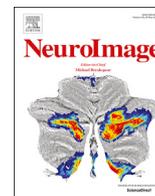




Contents lists available at ScienceDirect

NeuroImage

journal homepage: [www.elsevier.com/locate/neuroimage](http://www.elsevier.com/locate/neuroimage)

## Robust detrending, rereferencing, outlier detection, and inpainting for multichannel data



Alain de Cheveigné<sup>a,b,c,\*</sup>, Dorothée Arzounian<sup>a,b</sup>

<sup>a</sup> Laboratoire des Systèmes Perceptifs, UMR 8248, CNRS, France

<sup>b</sup> Département d'Etudes Cognitives, Ecole Normale Supérieure, PSL, France

<sup>c</sup> UCL Ear Institute, United Kingdom

### ARTICLE INFO

#### Keywords:

EEG  
MEG  
LFP  
ECoG  
Artifact  
ICA  
CSP  
DSS  
SNS  
CCA  
Sensor noise  
Detrending  
Weighted regression  
Robust statistics

### ABSTRACT

Electroencephalography (EEG), magnetoencephalography (MEG) and related techniques are prone to glitches, slow drift, steps, etc., that contaminate the data and interfere with the analysis and interpretation. These artifacts are usually addressed in a preprocessing phase that attempts to remove them or minimize their impact. This paper offers a set of useful techniques for this purpose: robust detrending, robust rereferencing, outlier detection, data interpolation (inpainting), step removal, and filter ringing artifact removal. These techniques provide a less wasteful alternative to discarding corrupted trials or channels, and they are relatively immune to artifacts that disrupt alternative approaches such as filtering. *Robust detrending* allows slow drifts and common mode signals to be factored out while avoiding the deleterious effects of glitches. *Robust rereferencing* reduces the impact of artifacts on the reference. *Inpainting* allows corrupt data to be interpolated from intact parts based on the correlation structure estimated over the intact parts. *Outlier detection* allows the corrupt parts to be identified. *Step removal* fixes the high-amplitude flux jump artifacts that are common with some MEG systems. *Ringing removal* allows the ringing response of the antialiasing filter to glitches (steps, pulses) to be suppressed. The performance of the methods is illustrated and evaluated using synthetic data and data from real EEG and MEG systems. These methods, which are mainly automatic and require little tuning, can greatly improve the quality of the data.

### Introduction

The very weak brain signals picked up by electroencephalography (EEG) or magnetoencephalography (MEG) have to compete with multiple sources of noise and artifact within the body, the environment, and the sensors or electrodes. Of particular concern are electrode-specific or sensor-specific sources, because they cannot be suppressed by combining channels linearly as in ICA, beamforming or other linear techniques (Parra et al., 2005; Debener et al., 2010). In EEG they include *slow drifts* at the electrode/gel/skin interface (Huigen et al., 2002; Kappenman and Luck, 2010), and in MEG the large amplitude steps that result from a slip in the flux-lock loop (Gross et al., 2013), as well as various other glitches of diverse nature. We will use the terms “artifact” and “noise” interchangeably as the distinction between them is not well defined.

Many techniques have been proposed to eliminate or palliate

artifacts, some of them well-established and included in standard guidelines and processing pipelines. These include temporal and spatial filtering, detrending, regression, rereferencing, rejection of corrupt data, and spatial interpolation. However, many of these methods are prone to failure for certain combinations of artifact, and in some cases they may make things worse. As an example, *high-pass filtering*, a standard method to deal with drifts, is highly sensitive to the presence of temporally-localized glitches that trigger ringing of the filter. Here we revisit the issue of preprocessing, and propose a methodology of robust estimation of regression parameters, including methods for *detrending*, *outlier detection*, *inpainting*, *rereferencing*, *step removal*, and *ringing artifact removal*. These methods have in common that they deal with artifacts that are not easily removed with standard spatial filtering techniques.

We first review some important artifacts that affect EEG and MEG data.

\* Corresponding author. DEC, ENS, 29 rue d'Ulm, 75230, Paris, France.

E-mail address: [Alain.de.Cheveigne@ens.fr](mailto:Alain.de.Cheveigne@ens.fr) (A. de Cheveigné).

<https://doi.org/10.1016/j.neuroimage.2018.01.035>

Received 15 September 2017; Received in revised form 4 December 2017; Accepted 15 January 2018

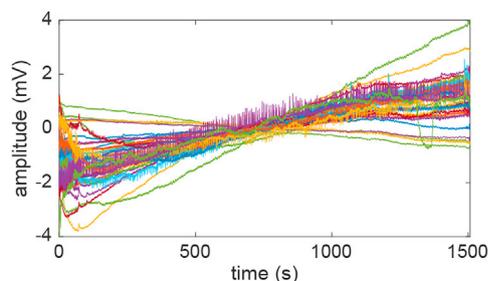
Available online 12 February 2018

1053-8119/© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

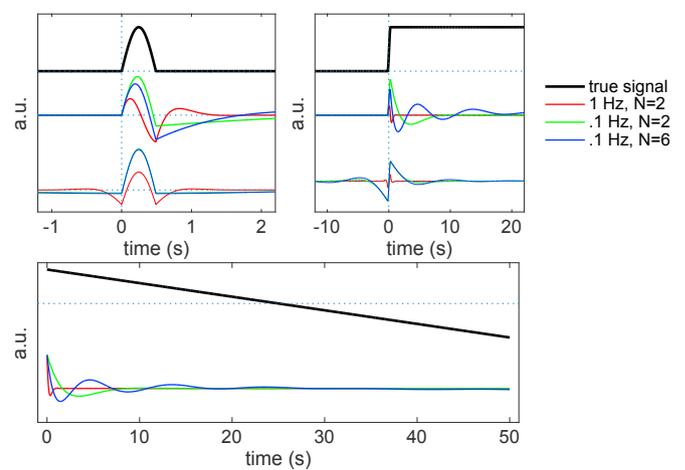
**Drifts in EEG.** The EEG signal often rides upon a slow drift signal produced at the skin/electrolyte/electrode interface (Huigen et al., 2002). Fig. 1 shows the time course of a typical sample of EEG data after removal of the mean from each channel. Each channel seems to follow its own drift pattern, albeit with some apparent inter-channel correlation. Slow drift can be minimized at the source, by piercing or abrading the outermost layer of the skin, the stratum corneum (Stjerna et al., 2010; Vanhatalo et al., 2005), but not all experimenters or subjects welcome this solution. Slow drifts may mask genuine cortical activity in the very low frequency range (Vanhatalo et al., 2005), and furthermore, if the data are epoched, the drift may misleadingly appear as a pattern reproducible over trials, a tendency that may be further reinforced by component analysis techniques that emphasize repeatable components (de Cheveigné and Parra, 2014). Drifts are usually dealt with by high-pass filtering or detrending, but unfortunately those methods themselves may create new artifacts as discussed in the next paragraphs.

**High-pass filtering artifacts.** High-pass filtering is effective to attenuate slow drifts (typical cutoffs range from 0.1 to 2 Hz) but it raises several issues. Most obvious is the removal of slow cortical activity, whether spontaneous (Vanhatalo et al., 2005), or stimulus-evoked (Southwell et al., 2017). Processing that blinds the investigator to such cortical activity is not ideal. Another problem, less often appreciated, is that transient features too may be distorted due to the filtering (Acunzo et al., 2012; Vanhatalo et al., 2005). Fig. 2 (top left) shows how a unipolar pulse (0.5 s duration) is affected by typical high-pass filters (Butterworth design) of various orders and cutoff. The pulse is attenuated relative to the original signal, and is followed by multipolar excursions that are *purely artifactual*, with a morphology that depends on the cutoff frequency, order, and type of filter. A transient cortical event would incur similar distortion, which is worrisome as the additional excursions might masquerade as distinct neural events. The point has been repeatedly made (Acunzo et al., 2012; Tanner et al., 2015, 2016; Widmann et al., 2015), but it escapes consideration in many studies.

