Noname manuscript No.
(will be inserted by the editor)

Topological Features of Electroencephalography are Robust to Re-referencing and Preprocessing

Jacob Billings* · **Ruxandra Tivadar*** · **Micah M. Murray** · **Benedetta Franceschiello** · **Giovanni Petri**

Received: date / Accepted: date

Abstract Electroencephalography (EEG) is among the most widely diffused, inexpensive, and adopted neuroimaging techniques. Nonetheless, EEG requires measurements against a reference site(s), which is typically chosen by the experimenter, and specific pre-processing steps precede analyses. It is therefore valuable to obtain quantities that are minimally affected by reference and pre-processing choices. Here, we show that the topological structure of embedding spaces, constructed either from multi-channel EEG timeseries or from their temporal structure, are subject-specific and robust to re-referencing and pre-processing pipelines. By contrast, the shape of correlation spaces, that is, discrete spaces where each point represents an electrode and the distance between them that is in turn related to the correlation between the respective timeseries, was neither significantly subject-specific nor robust to changes of reference. Our results suggest that the shape of spaces describing the observed configurations of EEG signals holds information about the individual specificity of the underlying individual's brain dynamics, and that

J. C. W. Billings
ISI Foundation,
Torino, Italy.
Department of Complex Systems,
Institute for Computer Science,
Czech Academy of Science,
Prague, Czechia.
ORCDI: 0000-0002-8186-6126

R. I. Tivadar
Laboratory for Investigative Neurophysiology,
Department of Radiology,
Lausanne University Hospital and University of Lausanne (CHUV-UNIL),
Lausanne, Switzerland
Department of Ophthalmology,
Fondation Asile des aveugles and University of Lausanne,
Lausanne, Switzerland
Cognitive Computational Neuroscience Group,
Institute for Computer Science,
University of Bern, Switzerland
ORCID: 0000-0002-6680-2917

M. M. Murray
Laboratory for Investigative Neurophysiology,
Department of Radiology,
Lausanne University Hospital and University of Lausanne (CHUV-UNIL),
Lausanne, Switzerland
Department of Ophthalmology,
Fondation Asile des aveugles and University of Lausanne,
Lausanne, Switzerland
EEG CHUV-UNIL Section,
CIBM Center for Biomedical Imaging,
Lausanne, Switzerland
Department of Hearing and Speech Sciences,
Vanderbilt University,
Nashville, TN, USA
ORCID: 0000-0002-7821-117X

B. Franceschiello
EEG CHUV-UNIL Section,
CIBM Center for Biomedical Imaging,
Lausanne, Switzerland
Laboratory for Investigative Neurophysiology,
Department of Radiology,
Lausanne University Hospital and University of Lausanne (CHUV-UNIL),
Lausanne, Switzerland
Department of Ophthalmology,
Fondation Asile des aveugles and University of Lausanne,
Lausanne, Switzerland
ORCID: 0000-0003-0754-5081

G. Petri
ISI Foundation,
Torino, Italy

ISI Global Science Foundation,
New York, NY, USA
ORCID: 0000-0003-1847-5031
E-mail: giovanni.petri@isi.it

temporal correlations constrain to a large degree the set of possible dynamics. In turn, these encode the differences between subjects' space of resting state EEG signals. Finally, our results and proposed methodology provide tools to explore the individual topographical landscapes and how they are explored dynamically. We propose therefore to augment conventional topographic analyses with an additional – topological – level of analysis, and to consider them jointly. More generally, these results provide a roadmap for the incorporation of topological analyses within EEG pipelines.

Keywords: Resting-state Electroencephalography, Topography, Topology, Network, Computational Modelling, Reference Electrode

1 Introduction

Electroencephalography (EEG) is a non-invasive neuroimaging technique measuring the electrical activity of the brain at the scalp (Biasiucci et al., 2019). EEG has several practical strengths as a neuroimaging tool (Michel Christoph M., 2009; Michel and Murray, 2012; Michel et al., 2004; Murray et al., 2008): it is temporally precise, cost-effective, easy to use, portable and compatible with other techniques, such as MRI and PET. Indeed, these strengths made EEG a primary tool for studying brain activity both from the clinical and research standpoints (Lepage et al., 2014). EEG primarily measures postsynaptic potentials of pyramidal neurons. The neurotransmitter release generated by excitatory or inhibitory action potentials results in local currents at the apical dendrites of the post-synaptic neuron, that in turn lead to current sources and sinks in the extracellular space. In biophysical terms, voltages refer to the exertion needed to move charge from one site to another. More practically, this means that voltage is the charge differential between a "chosen" electrode and a "reference" electrode (Tivadar et al., 2019; Biasiucci et al., 2019). EEG is the measurement of this voltage as it varies in time, and thus results in "time series" across different sites on the scalp. Two issues arise from the biophysical underpinnings of the EEG signal. First, the brain signals recorded at scalp level is given by the synchronous activity of multiple neurons that volume conduct. Therefore, a given electrode not only captures brain activity from directly beneath it, but to a certain extent from the entire brain. Second, measurements of voltage are referential, meaning that EEG time series (including event-related potentials (ERPs)) at a given electrode or scalp site will change when the reference changes, as there is no electrically neutral spot on the scalp or body surface. This has led to a long-standing debate in the EEG community, discussing which of the references is more informative for the analyses (Chella et al., 2016; Yao et al., 2019; Hu et al., 2019).

This issue concerns spontaneous data as well as pre-processed and post-processed averages, and functional connectivity data. Thus, referencing affects both temporal and spatial aspects of the recorded potentials (Chella et al., 2016). With regards to spatial aspects, the values of the electric field at the scalp will change when the reference changes, as a different baseline value (i.e. the voltage of the reference electrode) is being compared to every other electrode.

In terms of temporal aspects, a non-neutral reference introduces time-varying activity into the recordings of all electrodes, meaning that both the temporal waveforms, as well as their spectral properties suffer from distortions (Chella et al., 2016). Therefore, the variance around a mean voltage value (e.g., spectral power, amplitude, etc.), as well as other derived and associated measures, including results of statistical contrasts, will change when the reference changes. These facts have generally led to –and to some extent continue to result in– misinterpretation and misuse of EEG data, despite good quality in experimental design (Michel and Murray, 2012; Biasiucci et al., 2019; Tivadar et al., 2019).

To solve the reference issue, many in the EEG community have turned towards the characterization and analysis of properties of the electrical field at the scalp, such as topographical maps and spatial pattern analysis methods, as well as source localisation techniques (MWong, 2012; Michel and Murray, 2012; Michel et al., 2004; Grave de Peralta Menendez et al., 2000; Lehmann and Michel, 2011; Michel et al., 2001; Tenke and Kayser, 2005; Marinazzo et al., 2019). Treating the data from the entire electrode montage as a multivariate vector has several advantages over waveform-based analysis of voltage. First, the shape of the electrical field at the scalp will not change with a changing reference (Tivadar et al., 2019). Second, multivariate analyses also profit from the added information of high-density recordings. They can disentangle effects of strength from effects due to changes in sources' configuration or signal latency (Murray et al., 2008). Third, topographic information has direct neurophysiologic interpretability (Michel and Murray, 2012), as biophysical laws dictate that differences in topography indicate changes in the configuration of active cerebral sources (Vaughan, 1982; Lehmann, 1987). Nevertheless, traditional waveform-based analyses still predominate in the EEG community (Luck, 2014).

Here, we propose a new method of description of the EEG signal, which is robust across different pre-processing and reference choices. We first build representations of EEG data based on different types of signal embeddings. We then assess their robustness and discriminatory power, e.g. between different subjects and tasks, using recent topological data analysis tools. These tools have been shown to be useful in the analysis of neurophysiological data (Petri et al., 2014, 2013; Ibáñez-Marcelo et al., 2019a; Bassett and Sporns, 2017; Giusti et al., 2016; Varley et al., 2020; Billings et al.,

2021) because they are built to detect properties of datasets, e.g. point clouds or weighted networks, that are invariant under homeomorphic transformations, which include, deformations, rotations, contractions and any other continuous transformation Zomorodian and Carlsson (2005); Ghrist (2008).. The rationale for this ability is that they capture and quantify topology, that is, the shape of spaces in arbitrary dimensions, including discrete spaces obtained from signals, via their topological features, e.g. connected components, 1-dimensional holes, three-dimensional cavities, and the higher-dimensional analogues.

In this way, topological descriptions of EEG data have been shown to provide meaningful simplifications of high-dimensional data, by extracting low-dimensional summaries of the dataset's shape (Giusti et al., 2015), to capture mesoscale patterns of disconnectivity (Petri et al., 2014; Lee et al., 2012) and to explicitly encode interactions among many elements (i.e. nodes in a network, regions of the brain, etc.) (Battiston et al., 2020; Iacopini et al., 2019).

We find that embeddings constructed from multi-channel EEG timeseries, and from their temporal structure, are specific to subjects and robust to re-referencing and pre-processing pipelines. By contrast , spaces obtained from spatial correlations among electrodes, analogous to traditional functional connectivity Sporns (2013); Hutchison et al. (2013); Betzel et al. (2014), were weakly characteristic of subjects and lacked robustness to changes of reference. Our results highlight the utility of the proposed tools to explore individual topographical landscapes, and to observe how these landscapes dynamically change across time.

2 Methods

2.1 Overview of analysis pipeline

In the following sections, we summarise the pipeline we adopt to analyse EEG data and compare the results across multiple references and pre-processing choices. We start from raw EEG signals recorded on the scalp in a cohort of 21 subjects (Fig. 1a, see section 2.2). The recorded signals are then prepared using three different pre-processing pipelines (Fig 1b, see Section 2.3). For each pipeline, we consider different EEG references, which can result in signals with different waveforms and different relations to one another (Fig. 1c): for illustration, we show three EEG signal snippets for various references. For all pre-processing pipelines, subjects and references we compute three different representations: functional connectivity (Sec. 2.4), the Takens embedding (Sec. 2.5) and the direct temporal embedding (Sec. 2.6) (Fig. 1d). For all these representations, we extract summaries of their low-dimensional topological properties (Sec. 2.7) and use them to compute distances between them (Fig. 1e-f). Finally, we compare the results obtained from the previous analysis with a two synthetic benchmarks: one obtained by simulating EEG signals (Sec 2.8) and the other obtained by temporally reshuffling the EEG data (Sec. 2.9). The following sections provide details on the analysis steps described above.

2.2 Subjects

We tested twenty-one right-handed participants (18 male, 3 female, age range 21-39, mean age $\pm$ standard deviation: 25.76 ± 4.54 years). No participant had a history of or current neurological or psychiatric illness, according to self-report. Data from one participant was excluded due to excessive EEG artifacts, thus leaving 20 participants in the final sample (17 male, 3 female; aged 21-39). All participants provided written, informed consent to procedures approved by the cantonal ethics committee (CER-VD, Switzerland).

