How to Test the Quality of Reconstructed Sources in Independent Component Analysis (ICA) of EEG/MEG Data

Moritz Grosse-Wentrup*, Stefan Harmeling*, Thorsten Zander*†, Jeremy Hill‡ and Bernhard Schölkopf*
*Max Planck Institute for Intelligent Systems, Department Empirical Inference, 72076 Tübingen, Germany
†TU-Berlin, Institut für Psychologie und Arbeitswissenschaft, 10587 Berlin, Germany
‡Wadsworth Center, New York State Dept. of Health, Albany, NY 12201, USA
Email: {moritzgw,stefan.harmeling,bs}@tuebingen.mpg.de; tzander@gmail.com; jezhill@gmail.com

***Abstract*—We provide a simple method, based on volume conduction models, to quantify the neurophysiological plausibility of independent components (ICs) reconstructed from EEG/MEG data. We evaluate the method on EEG data recorded from 19 subjects and compare the results with two established procedures for judging the quality of ICs. We argue that our procedure provides a sound empirical basis for the inclusion or exclusion of ICs in the analysis of experimental data.**

***Keywords*-Independent Component Analysis; ICA; EEG; MEG.**

I. Introduction

Independent Component Analysis (ICA) is applied with great success in the analysis of EEG and MEG recordings, e.g. for artifact attenuation [1], [2], to separate sources into task-related and non task-related components [3]–[5], or as a pre-processing step for source localization [4], [6], [7]. Sorting reconstructed independent components (ICs) into cortical current sources and unwanted noise components, however, is often based on heuristics. Criteria invoked to identify cortical ICs include the dipolarity of source topographies [8] and the $1/f$-spectrum of reconstructed ICs [9]. Even though these criteria are neurophysiologically plausible, they lack a rigorous analytical justification. Furthermore, scientists need to be trained to visually identify neurophysiologically plausible ICs, introducing an undesirable subjective element into the analysis. A more objective criterion for judging the quality of ICs is their stability [10]. This criterion is based on the reasonable assumption that ICs representing noise components are unlikely to be reproducible on different data subsets. Splitting the data into multiple subsets, however, reduces the amount of data for each ICA fold, thereby diminishing the overall quality of the decomposition. Furthermore, this criterion may fail to reject ICs representing muscular artifacts, as these are often systematically correlated with experimental conditions and may thus remain stable across different data subsets.

We address this problem by deriving a method that quantifies the neurophysiological plausibility of reconstructed ICs. Given an unmixing matrix computed by ICA, we use EEG/MEG volume conduction models to derive an estimate of the corresponding mixing matrix under the assumption of mutually independent cortical current sources. Deviations between expected- and observed source topographies then enable us to decide whether or not ICs are neurophysiologically plausible. We evaluate this method on EEG data, acquired from 19 subjects during a neurofeedback paradigm, and compare the obtained results with the two sorting schemes discussed above.

This paper is organized as follows. In Section II-A, we introduce the general ICA model and derive an equivalence relation between the unmixing- and mixing matrices involving the data covariance matrix. We then discuss how this relation can be used in conjunction with volume conduction models to quantify the neurophysiological plausibility of ICs (Section II-B). We apply this method to empirical data in Section III, and conclude with a brief discussion of various applications of our method in Section IV.

II. Methods

A. The ICA mixing model

We begin by stating the general ICA mixture model

$$\boldsymbol{x} = A\boldsymbol{s} \tag{1}$$

with $\boldsymbol{s}, \boldsymbol{x} \in \mathbb{R}^N$ the N original sources and EEG/MEG observations, respectively. The columns of the full-rank mixing matrix $A \in R^{N \times N}$ describe the projection of each source to every recording channel. The original source components are assumed to be mutually statistically independent, i.e. $p(\boldsymbol{s}) = \prod_{i=1}^{N} p(s_i)$. Without loss of generality, we assume each source to have zero mean and unit variance. Throughout this section we disregard further assumptions on $p(\boldsymbol{s})$ (such as non-Gaussianity) that are required to carry out blind source separation (cf. [11], [12]). Instead, we assume that we have access to an oracle that provides us with an unmixing matrix $W \in \mathbb{R}^{N \times N}$ such that the reconstructed ICs $\boldsymbol{s} = W\boldsymbol{x}$ are mutually statistically independent. We then have $W = A^{-1}$ (up to scaling and permutation [13]) and can estimate the topographies associated with $\boldsymbol{s}$ as the columns of $A = W^{-1}$.

