

Reducing the Number of MEG/EEG Trials Needed for the Estimation of Brain Evoked Responses: A Bootstrap Approach

Cezary Sielużycki , Artur Matysiak , Reinhard König ,
and D. Robert Iskander , *Senior Member, IEEE*

Abstract—Objective: A common problem in magnetoencephalographic (MEG) and electroencephalographic (EEG) experimental paradigms relying on the estimation of brain evoked responses is the lengthy time of the experiment, which stems from the need to acquire a large number of repeated recordings. Using a bootstrap approach, we aim at reliably reducing the number of these repeated trials. **Methods:** To this end, we assessed five variants of non-parametric bootstrapping based on the classical signal-plus-noise model constituting the foundation of signal averaging in MEG/EEG. We explain which of these approaches should and which should not be used for the aforementioned purpose, and why. **Results:** We present results for two advocated bootstrap variants applied to auditory MEG data. The ensuing trial-averaged magnetic fields served as input to the estimation of cortical source generators, with spatio-temporal matching pursuit as an example of an inverse solution technique. We propose, for a wide range of trial numbers, a general framework to evaluate the statistical properties of the parameter estimates for source locations and related time courses. **Conclusion:** The proposed bootstrap framework offers a systematic approach to reduce the number of trials required to estimate the evoked response. The general validity of our findings is neither bound to any particular type of MEG/EEG data nor to any specific source localization method. **Significance:** Practical implications of this work relate to the optimization of acquisition time of MEG/EEG experiments, thus reducing stress for the subjects (especially for patients) and minimizing related artifacts.

Manuscript received July 1, 2020; revised November 19, 2020 and January 14, 2021; accepted February 14, 2021. Date of publication February 19, 2021; date of current version June 18, 2021. The work of Cezary Sielużycki has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant 665778, National Science Centre, Poland, POLONEZ 2 Programme, 2016/21/P/ST7/03929. The work of D. Robert Iskander was supported by the statutory funds of the Department of Biomedical Engineering at the Wrocław University of Science and Technology. This work was supported by the DAAD Personnel Exchange Programme between Germany and Poland, 2018–2019, 57393544. (Cezary Sielużycki and Artur Matysiak contributed equally to this work.) (Corresponding author: Cezary Sielużycki.)

Cezary Sielużycki and D. Robert Iskander are with the Department of Biomedical Engineering, Wrocław University of Science and Technology, 50-370 Wrocław, Poland (e-mail: cezary.sieluzycski@pwr.edu.pl).

Artur Matysiak and Reinhard König are with the Research Group Comparative Neuroscience, Leibniz Institute for Neurobiology.

This article has supplementary downloadable material available at <https://doi.org/10.1109/TBME.2021.3060495>, provided by the authors.

Digital Object Identifier 10.1109/TBME.2021.3060495

Index Terms—Bootstrap, EEG, MEG, parameter estimation, signal averaging.

I. INTRODUCTION

MAGNETOENCEPHALOGRAPHY (MEG) [1]–[3] and electroencephalography (EEG) [3], [4] are non-invasive techniques for studying the activity of the human brain with high temporal resolution of the order of milliseconds. They measure magnetic (MEG) or electric (EEG) fields stemming from neuronal activity of the brain which can be described in terms of electric currents flowing in the cortex and deeper regions. In MEG/EEG studies of event-related brain activity, the subjects are exposed to some sort of stimulus such as acoustic, visual, or somatosensory. Due to the minute amplitude of the brain response evoked by a single stimulus, however, the signal-to-noise ratio (SNR) of a single-trial MEG or EEG recording is poor compared to the background activity, which encompasses all other processes not time or phase locked to the external stimulus. To increase the SNR, the conventional approach is to use a large number of stimulus repetitions, of the order of 100 or so, in a typical experiment. In the conventional analysis of stimulus-evoked brain activities, the single-trial responses are then arithmetically averaged time-locked to stimulus onset.

The classical approach underlying signal averaging, commonly referred to as the signal-plus-noise (SPN) model and introduced in the seminal work of Dawson [5] in 1954, considers the evoked response to a series of stimulus repetitions as a deterministic phenomenon, whereas the background activity is treated as a zero-mean stochastic process referred to as noise. For a discrete-time measured signal s of length N in trial t and channel c , this can be expressed as

$$s_{t,c,n} = m_{c,n} + \varepsilon_{t,c,n}, \quad (1)$$

where m denotes the trial-independent evoked response, ε is the noise, and $n \in \{0, 1, \dots, N-1\}$ is a discrete time moment. Note that m in Eq. (1) is not indexed by trials t because of its assumed deterministic nature. This trial-independent character of m is, of course, a simplification (see, e.g., [6], [7] and references therein), yet a useful one, as it makes the estimation of m easier.

Following Eq. (1), for a Gaussian distribution of ε across trials, arithmetic averaging of single-trial signals s will make

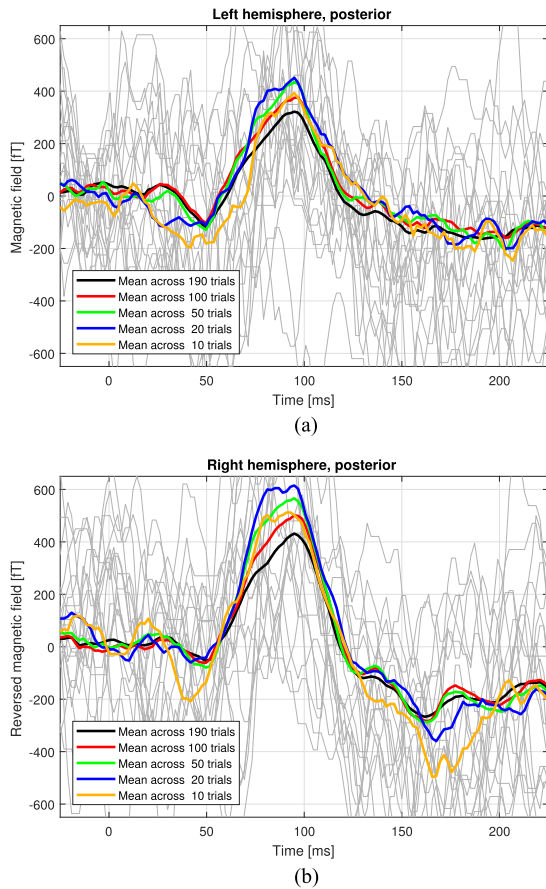


Fig. 1. Estimates of the auditory M100 wave derived by arithmetic averaging of single-trial recordings from the two posterior MEG channels with the largest absolute M100-peak magnitude over the left (a) and the right (b) hemisphere. The noisy single-trial time courses are shown in light gray; for clarity, only every 10th of the 190 artifact-free trials is shown. The colored lines are waveforms averaged across different numbers of trials, as indicated in the figure legend. For all averaged waveforms, the M100-peak magnitude is larger in the right hemisphere than in the left. Interestingly, the 190-trials waveform has the smallest M100-peak magnitude. Note that the magnetic fields recorded with posterior channels above the left and right hemisphere, respectively, have opposite polarity. To facilitate the comparison between the waveforms of the two channels, the time series of the channel above the right hemisphere have been amplitude reversed.

ε vanish proportionally to the square root of the number T of averaged trials and thus provide an estimator \hat{m} of m . This is visualized in Fig. 1 for an illustrative MEG data set from an auditory MEG experiment (see Section II-F). Data are shown for the two channels with the largest absolute magnetic field value above the left (a) and the right (b) hemisphere, respectively. The thin gray lines represent single-trial recordings, whereas the five thicker traces depict arithmetic means of different numbers of single trials, viz. for 10, 20, 50, 100, and 190 trials. The averaged waveforms have a morphology typical for an auditory MEG/EEG response using pure-tone stimulation, with a dominant large-amplitude deflection, the so-called M100/N100 response, occurring approximately 100 ms after stimulus onset [7]–[11]. Note, however, that averaging across different numbers of trials results in waveforms with different shapes and M100-peak amplitudes, and that even averaging only 10 single trials

leads to a prominent M100-peak amplitude, as it suppresses the background noise, and so increases SNR.

The motivation for this work goes back to the circumstance that MEG/EEG researchers oftentimes do not have the comfort of recording a large number of trials, especially when the experimental paradigm is complex. This is, for example, the case when the scientific question to be investigated requires that one or more stimulation parameters are altered in the course of an experimental session. It may then be necessary to substantially reduce the number of trials for each given set of parameters. Otherwise, the total examination time would be too long for the subject or patient, who might then introduce undesirable confounds, which, in turn, may deteriorate the quality of acquired data. Examples of those confounds include muscular tension, eye blinks, pronounced alpha rhythm due to sleepiness, involuntary head movements, etc. To avoid artifacts stemming from these disturbances, the total examination time should be kept reasonably short, which means that the number of trials for a given set of stimulus parameters cannot be large. Thus, in the majority of MEG/EEG studies, the experimenter has to trade off a large number of trials to obtain a high SNR against a smaller, suboptimal number of trials to minimize the burden for the subject.

