Bi-Smoothed Functional Independent Component Analysis for EEG Artifact Removal

Marc Vidal[1,2]*, Mattia Rosso[1], Ana M. Aguilera[2]*

[1] *Institute of Psychoacoustics and Electronic Music (IPEM), Ghent University, Belgium*
[2] *Department of Statistics and O.R. and IMAG, University of Granada, Spain*
* E-mail: aaguiler@ugr.es (A.M.A); marc.vidalbadia@ugent.be (M.V)

Abstract: Motivated by the challenge of mapping adverse artifactual events caused by body movements in electroencephalographic (EEG) signals, we present a functional independent component analysis based on the spectral decomposition of the kurtosis operator of a smoothed principal component expansion. A discrete roughness penalty is introduced into the orthonormality constraint of the covariance eigenfunctions to obtain a smoothed basis for the proposed independent component model. To select the tuning parameters, a cross-validation method incorporating shrinkage is used to enhance performance on functional representations with a large basis dimension. This method provides an estimation strategy to determine the penalty parameter and the optimal number of components. Our independent component approach is applied to real EEG data to estimate genuine brain potentials from a contaminated signal. As a result, it is possible to control high-frequency remnants of neural origin overlapping with artifactual sources to optimize their removal from the signal. An R package implementing our methods is available on CRAN.

Keywords: auditory-motor coupling task; EEG; functional data; motion artifacts; penalized splines; smoothed principal components.

1 Introduction

In the field of neurophysiology, electroencephalography (EEG) represents one of the few techniques providing a direct measure of bioelectrical brain activity, as oscillations in excitability of populations of cortical pyramidal cells [48] contribute to variations in the electrical potentials over the scalp. Oscillations are characterized by dominant intrinsic rhythms conventionally grouped into frequency bands, which are by now validated as markers of several neurocognitive phenomena [12]. However, despite the temporal resolution achievable with its high sampling rate, EEG is a technique that suffers from low signal-to-noise ratio. This is mainly due to the fact that the layers of tissue dividing the electrodes from the cortex act as a natural filter attenuating genuine brain activity, resulting in a combination of cortical and artifactual sources in the EEG signal. In addition, the dominant brain-related spectral features often overlap with artifactual activity in higher frequency bands [13], and particularly at lower frequencies most of the variance in the signal is explained by physiological sources outside the brain. For these reasons, analyzing EEG signals can ultimately be viewed as solving a source-separation problem with the goal of estimating brain potentials of interest.

Blind source separation techniques such as independent component analysis (ICA) are commonly used to address artifact detection and correction of EEG signals. The term ICA encompasses a broad scope of algorithms and theoretical rudiments aligned to the assumption of independence of the latent sources in the data. From the statistical perspective, it could be regarded as a refinement of principal component analysis (PCA) that goes beyond the variance patterns of the data, introducing high-order measures such as kurtosis or negentropy. This way, the data can be approximately represented in terms of a small set of independent variables, while in the PCA reduction, these variables are only assumed to be uncorrelated. An overview of statistical methodologies for ICA is provided in [30]. A comprehensive monograph of the subject can be found in [21].

The use of sampling units in form of functions that evolve on a continuum, rather than through vectors of measurements, has been popularized over the last two decades to solve a broad class of problems. Functional data analysis provides a natural generalization for a wide variety of statistical techniques that take advantage of the complete functional form of data by including relevant information related to smoothness and differentiability (see [20, 38, 47] for a systematic review of the topic). The extension of ICA to functional data has, however, not yet received the attention nor the prolific developments of other reduction techniques in this framework, such as functional principal component analysis (FPCA). A first attempt to develop an extension of the classic multivariate ICA model was investigated in [28] by exploiting the functional principal component decomposition. Functional ICA techniques were also implemented in [35], who defined the kurtosis operator and studied its discriminating properties under mixtures of Gaussian processes. More recently, [25, 46] developed a functional ICA model using an estimation procedure stemmed from the finite Karhunen-Loève (K-L) expansion [10, pp. 37], which is optimal in the least-squared error sense. We extend this model setup endowing the space with a new geometrical structure given by a Sobolev inner product in order to control the roughness of the K-L functions.

The use of functional data in brain imaging analysis has gained significant attention in recent years, despite the inherent complexity and computational demands involved in its processing. Electroencephalogram (EEG) data, for instance, provides a rich basis for applying a wide variety of functional data methods, ranging from the estimation of smoothed sample curves to advanced reduction and forecasting techniques (see, for example, [19, 29, 36, 41, 50]). Current research predominantly focuses on FPCA approaches to model artifact-free data. However, to the best of our knowledge, the efficiency of functional ICA techniques for handling data contaminated by physiological artifacts remains largely unexplored. In contrast, this issue has been extensively studied in the multivariate context; for instance, [44] provides a comprehensive comparison of state-of-the-art methods for artifact removal.

In this paper, we develop a methodology based on piecewise polynomial smoothing (B-splines) to separate and isolate neural activity from artifactual sources in EEG data. Given the transient nature and complex morphology of EEG signals, B-splines offer an effective alternative to represent the non-sinusoidal behavior of contaminated neural oscillations. The proposed functional ICA aims to improve brain signal estimation by removing smooth artifacts in a noise-free manner. Although wavelet-based and hybrid approaches have been shown to effectively denoise common artifacts (see, e.g., [8, 11, 13, 27]), our methodology uses a penalized spline (P-spline) approach [14, 16] for independent component estimation. P-splines

are computationally efficient and mathematically simpler. They have been successfully applied in dimension reduction [4] and the estimation of various functional regression models [2, 5–7], which supports their application in the proposed framework. One of the defining characteristics of our method is that the decomposition is regulated by the principal component eigendirections and further optimized using penalized estimators to ensure smoothness. For this reason, the proposed functional ICA can be conceived as a bi-smoothed estimation procedure. The end-user will finally appreciate how artifact extraction can be fine-tuned by regulating a single smoothing parameter, making it intuitive to improve the results through a visual inspection of the independent component scores.

The paper is organized as follows. We introduce our model in Section 2 and develop the smoothed FICA decomposition using basis expansion representations of functional data in Section 3. A method for selecting the tuning parameters is discussed in Section 4. To test the effectiveness of our model in recovering brain signals, Section 5 provides a simulation using real EEG data on single trial designs containing stereotyped artifacts. Section 6 shows how our smoothed FICA works in the context of event-related potentials designs. Finally, we conclude with a brief discussion in Section 7. The presented P-spline smoothed FICA is implemented in the R package *pfica* [45].

2 Smoothed functional independent component analysis

2.1 Preliminaries

Let $y_i = (y_{i1},\dots,y_{im_i})^{\mathrm{T}}$ $(i = 1,\dots,n)$ be a signal of n components digitized at the sampling points t_{ik} $(k = 1,\dots,m_i)$. Consider that the sample data is observed with error, so that it can be modeled as

$$y_{ik} = x_i(t_{ik}) + \varepsilon_{ik}, \qquad (1)$$

where x_i is the ith functional trajectory of the signal and ε_{ik} mutually independent measurement errors with zero means. The sample functions $x_1,\dots,x_n$ are assumed to be realizations of a random functional variable X in $L^2(T)$, the separable Hilbert space of square integrable functions from T to $\mathbb{R}$, endowed with the usual inner product $\langle f,g \rangle = \int_T f(t)g(t)\mathrm{d}t$, and the induced norm $\|f\| = \langle f,f \rangle^{1/2}$. Thorough the text, X is assumed to have zero mean and finite fourth moments, ensuring that moment operators are well-defined up to the fourth order. The sample covariance operator $\mathscr{C}_x$ is an integral operator with kernel $c(s,t) = n^{-1}\sum_{i=1}^n x_i(s)x_i(t)$; $s,t \in T$, admitting the Mercer decomposition

$$c(s,t) = \sum_{j=1}^{\infty} \eta_j \gamma_j(s)\gamma_j(t),$$

where $(\eta_j,\gamma_j)_j$ are, respectively, a positive sequence of eigenvalues in descending order and their associated orthonormal eigenfunctions. The functions $x_i(t)$ can be approximately represented by a truncated series of the K-L expansion

$$x_i^q(t) = \sum_{j=1}^{q} z_{ij}\gamma_j(t), \qquad (2)$$

where $z_{ij} = \langle x_i,\gamma_j \rangle$ are zero mean scores with $\mathrm{var}_n(z_j) = \eta_j$ and $\mathrm{cov}_n(z_j,z_{j'}) = 0$ for $j \neq j'$. These scores, referred to as the principal components, represent linear combinations of the functional random variable with maximum variance. Moreover, if the q term in (2) is optimally selected, the mean squared error is minimized, providing the best linear approximation to the original data [17, pp. 21]. Recently, a functional Varimax rotation has been introduced to enhance the interpretability of the most explanatory principal component scores [1].

