

Local dimension-reduced dynamical spatio-temporal models for resting state network estimation

Gilson Vieira · Edson Amaro · Luiz A. Baccalá

Received: 13 October 2014 / Accepted: 19 January 2015 / Published online: 3 February 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract To overcome the limitations of independent component analysis (ICA), today's most popular analysis tool for investigating whole-brain spatial activation in resting state functional magnetic resonance imaging (fMRI), we present a new class of local dimension-reduced dynamical spatio-temporal model which dispenses the independence assumptions that severely limit deeper connectivity descriptions between spatial components. The new method combines novel concepts of group sparsity with contiguity-constrained clusterization to produce physiologically consistent regions of interest in illustrative fMRI data whose causal interactions may then be easily estimated, something impossible under the usual ICA assumptions.

Keywords Resting state fMRI · Dynamical spatio-temporal models · Brain connectivity · Sparsity

1 Introduction

There is an ever-growing and pressing need for accurately describing how brain regions are dynamically interrelated in resting state fMRI [4]. Thanks to the nature of BOLD

signals, resting state interactions cannot be split into separate space and time descriptions, especially if the focus lies on characterizing spatial changes associated to a small number of regions of interest. The chief challenge is that any dynamical spatio-temporal model (DSTM) of fMRI datasets demands many parameters to describe what is also a large number of observed variables which, nonetheless, enjoy a great deal of spatial redundancy [3, 5, 37]. Estimating the spatial origin of signal variability for only relatively short sample sizes using DSTMs is problematic especially under the rather usual unfavourable signal-to-noise ratio (SNR) conditions [8, 24, 28, 34].

To circumvent limitations of modelling high-dimensionality systems, Wikle and Cressie [33] proposed dimension-reduced DSTMs aimed at capturing non-stationary spatial dependence under optimal state representations using Kalman filtering. In their formulation of DSTM, they invoke an a priori defined orthogonal basis to expand the redistribution kernel of a discrete time/continuous space, linear integro-difference equation (IDE) in terms of a finite linear combination of spatial components [33]. This idea was further supported in [14] and extended in [26] who considered parametrized redistribution kernels of arbitrary shape that meet homogeneity conditions in both space and time. Even though the base changes of [33] improve the understanding of high-dimensional processes, they by no means ensure sparse solutions which are key to achieving statistically robust dynamical descriptions.

Model robustness has alternatively been sought by indirect means as, for example, thru LASSO regression [29] and basis pursuit [6] for model selection and denoising, or sparse component analysis for blind source separation [39] and finally by iterative thresholding algorithms for image deconvolution and reconstruction [12, 17]. The latter methods seek sparsity by maximizing a penalized

G. Vieira (✉)
Inter-institutional Grad Program on Bioinformatics, University
of São Paulo, São Paulo, Brazil
e-mail: gilson.vieira@gmail.com

E. Amaro
LIM-44, Department of Radiology, Faculty of Medicine,
University of São Paulo, São Paulo, Brazil

L. A. Baccalá
Escola Politécnica, University of São Paulo, São Paulo, Brazil

loss function in a compromise between the goodness of fit and the number of basis elements that make up the signal. Recently, more attention has been given to group sparsity, where groups of variables are selected/shrunk simultaneously rather than individually (for a review see [2]). This is achieved by minimizing an objective function that includes a quadratic error term added to a regularization term that considers a priori beliefs or data-driven analysis to induce group sparsity [35, 36, 38].

The present paper extends the results in [31] about local dimension-reduced DSTMs (LDSTMs) involving state-space formulations that are suited to datasets of high dimensionality such as fMRI. LDSTMs take advantage of a sparsifying spatial wavelet transformation to represent the data thru fewer significant parameters which are then combined via sparsity and contiguity-constrained clustering to initialize the observation matrix and sources of a tailored expectation maximization (EM) algorithm. The main assumptions here are that the system is overdetermined (there exist more observed signals than sources) and that the columns of the observation matrix act as point-spreading functions (see Sect. 2). Finally, results are gauged using simulated data (Sect. 4) followed by further method illustration with directed connectivity disclosure using real fMRI resting state data.

2 Problem formulation

DSTM problems may be formulated as state-space models (see [9] for a comprehensive review of DSTM) where space-related measurements z_t depend on the dynamical evolution of a suitably defined source vector x_t through a linear gaussian model

$$\mathbf{x}_t = \sum_{l=1}^L \mathbf{H}_l \mathbf{x}_{t-l} + \mathbf{w}_t \quad (1)$$

$$\mathbf{z}_t = \mathbf{A} \mathbf{x}_t + \mathbf{v}_t, \quad (2)$$

where \mathbf{z}_t is an M dimensional column vector of observed signals at time t , \mathbf{x}_t is an K dimensional column vector of unknown source vectors, \mathbf{A} is an unknown $M \times K$ observation matrix, \mathbf{H}_l for $1 \leq l \leq L$ are unknown $K \times K$ matrices that describe source vector dynamics, \mathbf{w}_t is an innovation process and \mathbf{v}_t is an additive noise. Both \mathbf{w}_t and \mathbf{v}_t are assumed zero mean gaussian, respectively, with covariance \mathbf{Q} and \mathbf{R} . The \mathbf{H}_l matrices, the observation matrix \mathbf{A} together with \mathbf{Q} and \mathbf{R} and \mathbf{x}_t must be inferred from z_t . For added generality, Eq. (1) is presented in a slightly extended form compared to the corresponding model in [31].

Under the latter premises, the log-likelihood of model (1, 2) is given by

$$\begin{aligned} \log p(\mathbf{x}, \mathbf{A}, \mathbf{R}, \mathbf{H}_1, \dots, \mathbf{H}_L, \mathbf{Q} | \mathbf{z}) \\ = -\frac{T}{2} \log |\mathbf{R}| - \frac{1}{2} \sum_{t=1}^T (\mathbf{z}_t - \mathbf{A} \mathbf{x}_t)^T \mathbf{R}^{-1} (\mathbf{z}_t - \mathbf{A} \mathbf{x}_t) \\ - \frac{T-1}{2} \log |\mathbf{Q}| - \frac{1}{2} \sum_{t=L+1}^T \\ \times \left(\mathbf{x}_t - \sum_{l=1}^L \mathbf{H}_l \mathbf{x}_{t-l} \right)^T \mathbf{Q}^{-1} \left(\mathbf{x}_t - \sum_{l=1}^L \mathbf{H}_l \mathbf{x}_{t-l} \right), \end{aligned} \quad (3)$$

where $z = \text{vec}([\mathbf{z}_1 \cdots \mathbf{z}_T])$, $\mathbf{x} = \text{vec}([\mathbf{x}_1 \cdots \mathbf{x}_T])$ and vec stands for the column stacking operator [27].

The EM algorithm has long been the favourite tool to solve (1,2) for \mathbf{x}_t because (3) is sure to converge to at least a local maximum [13, 27]. The traditional EM algorithm starts with randomly generated solutions for all parameters and then proceeds by re-iterating its two main steps until the maximum of (3) is attained. It begins with the E-step where the unknown \mathbf{x}_t are replaced by their expected values given the data and current model parameter estimates. Under gaussian assumptions, the expected system \mathbf{x}_t are obtained via the Rauch–Tung–Striebel (RTS) smoother [25]. In the second algorithm step, the M-step, one estimates model parameters by maximizing the conditional expected likelihood from the previous E-step. In practice, EM algorithm performance degrades rapidly for high-dimensional systems under (1,2). Its solution may even become indeterminate and improper initialization, in fact, often deteriorates estimate quality.

