Magnetic Resonance Imaging 58 (2019) 97–107

Contents lists available at ScienceDirect

Magnetic Resonance Imaging

journal homepage: www.elsevier.com/locate/mri

Original contribution

Automatic classification and removal of structured physiological noise for resting state functional connectivity MRI analysis

Kangjoo Lee[a,b,*], Hui Ming Khoo[b,d], Constance Fourcade[c], Jean Gotman[b], Christophe Grova[a,b,c]

[a] Multimodal Functional Imaging Lab, Department of Biomedical Engineering, McGill University, Duff Medical Building, 3775 Rue University, Montreal, QC H3A 2B4, Canada
[b] Montreal Neurological Institute, McGill University, 3801 Rue University, Montreal, QC H3A 2B4, Canada
[c] Department of Physics and PERFORM Centre, Concordia University, 7200 Rue Sherbrooke St. W, Montreal, QC H4B 1R6, Canada
[d] Department of Neurosurgery, Osaka University, 2-2 Yamadaoka, Suita, Osaka Prefecture 565-0871, Japan

ARTICLE INFO

Keywords:
Physiological noise
Functional MRI
Sparse dictionary learning
Stepwise regression
Denoising
Classification

ABSTRACT

Resting state functional magnetic resonance imaging is used to study how brain regions are functionally connected by measuring temporal correlation of the fMRI signals, when a subject is at rest. Sparse dictionary learning is used to estimate a dictionary of resting state networks by decomposing the whole brain signals into several temporal features (atoms), each being shared by a set of voxels associated to a network. Recently, we proposed and validated a new method entitled Sparsity-based Analysis of Reliable K-hubness (SPARK), suggesting that connector hubs of brain networks participating in inter-network communication can be identified by counting the number of atoms involved in each voxel (sparse number k). However, such hub analysis can be corrupted by the presence of noise-related atoms, where physiological fluctuations in cardiorespiratory processes may remain even after band-pass filtering and regression of confound signals from the white matter and cerebrospinal fluid. Handling this issue might require manual classification of noisy atoms, which is a time-consuming and subjective task. Motivated by the fact that the physiological fluctuations are often localized in tissues close to large vasculatures, i.e. sagittal sinus, we propose an automatic classification of physiological noise-related atoms for SPARK using spatial priors and a stepwise regression procedure. We measured the degree to which the noise-characteristic time-courses within the mask are explained by each atom, and classified noise-related atoms using a subject-specific threshold estimated using a bootstrap resampling based strategy. Using real data from healthy subjects ($N = 25$), manual classification of the atoms by two independent reviewers showed the presence of sagittal sinus related noise in 65% of the runs. Applying the same manual classification after the proposed automatic removal method reduced this rate to 19%. A 10-fold cross-validation on real data showed good specificity and accuracy of the proposed automated method in classifying the target noise (area under the ROC curve = 0.89), when compared to the manual classification considered as the reference. We demonstrated decrease in k-hubness values in the voxels involved in the sagittal sinus at both individual and group levels, suggesting a significant improvement of SPARK, which is particularly important when considering clinical applications.

1. Introduction

Functional Magnetic Resonance Imaging (fMRI) measures the blood oxygenation level dependent (BOLD) signals, which involve hemodynamic fluctuations elicited by neuronal activity [1]. Functional connectivity (FC) that measures temporal correlation between the fMRI signals in remote voxels is widely used to study brain network organization, when subjects are at resting state [2]. Sparse dictionary learning (SDL) estimates a dictionary of resting state networks (RSNs) by decomposing the whole brain signals into several (N) characteristic temporal features, i.e. atoms, each being shared by a set of voxels associated to a RSN [3]. Using a sparse general linear model (GLM) [4], the BOLD signal in each voxel is represented by a voxel-specific linear combination of a few (k) atoms, where the whole brain activities can be characterized by the larger number (N) of atoms.

Recently, we developed and validated the SParsity-based Analysis of Reliable K-hubness (SPARK) method, suggesting that *connector hubs* of brain networks can be identified by counting the number (k) of RSNs

* Corresponding author
E-mail address: kangjoo.lee@mail.mcgill.ca (K. Lee).

https://doi.org/10.1016/j.mri.2019.01.019
Received 9 May 2018; Received in revised form 27 December 2018; Accepted 18 January 2019

involved in each voxel (i.e. *k*-hubness) within the sparse GLM framework [5]. Hub regions that are involved in multiple-networks, participating in inter-network connectivity, are called connector hubs [6]. Connector hubs are important for network integrity and higher-order brain functions in healthy brains [7]. They are more vulnerable than other brain regions, being primary targets of many neurological disorders [8,9]. SPARK provides unique information about connector hub structures by measuring a realistic number of RSNs involved in each voxel (*k*-hubness) and also identifying explicitly which RSNs are involved in each connector hub.

In fMRI, the BOLD signals involve also non-neuronal fluctuations from instrumental, physiological, and metabolic origins. Physiological fluctuations related to respiration (0.098–0.63 Hz), heartbeat (0.79–1.6 Hz), or vasomotion are structured biological processes [10]. They can modulate the detection sensitivity of the BOLD signal variance [11], while reducing the detection power of FC or even introducing spurious FC patterns [12,13]. Physiological fluctuations in respiration or cardiac processes may remain even after band-pass filtering and regression of confound signals from voxels in the white matter and cerebrospinal fluid [14]. Therefore, the measurement of *k*-hubness using SDL can be corrupted by the presence of these "noise-related" atoms, reflecting physiological fluctuations of non-neuronal origin [5]. Handling this issue might require manual classification of noise-related atoms, which is a tedious, time-consuming and subjective task. To address this issue, we focused on the fact that most physiological fluctuations are localized in tissues close to large vessels, such as the sagittal sinus [15]. Motivated by an automatic correction method suggested for independent component analysis (ICA) [15], we propose an automatic classification for SDL atoms using spatial priors, without considering external recordings of cardiorespiratory profiles or training datasets as required by conventional methods used to remove noise-related parts of the signal.

In this article, we will provide a theory of the proposed method in Section 2, the materials, implementations and validation scheme using real resting state fMRI data in Section 3, results in Section 4, followed by discussions and a conclusion of the study.

2. Theory

The main objective of the proposed automatic physiological noise classification was to improve the estimation of FC connector hubs in individual resting state fMRI. Using SPARK, connector hubs are actually identified by measuring *k*-hubness based on sparse GLM, i.e. counting the number (*k*) of RSNs involved in each voxel, and finding the voxels exhibiting $k > 1$ [5]. An important advantage of SPARK is that the reliability of connector hub identification is ensured statistically at the single subject level. To do so, the SPARK method consists of generating a large number (*B*) of resampled datasets for an individual fMRI, on which SDL is independently applied. Then, it collects the *B* sets of outputs from every SDL (i.e. *B* dictionaries) and estimates only the most reproducible patterns of FC across the resampled datasets. However, the outputs from each SDL may include several noise-related atoms due to the presence of non-neuronal fluctuations in fMRI data, therefore, the hub identification can be corrupted if one counts these noise-related atoms in *k*-hubness. The objective of the present work was to develop a plug-in noise classification procedure within the SPARK framework, taking advantage of this resampling-based strategy proposed in SPARK. Our idea was to classify and remove noise atoms from each SDL output, before collecting them and evaluating their statistical reproducibility. The summary of the proposed automatic classification of physiological noise-related atoms for SPARK is illustrated in Fig. 1.

In this section, we will give an overview of the sparse GLM using SDL in Step 1. The output of SDL includes *N* time-course atoms in the trained dictionary $\mathbf{\Omega}$ and the corresponding spatial maps $\mathbf{X}$ from individual fMRI data. In Step 2, we will introduce our proposed physiological noise classification strategy for the output of SDL using spatial prior. Motivated by Perlbarg et al. [15], we employed spatial priors to define a set of noise-characteristic time-courses, such as those in the sagittal sinus. For each SDL output, a stepwise regression procedure was performed to assess the contributions of the target noise to each atom and select a set of noise-related atoms. This stepwise regression procedure resulted in the estimation of a "frequency of selection score" for each of *N* atoms (*Fq* in Fig. 1). The *Fq* score reflected the degree of significance of each atom for describing the noise-characteristic signals defined using spatial prior. In Step 3, we will propose the parallel processes that perform Steps 1 and 2 independently for each resampled dataset within the SPARK framework. This step includes the main contributions of this study, allowing for the estimation of a subject-specific threshold (Fq_{thresh}) for *Fq* scores. This thresholding process finally resulted in the classification of noise-related atoms, determining which atoms should be discarded from the hub analysis. Finally, in Step 4, we will describe the method to extract statistically reproducible networks and hubs from the collection of "de-noised" outputs of SDL.

