A Non-Gaussian LCMV beamformer for MEG Source Reconstruction

Hamid R. Mohseni*§, Morten L. Kringelbach†§, Mark W. Woolrich‡, Tipu Z. Aziz§ and Penny Probert Smith*
*Institute of Biomedical Engineering, School of Engineering Science, University of Oxford, Oxford, UK
†Department of Psychiatry, University of Oxford, Warneford Hospital, UK
‡Oxford Centre for Human Brain Activity, Department of Psychiatry, University of Oxford
§Oxford Functional Neurosurgery, Nuffield Department of Surgery, John Radcliffe Hospital, Headington, Oxford, UK

***Abstract*—Evidence suggests that magnetoencephalogram (MEG) data have characteristics with non-Gaussian distribution, however, standard methods for source localisation assume Gaussian behaviour. We present a new general method for non-Gaussian source estimation of stationary signals for localising brain activity in the MEG data. By providing a Bayesian formulation for linearly constraint minimum variance (LCMV) beamformer, we extend this approach and show that how the source probability density function (pdf), which is not necessarily Gaussian, can be estimated. The proposed non-Gaussian beamformer is shown to give better spatial estimates than the LCMV beamformer, in both simulations incorporating non-Gaussian signal and in real MEG measurements.**

***Index Terms*—Magnetoencephalography, source reconstruction, non-Gaussian, LCMV-beamformer.**

I. Introduction

Magnetoencephalography (MEG) is a neuroimaging technique with excellent temporal resolution and reasonable spatial resolution. Unlike functional magnetic resonance imaging (fMRI) or positron emission tomography (PET), MEG provides a direct measurement of brain activity through recording the magnetic induction over the scalp produced by electrical activity in the neural cell assemblies [1]. What has been missing, however, is a way to precisely localise the sources of activity within the brain from the MEG measurements.

At the heart of this problem is the problem 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 the MEG signals. Some progress has been made using a number of different methods. One of the most successful and most common methods for localising brain activity with MEG is the LCMV beamformers [2]. However, as we show in this paper, its performance is optimal only in the presence of measurement whose probability density functions (pdfs) can be described using the first and second order statistics (i.e., Gaussian distribution).

The LCMV beamformer originally proposed for vector source localisation and later its important variation, scalar beamformer, was employed. It has been shown that these two types of beamformers are equivalent, in terms of output power and output SNR, if the beamformers pointing directions are optimised [3]. There are several other approaches that have been proposed to improve or extend the beamformer. Amongst them, we may refer to the graphical models for event-related field denoising and localisation [4], [5], which are modifications of [6] by incorporating the lead-field matrices and temporal information. In [5], in a Bayseian formulation, it has been shown that if the prior is assumed to be flat, the maximum a postriori (MAP) estimation is converted to the maximum likelihood estimation which in turn is equivalent to the LCMV beamformer. A Baysian paradigm has also been proposed to derive several MEG source localisation approaches including the LCMV beamformer [7]. We also later justify the LCMV beamformer from a Bayesian perspective in a way to be able to employ a non-Gaussian data distribution.

In this paper, we provide a framework to generalise the beamformer for non-Gaussian MEG data. The main motivation for this study is based on evidence which shows that MEG data has a non-Gaussian distribution [8], [9]. It is supported in [6], [10] which emphasise that evoked brain sources are often characterised by spikes or by modulated harmonic functions, leading to a non-Gaussian distribution. Moreover, there should be elements of the non-Gaussianity in the MEG data to be able to successfully employ the independent component analysis (ICA), which exploits non-Gaussian measures such as kurtosis [6], [10]. Therefore, a source localisation algorithm that captures the non-Gaussianity should perform better than the LCMV beamformer. The novel approach presented here uses Bayesian formulation of the beamformer, which reveals that the kernel density estimator [11], or any other multivariate density estimator, can be used to model the measurement pdf, and to develop a more accurate estimate of the source distribution.

