

# Sources of Variability in MEG

Wanmei Ou<sup>1</sup>, Polina Golland<sup>1</sup>, and Matti Hämäläinen<sup>2</sup>

<sup>1</sup> Department of Computer Science and Artificial Intelligence Laboratory, MIT, USA

<sup>2</sup> Athinoula A. Martinos Center for Biomedical Imaging, MGH, USA

**Abstract.** This paper investigates and characterizes sources of variability in MEG signals in multi-site, multi-subject studies. Understanding these sources will help to develop efficient strategies for comparing and pooling data across repetitions of an experiment, across subjects, and across sites. In this work, we investigated somatosensory MEG data collected at three different sites and applied variance component analysis and nonparametric KL divergence analysis in order to characterize the sources of variability. Our analysis showed that inter-subject differences are the biggest factor in the signal variability. We demonstrated that the timing of the deflections is very consistent in the early somatosensory response, which justifies a direct comparison of deflection peak times acquired from different visits, subjects, and systems. Compared with deflection peak times, deflection magnitudes have larger variation across sites; modeling of this variability is necessary for data pooling.

## 1 Introduction

Magnetoencephalography (MEG) is a noninvasive technique for investigating neuronal activity in the living human brain [4]. In contrast to functional magnetic resonance imaging (fMRI) which measures the hemodynamic changes associated with neuronal activity, MEG is directly related to the electric currents in neurons and thus has an excellent temporal resolution of milliseconds. Because of its potential in revealing the precise dynamic of neuronal activations, MEG is popular in neuroscience research, and it has started to move toward clinical applications such as presurgical planning for epileptic patients [8].

Testing interesting neurophysiological hypotheses often require a large number of subjects. However, the number of subjects or patients with a particular disease available at a certain location is often limited. Pooling data from multiple imaging centers is clearly helpful to overcome this limitation. At present, there are three different MEG systems, employing different sensor coil geometries. Therefore, it is important to assess possible variability in the data obtained from different systems.

In this work, we examine the data collected in a multi-site MEG study administered by the MIND institute. The goals of this study parallel the analogous projects in fMRI [14, 15]. Before pooling the MEG data, one must study the degree of consistency in the data generated from different systems and model the system bias in the combined data set. The MIND multi-site MEG project includes a calibration program to assess inter-trial, inter-visit, inter-subject, and inter-site variability, which are quantitatively explored in this paper.

The sources of variability can be studied either in the signal space (MEG sensor measurement) or in the source space (after solving the inverse problem).

Hämäläinen *et al.* [5] focused on the inter-scanner variability in the signal space using a minimum-norm estimate based extrapolation method. Closely matched extrapolated and true data demonstrated excellent reproducibility of MEG data across the three systems. On the other hand, the source estimates relate more directly to the neuronal phenomenon of interest. Weisend *et al.* [13] reported consistent source localization when using data from the same subject on different MEG systems. These two approaches were tailored to study inter-system variability only. Zou *et al.* [15] have performed an in-depth study in a multi-site fMRI project. Framed as a detection problem, they applied an expectation-maximization algorithm to assess the sensitivity and specificity from run, subject, and scanner. Because of its excellent temporal resolution, we focus on the timing and the magnitude of deflections when analyzing MEG data.

The main contribution of our work is the investigation of many possible sources of variability of the estimated current sources underlying the early somatosensory MEG responses. Due to limitation in our current registration algorithms, we defer the spatial characterization for a future study. The rich temporal information in MEG data enables us to extend the comparison to the single-trial level. Compared with the results of the prior work based on the averages of hundreds of trials [5, 13], our results reveal stronger consistency between the systems, and within each subject. We employ two approaches in characterizing the source of variability: the variance component analysis (VCA), which assumes a Gaussian model, and the nonparametric Kullback-Leibler (KL) divergence analysis to directly measure the differences between two sets of data. Our results show that the inter-subject difference is the strongest cause of variance. We also conclude that the peak time of early deflections is directly comparable across visits, subjects, and sites, but the variation in deflection magnitude across sites needs to be modeled for data pooling.