2.2 Functional ICA of a smoothed principal component expansion

The notion of independent components of a random vector cannot be immediately extended to the case of Hilbert-valued random elements due to the fact that a probability density function is not generally defined in this context [15]. In the sequel, we consider the definition of independence introduced in [18], which establishes that a functional random variable has independent components if after projecting onto a finite subset of a given orthonormal basis of the space we obtain a set of independent random variables. Then, the aim of functional independent component analysis (FICA) is to find a linear operator Γ, such that for a given orthonormal basis, say, ϕ_j $(j = 1,\dots,q)$ in $L^2(T)$, the variables $\langle \Gamma X, \phi_j \rangle$ are mutually independent. By considering X as a random element induced by a Gaussian measure, a functional principal component analysis (FPCA) would suffice to obtain the independent components [10, pp. 40]. However, in more general settings the operator Γ is required to achieve the goals of the model. This raises the question of how to choose an appropriate basis for X, as the results of the model depend significantly on this choice. For instance, while FPCA provides a natural basis, alternative bases may better align with the independence assumption required for FICA.

In this paper, the sample x_i is approximated by a smoothed FPCA representation obtained by introducing an orthonormality constraint with respect to the weighted Sobolev inner product

$$\langle f,g \rangle_\lambda = \langle f,g \rangle + \lambda \langle Rf, Rg \rangle, \qquad (3)$$

where R is an operator with the action $Rf(t) = \mathrm{d}^2 f(t)/\mathrm{d}t^2, f \in \mathrm{dom}(R)$ that measures the roughness of the curves, and λ is a nonnegative penalty parameter. Notice that, when $\lambda = 0$, (3) is simplified to the usual inner product, meaning that x_i can be uniquely represented by the K-L basis, i.e. the eigenfunctions of $\mathscr{C}_x$. To estimate the smoothed principal components, Silverman [42] proposed the following variance maximization problem

$$\gamma_{\lambda,j} = \mathrm{argmax}\,\frac{\mathrm{var}(\langle \gamma,x \rangle)}{\|\gamma\|^2 + \lambda \langle R\gamma, R\gamma \rangle} = \max\frac{\langle \gamma, \mathscr{C}_x \gamma \rangle}{\|\gamma\|_\lambda^2}, \qquad (4)$$

subject to the constraint $\langle \gamma, \gamma_{\lambda,k} \rangle_\lambda = 0$ for all $k < j$, where γ is a function assumed in a closed linear subspace of $L^2(T)$ with square-integrable second derivatives. We emphasize that, the problem of finding $\gamma_{\lambda,j}$ depends on the sample size n plus the selection of the penalization parameter λ. In [37], the authors established the existence of the solutions of the optimization problem (4) for any $\lambda \geq 0$. Silverman [42] proved the consistency of the estimators as $n \to \infty$ and $\lambda \to 0$. Generalized consistency and asymptotic distributions of the estimators have been derived in [24], using expansions of the perturbed eigensystem of a sample smoothed covariance operator.

The functions $(\gamma_{\lambda,j})$ form a complete orthonormal system in the subspace endowed by $\langle \cdot,\cdot \rangle_\lambda$, making this basis non-compatible for our independent component model in $L^2(T)$. However, [34] generalized Silverman's method providing the following equivalents FPCA.

Proposition 1. *Given a sample x_i $(i = 1,\dots,n)$ of a functional variable with trajectories in $L^2(T)$, there exists a positive definite operator S^2 such that the following FPCA decompositions are equivalent:*

1. *The FPCA of $S^2(x_i)$ with respect to $\langle \cdot,\cdot \rangle_\lambda$, $S^2(x_i) = \sum_j z_{ij}\gamma_{\lambda,j}$.*
2. *The FPCA of $S(x_i)$ with respect to $\langle \cdot,\cdot \rangle$, $S(x_i) = \sum_j z_{ij}S^{-1}(\gamma_{\lambda,j})$.*
3. *The FPCA of X with respect to $\langle \cdot,\cdot \rangle_S$, $x_i = \sum_j z_{ij}S^{-2}(\gamma_{\lambda,j})$, with $\langle f,g \rangle_S = \langle S(f),S(g) \rangle = \langle S^2(f),S^2(g) \rangle_\lambda$.*

Therefore, the eigenfunctions of the covariance operator $\mathscr{C}_{Sx} = S\mathscr{C}_x S$ of the smoothed sample $S(x_i)$ are given by $\beta_j = S^{-1}(\gamma_{\lambda,j})$, where $\gamma_{\lambda,j}$ are obtained by the penalized estimation procedure set out for (4). Then, the basis β_j is orthonormal with respect to the

usual inner product in $L^2(T)$, so that the smooth sample data $S(x_i)$ can be approximated by its truncated K-L expansion

$$\chi_i^q(t) = \sum_{j=1}^{q} z_{ij}\beta_j, \tag{5}$$

where $z_{ij} = \langle \beta_j, S(x_i) \rangle = \langle \gamma_{\lambda,j}, x_i \rangle$, and $\chi_i^q(t)$ denotes a q-dimensional orthonormal representation of the smoothed sample data $S(x_i)$ in $L^2(T)$. The functional ICA version proposed in this paper uses the elements of this expansion to estimate the independent components of the original data.

Our main assumption is that the target functions lie within the space spanned by the first q eigenfunctions of the operator $\mathscr{C}_{Sx}$, which possesses a smooth second-order structure represented by the primary modes of variation in the empirical data. Thus, in such eigensubspace, it is expected to gain some accuracy in the forthcoming results due to the attenuation of the higher oscillation modes corresponding to the small eigenvalues of $\mathscr{C}_{Sx}$. Henceforth, we denote by $\mathscr{M}^q = \text{span}(\beta_1, \dots, \beta_q)$ the subspace spanned by the q first eigenfunctions of $\mathscr{C}_{Sx}$. $\mathscr{M}^q$ is assumed to preserve the inner product in $L^2(T)$.

Most of the multivariate ICA methods require the standardization of the observed data with the inverse square root of the covariance matrix in order to remove any linear dependencies and normalize the variance along its dimensions. In infinite-dimensional spaces, however, covariance operators are not generally invertible. As long as our signal is represented in $\mathscr{M}^q$, no regularization is strictly needed and under moderate conditions, the inverse of the covariance operator and its square root can be well defined. Let $\tilde{\chi}^q = \mathscr{C}_{\chi^q}^{-1/2}(\chi^q)$ denote a standardized functional sample, therefore a sample whose covariance operator naturally satisfies to be the identity of the space.

As an extension of the multivariate case, the sample kurtosis operator of the standardized data is usually defined as

$$\mathscr{K}_{\tilde{\chi}^q}(h)(s) = \frac{1}{n} \sum_{i=1}^{n} \langle \tilde{\chi}_i^q, \tilde{\chi}_i^q \rangle \langle \tilde{\chi}_i^q, h \rangle \tilde{\chi}_i^q(s) = \langle k(s, \cdot), h \rangle, \tag{6}$$

where $k(s,t) = n^{-1} \sum_{i=1}^{n} \|\tilde{\chi}_i^q\|^2 \tilde{\chi}_i^q(s) \tilde{\chi}_i^q(t)$ denotes the kurtosis kernel function of $\tilde{\chi}^q$, and h the function in $\mathscr{M}^q$ to be transformed. We assume that the kurtosis operator is positive-definite, self-adjoint and unitary equivariant (see [25]). Since (6) is trace-class by finite-dimensional dependency, its kernel admits the Mercer eigendecomposition

$$k(s,t) = \sum_{l=1}^{q} \rho_l \psi_l(s) \psi_l(t),$$

where $(\rho_l, \psi_l)_{l=1}^{q}$ are a positive sequence of eigenvalues and related eigenfuntions, respectively. With this, we can define the independent components of χ_i^q as mutually independent variables with maximum kurtosis given by

$$\zeta_{il,\tilde{\chi}^q} = \langle \tilde{\chi}_i^q, \psi_l \rangle.$$

Challenging questions arise on how the Karhunen-Loève Theorem might be applied in this context. Intuitively, we note that this procedure leads to the expansion $\tilde{\chi}_i^q(t) = \sum_{l=1}^{q} \zeta_{il,\tilde{\chi}^q} \psi_l(t)$ which can be approximated in terms of r eigenfuntions ψ_l of interest, e.g. those associated with the independent components with extreme kurtosis values. One can also consider $\{\psi_1, \dots, \psi_q\}$ as a basis of projection for either x, χ^q or $\tilde{\chi}^q$, in view of the fact that it preserves the four-order structure of the standardized data.

