given in Ombao and Bellegem (2008). Partial coherence between two signals can be defined as the conditional coherence when all other signals have been observed. In other words partial coherence is the remaining coherence that cannot be explained by all other signals. It is characterized via the inverse of the spectral matrix. To handle estimation problems when the spectral matrix has a poor condition number, shrinkage estimators are proposed in Fiecas et al. (2010) and Fiecas and Ombao (2011). Partial directed coherence (PDC) can be defined as the conditional Granger causality from one signal to another normalized by the total causal outflow from the first signal and is developed under the context of a vector autoregressive model. PDC was introduced in the seminal papers of Saito et al. (1981) and Baccala and Sameshima (2001), and was later used by Hu, Fortin and Ombao (2019) and Hu et al. (2020) in the context of modeling high dimensional signals.

In the following, we provide the parametric definitions of these time-varying spectral measures of connectivity in the context of a tv-VAR model. First, we define the following time varying quantities of interest at frequency ω with sampling frequency ω_s:

$$\Phi(t,\omega) = I_P - \sum_{\ell=1}^{K} \Phi_{t,\ell} \exp\left(-\sqrt{-1}2\pi\ell\omega/\omega_s\right), \tag{25}$$

$$H(t,\omega) = \Phi(t,\omega)^{-1}, \tag{26}$$

$$S(t,\omega) = H(t,\omega)\Sigma_E H(t,\omega)^*, \tag{27}$$

$S(t,\omega)$ is called the spectral matrix of the time series, $S_{i,j}(t,\omega)$ represents the cross spectrum between tetrode i and tetrode j at frequency ω and time t, $H(t,\omega)$ is the time-varying and frequency-specific transfer function matrix. Using the previous notation in Equations 25, 26 and 27 we can provide the parametric definitions of the spectral measures of dependence defined above between tetrode i and tetrode j at frequency ω and time t:

$$\rho_{i,j}^2(t,\omega) = \frac{|S_{i,j}(t,\omega)|^2}{S_{i,i}(t,\omega)S_{j,j}(t,\omega)}, \tag{28}$$

$$\pi_{i,j}(t,\omega) = \frac{\left|\Phi_{i,j}(t,\omega)\right|}{\sqrt{\sum_{k=1}^{P}\left|\Phi_{k,j}(t,\omega)\right|^2}}. \tag{29}$$

The previous Equations 28 and 29 represent respectively coherence and partial directed coherence. To estimate these

spectral connectivity measures at time t, first a tv-VAR model is used to estimate in real-time the time-varying parameters, then it suffices to plug in the above formulas to get the spectral connectivity measures. Therefore, online estimates of the tv-VAR model parameters will naturally provide online plugin estimator for spectral connectivity:

$$\widehat{\rho}^2_{i,j}(t,\omega) = \frac{|\widehat{S}_{i,j}(t,\omega)|^2}{\widehat{S}_{i,i}(t,\omega)\widehat{S}_{j,j}(t,\omega)}, \tag{30}$$

$$\widehat{\pi}_{i,j}(t,\omega) = \frac{\left|\widehat{\Phi}_{i,j}(t,\omega)\right|}{\sqrt{\sum_{k=1}^{P}\left|\widehat{\Phi}_{k,j}(t,\omega)\right|^2}}. \tag{31}$$

4. SIMULATION STUDIES

We now investigate the ability of the SOPE method to estimate in real-time the time-varying parameters. The key metrics are computational time and also mean-squared error of the connectivity measures.

4.1 Computational time

As it was mentioned in the previous sections, when compared to the KF, the proposed SOPE method has the advantage of being computationally faster, because it does not need to keep track of a covariance matrix of the vectorized parameters. The average execution time per iteration (in milliseconds) to estimate the tv-VAR parameters is reported in the tables 1 and 2.

Table 1. All durations are in milliseconds. P represents the dimension and K represents the order of the tv-VAR model.

Execution time per iteration for Kalman filter

	$P=2$	$P=5$	$P=10$	$P=15$	$P=20$	$P=25$
$K=1$	0.24	0.14	0.31	1.61	7.89	28.58
$K=3$	0.11	0.25	3.60	27.35	116.66	434.45
$K=5$	0.09	0.38	14.10	102.29	517.96	1827.42

Table 2. All durations are in milliseconds. P represents the dimension and K represents the order of the tv-VAR model.

Execution time per iteration for SOPE

	$P=20$	$P=50$	$P=100$	$P=150$	$P=200$	$P=250$
$K=1$	0.11	0.23	0.46	1.39	2.19	3.47
$K=3$	0.19	0.96	4.63	13.77	22.70	38.37
$K=5$	0.34	2.04	13.79	37.96	70.34	114.85

Typically, brain signals such as EEG and LFP have high time resolution (roughly from 256Hz to 2000Hz). Therefore, any method that aims to be applied in real-time needs to have an execution time in the order of a few milliseconds per iteration at most, in order to be able to update the estimates for every observation. Clearly, Tables 1 and 2 show that the SOPE method is computationally faster than the Kalman filter by orders of magnitude. The SOPE method can handle higher dimensions in the order of 256 (dense EEG setting), as opposed to the Kalman filter that can barely handle small dimensions below 25. Furthermore, the SOPE implementation uses the Python language which is known to be slow when compared with lower level languages such as C. Therefore, an implementation of the proposed SOPE method in a lower level programming language will be useful for applications to even higher dimensions.

4.2 Parameter estimation

In the following, we start by providing a simple example to show the quality of the estimates obtained using the SOPE method, we then show a more rigorous comparison between the Kalman filter and the SOPE method in terms of average Sum of Square Errors (SSE).

In the following example, a tv-VAR model with dimension $P = 5$ and order $K = 1$ is used. The five components are split in two groups of size three and two, the two groups could potentially be represented by individual tv-VAR models of dimension three and two respectively. Every entry of the coefficient matrix $\Phi(t)$ is either a cosine function with random amplitude and random phase shift or a zero function, i.e., $\Phi(t)_{i,j} = A_{i,j}\cos\left(\pi\frac{t}{T} + B_{i,j}\right)$ if i and j are in the same group and $\Phi(t)_{i,j} = 0$ otherwise, where $A_{i,j}$ is the random amplitude and $B_{i,j}$ is the random phase shift. In order to ensure stationarity at every time t the amplitudes are designed to be higher when $i = j$ and smaller when $i \neq j$.

A straightforward implementation of Algorithm 1 in Section 2.4 with identity noise covariance matrix provides the following estimates (only 3×3 submatrix, refer to Appendix Section 7.3 for the entire parameter matrix) (with 0.025, 0.5 and 0.975 quantiles represented by dotted lines based on 1000 sample estimates), see Figure 3.

After obtaining the tv-VAR model parameter estimates, it is possible to compute the connectivity measures of interest, for example coherence and PDC as defined in Section 3, refer to Figure 4 to see the 3×3 submatrix estimates and Appendix Section 7.3 for the full matrix connectivity results. Equations 30 and 31 clearly show the nonlinear dependence between the tv-VAR parameter estimates and the measures of connectivity estimates. A small noise on the parameter estimates can result in a large deviation of the estimator of connectivity from the true values. Therefore, having a method that is able to control for the smoothness will result in a significant and practical advantage for practitioners.

The SOPE approach is computationally faster and hence can be applied to higher dimensions. Moreover, it can also provide estimates that compete with the Kalman filter in terms of mean square error. Thus, the SOPE approach does not need to trade shorter computational time with estimation accuracy.

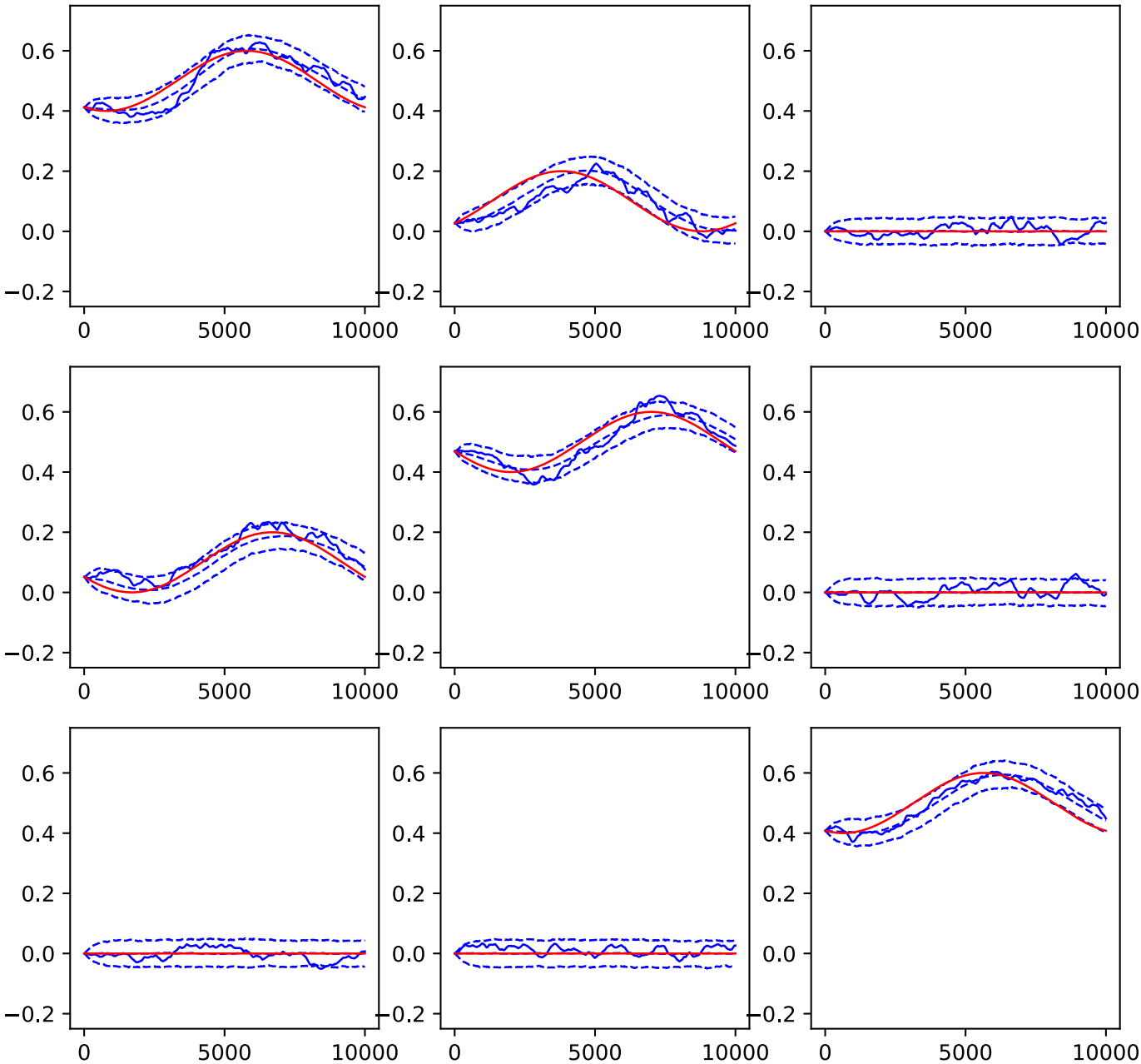

Figure 3. SOPE results for tv-VAR model: the plot in i^{th} row and j^{th} column represent the results for $\left(\Phi_1(t)\right)_{2+i,2+j}$ (red line) and $\left(\widehat{\Phi}_1(t)\right)_{2+i,2+j}$ (blue line), the 2.5, 50 and 97.5 percentiles (based on $B = 1000$ samples) are represented in dotted blue lines.

Both the Kalman filter and the SOPE algorithms have parameters that need to be tuned, the state covariance matrix for the Kalman filter and the regularization parameters α and β for the SOPE. In the following simulation, we select the optimal parameters that minimize the MSE for the tv-VAR model parameter estimates (for $P = 3$ and $K = 2$). For simplicity we take the state covariance as $\Sigma_Q = \sigma^2 I$, experimentally a 0.9 value for β seem to work very well in practice. Figures 5 and 6 show the average MSE as a function of the hyperparameters σ and α:

A few remarks about the results in the previous Figures are as follows. First, both approaches have similar MSE per parameter that is between 0.006 and 0.007 with Kalman filter being slightly better. Second, the optimal σ is in the order of 10^{-5} and the optimal α is in the order of 5000, which means that the Kalman filter is much more sensitive to the choice of σ than the SOPE is to the choice of α.

Given the previous simulations that assess the behaviour

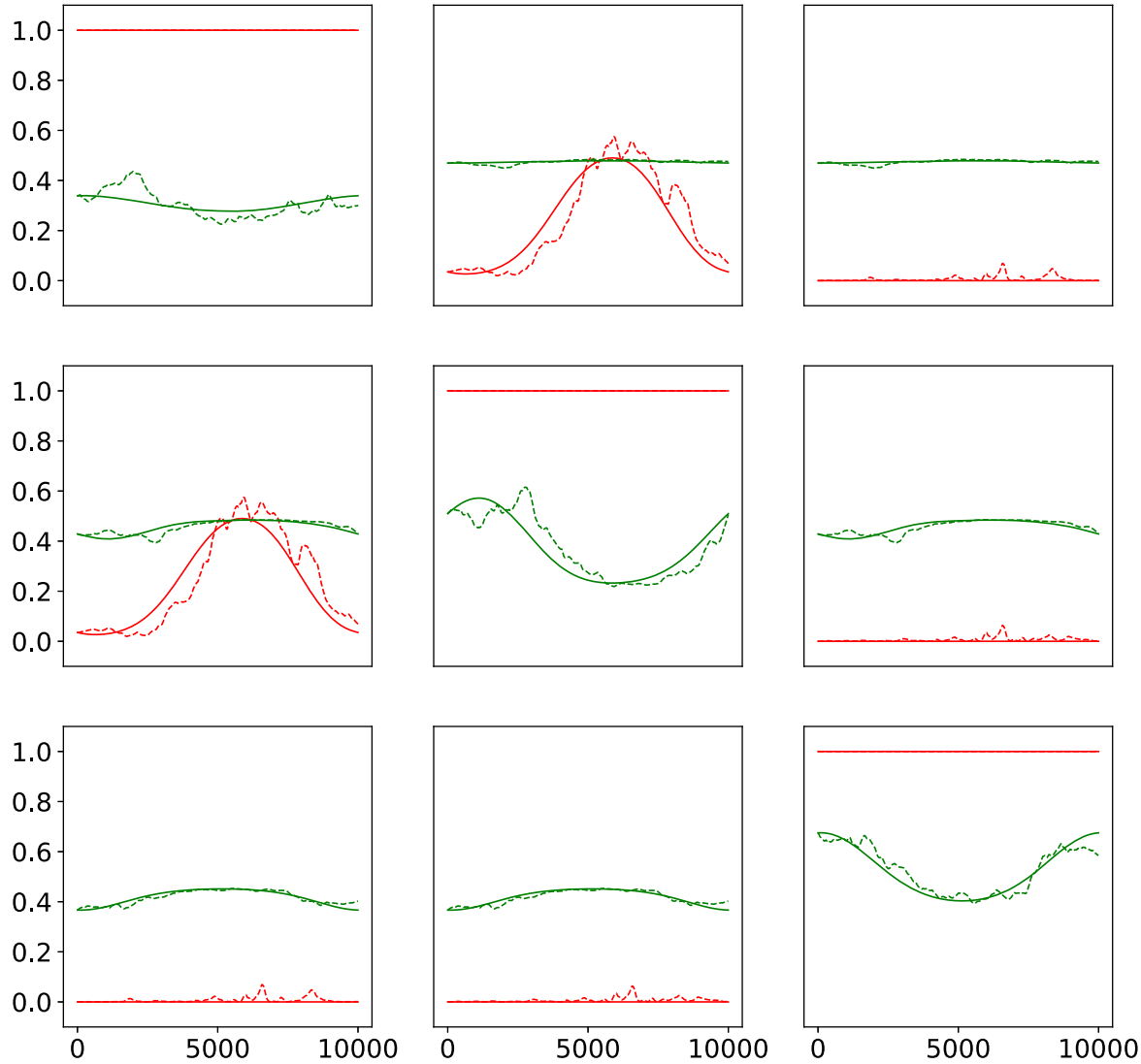

Figure 4. SOPE results for tv-VAR model: the plot in i^{th} row and j^{th} column represent the connectivity estimates between tetrode $2+i$ and $2+j$. Coherence is in red, while PDC is in green, solid lines represent the true quantity and dotted lines represent the estimates (for Delta band).