Another important aspect to consider is the so-called habituation effect, which manifests itself in a diminishing evoked-response peak amplitude, and possibly also in altering the related peak latency, with consecutive stimuli. This may, paradoxically, result in a smaller SNR of the trial-averaged response for a large set of trials compared to a smaller set. Such a scenario is possibly present in Fig. 1, which shows—for the two hemispheres—larger peak amplitudes with decreasing number of trials used for averaging from 190 to 20. It is, of course, difficult to judge whether the blue trace, reflecting the mean from the first 20 trials, is indeed the optimal estimate of the response. The somewhat rough morphology of the large deflection of the two blue traces at about 90 ms reflecting the auditory M100 wave possibly indicates contributions from noise rather than the signal of interest. Therefore, estimating an optimal (in some sense) number of trials for signal averaging is not trivial. It is not necessarily the case that the more trials the better. In consequence, it is difficult to propose a pseudo-ground-truth estimate which could further be used for assessing the bias or strictly speaking—given the lack of a true ground truth—the difference between estimates derived from different numbers of trials and those obtained for such a pseudo-reference. Interestingly, although the literature provides some recommendations for the number of trials needed in particular types of MEG/EEG experiments (see, e.g., [1]–[4], [12]–[14] and references therein), to the best of our knowledge there is a gap in the literature when it comes to evaluating that number systematically.

A classical probabilistic approach in search of an optimum number of trials in an MEG/EEG experiment would be to repeat the examination many times and derive measures of dispersion for the parameters in question from those numerous examinations. For trial-averaged signals, where, for each examination, the mean signal would be derived from trial sets of different sample sizes, the dispersion for a given number of trials

could be assessed on the basis of the multiple repetitions of an examination. However, such an approach would be extremely costly both in terms of time and money, because performing a measurement, say, 1000 times would go well beyond available resources. To overcome the immense impracticality of such a theoretical framework, we propose to employ non-parametric bootstrap techniques to simulate multiple examinations. This allows estimating the dispersion of the studied parameters such as source locations [15]–[17] and characteristics of related time courses.

All these aspects create a complex trade-off scenario, putting the experimenter into a quandary. In consequence, the aim of this work is to provide a framework for reducing the number of trials needed for reliable estimates of key waveform parameters such as peak amplitudes and latencies [14], [18], [19]. In studies of evoked activity these parameters are usually not limited to signals observed on the sensor level; rather, they often concern the spatial location of current sources in the brain and their corresponding time courses. To infer the locations of intracranial current sources from the magnetic/electric field recorded extracranially with MEG/EEG, one must solve the so-called inverse problem (see, e.g., [3], [4], [20] and references therein). Source localization strategies differ in their performance with respect to, for example, the noise level of the recorded magnetic/electric fields and the number of underlying current sources, and, consequently, their outcomes depend on the number of trials contributing to the averaged magnetic/electric field. Since the focus of this work was on reducing the number of trials used for averaging, we abstained from evaluating the impact of different source localization methods on the results, and have confined ourselves to using the spatio-temporal matching pursuit (STMP) [21] as a suitable example of the numerous source-localization approaches instead. Further, we do not provide an absolute answer regarding the number of trials, but propose a framework for the statistical evaluation aimed at obtaining a reliable and cost-effective solution. This framework can be applied to any signal processing pipeline using different source-localization strategies, in the same vein as sample size estimation methods can be used in a variety of statistical inference problems.

II. METHODS

This section comprises five different bootstrap approaches, four already existing ones and a novel one. We briefly introduce the approaches and provide recommendations as to their applicability to reduce the number of trials needed for the estimation of brain evoked responses by means of arithmetic signal averaging in MEG/EEG based on the classical SPN model.

Bootstrap is a technique of random sampling from available data [22]–[24], even though different kinds of random sampling used in the literature are referred to as bootstrap. Various authors use the term to name either random sampling with replacement [24] (non-parametric bootstrap), random sampling without replacement [25] (for example, jackknife), or even sampling from a given distribution [26] (parametric bootstrap).

However, strictly speaking, bootstrap denotes drawing with replacement [23], and this is the way we use it in this work.

Multiple samplings with bootstrap are meant to strengthen the evaluation power. This is so because small sample sizes (here, the sizes of sets of trials) raise concerns about the potentially poor quality of the estimates of the parameters of interest (here, source locations and corresponding waveform parameters), even if the available sample is representative for the true distribution. In particular, for a small sample size, these estimates cannot be derived with meaningful confidence intervals unless asymptotically derived assumptions regarding the distribution of the estimator, such as the invocation of the Central Limit Theorem, are made. Bootstrap techniques are known to be a solution to this problem in that they enable assessing an approximation of the true distribution with sufficient resolution.

In Section II-A to II-D, we review several existing types of bootstrapping and discuss which of them are applicable in the given framework and which of them should not be used. Next, in Section II-E, we introduce a new and alternative bootstrapping approach that can be used for MEG/EEG time series.

A. Bootstrapping in Time Domain

The classical and most commonly applied bootstrapping technique relies on the assumption that consecutive samples of a signal in time are independent and identically distributed (IID). However, since consecutive discrete-time samples $s_{t,c,n}$ of s from Eq. (1) are not IID in any trial t or channel c , one might hope to obtain an IID residuum r by subtracting the estimated model \hat{m} —the arithmetic average over trials—of m from s :

$$r_{t,c,n} = s_{t,c,n} - \hat{m}_{c,n}. \quad (2)$$

Then, r would be resampled with replacement along its third dimension in order to create a new residual r^* and a new realization s^* by adding r^* to the model \hat{m} [23]:

$$s_{t,c,n}^* = \hat{m}_{c,n} + r_{t,c,n}^* = \hat{m}_{c,n} + r_{t,c,n}^*. \quad (3)$$

However, in applications such as estimating stimulus-evoked responses in MEG/EEG, this approach is problematic because time-domain residuals—computed by subtracting a model of an evoked magnetic/electric field from the measured raw data—are not IID, as they contain rhythmic background activity. This is because the evoked response in the classical view expressed by Eq. (1) is independent of the rhythmic background activity. Hence, when the model solely addresses the evoked response—which is justified when one aims at modeling the response—the residuals will naturally contain the rhythmic background activity. This rhythmic activity takes form of the so-called delta (0.5–3 Hz), theta (4–7 Hz), alpha (8–13 Hz), mu (7.5–12.5 Hz), sensorimotor (12.5–15.5 Hz), beta (16–31 Hz), and gamma (32–100 Hz) waves. Each of the spontaneous oscillations is a quasi-periodic process with a non-zero autocorrelation in time and, hence, it cannot be considered an IID process. Resampling these spontaneous rhythms would spoil their intrinsic autocorrelation. In consequence, newly created bootstrap realizations would not be realistic. Therefore, we do not recommend bootstrapping

time samples within trials in studies of evoked activity with MEG/EEG.

B. Bootstrapping in Frequency Domain

To overcome this problem, one might be tempted to bootstrap in the frequency domain [27] rather than in the time domain. Here, the idea is to bootstrap the complex coefficients of the frequency spectrum obtained from the discrete Fourier transform of r . A new realization s^* is then created by adding the new residuum r^* , computed with the inverse Fourier transform of the bootstrapped frequency spectrum of r , to the model \hat{m} . However, this approach, too, would destroy autocorrelation in time because ε in Eq. (1) or r in Eq. (2) obey approximately a $1/f$ distribution [28]. Moreover, given this $1/f$ characteristic of the background activity, it is implausible to assume that all frequencies are IID. Thus, this way of bootstrapping for analyses of evoked activity acquired with MEG/EEG is not recommended either.

C. Bootstrapping in Phase Spectrum

With the above problems in mind, one might advocate bootstrapping in the domain of the phase spectrum [29], [30]. To adhere to the fundamental assumption of bootstrap, in this approach one would first need to investigate the residuals for being IID in the domain of the phase spectrum. To do so, one could subtract the mean across trials—which would serve as the model—from each trial and examine the resulting residuals for their potential of being considered IID in the domain of phase spectrum.

