

Document downloaded from:

<http://hdl.handle.net/10251/157360>

This paper must be cited as:

Safont Armero, G.; Salazar Afanador, A.; Vergara Domínguez, L.; Rodríguez Martínez, A. (2019). Nonlinear estimators from ICA mixture models. *Signal Processing*. 155:281-286. <https://doi.org/10.1016/j.sigpro.2018.10.003>



The final publication is available at

<https://doi.org/10.1016/j.sigpro.2018.10.003>

Copyright Elsevier

Additional Information

"NOTICE: this is the author's version of a work that was accepted for publication in *Signal Processing*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Signal Processing*, 155, (2019) <https://doi.org/10.1016/j.sigpro.2018.10.003>"

Nonlinear estimators from ICA Mixture Models

Gonzalo Safont¹, Addison Salazar¹, Luis Vergara¹, Alberto Rodríguez²

¹Universitat Politècnica de València, Instituto of Telecomunicaciones y Aplicaciones Multimedia

²Universidad Miguel Hernández de Elche, Departamento de Ingeniería de Comunicaciones

Abstract

Independent Component Analyzers Mixture Models (ICAMM) are versatile and general models for a large variety of probability density functions. In this paper we assume ICAMM to derive new MAP and LMSE estimators. The first one (MAP-ICAMM) is obtained by an iterative gradient algorithm, while the second (LMSE-ICAMM) admits a closed-form solution. Both estimators can be combined by using LMSE-ICAMM to initialize the iterative computation of MAP-ICAMM. The new estimators are applied to the reconstruction of missed channels in EEG multichannel analysis. The experiments demonstrate the superiority of the new estimators with respect to: Spherical Splines, Hermite, Partial Least Squares, Support Vector Regression, and Random Forest Regression.

Keywords: ICA; nonlinear estimators; LMSE; MAP; EEG reconstruction.

1. Introduction

Estimation is one of the fundamental problems in statistical signal processing [1]. It is an essential part of many fields like spectral analysis, coding, time series analysis, prediction, interpolation, smoothing. Moreover, it appears in many areas of application. In spite of the huge amount of previous work in statistical estimation methods, new developments are still possible if new statistical models appear, so that new maximum a posteriori (MAP) or least mean square error (LMSE) solutions can be found. In this paper we consider the Independent Component Analyzers Mixture Model (ICAMM) [2-4] of the multivariate probability density function (MPDF) of the observations. ICAMM is a versatile model which encompasses most of the usual MPDF models, including both non-Gaussian and Gaussian Mixture Models. This generality implies that optimal estimators assuming an underlying ICAMM can be very attractive options in a large variety of scenarios. Therefore, we have derived the MAP and LMSE estimators of missing data, considering that the underlying MPDF is properly captured by ICAMM. The MAP estimator will be obtained by an iterative algorithm, and will be called MAP-ICAMM. On the other hand, the LMSE estimator is the expected value of the missing data conditioned to the available data, which we have calculated assuming ICAMM; the corresponding estimator will be called LMSE-ICAMM. The new estimator has been assessed in the reconstruction of

missing data of electroencephalographic (EEG) signals measured on subjects while performing a memory and learning task. They have been compared with the method of splines [5], a deterministic method which approximates complex function stepwise by local polynomials. In particular, we have considered the following state-of-the-art methods: Spherical Splines ([6]), which is the most commonly used method for interpolation in EEG processing in many EEG processing software, such as EEGLAB [7]; Hermite interpolation ([8]), which is a benchmark smoothing and interpolation method; Partial Least Squares ([9]), a preferred tool for ill-posed linear estimation problems; Support Vector Regression ([10]), one of the most popular machine learning tools for regression; and Random Forest Regression ([11]), which is a recent multidimensional interpolation technique with good performance across different applications (see [12] and the references within).

In the next section we present the analytical derivation of the new proposed estimators (MAP-ICAMM and LMSE-ICAMM). Then Section 3 is devoted to the mentioned EEG data application. Conclusion section ends the paper.

