Supplementary Material:
Sevoflurane Alters Spatiotemporal Functional Connectivity Motifs That Link Resting-State Networks During Wakefulness

MohammadMehdi Kafashan,1, ShiNung Ching,1,2 and Ben Julian A. Palanca3*

1 Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO, USA, 2 Division of Biology and Biomedical Science, Washington University in St. Louis, St. Louis, MO, USA, 3 Department of Anesthesiology, Washington University School of Medicine in St. Louis, St. Louis, MO, USA

Correspondence*:
Ben Julian A. Palanca
Department of Anesthesiology, Washington University School of Medicine in St. Louis, 660 S Euclid Ave, St. Louis, MO 63110, US, palancab@wustl.edu

1 DYNAMIC CORRELATION MODEL

Let \( z_k, k = 1, 2, \ldots \) be a vector time series of correlation values between pairs of brain regions. We assume that \( z_k \) arises from an underlying state \( x_k \in \mathbb{R}^N \) via

\[
 z_k = F^{-1}(x_k),
\]

where the function \( F^{-1} \) is vector-valued inverse Fisher transform which can be written as

\[
 F^{-1}\left( \begin{bmatrix} x^1 \\ \vdots \\ x^N \end{bmatrix} \right) = \begin{bmatrix} \tanh(x^1) \\ \vdots \\ \tanh(x^N) \end{bmatrix}.
\]

Such a function bounds the state within \([-1, 1]\) and is the standard transformation of the cross-correlation into a normally distributed variable. We assume that \( x_k \) obeys a standard, linear state space model of the
form

\[ x_k = A_k x_{k-1} + w_{k-1}, \]
\[ y_k = F^{-1}(C_k x_k + v_k). \]

(3)

Here, the state noise process \( w_k \in \mathbb{R}^N \) is an \( N \)-dimensional random vector with multivariate Gaussian distribution having zero mean and covariance matrix \( Q_w \in \mathbb{R}^{N \times N} \), the noise vector \( v_k \in \mathbb{R}^N \) is a zero-mean multivariate Gaussian random vector with covariance matrix \( R_v \in \mathbb{R}^{N \times N} \). Bounded nature of correlation coefficients result in a non-Gaussian noise for the observation vector \( y_k \). The covariance \( Q_w \) determines the extent to which correlation can change in successive measurements, while \( R_v \) simply characterizes measurement noise. In this sense, the vector \( y_k \) consists of noisy observations of pairwise correlation values. We assume that state noise and measurement noise are uncorrelated.

We seek the optimal filter for obtaining the state estimate \( \hat{z}_k \) in the sense of minimum mean-squared error (MMSE), i.e.,

\[ \min_{\hat{z}_k} \mathbb{E} \left[ \|\hat{z}_k - z_k\|_2^2 \right]. \]

(4)

Central to this problem is the calculation of the probability density function (p.d.f.) of the state vector at any given time \( k \), conditioned on \( Y_k = \{y_0, y_1, \ldots, y_k\} \), the set of all the past observations. The presence of the nonlinearity (2) complicates this calculation, but only slightly since \( F \) is smooth and invertible. Thus, it follows immediately that we can obtain a surrogate measurement

\[ d_k = F(y_k), \]

(5)

such that \( d_k \) is linear in the state \( x_k \). Note that (1) directly yields the p.d.f. of \( z_k \) from that of \( x_k \). Thus, the problem (4) reduces to the classical Kalman filter to obtain the (Gaussian) p.d.f. of \( x_k \) given \( Y_k \) (Kalman, 1960b,a), based on the measurements \( d_k \).

It follows directly from (3)-(5) that the p.d.f. \( p(y_k | x_k) \) can be written as

\[ p(y_k | x_k) = \frac{1}{|J_k|} p(v_k = F(y_k) - C_k x_k), \]

(6)

where \( p(v_k = F(y_k) - C_k x_k) \) is p.d.f. of \( v_k \), a zero-mean multivariate Gaussian with covariance matrix \( R_v \). A standard Bayesian approach thus yields the posterior density \( p(x_k | Y_k) \), which is Gaussian with
covariance matrix ($\Sigma$) and mean ($\mu$) as
\[
\Sigma = \left(\begin{array}{c}
P_{k|k-1}^{-1} + C_k^T R_v^{-1} C_k
\end{array}\right)^{-1},
\]
\[
\mu = \Sigma \times \left(\begin{array}{c}
P_{k|k-1}^{-1} \hat{x}_{k|k-1} + C_k^T R_v^{-1} F(y_k)
\end{array}\right).
\]
(7)

The Kalman update equations are thus
\[
\hat{x}_{k|k-1} = A_k \hat{x}_{k-1|k-1},
\]
\[
P_{k|k-1} = A_k \tilde{P}_{k-1|k-1} A_k^T + Q_w,
\]
\[
\hat{x}_{k|k} = \left(\begin{array}{c}
P_{k|k-1}^{-1} + C_k^T R_v^{-1} C_k
\end{array}\right)^{-1}
\]
\[
\times \left(\begin{array}{c}
P_{k|k-1}^{-1} \hat{x}_{k|k-1} + C_k^T R_v^{-1} F(y_k)
\end{array}\right),
\]
\[
P_{k|k} = \left(\begin{array}{c}
P_{k|k-1}^{-1} + C_k^T R_v^{-1} C_k
\end{array}\right)^{-1}.
\]
(8)

