



Non-Gaussian probabilistic MEG source localisation based on kernel density estimation[☆]

Hamid R. Mohseni^{a,b,*}, Morten L. Kringelbach^{b,c,d,e}, Mark W. Woolrich^c, Adam Baker^{a,c},
Tipu Z. Aziz^d, Penny Probert-Smith^a

^a Institute of Biomedical Engineering, School of Engineering Science, University of Oxford, Oxford, UK

^b Department of Psychiatry, University of Oxford, Warneford Hospital, UK

^c Oxford Centre for Human Brain Activity (OHBA), Department of Psychiatry, University of Oxford, UK

^d Oxford Functional Neurosurgery, Nuffield Department of Surgery, John Radcliffe Hospital, Oxford, UK

^e CFIN/MindLab, Aarhus University, Aarhus, Denmark

ARTICLE INFO

Article history:

Accepted 7 September 2013

Available online 20 September 2013

Keywords:

Magnetoencephalography

Source reconstruction

Non-Gaussian

Beamformer

Null-beamformer

ABSTRACT

There is strong evidence to suggest that data recorded from magnetoencephalography (MEG) follows a non-Gaussian distribution. However, existing standard methods for source localisation model the data using only second order statistics, and therefore use the inherent assumption of a Gaussian distribution. In this paper, we present a new general method for non-Gaussian source estimation of stationary signals for localising brain activity from MEG data. By providing a Bayesian formulation for MEG source localisation, we show that the source probability density function (pdf), which is not necessarily Gaussian, can be estimated using multivariate kernel density estimators. In the case of Gaussian data, the solution of the method is equivalent to that of widely used linearly constrained minimum variance (LCMV) beamformer. The method is also extended to handle data with highly correlated sources using the marginal distribution of the estimated joint distribution, which, in the case of Gaussian measurements, corresponds to the null-beamformer. The proposed non-Gaussian source localisation approach is shown to give better spatial estimates than the LCMV beamformer, both in simulations incorporating non-Gaussian signals, and in real MEG measurements of auditory and visual evoked responses, where the highly correlated sources are known to be difficult to estimate.

© 2013 The Authors. Published by Elsevier Inc. All rights reserved.

Introduction

Magnetoencephalography (MEG) is emerging as a valuable neuro-imaging technique due to its excellent (millisecond) temporal resolution and reasonable spatial resolution. Unlike functional magnetic resonance imaging (fMRI) or positron emission tomography (PET), MEG provides direct measurement of brain activity through recording the magnetic induction over the scalp produced by electrical activity in neural cell assemblies, and thus opens the possibility of exploring the underlying dynamics of neural networks that occurs at a millisecond time scale (Hämäläinen et al., 1993; Hansen et al., 2010).

There is evidence that MEG normally follows a non-Gaussian distribution (Elul et al., 1975; Lee et al., 2000). This observation is supported by Nagarajan et al. (2005, 2006), who demonstrated that evoked brain sources are often characterised by spikes or by modulated harmonic functions, leading to a non-Gaussian distribution. Other evidence for

non-Gaussianity of MEG data, is given by the successful employment of independent component analysis (ICA), which separate sources based on non-Gaussian measures such as kurtosis, meaning that there are elements of non-Gaussianity in the MEG data (Nagarajan et al., 2005, 2006). Furthermore, it is possible that MEG data may show a Gaussian distribution, while the underlying sources may not be Gaussian due to the fact that superposition of many non-Gaussian sources can resemble a Gaussian distribution, according to the central limit theorem.

The scope for observing and interpreting neural activity depends on the resolution with which sources of activity can be localised within the brain from the MEG measurements at the skull. At the heart of this problem is the challenge of finding the optimal solution of the so-called inverse problem, which can only be solved by introducing a priori assumptions on the generation of MEG signals. One of the most successful methods for localising brain activity with MEG is a set of array processing methods known as *beamformers*, which have been widely used not only in MEG (Robinson et al., 1999; Van Veen et al., 1997), but also in communications and radar signal processing applications (Van Veen and Buckley, 1988). Perhaps the most common beamformer used in MEG imaging is linearly constrained minimum variance (LCMV) spatial filter, which is designed to minimise the power of the measured signal originating from all locations except the location(s) of interest

[☆] This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-No Derivative Works License, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

* Corresponding author.

(Van Veen and Buckley, 1988). However, as we show in this paper, its performance is only optimal in the presence of measurement whose pdfs can be described using only first and second order statistics (i.e., Gaussian distribution).

Although originally proposed for vector source localisation, the LCMV beamformer has since been adapted for scalar beamforming. It has been shown that these two types of beamformers are equivalent, in terms of output power and output SNR, if their pointing directions are optimised (Sekihara et al., 2006). There are several other approaches that have been proposed to improve or extend the LCMV beamformer. Among them are graphical models for event-related field denoising and localisation (Zumer et al., 2007, 2008), which provide a modification of the method presented in Nagarajan et al. (2006) by incorporating the lead-field matrices and temporal information. Zumer et al. (2008) have shown using a Bayesian formulation that, if the prior is assumed to be uniformly distributed, the maximum a posteriori (MAP) estimation converts to the maximum likelihood estimation which is in turn equivalent to the LCMV beamformer. These graphical models also typically assume Gaussian distributions for sources and noise. A Bayesian paradigm has also been proposed to derive several MEG source localisation approaches including the LCMV beamformer (Wipf and Nagarajan, 2009). Here we also justify the LCMV beamformer from a Bayesian perspective to be able to employ kernel density estimators in the source estimation.

Since the set of weights used by the beamformer relies on the data covariance matrix, the spatial resolution and accuracy of localisation are dependent on the estimation of the covariance matrix. Several methods have been suggested to provide a robust regularisation especially when the covariance matrix is (nearly) rank deficient through techniques such as diagonal loading, which can improve the robustness of the filter (Cox et al., 1987; Hudson, 1981). Other methods include adaptive iterative algorithms such as expectation maximisation (Friston et al., 2002), eigenspace beamformer (Sekihara et al., 2002), relevance vector machines (Wipf and Nagarajan, 2007) and Bayesian principal component analysis (Woolrich et al., 2011). The latter also provides a robust estimate of the rank of the covariance matrix. However, since all these methods only improve the estimated covariance matrix, the data still is assumed to follow a Gaussian distribution.

Anatomical and spatial information have also been used to improve the beamformer source localisation by incorporating realistic constraints on the orientation of the sources (Hillebrand, 2003; Limpi et al., 2006). In addition, statistical analyses have been specifically designed for the beamformer power spectrum to provide a reliable inference about the neural activity (Barnes and Hillebrand, 2003; Brookes et al., 2004, 2005).

Given the fact that MEG data follow non-Gaussian distributions, we hypothesise that source localisation algorithms that are able to capture non-Gaussianity should perform better than a Gaussian based method such as the LCMV beamformer. In this paper we provide a novel approach for non-Gaussian MEG source localisation. The approach presented here combines a Bayesian formulation of the beamformer with a model of the measurement pdf based on the kernel density estimator (Parzen, 1962), to provide a more accurate estimate of the source pdf.

Standard beamformer is sensitive to temporally correlated sources such that moderately correlated sources are generally poorly identified (Baillet et al., 2001; Sekihara et al., 2006). A number of studies have tried to address the problem of correlated sources in MEG source localisation. Wipf and Nagarajan (2007) showed that a Bayesian method for learning sparse models has the potential to remove the undesirable effects of correlation between sources. A dual source beamformer technique has also been employed to image temporally correlated sources (Brookes et al., 2007). An alternative method to circumvent this problem is the so called *null-beamformer* technique, which has been employed to suppress the activity from regions that are known to have interfering activity (Dalal et al., 2006; Haykin, 2002; Hui and Leahy, 2006). The null-beamformer has been further validated in auditory MEG source localisation (Popescu et al., 2008), and has also been

employed for MEG recordings concurrent to deep brain stimulation (Mohseni et al., 2010, 2012).

In light of the correlated source problem, we extend our method to provide a Bayesian formulation of the null-beamformer for non-Gaussian data. In this approach, the joint distribution of the correlated sources is estimated and the marginal distribution is employed to obtain the pdf of source of interest. It is shown that both the power and time-series obtained using this method are equivalent to those obtained from the null-beamformer when the measurement pdf is assumed to be Gaussian.

The organisation of the paper is as follows. First, in the **Problem formulation** section the problem is formulated. In the **LCMV-beamformer and its Bayesian derivation** section the LCMV beamformer is briefly reviewed and its solution is justified from a Bayesian perspective. In **The Non-Gaussian probability distribution (PD) beamformer** section the method is described in a Bayesian framework with a non-informative prior, and the details of an efficient implementation are given. In the **Non-Gaussian probability marginal distribution for correlated sources** section, in order to tackle the problem of making accurate source estimations from data produced by highly correlated sources, the proposed method is extended using a marginal distribution and its relation to the null-beamformer is explored. In **A kernel based estimation of the measurement pdf** section, a kernel based method for the estimation of the measurement pdf is explained. This is then applied to both simulated and real MEG data recorded in auditory and visual paradigms as presented in the **Experimental results** section. The sensitivity and efficiency of the methods are also described with regard to their free parameters, especially the diagonal loading factor λ in the beamformer and the bandwidth of the kernel h in the proposed method. In **Summary and discussion** section, the method is summarised and discussed. Finally, in **Appendix A**, we present some theoretical analyses of the method including the impact of noise, relation to the beamformer with a general linear constraint and an analysis of the convergence of the algorithm.

Methods

Problem formulation

Let the vector $\mathbf{y}_t \in \mathbb{R}^N$ be the measurement recorded at time samples $t \in \{1, \dots, T\}$ from N sensor sites. Suppose that \mathbf{y}_t is composed of the magnetic fields due to active current dipoles plus noise:

$$\mathbf{y}_t = \mathbf{F}\mathbf{s}_t + \mathbf{n}_t \quad (1)$$

where $\mathbf{F} = [\mathbf{f}_1 \dots \mathbf{f}_q] \in \mathbb{R}^{N \times q}$ is the matrix of the lead-fields and $\mathbf{s}_t = [\mathbf{s}_{1,t} \dots \mathbf{s}_{q,t}]^T \in \mathbb{R}^q$ is the vector of q sources. Here, $\mathbf{n}_t \in \mathbb{R}^N$ is the additive noise which is independent from the sources.

Suppose that \mathbf{y}_t , \mathbf{s}_t and \mathbf{n}_t are zero-mean stationary processes with probability density functions (pdfs) $g_{\mathbf{y}}(\cdot)$, $g_{\mathbf{s}}(\cdot)$ and $g_{\mathbf{n}}(\cdot)$, respectively. Furthermore, assume that each source $s_{i,t} \in \mathbb{R}$, $i \in \{1, \dots, q\}$, which is a stationary process, has the pdf $\mathcal{G}_{s_i}(\cdot)$, and that the lead-field vector of each source $\mathbf{f}_i \in \mathbb{R}^N$ is deterministic and known (Mosher et al., 1999).

It is also assumed, for simplicity, that the sources $s_{i,t}$, $i \in \{1, \dots, q\}$ are one-dimensional, whereby the method is called scalar source localisation. However, in reality the current dipoles are multidimensional and a vector source localisation, which assumes that the sources are two or three dimensional, is more appropriate.

LCMV-beamformer and its Bayesian derivation

To set the scene and better clarify the method proposed, we first present the LCMV beamformer and its justification within a Bayesian framework.

To locate the source $s_{k,t}$ at a particular location k and time sample t , the approach known as the LCMV beamformer is often used for MEG

source analysis. It is a linear filter that constrains the lead-field vector \mathbf{f}_k to pass the signal at a location of interest, while minimising the covariance of the measurement \mathbf{R}_y :

$$\underset{\mathbf{w}}{\operatorname{argmin}} \quad \mathbf{w}^T \mathbf{R}_y \mathbf{w}, \quad \text{subject to: } \mathbf{w}^T \mathbf{f}_k = 1. \quad (2)$$

Here, \mathbf{w} is a vector of weights and its closed-form solution using Lagrange multiplier method is given by $\mathbf{w}^T = (\mathbf{f}_k^T \mathbf{R}_y^{-1} \mathbf{f}_k)^{-1} \mathbf{f}_k^T \mathbf{R}_y^{-1}$. The estimated time-series $\hat{s}_{k,t}$ at time t and the estimated power \mathcal{P}_{s_k} are then given by:

$$\hat{s}_{k,t} = (\mathbf{f}_k^T \mathbf{R}_y^{-1} \mathbf{f}_k)^{-1} \mathbf{f}_k^T \mathbf{R}_y^{-1} \mathbf{y}_t \quad (3)$$

$$\mathcal{P}_{s_k} = (\mathbf{f}_k^T \mathbf{R}_y^{-1} \mathbf{f}_k)^{-1}. \quad (4)$$

Now we justify Eqs. (3) and (4) from a Bayesian perspective. One may rewrite Eq. (1) as:

$$\mathbf{y}_t = \mathbf{f}_k s_{k,t} + \eta_t \quad (5)$$

where $\eta_t = \sum_{i \neq k} \mathbf{f}_i s_{i,t} + \mathbf{n}_t$ is the interference coming from all other sources plus noise. We then assume that the pdf for η_t , $g_\eta(\cdot)$, can be approximated with a Gaussian function, i.e., $\eta_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_\eta)$. In the Bayesian framework, the posterior pdf of the source of interest $p(s_k | \mathbf{y}_t)$ is estimated and its expected value is considered as the estimation of the time-series $\hat{s}_{k,t}$. This is accomplished via Bayes' rule given by:

$$p(s_k | \mathbf{y}_t) \propto p(\mathbf{y}_t | s_k) p(s_k). \quad (6)$$

In Eq. (5), since we assumed that the interference plus noise distribution $g_\eta(\cdot)$ is Gaussian the likelihood $p(\mathbf{y}_t | s_k)$ is also Gaussian:

$$p(\mathbf{y}_t | s_k) \propto \exp[-(\mathbf{y}_t - \mathbf{f}_k s_k)^T \mathbf{R}_\eta^{-1} (\mathbf{y}_t - \mathbf{f}_k s_k)]. \quad (7)$$

By assuming an uninformative (uniform) prior on the source pdf $p(s_k)$, using Bayes rule (Eq. (6)), it can be shown that the posterior $p(s_k | \mathbf{y}_t)$ is also Gaussian and its mean (the estimated time-series) and variance (the estimated source power) are given by [see also (Dogandzic and Nehorai, 2000; Zumer et al., 2008) for when \mathbf{R}_η is assumed to be known]:

$$\hat{s}_{k,t} = (\mathbf{f}_k^T \mathbf{R}_\eta^{-1} \mathbf{f}_k)^{-1} \mathbf{f}_k^T \mathbf{R}_\eta^{-1} \mathbf{y}_t \quad (8)$$

$$\mathcal{P}_{s_k} = (\mathbf{f}_k^T \mathbf{R}_\eta^{-1} \mathbf{f}_k)^{-1}. \quad (9)$$

Eqs. (8) and (9) are equivalent to Eqs. (3) and (4), except for the fact that in the former \mathbf{R}_η has been used instead of \mathbf{R}_y . If we assume that the power of the source of interest is considerably smaller than the power of all other sources plus noise, we can approximate $\mathbf{R}_y \approx \mathbf{R}_\eta$, making the beamformer and Bayesian solutions equivalent. Given the fact that brain activity generally has distributed sources, this approximation is acceptable in the MEG applications. However, in the event that this approximation is not true, we may face inaccurate and biased results.

It is notable that, in Eq. (2), minimising the power of the noise plus interference \mathbf{R}_η is better known as the minimum variance distortionless response (MVDR) beamformer. This should potentially perform better than minimising the measurement power \mathbf{R}_y alone (Ehrenberg et al., 2010; Van Trees, 2004). Nevertheless, the main advantage of the LCMV beamformer over the Bayesian solution is that \mathbf{R}_y can be effectively estimated from the measurement, while estimation of \mathbf{R}_η in most applications is not possible (Wipf and Nagarajan, 2007).

The Non-Gaussian probability distribution (PD) beamformer

We start by assuming that the distribution over η , $g_\eta(\cdot)$, is known. Using this with Eq. (5), gives $p(\mathbf{y}_t | s_k) = g_\eta(\mathbf{y}_t - \mathbf{f}_k s_k)$. By inserting this into Eq. (6) and assuming that the prior is uniformly distributed, we have:

$$p(s_k | \mathbf{y}_t) \propto g_\eta(\mathbf{y}_t - \mathbf{f}_k s_k). \quad (10)$$

As with the classic Gaussian LCMV beamformer, we assume that the source power is considerably smaller than the interference plus noise power, and that the pdf of η can be approximated by the pdf of \mathbf{y} ; i.e., we assume $g_\eta(\cdot) \approx g_y(\cdot)$. As explained in the previous section, this is equivalent to implicitly assuming that the power at one voxel is considerably smaller than the power at all other voxels within the brain plus noise. This assumption is valid in most applications of MEG source localisation. Moreover, the main benefit of this assumption is that the data pdf can be estimated from the measurements, while estimation of the pdf of the noise plus interference is intractable in most cases. Furthermore, in Appendix A it is shown that we can exactly estimate the source pdf in the presence of uncorrelated sources and noise free signal. This means that regardless of the power of the interference, if the noise power is small enough, this assumption does not have any negative impact on the results, for example in ERFs obtained from averaging over many trials.

Based on the assumption that $g_\eta(\cdot) \approx g_y(\cdot)$, Eq. (10) is given as

$$p(s_k | \mathbf{y}_t) \propto g_y(\mathbf{y}_t - \mathbf{f}_k s_k). \quad (11)$$

Here, $g_y(\cdot)$ is obtained using the mechanism that is explained in [A kernel based estimation of the measurement pdf](#) section. We then may use the expected value of the posteriori as the estimation of the time-series $\hat{s}_{k,t}$:

$$\hat{s}_{k,t} = \int s_k p(s_k | \mathbf{y}_t) ds_k. \quad (12)$$

Other values including the mode of $p(s_k | \mathbf{y}_t)$ could also be used instead.

