Explorative data analysis for changes in neural activity

2013 J. Neural Eng. 10 026018

(http://iopscience.iop.org/1741-2552/10/2/026018)

J. Neural Eng. **10** (2013) 026018 (16pp) doi:10.1088/1741-2560/10/2/026018

Explorative data analysis for changes in neural activity

Duncan A J Blythe[1,2], **Frank C Meinecke**[1], **Paul von Bünau**[1] **and Klaus-Robert Müller**[1,3]

[1] Department of Computer Science, Berlin Institute of Technology, Sekretariat FR 6-9 Franklinstr 28/29, D-10587 Berlin, Germany
[2] Bernstein Centre for Computational Neuroscience, Humboldt-Universität zu Berlin Philippstr 13, Haus 6, D-10115 Berlin, Germany
[3] Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea

E-mail: duncan.blythe@bccn-berlin.de

Received 7 January 2013
Accepted for publication 14 February 2013
Published 15 March 2013
Online at stacks.iop.org/JNE/10/026018

Abstract
Neural recordings are non-stationary time series, i.e. their properties typically change over time. Identifying specific changes, e.g., those induced by a learning task, can shed light on the underlying neural processes. However, such *changes of interest* are often masked by strong *unrelated changes*, which can be of physiological origin or due to measurement artifacts. We propose a novel algorithm for disentangling such different causes of non-stationarity and in this manner enable better neurophysiological interpretation for a wider set of experimental paradigms. A key ingredient is the repeated application of Stationary Subspace Analysis (SSA) using different temporal scales. The usefulness of our explorative approach is demonstrated in simulations, theory and EEG experiments with 80 brain–computer interfacing subjects.

(Some figures may appear in colour only in the online journal)

1. Introduction

In analyzing multivariate time series, as, e.g., recorded in neurophysiological experiments, we face a challenge: artifacts, signals resulting from different types of brain activity—task-relevant and task-irrelevant—and changes thereof are observed as a highly variable stream of data.

Some of this variability can be attributed to noise [26], some to learning and plasticity [20] or also to unknown latent variables [11]; in other words, the data exhibit changes on many scales. Explorative data analysis methods such as principal component analysis (PCA) [25], independent component analysis (ICA) [12] and other projection methods can contribute to disentangling confounding trends from the data [22]. However, they generally assume an underlying stationary distribution of the data, so if distribution changes occur, say, due to learning processes or due to lapses of attention, they will yield suboptimal results.

This paper addresses the question of finding *changes* in multivariate neural data which occur over time; we propose a model that enables for the first time the extraction of both global behavioral condition changes and experimental condition-specific changes. To further illustrate this abstract scenario, a concrete example from the field of brain–computer interfacing (BCI) is as follows: the BCI user is instructed to perform a certain number of motor imagery commands, for example, imaginations of left hand, right hand and foot movements. Because the user *learns* to use the BCI over time, the data under each of these conditions may change over time: these are changes *specific* to each of the conditions[4]. Moreover, users often become tired while using the BCI. This results in higher levels of alpha activity (≈7–15 Hz) present under *all* conditions as the experiment progresses [5]. This increase in alpha would correspond to a *background* change in neural activity. We would like to model these two types of changes separately so they may be examined in isolation.

[4] Users can in fact learn different types of imaginations at different speeds, similarly to the ability of a tennis player to learn his fore- and backhand strikes to differing levels of proficiency.

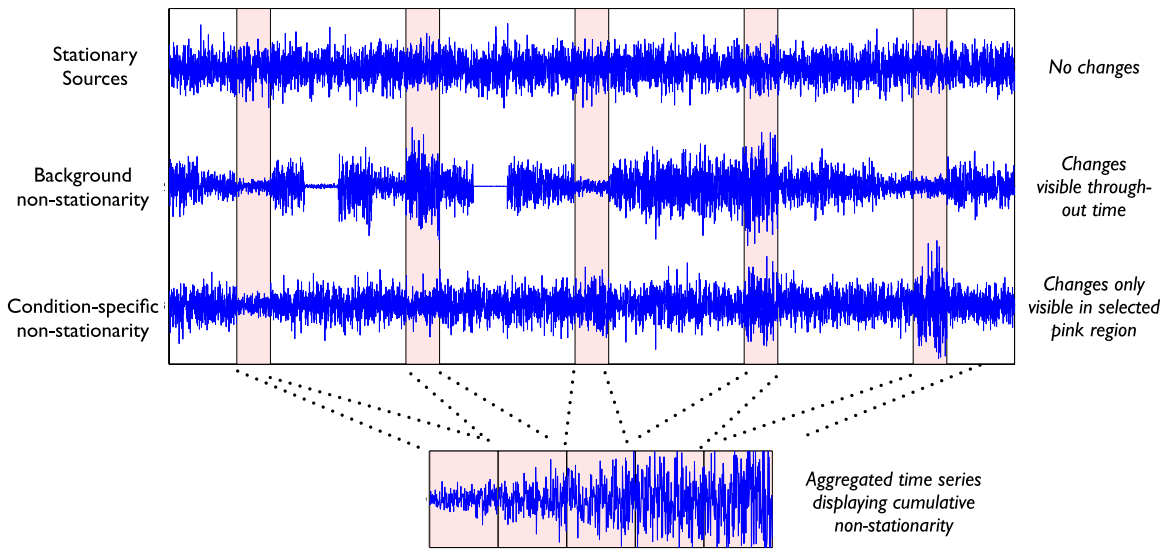

Figure 1. The problem setting is illustrated: changes specific to the data under one condition are masked by stronger changes taking place over all conditions. The top time series displays the data from all conditions, with a selected target condition marked in red.

In principle, one might think that the task of finding changes in the time domain of a data set has already been solved; one may simply optimize an objective function which measures change over time on some subset of the data, as proposed for the Stationary Subspace Analysis (SSA) [33]. However, in practice, these changes may not *all* be relevant to the data analytic task. This is illustrated in figure 1; suppose the data are split into experimental conditions and we are interested in changes which are particular to only *one* of these conditions. The first time series displays data from all conditions with the data from the target condition of interest highlighted. We see that the condition of interest displays a weak change, not contained in the other conditions. However, we also observe that there are other *stronger* changes (for example, related to artifacts) under all conditions. Fortunately, the weaker change of interest, which is specific to the highlighted condition, has a different *origin* than the changes affecting all conditions: as we will see below, this situation will play an important role in our approach below[5].

Scenarios where similarly complex changes are hidden in the data arise in many experimental paradigms in neurophysiology [23, 29]. So far, this fact has been ignored since there was no sound algorithmic solution to address the issue. The presented framework therefore aims to contribute to a better understanding of complex experimental paradigms which involve distributional changes *by construction*, e.g., paradigms focusing on learning and plasticity [20].

We present our new model more formally in section 2 and outline a method for removing the background changes in section 2.2, for which we provide theoretical guarantees in section 2.3. We then evaluate the quality of the estimation method in simulations in section 3. Lastly, the analysis of the BCI data sets is covered in section 4.

[5] In high-dimensional data setups, the fact that these different types of changes have the same origin has very low probability. Moreover, neurophysiological considerations imply that the two sources of change are more likely to originate from different points in the brain.

2. Model and parameter estimation

In this section, we introduce the generative model underlying our approach and a method for estimating its parameters, as well as theoretical results that provide us with performance guarantees.

2.1. Generative model

The generative model is an extension of the SSA linear mixing model for a multivariate non-stationary time series [33][6]. In this model, the observed multivariate data, $\mathbf{x}(t)$, are generated as a linear superposition of two groups of latent (only indirectly observable) sources. The first group, $\mathbf{s}^{\mathfrak{s}}(t) = [s_1(t), \ldots, s_d(t)]^\top$, is stationary, and the second group, $\mathbf{s}^{\mathfrak{n}}(t) = [s_{d+1}(t), \ldots, s_D(t)]^\top$, is non-stationary. In this context, stationarity of a time series is defined as the first two moments (mean and spatial covariance) being constant over time; the so-called weak stationarity is usually defined to include the temporal cross-covariance being constant over time; however, in the current contribution, we concentrate on the spatial covariance (at time lag 0)[7]. Moreover, for the purpose of our final electroencephalography (EEG) analysis (see section 4), we will ignore the mean (since the data are high-pass-filtered) and focus only on changes in the covariance matrix, which equates to the signal power when projected to a univariate time series. The observed D-variate data, $\mathbf{x}(t)$, are generated as a linear transformation of the two groups of stationary and non-stationary sources by an unknown square mixing matrix A, given in terms of the $D \times d$ and $D \times (D-d)$ rectangular matrices $A^{\mathfrak{s}}$ and $A^{\mathfrak{n}}$.

[6] Related mixing models often applied in neuroscience include the ICA model and the PCA model. For background on linear mixing models and source separation, see [12, 33].

[7] In this paper, we consider time series with vanishing autocorrelations; thus, wherever 'covariance' is printed, we mean to refer to a spatial covariance.

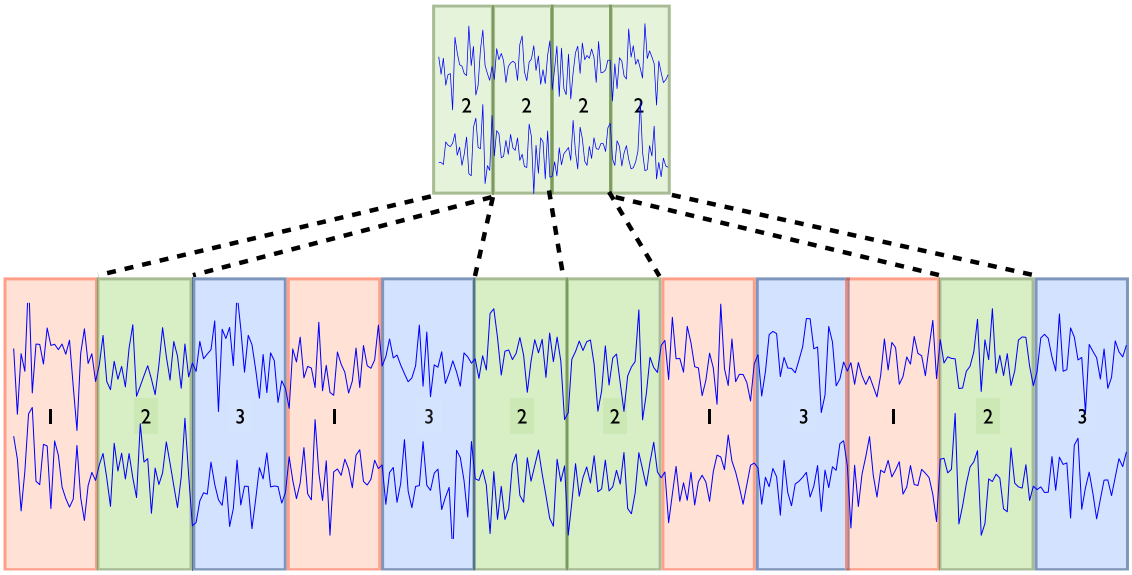

Figure 2. The block trial structure of the multivariate data which we model is displayed. Trials recorded under separate behavioral conditions are depicted as colored blocks. A trial from any condition occurs with equal probability. How the data from the second condition are formed is illustrated above.

$$\mathbf{x}(t) = A\mathbf{s}(t) = [A^{\mathfrak{s}}\ A^{\mathfrak{n}}]\begin{bmatrix}\mathbf{s}^{\mathfrak{s}}(t)\\ \mathbf{s}^{\mathfrak{n}}(t)\end{bmatrix}. \quad (1)$$

The matrix A is called the mixing matrix because the entry A_{ij} determines how the jth source contributes to the ith dimension of the observed data. If the sensors corresponding to the measurements are spatially distributed, then the columns of the mixing matrix can be interpreted visually as patterns, e.g., scalp maps in the EEG analysis. The columns of the matrices $A^{\mathfrak{s}}$ and $A^{\mathfrak{n}}$ span the stationary and non-stationary subspaces, respectively. Note that the column spaces of these matrices need not be orthogonal. The aim of SSA is to invert this mixing, i.e. to find a demixing matrix which allows us to recover the two groups of sources from the recorded mixed signals.

