MOX-Report No. 109/2023

Anatomically compliant modes of variations: new tools for brain connectivity

Clementi, L.; Arnone, E.; Santambrogio, M.D.; Franceschetti, S.; Panzica, F.;
Sangalli, L.M.

Anatomically compliant modes of variations: new tools for brain connectivity

Letizia Clementi[1,2,3], Eleonora Arnone[4], Marco D. Santambrogio[2], Silvana Franceschetti[5], Ferruccio Panzica[5], Laura M. Sangalli[1*],

1 MOX - Department of Mathematics, Politecnico di Milano, Milan, Italy
2 Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy
3 CHDS, Center for Health Data Science, Human Technopole, Milan, Italy
4 Department of Management, University of Turin, Turin, Italy
5 Fondazione IRCCS Istituto Neurologico "C. Besta", Milan, Italy

* laura.sangalli@polimi.it

Abstract

Anatomical complexity and data dimensionality present major issues when analysing brain connectivity data. The functional and anatomical aspects of the connections taking place in the brain are in fact equally relevant and strongly intertwined. However, due to theoretical challenges and computational issues, their relationship is often overlooked in neuroscience and clinical research. In this work, we propose to tackle this problem through Smooth Functional Principal Component Analysis, which enables to perform dimensional reduction and exploration of the variability in functional connectivity maps, complying with the formidably complicated anatomy of the grey matter volume. In particular, we analyse a population that includes controls and subjects affected by schizophrenia, starting from fMRI data acquired at rest and during a task-switching paradigm. For both sessions, we first identify the common modes of variation in the entire population. We hence explore whether the subjects' expressions along these common modes of variation differ between controls and pathological subjects. In each session, we find principal components that are significantly differently expressed in the healthy vs pathological subjects (with p-values < 0.001), highlighting clearly interpretable differences in the connectivity in the two subpopulations. For instance, the second and third principal components for the rest session capture the imbalance between the Default Mode and Executive Networks characterizing schizophrenia patients.

Introduction

Functional Magnetic Resonance Imaging (fMRI) based studies are standard in neuroscience to explore functional connections in the brain and how they can be affected by psychiatric and neurological diseases [1]. Neurodegenerative pathologies, indeed, affect both anatomical and functional connections in the brain, resulting also in modifications in their relationship (see, e.g., [2,3]). Nonetheless, the analysis of this type of data still presents relevant theoretical and technical challenges, due for instance to the data dimensionality and to the complicated anatomy of the brain.

The high dimensionality of fMRI data poses major statistical and computational issues, yet to be fully addressed [4–6]. Various alternative approaches to dimensionality

reduction of fMRI data have been proposed in the literature, including adaptations of Principal Component Analysis (PCA) to the context of large-scale datasets (see, e.g., [7]). Moreover, Blood-Oxygenation Level Dependent (BOLD) signals have also been considered as functional data [8–11], and analysed through functional PCA methods (see, e.g., [12,13]). However, high dimensionality is not the only challenge when applying PCA to the analysis of fMRI data.

Other critical issues are posed for instance by anatomical complexity. Indeed, the brain cortex has a highly complicated anatomy, characterized by the presence of gyri and sulci, fundamental to its correct functioning. Of course, fMRI signals observed within the same brain area are usually related. It should though be noted that areas carrying out distinct functions may be close in three-dimensional space, due to the aforementioned highly-folded structure of the brain. Because of these aspects, the application of classical multivariate dimensional reduction techniques, as well as of state-of-the arts methods for functional data, such as functional PCA, may lead to unsatisfactory or misleading results. The former techniques may present unsatisfactory results, since they disregard the spatial structure in the data, and are hence incapable of borrowing information from signals observed at nearby locations, where they are correlated. The latter techniques, based on functional data analysis, may instead produce artifacts, since these methods rely on the Euclidean metrics, and hence neglect the highly non-Euclidean brain anatomy, and its effects on the BOLD signals. See also the discussions in [14–16].

For these reasons, [15] proposed a form of PCA able to handle observations located over a complicated two-dimensional manifold, and applied it to Functional Connectivity (FC) maps computed over the cortical surface of the brain. In particular, they consider a penalization term that permits to include the manifold geometry into the estimation problem, thus appropriately accounting for the complicated shape of the cerebral cortex [17]. The extracted eigenfunctions may hence be treated with classical Functional Data Analysis tools. Others advanced data analysis methods capable of complying with the complex anatomy of the cortex have been proposed in, e.g., [14,18–20].

In this work, we aim to expand the approach proposed by [15] to data observed within volumetric domains with complicated shapes, considering FC maps computed within the grey matter volume. We propose a method to compute principal components of connectivity maps, where we appropriately consider domain-specific distances, within the grey matter volume, respecting the highly non-Euclidean anatomy of the brain. In particular, the method features a regularizing term, defined over the grey matter volume, that leads to the computation of distances within this volume, and not across sulci or other concavities in the grey matter. Moreover, to address the computational issues linked to the dimensionality of data, the considered technique leverages an appropriate finite element discretization, defined over the tetrahedral mesh representing the grey matter volume, over which the fMRI data are mapped. By merging approaches and ideas from statistics and from scientific computing, the considered Smooth Functional Principal Component Analysis (SF-PCA) tackles the challenges posed by both anatomical complexity and data dimensionality, permitting to explore the variability in neuroimaging signals associated with neural connectivity in the grey matter, complying with its complicated structure, at a sustainable computational cost.

To prove the validity of SF-PCA in this context, we analyze data from an experimental population composed of subjects affected by schizophrenia and by controls. Schizophrenia is a severe psychiatric condition characterized by hallucinations, delusions, and impaired cognitive function, hypothesized to result from abnormal anatomical neural connectivity, and a consequent decoupling of the brain's integrative thought processes [21–24]. The impact of this pathology on brain connectivity has been

widely explored, both through resting-state and task-based fMRI experiments, leading to a well-defined description of the impact of schizophrenia on brain connections [25–27]. For this purpose, we consider the FC maps computed on fMRI data acquired during resting-state and also while carrying out a task based on the switching paradigm. We assess the adequacy of the proposed SF-PCA method to analyze FC maps computed on grey matter volume domains. In particular, we first consider the entire population, including both controls and subjects affected by schizophrenia, and we extract the main common modes of variation of brain connectivity, identified by the Smooth Functional Principal Components (SF-PCs) of the FC maps. We then explore whether the corresponding SF-PCs scores, that describe the subjects' expression along these common modes of variation, differ significantly between the healthy and pathological subgroups. A significant difference in the expression of the main modes of variation would indeed highlight differences in the expression of the associated connectivity patterns. This aspect is further explored by Linear Discriminant Analysis of the SF-PCs scores, considering the Area Under the Curve of its receiver operating characteristic curve [28]. A comparison with standard MV-PCA highlights the advantages of the proposed method in terms of interpretability of the obtained principal components. Finally, we discuss how the differences in the expression of the SF-PCs corroborate previous findings in literature and also shed new light on connectivity impairment in schizophrenia.