II. Methods

1. Problem formulation

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

$$\mathbf{y}_t = \mathbf{F}\mathbf{s}_t + \mathbf{n}_t \tag{1}$$

where $\mathbf{F} = [\mathbf{f}_1 \ \ldots \ \mathbf{f}_q] \in \mathbb{R}^{N\times q}$ is the matrix of the lead-fields and $\mathbf{s}_t = [s_{1,t} \ \ldots \ s_{q,t}]^T \in \mathbb{R}^q$ 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 processe with pdfs $g_{\mathbf{y}}(.)$, $g_{\mathbf{s}}(.)$ and $g_{\mathbf{n}}(.)$, respectively. Furthermore, assume that each source $s_{i,t} \in \mathbb{R}$, $i \in \{1, \ldots, q\}$, which is a stationary process, has the pdf $g_{s_i}(.)$, and that the lead-field vector of each source $\mathbf{f}_i \in \mathbb{R}^N$ is deterministic and known using the Biot-Savart law [12].

2. 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 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 the location of interest, whilst minimising the covariance of the measurement $\mathbf{R_y}$:

$$\underset{\mathbf{w}}{\operatorname{argmin}}\ \mathbf{w}^T\mathbf{R_y}\mathbf{w}, \quad \textit{subject to}:\ \mathbf{w}^T\mathbf{f}_k = 1 \tag{2}$$

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

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

$$\mathcal{P}_{s_k} = (\mathbf{f}_k^T\mathbf{R_y}^{-1}\mathbf{f}_k)^{-1} \tag{4}$$

Now we examine equations (3) and (4) from a Bayesian perspective. One may rewrite equation (1) as:

$$\mathbf{y}_t = \mathbf{f}_k s_{k,t} + \boldsymbol{\eta}_t \tag{5}$$

where $\boldsymbol{\eta}_t = \sum_{i\neq k}\mathbf{f}_i s_{i,t} + \mathbf{n}_t$ is the interference coming from other sources plus noise. We then assume that the pdf for $\boldsymbol{\eta}_t$, $g_{\boldsymbol{\eta}}(.)$, can be approximated with a Gaussian function, i.e. $\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{\boldsymbol{\eta}})$. In the Bayes 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:

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

In equation (5), since we supposed that the interference plus noise distribution $g_{\boldsymbol{\eta}}(.)$ is Gaussian, the likelihood $p(\mathbf{y}_t|s_k)$ is also Gaussian. By assuming further that the prior $p(s_k)$ is uniform (absolutely no knowledge about the source pdf), it can be shown that the posterior pdf is expressed as (see e.g., [13]):

$$\begin{aligned} p(s_k|\mathbf{y}_t) &\propto \exp\Big[-(\mathbf{y}_t - \mathbf{f}_k s_k)^T\mathbf{R}_{\boldsymbol{\eta}}^{-1}(\mathbf{y}_t - \mathbf{f}_k s_k)\Big] \\ &\propto \exp\Big[-\big(s_k - (\mathbf{f}_k^T\mathbf{R}_{\boldsymbol{\eta}}^{-1}\mathbf{f}_k)^{-1}\mathbf{f}_k^T\mathbf{R}_{\boldsymbol{\eta}}^{-1}\mathbf{y}_t\big)^T \\ &\quad (\mathbf{f}_k^T\mathbf{R}_{\boldsymbol{\eta}}^{-1}\mathbf{f}_k)^{-1}\big(s_k - (\mathbf{f}_k^T\mathbf{R}_{\boldsymbol{\eta}}^{-1}\mathbf{f}_k)^{-1}\mathbf{f}_k^T\mathbf{R}_{\boldsymbol{\eta}}^{-1}\mathbf{y}_t\big)\Big] \end{aligned} \tag{7}$$

Therefore, the estimated power and the estimated time-series are given by:

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

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

Equations (8) and (9) are equivalent to equations (3) and (4), expect for the fact that in the former $\mathbf{R}_{\boldsymbol{\eta}}$ 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 other sources plus noise, we can approximate $\mathbf{R}_{\boldsymbol{\eta}} \approx \mathbf{R_y}$, making the beamformer and Bayesian solutions equivalent.