In the same vein, Fig. 2 (top right) shows how a step (0.5 s rise time), such as might occur in the cortical response to a stimulus onset (Southwell et al., 2017), is affected by the same filters. The sustained plateau is lost and replaced by spurious signal excursions of both signs that would completely obfuscate the pattern of cortical electrical response. An offset would trigger a similarly misleading “offset response” (not shown). Zero-phase filtering is sometimes recommended as conducive to less distortion, but it introduces an additional issue. Fig. 2 (top right, bottommost lines) shows the response of



**Fig. 1.** Sample of 40-channel EEG data with slow drifts. Data were acquired with a BioSemi system at a rate of 2048 Hz in the calibration phase of an experiment investigating auditory perception and brain state. The mean of each channel was subtracted before plotting.



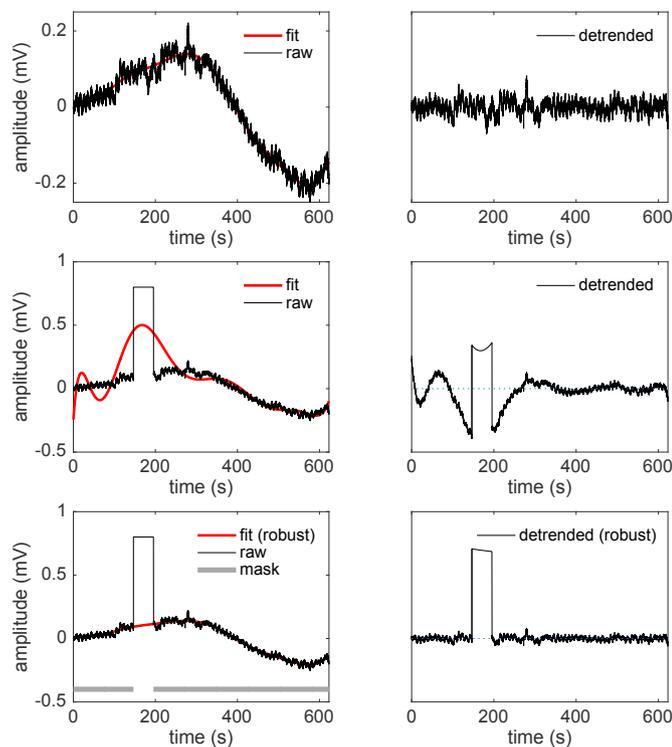
**Fig. 2.** Top: pulse (left, black) and step (right black) signals and corresponding filter outputs for 3 different high-pass filters. Lowermost curves are for non-causal filters (Matlab filtfilt function). Bottom: descending ramp (black) and filter output for 3 different Butterworth high-pass filters with cutoff frequency and order (N) indicated in the legend.

zero-phase filters to the same step. The presence of spurious deflections is now made worse by the fact that some of them occur *before* the event that triggered them. Applied to neural data, non-causal filters may confuse our understanding of the sequence of events within the brain and their causal relations.

Fig. 2 (bottom) illustrates a high-pass filter response to a descending ramp (analogous to the EEG data in Fig. 1). After a transient response (due to the implicit step at the onset) the signal is close to zero, as desired. However the transient fluctuations take time to subside, especially for a filter with a lower cutoff and/or higher order. They are of particular concern if the high-pass filter is applied after cutting the data into epochs (not recommended), as the fluctuations can easily be mistaken for a reproducible neural response. Such fluctuations may be reduced by including more data at the beginning of each epoch, to give time for the transient to subside, however the response of an infinite impulse response (IIR) filter never completely dies away. Together, these issues provide strong motivation to consider alternatives to high-pass filtering.

**Detrending artifacts.** In *detrending*, a smooth function, for example a low order polynomial, is fit to the data and then subtracted from it. Fig. 3 (top left) shows one channel of an EEG signal contaminated with a slow drift, together with a polynomial fit of order 10 (red), and Fig. 3 (top right) shows the same signal after subtraction of the fit. Slow fluctuations have disappeared, leaving rapid fluctuations on several time scales. The order of the polynomial controls the scale of the fluctuations that are removed.

Detrending is unfortunately sensitive to the presence of *glitches*, that are common in EEG and MEG. Fig. 3 (middle left) shows the same data to which a glitch has artificially been added. The fit (red) is affected by the glitch, with the result that the detrended signal now shows high-amplitude fluctuations over an interval that extends outside the glitch-contaminated part (Fig. 3 middle right). However, if the extent of the glitch is known, the fit can be restricted to the intact part (indicated by the grey bar). Fig. 3 (bottom left, red) illustrates such a fit, and Fig. 3 (bottom right) shows resulting detrended data. The glitch is of course still present, but the rest of the data are nicely detrended without spurious fluctuations. This approach is described in more detail in this paper.



**Fig. 3.** Top left: sample of EEG signal (black) and order-10 polynomial fit (red). Right: detrended data. Middle left: same EEG data as top with an artificial glitch (black) and polynomial fit (red). Right: “detrended” data. Bottom: same data with robust polynomial fit (red). The fit was weighted using the weighting function (mask) symbolized in grey. Right: detrended data. Data are from the same dataset as Fig. 1.

**Rereferencing.** EEG measures the potential at each electrode *relative* to some other electrode (Picton et al., 2000). The signal recorded on any channel obviously depends on the position of the reference electrode on the skull, and any noise introduced at the reference electrode (e.g. contact noise) contaminates all measured signals. To reduce the dependency on reference electrode position it is common to rereference to the average, i.e. subtract from each electrode the average over all electrodes (Nunes and Srinivasan, 2006; Keil et al., 2014) or some other linear combination (Yao, 2001). Noise at the physical reference electrode is cancelled, but noise on any other electrode is now injected, via the average, into *all* channels. This paper suggests a means to mitigate this issue.

**Glitches.** EEG and MEG are susceptible to temporally-localized glitches, due to muscle artifacts or motion of the electrodes or leads. Channel-specific artifacts (that affect one rather than multiple electrodes) are particularly troublesome as they cannot be suppressed by linear techniques without effectively discarding the contaminated channel (de Cheveigné, 2016). They can interfere with data-driven linear techniques, as illustrated in Fig. 3 (middle), and it is useful to identify them prior to other analysis.

Channel specific artifacts are usually uncorrelated across channels, in contrast to genuine brain activity that tends to be correlated across channels because of current spread (de Cheveigné and Simon, 2008). This suggests a way to *detect* them: (a) project a channel on the subspace spanned by all other channels, (b) estimate the statistics of the residual (channel minus projection), (c) find the outliers of this distribution. However this scheme runs into a practical difficulty: artifacts on other

channels may corrupt the projection, and thus wrongly trigger the detection of artifacts on the channel of interest. To be effective, the algorithm should track the multiple combinations of intact/corrupt channels that can occur in real data, so as to always project on intact data. This paper describes such an algorithm.

**Inpainting** is a term used in the field of image processing to designate the process by which a missing portion of an image is replaced using a model constrained by the intact parts (Bertalmio et al., 2000), and the same idea has been used for other signals such as audio (Adler et al., 2012). Here we apply the concept to electrophysiological data. Supposing the extent of an outlier or glitch (as in Fig. 3 middle) is known (for example from the previous algorithm), it can be replaced using a model constrained by intact data. This paper proposes such a process. Inpainting and outlier detection are closely related.

**MEG steps.** MEG has the advantage over EEG that it is not prone to the slow drift that arises at the electrode-skin contact in EEG. Slow nuisance signals may be present, due to environmental sources such as moving vehicles or machinery, but they are usually correlated across sensors and can be projected out of the data with linear methods. However, MEG is prone to sudden changes in the operating point of the flux lock loop associated with the SQUID sensors (Oswal et al., 2016; Gross et al., 2013), that produce large amplitude steps, against which high-pass filtering is particularly ineffective. This paper proposes a reliable method to remove these steps.

**Antialiasing filter artifacts.** An event such as a flux jump (as mentioned above) or stimulus artifact (for example as occurs in cochlear or deep brain stimulation) necessarily triggers a ringing artifact due to the antialiasing filter that precedes conversion of the analog EEG or MEG signal to a digital representation. The ringing extends beyond the interval containing the event, and may remain after the event has been removed (as in the case of MEG steps). This paper proposes a method to address this issue.