The algorithm for bootstrapping the phase spectrum of a single-trial, single-channel residual $r_{t,c,:}$ from Eq. (2) starts with calculating its spectrum $R_{t,c,:}$ using the discrete Fourier transform¹

$$R_{t,c,k} = \sum_{n=0}^{N-1} r_{t,c,n} e^{-\frac{2\pi jkn}{N}}, \quad k \in \{0, 1, \dots, N-1\}, \quad (4)$$

with its amplitude spectrum coefficients $A_{t,c,k} = |R_{t,c,k}|$ and the related phases

$$\Phi_{t,c,k} = \angle(R_{t,c,k}) \in \{\Phi_{t,c,0}, \Phi_{t,c,1}, \dots, \Phi_{t,c,N-1}\}, \quad (5)$$

where k denotes consecutive discrete frequencies.

The scheme considers bootstrapping $\Phi_{t,c,k}$ to obtain Φ_{t,c,k^*} in a manner that preserves the antisymmetry of the phase spectrum. For this, the first element of the phase spectrum, $\Phi_{t,c,0}$, is retained, whereas the elements $\Phi_{t,c,1}, \Phi_{t,c,2}, \dots, \Phi_{t,c, \lfloor \frac{N-1}{2} \rfloor + \delta}$ are resampled with replacement using a pseudorandom number generator, resulting in a set of bootstrapped phase spectrum estimates of the form

$$\begin{aligned} \Phi_{t,c,k^*} \in \{ & \Phi_{t,c,0}, \Phi_{t,c,1^*}, \Phi_{t,c,2^*}, \dots, \Phi_{t,c,(\lfloor \frac{N-1}{2} \rfloor + \delta)^*}, \\ & -\Phi_{t,c,(\lfloor \frac{N-1}{2} \rfloor)^*}, \dots, -\Phi_{t,c,2^*}, -\Phi_{t,c,1^*} \}, \end{aligned} \quad (6)$$

¹The subscript $:$ in $r_{t,c,:}$ denotes all discrete-time instants, hence $r_{t,c,:}$ is the c th row in the spatio-temporal $C \times N$ matrix for a trial t with C being the number of MEG/EEG channels and N the number of discrete-time instants, whereas in $R_{t,c,:}$ it denotes all discrete frequencies obtained from the Fourier transform.

where

$$\delta = \begin{cases} 0 & \text{for } N \text{ odd,} \\ 1 & \text{for } N \text{ even.} \end{cases} \quad (7)$$

The commonly used bootstrap asterisk notation in the subscripts indicates here that bootstrapping is performed in the third dimension of frequency samples. The bootstrapped phase spectral estimates Φ_{t,c,k^*} are then combined with the amplitude counterparts $A_{t,c,k}$ to form the complex bootstrap spectral estimates of the form

$$R_{t,c,k^*} = A_{t,c,k} e^{j\Phi_{t,c,k^*}}. \quad (8)$$

Finally, the discrete inverse Fourier transform is used to obtain the bootstrapped residual, i.e.,

$$r_{t,c,n}^* = \frac{1}{N} \sum_{k=0}^{N-1} R_{t,c,k^*} e^{\frac{2\pi jkn}{N}}, \quad n \in \{0, 1, \dots, N-1\}. \quad (9)$$

To form a new bootstrapped realization of the signal s , the bootstrapped residual is added to the earlier estimated model \hat{m} , i.e., $s^* = \hat{m} + r^*$.

The procedure is repeated B times, to facilitate estimating the characteristics of the signal model such as its mean, variance, and confidence intervals, without the need for repeating the experiment and collecting multiple realizations of the signal s .

Since an established statistical test for assessing if the process in question is IID is lacking, we propose to check whether the process—here, the series of phases—is:

- 1) independent, using, for example, the Wald–Wolfowitz runs test for independence,
- 2) stationary in mean, using, for example, the Kwiatkowski–Phillips–Schmidt–Shin test for trend-stationarity,
- 3) stationary in variance, using, for example, the White test for homoscedasticity.

When there are no reasons to reject any of the null hypotheses H_0 of the aforementioned tests, one could assume that the residual for a given trial is an IID process in the domain of phase spectrum. When applying this routine to single-sensor MEG data from the auditory experiment considered in this paper (see Section II-F), where we used the trial-averaged waveform as the model, H_0 s were not rejected for the residuals in the domain of phase spectrum in the vast majority of cases. With this approach, the $1/f$ frequency characteristics of the residuals is preserved. The exact pattern of autocorrelation will change, as the phase relations between various frequencies will be altered, but such a variability in autocorrelation of the residuals is expected anyway, whenever multiple realizations of the signal are acquired.

So far, we only considered bootstrapping either phases or complex amplitudes within a trial for a single MEG/EEG channel. The application of this approach to more than one MEG/EEG channel, but independently for each channel, would destroy the spatial covariance pattern of the residuals, i.e., the noise, across channels. One could therefore bootstrap a residual in all channels identically by keeping the phase of a given frequency constant across all channels. This would mean that, at least for simple periodic activity, there are either several generators of that frequency that are perfectly synchronized in phase or there is only one frequency-specific generator. However, both

scenarios are rather unlikely to occur in the real brain. Thus, in order to address the spatial characteristics of noise correlation in MEG/EEG, we advocate the two approaches presented in the following two subsections, where we consider all MEG/EEG channels simultaneously.

D. Bootstrapping Trial Labels (BTL)

When dealing with estimating current source locations in MEG/EEG, information from multiple channels needs to be taken into account. Instead of bootstrapping a multi-channel residual r within a single trial one can bootstrap single-trial MEG/EEG data across trials, without changing the internal structure of the single-trial recordings. This technique is sometimes referred to as the block bootstrap [24]. In such an approach, one only draws the numeric labels of trials. Using the notation from the left-hand side of Eq. (1), this can be formalized as

$$s_{t,c,n}^* = s_{t^*,c,n}. \quad (10)$$

This means that for each channel c and time n in the bootstrap trial, we choose $s_{c,n}$ by sampling from $\{s_{1,c,n}, \dots, s_{T,c,n}\}$, where T denotes the number of all trials.

Noise realizations from different trials can safely be considered IID because such realizations are well separated from each other in time, which justifies the assumption that they are mutually independent. Hence, as long as one adheres to the SPN model of Eq. (1), there is no need to search for a bootstrap model $\hat{m}_{c,n}$ according to Eq. (2) because subtracting a process supposed to be deterministic across trials is redundant. This is because bootstrapping entire trials is equivalent to bootstrapping residuals created by subtracting a trial-independent representation of the evoked response before bootstrapping and adding it later to each trial after bootstrapping. This is an important advantage that frees one from seeking a suitable model for bootstrapping. Given these qualities, this is certainly a recommendable approach, and it has been used by Darvas *et al.* [16] to assess the accuracy of source localization in MEG. However, they did not investigate how many trials are needed to obtain a stable localization.

E. Bootstrapping Complex Amplitudes (BCA) Across Trials

Here, we propose a new type of bootstrapping applicable across trials rather than within a trial. In this approach, the complex amplitudes, which encode the modulus and phase, of Fourier transformed single-trial recordings are bootstrapped across trials such that the exchanging (sampling with replacement) is done independently for each given frequency. Yet, the exchange for a given frequency is realized identically in all MEG/EEG channels, that is, trial labels are not bootstrapped across channels.

Since bootstrapping in this approach provides a different trial-to-trial variation profile of complex amplitudes for each frequency, it is very unlikely that a trial recomposed via the inverse Fourier transform will be the same as any of the original trials. Details on this likelihood are given in Sect. I of Supplementary Material (SM). In consequence, BCA is practically completely

unaffected by the likelihood of repeating a single trial, which, however, may be different in the BTL approach presented in Section II-D. This helps avoiding a diminished SNR, which may happen in BTL when a particular trial label is randomly selected a couple of times before computing the recomposed mean signal, because averaging a few instances of the same trial does not reduce noise in the mean signal. Since the bootstrapping in BCA is realized for each frequency independently, each frequency component is given a new and individual profile of complex-amplitude variation across trials. As a result, the new complete (that is, recomposed of all the frequency components) mean signal should generally have better SNR, compared to that of the mean signal obtained from BTL.

In analogy to Eq. (8) and Eq. (9), this method can be formalized as

$$s_{t,c,n}^* = \frac{1}{N} \sum_{k=0}^{N-1} S_{t,c,k}^* e^{\frac{2\pi jkn}{N}}, \quad n \in \{0, 1, \dots, N-1\}, \quad (11)$$

where

$$S_{t,c,k}^* = A_{t^*,c,k} e^{j\Phi_{t^*,c,k}}, \quad (12)$$

with the original $S_{t,c,:}$ being the spectrum of $s_{t,c,:}$ from Eq. (1) obtained with the discrete Fourier transform, and each t^* being identical in $A_{t^*,c,k}$ and the related $\Phi_{t^*,c,k}$.