The minimisation of the noise plus interference $\mathbf{R}_{\boldsymbol{\eta}}$ in equation (2) is better known as the minimum variance distortionless response (MVDR) beamformer. This should theoretically perform better than minimising the measurement power $\mathbf{R_y}$ (LCMV beamformer) [14], [15]. However, the difficulty for estimating the noise power in most applications means that the LCMV beamformer is generally preferred [16].

It is notable that if one uses $\mathbf{R}_{\boldsymbol{\eta}} + \alpha\mathbf{f}_k\mathbf{f}_k^T$ (where α is an arbitrary number) instead of $\mathbf{R}_{\boldsymbol{\eta}}$ in equations (8) and (9), the result does not change. This means that if signal and noise are independent, we have $\mathbf{R_y} = \mathbf{R}_{\boldsymbol{\eta}} + \alpha\mathbf{f}_k\mathbf{f}_k^T$, and therefore the LCMV-beamformer and Bayesian solutions (MVDR beamformer) are the same.

3. The Non-Gaussian Probability Distribution Beamformer

The LCMV beamformer assumes that the data is zero-mean. It minimises only the second order statistics and ignores the higher orders. Our aim is to estimate the posteriori using an arbitrary pdf, and not necessarily a Gaussian one.

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

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

As with the classic Gaussian LCMV beamformer, which replaces the interference plus noise power with the total power, we assume that the pdf of $\boldsymbol{\eta}$ can be approximated by the pdf of $\mathbf{y}$; i.e., $g_{\boldsymbol{\eta}}(.) \approx g_{\mathbf{y}}(.)$. This is equivalent to implicitly assume that the power of one voxel is considerably smaller than the power of all other voxels within the brain plus noise, which is true in the most application of the MEG source localisation. This assumption therefore gives equation (10) as:

$$p(s_k|\mathbf{y}_t) \propto g_{\mathbf{y}}(\mathbf{y}_t - \mathbf{f}_k s_k) \tag{11}$$

Here, $g_{\mathbf{y}}(.)$ is obtained using the mechanism that is explained in Section II.4. We may then use the expected value of the posteriori as the estimation of the time-series $\hat{s}_{k,t}$:

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

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

In addition to the posteriori distribution $p(s_k|\mathbf{y}_t)$, it is necessarily to obtain the source pdf $g_{s_k}(s_k)$ in order to estimate the source activity, which can be represented by the variance (or fourth order moments) of $g_{s_k}(s_k)$. In this study, we use $g_{s_k}(s_k) \equiv p(s_k|E\{\mathbf{y}\})$ as estimation of the source pdf, where $E\{.\}$ is the expectation operator over $g_{\mathbf{y}}(.)$. Using equation (11) and the fact that the measurement is zero mean, we have

$$g_{s_k}(s_k) \propto g_{\mathbf{y}}(\mathbf{f}_k s_k) \tag{13}$$

Here, we have replaced $g_{\mathbf{y}}(-\mathbf{f}_k s_k)$ by $g_{\mathbf{y}}(\mathbf{f}_k s_k)$, since the minus sign does not have any impact on the estimated power.

Alternatively, one may use $g_{s_k}(s_k) \equiv p(s_k|\mathbf{y}_1, \dots \mathbf{y}_n)$, leading to the relationship $g_{s_k}(s_k) \propto \prod_{t=1}^{T} g_{\mathbf{y}}(\mathbf{y}_t - \mathbf{f}_k s_k)$ (based on the independency of measurement samples). However, estimation of the source pdf using (13) is much faster especially when it is required to estimate the power of a large number of points inside the brain volume.

For the convenience of the reader, 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 $\mathbf{f}_k$.