2.3 Recording Procedure and Pre-processing Pipelines

Participants sat in a sound-attenuated darkened room (WhisperRoom MDL 102126E), and were first tested using a multisensory paradigm. Event-related potentials from this dataset have already been published in (Tivadar et al., 2018). After the experimental paradigm, participants were asked whether they would like a break before the resting-state recording was initiated. Participants were then asked to close their eyes and instructed not to engage in any specific physical or mental activity for 3 minutes. Continuous EEG was recorded at 1024Hz with a 128-channel BioSemi ActiveTwo AD-box (www.biosemi.com). No online filters were used. Online references were the typical BioSemi CMS and DRL electrodes, which form a feedback loop that drives the average potential of the subject as close as possible to the amplifier "zero". Data were offline re-referenced to at least three different references at different pre-processing steps ($filtered$, $clean$, and $cleanint$, described in more detail later, Figure 1b). We chose those electrodes on the $N = 128$ BioSemi cap that were closest to and most representative of the typical classical external electrodes used for referencing (Chella et al., 2016). Specifically, we used the average reference as well as C17, A23 and D24/B14 as representative of nose, inion, and linked-mastoids/earlobes references, respectively. Prior to cleaning, a 2nd order Butterworth filter (-12dB/octave roll-off; 0.1Hz high-pass; 60Hz low-pass; 50Hz notch) was applied, which was computed linearly in both forward and backward directions to remove phase shifts. Thus, by filtering, any activity lower than 0.1Hz and higher than 60Hz was removed, together with 50Hz activity which is typical of electrical noise (i.e. power line noise). These filtered data (denoted $filtered$ in the following) constitute the first pre-processing pipeline we will consider. Next, we further pre-processed the filtered

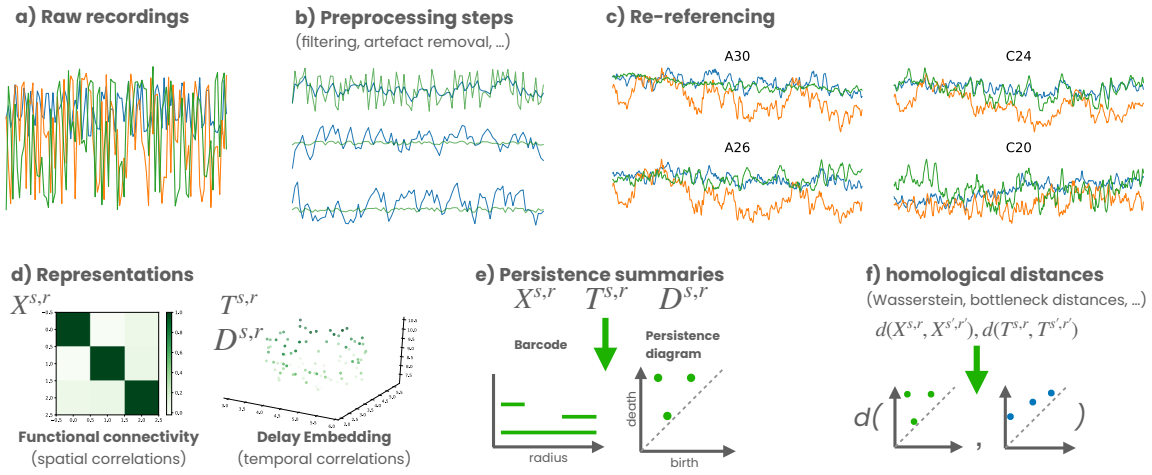

Fig. 1 Overview of analysis pipeline. The standard pipeline of EEG analysis follows these steps: a) raw signals are recorded from scalp EEG electrodes. b) the signals are filtered in order to remove noisy or uninteresting frequency bands (here, any activity <0.1Hz and >60Hz, as well as 50Hz electrical line noise) ($filtered$); recorded signals are then cleaned to remove artefacts (i.e. blinks, eye movements, muscle artefacts, heart rate artefacts, electrode pops - i.e. single or multiple sharp waveforms that appear after a sudden change in impedance, electrode drifts due to sweat,etc.) ($clean$); interpolated to account for technical issues (i.e. "dead" electrodes), electrode drifts due to sweat or electrode bridging (when electrolyte gel spreads between adjacent electrodes), etc. ($cleanint$). c) pre-processed data are then referenced to one of the electrodes or, in some case to the average value across all channels. Different reference choices can result in different effects: for illustration, we show here three intervals of EEG signals for four different references. Note how the relation among series can change depending on the choice of the reference. d) In this study we investigate three data representations: 1) X is obtained by considering the Pearson correlations among channels and results in a description of spatial correlations, 2) T is the Takens (or delay) embedding starting from the multivariate EEG timeseries and explicitly encoding temporal correlations within the signals, 3) D is a variant of T wherein the EEG timeseries is directly embedded, that is, without the imposition of time-delay vectors, effectively equivalent to considering the brain configuration space. e) We analyse the three types of embeddings using persistent homology, which quantitatively captures the shape of generic spaces in the form of barcodes or persistence diagrams. f) Finally, we can associate a distance between spaces by measuring distances between persistence diagrams themselves.

data, by cleaning them ($clean$ dataset). Data quality was thus controlled first via visual artifact detection and then via ICA decomposition in Matlab (R2020a) using the EEGLAB toolbox (Delorme and Makeig, 2004), in order to exclude any remaining transient noise, muscle artefacts, heart beat artefacts and lateral eye-movements or blinks. These data were also referenced to three references, excluding the average reference. This constitutes the second pre-processing pipeline. Finally, for the third pipeline ($cleanint$), we further inspected data from artefact-contaminated electrodes. These electrodes were interpolated using 3D spherical splines (Perrin et al., 1989), which take into account all of the electrode sites. We then re-referenced our dataset to all the four references specified above, including the average reference. To summarize, the $filtered$ and $clean$ data were only referenced to C17, A23, D24B14, while the $cleanint$ data was referenced to the previously named electrodes, and additionally to the average reference. Lastly, all timeseries were down-sampled to 64 samples per second to reduce computational load and to reflect the absence of spectral content >60 Hz.

2.4 Functional connectivity metric embedding construction

We compute functional connectivity networks using Pearson correlations. More precisely, for each combination of subject s and reference r, we compute the correlation matrix $\mathbf{C}^{s,r}$, in which entry $c_{ij}^{s,r}$ corresponds to the Pearson correlation between the timeseries of channels i and j with respect to a subject s and reference r. To each subject-reference pair (s,r) we associate to $\mathbf{C}^{s,r}$ a discrete metric space $X^{s,r}$, obtained by mapping the timeseries corresponding to each channel i to a point $p_i \in X^{s,r}$. Distances between points are given by $d_C^{s,r}(p_i,p_j) = 1 - c_{ij}^{s,r}$. This defines a metric space $(X^{s,r}, d_C^{s,r})$ for each subject-reference pair (s,r).

2.5 Delay embedding construction

Despite the presence of unmodeled noise when re-referencing EEG potentials, the EEG signal samples from the brain's dynamical state space. To reconstruct the underlying dynamical system, we compute the Takens embedding of each volunteer's re-referenced multichannel EEG recordings. As before, for each combination of subject s and reference r, we compute the Takens embedding $T^{s,r}$ as follows: for a single time series $x(t)$ we build a d-dimensional point cloud defined as $\{x(t_0), x(t_0+\tau), \ldots, x(t_0+(d-1)\tau)\}$ for all t_0 in the

time series, and where τ is a delay and d the embedding dimension. Standard techniques are adopted to choose the pair (τ, d). Distances between points in Takens embeddings are computed using the canonical Euclidean distance, d_E, as prescribed by Takens' embedding theorem (Noakes, 1991). It is possible to generalize the embedding to the case of I time series, which in turn results in a $d \times I$ dimensional embedding. We therefore consider the metric spaces $(T^{s,r}, d_E^{s,r})$. (Myers et al., 2019).

2.6 Direct embedding construction

Following Takens' finding that the delay embedding characterizes a system's dynamical state space, Deyle and Sugihara (2011) found similar properties when directly embedding the temporal evolution of the I dimensional vectors from multichannel recordings (Deyle and Sugihara, 2011). As before, for each combination of subject s and reference r, we compute the direct multichannel embedding $D^{s,r}$ as follows: for a multichannel time series $\boldsymbol{x}(t, i)$ we build a I-dimensional point cloud defined as $\{x(t_0, i_0), x(t_0, i_1), \ldots, x(t_0, i_I)\}$ for all t_0 in the time series, and where I is the number of recording channels. Here again, distances between points are computed using the canonical Euclidean distance, $d_E^{s,r}$. Thus we develop the metric spaces of the direct embedding $(D^{s,r}, d_E^{s,r})$.

The benefit of the direct embedding comes from its interpretability. Each embedded point corresponds to the vector of signals from all electrodes at a specific time point. The embedding therefore corresponds to the configuration space obtained from all instantaneous EEG topographies.

One downside of the direct embedding is that the number of embedded points increases by a factor of N/τ where N is the total number of time points in the recording. In order to decrease computation time, we utilize the sparsification method of Nicholas J. Cavanna et al. (2015) to generated sparse distance matrices having 10% of the original number of edges ($\epsilon = 0.3$).

2.7 Topological distances between spaces

We compute distances between spaces corresponding to the various embeddings using persistent homology (Edelsbrunner and Harer, 2008; Ibáñez-Marcelo et al., 2019b). More precisely, we perform standard persistent homology analysis on the Rips-Vietoris filtrations defined over the points in $X^{s,r}$, $T^{s,r}$, or $D^{s,r}$. Persistent homology works by studying the evolution of topological features (connected components, 1-dimensional cycles, 3d-cavities, etc.) along a series of progressively finer simplicial complex approximations. A simplicial complex can be intuitively imagined as a higher-dimensional version of a graph, that in addition to edges (that is, pairs of points or vertices, called 1-simplices) also allows for other elementary bricks composed by groups of $k+1$ points, called k-simplices ($k \geq 2$). In our case, however, we need a way to go from metric spaces to these simplicial complex approximations. We do this using the Rips-Vietoris construction. It works as follows: given a set of points $\{p_0, p_1, \ldots, p_n\}$ in a metric space M and an arbitrary radius r, for each point p_i we consider its neighbourhood $\mathbf{B}(p_i, r)$ of radius r; we define simplices in the Rips-Vietoris complex $RV(M, r)$ at distance r as follows: whenever $\mathbf{B}(p_i, r) \cap \mathbf{B}(p_j, r) \neq \emptyset$ for some i, j we add the 1-simplex $[p_i, p_j]$; whenever three points p_i, p_j, p_k all have non-empty pairwise intersections we add the 2-simplex $[p_i, p_j, p_k]$, and so on for higher dimensions. The collection of all these simplices constitutes $RV(M, r)$.