2. Estimators based on ICAMM

Let us consider an observation vector \mathbf{x} of dimension $(M \times 1)$. Without loss of generality, this vector can be defined as being composed by a vector of known components, \mathbf{y} , and a vector of unknown components, \mathbf{z} , in the form

$$\mathbf{x} = \begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix} = \mathbf{P}_y \mathbf{y} + \mathbf{P}_z \mathbf{z} \quad , \quad (1)$$

where \mathbf{z} is a vector of size $(M_{unk} \times 1)$ and therefore \mathbf{y} is a vector of size $((M - M_{unk}) \times 1)$. \mathbf{P}_y and \mathbf{P}_z are rectangular diagonal matrices respectively equal to the first $M - M_{unk}$ columns and the last

M_{unk} columns of the identity matrix of dimension $(M \times M)$, so that $\mathbf{P}_y \mathbf{y} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_{M_{unk}} \end{bmatrix}$ and

$\mathbf{P}_z \mathbf{z} = \begin{bmatrix} \mathbf{0}_{M - M_{unk}} \\ \mathbf{z} \end{bmatrix}$, where $\mathbf{0}_i$ is a zero vector of size $(i \times 1)$. The goal is to estimate \mathbf{z} from \mathbf{y} . Let us

assume that the MPDF $p(\mathbf{x})$ is modeled with a K -class ICAMM. If \mathbf{x} belongs to class C_k $k = 1 \dots K$, then

$$\mathbf{x} = \mathbf{A}_k \mathbf{s}_k + \mathbf{b}_k \quad , \quad (2)$$

where \mathbf{s}_k is a vector of size $(M \times 1)$ that contains statistically independent components (also known as ‘‘sources’’); \mathbf{A}_k is the mixing matrix; and \mathbf{b}_k is a bias vector. Notice that \mathbf{A}_k is a

square matrix that can be inverted to obtain $\mathbf{W}_k = \mathbf{A}_k^{-1}$, the de-mixing matrix of class k . Hence $p(\mathbf{x})$ can be expressed as a mixture of K components respectively corresponding to the K classes.

$$p(\mathbf{x}) = p(\mathbf{z}, \mathbf{y}) = \sum_{k=1}^K p(\mathbf{z}, \mathbf{y} | C_k) P(C_k) = \sum_{k=1}^K |\det \mathbf{W}_k| p(\mathbf{s}_k) P(C_k) \quad , \quad (3)$$

where $P(C_k)$ is the prior probability of class k .

Given a training set of observation vectors $\mathbf{x}^{(n)}$ $n=1, \dots, N$, we can estimate the model parameters \mathbf{W}_k , \mathbf{b}_k and $P(C_k)$, using one of the many existing methods (see [13] for a general procedure).

1.1. MAP estimator (MAP-ICAMM)

Let us consider the MAP estimator of \mathbf{z} from \mathbf{y}

$$\mathbf{z}_{MAP} = \underset{\mathbf{z}}{\operatorname{arg\,max}} \log p(\mathbf{z}, \mathbf{y}) = \underset{\mathbf{z}}{\operatorname{arg\,max}} L(\mathbf{z}, \mathbf{y}) \quad . \quad (4)$$

This maximization requires the calculation of the derivative of $L(\mathbf{z}, \mathbf{y})$ with respect to \mathbf{z} . By taking the log of (3) and considering that the components of \mathbf{s}_k are independent, we arrive to

$$\begin{aligned} \frac{\partial L(\mathbf{z}, \mathbf{y})}{\partial \mathbf{z}} &= \frac{1}{p(\mathbf{z}, \mathbf{y})} \sum_{k=1}^K |\det \mathbf{W}_k| P(C_k) \sum_{m=1}^M \frac{\partial p(\mathbf{s}_k)}{\partial s_{km}} \frac{\partial s_{km}}{\partial \mathbf{z}} = \\ &= \sum_{k=1}^K p(C_k | \mathbf{z}, \mathbf{y}) \sum_{m=1}^M \frac{\partial \log p(s_{km})}{\partial s_{km}} \frac{\partial s_{km}}{\partial \mathbf{z}} \end{aligned} \quad , \quad (5)$$

where s_{km} is the m -th source of class k . The value s_{km} can be obtained as $s_{km} = \mathbf{w}_{km}^T (\mathbf{x} - \mathbf{b}_k)$, with \mathbf{w}_{km}^T being the m -th row of \mathbf{W}_k . Thus, its derivative is equal to

$$\frac{\partial s_{km}}{\partial \mathbf{z}} = (\mathbf{w}_{km}^T \mathbf{P}_z)^T = \mathbf{P}_z^T \mathbf{w}_{km} \quad . \quad (6)$$

