Statistical model identification in electromagnetic source analysis

Waldorp, L.J.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
1 Introduction

With the advent of noninvasive techniques to study the human brain, correlating brain structure and function (functional mapping) and clinical assessment of brain pathologies have become increasingly popular [136], [34]. An example of functional mapping is temporal and spatial localization of visual information processing of spatial frequencies [61]. An example of clinical assessment is localization of brain lesions in patients [59].

Several (functional) noninvasive techniques are available. To name a few: positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and electro- and magnetoencephalography (EEG and MEG). PET measures blood flow and has a spatio-temporal resolution of approximately 4-6mm and 30-45sec. fMRI measures oxygenation of blood and has a spatio-temporal resolution of approximately 1-3mm and 1-5sec (e.g., [119], [32]). EEG and MEG measure the electromagnetic activity and have a spatio-temporal resolution of approximately 2-15mm and 1ms [42], [119]. Although EEG and MEG have a low spatial resolution compared to PET and fMRI, their temporal resolution makes them ideally suited for measuring cognitive processes [34], [47]. Processing visual information on spatial frequencies, for example, has a time span of only several hundreds of milliseconds [61].

EEG and MEG refer to sensor measurements of the potential and magnetic field respectively. Usually, multiple sensors at multiple time points (samples) are measured. Since it is generally believed that synchronous electrical activity of millions of neurons in the cortex generates the EEG and MEG [118], [42], describing the measured EEG and MEG in terms of underlying sources seems warranted. Identifying the underlying sources of the measured EEG and MEG is the objective of electromagnetic source analysis (EMSA). One of the problems associated with this technique is that there is in general no unique solution, i.e. there are infinitely many source configurations that generate exactly the same EEG and MEG [79], [42]. Only
1. Introduction

by imposing constraints on the solution can the problem be solved. Some of these constraints are: the type of sources (e.g., dipoles, multipoles), and the number of sources.

In [49, Chap. 3] and [114] it is shown that assuming the incorrect number of sources leads to inaccurate parameter estimates. Consequently, an accurate estimate of the number of sources is required in EMSA. This thesis is mainly concerned with determining the sources generating EEG and MEG.

1.1 Electromagnetic source analysis

The objective of electromagnetic source analysis (EMSA) is to obtain an estimate of the sources generating the EEG and MEG data. This analysis can be split-up into two parts: the forward and inverse problem. The forward problem refers to the computation of EEG and MEG given the head and source parameters [90]. The inverse problem refers to the estimation of the source (and head) parameters given the EEG and MEG. First some background on the generators of EEG and MEG and their models is discussed. Then the inverse problem is described.

1.1.1 Generators

EEG and MEG can be measured time-locked to a stimulus. This EEG and MEG then contain both a response (signal) related to the stimulus (event) and ongoing, unrelated activity (noise) [127]. The time-locked EEG and MEG are sometimes referred to as event-related potential or field (ERP or ERF), respectively (e.g. [35], [47]).

This ERP or ERF is relatively weak compared to other ongoing activity [69]. Therefore, an average of several replications (trials) of the experiment is required, which increases the signal to noise ratio (SNR) [35]. Three important assumptions of averaging the data are: (i) the signal (ERP and ERF) is exactly the same at each trial [127], (ii) the ongoing EEG and MEG are completely unrelated to the event [91], and (iii) the ongoing EEG and MEG on each trial is independent of that of any other trial [69]. In this thesis the averaged EEG and MEG (ERP and ERF) are of primary interest.