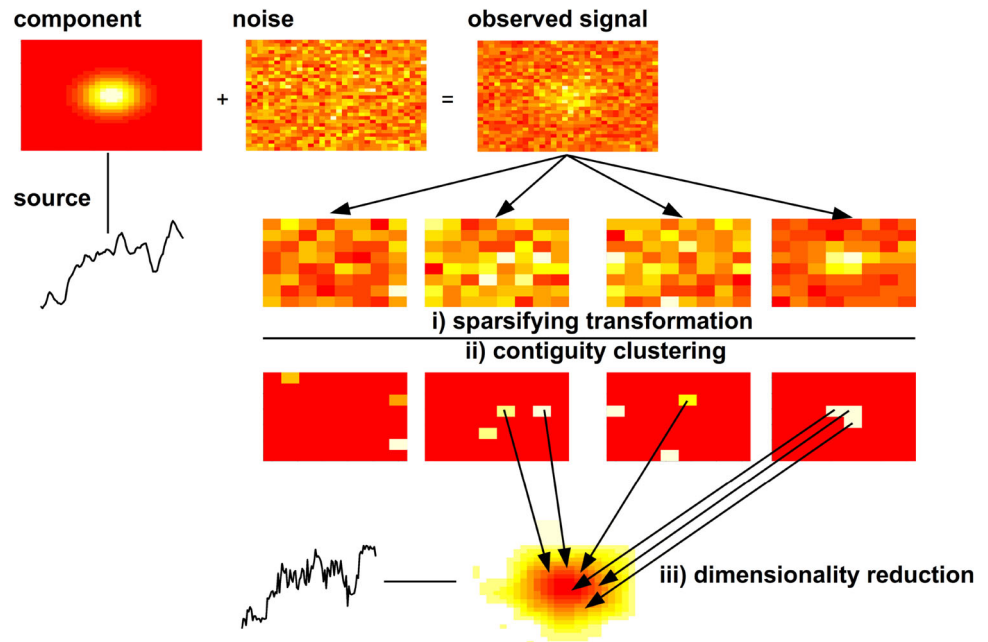
To achieve robust EM solutions, we take into account two common neuroscientist concerns as to what constitute meaningful brain activity components: (a) \mathbf{x}_t be an economic (i.e. compact/low dimensional) dynamical representation of the brain resting state fMRI dataset as a whole and (b) solutions must be spatially localized, i.e. their associated activation areas mathematically reflect point-spreading functions. We show that the latter assumptions not only allows estimating (1,2) parameters but also \mathbf{x}_t using the simpler Local Sparse Component Analysis discussed in [32] on \mathbf{z}_t . The nutshell description of the present algorithm is represented in Fig. 1. The aim is to find initial estimators for the observation matrix and system states which are used to initialize a EM algorithm for maximization of (3).

3 Algorithm details

3.1 Sparsifying spatial wavelet transformation

Given $\{\phi_m\}_{1 \leq m \leq M}$ an wavelet basis in \mathbf{R}^M , the first step is to calculate the wavelet representation of the matrix of observations $\mathbf{Z} \equiv (z_{m,t})_{m,t}$ for $1 \leq m \leq M$ and $1 \leq t \leq T$

Fig. 1 The main algorithm consists of (i) the application of a sparsifying spatial wavelet transformation, resulting into a description in terms of wavelet coefficient time series, (ii) contiguity-constrained clustering of the time series of wavelet coefficients by grouping only nearby coefficients and (iii) estimation of the observation matrix and system states by linear dimensionality reduction of the identified clusters



$$\hat{\mathbf{Z}} \equiv (\hat{z}_{m,t})_{m,t} = (\langle \mathbf{z}_t, \phi_m \rangle)_{m,t} = \Phi \mathbf{Z}, \tag{4}$$

where Φ is the $M \times M$ orthonormal matrix, whose rows are the ϕ_m 's. With obvious notation, $\mathbf{Z} = \mathbf{S} + \mathbf{V}$, where $\mathbf{S} = \mathbf{A}\mathbf{X}$, and $\hat{\mathbf{Z}} = \hat{\mathbf{S}} + \hat{\mathbf{V}}$. The transform Φ should be chosen such that a tailored clustering of the rows of $\hat{\mathbf{S}}$ provides the elements that approximate the rows of \mathbf{X} . But before this step, $\hat{\mathbf{S}}$ must be estimated using the sparsity assumption which implies finding a sparse representation of $\hat{\mathbf{Z}}$ that captures its intrinsic degrees of freedom.

By considering that $\mathbf{s}_t = \mathbf{A}\mathbf{x}_t$ admits a sparse representation lying in $\mathbf{B}_{1,1}^s$, a particular kind of Besov space [23], approximating \mathbf{z}_t by $\mathbf{s}_t \in \mathbf{B}_{1,1}^s$, can be expressed by adding a penalization term to $\|\mathbf{z}_t - \mathbf{s}_t\|_2^2$ requiring that $\|\mathbf{s}_t\|_{s,1}$ be small, where $\|\mathbf{s}_t\|_{s,1}$ is the $\mathbf{B}_{1,1}^s$ norm of \mathbf{s}_t . In other words, we want to minimize the following function:

$$f(\mathbf{s}_t) = \|\mathbf{z}_t - \mathbf{s}_t\|_2^2 + \|\mathbf{s}_t\|_{s,1} = \|\mathbf{z}_t - \mathbf{s}_t\|_2^2 + \sum_m \lambda_m |\hat{s}_{m,t}|, \tag{5}$$

where $\hat{s}_{m,t} = \langle \mathbf{s}_t, \phi_m \rangle$ and $\lambda_m > 0$ for $1 \leq m \leq M$ are regularization parameters [12].

For each t , the above function is coercive and strictly convex which means that it has a unique global minimum. If $\lambda_m = \lambda$, the minimum value of (5) is obtained via the soft-thresholding operator [15]

$$\hat{s}_{m,t} = \text{sign}(\hat{z}_{m,t}) \max(|\hat{z}_{m,t}| - \lambda, 0). \tag{6}$$

Since $\hat{s}_{m,t}$ can be zero for some values of t but not for others, the estimator (6) does not ensure sparsity of s_t over

time even for large λ values. To overcome this problem, we propose tying $\hat{s}_{m,t}$ for $1 \leq t \leq T$ together and using a recently introduced group-separable regularizer for the functional (5) but in the wavelet domain

$$\min_{\hat{\mathbf{s}}^m} \frac{1}{2} \|\hat{\mathbf{z}}^m - \hat{\mathbf{s}}^m\|_2^2 + \lambda_m \|\hat{\mathbf{s}}^m\|_2, \tag{7}$$

where $\hat{\mathbf{z}}^m$ and $\hat{\mathbf{s}}^m$ are the m -th rows of $\hat{\mathbf{Z}}$ and $\hat{\mathbf{S}}$, respectively. Given λ_m , solving (7) is achieved by the vector soft-thresholding operator [7, 35]

$$\hat{\mathbf{s}}^m = \frac{\max(\|\hat{\mathbf{z}}^m\|_2 - \lambda_m, 0)}{\|\hat{\mathbf{z}}^m\|_2} \hat{\mathbf{z}}^m. \tag{8}$$

In practice, we still need to estimate λ_m in (8) for signal denoising. Since Φ is orthogonal, if $\mathbf{R} = \sigma^2 \mathbf{I}_{M \times M}$, then $\hat{\mathbf{v}}^m \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{T \times T})$, where $\hat{\mathbf{v}}^m$ is the m -th row of $\hat{\mathbf{V}}$. For very large datasets, this assumption is quite strong but commonly employed in literature. As \mathbf{z}_t is sparse under Φ , most of $\{\hat{s}_{m,t}\}_{\forall m}$ must be zero. Provided that fifty percent of $\{\hat{s}_{m,t}\}_{\forall m}$ are zero, the following unbiased estimator for σ^2 can be defined

$$\hat{\sigma}^2 = \text{median}_{\forall m} \text{VAR}\{\hat{z}_{m,t}\}, \tag{9}$$

where VAR denotes temporal sample covariance.

