Spatial Patterns for Discriminative Estimation (SP♠DE)

A. Llera[1,2,5], R. Chauvin[1,2], P. Mulders[1,2,3], J. Naaijen[1,2], M. Mennes[1,2], C. F. Beckmann[1,2,4]

1- Donders Institute for Brain, Cognition and Behavior, Centre for Cognitive Neuroimaging, 6500 GL, Nijmegen, The Netherlands.

2- Department of Cognitive Neuroscience, Radboud University Medical Centre, 6525 EN, Nijmegen, The Netherlands.

3- Department of Psychiatry, Radboud University Medical Centre, 6500 HB, Nijmegen, The Netherlands.

4- Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), University of Oxford, Oxford, OX3 9DU, U.K.

5- Karakter, Child and Adolescent Psychiatry, University Center, 6525 GC Nijmegen, The Netherlands.

Abstract

Functional connectivity between brain regions is modulated by cognitive states or experimental conditions. Such connectivity variations can be appreciated in covariance or correlation matrices obtained from fMRI data. A multivariate methodology that can capture fMRI connectivity maps in light of different experimental conditions would be of primary importance to learn about the specific roles of the different brain areas involved in the observed connectivity variations. Here we adapt, extend, and evaluate a supervised dimensionality reduction model to fMRI data that provides a full multivariate characterization by achieving optimal discriminative linear spatial filters in terms of variance and provides interpretable spatial maps directly reflecting the brain areas involved in the accounted connectivity changes[1,2]. We demonstrate the strength of such approach for fMRI data by using data from the Human connectome project to show that the model provides close to perfect discrimination between different fMRI tasks at low dimensionality. Additionally, the straightforward interpretability of the model is demonstrated by the obtained linear filters relating to anatomical areas well known to be involved on each considered fMRI task, and by testing the robustness of both, discriminatory and spatial reproducibility, with respect to the number of subjects and of the fMRI time-points considered. We additionally show that such an approach can provide a complementary/alternative view to traditional task fMRI analyses by looking at changes in the covariance structure as a substitute to changes in the mean signal as the general linear model (GLM) analyses. We use 2-back and 0-back working memory task fMRI data from the Human connectome project to show that, at the group level, SPADE discovers connectivity changes in memory related areas during the 2-back periods, while attention network areas are involved during the 0-back task periods. We conclude that SPADE is a robust tool to investigate brain connectivity alterations across induced cognitive changes and has the potential to be used in pathological or pharmacological cohort studies. A publicly available toolbox is provided to facilitate the end use and further development of this methodology able to extract Spatial Patterns for Discriminative Estimation (SP♠DE).

Introduction

Functional magnetic resonance imaging (fMRI) is the primary tool used to investigate how the brain (re)acts under different mental states and the biological underpinnings of brain disorders. Most commonly, these state or group-differences are investigated using generalized linear models (GLM) [3] to detect brain regions that differentiate between experimental conditions. The GLM is for example the standard tool used to analyze task fMRI data incorporating a temporal model for the induced brain activity [4] that allows obtaining a supervised voxel-wise characterization of changes in the blood-oxygen-level dependent (BOLD) response [3]. Yet, while powerful and straightforward to interpret, the GLM has two crucial disadvantages. Foremost, while fMRI data is multivariate in nature, the GLM is applied as a mass univariate approach (i.e., independently to each voxel in the brain) and does not take into account interactions between different regional effects. Second, it ignores the variance in the data by looking at single mean BOLD changes across conditions. To overcome these issues, one can use multivariate models that allow studying the second order interactions between different brain areas [5,6]. Although these models overcome some of the issues with univariate modelling, they are not specifically tailored for state/group characterization. Further inclusion into models for state/group identification, i.e.

regression or classification, can provide strong discrimination but complicates the interpretation of the results. Consequently, to improve interpretability of results obtained from analyzing multivariate fMRI data while optimizing discrimination between two conditions or groups of fMRI time-series, it would be beneficial to use a multivariate approach to extract brain patterns that maximize discriminative properties. To that end we consider a model to extract Spatial Patterns for Discriminative Estimation (SP♠DE). SPADE is a covariance based discriminative approach that provides optimal linear discriminative filters in terms of variance[1]. Although the basics of the approach are commonly used in EEG-based brain-computer interfaces [2,7], its application for fMRI data is challenging since it requires a large number of observations to compute well-posed covariance matrices. Nevertheless, we introduced preliminary results in [8] and further work has been done since then by applying it to cohorts of neurodevelopmental disorders [9–12]. Recently we have also considered it to study the effect of stress induction [13]. Although recent advances in data acquisition and large-scale data collection allow collecting a large number of observations in relatively short time periods, this is not yet the standard, especially in clinical acquisitions. To partially overcome this issue we expand the original methodology formulation [1] by using a regularized version of the covariance matrices[14] and by embedding the process into a novel permutation testing statistical approach for model order selection; note here that, according to [15], the model order selection step is of primary importance for the proper interpretation of the spatial maps associated with the learned spatial filters.

Here, we make use of high-quality fMRI data from the Human Connectome Project (HCP) [16,17] to uncover and evaluate the potential power of the method for discriminative estimation on fMRI data. We consider resting state, motor, and working memory task fMRI data from 1063 subjects and demonstrate that we are able to learn low-dimensional representations of the data that provide close to perfect discrimination between the different mental states. The availability of these simple and well-known tasks allows to validate that the learned filters are robust and specific to brain areas well-known to be involved during their respective paradigms [15,18], validating the spatial interpretation of the obtained results on fMRI data. Further, the highly temporally sampled fMRI data of the HCP allows also considering the classification and spatial associations' robustness as a function of the number of subjects and time points included. Herewith, we can identify the scenarios in which the current approach is useful to address general questions using fMRI.

Methods

Dataset & preprocessing

In this work we use resting state and task fMRI data from the Human Connectome Project (HCP) [16,17]. For each subject, we consider the motor task (MT), the working memory task (WM), and the first of the two available resting state sessions (RS). We use the HCP1200 release which includes data from 1200 subjects and only consider data from subjects for whom all three data modalities (RS, MT and WM) were available resulting in a total of 1063 subjects. In all cases we used the fully preprocessed data in MNI152 space as delivered in the HCP1200 release. For the full preprocessing details we refer the reader to [19].

Spatial Patterns for Discriminative Estimation (SP♠DE)

SPADE indexes connectivity changes from two sets of fMRI data using the simultaneous diagonalization of two covariance matrices (Fukunaga, 1990); full analytical details are provided in the supplementary material. Although ideally one would like to compute full brain spatial covariance matrices, this is not computationally achievable yet due to memory constraints, and a spatial dimensionality reduction must be performed in order to apply the algorithm to fMRI data. For each subject and for each of the three considered fMRI data modalities we performed a spatial dimensionality reduction into 165 regions of interest (ROI) from a functional parcellation [20] by extracting the mean time-series across the voxels in each ROI. Each ROI time-series was then independently demeaned and divided by its standard deviation before further processing.

Then, considering such ROI time-series gathered under two different fMRI modalities (e.g., resting state versus working memory task fMRI), we compute a regularized covariance matrix[14] per subject and condition and average then across subjects to obtain a unique covariance matrix per modality; these two average covariance matrices are then used to estimate their simultaneous diagonalization[1]. This results in a set of discriminative

linear spatial filters that optimally separate the two initial modalities in terms of variance. We address the model order selection, i.e. the number of spatial filters selected to optimize discrimination, using permutation testing. Briefly, we select the filters that explain significantly more variance for a given condition than the variance obtained when randomizing the data across conditions. Full details are provided in the supplementary material. An important feature of the SPADE model is that connectivity changes can be summarized for each basis vector as a unique spatial map that provides a spatial weight for each ROI. This results in a straightforward interpretation of the relevant changes in connectivity. Since the simultaneous diagonalization assumes no implicit noise model, the estimated basis vectors cannot be directly interpreted back to the brain, and the interpretable associated spatial maps are obtained using structural coefficients [15,18]. Brief details are provided in the supplementary material and for a more extensive explanation we refer the reader to [15]. Note that since for a proper spatial interpretation the noise model is estimated from the selected relevant filters[15], the model order selection we introduce here is of most importance to obtain a proper interpretation of the resulting spatial filters.