2.1. Step 1: sparse dictionary learning of whole brain using sparse GLM

Step 1 includes the analysis of whole brain resting state fMRI data based on the sparse GLM, which was proposed in our previous work [4,5]. Let us denote the numbers of time-points *T*, voxels in the gray matter of the brain *V*, and functional networks *N*. The sparse GLM models the whole brain fMRI data $\mathbf{Y} \in \mathbb{R}^{T\times V}$ using a data-driven RSN dictionary $\mathbf{\Omega} \in \mathbb{R}^{T\times N}$ and a sparse coefficient matrix $\mathbf{X} \in \mathbb{R}^{N\times V}$ with the corresponding noise $\mathbf{E} \in \mathbb{R}^{T\times V}$:

$$\mathbf{y}_i = \mathbf{\Omega}\mathbf{x}_i + \mathbf{e}_i, \quad \text{subject to } \forall\, i\text{: } \|\mathbf{x}_i\|_0 \le k_i, \tag{1}$$

where $\mathbf{y}_i$ is a time-course measured in a voxel *i* and each column of $\mathbf{\Omega}$ is an atom $\boldsymbol{\omega}_j$, a network-characteristic temporal feature that is shared by a subset of voxels [4]. $\|\mathbf{x}_i\|_0$ denotes the L_0 norm of a vector $\mathbf{x}_i$, i.e. the number of non-zero elements. Each column of $\mathbf{X}$ ($\mathbf{x}_i$) is a sparse code for the voxel *i*, where the number of non-zeros in $\mathbf{x}_i$ defines the sparsity level k_i and the coefficients indicate the signal amplitudes of the atoms in each voxel. $\mathbf{\Omega}$ and $\mathbf{X}$ that best describe the data can be estimated using a SDL algorithm, such as K-SVD [16], and the automatic estimation of a voxel-specific level of sparsity k_i ($\ll N$) using the minimum description length (MDL) criteria [5]. MDL criteria is a model order selection method that can be effectively applied for sparse models of functional brain data, by finding a model parameter that gives a best tradeoff between the complexity of the model and goodness-of-fit [17]. See Appendix A for details on the estimation of a data-driven dictionary using SDL algorithms. After the estimation of $\mathbf{\Omega}$ and $\mathbf{X}$, *hubness* of the voxel *i* can be actually quantified by the sparsity k_i, as it indicates the number of RSN atoms associated to the voxel. However, the estimation of *k*-hubness may be problematic if the estimated dictionary $\hat{\mathbf{\Omega}}$ includes some atoms representing structured physiological noise.

2.2. Step 2: evaluation of noise-characteristics in the atoms using spatial priors and stepwise regression

The objective of this step was to select from $\hat{\mathbf{\Omega}}$ (the output of SDL in Step 1) a subset of atoms $\hat{\mathbf{\Omega}}_{noise}$ that were associated to the target noise, while preserving meaningful signals in $\hat{\mathbf{\Omega}}_{sig}$, such as $\hat{\mathbf{\Omega}} = \hat{\mathbf{\Omega}}_{sig} \cup \hat{\mathbf{\Omega}}_{noise}$. This step was inspired by the method to classify and remove structured noise from fMRI after ICA decomposition (CORSICA) [15]. The proposed procedure in Step 2 consisted of three sub-steps: (Step 2.1) Extracting several noise-characteristic signals within a spatial prior mask, (Step 2.2) Stepwise regression procedure to select atoms that explains the noise-characteristic signals extracted from the mask, and (Step 2.3) Determining which SDL atoms contribute significantly to the noise-characteristic signals within the mask using the information given by Steps 2.1 and 2.2. The summary of the proposed stepwise regression procedure in Steps 2 is illustrated in Fig. 2.

Step 2.1. Extracting noise-characteristic signals using spatial prior:

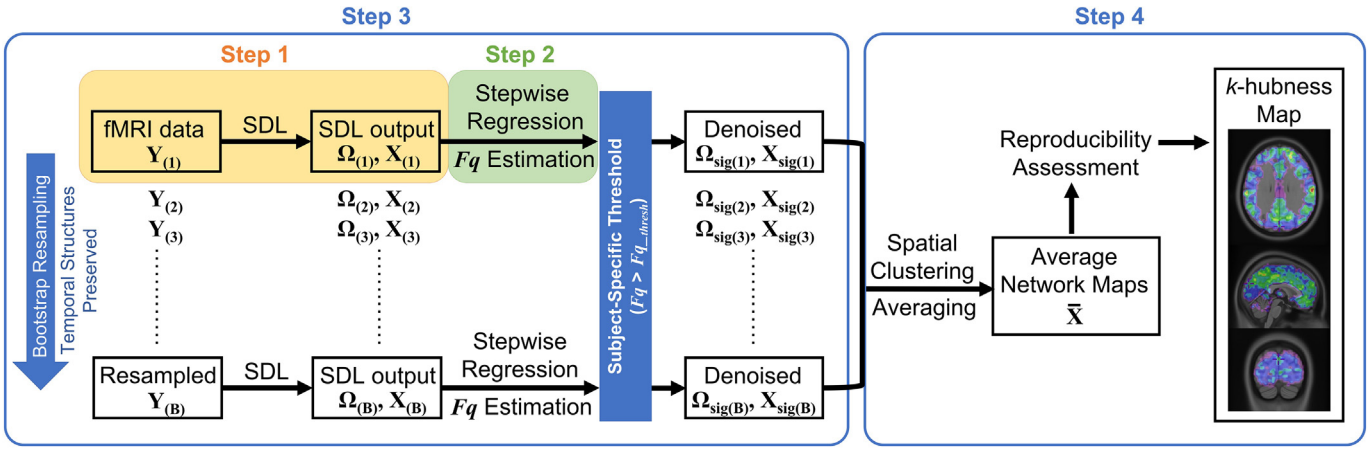

Fig. 1. Summary of the proposed automatic classification of physiological noise-related atoms for SPARK. **Y** denotes the time-points (T) by voxels (V) array of whole brain fMRI data. **Ω** denotes a $T \times N$ data-driven dictionary including N time-course atoms. **X** is a $N \times V$ sparse coefficient matrix, including the spatial maps of resting state networks corresponding to the atoms in **Ω**. B is the number of resampled datasets generated using a circular bootstrap resampling strategy, which was developed for resampling the BOLD signals with temporal dependencies. Fq is the frequency of selection score estimated for each atom in **Ω** using a stepwise regression procedure and spatial prior. Fq_{thresh} is a subject-specific threshold for Fq scores, estimated using the bootstrap resampling strategy. All atoms ω_j exhibiting a Fq score higher than the threshold Fq_{thresh} are classified as noise-related atoms and discarded from **Ω**, thus providing a de-noised dictionary $\mathbf{\Omega}_{sig}$ and the corresponding $\mathbf{X}_{sig}$. After collecting the B de-noised outputs and extracting the most reproducible patterns of FC across the resampled datasets, a k-hubness map is obtained by counting the number of consistent resting state networks involved in each voxel.

To employ an automatic strategy for classifying physiological noise for SDL similar to the one applied for ICA [15], we first modelled noise-characteristic signals through their contribution within a spatial prior mask. In this study, we targeted the sagittal sinus noise as an example, because it was one of the most common types of physiological noise atoms observed in SPARK analysis [5]. We generated a data-driven spatial prior mask using real resting state fMRI data. After applying standard SPARK method on several sets of real data, a large number of estimated atoms (i.e. 1500 atoms) were collected and manually classified by visual inspection on the corresponding spatial maps. Among

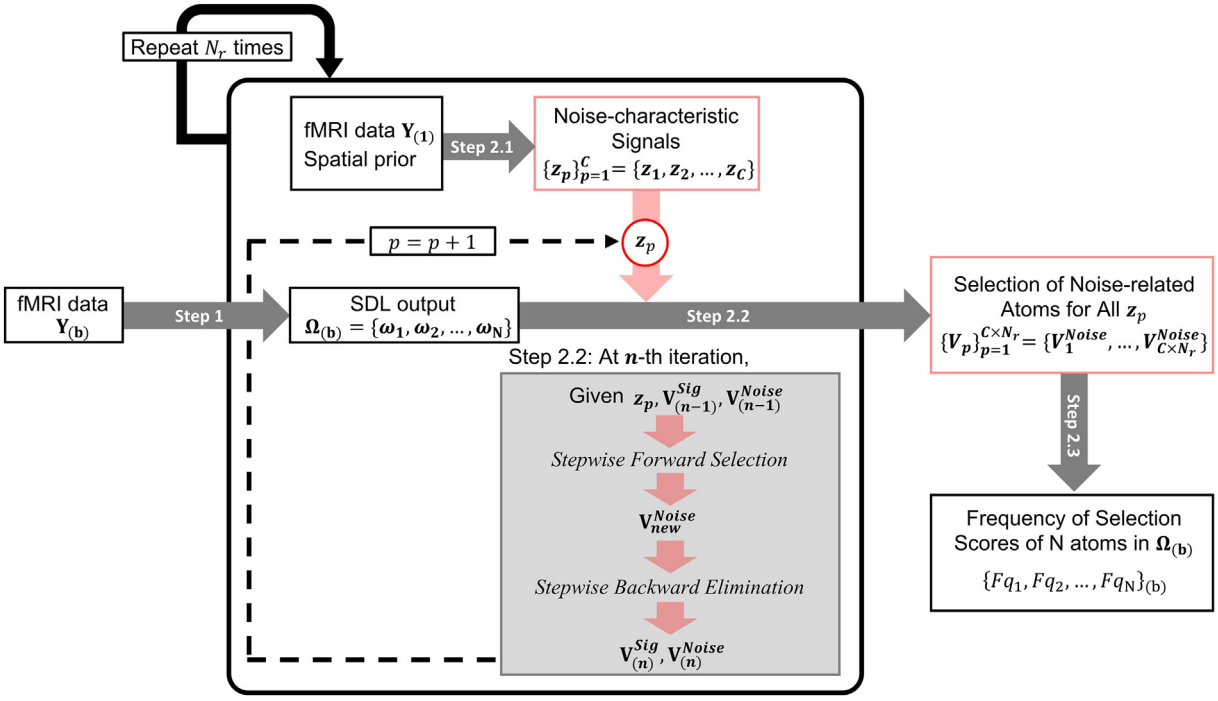

Fig. 2. Summary of the proposed stepwise regression procedure in Step 2. $\mathbf{Y}_{(b)}$ denotes a resampled fMRI data including the T-dimensional preprocessed BOLD time-series in V gray matter voxels. $\mathbf{\Omega}_{(b)}$ denotes the data-driven dictionary including N time-course atoms estimated from $\mathbf{Y}_{(b)}$. In Step 2.1, C noise-characteristic time-courses $\mathbf{z}_p$ were estimated within a spatial prior mask using the C-means clustering algorithm. In Step 2.2, the stepwise forward-backward regression procedure is performed over iterations. See Appendix B for details. This procedure stops when no more significant atom was found in the stepwise forward selection process, thus providing an estimate of $\mathbf{V}_p$, the set of atoms correlating best with the noise-characteristic $\mathbf{z}_p$. Repeating this stepwise regression sequentially for all $\{\mathbf{z}_p\}_{p=1}^{C}$ estimated within the spatial prior mask provided C subsets of selected atoms $\{\mathbf{V}_p\}_{p=1}^{C}$. After all, we repeated Steps 2.1 and 2.2 N_r times and collected $\{\mathbf{V}_p\}_{p=1}^{C}$ estimated in each repetition, in order to reduce the impact of random initialization of the C-means clustering algorithm used in Step 2.1. Finally, Step 2.3 provided the output of Step 2, an N dimensional vector of the frequency of selection scores (Fq) estimated for N atoms, measuring how often each atom appeared across the collected sets of selection $\{\mathbf{V}_p\}_{p=1}^{C \times N_r}$.