In addition to the posteriori distribution $p(s_k | \mathbf{y}_t)$, it is necessarily to obtain the source pdf given all observations, defined by $g_{s_k}(s_k) = p(s_k | \mathbf{y}_1, \dots, \mathbf{y}_T)$, in order to estimate the source activity, which can be represented by the variance (or fourth order moments) of $g_{s_k}(s_k)$. Normally in this case, it is assumed that the measurement samples are independent which leads to $g_{s_k}(s_k) \propto \prod_{t=1}^T g_y(\mathbf{y}_t - \mathbf{f}_k s_k)$. However, we use $g_{s_k}(s_k) \equiv p(s_k | E\{\mathbf{y}\})$ as an estimation of the source pdf, where $E\{\cdot\}$ is the expectation operator over $g_y(\cdot)$. Source pdf estimation using $g_{s_k}(s_k) \equiv p(s_k | E\{\mathbf{y}\})$ is much faster especially when it is required to estimate the power of a large number of points inside the brain volume. Using Eq. (11) and the fact that the measurement is zero mean, we therefore have

$$g_{s_k}(s_k) \propto g_y(\mathbf{f}_k s_k). \quad (13)$$

Here, we have replaced $g_y(-\mathbf{f}_k s_k)$ by $g_y(\mathbf{f}_k s_k)$, since the minus sign does not have any impact on the estimated power. Based on the above equation the power of the source of interest is given by:

$$\mathcal{P}_{s_k} = \int s_k^2 g_{s_k}(s_k) ds_k. \quad (14)$$

An important property of the two equations above is that an optimal estimation of the source pdf (and consequently the source power) is obtained if the noise power is very small, even if the interference is large (please also see [Appendix A](#)).

For the convenience of the reader, the pseudo-code of the method is presented in [Algorithm 1](#). The algorithm provides the expected value $\hat{s}_{k,t}$ at time t and the power \mathcal{P}_{s_k} of a source at a location with lead-field \mathbf{f}_k . The inputs of this algorithm are measurements \mathbf{y}_t and the lead-field of the source of interest \mathbf{f}_k . The data distribution $g_{\mathbf{y}}(\mathbf{y})$ is a multidimensional non-Gaussian distribution, and it is estimated from \mathbf{y}_t using kernel density estimators (see [A kernel based estimation of the measurement pdf](#) section). The outputs of the algorithm are the time-series and power at the location of interest. Similar to the traditional beamformer, in practice the brain is divided into a number of discrete voxels and the algorithm is implemented separately for each. Note that for simplicity of notation, we have used the integral rather than the discrete sum, but the integrals which are over univariate pdfs should be calculated numerically.

Algorithm 1. Implementation of the method for the non-Gaussian PD Beamformer

```
% estimating the time-series  $\hat{s}_{k,t}$ 
for  $t = 1$  to  $T$  do
    estimate  $g_{\mathbf{y}}(\mathbf{y})$  as explained in “A kernel based estimation of the measurement pdf section” or any other approach
    set  $p(s_k|\mathbf{y}_t) \propto g_{\mathbf{y}}(\mathbf{y}_t - \mathbf{f}_k s_k)$  % Eq. (11)
    normalise  $p(s_k|\mathbf{y}_t) = \frac{p(s_k|\mathbf{y}_t)}{\int p(s_k|\mathbf{y}_t) ds_k}$ 
    set  $\hat{s}_{k,t} = \int s_k p(s_k|\mathbf{y}_t) ds_k$  % Eq. (12)
end for

% fast estimation of the power  $\mathcal{P}_{s_k}$  at a location with lead-field  $\mathbf{f}_k$ 
estimate  $g_{\mathbf{y}}(\mathbf{y})$  as explained in A kernel based estimation of the measurement pdf section or any other approach
set  $g_{s_k}(s_k) \propto g_{\mathbf{y}}(\mathbf{f}_k s_k)$  % Eq. (13)
normalise  $g_{s_k}(s_k) = \frac{g_{s_k}(s_k)}{\int g_{s_k}(s_k) ds_k}$ 
set  $\mathcal{P}_{s_k} = \int s_k^2 g_{s_k}(s_k) ds_k$  % Eq. (14)
```

Non-Gaussian marginal distribution for correlated sources

In the previous section it was assumed that the sources are independent, in practical applications however, this assumption may not be upheld. Here, we provide a framework that can be used for pdf estimation of correlated sources.

Suppose that we are interested in the activity of a set of voxels in the brain $\{s_{m1}, \dots, s_{mp}\}$, which are not necessarily independent from each other. Similar to the application of the null-beamformer, we assume that the location of the correlated sources is known. For example, in an auditory experiment, we roughly know that the two sources are located in the left and right auditory cortices ([Dalal et al., 2006](#); [Popescu et al., 2008](#)).

Let the lead-field matrix of the correlated sources be $\mathbf{F}_m = [\mathbf{f}_{m1} \dots \mathbf{f}_{mp}]$, and let $\mathbf{s}_m = [s_{m1} \dots s_{mp}]^T$. Our aim is to estimate the posteriori distribution and pdf of the k th source in the first set: $p(s_{mk}|\mathbf{y}_t)$ and $g_{s_{mk}}(s_{mk})$. Using the same argument given for derivation of Eq. (11), one may estimate the joint pdf of the posterior distribution of the correlated sources $\{s_{m1}, \dots, s_{mp}\}$ using:

$$p(s_{m1}, \dots, s_{mp}|\mathbf{y}_t) = p(\mathbf{s}_m|\mathbf{y}_t) \propto g_{\mathbf{y}}(\mathbf{y}_t - \mathbf{F}_m \mathbf{s}_m).$$

Having the joint distribution of dependent variables, the marginal distribution (obtained by integrating over other variables) can be employed to estimate the pdf of interest $p(s_{mk}|\mathbf{y}_t)$ regardless of their correlation, or precisely

$$p(s_{mk}|\mathbf{y}_t) \propto \int g_{\mathbf{y}}(\mathbf{y}_t - \mathbf{F}_m \mathbf{s}_m) d\mathbf{s}_{mk/} \quad (15)$$

where $\mathbf{s}_{mk/}$ is a vector that is obtained by eliminating the k th entry of \mathbf{s}_m . The time-series can thus be considered as the expected value of the conditional density of the desired source; $\hat{s}_{mk,t} = \int s_{mk} p(s_{mk}|\mathbf{y}_t) ds_{mk}$.

Following a similar approach to the estimation of the source pdf for independent sources (Eq. (13)), the pdf of source of interest can be obtained using

$$g_{s_{mk}}(s_{mk}) \propto \int g_{\mathbf{y}}(\mathbf{F}_m \mathbf{s}_m) d\mathbf{s}_{mk/}. \quad (16)$$

Algorithm 2 presents the pseudo-code for the marginal distribution method. This algorithm is similar to [Algorithm 1](#), but includes the extra steps of estimating the joint distribution and its marginalisation. In [Algorithm 2](#), the indices k and m are dropped for notational simplicity. This code shows how to estimate the power \mathcal{P}_s and time-series \hat{s}_t of a source with lead-field \mathbf{f} , which is not independent from the sources augmented in the vector \mathbf{s}_j . Note that the lead-fields of the sources in vector \mathbf{s}_j are augmented in the same order in the matrix \mathbf{F}_j . In this algorithm, $g_{s,s_j}(s, \mathbf{s}_j)$ is the joint pdf of the desired source and other correlated sources. Normalisation of the pdf can be applied after the marginalisation to slightly reduce the computational cost.

Algorithm 2. Implementation of the non-Gaussian MD Beamformer

```
% estimating the time-series  $\hat{s}_t$ 
for  $t = 1$  to  $T$  do
    estimate  $g_{\mathbf{y}}(\mathbf{y})$ 
    set  $p(s, \mathbf{s}_j|\mathbf{y}_t) \propto g_{\mathbf{y}}(\mathbf{y}_t - \mathbf{f}s - \mathbf{F}_j \mathbf{s}_j)$ 
    marginalise  $p(s|\mathbf{y}_t) = \int p(s, \mathbf{s}_j|\mathbf{y}_t) d\mathbf{s}_j$ 
    normalise  $p(s|\mathbf{y}_t) = \frac{p(s|\mathbf{y}_t)}{\int p(s|\mathbf{y}_t) ds}$ 
    set  $\hat{s}_t = \int s p(s|\mathbf{y}_t) ds$ 
end for

% fast estimation of the power  $\mathcal{P}_s$  for the source  $s$  with lead field  $\mathbf{f}$ 
% which is not independent from  $\mathbf{s}_j$  with lead-field matrix  $\mathbf{F}_j$ 
estimate  $g_{\mathbf{y}}(\mathbf{y})$ 
set  $g_{s,s_j}(s, \mathbf{s}_j) \propto g_{\mathbf{y}}(\mathbf{f}s + \mathbf{F}_j \mathbf{s}_j)$ 
marginalise  $g_s(s) = \int g_{s,s_j}(s, \mathbf{s}_j) d\mathbf{s}_j$ 
normalise  $g_s(s) = \frac{g_s(s)}{\int g_s(s) ds}$ 
set  $\mathcal{P}_s = \int s^2 g_s(s) ds$ 
```

Relation to the null-beamformer

The null-beamformer is an effective method to remove sources correlated with the source of interest. Assume again that the unwanted sources have lead-field matrix \mathbf{F}_j and the source of interest has lead-field \mathbf{f} . The formulation is the same as that of the LCMV-beamformer, except that it has an additional constraint that nulls the interference through enforcing $\mathbf{w}^T \mathbf{F}_j = 0$, where 0 is a vector of zeros. Therefore,

$$\underset{\mathbf{w}}{\text{argmin}} \mathbf{w}^T \mathbf{R}_y \mathbf{w}, \quad \text{subject } E: \mathbf{w}^T \mathbf{f} = 1 \text{ and } \mathbf{w}^T \mathbf{F}_j = 0. \quad (17)$$

The closed-form solution of \mathbf{w} is again obtained using the Lagrange multiplier method, whereby the estimated time-series and power are given by:

$$\hat{s}_t = \mathbf{c}^T (\mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j)^{-1} \mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{y}_t \quad (18)$$

$$\mathcal{P}_s = \mathbf{c}^T (\mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j)^{-1} \mathbf{c} \quad (19)$$

where $\mathbf{F} = [\mathbf{f} \mathbf{F}_j]$ and $\mathbf{c} = [1, \mathbf{0}]^T$.

The solution proposed in the non-Gaussian MD beamformer is the same as the null-beamformer (Eqs. (18) and (19)), if a Gaussian pdf is fitted to the measurements. This is proved in [Appendix B](#), where we show that if $g_{\mathbf{y}}(\cdot)$ is a zero-mean Gaussian distribution with covariance

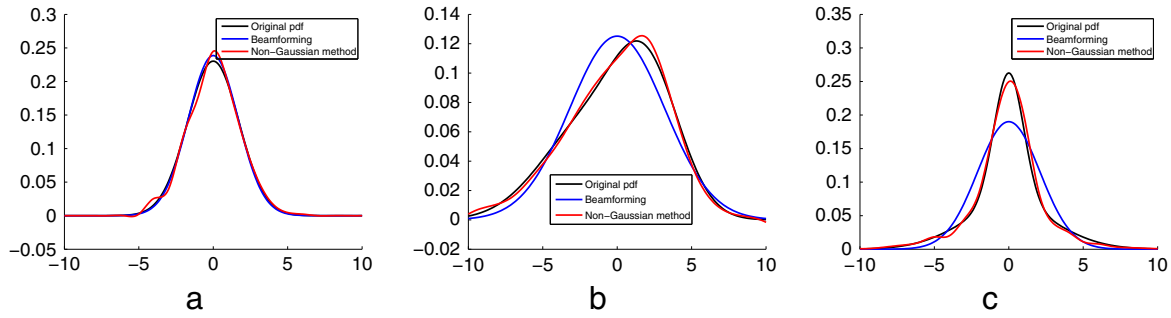


Fig. 1. Examples of the simulated source pdfs (black lines) and their estimation using the LCMV beamformer (blue lines) and the non-Gaussian method (red lines).

matrix \mathbf{R}_y , then the conditional density $p(s|\mathbf{y}_t)$ given by Eq. (15) is also Gaussian with mean value given by Eq. (18) and covariance matrix given by Eq. (19). Furthermore, the marginalised distribution $s_g(s)$ given by Eq. (16) is zero-mean Gaussian with covariance matrix equivalent to Eq. (19).

It is notable that the non-Gaussian MD source localisation is also related to the beamformer with general linear constraint (Van Trees, 2004) under the Gaussianity assumption (see Appendix C for further details).

A kernel based estimation of the measurement pdf

We have shown how the source pdf can be estimated from the data pdf. Thus, to implement the method, estimation of the measurement pdf $g_y(\cdot)$ is needed from the set of discrete observations \mathbf{y}_t , $t \in \{1, \dots, T\}$. The success of the method depends somehow on the quality of this estimation.

Since the dimension of \mathbf{y}_t is large (on the order of hundreds), we employ kernel based methods, which are fast estimators requiring no optimisation procedure. Kernel based estimation, which is one of the most common methods, assumes that the data pdf may be modelled as the sum of T kernels centred at each observation (Silverman, 1999).

Eq. (13) shows that we require $g_y(\mathbf{f}s)$ to be able to estimate the source pdf, which can be obtained using the kernel density estimator as:

$$g_y(\mathbf{f}s) = \sum_{t=1}^T \frac{1}{h^N} K\left(\frac{\mathbf{y}_t - \mathbf{f}s}{h}\right) \quad (20)$$

where $K: \mathbb{R}^N \rightarrow \mathbb{R}$ is the kernel, h is a scaling factor known as bandwidth, N is the number of sensors, and h^N indicates h to the power

N . This equation implies that constructing the whole measurement pdf $g_y(\mathbf{y})$ for any given value of \mathbf{y} is unnecessary, and instead, we only need to estimate $g_y(\mathbf{f}s)$ which is a univariate function. Therefore, the pdf at a location with lead-field matrix \mathbf{f} , based on the Eqs. (13) and (20) is given by:

$$g_s(s) = \frac{\sum_{t=1}^T \frac{1}{h^N} K\left(\frac{\mathbf{y}_t - \mathbf{f}s}{h}\right)}{\int \sum_{t=1}^T \frac{1}{h^N} K\left(\frac{\mathbf{y}_t - \mathbf{f}s}{h}\right) ds} \quad (21)$$

Note also that the same kernel based estimation can also be employed for the non-Gaussian method for correlated sources using Eq. (16).

To better clarify the algorithm, suppose that at time instant t and at a location with lead-field \mathbf{f} , we are interested in estimating the source pdf and consequently the source time-series s_t . The source pdf is approximated by a set of discrete points within a limited interval (e.g., for s from -10 to 10 with an increment of 0.1), and the value of each point is estimated using Eq. (20). This gives us a scaled version of the source pdf and should be normalised by its integral. The integral is calculated numerically based on the above set of points, resulting in an estimation of the source pdf. This procedure may be repeated for each location and for each time instant.

The kernel itself and the bandwidth both need to be chosen; we use the common Gaussian kernel, which is given by (Silverman, 1999):

$$K(\mathbf{x}) \propto \exp\left(-\frac{1}{2} \mathbf{x}^T \mathbf{R}_{ker}^{-1} \mathbf{x}\right) \quad (22)$$

where \mathbf{R}_{ker} is the covariance of the kernel.

The choice of covariance matrix of the kernel is important in determining the spatial resolution. The simplest choice is the identity matrix

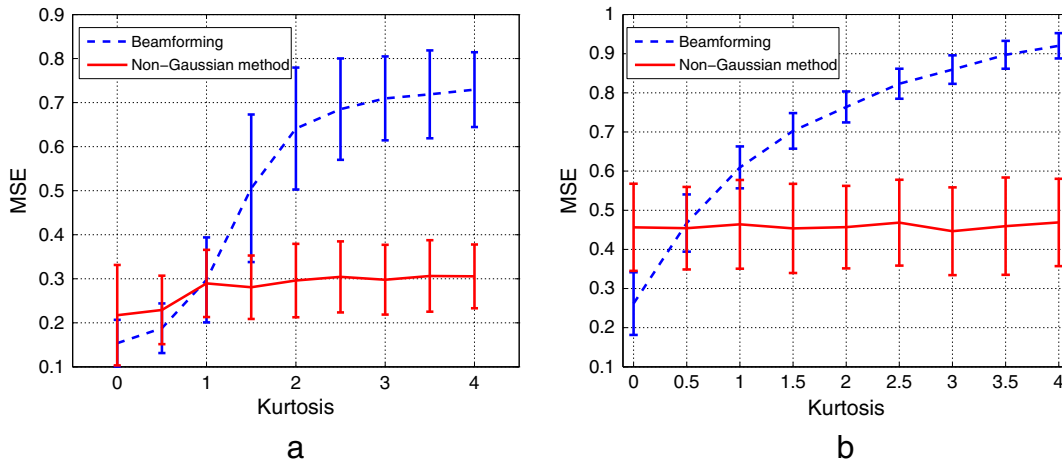


Fig. 2. Comparison between the LCMV beamformer and the non-Gaussian method in the simulated data when (a) one of the sources is non-Gaussian and (b) the additive noise is non-Gaussian. The method proposed is insensitive to the non-Gaussianity of the source plus noise.

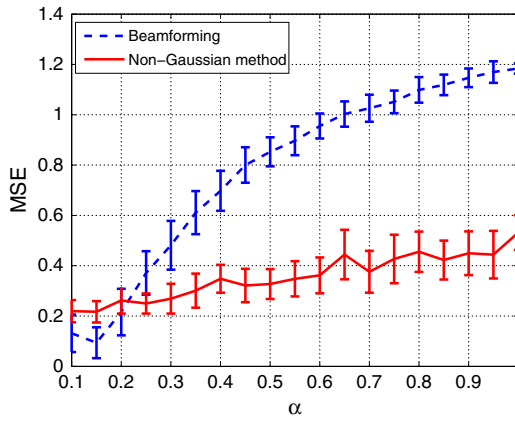


Fig. 3. MSE versus correlation between sources for the proposed method and for the classic LCMV beamforming approach in a simulation experiment. The LCMV beamformer is more sensitive to the correlation between sources than the proposed method.