It is generally believed that the largest contribution to EEG and MEG is due to postsynaptic potentials (PSPs) in the cortex [118]. To generate EEG and MEG, activity of the neurons should be synchronized. In this case synchronization requires that activity is simultaneous and that the neurons are oriented in parallel [42]. Since PSPs are slow (up till several hundred milliseconds) compared to the action potential (in the order of tens of milliseconds) [134], PSPs are more likely candidates of the generators of EEG and MEG. Moreover, PSPs are from the cell bodies and dendrites, whereas action potentials propagate through axons [34]. The orientations of cell bodies and dendrites allow for parallel orientation, whereas the orientation of axons is much more random [49]. The reason that the cortex is mainly responsible for EEG and MEG, then, is threefold: (1) the pyramidal cells, which are oriented
1. Introduction

in parallel, are mainly located in the cortex, (2) a large number of interconnections in the cortex facilitates synchronous activation, and (3) the cortex is relatively close to the EEG and MEG sensors [86].

The activity of neuronal clusters from the cortex can be modeled (source model) inside a head model. Assumptions on the head model include the conductivities of the compartments of the head and its geometry [90]. Assumptions on the source model include the type of source and the number of sources. The electromagnetic activity of a cluster of neurons is most often represented by a current dipole [42]. A dipole is a short element of current, characterized by its location, orientation, and strength. Consequently, a dipole can be represented by a vector pointed in the direction of the current [134]. The strength or amplitude of a dipole is defined as the product of the current and its length, and, therefore, has the units of Ampère-meter. There are several reasons why the dipole model is popular: (1) the current pattern from individual cortical neurons (pyramidal cells) resembles that of a dipole [118], (2) the pattern from a single dipole resembles that of a complex pattern of multiple dipoles observed from a distance [81], and (3) all source configurations (e.g., higher-order multipoles) can be described as a combination of dipoles [90]. The last reason refers to a linearity property, which is known as the superposition principle: the forward model of several dipoles equals the addition of forward models from several dipoles [42].

1.1.2 Inverse problem

The inverse problem is to compute the source (and head) parameters given the EEG and MEG data. A difficulty of the inverse problem is that there is in general no unique solution [42], [79], [98]. There are electrically and magnetically ‘silent’ sources, i.e. sources that produce no electric or magnetic field outside the head model. An example of an electrically silent source is a current loop [42]. An example of a magnetically silent source is a radial dipole in a spherically symmetric head model [98]. Consequently, a silent source can always be added to the source model which will result in the same EEG or MEG.

Contrary to silent sources, the dipole parameters of non-silent sources can be estimated. There are several approaches to estimating the dipole parameters. Each approach has its own set of assumptions on the head and source model. An overview of these approaches can be found in e.g. [42], [49], [8]. Two approaches are used in the present thesis: (1) the equivalent current dipole (ECD), where only a few sources are assumed to be active, and (2) the distributed source model (DSM) approach, where a great many sources are assumed to be active.

In the first approach it is assumed that only a few (1-4, say) sources are required to describe the EEG and MEG. It is assumed that each dipole represents a cluster of neurons. Either a single sample is analyzed (instantaneous analysis) or several samples are analyzed simultaneously (spatio-temporal analysis). In the latter analysis a choice has to be made whether the locations and/or orientations are allowed to vary across samples [75], [99], or not [55], [77]. The nonlinear and linear parameters
1. Introduction

of the dipole(s) can be estimated by, for example, least squares (LS) or maximum likelihood (ML). By iteratively searching the parameter space the minimum of the LS or ML function can be found. The LS or ML estimates are obtained at this minimum. The advantage of using LS or ML is that the statistical properties are optimal under certain conditions [3], although this is unclear if the model is incorrect. In Chap. 3 this latter issue is addressed. There are several disadvantages: it is in general difficult in nonlinear optimization to establish that a minimum is indeed the global minimum [75], and an estimate of the number of sources is required [114], [49, Chap. 3], [77], [8]. In Chap. 4 and 5 methods are proposed to obtain such an estimate.

In the second approach, the distributed source model, the brain model is divided into many small volume elements (voxels), and a source (usually a dipole) is assumed in each voxel [42]. Hence, the locations are assumed known and only the linear orientation and amplitude parameters are estimated. This approach has the advantage that the estimate of the linear (orientation and/or amplitude) parameters is computationally very efficient [75]. Additionally, a cluster of activity being represented by many small dipoles seems more physiologically plausible than a cluster being represented by a single dipole. There is also a major disadvantage: there are more parameters (orientations/amplitudes) than there are observations (sensors). This leads to biased estimates, which can only be resolved by using linear combinations of the solution, as will be shown in Chap. 6.