Materials and methods

This section describes the fMRI dataset employed in the present study, the experimental protocol, and the population enrolled, alongside the preprocessing and analysis pipelines.

Data

The data analyzed in this work has been originally collected and made available by [29]. Participants were reached through a patient-oriented strategy involving local clinics and online portals, different from the methods employed to recruit healthy volunteers, who were enlisted by community advertisements from the Los Angeles area. After receiving a verbal explanation of the study, participants gave written informed consent following procedures approved by the Institutional Review Boards at UCLA and the Los Angeles County Department of Mental Health [29]. All data was collected before 2016 (the date on which the acquisition started is undisclosed) and first published on 27 January 2016. In this work we are employing the preprocessed version of the dataset by [30], made available by the Consortium for Neuropsychiatric Phenomics at UCLA. The last update of the preprocessed version of the dataset is dated 21st April 2020. In none of the stages of this work, we had access to information that could identify the participants in the study. The fMRI data were acquired through a 3T Siemens Trio scanner. T1-weighted high-resolution anatomical scans were collected with the following parameters: slice thickness=1mm for 176 slices, TR=1.9s, TE=2.26ms, matrix=256x256, FOV=250mm. The T2*-weighted BOLD fMRI sequence parameters instead are: slice thickness=4mm for 34 slices, TR=2s, TE=30ms, flip angle=90°, matrix=64×64, FOV=192mm. For both the considered sessions, the authors acquired 34 volumes. The data are made available as coregistered to the skull-stripped ICBM 152 Nonlinear Asymmetrical template version 2009c [31,32]. The registration to the template allows to indirectly account for structural dissimilarities characterizing the two groups, among which the atrophy affecting schizophrenia patients.

The complete preprocessing and alignment pipeline, firstly proposed and

standardized in [33], is described in [30]. Nonetheless, for better clarity, the main steps performed can be summarized in three main blocks:

- **T1w preprocessing**: bias field is addressed via ANTs N4BiasFieldCorrection v2.1.08 [34], skullstripping using antsBrainExtraction.sh v2.1.0, and coregistration to the ICBM 152 Nonlinear Asymmetrical template version 2009c9 [31,32] through ANTs v2.1.0 [35].

- **Surface estimation** from the bias field corrected T1w, through antsBrainExtraction. The skull-stripping is then performed via FreeSurfer v6.0.0 [36].

- **EPI preprocessing**: motion artifacts and geometric distortions are removed through MCFLIRT v5.0.9 [37]. Skull-stripping is perfomed via BET and 3d AutoMask, finally coregistration to T1w is applied via FreeSurfer. The EPI preprocessing is performed in a single step using antsApplyTransformations v2.1.0, a tool performing a non-linear registration step addressing motion correcting, transformation to T1 weighted space and MNI template warp.

Participants

The dataset contains the fMRI scans of patients affected by either schizophrenia, bipolar disorder, or ADHD plus the control group. The fMRI were acquired during rest and while performing different tasks. Exclusion criteria were: left-handedness, pregnancy, history of head injury with loss of consciousness, diagnoses ascribable to more than one patient group or contraindications to undergo fMRI scanning.

For the present study, we select a subsample of the dataset, referring to 120 subjects from the control (CTRL) group, and 50 affected by schizophrenia (SCHZ). We analyze their resting state (REST) fMRI. Among the available task-based fMRIs, we consider those referring to the task-switching (SWITCH) paradigm. A control subject did not perform the switching task, thus the resulting population comprises 120 CTRL fMRIs for the REST session and 119 fMRIs for the SWITCH one. Instead, all the 50 SCHZ subjects participated in both sessions. Median age is 28 years (interquartile range: 15 years) for CTRL and 37 years (interquartile range: 14 years) for SCHZ. The percentage of males is 51.26% for CTRL and 76.00% for SCHZ.

Experimental Paradigm

The experimental paradigm consists of two separate sessions. During each session, the fMRI is acquired while the subject performs a task. All the procedures were approved by the Institutional Review Boards at UCLA and the Los Angeles County Department of Mental Health. In the present work, as above mentioned, we consider two fMRI sessions, acquired during:

- REST: the participants were asked to relax and keep their eyes open. They were not presented any stimuli or asked to respond during the scan. This session lasted 304 seconds.

- SWITCH: the participants were presented an image plus an instruction cue describing how to react to the image. Each image included a shape and a colour, from a set of four (either a red triangle, a red circle, a green triangle, or a green circle). Participants were given instruction to respond either to the stimuli's color (possible cues: "COLOR" or "C") or shape (possible cues: "SHAPE" or "S"). On

33% of trials, the instructions switched, such that participants were instructed to switch from responding from shape to color, or vice versa. On the remaining 67% of trials, the instructions remained the same, but the cues changed (e.g. from "SHAPE" to "S"). This task was designed to measure the changes in reaction time between trials requiring versus not requiring a switch in responding.

Analysis

This subsection describes in detail the analysis pipeline, providing also indications on the Matlab [38] and `R` [39] software used.

Brain mesh creation and data preprocessing

In order to apply Smooth Functional Principal Component Analysis (SF-PCA) we need a Finite Element mesh that discretizes the common brain volume, given by the ICBM 152 Nonlinear Asymmetrical template (version 2009c) [31,32], to which all subjects' data have been aligned, as detailed in the preprocessing steps above. With some abuse of notation, we denote by $\mathcal{D}$ both the original template domain, as well as the tetrahedral mesh discretizing it. The mesh $\mathcal{D}$ is computed through the `Brain2Mesh` Matlab toolbox (version 0.80) [40,41]. The number of mesh nodes is automatically computed by the `Brain2Mesh` tool, with an algorithm that considers the tissues to be segmented and the desired mesh density. In this work, we selected the finest mesh possible with respect to the data resolution. Specifically, approximately 46 000 voxels of the original fMRI contain information regarding brain activity, including both the grey and the white matter. Since we only consider the grey matter, this leads to a mesh of 36 035 nodes. Each mesh node is associated with a cerebral region according to the Hammers' atlas (rigidly registered to the template) [42,43] and to a voxel of the fMRI. Thus, each node in the grey matter mesh corresponds to a BOLD signal, that we highpass filter at 0.01 Hz to remove the linear drift and low-frequency noise [1].