3 Basis expansion estimation using a P-spline penalty

In order to estimate the independent components from noisy discrete observations in (1), it will be assumed that the trajectories belong to a finite-dimensional space of $L^2(T)$ spanned by a set of B-spline

basis functions $\{\phi_1(t), \dots, \phi_p(t)\}$. Then, each sample curve can be expanded as

$$x_i(t) = \sum_{j=1}^{p} a_{ij}\phi_j(t), \tag{7}$$

or, in matrix form, $x = A\phi$, where A is a coefficient matrix $A = (a_{ij}) \in \mathbb{R}^{n \times p}$ and $\phi = (\phi_1, \dots, \phi_p)^{\mathrm{T}}$, $x = (x_1, \dots, x_n)^{\mathrm{T}}$ denote vector-valued functions. The basis coefficients for each sample curve can be found by least squares approximation minimizing the mean squared error

$$\text{MSE}(a_i \mid x_i) = (x_i - \Phi_i a_i)^{\mathrm{T}}(x_i - \Phi_i a_i),$$

where $\Phi_i = \{\phi_j(t_{ik})\} \in \mathbb{R}^{m_i \times p}$ and $a_i = (a_{i1}, \dots, a_{ip})^{\mathrm{T}}$. For general guidance on both definition knots and order of B-splines, we refer the reader to [38]. Although in this paper a non-penalized least squares approximation is assumed, [3] give a detailed account of how to estimate the basis coefficients using different roughness penalty approaches (continuous and discrete) in terms of B-splines.

The next step consists of smoothing the sample curves in terms of the smoothed principal components and associated weight functions β_j in (5). To do so, we next derive the P-spline FPCA approach developed in [4] that incorporates a discrete penalty based on d-order differences of adjacent B-spline coefficients (P-spline penalty) in the orthonormality constraint. Let us consider the B-spline basis expansion of the covariance eigenfunctions $\gamma(t) = \phi(t)^{\mathrm{T}}b$, with $b = (b_1, \dots, b_p)^{\mathrm{T}}$ being its vector of basis coefficients, and a discrete P-spline roughness penalty function defined by $\text{PEN}_d(\gamma) = b^{\mathrm{T}}P_db$, where $P_d \in \mathbb{R}^{p \times p}$ is the penalty matrix $P_d = \Delta_d^{\mathrm{T}}\Delta_d$, with Δ_d being a matrix representation of the d-order difference operator R. Throughout the paper, we assume second order differences defining the penalty function $b^{\mathrm{T}}P_2b = (b_1 - 2b_2 + b_3)^2 + \dots + (b_{p-2} - 2b_{p-1} + b_p)^2$. This way, the inner product in (3) is given in terms of B-splines expansions as

$$\langle f, g \rangle_\lambda = \mathbf{f}^{\mathrm{T}}\mathscr{G}\mathbf{g} + \lambda \mathbf{f}^{\mathrm{T}}P_2\mathbf{g},$$

with $f = \phi^{\mathrm{T}}\mathbf{f}$, $g = \phi^{\mathrm{T}}\mathbf{g}$, and $\mathscr{G} = (\langle \phi_j, \phi_{j'} \rangle), (j, j' = 1, \dots, p)$. Then, the maximization problem in (4) is equivalent to solve the following matrix problem:

$$b_{\lambda,j} = \text{argmax} \frac{b^{\mathrm{T}}\mathscr{G}\Sigma_A\mathscr{G}b}{b^{\mathrm{T}}(\mathscr{G} + \lambda P_2)b}, \tag{8}$$

subject to the constraint $b^{\mathrm{T}}(\mathscr{G} + \lambda P_2)b_{\lambda,k} = 0$ for all $k < j$, where $\Sigma_A = n^{-1}A^{\mathrm{T}}A$ and $\lambda \geq 0$ is the penalty parameter used to control the trade-off between maximizing the sample variance and the strength of the penalty.

Because B-spline basis are non-orthonormal with respect to the usual L^2 geometry, we can apply Cholesky factorization of the form $LL^{\mathrm{T}} = \mathscr{G} + \lambda P_2$ to find a non-singular matrix that allows us to operate in terms of the B-spline geometrical structure. Then, finding the weight coefficients corresponds to solve the eigenvalue problem

$$L^{-1}\mathscr{G}\Sigma_A\mathscr{G}(L^{-1})^{\mathrm{T}}v_j = \eta_j v_j, \tag{9}$$

where $v_j = L^{\mathrm{T}}b_{\lambda,j}$ and the coefficients of $\gamma_{\lambda,j}$ are $b_{\lambda,j} = (L^{-1})^{\mathrm{T}}v_j$. Therefore, we have obtained a set of orthonormal functions with respect to the inner product $\langle \cdot, \cdot \rangle_\lambda$. The jth smoothed principal component is then given by

$$z_j = A\mathscr{G}b_{\lambda,j} = A\mathscr{G}(L^{-1})^{\mathrm{T}}v_j.$$

Thus, the problem is reduced to the multivariate PCA of the matrix $A\mathscr{G}(L^{-1})^{\mathrm{T}}$ in $\mathbb{R}^q$ (see [4] for a detailed study). From the results in [33, 34] we deduce in this paper the expression of the smoothing operator S that provides the equivalence between this multivariate PCA and the FPCA of the smoothed data $S(x_i)$ in $L^2(T)$.

Proposition 2. *Given the basis expansion (7) for a random sample x_i $(i = 1, \dots, n)$ of curves in $L^2(T)$, the PCA of the matrix $A\mathscr{G}(L^{-1})^{\mathrm{T}}$ with the usual inner product in $\mathbb{R}^p$ is equivalent to all*

FPCA in Proposition 1 *with the operator* S^2 *defined as* $S^2(f) = \phi(t)^{\mathrm{T}}(\mathscr{G} + \lambda P_d)^{-1}\mathscr{G}\mathbf{f}$, *with* $f = \phi(t)^{\mathrm{T}}\mathbf{f}$.

Proof: Define, for all $f = \phi(t)^{\mathrm{T}}\mathbf{f}$, $g = \phi(t)^{\mathrm{T}}\mathbf{g}$, the new inner product $\langle f, g \rangle_K = \mathbf{f}^{\mathrm{T}}K\mathbf{g}$ where $K = D^{\mathrm{T}}D$, with $D = L^{-1}\mathscr{G}^{\mathrm{T}}$. Proposition 2 in [33] proved that the PCA of matrix AD^{T} with the usual inner product in $\mathbb{R}^p$ is equivalent to FPCA of x_i with respect to $\langle \cdot, \cdot \rangle_K$. That is, $x_i = \sum_j z_{ij}f_j$ with $f_j = \phi^{\mathrm{T}}D^{-1}v_j$, with v_j being the eigenvectors of the matrix AD^{T}. Then, from Proposition 1 in this paper, we have that $\langle S^2(f), S^2(g) \rangle_\lambda = \langle f, g \rangle_K$. If we suppose that there exists a matrix C such that $S^2(f) = \phi^{\mathrm{T}}C\mathbf{f}$, then $\langle S^2(f), S^2(g) \rangle_\lambda = \mathbf{f}^{\mathrm{T}}C^{\mathrm{T}}(\mathscr{G} + \lambda P_d)C\mathbf{g} = \mathbf{f}^{\mathrm{T}}D^{\mathrm{T}}D\mathbf{g}$. As a consequence, $C^{\mathrm{T}}LL^{\mathrm{T}}C = D^{\mathrm{T}}D$, so that $L^{\mathrm{T}}C = RD$ with R being an orthonormal matrix ($RR^{\mathrm{T}} = I_p$). Therefore, $S^2(f) = \phi^{\mathrm{T}}\{(L^{-1})^{\mathrm{T}}RD\}\mathbf{f}$. On the other hand, from Proposition 1, we have that $\gamma_j = S^2(f_j)$ which implies that $(L^{-1})^{\mathrm{T}}v_j = (L^{-1})^{\mathrm{T}}RDD^{-1}v_j$. As a consequence we obtain that $R = I_p$ and $S^2(f) = \phi^{\mathrm{T}}\{(L^{-1})^{\mathrm{T}}D\}\mathbf{f} = \phi^{\mathrm{T}}\{(\mathscr{G} + \lambda P_d)^{-1}\mathscr{G}\}\mathbf{f}$. □

As a result, the principal components (scores) of $S(x_i)$ are given by $Z = A\mathscr{G}(L^{-1})^{\mathrm{T}}V$ where V is the matrix whose columns are the eigenvectors v_j verifying Equation (9), and thus the eigenfunctions β_j are $\beta_j = S^{-1}(\gamma_{\lambda,\cdot j})$.

