ASSESSING RELIABILITY OF ICA PROJECTIONS – A RESAMPLING APPROACH

Frank Meinecke[1,2], *Andreas Ziehe*[1], *Motoaki Kawanabe*[1] *and Klaus-R. Müller*[1,2*]

[1]GMD FIRST.IDA, Kekuléstr. 7, 12489 Berlin, Germany
[2]University of Potsdam, Am Neuen Palais 10, 14469 Potsdam, Germany
{meinecke,ziehe,nabe,klaus}@first.gmd.de

ABSTRACT

When applying unsupervised learning techniques like ICA or temporal decorrelation for BSS, a key question is whether the discovered projections are reliable. In other words: can we give *error bars* or can we assess the *quality* of our separation? We use resampling methods to tackle these questions and show experimentally that our proposed variance estimations are strongly correlated to the separation error. We demonstrate that this reliability estimation can be used to choose an appropriate ICA-model, to enhance significantly the separation performance, and, most important, to mark the components that can really have a physical meaning. An application to data from an MEG[1]-experiment underlines the usefulness of our approach.

1. INTRODUCTION

Blind source separation (BSS) techniques have found widespread use in various application domains, e. g. acoustics, telecommunication or biomedical signal processing. [1, 2, 3, 4, 5, 6, 7, 8]). BSS is a statistical technique to reveal m unknown source signals $s_j(t)$ when only mixtures of them can be observed. For a linear mixture model, each of the $n \geq m$ observed signals $x_i(t)$ is assumed to be generated by

$$x_i(t) = \sum_{j=1}^{m} A_{ij} s_j(t).$$

In the following we will work in the framework of Independent Component Analysis (ICA), that means, we assume the source signals $s_j(t)$ to be statistically independent.

This is a typical unsupervised learning problem, since there exists no further knowledge about the source signals or the mixing matrix. Unfortunately, all unsupervised techniques suffer from the same fundamental dilemma: When applied to an arbitrary data set, they will always come up with some answer that is found within their model class, regardless of the applicability of the used model to this data.

* To whom the correspondence should be addressed

[1] magnetoencephalography

In the ICA-case, for example, the chosen algorithm will give an estimate for the mixing matrix (or the sources $s_j(t)$) even if the observed data contains no structure at all (e.g. if all $x_i(t)$ are Gaussian and iid in time); of course, the estimated projections are arbitrary in this case and totally useless.

One of the main questions of all unsupervised learning scenarios is therefore whether the result of the algorithm is reliable and displays inherent properties of the data or whether it is just a random result without meaning.

In this paper, we first show in general, how a reliability estimation for unsupervised learning algorithms can be carried out by resampling methods (section 2). Once we are able to estimate the reliability of a solution, we can use this information for model selection purposes as well as to improve the used algorithm. In section 3 we apply the proposed resampling techniques to ICA. We show how these techniques enable us to *select* a good *ICA-algorithm*, to *improve* the separation performance and to find potentially *meaningful* projection directions. We will give an algorithmic description of the resampling methods and show excellent experimental results. We conclude with a brief discussion.

2. RESAMPLING TECHNIQUES FOR UNSUPERVISED LEARNING

2.1. The Resampling Idea

In the typical unsupervised learning scenario we are given data $\{x_1, x_2, ...\}$ and we try to learn a set of parameters $\{\theta_1, \theta_2, ...\}$. Each of the parameters then is a function of the given data set: $\theta_i = \theta_i(x_1, x_2, ...)$.

The objective of resampling techniques is to produce surrogate data sets from the original data that eventually allow to approximate the variance of each parameter by a repeated learning of the parameter[2]. We will denote the resampling variance by $Var(\theta_i^*)$.

[2] In fact, to calculate the variance, we have to define a pairwise distance in the parameter space; the variance estimate is then given by the mean squared distance between the true solution and its replicas

2.2. The Resampling Recipe

The most popular resampling methods are the Jackknife and the Bootstrap (see e.g. [9]). The Jackknife produces surrogate data sets by just deleting one item each time from the original data set. There are generalizations of this approach like cross-validation which can delete more than one item each time. A more general approach is the Bootstrap. Consider a block of, say, N data points. For obtaining one bootstrap sample, we draw randomly N elements from the original data i.e. some data points might occur several times while others don't occur at all in the bootstrap sample.