For the remainder of the paper we will denote the full process involving spatial dimensionality reduction of two fMRI data modalities, spatial covariance regularization, simultaneous diagonalization, model order estimation by permutation testing, and estimation of the associated interpretable spatial maps [15] as SPADE. To facilitate the end use application of these tools to the community, a toolbox providing full automatized estimation will be made publicly available through Git-Hub upon publication.

SPADE evaluations

We use the resting state (RS), motor task (MT) and working memory task (WM) fMRI data from the HCP sample and apply the introduced SPADE methodology to find brain connectivity differences between each pair of fMRI modalities independently: RS vs MT, RS vs WM and MT vs WM. To quantify the quality of the learned filters we perform a discriminative analysis to distinguish between fMRI modalities using a ten-fold cross validation approach for each of the three scenarios (RS vs MT, RS vs WM and MT vs WM) independently. At each fold, SPADE discriminative filters are learned from the covariance matrices of the two selected modalities using data from 90% of the subjects and the number of filters is selected using permutation testing. Then we compute the projection of these subjects' data into the newly learned basis, compute the logarithmic variance of the resulting time-series as features, and use them to train a Linear Discriminant Analysis (LDA) classifier [21] to distinguish between the two included modalities (e.g. RS vs MT). The filters are then used to project the remaining unseen data (~10% subjects) and extract the log-variance of these projections as features to test the classifier quality. Each classifier is evaluated using the accuracy which we define as the percentage of correctly classified samples. For comparison we also evaluate the results obtained using a linear support vector machine (SVM) classifier in the full covariance space, i.e. feature space of dimension (165*166)/2= 13695. We visually evaluate the spatial extent of the SPADE linear filters using structural coefficients for visualization of the associated spatial maps [15,18]. Finally, we consider the effect of a reduced number of subjects and temporal observations in the performance of the model. For each considered number of subjects (N) and time-points (T), we performed 100 different realizations of the SPADE analyses, where at each realization a subset of N subjects was selected randomly and the temporal down-sampling was performed by selecting T equally temporally spaced samples from the full original fMRI time-series i.e. simulating higher TR acquisitions.

In addition, we perform a secondary evaluation where we apply the SPADE model to a scenario previously studied using a GLM analysis. Using the WM task data from the HCP sample, we compared the 0-back memory task periods to the 2-back memory task periods at the group level using the SPADE approach. As with the previous evaluations, in this case we also test the discrimination power using 10-fold cross-validation and report the spatial maps associated to the top four spatial maps for comparison to the results presented in [22].

Results

To illustrate the discriminative performance of the SPADE methodology, Figure 1 presents two-dimensional representations of fMRI data obtained using the SPADE model. For each comparison between two fMRI modalities, we plotted one random fold from the ten-fold-cross-validation used to evaluate discrimination and divide the data into training and testing datasets. Note that by construction, the first and the last SPADE filters

maximize variance for one class while minimizing it for the other one. Consequently, the discriminative power of the algorithm usually comes from the combination of pairs of filters at both extremes of the eigen-spectrum. In the first row we can appreciate, for each of the three comparisons, a very clear two-dimensional separation between the training features obtained by combining the two extreme SPADE filters. The second row clearly shows the generalization ability of the SPADE model on the test-set data. We observed that in all cases the classification using uniquely one pair of filters (two-dimensional feature space) is very high, between 98.6% and 99.76% respectively; exact statistics are provided in the second column of Table 1.

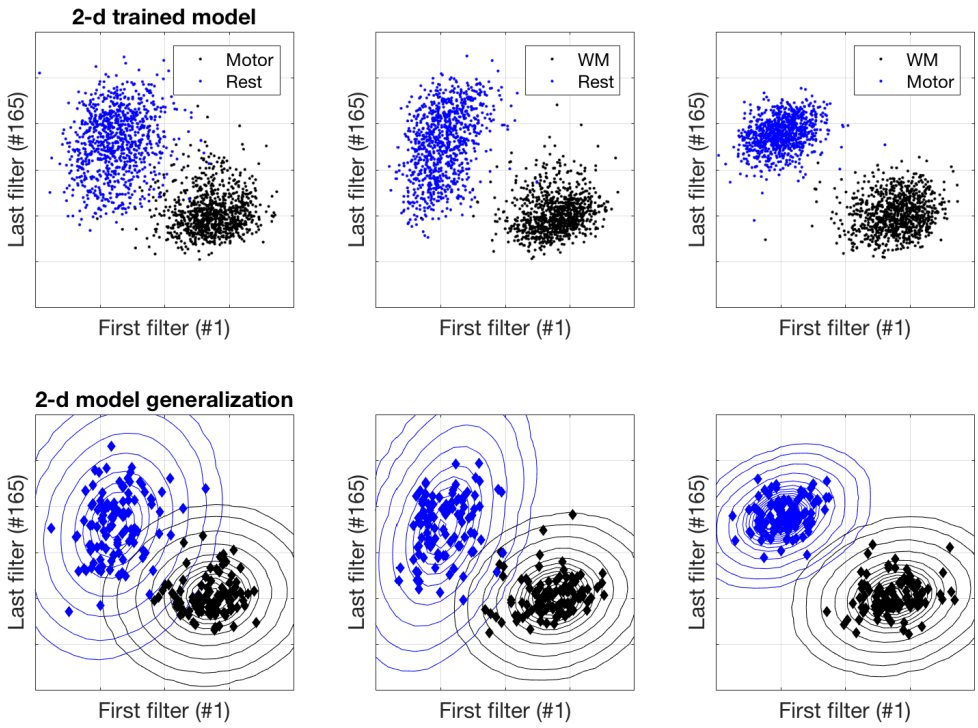


Figure 1: Two-dimensional fMRI data representations obtained using SPADE. The columns represent three different comparisons, MT vs RS, WM vs RS and WM vs MT task fMRI respectively. In the first row we present the logarithmic variance of the training fMRI data projected onto the first and last SPADE filters learned to discriminate each pair of fMRI modalities as indicated in the figure legends. Dots represent subjects and blue and black color encode the two different fMRI tasks considered on each scenario. In all three cases we can see that the training set is correctly separated in two visually distinguishable sets. In the second row we represent the learned two-dimensional training feature distributions presented in the first row as two Gaussian densities, and visualize the logarithmic variance of the testing fMRI data projections into the new basis as color coded diamonds. Appreciate the generalization ability of the SPADE model that provides an optimal two-dimensional classification space.

However, to fully evaluate the discriminative performance of the SPADE model we applied model order selection via permutation testing and selected filters at significance Bonferroni corrected level $p<0.05/165 = 3.3\times10^{-4}$ (see previous section and supplementary material for complete details). Although in practice the dimensionality was computed on each fold independently, for illustration we provide the exact set of filters selected when using the full sample at each comparison in Table 1 column 'SPADE selected filters'. On each of the three comparisons the classification performance obtained using SPADE at the dimensionality selected using the permutation testing strategy is presented in Table 1 column 'SPADE accuracy'. For comparison we also provide the classification results obtained using a SVM in the full covariance space in the last column of Table 1. Although the mean improvement with respect to the two-dimensional scenario or the SVM is relatively small, the improvement it is clearly notable in the reduced standard deviation of the results across folds. These results highlight that the top filters are the most relevant for discrimination but using the model order selection generalizes better across folds. Further, the comparison with the SVM provides remarkable results, especially in light of the reduced dimension in which SPADE classification is performed and the linear nature of the SPADE filters that allow for a straightforward spatial interpretation.

Comparison	SPADE-2D accuracy	SPADE Selected filters	SPADE accuracy	SVM accuracy
RT vs MT	98.64 (0.95) %	1-6 and 155-165	99.3 (2*10^-3) %	98.9 (0.37) %
RT vs WM	99.34 (0.33) %	1-2 and 152-165	99.77 (6*10^-5) %	99.8 (0.40) %
WM vs MT	99.76 (0.46) %	1-4 and 153-165	99.63 (2*10^-4) %	99.58 (0.35) %
mean	99.24 (0.58) %		99.56 (1.3*10^-3) %	99.42 (0.37) %

Table 1: Summary results of the SPADE analyses. The second row presents results when comparing resting state (RT) with motor task (MT) fMRI, the third row shows RT with working memory (WM) task fMRI, and the forth row compares WM and MT fMRI. The fifth row shows the mean results across the three comparisons. The second column (SPADE 2-D) presents the classification results when discriminating each pair of fMRI modalities using a SPADE two-dimensional projection of the data. The column 'SPADE selected filters' shows the eigen-spectrum location of the filters selected by the model order selection. The column 'SPADE accuracy' shows the classification results obtained using the model order selection and the column 'SVM accuracy' shows the results obtained using a SVM in full covariance space (i.e. feature space of dimension (165*166)/2= 13695).

