TWO MODELS FOR FUSION OF MEDICAL IMAGING DATA: COMPARISON AND CONNECTIONS

Yuri Levin-Schwartz[1], *Vince D. Calhoun*[2,3], *and Tülay Adalı*[1]

[1]Department of Computer Science and Electrical Engineering,
University of Maryland, Baltimore County, Baltimore, MD 21250 USA
[2]The Mind Research Network, Albuquerque, NM 87106 USA
[3]Department of Electrical and Computer Engineering,
University of New Mexico, Albuquerque, NM 87131 USA

ABSTRACT

Exploitation of complementary information is the principal reason for collecting data from multiple neurological sensors. Since little is known about the latent processes underlying neural function, it is important to minimize the assumptions placed on the data when performing a joint analysis. This motivates the use of data-driven fusion methods, such as independent vector analysis (IVA), for the analysis of neurological data. For neural datasets, the complementary information exploited by fusion methods may be in the form of similar spatial activation across datasets, the spatial IVA (sIVA) model, or similar subject relations across datasets, the transposed IVA (tIVA) model. Despite the potential power of these two models, no study has investigated how the differences in the modeling assumptions of sIVA and tIVA inform the fusion of real neuro-imaging data. In this paper, we utilize a unique set of multitask functional magnetic resonance imaging data from 271 subjects to directly compare the sIVA and tIVA models and visualize their differences using a novel technique, global difference maps. Through this application, we note important similarities between the results from the two methods that increase our confidence in their overall performance, though differences in modeling assumptions result in certain differences in the decompositions.

Index Terms— FMRI, Data Fusion, Independent Vector Analysis

1. INTRODUCTION

In many fields, the collection of data from multiple sensors has become common, since each sensor is expected to provide a different, yet complementary view of the system under study [5, 13, 14]. Such data can be multiset, $i.e.$, of the same type, such as multiple color channels in a video sequence or financial information from multiple stocks, or multimodal, $i.e.$, of different types, such as functional magnetic resonance imaging (fMRI) data and electroencephalogram (EEG) data collected under similar conditions. However, in both cases, the goal is to maximize the use of available information for the given task through optimal utilization of each dataset. Since, in general, little is known about the relationships between datasets, it is vital to minimize the underlying assumptions placed on the data, letting it "speak for itself." For this reason, data-driven methods, such as independent vector analysis (IVA), a recent and flexible multiset extension of independent component analysis (ICA), have proven useful for the fusion of multiple dataset in many areas, see $e.g.$, [9, 11, 15, 24]. However, the performance of these methods is intimately tied the manner in which the complementary information across the datasets is expressed.

For neural datasets, such as fMRI, the complementary information exploited by IVA is either in the form of similar spatial activation across datasets or similar subject relations across datasets. Exploitation of the correspondences across datasets through similar spatial activation is referred to as the spatial IVA (sIVA) model and is generally used in the analyses of fMRI data across multiple subjects or multiple tasks [16, 21]. On the other hand, if no such spatial relations across the datasets exist, such as for multimodal fusion, the transposed IVA (tIVA) model, which exploits similarities across subject relations, is used [1]. However, most analyses show limited success with the tIVA model, since the number of samples, subjects, is usually limited [1].

In this paper, we make use of a unique set of data that enables direct comparison of the sIVA model with the tIVA model, without the number of subjects severely limiting tIVA. This data, fMRI data drawn from 271 subjects, 121 patients with schizophrenia and 150 healthy controls, during the performance of three tasks, grants us the unprecedented ability to study how the differences in the modeling assumptions of sIVA and tIVA inform the fusion of real neuro-imaging data.

*This work was supported in part by NIH-NIBIB R01 EB 005846 and NSF-CCF 1618551.