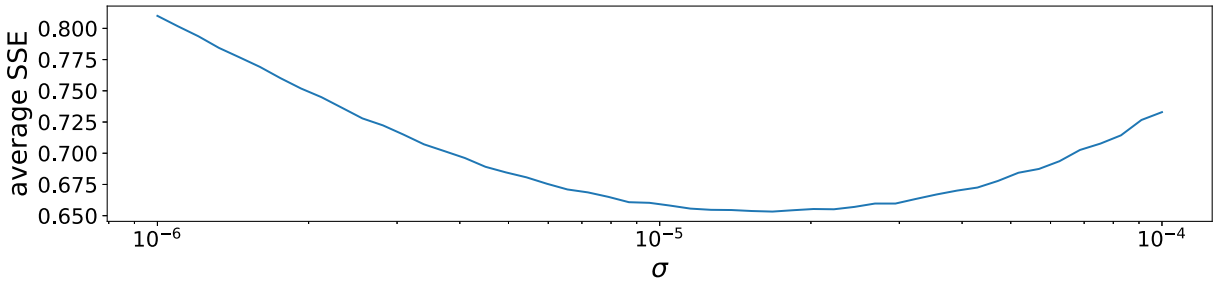

Figure 5. Kalman filter: average MSE per parameter as a function of σ for a tv-VAR model P=3 and K=2.

of the estimators in terms of the MSE with respect to the selection of the hyperparameters, we show here that we can use such knowledge about the hyperparameters and transfer it to another similar but different problem. Thus, using the optimal hyperparameters from the example above, we apply the KF and the SOPE methods to a tv-VAR model with $P = 4$ and $K = 3$ in order to investigate the parameter estimates for both models. We report bellow the results for

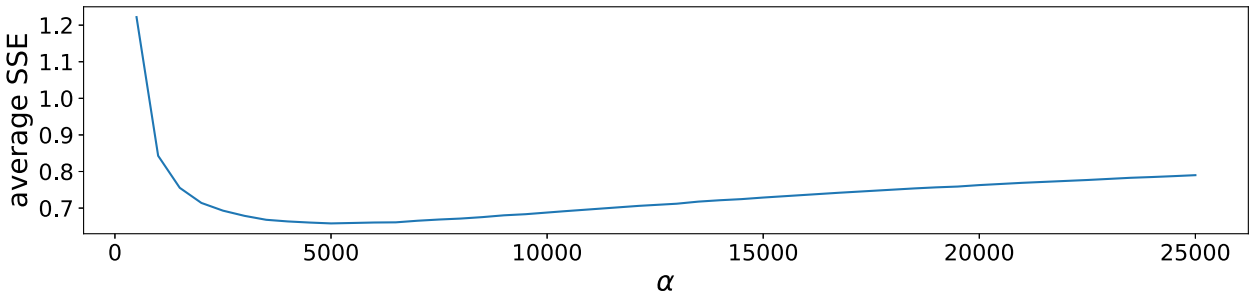

Figure 6. SOPE: average MSE per parameter as a function of α for a tv-VAR model P=3 and K=2.

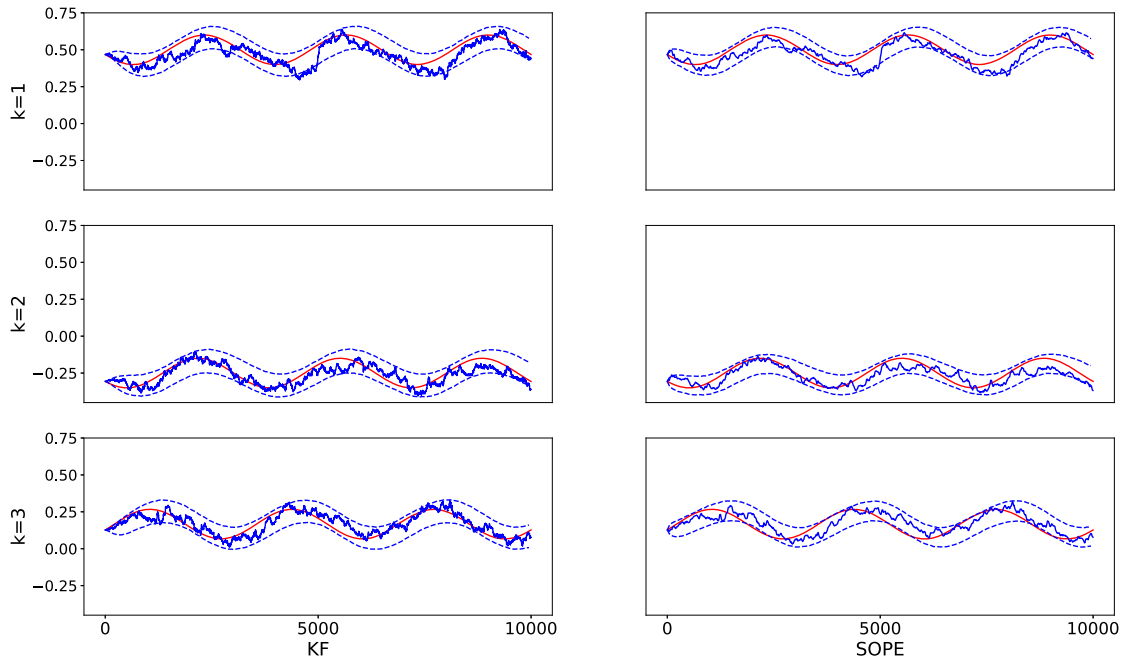

Figure 7. Parameter estimates for a tv-VAR model with $P = 4$ and $K = 3$, pair of tetrodes $(2, 2)$. True parameters are in solid red lines, estimated parameters are in solid blue lines and $2, 5$ and 97.5 percentiles (based on 5000 samples) are in dotted blue lines. KF on the left column and SOPE on the right column. The results are based on the optimal $\sigma \sim 10^{-5}$ and $\alpha \sim 5000$.

the pairs of tetrodes (2, 2) and (4, 2).

The previous example clearly shows that the SOPE method can provide similar results to the Kalman filter, even when the hyperparameters have been selected on another example. However, the Kalman filter does not assume anything about the smoothness of the parameters being estimated. Therefore, it is natural to wonder how would the SOPE method behave in the presence of abrupt changes such as a discontinuity.

4.3 Robustness of the SOPE approach

In order to assess the robustness of the SOPE method to abrupt changes, tv-VAR model order misspecification and to model misspecification in general, we carry on extensive simulation studies. The following scenarios are proposed:

1. The smoothness assumption is not always respected (e.g., presence of discontinuities)
2. The time varying VAR model order K is misspecified (over estimated and under estimated)
3. The entire model is misspecified (e.g., mixture of AR(2) processes)

4.3.1 Abrupt parameter changes

In the following example, we propose to compare the KF and the SOPE over 5000 repetitions in the setting where tv-VAR model includes four discontinuities (at times $t = 1000, t = 4000, t = 7000$ and $t = 8000$) for some discontinuous parameters, which are $\Phi_1(t)_{2,2}, \Phi_2(t)_{2,2}, \Phi_3(t)_{2,2}$. Figure 9 shows how both approaches estimate the discontinuous parameter.

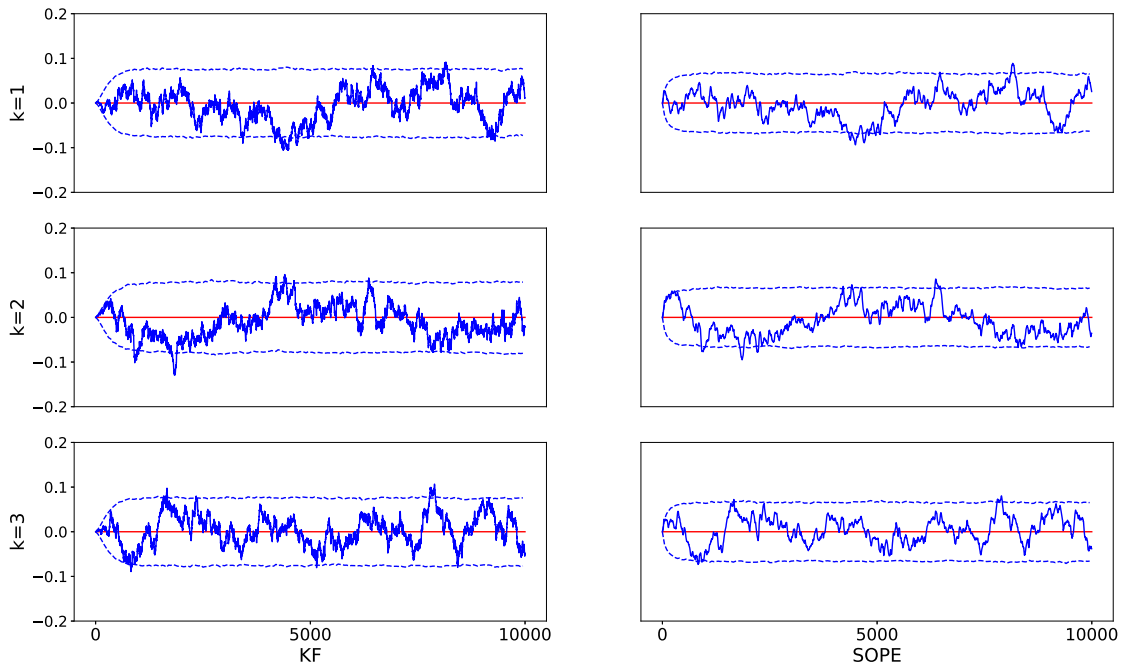

Figure 8. Parameter estimates for a tv-VAR model with $P = 4$ and $K = 3$, pair of tetrodes $(4, 2)$. True parameters are in solid red lines, estimated parameters are in solid blue lines and $2, 5$, 50 and 97.5 percentiles (based on 5000 samples) are in dotted blue lines. KF on the left column and SOPE on the right column. The results are based on the optimal $\sigma \sim 10^{-5}$ and $\alpha \sim 5000$.

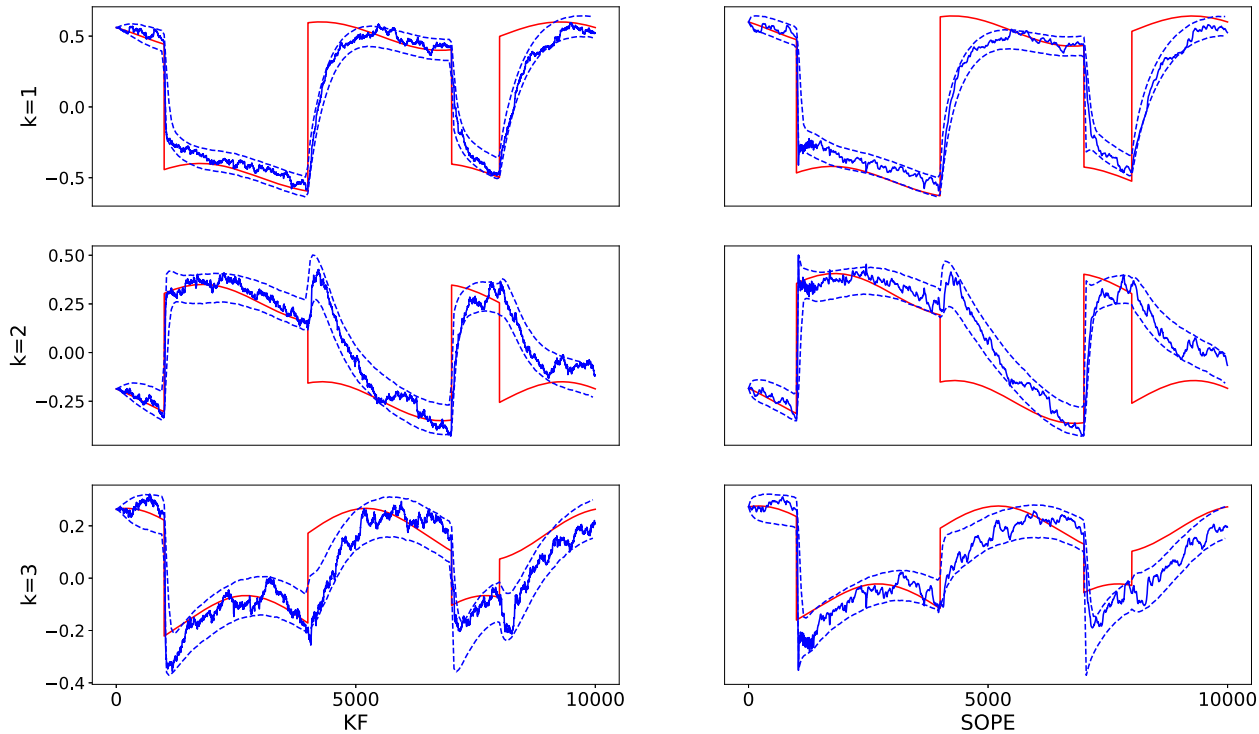

Figure 9. Parameter estimates for a tv-VAR model with with discontinuities, $P = 3$ and $K = 2$, pair of tetrodes $(2, 2)$. True parameters are in solid red lines, estimated parameters are in solid blue lines and $2, 5$, 50 and 97.5 percentiles (based on 5000 samples) are in dotted blue lines. KF on the left column and SOPE on the right column.

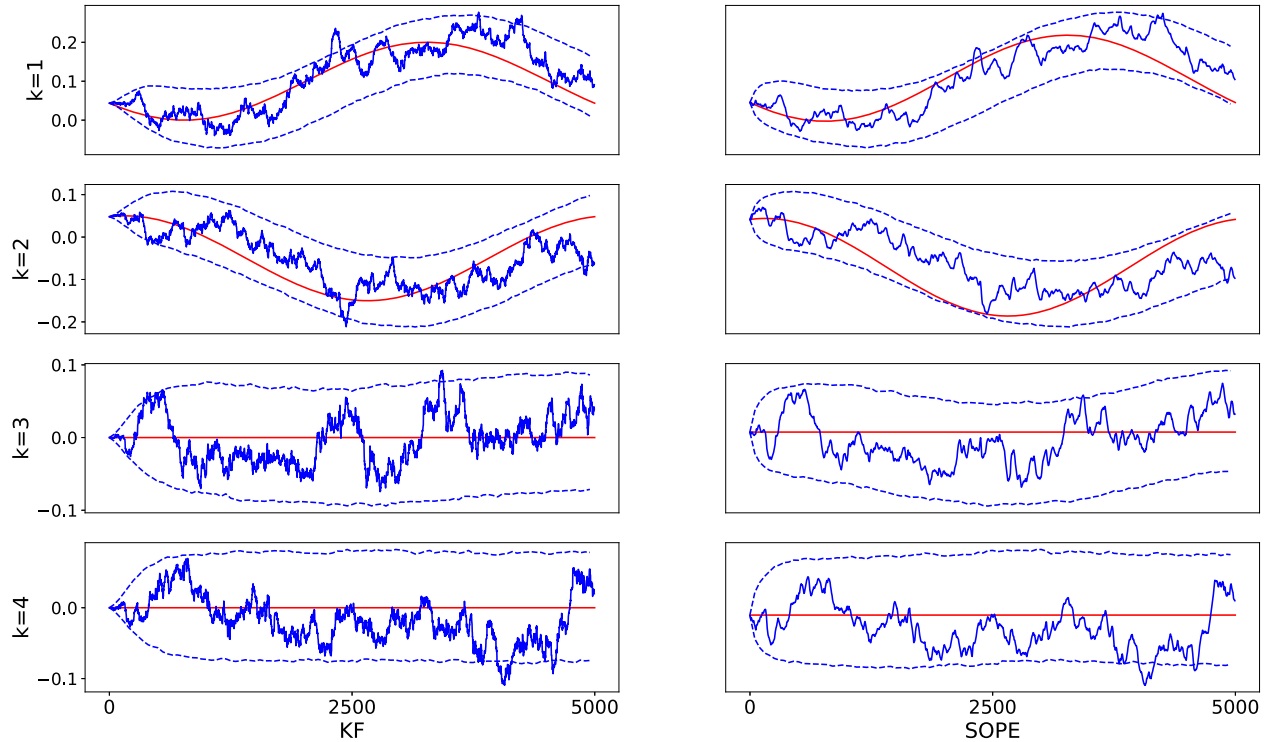

Figure 10. Parameter estimates for a tv-VAR model with with model misspecification (larger K), $P = 4$, $K = 2$ and $\widehat{K} = 4$, pair of tetrodes $(1, 3)$. True parameters are in solid red lines, estimated parameters are in solid blue lines and $2, 5$ and 97.5 percentiles (based on 5000 samples) are in dotted blue lines. KF on the left column and SOPE on the right column.

