NeuroImage 277 (2023) 120219

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

An *in–vivo* validation of ESI methods with focal sources

Annalisa Pascarella[a], Ezequiel Mikulan[b], Federica Sciacchitano[d], Simone Sarasso[b], Annalisa Rubino[e], Ivana Sartori[e], Francesco Cardinale[e], Flavia Zauli[b], Pietro Avanzini[f], Lino Nobili[g,h], Andrea Pigorini[c,1], Alberto Sorrentino[d,1,*]

[a] *CNR - IAC, Rome, Italy*
[b] *Department of Biomedical and Clinical Sciences "L. Sacco",Università degli Studi di Milano, Milan, Italy*
[c] *Department of Biomedical, Surgical and Dental Sciences, Università degli Studi di Milano, Milan, Italy*
[d] *Department of Mathematics, Università degli Studi di Genova, Genoa, Italy*
[e] *Department of Neurosciences, Center for Epilepsy Surgery "C. Munari", Hospital Niguarda, Milan, Italy*
[f] *CNR - Istituto di Neuroscienze, Parma, Italy*
[g] *Child Neuropsychiatry Unit, IRCCS "G. Gaslini" Institute, Genoa, Italy*
[h] *DINOGMI, Università degli Studi di Genova, Genoa, Italy*

ARTICLE INFO

Keywords:
ESI
EEG
Inverse methods

ABSTRACT

Electrophysiological source imaging (ESI) aims at reconstructing the precise origin of brain activity from measurements of the electric field on the scalp. Across laboratories/research centers/hospitals, ESI is performed with different methods, partly due to the ill-posedness of the underlying mathematical problem. However, it is difficult to find systematic comparisons involving a wide variety of methods. Further, existing comparisons rarely take into account the variability of the results with respect to the input parameters. Finally, comparisons are typically performed using either synthetic data, or in-vivo data where the ground-truth is only roughly known. We use an *in-vivo* high-density EEG dataset recorded during intracranial single pulse electrical stimulation, in which the true sources are substantially dipolar and their locations are precisely known. We compare ten different ESI methods, using their implementation in the MNE-Python package: MNE, dSPM, LORETA, sLORETA, eLORETA, LCMV beamformers, irMxNE, Gamma Map, SESAME and dipole fitting. We perform comparisons under multiple choices of input parameters, to assess the accuracy of the best reconstruction, as well as the impact of such parameters on the localization performance. Best reconstructions often fall within 1 cm from the true source, with most accurate methods hitting an average localization error of 1.2 cm and outperforming least accurate ones erring by 2.5 cm. As expected, dipolar and sparsity-promoting methods tend to outperform distributed methods. For several distributed methods, the best regularization parameter turned out to be the one in principle associated with low SNR, despite the high SNR of the available dataset. Depth weighting played no role for two out of the six methods implementing it. Sensitivity to input parameters varied widely between methods. While one would expect high variability being associated with low localization error at the best solution, this is not always the case, with some methods producing highly variable results and high localization error, and other methods producing stable results with low localization error. In particular, recent dipolar and sparsity-promoting methods provide significantly better results than older distributed methods. As we repeated the tests with "conventional" (32 channels) and dense (64, 128, 256 channels) EEG recordings, we observed little impact of the number of channels on localization accuracy; however, for distributed methods denser montages provide smaller spatial dispersion. Overall findings confirm that EEG is a reliable technique for localization of point sources and therefore reinforce the importance that ESI may have in the clinical context, especially when applied to identify the surgical target in potential candidates for epilepsy surgery.

* Corresponding author at Department of Mathematics, Università degli Studi di Genova, Via Dodecaneso 35, 16146 Genova, Italy.
E-mail address: sorrentino@dima.unige.it (A. Sorrentino).
[1] These authors contributed equally to the manuscript.

https://doi.org/10.1016/j.neuroimage.2023.120219.
Received 15 December 2022; Received in revised form 5 May 2023; Accepted 2 June 2023
Available online 10 June 2023.

1. Introduction

Electrophysiological source imaging (ESI) is a procedure that allows reconstructing the neural activity sources from recordings of the electric potential, usually obtained at the scalp. ESI is a key element in multiple frameworks related to the analysis of EEG data, including the identification of brain regions involved in specific tasks (Bidelman, 2018; Montani et al., 2019) and the estimation of connectivity in task-related or spontaneous activity (He et al., 2019). Moreover, recent evidence (Baroumand et al., 2022; Brodbeck et al., 2010, 2011; Kaiboriboon et al., 2012) suggests that ESI could be considered a valuable tool in the context of pre-surgical evaluation of epileptic patients where its accuracy and reliability are of paramount importance.

Despite being useful, the application of ESI is still not straightforward for the non-expert especially due to the need for subjective choices: among these, the choice of the ESI method among the many available options as well as, the choice of the method's parameters, which is often overlooked despite its great importance. Overall, these choices have a non-negligible impact on the resulting source estimate and therefore need to be carefully performed.

Indeed, it is well known that the inverse problem is ill-posed (Dassios and Hadjiloizi, 2009), mostly due to the non-uniqueness of the solution; *a priori* information is needed to overcome such ill-posedness, however, different methods will incorporate different a priori information/constraints on the solution, resulting in different source estimates: for example, MNE (Hämäläinen, and Ilmoniemi, 1994) is designed to keep the overall 2-norm of the solution low, while linearly constrained minimum variance (LCMV) beamformers (Van Veen et al., 1997) are designed to filter the signal from a specific location while suppressing interference from other locations. As a consequence, different methods will produce different estimates.

In addition, most ESI methods require the user to select one or more parameters. For instance, many of them require to choose a *regularization* parameter, typically but not necessarily related to the Signal-to-Noise Ratio (SNR) of the data. This choice typically impacts the spatial spread of the estimated neural current, its peak location, as well as the smearing of the time courses. Some methods also envisage a *depth weighing* parameter that should fight the well-known bias towards superficial sources that affects, among others, MNE estimates (Lin et al., 2006). Again, by tuning this parameter one can typically tune the depth of the estimated sources, however, what the correct value would be is not easy to determine.

Overall, selection of the most appropriate ESI method and of the optimal parameter(s) is really difficult for non–trained users. Indeed, even if a few very recent methods can automatically set the regularization parameter (Bertrand et al., 2019; Cai et al., 2021; Sun et al., 2022), most used methods still rely on subjective user choices.

In this work we evaluate and compare the performances of ten different ESI methods under multiple values of the input parameter(s), in the case of high-SNR focal sources whose locations are known exactly. Our aim is to characterize the accuracy of different methods, as well as their robustness with respect to the choice of the input parameters. The major strengths of this study consist in: (1) comparing a rather large number of ESI methods applied over the same *in-vivo* open dataset [15]; (2) using a unique ground-truth consisting of known single dipolar millimetric sources derived from the artifact created by the electrical stimulation of two adjacent intracranial electrodes (i.e., de facto, an *in-vivo* phantom head); (3) Applying ESI to high-density EEG scalp recordings (256 channels), thus allowing the exploration of multiple spatial sub-samplings; (4) investigating the stability of the results with respect to the input parameters (regularization, etc.). Altogether, we took advantage of state-of-the-art dataset and procedures to provide the most comprehensive evaluation and comparison among ESI methods to date.

Evaluation and comparison of different ESI methods is often not an easy task, as the true sources of experimental recordings are never known exactly. There are two typical workarounds to this problem. One is to assess the reconstruction error of ESI using synthetic data generated either via a synthetic forward model (Becker et al., 2015, 2016; Chowdhury et al., 2016; Grova et al., 2006; Samuelsson et al., 2021; Yao and Dewald, 2005) or using a hardware phantom (Baillet et al., 2001; Leahy et al., 1998); this approach can result in reasonable comparisons between different methods but can hardly be used to give realistic estimates of localization accuracy in experimental scenarios, as the data generation process, particularly the forward model, is necessarily simplified. The second possibility, applied in a growing number of studies, is to evaluate the accuracy of localization by using either intracranially-recorded interictal/ictal brain activity or post-surgical outcomes derived from epileptic patients (Brodbeck et al., 2011; Koessler et al., 2010; Luria et al., 2020; Mégevand et al., 2014, 2014; Pellegrino et al., 2020); this approach overcomes the limitations of synthetic data but has its own drawbacks, including the fact that resolution is limited to the size of the resected area and the limited spatial sampling of intracranial data, even though better resolution can be achieved combining both electrophysiological and surgical options (Sohrabpour et al., 2020).

In this study we overcome the limitations of both approaches by exploiting a recently published EEG dataset of high-density (256 channels) scalp recordings combined with a ground truth single dipolar source systematically provided through a brief current injection between two adjacent intracranial electrodes whose position is known with millimetric precision (Mikulan et al., 2020). This procedure is capable of generating real data of scalp recorded electrical signals originating from precisely known locations inside the human brain, thus representing an ideal benchmarking scenario for validating and comparing different ESI methods. Similar approaches have been previously reported e.g. in (Cohen et al., 1990; Unnwongse et al., 2023), however, considering only a much smaller number of electrodes and not comparing different ESI methods. Specifically, the dataset used here consisted in scalp EEG recordings collected during Single Pulse Electrical Stimulation (SPES), employed for brain mapping and for the identification of abnormal cortical excitability in epileptic patients implanted with Stereo-EEG (SEEG) leads (Cardinale et al., 2019; David et al., 2013; Matsumoto et al., 2004; Valentin et al., 2002). Data were originally collected at Niguarda Hospital, Milan, Italy. During SPES, a brief current pulse is injected between two adjacent leads, producing an electrical current whose location can be accurately determined. Since this electrical current is strong enough to produce a visible voltage signal on scalp hd-EEG, the procedure generates experimental data of scalp potentials originating from known locations inside the brain. The resulting dataset is characterized by a very high signal-to-noise ratio and is ideally suited to evaluate in vivo the performance of ESI in the case of focal activity.

Using the MNE-Python package, we evaluate and compare ten different ESI methods, thus possibly providing the most extensive comparison thus far: we test dipole fitting, wMNE (Hämäläinen and Ilmoniemi, 1994), sLORETA (Pascual-Marqui et al., 2002), eLORETA (Pascual-Marqui et al., 2006), dSPM (Dale et al., 2000), LCMV beamformer (Van Veen et al., 1997), RAP-MUSIC (Mosher and Leahy, 1999), Gamma Map (Wipf and Nagarajan, 2009), irMxNE (Strohmeier et al., 2014) and SESAME (Viani et al., 2020). For each method under consideration, we test several values for each input parameter, so as to verify (i) the optimal reconstruction attainable by an expert user who is capable of setting the parameter values correctly and (ii) to what extent the method is tolerant with respect to misspecifications of these values. We consider the above questions under different subsampling of the High-Density configuration, starting from the full 256 montage, down to 128, 64 and 32 channels.