The derivative of $\log p(s_{km})$, also known as the score function, can be calculated explicitly for many common probability density functions. This requires prior knowledge of the source PDF which could limit the applicability of the algorithm. To reach general applicability, we will assume that the probability density of each source is to be estimated using a nonparametric kernel density estimator with a Gaussian kernel. This requires labelling of the training samples which can be made, once the model parameters have been estimated, by selecting the class k

which maximizes the posterior probability $P(C_k | \mathbf{x}_n) = P(\mathbf{x}_n | C_k)P(C_k) / \sum_{k'=1}^K P(\mathbf{x}_n | C_{k'})P(C_{k'})$ [3]. Let us call $\mathbf{x}_k^{(l)}$ $l=1\dots L_k$ to the subset of the training samples assigned to class k , and $\mathbf{s}_k^{(l)} = \mathbf{W}_k(\mathbf{x}_k^{(l)} - \mathbf{b}_k)$, the corresponding source vectors. The non-parametric estimator of $p(s_{km})$ is given by

$$p(s_{km}) = \frac{1}{a_0} \sum_{l=1}^{L_k} e^{-\frac{1}{2h^2}(s_{km} - s_{km}^{(l)})^2}, \quad (7)$$

where $s_{km}^{(l)}$ is the m -th component of vector $\mathbf{s}_k^{(l)}$, h is called the bandwidth of the nonparametric estimator, and a_0 is a scaling constant calculated so that $\int_{-\infty}^{\infty} p(s_{km}) ds_{km} = 1$. For the Gaussian kernel of (7), this scaling constant is $a_0 = \sqrt{2\pi} h L_k$. The derivative of $\log p(s_{km})$ is

$$\frac{\partial \log p(s_{km})}{\partial s_{km}} = \frac{\sum_{l=1}^{L_k} \left(\frac{s_{km} - s_{km}^{(l)}}{h} \right) e^{-\frac{1}{2h^2}(s_{km} - s_{km}^{(l)})^2}}{\sum_{l=1}^{L_k} e^{-\frac{1}{2h^2}(s_{km} - s_{km}^{(l)})^2}}. \quad (8)$$

Hence, the derivative of $L(\mathbf{z}, \mathbf{y})$ can be calculated by replacing (6) and (8) into (5). Then, maximization of $L(\mathbf{z}, \mathbf{y})$ could be achieved by using classical gradient algorithms, although Newton methods are preferable to get fast convergence. The problem of the Newton methods is that computation of the second derivative is required. To overcome this inconvenience, a family of quasi-Newton methods have been proposed. In particular, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [14] is generally accepted as the quasi-Newton method yielding the best compromise between fast convergence and computational payload, even for non-smooth optimization. That is the option that we have selected to obtain $\mathbf{z}_{MAP} = \underset{\mathbf{z}}{\text{max}} L(\mathbf{z}, \mathbf{y})$. We have named this estimator as MAP-ICAMM. A pseudocode summary of the algorithm is included in the following:

Input: Training set of complete vectors $\mathbf{x}^{(n)}$ $n = 1 \dots N$, vector of known components \mathbf{y} , initial value for the unknown components $\mathbf{z}(0)$

Output: MAP prediction, \mathbf{z}_{MAP}

- 1: Estimate ICAMM parameters $\mathbf{b}_k \mathbf{W}_k P(C_k)$, $k = 1 \dots K$ from the training set [13]
 - 2: Label training samples [3] and compute $\mathbf{s}_k^{(l)} = \mathbf{W}_k (\mathbf{x}_k^{(l)} - \mathbf{b}_k)$ $l = 1 \dots L_k$ $k = 1 \dots K$
 - 3: **for** $i = 1 \dots I$
 - 4: Compute $\frac{\partial L(\mathbf{z}(i-1), \mathbf{y})}{\partial \mathbf{z}(i-1)}$ (5)-(8)
 - 5: **end for**
 - 6: $\mathbf{z}_{MAP} = \mathbf{z}(I)$
-

1.2. LMSE estimator (LMSE-ICAMM)

The general solution to the LMSE criterion is the conditional expectation of unknown data with respect to known data, that is,

$$\mathbf{z}_{LMSE} = E[\mathbf{z} | \mathbf{y}] = \int \mathbf{z} p(\mathbf{z} | \mathbf{y}) d\mathbf{z} \quad . \quad (9)$$

Considering the mixture model in (3) and using the chain rule, \mathbf{z}_{LMSE} can be expressed as

$$\mathbf{z}_{LMSE} = \int \mathbf{z} p(\mathbf{z} | \mathbf{y}) d\mathbf{z} = \sum_{k=1}^K \int \mathbf{z} p(\mathbf{z} | \mathbf{y}, C_k) d\mathbf{z} \cdot P(C_k | \mathbf{y}) = \sum_{k=1}^K E[\mathbf{z} | \mathbf{y}, C_k] \cdot P(C_k | \mathbf{y}). \quad (10)$$