Having estimated the weight functions coefficients and principal components scores, consider the smooth principal component expansion in (5) is truncated at the q-term. Then, the column vector of smoothed sample curves is given by $\chi^q(t) = Z^q\beta(t)$, where $Z^q = (z_{ij}) \in \mathbb{R}^{n \times q}$ is the matrix whose columns are the first q principal components scores with respect to the basis of smooth principal component weight functions $\beta(t) = (\beta_1(t), \ldots, \beta_q(t))^{\mathrm{T}}$.

With the above results, the functional independent components are computed from the smoothed principal component approximation of functional data. Following the ICA pre-processing steps, we first standardize the approximated curves defining the whitening operator as $\mathscr{C}_{x^q}^{-1/2}\{\chi^q(t)\} = \tilde{\chi}^q(t) = \tilde{Z}^q\beta(t)$, with $\tilde{Z}^q = Z^q\Sigma_{Z^q}^{-1/2}$ being the matrix of standardized principal components and $\Sigma_{Z^q}^{-1/2} = \sqrt{n}\{(Z^q)^{\mathrm{T}}Z^q\}^{-1/2}$, the inverse square root of the covariance matrix of Z^q. The described whitening transformation is essentially an orthogonalization of the coefficient part of χ^q, so the matrix $\tilde{Z}^q \in \mathbb{R}^{n \times q}$ naturally satisfy $\Sigma_{\tilde{Z}^q} = I_q$.

Then, the kurtosis operator (6) of the standardized curves $\tilde{\chi}^q(t)$ is given in matrix form by

$$\mathscr{K}_{\tilde{\chi}^q}(h) = \frac{1}{n}(\tilde{Z}^{q^{\mathrm{T}}}D_{\tilde{Z}^q}\tilde{Z}^q\mathbf{h})^{\mathrm{T}}\beta(t), \ \ \forall h = \beta(t)^{\mathrm{T}}\mathbf{h},$$

where $D_{\tilde{Z}^q} = \mathrm{diag}(\tilde{Z}^q\tilde{Z}^{q^{\mathrm{T}}})$. The eigenanalysis of this kurtosis operator leads to the diagonalization of the kurtosis matrix of the standardized principal components $\tilde{Z}^q$,

$$\Sigma_{4,\tilde{Z}^q}u_l = \rho_l u_l \quad (l = 1, \ldots, q), \tag{10}$$

where $\Sigma_{4,\tilde{Z}^q} \in \mathbb{R}^{q \times q}$ is defined as

$$\Sigma_{4,\tilde{Z}^q} = \frac{1}{n}\sum_{i=1}^{n}\|\tilde{z}_i^q\|^2\tilde{z}_i^q\tilde{z}_i^{q^{\mathrm{T}}} = \frac{1}{n}\tilde{Z}^{q^{\mathrm{T}}}D_{\tilde{Z}^q}\tilde{Z}^q,$$

with $\tilde{z}_i^q$ being the column vector $q \times 1$ with the ith row of the matrix $\tilde{Z}^q$. The eigenproblem (10) is not restricting to assume that $\Sigma_{4,\tilde{Z}^q}$ is uniquely determined. In fact, other kurtosis matrices can be considered (see, e.g., [23, 26]). This way, the P-spline smoothed functional ICA of x in $L^2(T)$ is obtained from the multivariate ICA of Z^q in $\mathbb{R}^q$. The resulting weight functions are now $\psi_l(t) = \beta(t)^{\mathrm{T}}u_l$ ($l = 1, \ldots, q$), where the coefficients vectors u_l are the eigenvectors of the predefined kurtosis matrix. Then, the independent components can be calculated as $\zeta_{l,\tilde{\chi}^q} = \tilde{Z}^q u_l$. Finally, the operator Γ defining

the FICA model is

$$\Gamma(\chi_i^q) = \beta^{\mathrm{T}}U^{\mathrm{T}}\Sigma_{Z^q}^{-1/2}z_i^q,$$

with z_i^q being the column vector $q \times 1$ with the ith row of Z^q and $U \in \mathbb{R}^{q \times q}$ the matrix of eigenvectors of the kurtosis matrix $\Sigma_{4,\tilde{Z}^q}$.

4 Parameter tuning

The problem concerning the estimation of the smoothed independent component curves lies in finding an optimal truncation point q, as well as a suitable penalty parameter. As q approaches p, more of the higher oscillation modes of the sample are induced in the estimation. Otherwise, we are denoising the data from its second and fourth-order structure simultaneously. From this perspective, it is desirable to increase the value of q such that the interesting latent functions of the whitened space can be captured by the kurtosis operator. Observe that this kind of regularization is not exactly the same as the one providing the P-spline penalization of the roughness of the weight functions. Thus, if the original curves are observed with independent error, and the error is persistent in the functional approximation, it may overlap the estimation of the kurtosis eigenfunctions. In such cases, smoothing would be appropriate to mitigate the influence of noise. Once the value of q is selected, components with extreme kurtosis can be examined, unlike in FPCA, where only components associated with large eigenvalues are typically considered.

4.1 Penalty parameter selection

Leave-one-out cross-validation [39] is generally used to select the penalty parameter in order to achieve a suitable degree of smoothness on the weight functions, but also to induce the truncation point q. In a more explicit and condensed form, this procedure in our model lies in finding a value of λ that minimizes

$$\mathrm{CV}_q(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left\|x_i - \chi_i^{q(-i)}\right\|^2, \tag{11}$$

where $\chi_{i,}^{q(-i)} = \sum_{l=1}^{q}z_{il}^{(-i)}\beta_l^{(-i)}(t)$ is the reconstruction of the ith curve x_i in terms of the q first smoothed principal components by leaving out it in the estimation process. We found, however, that cross-validation was not sensitive for a reasonably large basis dimension, forcing us to reformulate the strategy. To address this problem, the penalty parameter might be subjectively chosen although this can lead to the bias and poor extraction of the artifactual sources. Hence, for the results presented in this paper, we propose a novel adaptive approach which consists in replacing (11) by

$$\mathrm{BCV}_q(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left\|\chi_i^{q;\lambda(-i)} - \chi_i^{q;\lambda+\ell(-i)}\right\|^2, \tag{12}$$

where $\chi_i^{q;\lambda(-i)}$ is a smoothed representation of x_i for some λ and $\ell > 0$ a value that increases the penalty in the second term of the norm, assume $\ell = 0.1$. Then, for a fixed q, (12) is iterated for each λ in a given grid to find the one that minimizes $\mathrm{BCV}_q(\lambda)$. Among all the q considered in the estimation process, we select the truncation point that minimizes this function.