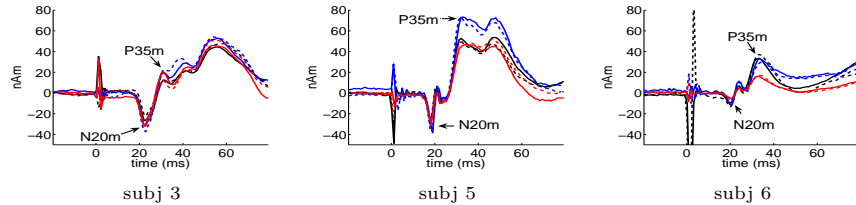
In the next section, we describe the multi-site MEG data and possible sources of variability. We then present the analysis methods in Section 3 and results in Section 4, followed by conclusions.

## 2 Data and Sources of Variability

In this work, we analyze MEG data acquired by the MEG Consortium supported by the MIND Institute. Six normal subjects were scanned at three different MEG sites with two visits to each site<sup>3</sup>. Each visit comprised experiments with three different types of stimuli: auditory, somatosensory, and visual. The three MEG systems employed were the 306-channel Neuromag VectorView system at Massachusetts General Hospital (Boston, MA), the 248-channel 4D Neuroimaging Magnes 3600 WH system at University of Minnesota (Minneapolis, MN), and the 275-channel VSM MedTech Omega275 system at the MIND Imaging Center (Albuquerque, NM). We will subsequently refer to the three MEG systems as MGH, UMN, and MIC, respectively. Anatomical images were collected for each subject with a Siemens Avanto 1.5 T scanner at the MGH site.

We analyze the data from somatosensory median-nerve stimulation, with on average  $N_{\text{tot}} = 300$  trials per visit after rejecting trials with eye-movement and

<sup>3</sup> Subject 2 and 4 had scans in two out of the three sites.



**Fig. 1.** SI dipole timecourses in three subjects over two visits in three sites estimated from the average timecourse for each visit. Solid line denotes visit 1 and dashed line denotes visit 2. Curves with different colors present signals obtained from different MEG systems: MGH (black), MIC (blue), and UNM (red).

other artifacts. It has been shown that this simple stimulus activates a complex cortical network [6]. The first activation of the contralateral primary sensory cortex (SI) peaks around 20 ms and continues over 100 ms; then the secondary sensory cortex (SII) activates bilaterally around 70 ms and lasts up to 200 ms, during which the posterior parietal cortex may also activate. Whether SI and SII form a sequential architecture or parallel architecture is still a topic of active debate [7, 11]. Although the SI-SII network exhibits robust activation, there is significant variation from trial to trial especially for SII due to physiological noise. In this initial study, we focus on two prominent and stable early deflections in SI: N20m and P35m, illustrated in Fig. 1.

Due to the structure of the data, variation can be assessed across different trials within a single visit, between visits at a single site, among subjects, and among MEG systems. MEG system variation is a result of hardware differences (number of sensors, sensor type/position, and magnetically shielded rooms) and software differences (methods of noise cancellation and filtering parameters). Subject variation reflects differences in neuronal mechanisms [6] and brain anatomy. Changes in environmental noise and in the relative head position contribute to variation between data obtained from two visits. Variations in the neuronal state and the subject’s movement often lead to variation across trials. Assessing the contributions of these different sources of variability will help to improve design and analysis of future multi-site studies.

As an illustration, Fig. 1 presents example SI dipole timecourses, estimated from the averages signals of all trials, 300 on average, over a visit. We will describe how to obtain these timecourses in the next section. The peak time and magnitude of N20m and P35m in Subject 3 match across visits and sites. Subject 5’s responses match except for one site where the subject received stimulus of different strength due to different stimulation electrodes used. We believe the magnitude mismatch in Subject 6 is due to physiological variation in the signal. Since the physiological variation has different effect in each site for this subject, it will be summarized as site variation in VCA. Future studies are needed to separate the physiological variation from site variation by including a control group being scanned at the same site with a long time interval.