In summary, electrophysiological data are plagued by many artifacts that interfere with the interpretation of the data, and that may reduce the effectiveness of processing algorithms needed to assist with this interpretation. Classic methods such as filtering are particularly prone to artifacts and signal distortion effects. This paper describes a panoply of alternative techniques to deal with these issues. These techniques have in common that they target artifacts that are not easily removed with standard spatial filtering techniques such as ICA. They are complementary with those techniques: once artifacts have been suppressed, techniques such as ICA may be more effective.

## Methods

### Signal models

We use three signal models. The first and last are applicable to a single channel independently from others, the second is applicable to multiple channels taken together.

For the first signal model, the data of channel  $n$  of a data matrix are expressed as:

$$x_n(t) = y_n(t) + f_n(t), \quad (1)$$

where  $f_n(t)$  is the projection of  $x_n(t)$  on some basis of functions (for example low-order polynomials) chosen to represent trends, and the residual  $y_n(t)$  is a “detrended” signal. We assume that this model holds for all values of  $t$  except those for which a weighting function  $w_n(t)$  is zero.

The second model is best described by first defining a simpler model that states that each channel can be expressed as a weighted sum of the other channels:

$$x_n(t) = \sum_{n' \neq n} b_{nn'} x_{n'}(t). \tag{2}$$

Equation (2) implies linear dependency between channels (rank of the data  $< N$ ) but is a stronger condition. The idea expressed by this model is that no single channel is “indispensable”, as any channel can be reconstructed from the other channels. We will further assume that there are multiple ways to reconstruct the channel, based on different subsets of the other channels, i.e. Eq. (2) also holds for subsets of  $\{n' \neq n\}$ .

This simple model is now modified by assuming the presence of temporally sparse glitches, labeled by zero values of a weighting matrix  $w_n(t)$ . This may invalidate the baseline model (Eq. (2)), however, we make three assumptions. The first is that for every  $t$  such that channel  $n$  is valid (i.e.  $w_n(t) \neq 0$ ), there exists some subset  $\mathcal{O}_n$  of valid channels  $\{n' \neq n\}$  for which Eq. (2) holds. In other words, glitch-free parts of channel  $n$  can be reconstructed based on the other channels, although the “recipe” to do so might be different at different  $t$ , i.e. there are several such subsets  $\mathcal{O}_n^k$ . The one that is valid at  $t$  is notated as  $\mathcal{O}_n^{k(t)}$ . The second assumption is that at least one of these subsets  $\mathcal{O}_n^k$  is also valid during the glitch that affects channel  $n$ . In other words it is possible to fix samples within the glitch-contaminated segment based on intact channels. The third assumption is that there are enough intact data so that all the required coefficients  $\{b_{nn'}\}_k$  in Eq. (2) can be estimated reliably. These assumptions are plausible if the glitches are temporally sparse and don't occur at the same time on all channels.

For the third model, each channel is expressed as:

$$x_n(t) = y_n(t) + f_n(t) \tag{3}$$

where  $f_n(t)$  is now *piecewise constant*. This model will allow us to address steps in MEG data.

### Robust detrending

This algorithm uses the first model. The slow trend on each channel  $n$  is estimated robustly using a weighting matrix  $w_n(t)$ , and then subtracted. The weighting matrix can be predetermined or else estimated iteratively with the following algorithm:

**Data:** corrupted N-channel data  $x_n(t)$ , weighting matrix  $w_n(t)$

**Result:** reconstructed data  $\bar{x}_n(t)$

**for**  $n \leftarrow 1$  **to**  $N$  **do**

Partition time axis into  $K$  subsets  $\{T_k\}$  such that channels  $\mathcal{O}_n^{k(t)}$  are valid over  $T_k$ ;

**for**  $k \leftarrow 1$  **to**  $K$  **do**

Select subset  $T'_k \subseteq T_k$  of times for which channel  $n$  is also valid ;

On  $T'_k$  estimate the matrix  $\mathbf{M}$  that projects  $x_n(t)$  on  $[x_{n'}(t)]_{n' \in \mathcal{O}_n^k}$  ;

Apply matrix  $\mathbf{M}$  to  $[x_{n'}(t)]_{n' \in \mathcal{O}_n^k}$  to obtain  $\bar{x}_n(t)$  for  $t \in T_k$  ;

*Optionally:*  $\bar{x}_n(t) \leftarrow x_n(t)$  for  $t \in T'_k$  ;

**end**

**end**

**Data:** EEG signal  $x(t)$

**Result:** detrended EEG  $y(t)$ , outlier map  $w(t)$

Initialize  $w(t) \leftarrow 1$ ;

**while not finished do**

$f(t) \leftarrow x(t)$  fit to basis functions using weights  $w(t)$  ;

$d(t) \leftarrow |x(t) - f(t)|$  ;

**if**  $|d(t)|/STD(|d(t)|) > threshold$  **then**

$w(t) \leftarrow 0$

**else**

$w(t) \leftarrow 1$

**end**

**end**

At each iteration the data are projected onto the basis functions, statistics of the residual  $d(t)$  are estimated, and samples larger than a threshold are flagged as outliers to be discounted in the next iteration. The loop is terminated after a few iterations, or if  $w(t)$  does not change between iterations. The weights  $w(t)$  can be initialized to values other than 1 if prior knowledge is available. A quality-of-fit score can be defined as  $\sum_{w(t) \neq 0} d(t)^2 / \sum_{w(t) \neq 0} x(t)^2$ . For multichannel data the algorithm is applied independently to each channel.

### Inpainting

This algorithm uses the second signal model to estimate missing portions of multichannel data  $x_n(t)$  (labeled by zero values of a weighting matrix  $w_n(t)$ ). According to that model, the signal  $x_n(t)$  of channel  $n$  can be approximated for every  $t$  as a weighted sum  $\bar{x}_n(t)$  of a subset  $\mathcal{O}_n^{k(t)}$  of the other channels. The subset may differ for different  $t$  according to which other channels are intact at that time. The algorithm estimates projection parameters based on intact portions of the data, and applies them to reconstruct the corrupted portions. More precisely:

In brief, corrupt values of each channel are reconstructed from channels that are intact at that time. Subsets  $T_k$  or  $T'_k$  do not need to be formed of contiguous samples, but the subset  $T'_k$  must be large enough to reliably estimate the projection matrix  $\mathbf{M}$ . If the step labeled as *optional* is included, the output  $\bar{x}_n(t)$  is identical to the original signal except during the glitch. If that step is omitted, the output is everywhere equal to a weighted sum of other channels ( $n' \neq n$ ), i.e. all output samples may differ from the input. The rank of the data is not usually reduced by processing.

#### Outlier detection

The previous algorithm requires a map  $w_n(t)$  indicating which parts of the data are corrupt. If that information is missing it can be derived from the data using an adapted version of the algorithm. Starting with an initial estimate of the weights  $w_n(t)$ , for example all ones, each channel  $n$  is approximated on the basis of other channels ( $n' \neq n$ ). The approximation  $\bar{x}_n(t)$  is then compared to the actual signal  $x_n(t)$ , and the times at which the mismatch exceeds a threshold are flagged as corrupt. More precisely:

**Data:** signal  $x_n(t)$

**Result:** weighting matrix  $w_n(t)$

Initialize  $w_n(t) \leftarrow 1$  for all  $n$ ;

**while not finished do**

$\bar{x}_n(t) \leftarrow$  inpainting algorithm applied to  $x_n(t)$ ,  $w_n(t)$

$d(t) \leftarrow |\bar{x}_n(t) - x_n(t)|$ ;

**if**  $d(t)/STD(d) > threshold$  **then**

        |  $w_n(t) \leftarrow 0$

**else**

        |  $w_n(t) \leftarrow 1$

**end**

**end**

The algorithm is iterated a few times to refine the outlier map  $w_n(t)$ . The value of *threshold* is not critical, a value of 1 is usually adequate. A quality-of-fit score can be defined as  $\sum_{n, w_n(t) \neq 0} d_n(t)^2 / \sum_{n, w_n(t) \neq 0} x_n(t)^2$ .

#### Robust rereferencing

It is common in EEG to rereference the data by subtracting the average of signals over all channels  $m(t) = (1/N) \sum_n x_n(t)$ . As explained in the Introduction, a glitch on one channel can corrupt this average, and thus contaminate all channels. This can be avoided by replacing  $x_n(t)$  in the formula for the mean by  $\bar{x}_n(t)$  obtained from the inpainting algorithm. Alternatively, a simple expedient is to replace the mean over channels by a weighted mean:  $m(t) = \sum_n w_n(t) x_n(t) / \sum_n w_n(t)$ . Analogous schemes for robust rereferencing were proposed by [Lepage et al. \(2014\)](#); [Bigdely-Shamlo et al. \(2015\)](#).