In summary, based on a unique open dataset in which the sources of the EEG activity within the brain are known - i.e. a ground truth for the inverse solution methods - we compared the performances of the most commonly used ESI methods and, for each method we optimized the input parameters.

2. Methods

2.1. Description of the data

The dataset used in this study is publicly available and has been described in (Mikulan et al., 2020); in the following subsections we briefly summarize the main relevant features.

2.1.1. Electrical stimulation

Subjects had implanted intracranial shafts for the pre-surgical evaluation of drug–resistant focal epilepsy; electrode positions were therefore established based on clinical needs, and ranged for each subject from superficial to deep locations: the distance from the closest scalp EEG sensor ranged from 28 to 64 mm. Electrical currents were delivered through platinum-iridium semi flexible multi-contact intracerebral electrodes (diameter: 0.8 mm; contact length: 2 mm, inter-contact distance: 1.5 mm; Dixi Medical, Besancon, France). Currents lasted 0.5 ms, had intensities ranging between 0.1 mA and 5 mA and were repeated either every 2 s (for 1 mA and 5 mA) or every 1 s (otherwise). The number of recorded trials was either 40 (for 1 mA and 5 mA) or 60 (otherwise).

Electrode positions were measured by co-registering the post-implant CT (O-arm 1000 system, Medtronic) to the pre-implant MRI by means of the FLIRT software (Jenkinson and Smith, 2001). The location of every single lead was assessed using Freesurfer (Dale et al., 1999), 3D Slicer (Fedorov et al., 2012) and SEEG assistant (Narizzano et al., 2017). When the EEG digitization MRI was different from the pre-implant MRI, transformation of the SEEG space to the EEG space was performed using an affine transformation between MRIs calculated with the ANTs software (Avants et al., 2011). Normalized contacts' coordinates were estimated through a non-linear registration between the subject's skull-stripped MRI and the skull-stripped MNI152 template (Fonov et al., 2009), using ANTs' SyN algorithm. The accuracy of the normalization procedure was verified by visual inspection.

2.1.2. High density EEG recordings

256 EEG channels (Geodesic Sensor Net; HydroCel CleanLeads) were recorded with an EGI NA-400 amplifier (Electrical Geodesics, Inc; Oregon, USA) at a sampling frequency of 8,000 Hz, using a custom-made acquisition software, based on EGI's AmpServerPro SDK and written in C++ and Matlab. No software filters were used during acquisition. The location of EEG electrodes and anatomical fiducials were digitized with a SofTaxicOptic system (EMS s.r.l., Bologna, Italy), coregistered with a pre-implant MRI (Achieva 1.5 T, Philips Healthcare).

2.1.3. Generation of evoked responses

Raw data were high-pass filtered at 0.1 Hz (FIR filter; zero phase; Hamming window; automatic selection of length and bandwidth); for two subjects, data were also notch filtered at 50, 100, 150 and 200 Hz due to the presence of line noise. After rejection of bad channels through visual inspection, epochs were generated from -300 ms to 50 ms with respect to the electrical stimulation.

Evoked potentials were generated by averaging across all epochs produced by stimulation of the same contact pair. This resulted in very clear dipolar patterns produced by a single source, with a high Signal-to-Noise Ratio (SNR). In Fig. 1 we report an example of a butterfly plot of EEG data in the time window [-0.5; 1]ms. Overall, the dataset analyzed in this study comprises 7 subjects for a total of 61 single-source potentials.

2.2. Forward model

The forward model is a BEM model with realistic geometry. The model comprises three compartments and was set up using the function `make_bem_model` with *ico* set to 4, corresponding to a downsampling of the Freesurfer triangulations to 5120 triangles; conductivities were automatically set to 0.3, 0.006 and 0.3 S/m, for the brain, skull and scalp compartments, respectively. The source space was built using

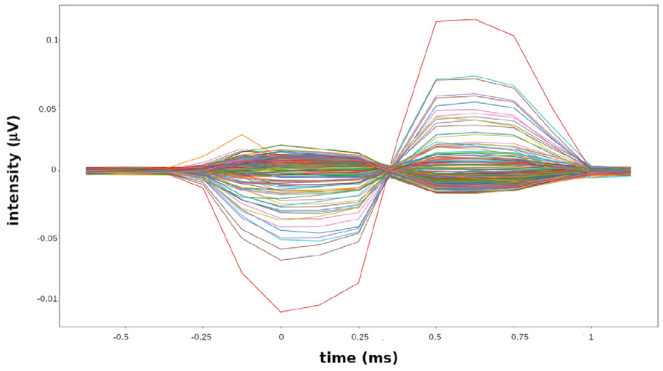


Fig. 1. An example of butterfly plot of averaged EEG data related to the stimulation artifact.

4098 locations in each hemisphere, for a total of 8196 available sources, with an average spacing of 4.9 mm.

2.3. Montages

We test inverse methods using four different sensor montages: the full montage contains 256 channels; then we repeatedly halve the number of channels, using channels corresponding to EGI's 128, 64, and 32 montages. Note that in the case of 256 channels the effective number of channels is smaller than the nominal number, due to the removal of bad channels. On average the number of bad channels that were removed is 46 ± 23, mainly located over the neck and the cheeks of the subjects, but, in some cases, also in the areas where the external part of intracranial electrodes were too dense to fit the hd-EEG net over the subject's head.

2.4. Inverse methods

Source localization was carried out using ten different inverse methods. Nine of them are available as open source code within the MNE-Python package[2] (Gramfort et al., 2014): dipole fitting, dSPM (Dale et al., 2000), eLORETA (Pascual-Marqui et al., 2006), Gamma Map (Wipf and Nagarajan, 2009), Linearly Constrained beamformer (Van Veen et al., 1997), Mixed Norm Estimate (Gramfort et al., 2012), MNE (Hämäläinen and Ilmoniemi, 1994), RAP-MUSIC (Mosher and Leahy, 1999), sLORETA (Pascual-Marqui et al., 2002). In addition to these nine inverse algorithms, we also used SESAME (Sommariva and Sorrentino, 2014; Viani et al., 2020), a Bayesian multi–dipole modeling algorithm currently listed as a plug-in of MNE-Python. Our choice of working with MNE-Python was motivated by the following reasons: it contains the most used ESI methods; it contains the largest set of methods; finally, it is written in Python, a freely available programming language. Valid alternatives would have been represented by Fieldtrip (Oostenveld et al., 2010) and Brainstorm (Tadel et al., 2011). Along the paper, we will refer to each inverse method with its short name as listed in Table 1.

In the analysis below we split the inverse methods in three classes according to the following classification: we call *distributed* methods those methods that are based on a distributed source model, and have no sparsity-promoting penalty terms, i.e. MNE, dSPM, LCMV, SLOR, ELOR; we call *dipolar* methods those based on strictly dipolar models such as DF, RAP and SSM we call *sparsity-promoting* methods those methods based on a distributed source model but with a sparsity-promoting penalty term, i.e irMxNE and GM. This non-standard classification is motivated by the fact that irMxNE and GM provide in output the esti-

[2] https://mne.tools

Table 1
Short name of inverse methods used in the study.

Method	Short name
Dipole Fitting	DF
dSPM	DSPM
eLORETA	ELOR
Gamma Map	GM
Linearly Constrained beamformer	LCMV
Minimum Norm Estimate	MNE
Mixed Norm Estimate	irMxNE
RAP–MUSIC	RAP
SESAME	SSM
sLORETA	SLOR

Table 2
Parameters used for each inverse method.

Method	Parameters	
	Depth	**Other**
DF	-	-
DSPM	0, 1, 2, 3, 4, 5	λ= 0.04, 0.0625, 0.11, 0.25, 1
ELOR	-	λ= 0.04, 0.0625, 0.11, 0.25, 1
GM	0, 1, 2, 3, 4, 5	α = 0.25, 0.5, 0.75, 1, 1.25
LCMV	0, 1, 2, 3, 4, 5	λ = 1, 0.1, 0.5, 0.01, 0.05
MNE	0, 1, 2, 3, 4, 5	λ= 0.04, 0.0625, 0.11, 0.25, 1
irMxNE	0, 1, 2, 3, 4, 5	α = 10, 30, 50, 70, 90
RAP	-	-
SSM	-	fs=0.07, 0.13, 0.2, 0.27, 0.33
SLOR	0, 1, 2, 3, 4, 5	λ= 0.04, 0.0625, 0.11, 0.25, 1

mated number of sources and the source locations, like purely dipolar methods do.

All methods, except DF and SSM, need a noise covariance matrix that was estimated from the pre-stimulus interval between -250 ms and -50 ms using the `compute_covariance` function in *auto* mode, in which cross-validation is used (Engemann and Gramfort, 2015).

2.4.1. Regularization parameters

All ESI methods under analysis require the user to choose the value of one or more input parameters. In the following, we evaluate the performances of the methods when different values of the parameters are used. Since performance evaluation also aims at quantifying the sensitivity of the source estimate with respect to the input parameter, in order to have a meaningful comparison across methods it is important to use comparable values and interval ranges, as much as possible. Common parameters vary in the same interval and with same values for each method. For parameters that are specific to one method, we used the same relative range of variation.

Both DF and RAP require the number of dipoles to be estimated as input parameters. For DF we use the a-priori information that the data was generated by a single source, while for RAP this information comes from the inspection of the singular values of the data covariance matrix (Mosher and Leahy, 1999).

Four ESI methods (MNE, dSPM, sLORETA, eLORETA) require the value of the regularization parameter as input, denoted as λ in the following. Typically, the value of the regularization parameter is chosen based on the SNR of the data: cleaner data correspond to smaller values of the regularization parameter, while noisy data require larger values. In this study we test five different values of λ ranging from 0.04, that corresponds to extremely clean data, up to 1, which has been recently shown to be an upper limit guaranteeing good accuracy of the reconstructions (Samuelsson et al., 2021). irMxNE requires to set α, a value between 0 and 100 for which we choose to use uniformly spaced points avoiding the extrema. SSM requires setting the noise variance, as a fraction of the peak signal: here we used uniformly spaced values between 0.07 (corresponding to low noise) and 0.33 (corresponding to high noise). GM requires to set the noise variance α of the whitened data: while the theoretical noise variance of whitened data is 1, there are some indications that a slightly larger values might perform better; we chose to use uniformly spaced points between 0.25 and 1.25. LCMV requires to set the regularization parameter λ for inversion of the covariance matrix: here we test five values, logarithmically spaced.

Finally, six methods also take in input a depth-weighting parameter, that aims to reduce the bias towards superficial sources: for this parameter we test five linearly spaced values between 0 (no weight) and 5 (Lin et al., 2006).

The different parameters we use and the corresponding values are reported in Table 2.