Connectivity Map

For each subject, we compute the Functional Connectivity (FC) maps through a seed-based technique, a straightforward approach to FC computation [1]. Previous research has shown the effectiveness of this method, also in comparison to more sophisticated approaches (see, e.g., [44,45]). The seeds, named regions of interest, are selected regions which, according to previous findings, are of particular interest with respect to the considered task. In particular, we consider the following seeds:

- Seed for REST session: we consider as the region of interest the anterior part of the cingulate gyrus, in the left hemisphere, that corresponds to the label number 24 of the Hammer's atlas [27]. In the considered brain discretization, this seed region is discretized in $m_{seed} = 113$ nodes. Previous studies observed anatomical abnormalities in the grey matter of patients affected by schizophrenia [46] in association with a loss of connectivity in this region [47,48]. Moreover, this region is hypothesized to play a pivotal role in the onset and progression of the pathology [49].

- Seed for SWITCH session: we consider as the region of interest the middle frontal gyrus of the right hemisphere, corresponding to the Hammer's atlas label number 29. This seed region is discretized in $m_{seed} = 1007$ nodes. The strength of functional connections between the right and left middle frontal gyrus has been found to predict performances in task-switching in subjects affected by cognitive impairment [50,51].

For each subject, we compute the FC maps through the following preprocessing steps:

1. we associate to each node $\mathbf{p}_j$ (for $j = 1, \ldots, m$ and $m = 36\,035$), of the grey matter volume $\mathcal{D}$, its BOLD signal $\mathbf{t}_j$;

2. we compute the mean BOLD signal $\bar{\mathbf{t}}$ over the seed region;

3. we associate to each node $\mathbf{p}_j$ a single value of correlation ρ_j, that corresponds to the Pearson correlation coefficient between the mean time series over the seed region $\bar{\mathbf{t}}$ and the BOLD $\mathbf{t}_j$ in $\mathbf{p}_j$;

4. we compute the z-Fisher's transform of the Pearson's correlation in each node, in order to stabilize variance [1].

Finally, for both the REST and SWITCH, we organize the extracted z-scores in a matrix $\mathbf{Z}$ of size $n \times m$, where n is the number of subjects enrolled in the study and m the number of nodes in the mesh. As indicated above, the number of subjects n is 170 for REST and 169 for SWITCH. The number of mesh nodes m is $36\,035$ for each session, as it does not depend on the considered seed.

Smooth Functional Principal Component Analysis

In our data setting, each row of the data matrix $\mathbf{Z}$ is the noisy evaluation of the connectivity map for one subject, defined over the grey matter volume $\mathcal{D}$, and evaluated at the m nodes $\mathbf{p}_j$ of its tetrahedral mesh discretization. Our goal is to extract some components that describe the strongest modes of variations in the computed connectivity maps, that is in the data matrix $\mathbf{Z}$, similarly to the standard MultiVariate Principal Component Analysis (MV-PCA). However, differently from MV-PCA, where the components are vectors in $\mathbb{R}^m$, we would like these components to be functions f defined over the whole domain $\mathcal{D}$, i.e., $f : \mathcal{D} \to \mathbb{R}$, and to be smooth. This would permit to appropriately consider the spatial structure in the observed map. To do so, we embed our analysis in the framework of Functional Data Analysis [8–11]. In particular, we consider the Smooth Functional Principal Component Analysis (SF-PCA) method, previously proposed by [15] for the analysis of functional maps referring to the cortical surface, and we extend it to the case of data referred to the grey matter.

The proposed SF-PCA approach relies on the computation of the principal components as a low-rank approximation of the data matrix [52], and proceeds to estimate the components sequentially, starting from the first one. Specifically, for each component, we consider a rank-1 approximation to a PCA problem for zero-centered random variables. With abuse of notation, we continue to denote by $\mathbf{Z}$ the zero-centered data matrix. The rank-1 approximation problem involves the minimization of the functional

$$\sum_{j=1}^{m} \sum_{i=1}^{n} \{z_{ji} - s_i f(\mathbf{p}_j)\}^2$$

where z_{ji} is the value of the zero-centered connectivity map, evaluated at the mesh node $\mathbf{p}_j$ for the i-th subject, f is the component we aim to estimate, and $\mathbf{s} = (s_1, \ldots, s_n)^\top$ is the associated score vector. The above functional does not account for the desired smoothness of the PC f. Therefore, following a common approach in regularized regression, we add to the functional the term $\mathbf{s}^\top \mathbf{s} \int_{\mathcal{D}} (\Delta f)^2$, which penalizes functions with high oscillations. Specifically, the operator Δf is the Laplacian of f, defined as

$$\Delta f = \frac{\partial^2 f}{\partial \mathbf{p}_x^2} + \frac{\partial^2 f}{\partial \mathbf{p}_y^2} + \frac{\partial^2 f}{\partial \mathbf{p}_z^2}$$

and measures the local curvature of the function f, as defined on the domain $\mathcal{D}$. Therefore, the first principal component $\hat{f}$ and the first score vector $\hat{\mathbf{s}}$ are computed minimizing the functional:

$$\sum_{j=1}^{m}\sum_{i=1}^{n}\{z_{ji} - s_i f(\mathbf{p}_j)\}^2 + \lambda \mathbf{s}^\top \mathbf{s} \int_{\mathcal{D}}(\Delta f)^2 \tag{1}$$

where $\lambda > 0$ is a tuning parameter, usually referred to as smoothing parameter, that regulates the relative magnitude of the two terms in the functional. In particular, large values of λ produce a very smooth functional component, while small values of λ return a noisier principal component, closer to the multivariate one. It should be noticed that the inclusion in the estimation functional of the regularizing term $\mathbf{s}^\top \mathbf{s} \int_{\mathcal{D}}(\Delta f)^2$, that is defined over the spatial domain $\mathcal{D}$, makes the estimation problem domain-specific. Indeed, to return a smooth PC f, the method borrows information from data observed at nearby locations, i.e., data z_{ji} observed at spatial locations $\mathbf{p}_j$ that are close in the grey matter volume $\mathcal{D}$, where the distance is naturally computed within the domain $\mathcal{D}$. Because of this, the method returns estimated PC functions that comply with the complex brain shape and do not artificially smooth data at locations that are close in 3D space, but far away in the grey matter volume, as they are for instance separated by a sulcus.

The functional in Eq (1) is written in a continuous form for f on the grey matter volume. However, to carry out the minimization problem, such a problem is discretized with linear finite elements. In particular, we represent the PC function f by a linear finite element function, defined on the grey matter mesh. A linear finite element function is defined by the values that it assumes at the mesh nodes and it is linear over each mesh element, i.e., over each tetrahedron. Therefore, each linear finite element function is uniquely associated with a vector of dimension m, the number of mesh nodes. The discretization of the integral of the square of the Laplacian of f, in the second term of the functional (1), is computed following a mixed finite element approach, similarly to the manifold case in [15]. Moreover, the minimization problem is solved in a 2-step iterative algorithm: in the first step f is kept fixed and the optimization is performed in $\mathbf{s}$, while in the second step $\mathbf{s}$ is kept fix and the optimization is performed in f. As initial value for f, the first multivariate PC is used. Finally, as mentioned above, the subsequent components are computed with the same 2-step iterative algorithm, after subtracting the previous principal components from the data matrix $\mathbf{Z}$.

