Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 58-15 (2024) 414–419

A latent representation of brain networks based on EEG ⋆

Lucia Falconi * **Giulia Cisotto** ** **Mattia Zorzi** ***

* *Automatic Control Laboratory, ETH Zurich, Physikstrasse 3, 8092 Zurich, Switzerland (email: lfalconi@ethz.ch)*

** *Dept. of Informatics, Systems, and Communication, University of Milano-Bicocca, Milan, Italy; Dept. of Information Engineering, University of Padova, Italy (e-mail: giulia.cisotto@unimib.it)*

*** *Dept. of Information Engineering, University of Padova, Italy (e-mail: zorzimat@dei.unipd.it)*

Abstract: Electroencephalography (EEG) is one of the most popular techniques to investigate normal as well as pathological cerebral mechanisms, as it allows to measure, non-invasively and in real-time, the brain activity. However, modeling EEG is still extremely challenging, because of its high-dimensionality, low signal-to-noise ratio, and high individual variability. This paper proposes a novel latent representation to study brain networks using EEG by means of a robust dynamic factor analysis (RDFA) approach. We investigate the ability of this latent representation to discriminate between two groups of subjects, *i.e.* alcoholic and healthy.
By RDFA, we can extract a limited number of highly explanatory factors, as low as 8, significantly discriminating between the two groups. Also, we show that different brain patterns can be identified across different stimulation scenarios and EEG locations. Although preliminary, this work could give support to domain experts while providing some clinically-meaningful insights to identify common patterns as well as individual characteristics in different groups of healthy and pathological subjects.

Keywords: EEG; system identification; dynamic factor analysis.

1. INTRODUCTION

Electroencephalography (EEG) is an electrophysiological technique to record brain activity in a non invasive way, with very high temporal resolution (*i.e.* milliseconds) and convenient acquisition setups (relatively cheap and fast to implement). Unfortunately, these advantages come with two relevant weaknesses: EEG suffers from a poor spatial resolution and a very low signal-to-noise ratio (SNR). This makes difficult to properly model EEG and easily generalize any accurate model over new subjects and new tasks (Croce et al., 2020). Consequently, one of the major challenges in EEG is to overcome the abovementioned issues and identify common patterns across different individuals and across different tasks performed by the same subject. A flourishing literature of models for EEG has been produced in the last decades (Pardey et al., 1996; Saeidi et al., 2021; Craik et al., 2019; Lawhern et al., 2018). EEG data are typically acquired from multiple locations by multiple EEG sensors uniformly placed on the human scalp. Accordingly, these signals characterize a dynamic network model, called brain network, where each node corresponds to a specific EEG sensor (and thus a specific scalp location).

It is most often needed to decompose the EEG data, represented as a multi-channel time-series, into components which can retain the most relevant information about space and time variations of the brain activity, while discarding or depressing irrelevant changes due to noise or other interferences. This step is shared among quite different methods, *i.e.* the independent component analysis (ICA) step in many standard processing techniques, employed to extract brain-related information and remove artefacts (Pion-Tonachini et al., 2019), or the latent representation in a variational autoencoder (VAE) embedding different brain activities during the imagination of hands or feet movements (Zancanaro et al., 2023).

There is no consensus about the number of latent factors or the characterization of a proper latent representation of a multi-channel EEG time-series coming from healthy subjects. Limiting the number of latent variables and maximizing their effectiveness in discriminating among different subjects (e.g. healthy vs pathological) or tasks (*i.e.* performed by the same individual) is generally beneficial for EEG modeling (Celli et al., 2022). However, the most recent literature reports highly varying latent representations, both for their dimensionality and effectiveness in application studies (Jeng et al., 2020).

⋆ This work was partially supported by Fondazione CARIPARO (Borse di Dottorato CARIPARO 2020) and the Italian Ministry of University and Research under the grant "Dipartimenti di Eccellenza 2023-2027" of the Department of Informatics, Systems and Communication of the University of Milano-Bicocca, Italy. G. Cisotto also acknowledges the financial support of PON "Green and Innovation" 2014-2020 action IV.6 funded by the Italian Ministry of University and Research to the University of Milano-Bicocca (Milan, Italy).

10.1016/j.ifacol.2024.08.564

Many paradigms for learning latent representations have been proposed in the control community, see for instance (Veedu et al., 2021; Zorzi and Sepulchre, 2016; Ning et al., 2015; Alpago et al., 2022; Maanan et al., 2018; You and Yu, 2024; Ramaswamy et al., 2022; Zorzi, 2019; Anderson et al., 2012; Zorzi and Chiuso, 2017). In this paper, we will focus our attention on dynamic factor models, which describe the variability among a large number of observed variables in terms of a potentially small number of latent factors, by assuming that the observed variables are mostly a linear combination of the latent factors (Deistler, 2019). The problem of learning such models as well as the number of latent factors is inherently fragile; in fact, small perturbations in the data can lead to a substantial variation of the number of latent factors (Ciccone et al., 2020, 2019).