To evaluate the spatial distributions of the filters associated to the discriminative results reported, Figure 2 summarizes the spatial maps obtained in all three comparisons; the full set of maps will be uploaded as supplementary material. Since each fold provides slightly different discriminative filters, for visual validation we performed SPADE analyses including all subjects in the learning phase. The obtained spatial filters were then transformed into interpretable spatial maps [15,18].

When looking at differences between resting state and motor task [23], in Figure 2 A) we observe that areas synchronizing to maximize variance for the motor task (left side, filters 1 and 2) involve bilateral supplementary motor cortex, caudate and operculum in the first filter, while primary motor cortex and contralateral cerebellar cortex involvement is present in the second filter. Areas identified by these filters are well known to be related to motor preparation, execution and control [24,25]. On the other side of the spectrum (right side, filters 164 and 165) we find the filters maximizing variance for resting state data while minimizing it for motor task data. The last filter (#165) involves sensory motor cortex, bilateral dorso-lateral PFC (attention network), cerebellum and several temporal areas including the temporal gyrus and temporal pole. On the other side, the penultimate filter (#164) resembles the default mode network (DMN), anterior cingulate cortex and thalamus.

When looking at the differences between resting state and working memory task, in Figure 2 B) we observe that the areas synchronizing to maximize variance for the working memory task in the first filter involve areas well known to working memory performance [26,27] as visual and precuneus area, similar to the dorsal pathway, and temporal-occipital areas, as present in the ventral pathway [28,29]. Filter number two involves areas similar to number one but includes bilateral hippocampus and thalamus and more temporal areas. On the other side of the eigen-spectrum, we find that the last filter (#165) involves the anterior cingulate cortex, cerebellum, frontal pole, thalamus, middle temporal gyrus and temporal pole, frontal opercula. Further, the penultimate filter (#164) involves again the DMN and the thalamus. We observe that the filters relating to increased variance during rest (Figure 2A and 2B) are in both cases informing about a stronger involvement of default mode network (DMN) related areas during the resting state and are extremely similar in both independent analyses (Figure 2 A and B right sides), showing the consistency of the SPADE algorithm.

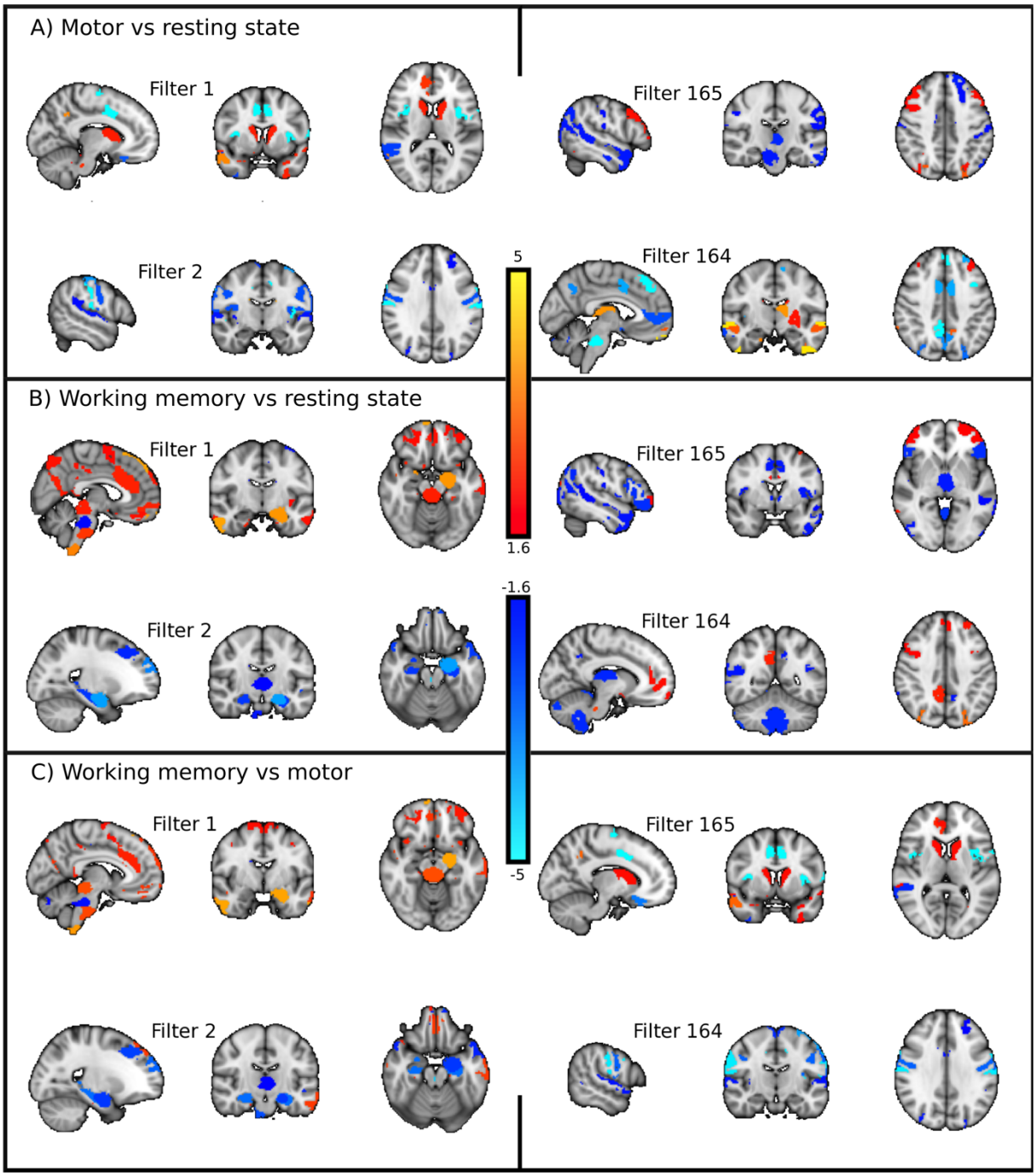


Figure 2: A) Resting state vs. motor task. B) Resting state vs. working memory task. C) Working memory vs. motor task fMRI. For each of the three scenarios we present a summary of the spatial maps associated with the discriminative spatial filters. On the left side we present filters 1 and 2, which maximize variance for A) motor, B) memory and C) memory. On the right side we present filters 165 and 164, which maximize variance for A) resting state, B) resting state and C) motor task. For visualization the value of the maps is converted to pseudo-z-statistics and thresholded at Z=1.6.

When looking at the differences between working memory and motor task (Figure 2 C), we observe that the spatial maps obtained are also very closely related to the ones that differentiated the tasks from the resting state; in particular, filters maximizing for WM (Figure 2 C, filters 1 and 2) resemble the ones maximizing variance for WM with respect to resting state (Figure 2 B filters 1 and 2); and filters maximizing variance for MT (Figure 2 C, filters 164 and 165) exhibit a similar spatial distribution to the ones maximizing variance for MT with respect to resting state (Figure 2 A filters 1 and 2). These results show that the filters extracted for each

modality involve areas well known to be required for the particular task performed and are highly specific since they can be obtained independently of the other covariance structure considered.

For the three considered comparisons, Figure 3 presents the mean and standard deviation of the classification performance results (top and bottom rows respectively) obtained over 100 random subjects subsets selections as a function of the number of subjects and of fMRI time points included (see section Methods for details). Clearly more subjects and longer scanning sessions provide stronger discriminability as shown by a higher accuracy and a lower standard deviation in these numerical experiments. However, we appreciate that in all cases above 100 subjects and a reasonable amount of time points (200 as a reference) are enough to reach high accuracy and a low standard deviation. Using 50 subjects still provides high accuracy in most cases but at a higher standard deviation cost. Dropping the number of subjects to 20 clearly decreases the model performance even when considering a high temporal sampling.