2.5. Performance evaluation

To quantify the source localization accuracy, we employ the Dipole Localization Error (DLE), which is defined as the distance between the estimated location and the putative dipole location, i.e. the medium point between the two electrodes in which current was fed. The estimated location is defined as follows for distributed methods (DSPM, ELOR, LCMV, MNE, SLOR) and for dipolar and sparsity-promoting methods (DF, GM, irMxNE, RAP, SSM). The distributed methods treat each time point independently; when applied to a time-series, they provide a (potentially) different intensity map/dipole location at each time step. For these methods we consider the solution at the peak latency and use the location corresponding to the peak intensity. For DF we consider the location of the equivalent dipole at the time point maximizing the goodness of fit in all analyzed time windows. The remaining methods work natively with time–series, and provide one intensity map/dipole location(s) for the whole analysis window. For these methods we use the location of the dipole estimated by applying the method to the window [−2; 2] ms; in case more than one dipole is estimated, we use the location of the estimated dipole with larger dipole moment. For all methods (except DF and RAP), we compute the DLE for the different parameter value combinations. In addition, for each method, we consider the mean solution, defined as the average of the estimated dipole location over all parameters combination; for this mean dipole we compute the corresponding DLE with the putative dipole location.

To quantify the spread of the estimate, we complement the DLE with a second metric, which is defined differently for distributed methods and for dipolar and sparsity-promoting methods:

- for distributed methods, we use the Spatial Dispersion (SD), defined as

$$\mathrm{SD} := \sqrt{\frac{\sum_{j=1}^{N_v} \left(d_j \, |S_j|\right)^2}{\sum_{j=1}^{N_v} |S_j|^2}} \tag{1}$$

  where N_v is the number of voxels, d_j is the distance between the j–th voxel and the global peak, and S_j is the value of the cortical map at the j–th voxel.
- for dipolar and sparsity-promoting methods, we use the Estimated Number of Dipoles (END), which can be considered as a "proxy" of the SD.

Throughout the Results section, we will assess the differences between the performance of each pair of methods for each montage with the use of the non-parametric Wilcoxon signed-rank test. The test is a paired difference test like the paired Student's *t*-test, with no assumptions on the distribution of the data. The test works by constructing the signed differences between value pairs, and ranking them according to their absolute value; for a full explanation, please see (Pestman, 2009), page 253. We set the significance threshold at 0.05 and use Bonferroni correction for multiple comparisons.

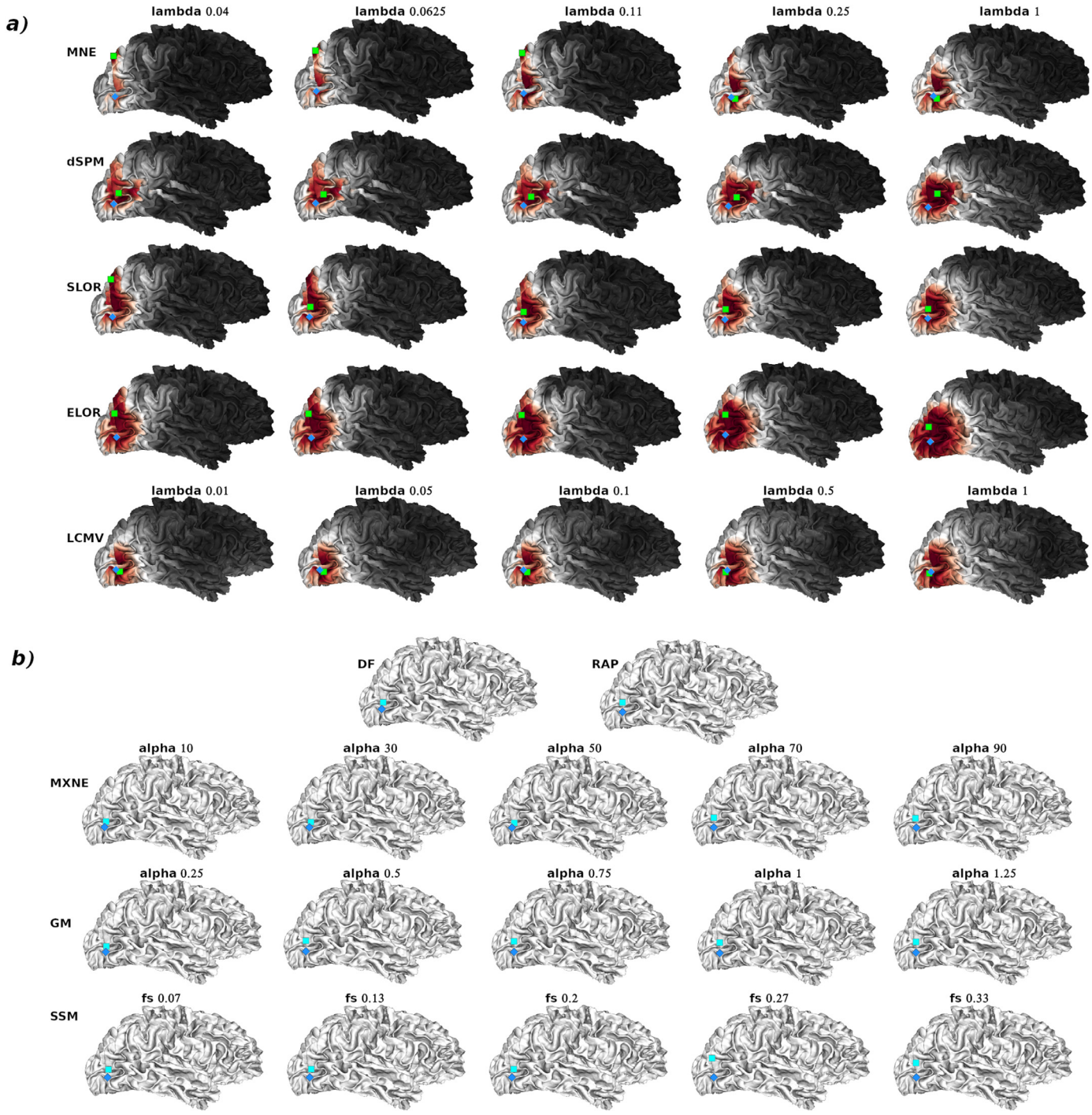


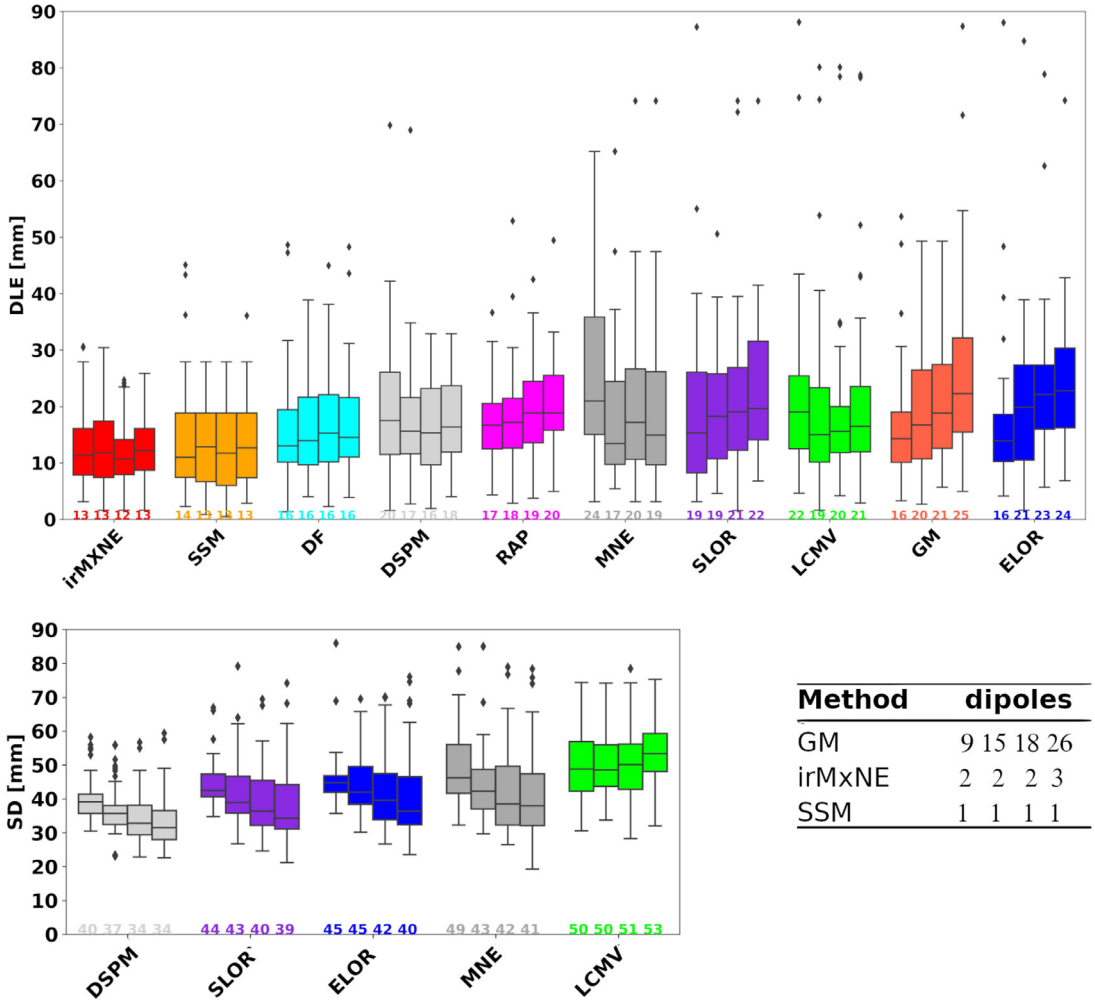
Fig. 2. Example case: exact source location (light blue diamond) and a) intensity map and estimated source location (green square) as obtained by the distributed methods (MNE, dSPM, SLOR, ELOR and LCMV) for five different values of the input parameter and b) estimated source location (cyan square) as obtained by the dipolar (DF, RAP and SSM) and sparsity-promoting methods (irMxNE, GM). For irMxNE, GM and SSM the best solution is shown for five different values of the input parameter. The inverse methods are applied to a session with 256 channels. The plot was done using the Visbrain suite (Combrisson et al., 2019). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Results

We present the results obtained by applying the ten inverse methods to all 61 sessions of the dataset. Fig. 2 shows an example of localization provided by the different inverse methods applied with five different values of a given input parameter, together with the exact location. We remark that, whenever multiple sources are estimated, only the strongest source is considered in the analysis.