If $\text{VAR}\{\hat{s}_{m,t}\} = 0$, we have that $\hat{z}_{m,t}$ are i.i.d normal variables, so

$$\frac{(N-1)\text{VAR}\{\hat{z}_{m,t}\}}{\sigma^2} \sim \chi_{N-1}^2 \tag{10}$$

implies that an interval with $(1 - \alpha)$ confidence for σ^2 is given by

$$\left[\frac{(N-1)\hat{\sigma}^2}{\chi_{1-\alpha/2, N-1}^2}, \frac{(N-1)\hat{\sigma}^2}{\chi_{\alpha/2, N-1}^2} \right], \quad (11)$$

where $\chi_{\xi, v}^2$ is the ξ -th percentile of the chi-square distribution with v degrees of freedom. Since $\|\hat{\mathbf{z}}^m\|_2 = (N-1)\text{V}\hat{\text{A}}\text{R}\{\hat{\mathbf{z}}_{m,t}\}$, (11) leads to λ_m given by

$$\lambda_m = \frac{(N-1)^2 \hat{\sigma}^2}{\chi_{\alpha/2, N-1}^2}, \quad (12)$$

with $\alpha = 0.05/M$.

3.2 Contiguity-constrained clustering

The next step consists of determining which time series of wavelet coefficients $\hat{\mathbf{s}}^m$ are associated to each spatial component $\mathbf{a}_k \mathbf{x}^k$, where \mathbf{a}_k is the k -th column of \mathbf{A} and \mathbf{x}^k is the k -th row of \mathbf{X} . For this, we use the spatial localization assumption. As the columns of the observation matrix are point-spreading functions, they should be perfectly described by wavelet coefficients forming localized spatial patterns. In this case, each spatial component can be determined using a clustering algorithm enforcing spatial contiguity. One way of achieving this is to apply complete linkage hierarchical clustering with the help of a dissimilarity measure that combines the time series temporal correlation and the physical distance between the wavelet coefficients. In this case, complete linkage hierarchical clustering is attractive because it yields relatively homogeneous clusters, a key property for subsequent accurate reduction of cluster dimensionality.

Clusterization begins with each $\hat{\mathbf{s}}^m$ defining a singleton cluster. At each step, it groups a pair (A, B) of clusters under the condition of minimizing the following distance function:

$$\text{dist}(A, B) = \max\{\psi(\hat{\mathbf{s}}^i, \hat{\mathbf{s}}^j) : i \in A, j \in B\}, \quad (13)$$

where

$$\psi(\hat{\mathbf{s}}^i, \hat{\mathbf{s}}^j) = \begin{cases} 1, & |\bar{\phi}_i - \bar{\phi}_j| > \max(2^{l_i}, 2^{l_j}) \\ 1 - |\text{cor}(\hat{\mathbf{s}}^i, \hat{\mathbf{s}}^j)|, & \text{otherwise,} \end{cases} \quad (14)$$

where $\text{cor}(\hat{\mathbf{s}}^i, \hat{\mathbf{s}}^j)$ denotes the correlation between $\hat{\mathbf{s}}^i$ and $\hat{\mathbf{s}}^j$, $\bar{\phi}_i = \int_{\mathbb{R}^d} s |\phi_i|^2 ds / \int_{\mathbb{R}^d} |\phi_i|^2 ds$ defines de center of mass of ϕ_i and l_i is the scale index of ϕ_i in the wavelet decomposition. Accordingly, the above dissimilarity measure combines the absolute value of the correlation coefficient and the physical distance between the wavelet coefficients. Clusterization stops when the minimal distance between the clusters is larger than r (i.e. $\min\{\text{dist}(A, B) : \forall A, \forall B\} > r$), for some appropriately

chosen r thus leading to a list of cluster memberships that characterize the system's spatial components.

Even though the dissimilarity measure (14) already establishes much of the structure that forms the spatial components of (17), one must decide when to stop clustering by an appropriate value of r . Note that the $\text{dist}(A, B)$ depends solely on the correlation between the wavelet coefficients in A and B . The Fisher z-transform of correlation coefficients, $0.5 \log_e(\frac{1+r}{1-r})$, follows a well-known statistic whose upper limit with an $(1 - \alpha/2)$ % confidence under the null hypothesis of independence is approximately

$$u = z_{(1-\alpha/2)} \sqrt{1/(N-3)}, \quad (15)$$

where $z_{(1-\alpha/2)}$ is the standard normal. Hence, we set the stopping value as

$$r = 1 - |(\exp(2u) + 1)/(\exp(2u) - 1)| \quad (16)$$

for $\alpha = 0.05$, which interestingly allows estimating the number of spatial components with reference neither to the actual noise level nor to the number of variables, but solely depending on sample size.

3.3 Within cluster dimensionality reduction

The next step consists of estimating the observation matrix \mathbf{A} and system states of (1, 2) by linear dimensional reduction of each spatial cluster identified in the previous step. After clustering the rows of $\hat{\mathbf{S}}$, the k -th spatial component $\mathbf{a}_k \mathbf{x}^k$ can be approximated by

$$\mathbf{Y}_k = \sum_{i \in I_k} \phi_i^{-1} \hat{\mathbf{s}}^i, \quad (17)$$

where \mathbf{Y}_k is an $M \times T$ data matrix, ϕ_i^{-1} is the i -th column of the inverse of $\Phi(\Phi^T$, for wavelet transforms) and I_k contains the indexes of the k -th cluster. We assume that the rows of \mathbf{Y}_k have zero mean, otherwise their mean value can be removed after (17).