So we need to compute $E[\mathbf{z} | \mathbf{y}, C_k]$ and $P(C_k | \mathbf{y})$. Regarding $E[\mathbf{z} | \mathbf{y}, C_k]$, let us first compute the conditional expectation of the sources $E[\mathbf{s}_k | \mathbf{y}, C_k]$. Considering (1) and (2) we may write

$$E[\mathbf{s}_k | \mathbf{y}, C_k] = E[\mathbf{W}_k (\mathbf{P}_y \mathbf{y} + \mathbf{P}_z \mathbf{z} - \mathbf{b}_k) | \mathbf{y}, C_k] = \mathbf{W}_k \mathbf{P}_y \mathbf{y} + \mathbf{W}_k \mathbf{P}_z E[\mathbf{z} | \mathbf{y}, C_k] - \mathbf{W}_k \mathbf{b}_k, \quad (11)$$

then we can solve for $E[\mathbf{z} | \mathbf{y}, C_k]$ by using the pseudoinverse $(\mathbf{W}_k \mathbf{P}_z)^+$

$$E[\mathbf{z} | \mathbf{y}, C_k] = (\mathbf{W}_k \mathbf{P}_z)^+ \cdot [E[\mathbf{s}_k | \mathbf{y}, C_k] - \mathbf{W}_k \mathbf{P}_y \mathbf{y} + \mathbf{W}_k \mathbf{b}_k] \quad . \quad (12)$$

Application of (12) to estimate $E[\mathbf{z} | \mathbf{y}, C_k]$ requires knowledge of $E[\mathbf{s}_k | \mathbf{y}, C_k]$. This later can be estimated using a variety of existing methods for interpolation in multidimensional irregular grids. Particularly simple is the nearest neighbours (NN) method of interpolation [15]. Given the observed vector \mathbf{y} , we select $R \leq L_k$ samples of the training subset $\mathbf{x}_k^{(l)}$ $l = 1 \dots L_k$ assigned to class

k , let us call these selected samples $\mathbf{x}_{ky}^{(1)} \dots \mathbf{x}_{ky}^{(R)}$, whose corresponding $\mathbf{y}_{ky}^{(1)} \dots \mathbf{y}_{ky}^{(R)}$, as defined in (1), are the R -NN to \mathbf{y} . Then, we can compute sample estimates of $E[\mathbf{s}_k | \mathbf{y}, C_k]$ in the form

$$\hat{E}[\mathbf{s}_k | \mathbf{y}, C_k] = \frac{1}{R} \sum_{r=1}^R \mathbf{s}_{ky}^{(r)} = \frac{1}{R} \sum_{r=1}^R \mathbf{W}_k (\mathbf{x}_{ky}^{(r)} - \mathbf{b}_k) \quad (13)$$

Regarding calculation of $P(C_k | \mathbf{y})$, this is just a problem of computing class posterior probabilities, so we can use any classifier trained with the vectors $\mathbf{y}^{(n)}$, $n = 1, \dots, N$ corresponding to the original training set $\mathbf{x}^{(n)}$, $n = 1, \dots, N$, as defined in (1). In particular an ICAMM classifier [3] could be tried. The estimator deduced from (10)-(13) will be called LMSE-ICAMM. A pseudocode resume of the algorithm is included in the following:

Algorithm 2. *Computation of the LMSE-ICAMM.*

Input: Training set of complete vectors $\mathbf{x}^{(n)} = \begin{bmatrix} \mathbf{y}^{(n)} \\ \mathbf{z}^{(n)} \end{bmatrix}$ $n = 1 \dots N$, vector of known components \mathbf{y}

Output: LMSE prediction, \mathbf{z}_{LMSE}

- 1: Estimate ICAMM parameters $\mathbf{b}_k \mathbf{W}_k P(C_k)$, $k = 1 \dots K$ from the training set [13]
 - 2: Label training samples [3] and compute $\mathbf{s}_k^{(l)} = \mathbf{W}_k (\mathbf{x}_k^{(l)} - \mathbf{b}_k)$ $l = 1 \dots L_k$ $k = 1 \dots K$
 - 3: Compute $\hat{E}[\mathbf{s}_k | \mathbf{y}, C_k]$ (13) and $E[\mathbf{z} | \mathbf{y}, C_k]$ (12)
 - 4: Compute $P(C_k | \mathbf{y})$ from a classifier trained with $\mathbf{y}^{(n)}$ $n = 1 \dots N$
 - 5: Compute \mathbf{z}_{LMSE} using (10)
-

3. Real data experiments

Electroencephalographic (EEG) signals are recordings of surface brain electrical activity taken at the scalp, where each sensor captures the combined signal from multiple neurons of the brain [16]. The study of EEG signals is a useful clinical tool because some illnesses, typically seizures and sleep disorders, produce abnormal electric patterns in the electrical activity of the brain that can be identified by an expert.