These resampling methods have some desirable properties, which make them very attractive; for example, it can be shown that for the iid case the bootstrap estimators of the distributions of many commonly used statistics are consistent [9]. This means that it is possible to get an absolute, unbiased estimator for the true variance of a learned parameter. We will therefore call this an *absolute* estimator.

However, we need another estimator, to decide whether the result of our learning algorithm displays inherent structure of the data set or whether it could be explained as a random result. To get such an estimator, we define a *null hypothesis* (e.g. $\mathcal{N}$ = *"The data contains no structure."*). The exact form of $\mathcal{N}$ depends on the learning scenario. Once $\mathcal{N}$ is defined, it is straightforward to obtain a conditional probability $P(Var(\theta_i^*)|\mathcal{N})$ for the resampling variance, given this hypothesis is true.

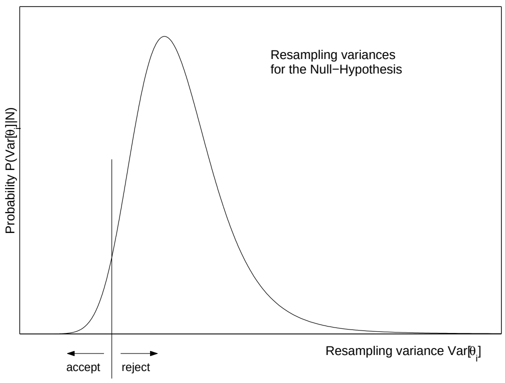

Fig. 1: The conditional probability $P(Var(\theta_i^*)|\mathcal{N})$ can easily obtained either analytically or numerically as a histogram. Demanding the likelihood $P(\mathcal{N}|Var(\theta_i^*))$ to be below a certain value defines a rejection threshold at the corresponding value of $Var(\theta_i^*)$.

For each measured resampling variance $Var(\theta_i^*)$ we can now ask for the likelihood $P(\mathcal{N}|Var(\theta_i^*))$ that it can be explained by the null hypothesis $\mathcal{N}$. If this likelihood exceeds a certain threshold, the result of the learning algorithm has to be rejected (cf. Fig. 1). Since this estimator measures the probability of a variance estimate relative to a null hypothesis, we will call it a *relative* estimator. Even if an unbiased *absolute* estimator for the variance of a solution may be hard to find, it is always quite easy to construct a relative estimator. In this case resampling can be done even with techniques that cannot be proved to provide consistent estimators.

Let us summarize the general resampling recipe:

1. Define a pairwise distance in the parameter space of the used unsupervised learning scenario.
2. Choose a resampling algorithm that reproduces the important properties of the data set; (e.g. in the ICA-case the resampling must leave the mixing matrix unchanged)
3. Define a null hypothesis to decide when a solution should be rejected.
4. Calculate the solution from the given data and the variance from the resampled data sets.

3. RESAMPLING ALGORITHMS FOR ICA

3.1. Application to ICA

In ICA, the parameters to be learned are the coefficients W_{ji} of the separating matrix that decomposes the given data $x_i(t)$ into independent components $y_j(t) = \sum W_{ji} x_i(t)$. Since both $\mathbf{A}$ and $\mathbf{s}(t)$ are unknown, it is impossible to recover the scaling or the order of the rows of the separating matrix $\mathbf{W}$. All that one can get are the projection *directions*. Consequently, we define the pairwise distance between two solutions (i.e. two separating matrices) as the angle difference between the respective projection directions. This provides a reliability check for each component.

The mixing/demixing process can be seen as a change of coordinates. From this point of view the data vector stays the same, but is expressed in different coordinate systems (passive transformation). Let $\{\mathbf{e}_i\}$ be the canonical basis of the estimated sources from the original data : $\mathbf{y} = \sum \mathbf{e}_i y_i$. Analogous, let $\{\mathbf{f}_j\}$ be the basis of the estimated sources on a bootstrap sample: $\mathbf{y}^* = \sum \mathbf{f}_j y_j^*$.