3.1. Localization with the best combination of input parameters

We start by considering the best accuracy attainable: for each method, for each session we consider the best solution across parameters (for more information on parameter values see Table 2), i.e. the one with the smallest Dipole Localization Error (DLE). The first row of Fig. 3 contains the boxplots of such optimal DLE (in mm) computed for all methods and montages. Each method is coded by a specific color,



Method	dipoles
GM	9 15 18 26
irMxNE	2 2 2 3
SSM	1 1 1 1

Fig. 3. Top: For each method, boxplot of minimum Dipole Localization Errors (DLE); from left to right, we show the boxplot obtained with 32, 64, 128, and 256 channels, with (below) the mean value. ESI methods are ordered based on the overall behaviour. Bottom: On the left boxplot of Spatial Dispersion for distributed methods (left); from left to right, we show the boxplot obtained with 32, 64, 128, and 256 channels, with (below) the mean value. On the right mean value of the Estimated Number of Dipoles for dipolar and sparsity-promoting methods for the four montages (32, 64, 128, 256).

and for each algorithm, from left to right, we show the boxplots obtained with 32, 64, 128, and 256 channels respectively. At the bottom of each boxplot we report its corresponding mean value (in mm). The second row of Fig. 3 contains, on the left, the spatial dispersion (SD) for distributed methods and, on the right, the average estimated number of dipoles (END) for dipolar and sparsity-promoting methods; both SD and END were computed at the smallest DLE.

The boxplots in Fig. 3 indicate that different methods feature quite different accuracy; in order to verify whether these differences are statistically significant, we performed pairwise Wilcoxon signed–rank tests. In Fig. 4, we report the results of the test assessing the difference between the performances of each pair of methods when the best solution across parameters is considered. The results indicate that the statistical difference between methods varies in the four different montages; in particular the best performances are obtained by: irMxNE and SSM with 32 and 64 channels and DF, irMxNE and SSM with 128 and 256 channels. Overall two methods, SSM and irMxNE, substantially outperform the others.

In Fig. 5 we report how many times (in percentage) each method provides a solution which is within a threshold δ of 5 mm from the best solution obtained across methods; we notice that changing the threshold does not modify the picture substantially. The methods with the highest percentage for 256 and 128 channels are irMxNE and SSM, where the smallest DLE is obtained more than 60%. SSM performs well also when 64 and 32 channels are considered. The smallest DLE is reached for almost half of the sessions by irMxNE (32 and 64 channels), MNE (256 channels) and DF (256 channels).

We finally investigate whether a small variance between the best source estimates obtained by different methods is an indicator of a good localization. In Fig. 6 we show a two-dimensional histogram counting how many times we observed a given average DLE (across methods) and a corresponding variability across methods (obtained by averaging all inter-method distances). We notice that a small variability does not always correspond to a small DLE, although there is a noticeable correlation between the two.

3.2. Impact of input parameters on source localization

To study the influence of input parameters on localization accuracy, for each montage and method, given a session, we compute the solutions corresponding to different combinations of parameter(s) values; we then consider the mean dipole and its DLE, as well as the standard deviation with respect to the mean dipole as:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N} DLE(r_i, \bar{r})^2}$$

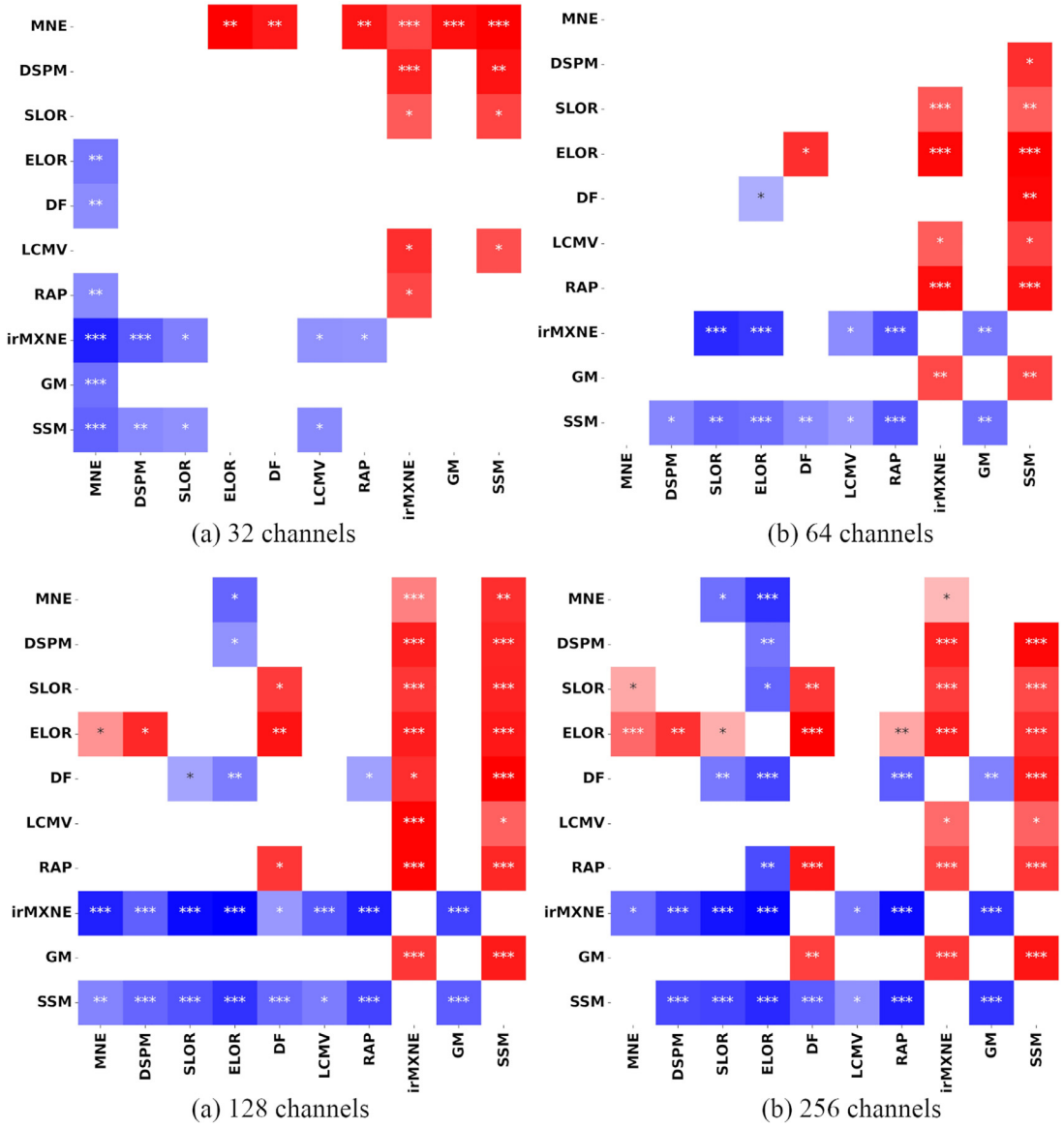


Fig. 4. Significance of pairwise Wilcoxon tests between the DLE of each and each other ESI method, for the four different montages: 32 (first row left), 64 (first row right), 128 (second row left), and 256 channels (second row right). A red square indicates that the method listed in the corresponding row is significantly worse than the one listed in the corresponding column, while a blue square indicates that the method listed in the corresponding row is significantly better than the one listed in the corresponding column. The asterisks are related to the corrected p-value: $*$ $p < 0.05$, $**$ $p < 0.005$, $***$ $p < 0.0005$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where N is the number of parameters combination (see Table 2), r_i is the location of the dipole estimated by the ith set of parameters and $\bar{r}$ is the location of mean dipole computed across all N parameters combination.

Fig. 7 shows the variability of the location of the estimated dipole across all parameters combinations. Among distributed methods, DSPM has a very low standard deviation, while among the dipolar and sparsity-promoting methods, SSM and GM are the ones with less spread around the mean dipole. Conversely, LCMV presents a very high variability of the solutions for the different input parameters at all montages. We remark that for DF and RAP we use the a–priori information of focal source to be detected.

Fig. 8 displays the influence of parameters on the solution of the different inverse methods. For each method we report, for each combination of input parameter values, the percentage of times the estimated location lies within 5 mm from the best solution obtained by the method across all parameter combinations. The depth parameter seems to affect mostly MNE and irMxNE, while having little impact on SLOR and GM. dSPM and LCMV give their best performance with a value of depth between 0 and 2. Overall, the most important parameter is the one related to the noise variance: for distributed methods, $\lambda = 1$ leads to the best performance; for irMxNE a value of alpha between 30 and 70 together with a value for depth equal to 1 gives the best result; SSM shows better performances for smaller values of the parameter, corresponding to higher SNR and GM seems to be not influenced by the setting of input parameters.

Novel extensions of irMxNE and Gamma-Map algorithms are equipped with some adaptive data–driven techniques for optimally tuning the regularization parameter (Cai et al., 2021; Deledalle et al., 2014). The current version of the MNE–python package allows to automatically set the regularization parameter α for irMxNE method exploiting the

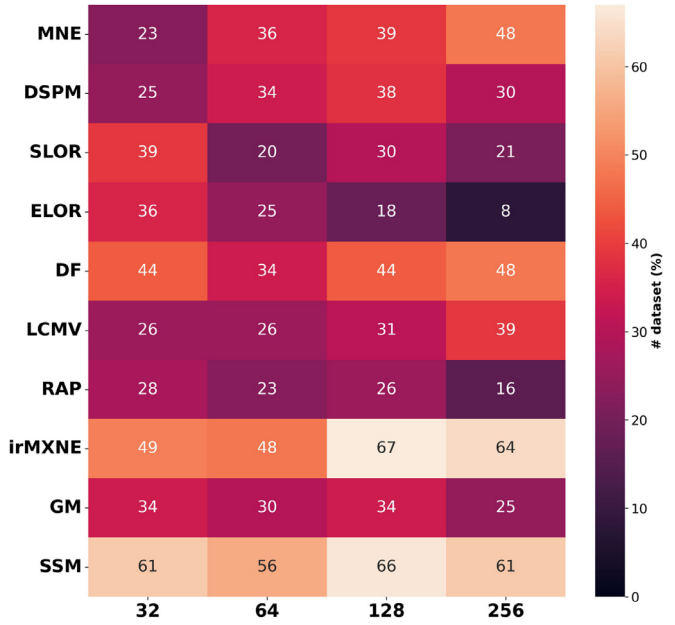


Fig. 5. Number of times (%) in which the minimum DLE for each montage is reached by each method by using a tolerance $\delta = 5$ mm.