The aim of this paper is to provide a latent representation of a brain network based on EEG signals which is robust to the aforementioned fragility. More precisely, such latent representation is based on the factor model inferred from the data by means of the robust dynamic factor analysis (RDFA) approach recently proposed in (Falconi et al., 2024). We compare the ability of the resulting latent representation to discriminate between two groups of subjects, *i.e.* with no-alcoholism history and with alcoholism dependency, using a public dataset. It turns out that the number of latent factors required for the latent representation is in line with other literature (Celli et al., 2022), i.e. such number is between 5 and 20.

The rest of the paper is organized as follows: in Section 2, we introduce the RDFA approach and the corresponding latent representation used in the following; in Section 3 we analyze the public dataset containing healthy as well as pathological EEG data using the RDFA approach. Finally, Section 4 concludes the paper highlighting some promising future perspectives.

Notation: Given a matrix M, we denote its transpose by $M^\top$ and by $M_{(i,j)}$ the element of M in the i-th row and j-th column. If M is a square matrix, $\mathrm{tr}(M)$ and $|M|$ denote its trace and its determinant, respectively. Let $\mathbf{Q}_m$ be the space of real symmetric matrices of size m; if $M \in \mathbf{Q}_m$ is positive definite or positive semidefinite, then we write $M \succ 0$ or $M \succeq 0$, respectively. If $M \in \mathbf{Q}_m$ then $\lambda_j(M)$ is the $j-th$ largest eigenvalue of M. We denote by $(\cdot)^*$ the complex conjugate transpose. $\Phi(e^{i\vartheta})$ for $\vartheta \in [-\pi, \pi]\}$ denotes a function defined on the unit circle $\{e^{i\vartheta} : \vartheta \in [-\pi, \pi]\}$, and the dependence on ϑ is dropped if necessary. If $\Phi(e^{i\vartheta})$ is positive (semi-)definite $\forall \vartheta \in [-\pi, \pi]$ we write $\Phi(e^{i\vartheta}) \succ 0$ ($\succeq 0$). Integrals are always defined from $-\pi$ to π with respect to the normalized Lebesgue measure $d\vartheta/2\pi$. We define the finite-dimensional space of trigonometric polynomials $\mathcal{Q}_{m,n} = \{\sum_{k=-n}^{n} R_k e^{-i\vartheta k}, R_k = R_{-k}^T \in \mathbb{R}^{m\times m}\}$.

2. ROBUST DYNAMIC FACTOR ANALYSIS

Let $y = \{y(t),\ t \in \mathbb{Z}\}$ be a stochastic zero mean Gaussian process of dimension m. We say that y is described by an AutoRegressive Moving-Average (ARMA) factor model if

$$y(t) = a(z)^{-1}(W_L(z)u(t) + W_D(z)w(t)) \tag{1}$$

where

$$a(z) = 1 + \sum_{k=1}^{p} a_k z^{-k}, \quad a_k \in \mathbb{R}$$

$$W_L(z) = \sum_{k=0}^{n} W_{L,k} z^{-k}, \quad W_{L,k} \in \mathbb{R}^{m\times f}$$

$$W_D(z) = \sum_{k=0}^{n} W_{D,k} z^{-k}, \quad W_{D,k} \in \mathbb{R}^{m\times m} \text{ diagonal.}$$

The processes $u = \{u(t),\ t \in \mathbb{Z}\}$ and $w = \{w(t),\ t \in \mathbb{Z}\}$ are normalized white Gaussian noises of dimension f and m, respectively, with $f \ll m$, which are independent. Model (1) has the following interpretation: y is the process containing the variables that we can observe; u is the process which describes the f latent factors not accessible to observations; $a(z)^{-1}W_L(z)$ is the factor loading transfer matrix, while $a(z)^{-1}W_L(z)u(t)$ is the latent variable; finally, $a(z)^{-1}W_D(z)w(t)$ is the idiosyncratic noise. The aim of a factor model, as the one in (1), is to explain the observed variables as the contribution of two terms: one which is common to all these variables and driven by a few latent factors; the other term is peculiar to each observed variable and it is given by the idiosyncratic noise. It is worth noting that the latent factors cannot be univocally characterized. Indeed, we can rewrite (1) as

$$y(t) = a(z)^{-1}(\tilde{W}_L(z)\tilde{u}(t) + W_D w(t))$$

where $\tilde{u}(t) = Uu(t)$, $\tilde{W}_L(z) = W_L(z)U^T$ and $U \in \mathbb{R}^{f\times f}$ is an arbitrary orthogonal matrix. Notice that

$$y_{MA}(t) = a(z)y(t) = W_L(z)u(t) + W_D(z)w(t) \tag{2}$$

is an MA process of order n whose spectral density

$$\Phi(z) = W_L(z)W_L(z)^* + W_D(z)W_D(z)^* = \Phi_L(z) + \Phi_D(z), \tag{3}$$

admits a "low rank plus diagonal" decomposition, because by construction we have that Φ_L has rank equal to f and Φ_D is diagonal.