The choice of r is of course problematic, as it requires picking a scale for the simplicial complex reconstruction. Persistent homology inverts the problem by scanning the properties of $RV(M, r)$ as a function of r. The ordered collection of $\{RV(M, r)\}_r$ is called a filtration of M (Figure 2a). The outputs of persistent homology are barcodes (and equivalently, persistence diagrams). These compressed summaries recapitulate the homological features of a space, describing how long certain topological features persist along the filtration (e.g. connected component, 1-dimensional holes, etc.) (Fig. 2b). Each bar corresponds to a specific topological feature and its appearance r_b and death r_d radii correspond to the radii at which that feature first appears and disappears, respectively. Persistence diagrams provide an equivalent description: each topological feature is represented in the 2-dimensional plot by a point with coordinates (r_b, r_d). We adopt the persistence diagram description, because it makes it easier to compute distances between them, and use those distances as a measure of similarity between the corresponding spaces (Edelsbrunner and Harer, 2008).

In $X^{s,r}$ spaces we use the Pearson distance as distance between points to construct the filtration(Fulekar, 2009). In $T^{s,r}$ and $D^{s,r}$ spaces we instead adopt the Euclidean distance.

Here, for computational reasons, we focus on the first two homological groups: H_0, that describes connected components, and H_1, that describes one-dimensional holes. Persistence diagrams are equipped with a metric themselves. We can therefore measure distances between them and use this homology-based distance as a topological distance between spaces (Reininghaus et al., 2015). We choose persistent homology as a descriptor for our study because it allows us to compare spaces with different numbers of points, dimensions and metric structure. More precisely, we define the homological distance between $\mathbf{C}^{s,r}$ and $\mathbf{C}^{s',r'}$ to be the sliced Wasserstein distance between the persistence diagrams corresponding to $X^{s,r}$ and $X^{s',r'}$. Similarly, we define the homological distance between Takens embeddings $T^{s,r}$ and $T^{s',r'}$

to be the Wasserstein distance between the corresponding persistence diagrams. The same applies for computing distances between multichannel embeddings $D^{s,r}$ and $D^{s',r'}$. For completeness, in the following we report the results for H_0 and H_1, although in this study the results from both dimensions typically agree.

2.8 Simulated EEG data

To demonstrate the performance of the present study's methods against a known benchmark, we also consider a synthetic dataset, (s^v, r^v), generated from the brain activity simulation software, The Virtual Brain (*tvb*) (version 2.2) (Leon et al., 2013; Schirner et al., 2021). Specifically, we modeled local neuronal dynamics as a generic 2-dimensional oscillator. Large-scale structural connectivity was taken from the default connectivity dataset having 76 nodes ('connectivity_76.zip'). The speed of signal propagation was set to $4ms^{-1}$. Coupling between nodes was set to the recommended linear function with a slope of 0.014. The integration scheme was specified as 'HeunStochastic', with integration rate set to 1024 Hz. Cortical activity was simulated across a surface mesh having 16384 vertices ('cortex_16384.zip'). Local cortical connectivity was specified using the prebuilt default parameters in the file 'local_connectivity_16384.mat'. The coupling strength between cortical vertices was set to a recommended global value of 2^{-10}. Simulated EEG data was monitored at 65 sites ('regionMapping_16k_76.txt'). EEG Data were simulated at a rate of 64 Hz for 4.27 minutes. Data from 5 different references was generated by setting aside 5 EEG channels, chosen at random, to be used as baselines; and then subtracting each of the remaining 60 channel data from each baseline, in turn. Thus, each simulated EEG dataset was of size [5 references x 60 channels x 16384 samples].

Using the above pipeline as a template, 8 'virtual volunteers' were modeled by adjusting the structural connectivity weight matrix, and by assigning a different level of additive noise to the integration scheme. Specifically, each edge of the connectivity matrix was multiplied by a random factor in [0.75, 1.25]. The default connectivity weights ranged between [0,3], with most weights assigned a value of 0. For each virtual volunteer, Gaussian noise having unit variance and a standard deviation between [0, 0.01] was added at each integration step.

2.9 Comparison with temporally reshuffled null model

We also consider an auxiliary randomized version of the (s, r) datasets: given each dataset

$$\mathbf{x}(t) = \{x(t_0), x(t_1), \ldots, x(t_T)\} \in \mathbb{R}^N,$$

we keep the vectors the same, but we reshuffle the temporal labeling. In this way, we preserve the statistical properties of the signal, but destroy the temporal correlations within it. In particular, this type of randomization conserves exactly the network of spatial correlations between electrodes. That is, those captured by the correlation networks $\mathbf{C}^{s,r}$. Similarly, the direct embedding is unaltered by reshuffling the temporal labels. For this reason, we only re-analyse the Takens embeddings constructed from the randomized data, T^*. The utility of the randomised model is that it provides a benchmark for the effect of reshuffling in addition to the observed inter-subject similarity. It also allow us to tease apart the role of the set of realised brain topographies (well described by $D^{s,r}$) from that of their temporal order.

3 Results

3.1 Effects of re-referencing on topology of functional representations.

We quantified the effects of changing the EEG reference on three different representations: functional connectivity, computed from spatial correlations between electrodes; direct embedding of the brain activations, representing the space of sampled configurations; and the dynamical landscape of brain activity, as reconstructed from temporal embeddings of the signals. More precisely, the first describes how different regions of the brain co-activate and it corresponds to how neuroimaging signals are often studied, both with fMRI (Bassett and Sporns, 2017; Petri et al., 2014) and EEG (Sakkalis, 2011; Ibáñez-Marcelo et al., 2019b). The direct embeddings instead cast EEG signals as points of a high-dimensional static point cloud, which effectively defines the space of realized activations (Donato et al., 2016). Finally, the Takens embeddings are used to reconstruct the structure of the dynamical attractors of dynamical systems and therefore capture the temporal properties of the system (Myers et al., 2019). We analysed data from $n = 20$ subjects, each re-referenced to $R = 4$ different references, using $filtered$, $cleaned$ and $cleaned\,interpolated$ data. For each pair (s, r) we computed the corresponding correlation $X^{s,r}$, Takens $T^{s,r}$ and direct $D^{s,r}$ embedding spaces. We then computed their persistence diagrams as described in Methods and measured the Wasserstein distances between them. In particular, we were interested in quantifying the changes induced by re-referencing data from the same subject. Figure 3a) shows the distances between the X spaces computed between all subjects and references, for H_0 (left) and H_1 (right). Figures 3 parts b) and c) show the same for T and D spaces, respectively.

Data calculated against different references and belonging to the same subject are grouped together in the heatmaps (increasing distances go from white to blue). Thus, diago-

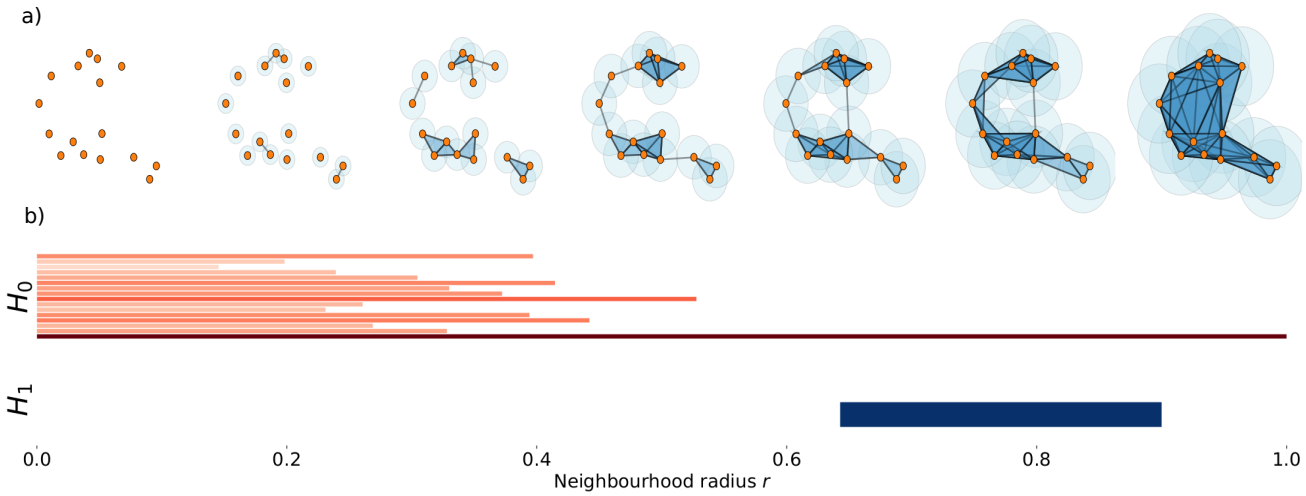

Fig. 2 Sketch of persistent homology computation. a) an example of a filtration of a data cloud in two-dimensions. As r increases the neighbourhoods become larger, they begin to overlap. When the neighbourhoods of two points overlap, an edge (1-simplex) is added. When three overlap in pairs, a full triangle (2-simplex) is added, and so on for higher-dimensional simplices. Therefore, as r increases, more and more simplices appear, making the Rips-Vietoris complex progressively denser and providing a sequence of simplicial approximations of the underlying topological space across radii. Persistent homology focuses on describing the shape (topological features) of the complexes in this sequence. For example, at the beginning ($r = 0$), the points all belong to components disconnected from each other. As r grows, the components begin to merge until only one component remains, which contains all points. Around $r \sim 0.65$, a 1-dimensional cycle appears in the simplicial complex and persists until around $r \sim 1$. A similar analysis can be performed for cavities (homological cycles) of arbitrary dimensions (three-dimensional holes bounded by triangles, four-dimensional cavities bounded by tetrahedra, etc.

b) The barcodes describing the lifetime of the various connected components (red bars), progressively merging into each other until only one survives (describing H_0), and the lifetime of the single 1-dimensional cycle described above (blue bar, describing H_1). Barcodes provide a summary of the topological properties of a space and can be used to compare them in a formal way. We show here the barcode persistence because it makes it easier to relate to the filtration. However, they are equivalent to persistence diagrams, which in turn are more amenable to compute (Wasserstein) distances between spaces.

nal blocks of short distances (see definition below) suggest that re-referencing induces mild changes in the topological structure of the spaces under study. It is important to specify what is meant by short distances. We choose here to use as the benchmark for these effects the distance between spaces corresponding to different individuals, but the same reference site (that is, $s \neq s'$ and $r = r'$). This choice is predicated on the finding that the brain waves of each individual are a unique biometric signature for that individual (Poulos et al., 1999; Marcel and R. Millan, 2007; Chan et al., 2018). The boxplots in Figure 3 show this in quantitative form: for each subject (dark coloured bars) we plot the distribution of distances between the persistence diagrams of spaces obtained from the different referencing. In addition, we plot (in lighter color) the distribution of distances between the spaces corresponding to different subjects.