We apply the described SF-PCA separately to the two sessions, REST and SWITCH, setting $\lambda = 10^{-1}$. Alongside, we extract the corresponding PC scores, that indicate the subjects' expressions on the common modes of variation. The method is implemented through the R/C++ library fdaPDE [53].

Group Comparison and SF-PCs interpretation

For each session, we carry out SF-PCA on the entire population, including healthy and pathological traces, to identify the common modes of variation. Once the main modes of variation, in the entire population, are identified, we look for differences in their expression between the two subgroups of healthy and pathological subjects. Indeed, a significant difference in the expression of the main modes of variation, i.e., on the PC scores, can highlight differences in the expressions of the associated connectivity patterns. To do so, we first compare the SF-PCs' scores distribution of the SCHZ and CTRL groups through a Wilcoxon test. We set the significance threshold at $\alpha = 0.01$. We further explore differences in the SF-PC scores through an Linear Discriminant Analysis (LDA) in a leave-one-out cross-validation; the metric employed for evaluation is the Area Under the Curve (AUC) of its receiving operating curve [28]. For a

comparison with previous approaches present in literature, we replicate such analysis with the standard multivariate PCA (MV-PCA).

We finally evaluate if the diagnosis-based differences highlighted by SF-PCA can provide meaningful insights into the pathology in analysis. To do so, we investigate if the modes of variation highlighted by the SF-PCs comply with previous findings in the state of the art.

Given the unbalance, in terms of sex and age, of the SCHZ and CTRL groups, we also evaluate if the emerged differences can be linked to such covariates. For this purpose, we compare the SF-PCs' scores distributions for the two sexes through Wilcoxon tests, and the Pearson correlation between the SF-PCs' scores and the age.

Results

Before describing the results of the proposed analysis pipeline, we show that simple depiction of raw connectivity fails to highlight differences between clinically different populations. Fig 1 depicts the mean connectivity maps of the two populations, CTRL and SCHZ, during the REST and the SWITCH task. Indeed, these pictures do not capture any visible difference in the two populations.
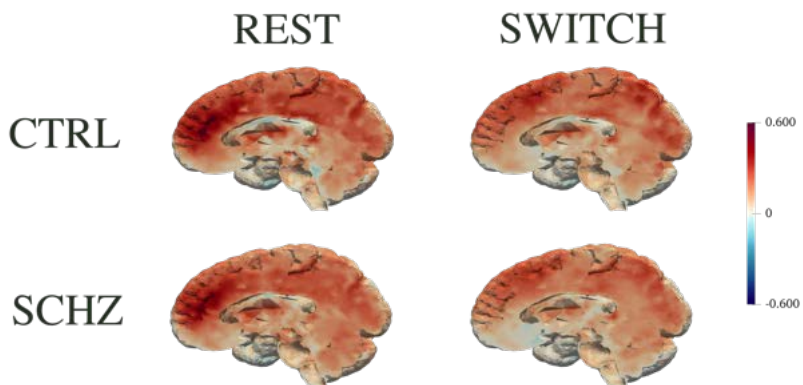

Fig 1. Mean connectivity maps of REST and SWITCH session. The first row depicts the mean z-score of the CTRL group for the two sessions, while the second depicts the mean z-score of the SCHZ group.

For both sessions, we extract four SF-PCs, named SF-PCj_REST and SF-PCj_SWITCH, for $j = 1, \ldots, 4$, respectively. We obtain a total explained variance of 20.40% for the REST session (cumulative explained variance for each component: 13.49%, 16.16%, 18.88%, 20.40%) and 20.54% for the SWITCH session (cumulative explained variance for each component: 15.24%, 18.83%, 19.38%, 20.54%).

The AUC resulting from the LDA of the scores of the four SF-PCs is 63.5% for REST and of 65.5% for SWITCH. Standard MV-PCA has the same discrimination power, having an AUC of 63.6% for REST and of 65.4% for SWITCH.

More generally, for the two considered sessions, SF-PCA and MV-PCA have the same classification performances also when considering other classification techniques, such as quadratic discriminant analysis, support vector machines and random forests. Random classification of each subject as healthy or affected by schizophrenia leads to an AUC of 54.6% for REST and of 53.6% for SWITCH.

To better investigate the distributions of the SF-PC scores in the two population subgroups, in Fig 2 and 3 we display the boxplots of the SF-PCs scores, plotted separately for CTRL and SCHZ, alongside with the results of the correspondent

Wilcoxon tests that compare the distributions of the scores for CTRL *vs* SCHZ groups, in REST and SWITCH sessions respectively. SF-PC2_REST, SF-PC3_REST and SF-PC2_SWITCH result significantly different, with p-values respectively: $p = 2 \cdot 10^{-4}$ for SF-PC2_REST, $p = 1 \cdot 10^{-3}$ for SF-PC3_REST, and $p = 6 \cdot 10^{-4}$ for SF-PC2_SWITCH. Similarly, for standard MV-PCA, the scores that are significantly different according to the Wilcoxon tests are: MV-PC3_REST ($p = 8 \cdot 10^{-4}$), MV-PC4_REST ($p = 2 \cdot 10^{-3}$), MV-PC2_SWITCH ($p = 6 \cdot 10^{-4}$), and MV-PC3_SWITCH ($p = 2 \cdot 10^{-3}$).
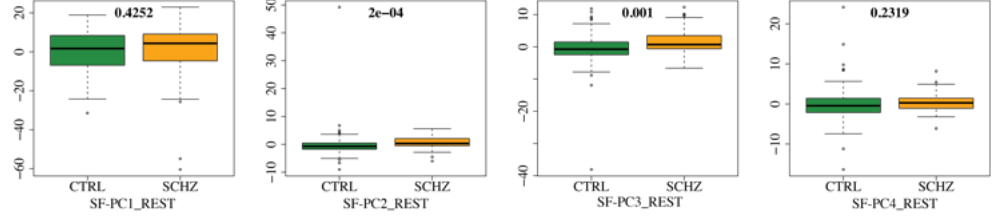

Fig 2. SF-PC scores for REST. Boxplots of scores of the first four SF-PC functions obtained by the proposed SF-PCA, computed on the entire population during REST, and plotted separately for CTRL and SCHZ groups. P-values of Wilcoxon tests that compares the distributions of the PCs scores for CTRL and SCHZ groups.
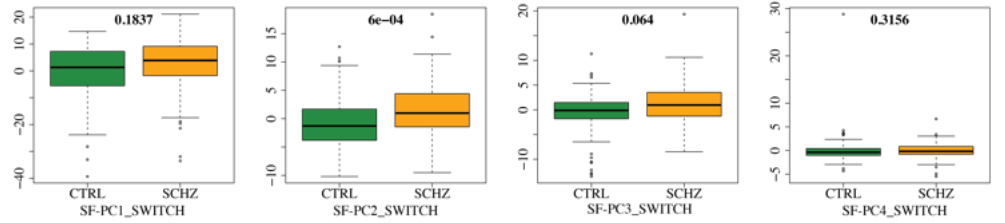

Fig 3. SF-PC scores for SWITCH. Same as Fig 2 but for SWITCH.