### 3 Analysis

This section describes preprocessing and two analysis methods used in this paper: variance component analysis (VCA) and the nonparametric KL divergence analysis. After registering the MEG data to the MRI scan with help of fiducial

on the scalp surface, we fitted the average signals from 18 to 35 ms after the stimulus onset with a single equivalent current dipole (ECD) using the Nelder-Mead simplex algorithm [10] (Fig. 1). In all data sets, the goodness of fit at major deflections was 70~98%, which is above the standard threshold. We chose dipole fitting rather than distributed source estimation [3, 4, 12], because it is reasonable to assume a single focal source, SI, in such an early response period.

Rich temporal information and consistent deflection timing in the average dipole timecourses encouraged us to investigate the degree of consistency at the single trial level. We extracted the SI response from each trial by projecting the single-trial data onto the field pattern of the dipole fitted to the averaged data.

The resulting single-trial responses were similar to the mean responses shown in Fig. 1. However, their low signal-to-noise ratio (SNR) caused ambiguity in identifying N20m and P35m. To enable reliable automatic detection of N20m and P35m, we employed a random sampling approach: we averaged  $N$  randomly sampled timecourses before applying the detector. We inspected the detection results with varying  $N$ , and found that  $N \geq 4$  provided sufficient SNR in the average response for accurate detection. Only a minor distortion in timing and amplitude was introduced from averaging of such small number of trials. In this work, we set  $N = 5$ , and refer to the averages of randomly sampled sets of five timecourses as “single-trial” experiments. This is in contrast to the commonly used approach of averaging over hundreds of trials, which loses much of temporal detail in the resulting timecourses. All the analysis results presented in this paper are based on  $10^5$  such “single-trials” per visit, subject, and site.

Our peak detection algorithm searches for extrema in the SI responses. To improve robustness of the method, we employ high-order derivatives estimated over a broad support. We experimented with several different robust detectors, including wavelet decomposition, arriving at qualitatively similar conclusions. Due to space limitation, we omit the details of the peak detection.

### 3.1 Model-Based Variance Component Analysis (VCA)

VCA is a common approach to quantifying sources of variability in data [2]. It models the observation as the sum of an unknown true mean  $\mu$  and errors introduced by each source. Each error term is assumed to be independently generated by a zero-mean Gaussian distribution with an unknown variance. The variance estimates provide a measure of how much each source contributes to the total variance in the data. Moreover, we compute the relative variability, by normalizing the variance estimates to sum to one, to compare results from different characteristics, such as peak time and magnitude.

We set the observation  $t_{ijkl}$  to be the peak time or the magnitude of N20m or P35m, corresponding to the observation from site  $i$ , subject  $j$ , visit  $k$ , and trial  $l$ <sup>4</sup>. Due to the structure of the data, we model  $t_{ijkl}$  as a cross-hierarchical combination of sources of variation from trials, visits, subjects, and sites:  $t_{ijkl} = \mu + a_i + b_j + c_{ijk} + d_{ijkl}$  where  $\mu$  is the true but unknown value of the observation, and  $a_i$ ,  $b_j$ ,  $c_{ijk}$ , and  $d_{ijkl}$  quantify the deviations from site  $i$ , subject  $j$ , visit  $k$ ,

<sup>4</sup> Each “trial” here again refers to an average of 5 randomly selected trials.

and trial  $l$ , respectively; they are assumed to be samples of independent Gaussian random variables. For example,  $\mathcal{N}(0, \sigma_b^2)$  describes the distribution of  $b_j$ , and  $\sigma_b^2$  indicates inter-subject variability.

To improve the robustness of the estimation, we take the Bayesian approach with weak priors on the parameters. The prior distributions are  $\mathcal{N}(0, 10^{10})$  for  $\mu$  (a common weak prior for the mean) and independent  $\text{Gamma}(0.01, 0.01)$  for the precision parameters:  $\frac{1}{\sigma_a^2}$ ,  $\frac{1}{\sigma_b^2}$ ,  $\frac{1}{\sigma_c^2}$ , and  $\frac{1}{\sigma_d^2}$ . We estimate the model parameters as medians of the corresponding posterior distributions. Due to the complex cross-hierarchy in the model, we employ Gibbs sampling, as implemented in BUGS software [1], to perform inference for parameters of interest. We use  $10^5$  burn-in samples, and inferences are based on another  $10^5$  samples.