$\mathbf{R}_{ker} = \mathbf{I}$, which results in a homogeneous kernel. An alternative efficient choice of covariance matrix of the kernel can be made with the eigenvectors associated with the largest eigenvalues of the measurement covariance matrix. To construct this kernel suppose that $\mathbf{R}_y = \mathbf{U}\Sigma\mathbf{U}^T$ is the singular value decomposition of the estimated measurement covariance matrix, and set

$$\mathbf{R}_{ker} = \mathbf{U}_{max}\mathbf{U}_{max}^T + \sigma\mathbf{I}$$

where \mathbf{U}_{max} is the matrix containing the eigenvectors associated to the largest eigenvalues, and σ is a regularisation parameter used to stabilise the calculation of the inverse, since $\mathbf{U}_{max}\mathbf{U}_{max}^T$ is rank deficient. This choice of kernel is especially helpful in our study, in which we employed the signal space separation artefact rejection algorithm (Taulu and Simola, 2006). This filtering reduces the rank of the data from number of sensors N to a much smaller number (order of tens). This kernel, which is homogeneous in the space that the data is distributed and is zero elsewhere, can improve the data pdf estimation and therefore results in a better source pdf estimation. In the above equation, \mathbf{R}_{ker} is $N \times N$, where N is the number of sensors, and \mathbf{U}_{max} is $N \times r$, where r is the rank of the measurement after applying the signal space separation filter. This choice of kernel can also be useful when some of the eigenvalues of data covariance matrix are considerably smaller than the

others, as is also the case with the multiple signal classification (MUSIC) algorithm (Moshier et al., 1992).

Computational complexity

Kernel based estimators are almost asymptotically unbiased for a large number of measurement samples (see Appendix D), however, the algorithm is computationally intensive if the number of measurement samples is large. The computational complexity of the proposed algorithm for calculation of the time-series and power at a single voxel is of order $O(T^2N^2D_s)$ and $O(TN^2D_s)$, respectively, where D_s is the number of points that are used to approximate the distribution of one dimensional source. Depending on the number of time samples T and number of points D_s , these orders are generally much bigger than the computational complexity of the LCMV beamformer which is of order $O(TN^2)$ for calculation of time-series and $O(N^2)$ for calculation of power.

For example, estimation of the time-series at each voxel, based on a segment of 30-second MEG data with sampling frequency of 250 Hz, requires approximately 7 s, using a desktop computer with 2.93 GHz CPU. This means that approximately 3 h is required to estimate the time-series for a set of voxels within the brain with a grid resolution of 12 mm. Compared to the standard beamformer, which needs few seconds to scan the whole brain, the computation time is substantially longer. However, given the fact that MEG source localisation is normally implemented offline and the algorithm may be readily parallelised, this high complexity should not present a barrier to its use.

Experimental results

Two sets of results are described. The first set uses simulations to investigate and quantify the errors of the methods proposed resulting from non-Gaussian sources and noise. The second set investigates the methods on real MEG data for resting state, auditory and visual stimuli. In each case the methods are compared with the LCMV and null beamformer.

Simulation experiments

To evaluate the new methods, we ran several simulation experiments. Each used a single layer realistic head model in which the brain was divided into a number of cells. The distance between adjacent cells was 5 mm. The number of samples was 400 and the number of sensors was set to 102 following the number of magnetometer sensors used in our MEG scanner.

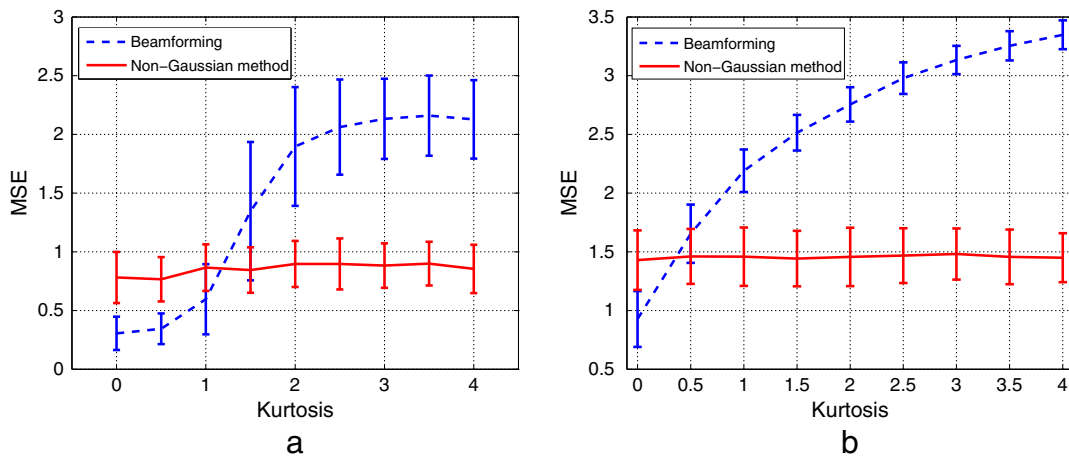


Fig. 4. Comparison between the classic null-beamformer and the proposed non-Gaussian method based on the marginalised distribution for highly-correlated sources when (a) one of the sources is non-Gaussian and (b) the additive noise is non-Gaussian. The method proposed is insensitive to the Gaussianity of the source plus noise.

The data were normalised such that their maximum value was equal to one. In the non-Gaussian method the pdf of each source was estimated over a compact interval between -10 and 10 . In all the simulations, the value of the bandwidth in the non-Gaussian method was set to $h = 1$, and the regularisation factor in the beamformer approach was set to $\lambda = 0.01\text{Tr}\{\mathbf{R}_y\}$ (1% of the trace of the measurement covariance matrix). Note that as long as the covariance matrix is full rank and not nearly rank deficient, the small value of λ has little impact on the results.

The first experiment illustrates the estimation of the source pdf for a single non-Gaussian source using the non-Gaussian PD beamformer. The next two examine quantitatively the use of this method on non-

Gaussian noise and correlated sources respectively. In each case the results are compared with the LCMV beamformer. The experiment in the [Investigation of null-beamformer and its non-Gaussian extension](#) section compares the performance of the non-Gaussian MD beamformer with the LCMV null-beamformer on the same data. The [Reconstruction of simulated sources](#) section extends the investigation of the non-Gaussian PD beamformer to image reconstruction for a set of uncorrelated Gaussian sources with non-Gaussian noise. The results from the LCMV beamformer and the SAKETINI method (Zumer et al., 2007) are also provided. Finally, in the [Reconstruction of simulated correlated sources](#) section, we show the reconstruction of the time-series using

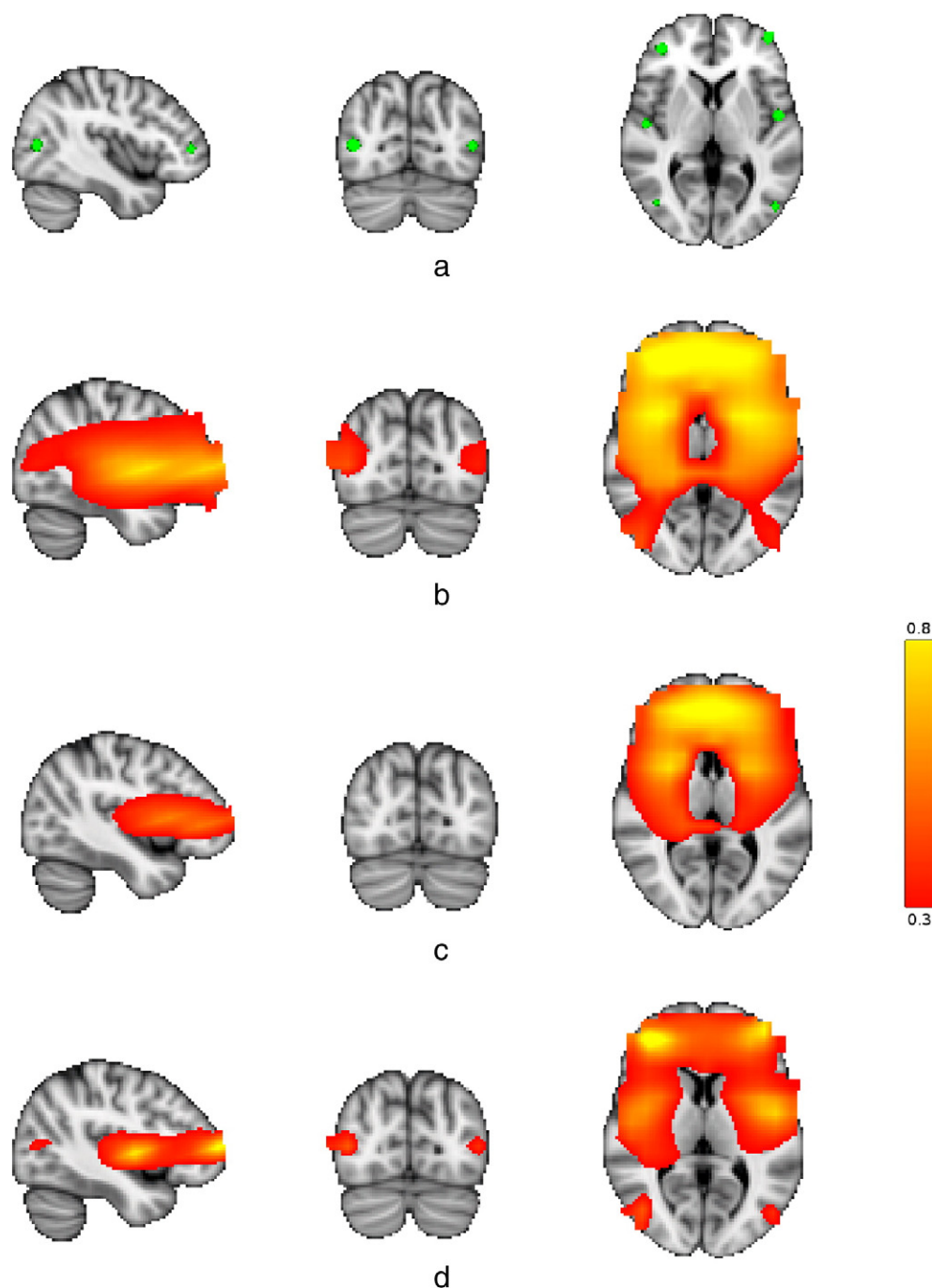


Fig. 5. An example for reconstruction of six simulated dipoles, (a) original location of dipoles, (b) estimated locations using the LCMV beamformer and (c) estimated locations using the SAKETINI method (Zumer et al., 2007) and (d) estimated locations using the proposed non-Gaussian approach. The non-Gaussian method reconstructed six sources more accurately than the other two methods which assume that the data is Gaussian.

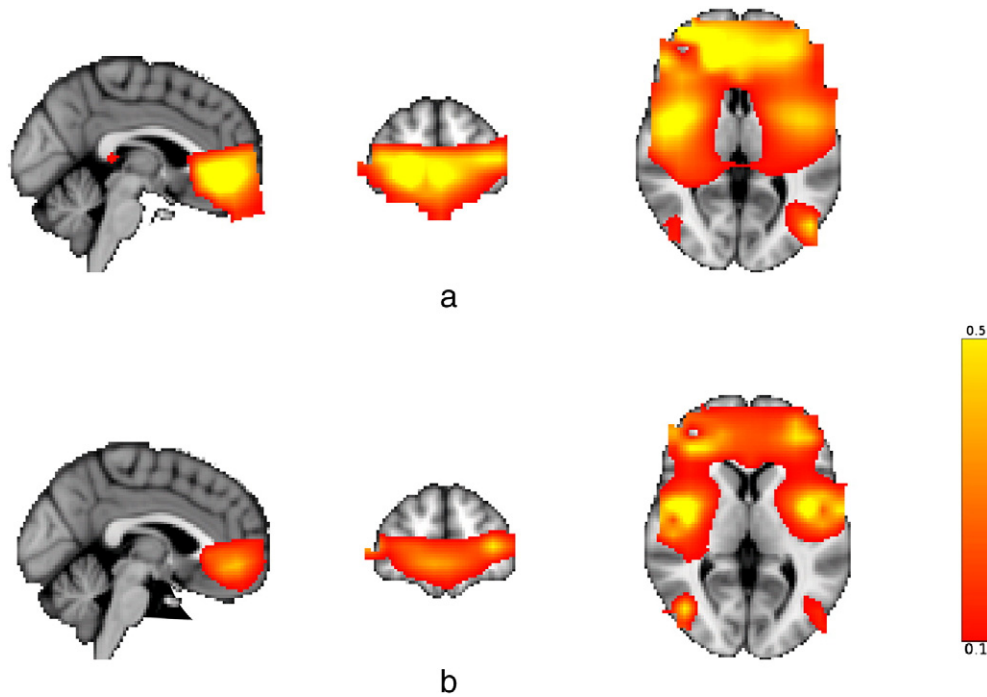


Fig. 6. An example of the reconstruction of six simulated dipoles when the two frontal sources are correlated using (a) the LCMV null beamformer and (b) the non-Gaussian method, while cancelling the left frontal source. The beamformer approach shows spurious activity between the two sources due to the correlation between sources, but the non-Gaussian approach shows more accurate results. There is activity around the left frontal source, using both approaches, because of the additive noise.

the non-Gaussian PD beamformer for non-Gaussian sources, comparing the results with the LCMV beamformer and Bayesian PCA method (Woolrich et al., 2011) as appropriate.

Impact of non-Gaussian source

In the first simulation experiment, the method was demonstrated through the reconstruction of a single source. Three different pdfs were assigned to the source in turn, the first Gaussian and the others non-Gaussian (described further in the next experiment). It was assumed that in addition there were 15 uncorrelated (Gaussian) sources whose locations were randomly allocated inside the brain volume. The pdfs of the source of interest are shown in Fig. 1. Zero-mean Gaussian white noise was also added to the simulated MEG data (after projection of the simulated sources through the lead-fields) to set $\text{SNR} = 8$ dB. SNR is defined in the sensor space as the ratio between mean power of the signal to the mean power of the noise across all sensors. The pdf of the first source was estimated using both Gaussian and non-Gaussian assumptions. Fig. 1(a) shows

the result for the Gaussian source where both beamformers give an accurate estimation. Figs. 1(b) and (c) show the results for two non-Gaussian sources. It is clear that the non-Gaussian PD beamformer increasingly outperforms the LCMV beamformer as the source pdf deviates more from a Gaussian distribution.

In the second simulation experiment, the mean squared error (MSE) resulting from non-Gaussianity of one source was quantified using a Monte Carlo simulation. This source was generated using a mixture model consisting of two zero-mean Gaussian components. The departure from Gaussianity of the source was quantified using the kurtosis k which is defined by $k = \frac{\mu_4}{\sigma^4} - 3$, where μ_4 is the fourth moment and σ is the standard deviation. When the kurtosis is zero, the data is Gaussian, and as it becomes larger, the pdf departs further from a Gaussian profile. Only when the variances of the two pdfs in the mixture model are identical, is the mixture also Gaussian. Fig. 2(a) presents the result of the MSE which is defined as the mean squared error between the original and constructed pdf for different kurtosis. The average SNR is -5 dB and the results were obtained using 1000 Monte Carlo simulations. It

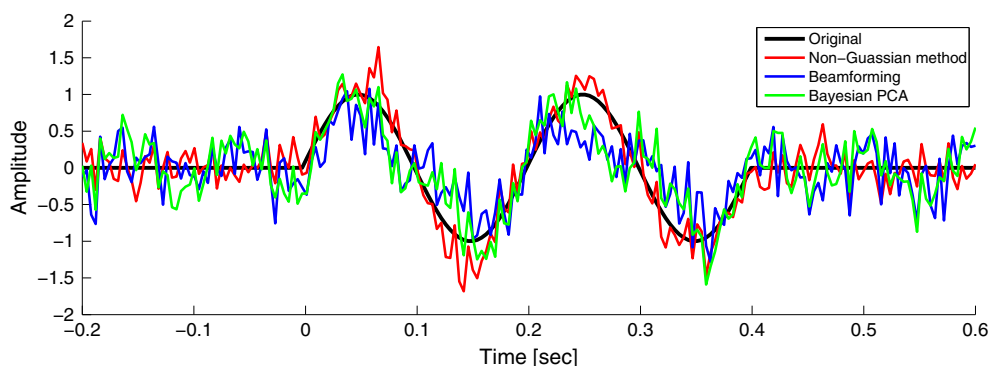


Fig. 7. Estimation of simulated time-series using the proposed non-Gaussian method (red), the classic LCMV beamformer (blue) and Bayesian PCA (green). The original waveform has been shown by a solid black line.

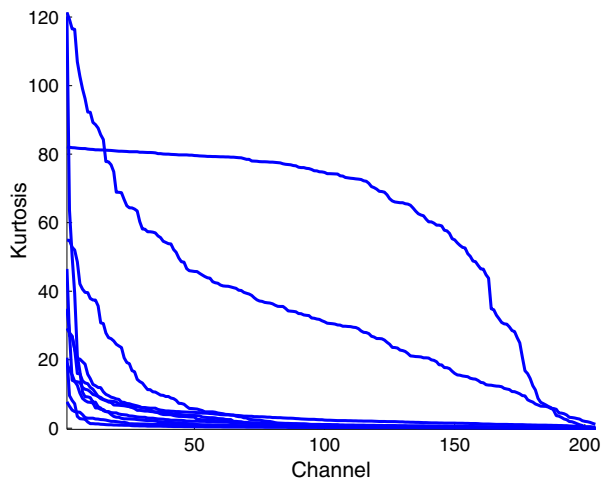


Fig. 8. Kurtosis versus channels in resting state experiments recorded from 10 subjects. Kurtosis can be very large in some cases indicating the high non-Gaussianity of the measured MEG.

is clear that as the kurtosis increases, the error in the LCMV beamformer increases rapidly. In contrast, the new non-Gaussian PD beamformer exhibits a flat error which means it is not sensitive to the shape of the pdf.

Impact of additive non-Gaussian noise

So far the effect of non-Gaussian sources has been considered where the noise was Gaussian. This simulation experiment was devoted to investigating additive non-Gaussian noise. In this case, all sources were Gaussian and the noise was generated as per the previous example by using a Gaussian mixture model with two zero-mean Gaussian components. The covariance matrix of one of the Gaussian components was the identity matrix \mathbf{I} and the covariance of the second Gaussian component was $\alpha\mathbf{I}$, where α was varied to gain different values of kurtosis. As demonstrated by Fig. 2(b), the non-Gaussian method is insensitive to the shape of the noise pdf whereas the MSE in the LCMV beamformer increases with increasing kurtosis.