However, our setting differs from the SSA model (given in equation (1)) in that we assume that the observations are generated as the *sum of two latent data generating systems*: a background system and a condition-specific system—where the latter contains the information of interest and the former introduces irrelevant distribution changes which we want to remove. In line with these assumptions, we model several time series corresponding to each experimental condition; this is at first glance inconsistent with all data being drawn in practice from a single empirical time series (corresponding to the samples recorded from the measurement device) but in fact, this correctly captures the fact that each condition is subject to separate parameters.

In order to obtain these empirical time series from the recorded data, we assume that for each condition, several trials (for example, corresponding to a single stimulus presentation or motor imagery, etc) are recorded; moreover, we assume that conditions occur with equal probability in the sequence of trials[8]. The trials may be assumed to be approximately stationary (although this is not essential) and non-stationarity exhibits itself between trials. Then, the data from the kth experimental condition may be obtained by concatenating all of the trials recorded under that condition.

Figure 2 illustrates the setup of the data with three conditions recorded. For instance, in a typical EEG motor imagery BCI session, several conditions are recorded: left imagery, right imagery and often foot imagery in addition, to cater for the possibility that discrimination between some pair of conditions is of higher quality than the remaining pairs. A trial for one of these conditions consists of the data recorded during a single imagined movement, for example, an imagined left-hand movement in the left condition[9].

These considerations allow us to plausibly model the observed data $\mathbf{x}_k(t)$ under condition k at time t (within condition time index) as generated by the sum

$$\mathbf{x}_k(t) = A\mathbf{s}(t) + B_k\mathbf{r}_k(t) \quad (2)$$

$$= [A^{\mathfrak{s}}\ A^{\mathfrak{n}}]\begin{bmatrix}\mathbf{s}^{\mathfrak{s}}(t)\\ \mathbf{s}^{\mathfrak{n}}(t)\end{bmatrix} + [B_k^{\mathfrak{s}}\ B_k^{\mathfrak{n}}]\begin{bmatrix}\mathbf{r}_{\mathbf{k}}^{\mathfrak{s}}(t)\\ \mathbf{r}_{\mathbf{k}}^{\mathfrak{n}}(t)\end{bmatrix}, \quad (3)$$

where $A\mathbf{s}(t)$ is the contribution of the background system and $B_k\mathbf{r}_k(t)$ is the condition-specific contribution[10]. The corresponding inverses are $A^{-1} = P = \left[\begin{smallmatrix}P^{\mathfrak{s}}\\ P^{\mathfrak{n}}\end{smallmatrix}\right]$ and $B_k^{-1} = Q_k = \left[\begin{smallmatrix}Q_k^{\mathfrak{s}}\\ Q_k^{\mathfrak{n}}\end{smallmatrix}\right]$. These matrices stay constant over time, whereas the sources, $\mathbf{s}(t)$, etc, change over time. Moreover, the matrix A is constant over all K conditions, whereas B_k is a variable which

[8] Randomization of conditions is a standard practice in many experimental designs.

[9] Note that in practice, the trials under each condition may have different lengths. This will generate a bias in estimation if not appropriately dealt with. One possible solution is to subsample segments of the longer condition so that all trials are forced to be of the same duration. This implies that there will be sections of unused data where perhaps beforehand there were not. This is not an obstacle to the application of our method.

[10] Similarly to the standard SSA model, $A^{\mathfrak{s}}$ and $A^{\mathfrak{n}}$ are of size $d \times D$ and $(D-d) \times D$ and $B_k^{\mathfrak{s}}$ and $B_k^{\mathfrak{n}}$ are of size $d_k \times D$ and $(D-d_k) \times D$ corresponding to the numbers of stationary sources in each system.

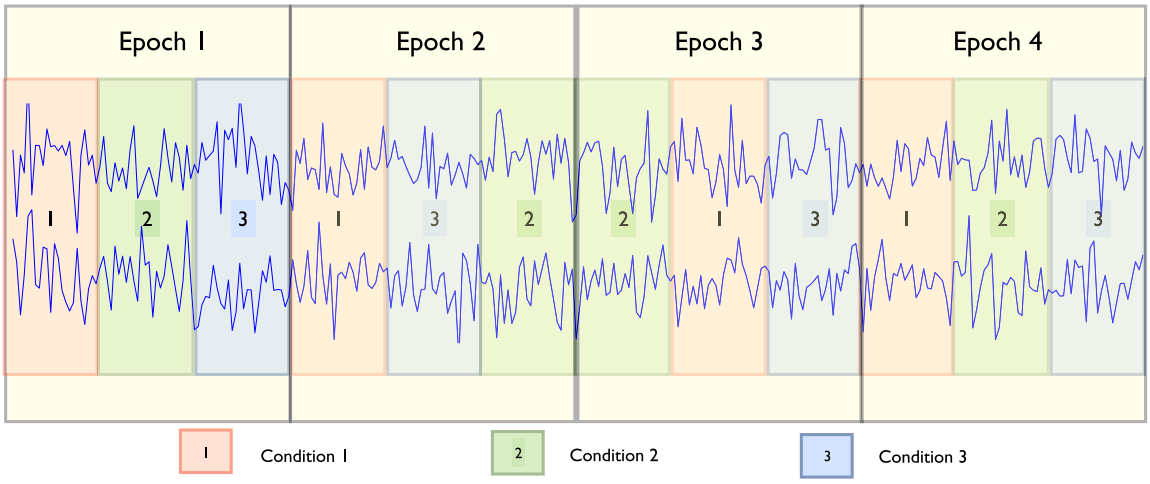

Figure 3. The SSA epoch structure of the multivariate empirical time series for estimating of the background stationary projection. Each epoch contains equal numbers of trials from each condition.

depends on the condition. The superscripts $\mathfrak{s}$ and n refer to the stationary and non-stationary contributions of each system; as stated above, stationarity refers to the *mean* and *covariance* remaining constant over time (not the time courses).

Note in particular at this point that $\mathbf{s}(t)$ (and $\mathbf{r}_k(t)$) are to be viewed as random time series variables; this implies that while $\mathbf{s}(t)$ represents the *same* random variable under two given conditions, the sample time series drawn from these random variables under two sample experimental conditions will be *different* entities. However, the time index t should be seen in each time series $\mathbf{x}_k(t)$ to represent the *same* quantity in order that we may, in practice, consider approximate simultaneity (relative to the time scale of the experiment) of samples drawn from separate experimental conditions.

The aim of our analysis is to remove the background non-stationary components $\mathbf{s}^{\mathfrak{n}}(t)$ from the signals in order to be able to analyze the distribution changes in a particular condition-specific system. To this end, we want to find a linear projection to the d stationary components of the background system. Even though in the final analysis we will only look at the data from one of the conditions, we use the whole data set for estimating this projection as we will see in the next section.

2.2. Parameter estimation

The parameters of the model (given in equation (2)) are the mixing matrix A of the background system and the mixing matrices $B_1, \ldots, B_K$ for each of the K condition-specific systems. For the purpose of removing the background non-stationary components, we are only interested in finding a matrix $P^{\mathfrak{s}} \in \mathbb{R}^{d\times D}$ which, when applied to the observed data, allows us to remove the non-stationary components of the background system, i.e.

$$\begin{aligned} P^{\mathfrak{s}}\mathbf{x}_k(t) &= P^{\mathfrak{s}}A^{\mathfrak{s}}\mathbf{s}^{\mathfrak{s}}(t) + \underbrace{P^{\mathfrak{s}}A^{\mathfrak{n}}}_{=0}\mathbf{s}^{\mathfrak{n}}(t) + P^{\mathfrak{s}}B_k\mathbf{r}_k(t) \\ &= P^{\mathfrak{s}}A^{\mathfrak{s}}\mathbf{s}^{\mathfrak{s}}(t) + P^{\mathfrak{s}}B_k\mathbf{r}_k(t) \end{aligned}$$

because the rows of $P^{\mathfrak{s}}$ are in the dual space of the non-stationary subspace spanned by $A^{\mathfrak{n}}$ such that $P^{\mathfrak{s}}A^{\mathfrak{n}} = 0$. Thus, the only matrix referred to by the model (2) that we estimate is a basis for the non-stationary subspace.

The projection matrix $P^{\mathfrak{s}}$, which recovers the stationary sources $\mathbf{s}^{\mathfrak{s}}(t)$, is estimated by applying SSA [33] in a particular way[11]. In brief, SSA finds the projection to the d stationary sources by minimizing the difference in distribution across epochs of the observed data. Special care is required here; however, since although we modeled the K conditions separately in equation (2), for the purpose of estimation of $P^{\mathfrak{s}}$, we consider *all* conditions as resulting from the same empirical time series. That is, the raw time series obtained by concatenating all trials from all conditions, in the order they appear (see figure 3), is divided into N epochs (using a chronological segmentation or a sliding window) for which epoch covariance matrices $\hat{\Sigma}_1, \ldots, \hat{\Sigma}_N$ are estimated[12]. The optimal stationary projection $\hat{P}^{\mathfrak{s}}$ is found by minimizing the difference between the projected epoch covariance matrices, as measured by the Kullback–Leibler (KL) divergence D_{KL} between Gaussians. The SSA objective function is the sum of the KL divergences between each epoch and the average epoch, whose mean and covariance we can set to 0 and I, respectively, without loss of generality. This leads to the objective function

$$\mathcal{L}(P) = \sum_{i=1}^{N} D_{\mathrm{KL}}[\mathcal{N}(0, P\hat{\Sigma}_i P^{\top})||\mathcal{N}(0, I)],$$

[11] Note that we will refer to $P^{\mathfrak{s}}$ as the 'projection to the background stationary sources'; this overloads a closely related use of 'projection matrix', referring to a square matrix A with $A^2 = A$.