Next, we note that the covariance matrix of (1) can be written as

$$\Sigma_{\boldsymbol{x}} = \mathrm{E}\{\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\} = \mathrm{E}\{A\boldsymbol{s}\boldsymbol{s}^{\mathrm{T}}A^{\mathrm{T}}\} = A\Sigma_{\boldsymbol{s}}A^{\mathrm{T}} \qquad (2)$$

with $\Sigma_{\boldsymbol{s}} \in \mathbb{R}^{N\times N}$ the source covariance matrix. Multiplying by W from the left we find that

$$W\Sigma_{\boldsymbol{x}} = WA\Sigma_{\boldsymbol{s}}A^{\mathrm{T}} = A^{\mathrm{T}}, \qquad (3)$$

as $W = A^{-1}$ and $\Sigma_{\boldsymbol{s}}$ the unit matrix by assumption of mutually independent sources with unit variance. We further note that the above relation holds for each individual IC, i.e.

$$\boldsymbol{w}_i^{\mathrm{T}}\Sigma_{\boldsymbol{x}} = \boldsymbol{a}_i^{\mathrm{T}} \qquad (4)$$

with $\boldsymbol{w}_i^{\mathrm{T}}$ the ith row of W and $\boldsymbol{a}_i$ the ith column of A. If the ICA model assumptions are fulfilled, we may thus multiply an IC's spatial filter by the data covariance matrix to obtain its associated source topography.

B. *Quantifying the neurophysiological plausibility of ICs*

For any data set recorded by EEG or MEG the data covariance matrix will not solely be determined by cortical current sources. Instead, noise sources as well as non-cortical artifacts, such as ocular- or muscular current sources, will also contribute to its shape. However, if the original sources can be recovered by a linear transformation, (4) holds independently of the origin of each source. Nevertheless, we argue in the following that (4) can be used to test whether a reconstructed source $s_i = \boldsymbol{w}_i^{\mathrm{T}}\boldsymbol{x}$ is likely to represent cortical current sources, or whether it may have been confounded by noise or non-cortical artifacts. Towards this goal, we first model the EEG/MEG data as

$$\boldsymbol{x} = L\boldsymbol{s}', \qquad (5)$$

with $L \in \mathbb{R}^{N\times K}$ a leadfield matrix that describes the projection of $K \gg N$ cortical current sources distributed throughout the brain to N recording channels [14]. Without loss of generality, we assume each source to have zero mean. We consider this model more realistic than the original ICA model (1), as it allows an arbitrary number of cortical columns to contribute to the brain's electromagnetic field. Under this model, the data covariance matrix is given by

$$\Sigma_{\boldsymbol{x}}^{\text{Model}} = L\Sigma_{\boldsymbol{s}'}L^{\mathrm{T}} \qquad (6)$$

with $\Sigma_{\boldsymbol{s}'} \in \mathbb{R}^{K\times K}$ the source covariance matrix of all cortical current sources. While L can be computed from biophysical volume conduction models [15], $\Sigma_{\boldsymbol{s}'}$ is in general unknown. In the absence of any further knowledge, we assume firstly that all cortical current sources are uncorrelated, and secondly that they contribute equally to the brain's electromagnetic field. We then have that $\Sigma_{\boldsymbol{s}'} = I$ the identity matrix. While the first assumption can be justified as a consequence of the ICA assumption of mutually independent sources, it remains to be established empirically whether the second assumption is warranted. Leaving this issue aside for the moment, we then have that

$$\Sigma_{\boldsymbol{x}}^{\text{Model}} = LL^{\mathrm{T}}. \qquad (7)$$

If $\Sigma_{\boldsymbol{x}}^{\text{Model}}$ is an accurate model of $\Sigma_{\boldsymbol{x}}$, we should find due to (4) that

$$\boldsymbol{w}_i^{\mathrm{T}}\Sigma_{\boldsymbol{x}} = \boldsymbol{w}_i^{\mathrm{T}}LL^{\mathrm{T}} = \boldsymbol{a}_i^{\mathrm{T}}. \qquad (8)$$

How well this relation is fulfilled in practice depends on whether $\Sigma_{\boldsymbol{x}}$ and $\Sigma_{\boldsymbol{x}}^{\text{Model}}$ are similar *in the direction that* $\boldsymbol{w}_i^T$ *points to*. As $\Sigma_{\boldsymbol{x}}^{\text{Model}}$ has been derived from the leadfields of cortical sources only, we argue that a violation of (8) indicates that $s_i = \boldsymbol{w}_i^{\mathrm{T}}\boldsymbol{x}$ represents a non-cortical source. In the following, we quantify such violations by the percentage of variance in the original topography that is not accounted for by the model-based topography, i.e.