To interpret the principal components and visualize their impact on different brain regions, we plot the median behaviour along each component, for subjects in the two populations. In particular, Fig 4 to 6 display in the top panels the $median(s_{CTRL}) \cdot PC_j$ and in the bottom panels the $median(s_{SCHZ}) \cdot PC_j$, where s are the scores relative to each principal component. Fig 4 and 5 depict such representations for SF-PCs and MV-PCs for REST; Fig 6 shows the same representation for the SF-PCs for SWITCH (whilst the MV-PCs for SWITCH are not reported for sake of space). These pictures show how each connectivity pattern extracted by the principal component is expressed in the two groups.
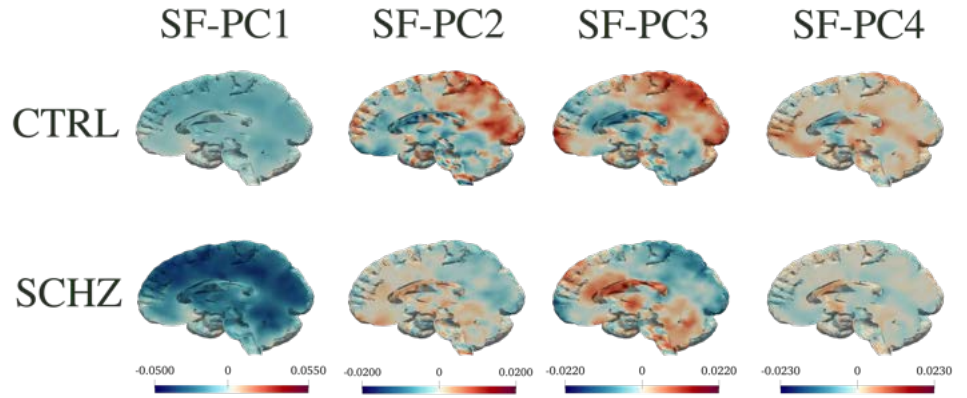
Fig 4. SF-PC functions for REST. Visualizations of the first four SF-PC functions obtained by the proposed SF-PCA, computed on the entire population during REST, and plotted separately for CTRL and SCHZ groups. Top: $median(s_{CTRL}) \cdot$ SF-PC. Bottom: $median(s_{SCHZ}) \cdot$ SF-PC. The figure highlights the median behaviour on each component, for subjects in the two populations.
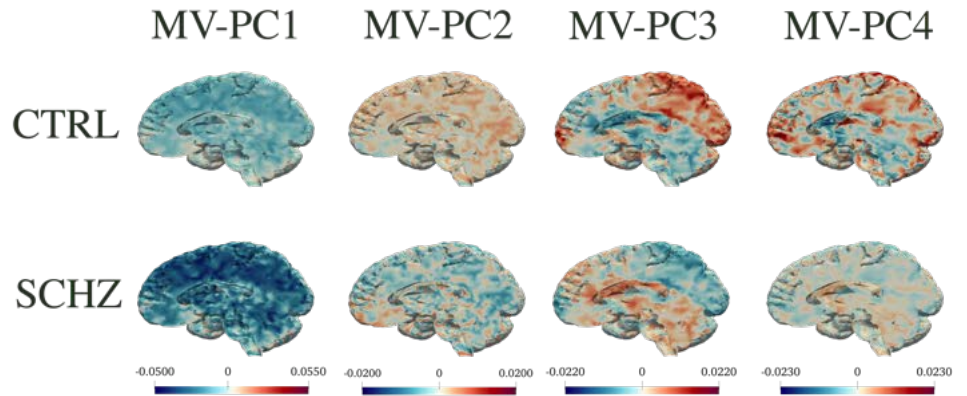

Fig 5. MV-PC for REST. Same as Fig 4 but for standard MV-PCA.
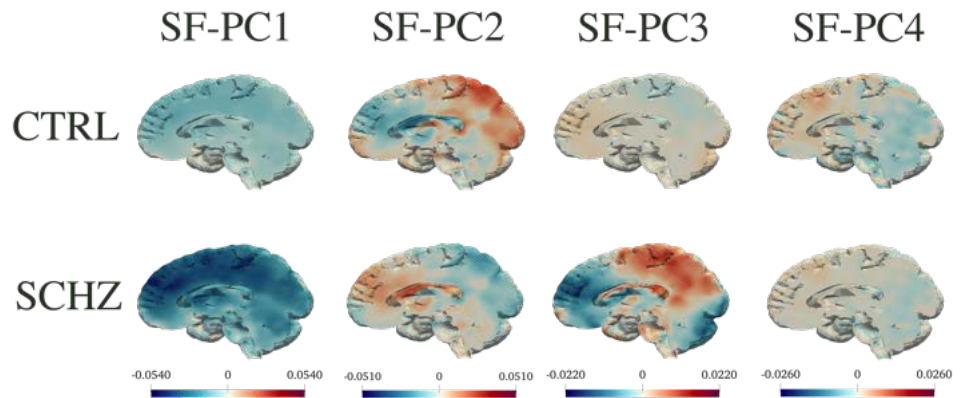

Fig 6. SF-PC functions for SWITCH. Same as Fig 4 but for SWITCH.

We now explore the distributions of the scores for males and females, and across age. For the REST session, the distributions of scores of the 4 computed SF-PCs are not significantly different in males vs females. For the SWITCH session, the distribution of the scores for SF-PC3_SWITCH is different in males vs females, with $p = 9 \cdot 10^{-3}$. For what concerns the Pearson correlation between the PCs scores and the subjects' age, the maximum correlation present in absolute value is $r = 0.33$.

Discussion

Healthy vs Affected by Schizophrenia: mean connectivity maps

Fig 1 depicts the mean maps of the two populations. This qualitative representation fails to detail the patterns characterizing different clinical groups. The only noticeable aspect is that the global functional connectivity for the SCHZ is slightly lower during REST as well as during the SWITCH task, coherently with what highlighted in literature (see, e.g., [50,54]). The compliance with literature suggests that our experimental population, although of limited size, is nevertheless representative of the pathology under study.