Importantly, this approach preserves the true nature of the spatial distribution of the phase range of a given frequency component. Note that in real multichannel MEG/EEG signals the spatial signature of a low-frequency component is characterized by a limited range of phases across channels, whereas a high-frequency component can vary in its phase more freely across space. In the BCA approach, what is changed for a given frequency component is its trial-to-trial variation profile of complex amplitudes. The complex amplitude does, of course, contain information on the phase, but since each frequency component is bootstrapped across trials, not across channels, the phase range of a given frequency phasor is preserved across space. In other words, when one considers the unit circle in a complex plane as the model of phase presentation, a low frequency in a given trial would occupy only a limited range of angles across MEG/EEG channels, whereas a high frequency would cover a fairly wide range of angles.

A natural outcome of this approach is that if the original trials carried some cross-frequency phase-amplitude coupling (see, e.g., [31], [32], and references therein), it would likely be dismantled due to bootstrapping the complex amplitudes independently for each frequency. This is because the original pattern of coupling between, say, the theta and the gamma rhythm would be altered, since after BCA each of the frequencies contributing to these rhythms would have a new arrangement of complex amplitudes across trials. However, as long as one adheres to the SPN model in search for the deterministic estimate of an evoked response, this outcome is irrelevant.

Insomuch as in the approach from Section II-D, one needs no model to be subtracted from the single-trial recordings before bootstrapping, which, here, stems from the linearity of the Fourier transform and of its inverse transform.

F. Illustrative MEG Dataset

We implemented the algorithms² described in Sections II-D and II-E in MATLAB (MathWorks, Natick, MA, USA) and applied them to an MEG dataset taken from an auditory experiment where a healthy male subject was exposed to simple acoustic stimuli. This dataset is well characterized and was used in different contexts in previous works [7], [10], [21], [33]. Details of the dataset are provided in Sect. II of SM. The study was approved by the Ethics Committee of the Otto von Guericke University Magdeburg.

For the current analysis, we first extracted $[-25; 226]$ -ms time windows from each trial, where 0 ms corresponds to stimulus onset. Next, this time window was downsampled by a factor of 2 resulting in 128 samples, with the time resolution of about 2 ms. This time window covered the most prominent feature of the evoked response in simple auditory MEG experiments, the M100 response (see Section I).

The five lines with different colors in Fig. 1 represent M100 waveforms obtained from averaging across five different numbers of single trials, for the two selected channels. The figure indicates that averaging across different numbers of trials leads to substantially different waveforms and M100 peak amplitudes of the averaged signal, and, thus, raises the question which of the mean waveforms is the “optimal” one. The averaged M100 waveforms are essentially characterized by two parameters: the peak amplitude and the corresponding peak latency. Averaging across the entire set of 190 trials (black trace) yielded an amplitude of 322 fT for the left (Fig. 1(a)) and 432 fT for the right hemisphere (Fig. 1(b)); the respective peak latency was 95 ms for both the left and the right hemisphere. This hemispheric asymmetry is a well-known effect in auditory MEG research (see, for example, [34], [35]). Shaw *et al.* [36] attributed this rightward bias of the auditory MEG signal to asymmetries in the underlying anatomy, i.e., to the local geometry of the highly convoluted cortex which results in an uneven MEG signal cancellation in the two hemispheres. In the subsequent analyses we will concentrate on the peak amplitude and latency for the estimated time courses of the current sources rather than for time courses of the magnetic fields recorded with a single MEG sensor. Note that source localization techniques take into account signals from essentially all MEG channels as input.

G. Illustrative Source Localization Approach: Spatio-Temporal Matching Pursuit

The application of bootstrap proposed for reducing the number of trials in MEG/EEG signal averaging is independent of a particular source localization approach. For this reason, we focus on one illustrative strategy, the spatio-temporal matching pursuit (STMP) [21]. Since we address both spatial and temporal aspects of the evoked response in this work, STMP seemed a natural choice because of its *simultaneous* rather than sequential approach to space and time [21]. In STMP, cortical locations of the current sources generating the measured magnetic fields and the corresponding time courses are found by maximizing

the dot product of the recorded multi-channel MEG data with a spatio-temporal dictionary composed of a spatial and a temporal part.

The spatial part consists of a large set of leadfield matrices (605 746 in the current implementation) and links source and channel locations using quasistatic approximations of the Maxwell equations, one for each possible location in the cortex. Source locations considered are defined by the vertices of the tessellated cortical surface [37]. Each of these vertices corresponds to a leadfield that geometrically links the spatial location of the vertex with the spatial locations of the MEG sensors. The forward solutions were computed using a realistic head model with the spherical-harmonics approximation [38] of the subject’s magnetic-resonance (MR) scan, as implemented in the FieldTrip toolbox [39].

The temporal part comprises a large set of complex-valued chirplets (chirp functions modulated in amplitude by Gaussian envelopes of different scales). These are meant to parametrize the time courses of active cortical sources. Due to the linearity of the quasi-static approximation, the morphology of each temporal atom is identical for the sensor and the source space. The temporal dictionary encompassed 31 409 chirplets.

Both spatial and temporal atoms are normalized in the respective dictionaries such that their ℓ^2 norms are equal to 1. This gives all atoms equal chances in the process of estimating the spatio-temporal activity of the brain.

STMP is an iterative procedure, in which every iteration provides a solution for the location of an individual cortical source, along with the time course of that source. The source locations are returned as elements of the spatial dictionary, from which one can easily derive their respective x , y , and z coordinates in space. The related time courses are returned as elements of the temporal dictionary, that is, the underlying chirplet functions used to estimate the time traces of the source activities. Details of the implementation of the STMP algorithm can be found in [21].

We modeled the neural sources underlying the magnetic field distribution of the M100 waveform by two equivalent current dipoles (ECDs), one in each hemisphere. Spatial locations of the two ECDs and their respective time courses were obtained in two STMP iterations. Thus, in one iteration, STMP resulted in a single $[x, y, z]$ location in space and the related time course. The latter we further examined in terms of its M100-peak amplitude and latency.

III. RESULTS

A. Source Localization Results

The first and second column of Fig. 2 show $[x, y, z]$ source locations obtained with the BTL bootstrapping strategy (Section II-D), whereas the third and fourth column of this figure show $[x, y, z]$ source locations based on the BCA approach (Section II-E). Further, the results are pairwise broken down into rows, one for the right (RH) and one for the left (LH) hemisphere; this order—first RH, then LH—reflects the fact that STMP addresses the stronger response first (which here occurred in the right hemisphere; see Fig. 1), as it is driven by an energetic criterion [21]. Moreover, Fig. 2 shows source locations for three

²The code is available at <https://github.com/CezarySieluzycycki/TBME2021>.