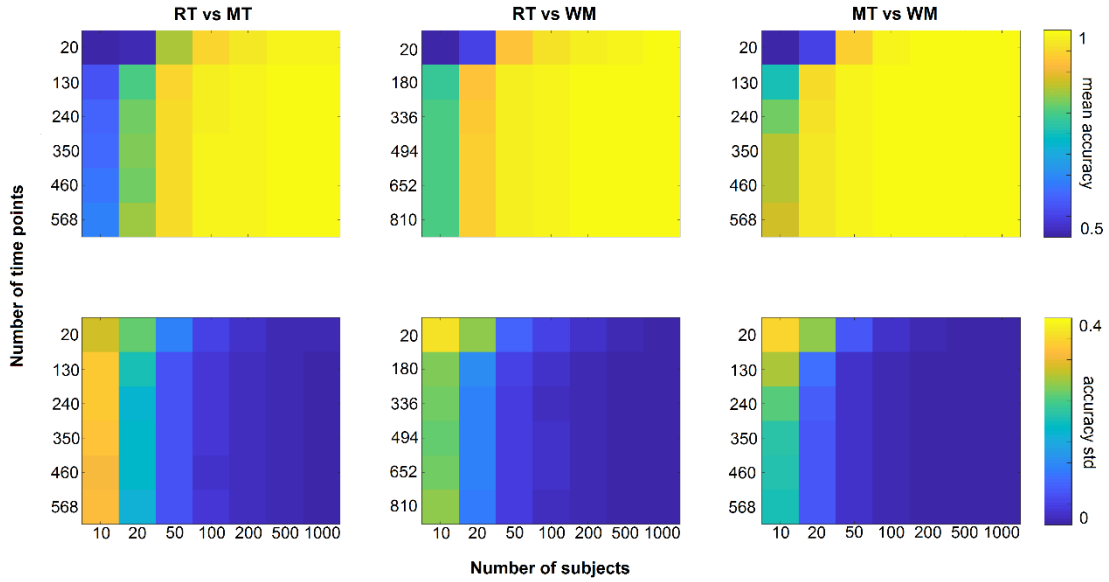


Figure 3: Top row presents the mean classification performance for the three comparisons indicated on each figure title, each as a function of the number of fMRI time samples (y-axis) and the number of subjects (x-axis). The bottom row shows the standard deviation around the classification accuracy presented in the top row.

The numerical results presented in figure 3 are interesting when the goal of a study is to discriminate between two groups. However, our main aim is to motivate SPADE as a tool that allows learning about the hidden brain processes that drive changes on brain functional connectivity as a function of cognitive changes. In figure 4 we consider the reproducibility of the spatial maps obtained through SPADE as a function of time points and number of subjects, for the MT vs WM case. The results for the other two scenarios, RT vs Wm and MT vs WM, are provided as supplementary material, subsection 'spatial reproducibility'. Each subfigure shows the mean correlation matrix between the significant spatial maps obtained from SPADE in the full sample (N=1063) and the spatial maps obtained at each of the 100 SPADE analyses on each subset of subjects, for the number of subjects indicated on each subfigure title (N) and the number of time-points indicated in the y-axis (T). The standard deviation across each entry of these matrices is shown in figure 5. The size of these matrices equals the number of significant selected filters indicated in Table 1. At the sub-sampled realizations, the analyses were performed at the same dimensionality than the full analyses, and the filters are not reordered to match the original order, showing the robustness in the order of their recovery at all the sub-samples. Consequently, the top left corner and bottom right corner of each subfigure relates to the spatial reproducibility of the two most discriminative filters, the associated to the first and the last filters learned through SPADE. Although these two filters can be expected to be reproduced considering just 50 subjects, figure 4 shows that increasing the number of subjects really boosts spatial reproducibility. Note for example that considering N=50 and T=240 we are able to recover the last filter and several top filters, and when considering approximately double number of time-points, N=460, or even more, N=568, we can slightly increase reproducibility; such reproducibility improvement is less than the obtained by doubling the number of subjects, i.e. N=100, T=240. A similar effect

can be found globally and it becomes clearer when considering the standard deviation around these results as presented in Figure 5. Increasing number of subjects provides stronger signal to noise ratio than increasing number of time-points. This simply means that the group effect will be easier found with more subjects, independently of the better single subject connectivity estimation through more fMRI samples. Although such results might seem a truism, it is a desirable characteristic of any robust method. Further, figures 4 and 5 show that although the mean spatial reproducibility for the top and bottom filters is very high when considering 100 subjects, we need to get to 200 subjects to strongly reduce the standard deviation. They also suggest that increasing the number of time-points above 350 does not provide clearly improved mean correlation reproducibility. Although the standard deviation around these mean values also decreases with higher T, such a decrease in standard deviation is negligible for the top and bottom filters.

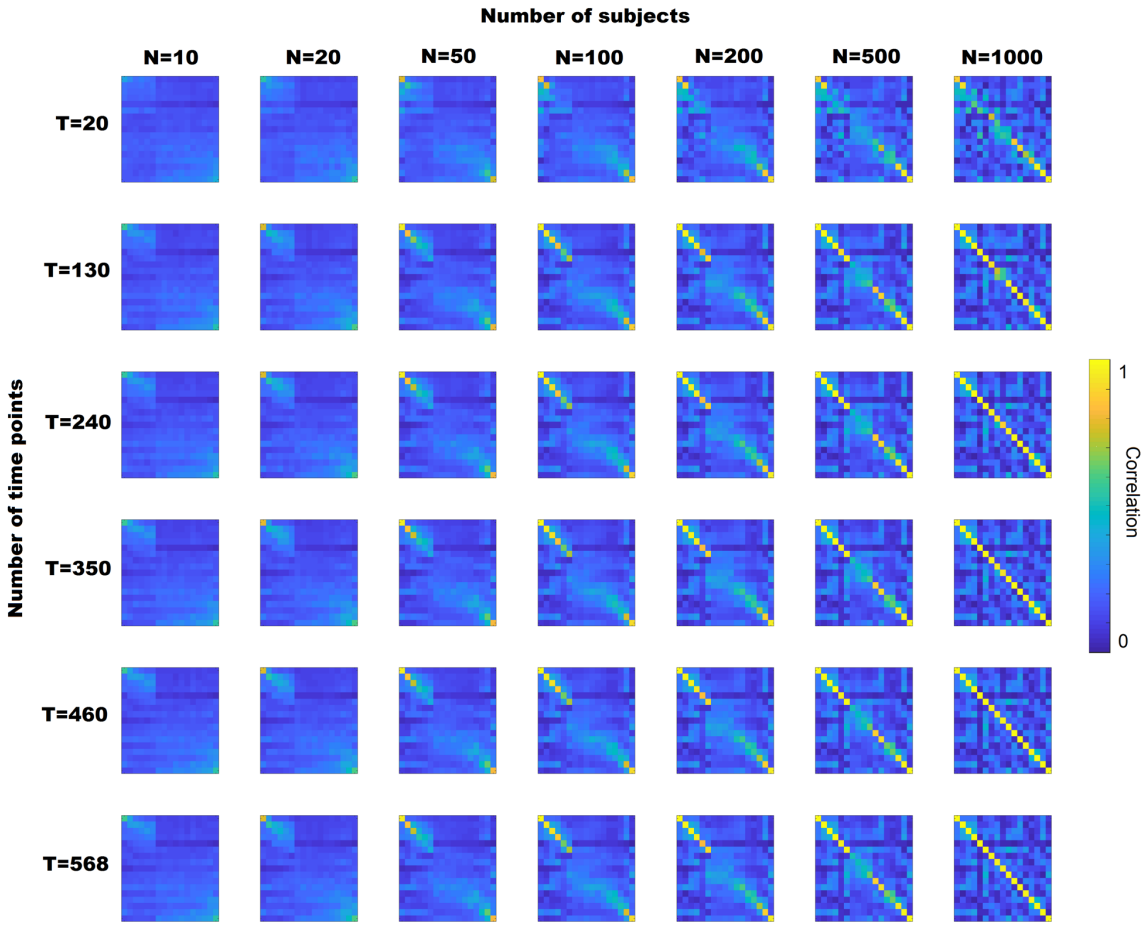


Figure 4: SPADE spatial maps reproducibility for RT vs MT. Each subfigure shows the mean correlation matrix between the significant spatial maps using the full sample (N=1063) and the ones obtained for N and T as indicated on each row and column. The mean is computed over 100 random selections of subjects (see Methods section for details). The size of each correlation matrix equals the number of significant selected filters as indicated in Table 1 and are accordingly ordered, with the top left corner and bottom right corner relating to the spatial reproducibility of the most important filters, the first (#1) and the last (#165) ones.