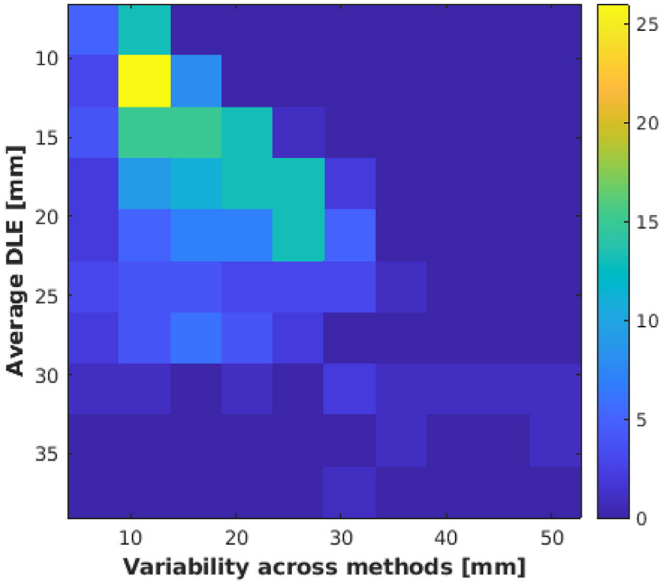


Fig. 6. Variability across methods and average DLE. The figure shows that concordance between methods is correlated with the average DLE, although there are cases of low variability and high DLE.

SURE algorithm (Deledalle et al., 2014). Here we compare the DLE obtained by this automatic parameter setting with those obtained with the optimal choice of input parameters. Fig. 10 shows the minimum DLE obtained in the four montages with or without the automatic setting of the regularization parameter α. In both cases, for each session, we consider the smallest DLE obtained by the best tuning of the depth weighting factor. We also report the results of the pairwise Wilcoxon signed–rank test assessing the difference between the performance of irMxNE method with or without the automatic setting of the regularization parameter α. We also investigate the impact of depth weighting factor when the SURE algorithm is used and find that the best performance is obtained by setting the depth parameter equal to 1 or 2. Similar results are showed in Fig. 8 when no automatic setting of α is performed.

3.3. Impact of covariates on source localization

We now proceed to investigate the impact on the accuracy of source localization of various covariates, namely: the number of EEG channels, the data SNR and the source depth.

Fig. 9 shows, for each montage, the boxplot of the smallest DLE computed across all methods for each session; the violet boxplot represents the smallest DLE computed across all montages and methods with a mean value of 5 mm; for more than half of the sessions we obtain a DLE < 5mm. The global mean value is 8 mm in the case of 32, 64 and 128 channels and 9 mm when 256 channels are considered. There is no statistical difference between the four montages.

We tested pairwise the effect of using different montages on DLE and SD in each ESI method. As far as the DLE is concerned, we obtained sig-

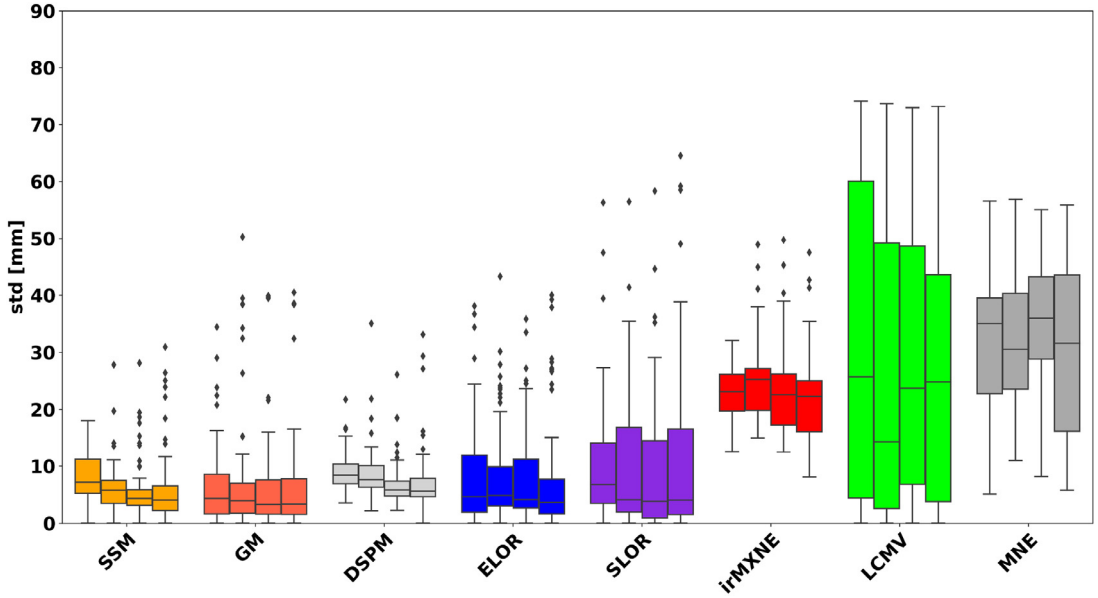


Fig. 7. From left to right for each method and montage we show the standard deviation of DLE computed by using the mean dipole over all parameter combinations..

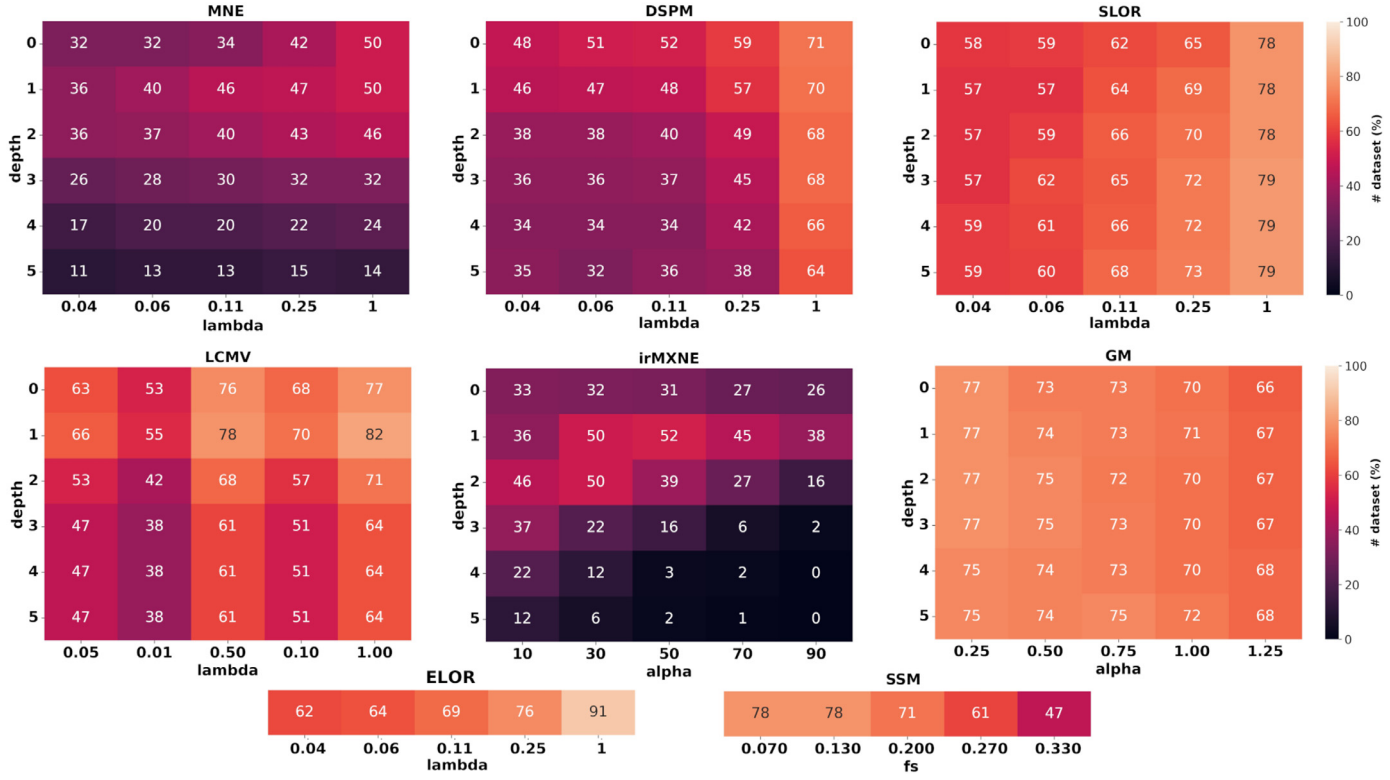


Fig. 8. Percentage of times (%) a specific combination of parameter values reaches within 5 mm from the best solution obtained by the method across all parameter combinations.

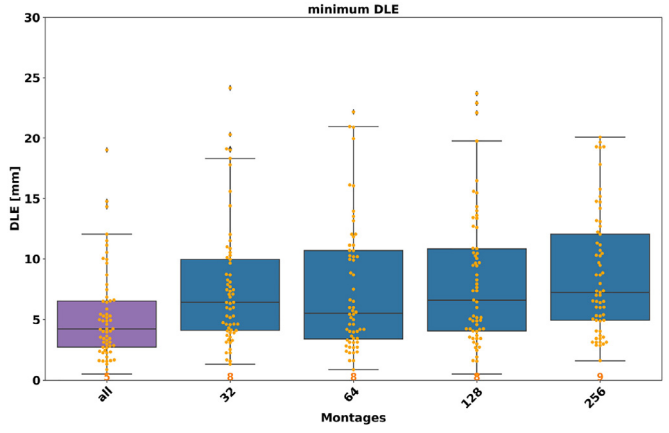


Fig. 9. Minimum DLE over all methods for each montage. The violet boxplot represents the minimum of DLE across all montages.

nificantly different values for many of the distributed methods (dSPM, ELOR, MNE, SLOR) as well as for GM and RAP. We note a positive effect of the montage also on the SD for the same imaging methods. The corrected significant p–values of these statistical tests are reported in Table 3.

In order to evaluate the performance of each method as a function of the depth of the true source and the SNR of the recorded session, we performed a mixed-effects linear regression analysis using DLE as the dependent variable, source depth and SNR as predictors, and subject as random factor (intercept). The mixed-effects approach was chosen due to the nested nature of the data (i.e. stimulation sites within subjects). The source depth was defined as the distance between the stimulation site and the closest hd–EEG electrode in the full montage. The SNR was computed as the ratio of the norm of the signal at peak and at baseline.

Table 3
For each method, significance of pairwise Wilcoxon signed–rank tests between the DLE and SD of each and each other montage. We report the corrected p-values. We remark that for RAP and GM there is no a corresponding SD value.