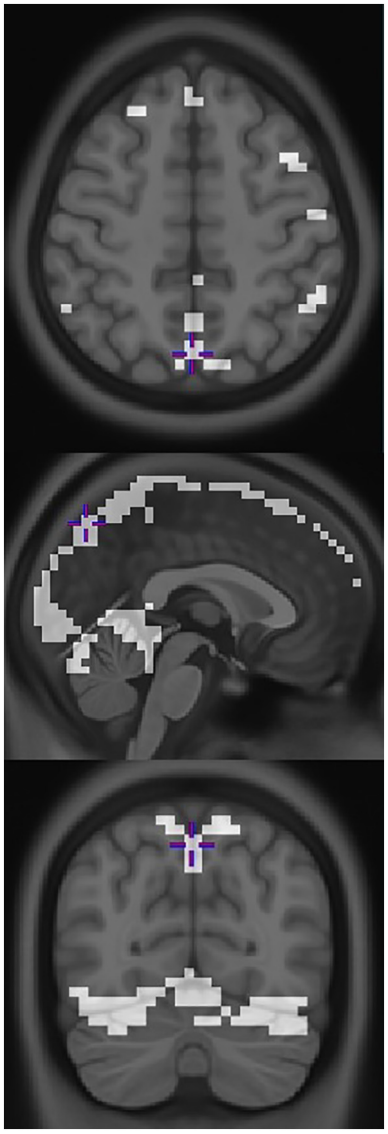

Fig. 3. The data-driven mask of prior including the sagittal sinus noise which was generated using real resting state fMRI data.

those maps, three atom maps that involved only the target region were carefully identified by visual inspection. The three atom maps were averaged in each voxel and converted to a binary map, thus providing a spatial prior mask (Fig. 3).

To model noise-characteristic signals through their contribution within the identified mask, we first reduced the number of time-courses in the mask by applying the standard C-means clustering [15]. The distance d between the signals in two voxels was measured by the temporal correlation between the signals of two voxels: $d^2(\mathbf{y}_i, \mathbf{y}_j) = (1 - \mathrm{corr}(\mathbf{y}_i, \mathbf{y}_j))^2$, where $\mathbf{y}_i$ and $\mathbf{y}_j$ are the preprocessed BOLD time-courses of voxels i and j. The cluster size C was determined by rounding the number of gray matter voxels in the mask divided by 10 to the nearest integer [15,18]. The signals in each cluster were then averaged to generate C time-courses characterizing physiological noise as $\{\mathbf{z}_p\}_{p=1}^{C}$.

Step 2.2. Stepwise regression procedure: After performing SDL as described in Step 1 (Section 2.1), the output of SDL ($\hat{\mathbf{\Omega}}$) involved N time-course atoms. Our objective was to classify and remove any noise-related atoms among the N atoms, while preserving the signal atoms associated with meaningful functional networks. To do so, we employed the stepwise regression procedure proposed in CORSICA [15], whose applicability to fMRI was already demonstrated using ICA. CORSICA proposed a two-step strategy based on an iterative stepwise forward (selection)-backward (elimination) process, which can be useful for detecting only the atoms that have most significant correlation to noise-characteristics signals, while reducing any false detection. Given C noise-characteristic time-courses ($\{\mathbf{z}_p\}_{p=1}^{C}$) extracted by the procedure described in Step 2.1, the proposed stepwise regression procedure was performed for each $\mathbf{z}_p$ and repeated for every $p = 1,\ldots,C$. The main objective of each stepwise regression procedure was to select, among N atoms in $\hat{\mathbf{\Omega}}$, a subset of atoms $\mathbf{V}_p$ that were significantly partially correlated with each noise-characteristic time-course $\mathbf{z}_p$. We repeated this for every $\{\mathbf{z}_p\}_{p=1}^{C}$, thus resulting in a total of C subsets of selected atoms $\{\mathbf{V}_p\}_{p=1}^{C}$. See the summary block-diagram in Fig. 2.

The stepwise regression procedure to estimate $\mathbf{V}_p$ for each $\mathbf{z}_p$ consisted of the stepwise forward selection and stepwise backward elimination processes, which were repeated over iterations. The procedure started with an empty set of selection of noise-related atom. At each iteration, the aim of the stepwise forward selection was to evaluate if any atom in the previous set of non-selected atoms was significantly partially correlated with $\mathbf{z}_p$ and to add the one showing the most significance as a new selected candidate. After that, the aim of backward process was to eliminate one or more atoms that were already selected in the previous iteration but were no longer significant in the updated set of selection, for which a new atom was added by the forward selection in current iteration. In other words, the backward elimination was performed to remove any redundancy that might occur in the new selection by adding the new atom in the forward process. At the end of each iteration (n), this procedure defined a set of $q_{(n)}$ selected noise-related atoms (partially correlated with $\mathbf{z}_p$) and a set of remaining $N - q_{(n)}$ non-selected atoms. This iterative procedure stopped when no more significant atom was found in the stepwise forward selection process, thus providing an estimate of $\mathbf{V}_p$, the set of atoms correlating best with the noise-characteristic $\mathbf{z}_p$. A detailed description of this stepwise forward-backward procedure is described in Appendix B. Repeating this stepwise regression sequentially for all $\{\mathbf{z}_p\}_{p=1}^{C}$ estimated within the spatial prior mask provided C subsets of selected atoms $\{\mathbf{V}_p\}_{p=1}^{C}$, exhibiting the noise characteristic time-courses within the mask.

Finally, as suggested in CORSICA, we repeated Steps 2.1 and 2.2 N_r times and collected $\{\mathbf{V}_p\}_{p=1}^{C}$ estimated in each repetition, in order to reduce the impact of random initialization of the C-means clustering algorithm as suggested in CORSICA. Consequently, this overall process resulted in $C \times N_r$ subsets of selected atoms, $\{\mathbf{V}_p\}_{p=1}^{C\times N_r}$ (e.g. $N_r = 50$).

Step 2.3. Evaluation of the association of SDL atoms to the noise-characteristic signals within the mask: The objective of Step 2.3 was to evaluate the significance of each of the N atoms in $\hat{\mathbf{\Omega}}$ for describing the noise-characteristic time-courses estimated within the spatial prior mask. The collected selection of noise-related atoms $\{\mathbf{V}_p\}_{p=1}^{C\times N_r}$, which was obtained by repeating the procedures described in Steps 2.1 and 2.2, was used to examine the significance of these N atoms. Within the collection $\{\mathbf{V}_p\}_{p=1}^{C\times N_r}$, some atoms might appear multiple times and more frequently than other atoms. We assumed that an atom with great probability of corresponding to a noise atom should occur in most $\mathbf{V}_p$ selections. To measure this occurrence, we obtained for each atom $\hat{\omega}_j$ a score Fq_j measuring the frequency of occurrence of that atom among all selections $\{\mathbf{V}_p\}_{p=1}^{C\times N_r}$. Therefore, the Fq score of an atom reflects the degree to which the noise-characteristic time-courses within the mask are explained by the atom. The frequency of selection score (Fq) was already suggested in CORSICA, providing a value between 0 and 1 [15]. This measure of Fq score actually provided us a strategy to select the SDL atoms that should be classified as noise. We sorted the N atoms with respect to the Fq scores and considered the atoms with supra-threshold scores to be related to the target noise. A threshold (Fq_{thresh}) is therefore critical to avoid any false detection.

2.3. Step 3: bootstrap resampling and estimation of subject-specific threshold of Fq score

The objectives of this step were (*i*) to assess the statistical reproducibility of the SDL atoms by generating a large number of statistically resampled datasets, as it has been suggested for SPARK, and (*ii*) to use this property to incorporate the stepwise regression procedure, finding an adaptive threshold Fq_{thresh} of the *Fq* scores exploiting the large distribution provided by our proposed bootstrap resampling strategy. This step actually includes the main contribution of the present study. Indeed, the way we considered the output of every stepwise regression (i.e. *Fq* scores) differs from the method proposed in CORSICA [15]. Moreover, it should be noted that objectives (*i*) and (*ii*) can be addressed using parallel computations in SPARK (see Fig. 1), thus decreasing significantly computation time.

Specifically, we generated B resampled datasets $\{\mathbf{Y}_{(b)}\}_{b=1}^{B}$ from the original data $\mathbf{Y}$ using a circular block bootstrap [19], which was developed for resampling the BOLD signals with temporal dependencies. The circular block bootstrap randomly selects multiple $h(\ll T) \times V$ blocks from $\mathbf{Y}$ and concatenates to build a new $T \times V$ time-series $\mathbf{Y}_{(b)}$. A choice of block length $h \sim \sqrt{T}$ is recommended to preserve the temporal structure in BOLD time-series [19]. For every resampled dataset $\mathbf{Y}_{(b)}$, SDL in Step 1 was applied to estimate N atoms as in the original implementation of SPARK, then the corresponding Fq_j scores of each atom j were estimated using the stepwise regression procedure described in Step 2. The B surrogates would provide $N \times B$ atoms, and thus a histogram of $N \times B$ *Fq* values. By doing this, instead of measuring only N *Fq* scores from the SDL result using the original data $\mathbf{Y}$, $N \times B$ scores can be measured from the collection of SDL results using B resampled datasets. Note that $\mathbf{Y}_{(b=1)}$ was the original data $\mathbf{Y}$ in our analysis, whereas as all the remaining datasets $\mathbf{Y}_{(b\neq 1)}$ were bootstrap resampled datasets. This provides a larger number of samples to determine statistically a threshold for *Fq* scores, furthermore, the presence of consistent level of score over B times actually represents its statistical reliability at the single subject level.