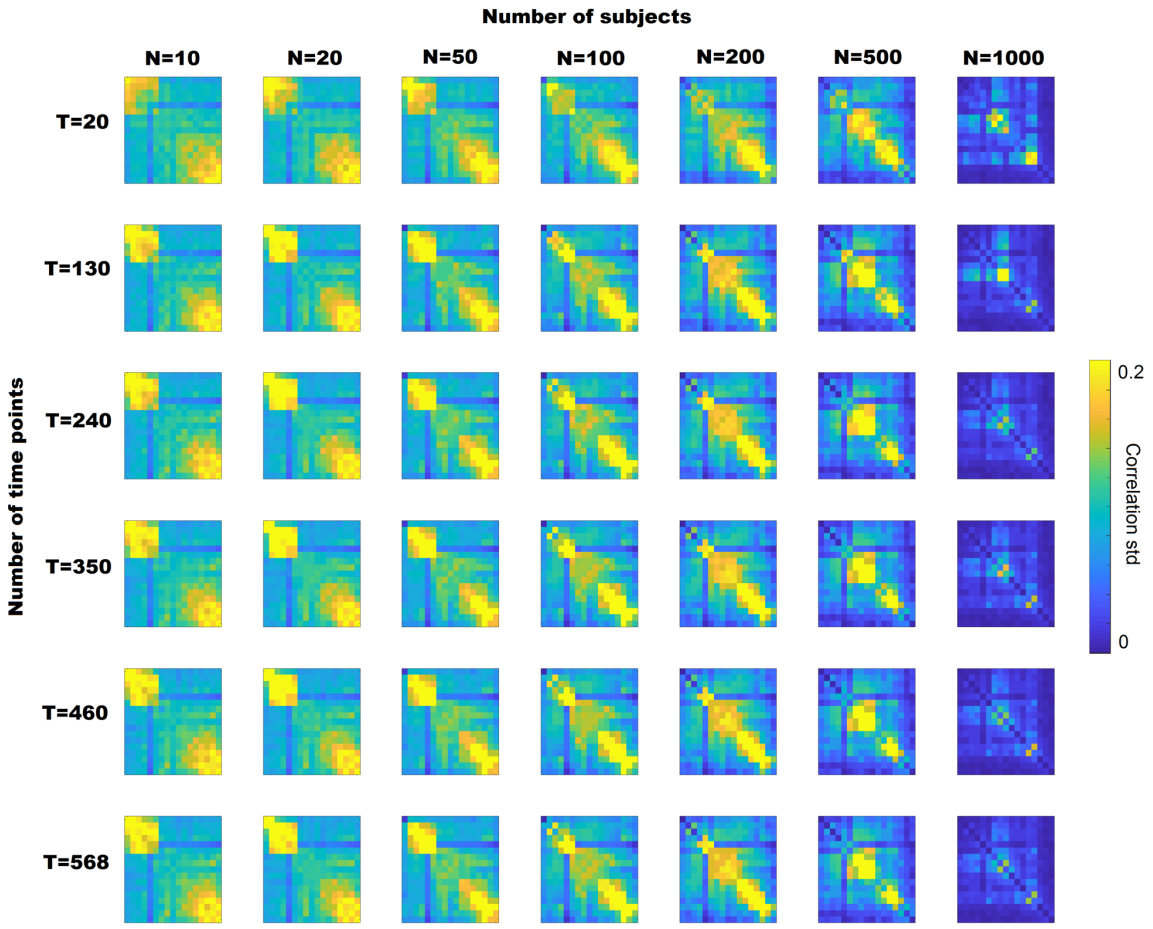


Figure 5: Standard deviation of SPADE spatial maps reproducibility for RT vs MT. Each subfigure shows the element wise standard deviation around the correspondent mean correlation value from the matrices shown in figure 4. The standard deviation is computed over 100 random selections of subjects (see Methods section for details).

In addition to the previously described comparisons, we performed another validation of the model by considering an analysis where we use the WM task fMRI data and compare the 2-back working memory task to the 0-back working memory task. A common approach to this type of data is a traditional example of task fMRI data analysis where conditions are compared by means of a GLM analysis, and the results of such analyses in this same dataset have been previously reported in [22]. In this study, increased activation in the lateral-prefrontal and dorsal parietal cortex was associated with 2-back vs. 0-back working memory task differences. Table 2 shows the classification of the 10-fold cross validation results for discrimination of the 0-back and the 2-back working memory tasks. We appreciate that a two-dimensional SPADE projection already provides a very high accuracy to identify which level of memory loading the task required (2nd column). Such low dimensional representation already provides higher discrimination than the obtained using a SVM in full covariance space (5th column). Further, in this scenario the two main filters provide a discrimination accuracy very close to the one obtained using all selected discriminant filters (3rd and 4th columns).

Comparison	SPADE-2D accuracy	SPADE Selected filters	SPADE accuracy	SVM accuracy
0 vs 2-back	93.9 (0.016) %	1-10 and 163-165	94.4 (0.015) %	91.3(0.018)%

Table 2: Summary results of the SPADE analyses in 2-back vs. 0-back working memory. The second column (SPADE 2-D) presents the classification results when discriminating each pair of fMRI modalities using a SPADE two-dimensional projection of the data. The column 'SPADE selected filters' shows the eigen-spectrum location of the filters selected by the model order selection. The column 'SPADE accuracy' shows the classification results obtained using the model order selection and the column 'SVM accuracy' shows the results obtained using an SVM in full covariance space (i.e. feature space of dimension (165*166)/2= 13695).

In Figure 6 we present a summary of the spatial maps learned using the SPADE model. The filter providing the most variance for the 2-back task, i.e. filter 1, involves left thalamus and hippocampus, bilateral putamen and caudate, temporal pole and language areas. The next filter providing most variance for the 2-back task, i.e. filter 2, involves left thalamus, bilateral hippocampus, brainstem, cingulate and paracingulate gyrus, bilateral inferior frontal gyrus and left inferior temporal gyrus. On the other side of the eigen-spectrum, the filter maximizing variance for the 0-back task, i.e. filter 165, involves brainstem, cingulate gyrus, left thalamus, frontal pole and temporal gyrus and the next strongest filter for the 0-back task, i.e. filter 164, includes bilateral thalamus, brainstem, cingulate gyrus, frontal pole and motor cortex. Summarizing, the networks identified for the 2-back task involve areas purely related to working memory task while the 0-back task involves areas more related to the attention and salience network. This clearly reflects that the 2-back task involves stronger memory processing than the 0-back case which involves sustained attention.

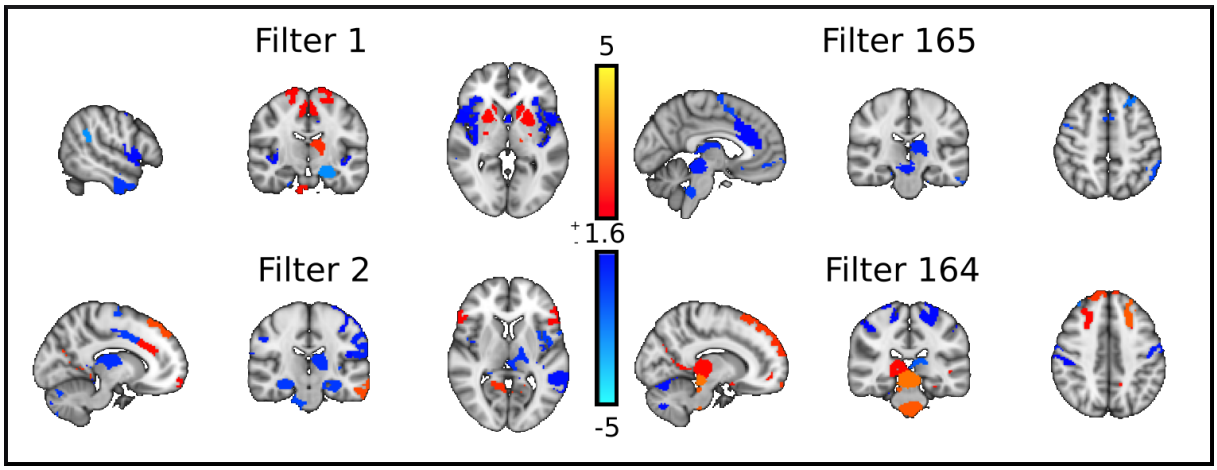


Figure 6: Spatial maps obtained using SPADE to compare the 2-back against 0-back working memory fMRI task data. Left side: filters 1 and 2 show the filters maximizing variance for 2-back memory task. Right side: filters 164 and 165 show the filters maximizing variance for 0-back memory task.