[12] The choice of epochs is a delicate issue. The same issues for our problem setting apply as for the SSA problem setting and details may be found in the SSA expository paper [33]. Whether the epoch length affects the stationarity of the data is an application-bound issue. In our BCI application below, we choose epochs so as to span several trials, thus reducing noise levels and including multiple instances of each condition but short enough so as to conform to the bounds derived in the mentioned paper [33].

which is minimized using an iterative gradient-based procedure to find $P^{\mathfrak{s}}$:

$$\hat{P}^{\mathfrak{s}} = \underset{PP^\top = I}{\operatorname{argmin}} \sum_{i=1}^{N} -\log\det(P\hat{\Sigma}_i P^\top).$$

Although SSA was originally developed to invert the model (1), it is also applicable to our problem setting. This is because the weaker the changes in the condition-specific non-stationary sources $\mathbf{r}_k^{\mathfrak{n}}(t)$, the more closely the model (2) is approximated by the SSA model. To achieve this, we design our epochs in such a way that each epoch contains the same number of trials from each of the K conditions; see figure 3 for an illustration. Thus, we obtain SSA epoch covariance matrices which are approximately equal to the covariance matrix of the background system up to an approximately *stationary* (i.e. constant over SSA epochs) contribution from the condition-specific systems. More formally, the true covariance matrix (covariance at lag 0 with size $D \times D$) of the ith SSA epoch Σ_i is given by

$$\Sigma_i = C_i + \bar{R}_i,$$

where C_i is the covariance matrix of the background system and $\bar{R}_i$ is the average of the covariance matrices over all conditions in the ith SSA epoch

$$\bar{R}_i = \frac{1}{K}\sum_{k=1}^{K} R_{i,k}, \tag{4}$$

with $R_{i,k}$ being the covariance matrix of the kth condition when considering only the condition-specific system in the ith SSA epoch. As we will show in the next section, the condition-specific part $\bar{R}_i$ converges to stationarity in the sense that it becomes constant over SSA epochs. This means that SSA will correctly identify the non-stationary directions in the background system which we want to remove.

2.3. Theoretical results

In this section, we describe a theoretical guarantee for the quality of the approach outlined in the previous section. The rather technical proof is deferred to the appendix; here, we focus on the intuitive interpretation of the result. In particular, proposition 2.1, which we prove in the appendix, states that the covariances resulting from forming SSA epochs containing several conditions will converge to stationarity in the limit of the numbers of conditions K and condition-specific stationary dimensions d_k, the number of condition-specific stationary sources. This allows the SSA algorithm to identify the desired part of the background system because the only remaining factor affecting the estimation is then the small sample effects affecting SSA.

Proposition 2.1. *Let $\bar{R}_i$ be, as defined in equation (4), the average condition-specific covariance in the ith epoch, K the number of conditions, N the number of epochs and D the number of dimensions, where the number of non-stationary dimensions d_n is fixed.*

Assume that the non-stationarity is bounded, i.e. for all i, k, corresponding to the ith epoch and kth condition, $\left\|R_{i,k} - \frac{1}{N}\sum_{i=1}^{N} R_{i,k}\right\|_{\mathcal{F}} < \Gamma$, for some constant Γ, where $||\cdot||_{\mathcal{F}}$ is the Frobenius norm on matrices. Assume finally that the columns of the matrices $B_k^{\mathfrak{n}}$, the condition-specific non-stationary subspaces, are also uniformly distributed with orthonormal columns[13].

Then, let P be any projection matrix. Then, for any $\delta > 0$, we have the following relation in probability, in K and D, of $\bar{R}_i$ to the average covariance over epochs $\frac{1}{N}\sum_{j=1}^{N}\bar{R}_j$ under the projection P:

$$\Pr\left(\frac{1}{N}\sum_{i=1}^{N}\left\|P\left(\bar{R}_i - \frac{1}{N}\sum_{j=1}^{N}\bar{R}_j\right)P^\top\right\|_{\mathcal{F}} > \mathcal{O}\left(\frac{1}{K}\right) + \mathcal{O}\left(\frac{1}{\sqrt{D}}\right) + \mathcal{O}\left(\frac{1}{\sqrt{\delta\sqrt{D}}}\right)\right) < \delta. \tag{5}$$

Proof. See the appendix. □

The term on the left-hand side of the inequality expresses the non-stationarity measured in the terms $\bar{R}_i$ over epochs; each term of the sum expresses the deviation of $\bar{R}_i$ from the average over SSA epochs. Note that this is a different measure from the measure of non-stationarity that we use for the *optimization* with SSA: the SSA objective function is invariant to changes in the parametrization of the data space, which is desirable for the estimation, whereas the measure in the proposition is not. However, for the purpose of deriving a convergence rate, the measure in the proposition is more convenient. Nevertheless, both measures are equivalent in the sense that each is zero if and only if the other is zero.

The term on the right-hand side of the inequality decreases in K and D: for complete convergence to be guaranteed under the assumptions of the proposition, note that both K (number of epochs) and D (number of dimensions) should be assumed to grow. Nevertheless, either D or K increasing guarantees a better estimate. That D may be assumed to be much bigger than d_n in practice is plausible, as often non-stationarities are confined to low-dimensional subspaces, for example, artifacts in EEG are often confined only to electrodes at the edge of the scalp.

Note also that the proposition is conditioned on two important assumptions: firstly that the non-stationary subspaces of the condition-specific systems are sampled uniformly at random and secondly that the non-stationarity is uniformly bounded over classes $\left(\left\|R_{i,k} - \frac{1}{N}\sum_{i=1}^{N} R_{i,k}\right\|_{\mathcal{F}} < \Gamma\right)$. These are plausible assumptions to make on the model. The first assumption is equivalent to assuming that we have no prior knowledge of the condition-specific system's mixing matrices, but may be weakened without making convergence impossible and the second assumption is necessary for the estimation to occur at all, but is nevertheless plausible; for example, in the BCI application (below), the condition-specific non-stationarities are much weaker than the background non-stationarities.

[13] They are a sample of Haar measure.

3. Simulations on synthetic data

In this section, we evaluate the proposed method in controlled simulations on synthetic data, where we can objectively assess the performance based on an artificial ground truth. Our analysis consists of two parts. First of all, we investigate how accurately our algorithm can identify the true background stationary components. In our model, these are the relevant components for the subsequent analysis. In the second part of this section, we show that these components are useful. To that end, we apply our method as a pre-processing step to change-point detection in a situation where there are irrelevant background non-stationary components, which we want to remove. We analyze the relative merits of our pre-processing in different scenarios.

3.1. Data generation

The synthetic data that we use in the controlled simulations are generated according to our model (equation (2)). At each time point, the observed D-variate sample $x_k(t)$ under the kth condition is a sum of contributions from the background system and the condition-specific system. The background variables are linearly transformed by a random orthogonal mixing matrix $A \in \mathbb{R}^{D\times D}$ that is kept fixed over all conditions. The condition-specific components are linearly transformed by an orthogonal mixing matrix $B_k \in \mathbb{R}^{D\times D}$ that is chosen randomly for each condition $1 \leqslant k \leqslant K$, where K is the total number of conditions. In the simulations, the number of stationary and non-stationary components is kept fixed to the same value for both systems.

The background system and the condition-specific system each comprises two groups of components: stationary and non-stationary. The mean and covariance matrix (at time lag 0) of the stationary components are the zero vector 0 and the identity I, respectively. The mean of the non-stationary components is also fixed to zero, whereas the covariance matrix (again at time lag 0) in each epoch is chosen according to two Markov models, one for the background system and one for the condition-specific system. Both Markov models consist of five states corresponding to five covariance matrices. The covariance matrices are diagonal with entries drawn randomly from a set of log-spaced values over the interval $[1/\nu, \nu]$ with $\nu > 1$; thus, a high ν corresponds to a high level of non-stationarity. For the background system, this parameter is ν_{shared} and for the condition-specific system, it is ν_{specific}. The probability of changing the state in the Markov model (i.e. switching to another covariance matrix) is parametrized by p_{shared} and p_{specific} (denoting precisely the probability of remaining in the current state, the probability if changing to each other state is the same) for the background system and the condition-specific system, respectively. Thus, a low p corresponds to few changes occurring. The shortest segment of stationary data possible consists of 50 samples, generated from a Gaussian with the respective covariance matrix; thus, changes may occur at most once every 50 data points.

To sum up, the key parameters of interest are the strength of the non-stationarity in both systems (ν_{shared} and ν_{specific}) and the probability of changes in the covariance matrix of the non-stationary sources (p_{shared} and p_{specific}).

3.2. Finding the background stationary sources

In the first set of experiments, we investigate how well our method can identify the true stationary components of the background system. We analyze the influence of the number of different conditions in two scenarios: low and high condition-specific degrees of non-stationarity. The total number of dimensions is set to 20 and the number of non-stationary directions is set to 1 and 2 for the condition-specific and background systems, respectively; the total number of samples is 10 000. The error is measured as the subspace angle between the true non-stationary subspace of the background system and the estimated one.

The results are shown in figure 4. In the left panel, we see the results for low condition-specific non-stationarity. As we increase the number of conditions, the error in finding the true stationary components (blue line) becomes smaller. This decrease in estimation error is in line with proposition 2.1 which states that the error is guaranteed to decrease given a larger number of conditions K; intuitively, this is because the non-stationary contributions of the condition-specific systems become more equally spread across the whole D-dimensional space since for every condition k, we choose a random mixing matrix B_k. This allows us to distinguish more clearly between the background and the condition-specific non-stationary directions and, in turn, find the background stationary components. The red line shows the error on the same data where we have removed the condition-specific contributions: this is the lowest possible error level on the condition-specific model for finite samples. We see that as soon as we have at least four different conditions, we can identify the true background stationary components as well as if there were no condition-specific non-stationarities.

The right panel shows the results in the case where the degree of non-stationarity in the condition-specific system is high relative to the background system. As we can see, this makes the desired stationary components of the background system more difficult to identify than in the previous case because the overall distribution is dominated by changes in the condition-specific system. We therefore need a larger number of conditions (around 11) in order to distinguish between the non-stationarity of the two systems.