It is plain to see that the set of intra-subject distances between re-referenced direct and temporal embeddings are generally shorter than distances measured between subjects. By contrast, we see that the set of intra-subject distances between re-referenced correlation spaces X are approximately the same value as inter-subject distances. Hypothesis testing using the two-sided Kolmogorov–Smirnov test ($p < 0.01$) confirms that the correlation metric spaces have statistically similar inter-subject embeddings and intra-subject embeddings.

Beyond statistical significance, we can also quantify how dissimilar intra-subject versus inter-subject distances are by computing the effect size between distributions, following Cohen's d method (Kelley and Preacher, 2012). We find that the magnitude of the effect sizes is consistently much larger for the direct embedding spaces D and the temporal embeddings space T than for the correlation spaces X (Figure 4) (Sawilowsky, 2009). Indeed, a simple t-test comparing the inter-subject and the intra-subject distances reveals that the two distributions are indistinguishable when drawn from the correlation spaces. Our results therefore imply that the topology of the brain configuration spaces and of the temporal embeddings retain individual-specificity after re-referencing of the EEG data. By contrast, correlation embeddings appear to be unspecific to a given individual subject. Observed differences between temporal embeddings and correlation embeddings are especially apparent in H_1. While the temporal embeddings, T and D, retain their individual specificity, the correlation embedding displays many instances wherein distances across participants are closer than distances across references for the same participant. Put another way, in many cases the correlation embedding exhibits more diversity when re-referencing a single EEG acquisition than when comparing

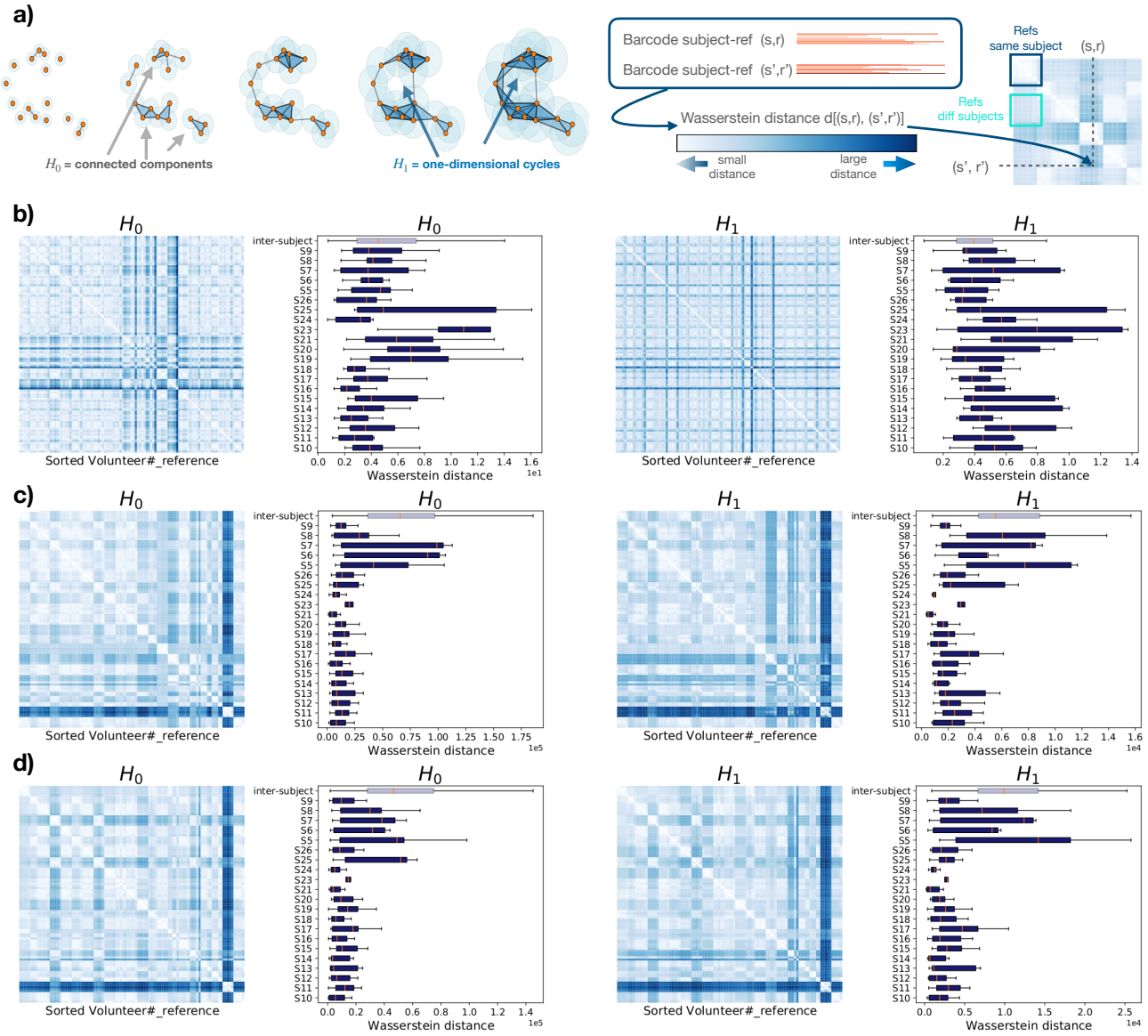

Fig. 3 Topological distance between spaces for different references. we show the results for the topological distances between the various embeddings for H_0, persistent connected components, and H_1, persistent one-dimensional cycles (panel a) left). We show distances between all (s, r) pairs as heatmaps (with distances growing from white to blue). That is, smaller differences between embeddings are coded by lighter colours. An example is shown in panel a) on the right. Rows and columns are ordered by ordered subject and reference (s, r) pairs, which results in references from each subject to be displayed in consecutive rows/columns. Therefore, the presence of diagonal blocks of short distances (lighter colored squares with respect to the off-diagonal blocks) implies that re-referencing induces small changes with respect to inter-subject variability. Panels b) c) and d) refer respectively to the $X^{s,r}$ $T^{s,r}$ and $D^{s,r}$ embedded spaces for the $cleanint$ datasets. For each embedding type, we compute distances between subjects and references for the first two homological groups, H_0 and H_1, using sliced Wasserstein distances between the corresponding persistence diagrams. It is easy to observe and modular block structure for the $T^{s,r}$ and $D^{s,r}$ spaces that is wider than for the $C^{s,r}$ spaces. The boxplots further support this result: we show the distribution of within-subject distances between spaces corresponding to different references $(d(X/T/D^{s,r}, X/T/D^{s',r'})|s = s' \forall r, r'$ (separated by subject, one dark coloured box for each subject, correpsonding to the diagonal blocks in the heatmaps) and compare it to the inter-subject distances $(d(X/T/D^{s,r}, X/T/D^{s',r'})|s \neq s', r = r'$, lighter color bar, corresponding to the off-diagonal terms in the heatmaps). For the temporal embeddings $T/D^{s',r'}$, the within-subject distances are smaller (KS test, $p < 0.01$) than the between-subject distances, while for the correlation-based embeddings $X^{s,r}$ the inter-subject distances are comparable with the within-subject distances. Results for other pipelines are reported in Figures A.1, A.2 and A.3.

multiple acquisitions from different participants (each pegged to the same reference). Moreover, this susceptibility to large changes in the correlation space after re-referencing is made plain through analysis of the simulated *tvb* data. Whereas intra-subject distances are always greater than inter-subject distances when observed through the correlation space. The temporal embeddings tend to preserve the expected intra-subject similarities. The one exception to this trend is found in the most abstract state space representation, T in H_1. Possibly this exception demonstrates the limits of the simulated dataset to produce complex EEG data that might preserve higher-order topological features of a subject's brain dynamics.

3.2 Role of different pre-processing pipelines

We next investigated whether the above results vary significantly for different choices of pre-processing. Indeed, EEG signals usually undergo a series of steps before being considered suitable for analysis. We repeated the analysis for all the pipelines.

Similarly to what we performed in the previous section, we show the results for effect sizes in Figure 4. We find that effect sizes are always larger (in magnitude) for the $T^{s,r}$ and $D^{s,r}$ spaces than for the corresponding $X^{s,r}$ spaces, and do not differ much across pre-processing pipelines ($d \sim 1 - 1.5$, considered a very large effect (Sawilowsky, 2009)). This result suggests that different pre-processing pipelines do induce some changes in the reconstructed topology, but also that these changes are negligible compared to what happens in the case of correlation networks. Additionally, we confirm the result that re-referencing induces changes in the reconstructed topology of D and T spaces that are much smaller than the inter-subject variability.

3.3 Comparison with the temporally reshuffled null model.

We showed that $T^{s,r}$ and $D^{s,r}$ spaces display stronger topological robustness with respect to those built from spatial correlations, while at the same time appearing to be more subject-specific. However, we have not ascertained whether these properties are a consequence of the statistics of the signals themselves, or rather they truly emerge from their temporal features. We tested these alternatives by comparing previous results to a null model in which we destroy temporal correlations by reshuffling the time labels of the instantaneous activity. Note that –by construction– the matrices $\mathbf{C}^{s,r}$ (and thus the $X^{s,r}$ spaces) remain exactly the same under this reshuffling. In other words, this means that we preserve exactly the spatial correlations, while destroying the temporal ones. The same argument holds for the $D^{s,r}$ embeddings; it is a direct embedding from the multi-channel EEG signals and therefore reshuffling the temporal labels does not change the point cloud. Here, we therefore only focus on the effects on $T^{s,r}$ spaces, by investigating the changes induced in the Takens embeddings constructed from the reshuffled series, $\tilde{T}^{s,r}$ spaces.

In Figure 5a) and b) we compare the results for H_0 and H_1. The first observation is that the distances between reshuffled datasets, corresponding to the same subject (across different references), appear to be much more heterogeneous than in the case of real data. This can be observed in several different ways. The distances between references of the same subject are much farther away from each other in the reshuffled case than in the real case (darker colored boxes versus the corresponding lighter colored boxes). Similarly, the inter-subject distances are typically larger in the reshuffled cases. This also holds for the distances between a specific real (s, r) pair and its reshuffled version (white box, labeled as real-rand in Fig. 5a-b). In fact, this latter distance distribution (real-rand) and the reshuffled inter-subject distance distribution (for all pipelines) have averages that are statistically indistinguishable (Mann-Whitney u test for equal mean, null hypothesis not rejected at $p < 0.01$), while the real inter-subject distance has a smaller average (significant on the same test). The same results hold for the other pre-processing schemes (Figure A.4).