We decided to approximate the distribution of these *Fq* scores using a *Beta* distribution, such as $Fq \sim Beta(a,b)$ for $a > 0$ and $b > 0$, because the *Fq* scores are continuous variables bounded to the [0,1] interval and the shape of *Fq* distributions from our individual real data was right-skewed (see Fig. 4). Based on a fitted *Beta* distribution, one can determine that an atom involves a significant proportion of noise if the null hypothesis $H_0 : Fq = 0$ is rejected with a significance level $\alpha = Pr(Beta(a,b) > u_{Beta}|H_0)$. A subject-specific threshold could then be determined as $\hat{Fq}_{thresh} = u_{Beta}$. Therefore, for each SDL analysis of the resampled data $\mathbf{Y}_{(b)}$, all atoms ω_j exhibiting a Fq_j score higher than the estimated threshold $\hat{Fq}_{thresh}$ were classified as noise-related atoms and discarded from the analysis, as follows:

$$\hat{\mathbf{\Omega}}_{noise(b)} = \{\hat{\omega}_j \in \hat{\mathbf{\Omega}}_{(b)} \,|Fq_j > \hat{Fq}_{thresh}\}, \forall\, b = 1, \ldots, B. \tag{2}$$

2.4. Step 4: extraction of statistically reproducible networks

After removing the noise-related atoms exhibiting *Fq* scores larger than Fq_{thresh}, we obtained from each resampled data $\mathbf{Y}_{(b)}$ a signal dictionary $\mathbf{\Omega}_{sig(b)}$ and a sparse coefficient matrix $\mathbf{X}_{sig(b)}$, whose dimensions were reduced to $(N - q_{(b)})$ by the number of removed atoms $q_{(b)}$. Step 4 aims at extracting statistically reproducible networks using the collection of outputs from the resampled datasets, as suggested in SPARK. Since the bootstrap resampling in Step 3 assumed that the resampled datasets were generated from the same distribution as that of the measured signals $\mathbf{Y}$, there should be consistent network patterns across the resampled datasets. Specifically, on the collection of all signal atom maps, $[\mathbf{X}_{sig(1)}; \mathbf{X}_{sig(2)}; \ldots \mathbf{X}_{sig(B)}]$, a C-means clustering algorithm with $C = N'$ can be applied as follows:

$$N' = N - \left[\frac{1}{B}\sum_{b=1}^{B}(q_{(b)})\right]. \tag{3}$$

Note that in our previous implementation of SPARK [5], there was no automatic removal of noise-related atoms and therefore, we considered $N' = N$.

After C-means clustering, the spatial maps were averaged in each cluster to generate an average matrix ($\bar{\mathbf{X}} \in \mathbb{R}^{N' \times V}$), which indicates the statistical reproducibility of the amplitudes of signal atoms in each voxel. In this matrix, low values were considered as Gaussian noise of inconsistent elements and removed using a p-value < 0.05, where the p-value was experimentally determined by our previous work [5]. Finally, the last step consisted of counting the number of non-zero elements in each column of $\bar{\mathbf{X}}$, thus providing an estimation of k-hubness, whereas only the number of meaningful networks without interference from physiological noise atoms were taken into account. In the remaining part of this paper, we will denote this entire process described in Steps 1–4 as " SPARK + Denoising ".

3. Material and methods

3.1. Dataset and preprocessing

We carried out our analyses and validation using the New York University test/retest dataset[1] available from the 1000 Functional Connectome Project, which provides three sessions of eyes-opened resting state fMRIs scanned using a 3T MRI from 25 healthy subjects (Age: 29.44 ± 8.6, 10M/15F). The EPI functional volumes were obtained (39 slices, TR/TE = 2000/25 *ms*; flip angle = 90°; matrix = 64 × 64; FOV = 192 *mm*; voxel size = 3 × 3 × 3 *mm*) with 5–16 months interval between sessions 1/2 and 30–45 min interval between sessions 2/3. T1 MPRAGE images were also provided (176 slices; TR/TE = 2500/4.35 *ms*; TI = 900 *ms*; flip angle = 8°; FOV = 256 *mm*) [20]. The data were preprocessed using NIAK 0.12.15 [18]. T1 images were preprocessed using the CIVET pipeline, including non-uniformity correction and non-linear coregistration to the Montreal Neurological Institute (MNI) template [21]. The EPI volumes were corrected for slice-timing and rigid-body motion was estimated within and between sessions using the median volume of the first session as reference. The median volume of a fMRI run was then coregistered with a T1 individual session. The functional data were resampled at a 4 *mm* isotropic resolution. Volumes with frame displacement greater than 0.5 *mm* were removed by scrubbing [22]. Slow time drifts (< 0.01 *Hz*) were filtered out and average signals in the white matter/ventricles and motion parameters were regressed out from the signals in each voxel. Spatial smoothing was finally applied with a Gaussian kernel (8 *mm* FWHM). As a result, a preprocessed data matrix $\mathbf{Y}$ consisting of $T = 197$ or less time samples (if some time-frames were removed by scrubbing) by $V = 17{,}388$ gray matter voxels was obtained for each run.

3.2. Standard SPARK procedure with manual classification of noise-related atoms

The 75 resting state fMRI runs were individually analyzed using standard SPARK procedure [5]. For each run, we generated $B = 300$ resampled datasets using the circular block bootstrap with random block lengths (h) varying between 10 and 30. We used a variant of K-SVD algorithm to estimate $N = 20$ atoms from each dataset [5]. The use of this algorithm for SDL was already proposed and validated in our previous work (see Appendix A). Two experienced reviewers (Reviewer 1: KL and Reviewer 2: HK) classified all atoms by visual inspection on the spatial maps using the following criteria: (i) Remove atoms that included spatial overlaps with large vessels (e.g. the sagittal sinus

[1] https://www.nitrc.org/projects/nyu_trt/

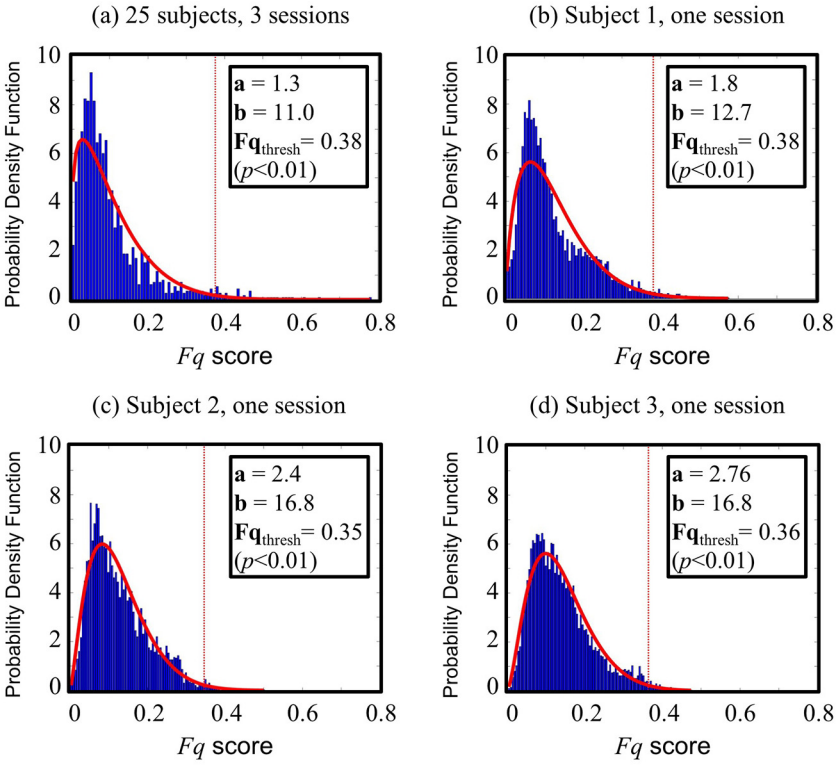

Fig. 4. The distribution of *Fq* scores estimated using real resting state fMRI data acquired from (a) 75 runs from 25 healthy subjects at three different sessions, (b–d) 300 bootstrap resampled runs generated using a resting state fMRI run obtained from an individual subject; three examples from three subjects are shown. In all cases (a–d), 20 SDL atoms were estimated from every resting state fMRI run, and a *Fq* score was estimated for each atom. Red curve indicates a fitted *Beta* distribution to this example datasets. Red dotted vertical line indicates a threshold (Fq_{thresh}) estimated using the fitted *Beta* distribution with 99% confidence intervals. *a* and *b* are the estimated parameters of the fitted Beta distribution for each data. The number of bins in each histogram is 100.

noise), scattered signals with peak intensities at the boundary of gray matter voxels, other structured noise (e.g. acquisition/motion artifacts) as suggested by Griffanti et al. [14], and also atoms involving extremely small number of active voxels (e.g. less than 30); (ii) Whenever both signal of interest and noise seemed to be present in an atom, we decided to be conservative and to keep those atoms in order not to lose any important signals. Finally, after removing the classified noise related atoms, a *k*-hubness map was estimated from each run.

3.3. Defining a spatial prior mask

We targeted the sagittal sinus noise, which was one of the most common types of physiological noise atoms estimated using SPARK [5]. We generated a data-driven spatial prior mask using the results in Section 3.2. All estimated atoms from the results (25 subjects × 3 fMRI runs × 20 atoms = 1500 atoms) were manually classified by visual inspection on the spatial maps. Among those atom maps, three atom maps that involved only the target region were carefully identified in agreement of three reviewers (KL, HK and CF). The three atom maps were averaged in each voxel and converted to a binary map, thus providing the binary spatial prior mask selected for our method (Fig. 3).