Using this, we can define a component-wise distance measure D_i as the angle differences between the directions of the independent component and the corresponding components learned on the resampled data:

$$D_i = \arccos\left(\frac{\mathbf{e}_i \cdot \mathbf{f}_i}{||\mathbf{e}_i|| \cdot ||\mathbf{f}_i||}\right).$$

To calculate this angle difference, recall that componentwise we have $y_j^* = \sum W_{jk}^* W_{ki}^{-1} y_i$. With $\mathbf{y}^* = \mathbf{y}$, this leads to: $\mathbf{f}_j = \sum \mathbf{e}_i (\mathbf{W}^* \mathbf{W}^{-1})_{ij}^{-1}$, i.e. $\mathbf{f}_j$ is the j-th column of $\mathbf{W}\mathbf{W}^{*-1}$. (In fact, we first have to permute the rows of W^* into the right order. However, we may circumvent this difficulty by the choice of our resampling algorithm.)

If the true mixing matrix is known, we can define the separation error E_i for each source in the same manner as angle difference between the true source and the corresponding learned independent component.

In this paper we illustrate our resampling method for two commonly used BSS algorithms.
The first method, JADE [6], is based on higher order statistics and uses the joint diagonalization of matrices obtained from 'parallel slices' of the fourth order cumulant tensor to identify the mixing matrix. The second one, TDSEP [7], relies on the time structure of signals and requires only second order statistics, i.e. enforcing temporal decorrelation between channels.

Time structure preserving bootstrap. For JADE the resampling can be applied in a straightforward manner. However, the simple bootstrap approach destroys time structure, but it can easily be generalized in such a way that temporal decorrelation methods like TDSEP are still applicable. The bootstrap resampling defines a series $\{a_t\}$ with each a_t indicating how often the data point $\mathbf{x}(t)$ has been drawn. Using this, we can calculate the resampled time-lagged correlation matrices as

$$C(\tau) = \frac{1}{T}\sum_{t=1}^{T} a_t \cdot \mathbf{x}(t)\mathbf{x}^T(t+\tau)$$

with $\sum a_t = T$ and $a_t \in \{0, 1, 2, ...\}$.

The filter trick. Another way of generating time structure preserving surrogate data is, for example, to apply a (random) linear filter $\mathcal{F}$ on the measured (mixed) data:

$$\mathcal{F}x_i(t) = \sum_{\tau=0}^{T} f_\tau \cdot x_i(t-\tau) = \sum_j A_{ij} \cdot \mathcal{F}s_j(t)$$

Since the mixing matrix A commutes with this filtering operator ($[\mathcal{F}, A] = 0$) and the filtered sources $s'_j(t) = \mathcal{F}s_j(t)$ are still mutually independent, the filtered signals $x'_i(t) = \mathcal{F}x_i(t)$ can be interpreted as linear mixtures of the filtered sources with the same mixing matrix A.

Null hypothesis. As null hypothesis in this ICA case we choose simply: $\mathcal{N} =$ *"The given data is generated by a Gaussian and iid source"*. Due to the symmetry of a normally distributed data set it is impossible, to estimate the projection directions, i.e. the result of the ICA algorithm will be completely produced by random fluctuations. In the following we will accept a solution only if $P(\mathcal{N}|Var(\theta_i^*)) \leq 5\%$. Note, that $P(Var(\theta_i^*)|\mathcal{N})$ depends strongly on the size of the data set, since random fluctuations are more dominant in small data sets. This may produce reasonable stable solutions even if the generating distribution fulfills $\mathcal{N}$. The relative likelihood estimator then prevents us from misinterpretation of the obtained result. For bigger data sets with more channels, however, the chosen null hypothesis may be too restricted to rule out certain results. In this ICA case, we face the difficulty that $\mathcal{N}$ makes a statement about *all* sources, but we want to obtain variance estimates for each single source. Depending on what we are looking for, it is often possible to refine the definition of $\mathcal{N}$; by using even several null hypotheses. The important fact is that low $P(\mathcal{N}|Var(\theta_i^*))$ does not necessarily mean that the result is meaningful, it only means that the result displays structure of the data that cannot be explained by $\mathcal{N}$. Usually, both, an absolute estimate and one (or several) relative estimates are needed to make a prediction about the reliability of the result of an unsupervised learning algorithm. First, all results that can be explained by the null hypothesis have to be discarded, then, the remaining results can be judged according to their absolute variance estimates.

3.2. The Resampling Algorithm

We will now give a short description of our resampling algorithm. To omit the permutation problem, we first separate the sources and then do the resampling only on the rotation part of the separating matrices obtained from the surrogate data sets.