In [49] it is outlined how to regard EMSA as a regression analysis with its associated assumptions. Let \( y_j \) be the vector with the EEG or MEG measurements of \( m \) sensors on trial \( j \). Let \( \bar{y} = \frac{1}{n} \sum_{j=1}^{n} y_j \) be the average, and let \( f(\theta) \) be the \( m \) vector with the biophysical function dependent on the dipole parameters collected in \( \theta \). The biophysical function also depends on the sensor positions but this is suppressed in the notation for convenience. Then symbolically the model for EEG and MEG is

\[
\hat{y} = f(\theta) + e,
\]

where \( e \) is the \( m \) vector of additive, random noise (also referred to as pure error). This model describes EEG or MEG at a single sample but can easily be extended to multiple samples (see Chap. 5). The averaged EEG or MEG is the dependent variable and the sensor locations are the independent variables. The importance of regarding the estimation of the dipole parameters as a regression analysis is that the relevance of associated concepts, such as bias, standard errors, consistency, and asymptotic normality, are immediate. This approach will be used extensively in this thesis. The assumptions of the regression analysis are: (i) the biophysical model is deterministic, (ii) the biophysical model is independent of the noise, (iii) the noise of each trial is independent of the noise from any other trial, (iv) the noise is uncorrelated and has equal variance across sensors (white noise) [49], and (v) the biophysical model is correct. Violation of the white noise assumption is investigated extensively in [49] and weighting methods are proposed to remedy this violation. In Chap. 2 an extension of weighting is given and investigated. In Chap. 4 and 5 the influence of weighting on model selection is investigated.
1. Introduction

Assumption (v) requires that, for instance, the number of sources is correct in the ECD approach, or that the sources are indeed active in the presumed area in the DSM approach. This can be attained by using techniques to identify a model. Two different techniques to identify a model are used for the two approaches, as discussed next.

1.2 Model selection in ECD

Model selection refers to the decision problem of selecting among a set of possible models the best approximating model of the data at hand [16], [142]. It is first explained what is meant by a model and how models can be related. Subsequently, two common approaches to model selection are briefly introduced.

1.2.1 Model

The parametric regression model for EEG and MEG was defined in the previous section in terms of means of the data. Often, though, a parametric model is defined in terms of distributions of the data: a parametric model is a set of distributions which depend on parameters and the distributions arise from the variations in the set of parameters [12], [38]. The distributions of the EEG and MEG data are generally multivariate since multiple sensors (and time samples) are measured. A single multivariate distribution of the EEG or MEG data arises from the parameterized mean and additive noise, under the assumptions of the previous section. Under the same assumptions, other distributions arise from using different parameter values for the parameterized mean. Consider, for instance, a single MEG sensor. Assume that the conditions of the previous section are satisfied and that the noise is normally distributed. Then two normal distributions with different locations on the range of MEG values arise from using either parameters for one or two sources. The selection of a one- or two-source model then uses the qualitative difference between the distributions of the data.

A model is said to be nested in another model if the set of distributions of one model is contained in the set of distributions of the other model [12]. In EMSA a one-source model is nested in a two-source model. If, on the other hand, the head models differ in two competing models, then the models are not necessarily nested. So, in general, models are not necessarily nested if the parameter space of one model contains the other. Tests on non-nested models are possible if there exists a function of the parameters of one model mapping them to the parameters of the competing model (e.g., [12], [38], [71]). For instance, given two different head models, a two-source model can be related to a one-source model by assuming that all parameters of the second source are zero.