3.4. Analysis 1: performance evaluation of the proposed method for SDL using 10-fold cross-validation

We conducted a 10-fold cross-validation to evaluate and validate the performance of the proposed stepwise regression procedure on the atoms estimated using SDL (Step 1 in Section 2.1), without proceeding to the next steps that incorporated the bootstrap resampling based strategy within the SPARK framework. We performed the automatic classification of sagittal sinus noise-related atoms using the proposed method with $N_r = 50$, and $N(= 20)$ *Fq* scores were estimated for each run. This cross-validation also aimed at determining a significance level (α) for estimating a threshold Fq_{thresh} of *Fq* scores, which is a critical parameter for the proposed classification of physiological noise-related atoms. To perform a 10-fold cross-validation, the 75 resting state fMRI runs were randomly partitioned into 10 equal sized subsets. For each round of cross-validation, one subset was used as a test dataset, while the remaining 9 subsets were used for training. This process was repeated 10 times. The goal of each round was (i) to estimate Fq_{thresh} from the distribution of *Fq* scores obtained from a training dataset and (ii) to use the estimated threshold for thresholding *Fq* scores of the atoms obtained from a test dataset and classifying noise-related atoms exhibiting a *Fq* score higher than the threshold. Since the generation of training and test datasets were performed including three fMRI runs from each subject, this validation was performed to take into account a mixed effect of intra- and inter-subject variability of the estimation of Fq_{thresh}.

In each round of cross-validation, only the scores estimated from the runs belonging to the training data were collected and used to draw the distribution of the *Fq* scores. The histogram of *Fq* scores was fitted to a *Beta* distribution using the maximum likelihood estimation of unknown parameters *a* and *b* [23]. Given a specific level of significance (α), a threshold Fq_{thresh} was estimated using the fitted $Beta(\hat{a}, \hat{b})$ distribution. Then, the atoms estimated from the test dataset were classified as noise-related atoms when $Fq > \hat{F}q_{thresh}$. After the 10 rounds of evaluation, we could collect the classification results for the 1500 atoms using the proposed method, because each subset was used exactly once as the test dataset in one round. We computed the sensitivity and specificity of the proposed method in comparison to the manual classification performed by reviewer 1. Varying the level of significance (α), this 10-fold procedure was repeated to measure the specificity and sensitivity for every possible threshold, thus providing a Receiver Operating Characteristic (ROC) curve.

3.5. Analysis 2. SPARK + Denoising: application of the proposed automatic classification and removal of physiological noise within the SPARK framework

The main objective of Analysis 2 was to compare the performance of SPARK in estimating functional hub structures with *versus* without the application of our proposed automatic classification of physiological

noise-related atoms: SPARK + Denoising (Section 3.5) *versus* SPARK (Section 3.2). In Analysis 2, we used the same datasets and SPARK parameters as used for the standard SPARK procedure in Section 3.2. For the application of our proposed automatic physiological noise classification for SPARK, all Steps 1–4 described in Section 2 were performed on each fMRI run. The histogram of $N(=20) \times B(=300)$ *Fq* scores collected across the resampled datasets was fitted to a *Beta* distribution. Then, Fq_{thresh} was estimated for each individual at 99% confidence intervals ($\alpha = 0.01$). The choice of $\alpha = 0.01$ was determined using the results from our cross-validation. After removing the classified noise-related atoms with *Fq* higher than $\hat{F}q_{thresh}$, a *k*-hubness map was estimated from each run. The group average *k*-hubness maps were obtained by averaging *k*-hubness values in each voxel across the 75 runs. Finally, reviewer 1 classified all those estimated atoms again from the results using the SPARK + Denoising method by visual inspection on the spatial maps using the criteria described in Section 3.2.

We also evaluated if the level of significance (α) used for estimating Fq_{thresh} from the distribution of *Fq* scores across subjects in our cross-validation could be actually used in the individual analysis using the entire SPARK + Denoising method for estimating Fq_{thresh} from the distribution of *Fq* scores within a subject. While our cross-validation estimated a threshold Fq_{thresh} from the distribution of *Fq* scores over all subjects in the training dataset, we wish to examine if such level of significance was also valid for the distribution of *Fq* scores over all bootstrap resampled datasets within a single subject. Varying $\alpha = 0.05$, 0.01 and 0.005, we performed Analysis 2 to compare the distribution of 10 estimated thresholds $\hat{F}q_{thresh}$ using SDL in our cross-validation and the distribution of 75 $\hat{F}q_{thresh}$ values estimated from 75 individual runs using the SPARK + Denoising method in Analysis 2.

4. Results

On real resting state fMRI data acquired from 25 healthy subjects at three different runs, we first evaluated the performance of the proposed stepwise regression procedure for SDL using the 10-fold cross-validation. Then, we plugged the entire procedure within the SPARK framework and evaluated the performance of the proposed stepwise regression procedure for every resting state fMRI run in the same datasets, when compared to the manual classification performed by human reviewers.

4.1. Result 1: 10-fold cross-validation

Fig. 5 shows the results from the cross-validation using the 1500 atoms (20 atoms × 25 subjects × 3 runs) estimated from the validation datasets. From the output of SDL obtained for this cross-validation, reviewer 1 classified 49/1500(3.3%) atoms as representing sagittal sinus noise, corresponding to 49/75(65%) of the analyzed fMRI runs, as presented in Fig. 6 (a). On these outputs of SDL, we applied the proposed stepwise regression procedure and performed our 10-fold cross validation. Collecting results from the 10-fold tests varying the threshold Fq_{thresh}, the area under the ROC curve (AUC) was 0.89 (Fig. 5 (a)). The ROC curve exhibited a prominent corner when $\alpha = 0.05$ was used: the proposed stepwise regression procedure could classify sagittal sinus noise-related atoms with a false positive rate (1-specificity) of 0.03 and a sensitivity of 0.9.

Next, we evaluated if the level of significance (α) used for estimating adaptive threshold Fq_{thresh} in our cross-validation could be actually used in the individual analysis using the SPARK + Denoising method (Table 1). Using this cross validation, the mean and standard deviation of the estimated 10 Fq_{thresh} values across the random 10 folds were 0.27 ± 0.003 using $\alpha = 0.05$, 0.38 ± 0.003 using $\alpha = 0.01$ and 0.42 ± 0.004 using $\alpha = 0.005$. Using the SPARK + Denoising method on the 75 individual resting state fMRI runs, the mean and standard deviation of the estimated 75 Fq_{thresh} values were 0.32 ± 0.05 using $\alpha = 0.05$, 0.42 ± 0.07 using $\alpha = 0.01$, and 0.46 ± 0.08 using $\alpha = 0.005$. Two-sided two-sample *t*-tests on the estimated $\hat{F}q_{thresh}$ values using the two analyses (cross-validation *versus* SPARK + Denoising method) showed a significant difference ($p < 0.05$) when the value of $\alpha = 0.05$ was used for the estimation of Fq_{thresh}. The use of $\alpha \leq 0.01$ showed no significant difference between them, suggesting good agreement between inter-subject variability assessed through cross-validation and distribution of threshold estimated for every single run among bootstrap sample. Therefore, the choice of $\alpha = 0.01$ was made in order to maximize specificity with little loss of sensitivity for our further analysis (specificity: 0.99 and sensitivity: 0.67), and to keep our classification conservative.

4.2. Result 2: SPARK with versus without the proposed automatic classification and removal of physiological noise-related atoms

We evaluated the proportion of runs including sagittal sinus noise-related atoms estimated from individual data using manual classification, when three different strategies were considered: SDL, SPARK, and SPARK with the proposed automatic physiological noise removal (SPARK + Denoising). We used $\alpha = 0.01$ for the results using the SPARK + Denoising method for this comparison. Reviewer 1 manually classified sagittal sinus noise atoms in the outputs of SDL and SPARK. The manual classification for the results from the output of SDL by reviewer 1 was actually used in the cross-validation. Reviewer 2 independently classified sagittal sinus noise-related atoms in the output of SPARK. As a result, when using SPARK, the reviewers 1 and 2 found 50/1500(3.3%) and 47/1500(3.1%) sagittal sinus noise-related atoms corresponding to 49/75(65.3%) and 46/75(61.3%) runs (Fig. 6 (a)). With respect to the sagittal sinus noise-related atoms, 90% of the two reviewers' classifications were in agreement (Fig. 6 (b)). Once the automatic classification and removal of the sagittal sinus noise-related atoms were applied using the SPARK + Denoising method on the same datasets, manual analysis was performed again by reviewer 1; the proportion of remaining sagittal sinus noise-related atoms remarkably decreased to 14/1460(1.06%) corresponding to 14/75(19%) runs.

We finally compared the group average *k*-hubness maps obtained using the three analyses: SPARK without any atom removal, SPARK with manual removal of sagittal sinus noise-related atoms, and SPARK with the proposed automatic removal of sagittal sinus noise-related atoms (see Fig. 7 (a)). An example of *k*-hubness maps estimated using the three analyses from an individual run is also shown in Fig. 7 (b). In both individual- and group-level comparisons, the *k*-hubness values at the voxels located at the regions involving the sagittal sinus were remarkably decreased when compared to those estimated without any noise removal, without affecting the measurement of *k*-hubness values in the remaining part of the cerebral cortex.