According to the approximation model,

$$\mathbf{Y}_k = \mathbf{a}_k \mathbf{x}^k + \mathbf{E}_k, \quad (18)$$

where \mathbf{E}_k is an $M \times T$ approximation error matrix, and one must find \mathbf{a}_k and \mathbf{x}^k minimizing the approximation error

$$\min_{\mathbf{a}_k, \mathbf{x}^k} \|\mathbf{Y}_k - \mathbf{a}_k \mathbf{x}^k\|_F, \quad (19)$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

In fact, each spatial component $\mathbf{a}_k \mathbf{x}^k$ is a rank-one $M \times T$ matrix given by the first singular value of \mathbf{Y}_k , i.e.

$$\mathbf{Y}_k \approx \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T, \quad (20)$$

where σ_1 is the largest singular value of \mathbf{Y}_k , and where \mathbf{u}_1 and \mathbf{v}_1 are, respectively, the left-singular vector and the

right-singular vectors associated to σ_1 . With no loss of generality, we consider that the norm of \mathbf{a}_k equals one leading to

$$\mathbf{a}_k = \mathbf{u}_1 \tag{21}$$

$$\mathbf{x}^k = \sigma_1 \mathbf{v}_1^T. \tag{22}$$

3.4 LDSTM parameter estimation

The remainder of the algorithm consists of applying the traditional EM algorithm for \mathbf{x}_t estimation [27] using the estimators for \mathbf{x}^k and \mathbf{a}_k from previous section to set the initial values for \mathbf{x}_t and \mathbf{A} . Additionally, during the iterative process, \mathbf{A} matrix estimation is modified to accommodate linear equality constraints that ensure well-localized \mathbf{a}_k 's. This is done by solving the following least squares problem:

$$\min_{\mathbf{a}_k} \|\mathbf{a}_k \mathbf{x}^k - \mathbf{Z}\|_2^2 \tag{23}$$

subject to $\mathbf{C}\mathbf{a}_k = 0$,

where $\mathbf{C} = (c_{i,j})_{i,j}$ is an $M \times M$ matrix with $c_{i,i} = 1$ if $\text{VAR}(s_{i,t}^k) > 0$ and $c_{i,j} = 0$ otherwise.

4 Numerical illustration

Using simulated data to examine algorithm performance under different conditions, we created a vector time series corresponding to points on a discretized one-dimensional space consisting of $M = 256$ space points whose activity evolves in over a period of $T = 500$ points each. The observation matrix that we used (Fig. 2a) consists of the columns of

$$\mathbf{A} = [\mathbf{f}_{80} \mathbf{f}_{180} \mathbf{f}_{100}],$$

where $\mathbf{f}_\mu = [f_{1,\mu}, \dots, f_{M,\mu}]^T$ with $f_{i,\mu} = f(i - \mu)$ and f following a discretized Gaussian point-spread function. The

observations were corrupted by white Gaussian noise with covariance matrix

$$\mathbf{R} = \sigma^2 \mathbf{I}_{128,128},$$

with σ^2 accounting for the SNR level defined as $\text{SNR} = 10 \log_{10}(\text{VAR}(\mathbf{s})/\sigma^2)$ where $\mathbf{s} = \text{vec}([\mathbf{A}\mathbf{x}_1 \cdots \mathbf{A}\mathbf{x}_N])$. The dynamics of the spatial components evolved according to a first-order autoregressive model ($L = 1$) with

$$\mathbf{H}_1 = \begin{bmatrix} 0.5 & -0.5 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

and

$$\mathbf{Q} = \begin{bmatrix} 1 & 0.5 & 0 \\ 0.5 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Figure 2b shows the sample variance for a simulated DSTM using the above parameters under $\text{SNR} = -19\text{db}$.

We used Daubechies (D2) functions to transform the data and gauged performance by executing 100 Monte Carlo simulations leading to the mean and deviation results as shown in Fig. 3. Algorithm effectiveness was evaluated in terms of how well sources were recovered, as measured by their correlation to the estimated \mathbf{x}_t , and by how well \mathbf{H}_l and \mathbf{Q} could be estimated as evaluated by computing the connectivity between states using Partial Directed Coherence (PDC) [1].

4.1 Simulation results

The mean absolute values of the correlation coefficient between the simulated and estimated sources versus SNR in Fig. 3a show that LDSTM outperforms traditional EM, with very good results for all the three sources even under very unfavourable SNR. Figure 3b shows PDC from \mathbf{x}_2 towards \mathbf{x}_1 for different SNR levels compared to the

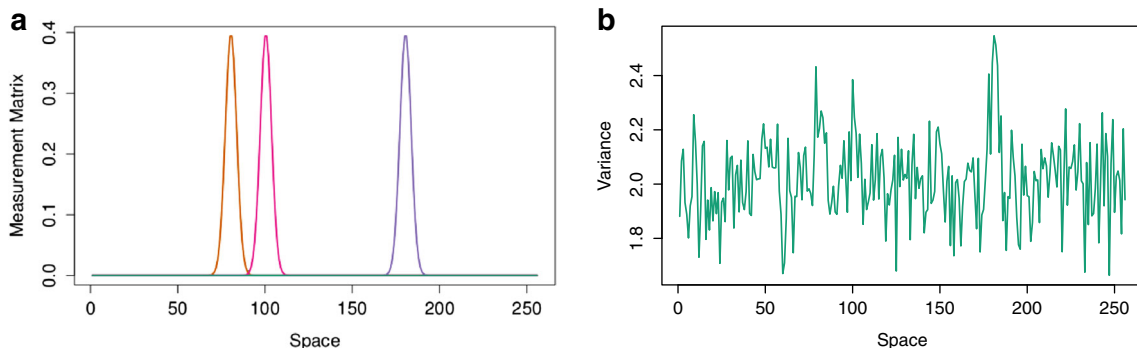


Fig. 2 a Measurement matrix \mathbf{A} and b sample variance of the example model with $N = 500$ and $\text{SNR} = -19$ db

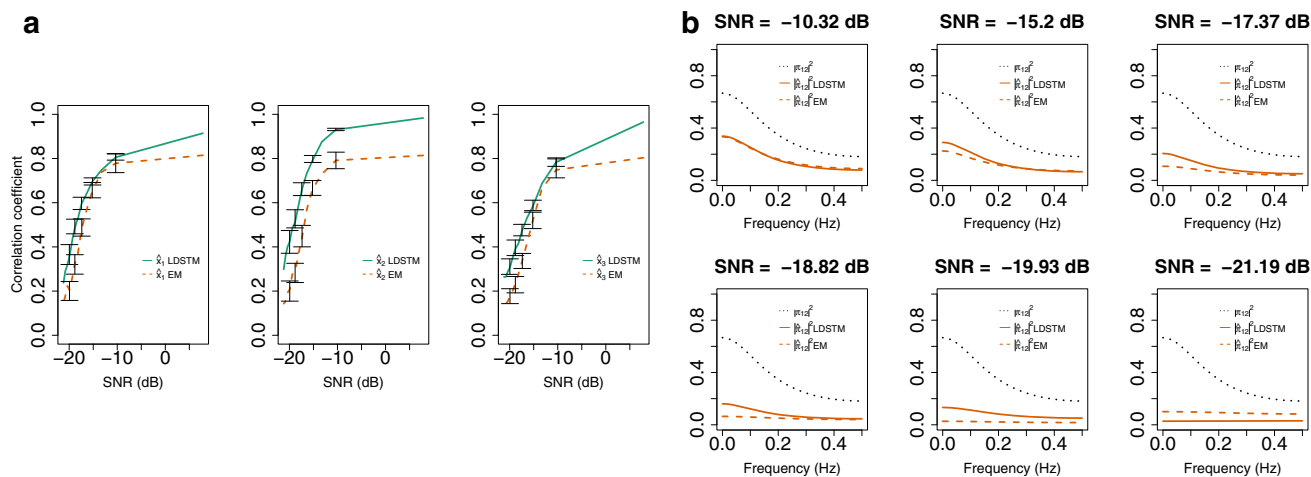


Fig. 3 a Efficiency comparison between LDSTM (solid lines) and EM (dashed lines) in recovering source temporal information. Lines represent the mean correlation between the simulated hidden state $x_{k,t}$ and the estimated hidden state $\hat{x}_{k,t}$ across 100 simulations. Vertical

error bars denote the 95 % confidence interval of the mean value. **b** Dotted lines represent the theoretical PDC of x_2 towards x_1 together with estimated PDC values of x_2 towards x_1 using LDSTM (solid) and EM (dash)

5 Real fMRI data

corresponding EM estimates. Correct PDC patterns were obtained whose magnitude decreases as SNR decreases but whose overall shape remains.

