MULTI-KERNEL CHANGE DETECTION FOR DYNAMIC FUNCTIONAL CONNECTIVITY GRAPHS

Georgios Vasileios Karanikolas,[1] *Olaf Sporns,*[2] *and Georgios B. Giannakis* [1]

[1] Dept. of ECE and DTC, University of Minnesota, USA   Emails: {karan029,georgios}@umn.edu
[2] Dept. of Psychological and Brain Sciences, Indiana University, USA   Email: osporns@indiana.edu

ABSTRACT

Dynamic functional connectivity (dFC) analyses of fMRI time-courses are typically performed using sliding-window based schemes. Such approaches not only inherently confine analysis to a single time-scale, but also do not generally lend themselves to accurate change-time estimates of the dynamically evolving graph topology. Change point detection methods on the other hand, offer the potential to overcome both limitations. However, the approaches employed so far in the dFC context are limited to detecting changes in linear relationships among time-courses corresponding to distinct regions of the brain. The present work puts forth a novel multi-kernel change point detection approach with the goal of capturing changes in the generally nonlinear relationships among time-courses, and thus in the topologies of the corresponding dynamically evolving FC graphs. The approach is tested on dynamic causal model (DCM) based synthetic resting-state fMRI data.

Index Terms— fMRI, change detection, kernel-based regression, multiple kernel learning

1. INTRODUCTION

Functional magnetic resonance imaging (fMRI) is an approach to assessing brain activity that has provided valuable insights with regards to brain functionality [7]. The focus of many contemporary fMRI studies has been on the so-termed functional connectivity networks. These networks feature brain regions as nodes and provide a representation of the statistical dependencies between them [16].

The assumption that dependencies remain static for the duration of the scan has remained prevalent until recently. Dynamic (d)FC, however, challenges this assumption allowing for time-varying dependencies [8].

Most dFC studies, though, rely on sliding-window based approaches [8]. Their adherence to a single time scale as well as their difficulty to coping with piecewise static setups and the corresponding detection of change points renders them inherently unsuitable for locating changes in the interregional dependencies.

Change detection approaches on the other hand can potentially overcome this limitation. Existing methods in the context of dFC, however, rely on linear models and hence cannot generally capture changes in nonlinear dependencies. These methods include a graphical lasso based approach [4] and one based on Pearson correlation coefficients [10].

Further impetus for considering nonlinear dependencies comes from observations suggesting that the relationship between the blood-oxygen-level-dependent (BOLD) response and the underlying neural activity may be nonlinear [2].

In this work, a kernel-based nonlinear regression approach for change detection is put forth, with the goal of detecting changes in the nonlinear dependencies between brain regions, in addition to linear ones. The choice of the kernel that markedly influences the effectiveness of any kernel-based method, is performed in a data driven fashion via multi-kernel learning. A pseudolikelihood based test statistic for detecting the presence and estimating the location of a single change point (CP) is then developed. A variant of the binary segmentation method [17], that makes use of the aforementioned statistic, is employed when the number of change points is unknown.

2. NONLINEAR MODELS FOR CHANGE DETECTION

Consider a network consisting of $|\mathcal{V}|$ nodes, with $\mathcal{V}$ denoting the corresponding set. Nodes, in our context, correspond to regions of the brain. Hereafter, we will thus use the terms node and region interchangeably. Let furthermore $\mathbf{x}_i := [x_i[1], \ldots, x_i[T]]^\top$ denote the time course over T slots corresponding to the i-th region ($\top$ stands for transposition). Moreover, define $\boldsymbol{\chi}[t] := [x_1[t], \ldots, x_{|\mathcal{V}|}[t]]^\top$ as the vector gathering the measurements at all nodes at time t, and let $\boldsymbol{\chi}_{\setminus i}[t]$ comprise the measurements at all nodes, except node i, for the same time instance. Finally, $\bar{\boldsymbol{\chi}}_{\setminus i}[t] := [\boldsymbol{\chi}_{\setminus i}^\top[t], \boldsymbol{\chi}^\top[t-1], \ldots, \boldsymbol{\chi}^\top[t-d]]^\top$ augments $\boldsymbol{\chi}_{\setminus i}[t]$ with d past network snapshots.