If we require a basis dimension p greater than sample size n, a shrinkage covariance estimator [40] can be considered for computing Σ_A. This method guarantees positive definiteness and consequently an estimation of the higher and important eigenvalues not biased upwards. The same strategy is used for $\mathrm{BCV}_q(\lambda)$. Recall the quadratic distances in (12). These are given in terms of basis functions by

$$\left\| \chi_i^{q;\lambda(-i)} - \chi_i^{q;\lambda+\ell(-i)} \right\|^2 = \int_T \left[\chi_i^{q;\lambda(-i)}(t) - \chi_i^{q;\lambda+\ell(-i)}(t) \right]^2 dt =$$

$$= \int_T \left[\sum_{l=1}^q z_{il}^{\lambda(-i)} \sum_{j=1}^p b_{lj}^{\lambda(-i)} \phi_j(t) - \sum_{l=1}^q z_{il}^{\lambda+\ell(-i)} \sum_{j=1}^p b_{lj}^{\lambda+\ell(-i)} \phi_j(t) \right]^2 dt =$$

$$= \int_T \left[\sum_{j=1}^p e_{ij} \phi_j(t) \right]^2 dt = e_i^T \mathscr{G} e_i,$$

where $b_j = (b_{j1}, \ldots, b_{jp})^T$ is the vector of basis coefficients of the jth weight function β_j in the B-spline basis $\phi_j(t)$ and $e_i = (e_{i1}, \ldots, e_{ip})^T$ is a vector of residuals. Next, the matrix $\mathscr{E} = (e_{ij}) \in \mathbb{R}^{n \times q}$ is reconstructed via shrinkage. That is, first we compute $\text{cov}_S(\mathscr{E})$ where cov_S is a predefined shrinkage covariance estimator, then we apply Cholesky decomposition of the form $LL^T = \text{cov}_S(\mathscr{E})$. Finally, the basis coefficients of the reconstructed residual functions are $\hat{e}_i = (L^{-1})^T e_i$, and consequently now

$$\text{BCV}(\lambda)_q = \frac{1}{n} \sum_{i=1}^n \left\| \chi_i^{q;\lambda(-i)} - \chi_i^{q;\lambda+\ell(-i)} \right\|^2 = \hat{e}_i^T \mathscr{G} \hat{e}_i.$$

We call this method *baseline cross-validation*, as it operates across different reconstructions of x_i for a given baseline penalty parameter and a fixed q. This approach is more versatile and particularly useful when the original curves are extremely rough and approximated with a larger basis dimension, thus avoiding the least squares to collapse. Moreover, for a given q, it allows to score more than one λ as a result of the various relative minima it produces. The intuition behind baseline cross-validation is that there are several smoothing levels to endow the estimator with the ability for predictive modelling. These are given at evaluating "short distances" for a smoothing baseline λ in a given χ_i^q, which may be seen as a way of finding a trade-off for the global roughness of a q-dimensional basis. Note that, as the value of q increases, and despite the minimization of the mean squared error, it may be more difficult to find a smoothing balance between the elements of the basis due to a complex fabric of variability modes.

Algorithm BASELINE CROSS-VALIDATION

Input: A, ϕ_j $(j = 1, \ldots, p), \mathscr{G}, P_2, \lambda_k = (\lambda_1, \ldots, \lambda_m)^T$.
Output: $\lambda^\bullet$.
for each λ in λ_k:
1: Calculate L^{-1} via Cholesky decomposition of the matrix $\mathscr{G} + \lambda P_2 = LL^T$ and for $\mathscr{G} + (\lambda + \ell)P_2 = LL^T$.
2: Diagonalize $L^{-1}\mathscr{G}\Sigma_{A_s}\mathscr{G}(L^{-1})^T$, where $\Sigma_{A_s} = \text{cov}_S(A)$, to obtain the coefficients of the eigenfunctions β_j, b_j and $b_{\ell,j}$ for the incremental smoothing case .
3: Calculate $Z^q = A^T \mathscr{G} b_j$, $Z_\ell^q = A^T \mathscr{G} b_{\ell,j}$ and $\mathscr{A} = b_j(Z^q)^T$, $\mathscr{A}_\ell = b_{\ell,j}(Z_\ell^q)^T$, where $\mathscr{A}, \mathscr{A}_\ell$ are the coefficient matrices of the smoothed principal component expansion in terms of ϕ_j.
4: $\mathscr{E} = \mathscr{A} - \mathscr{A}_\ell$ and reconstruct $\mathscr{E}$ via the covariance matrix $\text{cov}_S(\mathscr{E})$.
5: $\text{BCV}(\lambda) = n^{-1}\text{tr}(\hat{\mathscr{E}}^T \mathscr{G} \hat{\mathscr{E}})$, where $\hat{\mathscr{E}}$ is the reconstructed matrix of residual coefficients and $\text{tr}(\cdot)$ is an operator that sums the diagonal elements of a square matrix.
end for
$\lambda^\bullet \leftarrow \text{argmin}_\lambda \text{BCV}$.

---

5 Simulation study

A simulation study based on EEG data segments containing stereotyped artifacts was conducted to validate our methods for recovering brain sources. The data consists of 4 separate 64-channel recordings of a subject performing the following classes of self-paced repetitive movements: nodding, hand-tapping with a wide arm movement, eye-blinking and chewing. Recordings were performed in absence of sensory stimulation in a trial length 3 seconds sampled at 1 kHz, i.e., t_{ik} $(i = 1, \ldots, 64; k = 1, \ldots, 3000)$. More details on the preprocessing

steps and experimental conditions are deferred to the online supplementary material. In reconstructing the functional form of the sample paths, we sought a less smooth fitting to mimic the brain potential fluctuations. Accordingly, a basis of cubic B-spline functions of dimension $p = 230$ is fitted to all signal components minimizing the mean squared error to a negligible value.

The process of identifying artifactual functions is addressed by using topographic maps that roughly represent patterns of eigenactivity related to the distribution of bioelectric energy on the scalp. These maps are elaborated from the projection of the signal components $x_1, \ldots, x_{64}$ on to the basis of independent weight functions, i.e. $\zeta_{il,x} = \langle x_i, \psi_l \rangle$ $(i = 1, \ldots, 64; l = 1, \ldots, q)$, whose resulting score vectors $\zeta_{l,\chi} = (\zeta_{1l}, \ldots, \zeta_{nl})^T$ are depicted in the spatial electrode domain. Therefore, the aim is to examine how the kurtosis eigenfunctions contribute into x_i to discern possible patterns of artifactual activity. The components identified as artifacts will be considered for subtraction.

In order to simplify the burden of a manual selection, assume that all $\psi_1, \ldots, \psi_q$ obtained from the model correspond to a structure of latent artifactual eigenpatterns. Moreover, let $\chi_i^q(t) = \sum_{l=1}^q \zeta_{il,x} \psi_l(t)$ be an expansion of artifactual components and related artifactual eigenfunctions. Then, the artifact subtraction in terms of basis expansions is

$$x_i(t) - \chi_i^q(t) = \sum_{j=1}^p a_{ij} \phi_j(t) - \sum_{l=1}^q \zeta_{il,x} \sum_{j=1}^p (u_l^T b_j) \phi_j(t) =$$

$$= \sum_{j=1}^p d_{ij} \phi_j(t), \tag{13}$$

where d_{ij} are the cleaned (or residual) coefficients, with u_l being the vector of coefficients of the independent weight function ψ_l in terms of the principal eigenfunctions. Thus, given the model parameters q and λ, the procedure to estimate and remove smooth artifactual components from EEG functional data can succinctly be derived as follows:

Algorithm FUNCTIONAL ARTIFACT SUBTRACTION

Input: A, ϕ_j $(j = 1, \ldots, p), \mathscr{G}, P_2, \lambda, q$.
Output: d_j.
1: Calculate L^{-1} via Cholesky decomposition of the matrix $\mathscr{G} + \lambda P_2 = LL^T$.
2: Perform the PCA of $A\mathscr{G}(L^{-1})^T$. Obtain Z^q and the coefficients b_j of β_j.
   $\rightarrow$ if $p > n$ then diagonalize $L^{-1}\mathscr{G}\Sigma_{A_s}\mathscr{G}(L^{-1})^T$, where $\Sigma_{A_s} = \text{cov}_S(A)$.
3: Whiten Z^q: i.e. $\tilde{Z}^q = Z^q \Sigma_{Z^q}^{-1/2}$.
4: Fix a fourth-order matrix $\Sigma_{4,\tilde{Z}^q}$ and diagonalize it. Obtain the eigenvalues ρ_l and associated eigenvectors u_l $(l = 1, \ldots, q)$.
5: Calculate $\zeta_{il,x} = \langle x_i, \psi_l \rangle$ for $\psi_l(t) = \sum_{j=1}^q u_{lj}\beta_j(t)$.
6: Select the artifactual score vectors in $\zeta_{l,x}$. Expand the artifactual space as $\chi_i^q(t) = \sum_{l=1}^q \zeta_{il,x}\psi_l(t)$.
7: Subtract the artifactual coefficients in terms of ϕ_j using (13) and obtain the vector of coefficients d_j to reconstruct the functional brain signal.

Baseline cross-validation was performed on a given grid, selecting the value which minimizes $\text{BCV}_q(\lambda)$ for $q = 1, \ldots, j_0$ where j_0 is defined as the index entry corresponding to the first relative maximum of the first order differences of FPCA's eigenvalues $\Delta \eta_j$. We find that truncating at $q = j_0$ is a way of exploring independence in the high variability structure of the data. In analysing EEG signals, this entails major effectiveness at reducing the artifactual content to a few eigenfunctions, particularly for the low-frequency physiological activity such as blinks and movement-related artifacts. One may see this truncation rule as a measure to improve the accuracy in the estimation of certain artifacts, while preserving the modes of variability related to the rhythms of the latent brain processes. The log-distances using $\text{BCV}(\lambda)$ for each one of the datasets are shown in Figure 1. Further results are presented in Table 1.