Method	Channels	p-value	Channels	SD p-value
MNE	128 > 64	2.9×10^{-2}	64 > 256	4×10^{-2}
	32 > 256	9.8×10^{-3}	32 > 256	1.5×10^{-6}
	32 > 128	9.8×10^{-3}	32 > 128	2.3×10^{-5}
	32 > 64	3.1×10^{-4}	32 > 64	4×10^{-5}
DSPM	32 > 64	2.6×10^{-2}	32 > 64	9.9×10^{-6}
	32 > 128	5.9×10^{-3}	32 > 128	1.2×10^{-8}
			32 > 256	2.8×10^{-9}
			64 > 128	6.8×10^{-5}
			64 > 256	8.2×10^{-7}
SLOR	256 > 64	8.3×10^{-3}	64 > 256	1.6×10^{-5}
	128 > 64	9.4×10^{-3}	64 > 128	4.6×10^{-5}
			32 > 256	1.5×10^{-5}
			32 > 128	1.8×10^{-5}
ELOR	256 > 128	4.3×10^{-2}	128 > 256	8×10^{-3}
	256 > 64	7×10^{-3}	64 > 256	2.1×10^{-6}
	256 > 32	8.9×10^{-7}	64 > 128	4.1×10^{-4}
	128 > 64	1.5×10^{-2}	32 > 256	1.5×10^{-5}
	128 > 32	1×10^{-5}	32 > 128	1.8×10^{-2}
	64 > 32	2.8×10^{-3}		
RAP	256 > 64	6.4×10^{-3}		
	256 > 32	4.4×10^{-3}		
	256 > 128	8.2×10^{-3}		
	256 > 64	4.7×10^{-4}		
GM	256 > 32	2.5×10^{-5}		
	128 > 32	1.4×10^{-3}		

We found that the depth of the stimulation site impacts the performances of the following methods: MNE, sLORETA, eLORETA ($p < 1e-07$, Bonferroni corrected), RAP ($p < 0.05$, Bonferroni corrected); for all these

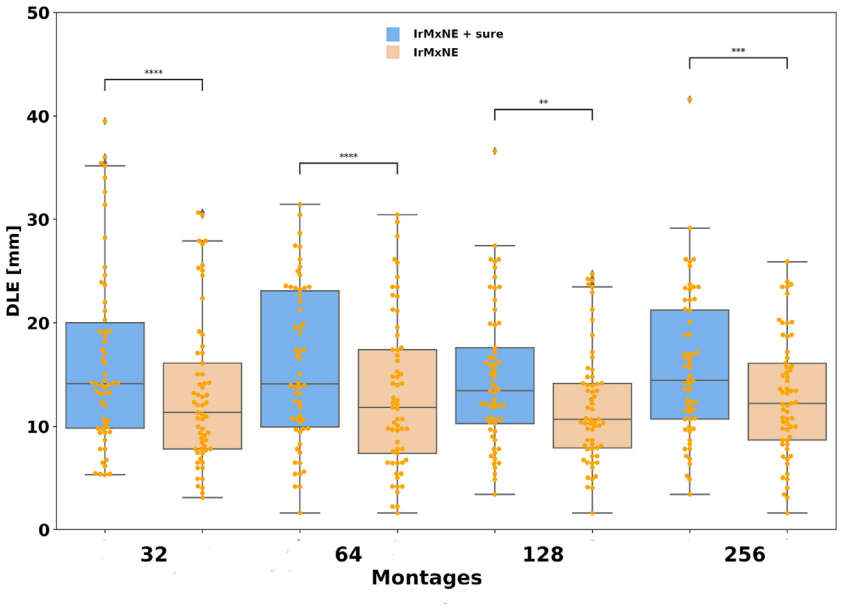


Fig. 10. Boxplot of Dipole Localization Errors (DLE) obtained in the four different montages by irMxNE method by using or not the automatic setting of regularization parameter α. The asterisks are related to the corrected p-value: $*$ $p < 0.05$, $**$ $p < 0.005$, $***$ $p < 0.0005$.

methods, as expected deeper sources tend to produce larger DLEs. On the other hand we found no impact of the SNR on the performance of the methods. In Fig. 11 we provide a pictorial view of the results by showing the regression lines of DLE as a function of source depth and as a function of SNR.

4. Discussion

The aim of this work was to evaluate and compare in vivo the localization accuracy of a relatively large set of ESI methods, as well as their stability with respect to the input parameter(s), under the hypothesis that the neural generators are focal sources and substantially dipolar sources.

4.1. Localization with the best combination of input parameters

When using the best combination of input parameters the results are encouraging: the best solution across methods is within 1 cm from the true source with very high probability, and several methods provide average reconstruction errors around 1 cm, with about 75% of cases falling within 2 cm. It is important to remark that in this study the source reconstruction procedure was completely automated: after application of the ESI method, only the stronger source was retained, sometimes leading to large errors. While in routine analysis it may be difficult to select the optimal combination of parameters, the user may leverage on prior knowledge and sometimes exclude some of the reconstructed sources, thus effectively obtaining a lower localization error than the one estimated in our study.

As expected, dipolar and sparsity-promoting methods provided better results than distributed methods, with irMxNE and SSM featuring the lowest DLE. In this respect, it may be worth recalling a substantial difference between the distributed and the dipolar and sparsity-promoting methods considered here: indeed SSM, RAP, GM and irMxNE compute a single estimate of source activity from a whole time window, thus potentially mitigating the impact of noise (albeit rather low in this dataset); on the other hand MNE, SLOR, ELOR, dSPM and LCMV provide a possibly different solution at each time point, and might be more affected by noise. We also notice that the performances of RAP and LCMV might be impacted by the short duration of the stimulation artifact, which entails that the number of samples used to calculate the data covariance matrix is smaller than the number of samples used in studies involving evoked responses.

We reckon the error we observe is most likely due to the combined effect of bias introduced by the ESI method, and forward modeling error. In particular, in this study we used the publicly available BEM forward model that does not take into account the spatial variability of the electrical conductivity in the brain. In addition, the BEM model ignores the presence of the burr holes; while this should have an almost negligible impact on the results (Lanfer et al., 2012), recent experimental evidence suggests that some impact might be expected (Unnwongse et al., 2023); future work might be devoted to confirm this. Finally, we recall that the single pulse stimulation is a squared wave lasting 1 millisecond, and therefore contains very high frequency components: future studies might be devoted to investigating whether the quasi-static approximation is still valid under these circumstances.

4.2. Impact of input parameters on source localization

In some ESI methods input parameters impact localization accuracy quite substantially, while in others their impact is much more contained. We quantified this variability by computing the standard deviation of the DLE across parameter combinations. Notice that, a priori, one would expect methods with high variability provide a more accurate best result, because more variability implies more chances of getting closer to the true solution at least once. By comparing Fig. 7 and 3 we see that this holds true for irMxNE and LCMV, that score quite well in terms of best solution and relatively high in terms of variability. Of course, one should prefer methods that have low DLE and also low variability. For instance, SSM has the lower standard deviation while being the second most accurate in terms of the best solution; MNE, on the other hand, has high variability but also high localization error, and is therefore the least recommended method for this type of data.

We also observed two unexpected results. First, for distributed methods we observed the best performances in correspondence of λ =1, i.e. the largest value of the regularization parameter: even though the reconstructions become more widespread, the peak gets closer to the true source. This result is puzzling because $\lambda = 1$ corresponds, in principle, to very noisy signals, while the input data are quite clean; on the other hand, it finds partial confirmation in Krishnaswamy et al. (2017) where authors use $\lambda = \frac{1}{9}$ for MEG data and $\lambda = 1$ for EEG data. We speculate this fact might be due to forward modeling errors, related to the volume conduction problem, that reduce the effective SNR of otherwise clean EEG data; in any case, more investigations are needed to clarify this point. Second, despite the presence of both deep and superficial

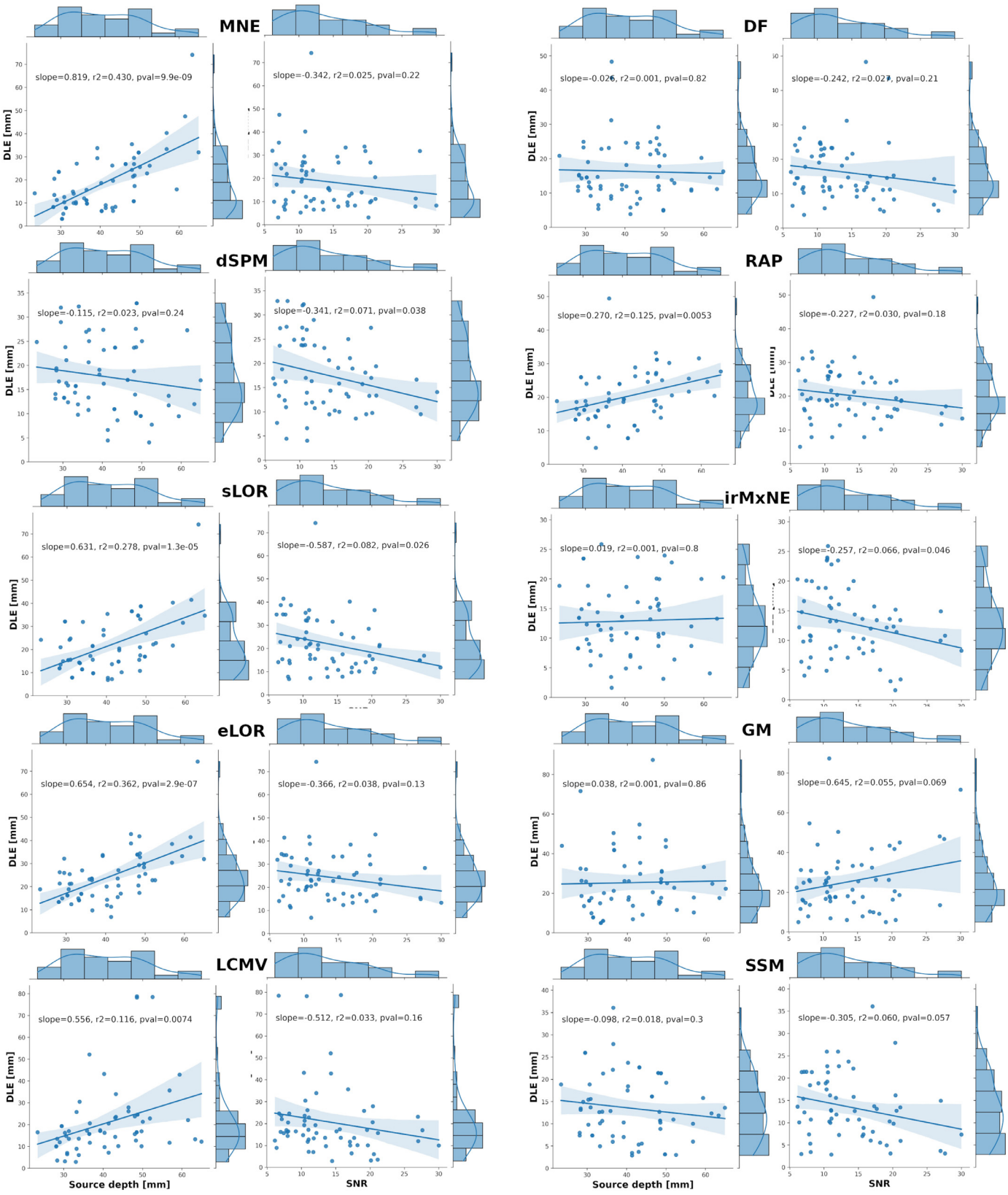


Fig. 11. For each method, regression lines of DLE as a function of source depth (left) and as a function of SNR (right).

sources in the dataset, the depth weighting parameter appears to have little or no impact for SLOR and GM; for MNE, dSPM, irMxNE and LCMV the best value is either zero or one.