### 3.2 Nonparametric KL Divergence Analysis

While VCA is powerful in quantifying variability, it cannot capture variation beyond second order statistics due to the Gaussian assumption. To overcome this limitation, we directly compare the distributions of the extracted parameters of the N20m and P35m deflections. We then employ the symmetrized KL divergence<sup>5</sup> [9] to quantify differences between the two distributions directly from the histogram. Results are presented as a distance matrix.

In this work, we construct the distributions separately for peak time and magnitude of a deflection. We defer the nonparametric KL divergence analysis to joint distributions of peak time and magnitude for future exploration.

To summarize, while VCA provides a generative model to quantify the variability of each source, it is limited by the Gaussian assumption. On the other hand, the nonparametric KL divergence analysis captures differences between distributions beyond second order statistics. However, it does not separate the variation due to different sources. Applying both approaches better characterizes the variability in the data set.

## 4 Results

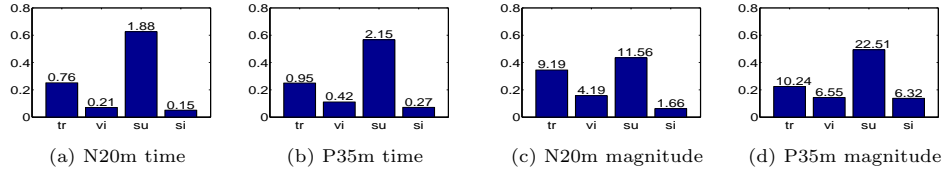
This section presents results obtained from VCA and the nonparametric KL divergence analysis.

### 4.1 Model-Based Variance Component Analysis

Fig. 2 presents VCA results for peak time and magnitude for N20m and P35m. The numbers on top of the bars denote the estimated variances of the corresponding sources, and the height of the bars is proportional to relative variability.

The relative variability of the peak time is similar for N20m and P35m. While the relative variability for subject and trial are about 60% and 25%, the relative site variability is less than 5%. This is not entirely surprising. Due to high temporal resolution, different MEG systems can precisely capture when the deflections occur. Therefore, peak timing is directly comparable for data

<sup>5</sup>  $D_{\text{sym}}(p_1||p_2) = \frac{1}{2}(D(p_1||p_2) + D(p_2||p_1))$ , where  $p_1$  and  $p_2$  are two probability distributions, and  $D$  denotes the KL divergence.



**Fig. 2.** Relatively variability (bar) and estimated variance (top of the bars) for N20m peak time (a), P35m peak time (b), N20m magnitude (c) and P35m magnitude (d), respectively. The sources of variability are trial (tr), visit (vi), subject (su), and site (si).

generated from the three systems. Little or no adjustment is required in pooling data across different systems and visits.

Both the estimated variance and the relative variability for N20m magnitude suggest that the site variability is small. On the other hand, the relatively large site variability in P35m magnitude suggests adjustment is needed for data pooling. If the error distributions from each source closely follow a Gaussian distribution, a simple approach is to subtract the estimated site bias  $b_j$  obtained from the current calibration study.

The variance estimates are larger in P35m than N20m for both peak time and magnitude. This observation agrees with the general understanding that deflections tend to vary more as they are further away from the stimulus onset because a more complex network is often involved in their generation and several connections can affect the signal timing and magnitude.