Preliminary results comparing both penalized and non-penalized estimation show that the smoothed FICA presumably attenuates the high-frequency potentials of neural origin, revealing the latent shape
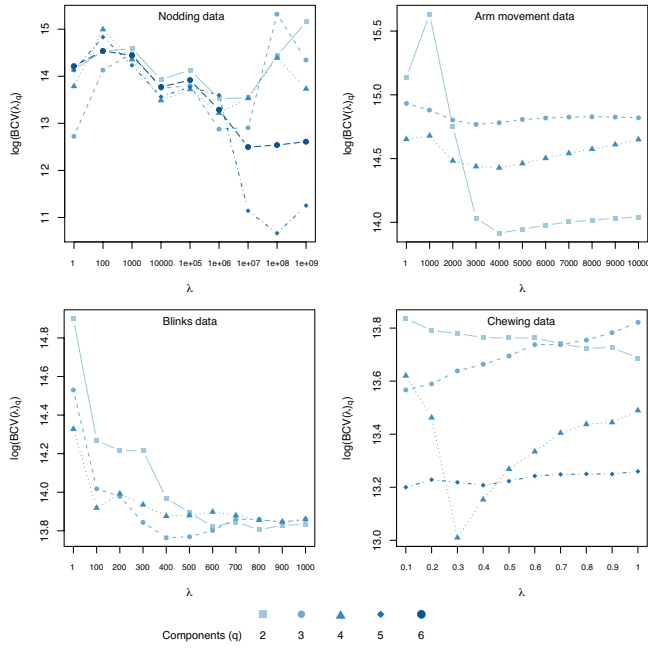
Fig. 1: The estimated log-BCV(λ) function for the first components of each EEG dataset containing different classes of artifacts.

Table 1 Summary of parameters and cumulative variance of the FICA model.

Trial	j_0	q	λ	log-BCV(λ)	var (%) λ	var (%) $\lambda = 0$
Nodding	6	5	10^8	10.66	99.40	94.43
Arm mov.	4	2	4000	13.91	75.85	62.42
Blinks	4	3	400.0	13.76	97.50	93.56
Chewing	5	4	0.300	13.01	68.23	68.03

of the artifact. More importantly, however, is that all topographic maps reflect well-known spatial activation of the artifactual content. A selection of eigenfunctions from each trial and their associated component scores are depicted in Figure 2. Physiological non-brain activity near the recording zone, such as blinks and large amplitude body movements, can be easily detected in controlled conditions using the proposed methodology. However, the coexistence of such artifacts may result in a non-linear distortion of them, e.g., via large changes of the impedance [31]. This could entail a more challenging situation, as algorithms based on linear mixing may not be that effective at a certain point. Nonetheless, the aforementioned artifacts enhance the role of smoothing due to their low-frequency trademark in the signal. In contrast, when artifacts are characterised by localised high-amplitude curves, as it is the case of the fourth artifactual eigenfunction (chewing), smoothing is not able to denoise effectively. We believe this happens for two reasons: first, the noise provided by the fourth-order structure of the model is essential to configure the shape of the artifact; second, the B-spline basis has a limited flexibility to smooth abrupt local contours. Hence, artifacts such as jaw clenching and chewing are quite sensible to smoothing and difficult to correct for subtraction. Interestingly, hybrid procedures combining spline interpolation and wavelet filtering have shown promising results trying to solve this problem in functional near-infrared spectroscopy research (see [32]).

It seems reasonable to conjecture that restricting q to the first FPCA terms decreases the odds of getting spurious artifactual functions, as they represent dominant modes of variability usually related to large artifacts. In such cases, the artifact subtraction with the smoothed components preserved the brain activity rhythms in the original form, while for $\lambda = 0$ it caused a reduction and a distortion of relevant potentials. However, BCV may tend to oversmooth slightly in a sense of an effective artifact removal, resulting in certain

artifactual residue after subtraction. This happens due to the complexity of the mixed sources and can be solved by examining other relative minima in our results. The plots for all channels and datasets comparing the effect of subtracting artifactual components are omitted for the sake of space. Online supplementary materials provide R code for its visualization.

Although our tests have provided good results by subtracting all smoothed components, further research is needed to corroborate their physiological validity. As reported in [9], reducing the dimensionality of the data with a PCA before applying ICA is not always beneficial, although in some cases may improve the signal-to-noise ratio of the large sources and their subsequent isolation. We see that our approach paves the way for developing measures of correlation, dipolarity, stability or sparsity in the functional data domain to fine-tune artifact selection. An important issue that remains open is whether the restriction imposed for the truncation point is beneficial or not to achieve better results.

6 Estimating brain signals from contaminated event-related potentials

To illustrate our methods, we reproduced a typical experimental scenario where a human participant had to perform full-arm movements synchronised to a periodic auditory stimulus. An EEG recording was performed during the task. Arguably, what we provide here is a paradigmatic example wherein the researcher needs to clean the signal from motion-related artifacts while preserving activity genuinely related to perceptual and motor brain processes. The subject was instructed to tap his hand on the table synchronizing with a steady auditory stimulus in one condition while listening to the same stimulus without any movement involved in the other. Disposing of a baseline, we could directly compare the outcome of our cleaning procedure with an uncontaminated experimental situation. We recorded 100 trials of 3 seconds per condition, divided into randomized blocks of 25 trials. The stimulus period was 750 milliseconds, i.e., 4 tappings in one of the conditions. Movements were intentionally exaggerated to maximize eventual movement-related artifacts. In this section, the same configuration for running the model ($p = 230; i = 1, \ldots, 64; k = 1, \ldots, 3000$) is preserved from the previous one.

The P-spline smoothed FICA is performed at each trial to obtain brain estimates by subtracting the artifactual components. Here, the complexity of the signal increases as it is assumed a mixture of artifacts and other brain processes due to the cognitive task. Figure 3 shows the grand-averaged results comparing both conditions before and after the artifact removal. A FPCA is performed on the averaged data to visualize the spatial distribution of the scores in the direction of the leading eigenvector before and after the removal. As expected, the activations where nearly coincident after the artifact removal and more prominent in the central region of the scalp. The upper left panel displays the EEG signal in some frontal channels where the movement-related artifact is prominently visible before the subtraction. Further evidence of such artifactual content is given in second row where the raw curves are shown in the other condition. Clearly, the pooled artifacts across the trials have here a different origin. The same panel shows the curves after subtracting the artifactual curves.