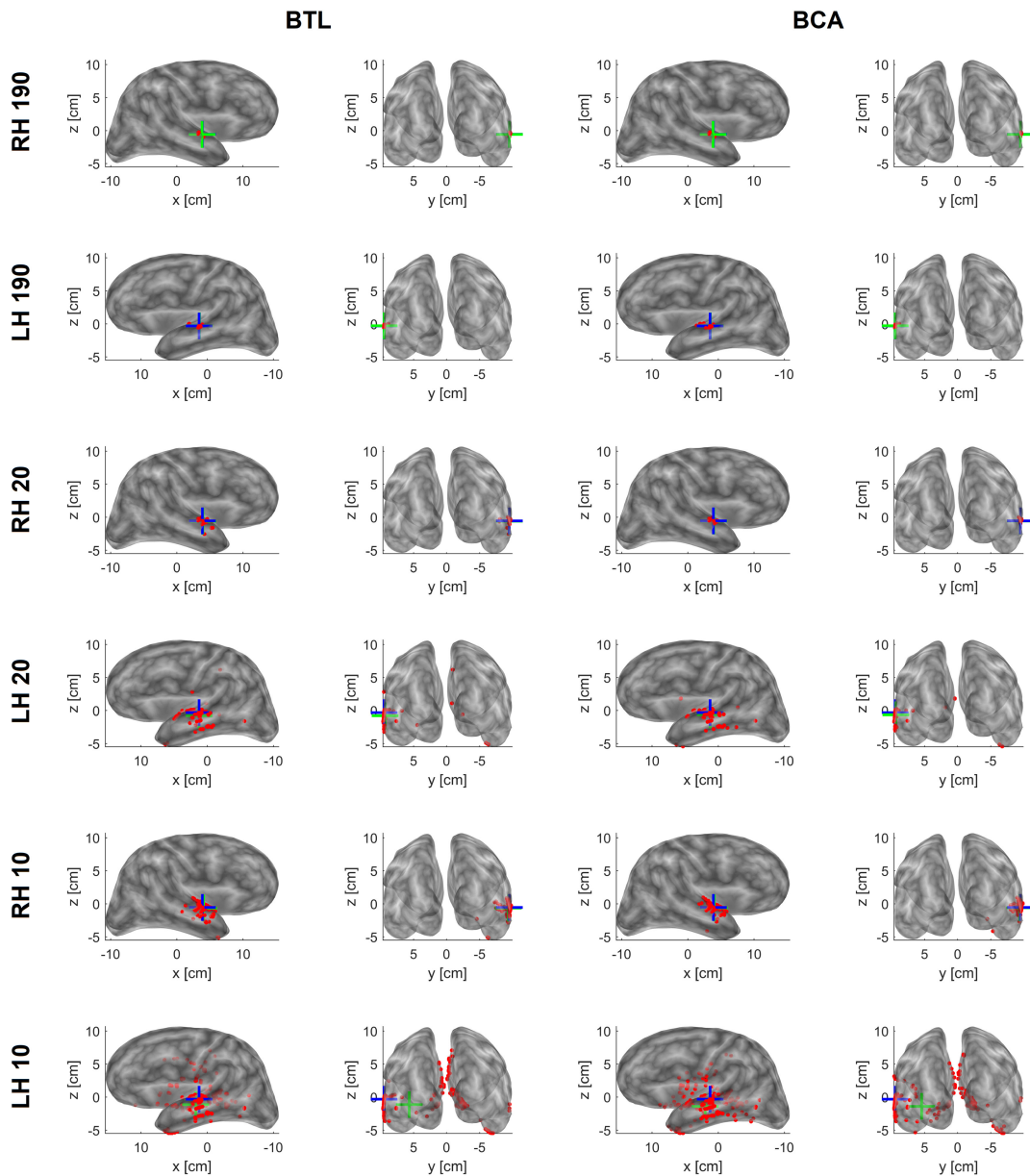


Fig. 2. Source locations (red dots) found for 1000 realizations of BTL (columns 1, 2) and for 1000 realizations of BCA (columns 3, 4). In the panels of the first two rows, all 190 trials were used for signal averaging; in the panels of the two middle rows only the first 20, and in the two bottom rows only the first 10 trials were taken into account. LH and RH denote left hemisphere and right hemisphere, respectively. Semitransparent inflated brains, with sagittal views of the left and right hemisphere (columns 1, 3), are shown for a better display of all source locations, along with an additional coronal view from the back of the brain (columns 2, 4). For each hemisphere, green crosses represent the mean locations across the red-dot locations of the individual solutions. Blue crosses mark, again for each hemisphere, the source locations obtained from the original set of 190 trials, i.e., without bootstrapping. Uncompressed images are available in the supplemental material (Fig. S8).

different mean signals, that is, for 190 (two top rows), 20 (two rows in the middle), and 10 trials (two bottom rows) averaged. In each individual panel, 1000 source locations are depicted as red dots, representing 1000 bootstrap solutions. The mean location of these 1000 solutions is indicated by the green cross, and the blue cross represents the inverse solution calculated from the original set of 190 trials, i.e., without bootstrapping.

Since the SNR of the MEG signal averaged across the original 190 trials is large, we observed that, for each of the two hemispheres and the two bootstrapping approaches, the individual source locations of the 1000 bootstrapping estimates were nearly

identical to each other, i.e., their spatial dispersion was very small (see panels in the two upper rows of Fig. 2). Further, in both the BTL and the BCA approach, the spatial distance between the green and the blue cross was marginal, i.e., the reproducibility of the bootstrapped source locations almost reached the quality of the solution with the original 190 trials. However, this picture changed when we moved to average signals derived from much smaller numbers of trials. Averaging across the first 10 trials resulted in a wide spread of source locations, as shown in the two bottom rows of Fig. 2. Hence, for both the BTL and the BCA approach, 10 trials were not sufficient to provide robust and

reliable estimates of source locations. Increasing the number of trials used for signal averaging from 10 to 20 (two middle rows of Fig. 2) already provides a very different picture, similarly for both BTL and BCA. Here, the spread of the individual solutions is clearly smaller than in the 10-trials case. Note also that the source locations for the average signals derived from 10 and 20 bootstrapped trials have a noticeably smaller spatial spread in the right hemisphere than those in the left hemisphere, for which the recorded MEG signal is weaker (see Fig. 1).

B. Systematic Analyses

To assess the dependence of the source locations and waveform parameters on the number of trials used for signal averaging systematically, we ran analyses for 19 different sizes of trial sets, starting from the first 10 trials, through the first 20, 30, and so on, up to all 190 trials. The statistics of the $[x, y, z]$ coordinates as well as of the 3D Euclidean distance d of the individual estimates of the $[x, y, z]$ source locations from the mean source locations are shown as box plots in Figs. S1–S4 of SM. The findings for the absolute M100-peak amplitudes of the source waveforms are displayed in Fig. S5 of SM, and those for the corresponding M100-peak latencies in Fig. S6 of SM. These six figures show results for the two sources and for all 19 sets of trials obtained with BTL (the two panels in the left columns of the figures) and BCA (the two panels in the right columns). Detailed description is given in SM.

It is evident from Figs. S1–S6 in SM, and in particular from the box plots of the absolute peak amplitude in Fig. S5, that with increasing numbers of trials taken for computing the average signal the parameter estimates tend to have smaller dispersion.

To provide more insight into the behavior of the estimated distributions of the parameters, we show in Fig. S7 of SM—using a Gaussian kernel of a standard width—the estimated probability density functions of the absolute peak amplitude of the source waveform for the two sources and for each set of trials obtained using BTL and BCA. We observed that the estimated mean and variance of the absolute M100 peak amplitude of the source in the right hemisphere decrease with increasing number of trials, for both BTL and BCA. This behavior is less pronounced for the source in the left hemisphere, where some bimodal character of estimated PDFs is visible. This observation most likely stems from a poorer SNR for this source, compared to the source in the right hemisphere. Note that knowledge about such a behavior of MEG parameter estimates is inaccessible without performing bootstrapping or extensively repeated experiments (Monte Carlo).

C. Dispersion of Source Location and Source Waveform Parameters

We already argued in Section I that any selection of “ground-truth” references for assessing the bias of these results would be somewhat arbitrary. Hence, we purposely abstained from evaluating such pseudo-biases and restricted ourselves to examining dispersions. Fig. 3 summarizes estimated dispersions of the various parameters of source-level measures. This figure shows the dependence of the standard deviation (SD) on the

number of trials used for signal averaging. Throughout the six panels, red symbols refer to the source in the right hemisphere, and blue symbols to the source in the left hemisphere. Circles reflect the BTL, and crosses the BCA approach. In addition, power law functions were fitted to the data.

Figs. 3(a)–(c) show SDs of source locations split into their $[x, y, z]$ components, respectively, and Fig. 3(d) shows SDs of 3D distances of the 1000 3D source locations from the mean 3D locations in the Euclidean space. We found for these two focal and spatially distinct sources that—as to their cortical location—already 60 trials were sufficient to achieve a solution with a small dispersion in space that lies well below the limits of the spatial resolution of MEG of the order of a few millimeters [40]–[44]. Notably, this threshold, which is indicated in Figs. 3(a)–(d) by the green vertical dash-dotted line, holds for all three directions of the coordinate system. For even smaller numbers of trials (below 50), SD strongly increased for the source in the left hemisphere, whereas for the source in the right hemisphere it remained constant even down to 20 trials, and only increased when 10 trials were used for signal averaging.

A more complex picture emerges from the SD of the M100-peak amplitudes of the estimated sources (Fig. 3(e)). The amplitude dispersion for the source in the right hemisphere stabilized for the average signal derived from 70 trials. In contrast, the dispersion of the source in the left hemisphere first decreased to a plateau that extends between 40 and 100 trials, before it decreased further and reached a constant dispersion at 140 trials. This observation of two distinct intervals is in line with the findings for the absolute M100-peak amplitude shown in Fig. S5(b), (d). There, we found in the same interval between 40 and 100 trials an increase of the median and large IQRs, followed by a distinct reduction in the IQR accompanied by a clear emergence of outliers when more than 110 trials were used to compute the average signal. Given that SD and IQR are different measures of dispersion, it is difficult to match the 140-trials threshold in SD in Fig. 3 with the 110-trials threshold in IQR in Fig. S5.

As for the M100-peak latency (Fig. 3(f)), with the exception of very small sizes of trial sets, dispersion stays within the circa 2-ms sampling resolution of the signal for the source in the right hemisphere, but not for the weaker source in the left hemisphere, for which it reaches levels of roughly 5 to 6 times the sampling resolution even for large trial-set sizes. Note that except for the latency estimates of the source in the right hemisphere (Wilcoxon test, $p = 0.243$), statistically significantly smaller standard deviations of the considered parameter estimates were obtained with the BCA method of bootstrapping than those achieved with the BTL method (Wilcoxon test, all $p < 0.05$ when the two samples of 19 standard deviations each were compared against each other, i.e., one 19-element sample vs the other).