We further support these observations by computing the corresponding effect sizes via Cohen's d, that is the effect sizes of difference between the distributions of distances $d(T^{s,r}, T^{s,r'})$ versus that of $d(T^{s,r}, \tilde{T}^{s,r})$. For both H_0 and H_1 and for all pipelines, the effect sizes across subjects are significantly smaller than zero (Figure 5c). Hence, destroying temporal correlations induces varied and heterogeneous changes in the topology of the resulting embeddings; changes that are much larger than those induced by pure re-referencing. Interestingly, for the tvb simulated data, signal changes related to re-referencing often alter the topology of the delay embedding as much, if not to a greater extent, than breaking the temporal correlations. This observation highlights the stationarity of tvb dynamics. That is, the range of temporal states is approximately as large as the range of spatial features.

Finally, we queried the magnitude of the changes induced by re-referencing with respect to inter-subject variability. Similarly to Figure 4, we computed the effect sizes for the distribution of intra-subject and inter-subject distances for the real and reshuffled data (Figure 5d).

That is, for H_0 all pipelines display effect sizes significantly different from zero for both real and reshuffled data (one-sample t-test for mean equal 0, $p < 0.01$). For H_1, only real data show an effect size significantly smaller than 0 (same test). Interestingly, when directly comparing the real

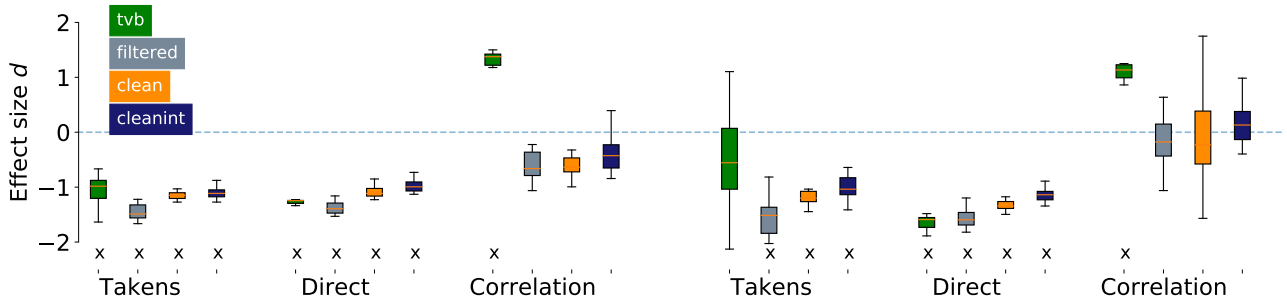

Fig. 4 Effect size distributions for within-subject versus inter-subject distances. For each subject s, we compute the Cohen's d to compare the collection of within-subject Wasserstein distances (computed across different references) versus the set of inter-subject Wasserstein distances. For each pre-processing step r, and types of embedding (T, D, X), we collect the Cohen's d values across all subjects and display them as a distribution. Instances where the set of within-subject distances are significantly ($p < 0.05$) different from inter-subject distances are marked with an 'x'. The effect of pre-processing is to remove idiosyncratic outliers. We find that for all studied pipelines, the temporal embeddings $T^{s,r} and D^{s,r}$ show larger differences (larger absolute effect size) with respect to the $X^{s,r}$ spaces, implying that temporal embeddings are more robust to re-referencing than functional connectivity.

and reshuffled effect sizes, we find significant differences (Mann-Whitney u test at $p < 0.01$ Bonferroni-corrected) only for the clean and filtered datasets. Note that this does not mean that in the cleanint datasets the actual distances are the same. Rather, the effect size between the intra-subject and inter-subject distribution is similar in the real and reshuffled cases in the presence of further pre-processing.

Overall, this pattern suggests that the differences observed above originate partially from the statistical properties of the signals, irrespective of the time ordering. More precisely, this concerns the relative difference across subjects with respect to different references of the same subject. However, temporal correlations are crucial to constrain the set of possible topologies describing a subject and to discriminate between subjects even when the statistical properties of the signals are similar.

4 Discussion

We studied the topological structure of different representations of resting-state EEG signals with a particular focus on how the choice of the reference alters the resulting topological observables. We examined three representations that capture different features of the data. The first one was the functional connectivity between electrodes, which we computed as Pearson correlations between electrode timeseries. When computing these correlations, time is integrated away and for this reason the resulting correlations capture the patterns of spatial coactivations among signals at different electrodes. The second type of representation was direct embeddings. In these embeddings, each time point is associated with the vector of instantaneous EEG signal. Together, all the vectors of instantaneous EEG activations define a space that captures the possible configurations of brain activations. Note that in this construction, relative temporal information is lost. Thus, the direct embedding does not encode brain dynamics, but rather the range of possible topographies. Finally, the third type of representation was given by Takens embeddings. These are constructed by concatenating instantaneous EEG vectors corresponding to successive time points. In this way, they reconstruct the properties of the attractor space of a dynamical system (Noakes, 1991; Myers et al., 2019).

We found that the extent of topological changes across correlation spaces corresponding to different references are often comparable with those measured between different subjects. We found that the direct embeddings and Takens embeddings exhibited limited changes across different references and were able to discriminate better across subjects.

The implications of these results are multi-fold. As mentioned above, for a fixed reference, the $D^{s,r}$ space is the configuration space (i.e. phase space) of EEG whole-brain activations (i.e. topographies (Tivadar et al., 2019)). The facts that the topology of signals does not change significantly across references and that it is subject-specific suggest that the overall shape of EEG configuration space holds information about the specificity of the underlying individual's brain dynamics, similarly to what has been observed for simpler dynamical systems (Donato et al., 2016).

However, by construction, the configuration spaces above neglect the role of time, i.e. the temporal ordering in which brain activity appears. It is reasonable then to ask whether the specific order of time points plays a role. The Takens embeddings explore exactly this question. In fact, as mentioned, they allow us to probe the brain's dynamical attractor space underlying the observed activations (Myers et al., 2019). If there was no information in the temporal ordering, the topological structures of $T^{s,r}$ and of its randomized $\tilde{T}^{s,r}$ spaces should be similar to one another. Instead, we observed a large standard deviation within embeddings when temporal correlations were removed. More precisely, we observed that a much larger and more heterogeneous set of topological

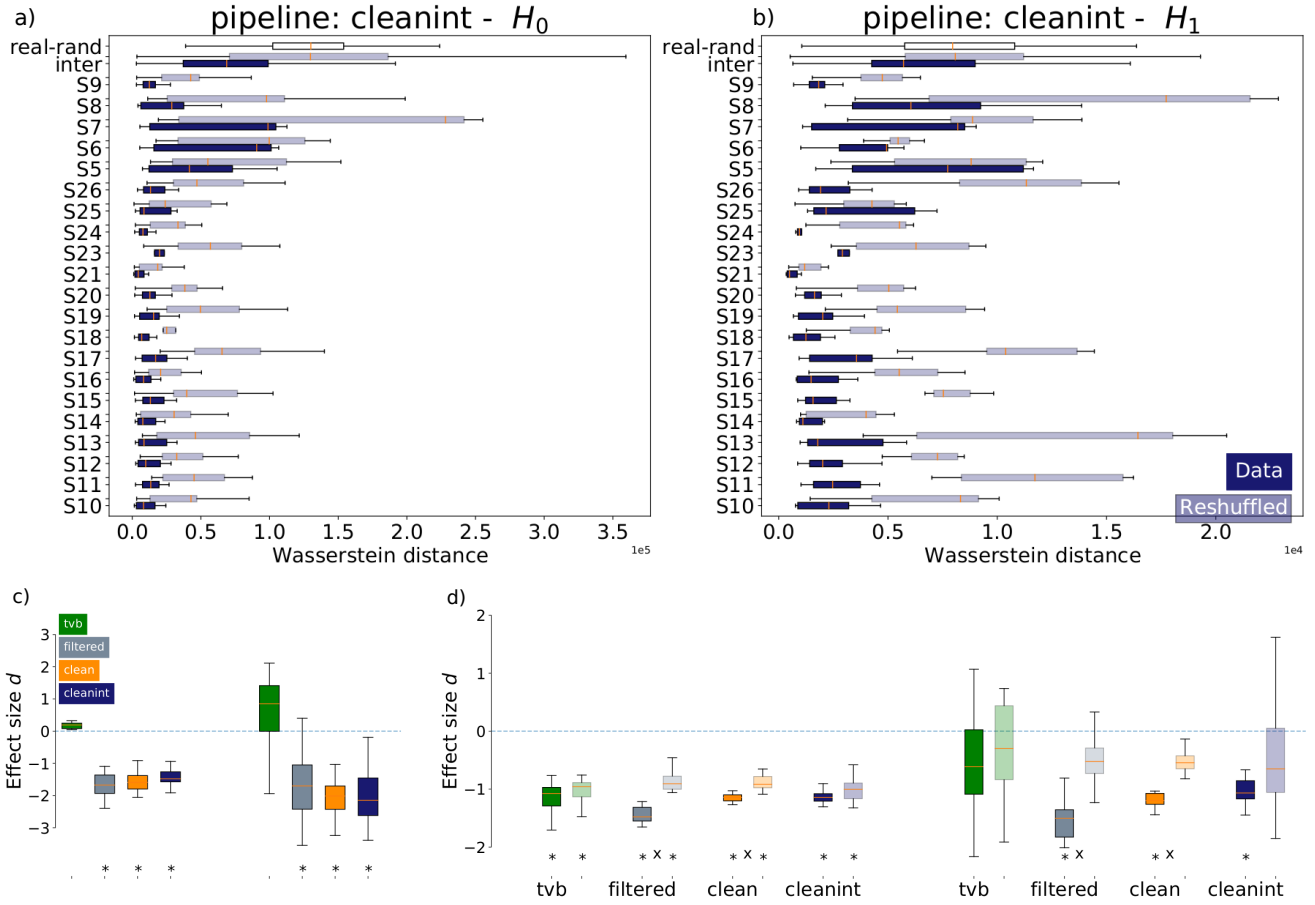

Fig. 5 Comparison of real with reshuffled data. a) and b) Distance distributions (respectively for H_0 and H_1) between references of individual subject (solid color), between the temporally reshuffled data (lighter color boxes). inter labels the inter-subject distance distribution for both real (solid) and reshuffled data (lighter color). real-rand (white box) represents the distribution of distances between a (s, r) pair and its temporally reshuffled version. Distances among references of the same subject have generally smaller mean and variance with respect to the reshuffled data. Moreover, the distances between a (s, r) pair and its randomized versions are often larger than those between different subjects. c) we confirm this by computing the corresponding effect size via Cohen's d, that is the effect sizes of the distances $d(T^{s,r}, T^{s,r'})$ versus that of $d(T^{s,r}, \tilde{T}^{s,r})$. For both H_0 and H_1 and for all pipelines, the effect sizes across subjects are significantly smaller than zero (varying between -1 and -2, considered to be very large effects, asterisks indicate significance on one-sample t-test at $p < 0.01$ Bonferroni corrected for multiple comparisons to reject the null hypothesis that the effect size mean is 0). This feature does not hold for tvb simulations for which the statistical properties of the signal's evolution across time vary to the same degree as the statistical properties of the signal's evolution across space. d) for all pipelines, effect sizes for the intrasubject distance distributions $d(T^{s,r}, T^{s,r'})$ are shown against the inter-subject distributions $d(T^{s,r}, T^{s',r'})$. Solid color boxes indicate comparison between real data, lighter color boxes indicate the same comparison for the reshuffled data. Asterisks indicate significance on one-sample t-test at $p < 0.01$ Bonferroni corrected for multiple comparisons to reject the null hypothesis that the effect size mean is 0; crosses indicate significance on Mann-Whitney u test at $p < 0.01$ Bonferroni corrected for multiple comparisons to reject the null hypothesis that the real and reshuffled samples have the same mean. In most cases, the shuffled data significantly blurs the idiosyncratic nature of individual EEG traces, causing rereferenced signals within a volunteer to look almost as different as the signal traces compared between volunteers.

spaces is explored when only spatial correlations and statistical properties of the signals are preserved. Conversely, this implies that temporal correlations constrain to a large degree the set of possible dynamics. These in turn shape how the space of activations differs across subjects' resting-state activity.