As a side note, it is interesting to compare the low standard deviation of dSPM with the high standard deviation of MNE. Indeed, we recall that the dSPM solution is obtained from the MNE solution through a "noise normalization" procedure. Apparently, such noise normalization contributes a little to reduce localization error but rather to reduce the dependence on the input parameters.

Overall, our results point out that some of the most largely used methods have a relatively strong dependence on the value of the input parameters: as such, objective and reliable criteria for choosing their values would be advisable in the future.

4.3. Impact of covariates on source localization

In our study, the best localization accuracy was often obtained with 32 channels, and we observed no major differences when using higher density montages; in fact, for some methods we found significant differences in the DLE obtained with different montages, and the lower DLE was almost systematically associated with lower density montage. On the other hand, increasing the number of channels leads to a reduced spatial dispersion for distributed methods, and several methods show reduced variability with respect to the input parameter when higher density montages are used: in this respect, higher density montages do provide better results in terms of increased stability.

The fact that the DLE does not decrease when using more sensors may be unexpected, however few considerations can help clarify this apparently counterintuitive facet. There are two specific features of the analyzed data set that make 32 channels sufficient: (i) the dipolar nature of the source under scrutiny and (ii) the high SNR of the data. Indeed, localizing a single current dipole amounts to estimating 6 parameters: 32 channels, corresponding to 32 equations or constraints, are more than enough provided good quality data (Michel and He, 2019). This is the case for our dataset, since the uniform spatial distribution of the 32 channels covers the whole head, and because the SNR of the data is very high. Therefore, given our specific experimental conditions, we do not expect a substantial gain in localization accuracy when adding more channels. This result is in agreement with literature on the topic (Baroumand et al., 2018; Sperli et al., 2006). From a mathematical perspective, higher density montages correspond to taller leadfield matrices featuring larger condition numbers, i.e. the problem becomes more ill-conditioned and more regularization might be required. As a partial confirmation, significant differences between montages were observed almost exclusively for distributed methods that are expected to suffer more from an ill-conditioned leadfield; the best DLE of SSM, DF, irMxNE and LCMV showed no significant dependence on the montage.

Overall, our results suggest that 32 channels are enough to reconstruct a focal source from high-SNR data; this might include single time points with one strong source, but also single topographies obtained by ICA, or specific frequencies. More complex configurations (or more noisy data sets), on the other hand, are likely to benefit from additional sensors. In this sense, the results of the present study do not generalize to experimental conditions such as those including multiple sources and/or sources with non-negligible spatial extent.

Similar considerations may explain the absence of significant dependence between localization accuracy and data SNR. Indeed, this result certainly does not hold in general, but it is a consequence of the high-SNR of the data (the minimum value here is 5dB) and of the focal nature of the sources.

Finally, for several distributed methods the localization accuracy is found worse for deeper sources; considering that this study was performed using the best solution across all parameters combination, our results suggest that depth weighing is not enough to remove the bias towards superficial sources. On the other hand, for all dipolar and sparsity-promoting methods except RAP no significant relationship between the DLE and the source depth is found. In general, deeper sources are more correctly localized by dipolar and sparsity-promoting methods.

4.4. Comparison with previous works

Several comparisons between ESI methods have been performed in recent years.

First of all, the same dataset used here was used in Mikulan et al. (2020), where an exemplar analysis with three ESI methods (MNE, ELOR and DSPM) was performed. Here, we considerably extended the comparison to include also more recent methods: noteworthy, we observed that recent methods such as SSM and irMxNE do outperform older ones, hereby confirming recent results (Luria et al., 2020; Samuelsson et al., 2021). In addition, we studied the impact of regularization and depth weighting parameters more in detail, highlighting similarities and differences between different methods.

Another study that relates quite directly to our current study is (Luria et al., 2020), where the authors compared retrospectively SSM, RAP and wMNE with the results of an ECD analysis on epileptic subjects; also in this case the reference source was a point source (even though it came from a former analysis and not a real signal source) and the authors find MNE to be the least accurate while SSM the most accurate.

There is also an increasing number of studies that find different methods have substantially similar, good performances. Beniczky et al. (2016) compared 5 ESI methods on ictal EEG data, and found a general agreement between methods, with MNE being the least accurate; although not statistically significant, this result represents a partial confirmation of our findings. In (Pellegrino et al., 2020) the authors compared DSPM, MNE, SLOR and cMEM (Chowdhury et al., 2013) (not tested here) in a clinical scenario, and found excellent performances for all of them. Their results showed sLOR was slightly but significantly better than dSPM, while in our findings SLOR was at times better than MNE but never better than dSPM. In (Tenney et al., 2014) the authors studied the accuracy of dipolar methods (ECD, MUSIC), imaging methods (MNE, sLORETA, SWARM) and different implementation of SAM beamformer as compared to intracranial EEG (iEEG) and the resection areas in a large cohort of pediatric patients with intractable epilepsy. The accuracy of all these methods was relatively similar when compared to the ground truth. The concordance or discordance of MUSIC with iEEG was the best predictor of long-term seizure outcome. In (de Gooijer-van de Groep et al., 2013) the authors recently compared interictal MEG spikes using MUSIC, SAM(g2), and sLORETA to interictal discharges recorded with iEEG. These three MEG methods showed similar concordance with iEEG but differed depending on the brain region in which the spike was located.

In other comparisons, the authors use datasets where more widespread activations are present. In (Koessler et al., 2010) the authors investigated the best conditions for locating the epileptogenic zone with hd–EEG and compared five different methods; their results highlighted that distributed methods are more appropriate to localize a widespread epileptogenic zone than a focal one, and that ictal spikes with focal scalp electric field are better localized by dipolar methods (ECD and MUSIC). This last finding is in line with our results. In (Mahjoory et al., 2017) the authors used MNE, eLOR and LCMV for source reconstruction and connectivity estimation from resting state data, and found relative agreement in source localizations, more than in connectivity estimation.

Most of these studies feature three important differences with respect to the one reported here. First, in previous studies the definition of true source was necessarily more vague and less accurate. Second, the data were analyzed by expert users who almost certainly had expectations, and could tune each method to provide a coherent picture; here, instead, each method was applied independently, and in an automated fashion. Finally, none of these studies makes explicit reference to the setting of the input parameters.

5. Conclusions

In this study we investigated in vivo the spatial accuracy of ESI methods, and its dependence on the input parameter(s) in the case of focal generators. Our data show good levels of accuracy of ESI techniques, with the best solution across parameters and methods within 1 cm from the true source. This is true also when ESI is applied to "conventional" (32 channels) rather than dense (64, 128, 256 channels) EEG recordings. We remark anyway that the absolute values of the localization error we report might be impacted by a number of factors, including the presence of burr holes that were not modeled in this study; in this sense, what is particularly interesting in our study is the comparative performance of different methods rather than their absolute one: recent dipolar and sparsity-promoting methods, particularly SSM and irMxNE, provide significantly better results than older distributed methods such as MNE, both in terms of a higher accuracy with the optimal parameter choice, and a lower sensitivity to the value of the input parameter; notably, our results confirm a substantial progress in the field. We also observed negligible impact of depth weighting for SLOR and GM, and a general preference for larger values of the regularization parameter λ in all distributed methods, a result that finds partial support in the literature.

Overall findings suggest that ESI methods can achieve very good localization accuracy with focal sources, and thus reinforce the importance that ESI may have in the clinical context, especially when applied to identify the precise target in candidates for resective epilepsy surgery.

Declaration of Competing Interest

None.

Credit authorship contribution statement

Annalisa Pascarella: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – review & editing. **Ezequiel Mikulan:** Conceptualization, Methodology, Writing – review & editing, Resources, Data curation. **Federica Sciacchitano:** Conceptualization, Methodology, Software. **Simone Sarasso:** Conceptualization, Methodology, Writing – review & editing, Resources, Data curation. **Annalisa Rubino:** Resources, Data curation. **Ivana Sartori:** Resources, Data curation. **Francesco Cardinale:** Resources, Data curation. **Flavia Zauli:** Resources, Data curation. **Pietro Avanzini:** Conceptualization, Methodology, Writing – review & editing. **Lino Nobili:** Conceptualization, Methodology, Writing – review & editing, Resources, Data curation. **Andrea Pigorini:** Conceptualization, Methodology, Writing – review & editing, Resources, Data curation, Supervision. **Alberto Sorrentino:** Conceptualization, Methodology, Writing – review & editing, Supervision.

Data availability

Data used in this study are publicly available. Code used in this study is available upon request.

Acknowledgments

This research has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreements No. 785907 and No. 945539 (Human Brain Project SGA2 and SGA3). Annalisa Rubino and Lino Nobili received funding from the Italian Ministry of Health, Targeted Research Grant No. RF-2010-2319316. This work was developed within the framework of the DINOGMI Department of Excellence of MIUR 2018–2022 (law 232/2016). Annalisa Pascarella acknowledges GNCS-INDAM for partially supporting this research.

References

Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C., 2011. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage 54 (3), 2033–2044.

Baillet, S., Riera, J.J., Marin, G., Mangin, J.F., Aubert, J., Garnero, L., 2001. Evaluation of inverse methods and head models for EEG source localization using a human skull phantom. Phys. Med. Biol. 46 (1), 77.

Baroumand, A.G., Arbune, A.A., Strobbe, G., Keereman, V., Pinborg, L.H., Fabricius, M., Rubboli, G., Madsen, C.G., Jespersen, B., Brennum, J., et al., 2022. Automated ictal EEG source imaging: a retrospective, blinded clinical validation study. Clin. Neurophysiol. 141, 119–125.

Baroumand, A.G., van Mierlo, P., Strobbe, G., Pinborg, L.H., Fabricius, M., Rubboli, G., Leffers, A.-M., Uldall, P., Jespersen, B., Brennum, J., et al., 2018. Automated EEG source imaging: a retrospective, blinded clinical validation study. Clin. Neurophysiol. 129 (11), 2403–2410.

Becker, H., Albera, L., Comon, P., Gribonval, R., Wendling, F., Merlet, I., 2015. Brain–source imaging: from sparse to tensor models. IEEE Signal Process. Mag. 32 (6), 100–112.

Becker, H., Albera, L., Comon, P., Gribonval, R., Wendling, F., Merlet, I., 2016. Localization of distributed EEG sources in the context of epilepsy: a simulation study. IRBM 37 (5–6), 242–253.

Beniczky, S., Rosenzweig, I., Scherg, M., Jordanov, T., Lanfer, B., Lantz, G., Larsson, P.G., 2016. Ictal EEG source imaging in presurgical evaluation: high agreement between analysis methods. Seizure 43, 1–5.