It is is clear from the figure above that the SOPE method is flexible enough to provide a wide range of estimates, from very smooth to very rough depending on the choice of the penalization hyperparameter. The SOPE has the additional advantage of being computationally fast which enables us to get similar quality results even in higher dimensions at a reasonable computational cost unlike the Kalman filter approach.

4.3.2 Vector autoregression order misspecification

In order for an estimation approach to be robust it needs to behave reasonably well when the order of the model is misspecified. In practice it is very difficult to approximate the order of a VAR model, and even more of a tv-VAR model.

In the following two examples we propose to fit a tv-VAR model ($P = 4$) with the wrong order K deliberately. First, when the order K is over estimated, i.e., $\widehat{K} = 4$ (fitted model in Equation 33) instead of $K = 2$ (true model in Equation 32). Figure 10 shows the tv-VAR parameter estimates in addition to the confidence bands based on 5000 repetitions.

$$X(t) = \sum_{\ell=1}^{K} \Phi_{t,\ell} X(t-\ell) + E(t), \tag{32}$$

$$X(t) = \sum_{\ell=1}^{\widehat{K}} \Phi_{t,\ell} X(t-\ell) + E(t). \tag{33}$$

Similarly, in the following example we propose to fit a tv-VAR model ($P = 4$) with again the wrong order K. However, with the order K being over estimated, i.e., $\widehat{K} = 2$ instead of $K = 4$. Figure 11 shows the estimates in addition to the confidence bands based on 5000 repetitions.

To see the impact of under estimating the tv-VAR model order on coherence and partial directed coherence, we fit the model then compute coherence and PDC, and report the results in Figure 12.

4.3.3 Model misspecification

The following example aims to present how the SOPE approach behaves when the entire model is misspecified. Often, brain signals can be modeled as a mixture of latent frequency-specific processes, as it has been proposed in Granados-Garcia et al. (2021). As such, it is reasonable to simulate data as a mixture of AR(2) processes. Thus, we propose the following formulation:

$$Y_1(t) = a(t) Z_L(t) + \big[1 - a(t)\big] Z_H(t) + \epsilon_1(t) \tag{34}$$

$$Y_2(t) = a(t) Z_L(t) + \big[1 - a(t)\big] Z_H(t) + \epsilon_2(t) \tag{35}$$

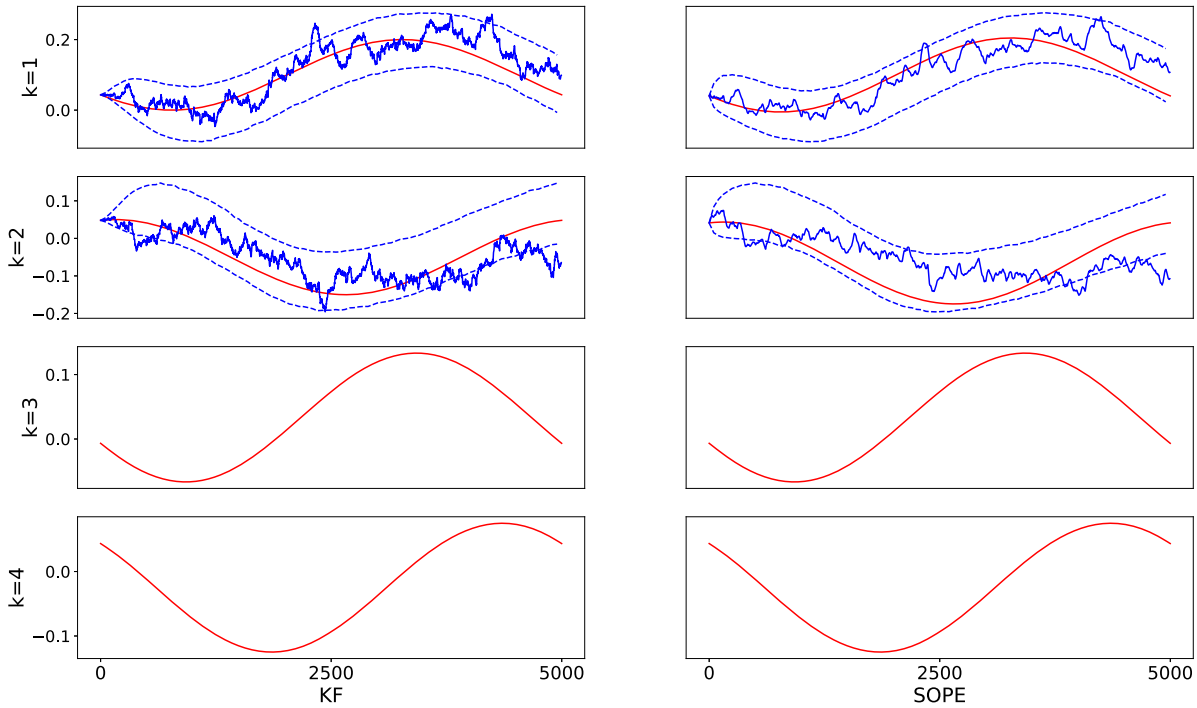

Figure 11. Parameter estimates for a tv-VAR model with with model misspecification (lower K), $P = 4$, $K = 4$ and $\widehat{K} = 2$, pair of tetrodes $(1, 3)$. True parameters are in solid red lines, estimated parameters are in solid blue lines and $2, 5$ and 97.5 percentiles (based on 5000 samples) are in dotted blue lines. KF on the left column and SOPE on the right column.

$$Y_3(t) = a(t)Z_H(t) + \big[1 - a(t)\big]Z_L(t) + \epsilon_3(t) \tag{36}$$

$$Y_4(t) = a(t)Z_H(t) + \big[1 - a(t)\big]Z_L(t) + \epsilon_4(t) \tag{37}$$

Let $Z(t)$, $\epsilon(t)$ and $A(t)$ denote the following:

$$\epsilon(t) = [\epsilon_1(t), \ldots, \epsilon_4(t)]' \tag{38}$$

$$Z(t) = [Z_L(t), Z_H(t)]' \tag{39}$$

$$A(t) = \begin{bmatrix} a(t) & 1 - a(t) \\ a(t) & 1 - a(t) \\ 1 - a(t) & a(t) \\ 1 - a(t) & a(t) \end{bmatrix} \tag{40}$$

Then, the previous model can be written in the following compact form:

$$Y(t) = A(t)Z(t) + \epsilon(t) \tag{41}$$

where $A(t)$ is the time varying mixing matrix composed of the activation function $a(t) = \frac{1}{1+\exp[40(t-T/2)/T]}$ and $Z(t)$ is the latent process composed of two AR(2) processes which spectrum is centered around low frequencies (100Hz) for $Z_L(t)$ and high frequencies (400Hz) for $Z_H(t)$, i.e., $[\phi_1^L = 1.34, \phi_2^L = -0.69]$ and $[\phi_1^H = -1.34, \phi_2^H = -0.69]$. Thus, the first two components share a low frequency AR(2) process in the beginning of the experiment and a high frequency AR(2) process at the end of the experiment. In contrast, the last two components share a high frequency AR(2) process in the beginning of the experiment and a low frequency AR(2) process at the end of the experiment.

After generating the data from the previous model as defined in Equation 41, a tv-VAR model of dimension $P = 4$ and order $K = 5$ is fitted then the coherence is estimated/computed based on the estimated parameters and the true model. The results are reported in Figures 13 and 14 respectively for low and high frequency coherence.

Despite the fact that the coherence estimates in Figures 13 and 14 are far from the true coherence, the SOPE approach can capture the phase transition in the mixture of AR(2) processes quite clearly. Indeed, in low frequency coherence the estimates display a clear transition from a low plateau to a higher one for the first tow components, as can be seen in Figure 13. However, for the last two components we observe the opposite transition from high frequency to low frequency, which is exactly what we expect. Similarly, we make the reverse observations for the high frequency in Figure 14.

5. ONLINE ESTIMATION OF CONNECTIVITY OF RAT LOCAL FIELD POTENTIALS

The ability to remember the order in which events occurred is fundamental to our daily life function. Considerable research shows that the hippocampus, a brain region

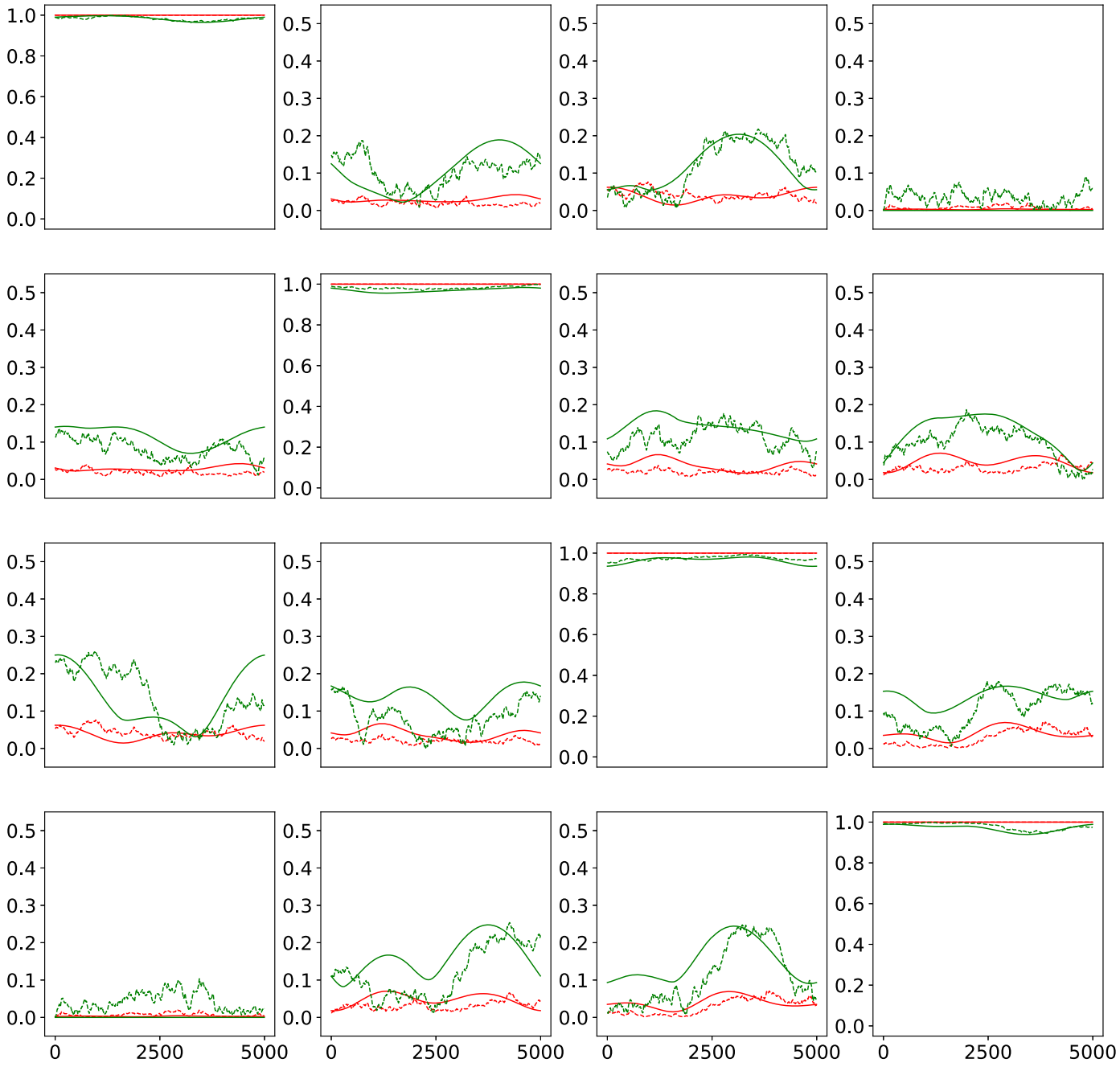

Figure 12. SOPE results for tv-VAR model under order misspecification $K = 4$ instead of $K = 2$: the plot in i^{th} row and j^{th} column represent the connectivity estimates between tetrode i and j. Coherence is in red, while PDC is in green, solid lines represent the true quantity and dotted lines represent the estimates (for Delta band).

strongly conserved across mammals, plays a crucial role in supporting that capacity. However, the underlying neural mechanisms are not well understood. To help address this issue, Allen et al. (2016) conducted an experiment in which they recorded neural activity from the CA1 subregion of the hippocampus as rats were tested on their memory of a sequence of five odors (lemon, rum, anise, vanilla, banana; see figure 15 and 16).

In the context of closed-loop neurofeedback systems, it is critical to have access to real-time monitoring of the brain activity. Hence, fast processing of EEG signals is mandatory to control the appropriate feedback. Our goal in this application is to demonstrate that the SOPE method has the ability to capture the interesting dynamics of brain connectivity in real-time, which will allow others to examine the effects of adjusting experimental parameters, includ-

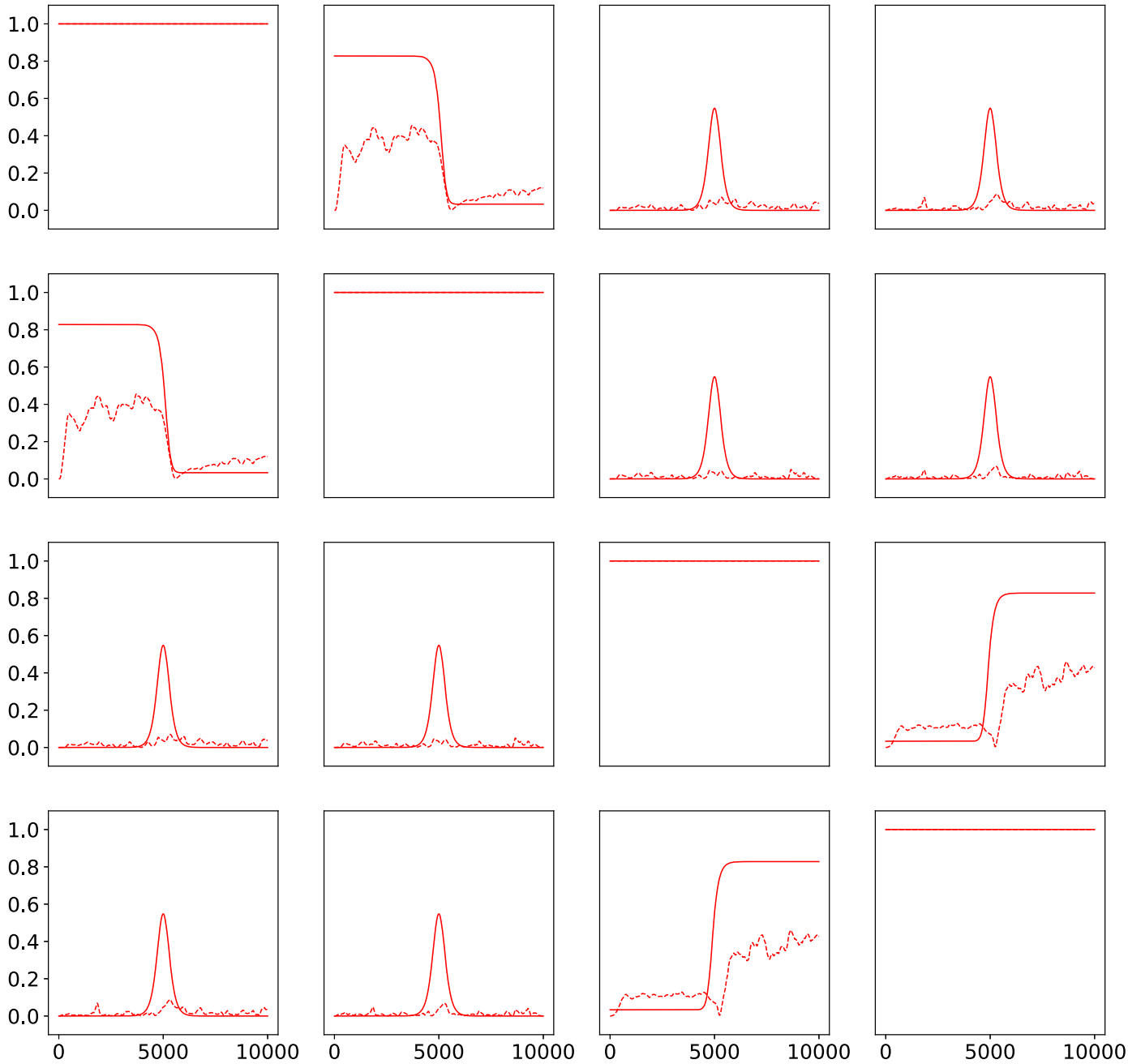

Figure 13. True and estimated time varying low frequency (50-150Hz) average coherence based on the tv-VAR model parameters with P=4, K=5, with initialization at zero. Solid lines represent the true coherence, while dotted lines represent the estimates.

ing stimulus intensity, to achieve the goals of the study or to provide the appropriate feedback for the system, see Nuno, Chung and Maharatna (2021).