There are several ways by which to intuitively understand these findings. One can think the configuration space of activations $D^{s,r}$ as a mountainous landscape, where each point represents a state that the brain has accessed during the recordings. The Takens embedding instead describes how an individual explores this configuration landscape, i.e. the order of states as explored by a subject; each point in the Takens embedding can be thought as composed by a set of successive positions in the configuration space (i.e. a path in the mountainous landscape across different states), that is, a trajectory through time. We found here that both the landscape and how it is explored –the set of trajectories– change across subjects. However, both of these features are very robust to re-referencing and pre-processing choices.

We claim that our results extend to the configuration spaces and the set of trajectories of topographies. While we

can only access information about topographies via their referenced instances, our results imply that different reference choices are essentially equivalent from a topological perspective. That is, for different references we found very similar topologies, which therefore implies that we are capturing the topology of the actual space of topographies. This claim is further supported when considering the properties of temporal versus spatial embeddings in a simulated virtual brain. Here, we know at the outset that EEG dynamics will trace a repetitive cycle, and that the orientation of this cycle is rotated from the perspective of a different reference. The topology of an temporal embedding of the cyclic dynamics is unperturbed after rotation. By contrast, drawing correlation distances between spatially-distinct electrodes causes some electrode pairs to look more-or-less similar after rotation—thus causing the topology of the correlation space to also change. The empirical and theoretical utility of the topological description of brain dynamics brings us to propose to augment conventional topographic analyses with this additional –topological– level of analysis.

Naturally, our work also leads to many novel questions, both technical and theoretical: do the topographical configuration space and/or its dynamical properties change under different conditions, e.g. wake versus sleep, altered states, performance of different tasks, etc.? Do shared topological structures emerge under such conditions that are stronger than inter-individual variability? Previous findings with fRMI and EEG suggest that during tasks the spatial correlation structure is already sufficient to discriminate between tasks, subjects (Ibáñez-Marcelo et al., 2019b,a) or altered states (Petri et al., 2014). It would be indeed important to ascertain which features of the topographic and topological spaces are preserved under different conditions, both analytically and empirically Haufe and Ewald (2019), as this would have direct impact on brain fingerprinting and on functional neuro-degeneration tracking among others (Bari et al., 2019; Rajapandian et al., 2020). This is an exciting endeavour that is currently at the forefront of our current ongoing investigations.

Funding

B.F. and M.M.M. are supported by the Fondation Asile des aveugles (grant number 232933 to M.M.M.). M.M.M. is also supported by The Swiss National Science Foundation (grant number 169206). R.T. is supported by the Swiss National Science Foundation (#320030_188737). G. P. and J.B. acknowledge support during the preparation of this work from Intesa Sanpaolo Innovation Center. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

Authors contribution

M.M.M., B.F. and G.P. conceived the study. R.T. gathered and pre-processed the data. J.B. structured and carried out the topological data analysis over the three embedding types. B.F., G.P.,J.B. and R.T. interpreted together the results. R.T. and J.B. wrote the first draft of the manuscript. B.F. supervised the neuroscientific contents and G.P. the topological ones. All authors contributed to the final draft and review.

References

Bari S, Amico E, Vike N, Talavage TM, Goñi J (2019) Uncovering multi-site identifiability based on resting-state functional connectomes. NeuroImage 202:115967

Bassett DS, Sporns O (2017) Network neuroscience. Nature neuroscience 20(3):353–364

Battiston F, Cencetti G, Iacopini I, Latora V, Lucas M, Patania A, Young JG, Petri G (2020) Networks beyond pairwise interactions: structure and dynamics. Physics Reports

Betzel RF, Byrge L, He Y, Goñi J, Zuo XN, Sporns O (2014) Changes in structural and functional connectivity among resting-state networks across the human lifespan. Neuroimage 102:345–357

Biasiucci A, Franceschiello B, Murray MM (2019) Electroencephalography. Current Biology 29(3):R80—-R85, DOI 10.1016/j.cub.2018.11.052, URL https://doi.org/10.1016/j.cub.2018.11.052

Billings J, Saggar M, Hlinka J, Keilholz S, Petri G (2021) Simplicial and topological descriptions of human brain dynamics. Network Neuroscience pp 1–20, DOI 10.1162/netn_a_00190, URL https://doi.org/10.1162/netn_a_00190, https://direct.mit.edu/netn/article-pdf/doi/10.1162/netn_a_00190/1916469/netn_a_00190.pdf

Chan HL, Kuo PC, Cheng CY, Chen YS (2018) Challenges and Future Perspectives on Electroencephalogram-Based Biometrics in Person Recognition. Frontiers in Neuroinformatics 12:66, DOI 10.3389/fninf.2018.00066, URL https://www.frontiersin.org/article/10.3389/fninf.2018.00066

Chella F, Pizzella V, Zappasodi F, Marzetti L (2016) Impact of the reference choice on scalp EEG connectivity estimation. Journal of neural engineering 13(3):36016

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including

independent component analysis. Journal of neuroscience methods 134(1):9–21

Deyle ER, Sugihara G (2011) Generalized theorems for nonlinear state space reconstruction. PLoS ONE 6(3):e18295, DOI 10.1371/journal.pone.0018295

Donato I, Gori M, Pettini M, Petri G, De Nigris S, Franzosi R, Vaccarino F (2016) Persistent homology analysis of phase transitions. Physical Review E 93(5):52138

Edelsbrunner H, Harer J (2008) Persistent homology-a survey. Contemporary mathematics 453:257–282

Fulekar MH (2009) Bioinformatics: applications in life and environmental sciences. Springer Science & Business Media

Ghrist R (2008) Barcodes: The persistent topology of data. DOI 10.1090/S0273-0979-07-01191-3, URL http://repository.upenn.edu/grasp{_}papershttps://repository.upenn.edu/grasp{_}papers/1

Giusti C, Pastalkova E, Curto C, Itskov V (2015) Clique topology reveals intrinsic geometric structure in neural correlations. Proceedings of the National Academy of Sciences 112(44):13455–13460

Giusti C, Ghrist R, Bassett DS (2016) Two's company, three (or more) is a simplex. Journal of computational neuroscience 41(1):1–14

Haufe S, Ewald A (2019) A simulation framework for benchmarking eeg-based brain connectivity estimation methodologies. Brain topography 32(4):625–642

Hu S, Yao D, Bringas-Vega ML, Qin Y, Valdes-Sosa PA (2019) The statistics of eeg unipolar references: derivations and properties. Brain topography 32(4):696–703

Hutchison RM, Womelsdorf T, Allen EA, Bandettini PA, Calhoun VD, Corbetta M, Della Penna S, Duyn JH, Glover GH, Gonzalez-Castillo J, et al. (2013) Dynamic functional connectivity: promise, issues, and interpretations. Neuroimage 80:360–378

Iacopini I, Petri G, Barrat A, Latora V (2019) Simplicial models of social contagion. Nature communications 10(1):1–9

Ibáñez-Marcelo E, Campioni L, Manzoni D, Santarcangelo EL, Petri G (2019a) Spectral and topological analyses of the cortical representation of the head position: Does hypnotizability matter? Brain and behavior 9(6):e01277

Ibáñez-Marcelo E, Campioni L, Phinyomark A, Petri G, Santarcangelo EL (2019b) Topology highlights mesoscopic functional equivalence between imagery and perception: The case of hypnotizability. NeuroImage 200:437–449

Kelley K, Preacher KJ (2012) On effect size. Psychological methods 17(2):137

Lee S, Kang H, Chung MK, Kim BN, Lee DS (2012) Persistent brain network homology from the perspective of dendrogram. IEEE transactions on medical imaging 31(12):2267–2277

Lehmann D (1987) Principles of spatial analysis. Handbook of electroencephalography and clinical neurophysiology: Methods of analysis of brain electrical and magnetic signals 1:309–354

Lehmann D, Michel CM (2011) EEG-defined functional microstates as basic building blocks of mental processes. Clinical Neurophysiology 122(6):1073–1074, DOI https://doi.org/10.1016/j.clinph.2010.11.003, URL http://www.sciencedirect.com/science/article/pii/S138824571000787X

Leon PS, Knock SA, Woodman MM, Domide L, Mersmann J, McIntosh AR, Jirsa V, Marinazzo D, Plesser HE (2013) The virtual brain: a simulator of primate brain network dynamics DOI 10.3389/fninf.2013.00010, URL www.frontiersin.org

Lepage KQ, Kramer MA, Chu CJ (2014) A statistically robust EEG re-referencing procedure to mitigate reference effect. Journal of Neuroscience Methods 235:101–116, DOI https://doi.org/10.1016/j.jneumeth.2014.05.008, URL http://www.sciencedirect.com/science/article/pii/S0165027014001629

Luck SJ (2014) An Introduction to the Event-Related Potential Technique. A Bradford Book, MIT press, URL https://books.google.com/books?id=SzavAwAAQBAJ

Marcel S, R Millan JD (2007) Person Authentication Using Brainwaves (EEG) and Maximum A Posteriori Model Adaptation. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(4):743–752

Marinazzo D, Riera JJ, Marzetti L, Astolfi L, Yao D, Sosa PAV (2019) Controversies in EEG Source Imaging and Connectivity: Modeling, Validation, Benchmarking

Michel CM, Murray MM (2012) Towards the utilization of EEG as a brain imaging tool. NeuroImage 61(2):371–385, DOI https://doi.org/10.1016/j.neuroimage.2011.12.039, URL http://www.sciencedirect.com/science/article/pii/S1053811911014418