5. Discussion and conclusions

We proposed an automatic classification and removal of physiological noise-related atoms for resting state fMRI connectivity analysis using sparse dictionary learning. Based on sparse dictionary learning for the sparse general linear model of resting state fMRI, the recently developed SPARK method suggested that connector hubs of brain networks can be identified by measuring *k*-hubness, the number of atoms involved in each voxel [5]. However, such hub analysis can be corrupted by the presence of atoms reflecting physiological noise in the dictionary. Manual classification of noise-related atoms is time-consuming and subjective to reviewer's experience and knowledge. Motivated by the fact that the physiological fluctuations are often localized in tissues close to large vessels, we developed an automatic classification of physiological noise-related atoms using spatial priors and stepwise regression. The proposed stepwise regression procedure provides the estimation of frequency of selection score for each atom, measuring

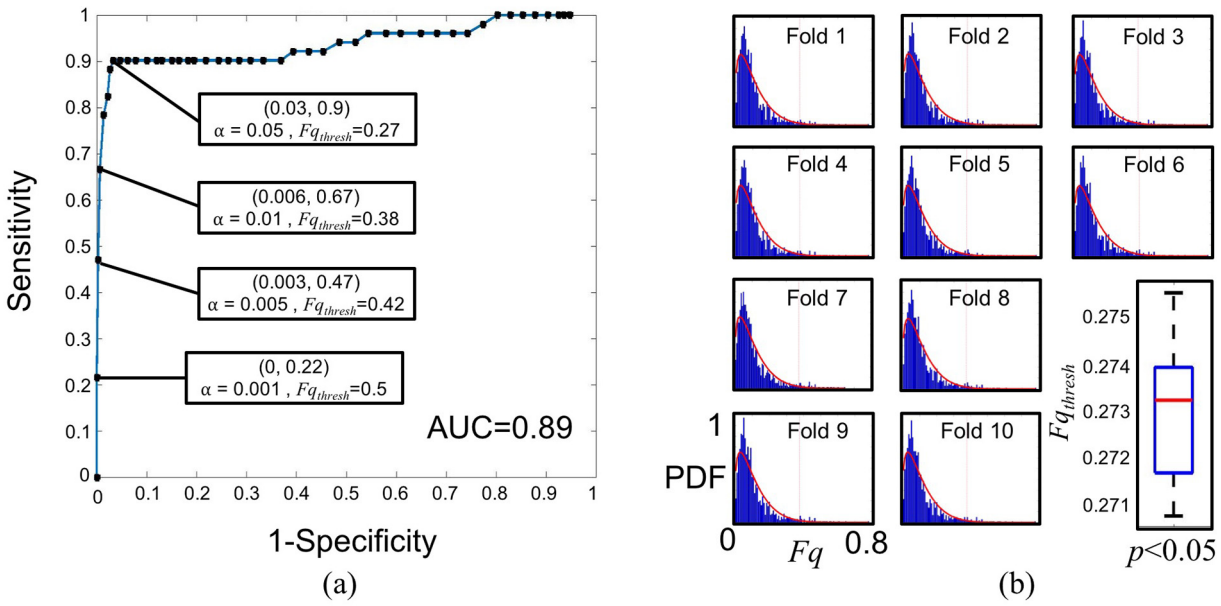

Fig. 5. 10-fold cross-validation results on 75 individual runs of resting state fMRI data acquired from 25 healthy subjects at three different runs. (a) ROC curve estimated by collecting results from the random 10-fold test datasets. Sensitivity = TP/(TP + FN). Specificity = TN/(TN + FP). T: true, F: false, P: positive, N: negative in comparison to the manual classification. At four different levels of α, the average of ten Fq_{thresh} values estimated over the random 10-folds are shown within the ROC curve. (b) The distribution of frequency of selection scores (Fq) in each of the ten training dataset. PDF: Probability density function. Red curve indicates a fitted *Beta* distribution to each training dataset. The number of bins in each histogram is 100. (Right-bottom) The box-plot of ten $\hat{Fq}_{thresh}$ values estimated from the ten rounds of cross-validation using the fitted *Beta* distribution with 95% confidence intervals (α = 0.05).

the degree to which each atom is partially correlated with noise-characteristic signals within a spatial prior mask. Our main contribution includes the estimation of adaptive threshold of the frequency of selection scores using the bootstrap resampling based strategy proposed in the SPARK framework. Due to the discrete nature of *k*-hubness and its small range, i.e. typically between 0 and 5, a change in *k*-hubness by the inclusion or exclusion of an atom may have a large impact on hubness measures, especially when compared to change in other continuous hubness metric, such as degree centrality ranging between 0 to several hundred. Although between-subject variability in the estimated thresholds were small (Table 1), we believe that such small variations may be important in the estimation of subject specific *k*-hubness using the output of this classification. In this regard, a choice of adaptive classification threshold for each individual is important for the hub analysis using SPARK. The proposed strategy provided the automatic classification of physiological noise-related atoms for individual fMRI data without requiring any external recording of physiological fluctuations nor training data. Finally, our proposed de-noising procedure suggested an improvement of the estimation of individually reliable connector hubs using *k*-hubness, whereas only the number of meaningful networks without interference from physiological noise atoms were taken into account.

In CORSICA, they suggested an automatic estimation of adaptive threshold by using Otsu's algorithm, which divides the histogram of Fq scores in two classes by minimizing the interclass variance [15]. In their study, using the scores assigned to each of the first 60 independent components estimated using ICA, they showed that the adaptive threshold allowed one to increase the sensitivity of the selection compared to that obtained with a fixed threshold. However, in our SDL based sparse GLM, the number of atoms (N), i.e. ~ 20 atoms, is often smaller than the number used for ICA and the distribution of Fq scores was actually skewed in real data (see Fig. 4). To address this, we proposed the bootstrap resampling based strategy to generate a larger number of Fq scores for each individual and estimated an individually adaptive threshold of Fq scores from the resampled datasets.

Our results and cross-validation using real data demonstrated that

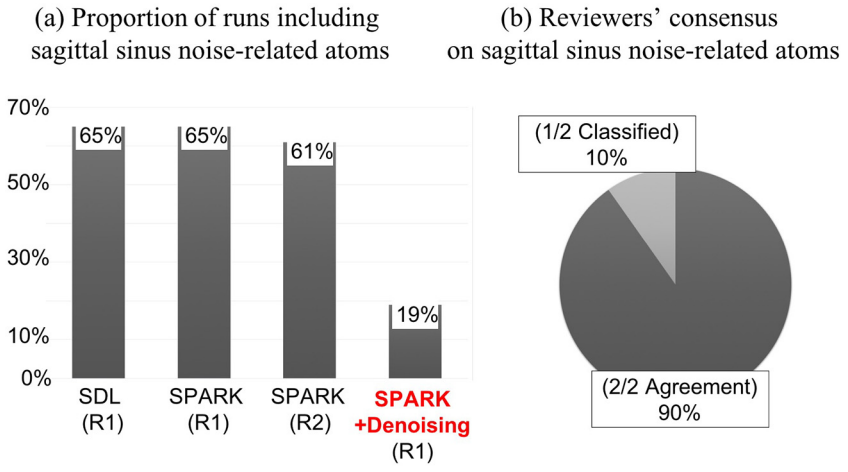

Fig. 6. Results of the automatic classification of the sagittal sinus noise-related atoms estimated using SPARK on individual resting state fMRI data. (a) The proportion of runs including sagittal sinus noise-related atoms estimated from individual data using manual classification, when three different strategies were applied: SDL, SPARK, and SPARK with the proposed automatic SPARK + Denoising method. Reviewer 1 (R1) manually classified sagittal sinus noise-related atoms in the outputs of SDL and SPARK. Reviewer 2 (R2) independently classified sagittal sinus noise-related atoms in the output of SPARK. Once the automatic classification and removal of the sagittal sinus noise-related atoms were applied for SPARK (SPARK + Denoising), manual noise classification was performed again by R1. (b) The two reviewers' sagittal sinus noise classifications for the output of SPARK were in agreement.

Table 1
The mean and standard deviation of the estimated thresholds Fq_{thresh} using (a) our cross-validation of the proposed stepwise regression based classification of physiological noise-related atoms from the output of SDL, and (b) the SPARK + Denoising method, which combined the proposed stepwise regression procedure combined with the bootstrap resampling based strategy within the SPARK framework. α values in the first column indicates the level of significance used for the estimation of $Fq_{threshold}$ from the fitted *Beta* distribution for each individual data. The second column presents the mean and standard deviation of 10 Fq_{thresh} values estimated using the method (a) in our 10-fold cross validation for the 75 resting state fMRI runs (25 subjects × 3 sessions). The third column presents the mean and standard deviation of 75 Fq_{thresh} values estimated for each of those 75 runs using the method (b). The last column indicates the results from two-sample *t*-test (two-sided, $p < 0.05$) between the values estimated using (a) and (b).

α	(a) Cross-validation Fq_{thresh}	(b) SPARK + Denoising Fq_{thresh}	Two-sample *t*-test ($p < 0.05$)
0.05	0.27 ± 0.003	0.32 ± 0.05	Significant
0.01	0.38 ± 0.003	0.42 ± 0.07	Not significant
0.005	0.5 ± 0.004	0.46 ± 0.08	Not significant

the proposed method could identify atoms including the target noise localized in the sagittal sinus. In our 10-fold cross-validation (Section 3.4), we first evaluated the performance of the proposed stepwise regression procedure using SDL. The area under the ROC curve was 0.89, suggesting good specificity of the proposed method in detecting atoms involving the sagittal sinus noise. The ROC curve also suggested the choice of significance level for estimating a threshold Fq_{thresh} of Fq scores from the outputs of SDL, such as $p < 0.05$. Furthermore, varying the level of significance α, we evaluated if the level of significance used for estimating Fq_{thresh} in our cross-validation could be actually used in the individual analysis using the SPARK + Denoising method. We found that the use of $\alpha \leq 0.01$ suggested a good level of confidence for both our cross-validation and individual data analysis using the SPARK + Denoising method. Based on these evaluations, the choice of $\alpha = 0.01$ was made in order to maximize specificity with little loss of sensitivity for our proposed method: we obtained the specificity of 0.99 and sensitivity of 0.67 in our analysis.