In this analysis, 21 tetrodes were selected from 23 originally, namely, $T1, T2, \ldots, T23$ except $T11$ and $T17$. Since it is difficult to showcase the value of an online-method using static figures, a video was created to demonstrate this point. Real-time visualization of the estimated brain connectivity (for both coherence and PDC) network can be accessed in the following link: real-time brain connectivity estimates In the following, we present the analysis of the rat LFP recordings in order to illustrate the benefits gained from controlling the smoothness of the connectivity estimates, and how this approach can lead to the discovery of interesting patterns that could have been missed otherwise.

As a starting point, a tv-VAR model of order $K = 1$ is to fit to the LFP data where the dimension $P = 21$, as it is the simplest model. In the next step, we compute the connec-

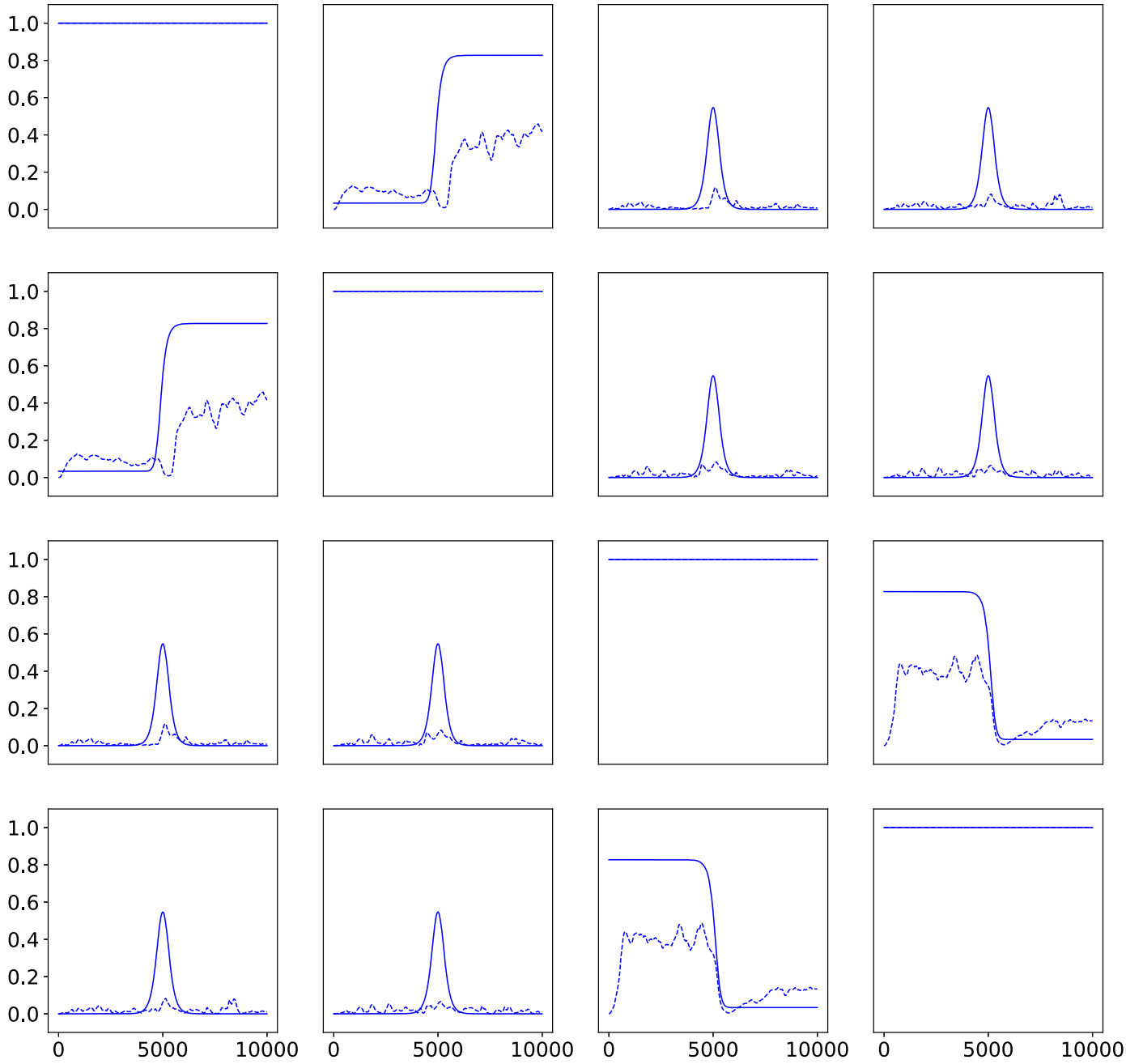

Figure 14. True and estimated time varying high frequency (350-450Hz) coherence based on the tv-VAR model with P=4, K=5, with initialization at zero. Solid lines represent the true coherence, while dotted lines represent the estimates.

tivity measures presented in Section 3 (coherence and partial directed coherence) for the slow gamma (20-40 Hz) and theta-alpha (4-12 Hz) oscillation bands. As previously mentioned, these measures of connectivity are very sensitive to the variance of the observation noise. Therefore, a small penalization for the tv-VAR parameters estimator's roughness will lead to a poor connectivity estimator. Figure 17 clearly shows the poor quality of such estimates using small penalization coefficient ($\alpha \sim 500$). However, when more adequate (higher) levels of smoothing are selected, the connectivity estimates for coherence seem to be strongly modulated by the sequence of odors for the slow gamma band, as can be seen in Figure 18 for the T6-T12 pair. Nevertheless, this amplitude modulation does not seem to be present for lower frequency bands such as the theta-alpha band, as can be seen in Figure 19. Furthermore, the above mentioned amplitude modulation is not present between all pairs of tetrodes (e.g., T12-T14 pair), which suggests some specific patterns in the

Figure 15. Rat with the implanted tetrodes.

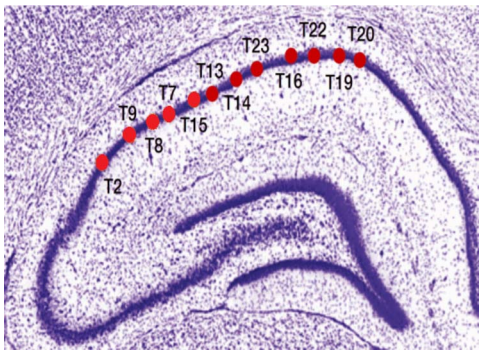

Figure 16. Tetrodes location in the hippocampus.

activity at the electrode tips that is driven by the sequence of odors (instead of potential electrophysiological artifacts or noise). The presence of this amplitude modulation of dependence between some tetrodes in the slow gamma band and not in the theta band supports the results found in Allen et al. (2016).

In both Figures 18 and 19, it can be seen that the choice of the penalization parameter α does not strongly influence the estimated connectivity as long as it is in the same order of magnitude, i.e., $\alpha \in O(10^4)$. This observation provides information regarding the low sensitivity of the SOPE method to the choice of the regularization parameter α.

The offline analysis of the tetrodes connectivity network presents some interesting dynamics after an odor is introduced. To analyze this activity, the average connectivity measures were computed for the following temporal milestones: immediately before the odor presentation (-250 ms) and immediately after the presentation (+250 ms). These values were then compared with the connectivity's 0.75-quantile threshold based on the empirical distribution which is described as follows.

The quantiles are computed based on the empirical distribution of the connectivity that is observed during the entire epoch which lasts for 50 seconds, at 1kHz sampling rate ($\widehat{\mathcal{C}}(t)$ and $\widehat{\mathcal{PDC}}(t), \forall t$). For every pair of tetrodes, the connectivity values are computed and then sorted from lowest to highest, the corresponding quantiles $(0.50, 0.75, 0.90)$ are computed, see Figure 18, top left subfigure as an example, where the 0.75-quantile for the T5-T20 tetrode pair is represented in dotted black line.

We present bellow only the results for the 0.75-quantile, the results for the 0.5 and 0.9 quantiles are kept for the Appendix, see Section 7.4. In Figures 20, 21, 22 and 23, the $T5$ and $T18$ tetrodes seem to play a central role as it appears to be associated with many changes in the graph structure. It also appears that the rum and banana odors seem to alter the connectivity between all tetrodes more often than the other odors. Results for other quantiles such as the 0.50 and the 0.90 are reported in the Appendix.

Coherence between the rat LFP tetrodes i and j can be viewed as a measure of the relative synchrony between the corresponding signals. More precisely, it is shown in Ombao and Bellegem (2008) to be the squared cross correlation between the components of the signals at some specific frequency band. On the other hand, partial directed coherence from tetrode i to tetrode j at some gven frequency band is a measure of the normalized (by the total amount of information flowing out from the jth tetrode) information flow from tetrode j to tetrode i. Hence, those connectivity measures capture different facets of brain connectivity: the first one captures the relative synchrony between tetrodes and the second captures the portion of the directed flow of information from one tetrode to another. From Figures 20, 21, 22 and 23, one can notice that coherence is a more conservative measure of connectivity (less connections in the networks), whereas partial directed coherence seems to be less conservative (more connections in the networks) by providing a more specific information about connectivity (directed and normalized).

The connectivity graphs in Figures 20, 21, 22 and 23, were obtained respectively from high frequency coherence, high frequency partial directed coherence, low frequency coherence and low frequency partial directed coherence. For every odor (lemon, rum, anise, vanilla and banana) the average connectivity networks were computed 250ms before the presentation of the odor and 250ms after the presentation of the odor. To assess the significance of those measures, the comparison with quantile values of the empirical connectivity networks before the start of the epoch was made. Therefore, an edge is drawn if the corresponding average connectivity was higher than the 75-percentile threshold. A green edge after the introduction of the odor, indicates an odor-induced increase in connectivity beyond the threshold. A gray edge after the introduction of the odor, indicates an odor-induced decrease in connectivity below the threshold.

The banana and rum odors introduced significant alterations in the connectivity structure, as can be seen in Figures 20 and 23. The rum odor seem to inhibit both coherence and PDC, see Figures 20 and 23. However, the banana seems to inhibit coherence at high frequency and increase PDC at low frequency, see 20 and 23.

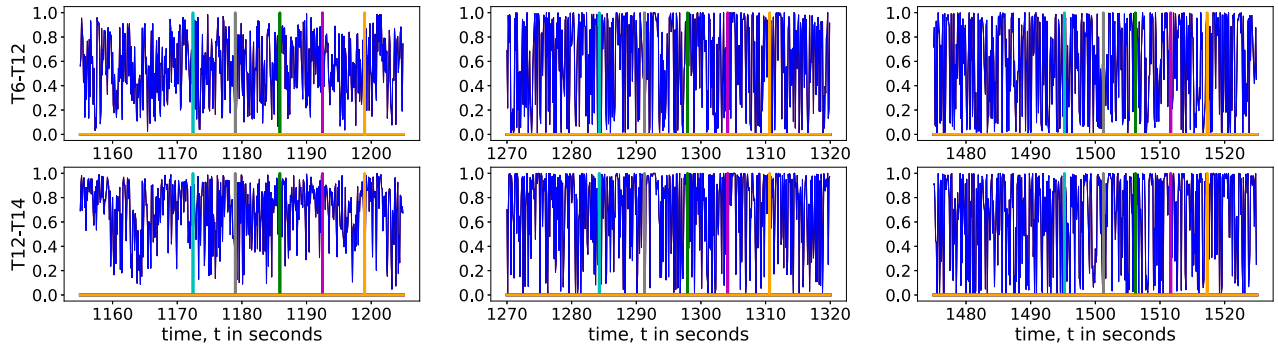

Figure 17. Coherence estimates using SOPE (slow gamma frequency band) during correct sequences of odors (lemon, rum, anise, vanilla, banana), vertical bars indicate the start of each odor. With regularization parameters $\alpha = 500$ *and* $\beta = 0.9$.

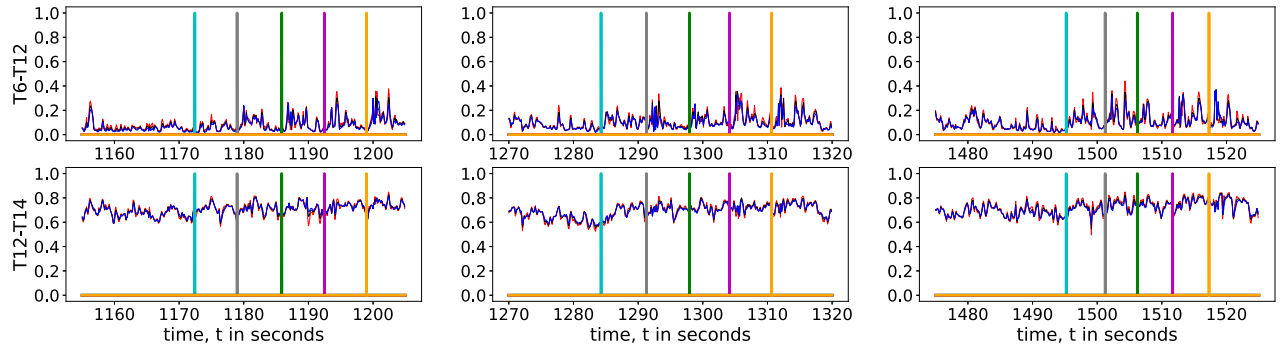

Figure 18. Coherence estimates using SOPE (slow gamma frequency band) during correct sequences of odors (in order: lemon, rum, anise, vanilla, banana), vertical bars indicate the start of each odor. With regularization parameters $\alpha = 1500$ *(red line),* $\alpha = 20000$ *(black line),* $\alpha = 25000$ *(blue line) and* $\beta = 0.9$.

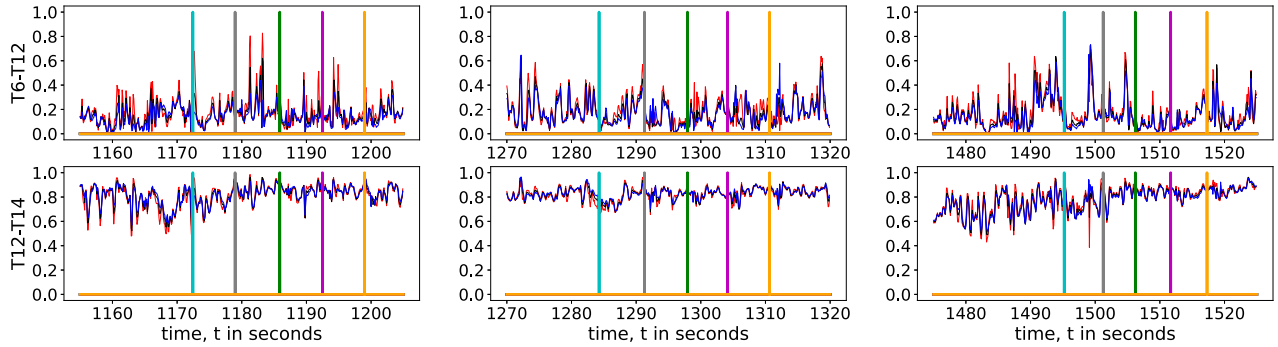

Figure 19. Coherence estimates using SOPE (theta frequency band) during correct sequences of odors (lemon, rum, anise, vanilla, banana), vertical bars indicate the start of each odor. With regularization parameters $\alpha = 15000$ *(red line),* $\alpha = 2000$ *(black line),* $\alpha = 25000$ *(blue line) and* $\beta = 0.9$.