3.3. Identification of the background non-stationary components in simulated EEG

We illustrate the application of SSA to artificial data generated by a realistic EEG forward model. The head model we use was calculated on the basis of the MRI scans of 152 human participants [14], the electrodes were placed on the head according to the international 10-5 system and the forward mapping from simulated dipoles to voltages at the electrodes was computed using the semi-analytic methods of the contribution of Nolte and Dassios [21]. Given this framework, we simulated two-background non-stationary sources: the first 2 cm underneath the electrode T7 close to the

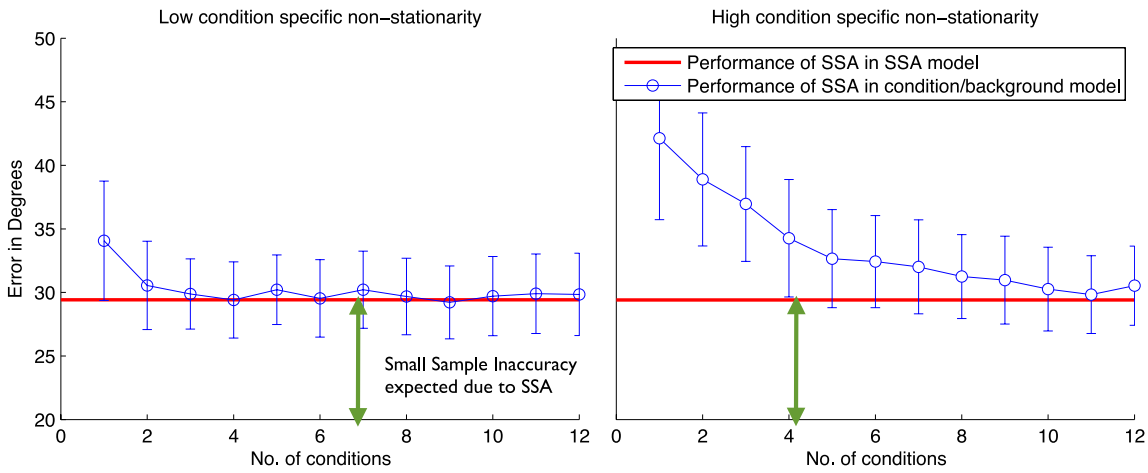

Figure 4. Performance of the proposed method measured in terms of the subspace angle (vertical axis) between the true and the estimated non-stationary subspaces of the background system, for different numbers of conditions (horizontal axis). The left and right panels show the results for the cases where the condition-specific non-stationarity is low and high, respectively. In each case, the red line denotes the case in which the condition-specific systems are replaced with stationary noise $d_k^s = D$ as a baseline result. The red line case thus represents the limit in potential accuracy of the method relative to accuracy in the standard SSA model.

edge of the montage to simulate the effects of a non-stationary EMG component and the second at the primary visual cortex, 2 cm under the electrode Oz, to simulate the effect of a non-stationary alpha source.

Subsequently, we modeled three scenarios. In the first scenario, for each of 13 conditions, in addition to the two-background non-stationary sources, a non-stationary source was positioned at random in the brain together with 12 stationary sources; thus, the data set has dimensionality 15, within a 118-dimensional (the number of electrodes) space. The non-stationarities were generated using a Markov model, giving rise to segments of simulated EEG data with varying power. The data were preprocessed using PCA [25] to eliminate the degenerate directions (PCA then yields a projection $U \in \mathbb{R}^{15\times 118}$ and a mixing matrix $V \in \mathbb{R}^{118\times 15}$). We then computed the estimated background non-stationary components using SSA and the estimated condition-specific components, for the first two conditions. More precisely, A^n and P^s were computed on the PCA-preprocessed data using SSA; thus, the scalp pattern corresponding to the most non-stationary background sources is given by composing A^n (of size $15 \times d^s$) with the mixing matrix yielded by PCA V (of size 118×15), i.e. we compute $(V \times A^n)$. Subsequently, SSA was applied to the condition-specific data preprocessed using PCA and projected by P^s (i.e. on $P^s U \mathbf{x_k}(t)$) in order to obtain a mixing matrix (which we will call C_k^n) corresponding to the non-stationary condition-specific patterns; thus, $B_k^n = A^s \times C_k^n$. The final corresponding scalp pattern is given by $V \times A^s \times C_k^n$. The results of this first scenario are displayed in figure 5 and show that SSA successfully locates the EMG and alpha component and succeeds in obtaining asymmetric non-stationary components between the first two classes, which match the true non-stationary condition-specific patterns. Moreover, we see that using SSA separately on each class does not deliver the background non-stationarities and the condition-specific non-stationarities successfully, whereas the estimation of the background non-stationarities over conditions using SSA is successful as described in section 2.2, following which the condition-specific non-stationary patterns are successfully found.

Although in general the background non-stationarities may only be reliably obtained given many conditions, we may still consider whether using SSA to remove background non-stationarities works when we have only two conditions and under which circumstances. This reflects the situation in our application to EEG data, where we consider motor imagery of two separate movements, described in section 4. Thus, in the second scenario, we reduce the number of conditions to 2 and simulate background non-stationary sources which are *stronger* in their non-stationarity than the condition-specific non-stationary sources. All other settings are kept constant. The results of this second scenario are displayed in figure 6 and show that in this case, when the background non-stationarities are strong, then these, and subsequently the condition-specific non-stationarities, may be obtained nonetheless, using the proposed method.

However, in the third and final scenario, we see that when the number of conditions is 2 and the background non-stationarities are weak, then obtaining patterns using SSA on each condition separately which nevertheless agree across the conditions, and thus may be assumed to correspond to background non-stationarities, occurs only after the most non-stationary components have been removed. Note that we remove more non-stationary directions here than actually are generated; this is achieved simply by setting the d parameter of the SSA computations to a different value than the true value. The results of this third scenario are displayed in figure 7. We will see, in the analysis of the BCI data in section 4, that both these second and third scenarios may arise.

3.4. Application to change-point detection

In this section, we demonstrate that the proposed method can be used to find *relevant* change points in the data from a

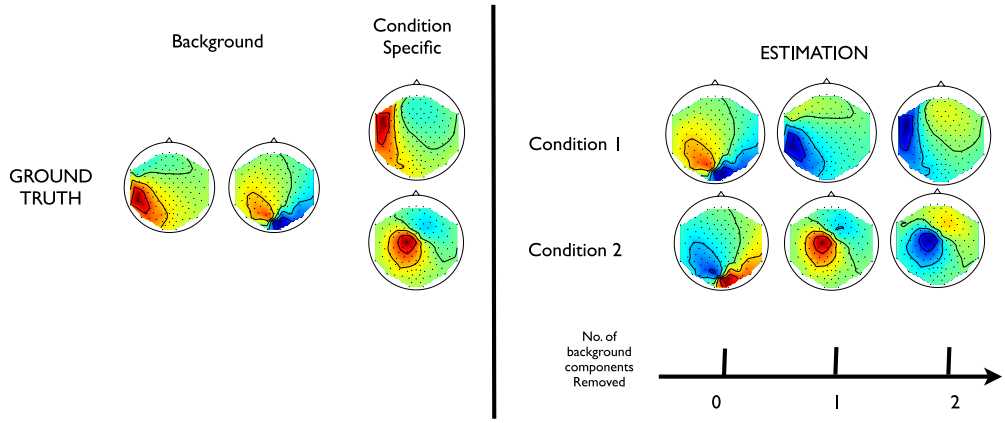

Figure 5. Scalp plots computed using SSA on the EEG forward model data are displayed. The left panel displays the true non-stationary patterns, and the right panel displays the estimated condition-specific non-stationary patterns after the removal of zero-, one- and two-background non-stationary directions. The plots show that SSA successfully obtains the patterns resulting from the background non-stationary alpha and background non-stationary EMG, and then allows one to concentrate on the condition-specific non-stationarities in each class. Note that after removing one-background non-stationary direction, SSA finds a difference between the conditions. This is because, for the two conditions studied, the condition-specific non-stationarities are strong. Nevertheless, SSA successfully obtains the correct background non-stationarity, so that the final condition-specific non-stationarities are correct.

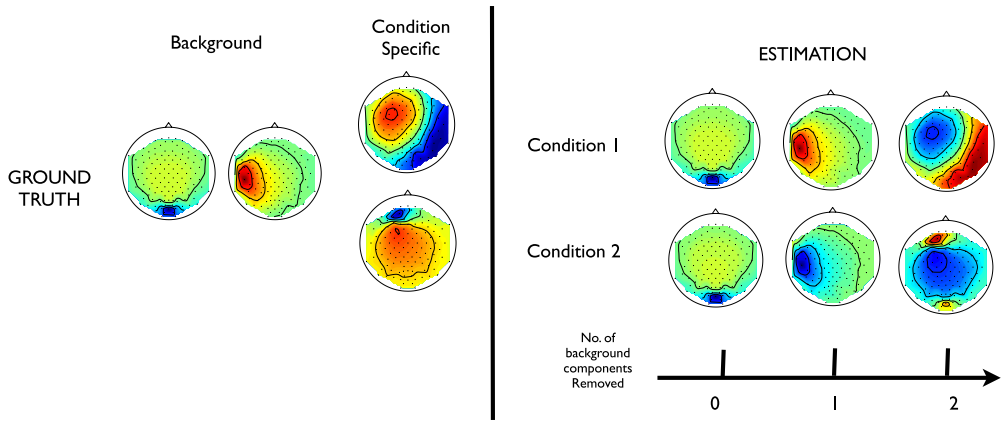

Figure 6. The results of the forward model simulation when the number of conditions is 2 and the background non-stationarities are stronger than the condition-specific non-stationarities are displayed. Here, using SSA, one obtains the background stationarities as estimated 'condition-specific' non-stationarities which agree across conditions after zero- and one-background non-stationarities have been removed. The condition-specific non-stationarities are then obtained after the removal of two-background non-stationary directions.

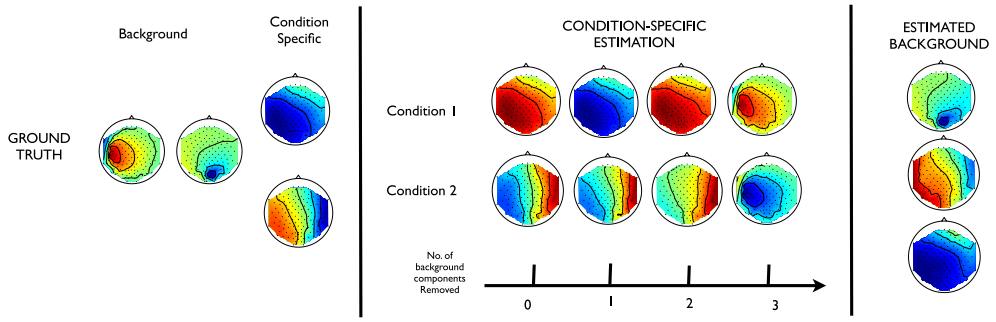

Figure 7. The results of the forward model simulation when the number of conditions is 2 and the condition-specific non-stationarities are stronger than the background non-stationarities are displayed. Here, after the removal of zero-, one- and two-background non-stationary directions, SSA finds *different* non-stationary patterns on each condition. However, after the removal of three-background non-stationary directions, SSA finds non-stationary patterns on each condition which agree, and arise, in fact, from the true background non-stationarities. The reason that the condition-specific non-stationarities are similar after the removal of zero- and one-background non-stationary directions is because the first estimated *background* non-stationary direction is nevertheless correct, the EMG component, so the condition-specific most non-stationary directions are preserved. These are then removed as the estimated second and third most non-stationary background directions.