#### Step removal

This algorithm uses the third signal model (piecewise constant) to locate and remove large amplitude steps such as SQUID jumps commonly

observed in MEG data ([Gross et al., 2013](#)). The position  $t_0$  of a single jump in a signal  $x(t)$  of duration  $T$  can be found reliably as:

$$t_0 = \underset{t \in [1, T]}{\operatorname{argmin}} (V_1^t(x) + V_t^T(x)), \quad (4)$$

where  $V_1^t(x) = \sum_{i=1}^t (x(i) - M_1^t(x))^2$  is the sum of squared deviations from the mean  $M_1^t(x) = (1/t) \sum_{i=1}^t x(i)$  of  $x(t)$  over  $[1, t]$ . As described, the algorithm suffers from a slight bias towards choosing a step position close to either extremity (0 or  $T$ ). This is because the variance of  $V$  increases as the number of samples used to calculate it decreases, making a spurious minimum in Eq. (4) more likely. To counteract this trend and avoid trivial splits, it is convenient to restrict the search to the interval  $[T_g, T - T_g]$  where  $T_g$  is a guard value. Likewise it is convenient to set a threshold on the ratio  $V_1^t(x) + V_t^T(x) / V_1^T(x)$  below which small amplitude “jumps” are ignored. Multiple jumps are located by recursing on each side of  $t_0$ . The jumps are then removed from the data by subtracting the mean of the data between jumps. *Robust* step removal can be implemented by including a weighting term  $w(t)$  in Eq. (4).

#### Ringling removal

After a step is removed by the previous algorithm, there remains a ringing artifact due to the step response of the antialiasing filter. If the characteristics of that filter are known, the ringing artifact can be removed by feeding a step to the same filter, and subtracting the response. A simple expedient to find the appropriate filter is to apply the Steiglitz-McBride iteration method ([Steiglitz and McBride, 1965](#)) (Matlab function `stmcb`) to a short segment of data following the step, yielding numerator and denominator coefficients of an IIR filter. Given that there is only a small number of coefficients (e.g. 8 each for numerator and denominator) there is little room for overfitting. This method is also effective to remove ringing associated with stimulus artifacts, for example when measuring responses to electrical stimulation ([Oswal et al., 2016](#)). In principle it is possible to average filter parameters estimates over events to get a more accurate estimate, however in practice the appropriate filter depends on the timing of the step relative to the sampling instants, which is unknown. The Steiglitz-McBride approach is easier for this reason.

In summary, a panoply of algorithms is available to address the issues raised in the Introduction. The next section illustrates their performance. These techniques target artifacts that are not easily removed with standard spatial filtering methods such as ICA, and thus are complementary with those methods.

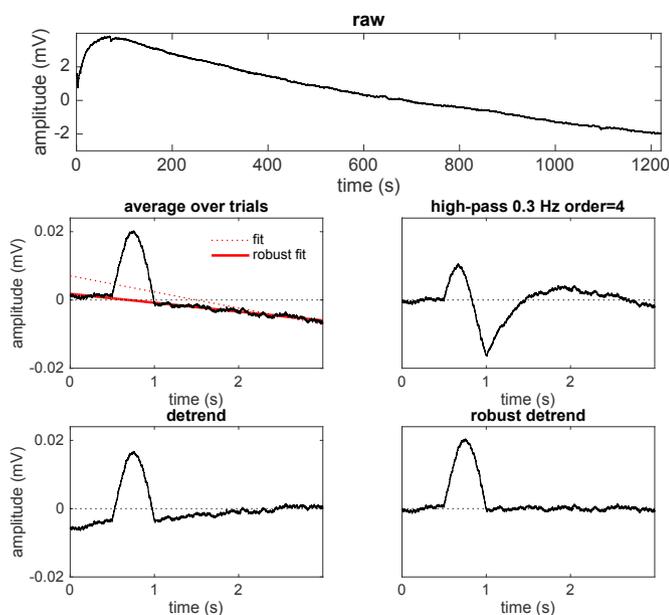
## Results

The algorithms are illustrated with synthetic data to clarify their properties, and with real data to evaluate their usefulness in a practical setting.

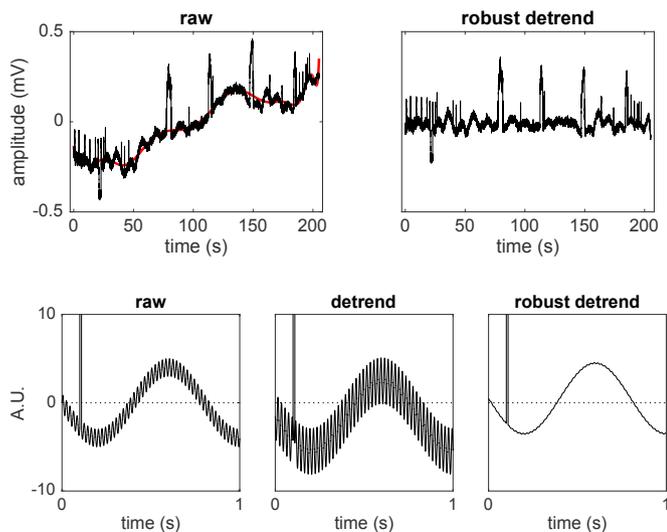
#### Robust detrending

[Fig. 3](#) (bottom) shows the result of applying robust detrending rather than standard detrending. After a few iterations the algorithm has estimated the position of the glitch (grey line) and performed a weighted polynomial fit (red). Subtracting the fit yields the detrended data (bottom right). Comparing to [Fig. 3](#) middle right, it appears that the algorithm has achieved a better fit to the non-glitch parts, yielding a more useful detrended signal.

[Fig. 4](#) (top) shows one channel of real EEG data on which are superimposed 200 repetitions of a synthetic unipolar pulse with a shape similar to that shown in [Fig. 2](#) (the low-amplitude pulses are not visible in the waveform at this scale). Averaging over trials reveals the pulse (black line in [Fig. 4](#)), but the linear trend that affects the raw data ([Fig. 4](#), top) also affects the average over trials ([Fig. 4](#), middle left). This can be



**Fig. 4.** Robust detrending of EEG data. Data consist of 200 repetitions of a synthetic unipolar pulse of duration 500 ms and amplitude  $20 \mu\text{V}$  superimposed on a real EEG signal (top). Middle left: trial average (black), linear fit (dotted red) and robust linear fit (full red) to trial average. Middle right: trial average of high-passed data (0.3 Hz cutoff, order 2). Bottom left: detrended trial average. Bottom right: robust detrended trial average. Data are from the same dataset as Fig. 1.



**Fig. 5.** Top left: raw EEG signal (black) and 30th order polynomial fit (red). Top right: detrended signal. Data are from the same dataset as Fig. 1. Bottom: Robust removal of a 50 Hz sinusoidal trend. Left: 1 Hz sinusoidal “signal” corrupted by 50 Hz artifact and a temporally-localized glitch (amplitude 100). Center: a 50 Hz sinusoidal function is fit to the data and subtracted. Right: same, but the fit is weighted by a weighting matrix that is zero at time of the glitch.

addressed by applying a high-pass filter *before* cutting the data into trials, as is common practice in EEG data analysis. However doing so also seriously distorts the shape of the pulse (Fig. 4, middle right). Filtering *after* cutting into trials has an even worse effect due to onset response artifacts that repeat on each trial (not shown). As an alternative to high-pass filtering, the trial-averaged data may be detrended by fitting a low-order polynomial to the averaged data and subtracting the fit (Fig. 4 bottom left). However, the presence of the pulse affects the fit (Fig. 4

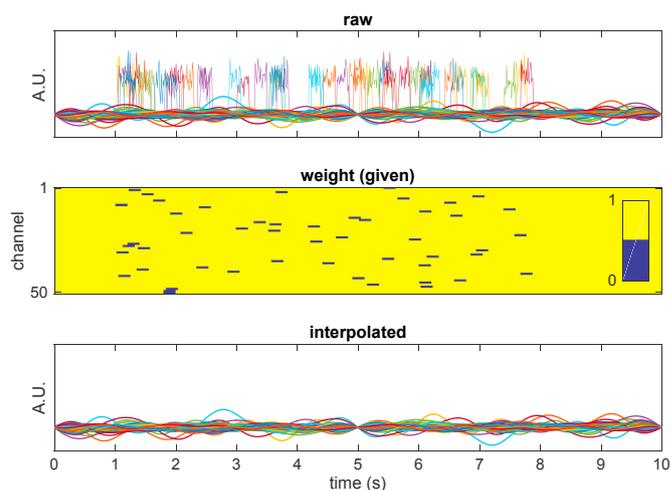
middle left, red dotted line), adding a reverse trend to the response (bottom left). In contrast, the *weighted* fit (Fig. 4 middle left, red full line) leads to a better estimate of the target signal (bottom right).