There are several applications of independent component analysis (ICA) in EEG signal processing. The first applications introduced the assumption that each EEG channel is composed by a mixture of non-Gaussian sources [17]. Current applications of ICA for EEG signal processing include source extraction, noise and artifact removal, and dynamic modeling of brain oscillations (see [17]-[19] and the references therein).

In the EEG application presented here, it was assumed that one or several channels have been partially corrupted or missing for some time and the proposed methods were used to recover the missing EEG samples. This kind of scenario is relatively common in EEG processing, for

instance, during artifact removal. In general, this data recovery step would benefit any application on EEG signals. As stated in the Introduction, we compared the performance of the proposed methods with five state-of-the-art methods: Spherical Splines [6], Hermite [8], Partial Least Squares (PLS) [9], Support Vector Regression (SVR) [10], and Random Forest Regression (RFR) [11]. As for our methods, we tested different initialization options for MAP-ICAMM by exhaustive simulation, arriving to the conclusion that initializing with the solution given by LMSE-ICAMM was in general the best option. Thus, in the following we have used the term MAP+LMSE-ICAMM.

The EEG signals used in this experiment were captured using an ActiveTwo system from BioSemi with active electrodes. The device recorded the signals while the subject was performing Sternberg memory task, a classical memory and learning neuropsychological test that measures multi-object short-term memory [20]. Electrodes were positioned following the 10-10 system, and 64 EEG channels were recorded at 512 Hz. For this particular experiment, we considered the signals from two healthy subjects (one male, one female) that performed a set of 120 trials of the experiment. The subjects answered correctly 98.60 % of the trials with an average response time of 1.23 s. The whole experiment lasted some 16 minutes per subject, during which $N = 491520$ EEG samples were recorded. Figure 1 displays several seconds of the captured EEG during one trial of Sternberg’s memory task.

The estimation process was set as follows. Following the definitions laid out in Section 2, we form observation vectors \mathbf{x} of dimension $M = 64$ by grouping the 64 samples (one for every channel) corresponding to a given instant n , hence we have a total of $N_r = 491520$ observation vectors. First, we split the signal into 10-second epochs. For each epoch, the initial nine seconds of the data were used as a training set, i.e., $N = 5120$, and the remaining second was used as a testing set. Every experiment consisted in assuming that one or more of the electrodes were unusable during one second during recording of the testing set. Then the “missing second” was reconstructed with the proposed methods. ICAMM was trained considering the $K = 2$, which corresponds to the two stages of Sternberg’s task: presentation of the stimuli and response to the prompt, using the training set and the method described in [13].

In every experiment we have computed two performance indicators: the signal-to-interference ratio (SIR), and the relative maximum error (RME):

$$SIR = \frac{\|\mathbf{z}\|^2}{\|\mathbf{z} - \hat{\mathbf{z}}\|^2}, \quad RME = \frac{\max |z_i - \hat{z}_i|}{\max |z_i|}, \quad i = 1 \dots M_{unk} \quad , \quad (13)$$

where $\hat{\mathbf{z}}$ is the estimation of \mathbf{z} provided by the corresponding method and $\|\cdot\|$ is the Euclidean norm. SIR is a typical measure of the quality of reconstruction of a signal from a mixture of signals plus noise (see for example [21]). It is equivalent to the inverse of the normalized

estimation error. Conversely, the relative maximum error is a measure of the maximum deviation that will be experienced from the actual data.

The average results for both subjects are shown in Figure 2. Every point in this figure shows the obtained value of SIR averaged over 1000 repetitions of the experiment, for a given number M_{unk} of missing channels. The axis on the RME plot (Figure 2.b) was inverted so that both indicators are better with higher values. All methods experienced a decrease in interpolation performance as the number of missing electrodes increased. This worsening was more marked for lower numbers of missing channels (<5) and then leveled off, with the exception of Spherical Splines, which degraded quickly as the number of missing channels increased. The worst results were returned by Spherical Splines, Hermite and RFR, with SVM and PLS yielding intermediate results. On the other hand, the proposed methods achieved the best results by a wide margin, with MAP+LMSE-ICAMM improving the result of LMSE-ICAMM. These trends were consistent across both indicators, with the results for the RME being more unstable due to the nonlinear nature of the indicator. The only exception was Hermite, whose RME results were better than its SIR results (yet still worse than those of the proposed methods).