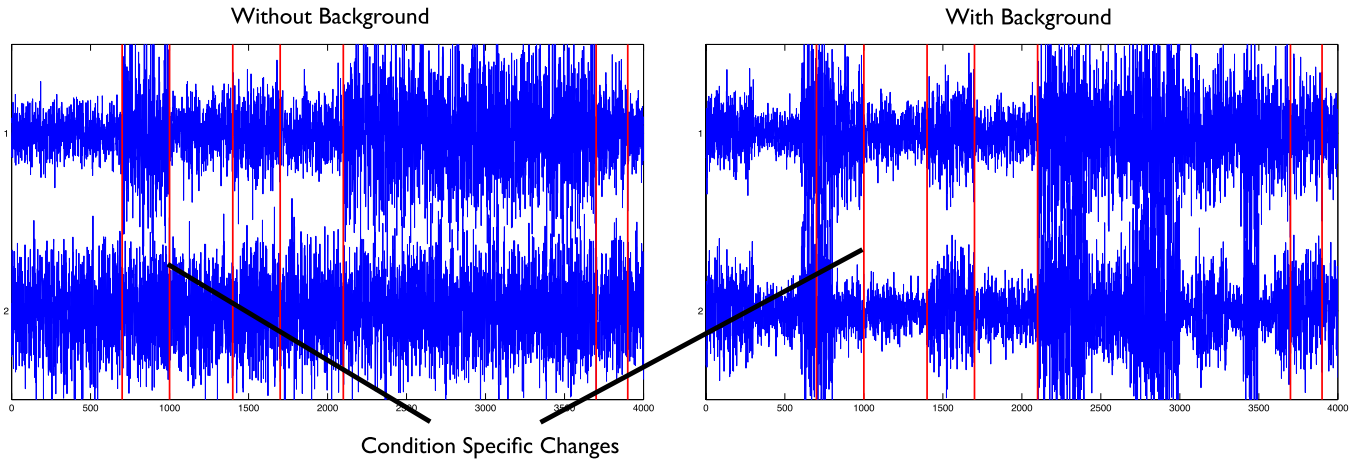

Figure 8. The change-point detection task is made more difficult when the change points of interest (red lines) in the condition-specific system (left) are masked by background change points (right). The left panel displays only the condition-specific system $B_k\mathbf{r_k}(t)$, whereas the right panel displays the entire data corresponding to $A\mathbf{s}(t) + B_k\mathbf{r_k}(t)$.

particular condition of interest, even when there are *irrelevant* distribution changes that are shared by all conditions. In reality, this situation occurs, e.g., when there are artifacts in the data that are independent of the experimental condition.

Change-point detection refers to the task of detecting points in time in which the distribution of a time series changes from one state to another; this task occurs potentially in sleep staging of EEG recordings [19] and in early warning systems for epilepsy patients using chronically implanted electrodes [10], among other applications [2, 13, 24]. There exists a wide range of different methods for change-point detection (also called time series segmentation). In our simulations, we use single linkage clustering based on a symmetrized KL divergence measure [15, 9] between segments of the time series.

Here, we test the hypothesis that the performance of change-point detection can benefit from removing non-stationary directions (feature extraction) that are not specific to the condition of interest. Thus, we suggest a *pre-processing* step which aims to boost the performance of change-point detection, when the model applies. Previous work [9] has shown that removing stationary directions in a pre-processing step improves the performance of change-point detection. Here, we go one step further and distinguish between relevant and irrelevant non-stationary changes by making use of data from other conditions during feature extraction to remove the background *non-stationary changes*. The problem setting is illustrated in figure 8; here, we see that infrequent and weak condition-specific changes are masked by stronger and more frequent background changes. This renders difficult detecting the condition-specific changes reliably. The solution we propose is to project the data to the background stationary sources.

The synthetic data are generated as described in section 3.1. The total number of dimensions is 10, the number of conditions is 10 and the number of non-stationary directions is 2 and 1 for the background and the condition-specific systems, respectively. In the simulations, we vary the strength of the non-stationarity in both systems (ν_{shared} and ν_{specific}), the likelihood of the relevant change points in the condition of interest (p_{specific}) and the likelihood of irrelevant change points in the background system (p_{shared}).

As a baseline, we directly apply the change-point detection algorithm to the data from the condition of interest and discard the remaining nine data sets. We compare this approach against the augmentation of this algorithm by our pre-processing (feature extraction). More specifically, we apply SSA to estimate the nine most stationary components *over all conditions* and then apply the change-point detection algorithm to these components on the data from the condition of interest. We measure the performance in terms of the area under the receiver–operator curve (AUC).

The results are shown in figure 9. The right column of the 3 × 3 grid shows the improvement in AUC due to our pre-processing for three different scenarios. When there are few background irrelevant change points (top row) *and* few condition-specific change points, the performance of change-point detection is relatively unaffected, there is an improvement using a pre-processing step only when the condition-shared changes are pronounced enough to yield reasonable estimation of the desired projection.

In the second row, we display the case in which there are many irrelevant background change points as well as many condition-specific change points of interest, which interfere with the detection of the relevant change points under the conditions of interest. Thus, in the middle panel of the right column, we see that the proposed pre-processing leads to a significant improvement in performance for almost all cases. However, when both systems have low non-stationarity (bottom left corner of the panel), it is hard to identify the informative background stationary components; hence, the performance does not increase.

The last simulations whose results are displayed in the bottom row confirm that it is the frequency of the changes in the distracting condition *shared* system which determine whether a high level of improvement is to be expected. Here, as in the second row, the frequency of background changes is high and the frequency of the desired condition-*specific*

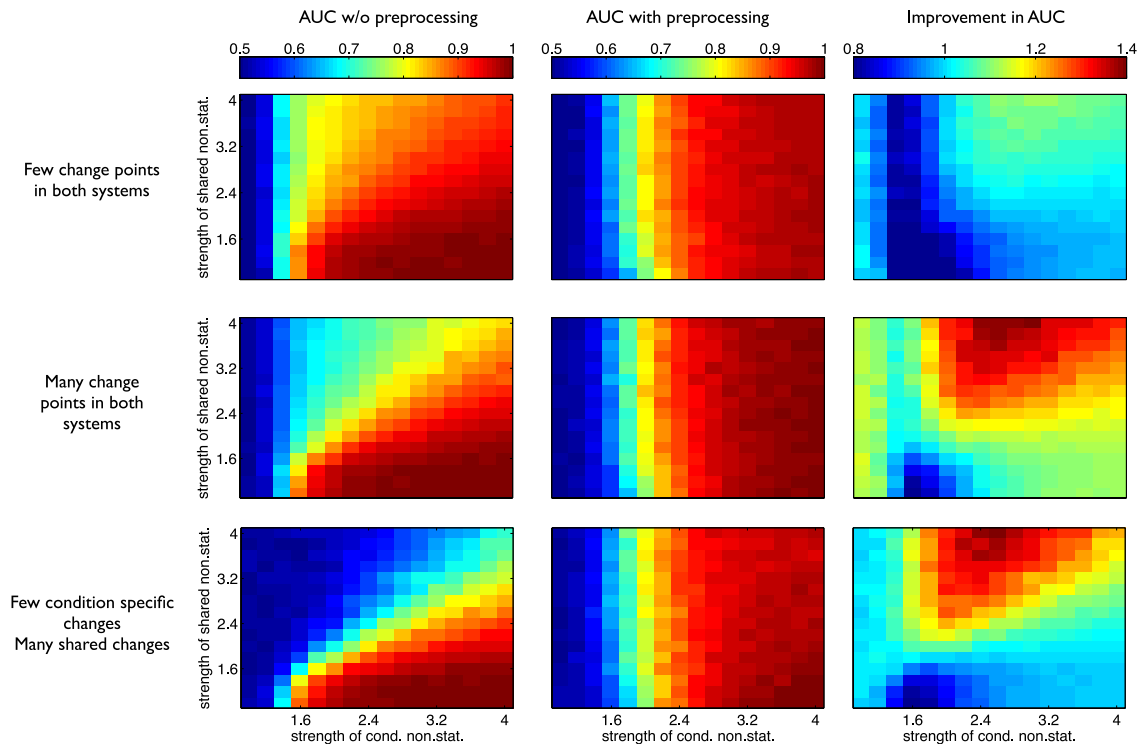

Figure 9. Results of the pre-processing for change-point detection. The rows of the 3×3 grid correspond to three different scenarios: few change points, many change points and few condition-specific changes but many background changes. From left to right, the columns shown the AUC without pre-processing (baseline), the AUC with pre-processing and a comparison between the two (improvement). In each panel, we vary the strength of the background non-stationarity along the vertical axis (ν_{shared}), and the strength of the condition-specific non-stationarity along the horizontal axis (ν_{specific}). The precise parameters for transition are in the top row $p_{\text{shared}} = 0.8$, $p_{\text{specific}} = 0.8$, in the middle row $p_{\text{shared}} = 0.6$, $p_{\text{specific}} = 0.6$ and in the bottom row $p_{\text{shared}} = 0.1$, $p_{\text{specific}} = 0.8$. The middle column pictures are similar because once the correct background non-stationary space has been correctly found and the background non-stationarities projected out (which is possible in all three cases), then the difficulty of the task is equal in all three simulations on this space.

changes is low. As in the second row, improvement is to be expected in all cases except the cases in which the strength of the background changes is not sufficient to yield reliable estimation of the shared stationary projection.

4. Application to BCI data

A *BCI* is a device which allows a human subject to perform a limited number of commands on a computer using only the activity observable using a brain imaging device, such as an fMRI scanner, NIRS imaging system or most commonly an EEG cap. In the motor imagery paradigm for EEG [7, 4], machine learning algorithms may be trained on data recorded during imaginations of hand movements corresponding to the desired direction of movement of a cursor on a computer screen. One practical problem with such a device is, however, that the distribution over scalp voltages in EEG recordings is highly non-stationary. This has the effect that BCIs based on machine learning degrade in their performance over time [31, 5, 8, 32]. To deal with this problem, to maintain high levels of performance, numerous machine learning methodologies have been proposed, based on unsupervised adaptation [31], robustification [5], penalization of non-stationarity [34, 28] and covariate shift adaptation [30]. An interesting *neuroscientific* question, however, is to diagnose the physical source of this non-stationarity. On the one hand, it may be the case that non-stationarity results to a high degree from artifactual sources, such as EMG activity, ocular activity or disturbances resulting from loose electrodes [34, 28, 27]. On the other hand, the fact remains that the activity in the EEG recording resulting from neural activity is also highly non-stationary, albeit to a lesser degree and often over a different timescale than the non-stationarity originating in artifacts. The targeted study of each of these non-stationary contributions has, until recently, been confounded by the fact that only a linear mixture of their respective activities may be observed at the scalp electrodes. In particular, the contribution of the neural component to the overall non-stationarity is typically weaker than the contribution of the artifactual activity, making, in particular, the targeted study of neural non-stationarity especially problematic. Nevertheless, the model studied in the present contribution presents the possibility of prying apart neural and artifactual non-stationarity. Given that this may be achieved, there exists the prospect of better understanding the neural changes which occur during EEG-based BCI use.