As a starting point, consider that the interregional depen-

This work was supported by NSF 1500713, 1514056, and NIH 1R01GM104975-01.

dencies are piecewise time invariant. In particular, they are assumed to remain unchanged for the duration of each segment. The time interval $[\tau_{n-1}, \tau_n - 1]$ is defined as the n-th segment. Furthermore, the dependencies for the n-th segment differ from those corresponding to segments $n-1$ and $n+1$. The beginning of each segment, let τ_{n-1}, hence, marks a change in the aforementioned dependencies and is thus called a change point. Hereafter, we will consider that the data is divided into N segments, and we will estimate the locations of the change points $\{\tau_n\}_{n=1}^{N-1}$, as well as N. By convention, we also have $\tau_0 = 1$ and $\tau_N = T$.

2.1. Kernel-based estimators

Focusing on the n-th segment and i-th node, let $f_i^{(n)}$ be a nonlinear function capturing the relation of $x_i[t]$ with $\bar{\boldsymbol{\chi}}_{\setminus i}[t]$. Note that a FC graph is simply a representation of the dependencies of each node on the rest of the nodes. The collection of functions $f_1^{(n)}, \ldots, f_{|\mathcal{V}|}^{(n)}$ hence provides an implicit description of the FC graph topology for the n-th segment. The corresponding regression model for $t \in [\tau_{n-1}, \tau_n - 1]$ is given by

$$x_i[t] = f_i^{(n)}(\bar{\boldsymbol{\chi}}_{\setminus i}[t]) + \epsilon_i^{(n)}[t] \tag{1}$$

with $\epsilon_i^{(n)}[t]$ capturing noise and unmodeled effects.

In this work, a reproducing kernel Hilbert space (RKHS) formulation will be employed for modeling $f_i^{(n)}$. Once a symmetric and positive semidefinite function κ is selected to act as a similarity measure, it induces a space of functions of the following form [13]

$$\mathcal{H} := \{\ f : f(\bar{\boldsymbol{\chi}}_{\setminus i}[t]) = \sum_{\tau=1}^{\infty} \beta_\tau\ \kappa(\bar{\boldsymbol{\chi}}_{\setminus i}[t], \bar{\boldsymbol{\chi}}_{\setminus i}[\tau])\ \}. \tag{2}$$

Functions κ satisfying the above properties are known as kernels. A typical example of such a function is the Gaussian kernel, defined as $\kappa_{\mathrm{G}}(\boldsymbol{\chi}_1, \boldsymbol{\chi}_2) := e^{-\|\boldsymbol{\chi}_1 - \boldsymbol{\chi}_2\|_2^2/(2\rho^2)}$.

In order to choose a function $f_i^{(n)}$ that optimally fits the data we will use kernel ridge regression. In particular, the optimization problem considered is of the form

$$\hat{f}_i^{(n)} = \underset{f \in \mathcal{H}}{\arg\min} \sum_{t=\tau_{n-1}}^{\tau_n - 1} (x_i[t] - f(\bar{\boldsymbol{\chi}}_{\setminus i}[t]))^2 + \lambda_i^{(n)} \|f\|_{\mathcal{H}}^2 \tag{3}$$

where $\|\cdot\|_{\mathcal{H}}$ stands for the norm of $\mathcal{H}$, and $\lambda_i^{(n)}$ is a regularization parameter; note that the second summand controls the smoothness of the estimated function.

Resorting to the representer theorem [14], the form of an optimal solution to (3) is given by

$$\hat{f}_i^{(n)}(\bar{\boldsymbol{\chi}}_{\setminus i}[t]) = \sum_{\tau=\tau_{n-1}}^{\tau_n - 1} \beta_{i\tau}^{(n)}\ \kappa(\bar{\boldsymbol{\chi}}_{\setminus i}[t], \bar{\boldsymbol{\chi}}_{\setminus i}[\tau]). \tag{4}$$