5.1 Image data acquisition

For further illustration purposes, we used fMRI images from seven healthy volunteers under a resting state protocol (approved by the local ethical committee and under individual informed written consent).

Whole-brain fMRI images (TR = 600 ms, TE = 33 ms, 32 slices, FOV = 247×247 mm, matrix size 128×128 , in-plane resolution 1.975×1.975 mm, slice thickness 3.5 mm with 1.8 mm of gap) were acquired on a 3T Siemens system using a Multiplexed Echo Planar Imaging sequence (multi-band accelerator factor of 4) [16]. To aid in the localization of functional data, high-resolution T1-weighted images were also acquired with an MPRAGE sequence (TR = 2500 ms, TE = 3.45 ms, inversion time = 1000 ms, 256×256 mm FOV, 256×256 in-plane matrix, $1 \times 1 \times 1$ mm voxel size, 7° flip angle).

5.2 LDSTM preprocessing

regions with preponderantly neuronal cell bodies. To further group analysis by temporal concatenation of the participants' fMRIs, individual grey matter images were registered to the 3-mm-thick Montreal Neurological Institute (MNI) template using a 12-parameter affine transform. To generate the spatial wavelet transformation, we used 3D Daubechies (D2) functions up to level 3. The model order for the dynamical component in (1) was defined by the Akaike information criterion.

5.3 ICA processing

To compare the LDSTM components with ICA, PICA was performed by multi-session temporal concatenation group ICA (using MELODIC in FSL). Preprocessing included slice time correction, motion correction, skull stripping, spatial smoothing (FWHM equals to 5 mm) and temporal high-pass filtering (allowing fluctuations above 0.005 Hz). The functional images were aligned into the standard space by applying 12 degrees-of-freedom linear affine transformation, and its time series were normalized to have variance of unity. The number of components was fixed at 30 to match the distinct pattern of resting state networks (RSN) usually found by other authors [4, 10].

5.4 Image results

5.4.1 LDSTM results

Figure 4 illustrates the advantage of wavelet transforming resting state fMRI datasets: the entropy in the image domain is much larger than that in the wavelet domain. This means that only a few wavelet coefficients are enough

Fig. 4 Fraction of cumulative energy in the image (*green*) and wavelet (*red*) domain for the resting state fMRI dataset. The *blue vertical line* crosses the fraction of cumulative energy represented by 10 % of the most energetic coefficients in the image (40 %) and wavelet (80 %) domains

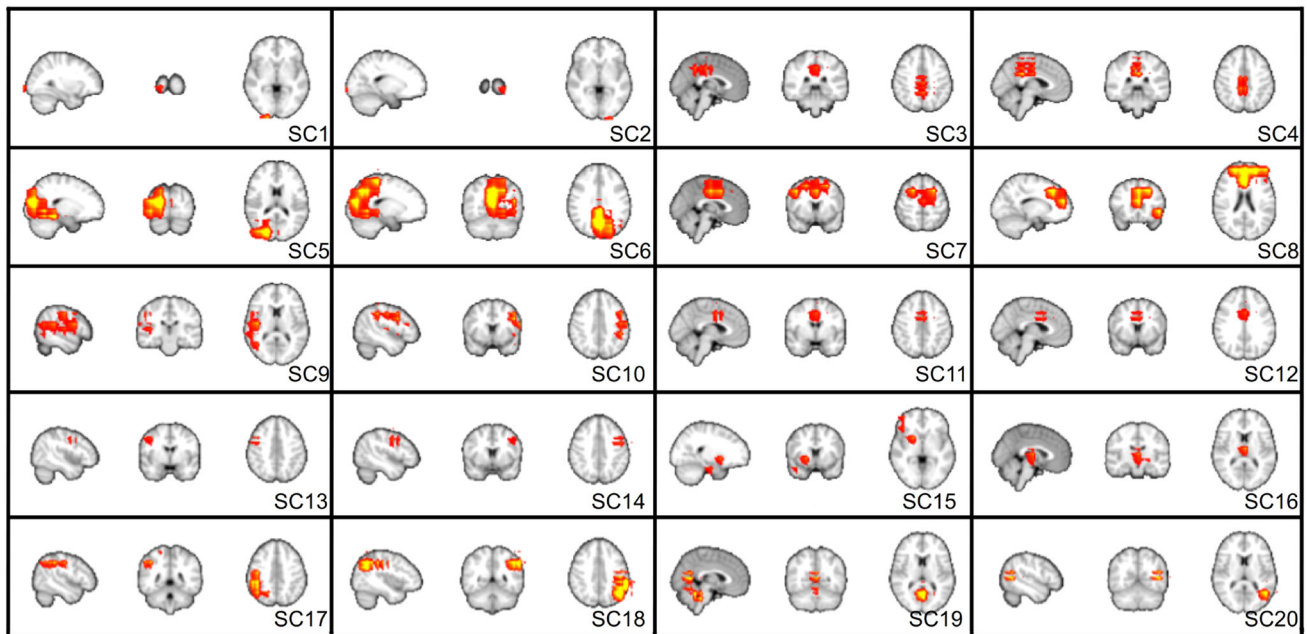
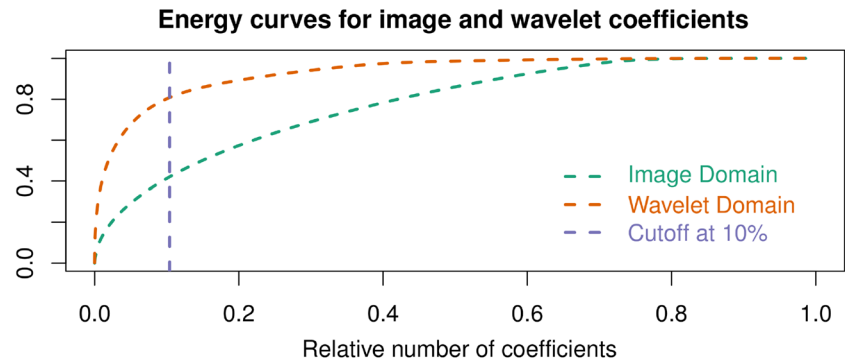


Fig. 5 Cortical and subcortical components identified by LDSTM

to account for much of the signal energy. In the example, 10 % of the wavelet coefficients explain 80 % of the image energy which is two times more than the 10 % of the most powerful image domain coefficients which represent just 40 %.

LDSTM analysis identified thirty nine well-localized spatial components comprising cortical (18), subcortical (2) and cerebellar (19) regions. Cortical and subcortical spatial components (\mathbf{a}_k 's) are shown in Fig. 5 which includes the following anatomical areas: occipital cortex (SC1 and SC2), lateral and superior occipital gyrus (SC5, SC6 and SC20), superior temporal gyrus (SC9 and SC10), precentral gyrus (SC13 and SC14), superior parietal gyrus (SC17 and SC18), precuneus (SC3 and SC19) and posterior cingulate (SC4), inferior frontal gyrus and anterior cingulate (SC7, SC8, SC11 and SC11)) and thalamus (SC15 and SC16). Cerebellar regions also form well-localized bilateral activity patterns as shown in Fig. 6.