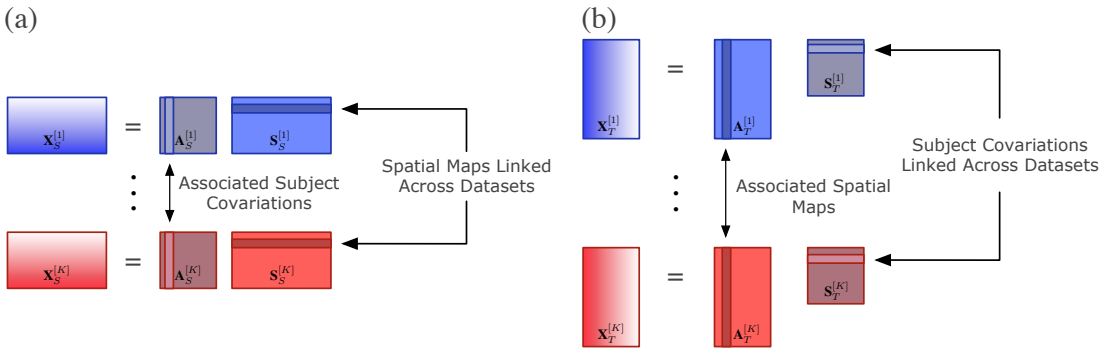

Fig. 1. Generative models for (a) sIVA and (b) tIVA. Note that for sIVA the spatial maps are linked across datasets, while for tIVA the subject covariates that are linked across datasets. Additionally, note that in both cases the shared information across datasets is fully accounted for in the decomposition.

We also propose the use of global difference maps (GDMs) as a simple but effective method to visually compare the results of multiple fusion algorithms and quantify the total discriminative power for each combination of algorithm and dataset. The remainder of the paper is organized as follows. Section 2 describes the mathematical framework for sIVA and tIVA, the fMRI tasks and their extracted features, as well as the construction of the global difference maps (GDM). We present our results in Section 3 and conclude the paper in Section 4.

2. MATERIALS AND METHODS

Due to the differences in timing among each of the tasks, it is difficult to directly fuse multitask fMRI data. Instead, the data from each subject and task is first analyzed through a multiple regression using the statistical parametric mapping toolbox (SPM) [23], where the regressors are created by convolving the hemodynamic response function (HRF) in SPM with the desired predictors. Then, features for each subject and task are generated by applying the appropriate linear contrasts to the regressor estimates. Such a reduction using lower-dimensional, yet still multivariate, representations of the data enables exploration of associations across these feature sets and facilitates the discovery of links across tasks and simplify the identification of biomarkers of disease, see *e.g.*, [6, 22].

2.1. SIVA

Consider the extension of the noiseless ICA model to K datasets, as

$$\mathbf{X}_S^{[k]} = \mathbf{A}_S^{[k]}\mathbf{S}_S^{[k]}, \quad 1 \leq k \leq K, \tag{1}$$

where the kth feature dataset of V voxels from N subjects, $\mathbf{X}_S^{[k]} \in \mathbb{R}^{N\times V}$, is a linear mixture of the N latent sources, $\mathbf{S}_S^{[k]} \in \mathbb{R}^{N\times V}$, through an invertible mixing matrix, $\mathbf{A}_S^{[k]} \in \mathbb{R}^{N\times N}$. Due to the inherent scaling and permutation ambiguities of ICA, running an ICA for each task individually and then aligning the results is both impractical and suboptimal. For this reason and to exploit the similarity of brain regions across the different tasks, one may apply sIVA to the collection of K tasks. SIVA seeks to estimate K demixing matrices, such that the estimated spatial maps are given by $\hat{\mathbf{S}}_S^{[k]} = \mathbf{W}_S^{[k]}\mathbf{X}_S^{[k]}$. The generative model for sIVA is shown in Figure 1a.

Note that the columns of the estimated mixing matrices, $\hat{\mathbf{A}}_S^{[k]}$, provide the loadings of the estimated components across subjects. Thus, the pth column of the kth estimated mixing matrix, $\hat{\mathbf{a}}_{S,p}^{[k]}$, represents the relative weights of the pth source estimate, $\hat{\mathbf{s}}_{S,p}^{[k]}$, for each subject. Since each dataset is reduced to a feature for each subject, it is possible to look for differences in the expression of certain components across two groups, which will yield "biomarkers of disease." Determination of these biomarkers is done through the performance of a two-sample t-test on the subject covariations, where one group is represented by the subject covariations from the patients with schizophrenia and the other by the subject covariations from the healthy controls [6].