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

% estimating the time-series $\hat{s}_{k,t}$
for $t = 1$ to T **do**
  estimate $g_{\mathbf{y}}(\mathbf{y})$
  set $p(s_k|\mathbf{y}_t) \propto g_{\mathbf{y}}(\mathbf{y}_t - \mathbf{f}_k s_k)$
  normalise $p(s_k|\mathbf{y}_t) = \frac{p(s_k|\mathbf{y}_t)}{\int p(s_k|\mathbf{y}_t) ds_k}$
  set $\hat{s}_{k,t} = \int s_k p(s_k|\mathbf{y}_t) ds_k$
end for

% fast estimation of the power $\mathcal{P}_{s_k}$ at a location with lead-field $\mathbf{f}_k$
estimate $g_{\mathbf{y}}(\mathbf{y})$
set $g_{s_k}(s_k) \propto g_{\mathbf{y}}(\mathbf{f}_k s_k)$
normalise $g_{s_k}(s_k) = \frac{g_{s_k}(s_k)}{\int g_{s_k}(s_k) ds_k}$
set $\mathcal{P}_{s_k} = \int {s_k}^2 g_{s_k}(s_k) ds_k$

4. A Kernel Based Estimation of the Measurement pdf

We have shown that how the source pdf can be estimated from the data pdf. To implement the method, the measurement pdf $g_{\mathbf{y}}(.)$ must be estimated from the set of discrete observations $\mathbf{y}_t, t \in \{1, \dots, T\}$. The success of the method depends 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 requiring no optimisation procedure. Kernel based estimation, which is one of the most common methods in the multivariate density estimation techniques, assumes that the pdf is the sum of T kernels. In other words, we place the centre of each kernel at a single observation [17]; hence T is equal to the number of observations.

From equation (13), the pdf at a location with lead-field matrix $\mathbf{f}_k$ is given as follows:

$$g_s(s) = \frac{\sum_{t=1}^{T} \frac{1}{h^N} K(\frac{\mathbf{y}_t - \mathbf{f}_k s}{h})}{\int \sum_{t=1}^{T} \frac{1}{h^N} K(\frac{\mathbf{y}_t - \mathbf{f}_k s}{h}) ds} \tag{14}$$

where $K : \mathbb{R}^N \to \mathbb{R}$ is the kernel and h is a scaling factor known as the bandwidth. The most common kernel in the pdf estimation is given by [17]:

$$K(\mathbf{x}) \propto \exp(-\frac{1}{2}\mathbf{x}^T \mathbf{R}_{ker}^{-1} \mathbf{x}) \tag{15}$$

where $\mathbf{R}_{ker}$ is the covariance of the kernel. The simplest choice for the covariance of the kernel is the identity matrix $\mathbf{R}_{ker} = \mathbf{I}$, leading to a homogeneous kernel.

An alternative 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 $\mathbf{R}_{\mathbf{y}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ is the singular value decomposition of the estimated measurement covariance matrix, and set $\mathbf{R}_{ker} = \mathbf{U}_{max}\mathbf{U}_{max}^T + \sigma\mathbf{I}$. Here, σ is a small number used to stabilise the calculation of the inverse, because $\mathbf{U}_{max}\mathbf{U}_{max}^T$ is rank deficient. This choice may give better performance than choosing of $\mathbf{R}_{ker} = \mathbf{I}$, for instance when some of the eigenvalues of $\mathbf{R}_{\mathbf{y}}$ are much smaller than the others.

III. Experimental Results

1. Simulation Experiments

In the first simulation experiment, the method was demonstrated using the reconstruction of a source with three different pdfs which have been shown in Fig. 1. It was assumed that in addition to the source of interest, there were 15 uncorrelated Gaussian sources whose locations were randomly chosen inside the brain volume. Zero-mean Gaussian white noise was also added to the simulated MEG data (after multiplying the lead-fields and simulated sources) to set SNR = 8dB. SNR is defined in the sensor space as the ratio between the 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 (left) shows the result for the Gaussian source where both beamformers give an accurate estimation. Fig. 1 (middle) and (right) show the results for two non-Gaussian sources. It is clear that the non-Gaussian beamformer increasingly outperforms the LCMV beamformer as the source pdf deviates increasingly from a Gaussian distribution.