In the subsequent results, we assume an initial multivariate Gaussian prior $x_{0|-1}$ with mean $\hat{x}_{0|-1}$ and covariance of $P_{0|0}$. We note that, for the problem (4), (8) returns the optimal estimate. After obtaining state estimate $\hat{x}_{k|k}$, the correlation coefficients $\hat{z}_k$ are approximated using fourth order Taylor expansion of (1) around $\hat{x}_{k|k}$ as
\[
\hat{z}_k = \tanh(\hat{x}_{k|k}) - \tanh(\hat{x}_{k|k}) \times \text{sech}^2(\hat{x}_{k|k}) \times \text{diag}(P_{k|k}) +
\]
\[
\frac{1}{24}(22 \tanh(\hat{x}_{k|k}) \times \text{sech}^4(\hat{x}_{k|k}) - 2 \sinh(3\hat{x}_{k|k}) \times \text{sech}^5(\hat{x}_{k|k})) \times \text{diag}(P_{k|k}),
\]
(9)

where all the multiplication and powers of vectors are element-wise operations. Also, diag() operator applies to a square matrix and returns a vector containing the diagonal elements of the matrix.

A final obstacle in using the filter equations (8) is knowing the noise parameters that specify the state-space model. Since such information is difficult to define a priori, a principled way to update them alongside the state estimates is required. Here, we face the challenge of estimating noise parameters alongside our online filtering. To address this issue we use an EM algorithm (Ghahramani and Hinton, 1996; Shumway and Stoffer, 2010) to jointly estimate the state and model parameters from observed correlation data. Assuming diagonal covariance matrices for model parameters, we can decouple the update procedures for each edge, resulting in a reduction of the computational burden typically associated with EM methods. Briefly speaking, after initializing the model parameter (process and measurement noise), we perform the E-step, filtering and smoothing, and then updating the parameter in the M-step. The E and M steps alternate
to convergence. In order to obviate computational issues associated with this algorithm, a maximum iteration number is set alongside the convergence criterion.

2 DETERMINING THE NUMBER OF MOTIFS USING AIC

Selecting the number of clusters, K, in unsupervised learning is one of the key problems. The correct choice of K is often ambiguous. There exist several heuristic methods to facilitate the selection of k such as "The Elbow Method" by (Thorndike, 1953) and a technique based on the use of hierarchical clustering by (Salimi-Khorshidi et al., 2011). Another set of approaches for determining k are information criteria methods. In this paper, we adopt the Akaike information criterion (AIC) (Akaike, 1974) approach. AIC rewards goodness of fit, assessed by the likelihood function, but it also includes a penalty that is an increasing function of K. Figure 1 shows AIC versus K for both 0% and 1.2% vol sevoflurane. In this plot, AIC values are normalized to have maximum value of unity for each condition. It can be seen from this figure that K = 8 is the optimal choice of K for both condition.
Figure 2: (A) Spatial motifs and (B) temporal correlation trajectories associated with each cluster for synthesized data. Green line shows the average correlation over time and space for each cluster.

3 SPATIOTEMPORAL ANALYSES OF SYNTHESIZED DATA

Here, we demonstrate the efficacy and utility of the approach in a simple example of a data set with 9 locations and 160 time windows. We assume there exists 5 spatial motifs shown in Figure 2A and each motif has the same temporal correlation trajectory with the same mean over time (Green lines shown in Figure 2B). Before decomposing the correlation trajectories, we added noise drawn from uniform distribution inside interval $(-0.1, 0.1)$ to all region pair correlations over time. Figure 2 shows spatiotemporal decomposition of correlation time series. Spatial and temporal motifs are shown in Figures 2A and 2B, respectively. It can be conclude from these plots that temporal information can play an important role for extracting meaningful motifs even though if all the region pairs have the same static correlation value.

4 ROBUSTNESS AS A FUNCTION OF INTERSUBJECT VARIABILITY

To see how consistent the spatiotemporal motifs are for different conditions (0% and 1.2%), we look at how spatial motifs at 0% and 1.2% obtained from each individual, without concatenating data from all subjects for each condition, are similar. Figure 3A demonstrates the average cardinality of each motifs over different individual. Error bars stand for SEM of the average cardinality. Figure 3B represents average similarity...
Figure 3. (A) Bar plot of the cardinality percentage of each motif for both condition (0% and 1.2%) for a fixed window size, 11 seconds, over different individuals. Error bars stand for SEM. Average similarity (B) and SEM of average similarity (C) between motifs over different individuals, by computing the correlation of spatial component of each motif.

between motifs at 0% and 1.2% obtained from each individual for 11 seconds window size. Standard error of the mean of the average similarity is shown in Figure 3C. We can see from this figures that the similarity between first, second and eighth motifs are robust over different individual.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Ghahramani, Z. and Hinton, G. E. (1996). Parameter estimation for linear dynamical systems. Technical report, Technical Report CRG-TR-96-2, University of Toronto, Dept. of Computer Science.

Kalman, R. E. (1960a). Contributions to the theory of optimal control. *Bol. Soc. Mat. Mexicana*, 5(2):102–119.

Kalman, R. E. (1960b). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45.

Salimi-Khorshidi, G., Nichols, T. E., Smith, S. M., and Woolrich, M. W. (2011). Using gaussian-process regression for meta-analytic neuroimaging inference based on sparse observations. *Medical Imaging, IEEE Transactions on*, 30(7):1401–1416.

Shumway, R. H. and Stoffer, D. S. (2010). *Time series analysis and its applications: with R examples*. Springer Science & Business Media.

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, 18(4):267–276.