Impact of correlation between sources

This experiment assesses the effect of correlated sources on the LCMV beamformer and the non-Gaussian PD beamformer. Here, the sources and the additive noise were all assumed to be Gaussian. Two correlated sources s_1 and s_2 were generated using standard Gaussian distribution: $s_1 \sim \mathcal{N}(0, 1)$, $s_2 \sim (1 - \alpha)\mathcal{N}(0, 1) + \alpha s_1$, where α is a measure of correlation between sources; $\alpha = 0$ means the sources are uncorrelated and $\alpha = 1$ means they are completely correlated. The locations of the two sources were randomly and uniformly set within the brain in the Monte Carlo simulation to estimate the average impact of the method over all locations. Fig. 3 shows that even for Gaussian sources the non-Gaussian method shows a better reconstruction of the desired source pdf than the LCMV beamformer as the correlation between sources increases. Even with moderate correlation between

sources, the MSE from the novel non-Gaussian method does not increase very significantly.

Investigation of null-beamformer and its non-Gaussian extension

Next, we investigated the performances of the null-beamformer and the non-Gaussian MD beamformer designed to cope with correlated sources using the same simulation configuration. The results are presented in Fig. 4. In each case, we assumed that the null location is known. Fig. 4(a) shows the results when one source is non-Gaussian and Fig. 4(b) shows the results when the additive noise is non-Gaussian. Both methods are capable of cancelling out the correlated source, but the non-Gaussian MD beamformer is better at dealing with the non-Gaussianity of the source and additive noise.

Reconstruction of simulated sources

In the next simulation experiment, we simulated six uncorrelated Gaussian sources with non-Gaussian noise to compare the source localisation accuracy of the LCMV beamformer and the non-Gaussian PD beamformer. The noise was created as in the other experiments with the kurtosis equal to 3. The location of the sources is shown in Fig. 5(a) and marked by green dots. The sources in this example were also reconstructed using the SAKETINI method (Zumer et al., 2007), which is a graphical model and assumes that a segment of the noise and the number of sources are known. The noise segment was generated with the same distribution as that of the additive noise used to simulate the measurements. The number of sources was set to 6 and the algorithm was implemented using the NUTMEG toolbox with the recommended settings (Dalal et al., 2004).

The results using the LCMV beamformer, SAKETINI and the non-Gaussian PD beamformer are shown in Figs. 5(c–d). The source activities estimated using the three methods have been normalised between 0 and 1. It is clear that the new method gives more accurate reconstruction, particularly, with higher spatial precision for frontal and temporal sources.

Reconstruction of simulated correlated sources

We also investigated the impact of correlated sources on the source reconstruction. We used the same simulation settings as the previous example, but the two frontal sources are chosen to be completely correlated, i.e., the same time-series. If we reconstruct these correlated sources without the nulling technique, a spurious source between the two correlated sources is estimated. We therefore assume that the location of one of the correlated sources (left frontal source) is known, and then place a null at its location. The results using LCMV and the non-Gaussian method are shown in Fig. 6, in which due to non-Gaussian noise, the LCMV beamformer shows inferior localisation compared to the non-Gaussian method.

Estimation of simulated time-series

In the last simulation experiment, we examined the accuracy of estimation of the underlying time-series. The location of the desired source is the right frontal source as shown in Fig. 5(a). The shape of the original time-series is shown in Fig. 7 with a solid black line. It consists of two cycles of a sinusoidal wave within a period of no activity to represent an event-related field potential. We also compare the proposed method with a recent method based on the Bayesian PCA for estimation of the



Fig. 9. Histograms of the first seven channels for data in a 100–150 ms interval after stimulus onset in a visual paradigm. The histogram should be in the form of Gaussian function (bell shape) to effectively employ the classical LCMV beamformer.

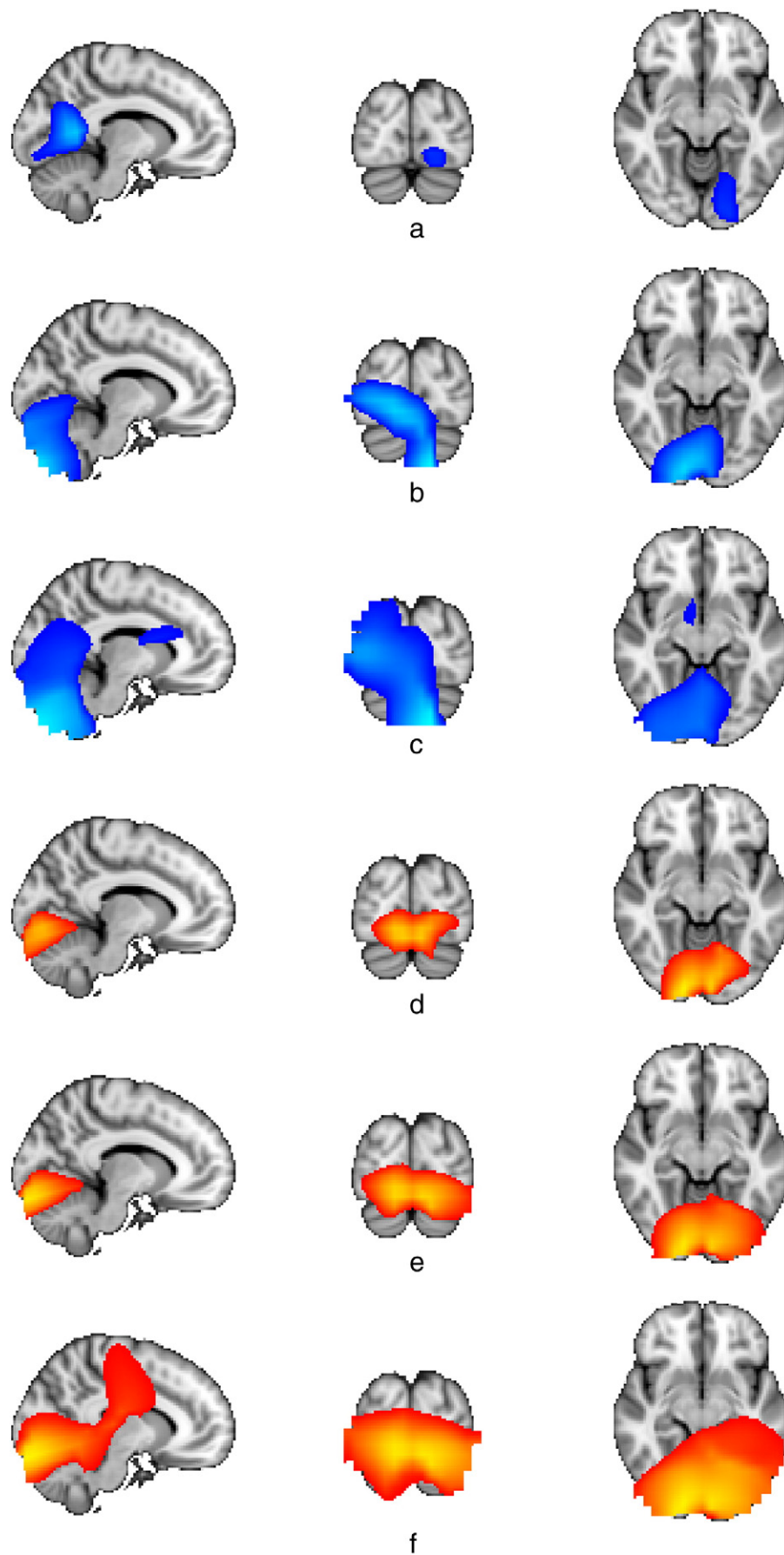


Fig. 10. Reconstruction of neural activity in the visual paradigm using beamforming with (a) $\lambda = 0.01\%$, (b) $\lambda = 0.01\%$ and (c) $\lambda = 1\%$ of trace of data covariance matrix, and using non-Gaussian method with (d) $h = 1$, (e) $h = 5$ and (f) $h = 10$.

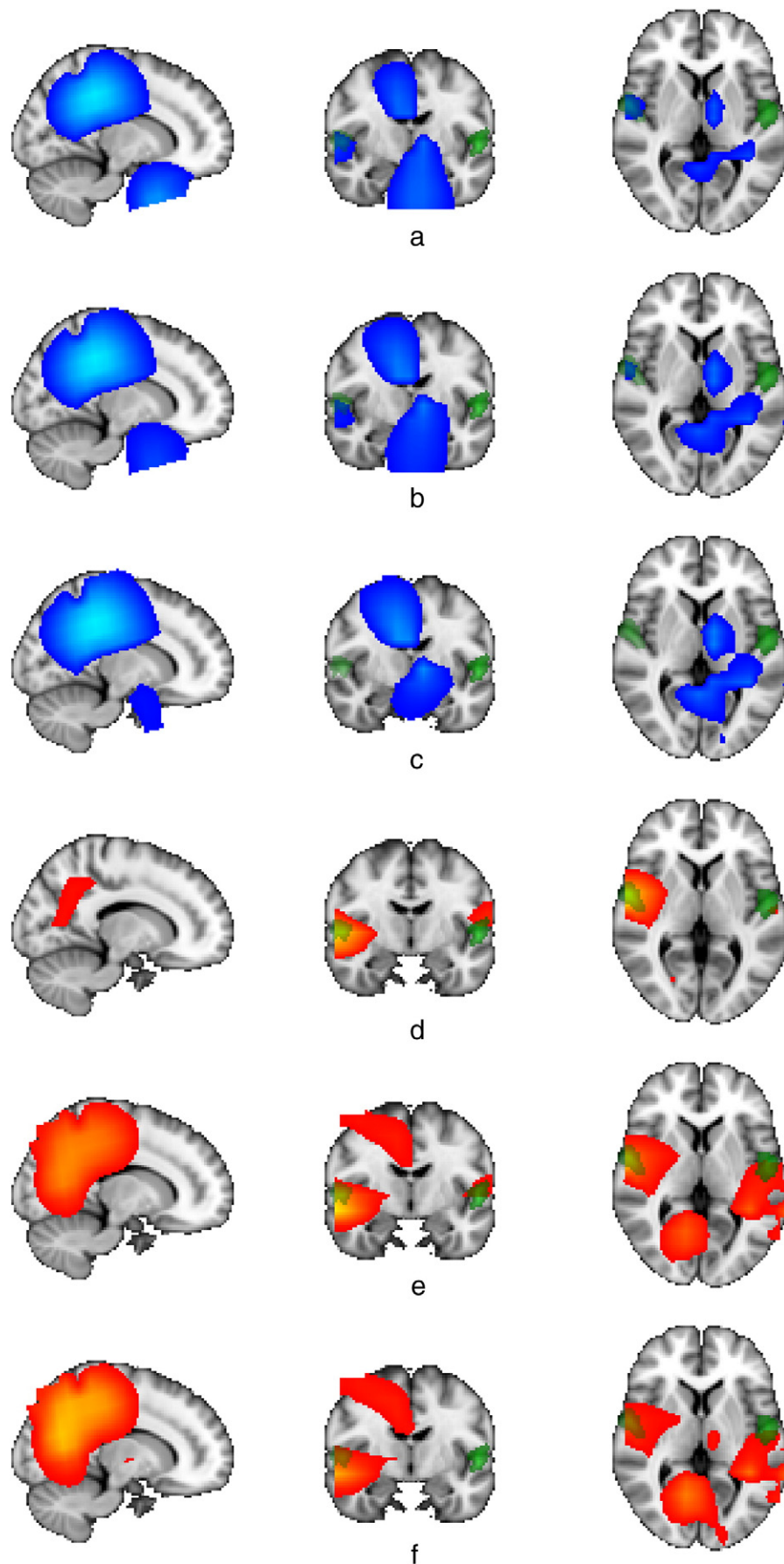


Fig. 11. Reconstruction of neural activity in the auditory paradigm using LCMV beamformer with (a) $\lambda = 0.01\%$, (b) $\lambda = 0.1\%$ and (c) $\lambda = 1\%$ of trace of data covariance matrix, and using the proposed non-Gaussian method with (d) $h = 1$, (e) $h = 5$ and (f) $h = 10$.

covariance matrix. This method particularly is useful when there are few sources and therefore the covariance matrix can be represented by a few eigenvectors (for further details please refer to (Woolrich et al., 2011)).

The estimated time-series using the non-Gaussian PD beamformer, the LCMV beamformer and Bayesian PCA were plotted using red, blue and green lines, respectively, in Fig. 7. These results imply that the non-Gaussian methods outperform the LCMV when the observations are non-Gaussian. The proposed method also outperforms the Bayesian PCA approach when the signal is corrupted with an additive non-Gaussian noise. This is because the Bayesian PCA only uses the second order statistics and therefore still assumes that the observations are Gaussian.

Experiments using MEG data

Data acquisition

Ethical approval of the research methods was obtained on 18 July 2008 from Oxfordshire Research Ethics Committee (reference: 08/H0604/58). The MEG data were recorded from three cohorts of 12 participants, all of whom were given informed written consent before taking part in the studies.

We compared the performance of the non-Gaussian method with the LCMV beamformers on data from three separate MEG experiments. The resting state experiment used ten right-handed participants (4 males, age 22–45 years) under the condition of eyes open with fixation. The visual experiment used a 27-year-old right-handed female with normal vision or vision corrected to normal. The auditory experiment used a 24-year-old right-handed participant with normal hearing.

All MEG recordings were performed using a 306 channel Elekta Neuromag system with 102 magnetometers and 102 pairs of planar gradiometers. Data were recorded at a sampling rate of 1000 Hz with a 0.1 Hz high pass filter. Prior to acquisition, a three-dimensional digitizer (Polhemus Fastrack) was used to localise the participants' head shape relative to the position of the head-coils, with respect to three anatomical landmarks which could be registered on the MRI scan (the nasion, and the left and right preauricular points). A structural MRI was also acquired for each participant.

A single layer realistic head model was used in the source analysis of both the visual and auditory experiments. Head movements were kept to a minimum, and head positions were localised immediately before the start of the experiment. Signal space separation or MaxFilter, which is a commercial pre-processing software package recommended by the Elekta company, was applied to the continuous data (Taulu and Simola, 2006). The continuous measurements were then linearly filtered in a bandpass range of 1–40 Hz. The data were segmented and a baseline correction applied (using a 200 ms segment of data before the stimuli). Finally, the trials were visually inspected and those with large variations were removed.

Investigation of Gaussianity of MEG data

In the first experiment, we examined the Gaussianity of real MEG data recorded from ten normal participants in the resting state through estimating the kurtosis. The kurtosis of the gradiometer signals is plotted in Fig. 8 (sorted in order of descending kurtosis for each subject for clarity). It is clear that even for resting state data, the kurtosis in some cases can become very large meaning that the data pdf is highly non-Gaussian. As a further test, the Jarque–Bera goodness-of-fit test (Jarque and Bera, 1987) was applied to both skewness and kurtosis. The test implied that only two channels in one subject and one channel in another subject were Gaussian distributed with $p < 0.05$. We conclude therefore that an assumption of Gaussianity is not valid for real MEG data.

In the following experiments, we compared the performance of the methods for source reconstruction and investigated the sensitivity of

the results with regard to two key parameters λ and the kernel bandwidth h .

Visual paradigm

The second MEG experiment used a visual paradigm where a series of human and animal faces were presented to the participant. Each image was presented for 300 ms and the time interval between images was 1500 ms. We were interested in localising the first significant peak of averaged epoched data which occurs around 100–150 ms after the onset of visual stimuli. The origin of this peak is known to be located in bilateral regions of the primary visual cortex. Fig. 9 shows the histogram of this peak for the first five channels; none matches a Gaussian distribution. Based on the results from the simulation experiments, we therefore would expect the non-Gaussian method to outperform the standard beamforming approaches.

This was confirmed in Fig. 10. Figs. 10(a)–(c) show the output power of the beamformer (trace of the estimated covariance matrix as in Eq. (4)) normalised by the norm of the associated lead-fields; i.e., normalised by the power of projected white Gaussian noise (see (Van Veen et al., 1997)). Figs. 10(d)–(f) show the estimated power using the non-Gaussian PD beamformer, which was also normalised by the norm of the associated lead-fields. In both methods, the same anatomical plane is displayed and it includes the results for different values of λ and h . For the LCMV beamformer, we see from the figure that if a small value for λ is chosen, the peak of the spectrum is in a frontal part of the visual cortex and if a large value is chosen, the peak of reconstructed source is in the cerebellum and not in the visual cortex. In contrast the non-Gaussian method reconstructs activity in both right and left primary visual cortex for moderate and small values of h . It also has better spatial resolution.

Auditory paradigm

The third MEG experiment was particularly useful to compare the performance of both sets of beamformers: the LCMV beamformer with the non-Gaussian PD beamformer and the null-beamformer with the non-Gaussian MD beamformer. It used a mismatch negativity auditory paradigm, where participants were presented with trains of 7 different tones repeated randomly. The frequency of the tones increased from 500 to 800 Hz in steps of 50 Hz. The number of times that the same tone was presented in the sequence varied pseudo-randomly between one and eleven. The probability that the same tone was presented once or twice was 2.5%, three and four times was 3.75% and five to eleven times was 12.5%. Stimuli were presented binaurally via headphones for 15 min. The duration of each tone was 70 ms, with 5 ms rise and fall times, and the inter-stimulus interval was 500 ms. Here, the first tone of each train is considered as deviant and the rest as the standard tone. About 250 deviant trials were presented.

We were only interested in localising a significant peak known as N100 which occurs 100 ms after the stimulus onset. The source of this peak is known to be in bilateral regions of the primary auditory cortex. Traditional beamforming methods have been shown to have difficulty in reconstructing these bilateral sources correctly.