The absence of artificial stochastic model constraints permitted exposing the dynamic connectivity between the identified components. Figure 7a summarizes the connectivity network estimated using PDC applied to the reconstructed system components. In addition, PDC also highlights that resting state connectivity is present mainly at low frequencies (Fig. 7b), corroborating several studies of resting state brain connectivity [4].

5.4.2 ICA results

Among the 30 component maps obtained by performing a PICA across all participants, 14 components were considered artifactual components due to scanner and physiological noise. Their signal variances are related to cerebrospinal fluid and white matter, head motion and large vessels. Figure 8 depicts fourteen functional components related to previously report resting state studies. They

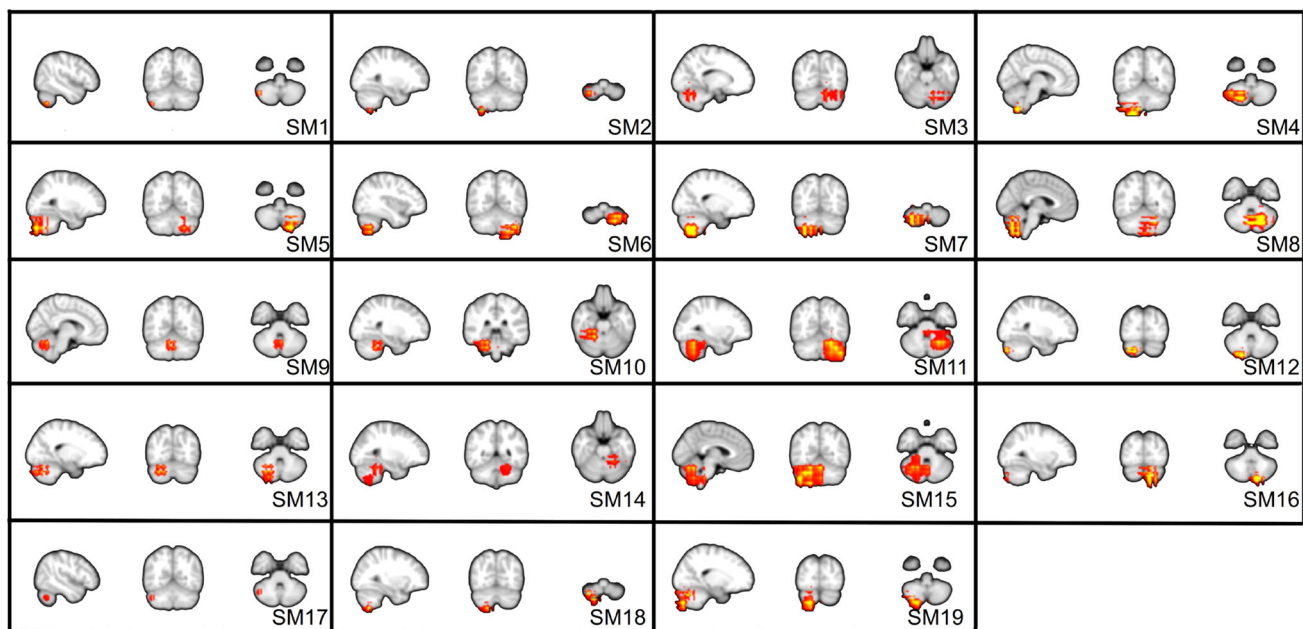


Fig. 6 Cerebellum components identified by LDSTM

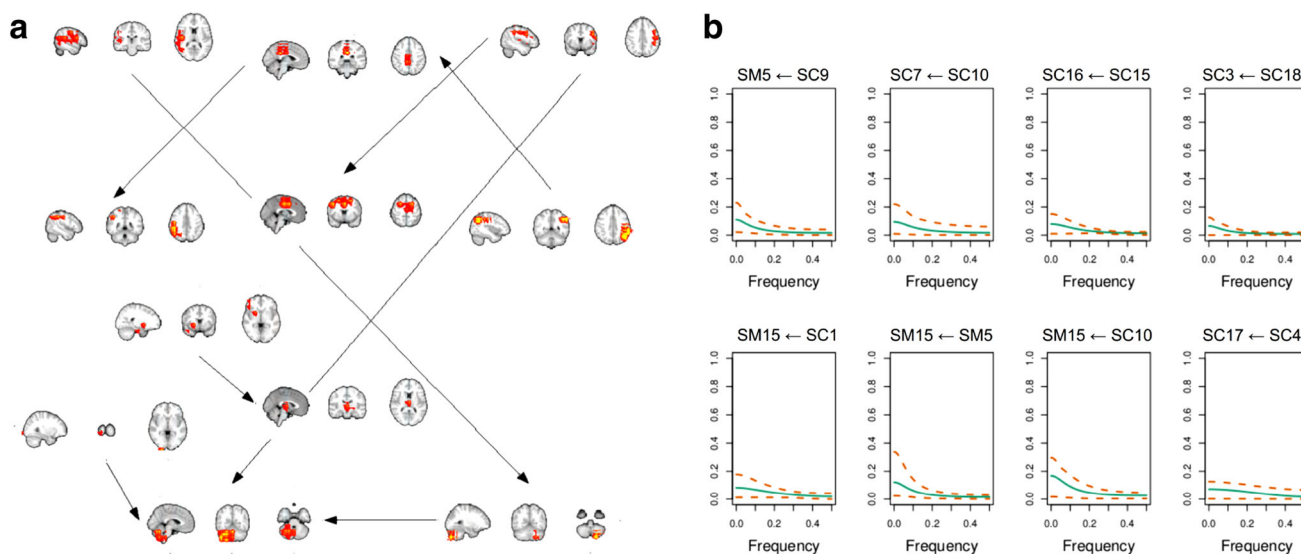


Fig. 7 fMRI resting state analysis using LDSTM. *Numbers* represent different components. Components numbered twice represent two components located at the same region. **a** Connectivity map showing

comprise the default mode network (IC2, IC9, IC10) and brain regions involved in visual (IC1, IC4), auditory/motor (IC5), sensory/motor (IC8), attentional (IC7, IC6, IC12, IC13) and executive functions (IC7, IC11, IC14). In addition, we found 2 components rarely reported in resting state studies. One is a cerebellum component (IC16) and the other is a brainstem component (IC15).

components whose system states are connected via the PDC. **b** PDC plots for each arrow drawing in **a**. *Dashed lines* denote the 95 % confidence interval of the mean value (*solid lines*)

6 Discussion

Local dimension-reduced modelling (LDSTM) as presented here addresses an approach to source estimation and localization in resting state fMRI data analysis that dispenses with artificial stochastic model assumptions, such as those used in classical blind source separation (principal