As an additional example, Fig. 5 (top left) shows one channel of raw EEG with multiple glitches, together with a robust polynomial fit of order 30 (red line). Fig. 5 (top right) shows the detrended signal. The time scale of the variations that are included in the trend depends on the order of the polynomial, a higher order allowing finer fluctuations to be removed (see Discussion for caveats). As a final example, Fig. 5 (bottom left) shows a synthetic signal corrupted by both a 50 Hz sinusoidal artifact and a temporally-localized glitch (amplitude 100, shown truncated). An effective way to remove the sinusoidal artifact is to fit a quadrature pair of 50 Hz sinusoids (in lieu of “trend”) and subtract the fit, equivalent to removing that component in the Fourier domain, but the presence of the glitch disrupts this process and leads instead to a *greater* sinusoidal artifact amplitude (Fig. 5, bottom center). In contrast the weighted fit produces a clean signal (bottom right). Notch filtering is prone to similar effects, that weighted detrending conveniently avoids. This last example shows that robust detrending is not restricted to slow trends.

It is worth remarking that the same algorithm can be used instead to robustly follow the *slow* variations in the signal, analogous to a low-pass filter. The advantage over a standard lowpass filter is again that the result is not affected by localized glitches. These persist of course in the “smoothed” signal, but the shape of the non-glitch portions is not affected.

**Inpainting.** Fig. 6 (top) shows a 50-channel signal  $[x_n(t)]$  produced by mixing 10 uncorrelated “sources”, consisting of sinusoids of different frequencies, via a  $10 \times 50$  random mixing matrix. The signal of each channel was corrupted by adding a randomly-placed “glitch” of duration 0.2 s. The positions of these glitches are indicated by the weighting matrix  $[w_n(t)]$  (Fig. 6 middle). Applying the inpainting algorithm produces a cleaned signal (Fig. 6 bottom) indistinguishable from the glitch-free original (not shown). To achieve this result, the algorithm estimated the correlation structure of intact portions ( $w_n(t) \neq 0$ ), and used them to design a set of projection matrices to reconstruct corrupt portions ( $w_n(t) = 0$ ). In this example there were 146 distinct configurations of intact/corrupt channels, and thus the same number of projection matrices. Note that none of the channels was glitch-free, and that many glitches overlapped in time.

**Outlier detection** In the previous example, the weighting matrix  $w_n(t)$  was known in advance. If no prior knowledge is available, a weighting matrix can be estimated from the data based on the assumption that glitches are uncorrelated across channels. Fig. 7 (top) shows the



**Fig. 6.** EEG signal inpainting. Top: 50-channel synthetic signal (rank 10) corrupted with randomly-placed “glitches” (thin lines). Middle: weighting matrix, zero at the positions of the glitches. Bottom: signal interpolated by the inpainting algorithm.

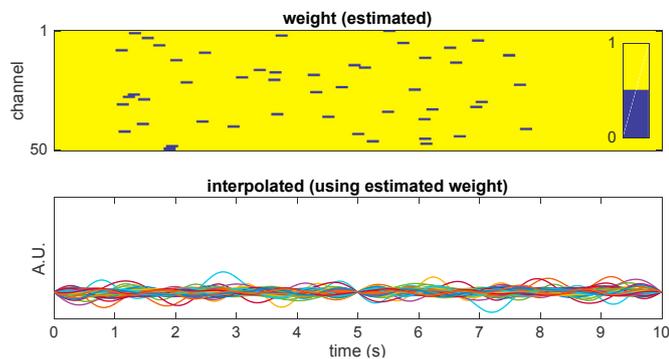


Fig. 7. Outlier detection. Top: weight mask estimated from the signal plotted in Fig. 6 (top). Bottom: signal interpolated using the estimated weighting mask.

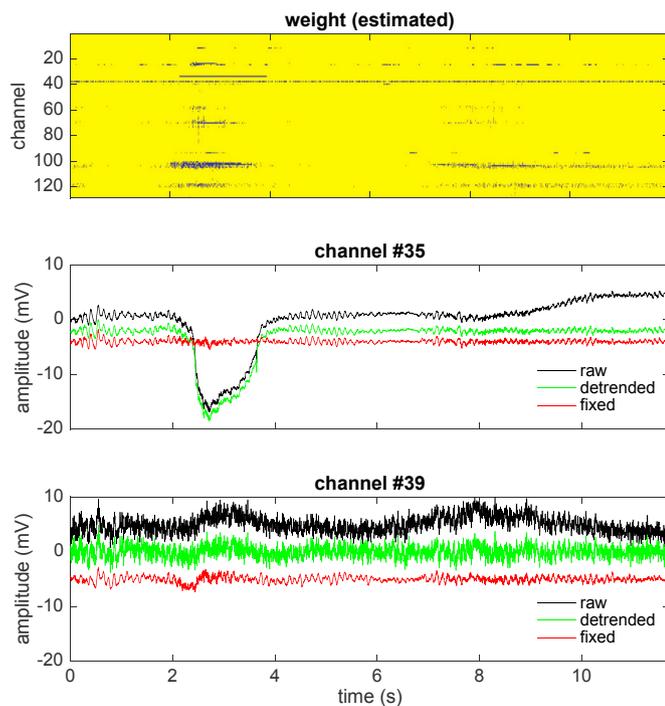


Fig. 8. Outlier detection. Top: weight mask estimated from a segment of 128-channel EEG (detrended). Middle: one channel of EEG, showing the raw, detrended, and interpolated signal, offset vertically for clarity. Bottom: same for another channel. Data are from the same dataset as Fig. 1.

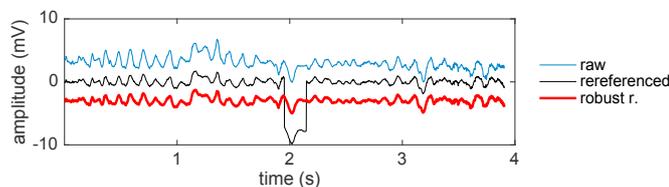


Fig. 9. Robust rereferencing. Single EEG channel before (blue) and after standard rereferencing (black) and robust rereferencing (red).

weighting matrix estimated by the outlier detection algorithm (threshold = 1, 10 iterations) from the data plotted in Fig. 6 (top). Based on this estimate, derived blindly from the data, the data can be effectively denoised (Fig. 7 bottom), yielding a result almost identical to the original (not shown).

As another example, a segment of 128-channel EEG data was first detrended and then processed with the outlier detection algorithm

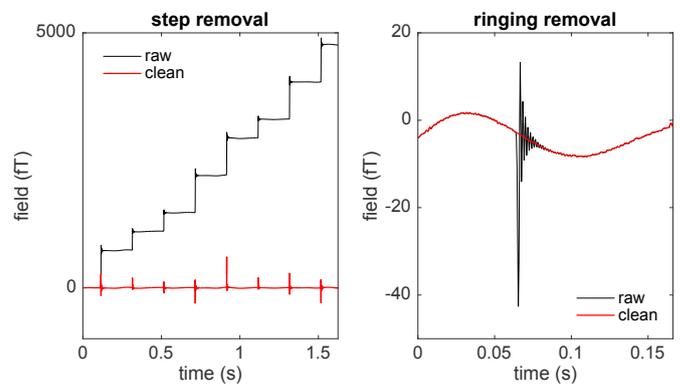


Fig. 10. Step and ringing artifact removal. Left: one channel of MEG data in response to simulated deep brain stimulation (Oswal et al., 2016) (black) and the same signal after automatic step removal (red). Right: one channel of MEG from same data set showing a stimulus artifact (black) and the same signal after ringing removal (red). Data were recorded on a 275-channel CTF system at a 2400 Hz sampling rate (Oswal et al., 2016).