The results of the source reconstruction are presented in Fig. 11 using the non-Gaussian PD and LCMV beamformers. In all the figures, the small green volume is the mask representing the primary auditory cortex presented using the Juelich Histological Atlas (Morosan et al., 2001). Figs. 11(a)–(c) show the results using the LCMV beamformer (blue regions) with λ equal to 0.01%, 0.1% and 1% of the trace. It is clear that, depending on the value of λ , the LCMV beamformer has only reconstructed a source primarily in the left or right auditory cortex, leaving out the source in the other cortex. In addition, it has placed an extraneous source in the middle of the brain (close to the precuneus and posterior cingulate cortex), and one deep in the brain. These extraneous sources are likely to arise from correlations between genuine sources. Figs. 11(d)–(f) show the results from the non-Gaussian method for $h = 1, 5$ and 10, respectively. It shows better reconstruction

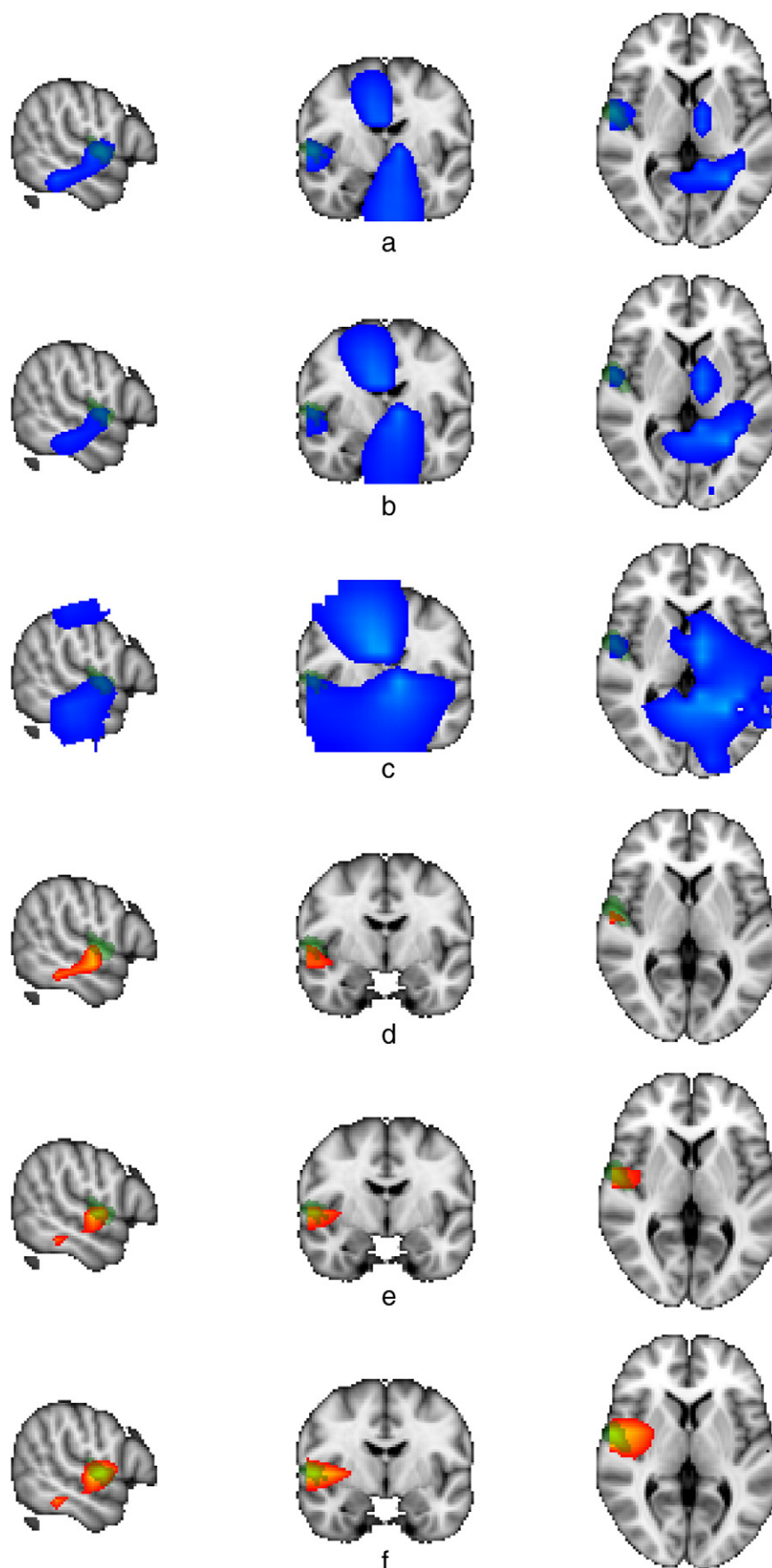


Fig. 12. Reconstruction of neural activity in the auditory paradigm when the right-hand side source is cancelled out using the classic null-beamformer with (a) $\lambda = 0.01\%$, (b) $\lambda = 0.1\%$ and (c) $\lambda = 1\%$ of trace of data covariance matrix, and when the right-hand side source is cancelled out using the proposed marginalised based non-Gaussian method with (d) $h = 1$, (e) $h = 5$ and (f) $h = 10$.

of both left and right sources in the primary auditory cortex (especially for $h = 5$). It does not show any source deep in the brain, but for higher values of h it does present a potentially incorrect source in the middle of the brain due to the high correlation between sources. These observations also show that the sensitivity of the new method to the choice of bandwidth h was no higher than the sensitivity of the LCMV beamformer to regularisation factor λ .

Since the two sources are partially correlated, the classic null-beamformer and the non-Gaussian MD method were employed to cancel out one of the correlated sources. Figs. 12(a)–(c) show the results from the classic null-beamformer when a null has been placed at the right auditory cortex with three different values of λ . The null-beamformer shows activity near the left auditory cortex and has reduced the large source of activity in the middle of the brain for small values for λ but shows significant areas of extraneous activity as λ is increased. Figs. 12(d)–(f) show the equivalent results using the non-Gaussian approach. It is clear that the source in the left auditory cortex had been located and other extraneous sources reduced. The non-Gaussian method shows less sensitivity to its parameter h than the null-beamformer does to its parameter λ .

Fig. 13 presents similar results to the previous figure, but here the null has been placed in the left auditory cortex. The results of the null-beamformer with different values of λ are shown in Figs. 13(a)–(c). While the large sources of activity in the middle and deep structure of the brain have largely been suppressed, the source in the right auditory cortex is not at the correct location. In contrast, Figs. 13(d)–(f) demonstrate the superior results from the non-Gaussian method, which localises the source of activity very close to the expected location. The extraneous sources have also been completely removed.

Resting state

One of the resting state networks which has been well identified in both fMRI and MEG data is the sensorimotor network (see (Brookes et al., 2011) for more details). We use this network as a ground truth to further investigate the performance of the proposed method. A 30-second segment of raw MEG data were filtered in the beta band (13–30 Hz) and the time-series at each voxel were estimated. The correlation between the Hilbert envelope of the time course of the seed source, located at the right sensorimotor cortex, with the envelopes for the time courses at each voxel was then estimated. The results for the LCMV beamformer and the proposed non-Gaussian method are presented in Fig. 14. The activity at the left sensorimotor cortex, as reconstructed by the LCMV beamformer, does not show any correlation with the right sensorimotor cortex, while the activity, reconstructed by the non-Gaussian method, shows a correlation of about 0.4. This example also demonstrates the better performance of the proposed method over the standard LCMV beamformer.

Summary and discussion

In this paper, we introduced a probabilistic formulation for MEG source localisation which allows non-Gaussian measurements to be modelled. An extension of the method to handle correlated sources through the marginalisation of the estimated joint pdf was also presented. The implementation of the new method showed that it outperforms existing LCMV beamforming methods in terms of spatial and temporal source estimations in data from simulations and real MEG experiments. In particular, in the simulation experiments, we showed that the non-Gaussian methods are superior to the LCMV beamformers in coping with non-Gaussian pdfs for source and additive noise, and for correlated sources. In addition, the mean square error in estimating non-Gaussian time-series was also less. The improved results for simulation experiments were repeated when the method was applied to real MEG data. The non-Gaussian beamformer was again superior in

localising visual and auditory paradigms, where the latter is known to contain sources with high correlation.

Here, we used the kernel density estimator to estimate the data and consequently the source distribution. Based on Algorithm 1, it is clear that any other multivariate density estimator such as mixture of Gaussian functions or multivariate Edgeworth series (Davis, 1976) could also be employed. Analysis of the impact of these approaches on the source localisation is beyond the scope of this paper, and we plan to evaluate their performances in the near future.

One criticism of this approach is that estimation of high dimensional density normally is not feasible because of the lack of enough observations. However, we present the following arguments to show that kernel density estimators are suitable in MEG source localisation:

- Importantly, although in theory we assume that the multivariate data distribution $g_y(\mathbf{y})$ is known, in practice we do not estimate this multivariate distribution and only estimate a univariate function $g_y(\text{mathbfbf{y}})$ – see A kernel based estimation of the measurement pdf section and specifically Eq. (20). The reason for this simplification is that this method assumes, as the traditional beamformer, that all sources are independent and thus it is not required to estimate the inter-dependency of sources. However, it is notable that in the proposed approach for correlated sources, when we need to estimate the multivariate joint distribution, it is required that we estimate the inter-dependency of sources. In this case, the number of correlated sources to be estimated should not be overly high, because it may not be feasible to estimate a high dimensional distribution.
- Although we suggest the use of kernel density estimators in the application of the MEG source analysis, the problem of high-dimensionality in multivariate density estimators is not limited to this method and has been discussed in other applications. For instance, for the purposes of statistical discrimination, Scott (1992) and Scott and Sain (2004) argued that kernel methods are powerful tools even over dozens of dimensions. This is because if the bandwidth is very small, the kernel estimate is essentially a nearest-neighbour classification, and if the band-width is very large, the result is the Fisher's linear discriminant analysis. Thus, at the extremes, kernel density analysis mimics two successful algorithms and an appropriately chosen value for the bandwidth should outperform both algorithms. Here, we also observed that if we select the covariance of the kernel equal to the covariance of the observed data then the kernel density estimator and the LCMV beamformer show similar results.
- Other factors including SNR and correlation between sources, generally have more impact on a MEG source reconstruction algorithm than the number of observations. For example, only one sample is enough to exactly estimate the location of one source in a high SNR data set, while it is impossible to estimate the covariance matrix based on only one sample. On the other hand, if we have a very low SNR, the exact estimation of the covariance matrix or data distribution, obtained from a large number of observations, does not necessarily help to localise the source activity.

We may conclude from the above points that the ultimate aim of beamforming is source reconstruction and not estimation of the covariance matrix or data distribution. This has again been discussed in the statistical discrimination applications by Friedman (1997), who argued that the optimal bandwidth for kernel discrimination is normally much larger than for optimal density estimation. Therefore, the bandwidth or form of the kernel (or regularisation factor in the beamformer) should be chosen such that the best source localisation and not the best density (or covariance matrix) estimation is achieved.

Comparison to the other related methods

To date, several methods have been proposed to tackle or exploit the non-Gaussianity of the MEG data set. Specifically, Nagarajan et al. (2005, 2006) proposed a generative model for ERF data that uses pre-stimulus

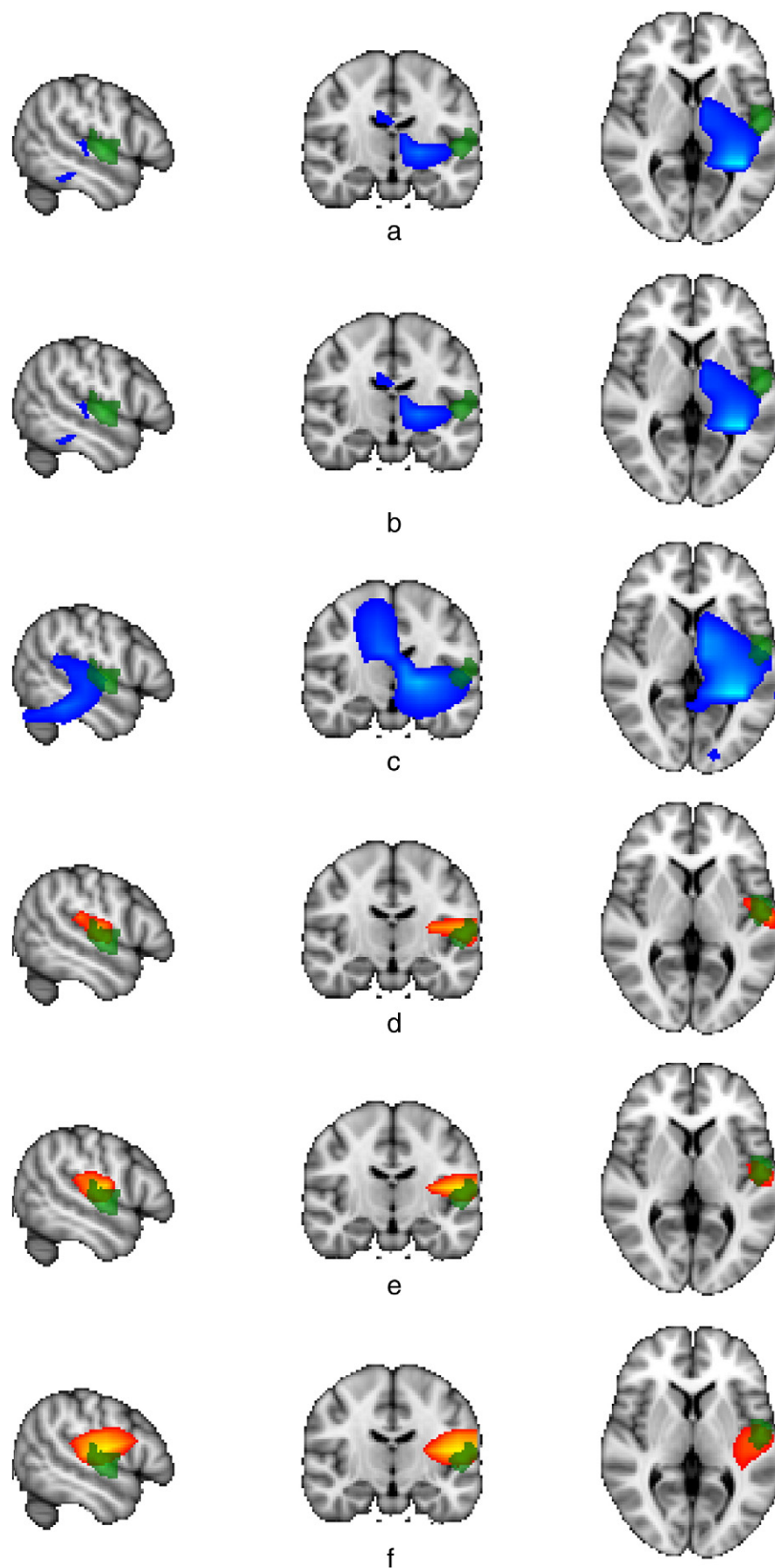


Fig. 13. Reconstruction of neural activity in the auditory paradigm when the left-hand side source is cancelled out using the classic null-beamformer with (a) $\lambda = 0.01\%$, (b) $\lambda = 0.1\%$ and (c) $\lambda = 1\%$ of trace of data covariance matrix, and using the proposed marginalised based non-Gaussian method with (d) $h = 1$, (e) $h = 5$ and (f) $h = 10$.

data segments to learn noise and interference components, thereby allowing for the rejection of noise and interference from post-stimulus segments. This method, which is based on factor analysis, models the source activity by a mixture of Gaussian functions to incorporate the non-Gaussianity of the evoked brain sources. Although this method assumes that the sources of interest are non-Gaussian, other distributions including the noise and artefacts are assumed to be Gaussian. Conversely in the approach proposed here we assumed that the measurements are non-Gaussian, which includes both the source and noise components. Moreover, in Nagarajan et al. (2005, 2006), it is assumed that the noise and artefact in the pre-stimulus and post-stimulus segments are statistically the same, whereas in the proposed approach we do not exploit information obtained from pre-stimulus segments. It is also clear that the method proposed in Nagarajan et al. (2005, 2006) can be used for the purpose of noise and artefact rejection before employing any source reconstruction method including our proposed approach.

The method described in Nagarajan et al. (2005, 2006) has been extended by Zumer et al. (2007, 2008) by incorporating the lead-field of the source of interest and also by including temporal basis functions for modelling the MEG signals. In contrast to the method in Nagarajan et al. (2005, 2006) and the method presented here, Zumer et al. (2007, 2008) assume that all distributions are Gaussian. This is partly to require fewer parameters to be estimated, which in turn leads to a robust estimation with less dependency on the initialisation step. They have also related their approach to the traditional beamformer and demonstrated that the beamformer is algebraically equivalent to the maximum likelihood approach based on a uniformly distributed prior (Zumer et al., 2007). Here we have verified this finding and used this result to extend the method to non-Gaussian distributions (see Eq. (8)). Moreover, the proposed non-Gaussian method maintains one of the potential benefits of the LCMV beamformer, in that it is less dependent on a generative model describing the MEG data as a function of dipole sources and their corresponding lead fields.

ICA is another method that uses the non-Gaussianity of MEG data and has been extensively employed for artefact rejection or dimensionality reduction. As in the method presented here, ICA assumes that the data is stationary but not Gaussian. It is also clear that, in contrast to

our method, ICA does not use spatial information, specifically the lead-field matrices, and also uses measures of the non-Gaussianity, such as kurtosis, rather than the data distribution.

Markov Chain Monte Carlo (MCMC) is another method that allows for non-Gaussianity (Jun et al., 2005). MCMC is an alternative to variational Bayes and its advantages over variational methods have been shown by Nummenmaa et al. (2007), when the posterior distribution is not uni-modal. As with our proposed method, MCMC uses sample points to estimate the distribution, which leads to a computationally intensive method. The main difference between MCMC and our approach is that MCMC assumes that the posterior distribution of the parameters of the model is non-Gaussian rather than the distribution of the measurements. This prevents the algorithm from being trapped in local minima, yet does not address the non-Gaussianity of the MEG signals.

In the case of measurement data from a large number of sensors, it has been argued that dimensionality reduction is often necessary (Zumer et al., 2007). The dimensionality reduction used here was performed using the signal space separation algorithm by projecting the data into the space that represents the internal volume of the sensors (see (Taulu and Simola, 2006) for more details). Other methods for dimensionality reduction as well as noise and artefact rejection such as PCA and ICA, can also be employed to improve the current approach. PCA may fail in low SNR data and give erroneous results, and other approaches like factor analysis (Nagarajan et al., 2006) and Bayesian PCA (Woolrich et al., 2011) seem more appropriate, where a robust estimate of the noise covariance is made.