2.2. TIVA

If the K datasets share little or no similarity across spatial maps, such as for multimodal fusion [1], then the sIVA model cannot be used. However, there may exist connections across datasets in terms of the expression of the different spatial maps across subjects, $i.e.$, subject covariations, that can be exploited to determine a successful decomposition. This objective, the estimation of maximally similar subject covariations across different datasets, forms the fundamental goal of tIVA and is achieved by first transposing the model in (1), such that

$$\begin{aligned}\mathbf{X}_T^{[k]} = \left(\mathbf{X}_S^{[k]}\right)^{\mathrm{T}} &= \left(\mathbf{S}_S^{[k]}\right)^{\mathrm{T}}\left(\mathbf{A}_S^{[k]}\right)^{\mathrm{T}} \\ &= \mathbf{A}_T^{[k]}\mathbf{S}_T^{[k]}, \quad 1 \leq k \leq K. \end{aligned} \tag{2}$$

Note that through this transposition, the roles of samples and observations are reversed with respect to sIVA. Unlike sIVA, tIVA seeks to estimate K demixing matrices, such that the estimated subject covariations are given by $\hat{\mathbf{S}}_T^{[k]} = \mathbf{W}_T^{[k]}\mathbf{X}_T^{[k]}$. The generative model for tIVA is shown in Figure 1b. Additionally, note that a lack of similarity across spatial maps is not the sole reason to use the tIVA model, since its different modeling assumptions may provide additional advantages over the sIVA model. Finally, note that, since $V \gg N$, K, the computational complexity is approximately $\mathcal{O}(VN^2)$, $\mathcal{O}(KVN^2 + K^2VN)$, and $\mathcal{O}(KN^3 + K^2N^2)$ per iteration for ICA, sIVA, and tIVA, respectively.

2.3. Order Selection

Note that the tall nature of the matrices in Figure 1b, necessitates the performance of dimension reduction prior to IVA, where the order is no larger than N. The selection of an appropriate order for the transposed model, which is vital to the success of tIVA, is an open question and has received considerably less attention than the problem of order selection for the sIVA model, see $e.g.$, [7, 8, 12, 25, 26]. In this paper, we use the order selection method in [17], since it is the only method, to our knowledge, that addresses the issue of order selection for the tIVA model and can be readily applied to the sIVA model. Using this technique, an order of 24 was estimated and used for both tIVA and sIVA.

2.4. Algorithm Selection

As with all data-driven methods, the use of the appropriate algorithm is intimately tied to the success of both tIVA and sIVA. Due to the fact that the subjects specify the samples for tIVA, a significant number of subjects is needed to have sufficient statistical power. This has, thus far, limited the number of algorithms that can be used for tIVA to those that solely exploit second-order statistics (SOS) [1, 3]. However, since we have a relatively large number of subjects in this study, we can employ an IVA algorithm that exploits both SOS as well as higher-order statistics (HOS). To this end, we use IVA with a generalized Gaussian distribution prior (IVA-GGD) [4] as the IVA algorithm in this work. The multivariate generalized Gaussian distribution (MGGD) covers a wide range of unimodal distributions through the value of a shape parameter β, such that the MGGD reduces to a multivariate Gaussian for $\beta = 1$, is super-Gaussian for $\beta < 1$, is a multivariate Laplacian for $\beta = 0.5$, and is sub-Gaussian for $\beta > 1$. The IVA-GGD algorithm requires a user-specified set of shape parameter values and then selects the most appropriate ones from the list for a given problem. The shape parameters that we used were: 0.5, 1, and 5. The reason for selecting this set of parameters is the desire to use the same IVA algorithm for both sIVA and tIVA, the fact that the Laplacian distribution provides a good approximation for the spatial maps [19], and sub-Gaussian distributions are a good approximation of subject covariations [17], $i.e.$, the sources in the tIVA model.

2.5. FMRI Tasks and Extracted Features

The data used in this study is from the Mental Illness and Neuroscience Discovery Clinical Imaging Consortium Collection (available at http://coins.mrn.org/dx) and were obtained from 150 healthy controls and 121 patients with schizophrenia. We next briefly introduce the tasks used in this study as well as the multivariate features extracted from each task.

2.5.1. Auditory Oddball Task (AOD)

This auditory task involved subjects listening to three different types of auditory stimuli: standard (1 kHz tones occurring with probability 0.82), novel (computer generated, complex sounds occurring with probability 0.09), and target (1.2 kHz tones with probability 0.09, to which a right thumb button press was required), in a pseudo-random order [10]. For this task, the regressor was created by modeling the target and standard stimuli as delta functions convolved with the default SPM HRF in addition to their temporal derivatives [20]. Subject averaged contrast images between the target versus the standard tones were used as the feature for this task.