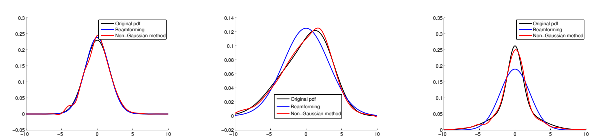
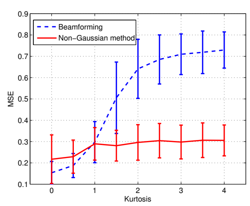


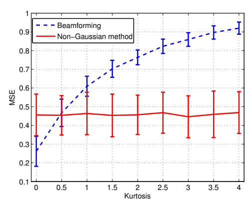
Fig. 1: Examples of the simulated source pdfs (black lines) and their estimations using LCMV beamformer (blue lines) and the non-Gaussian method (red lines).

In the next simulation experiment, the mean square 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 described using 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, while as it becomes larger, the pdf departs further from a Gaussian profile. When the variances of two distributions are identical, the mixture is also Gaussian. Fig. 2(a) presents the result of the mean squared error (MSE) between the original and constructed pdf for different kurtosis. The average SNR is $-5dB$ and the results were obtained using 1000 Monte Carlo simulations. It is clear that as the kurtosis increases, the error in the LCMV beamformer increases rapidly. In contrast, the new non-Gaussian beamformer exhibits a flat error which means it is not sensitive to the shape of the pdf.

The next experiment investigated additive non-Gaussian noise. In this case, all sources were Gaussian and the noise was generated in a similar way to 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 $\sigma\mathbf{I}$, where σ 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.



(a)



(b)

Fig. 2: Comparison between LCMV beamformer and the non-Gaussian methods 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.

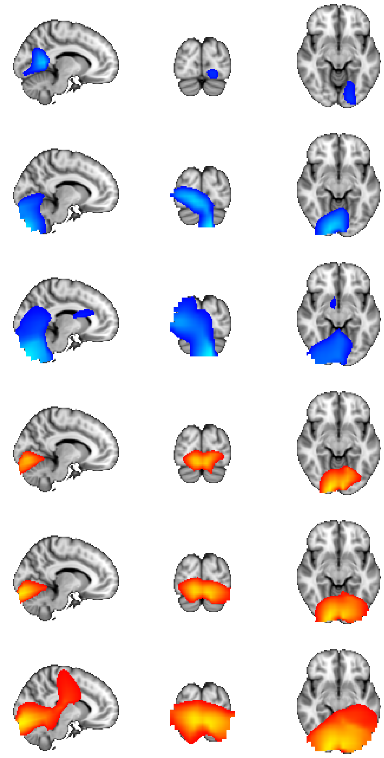


Fig. 3: Reconstruction of neural activity in the visual paradigm using beamforming with $\lambda = 0.01\%$, $\lambda = 0.1\%$ and $\lambda = 1\%$ (blue volumes from top to bottom) of trace of data covariance matrix, and using non-Gaussian method with $h = 1$, $h = 5$ and $h = 10$ (red volumes from top to bottom).

2. *Experiments using MEG Data*

In the following experiments, we compared the performance of the methods for source reconstruction and investigated the sensitivity of the results with regards to two key parameters, the regularisation factor λ and the kernel bandwidth h.

In this experiment, we used a visual paradigm where a series of human and animal faces were presented to the participant. Each image was presented for 300ms and the time interval between images was 1500ms. We were interested in localising a significant peak which occurs around 100–150ms after the stimulus onset. The origin of this peak is known to be located in bilateral regions of the primary visual cortex.

Figs. 3(blue volumes) show the output power of the beamformer (trace of estimated covariance matrix as in equation (4)) normalised by the norm of the associated lead-fields; i.e. normalised by the power of projected white Gaussian noise (see [2]). Figs. 3(red volumes) show the estimated power using the non-Gaussian 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 observe that if a small value for λ is chosen, the peak of the spectrum is in a frontal part of visual cortex, whereas if a large value is chosen, the peak of the spectrum is located incorrectly in the cerebellum rather than visual cortex. In contrast, the non-Gaussian method reconstructs activity in both the right and left primary visual cortex for moderate and small values of h. It also has better spatial resolution.