Finally, we assumed that the prior of the source is uniformly distributed (no prior information on s) which leads to a simple form of posterior distribution. This is in line with several other methods in the estimation theory literature including pseudo-inverse, maximum likelihood, LCMV beamformer, particle filters, and minimum norm source estimation. Nevertheless, the proposed method can be extended by using a prior obtained from anatomical information. One way, as described in Limpiti et al. (2006), is to use a cortical patch as a source model for representing spatially distributed neural activity based on a set of basis functions. Similarly, we may also incorporate prior assumptions about the source configuration obtained

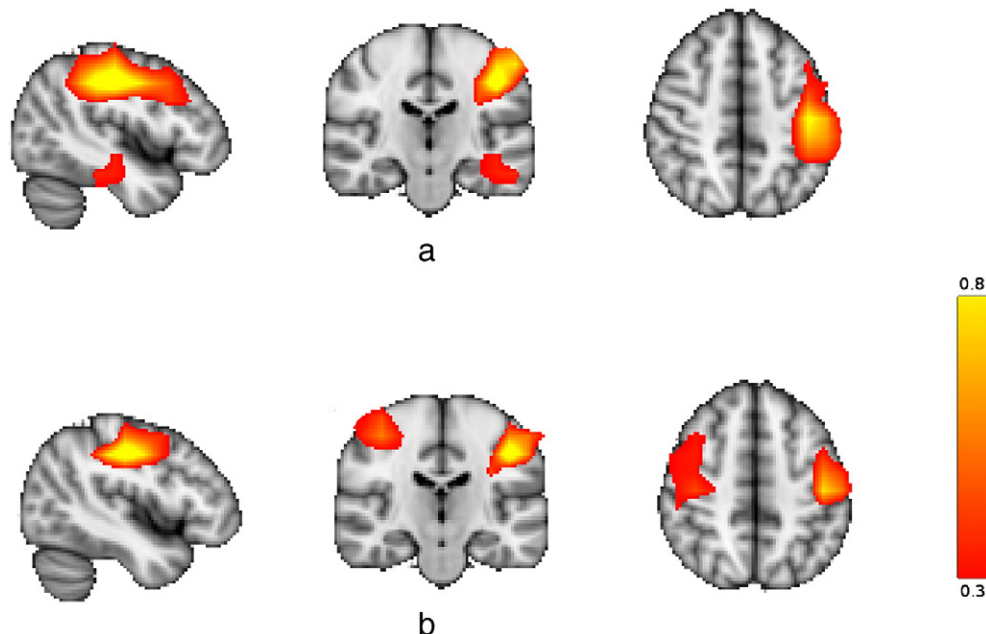


Fig. 14. Correlation between the Hilbert envelope of the right sensorimotor cortex and all other voxels for MEG data filtered within the beta band, (a) results from the LCMV beamformer and (b) results from the non-Gaussian method. The non-Gaussian method shows a correlation with the left sensorimotor cortex, while the LCMV beamformer does not show any correlation.

from structural MRI. Such models have already been developed in a Bayesian framework in the context of distributed source localisation (Mattout et al., 2006).

The method introduced in this paper has shown clear advantages over the LCMV beamformer, which uses second order statistics to represent non-Gaussian data. Through a stochastic formulation and kernel density estimation, the method offers a general representation of the source pdf. Although the implementation presented uses a non-informative prior on the source pdf, the method allows a prior to be incorporated. The method currently implemented is based on scalar sources; however the general formulation is readily applicable to vector source localisation, although the complexity would be greater. We believe that the methods introduced here provide a useful contribution to MEG imaging by relaxing a strong assumption in the LCMV and null-beamformer.

Acknowledgments

This work is funded by The Wellcome Trust and Engineering and Physical Sciences Research Council UK (EPSRC) under grant number WT 088877/Z/09/Z. M. L. Kringelbach is funded by the TrygFonden Charitable Foundation and T. Z. Aziz is funded by the Medical Research Council, the Norman Collisson Foundation and the Charles Wolfson Charitable Trust.

Conflict of interest

There is no conflict of interest.

Appendix

In this section, the theoretical analysis of the proposed method is presented. Appendix A shows that the source pdf can be exactly estimated if the noise power is very small and sources are independent. These results assure that a reliable approximation of the source pdf is obtained in the case of a high SNR data set. Appendix B shows that the proposed method is equivalent to the null-beamformer in the case of Gaussian measurements. Appendix C shows the relation of the proposed method to the well-known beamformer with general linear constraint, which is a generalisation of the null-beamformer. Appendix D presents the convergence analysis and convergence rate of the algorithm based on the multivariate kernel density estimation.

A. Impact of the noise

Below, we show that Eq. (13) gives the exact source pdf, if the noise power is considerably lower than the signal and interference power. We can say that the noise power tends to zero, if $\int_{|\mathbf{n}| > \delta} g_{\mathbf{n}}(\mathbf{n}) d\mathbf{n} \rightarrow 0$ for all $\delta > 0$, where $|\cdot|$ denotes the Frobenius norm. An example of such noise is the multivariate Gaussian whose norm of the covariance matrix tends to zero.

Theorem 1. In addition to the assumptions stated in Problem formulation section, suppose that the noise power tends to zero, and the lead-field vectors \mathbf{f}_i , $i = 1, \dots, q$, are mutually independent. Then $g_{s_k}(s_k)$ can be exactly estimated by $\frac{g_{\mathbf{y}}(\mathbf{f}_k s_k)}{\int g_{\mathbf{y}}(\mathbf{f}_k s) ds}$, where the normalising constant is also given by $\int g_{\mathbf{y}}(\mathbf{f}_k s) ds = g_{s_k}(0) \int g_{\mathbf{n}}(\mathbf{F}s) ds$.

Proof

Lemma 1. Based on the assumption in Theorem 1 and pseudo-convolution, we have

$$g_{\mathbf{y}}(\mathbf{y}) \rightarrow \begin{cases} \alpha g_{\mathbf{s}}(\mathbf{s}^*) & \mathbf{y} \in \mathcal{R}(\mathbf{F}) \text{ whereby } \mathbf{y} = \mathbf{F}\mathbf{s}^* \\ 0 & \mathbf{y} \notin \mathcal{R}(\mathbf{F}) \end{cases} \quad (23)$$

where $\alpha = \int g_{\mathbf{n}}(\mathbf{F}\mathbf{s}) ds$, and range of a matrix \mathbf{F} is $\mathcal{R}(\mathbf{F}) = \{\mathbf{u} | \exists \mathbf{s}^* \in \mathbb{R}^q \text{ such that } \mathbf{u} = \mathbf{F}\mathbf{s}^*\}$

Proof. Suppose there is a \mathbf{s}^* such that $\mathbf{y} = \mathbf{F}\mathbf{s}^*$, so

$$\alpha g_{\mathbf{s}}(\mathbf{s}^*) = g_{\mathbf{s}}(\mathbf{s}^*) \int g_{\mathbf{n}}(\mathbf{F}\mathbf{s}) ds = g_{\mathbf{s}}(\mathbf{s}^*) \int g_{\mathbf{n}}(\mathbf{F}(\mathbf{s}^* - \mathbf{s})) ds. \quad (24)$$

Therefore using Eq. (1):

$$|g_{\mathbf{y}}(\mathbf{y}) - \alpha g_{\mathbf{s}}(\mathbf{s}^*)| = \left| \int g_{\mathbf{s}}(\mathbf{s}) g_{\mathbf{n}}(\mathbf{y} - \mathbf{F}\mathbf{s}) ds - g_{\mathbf{s}}(\mathbf{s}^*) \int g_{\mathbf{n}}(\mathbf{F}(\mathbf{s}^* - \mathbf{s})) ds \right| \leq \int |g_{\mathbf{s}}(\mathbf{s}) - g_{\mathbf{s}}(\mathbf{s}^*)| g_{\mathbf{n}}(\mathbf{y} - \mathbf{F}\mathbf{s}) ds. \quad (25)$$

Now Let $\delta > 0$, and split the region of integration into two regions $|\mathbf{s} - \mathbf{s}^*| \leq \delta$ and $|\mathbf{s} - \mathbf{s}^*| > \delta$ such that $\int = \int_{|\mathbf{s} - \mathbf{s}^*| \leq \delta} + \int_{|\mathbf{s} - \mathbf{s}^*| > \delta} = I_1(\mathbf{y}) + I_2(\mathbf{y})$. Then we have:

$$|I_1(\mathbf{y})| \leq \max_{|\mathbf{s} - \mathbf{s}^*| \leq \delta} |g_{\mathbf{s}}(\mathbf{s}) - g_{\mathbf{s}}(\mathbf{s}^*)| \int_{|\mathbf{s} - \mathbf{s}^*| \leq \delta} g_{\mathbf{n}}(\mathbf{y} - \mathbf{F}\mathbf{s}) ds \quad (26)$$

by continuity of $g_{\mathbf{s}}(\mathbf{s})$, the above value is small for δ that is sufficiently small. On the other hand, for δ fixed we have

$$|I_2(\mathbf{y})| \leq \int_{|\mathbf{s} - \mathbf{s}^*| > \delta} (|g_{\mathbf{s}}(\mathbf{s})| + |g_{\mathbf{s}}(\mathbf{s}^*)|) g_{\mathbf{n}}(\mathbf{y} - \mathbf{F}\mathbf{s}) ds \leq 2M \int_{|\mathbf{s} - \mathbf{s}^*| > \delta} g_{\mathbf{n}}(\mathbf{y} - \mathbf{F}\mathbf{s}) ds \quad (27)$$

where M is the maximum of $g_{\mathbf{s}}(\mathbf{s})$. Noting that because $|\mathbf{s} - \mathbf{s}^*| > \delta$ there is a fixed ϵ such that $|\mathbf{y} - \mathbf{F}\mathbf{s}| = |\mathbf{F}(\mathbf{s}^* - \mathbf{s})| > \epsilon\delta$. This means $I_2(\mathbf{y})$ tends to zero by explicit assumption about the noise ($\int_{|\mathbf{n}| > \delta} g_{\mathbf{n}}(\mathbf{n}) d\mathbf{n} \rightarrow 0$).

Now suppose there is no \mathbf{s}^* such that $\mathbf{y} = \mathbf{F}\mathbf{s}^*$, but that there is a $\delta > 0$ such that $|\mathbf{s}^* - \mathbf{s}| > \delta$ and there is $\epsilon > 0$ such that $|\mathbf{F}(\mathbf{s}^* - \mathbf{s})| > \epsilon\delta$, for all $\mathbf{s} \in \mathbb{R}^q$. Therefore

$$g_{\mathbf{y}}(\mathbf{y}) \leq M \int_{|\mathbf{s} - \mathbf{s}^*| > \delta} g_{\mathbf{n}}(\mathbf{y} - \mathbf{F}\mathbf{s}) ds \quad (28)$$

where M again is the maximum of $g_{\mathbf{s}}(\mathbf{s})$. Using the same argument as in Eq. (27), $g_{\mathbf{y}}(\mathbf{y})$ tends to zero. ■

Using the above Lemma it is now shown that $g_{\mathbf{y}}(\mathbf{f}_k s_k)$ is the exact estimation of $g_{s_k}(s_k)$. If we set $\mathbf{s}^* = [0 \dots 0 \dots s_k \dots 0 \dots 0]^T$, then $\mathbf{y} = \mathbf{F}\mathbf{s}^*$ gives $\mathbf{y} = \mathbf{f}_k s_k$. Using Lemma 1:

$$g_{\mathbf{y}}(\mathbf{f}_k s_k) \rightarrow \alpha g_{\mathbf{s}}(\mathbf{s}^*) = \alpha g_{s_k}(s_k) g_{s_k}(0) = Z g_{s_k}(s_k) \quad (29)$$

where $Z = \alpha g_{s_k}(0) = g_{s_k}(0) \int g_{\mathbf{n}}(\mathbf{F}\mathbf{s}) ds$. Eq. (29) gives $g_{s_k}(s_k) = g_{\mathbf{y}}(\mathbf{f}_k s_k) / Z$, where Z also is a normalising constant $Z = \int g_{\mathbf{y}}(\mathbf{f}_k s_k) ds_k$. ■

Given this result we can now investigate the estimated source pdfs in a noisy environment.

First, note that the relationship between the measurement pdf and the source pdf in addition to the noise pdf based on Eq. (1) is given by the following equation (the pdf of the sum of two random variables (Grimmett and Stirzaker, 2001)):

$$g_{\mathbf{y}}(\mathbf{y}) = \int_{\mathbb{R}^q} g_{\mathbf{s}}(\mathbf{s}) g_{\mathbf{n}}(\mathbf{y} - \mathbf{F}\mathbf{s}) ds. \quad (30)$$

From Eqs. (13) and (30) the pdf of k th source $g_{s_k}(s_k^*)$ is given by:

$$g_{s_k}(s_k^*) \propto g_{\mathbf{y}}(\mathbf{f}_k s_k^*) = \int g_{\mathbf{s}}(\mathbf{s}) g_{\mathbf{n}}(\mathbf{f}_k s_k^* - \mathbf{F}\mathbf{s}) ds = \iint g_{s_k}(s) g_{s_j}(s_j) g_{\mathbf{n}}(\mathbf{f}_k s_k^* - \mathbf{f}_j s_j - \mathbf{F}(\mathbf{s}_k - \mathbf{s}_j)) ds ds_j \quad (31)$$

where we assumed that s_k is independent from the other sources. By changing the variable $s_k^* - s_k = u$, we have

$$\begin{aligned} g_y(\mathbf{f}s_k^*) &= \int \int g_{s_k}(s_k^* - u) g_{s_j}(\mathbf{s}_j) g_n(\mathbf{f}u - \mathbf{F}_j \mathbf{s}_j) d\mathbf{s}_j du \\ &= \int g_{s_k}(s_k^* - u) g_y(\mathbf{f}u) du = g_{s_k}(s_k^*) * g_y(\mathbf{f}s_k^*) \end{aligned} \quad (32)$$

where $g_y(\mathbf{f}u) = \int g_{s_j}(\mathbf{s}_j) g_n(\mathbf{f}u - \mathbf{F}_j \mathbf{s}_j) d\mathbf{s}_j$ and $*$ is the convolution operator. Therefore, Eq. (32) states that the estimation of $g_{s_k}(s_k^*)$ is proportional to $g_{s_k}(s_k^*) * g_y(\mathbf{f}s_k)$. This means that the estimated source pdf is the original pdf that is convolved with an unknown function which is based on the pseudo-convolution of the noise and interference pdfs. This means that in contrast to the noise-free environment, interference can also distort the estimated pdf.

It is clear that, if the noise pdf is known for example from pre-stimulus trials, one can employ the deconvolution method to remove the noise and then estimate $g_{s_k}(s_k)$ from a noise-free data set.

B. Relation to the null-beamformer

Here, it is shown that if the measurement is Gaussian, $p(s | \mathbf{y}_t)$ given by Eq. (15) is Gaussian with variance equal to Eq. (19) and mean equal to Eq. (18). Similarly, it can be shown that $g_s(s)$ given by Eq. (16) is a zero-mean Gaussian with variance equal to Eq. (19).

Using block matrix inversion, the output power of the null-beamformer, Eq. (19), can be expressed as:

$$\begin{aligned} \mathcal{P}_s &= \mathbf{c}^T (\mathbf{F}^T \mathbf{R}_y^{-1} \mathbf{F})^{-1} \mathbf{c} = \mathbf{c}^T \left(\begin{bmatrix} \mathbf{f}^T \\ \mathbf{F}_j^T \end{bmatrix} \mathbf{R}_y^{-1} \begin{bmatrix} \mathbf{f} & \mathbf{F}_j \end{bmatrix} \right)^{-1} \mathbf{c} \\ &= [1 \ 0] \begin{bmatrix} \mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{f} & \mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{F}_j \\ \mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{f} & \mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= \left(\mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{f} - (\mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{F}_j) (\mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j)^{-1} (\mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{f}) \right)^{-1} \end{aligned} \quad (33)$$

which is the Schur complement. The time-series, Eq. (18), using block matrix inversion is also expanded as:

$$\begin{aligned} \hat{s}_t &= [1 \ 0] \begin{bmatrix} \mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{f} & \mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{F}_j \\ \mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{f} & \mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{f}^T \\ \mathbf{F}_j^T \end{bmatrix} \mathbf{R}^{-1} \mathbf{y}_t \\ &= \left(\mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{f} - \mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{F}_j (\mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j)^{-1} \mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{f} \right)^{-1} \mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{y}_t \\ &\quad + \left(\mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{f} - \mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{F}_j (\mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j)^{-1} \mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{f} \right)^{-1} \\ &\quad \mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{F}_j (\mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j)^{-1} \mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{y}_t \\ &= \left(\mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{f} - \mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{F}_j (\mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j)^{-1} \mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{f} \right)^{-1} \\ &\quad \left(\mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{y}_t + \mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{F}_j (\mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j)^{-1} \mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{y}_t \right) \end{aligned} \quad (34)$$

where $-T$ denotes transpose and inverse operation.