Michel CM, Thut G, Morand S, Khateb A, Pegna AJ, Grave de Peralta R, Gonzalez S, Seeck M, Landis T (2001) Electric source imaging of human brain functions. Brain Research Reviews 36(2):108–118, DOI https://doi.org/10.1016/S0165-0173(01)00086-8, URL http://www.sciencedirect.com/science/article/pii/S0165017301000868

Michel CM, Murray MM, Lantz G, Gonzalez S, Spinelli L, Grave de Peralta R (2004) EEG source imaging. Clinical Neurophysiology 115(10):2195–2222, DOI https://doi.org/10.1016/j.clinph.2004.06.001, URL http://www.sciencedirect.com/science/article/pii/S1388245704002135

Michel Christoph M TKDBLRRGJW (2009) Electrical Neuroimaging. DOI 10.1017/CBO9780511596889

Murray MM, Brunet D, Michel CM (2008) Topographic ERP analyses: a step-by-step tutorial review. Brain topography 20(4):249–264, DOI 10.1007/s10548-008-0054-5, URL https://doi.org/10.1007/s10548-008-0054-5

MWong PKH (2012) Introduction to brain topography. Springer Science & Business Media

Myers A, Munch E, Khasawneh FA (2019) Persistent homology of complex networks for dynamic state detection. Physical Review E 100(2):22314

Nicholas J Cavanna, Mahmoodreza Jahanseir, Don Sheehy (2015) A geometric perspective on sparse filtrations. In: CCCG, Queen's University, Ontario, Canada

Noakes L (1991) The Takens embedding theorem. International Journal of Bifurcation and Chaos 1(04):867–872

Grave de Peralta Menendez R, Gonzalez Andino S, Morand S, Michel C, Landis T (2000) Imaging the electrical activity of the brain: Electra. Human Brain Mapping 9(1):1–12

Perrin F, Pernier J, Bertrand O, Echallier JF (1989) Spherical splines for scalp potential and current density mapping. Electroencephalography and clinical neurophysiology 72(2):184–187

Petri G, Scolamiero M, Donato I, Vaccarino F (2013) Topological strata of weighted complex networks. PloS one 8(6):e66506

Petri G, Expert P, Turkheimer F, Carhart-Harris R, Nutt D, Hellyer PJ, Vaccarino F (2014) Homological scaffolds of brain functional networks. Journal of The Royal Society Interface 11(101)

Poulos M, Rangoussi M, Alexandris N (1999) Neural network based person identification using EEG features. In: 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258), IEEE, vol 2, pp 1117–1120

Rajapandian M, Amico E, Abbas K, Ventresca M, Goñi J (2020) Uncovering differential identifiability in network properties of human brain functional connectomes. Network Neuroscience 4(3):698–713

Reininghaus J, Huber S, Bauer U, Kwitt R (2015) A stable multi-scale kernel for topological machine learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4741–4748

Sakkalis V (2011) Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. Computers in biology and medicine 41(12):1110–1117

Sawilowsky SS (2009) New effect size rules of thumb. Journal of Modern Applied Statistical Methods 8(2):26

Schirner M, Domide L, Perdikis D, Triebkorn P, Stefanovski L, Pai R, Popa P, Valean B, Palmer J, Langford C, Blickensdörfer A, van der Vlag M, Diaz-Pier S, Peyser A, Klijn W, Pleiter D, Nahm A, Schmid O, Woodman M, Zehl L, Fousek J, Petkoski S, Kusch L, Hashemi M, Marinazzo D, Mangin JF, Flöel A, Akintoye S, Stahl BC, Cepic M, Johnson E, Deco G, McIntosh AR, Hilgetag CC, Morgan M, Schuller B, Upton A, McMurtrie C, Dickscheid T, Bjaalie JG, Amunts K, Mersmann J, Jirsa V, Ritter P (2021) Brain modelling as a service: The virtual brain on ebrains. 2102.05888

Sporns O (2013) Network attributes for segregation and integration in the human brain. Current opinion in neurobiology 23(2):162–171

Tenke CE, Kayser J (2005) Reference-free quantification of EEG spectra: combining current source density (CSD) and frequency principal components analysis (fPCA). Clinical Neurophysiology 116(12):2826–2846

Tivadar RI, Retsa C, Turoman N, Matusz PJ, Murray MM (2018) Sounds enhance visual completion processes. NeuroImage 179:480–488

Tivadar RI, Murray MM, Tivadar RI, Murray MM (2019) A primer on electroencephalography and event-related potentials for organizational neuroscience. Organizational Research Methods 22(1):69–94, DOI 10.1177/1094428118804657, URL http://journals.sagepub.com/doi/10.1177/1094428118804657

Varley TF, Denny V, Sporns O, Patania A (2020) Topological analysis of differential effects of ketamine and propofol anesthesia on brain dynamics. bioRxiv

Vaughan HG (1982) The neural origins of human event-related potentials. Annals of the New York Academy of Sciences

Yao D, Kin, Yun, Hu, Shiang, Dong, Li, Vega, Maria, Sosa, Pedro A Valdés (2019) Which {Reference} {Should} {We} {Use} for {EEG} and {ERP} practice? Brain Topography 32(4):530–549, DOI 10.1007/s10548-019-00707-x, URL http://link.springer.com/10.1007/s10548-019-00707-x

Zomorodian A, Carlsson G (2005) Computing persistent homology. Discrete & Computational Geometry 33(2):249–274

A Appendix

A.1 Results for additional preprocessing pipelines

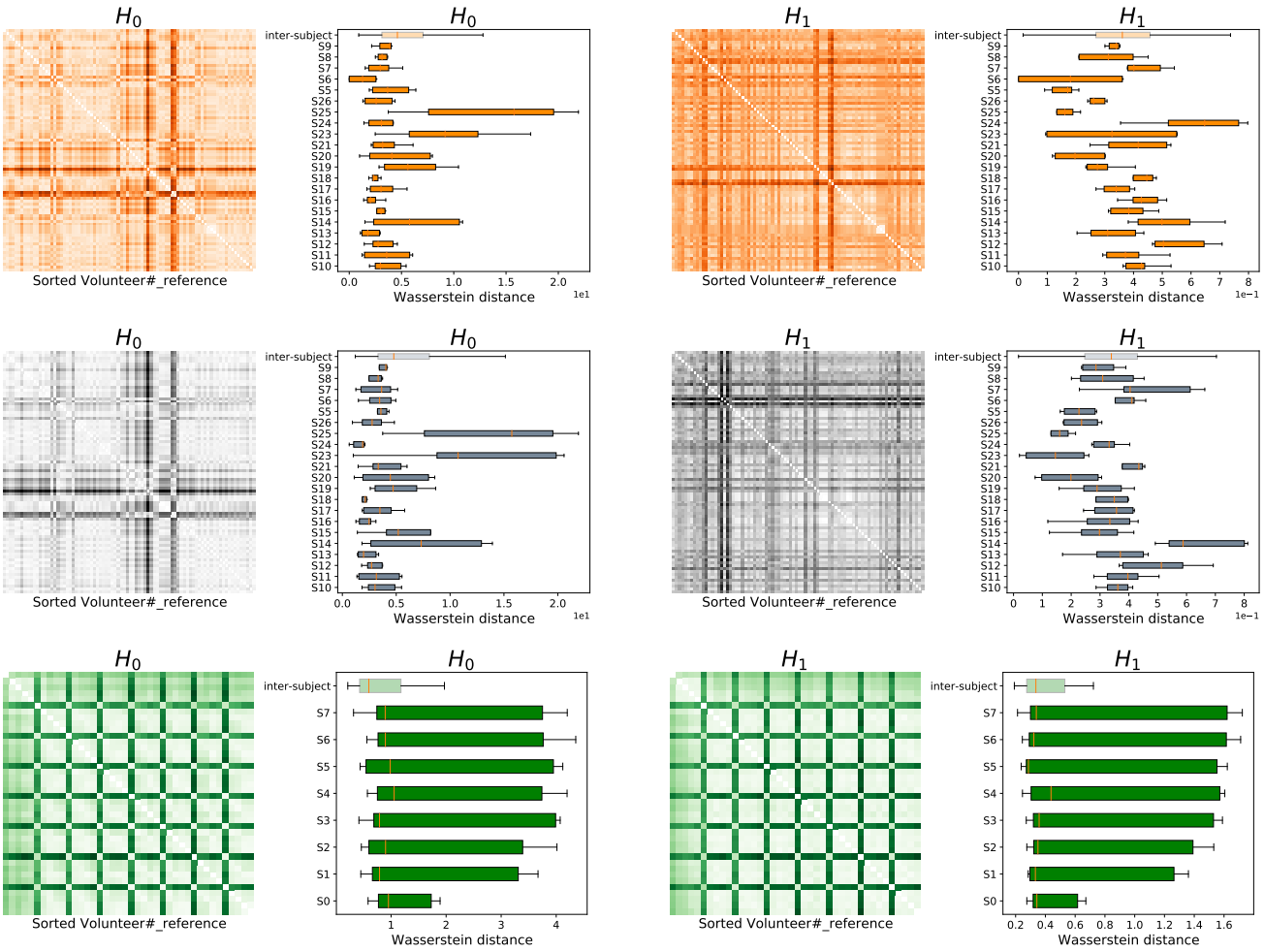

Fig. A.1 Effect of different preprocessing pipelines on correlation spaces $X^{s,r}$. Additional pipeline results for Figure 3: (top row) clean pipeline. (middle row) filtered pipeline. (bottom row) tvb pipeline.