Assume to collect a realization $\mathrm{y}^N = \{\mathrm{y}(1) \dots \mathrm{y}(N)\}$ of the process y defined by (1); we face the problem of estimating the parameters of the factor model (1), i.e. Φ_L, Φ_D and $a(z)$, and the number of factors f, assuming that the orders p and n are given. To address this problem, the idea is to estimate first the parameters of the filter a, and then the spectral densities Φ_L and Φ_D by preprocessing y^N through a. More precisely, the ARMA factor model is estimated by the two steps:

(i) *AR dynamics estimation.* Given the realization y^N, we estimate the p parameters of the filter a by applying the maximum likelihood estimator proposed in (Crescente et al., 2020, Section II.b). In doing so, we are estimating an AR process whose spectral density is $a(z)^{-1}a(z)^{-*}I_m$.

(ii) *MA dynamics factor analysis.* Let $\mathrm{y}_{MA}^N = \{y_{MA}(1) \dots y_{MA}(N)\}$ be the finite length trajectory obtained by filtering the trajectory y^N through $a(z)I_m$, *i.e.* according to (2), with zero initial conditions. Then, we compute the truncated periodogram $\hat{\Phi} \in \mathcal{Q}_{m,n}$ from the time series y_{MA}^N (Stoica et al. (2005)) which represents a preliminary estimate of Φ defined in (3). If $\hat{\Phi}(e^{i\vartheta})$ is not positive definite for $\vartheta \in [-\pi, \pi)$, then it will be corrected by adding εI_m, with $\varepsilon > 0$ taken small, in such a way that $\hat{\Phi}(e^{i\vartheta}) + \varepsilon I_m \succeq 0$.

Then, we estimate Φ_L and Φ_D according to the following robust approach proposed in (Falconi et al., 2024, Section III-V):

$$\begin{aligned} \min_{\Phi,\Phi_L,\Phi_D\in\mathcal{Q}_{m,n}} \quad & \operatorname{tr}\int \Phi_L \\ \text{st} \quad & \Phi_L + \Phi_D = \Phi, \\ & \Phi \succ 0 \text{ a.e.}, \;\; \Phi_L, \Phi_D \succeq 0, \\ & \Phi_D \text{ diagonal}, \\ & \mathcal{S}_{IS}(\Phi||\hat{\Phi}) \leq \delta. \end{aligned} \tag{4}$$

Here, the objective represents the nuclear norm of Φ_L and promotes a solution for Φ_L having low rank. The first three constraints impose that Φ_L and Φ_D provide a genuine spectral density decomposition of type (3). The last constraint, where $\mathcal{S}_{IS}(\Phi||\hat{\Phi})$ is the Itakura-Saito divergence (Ferrante et al., 2012)

$$\mathcal{S}_{IS}(\Phi||\hat{\Phi}) := \int \log|\hat{\Phi}\Phi^{-1}| + \operatorname{tr}(\hat{\Phi}^{-1}\Phi - I_m),$$

imposes that Φ belongs to a set "centered" in the nominal spectral density $\hat{\Phi}$ with prescribed tolerance δ. The latter is estimated from the data according to a resampling-based method, see (Falconi et al., 2024, Section II.A) for more details: δ is chosen in such a way the resulting confidence region contains the true spectrum with a prescribed probability α. In this paper we set $\alpha = 0.5$; the MA order $n = 2$ and the AR order $p = 2$ have been chosen small. However, one could select them empirically to trade-off parsimony of the model and goodness-of-fit.

Let Φ_L be the low rank estimated spectral density obtained by applying the RDFA procedure to the dataset y_{MA}^N and

$$s_j(\Phi_L) = \int \frac{\lambda_j\big(\Phi_L(e^{i\vartheta})\big)}{\lambda_1\big(\Phi_L(e^{i\vartheta})\big)} \tag{5}$$

be the integral over the unit circle of the j-th largest normalized eigenvalue of Φ_L. We can determine the numerical rank of Φ_L, *i.e.* the number f of latent factors, as the largest integer j such that $s_j > c$, where $c \in [0, 1]$ is a user-chosen threshold.