When no classification of noise-related atoms was applied, the presence of sagittal sinus noise-related atoms in the output of SDL was consistently found in 65% of the runs using the manual classification (Fig. 6). Reviewer 1 found the similar sets of the sagittal sinus noise-related atoms in both levels of analyses (SDL and SPARK), demonstrating that the structured physiological noise associated with the sagittal sinus was consistently found across the resampled datasets. Notably, reviewer 1 found the sagittal sinus noise-related atoms in 90% agreement with those classified by reviewer 2 from the outputs of SPARK. Variability of the manual classification between the two reviewers resulted from several atoms that involved both the sagittal sinus noise and other cortical regions, such as parts of the visual or parietal cortices adjacent to the sagittal sinus. Given the criteria of the manual classification, both reviewers agreed and decided to be conservative and to keep those atoms in order not to lose any important signals, while different decisions were made for some atoms with mixed signals and noise. Considering the high agreement in the manual classification by the two reviewers and the lack of ground truth, we decided to use the classification results by reviewer 1 for the evaluation of performance of the proposed method in our further analyses. Therefore, reviewer 2 did not perform the manual classification on the results using the SPARK + Denoising method.

Indeed, our further analyses on individual subjects (Fig. 6) using the proposed SPARK + Denoising method demonstrated a good performance using the choice of $\alpha = 0.01$. When the SPARK + Denoising method was applied for individual resting state fMRI runs, the proportion of sagittal sinus noise-related atoms was remarkably reduced to 19%. Finally, our method suggests a de-noised estimation of *k*-hubness for each voxel, including only the number of meaningful networks without interference from physiological noise atoms. Using the SPARK + Denoising method, the resultant group average *k*-hubness map in Fig. 7 clearly showed decreases in *k*-hubness values at the voxels

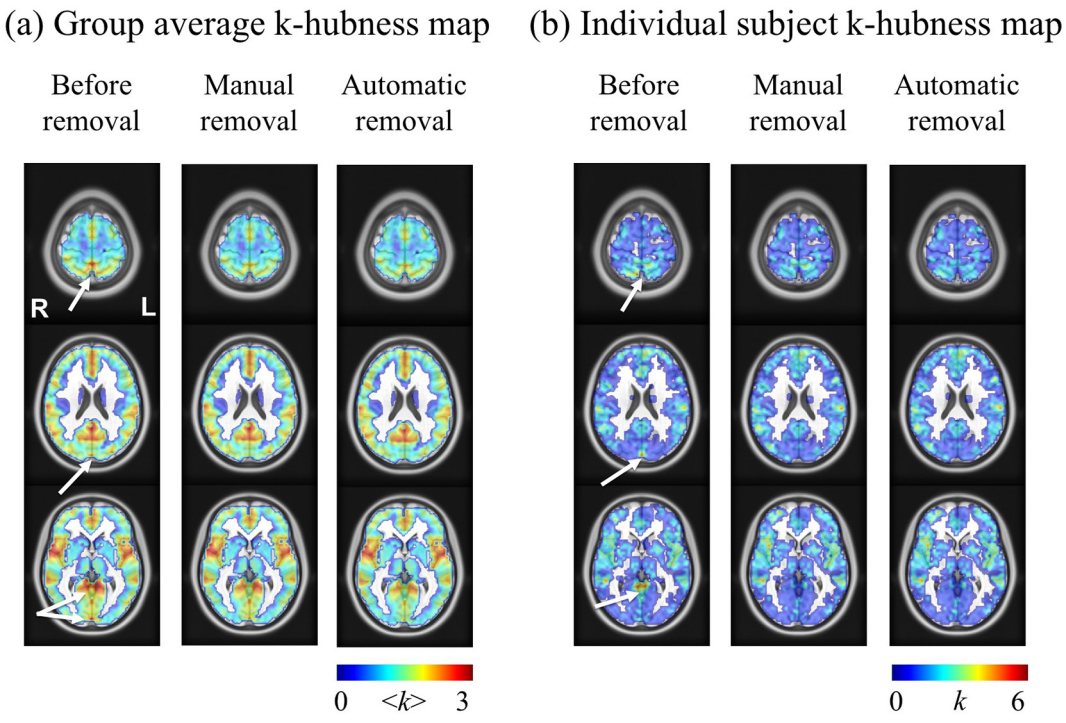

Fig. 7. Comparisons of the *k*-hubness maps obtained using the three analyses: SPARK without any atom removal, SPARK with manual removal of sagittal sinus noise-related atoms, and SPARK with the proposed automatic removal of sagittal sinus noise-related atoms. (a) Group average *k*-hubness maps using the three analyses and (b) an example of *k*-hubness maps from an individual run. $< k >$ is an average *k*-hubness measured in each voxel over the 75 runs (25 healthy subjects with three resting state fMRI sessions).

located at the regions involving the sagittal sinus, while the *k*-hubness values in the remaining part of the cerebral cortex were not affected. Even though the two reviewers conducted the manual classification of several types of noise-related atoms including the sagittal sinus noise as described in Section 3.2, we did not remove other types of classified noise-related atoms for generating the *k*-hubness maps in Fig. 7. Therefore, only the effect of classifying and removing the sagittal sinus noise-related atoms were taken into account for the comparisons in Fig. 7. The *k*-hubness maps estimated using SPARK with the proposed automatic noise removal method were similar to those estimated using SPARK with manual classification, suggesting a reliable performance in detecting atoms involving physiological noise present in the sagittal sinus.

Nevertheless, a few false detections occurred when both signals and noise were present in one atom. Classification of noise-related atoms from sparse dictionary learning assumes that signal- and noise-components present in data are sufficiently separated into different atoms. The performance of vK-SVD sparse dictionary learning algorithm for separating signal and noise was carefully validated in our previous work [4,5]. These validations included simple 2-dimensional box simulations varying the level of Gaussian noise, realistic simulations that introduced artificial hubs on real fMRI data varying the level of structured perturbation or noise, and finally using real fMRI data. However, if a significantly large proportion of noise is present in data after preprocessing, it is possible that such a clear separation could not be obtained. For example, in our manual classification (Section 3.2), reviewer 1 classified 0.03% of the estimated atoms (38/1500) as unclear (if they were signal- or noise-dominant) and decided to remove only 0.01% (20/1500) of them. Reviewer 2 classified 0.06% (88/1500) as unclear and decided to remove only 0.02% (30/1500) of them. This suggests that the performance of either manual or automatic classification depends on the presence of noise in acquired data. A final level of visual inspection on the automatically classified atoms using the proposed method may help confirm and improve the estimation of *k*-hubness and subsequent hub analysis. It would be also interesting to introduce additional recordings of cardiac and respiratory signals to constrain the dictionary training scheme to improve the performance of proposed method.

We tested a specific type of structured noise localized within the sagittal sinus. However, as suggested in Perlbarg et al. [15], this method has general applicability to multiple masks for different types of structures for the step-wise regression. For example, they manually designed two masks of interest on the individual T1-weighted images as prior information for CORSICA. One mask included the first three ventricles to detect especially respiration-related movements, and another mask included the brainstem to detect cardiac fluctuations. Assessing the impact of the choice of the spatial prior mask on *k*-hubness estimation could be of interest, for example, using an individual anatomical vascular MRI to define a mask of sagittal sinus noise; however, this was out of the scope of the present study.

Even if our sagittal sinus noise-related spatial prior was generated from the data, it was actually estimated by averaging three visually selected atom maps from three different fMRI sessions, among the 75 sessions (25 subjects/3 sessions each), therefore only very limited circularity was considered. We do not expect the choice of this spatial prior to bias our analysis, whereas available atlases of the vasculature [24] could also be considered to define this spatial prior and will be considered in our future investigations. It is also important to note that our sagittal sinus noise-related spatial prior was defined in the stereotaxic space (the MNI template space), and the proposed atom classification was also applied for SPARK analysis in the stereotaxic space. In our experience, the localization of sagittal sinus noise in the estimated dictionary atoms is consistent as long as the SPARK procedure is performed in the stereotaxic space, i.e. the MNI template space. This suggests that this prior map should work equally well for other datasets preprocessed in a strategy similar to our preprocessing procedure. However, if one wants to analyze resting state fMRI data in individual and/or native space using SPARK, it would be appropriate to estimate and identify adaptive spatial priors, for instance using individual MRI angiography.

In conclusion, we proposed the automatic classification and removal of physiological noise-related atoms for SDL from real resting state fMRI data, using the stepwise regression procedure within the SPARK framework with spatial priors. Our real data analyses using the 10-fold cross-validation for SDL atoms and individual subject analyses using the combined SPARK + Denoising method based on the bootstrap resampling based strategy showed a reliable and robust classification of sagittal sinus noise-related atoms. We clearly demonstrated decrease in *k*-hubness values in the voxels involved in the sagittal sinus at both individual and group levels, whereas only the number of meaningful networks without interference from physiological noise atoms were taken into account. These results suggest a more reliable and denoised estimation of functional connector hubs in resting state functional MRI connectivity by combining the proposed denoising strategy with SPARK.

Acknowledgment

This research was supported by the Canadian Institutes of Health Research MOP-130442 and 133619, Natural Sciences and Engineering Research Council of Canada Rgpin 35661 and Max E. Binz Fellowship. H.M.K. is supported by Mark Rayport and Shirley Ferguson Rayport fellowship (Canada) and was supported by the Preston Robb fellowship (Canada) and the Uehara Memorial Foundation research fellowship (Japan).

Appendix A. Application of sparse dictionary learning algorithms for the sparse general linear model

For the sparse GLM, $\mathbf{\Omega}$ and $\mathbf{X}$ that best describe the data can be estimated using a SDL algorithm such as K-SVD [16] that solves the following optimization problem:

$$\min_{\mathbf{\Omega},\mathbf{X}} \|\mathbf{Y} - \mathbf{\Omega X}\|_F^2, \quad \text{subject to } \forall\, i\text{: } \|\mathbf{x}_i\|_0 \le k_i. \tag{A.1}$$

Using SPARK, we previously proposed and validated the automatic estimation of a voxel-specific level of sparsity k_i ($\ll N$) for the sparse GLM using the MDL criteria [5]. The K-SVD algorithm usually finds a sparse solution using greedy algorithms, such as the matching pursuit [25] and orthogonal matching pursuit [26], which compute inner products between the BOLD signals in voxel i and dictionary columns with the sparsity constraint k_i. Such approaches assumes k_i as the maximum number of networks allowed in voxel i (i.e. $\|\mathbf{x}_i\|_0 \le k_i$). We recently proposed a variant of K-SVD which can estimate voxel-specific optimal levels of sparsity using MDL criteria [5]. Taking advantage of the simple thresholding approach and the variant of K-SVD algorithm, *hubness* of the voxel i can be actually quantified by the sparsity k_i, as it indicates the number of RSN atoms associated to the voxel.