In general these presented results show some overlap with [22]. More specifically, some areas of filters 1 and 2 are present in both analyses. However, SPADE also reveals additional discriminative areas not found by the direct univariate analysis, including for example the striatum or the superior frontal gyrus.

Discussion

In this work we showed that we can achieve optimal linear filtering of fMRI data by applying an initial spatial dimensionality reduction followed by simultaneous diagonalization of covariance matrices [1,2]. We used a regularized version of the covariance matrices [14], introduced a permutation strategy for model order selection, i.e. how many spatial filters optimize discrimination, and we denote the fully developed model as SPADE (Spatial Patterns for Discriminative Estimation). We validated the use of this strategy by using the high-quality HCP data from 1063 subjects. The presented analyses illustrate the power of SPADE filters for discrimination of

fMRI data, and although accurate classification can be achieved also using (for example) an SVM in a high-dimensional space (full covariance space), the power of the presented model is the intuitive interpretation of the results that are summarized as a small set of spatial maps that can be used to further develop our understanding of brain function. Contrary to the SVM, SPADE provides simple features that discriminate between the compared conditions and which we can easily visualize back to brain space. This is certainly a big advantage with respect to discrimination in full covariance, correlation/partial correlation space where the interacting factors between different spatial nodes make the interpretation more difficult.

Our analyses revealed a strong specificity of the spatial distribution of different fMRI modalities across the three considered modalities (rest, motor and working memory). Each of the discriminative spatial maps obtained through SPADE reflects brain areas well-known to be relevant for a particular modality, reflecting for example that a stronger engagement of the DMN is recorded during rest than during task fMRI. In the working memory tasks, filters reveal independent effects for the ventral-dorsal pathways; the sensory integration is represented in one filter and the higher-order memory storage and utilization in another filter. Indeed, these two pathways start by sending sensory information to the parietal and temporal areas, characterizing spatial versus categorical features related to the items to memorize. Both connect to the temporo-medial structure (TMS) via specific connectivity to the parahippocampus, enthorinal cortex, before converging on the hippocampus [30]. The TMS then connects through the temporal pole to the (prefrontal, pole) frontal and (anterior) cingulate to store and use the information. Regarding working memory, we observe the phonological loop with both Wernicke and Broca areas [31–33]. SPADE is extracting differentiating filters out of these networks, demonstrating the existence of independent dynamics that can be the subject of further studies.

Ultimately, the SPADE model can extract regions of connectivity variation, obtained across different scenarios using fMRI data. We have shown its ability when using big data by demonstrating that it is able to pick up on areas relevant to each fMRI scenario but also by showing its close to optimal discrimination between representative fMRI experiments. We consider that the model order estimation, as well as the classification generalization of the model, can be used as a proxy of the specificity of the resulting spatial maps. As a consequence, the quality of the spatial maps reported here for interpretation is very high. Clearly, this occurs due to the strong differences in paradigms that we differentiate; for example, resting state from motor task fMRI. Further, the presented extended numerical analyses identify the robustness of the results to the number of subjects and temporal sampling of the fMRI and can serve as a reference to identify which scenarios can be addressed though this methodology. Although high classification accuracy and a low standard deviation can be achieved using around 100 subjects, getting strongly reproducible spatial maps with low standard deviation was only achieved using 200 or more subjects. We observed that one side of the eigen-spectrum achieved higher spatial reproducibility with less data than the other side. A solution to this fact, that is part of current research, could be to learn these last filters as top filters of a new simultaneous diagonalization where the input covariance matrices to the algorithm are used in a reversed order.

To show the performance of SPADE to discriminate even between relatively similar paradigms, we applied it to compare 2-back working memory from 0-back working memory task fMRI data. We showed that the 2-back task involves areas well known to be involved in working memory while the 0-back did not involve for example the hippocampus, but instead, the identified networks involved areas related to the salient and attention networks, as the frontal pole or the dorsal anterior cingulate gyrus. In comparing our results with those presented in [22], we note that our analysis point to connectivity changes between 2 and 0-back task at areas known to be involved in working memory tasks that were not clearly identified using the GLM approach in [22]. This is not a surprising result since both analyses are looking at different statistics in the data; while [22] reports changes in the mean fMRI signal across conditions through a GLM analyses, the findings we report using SPADE refer to changes in covariation between brain areas. Consequently, although the results are different, the analysis we present here and that in [22] are not exclusive but rather complementary, with both having their advantages and limitations: while for memory reasons we cannot yet compute the SPADE solutions at the full brain, with a voxel-wise resolution, the computational load of performing the SPADE decomposition itself is much lower than that of a typical GLM. As an example of the computational advantage of the SPADE model we observe that after computing the ROI time-series for the 1063 subjects included in this study the computation of the full SPADE spatial maps took less than two minutes using a standard laptop.

Although working full brain voxel-wise would be optimal, it is to date computationally intractable and our choice of working in a parcellated full brain [20] could be replaced by working in a resting state network space or even voxel-wise when restricted to an anatomical ROI that might be relevant for a particular investigation. Here we considered a parcellated full brain for ease of results interpretation and visual validation. Further, since the simultaneous diagonalization introduced by [1] can be used for any pair of symmetric positive definite matrices, and given the growing neuroimaging interest in working with partial correlation matrices, we considered the extension of the SPADE algorithm to partial correlation matrices and observe that it did not improve in discrimination with respect to the original formulation. Nevertheless we believe that such an approach warrants more thorough investigation in cases where the initial spatial dimensionality reduction results in more highly correlated data, for example when using an ICA based strategy [5] or a small spatially continuous ROI; note here that in these two scenarios, although the methodology is identical, the interpretation changes drastically since the filters will relate to resting state networks directly or voxels being involved, respectively. An extra flexibility of the presented model is the possibility of extension to multi-class versions by using combinations of binary problems or with approximations to multiple simultaneous diagonalizations [34]. Another research direction we are investigating for this model includes the study of subject-wise deviation from group functional connectivity as a strategy to address normative modelling [35] on functional connectomes, or to get subject dependent co-varying networks loadings that would allow SPADE to complement the GLM results at the single subject level on task data. A final relevant line of research we consider is the non-trivial extension of SPADE to full brain. Although this is challenging due to the high spatial resolution of fMRI data, and the strong spatial correlation in the functional data, solving this limitation would make SPADE a very strong and specific model to be applied in even more relevant scenarios, including voxel-wise task data analyses.

To conclude, the SPADE model is able to capture and summarize the brain nodes involved in network covariation changes induced by fMRI experimental manipulation. As introduced here, the SPADE model can be used to address questions of potential interest to diverse scientific and industrial communities.

Acknowledgements

The research leading to these results has received funding from the developing Human Connectome Project through a Synergy Grant by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013), ERC Grant Agreement no. 319456. We further gratefully acknowledge support from the Netherlands Organization for Scientific Research (NWO) through VENI grant to J.N. (V1.Veni.194.032) and through VIDI grant to CFB (864.12.003). We also gratefully acknowledge funding from the Wellcome Trust UK Strategic Award (098369/Z/12/Z).

Data were provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

SUPLEMENTARY MATERIAL (SM)

Discriminative simultaneous covariance diagonalization.