After performing BSS, the estimated ICA-projections are used to generate surrogate data by resampling. On the whitened[3] surrogate data, the source separation algorithm is used again to estimate a rotation that separates this surrogate data.

To compare different rotation matrices, we use the fact that the matrix representation of the rotation group $SO(N)$ can be parameterized by $R(\alpha) = \exp\left(\frac{1}{2}\sum_{i,j}\alpha_{ij}\mathbf{M}^{ij}\right)$ with $\left(\mathbf{M}^{ij}\right)_{ab} = \delta_a^i\delta_b^j - \delta_a^j\delta_b^i$, where the matrices $\mathbf{M}^{ij}$ are generators of the group and the α_{ij} are the rotation parameters (angles) of the rotation matrix R. Using this parametrization we can easily compare different N-dimensional rotations by comparing the rotation parameters α_{ij}.

Since the sources are already separated, the estimated rotation matrices will be close to the identity matrix[4]. The quantity Var(α_{ij}) measures the instability of the separation with respect to a rotation in the (i,j)-plane. Since the reliability of a projection is bounded by the maximum angle variance of all rotations that affect this direction, we define the uncertainty of the i-th ICA-projection as: $U_i := \max_j \sqrt{Var(\alpha_{ij})}$.
Let us summarize the resampling algorithm:

1. Estimate the separating matrix $\mathbf{W}$ with some ICA algorithm. Calculate the ICA-projections $\mathbf{y} = \mathbf{W}\mathbf{x}$.
2. Produce k surrogate data sets from $\mathbf{y}$ and whiten these data sets.
3. For each surrogate data set: produce a set of rotation matrices by performing ICA.
4. Calculate variances of rotation parameters (angles) α_{ij}.
5. For each ICA component calculate the uncertainty $U_i = \max_j \sqrt{Var(\alpha_{ij})}$.

[3] The whitening transformation is defined as $\mathbf{x}' = \mathbf{V}\mathbf{x}$ with $\mathbf{V} = E[\mathbf{x}\mathbf{x}^T]^{-1/2}$.

[4] For the following interpretation it is important to perform the resampling when the sources are already separated, so that the α_{ij} are distributed around zero, because $SO(N)$ is a non-Abelian group; i.e. in general $R(\alpha)R(\beta) \neq R(\beta)R(\alpha)$.

3.3. Asymptotic Considerations for Resampling

Properties of resampling methods are typically studied in the limit when the number of bootstrap samples $B \to \infty$ and the length of signal $T \to \infty$ [9]. In the case of bootstrap resampling, as $B \to \infty$, the bootstrap variance estimator $U_i^*(B)$ computed from the α_{ij}^*'s converge to $U_i^*(\infty) := \max_j \sqrt{\mathrm{Var}_{\hat{F}}[\alpha_{ij}^*]}$ where α_{ij}^* denotes the resampled angle deviation and $\hat{F}$ denotes the distribution generating it. Furthermore, if $\hat{F} \to F$, $U_i^*(\infty)$ converges to the true variance $U_i = \max_j \sqrt{\mathrm{Var}_F[\alpha_{ij}]}$ as $T \to \infty$. This is the case, for example, if the original signal is i.i.d. in time. When the data has time structure, $\hat{F}$ does not necessarily converge to the generating distribution F of the original signal anymore. Although we cannot neglect this difference completely, it is small enough to use our scheme for the purposes considered in this paper, e.g. in TDSEP, where the α_{ij} depend on the variation of the time-lagged covariances $C_{ij}(\tau)$ of the signals, we can bound the difference $\Delta_{ij} = \mathrm{Var}_{\hat{F}}[\hat{C}_{ij}^*(\tau)] - \mathrm{Var}_F[\hat{C}_{ij}(\tau)]$ between the real variation and its bootstrap estimator as

$$|\Delta_{ij}| \leq \begin{cases} \frac{1}{T}\left\{\frac{2a^2}{1-a^2}M^2\left\{1+a^{2\tau}\right\}+2\tau a^{2\tau}\right\}, & j=i \\ \frac{1}{T}\frac{2a^2}{1-a^2}M^2, & j \neq i \end{cases}$$

if $\exists a < 1, M \geq 1, \quad \forall i: \quad |C_{ii}(\tau)| \leq M a^{\tau} |C_{ii}(0)|$. In our experiments, however, the bias is usually found to be much smaller than this upper bound.