Our procedure notably reduces the movement-related artifact and renders the signal more stationary. Indeed, differences are smaller in the non-movement condition but, in either case, our algorithm is capable of reducing artifactual content while retaining the brain activity intact. From our previous tests, one may expect some artifact residue at a trial level depending on the estimated λ and the diversity of source artifacts. We stress that as the response to the repeated stimulus is assumed to be invariant and small in terms of amplitude, averaging suppresses non-phase-locked activity and reveals the potential elicited by the stimulus [43]. Consequently, the attenuation of the roughness of the artifactual component functions will lead to a better estimation of brain potentials at averaging rather than the subtraction of rough components.
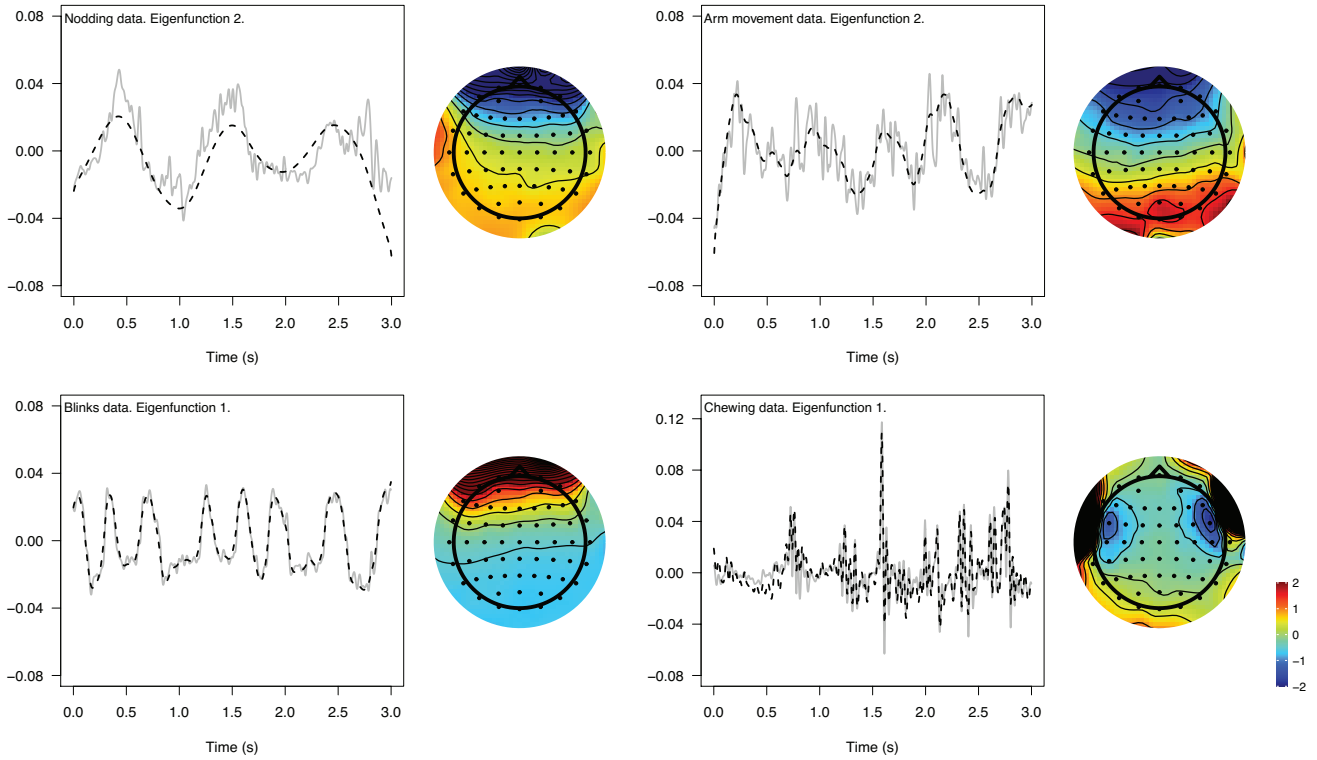
Fig. 2: Artifactual eigenfunctions selected from each trial. The unpenalized FICA (grey) and P-spline smoothed FICA (black dashed) decompositions are compared. The scalp maps represent the scores depicted in the spatial electrode domain obtained by projection of the smooth eigenfunctions in the original sample.

7 Discussion

The proposed independent techniques provide, to the best of our knowledge, the first functional framework for smoothed artifact extraction and removal in dense functional data approximated using a large number of knots. Shrinkage estimators have proven to be an effective starting point for estimating covariance operators in such data (see also [22]). To further enhance performance within this setting, we introduced a novel cross-validation method for model parameter selection. While computationally intensive, our approach has proven to outperform the lack of sensitivity of other existent methods. Overall, this allows the application of independent component techniques from a smoothing perspective somewhat more flexible when compared to other modelling strategies.

Although [25] established a form of Fisher consistency for the kurtosis operator decomposition, no asymptotic results have yet been derived for either the non-smoothed or smoothed independent components. It is reasonable to assume that the competitive performance observed may be partially attributed to asymptotic results from previous FPCA studies. In our empirical setting, however, the analysis of such properties must take into account the specific characteristics of the functional data and the penalized spline method employed, which introduces considerable technical complexity (see, e.g., [51] and [49]). These theoretical developments are beyond the scope of the present work, but we intend to pursue them in a separate study.

In our simulations, the kurtosis operator has proven to work well at capturing artifactual eigenfunctions with different frequency characteristics, at least under certain conditions. One of the strengths of our model is the double regularization, which allows us to circumvent the leak of brain activity and get clean movement-related artifacts. In essence, the degree of separation is defined through the space dimension, from more dependent (first q terms of the FPCA decomposition) to more independent ($q \rightarrow p$). Thus q acts as a regularization parameter to explore the variational component of the artifactual sources in the EEG signal, while λ provides more accurate estimations, particularly in using the first q terms of the K-L expansion. Further research is needed to determine how the model

parameter selection can optimize the removal of artifacts with a minimum loss of variance patterns associated to neural activity.
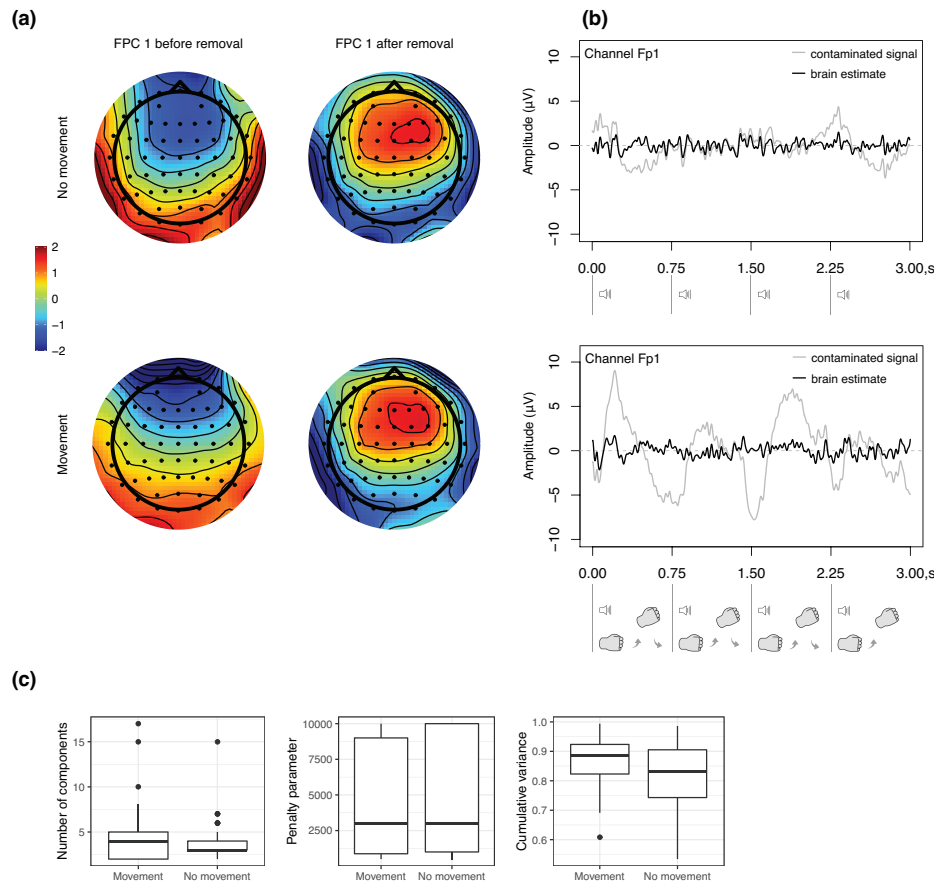
Fig. 3: (a) Topographic maps representing the leading functional principal component of the averaged trials before performing the P-spline smoothed FICA and after. (b) Grand-average across trials of a prefrontal channel where the artifactual activity is revealed. A descriptive scheme of the movement is provided at the bottom of the plots. (c) Box plots of the number of components, the selected penality parameter and the cummulative variance of the model.

8 References

1 Acal, C., Aguilera, A. M., and Escabias, M. (2020). New Modeling Approaches Based on Varimax Rotation of Functional Principal Components. *Mathematics*, 8(11):2085.

2 Aguilera, A., Aguilera-Morillo, M., and Preda, C. (2016). Penalized versions of functional pls regression. *Chemometrics and Intelligent Laboratory Systems*, 154:80–92.

3 Aguilera, A. M. and Aguilera-Morillo, M. C. (2013a). Comparative Study of Different B-Spline Approaches for Functional Data. *Mathematical and Computer Modelling*, 58:1568–1579.

4 — (2013b). Penalized PCA Approaches for B-Spline Expansions of Smooth Functional Data. *Applied Mathematics and Computation*, 219:7805–7819.

5 Aguilera-Morillo, M. and Aguilera, A. (2020). Multi-Class Classification of Biomechanical Data: A Functional LDA Approach Based on Multi-Class Penalized Functional PLS. *Statistical Modelling*, 20(6):592–616.

6 Aguilera-Morillo, M., Aguilera, A., and Durbán, M. (2017). Prediction of Functional Data with Spatial Dependence: a Penalized Approach. *Stochastic Environmental Research and Risk Assessment*, 31:7–22.