2.1 Latent representation via RDFA

The latent variable corresponding to the factor model (1) is defined as

$$y_L(t) = a(z)^{-1}W_L(z)u(t).$$

Recall that this latent representation is not unique, but the spectral density of y_L is unique and it is equal to

$$\Psi(z) = (a(z)^{-1}W_L(z))(a(z)^{-1}W_L(z))^*.$$

In what follows, we choose a latent representation whose latent factors can be ordered in terms of their degree of influence on y_L. Let $y_{L,j}(t)$, with $j = 1 \dots m$, denote the j-th component of y_L. Then, we have that

$$\operatorname{var}(y_{L,j}(t)) = \int (\Psi)_j$$

where $(\Psi)_j$ denotes the j-th element in the main diagonal of Ψ. Notice that the rank of Ψ is equal to the number of factors, *i.e.* f. Thus, Ψ admits the following pointwise eigenvalue decomposition

$$\Psi(e^{i\vartheta}) = \sum_{k=1}^{f} \lambda_k(\Psi(e^{i\vartheta}))\left(v_k(e^{i\vartheta})v_k(e^{i\vartheta})^*\right)_j \tag{6}$$

where $\lambda_k(\Psi(e^{i\vartheta})) > 0$ denotes the k-th largest eigenvalue of $\Psi(e^{i\vartheta})$), $v_k(e^{i\vartheta})$ is the corresponding eigenvector and $v_1(e^{i\vartheta}) \dots v_f(e^{i\vartheta})$ are the orthonormal vectors. Notice that, such decomposition is unique provided that $\lambda_k(\Psi(e^{i\vartheta})) \neq \lambda_o(\Psi(e^{i\vartheta}))$ for any $k \neq o$. Then,

$$\operatorname{var}(y_{L,j}(t)) = \sum_{k=1}^{f} \operatorname{var}_k(y_{L,j}(t))$$

and

$$\operatorname{var}_k(y_{L,j}(t)) := \int \lambda_k(\Psi)(v_k v_k^*)_j \geq 0 \tag{7}$$

denotes the portion of variance of $y_{L,j}$ explained by the k-th latent factor according to the latent representation induced by the decomposition in (6). In particular, the fraction of variance of $y_{L,j}$ explained by the o most influent factors is equal to

$$\sigma^2_{L,j}(o) = \frac{\sum_{k=1}^{o} \operatorname{var}_k(y_{L,j}(t))}{\operatorname{var}(y_{L,j}(t))}. \tag{8}$$

Clearly, $\sigma^2_{L,j}(1) \leq \sigma^2_{L,j}(2) \leq \ldots \leq \sigma^2_{L,j}(f) = 1$.

3. RESULTS AND DISCUSSION

3.1 Dataset

To test our approach, we used the public EEG dataset collected by the Henri Begleiter Neurodynamic Lab, State University of New York (Begleiter (1999), which includes 77 alcoholic patients and 44 healthy controls [1]. The experimental paradigm consisted in a standard visual stimulation where all subjects, regardeless which group they belong to, were randomly administered three different types of stimulation, namely *scenarios*: a single stimulation ($s = 1$), a double stimulation in a matched modality ($s = 2$), and a double stimulation in an unmatched modality ($s = 3$). All visual stimuli (*i.e.*, pictures) were taken from the 1980 Snodgrass and Vanderwart picture set (Snodgrass and Vanderwart, 1980). In this standard dataset of images, pictures were selected based on four variables that influence memory and cognitive processing: name agreement, image agreement, familiarity, and visual complexity. Healthy controls and alcoholic patients are likely to show different visual/cognitive processing capabilities. However, the clinical interpretation of the results is beyond the scope of this paper. The EEG data were acquired from 64 locations all over the scalp, following the International $10-10$ EEG Standard, with a sampling frequency of 256 Hz. We randomly extracted three 1-second long EEG segments from every subject in the dataset. Therefore, $44 \times 3 = 132$ segments from healthy controls and $77 \times 3 = 231$ segments from alcoholic patients are available for every scenarios.

3.2 Statistical analysis of the number of factors

Let $l \in \{A, C\}$ be a categorical variable used to discriminate between the two groups, *i.e.* alcoholic patients

[1] The dataset is availble at the UCI Machine Learning repository: `https://archive.ics.uci.edu/dataset/121/eeg+database`

(A) and healthy controls (C). We denote by $y_{l,i,s,r}$ the EEG segment corresponding to the experimental scenario $s \in \{1,2,3\}$, run $r \in \{1,2,3\}$ and individual i, where $i \in \{1,2,\dots,77\}$ for $l = A$ and $i \in \{1,2,\dots,44\}$ for $l = C$.