2.5.2. Sternberg Item Recognition Paradigm Task (SIRP)

In this visual task, the subjects had to remember a set of 1, 3, or 5 randomly chosen integers between 0 and 9. The subjects were shown the series of integers and had to indicate, with a button press with the right thumb, whether it was a member of the memorized set or not [10]. For this task, the regressor was created by convolving this probe response block for the three-digit set with the default SPM HRF [20]. This was done for both runs of the probe response and the average map was used as the feature for this task.

2.5.3. Sensory Motor Task (SM)

In this auditory task, the subjects were presented with a sequence of auditory stimuli in an increasing then decreasing step-wise manner. Each tonal change required a button press with the right thumb. For this task, the regressor was created by convolving the whole increase-and-decrease block with the default SPM HRF [20]. For each subject, the average map was used as the feature for this task.

2.6. GDMs

After the performance of a data fusion method on the three fMRI datasets, statistically significant biomarkers are found through a 2-sample t-test run on the subject covariations of each dataset, individually. However, for a given decomposition there may be multiple significant biomarkers, making a

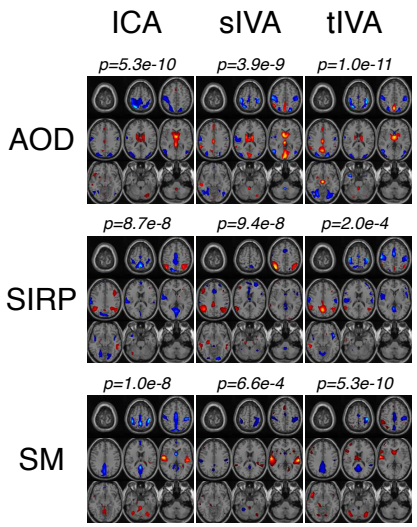

Fig. 2. GDMs for the AOD, SIRP, and SM tasks using the methods ICA, sIVA, and tIVA. The GDMs for the same method are in the columns, while the GDMs for the same dataset across methods are in the rows. These spatial maps correspond to z-maps thresholded at $z = 2.7$, where red and orange represent an increase in activation for controls versus patients and blue represent an increase in activation in patients over controls. Note that the p-value associated with each GDM, which assesses the significance of the decomposition, is shown above the corresponding spatial map.

summarization and comparison of different techniques difficult. For this reason, we propose to summarize the performance of a data fusion method through the use of GDMs, constructed as follows. For the M z-scored, statistically significant, at $p < 0.05$, biomarkers for each dataset, $\hat{\mathbf{s}}_m^{[k]}, 1 \leq m \leq M$, we construct the GDM for that method and dataset, $\hat{\mathbf{s}}_{GDM}^{[k]}$, as

$$\hat{\mathbf{s}}_{\text{GDM}}^{[k]} = \sum_{m=1}^{M} \frac{|T_m|}{\sum_{n=1}^{M} |T_n|} \hat{\mathbf{s}}_m^{[k]}, \tag{3}$$

where T_m is the t-statistic for the mth statistically significant biomarker and $|\cdot|$ is the absolute value operator. Thus, the GDM can be seen as a summary map that captures the whole difference between patients and controls for a given decomposition and dataset within that decomposition. Each significant biomarker is scaled by the value of its corresponding t-statistic, so it is weighed more if the component is better able to differentiate between patients and controls. It is important to note that we can quantify the discriminative power of a GDM, and thus indirectly the decomposition of a fusion method, by regressing the GDMs back onto the original task data, and performing a two-sample t-test on the resulting subject covariations. Finally note that, though not shown here, the GDMs are more significant than the original biomarkers due to the fact that they are composed entirely of regions that are able to differentiate between the two groups.

3. RESULTS AND DISCUSSION

Figure 2 contains the GDMs for both sIVA and tIVA using the IVA-GGD algorithm. In order to explore the additive value of each of these fusion models, we also ran ICA on each feature dataset separately using the entropy bound minimization (EBM) algorithm [18] and computed GDMs from the results. There are many interesting points to make in regards to the GDMs shown in Figure 2 and therefore about the methods themselves when applied to this data. The first point is that the GDMs for the AOD task are the most significant for all three methods. This suggests that the AOD feature dataset is bringing the most discriminative power to the analysis. Additionally, when comparing the GDMs derived from ICA with those from sIVA, our results become less significant, suggesting that there may be some cost to fusing datasets if they are very different. It is important to note that, in general, we observe the opposite trend when comparing the GDMs derived from ICA with those from tIVA, suggesting that tIVA is a better way of deriving statistically significant results [2].