Substituting (4) into (3) yields the final form of the optimization problem considered

$$\hat{\boldsymbol{\beta}}_i^{(n)} = \underset{\boldsymbol{\beta}_i^{(n)}}{\arg\min} \|\mathbf{x}_i^{(n)} - \mathbf{K}_{-i}^{(n)} \boldsymbol{\beta}_i^{(n)}\|^2 + \lambda_i^{(n)} {\boldsymbol{\beta}_i^{(n)}}^\top \mathbf{K}_{-i}^{(n)} \boldsymbol{\beta}_i^{(n)} \tag{5}$$

where $\mathbf{x}_i^{(n)} := [x_i[\tau_{n-1}], \ldots, x_i[\tau_n - 1]]^\top$ and $\boldsymbol{\beta}_i^{(n)} := [\beta_{i\tau_{n-1}}^{(n)}, \ldots, \beta_{i(\tau_n - 1)}^{(n)}]^\top$. Moreover, the kernel matrix entries are $[\mathbf{K}_{-i}^{(n)}]_{t\tau} := \kappa(\bar{\boldsymbol{\chi}}_{\setminus i}[t], \bar{\boldsymbol{\chi}}_{\setminus i}[\tau])$ for $t, \tau \in [\tau_{n-1}, \tau_n - 1]$.

The solution to (5) is obtained as $\hat{\boldsymbol{\beta}}_i^{(n)} = (\mathbf{K}_{-i}^{(n)} + \lambda_i^{(n)} \mathbf{I})^{-1} \mathbf{x}_i^{(n)}$ and after substituting back to (4), one obtains the optimal estimating function $\hat{f}_i^{(n)}$.

With $\{\hat{f}_i^{(n)}\}_{i=1}^{|\mathcal{V}|}$ obtained, the estimated residuals are given by $\hat{\epsilon}_i^{(n)}[t] = x_i[t] - \hat{f}_i^{(n)}(\bar{\boldsymbol{\chi}}_{\setminus i}[t])$ for $t \in [\tau_{n-1}, \tau_n - 1]$ and $i = 1, \ldots, |\mathcal{V}|$. These residuals will be used for detecting the presence and estimating the location of the change points, as it will become evident in the following sections.

2.2. Detecting a single change point

We will first consider the base case, where there exists at most one change point. Detecting the presence of a CP amounts to performing a hypothesis test of the following general form

$$\mathcal{H}_0 : \text{no CP in } [1, T]; \quad \mathcal{H}_1 : \text{a CP } \hat{\tau}_1 \text{ exists in } [1, T]. \tag{6}$$

Let now $\mathbf{f}^{(s)} := [f_1^{(s)} \ldots f_{|\mathcal{V}|}^{(s)}]^\top$ gather the functions describing the per-node dependencies for segment s, with $s \in \{0, 1, 2\}$; note that $s = 0$, by definition, corresponds to the entire data record, namely $t \in [1, \ldots, T]$. In our context, the sought after CP is indicative of a change to the aforementioned dependencies. The corresponding hypothesis test can thus be written as

$$\mathcal{H}_0 :\ \mathbf{f}^{(1)} = \mathbf{f}^{(2)} := \mathbf{f}^{(0)} \qquad \mathcal{H}_1 :\ \mathbf{f}^{(1)} \neq \mathbf{f}^{(2)} \tag{7}$$

According to $\mathcal{H}_1$ the dependencies, and hence the functions describing them, differ between the segments $[1 \ldots \tau_1 - 1]$ and $[\tau_1 \ldots T]$, for some τ_1 which will be estimated. In contrast, according to $\mathcal{H}_0$ no such τ_1 exists. Finally, note that $\{f_i^{(s)}\}$ belong to the general class of nonlinear functions, thereby allowing for capturing changes in nonlinear dependencies, in addition to linear ones.

As a first step in developing a test statistic for (7), consider the postulated (conditional) pseudolikelihood

$$p(\boldsymbol{\chi}[t] | \boldsymbol{\chi}[t-1], \ldots, \boldsymbol{\chi}[t-d]) = \prod_{i=1}^{|\mathcal{V}|} p(x_i[t] | \boldsymbol{\chi}_{\setminus i}[t], \boldsymbol{\chi}[t-1], \ldots, \boldsymbol{\chi}[t-d]). \tag{8}$$