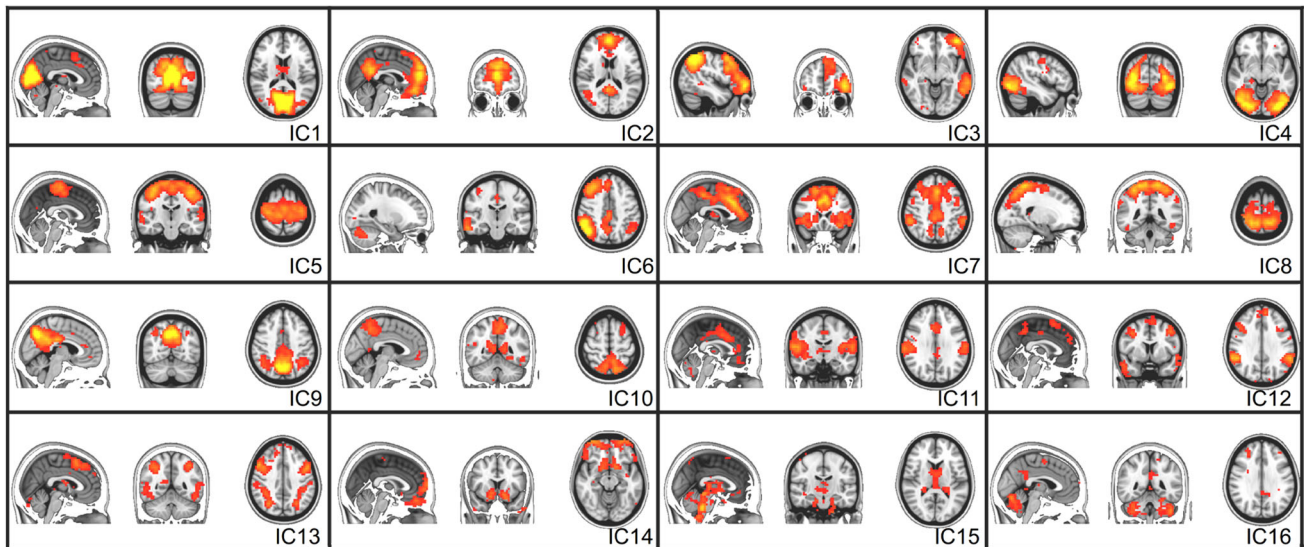


Fig. 8 ICA spatial components. The components are sorted according to their relative percentage of variance from *top left* to *bottom right*

component analysis (PCA), independent component analysis (ICA) and non-negative matrix factorization (NMF) [3, 18, 19, 21]). In addition to being sparse, the columns of the observation matrix act as point-spreading functions that allow system sources and their observation matrix to be identified via LSCA [32] of the whole fMRI dataset.

The cortical components identified by LDSTM (Fig. 5) reflect most of the data variability and coincide with traditional resting state regions observed across different individuals, data acquisition and analysis techniques. They comprise the default mode network (SC8) and brain regions involved in visual (SC1, SC2, SC5, SC6), motor (SC13, SC14, SC7) and attentional functions (SC9, SC10, SC17, SC18), indicating that most of the ICA components (Fig. 8) can in fact be decomposed into several local sparse components. However, the present results draw attention to the fact that they were obtained without any additional assumption, such as source independence and/or stationarity. All that was assumed was \mathbf{a}_k spatial localization, which goes along the line of [11]'s observation that ICA effectiveness for brain fMRI is linked to their ability to handle sparse sources rather than independent ones. This could be explained by pointing out that ICA pre-processing steps involve projecting the data into a reduced-dimensional subspace via the singular value decomposition which in turn confines the sources to regions of high signal variance.

PDC analysis shows a network where the information flows from regions in the superior parietal cortex (SPC) to regions in the cerebellum (CER) and anterior cingulate. As expected, the right SPC sends information to the left CER, and left SPC sends information to the right CER. Although the relationship between these structures is known, this stresses two main systems engaging in the mentioned

network. The connectivity between SPC and CER is in line with recent studies showing evidence of a cerebellar-parietal network involved in phonological storage [22]. In addition, visual-parietal-cerebellar interactions are expected by following studies of effective connectivity using fMRI [20]. We also observe a network running from the left to right parietal cortex passing through the posterior cingulate. Altogether, we believe that our results provide insight into the mechanisms of how the regions of the fronto-parietal network interact. It also highlights understudied aspects of the cerebellum in this network during resting state.

In our model, LDSTM identified approximately 50 % of components in the cerebellum. This result is surprising as the rate of cerebellar components identified in resting state using ICA is below 20 % in general [4]. Some of these regions seem to be related to noise sources for being located near cerebellar arteries and veins. The components SM1, SM2, SM12, SM17 and SM18 run in the superior surface of the cerebellum near to the superior cerebellar veins, while the components SM8 and SM9 extend into the end of the straight sinus near to internal cerebral veins. On the other hand, the idea that the cerebellum should present as many components as the cortex is encouraging. Many recent fMRI studies have shown that different cerebellar regions are critical for processing higher-order functions in different cognitive domains, in the same way as it occurs in the cortex [30]. In these studies, it is worth mentioning that cerebellar clusters are always smaller than those of corresponding functionality in the cortex. We believe that some differences between ICA and LDSTM may be explained in part by the features along the domain in which they represent the sources.

Since spatial wavelet analysis efficiently encodes the data neighbourhood information via a orthogonal transformation, the present method properly addresses a number of issues involving whole-brain connectivity estimation. The first one is associated to the lack of knowledge about the spatial localization of the sources. The method provides a data-driven approach to locate the main sources of data variability, thus avoiding the effects and uncertainties due to a priori region of interest delineation. The second aspect is that the new method naturally employs multi-scale transformations to create a compact model of the images, a feature of growing importance as higher-resolution images are sure to become available in the near future and whose computational processing load may be thereby substantially mitigated. Finally and most importantly, unlike ICA, the method permits deeper connectivity analysis between the identified spatial components as no independence assumption is made 'a priori'.

Various method extensions are possible, especially when it comes to estimate appropriate regularization parameter choice as a function of the amount of noise present in the data. In the present implementation, spatial noise is assumed homogeneous and normally distributed which implies a chi-squared distribution for wavelet coefficient variance. Examination of wavelet coefficients variance for real fMRI data, however, points to the need to consider heavy-tailed distributions, so that a more general approach is currently being developed to estimate wavelet domain noise variance from a finite mixture of exponential distributions that could then be used to quantify the level of data sparsity.

7 Conclusions

Here, an EM-based algorithm was presented for LDSTM identification. By projecting high-dimensional datasets into smoothness spaces, one can describe the system's spatial components via a reduced number of parameters. Further dimension reduction and denoising is obtained by soft-vector thresholding under contiguity-constrained hierarchical clustering. Finally, simulated results corroborate that the new algorithm can outperform the traditional EM approach even under mild conditions. Even with very large datasets as in the fMRI example, LDSTM shows promise in its ability to parcelate the human brain into well-localized physiologically plausible regions of spatio-temporal brain activation patterns.

Acknowledgments CNPq Grants 307163/2013-0 to L.A.B. We also thank NAPNA—Núcleo de Neurociência Aplicada from the University of São Paulo—and FAPESP Grant 2005/56464-9 (CInAPCe) during which time part of this work took place.