Looking at the GDMs, we note that most of the activated areas are similar for both tIVA and sIVA, increasing our confidence in both sets of results. We also note that the GDMs for sIVA show more parietal and temporal activation in the SIRP and SM datasets, respectively, than tIVA. This result seems reasonable, since greater spatial variability is expected for sIVA than for tIVA. Additionally, due to the fact that the activation is also present in the GDMs of ICA, tIVA may lose some inherent spatial variability of the sources. We also note that there is clear default mode network activation in the GDMs for both sIVA and tIVA in the SIRP dataset, though such activation is not present in the corresponding GDM from ICA. This result is particularly encouraging, since meaningful activation, not present in a single dataset analysis, has been drawn out with both fusion models. This, combined with the increase in sensorimotor activation for the patients over the controls, suggest that the patients found the task harder than the controls did. Finally, we note that tIVA retains more of the sensorimotor regions from the ICA results than sIVA does.

4. CONCLUSIONS

In this paper, we compare two different models for data fusion: the sIVA model and the tIVA model, for the analysis of multitask fMRI data. To facilitate the comparison between these two decompositions, we propose the use of GDMs to summarize the total discriminative power of each dataset within a decomposition. Through this application, we find that the regions in the GDMs are similar across methods, thus increasing our confidence in the overall result. We also find that although sIVA has, in general, higher spatial variability than tIVA, tIVA appears more sensitive to group differences.

5. REFERENCES

[1] Tülay Adalı, Yuri Levin-Schwartz, and Vince D. Calhoun. Multi-modal data fusion using source separation: Application to medical imaging. *Proceedings of the IEEE*, 103(9):1494 – 1506, 2015.

[2] Tülay Adalı, Yuri Levin-Schwartz, and Vince D. Calhoun. Multimodal data fusion using source separation: Two effective models based on ICA and IVA and their properties. *Proceedings of the IEEE*, 103(9):1478 – 1493, 2015.

[3] Matthew Anderson, Tülay Adalı, and Xi-Lin Li. Joint blind source separation with multivariate Gaussian model: Algorithms and performance analysis. *IEEE Transactions on Signal Processing*, 60(4):1672–1683, April 2012.

[4] Matthew Anderson, Geng-Shen Fu, Ronald Phlypo, and Tülay Adalı. Independent vector analysis, the Kotz distribution, and performance bounds. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3243–3247, 2013.

[5] Christoph Arndt and Otmar Loffeld. Information gained by data fusion. In *Lasers, Optics, and Vision for Productivity in Manufacturing I*, pages 32–40. International Society for Optics and Photonics, 1996.

[6] Vince D. Calhoun and Tülay Adalı. Feature-based fusion of medical imaging data. *IEEE Trans. Inf. Technol. Biomed.*, 13(5):711–720, Sep. 2009.

[7] Sharon Chen, Thomas J. Ross, Keh-Shih Chuang, Elliot A. Stein, Yihong Yang, and Wang Zhan. A new approach to estimating the signal dimension of concatenated resting-state functional MRI data sets. *Magnetic Resonance Imaging*, 28(9):1344–1352, 2010.

[8] Dietmar Cordes and Rajesh R. Nandy. Estimation of the intrinsic dimensionality of fMRI data. *NeuroImage*, 29(1):145–154, 2006.

[9] Arman Eshaghi, Sadjad Riyahi-Alam, Roghayyeh Saeedi, Tina Roostaei, Arash Nazeri, Aida Aghsaei, Rozita Doosti, Habib Ganjgahi, Benedetta Bodini, Ali Shakourirad, M. Pakravan, H. Ghana'ati, K. Firouznia, M. Zarei, A. R. Azimi, and M. A. Sahraian. Classification algorithms with multi-modal data fusion could accurately distinguish neuromyelitis optica from multiple sclerosis. *NeuroImage: Clinical*, 7:306–314, 2015.

[10] Randy L. Gollub, Jody M. Shoemaker, Margaret D. King, Tonya White, Stefan Ehrlich, Scott R. Sponheim, Vincent P. Clark, Jessica A. Turner, Bryon A. Mueller, Vince Magnotta, Daniel O'Leary, Beng C. Ho, Stefan Brauns, Dara S. Manoach, Larry Seidman, Juan R. Bustillo, John Lauriello, Jeremy Bockholt, Kelvin O. Lim, Bruce R. Rosen, S. Charles Schulz, Vince D. Calhoun, and Nancy C. Andreasen. The MCIC collection: a shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia. *Neuroinformatics*, 11(3):367–388, 2013.