Note that when $d = 0$ this form reduces to the regular pseudolikelihood [1]. If $\epsilon_i^{(n)}[t]$ are Gaussian and (approximately) uncorrelated, the pseudolikelihood (as per (8)) of the data $\mathbf{X} := [\mathbf{x}_1, \ldots, \mathbf{x}_{|\mathcal{V}|}]$ under the two hypotheses is given by

$$p(\mathbf{X};\mathcal{H}_0)=\prod_{i=1}^{|\mathcal{V}|}\prod_{t=1}^{T}(\sqrt{2\pi}\sigma_i^{(0)})^{-1}\exp\left(-\frac{{\epsilon_i^{(0)}}^2[t]}{2{\sigma_i^{(0)}}^2}\right) \tag{9a}$$

$$p(\mathbf{X};\tau_1,\mathcal{H}_1)=\prod_{i=1}^{|\mathcal{V}|}\prod_{t=1}^{\tau_1-1}(\sqrt{2\pi}\sigma_i^{(1,2)})^{-1}\exp\left(-\frac{{\epsilon_i^{(1)}}^2[t]}{2{\sigma_i^{(1,2)}}^2}\right)$$
$$\prod_{t=\tau_1}^{T}(\sqrt{2\pi}\sigma_i^{(1,2)})^{-1}\exp\left(-\frac{{\epsilon_i^{(2)}}^2[t]}{2{\sigma_i^{(1,2)}}^2}\right) \tag{9b}$$

where for $\mathcal{H}_0$, ${\sigma_i^{(0)}}^2 := \operatorname{var}(\epsilon_i^{(0)})$ and ${\sigma_i^{(1,2)}}^2$ is similarly defined for $\mathcal{H}_1$. Once the estimators $\hat{f}_1^{(s)}\dots\ \hat{f}_{|\mathcal{V}|}^{(s)}$ are obtained for $s\in\{0,1,2\}$, the corresponding estimated residual for the i-th node and t-th time instance is given by $\hat{\epsilon}_i^{(s)}[t]=x_i[t]-\hat{f}_i^{(s)}(\bar{\mathcal{X}}_{\setminus i}[t])$. For the variance estimates, we have ${\hat{\sigma}_i^{(0)}}^2=\frac{1}{T}\sum_{t=1}^{T}{\hat{\epsilon}_i^{(0)}}^2[t]$ and ${\hat{\sigma}_i^{(1,2)}}^2=\frac{1}{T}\left[\sum_{t=1}^{\tau_1-1}{\hat{\epsilon}_i^{(1)}}^2[t]+\sum_{t=\tau_1}^{T}{\hat{\epsilon}_i^{(2)}}^2[t]\right]$.

Substituting for the estimated quantities in the log-(pseudo)likelihood ratio $L(\mathbf{X};\tau_1) := \log(p(\mathbf{X};\tau_1,\mathcal{H}_1)/p(\mathbf{X};\mathcal{H}_0))$ yields an approximate form of the generalized likelihood ratio test statistic

$$\Lambda := \max_{\tau_1\in[L_{\min}+1,\dots,T-L_{\min}+1]}\ \frac{T}{2}\log\sum_{i=1}^{|\mathcal{V}|}\frac{{\hat{\sigma}_i^{(0)}}^2}{{\hat{\sigma}_i^{(1,2)}}^2} \tag{10}$$

where $L_{\min}$ is the minimum allowed segment length. The argument for which this maximum is achieved, call it $\hat{\tau}_1$, is the proposed change point.

Deciding whether a CP exists or not amounts to comparing Λ in (10) with a threshold, which for a prescribed maximum probability of false alarm (deciding $\mathcal{H}_1$ when $\mathcal{H}_0$ is in effect), is obtained via the distribution of Λ under $\mathcal{H}_0$. In order to estimate the latter, we rely on a model-based bootstrap scheme; see e.g. [12, Ch. 8].

2.3. Detecting multiple change points

In this subsection, we will examine the case where both the locations and the number of CPs are unknown. In particular, we will consider a variant of the binary segmentation approach [17], to which the hypothesis test in (7) serves as a building block.