To illustrate this possibility, we display scalp maps in figure 10 of a BCI subject. For each condition (left imagination and foot imagination), the most non-stationary pattern in that condition *subsequent* to the removal of a number m of shared non-stationary directions has been visualized. The figure displays these patterns for $m = 0, 1, \ldots, 6$. We observe in the results that the model studied above is applicable to the

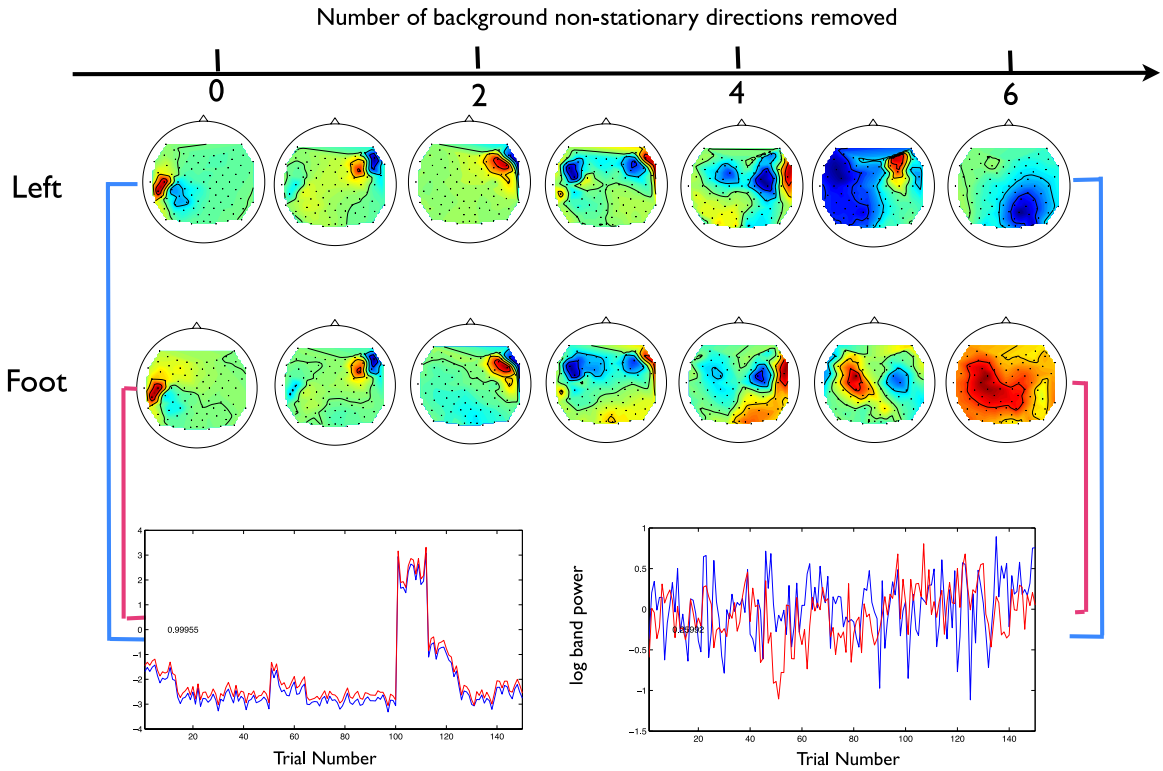

Figure 10. Scalp maps from a single BCI subject are displayed. The upper row displays maps from the first condition, whereas the second row from the second condition. The conditions are ordered in terms of how many joint non-stationary components were removed. The numbers removed are, from left to right, 0,1,...,6. The difference between the conditions becomes clearly visible only after five directions have been removed. The similarity between the preceding components is clearly visible, between conditions. The traces at the bottom, colored red and blue corresponding to the condition, display the log band power recorded on each trial over the course of the experiment. The highest components are highly correlated, the lower, uncorrelated. These band powers are calculated on the data $Q_k^n P^{\mathfrak{s}} U \mathbf{x_k}(t)$.

data set at hand: namely there is a common non-stationary component to both the condition systems. When no or few class-shared non-stationarities are removed, the patterns are similar between the two conditions. The structure of the displayed patterns indicates that the activity generating this non-stationarity corresponds to an EMG artifact. We see, however, by increasing the number of common non-stationary sources removed that differences in the patterns between the estimated condition non-stationarities emerge. For example, the patterns in the left-hand condition, after the removal of six background non-stationary sources, display high weights contralateral to the imagined hand and are smooth in their topography, in contrast to the most non-stationary patterns. These facts indicate that the patterns may contain information regarding *task-relevant neural* changes observable in the data from each condition.

We now extend our analysis to include the data from all 80 subjects recorded for the study at hand [6, 16]. In order to illustrate the usefulness of the proposed model for a general BCI motor imagery subject, we use the following heuristic to quantify to what extent the model is suitable for any given data set:

(1) We compute a similarity measure between each of the two condition-specific most non-stationary patterns for every choice of the dimensionality of background non-stationary subspace $d_{\mathfrak{s}}$ including the case where there are assumed to be *no* background non-stationary sources. An example of the results of such a computation are displayed in figure 11.
(2) We then compute the *difference* in the average similarity of the condition-specific non-stationary patterns for $d_{\mathfrak{s}} = 5, \ldots, 10$ and the similarity of the condition-specific patterns for $d_{\mathfrak{s}} = 0$.
(3) If this difference is *greater* than a predefined threshold, then we conclude that the model is applicable for this data set.

An important point to note here, which is important for the validity of the heuristic, is that in the null case in which the model does not apply, the similarity between the condition-specific patterns for $d_{\mathfrak{s}} = 0$ is on average smaller than the similarity between the condition-specific patterns for $d_{\mathfrak{s}} = 5, \ldots, 10$. The results of applying this heuristic to all 80 BCI motor imagery subjects is displayed in figure 12. They show for a significant fraction of subjects that there are clear background non-stationary components; this may be concluded from the fact that the most non-stationary components estimated on the conditions separately are identical, up to estimation error and noise. This implies that the model we propose in this paper is highly appropriate for studying the changes in distribution of the EEG of BCI subjects.

Note at this point that this heuristic is designed to indicate the usefulness of the model for the data set. A more principled approach must wait for further work due to the advanced statistical exposition and innovation such an approach would

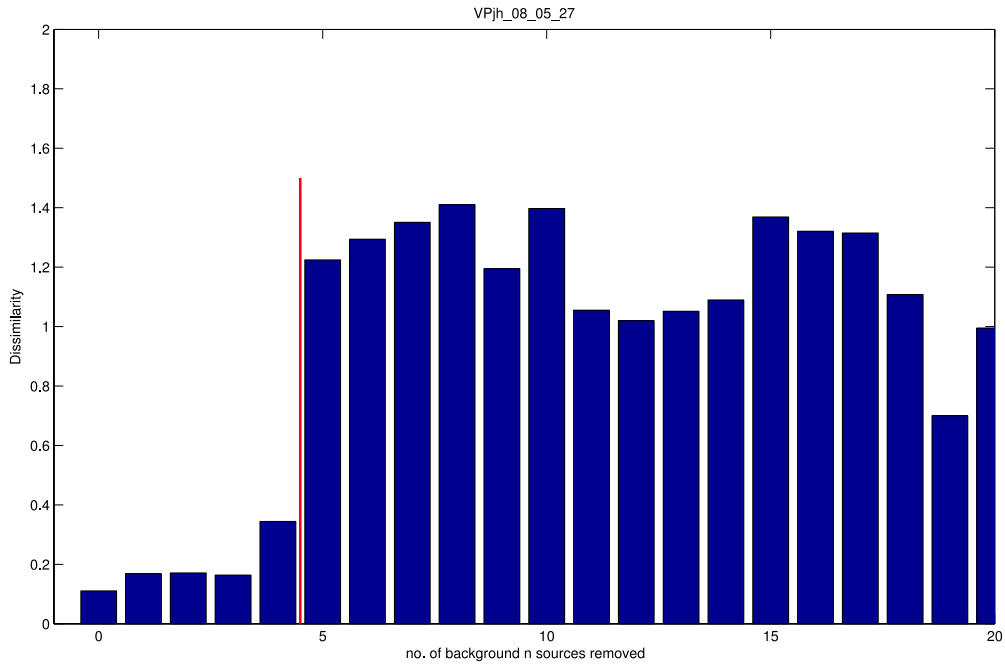

Figure 11. The results for a single subject (VPjh) of computing step 1 of the heuristic for determining the applicability of the model for a given data set are displayed. The *x*-axis displays the number of background sources removed and the *y*-axis displays the dissimilarity (1-norm between normalized patterns) between the condition specific non-stationary patterns on each of the two classes. The red line indicates the transition from a non-random level of dissimilarity to a random level, suggesting that the number of background non-stationary sources is 5.

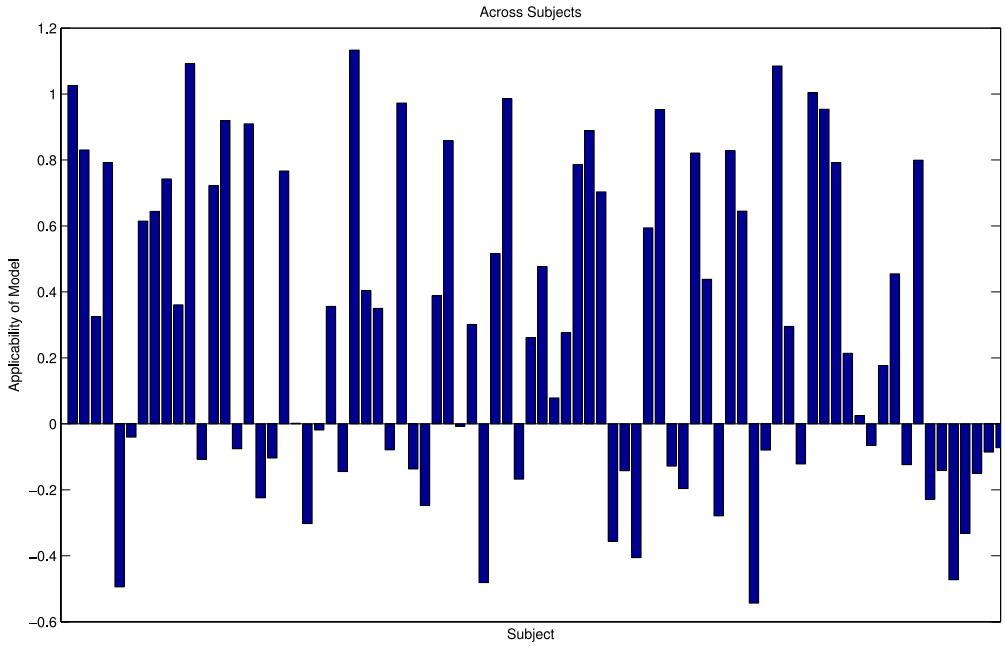

Figure 12. The results of applying the heuristic for estimating the applicability of the model are displayed. The *y*-axis shows the difference in the average similarity of the condition-specific non-stationary patterns for $d_s = 5, \ldots, 10$ and the similarity of the condition-specific patterns for $d_s = 0$. The *x*-axis displays the subject number. For a large proportion of the subjects in the data set, the model is applicable.

imply. For example, an approach which compares the similarity of vectors over different dimensionalities must correct for this discrepancy; see, for example, [1] for theorems facilitating this correction.