For any segment $y_{l,i,s,r}$, we estimated the corresponding factor model using the RDFA method (thus $y_{l,i,s,r}$ represents our dataset denoted by y^N in Section 2). Then, the number of its latent factors, estimated as described in Section 2 with threshold value $c = 0.05$ is denoted by $f_{l,i,s,r}$. The value of the parameter c was chosen empirically, as the one providing a reasonable performance. However, changing it in the range $[0.01\ 0.15]$, we did not notice a remarkable difference, as shown later in this section. To compare the estimated $f_{l,i,s,r}$ given by the RDFA methodology, we considered two well-established methods. The first one is the *principal component analysis* (PCA), which transforms the original data into new coordinates, called *principal components*, ordered for their decreasing relevance to explain the original data. Given the segment $y_{l,i,s,r}$, we can estimate the number of latent factors $f_{l,i,s,r}$ via PCA by determining the greatest index j such that $\lambda_j(\hat{\Sigma})/\lambda_1(\hat{\Sigma})$ is greater than the threshold c, where $\hat{\Sigma}$ is the sampled covariance matrix. Notice that, for a fair comparison, we use the same threshold c as in RDFA method. One major shortcoming of PCA is the lack of focus on time dependence. For this reason, we also considered the *Dynamic Principal Component Analysis* (DPCA) approach (Brillinger, 2001), which takes into account temporal correlations among data, too. The idea of DPCA is the same as its static counterpart, but the *dynamic* principal components are obtained from the pointwise eigen-decomposition of $\hat{\Gamma}(e^{j\vartheta})$, which is an estimate of the spectral density matrix obtained by the segment $y_{l,i,s,r},$. Here, we computed $\hat{\Gamma}(e^{j\vartheta})$, by windowing the segment $y_{l,i,s,r}$ with a 50 sample-long Bartlett window. Then, it is possible to compute $s_j(\hat{\Gamma})$ using the same definition in (5) and the number of latent factors $f_{l,i,s,r}$, by counting the number of s_j larger than the threshold c.

For the three methods, we computed the average value of latent factors across the three different runs as

$$\bar{f}_{l,i,s} = \frac{1}{3}\sum_{r=1}^{3} f_{l,i,s,r},$$

with $l \in \{A,C\}$, $i \in \{1,2,\dots,77\}$ for $l = A$ and $i \in \{1,2,\dots,44\}$ for $l = C$, and $s \in \{1,2,3\}$.

To evaluate the performance of these estimators, we assessed their ability to discriminate between alcoholic patients and healthy controls as our show-case problem.

Figure 1 shows the comparison between the number of latent factors to represent the brain network in different scenarios and groups of subjects, *i.e.* alcoholic patients or healthy controls, using RDFA, PCA and DPCA, respectively. Each box-plot includes the values of the following populations: $\bar{f}_{l,s} = \{\bar{f}_{l,1,s}, \bar{f}_{l,2,s},, \bar{f}_{l,3,s}, \dots\}$ with $l \in \{A,C\}$ and $s \in \{1,2,3\}$. Then, we compared $\bar{f}_{A,s}$ with $\bar{f}_{C,s}$ across the three different scenarios. The plots corresponding to RDFA suggest that the brain networks from alcoholic patients are characterized by a higher number of latent factors with respect to the healthy controls. The largest difference between groups seems to occur when

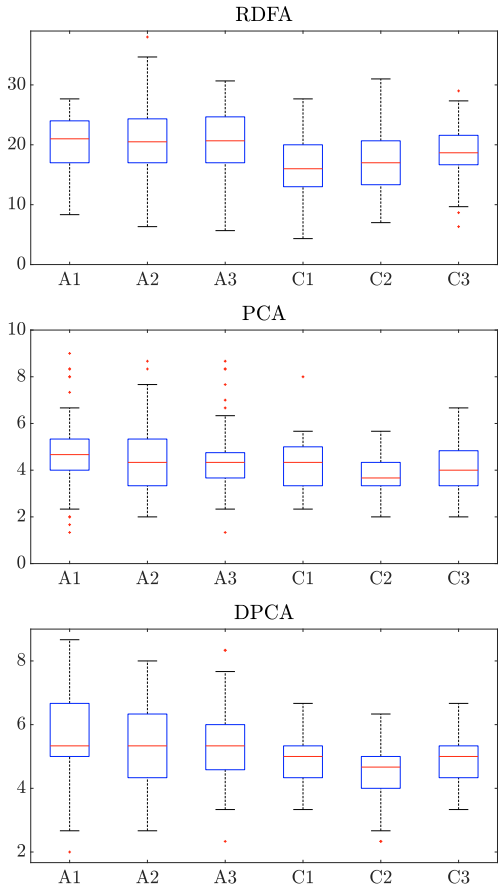

Fig. 1. Box-plot of the the populations $\bar{f}_{A,s}$ and $\bar{f}_{C,s}$ where the number of latent factors was estimated via RDFA, PCA and DPCA with a threshold of $c = 0.05$.

the individuals are administered with the simplest type of stimulation, *i.e.* with a single stimulus. This might be explained by the increased variability in the individual effects of a double stimulation (with the unmatched double stimulation in the patients group possibly leading to the most difficult effects to identify). This is also supported by the higher consistency of the average number of latent factors needed for the healthy controls group, on average approximately equal to 20, regardless the considered scenario. On the other hand, we observed that PCA and DPCA can represent brain networks using a significantly lower number of latent factors, *i.e.* less than 10. However, they seem not to provide similar discriminability as when RDFA is employed.