The process starts by assuming that there exist no CPs, that is $N=1$. This corresponds to the first stage, i.e. $k=1$. A hypothesis test using the statistic in (10) is subsequently performed. Let $\hat{\tau}_1^1$ denote the proposed CP, with the superscript indicating the stage. If $\mathcal{H}_1$ is accepted, the process continues to stage $k=2$ with the data segmented into $[1,\hat{\tau}_1^1-1]$ and $[\hat{\tau}_1^1,T]$. Otherwise the process stops with zero CPs discovered. At stage $k=2$, a test statistic is computed for each segment. Let $\hat{\Lambda}_1,\hat{\Lambda}_2$ denote the test statistics estimated for $[1,\hat{\tau}_1^1-1]$ and $[\hat{\tau}_1^1,T]$, respectively. We then pick the maximum over $\{\hat{\Lambda}_1,\hat{\Lambda}_2\}$. For the sake of description, assume that $\hat{\Lambda}_1=\max\{\hat{\Lambda}_1,\hat{\Lambda}_2\}$. A hypothesis test is then performed in $[1,\hat{\tau}_1^1-1]$, since this is the segment corresponding to the maximum. If $\mathcal{H}_0$ is accepted, the final segmentation is $[1,\hat{\tau}_1^1-1],[\hat{\tau}_1^1,T]$. Otherwise we procceed to stage $k=3$ with the data split into $[1,\hat{\tau}_1^2-1],[\hat{\tau}_1^2,\hat{\tau}_1^1-1],[\hat{\tau}_1^1,T]$.

We will now describe a general stage of the algorithm, let k. Note that at the beginning of the k-th stage k segments have been discovered. For each such segment, let n, we then estimate the corresponding test statistic Λ_n. The segment for which the maximum over the test statistics $\{\Lambda_n\}_{n=1}^k$ is achieved, is given by

$$n^* := \arg\max_{n=1,\dots,k}\ \Lambda_n. \tag{11}$$

A hypothesis test for the presence of a CP in segment n^* is then performed. If $\mathcal{H}_0$ is accepted, no additional CP is discovered and the process stops; otherwise the proposed CP $\hat{\tau}_{n^*}^k$ is added to the set of discovered CPs and the process continues to stage $(k+1)$.

2.4. Multi-kernel learning

Our analysis so far presumed that the selection of the kernel was made a priori. Choosing this similarity measure appropriately is key to obtaining meaningful estimators, since this choice "shapes" the space $\mathcal{H}$ (cf. (2)). In this work, multiple kernel learning [6] is employed for selecting an optimal combination of kernels, from a dictionary, based on the data.

Consider a dictionary comprising the kernels $\kappa_1,\dots,\kappa_P$ and let κ be a nonnegative combination of these, that is $\kappa=\sum_{p=1}^P\theta_p\kappa_p$ with $\theta_p\geq 0$ for $p=1,\dots,P$. This choice guarantees that as long as $\kappa_1,\dots,\kappa_P$ are valid kernels, so will be κ.

Our goal is to choose an optimal weight vector $\boldsymbol{\theta} := [\theta_1,\dots,\theta_P]$. Towards that end, the kernel matrix $\mathbf{K}_{-i}^{(n)}$ in (5) will be replaced by its multi-kernel counterpart, namely $\mathbf{K}_{-i}^{(n)}=\sum_{p=1}^P\theta_p\mathbf{K}_{p-i}^{(n)}$, and we will jointly minimize over $\boldsymbol{\theta}$ and $\boldsymbol{\beta}_i$. To this end, note that (5) has the same solution as

$$\arg\min_{\boldsymbol{\beta}_i^{(n)}}\|(1/\sqrt{\lambda_i^{(n)}})\mathbf{x}_i^{(n)}-\sqrt{\lambda_i^{(n)}}\boldsymbol{\beta}_i^{(n)}\|^2+{\boldsymbol{\beta}_i^{(n)}}^\top\mathbf{K}_{-i}^{(n)}\boldsymbol{\beta}_i^{(n)}. \tag{12}$$