$$r^2_{\text{Model}} = \left(\frac{\|\boldsymbol{w}_i^{\mathrm{T}}\Sigma_{\boldsymbol{x}} - \alpha_i\boldsymbol{w}_i^{\mathrm{T}}\Sigma_{\boldsymbol{x}}^{\text{Model}}\|_2}{\|\boldsymbol{w}_i^{\mathrm{T}}\Sigma_{\boldsymbol{x}}\|_2}\right)^2. \qquad (9)$$

Here, $\alpha_i \in \mathbb{R}$ is chosen by least-squares regression to minimize (9), thereby ensuring that r^2_{Model} is not confounded by different scaling of $\Sigma_{\boldsymbol{x}}$ and $\Sigma_{\boldsymbol{x}}^{\text{Model}}$. Sorting reconstructed ICs in ascending order according to their r^2_{Model}-values then allows us to rank ICs from most to least plausible. In the next section, we investigate empirically whether this ranking agrees with established methods for identifying cortical ICs.

III. Experiments

A. *Experimental data*

We recorded a 121-channel EEG, with active electrodes placed according to the extended 10-20 system, at 500 Hz from 19 healthy subjects during a neurofeedback paradigm. In this paradigm, subjects were trained in three 20-minute sessions to modulate parietal γ-range oscillations. All subjects gave informed consent in agreement with guidelines set by the Max Planck Society. A more detailed description of the data set is given in [16].

B. *ICA analysis*

For each subject's last two recording sessions, we first re-referenced the data to common average reference and high-pass filtered it with a third order Butterworth filter (cut-off frequency 3 Hz). We then computed the data covariance matrices $\Sigma_{\boldsymbol{x}}$ and $\Sigma_{\boldsymbol{x}}^{\text{Transfer}}$ on the data of the first and second recording session, respectively. The data of the first recording session was then separated into ICs by running the SOBI algorithm with default parameters [11]. As electrodes sometimes had to be switched off due to high impedances, the number of ICs varied between 112 and 119 across subjects. This resulted in a total of 2238 ICs.

We then computed a leadfield matrix $L \in \mathbb{R}^{121\times 15028}$, modeling the projection of $K = 15028$ current dipoles distributed throughout cortex to the $N = 121$ recording channels, using the Brainstorm toolbox [15]. Specifically, we used a four-shell spherical head model with standardized electrode locations. We then used this leadfield matrix to compute r^2_{Model} for each individual IC according to (9).

C. Empirical evaluation of IC ranking

We sorted the ICs of all subjects in ascending order according to their r^2_{Model}-values, and then compared this ranking with previously established methods for identifying plausible ICs. Firstly, we investigated whether ICs with small values of r^2_{Model} exhibit a dipolar topography, as dipolar topographies are considered typical for ICs representing mutually independent cortical current sources [8]. Secondly, we replaced $\Sigma_{\boldsymbol{x}}^{\text{Model}}$ in (9) by $\Sigma_{\boldsymbol{x}}^{\text{Transfer}}$ to obtain a measure of session-to-session stability of ICs. This score is subsequently termed r^2_{Transfer}. We then investigated whether ICs with small values of r^2_{Model} are also stable across recording sessions, as stable ICs are unlikely to represent noise sources [10].

D. Experimental results

The experimental results are illustrated in Figure 1. The blue line in Figure 1.A displays r^2_{Model} of every IC, sorted in ascending order. The steep slope of this line indicates that only a small percentage of ICs exhibit topographies that conform to those expected for purely cortical sources. The first column of Figure 1.B displays the topographies of a representative IC with a small r^2_{Model}-value, marked in Figure 1.A by a red *I*. Its original topography, shown in the first row, displays a clear dipolar structure that is very well reproduced by the model-based topography in the second row. In general, we found all ICs with a r^2_{Model}-value below 20% to exhibit a clear dipolar structure. Between 20% and 40% the majority of ICs still showed a dipolar topography, albeit less smooth than those with smaller r^2_{Model}-values. Above 40% source topographies appeared mostly cluttered (cf. the IC shown in the last column of Figure 1.B, marked in Figure 1.A by a red *IV*). As such, our method ranks those ICs as plausible that exhibit a dipolar topography considered typical for ICs representing mutually independent cortical current sources [8].