For the filter trick it is not so easy to show theoretically whether it can be used as a absolute reliability estimate. But, as stated before, there is no restriction for using it as a relative estimate. However, our experiments show numerically that the filter trick may provide good absolute variance estimates as well.

4. EXPERIMENTS

In every experiment that is reported here, we used both the bootstrap and the filter trick. Remarkably, the results are almost identical. The following figures show the results obtained by the filter - resampling (filter length 5), the filter coefficients are Gaussian random numbers and the used resampling size is B=100 in all cases.

4.1. Separation Error vs. Uncertainty Estimate

To show the practical applicability of the resampling idea to ICA, we compare the separation error E_i with the uncertainty U_i. The separation was performed on artificial 2D mixtures of speech and music signals and iid data sets of the same variance (1000 data points). To achieve different separation qualities, white Gaussian noise of different intensity has been added to the mixtures. Figure 2 relates the uncertainty to the separation error for JADE (TDSEP results look qualitatively the same). In Fig. 2 (a) we see the separation error distribution which has a strong peak for small values of our uncertainty measure, whereas for large uncertainties it tends to become flat, i.e. – as also seen from Fig. 2 (b) – the uncertainty reflects very well the true separation error.

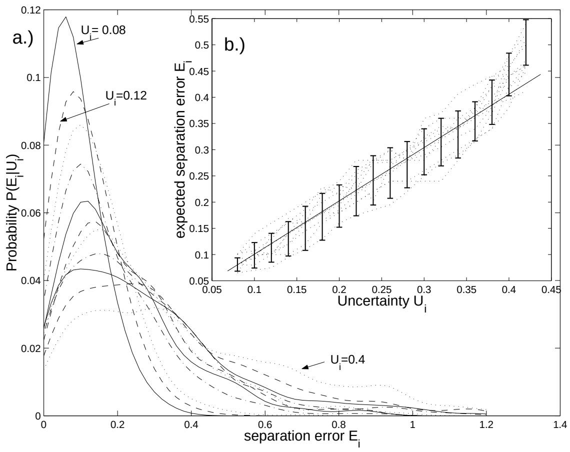

Fig. 2: (a) The probability distribution for the separation error for a small uncertainty is close to zero, for higher uncertainty it spreads over a larger range. (b) The expected error increases with the uncertainty.

4.2. Selecting the appropriate BSS algorithm

As our variance estimate is highly correlated with the (true) separation error, it seems promising to use it as a model selection criterion for: (a) selecting some hyperparameter of the BSS algorithm, e.g. choosing the lag values for TDSEP or (b) choosing between a set of different algorithms that rely on different assumptions about the data, i.e. higher order statistics (e.g. JADE, INFOMAX, FastICA, ...) or second order statistics (e.g. TDSEP). It could, in principle, be much better to extract the first component with one and the next with another assumption. To illustrate the usefulness of our reliability measure, we study a five-channel mixture of two signals of pure white Gaussian noise, two audio signals and one signal of uniformly distributed noise.

The reliability analysis for JADE gives the advice to rely only on channels 1, 3 and 5 (cf. Fig. 3 left). In fact, these are the channels that contain the audio signals and the uniformly distributed noise. The components 2 and 4 span the subspace that contains the Gaussian noise. For these channels the uncertainty is above the threshold that we fixed for the likelihood of the null hypothesis (horizontal line). The same analysis applied to the TDSEP-projections (time lags $0, \ldots, 20$) shows that TDSEP can give reliable estimates only for the two audio sources (cf. Fig. 3 right) which has

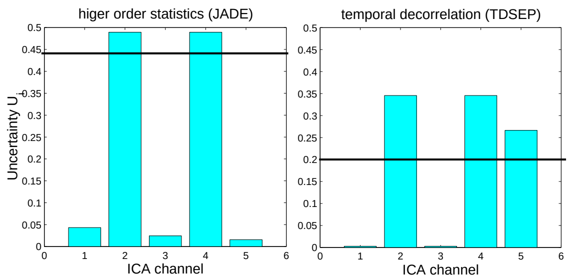

Fig. 3: Uncertainty of ICA projections of an artificial mixture using JADE and TDSEP. Resampling displays the strengths and weaknesses of the different models. The horizontal line indicates the rejection threshold for the null hypothesis.