Consider two sets of time-series, X and Y, measured at the same spatial locations during two different states of the system being measured, and define the spatial covariance matrices of each of these series as C_X and C_Y. Usually, the unsupervised simultaneous diagonalization of both covariances is achieved through a generalized eigenvalue decomposition, involving a whitening transformation of one of the covariance structures (e.g C_X), combined with a further rotation to also diagonalize the other covariance (C_Y). A different approach was introduced in [1] where the whitening transformation is performed with respect to the sum of both covariances $C := C_X + C_Y$; such whitening can be summarized by two matrices, a rotation W and a scaling diagonal matrix P, such that $I=P^TW^TCWP$. Individually applying these transformations to C_X and C_Y, one obtains two (non-diagonal) matrices $K_X= P^TW^TC_XWP$ and $K_Y= P^TW^TC_YWP$. Performing now the eigenvalue decomposition of (for example) K_Y we obtain matrices Z_Y and diagonal D_Y such that $Z_YK_YZ_Y^T=D_Y$. Defining $V:=WPZ_Y$, we have that $V^TC_XV=I-D_Y$ and $V^TC_YV=D_Y$. This means that V defines a basis for the data that diagonalizes simultaneously C_X and C_Y and, and by

construction, the sum of the variances of the projections of data X and Y into each of the new basis vectors adds to one; consequently, for basis directions where the projection of X has large variance (i.e. close to one), the projection of Y must have low variance (i.e. close to zero), and vice versa. The algorithmic imposition to maximize variance for one class and minimize it for the other one, makes of this process a supervised multivariate dimensionality reduction model, and the new learned basis is formed of spatial filters that provide optimal linear discrimination in terms of variance of the projected data (Blankertz et al., 2011). Note that for implementation purposes it all reduces to solving a generalized eigenvalue problem which can be easily accomplished using most software platforms, for example using the matlab command eig (C_x, C) or the python command scipy.linalg.eigh(C_x,C,eigvals_only=False)). In the recent literature, the process described in [1] is commonly used as a feature extractor for EEG based brain computer interfaces [2,7] and the features extracted for discrimination are usually the logarithmic of the variance of the projected data; by construction of the algorithm the most discriminative dimensions are the associated with the extreme eigenvectors since towards the center of the eigen-spectrum the variances for both conditions tend to be close to ½. In practice, for discrimination purposes the dimension is usually reduced to a few pairs of extreme eigenvectors and a low dimensional linear classifier suffices [7].

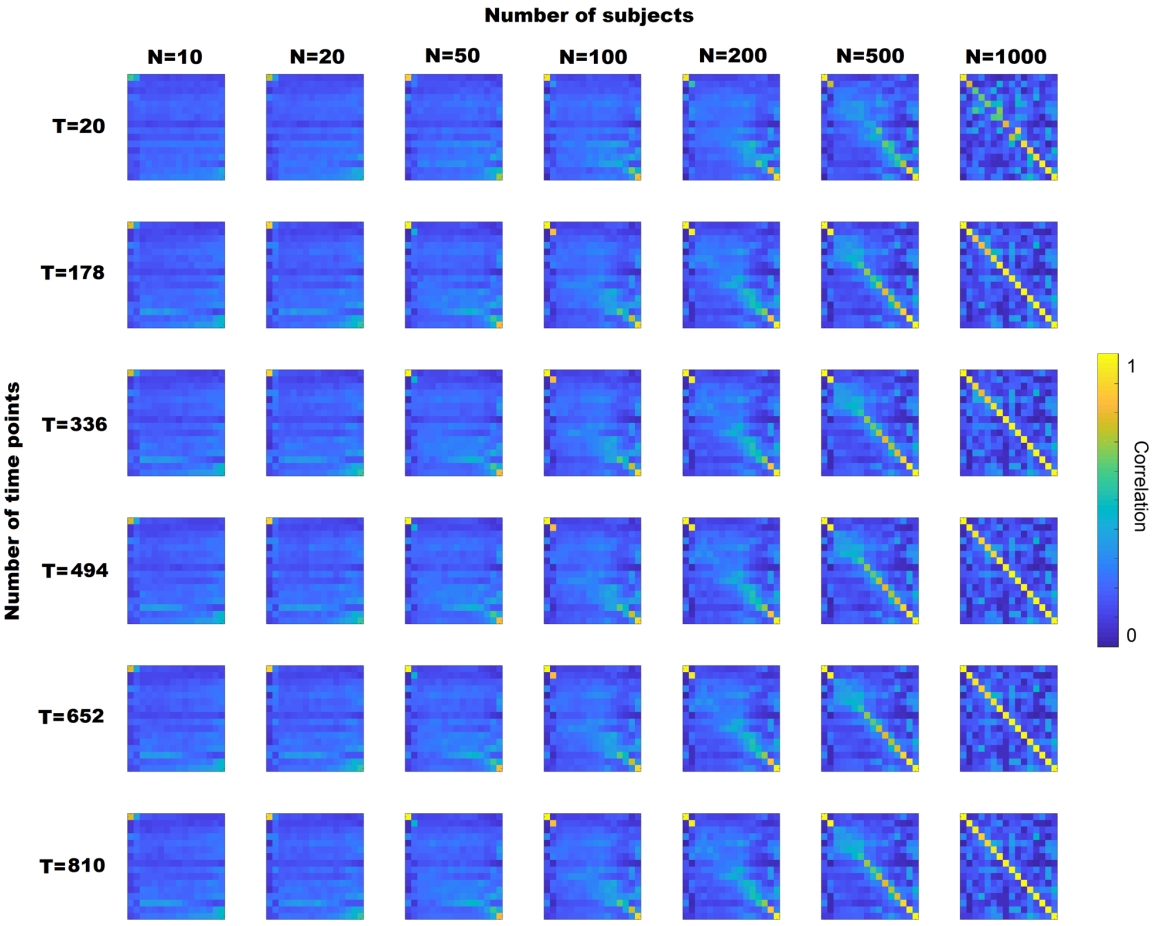
Interpretation of the linear filters

Since the simultaneous diagonalization assumes no implicit noise model, the spatial filters learned, i.e. the columns of V (V_k) cannot be directly interpreted and the interpretable spatial maps (A_k) are obtained from $A=C_{X|Y}VC_S^{-1}$ [15,18], where X|Y denotes the spatially concatenated X and Y data, and S the data projected to the new basis V, i.e. $S= V^T(X|Y)$. For completeness we note here that X|Y = AS + e, where e denotes the residual of the approximation X|Y = AS and for a more detailed description we refer the reader to [15,18].

Model order estimation

Although for classification purposes the SPADE model order selection is not critical due to the latter integration of a classifier, for interpretation purposes we are required to select which filters provide significant connectivity covariation information. To that end we address for the first time the model order estimation on this model by permutation testing to find a robust estimation of which filters explain more variance for a modality than for the other one with respect to a null distribution obtained by permutation. To that end a p-value is computed for each basis vector (column of V, V_k) by comparing the absolute difference between the variances of the data projections, Abs(Var(V'_k*X)- Var(V'_k*Y)), with a distribution of values obtained when the simultaneous diagonalization new basis (V) is computed using randomized groups of covariances. In the dataset set used in the main text the randomization is achieved by permuting half of the subject's covariance matrices across different fMRI modalities.

476 ***Spatial reproducibility***



477

478 Figure S1: SPADE spatial maps reproducibility for RT vs WM. Each subfigure shows the mean correlation matrix between the significant
479 spatial maps using the full sample (N=1063) and the ones obtained for N and T as indicated on each row and column. The mean is
480 computed over 100 random selections of subjects (see section Methods for details). The size of each correlation matrix equals the number
481 of significant selected filters as indicated in Table 1 and are accordingly ordered, with the top left corner and bottom right corner relating
482 to the spatial reproducibility of the most important filters, the first (#1) and the last (#165) ones.