D. Proposed Framework for an Experimental Design

Based on the findings outlined in Sections III-A, III-B, and III-C, we propose the following framework for the final design of an experiment constituting the core of a new study. This framework also applies to EEG. It encompasses standard MEG/EEG

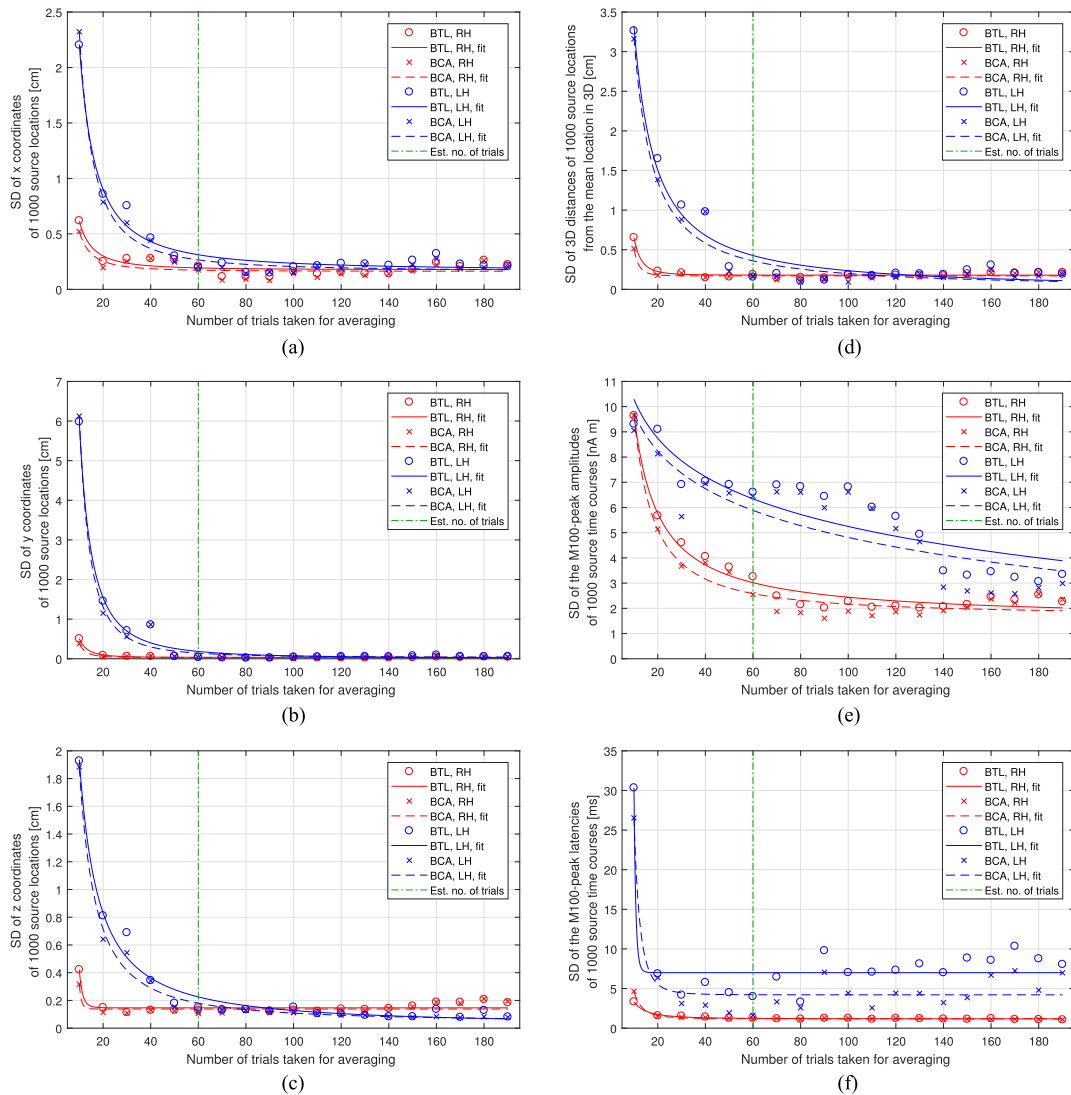


Fig. 3. Estimated standard deviations (SD) of source locations for x - (a), y - (b) and z - (c) coordinates, 3D Euclidean distances from the mean locations with respect to the origin of the head frame (d), the absolute peak amplitude of the source waveform (e) and its latency (f) for the two considered sources and for each considered number of trials obtained using the BTL (Section II-D) and the BCA (Section II-E) method of bootstrapping. Note the different scales of ordinates in (a), (b), and (c). Green dash-dotted vertical lines indicate the estimated number of trials needed for signal averaging.

routines for data acquisition and pre-processing, and the use of BTL (Section II-D) or BCA (Section II-E) as relevant bootstrap approaches (for a discussion, see Section IV-B):

- 1) Using the experimental paradigm in question, perform a lengthy MEG/EEG pre-measurement, with as many trials as possible, with one or a few subjects.
- 2) Choose the data analysis pipeline and pre-process the acquired data (artifact rejection, filtering, baseline correction, etc.) using that pipeline.
- 3) Bootstrap pre-processed single trials using BTL or BCA for various sizes of trial sets.
- 4) Evaluate the statistical properties of the parameter estimates of interest for each size of trial sets.
- 5) Based on this statistical evaluation, decide on a sufficient number of trial repetitions to be used in the final experiment to be run for the entire population of subjects, using a chosen criterion.

- 6) Run the final experiment, using the previously chosen pipeline to analyze the data.

We expect for the majority of MEG/EEG studies that the number of trials determined in this way is clearly smaller not only compared to the one used in the lengthy pre-experiment, but also to the 100–200 trials routinely used in MEG/EEG experiments.

IV. DISCUSSION

We have proposed a bootstrap framework with the aim to substantially reduce the number of event-related trials in MEG/EEG experiments focusing on brain evoked responses. To achieve this, we have adhered to the classical SPN model, expressed by Eq. (1), as to date, it is still the most commonly used model across a wide range of experimental paradigms. Even though several alternative models have been proposed (see, e.g., [6], [7], [33], [45]–[47] and references therein), the SPN model is

still omnipresent. This may be rooted in its simplicity, which translates to the sparseness concerning the number of considered parameters, thus deftly following the Ockham's razor principle. This sparseness is justified inasmuch as an unambiguous decomposition of the measured superposition of the brain response and the contaminating noise is not possible. This is because in a trial t , the matrix equation—be it Eq. (1) or any alternative approach—reflecting the measured superposition contains [at least] two matrix unknowns of the same size $C \times N$ (with C denoting the number of channels).

We have demonstrated the usefulness of the proposed bootstrap framework by its application to the estimation of cortical locations of current sources underlying the measured magnetic field and of selected characteristics of the corresponding time courses. To this aim, we have used an illustrative inverse-solution method, the spatio-temporal matching pursuit approach. We note that the applicability of the proposed bootstrap framework is not limited by STMP in any way, and that the framework can also be used with other inverse-solution strategies. This may, of course, lead to somewhat different results as to the number of trials eventually considered to be sufficient but we see no reasons to expect that the feasibility of the proposed bootstrap framework would not hold for those strategies.

A. Estimating the Number of Trials

The proposed bootstrap framework offers an insight into the quality of estimates for different numbers of trials, without the need of performing multiple experiments. This is indeed very useful information whenever one has to compromise on mutually contradicting requirements regarding experimental designs (and the related quality of estimation) and experimental cost. Importantly, this issue regarding costs is not only restricted to financial aspects of the measurement time per se, but also includes the artifact problem, which comes into play when the measurement session is prolonged beyond an acceptable time limit, when mere habituation becomes an important factor too (see Section I). The results presented in Section III-A–C, which are in line with these considerations, have led us to the procedure proposed in Section III-D. Further analyses, not shown in this work, indicated that similar results are obtained for only 200 instead of 1000 bootstrap realizations (see also [48]). Therefore, we suggest that whenever time is critical it may be sufficient to constrain the number of bootstrap realizations to about 200.

Darvas *et al.* [16] assessed dispersions of similar parameters like those we studied and also did so with bootstrap, though they did not evaluate different sizes of trial sets. Irrespective of dispersion they attempted to evaluate bias; however, their approach was based on a debatable assumption that the inverse solutions obtained for the mean signal of the original, non-bootstrapped set of trials can be treated as the ground truth.