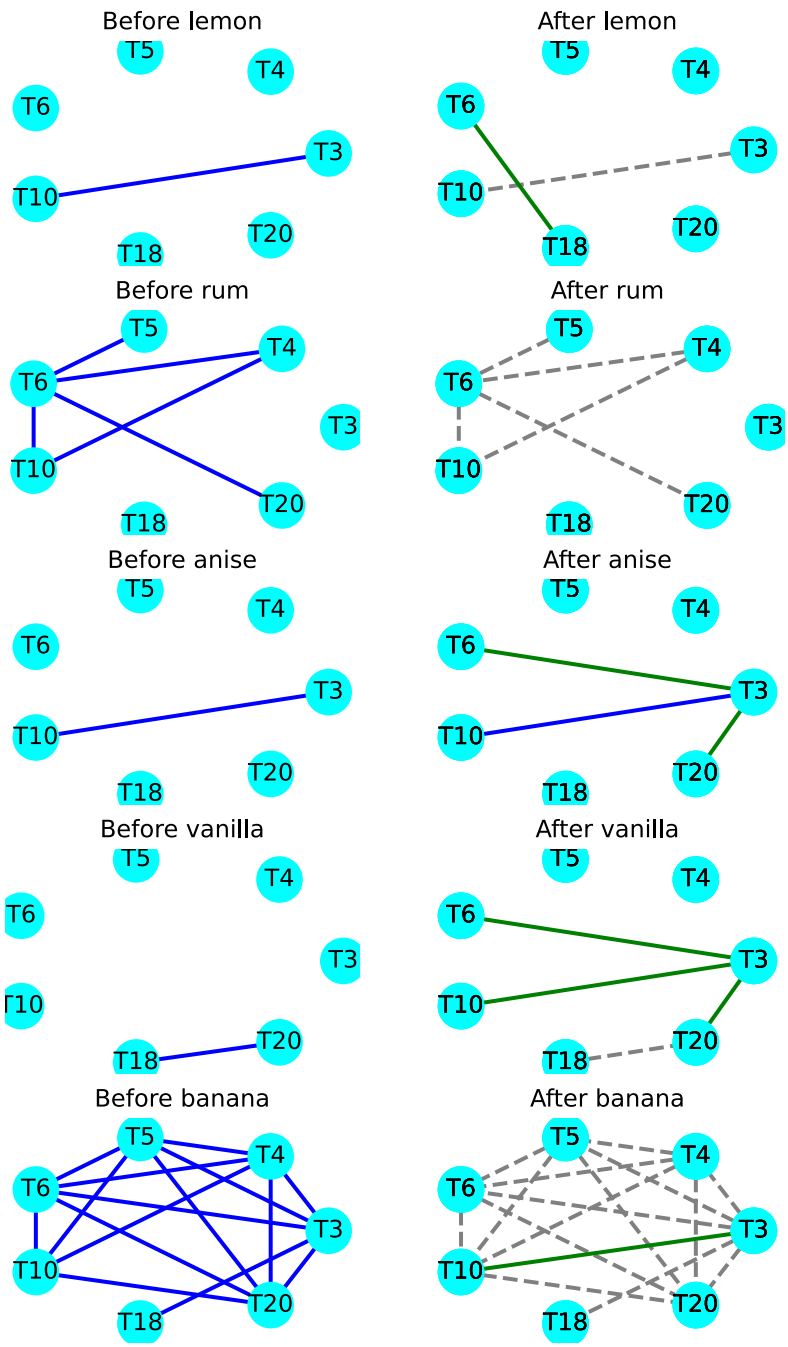

Figure 20. In the prior to odor presentation networks (based on the realizations at the start of the epoch) an edge is present if the high frequency coherence connectivity is higher than some threshold as defined by the 0.75 quantile (the quantile is computed based on the data before the odors are presented). On the post odor presentation networks (based on the realizations at the start of the epoch), a blue edge indicates that this connectivity was high prior to and it remained post odor presentation, a dotted gray edge indicates that this connectivity was high before and it is no longer the case, finally a green edge indicates that this connectivity was low before and it became high after the introduction of the banana odor.

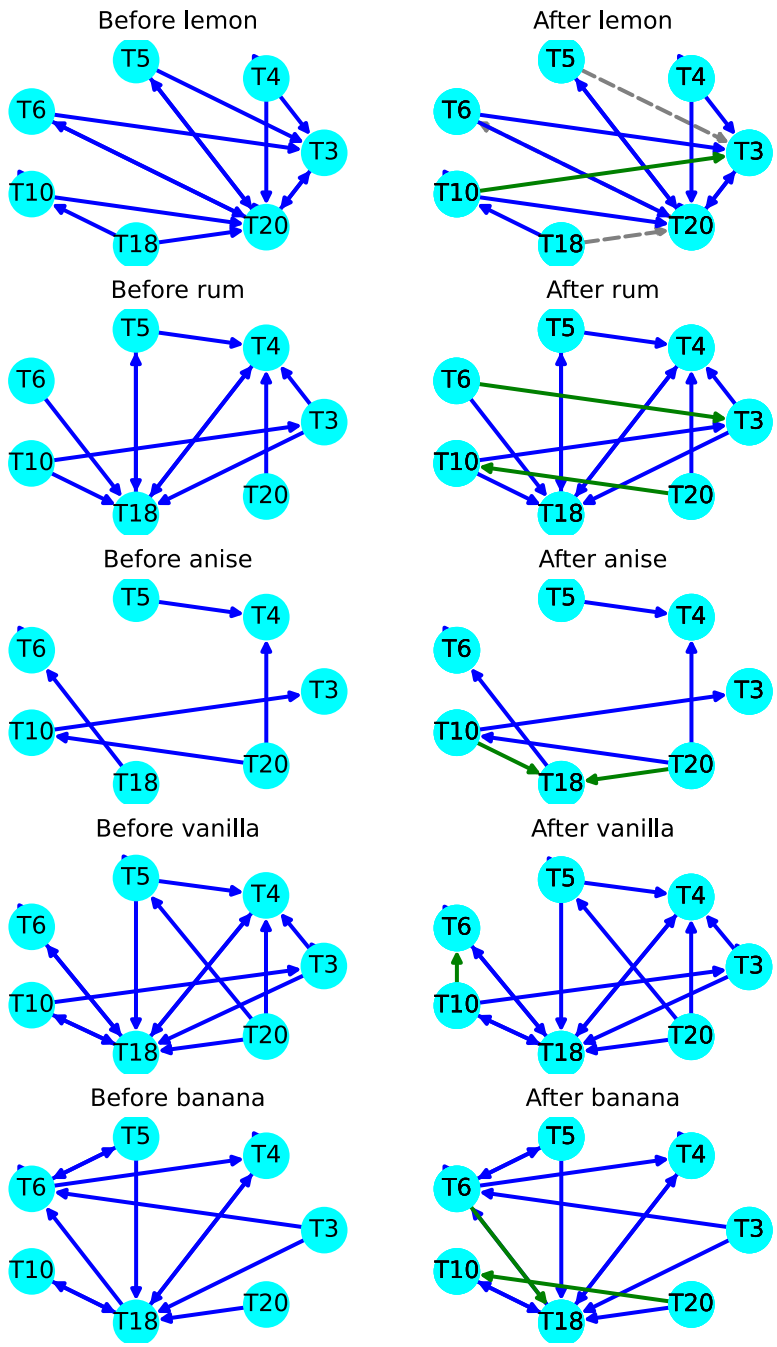

Figure 21. In the prior to odor presentation networks (based on the realizations at the start of the epoch) an edge is present if the high frequency PDC connectivity is higher than some threshold as defined by the 0.75 quantile (the quantile is computed based on the data before the odors are presented). On the post odor presentation networks (based on the realizations at the start of the epoch), a blue edge indicates that this connectivity was high prior to and it remained post odor presentation, a dotted gray edge indicates that this connectivity was high before and it is no longer the case, finally a green edge indicates that this connectivity was low before and it became high after the introduction of the banana odor.

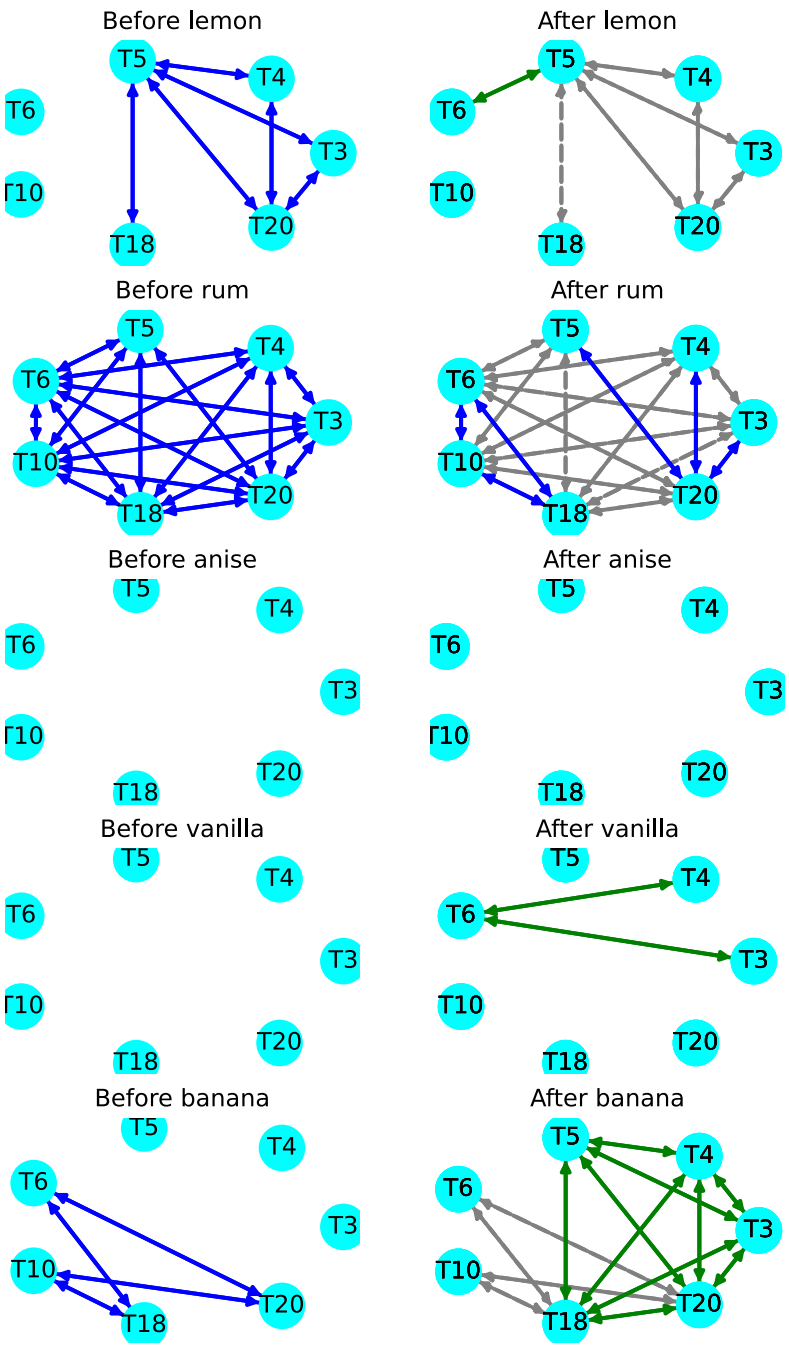

Figure 22. In the prior to odor presentation networks (based on the realizations at the start of the epoch) an edge is present if the low frequency coherence connectivity is higher than some threshold as defined by the 0.75 quantile (the quantile is computed based on the data before the odors are presented). On the post odor presentation networks (based on the realizations at the start of the epoch), a blue edge indicates that this connectivity was high prior to and it remained post odor presentation, a dotted gray edge indicates that this connectivity was high before and it is no longer the case, finally a green edge indicates that this connectivity was low before and it became high after the introduction of the banana odor.

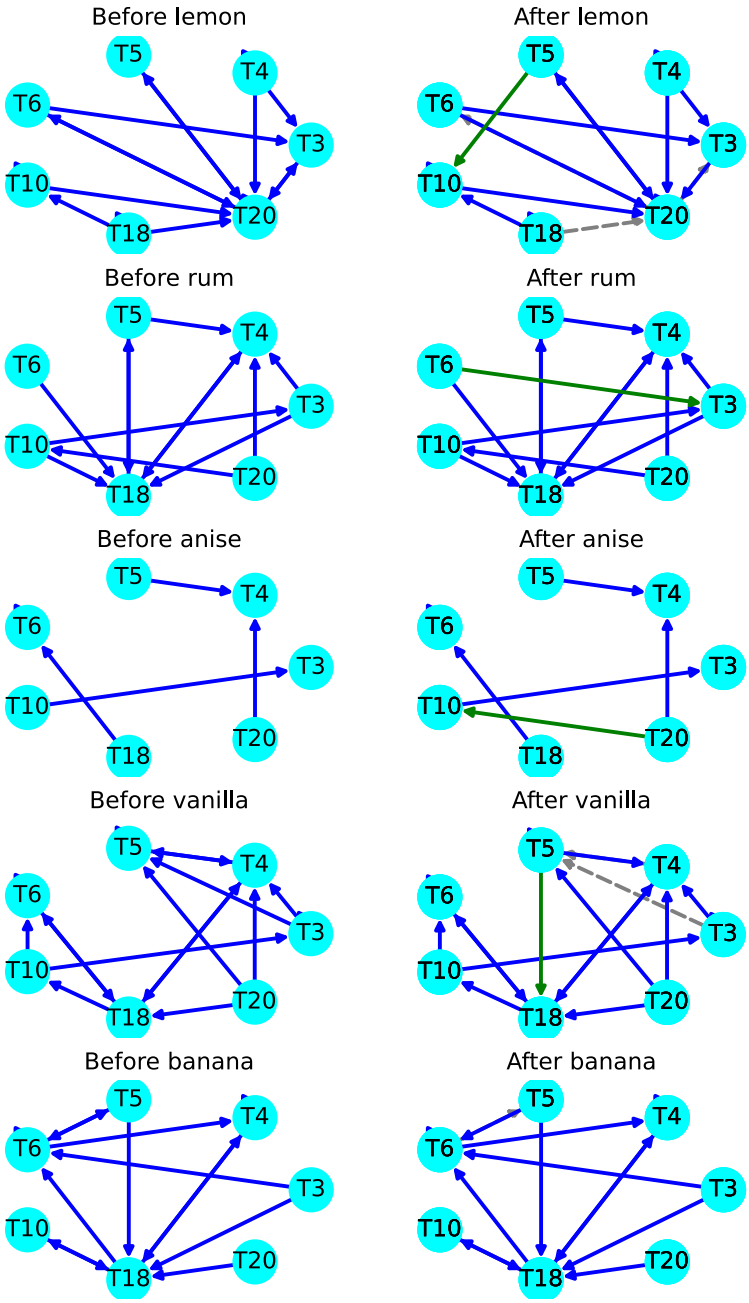

Figure 23. In the prior to odor presentation networks (based on the realizations at the start of the epoch) an edge is present if the low frequency PDC connectivity is higher than some threshold as defined by the 0.75 quantile (the quantile is computed based on the data before the odors are presented). On the post odor presentation networks (based on the realizations at the start of the epoch), a blue edge indicates that this connectivity was high prior to and it remained post odor presentation, a dotted gray edge indicates that this connectivity was high before and it is no longer the case, finally a green edge indicates that this connectivity was low before and it became high after the introduction of the banana odor.