The capability of our method to rank sources according to their neurophysiological plausibility is further illustrated by a comparison of the r^2_{Model}- and r^2_{Transfer}-values in Figure 1.A, where each IC is represented by a black square:

1) Low r^2_{Model} & low $r^2_{Transfer}$: In this case, both, the model- and the stability-based criterion, rank an IC as plausible (left/bottom quadrant of Figure 1.A). The first column of Figure 1.B displays the topographies of an IC representative of this situation. Here, all three topographies exhibit the same dipolar structure considered typical of cortical current sources.

2) Low r^2_{Model} & high $r^2_{Transfer}$: In this case, the model-based criterion ranks an IC as plausible, while the session-to-session criterion would lead to a rejection of this IC. As evident from the right/bottom quadrant of Figure 1.A, only a few ICs fall into this category. We found that these ICs exhibited a clear dipolar structure in the original recording session that was not reproducible in the second session (cf. the topographies in the second column of Figure 1.B). We interpret this situation as a noise source corrupting the second recording session.

3) High r^2_{Model} & low $r^2_{Transfer}$: The ICs falling into this category are ranked as stable across sessions, yet do not conform to the model-based topography. A very large percentage of ICs fall into this category (left/top quadrant of Figure 1.A). Inspecting these ICs revealed that their topographies exhibited foci over peripheral electrodes that are typical for ICs representing muscular artifacts. The third column of Figure 1.B displays an IC representative of this class. As muscular artifacts are often consistent across recording sessions, these non-cortical ICs are not rejected based on their stability. The model-based criterion, however, correctly identifies these ICs as not representing cortical current sources.

4) High r^2_{Model} & high $r^2_{Transfer}$: ICs ranked as non-plausible by both criteria (right/top quadrant of Figure 1.A) consistently exhibited cluttered topographies as the ones displayed in the last column of Figure 1.B. These ICs most likely represent non-physiological noise sources.

In summary, our model-based ranking assigned a high plausibility to ICs that were found to be stable across recording sessions and did not exhibit topographies typical of muscular sources.

IV. Conclusion

In this work, we have presented a simple method to quantify the neurophysiological plausibility of reconstructed ICs. We presented empirical evidence that our method assigns low values of r^2_{Model} to ICs that show a dipolar topography, are stable across recordings sessions, and are not focused on peripheral electrodes. We interpret this as strong evidence for the capability of our method to identify neurophysiologically plausible ICs. This result also provides an empirical justification for the assumptions made in Section II-B.

At present, the inclusion or exclusion of ICs in the analysis of experimental data is often based on heuristics. Our method, on the other hand, provides a sound empirical basis for such decisions. We hope that this will contribute to a more objective decision on whether or not to reject an IC that is also easier to communicate to fellow researchers than a decision based on visual inspection of source topographies. We advise to reject all ICs with r^2_{Model}-values greater than 20%, but note that this rejection threshold depends on the intended level of rigour.

Finally, we would like to point out that another potential application of our method is the systematic evaluation of the quality of different ICA algorithms and data processing pipelines. For instance, an ICA decomposition can be scored according to the mean of r^2_{Model}, with lower values indicating