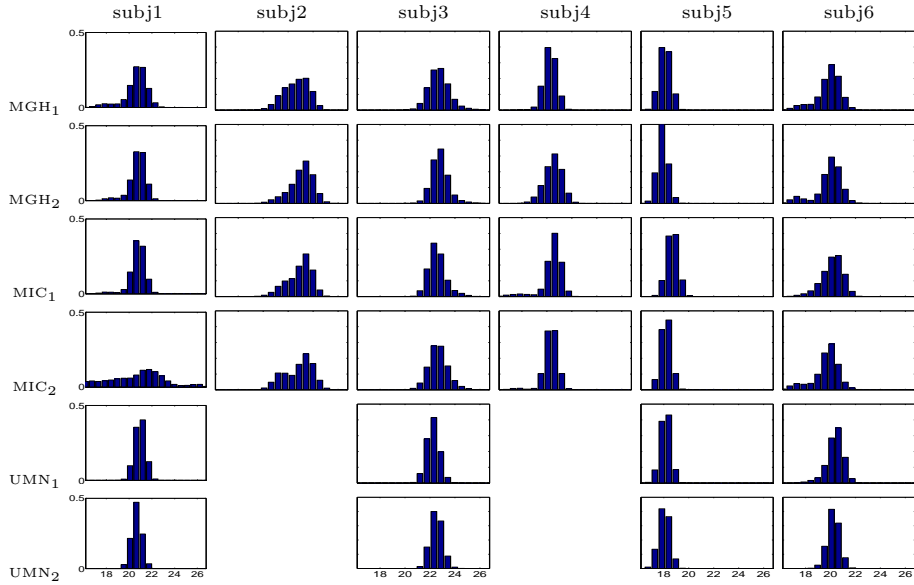
## 4.2 Nonparametric KL Divergence Analysis

To investigate components of variability that are not captured by the Gaussian model, we applied the nonparametric KL divergence analysis. Fig. 3 presents the normalized histograms of N20m peak time each composed of  $10^5$  random samples for data obtained from each visit, subject, and site. The consistency of single trial responses is reflected by the matched distributions across sites and visits, which is a much stronger evidence than consistent deflection peak time in the average response of hundreds of trials. For example, subject 2 exhibits consistently skewed distributions. Due to some presently unknown experimental problem, inconsistency occurs in subject 1’s second visit to MIC. This data is removed from further analysis<sup>6</sup>. There was a change in stimulus strength for subject 5 during his/her visit to MIC. We can observe small delay in one of the histograms, and note that further investigation is needed to understand the relationship between stimulus strength and deflection peak time.

The symmetrized KL divergence between the histograms is depicted in Fig. 4(a). Each row or column corresponds to one visit of a subject to a particular site. There are four or six rows in a  $s_{ui}$ - $s_{uj}$  block depending on whether subject  $i$  has MEG scans in two sites or all three sites<sup>7</sup>. Small KL divergence in the highlighted blocks along the diagonal further confirms that the N20m peak time is consistent within each subject and is independent of visit days and MEG systems. By capturing higher order statistics, the KL divergence analysis conveys

<sup>6</sup> This data was also excluded in VCA.

<sup>7</sup>  $s_{u1}$ - $s_{uj}$  block has five rows because subject 1’s second visit to MIC was discarded.



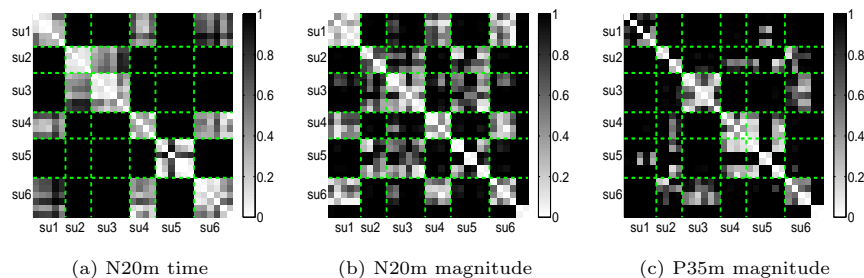
**Fig. 3.** Normalized histograms of N20m time from  $10^5$  random subsamples of the data.

a stronger message than VCA: N20m peak time is directly comparable across sites and visits. P35m’s peak time exhibits very similar behavior (not shown).