Statistical hypothesis testing. To support the preliminary observations of Section 3.2, we performed a statistical analysis using a hypothesis testing method to evaluate the statistical difference between groups in the different scenarios, using RDFA and its competitors.

Specifically, we applied the Wilcoxon rank-sum test to evaluate whether the populations

$$\bar{f}_{A,s} := \{\, \bar{f}_{A,i,s} \quad \text{s.t.} \quad i = 1,2,\dots,77\}$$
$$\bar{f}_{C,s} := \{\, \bar{f}_{C,i,s} \quad \text{s.t.} \quad i = 1,2,\dots,44\}$$

are likely to come from different distributions, taking $s = 1,2,3$, separately. The significance level for all tests was set to 0.05. Furthermore, we investigated the results

of this test with different values of the threshold c. The results are summarized in Table 1 which shows the test decision variable (H) and the p-value (p) for every test performed across the two groups of subjects for each value of c, each scenario s, for RDFA as well as the competitors. As expected, the number of RDFA-based latent factors is statistically significantly lower in the control group (*i.e.* the null hypothesis was rejected for all scenarios and c values), thus making the number of factors, itself, a useful feature to discriminate between alcoholic patients and healthy controls. The results obtained by RDFA are in line with those of DPCA, while PCA generally showed lower performance. However, the most relevant statistical difference was obtained by using RDFA (regardless the specific value of the parameter c) in the first scenario, *i.e.* with the simplest visual stimulation.

		RDFA		PCA		DPCA	
s	c	H	p	H	p	H	p
1	0.01	1	$9.0 \cdot 10^{-5}$	1	$4.0 \cdot 10^{-3}$	1	$1.1 \cdot 10^{-4}$
	0.05	1	$1.3 \cdot 10^{-4}$	0	$1 \cdot 10^{-1}$	1	$1.8 \cdot 10^{-3}$
	0.1	1	$4.1 \cdot 10^{-5}$	0	$1.6 \cdot 10^{-1}$	1	$1.1 \cdot 10^{-2}$
	0.15	1	$1.4 \cdot 10^{-4}$	0	$1.3 \cdot 10^{-1}$	1	$3.8 \cdot 10^{-3}$
2	0.01	1	$7.1 \cdot 10^{-4}$	1	$1.9 \cdot 10^{-3}$	1	$9.5 \cdot 10^{-5}$
	0.05	1	$1.2 \cdot 10^{-3}$	1	$4.4 \cdot 10^{-3}$	1	$1.0 \cdot 10^{-3}$
	0.1	1	$9.3 \cdot 10^{-4}$	1	$6.8 \cdot 10^{-3}$	1	$6.9 \cdot 10^{-3}$
	0.15	1	$2.5 \cdot 10^{-3}$	1	$5.0 \cdot 10^{-3}$	1	$9.2 \cdot 10^{-3}$
3	0.01	1	$8.5 \cdot 10^{-3}$	1	$4.7 \cdot 10^{-3}$	1	$2.8 \cdot 10^{-3}$
	0.05	1	$2.6 \cdot 10^{-2}$	0	$6.6 \cdot 10^{-2}$	1	$1.8 \cdot 10^{-2}$
	0.1	1	$1.7 \cdot 10^{-2}$	1	$7.5 \cdot 10^{-3}$	0	$5.0 \cdot 10^{-2}$
	0.15	1	$4.9 \cdot 10^{-2}$	1	$1.6 \cdot 10^{-2}$	0	$8.3 \cdot 10^{-2}$

Table 1. Comparison between groups via Wilcoxon rank-sum test using the number of latent factors as extracted by our RDFA and other two state-of-the-art methods (PCA and DPCA). The significance level was set to 0.05.

3.3 Latent representation via RDFA

For each experimental scenario, run and individual, we computed the fraction of variance of the latent variable corresponding to the j-th node explained by the o most influent factors according to the latent representation defined in Section 2.1. Let $\sigma^2_{l,i,s,r,j}(o)$ denote such fraction, defined in (8), corresponding to the factor model estimated from the segment $y_{l,i,s,r}$. Then, we computed this fraction averaged across the three different runs and individuals, for $l = A$ and $l = C$, respectively:

$$\bar{\sigma}^2_{A,s,j}(o) = \frac{1}{77 \cdot 3} \sum_{i=1}^{77} \sum_{r=1}^{3} \sigma^2_{l,i,s,r}(o),$$

$$\bar{\sigma}^2_{C,s,j}(o) = \frac{1}{44 \cdot 3} \sum_{i=1}^{44} \sum_{r=1}^{3} \sigma^2_{l,i,s,r}(o).$$