It is also relevant to mention that in the computation of LMSE-ICAMM in the experiment of Figure 2 we considered $R = L_k$ in (13). This simply implies computing the sample mean of all the training sources $\mathbf{s}_k^{(l)} = \mathbf{W}_k(\mathbf{x}_k^{(l)} - \mathbf{b}_k)$ $l=1\dots L_k$ corresponding to class k , ($\hat{E}[\mathbf{s}_k | \mathbf{y}, C_k] \approx \hat{E}[\mathbf{s}_k | C_k]$), i.e., there is no need for searching the R -NN, and computational load is significantly reduced. Actually, we have performed experiments to compare the results obtained for $R < L_k$ with those ones of $R = L_k$. The results for $R=10$ and $R = L_k$ are showed in Figure 3, where SIR has been averaged over 100 repetitions. It can be seen that the curves are practically identical and even slightly better if $R = L_k$. This is because the whole training subset $\mathbf{s}_k^{(l)}$ $l=1\dots L_k$ is considered to obtain the sample estimate in (13), thus reducing the variance of the estimate. Similar results were obtained with other values of R .

Finally, Figure 4 shows the estimation for a given time instant from a case with 32 missing channels, and Figure 5 shows the estimation error for the same time instant. The results are plotted as “scalp maps,” topographical maps of EEG data over the scalp of the subject. Electrodes are indicated by dots and missing channels are indicated by stars. In concordance with the results in Figure 2, LMSE-ICAMM and MAP+LMSE-ICAMM achieved a much lower estimation error than the rest of the methods. For instance, RFR, SVR and PLS all had a positive peak near the right occipital region that was much larger than that of the actual data, while LMSE-ICAMM and MAP+LMSE-ICAMM both correctly predicted that peak. Furthermore, the general distribution of the channel amplitudes (as indicated by the contour lines) with the proposed methods is different from that of the other considered method, with

MAP+LMSE-ICAMM achieving a distribution of amplitudes very similar to that of the true data.

4. Conclusions

We have presented two new statistical estimators. Both are derived assuming a non-Gaussian mixture model for the joint probability density of the observations (ICAMM). The versatility and generality of ICAMM allows a wide scope of applications were the new estimators, which are optimal for the assumed conditions, can outperform other methods. MAP-ICAMM requires an iterative algorithm, while LMSE-ICAMM admits a closed form solution. A natural way to combine both methods is to initialize MAP-ICAMM with LMSE-ICAMM, which has come to be the best option in the performed experiments. The experimental comparison has been made in the context of recovering missing EEG channels. The superiority of the new estimators with respect to several methods (Spherical Splines, Hermite, Random Forest Regression, Support Vector Regression and Partial Least Squares) has been evidenced. These results and the generality of the underlying model encourage the experimentation of the proposed methods in other areas of application.

Acknowledgements

This work was supported by Spanish Administration (Ministerio de Economía y Competitividad) and European Union (FEDER) under grant TEC2014-58438-R, and *Generalitat Valenciana* under grant PROMETEO II/2014/032.

References

- [1] G.A. Einicke, Smoothing, Filtering and Prediction - Estimating The Past, Present and Future, InTech, 2012.
- [2] P. Common and C. Jutten, Handbook of Blind Source Separation: Independent Component Analysis and Applications, Academic Press, 2010.
- [3] A. Salazar, On Statistical Pattern Recognition in Independent Component Analysis Mixture Modelling, Springer, 2013.
- [4] A. Salazar and L. Vergara, Independent Component Analysis (ICA): Algorithms, Applications and Ambiguities, Nova Science Publishers, 2018.
- [5] Y. Wang, Smoothing Splines: Methods and Applications, Taylor&Francis, 2011.
- [6] T.C. Ferree, "Spherical splines and average referencing in scalp electroencephalography," Brain topography, 19(1-2), pp.43-52, 2006.
- [7] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics," Journal of Neuroscience Methods, 134, pp. 9-21, 2004.