Based on this observation, we consider per segment n and node i the problem

$$\min_{\substack{\boldsymbol{\theta}\in\mathbb{R}^P:\boldsymbol{\theta}\succeq\mathbf{0}\\ \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_2\leq\Lambda}}\ \min_{\boldsymbol{\beta}_i^{(n)}}\ \lambda_i^{(n)}\boldsymbol{\beta}_i^{(n)\top}\boldsymbol{\beta}_i^{(n)}-2\boldsymbol{\beta}_i^{(n)\top}\mathbf{x}_i^{(n)}$$
$$+\sum_{p=1}^P\theta_p\boldsymbol{\beta}_i^{(n)\top}\mathbf{K}_{p-i}^{(n)}\boldsymbol{\beta}_i^{(n)} \tag{13}$$

where the constraint $\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_2\leq\Lambda$ effects ℓ_2 regularization on $\boldsymbol{\theta}$ [3]. An iterative algorithm that alternates between estimating $\boldsymbol{\theta}$ and $\boldsymbol{\beta}_i^{(n)}$ is used to solve (13); see [3, 11] for further details.

3. NUMERICAL TESTS

The proposed approach was evaluated on synthetic data generated using a dynamic causal model [5], albeit with a time-varying ground truth topology. The simulation setup for each individual segment resembles that of [15].

Specifically, with ψ_i denoting the neural time course for the i-th node, and u_i standing for the input to the same node, the dynamic causal neural network model for the n-th segment can be described by

$$\dot{\boldsymbol{\psi}}(t) = \delta \mathbf{A}^{(n)} \boldsymbol{\psi}(t) + \mathbf{u}(t) \tag{14}$$

where $\boldsymbol{\psi}(t) := [\psi_1(t), \ldots, \psi_{|\mathcal{V}|}(t)]^\top$, $\mathbf{u}(t) := [u_1(t), \ldots, u_{|\mathcal{V}|}(t)]^\top$, and $\mathbf{A}^{(n)}$ denotes the ground truth connectivity matrix for the n-th segment. Finally, the parameter δ regulates the neural lags; we will set $\delta = 20$ hereafter.

The inputs $\{u_i(t)\}$ were chosen so as to simulate resting state fMRI data; see [11] for further details on the input signals, as well as the choice of the DCM parameters. Each neural time course ψ_i obtained as a solution to (14) was subsequently fed into the nonlinear balloon model for vascular dynamics. Finally, the vectors $\{\mathbf{x}_\nu\}_{\nu=1}^{|\mathcal{V}|}$ were obtained by sampling the output of the latter with period $TR = 1$s.

The ground truth upper triangular connectivity matrices $\{\mathbf{A}^{(n)}\}_{n=1}^{N}$, with $\mathbf{A}^{(n)} \in \mathbb{R}^{10\times 10}$, featured a fixed number of randomly placed nonzero entries. Moreover, each nonzero entry was drawn from a uniform distribution over the interval $[0.25, 0.6]$. Since τ_n marks a CP in the topology, the connectivity matrices further satisfy $\mathbf{A}^{(n-1)} \neq \mathbf{A}^{(n)}$ and $\mathbf{A}^{(n)} \neq \mathbf{A}^{(n+1)}$ for all (valid) segments n.

Regarding the parameters of the method, the kernel dictionary comprised 10 Gaussian kernels with variances spanning the interval $[10^{-3}, 10^{2}]$, and a linear kernel. The regularization parameters $\{\lambda_i^{(n)}\}$ were chosen from the interval $[10^{-7}, 10^{2}]$ so as to minimize the 5-step prediction error; see e.g. [9] for further details. The model order was set to $d = 10$, whereas $L_{\min} = 25$. Finally, 250 bootstrap realizations were generated for estimating the distribution of Λ_{n^*}.

With $\boldsymbol{\tau} := [\tau_0, \ldots, \tau_N]$ denoting the (ground truth) vector of change points, the following temporal configurations were considered $\boldsymbol{\tau} = [1, 60, 110]$, $[1, 40, 70, 180]$, $[1, 60, 175, 250]$, $[1, 50, 90, 130, 180]$, $[1, 60, 120, 180, 240]$. For each configuration five realizations were generated.

The results indicate that the novel approach successfully detects the presence of CPs, while also accurately estimating their locations. Regarding the former, an average discrepancy $|N - \hat{N}|$ of 0.4, where $\hat{N} - 1$ stands for the estimated number of CPs, was achieved. For the latter, a mean absolute error $|\tau - \hat{\tau}|$ of 13 time points was obtained. Note that some discrepancy should inherently be expected given the presence of hemodynamics. Overall, 83% of the time points were assigned to the correct segment.