[11] John Carl Hoefs. Serum protein concentration and portal pressure determine the ascitic fluid protein concentration in patients with chronic liver disease. *The Journal of Laboratory and Clinical Medicine*, 102(2):260–273, 1983.

[12] Mingqi Hui, Rui Li, Kewei Chen, Zhen Jin, Li Yao, and Zhiying Long. Improved estimation of the number of independent components for functional magnetic resonance data by a whitening filter. *IEEE Journal of Biomedical and Health Informatics*, 17(3):629–641, May 2013.

[13] Alex P. James and Belur V. Dasarathy. Medical image fusion: A survey of the state of the art. *Information Fusion*, 19:4–19, 2014.

[14] Dana Lahat, Tülay Adalı, and Christian Jutten. Multimodal data fusion: An overview of methods, challenges, and prospects. *Proceedings of the IEEE*, 103(9):1449–1477, Sep. 2015.

[15] Jong-Hwan Lee, Te-Won Lee, Ferenc A. Jolesz, and Seung-Schik Yoo. Independent vector analysis (IVA) for group fMRI processing of subcortical area. *International Journal of Imaging Systems and Technology*, 18(1):29–41, 2008.

[16] Jong-Hwan Lee, Te-Won Lee, Ferenc A. Jolesz, and Seung-Schik Yoo. Independent vector analysis (IVA): Multivariate approach for fMRI group study. *NeuroImage*, 40(1):86–109, 2008.

[17] Yuri Levin-Schwartz, Yang Song, Peter J. Schreier, Vince D. Calhoun, and Tülay Adalı. Sample-poor estimation of order and common signal subspace with application to fusion of medical imaging data. *NeuroImage*, 134:486 – 493, 2016.

[18] Xi-Lin Li and Tülay Adalı. A novel entropy estimator and its application to ICA. In *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6, Sep. 2009.

[19] Martin J. Mckeown, Scott Makeig, Greg G. Brown, Tzyy-Ping Jung, Sandra S. Kindermann, Anthony J. Bell, and Terrence J. Sejnowski. Analysis of fMRI Data by Blind Separation Into Independent Spatial Components. *Human Brain Mapping*, 6:160–188, 1998.

[20] Andrew M. Michael, Stefi A. Baum, Jill F. Fries, Beng-Choon Ho, Ronald K. Pierson, Nancy C. Andreasen, and Vince D. Calhoun. A method to fuse fMRI tasks through spatial correlations: Applied to schizophrenia. *Human Brain Mapping*, 30(8):2512–2529, 2009.

[21] Mahdi Ramezani, Kris Marble, H. Trang, Ingrid S. Johnsrude, and Purang Abolmaesumi. Joint sparse representation of brain activity patterns in multi-task fMRI data. *IEEE Transactions on Medical Imaging*, 34(1):2–12, 2015.

[22] Mahdi Ramezani, Abtin Rasoulian, Tom Hollenstein, Kate Harkness, Ingrid Johnsrude, and Purang Abolmaesumi. Joint source based analysis of multiple brain structures in studying major depressive disorder. *Proc. SPIE*, 9034:1–6, 2014.

[23] SPM5. Statistical Parametric Mapping. http://www.fil.ion.ucl.ac.uk/spm/software/spm5, 2011.

[24] M. Statheropoulos, N. Vassiliadis, and A. Pappa. Principal component and canonical correlation analysis for examining air pollution and meteorological data. *Atmospheric Environment*, 32(6):1087–1095, 1998.

[25] Xiaoping Xie, Zhitong Cao, Xuchu Weng, and Dan Jin. Estimating intrinsic dimensionality of fMRI dataset incorporating an AR(1) noise model with cubic spline interpolation. *Neurocomputing*, 72(4):1042–1055, 2009.

[26] Grigori Yourganov, Xu Chen, Ana S. Lukic, Cheryl L. Grady, Steven L. Small, Miles N. Wernick, and Stephen C. Strother. Dimensionality estimation for optimal detection of functional networks in BOLD fMRI data. *NeuroImage*, 56(2):531–543, 2011.