For the neuroscientist, it is important to be able to observe these alterations in brain connectivity as the odor stimuli are presented (rather than retrospectively) because this could allow for adaptive tuning of the odor intensity as needed. From the above analyses, it is clear that the SOPE method can successfully control the level of smoothness of the connectivity estimates. The SOPE provides access to smooth (robust to noise and artifacts) real-time estimates that can capture reliably the brain connectivity. Using the SOPE method will allow practitioners to have immediate feedback on the efficacy of the stimulus (type and intensity) that is being applied. This would allow the selection of the experiment settings before the completion of the data acquisition process. Real-time visualization of the rat's hippocampus connectivity network is shown in: real-time visualization of the rat's hippocampus connectivity network.

6. CONCLUSION

We presented a new online parameter estimation method (SOPE) for locally stationary time series under the context of a tv-VAR model. It was shown that the proposed SOPE method can provide smooth and online estimates with a reasonable computational time that allows it to scale for higher dimensions unlike other competing methods. Furthermore, the SOPE method is well motivated since it is based on a penalized least squares/likelihood approach. As demonstrated, the penalization term can have a Bayesian interpretation, with different penalization terms corresponding to different prior choices.

When compared to the Kalman filter, the SOPE method can provide very similar estimates, but with the significant advantage of being computationally more efficient, and thus it can be applied to higher dimensional time series. Furthermore, this new method provides a meaningful way to control the smoothness of the estimates using only two hyperparameters (α and β), which are much simpler to tune unlike the Kalman filter parameters. Asymptotically our approach has theoretical guaranties that indicate its asymptotic behaviour depending on the choice of the penalization coefficients. The interest behind this approach is to increase the robustness of the estimates and to stabilize the connectivity measures (which often involve non-smooth functions of the parameters e.g, Coherence, as it was shown in Figures 17 and 19). Using the SOPE method to estimate coherence and partial directed coherence between different tetrodes present in the CA1 region of the hippocampus, the SOPE method was able to detect the interesting dynamics in connectivity and the influence of the odor stimulus on the connectivity structure – in real-time. Thus, the SOPE will be a useful tool for experiments requiring the tracking of changes in connectivity in real-time for online manipulation of experimental parameters or circuit activity.

7. APPENDIX

7.1 Generalized SOPE

If the innovation components are not iid ($\Sigma_E \neq I$), the algorithm presented in section 2.3 can be adapted similarly to the generalized least squares.

$$-log f\big(X(t)\big|b, U(t), \Sigma_E\big) \propto \big|\big|X(t) - bU(t)\big|\big|^2_{\Sigma_E^{-1}} \tag{42}$$

which in turn leads to the following problem:

$$\begin{aligned} \widehat{\Phi}(t) &= \underset{b\in\mathbb{R}^{P\times KP}}{\arg\max}\, f\big(X(t)\big|b, U(t)\big) f\big(b\big|\mathcal{I}_{t-1}\big) \\ \iff \widehat{\Phi}(t) &= \underset{b\in\mathbb{R}^{P\times KP}}{\arg\min}\, \big|\big|X(t) - bU(t)\big|\big|^2_{\Sigma_E^{-1}} + P(b) \end{aligned} \tag{43}$$

Solving Equation (26), is not straightforward. Assuming the covariance is known, a change of variable

$$\tilde{U} = \Big[\big(\Sigma_E^{-\frac{1}{2}} X(t-1)\big)', \ldots, \big(\Sigma_E^{-\frac{1}{2}} X(t-K)\big)'\Big]',$$

and

$$\tilde{\Phi}(t) = \Sigma_E^{-\frac{1}{2}} \Phi(t) \big(I_K \otimes \Sigma_E^{\frac{1}{2}}\big),$$

could be performed to decorrelate the time series components, which allows the problem to be solved in a similar way to the least squares problem, then another change of variable would be necessary to cancel the initial change of variable. In practice, the covariance matrix Σ_E is unknown and consequently it has to be estimated simultaneously with the parameters of interest $\Phi(t)$ based on the residuals:

$$\widehat{\Sigma}_E = \Sigma_T = \frac{T-1}{T}\Sigma_{T-1} + \frac{1}{T} R(T)R(T)' = \sum_{t=1}^{T} \frac{1}{T} R(t)R(t)' \tag{44}$$

this leads to the following general SOPE algorithm:

Algorithm 2 Smooth Online Parameter Estimation for tv-VAR models (general covariance)

procedure GetSmoothEstimates($X(1), \ldots, X(T)$)
  Initialize:
  $\widehat{\Phi}(K-1) =$ Least Squares
  $\widehat{\Phi}(K) = \widehat{\Phi}(K-1)$
  $\Sigma_K = I$ (noise covariance)
  $\alpha \in (0, \infty)$, $\beta \in [0, \ldots, 1)$
  for $t = K+1, \ldots, T$ **do**
    $\tilde{X} = \Sigma_{t-1}^{-1/2} X(t)$
    $\tilde{U} = \Big[\big(\Sigma_{t-1}^{-1/2} X(t-1)\big)', \ldots, \big(\Sigma_{t-1}^{-1/2} X(t-K)\big)'\Big]'$
    $\widehat{\tilde{\Phi}}(t) = \Big(\tilde{X}\tilde{U}' + \lambda\Big[\widehat{\tilde{\Phi}}(t-1) + \beta\big(\widehat{\tilde{\Phi}}(t-1) - \widehat{\tilde{\Phi}}(t-2)\big)\Big]\Big)^{-1} \big(\tilde{U}\tilde{U}' + \lambda I\big)^{-1}$
    $\widehat{\Phi}(t) = \Sigma_{t-1}^{1/2} \widehat{\tilde{\Phi}}(t) \big(I_K \otimes \Sigma_{t-1}^{-1/2}\big)$
    $R(t) = X(t) - \widehat{\Phi}(t) U(t)$
    $\Sigma_t = \frac{t-1}{t}\Sigma_{t-1} + \frac{1}{t} R(t)R(t)'$
  end for
end procedure

7.2 Infill asymptotics

Let $h = \frac{1}{T} \underset{T\to\infty}{\to} dt$. In the following, we use the dot notation as a shortcut: $\dot{b} = \frac{d}{du}b$ and $\ddot{b} = \frac{d^2}{du^2}b$ etc., thus, for properly specified coefficients $h(T)$, the following limits are derived:

(45)
$$\begin{cases} \lambda h^2 \underset{T\to\infty}{\to} c_1 \implies \lambda\big|\big|b(\frac{t}{T}) - b(\frac{t-1}{T})\big|\big|_F^2 \\ \quad = \lambda h^2 \big|\big|\frac{b(\frac{t}{T})-b(\frac{t-1}{T})}{h}\big|\big|_F^2 \qquad \to c_1\big|\big|\dot{b}(u)\big|\big|_F^2 \\ \lambda h^4 \underset{T\to\infty}{\to} c_2 \implies \lambda\big|\big|b(\frac{t}{T}) - 2b(\frac{t-1}{T}) + b(\frac{t-2}{T})\big|\big|_F^2 \\ \quad = \lambda h^4 \big|\big|\frac{b(\frac{t}{T})-2b(\frac{t-1}{T})+b(\frac{t-2}{T})}{h^2}\big|\big|_F^2 \qquad \to c_2\big|\big|\ddot{b}(u)\big|\big|_F^2 \end{cases}$$

hence, considering the overall estimation problem, one can deduce that the minimization problem of the sum of square errors will turn into a minimization of an integral provided that the penalization parameters grow at the adequate rate:

(46)
$$\sum_{t=1}^{T} \frac{1}{T}\Big[\big|\big|X(t) - b(t)U(t)\big|\big|_2^2 + \lambda\big|\big|b(t) - b(t-1)\big|\big|_F^2\Big] \to \int_{u=0}^{1} \big|\big|X(\lfloor uT\rfloor) - b(u)U(\lfloor uT\rfloor)\big|\big|_2^2 + c_1\big|\big|\dot{b}(u)\big|\big|_F^2 du$$

(47)
$$\sum_{t=1}^{T} \frac{1}{T}\Big[\big|\big|X(t)-b(t)U(t)\big|\big|_2^2+\lambda\big|\big|b(t)-2b(t-1)+b(t-2)\big|\big|_F^2\Big] \to \int_{u=0}^{1} \big|\big|X(\lfloor uT\rfloor) - b(u)U(\lfloor uT\rfloor)\big|\big|_2^2 + c_2\big|\big|\ddot{b}(u)\big|\big|_F^2 du$$

which leads to the following calculus of variations problems:

(48)
$$\widehat{\Phi} = \underset{b\in\mathcal{C}^2([0,1],\mathbb{R}^{P\times KP})}{\arg\min} \int_{u=0}^{1} \big|\big|X(\lfloor uT\rfloor) - b(u)U(\lfloor uT\rfloor)\big|\big|_2^2 + c_1\big|\big|\dot{b}(u)\big|\big|_F^2 du$$

(49)
$$\widehat{\Phi} = \underset{b\in\mathcal{C}^3([0,1],\mathbb{R}^{P\times KP})}{\arg\min} \int_{u=0}^{1} \big|\big|X(\lfloor uT\rfloor) - b(u)U(\lfloor uT\rfloor)\big|\big|_2^2 + c_2\big|\big|\ddot{b}(u)\big|\big|_F^2 du$$

in order to be rigorous here we need to state the boundary conditions of the problem. Of course, in practice we cannot have these boundary conditions since we need to solve this problem online. However, given previous estimates will automatically impose some boundary conditions.

The two problems in (29) and (30) involve a minimization over a space of functions. Now the problem can be stated in the framework of calculus of variations as follows. For penalty term in (9) we asymptotically get:

(50)
$$\begin{cases} \mathcal{L}(u,b,\dot{b}) & = \big|\big|X(\lfloor uT\rfloor) - b(u)U(\lfloor uT\rfloor)\big|\big|_2^2 + \\ & \quad c_1\big|\big|\dot{b}(u)\big|\big|_F^2 \\ \mathcal{J}[b] & = \int_0^1 \mathcal{L}(u,b,\dot{b})du \\ \widehat{\Phi} & = \underset{b\in B}{\arg\min}\, \mathcal{J}[b] \end{cases}$$

where the Lagrangian term is just the least squares term that corresponds to the likelihood of the observations plus a penalty term for the roughness of the function, and B is some smooth enough class of functions. To solve such problems, we consider a necessary condition for optimality also known as Euler-Lagrange equation of the involved Lagrangian:

$$\frac{\partial\mathcal{L}}{\partial b} - \frac{d}{du}\frac{\partial\mathcal{L}}{\partial\dot{b}} = 0\Big|_{b=\widehat{\Phi}}$$
$$\iff \ddot{\widehat{\Phi}}(u) = \frac{1}{c_1}\nabla_{\widehat{\Phi}}||X(\lfloor uT\rfloor) - \widehat{\Phi}U(\lfloor uT\rfloor)||_2^2.$$

similarly, using the second-order difference penalty term in (10) will lead asymptotically to the formalized problem:

(51)
$$\begin{cases} \mathcal{L}(u,b,\dot{b},\ddot{b}) & = \big|\big|X(\lfloor uT\rfloor) - b(u)U(\lfloor uT\rfloor)\big|\big|_2^2 + \\ & \quad c_2\big|\big|\ddot{b}(u)\big|\big|_F^2 \\ \mathcal{J}[b] & = \int_0^1 \mathcal{L}(u,b,\dot{b},\ddot{b})du \\ \widehat{\Phi} & = \underset{b\in B}{\arg\min}\, \mathcal{J}[b] \end{cases}$$

where again the Lagrangian term is just the least squares term that corresponds to the likelihood of the observations, plus a penalization term for the roughness (curvature) of the function. Similarly, the Euler-Lagrange equation becomes:

$$\frac{\partial\mathcal{L}}{\partial b} - \frac{d}{du}\frac{\partial\mathcal{L}}{\partial\dot{b}} + \frac{d}{du}\frac{\partial\mathcal{L}}{\partial\ddot{b}} = 0\Big|_{b=\widehat{\Phi}}$$
$$\iff \dddot{\widehat{\Phi}}(u) = -\frac{1}{c_2}\nabla_{\widehat{\Phi}}||X(\lfloor uT\rfloor) - \widehat{\Phi}U(\lfloor uT\rfloor)||_2^2.$$

To be able to use the results from the calculus of variations theory, some smoothness assumptions are necessary. in particular b must be sufficiently smooth (differentiable) and the Lagrangian term must be a smooth function of b, $\dot{b}$ and of $\ddot{b}$, which is obviously the case here since we deal with a quadratic function of b, $\dot{b}$ and of $\ddot{b}$. The smoothness of b can be controlled in simulation ($\mathcal{C}^2$, $\mathcal{C}^3$ etc.). In practice, however, this might be violated but we assume it is the case since we are concerned with locally stationary processes. The above conditions are necessary for deriving the Euler-Lagrange equation.

7.3 Parameter and connectivity matrices

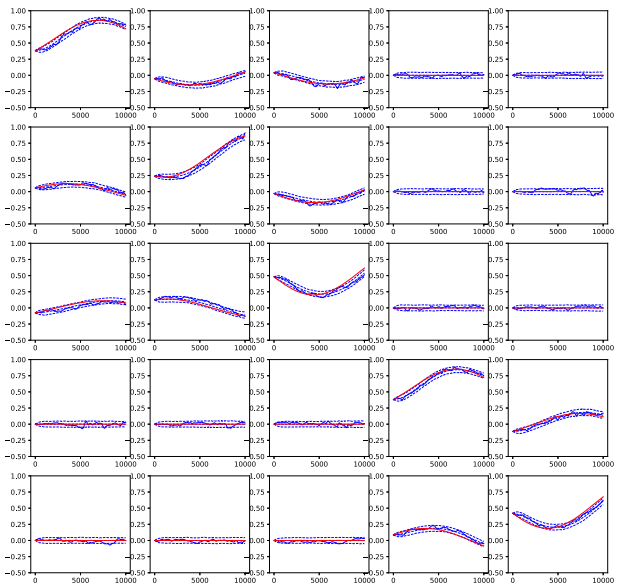

Figure 24. SOPE results for tv-VAR model: the plot in i^{th} row and j^{th} column represent the results for $\left(\Phi_1(t)\right)_{i,j}$ (red line) and $\left(\widehat{\Phi}_1(t)\right)_{i,j}$ (blue line), the 2.5, 50 and 97.5 percentiles (based on $B = 1000$ samples) are represented in dotted blue lines.

Figure 25. SOPE results for tv-VAR model: the plot in i^{th} row and j^{th} column represent the connectivity estimates between tetrode i and j. Solid lines represent the true quantity and dotted lines represent the estimates (for Delta band).