4. CONCLUSIONS

In this work, a novel kernel-based nonlinear change detection approach was introduced. The goal is that of capturing changes in nonlinear dependencies that linear methods may generally ignore. Multiple kernel learning was employed for choosing an optimal combination of kernels, from a dictionary, based on the data. An appropriate test statistic was devised and an algorithm for detecting the presence and estimating the location of an unknown number of change points was developed. Tests on DCM-based synthetic data demonstrated the potential of the proposed approach.

5. REFERENCES

[1] J. Besag, "Statistical analysis of non-lattice data," *Journal of the Royal Statistical Society. Series D (The Statistician)*, vol. 24, no. 3, pp. 179–195, 1975.

[2] R. B. Buxton, K. Uludag, D. J. Dubowitz, and T. T. Liu, "Modeling the hemodynamic response to brain activation," *NeuroImage*, vol. 23, pp. S220 – S233, 2004.

[3] C. Cortes, M. Mohri, and A. Rostamizadeh, "L2 regularization for learning kernels," in *Proc. Conf. on Uncertainty in Artificial Intelligence*, Arlington, VA, USA, 2009, pp. 109–116.

[4] I. Cribben, R. Haraldsdottir, L. Y. Atlas, T. D. Wager, and M. A. Lindquist, "Dynamic connectivity regression: Determining state-related changes in brain connectivity," *NeuroImage*, vol. 61, no. 4, pp. 907–920, Jul. 2012.

[5] K. J. Friston, L. Harrison, and W. Penny, "Dynamic causal modelling," *NeuroImage*, vol. 19, no. 4, pp. 1273–1302, 2003.

[6] M. Gönen and E. Alpaydın, "Multiple kernel learning algorithms," *Journal of Machine Learning Research*, vol. 12, pp. 2211–2268, 2011.

[7] S. A. Huettel, A. W. Song, and G. McCarthy, *Functional Magnetic Resonance Imaging*. Sinauer Associates, 2004.

[8] R. M. Hutchison, T. Womelsdorf, E. A. Allen, P. A. Bandettini, V. D. Calhoun, M. Corbetta, S. Della Penna, J. H. Duyn, G. H. Glover, J. Gonzalez-Castillo, D. A. Handwerker, S. Keilholz, V. Kiviniemi, D. A. Leopold, F. de Pasquale, O. Sporns, M. Walter, and C. Chang, "Dynamic functional connectivity: Promise, issues, and interpretations," *NeuroImage*, vol. 80, pp. 360–378, Oct. 2013.

[9] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. OTexts, 2014.

[10] S.-O. Jeong, C. Pae, and H.-J. Park, "Connectivity-based change point detection for large-size functional networks," *NeuroImage*, vol. 143, no. Supplement C, pp. 353 – 363, 2016.

[11] G. V. Karanikolas, G. B. Giannakis, K. Slavakis, and R. M. Leahy, "Multi-kernel based nonlinear models for connectivity identification of brain networks," in *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, Shanghai, China, Mar. 2016, pp. 6315–6319.

[12] S. N. Lahiri, *Resampling Methods for Dependent Data*. Springer Science & Business Media, 2013.

[13] B. Schölkopf and J. A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT press, 2002.

[14] B. Schölkopf, R. Herbrich, and A. J. Smola, "A generalized representer theorem," in *Computational Learning Theory: 14th Annual Conf. on Computational Learning Theory, COLT and 5th European Conf. on Computational Learning Theory, EuroCOLT*. Amsterdam, The Netherlands: Springer Berlin Heidelberg, Jul. 2001, pp. 416–426.

[15] S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, and M. W. Woolrich, "Network modelling methods for fMRI," *NeuroImage*, vol. 54, no. 2, pp. 875–891, 2011.

[16] O. Sporns, *Networks of the Brain*. MIT press, 2011.

[17] L. J. Vostrikova, "Detecting "disorder" in multidimensional random processes," *Soviet Mathematics: Doklady*, vol. 24, pp. 55–59, 1981.