been expected because the noise is iid. According to our measure, the estimation for the audio signals is more reliable in the TDSEP-case. Calculation of the separation error verifies this: TDSEP separates better by about 2 orders of magnitude (JADE: $E_1 = 6.1 \cdot 10^{-2}, E_3 = 3.3 \cdot 10^{-2}$, TDSEP: $E_1 = 8.2 \cdot 10^{-4}, E_3 = 3.7 \cdot 10^{-4}$). Finally, in our example, estimating the audio sources with TDSEP and after this applying JADE to the orthogonal subspace, gives the optimal solution since it combines the small separation errors E_3, E_4 for TDSEP with the ability of JADE to separate the uniformly distributed noise.

4.3. Blockwise uncertainty estimates

For a longer time series it is not only important to know which ICA channels are reliable, but also to know whether different parts of a given time series are more (or less) reliable to separate than others. To demonstrate these effects, we mixed two audio sources[5], where the mixtures are partly corrupted by white Gaussian noise. Reliability analysis is performed on windows of length 1000, shifted in steps of 250; the resulting variance estimates are smoothed. Fig. 4 shows again that the uncertainty measure is nicely correlated with the true separation error, furthermore the variance goes systematically up within the noisy part but also in other parts of the time series that do not seem to match the assumptions underlying the algorithm.[6] So our reliability estimates can eventually be used to improve separation performance by removing all but the 'reliable' parts of the time series. For our example this reduces the overall separation error by 2 orders of magnitude from $2.4 \cdot 10^{-2}$ to $1.7 \cdot 10^{-4}$. This moving-window resampling can detect instabilities of the projections in two different ways: Besides the resampling variance that can be calculated for each window, one can also calculate the change of the projection directions

[5] recording 10s at 8kHz $= 80,000$ data points

[6] For example, the peak in the last third of the time series can be traced back to the fact that the original time series are correlated in this region.

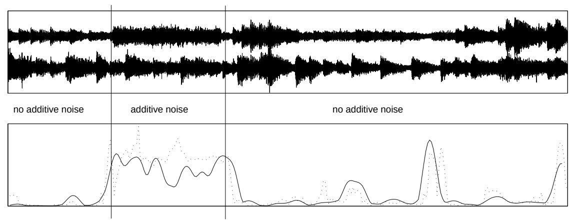

Fig. 4: Upper panel: mixtures, partly corrupted by noise. Lower panel: the blockwise variance estimate (solid line) vs the true separation error on this block (dotted line).

between two windows. The later has already been used successfully by Makeig et. al. [10].

5. ASSIGNING MEANING: APPLICATION TO BIOMEDICAL DATA

We now apply our reliability analysis to biomedical data that has been produced by an MEG experiment with acoustic stimulation. The stimulation was achieved by presenting alternating periods of music and silence, each of 30s length, to the subjects right ear during 30 min. of total recording time (for details see [11]). The measured DC magnetic field values, sampled at a frequency of 0.4 Hz, gave a total number of 720 sample points for each of the 49 channels. While previously [11] analyzing the data, we found that many of the ICA components are seemingly meaningless and it took some medical knowledge to find potential meaningful projections for a later close inspection. However, our reliability assessment can also be seen as indication for meaningful projections, i.e. *meaningful* components should have *low* variance. In the experiment, BSS was performed on the 23 most powerful principal components using (a) higher order statistics (JADE) and (b) temporal decorrelation (TDSEP, time lag 0..50). The results in Fig. 5 show that none

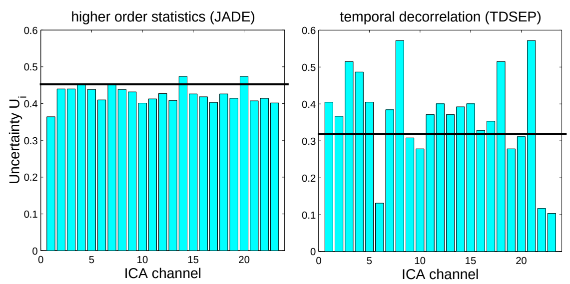

Fig. 5: Resampling on the biomedical data from MEG experiment shows: (a) no JADE projection is reliable (has low uncertainty) (b) TDSEP is able to identify three sources with low uncertainty. Horizontal line: rejection threshold

of the JADE-projections (left) have small variance whereas TDSEP (right) identifies three sources with a good reliability. In fact, these three components have physical meaning: while component 23 is an internal very low frequency signal (drift) that is always present in DC-measurements, interestingly component 6 shows a (noisy) rectangular waveform that clearly displays the $1/30s$ on/off characteristics of the stimulus (correlation to stimulus 0.7; see Fig. 6). The