483

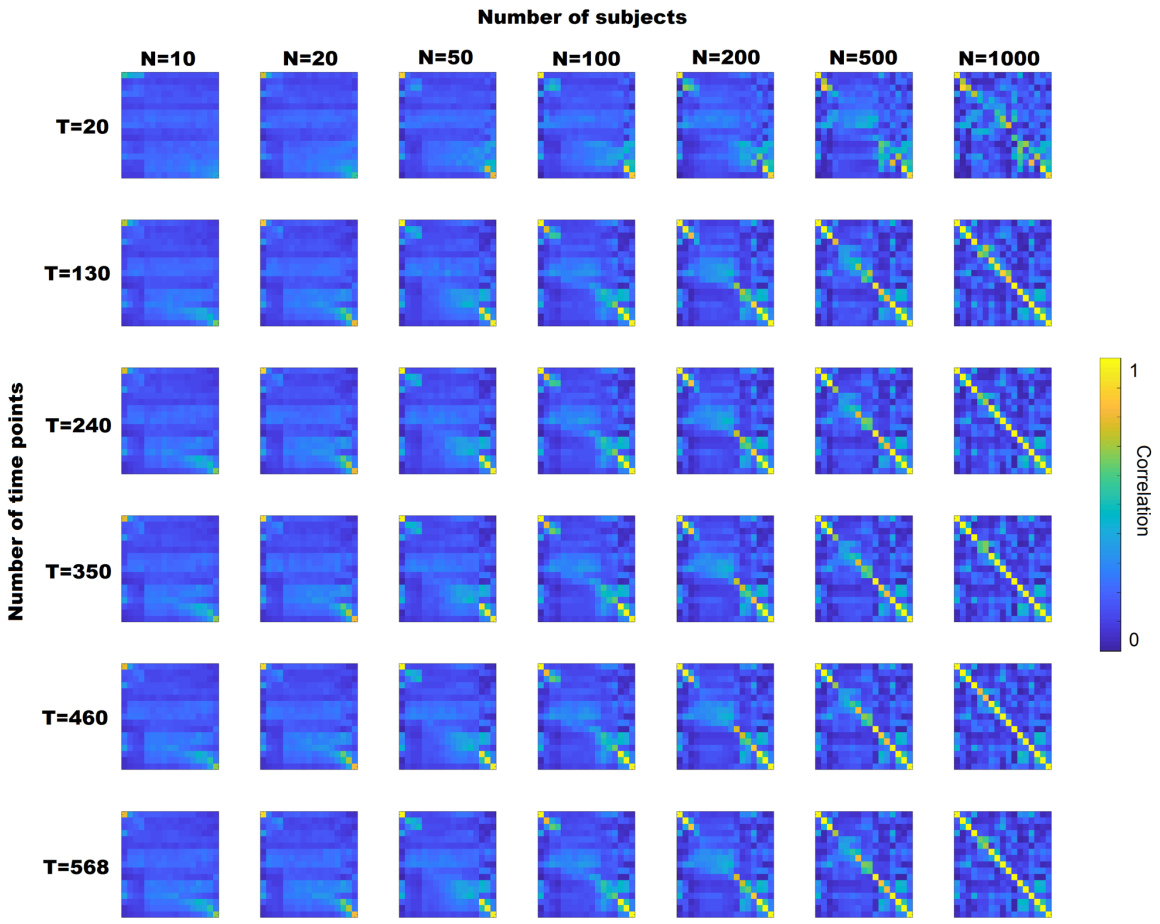


Figure S2: SPADE spatial maps reproducibility for MT vs WM. Each subfigure shows the mean correlation matrix between the significant spatial maps using the full sample (N=1063) and the ones obtained for N and T as indicated on each row and column. The mean is computed over 100 random selections of subjects (see section Methods for details). The size of each correlation matrix equals the number of significant selected filters as indicated in Table 1 and are accordingly ordered, with the top left corner and bottom right corner relating to the spatial reproducibility of the most important filters, the first (#1) and the last (#165) ones.

References

1. Fukunaga, K. *Introduction to Statistical Pattern Recognition*. *Pattern Recognition* **22,** (1990).

2. Ramoser, H., Müller-Gerking, J. & Pfurtscheller, G. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. Rehabil. Eng.* **8,** 441–446 (2000).

3. Friston, K., Holmes, A. & Worsley, K. Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* (1994).

4. Aguirre, G. K. & Detre, J. A. Experimental Design and the Relative Sensitivity of BOLD and Perfusion fMRI. **500,** 488–500 (2002).

5. Beckmann, C. F. & Smith, S. M. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans Med Imaging* **23(2),** 137–152 (2004).

6. Beckmann, C. F. & Smith, S. M. Tensorial extensions of independent component analysis for multisubject FMRI analysis. **25,** 294–311 (2005).

7. Blankertz, B., Lemm, S., Treder, M., Haufe, S. & Müller, K. R. Single-trial analysis and classification of

ERP components - A tutorial. *Neuroimage* **56,** 814–825 (2011).

8. A. Llera, R. Chauvyn, P. Mulders, M. Mennes, C. F. B. Rest vs Task fMRI connectivity: a simultaneous covariance diagonalization approach. in *23-th Organization for Human Brain Mapping (OHBM)* (2017). doi:10.13140/RG.2.2.11971.09765

9. Subbaraju, V., Belathur, M. & Sundaram, S. Identifying differences in brain activities and an accurate detection of autism spectrum disorder using resting state functional-magnetic resonance imaging : A spatial filtering approach. *Med. Image Anal.* **35,** 375–389 (2017).

10. Subbaraju, V. Identification of lateralized compensatory neural activities within the social brain due to autism spectrum disorder in adolescent males. **47,** 631–642 (2018).

11. Aradhya, A. M. S., Subbara, V., Sundaram, S. & Sundarara, N. Regularized Spatial Filtering Method (R-SFM) for detection of Attention Deficit Hyperactivity Disorder (ADHD) from resting-state functional Magnetic Resonance Imaging (rs-fMRI). 5541–5544 (2018).

12. Itani, S. & Thanou, D. Combining Anatomical and Functional Networks for Neuropathology Identification : A Case Study on Autism Spectrum Disorder.

13. Zhang, W. *et al.* Discriminating stress from rest based on resting-state connectivity of the human brain: A supervised machine-learning study. *Hum. Brain Mapp.* **(In Press),** (2020).

14. Lotte, F. & Guan, C. Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms. *IEEE Trans. Biomed. Eng.* **58,** 355–362 (2011).

15. Haufe, S. *et al.* On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* **87,** 96–110 (2014).

16. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: An overview. *Neuroimage* **80,** 62–79 (2013).

17. Elam, J. S. & Van Essen, P. D. Human Connectome Project. in *Encyclopedia of Computational Neuroscience* 1–4 (2013).

18. Parra, L. C., Spence, C. D., Gerson, A. D. & Sajda, P. Recipes for the linear analysis of EEG. *Neuroimage* **28,** 326–341 (2005).

19. Van Essen, D. C. *et al.* The Human Connectome Project: A data acquisition perspective. *NeuroImage* **62,** 2222–2231 (2012).

20. van Oort, E. S. B. *et al.* Functional parcellation using time courses of instantaneous connectivity. *Neuroimage* **170,** 31–40 (2018).

21. Bishop, C. M. *Pattern Recognition and Machine Learning*. *Pattern Recognition* **4,** (2006).

22. Barch, D. M. *et al.* NeuroImage Function in the human connectome : Task-fMRI and individual differences in behavior. *Neuroimage* **80,** 169–189 (2013).

23. Cole, M. W., Ito, T., Bassett, D. S. & Schultz, D. H. Activity flow over resting-state networks shapes cognitive task activations. *Nat. Neurosci.* **19,** 1718–1726 (2016).

24. Sahyoun, C., Floyer-Lea, A., Johansen-Berg, H. & Matthews, P. . Towards an understanding of gait control: brain activation during the anticipation, preparation and execution of foot movements. *Neuroimage* **21,** 568–575 (2004).

25. Smith, S. M. *et al.* Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci.* **106,** 13040–13045 (2009).

26. Owen, A. M., McMillan, K. M., Laird, A. R. & Bullmore, E. N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Hum. Brain Mapp.* **25,** 46–59 (2005).

27. Soto, D., Rotshtein, P. & Kanai, R. Parietal structure and function explain human variation in working memory biases of visual attention. *Neuroimage* **89,** 289–296 (2014).

28. Takahashi, E., Ohki, K. & Kim, D. S. Dissociation and convergence of the dorsal and ventral visual working memory streams in the human prefrontal cortex. *Neuroimage* **65,** 488–498 (2013).

29. Haak, K. V. & Beckmann, C. F. Objective analysis of the topological organization of the human cortical visual connectome suggests three visual pathways. *Cortex* **98,** 73–83 (2018).

30. Fernandez-Ruiz, A. & Oliva, A. Distributed Representation of 'What' and 'Where' Information in the Parahippocampal Region. *J. Neurosci.* **36,** 8286–8288 (2016).

31. Müller, N. G. & Knight, R. T. The functional neuroanatomy of working memory: Contributions of human brain lesion studies. *Neuroscience* **139,** 51–58 (2006).

32. Baddeley, A. Working memory: Looking back and looking forward. *Nat. Rev. Neurosci.* **4,** 829–839 (2003).

33. Baddeley, A. The episodic buffer: A new component of working memory? *Trends Cogn. Sci.* **4,** 417–423 (2000).

34. Grosse-wentrup, M., Member, S. & Buss, M. Multi-class Common Spatial Patterns and. **55,** 1–10 (2008).

35. Marquand, A. F., Rezek, I., Buitelaar, J. & Beckmann, C. F. Archival Report Understanding Heterogeneity in Clinical Cohorts Using Normative Models : Beyond Case-Control Studies. *Biol. Psychiatry* **80,** 552–561 (2016).