7 Aguilera-Morillo, M., Aguilera, A., and Escabias, M. (2013). Penalized Spline Approaches for Functional Logit Regression. *TEST*, 22:251–277.

8 Akhtar, M. T., Mitsuhashi, W., and James, C. J. (2012). Employing Spatially Constrained ICA and Wavelet Denoising, for Automatic Removal of Artifacts from Multichannel EEG Data. *Signal Processing*, 92:401–416.

9 Artoni, F., Delorme, A., and Makeig, S. (2018). Applying Dimension Reduction to EEG data by Principal Component Analysis Reduces the Quality of its Subsequent Independent Component Decomposition. *Neuroimage*, 175:176–187.

10 Ash, R. B. and Gardner, M. F. (1975). *Topics in Stochastic Processes*. Academic Press, New York.

11 Bajaj, N., Requena Carrión, J., Bellotti, F., Berta, R., and De Gloria, A. (2020). Automatic and Tunable Algorithm for EEG Artifact Removal Using Wavelet Decomposition with Applications in Predictive Modeling During Auditory Tasks. *Biomedical Signal Processing and Control*, 55:101624.

12 Buzsáki, G. (2006). *Rhythms of the Brain*. Oxford University Press.

13 Castellanos, N. P. and Makarov, V. A. (2006). Recovering EEG Brain Signals: Artifact Suppression with Wavelet Enhanced Independent Component Analysis. *Journal of Neuroscience Methods*, 158:300–312.

14 Currie, I. D. and Durban, M. (2002). Flexible Smoothing with P-Splines: a Unified Approach. *Statistical Modelling*, 2(4):333–349.

15 Delaigle, A. and Hall, P. (2010). Defining Probability Density for a Distribution of Random Functions. *The Annals of Statistics*, 38:1171–1193.

16 Eilers, P. H. C. and Marx, B. D. (1996). Flexible Smoothing with B-Splines and Penalties (with Discussion). *Statistical Science*, 11:89–121.

17 Ghanem, R. and Spanos, P. (1991). *Stochastic Finite Elements: A Spectral Approach*. Springer-Verlag, New York.

18 Gutch, H. W. and Theis, F. J. (2012). To Infinity and Beyond: On ICA over Hilbert Spaces. In Theis, F. J., Cichocki, A., Yeredor, A., and Zibulevsky, M., editors, *Latent Variable Analysis and Signal Separation*, pages 180–187. Springer Berlin Heidelberg.

19 Hasenstab, K., Scheffler, A., Telesca, D., Sugar, C. A., Jeste, S., DiStefano, C., and Sentürk, D. (2017). A Multi-Dimensional Functional Principal Components Analysis of EEG Data. *Biometrics*, 3(73):999–1009.

20 Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Willey, Chichester.

21 Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. John Wiley & Sons, Ltd, New York.

22 Ieva, F., Paganoni, A. M., and Tarabelloni, N. (2016). Covariance-Based Clustering in Multivariate and Functional Data Analysis. *Journal of Machine Learning Research*, 17:1–21.

23 Kollo, T. (2008). Multivariate Skewness and Kurtosis Measures with an Application in ICA. *Journal of Multivariate Analysis*, 99(10):2328–2338.

24 Lakraj, G. P. and Ruymgaart, F. (2017). Some Asymptotic Theory for Silverman's Smoothed Functional Principal Components in an Abstract Hilbert Space. *Journal of Multivariate Analysis*, 155:122–132.

25 Li, B., Bever, G. V., Oja, H., Sabolová, R., and Critchley, F. (2019). Functional Independent Component Analysis: an Extension of the Fourth-Order Blind Identification. Technical report, Univ. Namur.

26 Loperfido, N. (2017). A New Kurtosis Matrix, with Statistical Applications. *Linear Algebra and its Applications*, 512:1–17.

27 Mammone, N. and F.C.Morabito (2014). Enhanced Automatic Wavelet Independent Component Analysis for Electroencephalographic Artifact Removal. *Entropy*, 16(12):6553–6572.

28 Mehta, N. and Gray, A. (2009). FuncICA for Time Series Pattern Discovery. In *Proceedings of the 2009 SIAM International Conference on Data Mining*, pages 73–84.

29 Nie, Y., Wang, L., Liu, B., and Cao, J. (2018). Supervised Functional Principal Component Analysis. *Statistics and Computing*, 28:713–723.

30 Nordhausen, K. and Oja, H. (2018). Independent Component Analysis: A Statistical Perspective. *WIREs Computational Statistics*, 10:1–23.

31 Noury, N., Hipp, J. F., and Siegel, M. (2016). Physiological Processes Non-linearly Affect Electrophysiological Recordings During Transcranial Electric Stimulation. *Neuroimage*, 140:99–109.

32 Novi, S. L., Roberts, E., Spagnuolo, D., Spilsbury, B. M., Price, D. C., Imbalzano, C. A., Forero, E., Yodh, A. G., Tellis, G. M., Tellis, C. M., and Mesquita, R. C. (2020). Functional Near-Infrared Spectroscopy for Speech Protocols: Characterization of Motion Artifacts and Guidelines for Improving Data Analysis. *Neurophotonics*, 7(1):1–15.

33 Ocaña, F. A., Aguilera, A. M., and Escabias, M. (2007). Computational Considerations in Functional Principal Component Analysis. *Computational Statistics*, 22:449–465.

34 Ocaña, F. A., Aguilera, A. M., and Valderrama, M. J. (1999). Functional Principal Component Analysis by Choice of Norm. *Journal of Multivariate Analysis*, 71:262–276.

35 Peña, C., Prieto, J., and Rendón, C. (2014). Independent Components Techniques Based on Kurtosis for Functional Data Analysis. Working paper 14–10, Universidad Carlos III de Madrid.

36 Pokora, O., Kolacek, J., Chiu, T., and Qiu, W. (2018). Functional Data Analysis of Single-Trial Auditory Evoked Potentials Recorded in the Awake Rat. *Biosystems*, 161:67–75.

37 Qi, X. and Zhao, H. (2011). Some Theoretical Properties of Silverman's Method for Smoothed Functional Principal Component Analysis. *Journal of Multivariate Analysis*, 102:742–767.

38 Ramsay, J. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer, New York.

39 Rice, J. A. and Silverman, B. W. (1991). Estimating the Mean and Covariance Structure Nonparametrically when the Data are Curves. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(1):233–243.

40 Schäfer, J. and Strimmer, K. (2005). A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics. *Statistical Applications in Genetics and Molecular Biology*, 4:1–29.

41 Scheffler, A., Telesca, D., Q. Li, Q., Sugar, C. A., Distefano, C., Jeste, S., and Şentürk, D. (2018). Hybrid Principal Components Analysis for Region-Referenced Longitudinal Functional EEG Data. *Biostatistics*, 21:139–157.

42 Silverman, B. W. (1996). Smoothed Functional Principal Components Analysis by Choice of Norm. *The Annals of Statistics*, 24:1–24.

43 Tong, S. and Thakor, N. V. (2009). *Quantitative EEG Analysis Methods and Clinical Applications*. Artech House, Boston.

44 Urigüen, J. A. and Garcia-Zapirain, B. (2015). EEG Artifact Removal—State-of-the-Art and Guidelines. *Journal of Neural Engineering*, 12:1–23.

45 Vidal, M. and Aguilera, A. M. (2021). *pfica: Penalized Independent Component Analysis for Univariate Functional Data*. R package version 0.1.2.

46 Virta, J., Li, B., Nordhausen, K., and Oja, H. (2020). Independent Component Analysis for Multivariate Functional Data. *Journal of Multivariate Analysis*, 176:1–19.

47 Wang, J.-L., Chiou, J.-M., and H. G. Müller (2016). Functional Data Analysis. *Annual Review of Statistics and Its Application*, 3:257–295.

48 Wang, X. (2010). Neurophysiological and Computational Principles of Cortical Rhythms in Cognition. *Physiological reviews*, 90:1195–1268.

49 Xiao, L. (2019). Asymptotic Theory of Penalized Splines. *Electronic Journal of Statistics*, 13:747–794.

50 Xiao, L., Zipunnikov, V., Ruppert, D., and Crainiceanu, C. (2016). Fast Covariance Estimation for High-Dimensional Functional Data. *Statistics and Computing*, 26:409–421.

51 Zhang, X. and Wang, J.-L. (2016). From Sparse to Dense Functional Data and Beyond. *The Annals of Statistics*, 44:2281–2312.