One clear limitation of the data set is that only two conditions are available. Finding the background non-stationarities is thus only possible because the condition-specific non-stationarities are weaker than the background non-stationarities, for most subjects in this data set. Occasionally, the condition-specific non-stationarities are not weaker; see figures 12 and 13. Here, we observe that a strong condition-specific non-stationary component is present. For such a subject, the presence of more experimental conditions would guarantee that this component is nevertheless not neglected. However, despite this, we see that in a majority of cases, the condition-specific components are *not* stronger

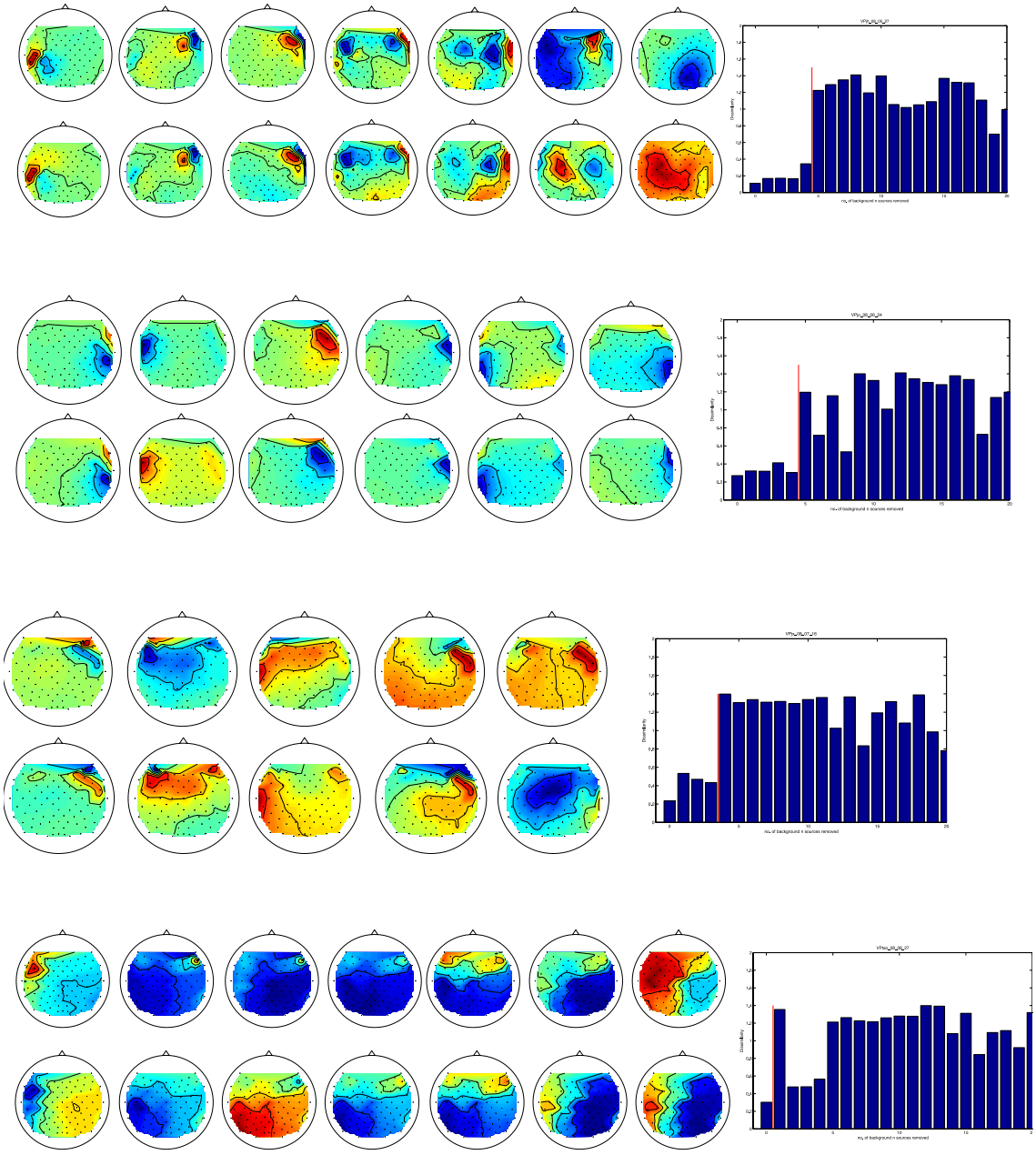

Figure 13. Three subjects (the top three: VPjh, VPjo and VPjx) with clear background non-stationary directions are displayed. The scalp maps are displayed on the left and the similarity between components on the right as per figure 11. The fourth and final subject (bottom: VPtao) displays background non-stationary directions but also a highly non-stationary condition-specific component. The component is located frontal right in the first condition, and has a localized topography. We conjecture that this is caused by an EMG artifact which is biased toward the first condition. The red line on the right displays the number of background non-stationary components estimated by the heuristic. This choice is visible in the similarity of the patterns displayed on the right; in the uppermost row, we display an extra pair of patterns to display the differentiation and class structure in the condition-specific patterns. In the bottom row, we display additional patterns to show that, in this case, the heuristic's choice is conservative.

than the background components and the successful removal of the background components is possible, and, thus that the model is suitable for this type of data set.

5. Concluding discussion

Neural signals are subject to change which is typically induced by the experimental paradigm. In addition, typically a number of artifactual signals are present in an experiment that hampers a possible interpretation: they may consist of drifts, or also other more complex factors that are common between behavioral conditions. This work has contributed with its explorative framework which enables the removal of these uninteresting non-stationary components. Thus, for the first time, condition-specific non-stationarities can be extracted and physiologically interpreted. In particular, this was demonstrated on simulated EEG examples from a realistic

forward model as well as on real experimental data from a large-scale BCI study.

An important assumption that needs to be made is that the condition-specific non-stationary patterns $B_k^{\mathfrak{n}}, B_{k'}^{n}$ should be independent for $k \neq k'$. If these are highly dependent, then no estimation of $P^{\mathfrak{s}}$, the projection to the background stationary sources, is possible. If independence does not hold, then the condition-specific features of interest may significantly overlap in their projection with the condition-shared non-stationarities, and information may be lost. The requirement for independence is, however, an intrinsic feature of the model and not a deficiency of the estimation method. If independence is not assumed in *some* form, then estimation by any method is impossible. However, as discussed in section 2.3, the assumption of independence of origin is realistic, in particular since the non-stationarities specific to each condition result from separate underlying physical systems. Note that it will often be possible to know *a priori* if these assumptions are likely to hold for a particular application. For example, in BCI, different brain regions govern the control of separate motor imaginations and thus the condition-specific non-stationarity patterns should be dissimilar from one another. If the assumption of independence of origin does not hold, then the estimation may still be possible under a slightly altered independence assumption. This is because it may be plausibly assumed that the evolution of non-stationarity in the *time domain* is independent in each condition-specific system. This should also guarantee consistency; we postpone the proof of this result to a later publication.

Another important assumption we make is that no condition should be allowed to contribute disproportionately to the non-stationarity in the overall data. As we argued earlier, this assumption is realistic in applications in which the background non-stationarity, for example, artifacts in EEG recordings are stronger than the condition-specific non-stationarities. Moreover, this assumption is related to assuming that the background non-stationarities are stronger than the condition-specific non-stationarities. If it is the case that one condition-specific system exhibits high non-stationarity, then the confounding influence of the background non-stationarity is reduced for this class in any case. Thus, making special allowance for background non-stationarity in such a case is less pressing.

After *removing* the background non-stationary activity from a data set, further analysis must be confined to a subspace of the orthogonal complement in data space of the removed directions. However, as we demonstrated in our change-point detection analysis, this subspace may yield important performance gains over analysis on the entire space.

In summary, we have shown that the model and methods we propose boost the performance of change-point detection and allows the data analyst to discover condition-specific changing components in BCI data. Future studies will apply the novel framework for analyzing experimental data to primate experiments that study learning and plasticity. In addition, we will study condition-specific non-stationarities in multimodal data [3] and in co-adaptive BCI [32]. It should also prove interesting to localize the respective sources of different non-stationary nature [17, 18].

Acknowledgments

The authors would like to thank Daniel Bartz, Franz Kiraly and Stefan Haufe for their valuable comments on the topic of this paper. DAJB was supported in part by a grant from the research training group GRK 1589/1 'Sensory Computation in Neural Systems' and by the Bernstein Focus Neuro-technology (BFNT), funded by the BMBF, FKZ 01GQ0850. FCM was supported by BMBF and DFG in the German-Japanese research project 'Robust Adaptive BCI in non-stationary environments' FKZ 01GQ1115. K-RM was supported in part by the World Class University Program through the National Research Foundation of Korea funded by the Ministry of Education, Science, and Technology, under grant R31-10008.

Appendix. Proof of proposition 1

Proof. The proof strategy is as follows: if we show that the statement holds for each of the non-stationary projections $Q_k^{\mathfrak{n}}$ simultaneously, where $Q_k = (B_k)^{-1}$, then clearly the statement holds for a random projection P. We need to show that the projection of the system by $Q_k^{\mathfrak{n}}$ is stationary, for all K, in the limit of K and D.

For any epoch indexed by i, any condition index k', where we define $\hat{R}_k = \frac{1}{N}\sum_i^N R_i^k$:

$$\begin{aligned}
&\left\| \frac{1}{K}\sum_{k=1}^{K}(Q_{k'}^{\mathfrak{n}}(R_i^k - \hat{R}_k)(Q_{k'}^{\mathfrak{n}})^\top \right\|_{\mathcal{F}} \\
&\quad \leqslant \frac{1}{K}\left\|R_i^{k'} - \hat{R}_{k'}\right\|_{\mathcal{F}} + \frac{1}{K}\sum_{k\neq k'}\left\|Q_{k'}^{\mathfrak{n}}\left(R_i^k - \hat{R}_k\right)Q_{k'}^{\mathfrak{n}}\right\|_{\mathcal{F}} \\
&\quad \leqslant \frac{1}{K}\Gamma + \frac{\kappa}{K}\sum_{k\neq k'}\|S^\top I_{d_n} S\|_{\mathcal{F}}.
\end{aligned}$$

The right-hand quantity in the first line is bounded above by the quantity on the right in the second line, $\frac{\kappa}{K}\sum_{k\neq k'}\|S^\top I_{d_n} S\|$, where $S^\top$ is a d_n-dimensional random projection matrix, I_{d_n} is a matrix of zeros apart from the left, top hand d_n rows and columns which are equal to the identity matrix and κ is a constant. The random variable on the right under the sum $\|S^\top I_{d_n} S\|$ has mean μ_D of order $\frac{1}{D}$ and standard deviation σ_D of order $\frac{1}{\sqrt{D}}$. This is because as D grows, for fixed d, the rows of the matrix P tend to independent samples from the distribution over one-dimensional subspaces. The dot product between random one-dimensional subspaces has density of order $\sqrt{D}x^{-1}(1-x^2)^{(D-3)/2}$ with moments of order $1/\sqrt{D}$ [1]. Each row in the matrix $P^\top I_{d_n}$ is a dot product between basis vectors of unit length of random one-dimensional subspaces. Thus, each row has norm bounded by constant order. Moreover, the effect of the truncated matrix in the second term is to truncate the entries in P, so that each entry of the entire matrix $P^\top I_{d_n} P$ may be approximated as dot products between random subspaces and the entries may be assumed independent. Thus, the mean of each term in the Frobenius norm is at most of the order of the variance of the dot product between spanning vectors of random one-dimensional subspaces and the standard deviation at most of order of the

square root of the fourth moment, which both decrease as $1/\sqrt{D}$ and $1/(D^{1/4})$. Thus, the sum term is distributed in the limit of K with standard deviation bounded by $\frac{\beta}{\sqrt{KD^{1/4}}}$ and mean by $\frac{q}{\sqrt{D}}$ for some constants q and β.