Appendix B. Stepwise regression procedure

Stepwise forward selection. At (n)-th iteration, given the previous sets of selected atoms $\mathbf{V}_{(n-1)}^{Noise}$ and non-selected atoms $\mathbf{V}_{(n-1)}^{Sig}$, we selected an

atom $\hat{\omega}_j$ from $\mathbf{V}_{(n-1)}^{Sig}$ to add it as a new selected candidate, as:

$$\mathbf{V}_{(n-1),+j}^{Noise} = \mathbf{V}_{(n-1)}^{Noise} \cup \hat{\omega}_j, \tag{B.1}$$

where $q_{(n-1),+j} = q_{(n-1)} + 1$. The significance of adding this atom $\hat{\omega}_j$ to the selection was tested by comparing $\mathbf{V}_{(n-1)}^{Noise}$ and $\mathbf{V}_{(n-1),+j}^{Noise}$ with respect to the noise $\mathbf{z}_p$ using the partial correlation based method suggested in Perlbarg et al. [15]. After testing all atoms $\hat{\omega}_j$ ($j = 1, \ldots, N - q_{(n-1)}$) in $\mathbf{V}_{(n-1)}^{Sig}$, the one showing the most significance (i.e. partial correlation higher than a statistically determined threshold) was added to a new selection $\mathbf{V}_{new}^{Noise}$ including $q_{(n-1)} + 1$ atoms. This stepwise regression procedure stopped when no more significant atom was found and $\mathbf{V}_{(n-1)}^{Noise} = \mathbf{V}_{new}^{Noise}$.

Stepwise backward elimination. In (n)-th iteration after the forward selection, we removed each atom $\hat{\omega}_j$ from $\mathbf{V}_{new}^{Noise}$, such as:

$$\mathbf{V}_{new,-j}^{Noise} = \mathbf{V}_{new}^{Noise} \setminus \{\hat{\omega}_j\}. \tag{B.2}$$

To test if each atom in $\mathbf{V}_{new}^{Noise}$ was still significant for describing the noise $\mathbf{z}_p$, we compared $\mathbf{V}_{new}^{Noise}$ and $\mathbf{V}_{new,-j}^{Noise}$ using the partial correlation based method. After testing all atoms $\hat{\omega}_j$ ($j = 1, \ldots, q_{(n-1)} + 1$) in $\mathbf{V}_{new}^{Noise}$, one or more atoms exhibiting a partial correlation lower than a statistically determined threshold were selected. The non-significant atoms were then removed from $\mathbf{V}_{new}^{Noise}$, finally determining $\mathbf{V}_{(n)}^{Noise}$ and $\mathbf{V}_{(n)}^{Sig}$.

The significance of an atom $\hat{\omega}_j$ in a sparse GLM was tested by evaluating the variance explained by the model including this atom (R_{large}^2 for a large model $\mathbf{V}_{large}$) with respect to the variance explained by the model without this atom (R_{small}^2 for a small model $\mathbf{V}_{small}$), as suggested in Perlbarg et al. for CORSICA [15]. When this strategy was applied for our method SPARK, at (n)-th iteration, the large model was $\mathbf{V}_{(n-1),+j}^{Noise}$ and the small model was $\mathbf{V}_{(n-1)}^{Noise}$ in the stepwise forward selection, whereas the large model was $\mathbf{V}_{new}^{Noise}$ and the small model was $\mathbf{V}_{new,-j}^{Noise}$ in the stepwise backward elimination.

In their method, a noise-characteristic time-course signal $\mathbf{z}_p$ is orthogonally projected onto the subspace spanned by the large model $\mathbf{V}_{large}$ and onto the subspace spanned by the small model $\mathbf{V}_{small}$. $\hat{\mathbf{z}}_p$ denotes the orthogonal projection of $\mathbf{z}_p$ onto the subspace spanned by a model $\mathbf{V}$. Then, the following partial correlation for a tested atom $\hat{\omega}_j$ at (n)-th iteration ($R_{(n),j}^2$) was computed as:

$$R_{(n),j}^2 = \frac{R_{large}^2 - R_{small}^2}{1 - R_{large}^2}(N - q - 1), \tag{B.3}$$

where $R_{\mathbf{V}}^2 = var(\hat{\mathbf{z}}_p)/var(\mathbf{z}_p)$ for a model $\mathbf{V}$. Perlbarg et al. suggested that this partial correlation value $R_{(n),j}^2$ quantifies the proportion of the total variance that is explained by the large model but not by the small model, and follows an F-distribution with $(1, N - q - 1)$ degrees of freedom. Therefore, in our method, an atom $\hat{\omega}_j$ estimated using SDL was considered significant in the large model if one can reject the null hypothesis H_0: $R_{(n),j}^2 = 0$ at p-value < 0.001, as the value suggested in CORSICA.

References

[1] Fox M D, Raichle M E. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat Rev Neurosci 2007;8(9):700–11.
[2] Greicius M D, Krasnow B, Reiss A L, Menon V. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proc Natl Acad Sci 2003;100(1):253–8.
[3] Eickhoff S B, Thirion B, Varoquaux G, Bzdok D. Connectivity-based parcellation: critique and implications. Hum Brain Mapp 2015;36(12):4771–92.
[4] Lee K, Tak S, Ye J C. A data-driven sparse GLM for fMRI analysis using sparse dictionary learning with MDL criterion. IEEE Trans Med Imaging 2011;30(5):1076–89.
[5] Lee K, Lina J-M, Gotman J, Grova C. SPARK: sparsity-based analysis of reliable k-hubness and overlapping network structure in brain functional connectivity. NeuroImage 2016;134:434–49.
[6] Sporns O, Honey C J, Kötter R. Identification and classification of hubs in brain networks. PloS one 2007;2(10):e1049.
[7] Bullmore E, Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neurosci 2009;10(3):186–98.
[8] Buckner R L, Sepulcre J, Talukdar T, Krienen F M, Liu H, Hedden T, et al. Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. J. Neurosci. 2009;29(6):1860–73.
[9] Crossley N A, Mechelli A, Scott J, Carletti F, Fox P T, McGuire P, et al. The hubs of the human connectome are generally implicated in the anatomy of brain disorders. Brain 2014;137(8):2382–95.
[10] McKeown M J, Hansen L K, Sejnowsk T J. Independent component analysis of functional MRI: what is signal and what is noise? Curr Opin Neurobiol 2003;13(5):620–9.
[11] Krüger G, Glover G H. Physiological noise in oxygenation-sensitive magnetic resonance imaging. Magn Reson Med 2001;46(4):631–7.
[12] Power J D, Plitt M, Laumann T O, Martin A. Sources and implications of whole-brain fMRI signals in humans. Neuroimage 2017;146:609–25.
[13] Thomas C G, Harshman R A, Menon R S. Noise reduction in BOLD-based fMRI using component analysis. Neuroimage 2002;17(3):1521–37.
[14] Griffanti L, Douaud G, Bijsterbosch J, Evangelisti S, Alfaro-Almagro F, Glasser M F, et al. Hand classification of fMRI ICA noise components. Neuroimage 2017;154:188–205.
[15] Perlbarg V, Bellec P, Anton J-L, Pélégrini-Issac M, Doyon J, Benali H. CORSICA: correction of structured noise in fMRI by automatic identification of ICA components. Magn Reson Imaging 2007;25(1):35–46.
[16] Aharon M, Elad M, Bruckstein A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Signal Process 2006;54(11):4311.
[17] Saito N. Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion. Wavelets geophys 1994;4:299–324.
[18] Bellec P, Carbonell F, Perlbarg V, Lepage C, Lyttelton O, Fonov V, et al. A neuroimaging analysis kit for Matlab and Octave. Proceedings of the 17th International Conference on Functional Mapping of the Human Brain. 2011.
[19] Bellec P, Marrelec G, Benali H. A bootstrap test to investigate changes in brain connectivity for functional MRI. Stat. Sin. 2008;18(4):1253.
[20] Zuo X-N, Kelly C, Adelstein J S, Klein D F, Castellanos F X, Milham M P. Reliable intrinsic connectivity networks: test-retest evaluation using ICA and dual regression approach. Neuroimage 2010;49(3):2163–77.
[21] Ad-Dabbagh Y, Lyttelton O, Muehlboeck J, Lepage C, Einarson D, Mok K, et al. The CIVET image-processing environment: a fully automated comprehensive pipeline for anatomical neuroimaging research. Proceedings of the 12th annual meeting of the organization for human brain mapping. 2006. p. 2266. Florence, Italy.
[22] Power J D, Barnes K A, Snyder A Z, Schlaggar B L, Petersen S E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 2012;59(3):2142–54.
[23] Hahn G J, Shapiro S S. Statistical models in engineering. Tech. rep.. 1967.
[24] Huck J, Wanner Y, Fan A P, Schmidt A-T, Grahl S, Schneider U, et al. High resolution atlasing of the venous brain vasculature from 7T quantitative susceptibility. bioRxiv 2018:444349.
[25] Mallat S, Zhang Z. Matching pursuits with time-frequency dictionaries. IEEE Trans Signal Process 1993;41(12):3397–415.
[26] Pati Y C, Rezaiifar R, Krishnaprasad P S. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers, 1993. IEEE; 1993. p. 40–4.