Now we derive the null-beamformer (above expanded equations) using the method proposed. Eq. (15) is rewritten as

$$p(s | \mathbf{y}_t) = \int p(s, \mathbf{s}_j | \mathbf{y}_t) d\mathbf{s}_j \propto \int p(\mathbf{f}s + \mathbf{F}_j \mathbf{s}_j | \mathbf{y}_t) d\mathbf{s}_j. \quad (35)$$

Using the Gaussianity of the data, $p(s | \mathbf{y}_t)$ is found as

$$\begin{aligned} p(s | \mathbf{y}_t) &\propto \int \exp \left[-(\mathbf{f}s + \mathbf{F}_j \mathbf{s}_j - \mathbf{y}_t)^T \mathbf{R}_y^{-1} (\mathbf{f}s + \mathbf{F}_j \mathbf{s}_j - \mathbf{y}_t) \right] d\mathbf{s}_j \\ &\propto \int \exp \left[-s^T \mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{f} s + s^T \mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{y}_t + \mathbf{y}_t^T \mathbf{R}_y^{-1} \mathbf{f} s - s^T \mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{F}_j \mathbf{s}_j \right. \\ &\quad \left. - \mathbf{s}_j^T \mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{f} s - \mathbf{s}_j^T \mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j \mathbf{s}_j + \mathbf{s}_j^T \mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{y}_t + \mathbf{y}_t^T \mathbf{R}_y^{-1} \mathbf{F}_j \mathbf{s}_j \right] d\mathbf{s}_j \\ &\propto \exp \left[-s^T \mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{f} s + s^T \mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{y}_t + \mathbf{y}_t^T \mathbf{R}_y^{-1} \mathbf{f} s \right] \times \\ &\quad \int \exp \left[-\left(\mathbf{s}_j - (\mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j)^{-1} \mathbf{F}_j^T \mathbf{R}_y^{-1} (\mathbf{f}s - \mathbf{y}_t) \right)^T (\mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j) \right. \\ &\quad \left. \left(\mathbf{s}_j - (\mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j)^{-1} \mathbf{F}_j^T \mathbf{R}_y^{-1} (\mathbf{f}s - \mathbf{y}_t) \right) \right] d\mathbf{s}_j \times \\ &\quad \exp \left[(\mathbf{f}s - \mathbf{y}_t)^T \mathbf{R}_y^{-1} \mathbf{F}_j (\mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j)^{-1} \mathbf{F}_j^T \mathbf{R}_y^{-1} (\mathbf{f}s - \mathbf{y}_t) \right] \end{aligned} \quad (36)$$

The last integral is an integration over a Gaussian distribution with random variable \mathbf{s}_j and therefore it will be a constant. We continue by eliminating the integral and combining the two exponentials:

$$\begin{aligned} p(s | \mathbf{y}_t) &\propto \exp \left[-s^T \left(\mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{f} - \mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{F}_j (\mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j)^{-1} \mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{f} \right) s \right. \\ &\quad \left. + s^T \left(\mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{y}_t + \mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{F}_j (\mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j)^{-1} \mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{y}_t \right) \right. \\ &\quad \left. + \left(\mathbf{y}_t^T \mathbf{R}_y^{-1} \mathbf{f} + \mathbf{y}_t^T \mathbf{R}_y^{-1} \mathbf{F}_j (\mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j)^{-1} \mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{f} \right) s \right] \\ &\propto \exp \left[-(s - \mathbf{a})^T \left(\mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{f} - \mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{F}_j (\mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j)^{-1} \mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{f} \right) (s - \mathbf{a}) \right] \end{aligned} \quad (37)$$

where $\mathbf{a} = (\mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{f} - \mathbf{f}^T \mathbf{R}_y^{-1} (\mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j)^{-1} \mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{f})^{-1} (\mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{y}_t + \mathbf{f}^T \mathbf{R}_y^{-1} \mathbf{F}_j (\mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{F}_j)^{-1} \mathbf{F}_j^T \mathbf{R}_y^{-1} \mathbf{y}_t)$. The above equation shows that $p(s | \mathbf{y}_t)$ is a Gaussian distribution and its variance is equal to Eq. (33). The best estimation of the time-series is its mean which is equal to Eq. (34).

C. Relation to the beamformer with general linear constraint

A beamformer with general linear constraint is given by the following formulation (Haykin, 2002):

$$\text{argmax } \mathbf{w}^T \mathbf{R}_y \mathbf{w}, \text{ subject to: } \mathbf{w}^T \mathbf{F} = \mathbf{c}^T \quad (38)$$

where $\mathbf{F} = [\mathbf{f}_1 \dots \mathbf{f}_q]$ and \mathbf{c} is a constraint vector. The solution of this problem is the same as those given by Eqs. (18), (19), except that, here, \mathbf{c} is an arbitrary vector.

Our aim is to present an estimation method for $g_s(s)$ that is equivalent to the beamformer with general linear constraint, when the measurement is Gaussian.

The method based on the marginal distribution given in the Non-Gaussian marginal distribution for correlated sources section, marginalises out all the parameters except the desired parameter. In other words, it integrates $g_s(\mathbf{s})$ over the subspace that is perpendicular to $\mathbf{c} = [1 \ 0]^T$. Therefore, we can employ the same procedure to first estimate $g_s(\mathbf{s}) \propto g_y(\mathbf{F}\mathbf{s})$, and then the desired pdf $g_s(s)$ for an arbitrary \mathbf{c} is obtained by integration of $g_s(\mathbf{s})$ over the subspace that is perpendicular to \mathbf{c} , or equivalently:

$$g_s(s) = \int_{\Omega_c} g_y(\mathbf{F}\mathbf{s}) d\mathbf{s}_j \quad (39)$$

where $\Omega_c = \{\mathbf{s} | \mathbf{c}^T \mathbf{s} - \mathbf{c}^T \mathbf{c} s = 0\}$, and $\int_{\Omega_c} g d\mathbf{s}_j$ is the integral of g over subspace Ω_c with appropriate variable of integration $d\mathbf{s}_j$.

The general solution of the method proposed (Eq. (39)) has the same solution to the beamformer with general linear constraint, when the measurement is Gaussian. This can simply be shown by rotating the coordinate such that the vector \mathbf{c} lies on the first axis. In this case, the rotated \mathbf{c} is equal to $[1 \ 0]^T$ and, thus, Eq. (16) can be used for this rotated coordinate. By rotating back to the original coordinate, Eq. (39) is obtained.

D. Convergence analysis

The convergence, bias and variance of the kernel based pdf estimation stated in Eq. (21) is given by the following theorem.

Theorem 2. Suppose the kernel K satisfies the following conditions:

$$\int K(\mathbf{u})d\mathbf{u} = 1, \quad \left| \int \mathbf{u}^T \mathbf{u} K(\mathbf{u})d\mathbf{u} \right| < \infty, \quad K(\mathbf{u}) = K(-\mathbf{u}) \quad (40)$$

- (i) At every point of continuity of $g_s(\cdot)$, if the noise power tends to zero then

$$\lim_{T \rightarrow \infty} \hat{g}_s(s) \rightarrow g_s(s) \quad (41)$$

- (ii) If $g_s(s)$ is twice differentiable and its second partial derivatives are bounded and continuous, then the bias is given by:

$$\begin{aligned} E\{\hat{g}_s(s)\} - g_s(s) \rightarrow \\ \frac{h^2}{2g_{yy}(\mathbf{f})} \left(\int \mathbf{u}^T \mathbf{G}_{gy}(\mathbf{f}s) \mathbf{u} K(\mathbf{u})d\mathbf{u} - \frac{g_y(\mathbf{f}s)}{g_y(\mathbf{f})} \int_{\Omega_f} \mathbf{u}^T \mathbf{G}_{gy/}(\mathbf{f}) \mathbf{u} K/(\mathbf{u})d\mathbf{u} \right) \\ + O(h^4) + O\left(\frac{1}{Th^{N-1}}\right) \end{aligned} \quad (42)$$

where the Hessian matrix \mathbf{G}_{gy} is given by $\mathbf{G}_{gy}(\mathbf{u}) = \left[\frac{\partial^2 g_y(\mathbf{u})}{\partial u_i \partial u_j} \right]_{i,j}$.

Here, $\mathbf{G}_{gy/}(\mathbf{f}) = \int \mathbf{G}_{gy}(\mathbf{f}s)ds$, $g_{y/}(\mathbf{f}) = \int g_y(\mathbf{f}s)ds$, $K/(\mathbf{u}) = \int K(\mathbf{f}s - \mathbf{u})ds$ and $\Omega_f = \{\mathbf{u} \mid \mathbf{f}^T \mathbf{u} - \mathbf{f}^T \mathbf{u} = 0\}$.

- (iii) The variance of the estimator $E\{(g_s(s) - \hat{g}_s(s))^2\}$ is given by

$$\begin{aligned} E\{(g_s(s) - \hat{g}_s(s))^2\} \rightarrow \\ \frac{g_y(\mathbf{f}s)}{Th^N g_{yy}^2(\mathbf{f})} \left(\int K^2(\mathbf{u})d\mathbf{u} + h \frac{g_y(\mathbf{f}s)}{g_y(\mathbf{f})} \left(\int_{\Omega_f} K/(\mathbf{u})d\mathbf{u} - 2 \int K(\mathbf{u})K/(\mathbf{u})d\mathbf{u} \right) \right) \\ + O\left(\frac{1}{Th^{N-2}}\right). \end{aligned} \quad (43)$$

Proof. For part (i) Please refer to a similar theorem in Parzen (1962).

We use the following Lemma to proof parts (ii) and (iii):

Lemma 2. Let $\mathcal{U}_1(\mathbf{y})$ and $\mathcal{U}_2(\mathbf{y})$ be two random variables and let $\{\mathbf{y}_1, \dots, \mathbf{y}_T\}$ be an independent and identically distributed sequence of random variables, then

$$E\left\{ \frac{1}{T} \sum_{t=1}^T \mathcal{U}_1(\mathbf{y}_t) \right\} \approx \frac{E\{\mathcal{U}_1\}}{E\{\mathcal{U}_2\}} + \frac{1}{T} \left(\frac{E\{\mathcal{U}_1\} \text{var}\{\mathcal{U}_2\}}{E\{\mathcal{U}_2\}^3} - \frac{\text{cov}\{\mathcal{U}_1, \mathcal{U}_2\}}{E\{\mathcal{U}_2\}^2} \right) \quad (44)$$

and

$$\begin{aligned} \text{var}\left\{ \frac{1}{T} \sum_{t=1}^T \mathcal{U}_1(\mathbf{y}_t) \right\} \approx \frac{1}{TE\{\mathcal{U}_2\}^2} \\ \left(\text{var}\{\mathcal{U}_1\} + \frac{E\{\mathcal{U}_1\}^2 \text{var}\{\mathcal{U}_2\}}{E\{\mathcal{U}_2\}^2} - 2 \frac{E\{\mathcal{U}_1\} \text{cov}\{\mathcal{U}_1, \mathcal{U}_2\}}{E\{\mathcal{U}_2\}} \right). \end{aligned} \quad (45)$$

Proof. See Proposition 31.8 in Port (1994). ■

Suppose $\mathcal{U}_1(\mathbf{y}) = \frac{1}{h^N} K\left(\frac{\mathbf{f}s - \mathbf{y}}{h}\right)$ (numerator of Eq. (21)) and $\mathcal{U}_2(\mathbf{y}) = \int \frac{1}{h^N} K\left(\frac{\mathbf{f}s - \mathbf{y}}{h}\right) ds$ (denominator of Eq. (21)), then:

$$E\{\mathcal{U}_2\} = E\left\{ \int \frac{1}{h^N} K\left(\frac{\mathbf{f}s - \mathbf{y}}{h}\right) ds \right\} = \iint \frac{1}{h^N} K\left(\frac{\mathbf{f}s - \mathbf{y}}{h}\right) g_y(\mathbf{y}) d\mathbf{s} d\mathbf{y}. \quad (46)$$

To calculate the above integral, we rotate the coordinate such that the line along the direction of \mathbf{f} lays on the first coordinate axis. This can be done simply by a rotation matrix Θ such that $\Theta \mathbf{f} = \|\mathbf{f}\| [1 \ 0 \dots 0]^T$, where $\|\mathbf{f}\|$ is the norm of \mathbf{f} . Suppose that the rotated pdf and the rotated kernel are $g_y^r(\mathbf{v}) = g_y(\Theta \mathbf{y})$ and $K^r(\mathbf{v}) = K(\Theta \mathbf{y})$, respectively, where $\mathbf{v} = [v_1, \dots, v_N]^T$ is the new coordinate. Thus, Eq. (51) is rewritten as:

$$\begin{aligned} E\{\mathcal{U}_2\} &= \iint \frac{1}{h^N} K^r\left(\frac{\|\mathbf{f}\|s - v_1}{h}, \frac{v_2}{h}, \dots, \frac{v_N}{h}\right) g_y^r(\mathbf{v}) d\mathbf{s} d\mathbf{v} \\ &= \int \frac{1}{\|\mathbf{f}\| h^{N-1}} K^r\left(\frac{v_2}{h}, \dots, \frac{v_N}{h}\right) g_y^r(\mathbf{v}) d\mathbf{v}. \end{aligned} \quad (47)$$

Here we used $u = \|\mathbf{f}\|s - v_1$ and we defined $K^r\left(\frac{v_2}{h}, \dots, \frac{v_N}{h}\right) = \int K^r\left(u, \frac{v_2}{h}, \dots, \frac{v_N}{h}\right) du$. Now let us define $\mathbf{v}_/ = [v_2, \dots, v_N]^T$, thus, $K^r\left(\frac{v_2}{h}, \dots, \frac{v_N}{h}\right) = K/(\mathbf{v}_/)$ and $g_y^r(\mathbf{v}) = g_y^r(v_1, \mathbf{v}_/)$. By change of variable $\mathbf{u}_/ = \frac{\mathbf{v}_/}{h}$, we have:

$$E\{\mathcal{U}_2\} = \iint \frac{1}{\|\mathbf{f}\|} K/(\mathbf{u}_/) g_y^r(u_1, h\mathbf{u}_/) du_1 d\mathbf{u}_/. \quad (48)$$

Using Taylor series expansion around origin for fixed u_1 , we approximate $g_y^r(u_1, h\mathbf{u}_/) \approx g_y^r(u_1, 0) + h\mathbf{u}_/^T \mathbf{J}_{gy}^r(u_1, 0) + \frac{h^2}{2} \mathbf{u}_/^T \mathbf{G}_{gy/}^r(u_1, 0) \mathbf{u}_/$, where $\mathbf{G}_{gy/}^r(u_1, \mathbf{u}_/) = \left[\frac{\partial^2 g_y^r(u_1, \mathbf{u}_/)}{\partial u_i \partial u_j} \right]_{i,j>1}$ and $\mathbf{J}_{gy}^r(u_1, \mathbf{u}_/) = \left[\frac{\partial g_y^r(u_1, \mathbf{u}_/)}{\partial u_i} \right]_{i>1}$. We continue as:

$$\begin{aligned} E\{\mathcal{U}_2\} &\approx \iint \frac{1}{\|\mathbf{f}\|} K/(\mathbf{u}_/) \left(g_y^r(u_1, 0) + h\mathbf{u}_/^T \mathbf{J}_{gy}^r(u_1, 0) + \frac{h^2}{2} \mathbf{u}_/^T \mathbf{G}_{gy/}^r(u_1, 0) \mathbf{u}_/ \right) \\ &du_1 d\mathbf{u}_/ = \frac{1}{\|\mathbf{f}\|} \int g_y^r(u_1, 0) du_1 + \frac{h^2}{2\|\mathbf{f}\|} \iint \mathbf{u}_/^T \mathbf{G}_{gy/}^r(u_1, 0) \mathbf{u}_/ K/(\mathbf{u}_/) du_1 d\mathbf{u}_/. \end{aligned} \quad (49)$$

This simplification is because of two identities $\int K/(\mathbf{u}_/) d\mathbf{u}_/ = 1$ and $\int K^r(\mathbf{u}_/) d\mathbf{u}_/ = 0$. The first identity is the result of $\int K/(\mathbf{u}_/) d\mathbf{u}_/ = \int \int K/(\mathbf{u}_/) du_1 d\mathbf{u}_/ = \int K(\mathbf{u})d\mathbf{u} = 1$ and the second identity is the result of $K/(\mathbf{u}_/) = \int K^r(u_1, \mathbf{u}_/) du_1 = \int K/(-u_1, -\mathbf{u}_/) du_1 = \int K/(\mathbf{u}_/) du_1 = K/(-\mathbf{u}_/)$.

If we rotate back to the original coordinate we have

$$\begin{aligned} \frac{1}{\|\mathbf{f}\|} \int g_y^r(u_1, 0) du_1 &= \int g_y(\mathbf{f}s) ds \\ \frac{h^2}{2\|\mathbf{f}\|} \iint \mathbf{u}_/^T \mathbf{G}_{gy/}^r(u_1, 0) \mathbf{u}_/ K/(\mathbf{u}_/) du_1 d\mathbf{u}_/ &= \frac{h^2}{2} \int_{\Omega_f} \mathbf{y}^T \mathbf{G}_{gy}(\mathbf{f}s) d\mathbf{s} K/(\mathbf{y}) d\mathbf{y} \end{aligned} \quad (50)$$

where $\mathbf{G}_{gy}(\mathbf{u}) = \left[\frac{\partial^2 g_y(\mathbf{u})}{\partial u_i \partial u_j} \right]_{ij}$, $\Omega_f = \{\mathbf{u} \mid \mathbf{u}^T \mathbf{f} - \mathbf{f}^T \mathbf{f} = 0\}$ and $K/(\mathbf{u}) = \int K(\mathbf{f}s - \mathbf{u})ds$. By substituting Eq. (50) into Eq. (49), we obtain:

$$E\{\mathcal{U}_2\} \approx \int g_y(\mathbf{f}s) ds + \frac{h^2}{2} \int_{\Omega_f} \mathbf{u}^T \mathbf{G}_{gy}(\mathbf{f}s) \mathbf{u} K/(\mathbf{u}) d\mathbf{u} \quad (51)$$

where $\mathbf{G}_{gy/} = \int \mathbf{G}_{gy}(\mathbf{f}s) ds$.