Thus by Chebyshev's inequality,

$$\Pr\left(\frac{1}{K}\sum_{k\neq k'}\left\|S^{\top}I_{d_n}S\right\|_{\mathcal{F}} > \epsilon/\kappa + \frac{q}{\sqrt{D}}\right) < \frac{\kappa^2\beta^2}{\epsilon^2 K\sqrt{D}},$$

which implies

$$\Pr\left(\frac{1}{K}\Gamma + \frac{\kappa}{K}\sum_{k\neq k'}\left\|S^{\top}I_{d_n}S\right\|_{\mathcal{F}} > \frac{1}{K}\Gamma + \epsilon + \frac{c}{\sqrt{D}}\right) < \frac{\kappa^2\beta^2}{\epsilon^2 K\sqrt{D}}.$$

Then by Boole's inequality, the probability that each one of these inequalities holds gives that for any k',

$$\Pr\left(\frac{1}{K}\Gamma + \frac{\kappa}{K}\sum_{k\neq k'}\left\|S^{\top}I_{d_n}S\right\|_{\mathcal{F}} > \frac{1}{K}\Gamma + \epsilon + \frac{c}{\sqrt{D}}\right) < \frac{\kappa^2\beta^2}{\epsilon^2 \sqrt{D}}.$$

Putting all of this together yields the estimate

$$\Pr\left(\left\|\frac{1}{K}\sum_{k=1}^{K} Q^{\mathrm{n}}_{k'}\left(R^k_i - \hat{R}_k\right)\left(Q^{\mathrm{n}}_{k'}\right)^{\top}\right\|_{\mathcal{F}} > \mathcal{O}\left(\frac{1}{K}\right) + \mathcal{O}\left(\frac{1}{\sqrt{D}}\right) + \epsilon\right) < \frac{\kappa^2\beta^2}{\epsilon^2\sqrt{D}}, \tag{A.1}$$

which is equivalent to

$$\Pr\left(\left\|\frac{1}{K}\sum_{k=1}^{K} Q^{\mathrm{n}}_{k'}\left(R^k_i - \hat{R}_k\right)\left(Q^{\mathrm{n}}_{k'}\right)^{\top}\right\|_{\mathcal{F}} > \mathcal{O}\left(\frac{1}{K}\right) + \mathcal{O}\left(\frac{1}{\sqrt{D}}\right) + \mathcal{O}\left(\frac{1}{\sqrt{\delta\sqrt{D}}}\right)\right) < \delta. \tag{A.2}$$

Finally, this implies the statement in the proposition since

$$\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{K}\sum_{k=1}^{K} Q^{\mathrm{n}}_{k'}\left(R^k_i - \hat{R}_k\right)\left(Q^{\mathrm{n}}_{k'}\right)^{\top}\right\|_{\mathcal{F}} \geqslant \frac{1}{N}\sum_{i=1}^{N}\left\|Q^{\mathrm{n}}_{k'}\left(\bar{R}_i - \frac{1}{N}\sum_{j=1}^{N}\bar{R}_j\right)\left(Q^{\mathrm{n}}_{k'}\right)^{\top}\right\|_{\mathcal{F}}. \tag{A.3}$$

□

References

[1] Absil P A, Edelman A and Koev P 2006 On the largest principal angle between random subspaces *Linear Algebra Appl.* **414** 288–94

[2] Andre-Obrecht R 1988 A new statistical approach for the automatic segmentation of continuous speech signals *IEEE Trans. Acoust. Speech Signal Process.* **36** 29–40

[3] Bießmann F, Plis S M, Meinecke F C, Eichele T and Müller K R 2011 Analysis of multimodal neuroimaging data *IEEE Rev. Biomed. Eng.* **4** 26–58

[4] Blankertz B, Dornhege G, Krauledat M, Müller K-R and Curio G 2007 The non-invasive Berlin brain–computer Interface: fast acquisition of effective performance in untrained subjects *NeuroImage* **37** 539–50

[5] Blankertz B, Kawanabe M, Tomioka R, Hohlefeld F, Nikulin V and Müller K R 2008 Invariant common spatial patterns: alleviating nonstationarities in brain–computer interfacing *Advances in Neural Information Processing Systems* vol 20 pp 113–20

[6] Blankertz B, Sannelli C, Halder S, Hammer E M, Kübler A, Müller K R, Curio G and Dickhaus T 2010 Neurophysiological predictor of SMR-based BCI performance *Neuroimage* **51** 1303–9

[7] Blankertz B, Tomioka R, Lemm S, Kawanabe M and Müller K-R 2008 Optimizing spatial filters for robust EEG single-trial analysis *IEEE Signal Process. Mag.* **25** 41–56

[8] Blankertz B and Vidaurre C 2009 Towards a cure for BCI illiteracy: machine learning based co-adaptive learning *BMC Neurosci.* **10** P85

[9] Blythe D A J, von Bünau P, Meinecke F C and Müller K-R 2012 Feature extraction for change-point detection using Stationary Subspace Analysis *IEEE Trans. Neural Netw. Learn. Syst.* **23** 631–43

[10] Celka P and Colditz P 2002 Time-varying statistical dimension analysis with application to newborn scalp EEG seizure signals *Med. Eng. Phys.* **24** 1–8

[11] Churchland M M, Yu B M, Ryu S I, Santhanam G and Shenoy K V 2006 Neural variability in premotor cortex provides a signature of motor preparation *J. Neurosci.* **26** 3697–712

[12] Comon P 1994 Independent component analysis, a new concept? *Signal Process.* **36** 287–314

[13] Csörgö H and Horvárth L 2009 20 Nonparametric methods change point problems *Handbook of Statistics* vol 7 ed P R Krishnaiah and C R Rao (New York: Elsevier) pp 403–25

[14] Fonov V S, Evans A C, Botteron K, Almli C R, McKinstry R C and Collins D L (The Brain Development Cooperative Group) 2011 Unbiased average age-appropriate atlases for pediatric studies *NeuroImage* **54** 313–27

[15] Gower J C and Ross G J S 1969 Minimum spanning trees and single linkage cluster analysis *J. R. Stat. Soc.* **18** 54–64

[16] Hammer E M, Halder S, Blankertz B, Sannelli C, Dickhaus T, Kleih S, Müller K-R and Kübler A 2012 Psychological predictors of SMR-BCI performance *Biol. Psychol.* **89** 80–86

[17] Haufe S, Nikulin V V, Ziehe A, Müller K-R and Nolte G 2008 Combining sparsity and rotational invariance in EEG/MEG source reconstruction *NeuroImage* **42** 726–38

[18] Haufe S, Tomioka R, Dickhaus T, Sannelli C, Blankertz B, Nolte G and Müller K-R 2011 Large-scale EEG/MEG source localization with spatial flexibility *NeuroImage* **54** 851–9

[19] Kohlmorgen J, Müller K R, Rittweger J and Pawelzik K 2000 Identification of nonstationary dynamics in physiological recordings *Biol. Cybern.* **83** 73–84

[20] Mandelblat-Cerf Y, Novick I, Paz R, Link Y, Freeman S and Vaadia E 2011 The neuronal basis of long term sensorimotor adaptation *J. Neurosci.* **31** 300–13

[21] Nolte G and Dassios G 2005 Analytic expansion of the EEG lead field for realistic volume conductors *Phys. Med. Biol.* **50** 3807–23

[22] Parra L C, Spence C D, Gerson A D and Sajdac P 2005 Recipes linear analysis of EEG *NeuroImage* **28** 326–41

[23] Paus T, Zatorre R J, Hofle N, Caramanos Z, Gotman J, Petrides M and Evans A C 2006 Time-related changes in neural systems underlying attention and arousal during the performance of an auditory vigilance task *J. Cogn. Neurosci.* **9** 392–408

[24] Pawelzik K, Kohlmorgen J and Müller K-R 1996 Annealed competition of experts for a segmentation and classification of switching dynamics *Neural Comput.* **8** 340–56

[25] Pearson K 1901 On lines and planes of closest fit to systems of points in space *Phil. Mag.* **2** 559–72
[26] Pijn J P, Neerven J V, Noest A and Lopes da Silva F H 1991 Chaos or noise in EEG signals; dependence on state and brain site *Electroencephalogr. Clin. Neurophysiol.* **79** 371–81
[27] Samek W, Kawanabe M and Vidaurre C 2011 Group-wise stationary subspace analysis - a novel method for studying non-stationarities *Proc. 5th Int. BCI Conf. (Graz)* ed G R Müller-Putz, R Scherer, M Billinger, A Kreilinger, V Kaiser and C Neuper (Verlag der Technischen Universität Graz) pp 16–20
[28] Samek W, Vidaurre C, Müller K-R and Kawanabe M 2012 Stationary common spatial patterns for brain computer interfacing *J. Neural Eng.* **9** 026013
[29] Shenoy P, Krauledat M, Blankertz B, Rao R P N and Müller K-R 2006 Towards adaptive classification for BCI *J. Neural Eng.* **3** 13–23
[30] Sugiyama M, Krauledat M and Müller K R 2007 Covariate shift adaptation by importance weighted cross validation *J. Mach. Learn. Res.* **8** 985–1005
[31] Vidaurre C, Kawanabe M, Bünau P von, Blankertz B and Müller K-R 2011 Towards an unsupervised adaptation of LDA for brain–computer interfaces *IEEE Trans. Biomed. Eng.* **58** 587–97
[32] Vidaurre C, Sannelli C, Müller K-R and Blankertz B 2011 Machine-learning-based coadaptive calibration for brain–computer interfaces *Neural Comput.* **23** 791–816
[33] von Bünau P, Meinecke F C, Király F J and Müller K-R 2009 Finding stationary subspaces in multivariate time series *Phys. Rev. Lett.* **103** 214101
[34] Wojcikiewicz W, Vidaurre C and Kawanabe M 2011 Stationary common spatial patterns: Towards robust classification of non-stationary EEG signals *IEEE Int. Conf. on Acoustics, Speech and Signal Processing* pp 577–80