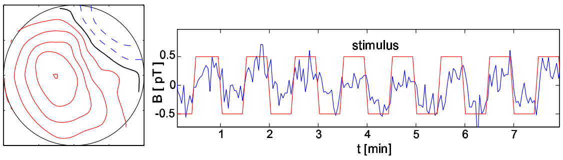

Fig. 6: Spatial field pattern and time course of TDSEP channel 6.

clear dipole-structure of the spatial field pattern in Fig. 6 underlines the relevance of this projection. The components found by JADE do not show such a clear structure and the strongest correlation of any component to the stimulus is about 0.3, which is of the same order of magnitude as the strongest correlated PCA-component before applying JADE.

6. DISCUSSION

We proposed a simple method based on resampling techniques to estimate the reliability of results that can be obtained from unsupervised learning algorithms. After briefly discussing the general resampling idea, we applied it to the ICA scenario and showed, that our technique really approximates the separation error, several directions are open(ed) for applications. First, we may like to use it for model selection purposes to distinguish between algorithms or to chose good hyperparameters (possibly even component-wise). Second, variances can be estimated on blocks of data and separation performance can be enhanced by using only low variance blocks where the model matches the data nicely. Finally reliability estimates can be used to find meaningful components. Here our assumption is that the more meaningful a component is, the more stably we should be able to estimate it. In this sense artifacts appear of course also as meaningful, whereas noisy directions are discarded easily, due to their high uncertainty.
Future research will focus on the application of resampling techniques to other unsupervised learning scenarios (e.g. for clustering [12]). We intend to consider also Bayesian modellings where often a variance estimate comes for free, along with the trained model.

7. REFERENCES

[1] Ch. Jutten and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1–10, 1991.

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

[3] G. Deco and D. Obradovic, *An information-theoretic approach to neural computing*, Springer, New York, 1996.

[4] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems (NIPS 95)*, D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo, Eds. 1996, vol. 8, pp. 882–893, The MIT Press.

[5] A. J. Bell and T. J. Sejnowski, "An information maximisation approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[6] J.-F. Cardoso and Antoine Souloumiac, "Blind beamforming for non Gaussian signals," *IEEE Proceedings-F*, vol. 14O, no. 6, pp. 362–370, Dec. 1994.

[7] A. Ziehe and K.-R. Müller, "TDSEP – an efficient algorithm for blind separation using time structure," in *Proc. Int. Conf. on Artificial Neural Networks (ICANN'98)*, L. Niklasson, M. Bodén, and T. Ziemke, Eds., Skövde, Sweden, 1998, pp. 675 – 680, Springer Verlag.

[8] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley, 2001.

[9] J. Shao and D. Tu, *The Jackknife and Bootstrap*, Springer, New York, 1995.

[10] S. Makeig, S. Enghoff, T.-P. Jung, and T. Sejnowski, "Moving-window ICA decomposition of EEG data reveals event-related changes in oscillatory brain activity," in *Proc. 2nd Int. Workshop on Independent Component Analysis and Blind Source Separation (ICA'2000)*, Helsinki, Finland, 2000, pp. 627–632.

[11] G. Wübbeler, A. Ziehe, B.-M. Mackert, K.-R. Müller, L. Trahms, and G. Curio, "Independent component analysis of non-invasively recorded cortical magnetic dc-fields in humans," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 5, pp. 594–599, 2000.

[12] V. Roth, T. Lange, M. Braun, and J. Buhmann, "A resampling based approach to cluster validation," 2001, submitted to NIPS.

Acknowledgments K.-R.M thanks Guido Nolte and the members of the Oberwolfach Seminar September 2000 in particular Lutz Dümbgen and Enno Mammen for helpful discussions and suggestions. K. -R. M and A. Z. acknowledge partial funding by the EU project (IST-1999-14190 – BLISS). We thank the biomagnetism group of PTB for providing the DC-MEG data.