(threshold = 1, 6 iterations). The estimated weight matrix is displayed in Fig. 8 (top). The pattern of outliers is quite diverse, with some channels affected most of the time (e.g. channel 39) and many channels some of the time (e.g. circa 2–3 s). Fig. 8 (middle) shows the time course of channel 35 before (black) and after robust detrending (green), and after outlier detection and inpainting (red). The signals are offset vertically for clarity. The invalid portion (between 2 and 4 s), visible as a horizontal line in the weights plot, is replaced by a combination of intact channels. Fig. 8 (bottom) shows similar data for channel 39 (weights visible as a horizontal dotted line across the full data segment). Here the algorithm has replaced a large number of samples, yielding an apparently cleaner signal (red trace).

**Robust rereferencing.** Fig. 9 (blue) shows the time course of one channel of a segment of 128-channel EEG data. Another channel (not shown) was affected by an artificial glitch of amplitude  $-1000$  mV, with the result that the channel average also shows a glitch (of amplitude  $\sim -8$  mV) that is injected into the rereferenced waveform (Fig. 9, black). Robust rereferencing avoids this problem (red).

**Step removal.** Fig. 10 (left) shows a sample of MEG data recorded with a phantom dipole source in an experiment that simulated conditions characteristic of deep brain stimulation (Oswal et al., 2016). Stimulation produces a high-amplitude magnetic pulse that can cause the electronics controlling the SQUID sensors to switch state, resulting in a sharp step of large amplitude. Such steps also occur spontaneously in MEG data (Gross et al., 2013). In this example, at each stimulus the MEG signal (black) is affected by a step that the algorithm removes (red).

**Ringing removal.** Visible in the previous example are a series of glitches due to the ringing response from the low-pass antialiasing filter, that remains after step removal (Fig. 10, right). That ringing response (black) is effectively removed by the ringing removal algorithm (red). Ringing artifacts from other events (e.g. stimulus artifacts) can also be removed if their timing is known.

## Discussion

### The importance of preprocessing

Artifacts are deleterious in four ways. First, they *mask* interesting activity; for example, line noise may mask oscillatory cortical activity near 50 or 60 Hz, and electrode drift may mask slow cortical potentials. Second, they may *masquerade* as brain activity; for example a slow trend might lend a ramp-like shape to an evoked response, or ocular or myogenic artifacts might masquerade as cortical activity if they correlate with stimulation or behavior (Yeung et al., 2004; Yuval-Greenberg et al., 2008). Third, the need to attenuate them may prompt the use of

processing such as high-pass filtering that entails deleterious *side effects*. Fourth, the artifacts may *impede analysis* by interfering with the processing; for example, high-amplitude artifacts may dominate the sums of squares involved in methods such as CSP, ICA, CCA, etc. Artifact removal is a prerequisite for good data analysis, and this paper offers a panoply of methods for the purpose. They are presented here together (rather than in separate papers) because they share concepts and processing, and because the interactions between methods need to be considered.

#### In which order?

Should a slow trend be removed before glitches, or after? Should they both be removed before spatial filtering, or after? At what stage should temporal filtering be applied, if at all? Linear operations may be swapped, but the data-dependent algorithms that determine their parameters depend on the order in which they are applied. Algorithms that rely on minimizing a sum of squares are sensitive to components with large variance: a trend or glitch may entice an algorithm to model it rather than sources of interest, and thus removing the artifact may make the algorithm more effective. Definitive guidelines are hard to set because of the variety of situations. As a rule of thumb, if algorithm  $\mathcal{B}$  is sensitive to an artifact that algorithm  $\mathcal{A}$  can remove, then  $\mathcal{A}$  should be applied before  $\mathcal{B}$ . A difficulty arises of course if  $\mathcal{A}$  is also sensitive to artifacts that  $\mathcal{B}$  can remove.

A typical EEG recording might be contaminated by a combination of *slow drift* specific to each electrode, *temporally localized glitches* also specific to each electrode, *eye blinks* shared across electrodes, *50 Hz and harmonics* shared across electrodes, *myogenic activity* that may be either electrode-specific or shared, *alpha activity* shared across electrodes, and so-on. A likely sequence might be: (a) discard pathological channels for which there is no useful signal, (b) apply robust detrending to each channel, (c) detect and interpolate temporally-local channel-specific glitches, (d) robust re-reference, (e) project out eye artifacts (e.g. using ICA or DSS), (f) fit and remove, or project out, 50 Hz and harmonics, (g) project out alpha activity, etc., (f) apply linear analysis techniques (ICA, CSP, etc.) to further isolate activity of interest.

A typical MEG recording might be contaminated by a combination of *SQUID jumps* specific to each sensor, *slow components* due to nearby vehicles and machinery (e.g. elevator), shared across sensors, *50 Hz and harmonics* from power lines and devices, shared across sensors, *sensor noise* specific to each sensor, *vibration artifacts* shared across sensors, *alpha activity* shared across sensors, *stimulus artifact*, etc. A likely sequence might be: (a) remove squid jumps, (b) remove stimulus artifacts, (c) remove associated antialiasing filter ringing artifacts, (d) isolate and remove slow environmental components, 50 Hz and harmonics, vibration artifacts, alpha activity, etc. using projection techniques or spatial filtering based on ICA, DSS, etc., (e) suppress sensor noise using the Sensor Noise Suppression (SNS) algorithm (de Cheveigné and Simon, 2008), (f) apply linear analysis techniques (ICA, DSS, CSP, etc.) to further isolate activity of interest.

#### Comparison with other tools

A standard practice is to *discard* data corrupted by an artifact, for example discard a channel or trial. This is justified if those data are completely invalid, but wasteful otherwise. Data loss may be unacceptable if artifacts are spatially and/or temporally dense. One motivation for this work is to minimize such data loss while still attaining high data quality. Another is to avoid manual intervention, in contrast to common practices of *visual inspection* and *manual editing*. This goal is not fully attained, as it may still be necessary to manually adjust the order of processing stages or their parameters.

*Temporal filtering* is widely used (high-pass to remove drift, lowpass to attenuate sensor noise and myogenic artifacts, notch to remove power line components, etc.) but prone to the pitfalls noted in the Introduction. An aim of this work is to offer effective alternatives. *Detrending*, a popular

alternative to high-pass filtering, is impaired by glitches. Robust detrending reduces this problem, giving the method a major advantage over filtering. Like non-causal filtering, detrending can raise causality concerns as the trend removed at time  $t$  reflects data that occurs after  $t$ . In practice the strong constraint imposed by the basis functions (e.g. smooth low-order polynomials) rules out the most egregious effects. Polynomial trend removal is implemented in toolboxes such as FieldTrip (Oostenveld et al., 2011), but without weighting. An alternative to fitting a polynomial is local linear regression, in which a linear fit is repeatedly applied to short segments that are moved forward in time, and the fits averaged. A robust version of this process is proposed in the matlabmk toolbox ([http://kutaslab.ucsd.edu/matlabmk\\_fn\\_docs/matlabmk/robust\\_locdetrend.html](http://kutaslab.ucsd.edu/matlabmk_fn_docs/matlabmk/robust_locdetrend.html)). The robust rereferencing scheme of Bigdely-Shamlo et al. (2015) detects and interpolates corrupt channels over their entire duration, in contrast to the method reported here that does so only for corrupt segments of each channel.

A *spatial filter* is another common tool, with coefficients that are either predetermined (e.g. rereferencing, or Laplacian) or calculated from the data (e.g. ICA, PCA, etc.). A spatial filter allows multiple noise sources to be cancelled, but this works only if the dimensionality of the noise space (number of distinct active sources) is less than that of the data (number of channels). Usually, the coefficients of the filter remain constant over the data set, in other words the filter is *time invariant*. In contrast, the inpainting and outlier detection algorithms described here apply distinct spatial filters over different subsets of the time axis, similar to the STAR algorithm (de Cheveigné, 2016). The ability to apply different filters at different times sets these methods apart from standard time-invariant spatial filter methods such as PCA, ICA, CSP, beamforming, etc.