Contrary to the classical accounts, in modern accounts of model selection it is recognized that the true (correct) model is unattainable (e.g., [17], [107], [16]). Instead of assuming that the true model is among the set of possible models, it is assumed that only a biased best approximate model is attainable (see Chap. 3).
The repercussions of the modern account can be found in asymptotic properties (consistency, see Shao [107]) and in the interpretation of the models as best but possibly biased approximations of the truth. In this thesis both the classical and the modern account of model selection are used.

1.2 Selection procedure

The selection of an approximate model is the decision to select a model based on some criterion function in combination with a decision rule. The criterion function is a function of the data; for instance, a hypothesis test. The decision rule refers to the rule which determines on the basis of the outcome of the criterion function which model is finally selected. In hypothesis testing, for instance, the decision rule is the nominal alpha, which determines when to reject or accept the hypothesis (more examples are given in Chap. 4 and 5). The combination of the criterion function and the decision rule is called a model selection procedure (MSP). An MSP is chosen such that conclusions on the parameters are warranted. Generally, two approaches can be discerned. The first approach is based on so-called information criteria and the second approach on hypothesis testing theory [38].

In the first approach, information criteria, model selection boils down to the balance between bias and variance of parameter estimates [11], [45], [142]. Bias refers to how close on average a parameter estimate is to the true (or best approximate) value, and variance refers to the precision (variability) of the parameter estimate. Decreasing the bias leads to more accurate descriptions of the data set at hand, whereas decreasing the variance leads to better generalization to other data [11], [31], [45]. Regarding a model as a set of distributions of the data (see Sec. 1.2.1), an MSP in this class can be translated to determining the ‘distance’ between two distributions by some measure [38]. One such measure is the Kullback-Leibler information criterion (KL) [17] (if the data are normally distributed the KL is the same as the mean squared (prediction) error (MSE) [25], [38]). With the KL the distance between the distribution of the best approximation and the distribution of the estimate can be minimized. The KL can be written in terms of the bias and variance of the parameter estimates (e.g., [14], [17], [45]). Consequently, minimizing the KL can be interpreted as minimizing the bias and variance of the parameter estimates simultaneously. Different MSPs resolve the issue of emphasizing bias or variance differently. For instance, the emphasis in the Akaike information criterion (AIC, e.g., [2]) is on minimal bias whereas the emphasis in the Bayesian information criterion (BIC, e.g., [102]) is on minimal variance.

The second approach is hypothesis testing theory. In the case of two hypotheses (the null and alternative hypothesis), the test is characterized by splitting up the sample space in two regions: one for which it is decided to accept the null hypothesis, and the other for which it is decided to reject the null hypothesis (critical region). A null hypothesis is, for instance, that two models (or some of their parameters) are the same, and the alternative hypothesis is that they are not the same. The decision rule of when to accept the null hypothesis is based on the balance between
the probability of two incorrect decisions. The first is the probability of rejecting
the null hypothesis when in fact it is true (Type I error, significance level), and the
second is the probability of accepting the null hypothesis when in fact it is false
(Type II error) [26]. The objective of constructing a test is then to minimize these
two incorrect decisions. The Neyman-Pearson lemma of obtaining a test which is
most powerful (i.e. the highest probability of correctly rejecting the null hypothesis,
which equals 1-Type II), minimizes the Type II error given a certain Type I error.
Tests constructed according to this principle are considered optimal. In this thesis
the tests, for example Hotelling's $T^2$ or the Wald test (see Chap. 4) are based on
this lemma.

A marked difference between the two approaches, information theory and hy-
pothesis testing, is how the decision is reached when to accept or reject a model.
In hypothesis testing the emphasis is on how to minimize the Type I and II er-
rors, whereas for the information criteria it is about how to minimize the bias and
variance of the parameter estimates. Each approach minimizes a different quantity
(although there are instances in which the two can be related, e.g., [1], [31]). The
question is, of course, which approach is best in the context of EMSA, and which
within each class is best. This question is explored in Chap. 4 and 5.