Entire Population: common modes of variation

Through SF-PCA we extract $K = 4$ SF-PCs for each session on the entire population, meaning that each subject is associated with 4 SF-PC_REST scores and 4 SF-PC_SWITCH scores. The percentage of explained variance (that is for both sessions approximately 20%) must be evaluated in the light of the present work's aim, which is individuating common patterns in connectivity and evaluating possible differences between the two population subgroups, rather than explaining the most part of variability in the dataset. In fact, a dimension reduction from an initial set of approximately 36 000 covariates to 4 would, inevitably, be ineffective, in terms of explained total variance. Even more so when taking into account the extent of interpersonal differences in neural data, and considering the heterogeneity of manifestation and pathology evolution in schizophrenia.

Healthy vs Affected by Schizophrenia: different expressions of the common modes of variation

The proposed SF-PCs has the same discrimination power as standard MV-PCA (about 63% for REST and 65% for SWITCH), but has the valuable advantage of identifying clearly interpretable patterns. In any case, we point out that our objective is not classification, but rather than the identification of clinically relevant patterns in the main modes of variation of functional connectivity. Discrimination on the PC scores indicates these patterns capture significant differences in the two subpopulations, highlighting aspects related to the pathology.

The SF-PCs appear differently expressed in the two subpopulations, as shown by Fig 4 and 6, which plot the median patterns in two groups. At this point, we assess the neurological interpretation of the SF-PCs, with a special emphasis on those diverging between groups. As it is common in classic MV-PCA, the first SF-PCs for REST and SWITCH, i.e., SF-PC1_REST and SF-PC1_SWITCH, depict a grand mean of the entire population. For what concerns the REST, two SF-PCs, SF-PC2_REST and SF-PC3_REST, present different distributions of the scores between CTRL and SCHZ, with p-values $p = 2 \cdot 10^{-4}$ and $p = 1 \cdot 10^{-3}$, respectively; see Fig 2. SF-PC2_REST shows a distinct pattern associated with well defined regions; see Fig 4. In particular, it involves, on the one hand, the superior parietal gyrus, and, on the other, the cuneus and the medial prefrontal cortex. The first region is part of the Executive Control Network,

while the latter has been associated with the Default Mode Network. SF-PC2_REST may thus suggest a difference in the correlation and uncoupling of such networks among the groups, a phenomenon known to affect schizophrenia patients (see, e.g., [55]). SF-PC3_REST captures additional contrasts among the areas highlighted by SF-PC2_REST, substantiating the importance of the modifications affecting the Default Mode Network in schizophrenia patients. Moreover, a hub of activation in the prefrontal region is highlighted for the SCHZ group. This brain region plays a pivotal role in the cognitive impairment, and in other manifestations of this pathology [56]. As a comparison, we note that standard MV-PCA, although reaching the same classification performances as SF-PCA, it fails to capture clearly identifiable patterns. For instance, while the involvement of the Default Mode Network is partially captured by MV-PC3_REST, no relevant pattern is highlighted by MV-PC2_REST; see Fig 5. Overall, all the extracted MV-PCs are highly noisy, making it difficult to propose interpretations.

For what concerns the modes of variations associated with the SWITCH paradigm, the scores of SF-PC2_SWITCH point out a difference in the SF-PC's expression, with a p-value $p = 6 \cdot 10^{-4}$; see Fig 3. The pattern identified by SF-PC2_SWITCH, depicted in Fig 6, mainly involves the frontal and parietal regions, coherently with previous findings regarding task-switching paradigm (see, e.g., [57]), showing how the mechanisms and connections involved in this activity are impaired by the pathology under study.

It thus emerges that the modes of variations computed through SF-PCA are informative of the pathology in analysis, as they successfully highlight relevant differences between the populations, as also confirmed by the compliance with previous literature.

Gender and age: no significant difference in the expressions of the common modes of variation

It is very interesting to note that the scores of the obtained PCs are not significantly different between males and females ($p > 0.05$), with the exception of SF-PC3_SWITCH ($p = 9 \cdot 10^{-3}$), and are also unrelated to age (max Pearson's correlation in absolute value $r = 0.33$). This supports that the signal captured by SF-PCA is related to the pathology under study, rather than these other covariates. Hence, the unbalance in the considerate population does not hinder the validity of the obtained results. SF-PC3_SWITCH indicates a hub of activation in the prefrontal region, which might be an interesting phenomenon to explore in further studies.

Conclusions

In the present work, we proposed to extend the use of SF-PCA method, previously validated for FC maps computed over the cortical surface, for the analysis of grey matter volume data. The FC maps considered refer to two separate fMRI sessions: one resting-state and one during a task based on the switching paradigm.

The adequacy of the proposed approach is assessed through classification performance and compliance of the extracted features with previous findings. During both sessions, the extracted SF-PCs scores significantly differ in the two subpopulations of CTRL and SCHZ. Linear Discriminant Analysis of the SF-PC scores discriminates the subjects' diagnosis with an above-chance Area Under the Curve. More significantly, the extracted features are coherent with previous literature. In particular, during REST, the SF-PCs highlight the disruption of the physiological alternation between the Default Mode and the Executive Control networks, one of the key mechanisms in schizophrenia's physiopathology. Analogously, SF-PC2_SWITCH identifies the regions known to be

involved in task-switching and impaired by schizophrenia. Unlike the standard MV-PCA, the proposed SF-PCA, succeeds in identifying readable and relevant patterns. We believe that the gained interpretability is a valuable asset for studies in the field.

Given the unbalance of the experimental population, in terms of sex and age, we have verified that the main modes of variations do not significantly differ across age and gender. The only resulting difference between genders, indicated by SF-PC3_SWITCH, does not overlap with phenomena related to the pathology; hence, the population unbalance does not invalidate our results. In clinical applications, it would nevertheless be desirable to have more balanced groups. Moreover, an improvement in the sample size and in data quality and resolution would certainly lead to the individuation of additional patterns and to higher explained variability and discrimination performances. In any case, we stress that the aim of the present work is not to derive clinical evidence about schizophrenia, but rather to propose an innovative data analysis method that can be profitably applied in neuroscience research. In particular, we have shown that SF-PCA is a valuable tool to explore the variability in functional connectivity maps in the complex tridimensional domain of the grey matter. Furthermore, by reducing the dimensionality of data and simultaneously maintaining their interpretability, SF-PCA also addresses some of the computational issues, very relevant in the field. For these reasons, SF-PCA may be especially useful for the study of brain connectivity in connection with neurological and psychiatric illnesses, as it may identify significant differences between subpopulations, and shed new light on the considered pathology. Moreover, the ability of the proposed method to include in the data analysis the volume brain anatomy constitutes a relevant addition with respect to the available literature. In future research, we aim to extend this valuable modeling feature to other data analysis techniques, commonly used in neuroimaging studies, such as independent component analysis and generalized linear models, and develop their anatomically compliant versions. Furthermore, following the approaches described in [58,59] for cortical surface data, we may aim at developing methods capable to handle subject-specific anatomies. This would enable to jointly analyse functional aspects and subject-specific structural changes induced by the considered pathology [60], such as atrophy.