ICA is reported to benefit from high-pass filtering (Debener et al., 2010; Winkler et al., 2015), and should thus benefit from detrending. More generally, the methods presented here can contribute to the success of component analysis in two ways. First, by taking care of high-variance components that would otherwise dominate the least-squares solutions involved in many methods, and second by taking care of numerous channel-specific artifacts that inflate the number of “sources” that need to be resolved from the limited number of sensors available (Debener et al., 2010). It should be noted that the methods presented here do not usually reduce the rank of the data. They thus complement standard methods such as ICA.

The combination of outlier detection and inpainting can be seen as a “robust” version of our earlier SNS algorithm which projects each channel on the subspace spanned by the others (de Cheveigné and Simon, 2008). It can also be seen as a generalization of our earlier STAR algorithm which only allows one corrupt channel at any time (de Cheveigné, 2016). It is similar in spirit to the Artifact Subspace Reconstruction (ASR) method (Kothe, 2013), and related to spatial interpolation methods such as Perrin et al. (1989) that require neighboring channels to be intact. Only between-channel correlation is exploited here, for inpainting and outlier detection; an alternative is to exploit *temporal* structure within each channel via an autoregressive or wavelet model, or both space and time via a multichannel autoregressive model (Lawhern et al., 2012). Many methods have been developed for image, audio and video inpainting and matrix completion (Candès and Recht, 2008) that may also be of use for electrophysiological data. These remain to be fully explored.

The standard approach to address SQUID jumps in MEG is to discard corrupted data (Gross et al., 2013), and popular toolboxes such as FieldTrip (Oostenveld et al., 2011) or Brainstorm (Tadel et al., 2011) offer tools to detect such jumps. Discarding is not an option if the steps are dense on a majority of channels (e.g. Oswal et al. (2016)). In contrast, the method described here removes the steps with no data loss. After steps are removed, *ringing artifacts* due to the antialiasing filter remain, and the same artifacts may be observed in both EEG and MEG in response to electric stimulation (cochlear implant, deep brain stimulation, TMS, etc.). The standard approach is to discard the ringing interval together with the artifact (Herring et al., 2015), but this is not an option if artifacts

are dense. As far as we know, our methodology has not been reported elsewhere.

Robust techniques are well developed in statistics and data mining (Aggarwal, 2016), and some have been applied to EEG or other signals (Pernet et al., 2011; Lepage et al., 2014; Bigdely-Shamlo et al., 2015; Krauledat et al., 2007). Common to these techniques is the weighting of the data according to their reliability. The weighting is often implicit in the definition of the robust statistic, or in the estimation process, in contrast to the explicit weighting matrix that we use here. An advantage of explicit weights is that a measure of reliability derived in one context (e.g. detrending) can be applied in another (e.g. rereferencing).

Outliers are detected here based on mismatch from a model (spatial or temporal), rather than on the basis of their extreme values as is common. To derive the model we use standard statistics such as the mean and standard deviation. More robust statistics (median, etc.) have been advocated (Leys et al., 2013), but using them is unlikely to make a large difference in this context. Binary weights are used throughout; the use of graded weights is supported by the implementation, but unlikely to offer a major benefit.

#### Caveats and failure scenarios

**Detrending** assumes a signal model (basis) that must be flexible to fit the trend, yet inflexible so as to not absorb fluctuations of interest. The choice of parameters (e.g. polynomial order) is critical to correctly set the time scale of the trend to be removed, and success in this respect is uncertain if target and trend share a similar time scale. Robust detrending can fail if a glitch is mistaken as a trend (for example if it has a long temporal extent), or a trend mistaken as a glitch. To avoid this, it may be useful to first detrend using a low-order model, and then a higher-order model with weights constrained by the first fit. Simple trends are easily fit with a low order polynomial, or the first few terms of a Fourier series. Fitting more complex trends may run into difficulties due to what is known as Runge's effect (polynomials) or Gibb's effect (sinusoids) (Platte et al., 2011). As an example, in Fig. 5 (top right) the 30-th order polynomial fit leaves a residual fluctuation that is not completely eliminated by a higher-order order fit (not shown). A potential solution is to use splines (Shumway and Stoffer, 2011), or weighted smoothing interpolation (Davies and Meise, 2008), but these solutions have not been explored here. A simpler expedient is to apply detrending to shorter segments for which a low-order model is adequate. The outcome of robust detrending is dependent on the choice of the threshold parameter (in addition to those of the trend model), and we know of no convergence proof.

**Inpainting** fails if the signal to reconstruct is not contained in the subspace spanned by the intact channels (signal model 2). This can happen if the number of sources active at that time is greater than that the number of intact channels. In the extreme, if there are no intact channels, the algorithm cannot proceed (in that case the data is left untouched). Inpainting may also fail if there are insufficient data to reliably estimate the reconstruction matrices. Inpainting also fails if the channel is entirely corrupt, as the projection matrices cannot be estimated. The best option is to discard the channel (and possibly interpolate from intact neighbors).

**Outlier detection** assumes that the intact data match model 2, outliers being defined as samples that do not match the model. A same data set might match multiple models. For example two partly overlapping glitches could be interpreted as a pulse common to both (the part that for which they overlap), together with shorter glitches specific to each. The algorithm can choose either, and the choice that it makes might be unexpected. There is no guarantee of convergence of the outlier detection, and indeed the solution is often observed to fluctuate. That said, the algorithm typically gives plausible and useful results on real data.

**Robust rereferencing** is contingent on the quality of the outlier detection that produces the mask, although the impact of an error on any particular channel is mitigated via the averaging process.

**Step removal** could in theory remove genuine step-like activity, but in practice there is never any ambiguity and the method is robust. It runs into difficulty if the sections between steps are not flat. In that case a combination of step removal and detrending may be required.

**Ringling removal** treats a short segment of data as a filter impulse response, that it then removes. Brain signal features that occur during that short segment might be removed as well, although the likelihood of this happening is limited by the constraints imposed by the low-order IIR filter. The risk could be further reduced by fitting all ringling artifacts with exactly the same filter, although this option was not explored.

The methods described in this paper deal only with *channel specific* artifacts; artifacts that impinge on multiple channels are not addressed at all.

#### Implementation

The routines are implemented in Matlab within the NoiseTools toolbox (<http://audition.ens.fr/adc/NoiseTools/>). They are based on closed-form solutions, possibly iterated a small number of times, and thus most methods are cheap in terms of computation. Inpainting and outlier detection are somewhat more costly because a large number of combinations of valid/invalid channels may need to be considered (potentially up to  $2^N$  for  $N$  channels, in typical practice a few thousand). It can happen that channels within  $\mathcal{C}_n^{k(t)}$  are rank-deficient over  $T_k$ ; to handle this situation, for each  $k$  those channels are submitted to a PCA, the PC series is truncated, and the projection is performed on the remaining PCs. The eigendecomposition required by this step is the main computational bottleneck. To limit the cost of eigendecomposition ( $\mathcal{O}(N^3)$ ), for each channel  $n$ , a subset of size  $M < N$  of channels “closest” to channel  $n$  is selected on the basis of physical proximity or correlation with  $x_n(t)$ . To further reduce computation, configurations that involve only a few data samples are skipped, and the corresponding data are reconstructed by serial interpolation as a weighted sum of neighboring intact samples.

The algorithms are described here in terms of batch processing, but the statistics that they require (mean, covariance, cross-correlation) can be calculated incrementally and updated in real time, potentially allowing the methods to be deployed in a real-time brain-computer interface (BCI) system. One such application, cognitive control of a hearing aid, is a primary motivation of this work.

#### Conclusion

This paper presents a set of methods to preprocess multichannel data such as EEG or MEG and improve their quality. These methods address ubiquitous sources of artifact that corrupt data and interfere with analysis and interpretation, and complement other methods such as temporal or spatial filtering, either as a replacement with better performance and fewer drawbacks, or as a complementary processing step to make them perform better. The methods include robust detrending, robust rereferencing, outlier detection, inpainting (interpolation), step detection and correction, and ringling artifact removal.