- [8] X. Han, "Direction-consistent tangent vectors for generating interpolation curves," *Journal of Computational and Applied Mathematics*, 346, pp. 237-246, 2019.
- [9] A.E. Stott, S. Kanna, D.P. Mandic, "Widely linear complex partial least squares for latent subspace regression," *Signal Processing*, 152, pp. 350-362, 2018.
- [10] A.E. Hassanien, M. Kilany, E.H. Houssein, H. AlQaheri, "Intelligent human emotion recognition based on elephant herding optimization tuned support vector regression," *Biomedical Signal Processing and Control*, 45, pp. 182-191, 2018.
- [11] L. Breiman, "Random forests," *Machine Learning*, 45(1), pp. 5-32, 2001.
- [12] A.J. Wyner, M. Olson, J. Bleich, and D. Mease, "Explaining the success of AdaBoost and random forests as interpolating classifiers," *The Journal of Machine Learning Research*, 18(1), pp. 1558-1590, 2017.
- [13] A. Salazar, L. Vergara, A. Serrano and J. Igual, "A General Procedure for Learning Mixtures of Independent Component Analyzers," *Pattern Recognition*, 43(1), pp. 69-85, 2010.
- [14] E.K.P. Chong, S.H. Zak, *An Introduction to Optimization*, John Wiley & Sons, 2013.
- [15] O. Kramer, *Dimensionality reduction with unsupervised nearest neighbors*, Springer, 2013.
- [16] E. Niedermeyer and F.L. da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, Lippincot Williams & Wilkins, 2004.
- [17] T.P. Jung and T.W. Lee, "Applications of Independent Component Analysis to Electroencephalography," in *Statistical and Process Models for Cognitive Neuroscience and Aging*, M.J. Wenger and C. Schuster (ed.), Lawrence Erlbaum, 2007.
- [18] G. Safont, A. Salazar, L. Vergara, E. Gomez, and V. Villanueva, "Mixtures of independent component analyzers for microarousal detection," in *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 752-755, Valencia, Spain, 2014.
- [19] G. Safont, A. Salazar, L. Vergara, E. Gomez, V. Villanueva, "Probabilistic Distance for Mixtures of Independent Component Analyzers," *IEEE Transactions on Neural Networks and Learning Systems*, 29(4), pp. 1161-1173, 2018.
- [20] S. Sternberg, "High-speed scanning in human memory," *Science*, 153(3736), pp. 652-654, 1966.
- [21] F. Lu, Z. Huang, W. Jiang, "Underdetermined blind separation of non-disjoint signals in time-frequency domain based on matrix diagonalization," *Signal Processing*, 91, pp. 1568-1577, 2011.

Figure 1. An example of seven seconds of EEG data from one trial of the experiment.

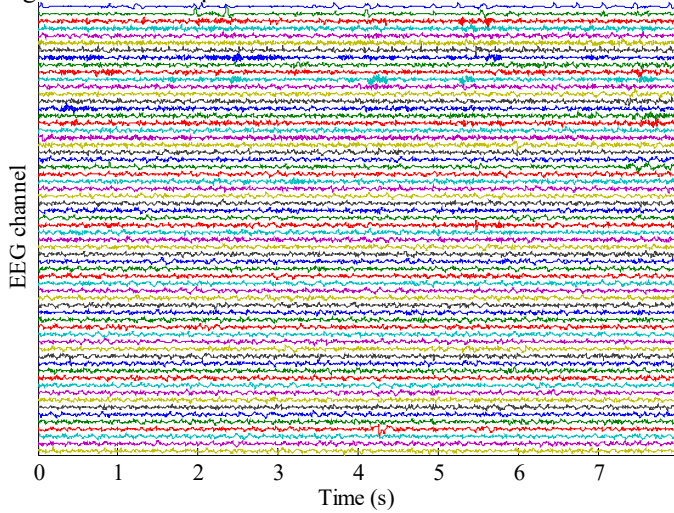


Figure 2. Average performance indicators for the different estimators in the EEG experiments. The axis on the RME plot is inverted so that both indicators are better with higher values.