IV. Conclusion

In this paper, we provided a framework for source reconstruction based on the observation distribution. We used the kernel based density estimator to estimate the data distribution because of their simple implementations. Based on Algorithm 1, any other multivariate density estimator such as mixture of Gaussian or multivariate Edgeworth series can also be employed. Analysis of the impact of these approaches on the performance of the localisation is beyond this paper and we consider it as a future work.

The implementation of the new method showed that it outperforms existing LCMV beamforming method in terms of spatial and temporal source estimation in data from simulation and from real MEG experiments. In summary, this new class of non-Gaussian beamformer shows great promise in enhancing the source localisation with MEG, which in turn may enhance our understanding of human brain function.

REFERENCES

[1] P. Hansen, M. L. Kringelbach, and R. Salmelin, *MEG. An introduction to methods*. Oxford: Oxford University Press, 2010.

[2] B. D. V. Veen, W. V. Dronglen, M. Yuchtman, and A. Suzuki, "Localization of brain electrical activity via linearly constrained minimum variance spatial filtering," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 9, pp. 867–880, 1997.

[3] K. Sekihara, S. S. Nagarajan, D. Poeppel, and A. Marantz, "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.*, vol. 49, no. 12, pp. 1534–1546, 2006.

[4] J. Zumer, H. Attias, K. Sekihara, and S. Nagarajan, "A probabilistic algorithm integrating source localization and noise suppression for MEG and EEG data," *Neuroimage*, vol. 37, no. 1, pp. 102–115, 2007.

[5] ——, "Probabilistic algorithms for MEG/EEG source reconstruction using temporal basis functions learned from data," *Neuroimage*, vol. 41, no. 3, pp. 924–940, 2008.

[6] S. Nagarajan, H. Attias, K. Hild, and K. Sekihara, "A graphical model for estimating stimulus-evoked brain responses from magnetoencephalography data with large background brain activity," *NeuroImage*, vol. 30, no. 2, pp. 400–416, 2006.

[7] D. Wipf and S. Nagarajan, "A unified bayesian framework for MEG/EEG source imaging," *NeuroImage*, vol. 44, no. 3, pp. 400–416, 2009.

[8] R. Elul, J. Hanley, and J. Q. Simmons, "Non-gaussian behavior of the EEG in down's syndrome suggests decreased neuronal connections," *Acta Neurologica Scandinavica*, vol. 51, pp. 21–28, 1975.

[9] T.-W. Lee, M. Lewicki, and T. Sejnowski, "ICA mixture models for unsupervised classification of non-Gaussian classes and automatic context switching in blind signal separation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 10, pp. 1078–1089, 2000.

[10] S. Nagarajan, H. Attias, K. Hild, and K. Sekihara, "Stimulus evoked independent factor analysis of meg data with large background activity," in *Adv. Neural Inf. Process. 18.*, 2005.

[11] E. Parzen, "On estimation of a probability density function and mode," *Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.

[12] J. Mosher, R. Leahy, and P. Lewis, "EEG and MEG: forward solutions for inverse methods," *Biomedical Engineering, IEEE Transactions on*, vol. 46, no. 3, pp. 245–259, 1999.

[13] K. Friston, D. Glaser, R. Henson, S. Kiebel, C. Phillips, and J. Ashburner, "Classical and Bayesian inference in neuroimaging: theory," *NeuroImage*, vol. 16, no. 2, pp. 465–483, 2002.

[14] H. Van Trees, *Detection, Estimation, and Modulation Theory, Radar-Sonar Signal Processing and Gaussian Signals in Noise, part IV*. Wiley-Interscience, 2004.

[15] L. Ehrenberg, S. Gannot, A. Leshem, and E. Zehavi, "Sensitivity analysis of MVDR and MPDR beamformers," in *Electrical and Electronics Engineers in Israel (IEEEI), 2010 IEEE 26th Convention of*, 2010, pp. 000 416–000 420.

[16] D. Wipf and S. Nagarajan, "Beamforming using the relevance vector machine," in *Proceedings of the 24th international conference on Machine learning*, 2007, pp. 1023–1030.

[17] B. W. Silverman, *Density estimation for statistics and data anaysis*. Chapman and Hall/CRC, 1999.