The fact that we observe hemispheric differences for the M100 parameters (i.e., peak amplitude and latency) of the two sources is nothing unusual. It is related to the observation reported in literature that the auditory cortices in the two hemispheres may well differ in their morphology [36], [49]. Thus, it may have implications on the selection of the considered threshold

for the sufficient number of trials. In choosing the criterion for this threshold, we have essentially followed the accepted thresholds for spatial and temporal resolution (sampling) of the MEG method [40]–[44]. Remarkably, the 60-trial threshold derived from the SDs of the source coordinates (Figs. 3(a)–(d)) appears to be a reliable result, even for small numbers of trials far below the “magic” number 100. Nonetheless, the different SDs obtained for the M100-peak amplitude of the two sources in our example imply that increased caution in the interpretation of the results must be exercised if—as is the case here—substantial differences in the SNRs underlying the source waveforms are present. Hence, we abstain from restricting the generality of the proposed approach by any particular selection, and leave the choice of the criterion for selecting the ultimate number to the user, because different experimental paradigms and processing strategies (like, e.g., a particular inverse-solution approach) may require different criteria, of which there are plentiful.

Finally, on a different note, the proposed framework may also be used to identify a subset of single trials different than those consisting of the first 10, first 20, first 30, ... trials we have been discussing here, i.e. a subset of those trials that show the largest stimulus-evoked response among all trials. This approach might be interesting especially with regard to a potential increase of the SNR. However, the single trials forming this subset will spread across the entire pool of trials, and would therefore require the acquisition of the full number of trials (typically 100 or more). Thus, this approach would neither lead to a reduction of the number of acquired trials nor of the measurement time and, thus, contradict the intended scope of our work.

B. Performance of BTL and BCA Bootstrapping Approaches

We have outlined in Section II that only two of the considered bootstrapping approaches are applicable to properly evaluate the number of trials for MEG/EEG signal averaging: the already known bootstrapping of trial labels (BTL, Section II-D) and the newly proposed bootstrapping of complex amplitudes across trials (BCA, Section II-E). In principle, both approaches can be applied in step 3 of the framework proposed in Section III-D. However, it is still worthwhile to have a closer look at the performance of the two approaches.

Pronounced differences between the results obtained from BTL and BCA are only found by comparing the standard deviations of the absolute M100-peak amplitude (Fig. 3(e)) and of the corresponding M100-peak latency (Fig. 3(f)), and, to a lesser extent, for the source coordinates for smaller sizes of trial sets (Figs. 3(a)–(d)). Taken together, our results show that the bootstrapped trials result in larger SNRs, which lead to smaller SDs, for the mean signals in the BCA approach. These better SNRs stem from the fact that the likelihood of a repeated single trial in the scheme of drawing with replacement, inherent to bootstrap, is way smaller in BCA than in BTL (see Section II-E). This property is a general advantage of the newly proposed BCA method.

The question remains what strategy one should use when collating information on parameter distributions from source

locations as well as absolute peak amplitude and peak latency. We do not advocate simply choosing the overall largest number of trials for which all distributions reach the smallest dispersion characteristics. Instead, we recommend, for each experimental paradigm (and possibly for each considered inverse solution), to examine not only the dispersion characteristics of the parameter estimators, but also whether adding additional trials to the measurement protocol results in a qualitative change of those characteristics.

C. Bootstrapping and Models of Evoked-Response Generation

In our approach, we assume the most commonly used model of evoked response generation, the signal-plus-noise model. Sayers [50] proposed an alternative view to the SPN model. He postulated that an increase of amplitude in the trial-averaged signal stems from a phase reset of the spontaneous brain activity manifesting itself in the form of oscillations (rhythms). The phase reset is supposed to be triggered by the occurrence of the stimulus. Hence, conceptually, for a short period of time following stimulus onset, the brain rhythms are synchronized in phase. Their amplitudes sum up, and, thus, form a waveform with a morphology that contributes to the one observed in the trial-averaged experimental data. The debate as to which of the two views—SPN model or phase-reset model—reflects the true mechanisms of evoked response generation has been intense (see, e.g., [47] and references therein). In our opinion, it is not possible to unambiguously identify one model or the other in real experimental data due to the fact that MEG (as well as EEG) measurements always register a superposition of two unknowns—the signal and the noise, which are hard to disentangle from each other. This is so because one deals with an underdetermined system of equations (more variables than linearly independent equations), as even for a large number of channels in modern MEG/EEG systems there may always be more current sources contributing to the measured magnetic/electric field.

Referring to the novel BCA method (Section II-E), one might argue that if the phase-reset model applies then we may not assume phase independence across trials and, hence, bootstrapping complex amplitudes across trials would be questionable, unless noise components do not obey phase reset. Following [50], one might conclude that since spontaneous processes are in phase only for a certain period of time in which they create the evoked response and out of phase elsewhere, the same process can be considered a signal in some period of time and noise elsewhere. If so, it is difficult to assume mutual independence of phase, or complex amplitudes, across trials, unless for some limited periods of time. Nonetheless, irrespective of these theoretical considerations, given that the mechanisms of evoked responses generation are intensely debated [47], [51], we note that BCA, when applied to the data we used, yields results similar to those from the simpler BTL approach presented in Section II-D. Therefore, we leave it to the preference of the user which of the two approaches they would choose.

D. Future Real-Time Application

If hardware limitations (such as the speed of a CPU) were not an issue, an ideal scenario would be to identify a small but meaningful number of individual trials already during data acquisition. In such a scenario, one would evaluate the quality of estimates in real time, sequentially running consecutive trials until the dispersion measures like those depicted in Fig. 3 reach stability according to a chosen criterion. Although attempts to localize MEG/EEG sources in real time have been proposed [52], [53], they significantly compromise on the spatial resolution of the solution space due to computational burden. As demonstrated in this work, employing bootstrap provides statistical evaluation of a meaningful threshold when reducing the number of trials. Even if not satisfactorily computable in the real time of an experiment today, it may—with the fast pace of CPU development—be realistic in a not so distant future.

V. CONCLUSION

The aim of this work was to develop a systematic framework that allows MEG/EEG experiments to be performed with a number of stimulus repetitions as small as possible to minimize the time load for the subjects and thus reduce related artifacts due to muscular tension, involuntary head movements, eye blinks, pronounced alpha rhythm due to sleepiness, etc. We have shown that the bootstrap technique is a powerful tool for optimizing the number of single trials used for signal averaging in MEG to estimate parameters like spatial coordinates of source locations³ or peak attributes of respective time courses. Information from bootstrap resampling can further be utilized for the estimation of the required confidence interval or for statistical inference to differentiate between given parameters of interest with a desired statistical power. We kept the framework as general as possible, since our motivation was to provide a universal approach rather than a definite answer to the quest for the optimal number of trials, which for any experimental paradigm or processing pipeline (including source localization) might be different. The relevant literature either provides only some general recommendations for this number, which are based on tacit priors, or, for more complex experimental designs, tends to use the “it depends” argument. To the best of our knowledge, this work is the first attempt to systematically assess how the reduction of the number of trials influences the quality of estimates of spatial and temporal parameters of the current sources underlying brain evoked activity.

REFERENCES

- [1] P. Hansen *et al.*, *MEG: An Introduction to Methods*. London, U.K.: Oxford Univ. Press, 2010.
- [2] S. Supek and C. J. Aine, Eds., *Magnetoencephalography: From Signals to Dynamic Cortical Networks*. Berlin, Germany: Springer, 2014.
- [3] R. Hari and A. Puce, *MEG-EEG Primer*. London, U.K.: Oxford Univ. Press, 2017.

³For distributed-source localization methods, which do not return a single vertex per source, several different strategies could be proposed. The simplest would be selecting the vertex corresponding to the geometric center of a distributed source. However, these aspects are beyond the scope of this work.