1.3 Model determination in DSM

Since in the DSM approach more parameters are estimated than there are sensors,
the model selection techniques described in the previous section cannot be applied in
this case. However, hypotheses on, for instance, the presence of activity in a certain
brain area can be tested. Conventionally, the estimated amplitude in each voxel
is tested whether it is significantly larger than zero (with or without Bonferroni-
corrected significance levels) [22]. This technique does not yield an estimate of
the best (approximating) model of the data, but instead, merely determines if the
hypothesized area(s) of activity are indeed significant. Unfortunately, since the
estimates of the amplitudes are biased, the statistical interpretation of these tests
is unclear. Additionally, the restrictions (which possibly lead to the bias in the
estimate) make it difficult to determine the degrees of freedom required for testing
[115]. In Chap. 6 an alternative to the conventional approach is discussed that
yields hypothesis tests that are statistically easier to interpret.

1.4 Abstracts of the chapters

2. Electromagnetic source analysis with a parametric covariance model

Estimated generalized least squares (EGLS) electromagnetic source analysis is used
to downweight noisy and correlated data. Standard EGLS requires many trials to
accurately estimate the noise covariances and thus the source parameters. Alternati-
vively, the noise covariances can be modeled parametrically. Only the parameters of
the model describing the noise covariances need to be estimated, and therefore, less
1. Introduction

trials are required. This method is referred to as Parametric EGLS (PEGLS). In this paper PEGLS is developed and its performance is tested in a simulation study and in a pseudo-empirical study.

3. Consequences of model misspecification in EEG and MEG source analysis

If the correct head model and number of sources are known in the analysis of EEG and MEG data, then a nonlinear least squares (NLS) estimate has known asymptotic properties: the NLS estimate is consistent and asymptotically normal. These results are reasonably well approximated under certain conditions in finite samples: the NLS estimate is (nearly) unbiased and is approximately normally distributed. It is, however, not clear if these asymptotic properties apply if the model is misspecified (incorrect). In this chapter it is shown that consistency can be redefined to a model that approximates the true model best, and that an NLS estimate is consistent in this sense and asymptotically normal. Additionally, sufficient conditions are given that lead to overestimation of the size of confidence intervals if the usual parameter covariance matrix is used instead of the 'sandwich' covariance matrix.

4. Model selection in instantaneous EMSA

In electromagnetic source analysis it is necessary to determine how many sources are required to describe the EEG or MEG adequately. Model selection procedures (MSPs, or goodness of fit procedures) give an estimate of the required number of sources. Existing and new MSPs are evaluated in different source and noise settings: two sources which are close or distant, and noise which is uncorrelated or correlated. The commonly used MSP residual variance is seen to be ineffective, that is it often selects too many sources. Alternatives like the adjusted Hotelling’s test, Bayesian information criterion, and the Wald tests on source amplitudes or locations are seen to be effective. The adjusted Hotelling’s test is recommended if a conservative approach is taken, and MSPs such as Bayes information criterion or the Wald test on source amplitudes are recommended if a more liberal approach is desirable. The MSPs are applied to empirical data (visual evoked fields).

5. Model selection in spatio-temporal EMSA

The model selection procedures of Chap. 4 are extended to a spatio-temporal analysis. It is seen that the residual variance tends to overestimate the number of sources. Moreover, the Akaike information criterion and the Wald test on amplitudes and locations have the highest probabilities of selecting the correct number of sources. The Wald test on amplitudes has the advantage that it can be tested which source is active at which time sample.

6. Hypothesis testing in the distributed source model

In the linear approach to EEG and MEG source analysis, the head model is divided into many voxels and a dipolar source is assumed in each of these voxels (distributed source model). Then only the linear moment parameters of the dipoles have to be estimated. Since there are more parameters than sensors in this approach, the
estimates are biased. Moreover, it is difficult to use standard hypothesis tests, since there are either no degrees of freedom, or they are difficult to determine. However, an unbiased linear combination of the estimated parameters does exist. It can be tested whether this linear combination differs from zero or whether there is a difference between conditions. The merits of this approach are discussed and the method is applied to simulated data.