Finally, the present approach could be extended to different ways of computation of FC maps, for instance through nonlinear metrics, or by considering activation maps. In future works, we plan to apply the hereby validated method to further pathologies and to employ the extracted SF-PCs to evaluate the correlation between connectivity and the outcomes of clinical assessment tests. We also aim to approach only pathological samples, to be able to explore in more detail the heterogeneity characterizing psychiatric illnesses. Finally, we plan to assess the proposed approach in clinical settings, to monitor patients enrolled in rehabilitation programs.

Acknowledgments

References

1. Ombao H, Lindquist M, Thompson W, Aston J. Handbook of neuroimaging data analysis. Chapman and Hall/CRC; 2016.

2. Dong D, Wang Y, Chang X, Luo C, Yao D. Dysfunction of large-scale brain networks in schizophrenia: a meta-analysis of resting-state functional connectivity. Schizophrenia bulletin. 2018;44(1):168–181.

3. Martí-Juan G, Sastre-Garriga J, Martinez-Heras E, Vidal-Jordana A, Llufriu S, Groppa S, et al. Using The Virtual Brain to study the relationship between structural and functional connectivity in patients with multiple sclerosis: a multicenter study. Cerebral Cortex. 2023;33(12):7322–7334.

4. Thirion B, Pinel P, Mériaux S, Roche A, Dehaene S, Poline JB. Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. Neuroimage. 2007;35(1):105–120.

5. Lee KJ. Computational issues in using Bayesian hierarchical methods for the spatial modeling of fMRI data. Retrieved from the University of Minnesota Digital Conservancy. 2010;.

6. Cohen JD, Daw N, Engelhardt B, Hasson U, Li K, Niv Y, et al. Computational approaches to fMRI analysis. Nature neuroscience. 2017;20(3):304–313.

7. Smith SM, Hyvärinen A, Varoquaux G, Miller KL, Beckmann CF. Group-PCA for very large fMRI datasets. Neuroimage. 2014;101:738–749.

8. Ramsay JO, Silverman BW. Functional Data Analysis. Springer Series in Statistics. Springer; 2005. Available from: https://books.google.it/books?id=mU3dop5wY_4C.

9. Ramsay JO, Silverman BW. Applied functional data analysis. Springer Series in Statistics. Springer-Verlag, New York; 2002. Available from: https://doi.org/10.1007/b98886.

10. Ferraty F, Vieu P. Nonparametric functional data analysis: theory and practice. vol. 76. Springer; 2006.

11. Kokoszka P, Reimherr M. Introduction to Functional Data Analysis. Chapman & Hall/CRC Texts in Statistical Science. Chapman & Hall/CRC; 2017.

12. Viviani R, Grön G, Spitzer M. Functional principal component analysis of fMRI data. Human brain mapping. 2005;24(2):109–129.

13. Margaritella N, Inácio V, King R. Parameter clustering in Bayesian functional principal component analysis of neuroscientific data. Statistics in Medicine. 2021;40(1):167–184.

14. Chung MK, Robbins SM, Dalton KM, Davidson RJ, Alexander AL, Evans AC. Cortical thickness analysis in autism with heat kernel smoothing. NeuroImage. 2005;25(4):1256–1265.

15. Lila E, Aston JA, Sangalli LM. Smooth principal component analysis over two-dimensional manifolds with an application to neuroimaging. The Annals of Applied Statistics. 2016;10(4):1854–1879.

16. Lila E, Aston JA. Functional random effects modeling of brain shape and connectivity. arXiv preprint arXiv:200906059. 2020;.

17. Lila E, Aston JA, Sangalli LM. Functional data analysis of neuroimaging signals associated with cerebral activity in the brain cortex. In: Functional Statistics and Related Fields. Springer; 2017. p. 169–172.

18. Chung MK, Hanson J, Pollak S. Statistical analysis on brain surfaces. Handbook of Neuroimaging Data Analysis. 2016;233:46–57.

19. Seo S, Chung MK, Vorperian HK. Heat kernel smoothing using Laplace-Beltrami eigenfunctions. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010: 13th International Conference, Beijing, China, September 20-24, 2010, Proceedings, Part III 13. Springer; 2010. p. 505–512.

20. Huang SG, Lyu I, Qiu A, Chung MK. Fast polynomial approximation of heat kernel convolution on manifolds and its application to brain sulcal and gyral graph pattern analysis. IEEE transactions on medical imaging. 2020;39(6):2201–2212.

21. Damoiseaux JS. Effects of aging on functional and structural brain connectivity. Neuroimage. 2017;160:32–40.

22. Rogers BP, Morgan VL, Newton AT, Gore JC. Assessing functional connectivity in the human brain by fMRI. Magnetic resonance imaging. 2007;25(10):1347–1357.

23. Noble S, Scheinost D, Constable RT. A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. Neuroimage. 2019;203:116157.

24. Edison P. Brain Connectivity: Structural and Functional Neuronal Integrity and Its Relationship with Pathological Substrates; 2020.

25. Zhou Y, Liang M, Tian L, Wang K, Hao Y, Liu H, et al. Functional disintegration in paranoid schizophrenia using resting-state fMRI. Schizophrenia research. 2007;97(1-3):194–205.

26. Sheffield JM, Barch DM. Cognition and resting-state functional connectivity in schizophrenia. Neuroscience & Biobehavioral Reviews. 2016;61:108–120.

27. Woodward ND, Rogers B, Heckers S. Functional resting-state networks are differentially affected in schizophrenia. Schizophrenia research. 2011;130(1-3):86–93.

28. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. vol. 112. Springer; 2013.

29. Poldrack RA, Congdon E, Triplett W, Gorgolewski K, Karlsgodt K, Mumford J, et al. A phenome-wide examination of neural and cognitive function. Scientific data. 2016;3(1):1–12.

30. Gorgolewski KJ, Durnez J, Poldrack RA. Preprocessed consortium for neuropsychiatric phenomics dataset. Faculty of 1000 Ltd. 2017; p. 6:1262. doi:https://doi.org/10.12688/f1000research.11964.2.

31. Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL, et al. Unbiased average age-appropriate atlases for pediatric studies. NeuroImage. 2011;54(1):313–327.

32. Fonov VS, Evans AC, McKinstry RC, Almli CR, Collins D. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage. 2009;47(1):S39–S41.

33. Esteban O, Markiewicz CJ, Blair RW, Moodie CA, Isik AI, Erramuzpe A, et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nature methods. 2019;16(1):111–116.

34. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. IEEE transactions on medical imaging. 2010;29(6):1310–1320.

35. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical image analysis. 2008;12(1):26–41.

36. Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis: I. Segmentation and surface reconstruction. Neuroimage. 1999;9(2):179–194.

37. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage. 2002;17(2):825–841.

38. MATLAB. version 9.10 (R2010a). Natick, Massachusetts: The MathWorks Inc.; 2021.

39. R Core Team. R: A Language and Environment for Statistical Computing; 2022. Available from: https://www.R-project.org/.

40. Tran AP, Yan S, Fang Q. Improving model-based functional near-infrared spectroscopy analysis using mesh-based anatomical and light-transport models. Neurophotonics. 2020;7(1):015008–015008.

41. Fang Q. Brain2mesh: a one-liner for 3D brain mesh generation; 2022. Available from: https://github.com/fangq/brain2mesh.

42. Hammers A, Allom R, Koepp MJ, Free SL, Myers R, Lemieux L, et al. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Human brain mapping. 2003;19(4):224–247.

43. Gousias IS, Rueckert D, Heckemann RA, Dyet LE, Boardman JP, Edwards AD, et al. Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest. Neuroimage. 2008;40(2):672–684.

44. Uddin LQ, Clare Kelly A, Biswal BB, Xavier Castellanos F, Milham MP. Functional connectivity of default mode network components: correlation, anticorrelation, and causality. Human brain mapping. 2009;30(2):625–637.

45. Joel SE, Caffo BS, Van Zijl PC, Pekar JJ. On the relationship between seed-based and ICA-based measures of functional connectivity. Magnetic resonance in medicine. 2011;66(3):644–657.

46. Shukla DK, Wijtenburg SA, Chen H, Chiappelli JJ, Kochunov P, Hong LE, et al. Anterior cingulate glutamate and GABA associations on functional connectivity in schizophrenia. Schizophrenia bulletin. 2019;45(3):647–658.

47. Bersani FS, Minichino A, Fojanesi M, Gallo M, Maglio G, Valeriani G, et al. Cingulate Cortex in Schizophrenia: its relation with negative symptoms and psychotic onset. A review study. Eur Rev Med Pharmacol Sci. 2014;18(22):3354–3367.

48. Schultz CC, Koch K, Wagner G, Nenadic I, Schachtzabel C, Güllmar D, et al. Reduced anterior cingulate cognitive activation is associated with prefrontal–temporal cortical thinning in schizophrenia. Biological psychiatry. 2012;71(2):146–153.

49. Tamminga CA, Vogel M, Gao X, Lahti AC, Holcomb HH. The limbic cortex in schizophrenia: focus on the anterior cingulate. Brain research Brain research reviews. 2000;31(2-3):364–370.

50. Baniqued PL, Low KA, Fletcher MA, Gratton G, Fabiani M. Shedding light on gray (ing) areas: Connectivity and task switching dynamics in aging. Psychophysiology. 2018;55(3):e12818.

51. Aron AR, Monsell S, Sahakian BJ, Robbins TW. A componential analysis of task-switching deficits associated with lesions of left and right frontal cortex. Brain. 2004;127(7):1561–1573.

52. Jolliffe IT. Principal component analysis. Springer; 2002.

53. Arnone E, Sangalli LM, Lila E, Ramsay J, Formaggia L, de Falco C. fdaPDE: Functional Data Analysis and Partial Differential Equations (PDE); Statistical Analysis of Functional and Spatial Data, Based on Regression with PDE Regularization; 2022.

54. Li H, Ou Y, Liu F, Chen J, Zhao J, Guo W, et al. Reduced connectivity in anterior cingulate cortex as an early predictor for treatment response in drug-naive, first-episode schizophrenia: a global-brain functional connectivity analysis. Schizophrenia research. 2020;215:337–343.

55. Hu ML, Zong XF, Mann JJ, Zheng JJ, Liao YH, Li ZC, et al. A review of the functional and anatomical default mode network in schizophrenia. Neuroscience bulletin. 2017;33(1):73–84.

56. Sakurai T, Gamo NJ, Hikida T, Kim SH, Murai T, Tomoda T, et al. Converging models of schizophrenia–Network alterations of prefrontal cortex underlying cognitive impairments. Progress in neurobiology. 2015;134:178–201.

57. Dove A, Pollmann S, Schubert T, Wiggins CJ, Von Cramon DY. Prefrontal cortex activation in task switching: an event-related fMRI study. Cognitive brain research. 2000;9(1):103–109.

58. Lila E, Aston JA. Statistical analysis of functions on surfaces, with an application to medical imaging. Journal of the American Statistical Association. 2020;115(531):1420–1434.

59. Lila E, Aston JA. Functional random effects modeling of brain shape and connectivity. The Annals of Applied Statistics. 2022;16(4):2122–2144.

60. Good CD, Johnsrude IS, Ashburner J, Henson RN, Friston KJ, Frackowiak RS. A voxel-based morphometric study of ageing in 465 normal adult human brains. Neuroimage. 2001;14(1):21–36.

MOX Technical Reports, last issues

Dipartimento di Matematica

Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

106/2023 Fontana, N.; Savaré, L.; Ieva, F.
Integrating state-sequence analysis to uncover dynamic drug-utilization patterns to profile heart failure patients

108/2023 Arnone, E.; Negri, L.; Panzica, F.; Sangalli, L.M.
Analyzing data in complicated 3D domains: smoothing, semiparametric regression and functional principal component analysis

105/2023 Cicci, L.; Fresca, S.; Guo, M.; Manzoni, A.; Zunino, P.
Uncertainty quantification for nonlinear solid mechanics using reduced order models with Gaussian process regression

103/2023 Dimola N.; Kuchta M.; Mardal K.A.; Zunino P.
Robust Preconditioning of Mixed-Dimensional PDEs on 3d-1d domains coupled with Lagrange Multipliers

104/2023 Possenti, L.; Gallo, A.; Vitullo, P.; Cicchetti, A.; Rancati, T.; Costantino, M.L.; Zunino, P.
A computational model of the tumor microenvironment applied to fractionated radiotherapy

101/2023 Formaggia, L.; Zunino, P.
Hybrid dimensional models for blood flow and mass transport

100/2023 Vitullo, P.; Cicci, L.; Possenti, L.; Coclite, A.; Costantino, M.L.; Zunino, P.
Sensitivity analysis of a multi-physics model for the vascular microenvironment

98/2023 Lespagnol, F.; Grandmont, C.; Zunino, P.; Fernandez, M.A.
A mixed-dimensional formulation for the simulation of slender structures immersed in an incompressible flow

96/2023 Bonetti, S.; Botti, M.; Antonietti, P.F.
Robust discontinuous Galerkin-based scheme for the fully-coupled non-linear thermo-hydro-mechanical problem

95/2023 Barnafi, N. A.; Regazzoni, F.; Riccobelli, D.
Reconstructing relaxed configurations in elastic bodies: mathematical formulation and numerical methods for cardiac modeling