Figure 2 (top panel) shows $\bar{\sigma}^2_{A,s,j}(o)$ and $\bar{\sigma}^2_{C,s,j}(o)$ as a function of $o \in \{1 \dots 20\}$, for the node j corresponding to POz for the first scenario (*i.e.* $s = 1$). First, we notice that the curve corresponding to group C is always above the one corresponding to group A. This result is in line with the analysis in Subsection 3.2, as the number of estimated factors for the group A is higher. In other words, more factors are necessary to explain the variability of the nodes for alcoholic patients. Moreover, we notice that the fraction of variance explained by the most influent factor, *i.e.* $o = 1$, although the most consistent (around 0.4), is similar across groups. This suggests that it might describe a common "baseline" dynamics, without discriminating any individual characteristic related to the task or the presence of the pathology. As o increases, the difference between $\bar{\sigma}^2_{A,s,j}(o)$ and $\bar{\sigma}^2_{C,s,j}(o)$ becomes larger, with $\bar{\sigma}^2_{C,s,j}(o)$ stably exceeding $\bar{\sigma}^2_{A,s,j}(o)$ and reaching the largest distance at $o = 8$. Thus, we explored the amount of variance explained by the first $o = 8$ factors accross the different scenarios in three different representative electrodes, *i.e.* FCz, CPz, and POz, located in quite different positions over the scalp (see Figure 2 bottom panels). As expected, POz always shows the highest values, being located over the visual cortex, *i.e.* the brain region mostly involved in the processing of visual stimuli (Renold et al., 2014). Then, we found that both FCz and CPz replicated the behaviour observed at POz, *i.e.,* with $\bar{\sigma}^2_{C,s,j}(8)$ significantly exceeding $\bar{\sigma}^2_{A,s,j}(8)$, no matter which scenario is considered.

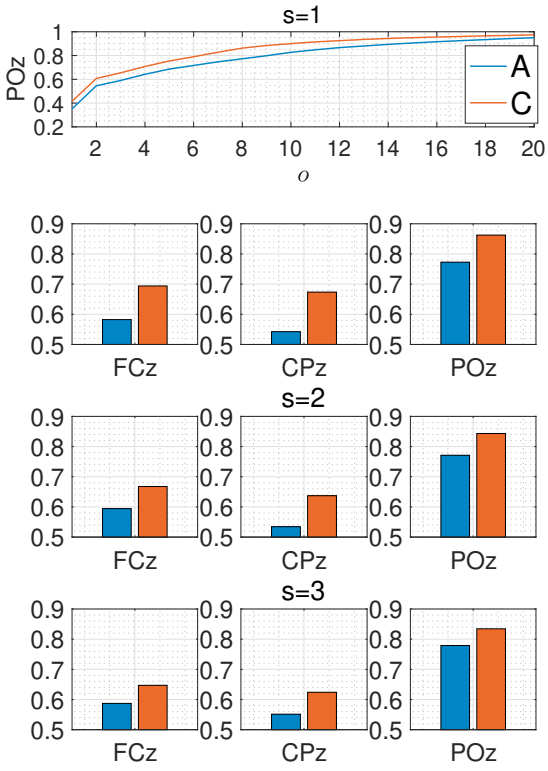

Fig. 2. *Top panel for* $s = 1$: fraction of variance explained for the node corresponding to POz as a function of the number of the most influent factors. *Bottom panels for* $s = 1, 2, 3$: bar plots representing the fraction of variances explained by the 8 most influential factors for the nodes corresponding to FCz, CPz, and POz.

Interestingly, the largest difference occurs for $s = 1$ (*i.e.* the scenario with the simplest visual stimulation) which is in line with the previous analysis. Finally, the explanatory value of the first 8 factors changes across the three representative electrodes we chose, with a similar pattern in the two groups across the scenarios. This might give some clinically-meaninful insights that could deserve further discussions with the domain experts.

4. CONCLUSIONS

We applied the RDFA method introduced in Falconi et al. (2024) to provide a latent representation of a brain network obtained from EEG signals, and we investigated its ability to discriminate between alcoholic patients and healthy controls. We found that the proposed latent representation shows a "baseline" dynamics common to all individuals (as the most influent factor, accounting for about 0.4 variance, is similar between groups, regardless any individual characteristic). However, we were also able to show that the brain networks from alcoholic patients are characterized by a different number of latent factors than the healthy controls, *i.e.* typically higher. The difference between the two groups is particularly evident when individuals are administered with a simpler stimulation. Further investigations could be carried on in the future to study the trade-off between the number of latent factors used in the latent representations and the effectiveness of a clusterization or a classification of the two different groups of individuals.

REFERENCES

Alpago, D., Zorzi, M., and Ferrante, A. (2022). A scalable strategy for the identification of latent-variable graphical models. *IEEE Trans. Automatic Control*, 67(7), 3349–3362.

Anderson, B., Deistler, M., Chen, W., and Filler, A. (2012). Autoregressive models of singular spectral matrices. *Automatica*, 48(11), 2843–2849.