Now $E\{\mathcal{U}_1\mathcal{U}_2\}$ is given by:

$$\begin{aligned} E\{\mathcal{U}_1\mathcal{U}_2\} &= E\left\{\frac{1}{h^N}K\left(\frac{\mathbf{fs}-\mathbf{y}}{h}\right)\int\frac{1}{h^N}K\left(\frac{\mathbf{fs}-\mathbf{y}}{h}\right)ds\right\} \\ &= \int\frac{1}{h^{2N}}K\left(\frac{\mathbf{fs}-\mathbf{y}}{h}\right)\int K\left(\frac{\mathbf{fu}-\mathbf{y}}{h}\right)d\mathbf{g}_{\mathbf{y}}(\mathbf{y})d\mathbf{y}. \end{aligned} \quad (52)$$

Using $u = v + s$ for inner integral and then using $\mathbf{u} = \frac{\mathbf{fs}-\mathbf{y}}{h}$, we obtain:

$$\begin{aligned} E\{\mathcal{U}_1\mathcal{U}_2\} &= \int\frac{1}{h^{2N}}K\left(\frac{\mathbf{fs}-\mathbf{y}}{h}\right)\int K\left(\frac{\mathbf{fv}+\mathbf{fs}-\mathbf{y}}{h}\right)d\mathbf{v}g_{\mathbf{y}}(\mathbf{y})d\mathbf{y} \\ &= \int\frac{1}{h^N}K(\mathbf{u})\int K\left(\frac{\mathbf{fv}}{h}+\mathbf{u}\right)d\mathbf{v}g_{\mathbf{y}}(\mathbf{fs}-\mathbf{uh})d\mathbf{u} \\ &= \int\frac{1}{h^{N-1}}K(\mathbf{u})\int K(\mathbf{fv}+\mathbf{u})d\mathbf{v}g_{\mathbf{y}}(\mathbf{fs}-\mathbf{uh})d\mathbf{u} \\ &= \int\frac{1}{h^{N-1}}K(\mathbf{u})K_{/}(\mathbf{u})g_{\mathbf{y}}(\mathbf{fs}-\mathbf{uh})d\mathbf{u} \end{aligned} \quad (53)$$

where $K_{/}(\mathbf{u}) = \int K(\mathbf{fs}-\mathbf{u})ds = \int K(\mathbf{fs}+\mathbf{u})ds$. Using Taylor series expansion $g_{\mathbf{y}}(\mathbf{fs}-\mathbf{uh}) \approx g_{\mathbf{y}}(\mathbf{fs}) + \mathbf{h}\mathbf{u}^T \mathbf{J}_{g_{\mathbf{y}}}(\mathbf{fs}) + \frac{1}{2}\mathbf{h}^2 \mathbf{u}^T \mathbf{G}_{g_{\mathbf{y}}}(\mathbf{fs})\mathbf{u}$ the above equation is rewritten and simplified to:

$$E\{\mathcal{U}_1\mathcal{U}_2\} \approx \frac{1}{h^{N-1}}\left(g_{\mathbf{y}}(\mathbf{fs})\int K(\mathbf{u})K_{/}(\mathbf{u})d\mathbf{u} + \frac{1}{2}h^2\int \mathbf{u}^T \mathbf{G}_{g_{\mathbf{y}}}(\mathbf{fs})\mathbf{u}K(\mathbf{u})K_{/}(\mathbf{u})d\mathbf{u}\right). \quad (54)$$

Similarly, it can be shown that

$$\begin{aligned} E\{\mathcal{U}_1\} &= E\left\{\frac{1}{h^N}K\left(\frac{\mathbf{fs}-\mathbf{y}}{h}\right)\right\} \\ &\approx g_{\mathbf{y}}(\mathbf{fs}) + \frac{h^2}{2}\int \mathbf{u}^T \mathbf{G}_{g_{\mathbf{y}}}(\mathbf{fs})\mathbf{u}K(\mathbf{u})d\mathbf{u} \end{aligned} \quad (55)$$

and

$$\begin{aligned} E\{\mathcal{U}_1^2\} &= E\left\{\frac{1}{h^{2N}}K^2\left(\frac{\mathbf{fs}-\mathbf{y}}{h}\right)\right\} \\ &\approx \frac{1}{h^N}\left(g_{\mathbf{y}}(\mathbf{fs})\int K^2(\mathbf{u})d\mathbf{u} + \frac{h^2}{2}\int \mathbf{u}^T \mathbf{G}_{g_{\mathbf{y}}}(\mathbf{fs})\mathbf{u}K^2(\mathbf{u})d\mathbf{u}\right) \end{aligned} \quad (56)$$

and

$$\begin{aligned} E\{\mathcal{U}_2^2\} &= E\left\{\frac{1}{h^{2N}}\left(\int K\left(\frac{\mathbf{fs}-\mathbf{y}}{h}\right)ds\right)^2\right\} \\ &\approx \frac{1}{h^{N-1}}\left(\int p(\mathbf{fs})ds\int_{\Omega_t} K_{/}^2(\mathbf{u})d\mathbf{u} + \frac{h^2}{2}\int_{\Omega_t} \mathbf{u}^T \mathbf{G}_{g_{\mathbf{y}}/}(\mathbf{fs})\mathbf{u}K_{/}^2(\mathbf{u})d\mathbf{u}\right). \end{aligned} \quad (57)$$

Therefore, using Eqs. (55) and (56) we have:

$$\begin{aligned} \text{var}(\mathcal{U}_1) &= E\{\mathcal{U}_1^2\} - E\{\mathcal{U}_1\}^2 \\ &\approx \frac{1}{h^N}g_{\mathbf{y}}(\mathbf{fs})\int K^2(\mathbf{u})d\mathbf{u} + \frac{1}{2h^{N-2}}\int \mathbf{u}^T \mathbf{G}_{g_{\mathbf{y}}}(\mathbf{fs})\mathbf{u}K^2(\mathbf{u})d\mathbf{u} \end{aligned} \quad (58)$$

and using Eqs. (51) and (57):

$$\begin{aligned} \text{var}(\mathcal{U}_2) &\approx \frac{1}{h^{N-1}}\int g_{\mathbf{y}}(\mathbf{fs})ds\int_{\Omega_t} K_{/}^2(\mathbf{u})d\mathbf{u} \\ &+ \frac{1}{2h^{N-3}}\int_{\Omega_t} \mathbf{u}^T \mathbf{G}_{g_{\mathbf{y}}/}(\mathbf{fs})\mathbf{u}K_{/}^2(\mathbf{u})d\mathbf{u} \end{aligned} \quad (59)$$

and using Eqs. (55), (51) and (54):

$$\begin{aligned} \text{cov}(\mathcal{U}_1, \mathcal{U}_2) &= E\{\mathcal{U}_1\mathcal{U}_2\} - E\{\mathcal{U}_1\}E\{\mathcal{U}_2\} \\ &\approx \frac{1}{h^{N-1}}g_{\mathbf{y}}(\mathbf{fs})\int K(\mathbf{u})K_{/}(\mathbf{u})d\mathbf{u} + \frac{1}{2h^{N-3}}\int \mathbf{u}^T \mathbf{G}_{g_{\mathbf{y}}}(\mathbf{fs})\mathbf{u}K(\mathbf{u})K_{/}(\mathbf{u})d\mathbf{u}. \end{aligned} \quad (60)$$

Finally, using Lemma 2 and the fact that $\frac{1}{s+\delta} \approx \frac{1}{s} - \frac{\delta}{s^2}$, for a sufficiently small δ , the bias and variance of the estimator are obtained. ■

Thus, the kernel based estimations are almost asymptotically unbiased, if $h \rightarrow 0$ and $Th^{N-1} \rightarrow \infty$ as $T \rightarrow \infty$. Furthermore, the estimator is consistent (variance is zero) provided $Th^N \rightarrow y$, as $T \rightarrow \infty$. This means that as with many smoothing problems, small bandwidths give small bias and large variance, whereas large bandwidths give large bias and small variance.

The theoretical optimal bandwidth is obtained by minimising the integrated asymptotic mean square error (IMSE), which is defined by Cheng and Sfring (1981):

$$\text{IMSE} = E\left\{\int (g_s(s) - \hat{g}_s(s))^2\right\} \quad (61)$$

which in turn may be written as the sum of the integrated variance and the integrated squared bias. Therefore, using Theorem 2, the resulting expression is of the form of:

$$\text{IMSE} = ah^4 + \frac{b}{Th^N} + O(h^6) + O\left(\frac{1}{Th^{N-1}}\right) \quad (62)$$

where a and b are constants. The optimal bandwidths are obtained by differentiating the above equation with respect to the bandwidth h and equating to zero. This shows that h is of order $T^{\frac{4}{N+4}}$. Substituting h into Eq. (62) shows that the best convergence rate in the sense of IMSE is in the order of $T^{-\frac{4}{N+4}}$. It is expected, therefore, that the convergence properties are better when the number of sensors are reduced while the number of observations is fixed.

E. Computational complexity of the algorithm

The computational complexity of the proposed method can be derived as follows. The part of Algorithm 1, which has the largest order of complexity, is given in Eq. (20); other parts of the algorithm such as normalisation have a smaller order and can be ignored. Eq. (20) is the sum of T kernels, each of which has a computational complexity of order $O(N^2)$, and therefore this equation has a computational complexity of order $O(N^2T)$. For estimation of the power, this term should be calculated once for each point of the discrete source distribution, therefore the computational complexity for estimation of the power is of order $O(TN^2D_s)$. For estimation of the time-series this term should be calculated for each time sample and thus its computational complexity is of order $O(T^2N^2D_s)$.

The computational complexity of the LCMV beamformer at each voxel is given by Eqs. (3) and (4). The calculation of $\mathbf{f}^T \mathbf{R}^{-1} \mathbf{f}$ is of order $O(N^2)$ (its inverse only needs one operation) and $\mathbf{f}^T \mathbf{R}_t \mathbf{f}$ is of order $O(N^2T)$. Thus, the computational complexity of the estimation of power is of order $O(N^2)$ and the computational complexity of the estimation of the time-series is of order $O(N^2T)$.

References

- Baillet, S., Mosher, J.C., Leahy, R.M., 2001. Electromagnetic brain mapping. *IEEE Signal Proc. Mag.* 18 (6), 14–30.
- Barnes, G.R., Hillebrand, A., 2003. Statistical flattening of MEG beamformer images. *Hum. Brain Mapp.* 18 (1), 1–12.
- Brookes, M.J., Gibson, A.M., Hall, S.D., Furlong, P.L., Barnes, G.R., Hillebrand, A., Singh, K.D., Holliday, I.E., Francis, S.T., Morris, P.G., 2004. A general linear model for MEG beamformer imaging. *NeuroImage* 23 (3), 936–946.
- Brookes, M.J., Gibson, A.M., Hall, S.D., Furlong, P.L., Barnes, G.R., Hillebrand, A., Singh, K.D., Holliday, I.E., Francis, S.T., Morris, P.G., 2005. GLM-beamformer method demonstrates

- stationary field, alpha ERD and gamma ERS co-localisation with fMRI BOLD response in visual cortex. *NeuroImage* 1, 302–308.
- Brookes, M.J., Stevenson, C.M., Barnes, G.R., Hillebrand, A., Simpson, M.I., Francis, S.T., Morris, P.G., 2007. Beamformer reconstruction of correlated sources using a modified source model. *NeuroImage* 34 (4), 1454–1465.
- Brookes, M.J., Hale, J.R., Zumer, J.M., Stevenson, C.M., Francis, S.T., Barnes, G.R., Owen, J.P., Morris, P.G., Nagarajan, S.S., 2011. Measuring functional connectivity using MEG: methodology and comparison with fMRI. *NeuroImage* 3, 1082–1104.
- Cheng, P., Serfling, R., 1981. Asymptotic mean integrated squared errors of some nonparametric density estimators (corresp.). *IEEE Trans. Inf. Theory* 2, 239–242.
- Cox, H., Zeskind, R.M., Owen, M.M., 1987. Robust adaptive beamforming. *IEEE Trans. Acoust. Speech Signal Process. ASSP-35* (Suppl. 1), 1365–1376.
- Dalal, S.S., Zumer, J., Agrawal, V., Hild, K., Sekihara, K., Nagarajan, S., 2004. Nutmeg: a neuromagnetic source reconstruction toolbox. *Neuro. Clin. Neurophysiol.* 2004, 52.
- Dalal, S.S., Sekihara, K., Nagarajan, S.S., 2006. Modified beamformers for coherent source region suppression. *IEEE Trans. Biomed. Eng.* 53 (7), 1357–1363.
- Davis, A.W., 1976. Statistical distributions in univariate and multivariate Edgeworth populations. *Biometrika* 63 (3), 661–670.
- Dogandzic, A., Nehorai, A., 2000. Estimating evoked dipole responses in unknown spatially correlated noise with EEG/MEG arrays. *IEEE Trans. Signal Process.* 1, 13–25.
- Ehrenberg, L., Gannot, S., Leshem, A., Zehavi, E., 2010. Sensitivity analysis of MVDR and MPDR beamformers. *Electrical and Electronics Engineers in Israel (IEEEI)*, IEEE 26th Convention of, pp. 416–420.
- Elul, R., Hanley, J., Simmons, J.Q., 1975. Non-Gaussian behavior of the EEG in down's syndrome suggests decreased neuronal connections. *Acta Neurol. Scand.* 51, 21–28.
- Friedman, J., 1997. On bias, variance, 0/1 loss, and the curse-of-dimensionality. *Data Min. Knowl. Disc.* 1 (1), 55–77.
- Friston, K., Glaser, D., Henson, R., Kiebel, S., Phillips, C., Ashburner, J., 2002. Classical and Bayesian inference in neuroimaging: applications. *NeuroImage* 2, 484–512.
- Grimmett, G., Stirzaker, D., 2001. *Probability and Random Processes*. Oxford University Press.
- Hämäläinen, M., Hari, R., Ilmoniemi, R.J., Knuutila, J., Lounasmaa, O.V., 1993. Magnetoencephalography theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Mod. Phys.* 413–497. <http://dx.doi.org/10.1103/RevModPhys.65.413>.
- Hansen, P., Kringelbach, M.L., Salmelin, R., 2010. *MEG. An Introduction to Methods*. Oxford University Press, Oxford.
- Haykin, S., 2002. *Adaptive Filter Theory*. Prentice Hall.
- Hillebrand, A., 2003. The use of anatomical constraints with MEG beamformers. *NeuroImage* 4, 2302–2313.
- Hudson, J.E., 1981. *Adaptive Array Principles*. Peter Peregrinus, London, U.K.
- Hui, H.B., Leahy, R., 2006. Linearly constrained MEG beamformers for MVAR modeling of cortical interactions. *IEEE Inter. Symp. on Biomedical Imaging: Nano to Macro*, Buenos Aires, Argentina.
- Jarque, C.M., Bera, A.K., 1987. A test for normality of observations and regression residuals. *Int. Stat. Rev.* 55 (2), 163–172.
- Jun, S., George, J., Paré-Blagoev, J., Plis, S., Ranken, D., Schmidt, D., Wood, C., 2005. Spatio-temporal Bayesian inference dipole analysis for MEG neuroimaging data. *NeuroImage* 28 (1), 84–98.
- Lee, T.-W., Lewicki, M., Sejnowski, T., 2000. ICA mixture models for unsupervised classification of non-Gaussian classes and automatic context switching in blind signal separation. *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (10), 1078–1089.
- Limpiti, T., Van Veen, B., Wakai, R., 2006. Cortical patch basis model for spatially extended neural activity. *IEEE Trans. Biomed. Eng.* 53 (9), 1740–1754.
- Mattout, J., Phillips, C., Penny, W.D., Rugg, M.D., Friston, K.J., 2006. MEG source localization under multiple constraints: an extended Bayesian framework. *NeuroImage* 30 (3), 753–767.
- Mohseni, H.R., Kringelbach, M., Smith, P.P., Stein, J., Aziz, T.Z., 2010. Application of null-beamformer to source localization of MEG data in deep brain stimulation. *IEEE EMBC Conference*, Buenos Aires, Argentina.
- Mohseni, H.R., Smith, P.P., Parsons, C.E., Young, K.S., Hyam, J.A., Stein, A., Stein, J.F., Green, A.L., Aziz, T.Z., Kringelbach, M.L., 2012. MEG can map short and long-term changes in brain activity following deep brain stimulation for chronic pain. *PLoS One* 7 (6), e37993.
- Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., Zilles, K., 2001. Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage* 13 (4), 684–701.
- Mosher, J.C., Lewis, P.S., Leahy, R.M., 1992. Multiple dipole modeling and localization from spatio-temporal MEG data. *IEEE Trans. Biomed. Eng.* 39 (6), 541–557.
- Mosher, J., Leahy, R., Lewis, P., 1999. EEG and MEG: forward solutions for inverse methods. *IEEE Trans. Biomed. Eng.* 3, 245–259.
- Nagarajan, S., Attias, H., Hild, K., Sekihara, K., 2005. Stimulus evoked independent factor analysis of MEG data with large background activity. *Adv. Neural Inf. Process.* 18.
- Nagarajan, S., Attias, H., Hild, K., Sekihara, K., 2006. A graphical model for estimating stimulus-evoked brain responses from magnetoencephalography data with large background brain activity. *NeuroImage* 30 (2), 400–416.
- Nummenmaa, A., Auranen, T., Hamalainen, M., Jaaskelainen, I., Lampinen, J., Sams, M., Vehtari, A., 2007. Hierarchical Bayesian estimates of distributed meg sources: theoretical aspects and comparison of variational and MCMC methods. *NeuroImage* 35 (2), 669–685.
- Parzen, E., 1962. On estimation of a probability density function and mode. *Ann. Math. Stat.* 33 (3), 1065–1076.
- Popescu, M., Popescu, E., Chan, T., Blunt, S.D., Lewine, J.D., 2008. Spatio temporal reconstruction of bilateral auditory steady-state responses using MEG beamformers. *IEEE Trans. Biomed. Eng.* 55 (3), 1092–1102.
- Robinson, S., Vrba, J., et al., 1999. Functional neuroimaging by synthetic aperture magnetometry (SAM). *Recent Adv. Biomagnetism* 302–305.
- Scott, D., 1992. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley, New York.
- Scott, D., Sain, S., 2004. Multi-dimensional density estimation. *Handbook of Statistics*, pp. 1–39 (August).
- Sekihara, K., Nagarajan, S., Poeppel, D., Marantz, A., Miyashita, Y., 2002. Application of an MEG eigenspace beamformer to reconstructing spatio-temporal activities of neural sources. *Hum. Brain Mapp.* 15 (4), 199–215.
- Sekihara, K., Nagarajan, S.S., Poeppel, D., Marantz, A., 2006. Performance of an MEG adaptive-beamformer technique in the presence of correlated neural activities: effects on signal intensity and time-course estimates. *IEEE Trans. Biomed. Eng.* 49 (12), 1534–1546.
- Silverman, B.W., 1999. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall/CRC.
- Taulu, S., Simola, J., 2006. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Phys. Med. Biol.* 7, 1759.
- Van Trees, H., 2004. *Detection, Estimation, and Modulation Theory, Radar–Sonar Signal Processing and Gaussian Signals in Noise, Part IV*. Wiley-Interscience.
- Van Veen, B.D., Buckley, K., 1988. Beamforming: a versatile approach to spatial filtering. *IEEE ASSP Mag.* 5 (2), 4–24.
- Van Veen, B.D., Dronglen, W.V., Yuchtman, M., Suzuki, A., 1997. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans. Biomed. Eng.* 44 (9), 867–880.
- Wipf, D., Nagarajan, S., 2007. Beamforming using the relevance vector machine. *Proceedings of the 24th International Conference on Machine Learning*, pp. 1023–1030.
- Wipf, D., Nagarajan, S., 2009. A unified Bayesian framework for MEG/EEG source imaging. *NeuroImage* 44 (3), 400–416.
- Woolrich, M., Hunt, L., Groves, A., Barnes, G., 2011. MEG beamforming using Bayesian PCA for adaptive data covariance matrix regularisation. *NeuroImage* 57 (4), 1466–1479.
- Zumer, J., Attias, H., Sekihara, K., Nagarajan, S., 2007. A probabilistic algorithm integrating source localization and noise suppression for MEG and EEG data. *NeuroImage* 37 (1), 102–115.
- Zumer, J., Attias, H., Sekihara, K., Nagarajan, S., 2008. Probabilistic algorithms for MEG/EEG source reconstruction using temporal basis functions learned from data. *NeuroImage* 41 (3), 924–940.