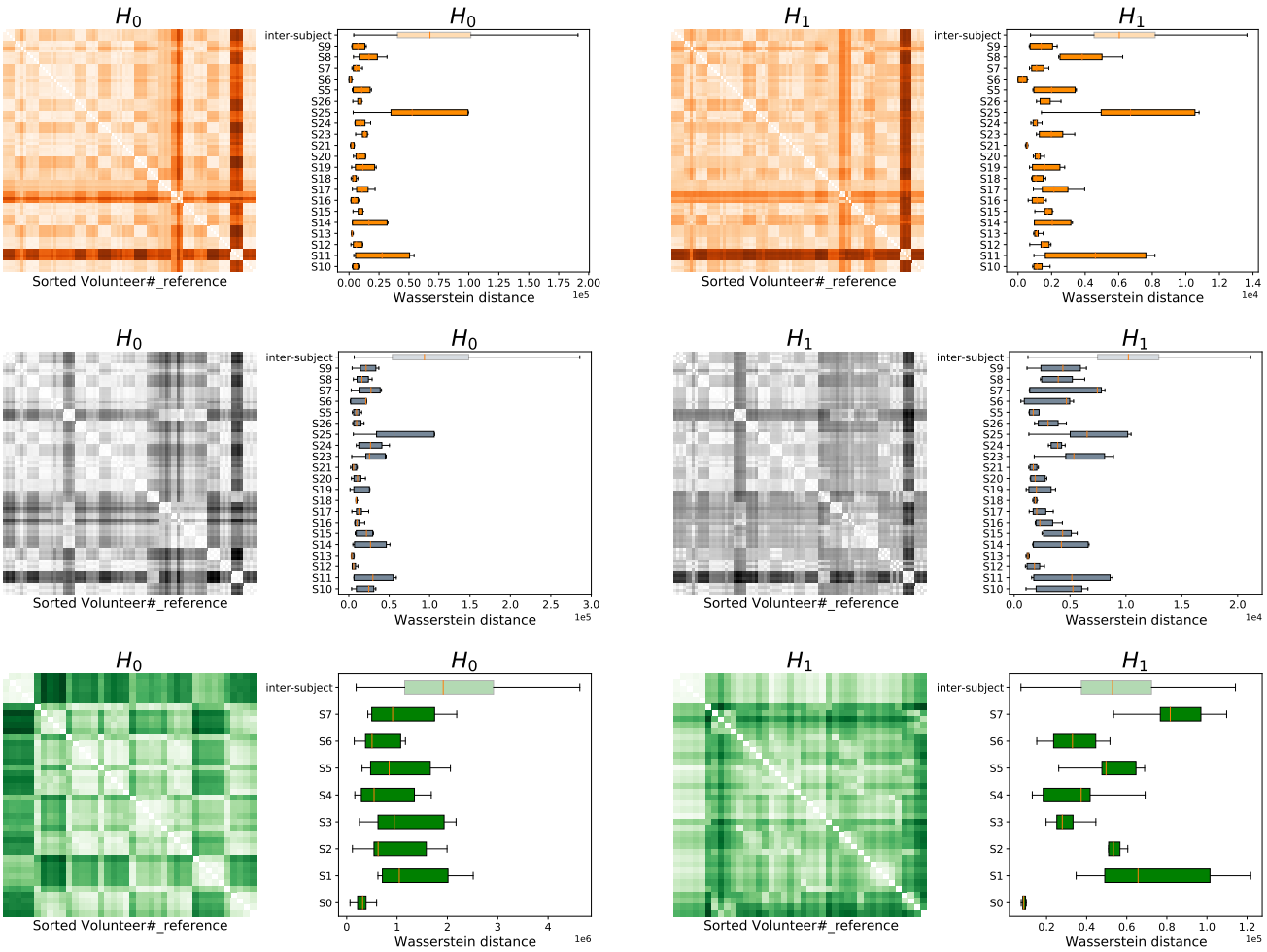

Fig. A.2 Effect of different preprocessing pipelines on Takens embedding spaces $T^{s,r}$. Additional pipeline results for Figure 3: (top row) clean pipeline. (middle row) filtered pipeline. (bottom row) tvb pipeline.

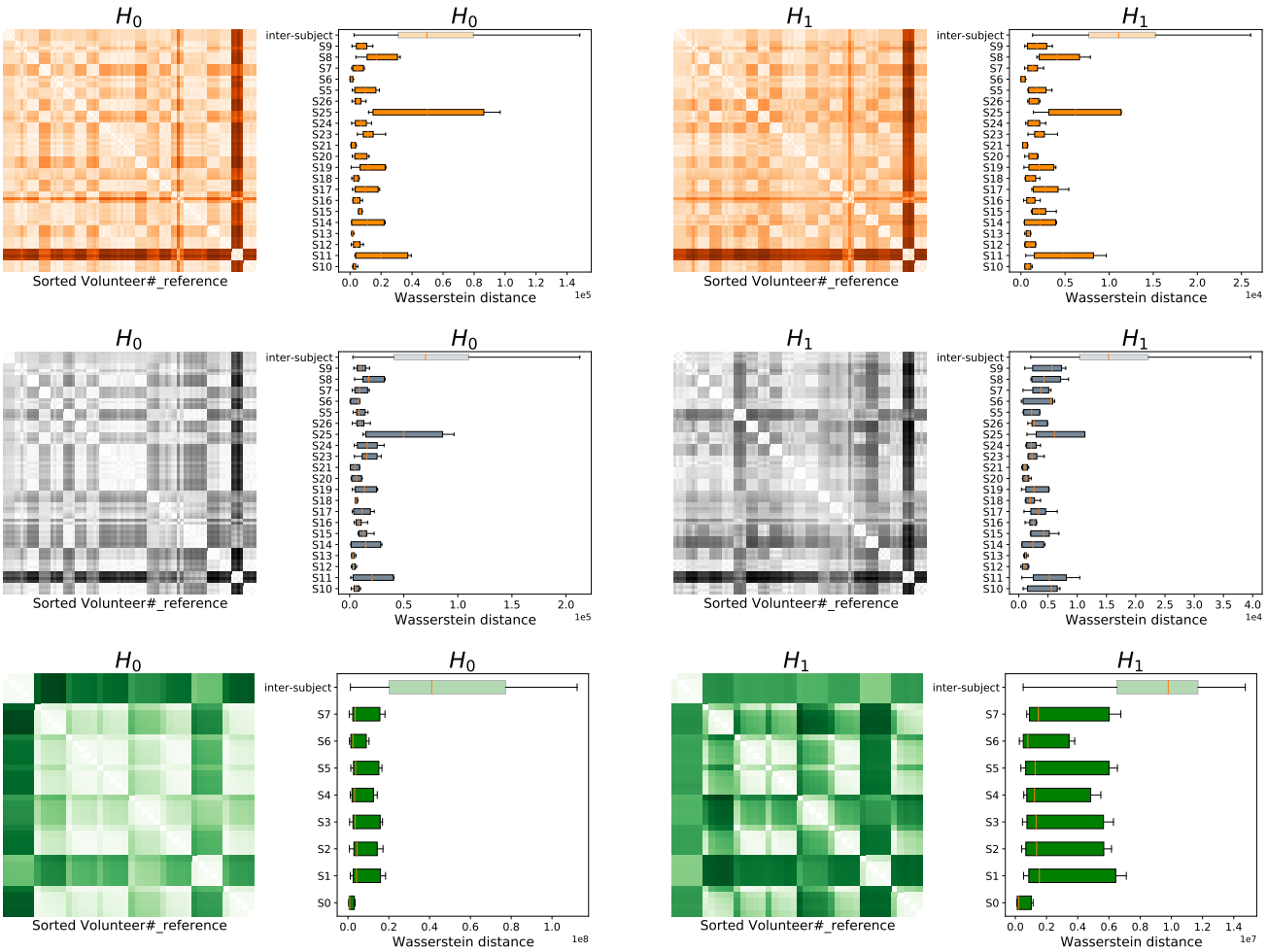

Fig. A.3 Effect of different preprocessing pipelines on direct temporal embedding spaces $D^{s,r}$. Additional pipeline results for Figure 3: (top row) clean pipeline. (middle row) filtered pipeline. (bottom row) tvb pipeline.

Results for $T^{s,r}$ spaces from temporally shuffled timeseries

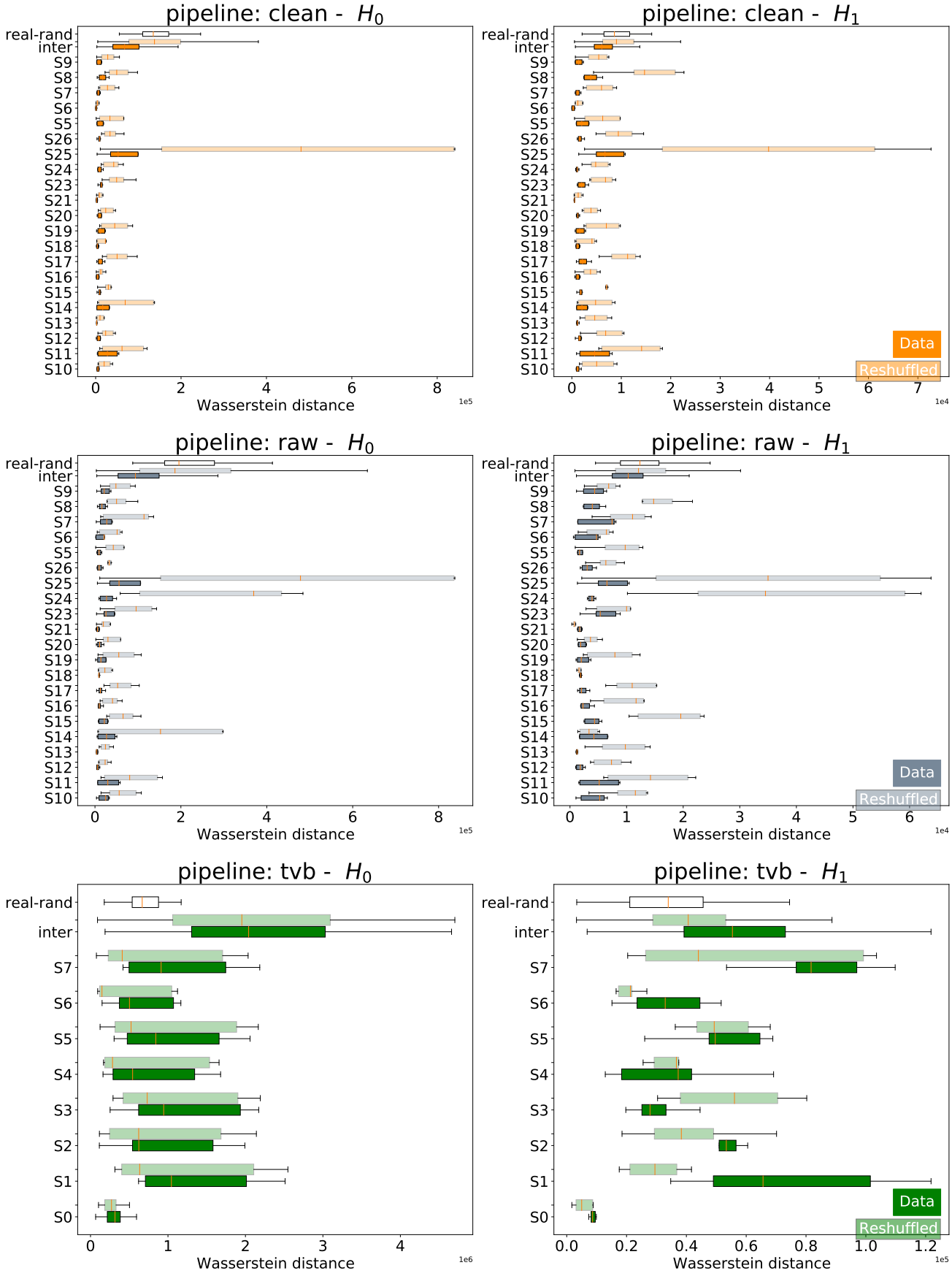

Fig. A.4 Effect for temporally reshuffled timeseries for clean, filtered, and tvb pipelines.