Begleiter, H. (1999). EEG Database. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5TS3D.

Brillinger, D.R. (2001). *Time series: data analysis and theory*. SIAM.

Celli, M., Mazzonetto, I., Zangrossi, A., Bertoldo, A., Cona, G., and Corbetta, M. (2022). One-year-later spontaneous eeg features predict visual exploratory human phenotypes. *Communications Biology*, 5(1), 1361.

Ciccone, V., Ferrante, A., and Zorzi, M. (2019). Factor models with real data: A robust estimation of the number of factors. *IEEE Trans. Automatic Control*, 64(6), 2412–2425.

Ciccone, V., Ferrante, A., and Zorzi, M. (2020). Learning latent variable dynamic graphical models by confidence sets selection. *IEEE Trans. Automatic Control*, 65(12), 5130–5143.

Craik, A., He, Y., and Contreras-Vidal, J.L. (2019). Deep learning for electroencephalogram (eeg) classification tasks: a review. *J. neural engineering*, 16(3), 031001.

Crescente, F., Falconi, L., Rozzi, F., Ferrante, A., and Zorzi, M. (2020). Learning AR factor models. In *IEEE CDC*, 274–279.

Croce, P., Quercia, A., Costa, S., and Zappasodi, F. (2020). Eeg microstates associated with intra-and inter-subject alpha variability. *Scientific reports*, 10(1), 2469.

Deistler, M. (2019). Singular arma systems: A structure theory. *Numerical Algebra, Control & Optimization*, 9(3).

Falconi, L., Ferrante, A., and Zorzi, M. (2024). A Robust Approach to ARMA Factor Modeling. *IEEE Trans. Automatic Control*, 69(2).

Ferrante, A., Masiero, C., and Pavon, M. (2012). Time and spectral domain relative entropy: A new approach to multivariate spectral estimation. *IEEE Trans. Automatic Control*, 57(10), 2561–2575.

Jeng, P.Y., Wei, C.S., Jung, T.P., and Wang, L.C. (2020). Low-dimensional subject representation-based transfer learning in eeg decoding. *IEEE J. Biomedical and Health Informatics*, 25(6), 1915–1925.

Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., and Lance, B.J. (2018). Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *J. neural engineering*, 15(5), 056013.

Maanan, S., Dumitrescu, B., and Giurcăneanu, C. (2018). Maximum entropy expectation-maximization algorithm for fitting latent-variable graphical models to multivariate time series. *Entropy*, 20, 76.

Ning, L., Georgiou, T.T., Tannenbaum, A., and Boyd, S.P. (2015). Linear models based on noisy data and the Frisch scheme. *SIAM Review*, 57(2), 167–197.

Pardey, J., Roberts, S., and Tarassenko, L. (1996). A review of parametric modelling techniques for eeg analysis. *Medical engineering & physics*, 18(1), 2–11.

Pion-Tonachini, L., Kreutz-Delgado, K., and Makeig, S. (2019). Iclabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, 198, 181–197.

Ramaswamy, K., Bottegal, G., and Van den Hof, P. (2022). Learning linear modules in a dynamic network with missing node observations. *arXiv preprint arXiv:2208.10995*.

Renold, H., Chavarriaga, R., Gheorghe, L., and Millán, J.d.R. (2014). Eeg correlates of active visual search during simulated driving: An exploratory study. In *IEEE SMC*, 2815–2820.

Saeidi, M., Karwowski, W., Farahani, F.V., Fiok, K., Taiar, R., Hancock, P., and Al-Juaid, A. (2021). Neural decoding of eeg signals with machine learning: A systematic review. *Brain Sciences*, 11(11), 1525.

Snodgrass, J.G. and Vanderwart, M.L. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol. Hum. Learn.*, 6(2), 174–215.

Stoica, P., Moses, R.L., et al. (2005). *Spectral analysis of signals*. Prentice Hall, NJ.

Veedu, M.S., Doddi, H., and Salapaka, M.V. (2021). Topology learning of linear dynamical systems with latent nodes using matrix decomposition. *IEEE Trans. Automatic Control*, 67(11), 5746–5761.

You, J. and Yu, C. (2024). Sparse plus low-rank identification for dynamical latent-variable graphical AR models. *Automatica*, 159, 111405.

Zancanaro, A., Zoppis, I., Manzoni, S., Cisotto, G., et al. (2023). vEEGNet: A new deep learning model to classify and generate EEG. In *In Proc. ICT4AWE*, 245–252.

Zorzi, M. and Sepulchre, R. (2016). AR identification of latent-variable graphical models. *IEEE Trans. Automatic Control*, 61(9), 2327–2340.

Zorzi, M. (2019). Empirical Bayesian learning in AR graphical models. *Automatica*, 109.

Zorzi, M. and Chiuso, A. (2017). Sparse plus low rank network identification: A nonparametric approach. *Automatica*, 76, 355–366.