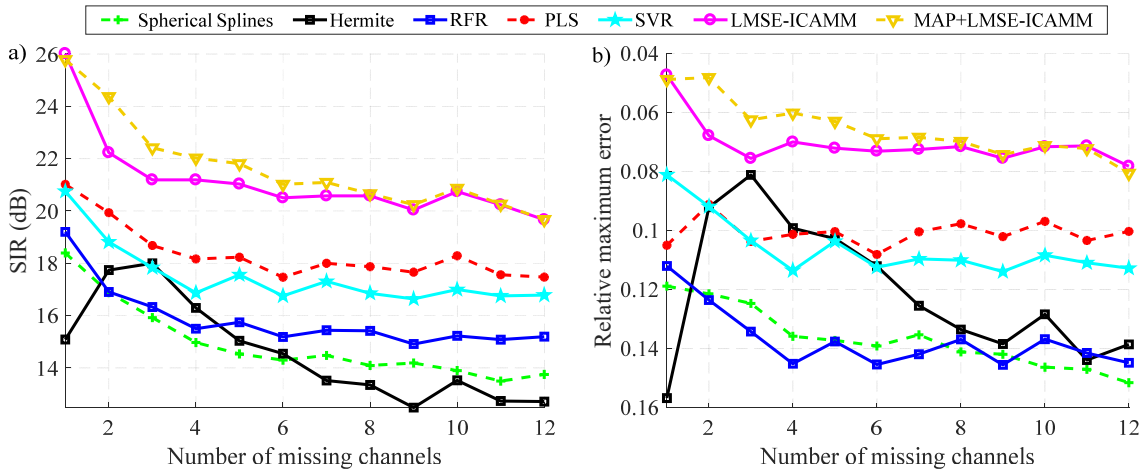


Figure 3. SIR indicator for LMSE-ICAMM considering $R = L_k$ and $R = 10$ in (13)

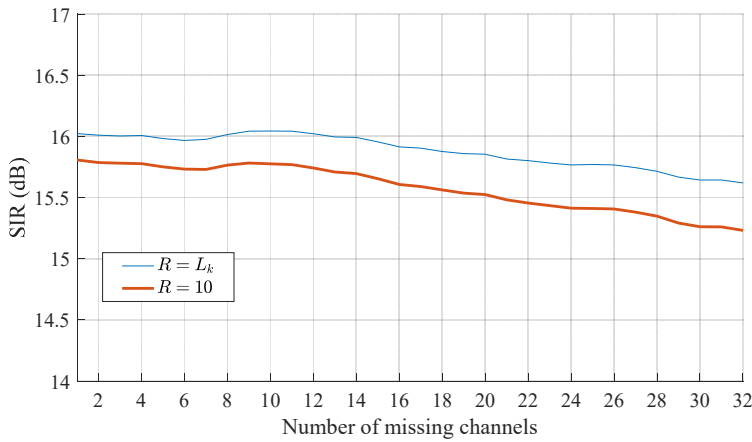


Figure 4. Scalp maps for a case with 32 missing channels. Missing channels are indicated by stars (*), while known channels are marked by dots (•).

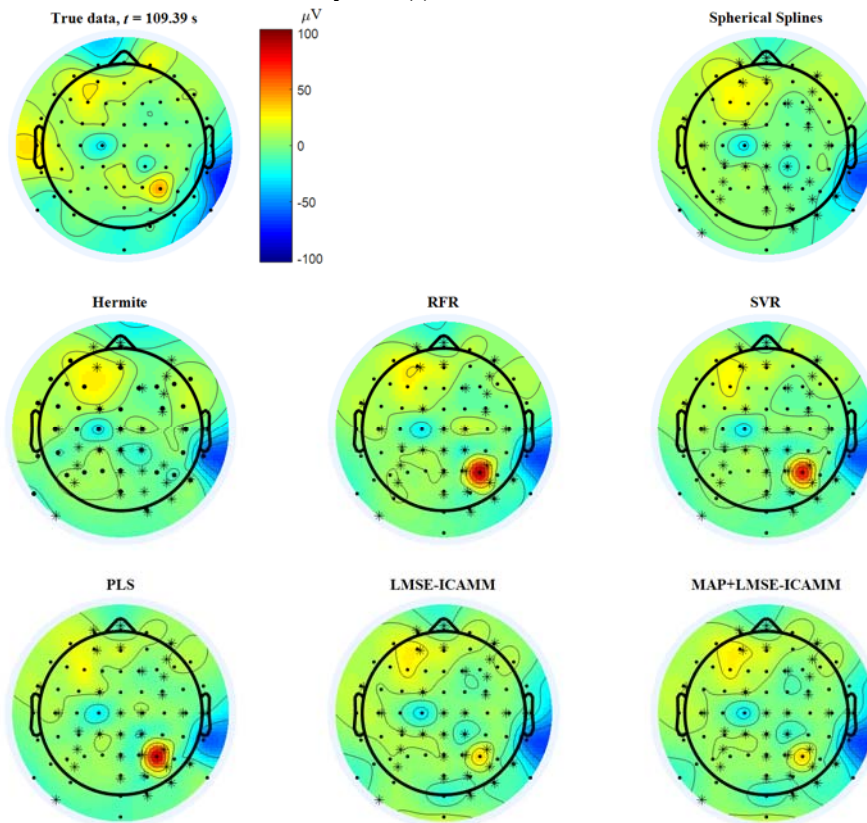


Figure 5. Estimation error for the same case with 32 missing channels shown in Figure 4. Missing channels are indicated by stars (*), while known channels are marked by dots (•).