The site effect is much more pronounced when we consider the deflection magnitude. Change in the stimulus strength is clearly reflected by the larger KL divergence between subject 5’s data obtained from MIC and all other sites. In general, there is more site variability in P35m magnitude than in N20m magnitude for a single subject, with subject 1 being the most prominent example. While these results agree with the general trend observed in VCA, they imply site variability may be larger than that estimated by the model-based analysis. This suggests a refined VCA that relaxes the Gaussian assumption is necessary to accurately capture the variability in this MEG data.

## 5 Conclusions

Our study demonstrated that inter-subject effect is the largest contributor to the variability of the MEG data. We analyzed variability due to site, subject, visit, and trial effects in MEG data using variance components analysis and nonparametric KL divergence analysis. The two analysis methods established that we can directly compare deflection peak time across systems and visits. However, system effects on the deflection magnitude should be modeled for data pooling. Subject 6’s data suggests that the site effect may originate from that subject’s physiological variability. Our random sampling approach illustrated that the timing of deflections is highly consistent even at the single-trial level. Hence, average across a large number of trials is not necessary. Histograms built upon averages of a small number of single trials can better capture the inter-subject differences. The increased sensitivity of such an approach can be helpful in studying differences in MEG responses between normal subjects and clinical populations.



**Fig. 4.** (a) Symmetrized pairwise KL divergence of histograms in presented Fig. 3. (b) and (c) Symmetrized KL divergence for N20m magnitude and P35m magnitude, respectively.

*Acknowledgments* This work was supported in part by NIH NIMIC U54-EB005149, NCRN mBIRN U24-RR021382, P41-RR13218 and P41-RR14075 grants, by the NSF CAREER Award 0642971, and by U.S. DOE Award Number DE-FG02-99ER62764 to the MIND Institute. Wanmei Ou is partially supported by the NSF graduate fellowship.

## References

1. Bayesian inference using Gibbs sampling. <http://www.mrc-bsu.cam.ac.uk/bugs/>.
2. G. Box and G. Tiao. Bayesian inference in statistical analysis. *Wiley*, 1992.
3. A. Dale and M. Sereno. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. *J. Cog. Neurosci.* 5:162–176, 1993.
4. M.S. Hämäläinen, *et al.* Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65:413-497, 1993.
5. M.S. Hämäläinen, *et al.* Comparison of signals acquired with different MEG systems using an extrapolation method based on minimum-norm estimates. *In Proc. Biomag*, 2006.
6. R. Hari, and N. Forss. Magnetoencephalography in the study of human somatosensory cortical processing. *Phil. Trans. R. Soc. Lond. B*, 354:1145–1154, 1999.
7. J. Kass, *et al.* Multiple representations of the body within the primary somatosensory cortex of primates. *Science*, 204:521–523, 1979.
8. S. Knake, *et al.* The value of multichannel MEG and EEG in the presurgical evaluation of 70 epilepsy patients. *Epilepsy Res.*, 69:80–86, 2006.
9. S. Kullback and R.A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.
10. J.A. Nelder and R. Mead. A Simplex Method for Function Minimization. *Comput. J.* 7:308-313, 1965.
11. M. Rowe, *et al.* Parallel organization of somatosensory cortical areas I and II for tactile processing. *Clin. Exp. Pharmacol. Physiol.*, 23:931–938, 1996.
12. K. Uutela, M.S. Hämäläinen, and E. Somersalo. Visualization of magnetoencephalographic data using minimum current estimates. *NeuroImage* 10:173-180, 1999
13. M.P. Weisend, *et al.* Paving the way for cross-site pooling of magnetoencephalography (MEG) data. *In Proc. Biomag*, 2006.
14. A. Yendiki, *et al.* Multi-Site Characterization of an fMRI Working Memory Paradigm: Reliability of Activation Indices. *in Proc. HBM*, 2006.
15. K.H. Zou, *et al.* Reproducibility of functional MR imaging: preliminary results of prospective multi-institutional study performed by Biomedical Informatics Research Network. *Radiology*, 237:781–789, 2005.