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Baccala LA, de Brito CSN, Takahashi DY, Sameshima K (2013) Unified asymptotic theory for all partial directed coherence forms. *Philos Trans R Soc A* 371(1997):20120158
- Bach F, Jenatton R, Mairal J, Obozinski G (2011) Structured sparsity through convex optimization. arXiv e-print [arXiv:1109.2397](https://arxiv.org/abs/1109.2397)
- Beckmann CF, Smith SM (2004) Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans Med Imaging* 23(2):137–152
- Biswal BB, Mennes M, Zuo X-N, Gohel S, Kelly C, Smith SM, Beckmann CF, Adelstein JS, Buckner RL, Colcombe S, Dogonowski A-M, Ernst M, Fair D, Hampson M, Hoptman MJ, Hyde JS, Kiviniemi VJ, Ktter R, Li S-J, Lin C-P, Lowe MJ, Mackay C, Madden DJ, Madsen KH, Margulies DS, Mayberg HS, McMahon K, Monk CS, Mostofsky SH, Nagel BJ, Pekar JJ, Peltier SJ, Petersen SE, Riedl V, Rombouts SAR, Rypma B, Schlaggar BL, Schmidt S, Seidler RD, Siegle GJ, Sorg C, Teng G-J, Veijola J, Villringer A, Walter M, Wang L, Weng X-C, Whitfield-Gabrieli S, Williamson P, Windischberger C, Zang Y-F, Zhang H-Y, Castellanos FX, Milham MP (2010) Toward discovery science of human brain function. *Proc Natl Acad Sci USA* 107(10):4734–4739
- Blumensath T, Jbabdi S, Glasser MF, Van Essen DC, Ugurbil K, Behrens TEJ, Smith SM (2013) Spatially constrained hierarchical parcellation of the brain with resting-state fMRI. *NeuroImage* 76:313–324
- Chen SS, Donoho DL, Saunders MA (1998) Atomic decomposition by basis pursuit. *SIAM J Sci Comput* 20(1):33–61
- Combettes PL, Wajs VR (2005) Signal recovery by proximal forward-backward splitting. *Multiscale Model Simul* 4(4):1168–1200
- Cortes J (2009) Distributed kriged kalman filter for spatial estimation. *IEEE Trans Autom Control* 54(12):2816–2827
- Cressie N, Wikle CK (2011) *Statistics for spatio-temporal data*. Wiley, Hoboken
- Damoiseaux JS, Rombouts SAR, Barkhof F, Scheltens P, Stam CJ, Smith SM, Beckmann CF (2006) Consistent resting-state networks across healthy subjects. *Proc Natl Acad Sci* 103(37):13848–13853
- Daubechies I, Roussos E, Takerkart S, Benharrosh M, Golden C, D'Ardenne K, Richter W, Cohen JD, Haxby J (2009) Independent component analysis for brain fMRI does not select for independence. *Proc Natl Acad Sci USA* 106(26):10415–10422
- Daubechies I, Defrise M, De Mol C (2003) An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. arXiv e-print [math/0307152](https://arxiv.org/abs/math/0307152)
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B* 39(1):1–38
- Dewar M, Scerri K, Kadiramanathan V (2009) Data-driven spatio-temporal modeling using the integro-difference equation. *IEEE Trans Signal Process* 57(1):83–91
- Donoho DL, Johnstone IM, Kerkycharian G, Picard D (1995) Wavelet shrinkage: asymptopia? *J R Stat Soc Ser B (Methodol)*

- 57(2):301–369 ArticleType: research-article/Full publication date: 1995/Copyright 1995 Royal Statistical Society
16. Feinberg DA, Moeller S, Smith SM, Auerbach E, Ramanna S, Glasser MF, Miller KL, Ugurbil K, Yacoub E (2010) Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. *PLoS One* 5(12):e15710
 17. Figueiredo MAT, Nowak RD (2003) An EM algorithm for wavelet-based image restoration. *IEEE Trans Image Process* 12(8):906–916
 18. Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ (1993) Functional connectivity: the principal-component analysis of large (PET) data sets. *J Cereb Blood Flow Metab* 13(1):5–14
 19. Georgiev P, Theis F, Cichocki A, Bakardjian H (2007) Sparse component analysis: a new tool for data mining. In: Pardalos PM, Boginski VL, Vazacopoulos A (eds) *Data mining in biomedicine*, number 7 in Springer optimization and its applications. Springer, New York, pp 91–116
 20. Kellermann T, Regenbogen C, De Vos M, Mwangi C, Finkelmeyer A, Habel U (2012) Effective connectivity of the human cerebellum during visual attention. *J Neurosci* 32(33):11453–11460
 21. Lohmann G, Volz KG, Ullsperger M (2007) Using non-negative matrix factorization for single-trial analysis of fMRI data. *Neuroimage* 37(4):1148–1160
 22. Macher K, Bhringer A, Villringer A, Pleger B (2014) Cerebellar-parietal connections underpin phonological storage. *J Neurosci* 34(14):5029–5037
 23. Mallat SG (2009) *A wavelet tour of signal processing the sparse way*. Elsevier/Academic Press, Amsterdam; Boston
 24. Mardia KV, Goodall C, Redfern EJ, Alonso FJ (1998) The kriged kalman filter. *Test* 7(2):217–282
 25. Rauch HE, Striebel CT, Tung F (1965) Maximum likelihood estimates of linear dynamic systems. *J Am Inst Aeronaut Astronaut* 3(8):1445–1450
 26. Scerri K, Dewar M, Kadiramanathan V (2009) Estimation and model selection for an IDE-based spatio-temporal model. *IEEE Trans Signal Process* 57(2):482–492
 27. Shumway RH, Stoffer DS (1982) An approach to time series smoothing and forecasting using the em algorithm. *J Time Ser Anal* 3(4):253–264
 28. Stoodley CJ, Schmahmann JD (2009) Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. *Neuroimage* 44(2):489–501
 29. Theophilides CN, Ahearn SC, Grady S, Merlino M (2003) Identifying west Nile virus risk areas: the dynamic continuous-area space-time system. *Am J Epidemiol* 157(9):843–854
 30. Tibshirani R (1994) Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B* 58:267–288
 31. Vieira G, Amaro E, Baccala LA (2014) Local dimension-reduced dynamical spatio-temporal models for resting state network estimation. In: Hutchison D, Kanade T, Kittler J, Kleinberg JM, Kobsa A, Mattern F, Mitchell JC, Naor M, Nierstrasz O, Pandu Rangan C, Steffen B, Terzopoulos D, Tygar D, Weikum G, Iezak D, Tan A-H, Peters JF, Schwabe L (eds) *Brain informatics and health*. Springer International Publishing, Cham, pp 436–446
 32. Vieira G, Amaro E, Baccala LA (2014) Local sparse component analysis for blind source separation: an application to resting state fmri. In: *Proceedings of IEEE EMBS conference*, IEEE
 33. Wikle CK, Cressie N (1999) A dimension-reduced approach to space-time kalman filtering. *Biometrika* 86(4):815–829
 34. Woolrich MW, Jenkinson M, Michael Brady J, Smith SM (2004) Fully bayesian spatio-temporal modeling of FMRI data. *IEEE Trans Med Imaging* 23(2):213–231
 35. Wright SJ, Nowak RD, Figueiredo MAT (2009) Sparse reconstruction by separable approximation. *IEEE Trans Signal Process* 57(7):2479–2493
 36. Yuan M, Lin Y (2006) Model selection and estimation in regression with grouped variables. *J R Stat Soc* 68(1):49–67
 37. Zalesky A, Fornito A, Harding IH, Cocchi L, Ycel M, Pantelis C, Bullmore ET (2010) Whole-brain anatomical networks: Does the choice of nodes matter? *NeuroImage* 50(3):970–983
 38. Zhao P, Rocha G (2009) The composite absolute penalties family for grouped and hierarchical variable selection. *Ann Stat* 37(6A):3468–3497 arXiv e-print [arXiv:0909.0411](https://arxiv.org/abs/0909.0411)
 39. Zibulevsky M, Pearlmutter BA (2001) Blind source separation by sparse decomposition in a signal dictionary. *Neural Comput* 13(4):863–882