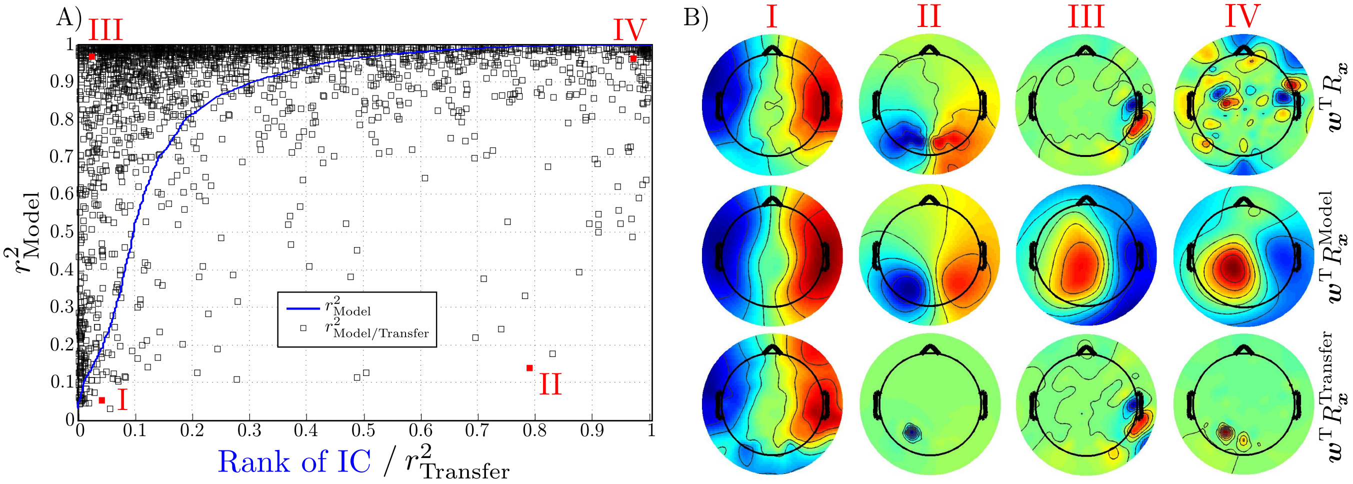

Figure 1. A) r^2_{Model} of all ICs sorted in ascending order (blue line) and relation of r^2_{Model} with r^2_{Transfer} of individual ICs (black squares). B) Representative examples of IC topographies with the original source topography in the top row, the model-based topography in the second row, and the session-to-session topography in the third row. Topographies plotted with EEGLAB [6].

a better decomposition. Once a leadfield matrix has been computed for a given setup, our method is computationally cheap and fully automatic. This makes it feasible to automatically optimize pre-processing pipelines or parameters of ICA algorithms with the goal of extracting a maximum of neurophysiologically plausible ICs from a given data set.

References

[1] T. Jung, S. Makeig, C. Humphries, T. Lee, M. McKeown, V. Iragui, and T. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 2, pp. 163–178, 2000.

[2] B. McMenamin, A. Shackman, J. Maxwell, D. Bachhuber, A. Koppenhaver, L. Greischar, and R. Davidson, "Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG," *NeuroImage*, vol. 49, pp. 2416–2432, 2010.

[3] S. Makeig, T. Jung, A. Bell, D. Ghahremani, and T. Sejnowski, "Blind separation of auditory event-related brain responses into independent components," *Proceedings of the National Academy of Sciences*, vol. 94, no. 20, p. 10979, 1997.

[4] S. Makeig, M. Westerfield, T. Jung, S. Enghoff, J. Townsend, E. Courchesne, and T. Sejnowski, "Dynamic brain sources of visual evoked responses," *Science*, vol. 295, no. 5555, p. 690, 2002.

[5] N. Xu, X. Gao, B. Hong, X. Miao, S. Gao, and F. Yang, "BCI competition 2003-data set IIb: enhancing P300 wave detection using ICA-based subspace projections for BCI applications," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1067–1072, 2004.

[6] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[7] L. Zhukov, D. Weinstein, and C. Johnson, "Independent component analysis for EEG source localization in realistic head models," *IEEE Engineering in Medicine and Biology Magazine*, vol. 19, no. 3, pp. 87–96, 2000.

[8] A. Delorme, J. Palmer, J. Onton, R. Oostenveld, and S. Makeig, "Independent EEG sources are dipolar," *PLoS One*, vol. 7, no. 2, p. e30135, 2012.

[9] M. Grosse-Wentrup and B. Schölkopf, "High gamma-power predicts performance in sensorimotor-rhythm brain-computer interfaces," *Journal of Neural Engineering*, vol. 55, pp. 1991–2000, 2012.

[10] J. Himberg and A. Hyvarinen, "ICASSO: Software for investigating the reliability of ICA estimates by clustering and visualization," in *IEEE 13th Workshop on Neural Networks for Signal Processing (NNSP'03)*. IEEE, 2003, pp. 259–268.

[11] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[12] T. Lee, M. Girolami, and T. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources," *Neural Computation*, vol. 11, no. 2, pp. 417–441, 1999.

[13] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[14] S. Baillet, J. Mosher, and R. Leahy, "Electromagnetic brain mapping," *IEEE Signal Processing Magazine*, vol. 18, no. 6, pp. 14–30, 2001.

[15] J. Mosher, S. Baillet, F. Darvas, D. Pantazis, E. Yildirim, and R. Leahy, "Brainstorm electromagnetic imaging software," in *5th International Symposium on Noninvasive Functional Source Imaging within the Human Brain and Heart (NFSI 2005)*, 2005.

[16] M. Grosse-Wentrup and B. Schölkopf, "A brain-computer interface based on operant conditioning of parietal gamma-oscillations," (under review).