- [4] D. L. Schomer and F. H. Lopes da Silva, Eds., *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. London, U.K.: Oxford Univ. Press, 2017.
- [5] G. D. Dawson, "A summation technique for the detection of small evoked potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 6, no. 1, pp. 65–84, 1954.
- [6] J. C. de Munck, F. Bijma, P. Gaura, C. A. Sieluzycski, M. I. Branco, and R. M. Heethaar, "A maximum-likelihood estimator for trial-to-trial variations in noisy MEG/EEG data sets," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 12, pp. 2123–2128, Dec. 2004.
- [7] C. Sieluzycski, R. König, A. Matysiak, R. Kus, D. Ircha and P. J. Durka, "Single-trial evoked brain responses modeled by multivariate matching pursuit," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 1, pp. 74–82, Jan. 2009.
- [8] A. Salajegheh *et al.*, "Systematic latency variation of the auditory evoked M100: From average to single-trial data," *NeuroImage*, vol. 23, no. 1, pp. 288–295, 2004.
- [9] A. Link *et al.*, "Comparing a template approach and complex bandpass filtering for single-trial analysis of auditory evoked M100," *Biomedizinische Technik*, vol. 52, no. 1, pp. 106–110, 2007.
- [10] C. Sieluzycski *et al.*, "Multivariate matching pursuit in the analysis of single-trial latency of the auditory M100 acquired with MEG," *Int. J. Bioelectromagn.*, vol. 11, no. 4, pp. 155–160, 2009.
- [11] N. Zacharias *et al.*, "The M100 component of evoked magnetic fields differs by scaling factors: Implications for signal averaging," *Psychophysiology*, vol. 48, no. 8, pp. 1069–1082, 2011.
- [12] J. Gross *et al.*, "Good practice for conducting and reporting MEG research," *NeuroImage*, vol. 65, no. 100, pp. 349–363, 2013.
- [13] T. Grootswagers *et al.*, "Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time series neuroimaging data," *J. Cogn. Neurosci.*, vol. 29, no. 4, pp. 677–697, 2017.
- [14] M. A. Boudewyn *et al.*, "How many trials does it take to get a significant ERP effect? It depends," *Psychophysiology*, vol. 55, no. 6, Jun. 2018, Art. no. e13049.
- [15] C. Braun *et al.*, "Confidence interval of single dipole locations based on EEG data," *Brain Topogr.*, vol. 10, no. 1, pp. 31–39, 1997.
- [16] F. Darvas *et al.*, "Investigations of dipole localization accuracy in MEG using the bootstrap," *NeuroImage*, vol. 25, no. 2, pp. 355–368, 2005.
- [17] C. Papadelis and A. A. Ioannides, "Localization accuracy and temporal resolution of MEG: A phantom experiment," *Int. Congr. Ser.*, vol. 1300, pp. 257–260, 2007.
- [18] R. Huffmeijer *et al.*, "Reliability of event-related potentials: The influence of number of trials and electrodes," *Physiol. Behav.*, vol. 130, pp. 13–22, 2014.
- [19] A. G. Fischer *et al.*, "Comparing the error-related negativity across groups: The impact of error- and trial-number differences," *Psychophysiology*, vol. 54, no. 7, pp. 998–1009, 2017.
- [20] M. Hämäläinen *et al.*, "Magnetoencephalography-theory, instrumentation, and applications to non-invasive studies of the working human brain," *Rev. Modern Phys.*, vol. 65, no. 2, pp. 413–497, 1993.
- [21] P. Kordowski *et al.*, "Simultaneous spatio-temporal matching pursuit decomposition of evoked brain responses in MEG," *Biol. Cybern.*, vol. 111, no. 1, pp. 69–89, 2017.
- [22] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. London, U.K.: Chapman & Hall, 1993.
- [23] A. M. Zoubir and D. R. Iskander, *Bootstrap Techniques for Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [24] A. M. Zoubir and D. R. Iskander, "Bootstrap methods and applications: A tutorial for the signal processing practitioner," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 10–19, Jul. 2007.
- [25] B. Efron, *The Jackknife, the Bootstrap, and Other Resampling Plans*. Philadelphia, PA, USA: SIAM, 1982.
- [26] W. Ebisuzaki, "A method to estimate the statistical significance of a correlation when the data are serially correlated," *J. Climate*, vol. 10, no. 9, pp. 2147–2153, Sep. 1997.
- [27] C. Kirch and D. N. Politis, "TFT-bootstrap: Resampling time series in the frequency domain to obtain replicates in the time domain," *Ann. Statist.*, vol. 39, no. 3, pp. 1427–1470, 2011.
- [28] J. C. de Munck, "Estimating stationary dipoles from MEG/EEG data contaminated with spatially and temporally correlated background noise," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1565–1572, Jul. 2002.
- [29] J. Theiler *et al.*, "Testing for nonlinearity in time series: The method of surrogate data," *Phys. D*, vol. 58, no. 1–4, pp. 77–94, 1992.
- [30] M. X. Cohen, "Assessing transient cross-frequency coupling in EEG data," *J. Neurosci. Methods*, vol. 168, no. 2, pp. 494–499, 2008.
- [31] J. Daume *et al.*, "Phase-amplitude coupling and long-range phase synchronization reveal frontotemporal interactions during visual working memory," *J. Neurosci.*, vol. 37, no. 2, pp. 313–322, 2017.
- [32] M. J. Hülsemann *et al.*, "Quantification of phase-amplitude coupling in neuronal oscillations: Comparison of phase-locking value, mean vector length, modulation index, and generalized-linear-modeling-cross-frequency-coupling," *Front. Neurosci.*, vol. 13, no. 573, pp. 1–15, 2019.
- [33] C. Sieluzycski and P. Kordowski, "Maximum-likelihood estimation of channel-dependent trial-to-trial variability of auditory evoked brain responses in MEG," *Biomed. Eng. Online*, vol. 13, no. 75, pp. 1–19, 2014.
- [34] A. M. Galaburda *et al.*, "Right-left asymmetries in the brain," *Science*, vol. 199, no. 4331, pp. 852–856, 1978.
- [35] J. Hine and S. Debener, "Late auditory evoked potentials asymmetry revisited," *Clin. Neurophysiol.*, vol. 118, no. 6, pp. 1274–1285, 2007.
- [36] M. E. Shaw *et al.*, "How anatomical asymmetry of human auditory cortex can lead to a rightward bias in auditory evoked fields," *NeuroImage*, vol. 74, pp. 22–29, 2013.
- [37] F. Ségonne *et al.*, "A hybrid approach to the skull stripping problem in MRI," *NeuroImage*, vol. 22, no. 3, pp. 1060–1075, 2004.
- [38] G. Nolte, "The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors," *Phys. Med. Biol.*, vol. 48, no. 22, pp. 3637–3652, 2003.
- [39] R. Oostenveld *et al.*, "FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data," *Comput. Intell. Neurosci.*, vol. 2011, pp. 1–9, 2011.
- [40] D. S. Barth *et al.*, "Magnetic localization of a dipolar current source implanted in a sphere and a human cranium," *Electroencephalogr. Clin. Neurophysiol.*, vol. 63, no. 3, pp. 260–262, 1986.
- [41] T. Yamamoto *et al.*, "Magnetic localization of neural activity in the human brain," *Proc. Nat. Acad. Sci. USA*, vol. 85, no. 22, pp. 8732–8736, 1988.
- [42] R. M. Leahy *et al.*, "A study of dipole localization accuracy for MEG and EEG using a human skull phantom," *Electroencephalogr. Clin. Neurophysiol.*, vol. 107, no. 2, pp. 159–173, 1998.
- [43] J. Vrba and S. E. Robinson, "Signal processing in magnetoencephalography," *Methods*, vol. 25, no. 2, pp. 249–271, 2001.
- [44] L. Kaufman and Z.-L. Lu, "Basics of neuromagnetism and magnetic source imaging," in *Magn. Source Imag. Hum. Brain*, Z.-L. Lu and L. Kaufman, Eds. Mahwah, NJ: Lawrence Erlbaum Associates, 2003, pp. 1–42.
- [45] W. Truccolo *et al.*, "Trial-to-trial variability of cortical evoked responses: Implications for the analysis of functional connectivity," *Clin. Neurophysiol.*, vol. 113, no. 2, pp. 206–226, 2002.
- [46] S. D. Georgiadis *et al.*, "Single-trial dynamical estimation of event-related potentials: A kalman filter-based approach," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 8, pp. 1397–1406, Aug. 2005.
- [47] J. C. de Munck and F. Bijma, "How are evoked responses generated? The need for a unified mathematical framework," *Clin. Neurophysiol.*, vol. 121, no. 2, pp. 127–129, 2010.
- [48] J. P. Rosenfeld *et al.*, "Bootstrapping the P300 in diagnostic psychophysiology: How many iterations are needed?" *Psychophysiology*, vol. 54, no. 3, pp. 366–373, 2017.
- [49] J. Benner *et al.*, "Prevalence and function of heschl's gyrus morphotypes in musicians," *Brain Struct. Funct.*, vol. 222, no. 8, pp. 3587–3603, 2017.
- [50] B. Sayers *et al.*, "The mechanism of auditory evoked EEG responses," *Nature*, vol. 247, no. 5441, pp. 481–483, 1974.
- [51] N. Yeung *et al.*, "Detection of synchronized oscillations in the electroencephalogram: An evaluation of methods," *Psychophysiology*, vol. 41, no. 6, pp. 822–832, 2004.
- [52] C. Dinh *et al.*, "Real-time MEG source localization using regional clustering," *Brain Topogr.*, vol. 28, no. 6, pp. 771–784, 2015.
- [53] L. Duque-Muñoz *et al.*, "Simplified EEG inverse solution for BCI real-time implementation," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2016, pp. 4051–4054.