7.4 Connectivity networks prior to and post odor presentation

Connectivity networks for 0.5 and 0.95 quantiles:

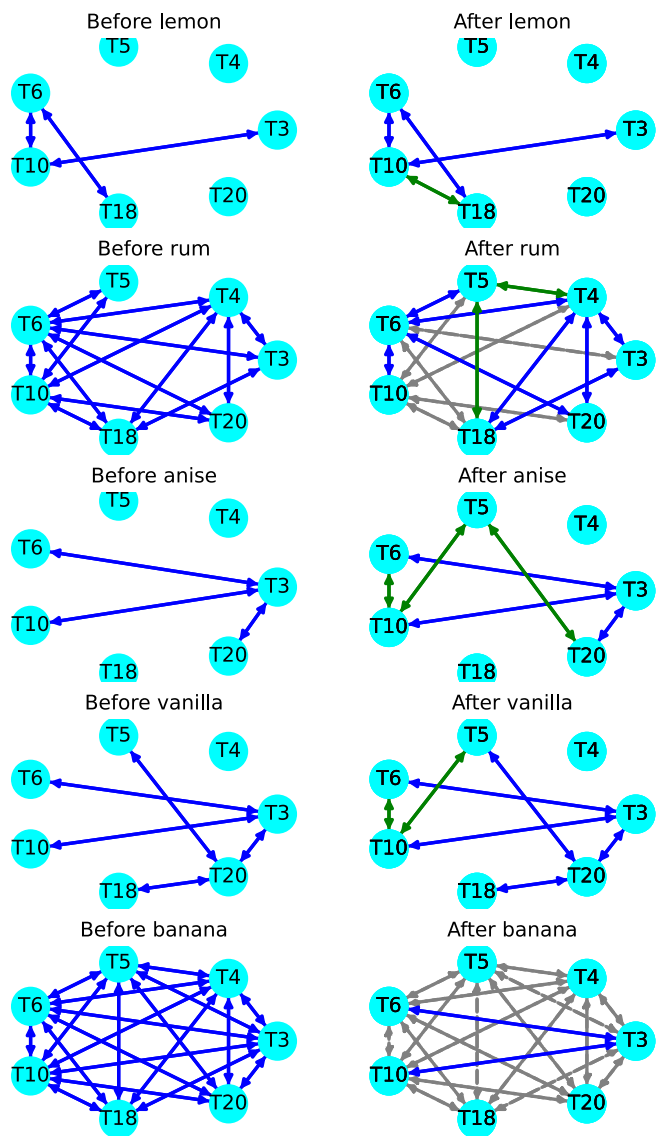

Figure 26. In the before-odor networks an edge is present if the high frequency coherence connectivity is higher than some threshold as defined by the 0.5 quantile. On the post odor presentation networks (based on the realizations at the start of the epoch), a blue edge indicates that this connectivity was high prior to and it remained post odor presentation, a dotted line gray edge indicates that this connectivity was high before and it is no longer the case and finally a green edge indicates that this connectivity was low before and it became high after the introduction of the smell.

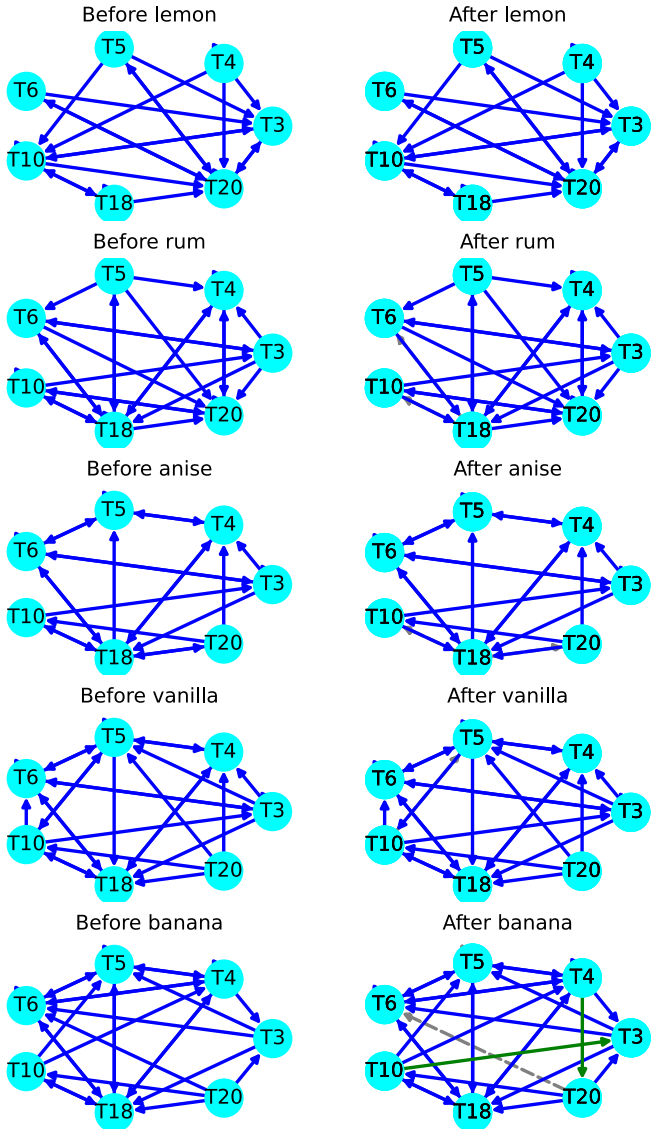

Figure 27. In the before-odor networks an edge is present if the high frequency PDC connectivity is higher than some threshold as defined by the 0.5 quantile. On the post odor presentation networks (based on the realizations at the start of the epoch), a blue edge indicates that this connectivity was high prior to and it remained post odor presentation, a dotted line gray edge indicates that this connectivity was high before and it is no longer the case and finally a green edge indicates that this connectivity was low before and it became high after the introduction of the smell.

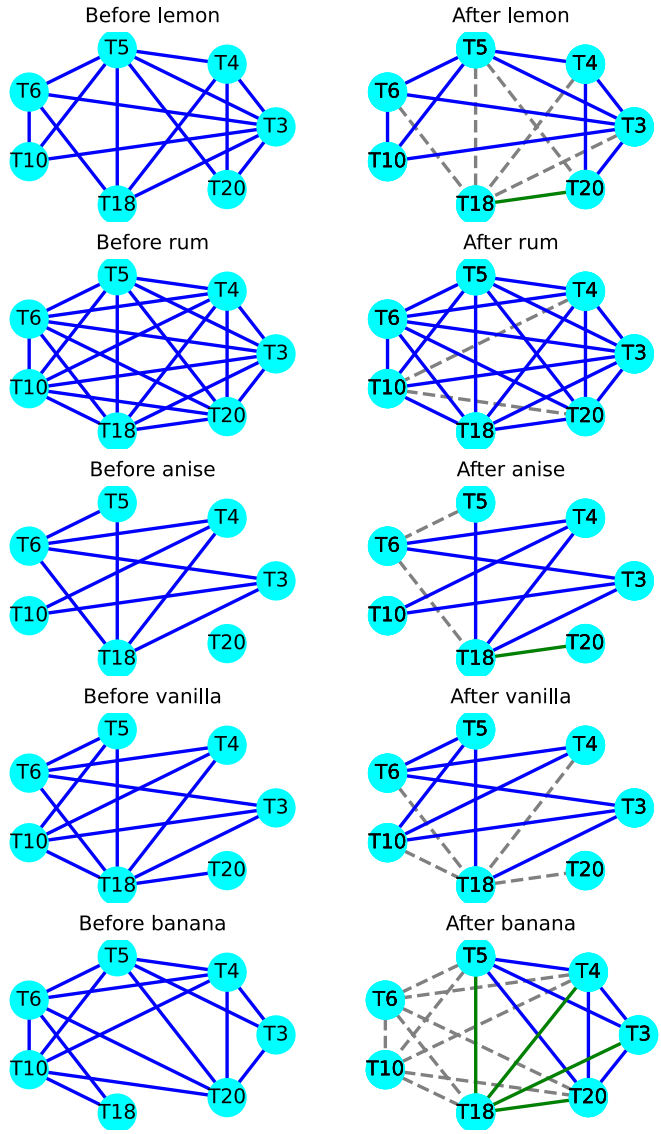

Figure 28. In the before-odor networks an edge is present if the high frequency coherence connectivity is higher than some threshold as defined by the 0.5 quantile. On the post odor presentation networks (based on the realizations at the start of the epoch), a blue edge indicates that this connectivity was high prior to and it remained post odor presentation, a dotted line gray edge indicates that this connectivity was high before and it is no longer the case and finally a green edge indicates that this connectivity was low before and it became high after the introduction of the smell.

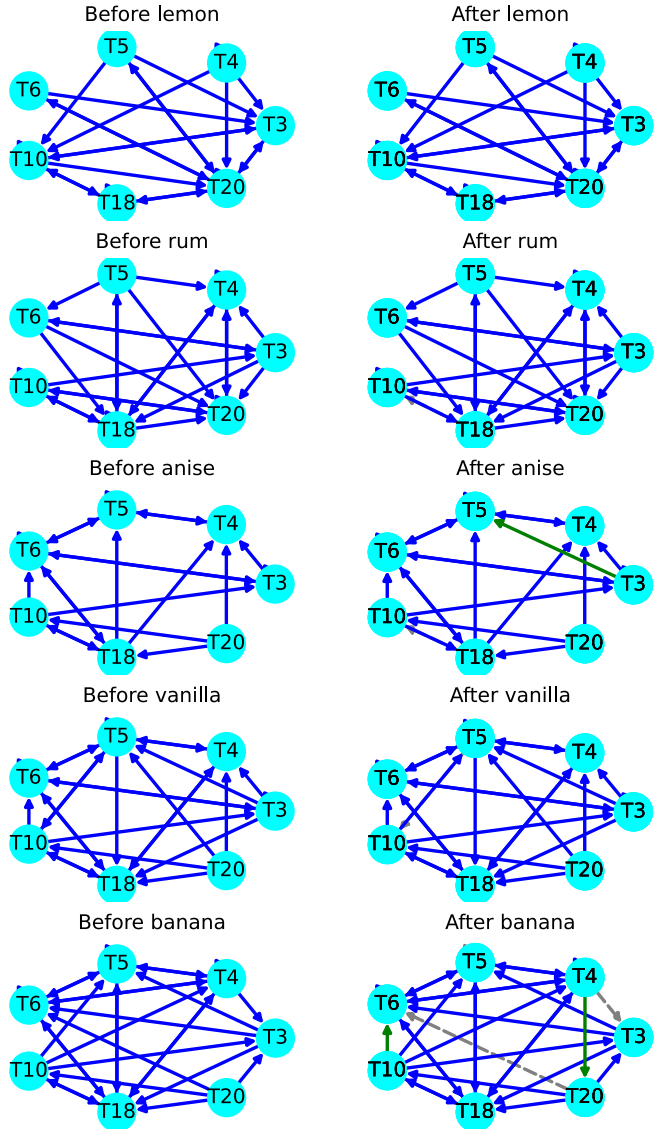

Figure 29. In the before-odor networks an edge is present if the low frequency PDC connectivity is higher than some threshold as defined by the 0.5 quantile. On the post odor presentation networks (based on the realizations at the start of the epoch), a blue edge indicates that this connectivity was high prior to and it remained post odor presentation, a dotted line gray edge indicates that this connectivity was high before and it is no longer the case and finally a green edge indicates that this connectivity was low before and it became high after the introduction of the smell.

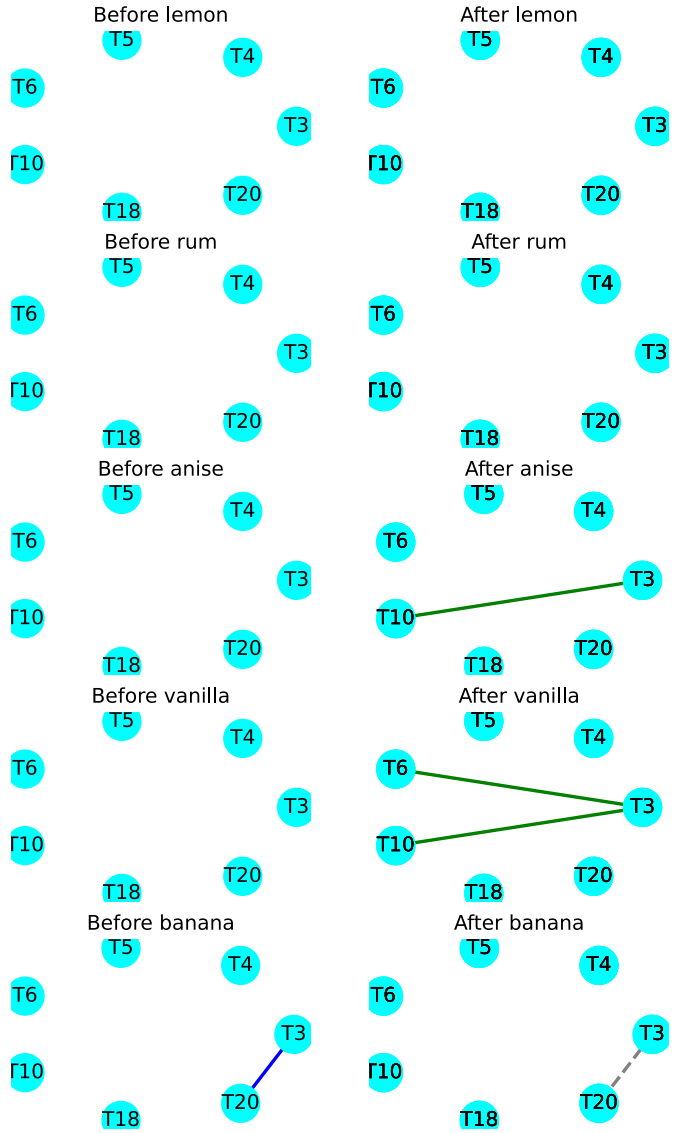

Figure 30. In the prior to odor presentation networks (based on the realizations at the start of the epoch) an edge is present if the high frequency coherence connectivity is higher than some threshold as defined by the 0.9 quantile. On the post odor presentation networks (based on the realizations at the start of the epoch), a blue edge indicates that this connectivity was high prior to and it remained post odor presentation, a dotted line gray edge indicates that this connectivity was high before and it is no longer the case and finally a green edge indicates that this connectivity was low before and it became high after the introduction of the smell.

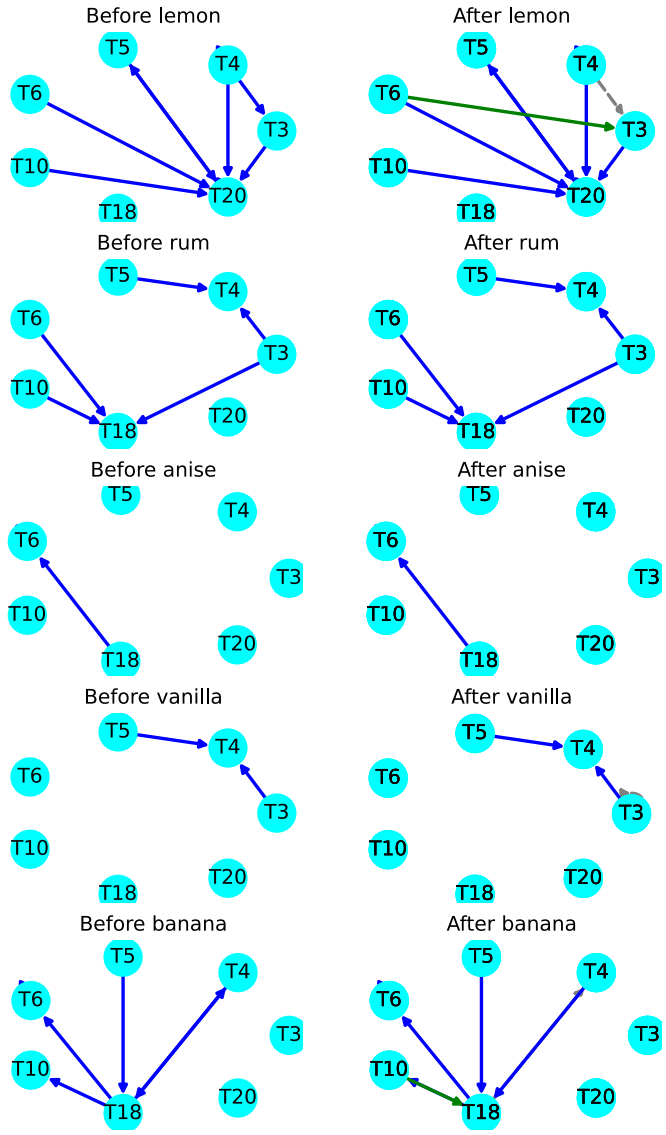

Figure 31. In the prior to odor presentation networks (based on the realizations at the start of the epoch) an edge is present if the high frequency PDC connectivity is higher than some threshold as defined by the 0.9 quantile. On the post odor presentation networks (based on the realizations at the start of the epoch), a blue edge indicates that this connectivity was high prior to and it remained post odor presentation, a dotted line gray edge indicates that this connectivity was high before and it is no longer the case and finally a green edge indicates that this connectivity was low before and it became high after the introduction of the smell.