#### Acknowledgements

This work was supported by the EU H2020-ICT grant 644732 (COCOHA), and grants ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL\*. Vladimir Litvak kindly pointed us to the MEG data of the study of Oswal et al. (2016) made available at [https://figshare.com/articles/phantom090715\\_BrainampDBS\\_20150709\\_01\\_ds.zip/4042911/3](https://figshare.com/articles/phantom090715_BrainampDBS_20150709_01_ds.zip/4042911/3), and helped test the step removal and ringling removal algorithms.

#### References

- Acunzo, D.J., MacKenzie, G., van Rossum, M.C.W., 2012. Systematic biases in early ERP and ERF components as a result of high-pass filtering. *J. Neurosci. Meth.* 209, 212–218.

- Adler, A., Emiya, V., Jafari, M.G., Elad, M., Gribonval, R., Plumbley, M.D., 2012. Audio inpainting. *IEEE Trans. Audio Speech Lang. Process.* 20, 922–932.
- Aggarwal, C.C., 2016. *Outlier Analysis*. Springer.
- Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C., 2000. Image inpainting. In: *Proc. SIGGRAPH00*, p. 417–424.
- Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.M., Robbins, K.A., 2015. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Front. Neuroinf.* 9, 16.
- Candès, E., Recht, B., 2008. Exact low-rank matrix completion via convex optimization. In: 2008 46th Annual Allerton Conference on Communication, Control, and Computing 6:1–49.
- Davies, P.L., Meise, M., 2008. Approximating data with weighted smoothing splines. *J. Nonparametric Statistics* 20, 207–228.
- de Cheveigné, A., 2016. Sparse time artifact removal. *J. Neurosci. Meth.* 262, 14–20.
- de Cheveigné, A., Parra, L.C., 2014. Joint decorrelation, a versatile tool for multichannel data analysis. *NeuroImage* 98, 487–505.
- de Cheveigné, A., Simon, J.Z., 2008. Sensor noise suppression. *J. Neurosci. Meth.* 168, 195–202.
- Debener, S., Thorne, J., Schneider, T.R., Viola, F.C., 2010. Using ica for the analysis of multi-channel eeg data. In: Ullsperger, M., Debener, S. (Eds.), *Simultaneous EEG and fMRI: Recording, Analysis, and Application*. Oxford Scholarship Online.
- Gross, J., Baillet, S., Barnes, G.R., Henson, R.N., Hillebrand, A., Jensen, O., Jerbi, K., Litvak, V., Maess, B., Oostenveld, R., Parkkonen, L., Taylor, J.R., van Wassenhove, V., Wibral, M., Schoffelen, J.M., 2013. Good practice for conducting and reporting MEG research. *NeuroImage* 65, 349–363.
- Herring, J.D., Thut, G., Jensen, O., Bergmann, T.O., 2015. Attention modulates TMS-locked alpha oscillations in the visual cortex. *J. Neurosci.* 35, 14435–14447.
- Huigen, E., Peper, A., Grimbergen, C.A., 2002. Investigation into the origin of the noise of surface electrodes. *Med. Biol. Eng. Comput.* 40, 332–338.
- Kappenman, E.S., Luck, S.J., 2010. The effects of electrode impedance on data quality and statistical significance in ERP recordings. *Psychophysiology* 47, 888–904.
- Keil, A., Debener, S., Gratton, G., Junghöfer, M., Kappenman, E.S., Luck, S.J., Luu, P., Miller, G.A., Yee, C.M., 2014. Committee report: publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology* 51, 1–21.
- Kothe, C., 2013. *The Artifact Subspace Reconstruction Method*. <http://sccn.ucsd.edu/eeglab/plugins/asr.pdf>.
- Krauledat, M., Dornhege, G., Blankertz, B., Müller, K.R., 2007. Robustifying EEG data analysis by removing outliers. *Chaos Complex. Lett.* 2, 259–274.
- Lawhern, V., Hairston, W.D., McDowell, K., Westerfield, M., Robbins, K., 2012. Detection and classification of subject-generated artifacts in EEG signals using autoregressive models. *J. Neurosci. Meth.* 208, 181–189.
- Lepage, K., Kramer, M., Chu, C., 2014. A statistically robust EEG re-referencing procedure to mitigate reference effect. *J. Neurosci. Meth.* 30, 101–116.
- Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L., 2013. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* 49, 764–766.
- Nunes, P., Srinivasan, R., 2006. *Electric Fields of the Brain: the Neurophysics of EEG*. Oxford University Press.
- Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* <https://doi.org/10.1155/2011/156869>.
- Oswal, A., Jha, A., Neal, S., Reid, A., Bradbury, D., Aston, P., Limousin, P., Foltynie, T., Zrinzo, L., Brown, P., Litvak, V., 2016. Analysis of simultaneous MEG and intracranial LFP recordings during Deep Brain Stimulation: a protocol and experimental validation. *J. Neurosci. Meth.* 261, 29–46.
- Parra, L.C., Spence, C.D., Gerson, A.D., Sajda, P., 2005. Recipes for the linear analysis of EEG. *NeuroImage* 28, 326–341.
- Pernet, C., Chauveau, N., Gaspar, C., Rousset, G., 2011. Limo eeg: a toolbox for hierarchical linear modeling of electroencephalographic data. *Comput. Intell. Neurosci.* <http://eprints.gla.ac.uk/56173/>.
- Perrin, F., Pernier, J., Bertrand, O., Echallier, J.F., 1989. Spherical splines for scalp potential and current density mapping. *Electroencephalogr. Clin. Neurophysiol.* 72, 184–187.
- Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson, R., Miller, G.A., Ritter, W., Ruchkin, D.S., Rugg, M.D., Taylor, M.J., 2000. Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology* 37, 127–152.
- Platte, R.B., Trefethen, L.N., Kuijlaars, A.B.J., 2011. Impossibility of fast stable approximation of analytic functions from equispaced samples. *SIAM Rev.* 53, 308–318.
- Shumway, R.H., Stoffer, D.S., 2011. *Time Series Analysis and its Applications*. Springer.
- Southwell, R., Baumann, A., Gal, C., Barascud, N., Friston, K., Chait, M., 2017. Is predictability salient? a study of attentional capture by auditory patterns. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372 <https://doi.org/10.1098/rstb.2016.0105>.
- Steiglitz, K., Mcbride, L., 1965. A technique for the identification of linear systems. *Automatic Control, IEEE Transactions on* 10, 461–464.
- Stjerna, S., Alatalo, P., Mäki, J., Vanhatalo, S., 2010. Evaluation of an easy, standardized and clinically practical method (SurePrep) for the preparation of electrode-skin contact in neurophysiological recordings. *Physiol. Meas.* 31, 889–901.
- Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M., 2011. Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.*
- Tanner, D., Morgan-Short, K., Luck, S.J., 2015. How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology* 52, 997–1009.
- Tanner, D., Norton, J.J., Morgan-Short, K., Luck, S.J., 2016. On high-pass filter artifacts (they're real) and baseline correction (it's a good idea) in ERP/ERMF analysis. *J. Neurosci. Meth.* 266, 166–170.
- Vanhatalo, S., Voipio, J., Kaila, K., 2005. Full-band EEG (FbEEG): an emerging standard in electroencephalography. *Clin. Neurophysiol.* 116, 1–8.
- Widmann, A., Schröger, E., Maess, B., 2015. Digital filter design for electrophysiological data - a practical approach. *J. Neurosci. Meth.* 250, 34–46.
- Winkler, I., Debener, S., Müller, K.R., Tangermann, M., 2015. On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2015-November:4101–4105*.
- Yao, D., 2001. A method to standardize a reference of scalp EEG recordings to a point at infinity. *Physiol. Meas.* 22, 693–711.
- Yeung, N., Bogacz, R., Holroyd, C.B., Cohen, J.D., 2004. Detection of synchronized oscillations in the electroencephalogram: an evaluation of methods. *Psychophysiology* 41, 822–832.
- Yuval-Greenberg, S., Tomer, O., Keren, A.S., Nelken, I., Deouell, L.Y., 2008. Transient induced gamma-band response in EEG as a manifestation of miniature saccades. *Neuron* 58, 429–441.