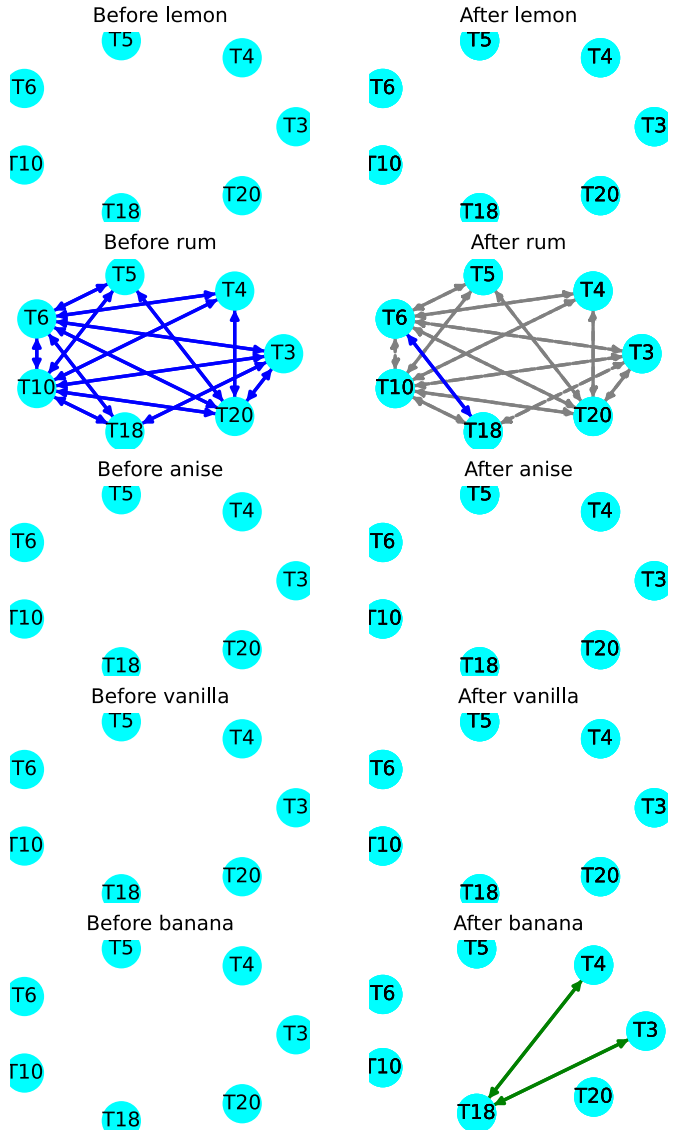

Figure 32. In the prior to odor presentation networks (based on the realizations at the start of the epoch) an edge is present if the high frequency coherence connectivity is higher than some threshold as defined by the 0.9 quantile. On the post odor presentation networks (based on the realizations at the start of the epoch), a blue edge indicates that this connectivity was high prior to and it remained post odor presentation, a dotted line gray edge indicates that this connectivity was high before and it is no longer the case and finally a green edge indicates that this connectivity was low before and it became high after the introduction of the smell.

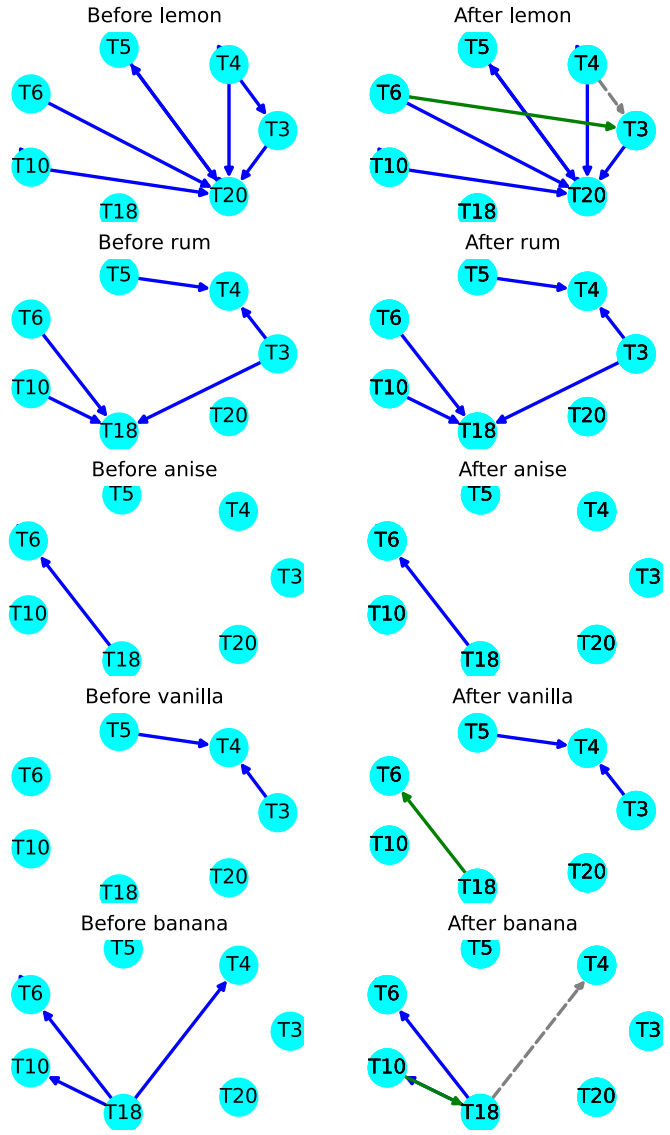

Figure 33. In the prior to odor presentation networks (based on the realizations at the start of the epoch) an edge is present if the low frequency PDC connectivity is higher than some threshold as defined by the 0.9 quantile. On the post odor presentation networks (based on the realizations at the start of the epoch), a blue edge indicates that this connectivity was high prior to and it remained post odor presentation, a dotted line gray edge indicates that this connectivity was high before and it is no longer the case and finally a green edge indicates that this connectivity was low before and it became high after the introduction of the smell.

Received 26 March 2021

REFERENCES

Allen, T. A., Salz, D. M., McKenzie, S. and Fortin, N. J. (2016). Nonspatial Sequence Coding in CA1 Neurons. *Journal of Neuroscience* **36** 1547-1563.

Baccala, L. and Sameshima, K. (2001). Partial directed coherence: A new concept in neural structure determination. *Biological Cybernetics* **84** 463-474.

Basu, S., Li, X. and Michailidis, G. (2019). Low Rank and Structured Modeling of High-dimensional Vector Autoregressions. *IEEE Transactions on Signal Processing* **67** 1207-1222. MR3925874

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Annals of Statistics* **43** 1535-1567. MR3357870

Brockwell, P. J. and Davis, R. A. (1986). *Time Series: Theory and Methods*. Springer-Verlag. MR0868859

Bruce, S. A., Hall, M. H., Buysse, D. J. and Krafty, R. T. (2018). Conditional adaptive Bayesian spectral analysis of nonstationary biomedical time series. *Biometrics* **74** 260-269. MR3777946

Cole, M. W., Bagic, A., Kass, R. and Schneider, W. (2010). Prefrontal Dynamics Underlying Rapid Instructed Task Learning Reverse with Practice. *Journal of Neuroscience* **30** 14245–14254.

Cribben, I., Haraldsdottir, R., Atlas, L. Y., Wager, T. D. and Lindquist, M. A. (2012). Dynamic connectivity regression: determining state-related changes in brain connectivity. *Neuroimage* **61** 907-920.

Dahlhaus, R. (1997). Fitting Time Series Models to Nonstationary Processes. *The Annals of Statistics* **25** 1-37. MR1429916

Dahlhaus, R. (2012). Locally Stationary Processes. *Handbook of Statistics* **30** 351-413.

Dahlhaus, R. and Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes. *The Annals of Statistics* **34** 1075-1114. MR2278352

Dahlhaus, R. and Subba Rao, S. (2007). A recursive online algorithm for the estimation of time-varying ARCH parameters. *Bernoulli Society for Mathematical Statistics and Probability* **13** 389-422. MR2331257

Emami, M. and Taban, M. (2018). A novel intelligent adaptive Kalman Filter for estimating the Submarine's velocity: With experimental evaluation. *Ocean Engineering* **158** 403-411.

Fiecas, M. and Ombao, H. (2011). The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. *The Annals of Applied Statistics* **5** 1102-1125. MR2840188

Fiecas, M. and Ombao, H. (2016). Modeling the Evolution of Dynamic Brain Processes During an Associative Learning Experiment. *Journal of the American Statistical Association* **111** 1440-1453. MR3601700

Fiecas, M., Ombao, H., Linkletter, C., Thompson, W. and Sanes, J. (2010). Functional connectivity: shrinkage estimation and randomization test. *Neuroimage* **4** 3005-3014.

Friston, K. J. (2011). Functional and Effective Connectivity: A Review. *Brain Connectivity* **1** 13-36.

Gelfand, I. M. and Fomin, S. V. (1964). *Calculus of Variations*. Prentice-Hall.

Gorrostieta, C., Ombao, H., Bédard, P. and Sanes, J. N. (2012). Investigating brain connectivity using mixed effects vector autoregressive models. *NeuroImage* **59** 3347-3355.

Granados-Garcia, G., Fiecas, M., Babak, S., Fortin, N. J. and Ombao, H. (2021). Brain waves analysis via a non-parametric Bayesian mixture of autoregressive kernels. *Computational Statistics & Data Analysis* 107409.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* **37** 424-438.

Hager, W. W. (1989). Updating the Inverse of a Matrix. *SIAM Review* **31** 221-239. MR0997457

Hu, L., Fortin, N. J. and Ombao, H. (2019). Modeling High-Dimensional Multichannel Brain Signals. *Statistics in Biosciences* **11** 91–126.

Hu, L., Guindani, M., Fortin, N. J. and Ombao, H. (2020). A hierarchical bayesian model for differential connectivity in multi-trial brain signals. *Econometrics and Statistics* **15** 117-135. MR4124739

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME–Journal of Basic Engineering* **82** 35-45. MR3931993

Kolmogorov, A. N. (1939). Sur l'interpolation et l'extrapolation des suites stationnaires. *C.R. de l'Acad. Sci., Paris* **208**.

Li, L., Pluta, D., Shahbaba, B., Fortin, N., Ombao, H. and Baldi, P. (2019). Modeling Dynamic Functional Connectivity with Latent Factor Gaussian Processes. 8263-8273.

Li, Z., Rosen, O., Ferrarelli, F. and Krafty, R. T. (2021). Adaptive Bayesian Spectral Analysis of High-Dimensional Nonstationary Time Series. *Journal of Computational and Graphical Statistics* **0** 1-14. MR4313477

Lindquist, M. A. (2008). The Statistical Analysis of fMRI Data. *Statistical Science* **23** 439-464. MR2530545

Lippuner, D. and Moschytz, G. (2004). The Kalman filter in the context of adaptive filter theory. *International Journal of Circuit Theory and Applications* **32** 223-253.

Lütkepohl, H. (1991). *New Introduction to Multiple Time Series Analysis*. Springer. MR1125072

Mehra, R. K. (1972). Approaches to Adaptive Filtering. *IEEE Transactions on Automatic Control* **AC-17** 693-698.

Nason, G. P., Von Sachs, R. and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society Series B* **62** 271-292. MR1749539

Nuno, R. A. G., Chung, C. H. R. and Maharatna, K. (2021). Hardware architecture for real-time EEG-based functional brain connectivity parameter extraction. *Journal of Neural Engineering* **18**.

Ombao, H. and Bellegem, S. V. (2008). Evolutionary Coherence of Nonstationary Signals. *IEEE Transactions on Signal Processing* **56** 2259-2266. MR2516630

Ombao, H., Sachs, R. and Guo, W. (2005). SLEX Analysis of Multivariate Nonstationary Time Series. *Journal of the American Statistical Association* **100** 519-531. MR2160556

Ombao, H., Fiecas, M., Ting, C.-M. and Low, Y. F. (2018). Statistical models for brain signals with properties that evolve across trials. *NeuroImage* **180** 609-618.

Pagnotta, M. F. and Plomp, G. (2018). Time-varying MVAR algorithms for directed connectivity analysis: Critical comparison in simulations and benchmark EEG data. *PloS one* **13** e0198846.

Park, T., Eckley, I. A. and Ombao, H. C. (2014). Estimating Time-Evolving Partial Coherence Between Signals via Multivariate Locally Stationary Wavelet Processes. *IEEE Transactions on Signal Processing* **62** 5240-5250. MR3268108

Plackett, R. L. (1950). Some theorems in least squares. *Biometrika* **37 1-2** 149–57. MR0036980

Priestley, M. B. (1965). Evolutionary Spectra and Non-Stationary Processes. *Journal of the Royal Statistical Society* **27** 204-237. MR0199886

Saito, Y., Harashima, H., Yamaguchi, N. and Fujisawa, K. (1981). Recent advances in EEG and EMG data processing. *Chap. Tracking of information within multichannel EEG record-causal analysis in EEG* 133–146.

Samdin, S., Ting, C.-M., Ombao, H. and Salleh, S. (2016). A Unified Estimation Framework for State-Related Changes in Effective Brain Connectivity. *IEEE Transactions on Biomedical Engineering* **64** 844 – 858.

Shappell, H., Caffo, B. S., Pekar, J. J. and Lindquist, M. A. (2019). Improved state change estimation in dynamic functional connectivity using hidden semi-Markov models. *NeuroImage* **191** 243-257.

Shumway, R. H. and Stoffer, D. S. (2005). *Time Series Analysis and Its Applications*. Springer-Verlag. MR2228626

Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica* **48** 1-48.

Sorenson, H. W. (1970). Least-squares estimation: from Gauss to Kalman. *IEEE Spectrum* **7** 63-68.

Stigler, S. M. (1981). Gauss and the Invention of Least Squares. *The Annals of Statistics* **9** 465-474. MR0615423

Temkin, O. (1947). GALL AND THE PHRENOLOGICAL MOVEMENT. *Bulletin of the History of Medicine* **21** 275-321.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society* **58** 267-288. MR1379242

Ting, C.-M., Ombao, H., Samdin, S. B. and Salleh, S. H. (2018). Estimating Dynamic Connectivity States in fMRI Using Regime-Switching Factor Models. *IEEE Transactions on Medical Imaging* **37** 1011-1023.

Ting, C.-M., Samdin, S., Tang, M. and Ombao, H. (2020). Detecting Dynamic Community Structure in Functional Brain Networks Across Individuals: A Multilayer Approach. *IEEE Transactions on Medical Imaging* 1-1.

Wang, Y., Ting, C.-M. and Ombao, H. (2016). Modeling Effective Connectivity in High-Dimensional Cortical Source Signals. *IEEE Journal of Selected Topics in Signal Processing* **10** 1315-1325.

Wiener, N. (1949). *The Extrapolation, Interpolation rand Smoothing of Stationary Time Series*. Wiley. MR0031213

Xu, Y. and Lindquist, M. A. (2015). Dynamic connectivity detection: an algorithm for determining functional connectivity change points in fMRI data. *Frontiers in neuroscience* **9** 285.

Young, P. (2011). *Recursive Approaches to Time Series Analysis*. Springer. MR3024956

Yu, Z., Prado, R., Quinlan, E. B., Cramer, S. C. and Ombao, H. (2016). Understanding the Impact of Stroke on Brain Motor Function: A Hierarchical Bayesian Approach. *Journal of the American Statistical Association* **111** 549-563. MR3538686

Zhao, W. and Prado, R. (2020). Efficient Bayesian PARCOR Approaches for Dynamic Modeling of Multivariate Time Series. *Journal of Time Series Analysis* **41** 759–784. MR4176359

Zhou, B., Moorman, D. E., Behseta, S., Ombao, H. and Shahbaba, B. (2016). A Dynamic Bayesian Model for Characterizing Cross-Neuronal Interactions During Decision-Making. *Journal of the American Statistical Association* **111** 459-471. MR3538679

Anass El Yaagoubi Bourakna
Statistics Program
King Abdullah University of Science and Technology
E-mail address: anass.bourakna@kaust.edu.sa

Marco Pinto
Oslo Metropolitan University
E-mail address: Marco.Pinto@oslomet.no

Norbert Fortin
University of California Irvine
E-mail address: norbert.fortin@uci.edu

Hernando Ombao
Statistics Program
King Abdullah University of Science and Technology
E-mail address: hernando.ombao@kaust.edu.sa