Bertrand, Q., Massias, M., Gramfort, A., Salmon, J., 2019. Handling correlated and repeated measurements with the smoothed multivariate square-root lasso. Adv. Neural Inf. Process. Syst. 32.

Bidelman, G.M., 2018. Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. Neuroimage 175, 56–69.

Brodbeck, V., Spinelli, L., Lascano, A.M., Pollo, C., Schaller, K., Vargas, M.I., Wissmeyer, M., Michel, C.M., Seeck, M., 2010. Electrical source imaging for presurgical focus localization in epilepsy patients with normal MRI. Epilepsia 51 (4), 583–591.

Brodbeck, V., Spinelli, L., Lascano, A.M., Wissmeier, M., Vargas, M.-I., Vulliemoz, S., Pollo, C., Schaller, K., Michel, C.M., Seeck, M., 2011. Electroencephalographic source imaging: a prospective study of 152 operated epileptic patients. Brain 134 (10), 2887–2897.

Cai, C., Hashemi, A., Diwakar, M., Haufe, S., Sekihara, K., Nagarajan, S.S., 2021. Robust estimation of noise for electromagnetic brain imaging with the champagne algorithm. Neuroimage 225, 117411.

Cardinale, F., Rizzi, M., Vignati, E., Cossu, M., Castana, L., d'Orio, P., Revay, M., Costanza, M.D., Tassi, L., Mai, R., et al., 2019. Stereoelectroencephalography: retrospective analysis of 742 procedures in a single centre. Brain 142 (9), 2688–2704.

Chowdhury, R.A., Lina, J.M., Kobayashi, E., Grova, C., 2013. Meg source localization of spatially extended generators of epileptic activity: comparing entropic and hierarchical bayesian approaches. PLoS ONE 8 (2), e55969.

Chowdhury, R.A., Merlet, I., Birot, G., Kobayashi, E., Nica, A., Biraben, A., Wendling, F., Lina, J.-M., Albera, L., Grova, C., 2016. Complex patterns of spatially extended generators of epileptic activity: comparison of source localization methods cMEM and 4-exso-MUSIC on high resolution EEG and MEG data. Neuroimage 143, 175–195.

Cohen, D., Cuffin, B.N., Yunokuchi, K., Maniewski, R., Purcell, C., Cosgrove, G.R., Ives, J., Kennedy, J.G., Schomer, D.L., 1990. Meg versus eeg localization test using implanted sources in the human brain. Annal. Neurol.: Off. J. Am. Neurol. Assoc. Child Neurol. Soc. 28 (6), 811–817.

Combrisson, E., Vallat, R., O'Reilly, C., Jas, M., Pascarella, A., Saive, A.-l., Thiery, T., Meunier, D., Altukhov, D., Lajnef, T., et al., 2019. Visbrain: a multi-purpose GPU-accelerated open-source suite for multimodal brain data visualization. Front. Neuroinform. 13, 14.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. segmentation and surface reconstruction. Neuroimage 9 (2), 179–194.

Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., Halgren, E., 2000. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron 26 (1), 55–67.

Dassios, G., Hadjiloizi, D., 2009. On the non-uniqueness of the inverse problem associated with electroencephalography. Inverse Probl. 25 (11), 115012.

David, O., Job, A.-S., De Palma, L., Hoffmann, D., Minotti, L., Kahane, P., 2013. Probabilistic functional tractography of the human cortex. Neuroimage 80, 307–317.

Deledalle, C.-A., Vaiter, S., Fadili, J., Peyré, G., 2014. Stein unbiased gradient estimator of the risk (SUGAR) for multiple parameter selection. SIAM J. Imaging Sci. 7 (4), 2448–2487. doi:10.1137/140968045.

Engemann, D.A., Gramfort, A., 2015. Automated model selection in covariance estimation and spatial whitening of MEG and EEG signals. Neuroimage 108, 328–342. doi:10.1016/j.neuroimage.2014.12.040. https://www.sciencedirect.com/science/article/pii/S1053811914010325

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al., 2012. 3D slicer as an image computing platform for the quantitative imaging network. Magn. Reson. Imaging 30 (9), 1323–1341.

Fonov, V.S., Evans, A.C., McKinstry, R.C., Almli, C.R., Collins, D.L., 2009. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. Neuroimage (47) S102.

Gramfort, A., Kowalski, M., Hämäläinen, M., 2012. Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. Phys. Med. Biol. 57 (7), 1937.

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L., Hämäläinen, M.S., 2014. MNE software for processing MEG and EEG data. Neuroimage 86, 446–460.

de Gooijer-van de Groep, K.L., Leijten, F.S.S., Ferrier, C.H., Huiskamp, G.J.M., 2013. Inverse modeling in magnetic source imaging: comparison of MUSIC, SAM (g2), and sLORETA to interictal intracranial EEG. Hum. Brain Mapp. 34 (9), 2032–2044.
Grova, C., Daunizeau, J., Lina, J.-M., Bénar, C.G., Benali, H., Gotman, J., 2006. Evaluation of EEG localization methods using realistic simulations of interictal spikes. Neuroimage 29 (3), 734–753.
Hämäläinen, M.S., Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Med. Biol. Eng. Comput. 32 (1), 35–42.
He, B., Astolfi, L., Valdés-Sosa, P.A., Marinazzo, D., Palva, S.O., Bénar, C.-G., Michel, C.M., Koenig, T., 2019. Electrophysiological brain connectivity: theory and implementation. IEEE Trans. Biomed. Eng. 66 (7), 2115–2137.
Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5 (2), 143–156.
Kaiboriboon, K., Lüders, H.O., Hamaneh, M., Turnbull, J., Lhatoo, S.D., 2012. EEG source imaging in epilepsy–practicalities and pitfalls. Nature Rev. Neurol. 8 (9), 498.
Koessler, L., Benar, C., Maillard, L., Badier, J.-M., Vignal, J.P., Bartolomei, F., Chauvel, P., Gavaret, M., 2010. Source localization of ictal epileptic activity investigated by high resolution EEG and validated by SEEG. Neuroimage 51 (2), 642–653.
Krishnaswamy, P., Obregon-Henao, G., Ahveninen, J., Khan, S., Babadi, B., Iglesias, J.E., Hämäläinen, M.S., Purdon, P.L., 2017. Sparsity enables estimation of both subcortical and cortical activity from MEG and EEG. Proc. Natl. Acad. Sci. 114 (48), E10465–E10474.
Lanfer, B., Scherg, M., Dannhauer, M., Knösche, T.R., Burger, M., Wolters, C.H., 2012. Influences of skull segmentation inaccuracies on EEG source analysis. Neuroimage 62 (1), 418–431.
Leahy, R.M., Mosher, J.C., Spencer, M.E., Huang, M.X., Lewine, J.D., 1998. A study of dipole localization accuracy for MEG and EEG using a human skull phantom. Electroencephalogr. Clin. Neurophysiol. 107 (2), 159–173.
Lin, F.H., Witzel, T., Ahlfors, S.P., Stufflebeam, S.M., Belliveau, J.V., Hamalainen, M.S., 2006. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. Neuroimage 31, 160–171.
Luria, G., Duran, D., Visani, E., Sebastiano, D.R., Sorrentino, A., Tassi, L., Granvillano, A., Franceschetti, S., Panzica, F., 2020. Towards the automatic localization of the irritative zone through magnetic source imaging. Brain Topogr. 33 (5), 651–663.
Mahjoory, K., Nikulin, V.V., Botrel, L., Linkenkaer-Hansen, K., Fato, M.M., Haufe, S., 2017. Consistency of EEG source localization and connectivity estimates. Neuroimage 152, 590–601.
Matsumoto, R., Nair, D.R., LaPresto, E., Najm, I., Bingaman, W., Shibasaki, H., Lüders, H.O., 2004. Functional connectivity in the human language system: a cortico-cortical evoked potential study. Brain 127 (10), 2316–2330.
Mégevand, P., Spinelli, L., Genetti, M., Brodbeck, V., Momjian, S., Schaller, K., Michel, C.M., Vulliemoz, S., Seeck, M., 2014. Electric source imaging of interictal activity accurately localises the seizure onset zone. J. Neurol. Neurosurg. Psychiatry 85 (1), 38–43.
Michel, C.M., He, B., 2019. Eeg source localization. Handb. Clin. Neurol. 160, 85–101.
Mikulan, E., Russo, S., Parmigiani, S., Sarasso, S., Zauli, F.M., Rubino, A., Avanzini, P., Cattani, A., Sorrentino, A., Gibbs, S., et al., 2020. Simultaneous human intracerebral stimulation and HD-EEG, ground-truth for source localization methods. Sci. Data 7 (1), 1–8.
Montani, V., Chanoine, V., Stoianov, I.P., Grainger, J., Ziegler, J.C., 2019. Steady state visual evoked potentials in reading aloud: effects of lexicality, frequency and orthographic familiarity. Brain Lang. 192, 1–14.
Mosher, J.C., Leahy, R.M., 1999. Source localization using recursively applied and projected (RAP) MUSIC. IEEE Trans. Signal Process. 47 (2), 332–340.
Narizzano, M., Arnulfo, G., Ricci, S., Toselli, B., Tisdall, M., Canessa, A., Fato, M.M., Cardinale, F., 2017. Seeg assistant: a 3dslicer extension to support epilepsy surgery. BMC Bioinform. 18 (1), 124.
Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2010. Fieldtrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011.
Pascual-Marqui, R.D., Pascual-Montano, A.D., Lehmann, D., Kochi, K., Esslen, M., Jancke, L., Anderer, P., Saletu, B., Tanaka, H., Hirata, K., et al., 2006. Exact low resolution brain electromagnetic tomography (eLORETA). Neuroimage 31 (Suppl 1).
Pascual-Marqui, R.D., et al., 2002. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. Methods Find Exp. Clin. Pharmacol. 24 (Suppl D), 5–12.
Pellegrino, G., Hedrich, T., Porras-Bettancourt, M., Lina, J.-M., Aydin, U., Hall, J., Grova, C., Kobayashi, E., 2020. Accuracy and spatial properties of distributed magnetic source imaging techniques in the investigation of focal epilepsy patients. Hum. Brain Mapp..
Pestman, W., 2009. Mathematical statistics, ser.
Samuelsson, J.G., Peled, N., Mamashli, F., Ahveninen, J., Hämäläinen, M.S., 2021. Spatial fidelity of MEG/EEG source estimates: ageneral evaluation approach. Neuroimage 224, 117430.
Sohrabpour, A., Cai, Z., Ye, S., Brinkmann, B., Worrell, G., He, B., 2020. Noninvasive electromagnetic source imaging of spatiotemporally distributed epileptogenic brain sources. Nat. Commun. 11 (1), 1946.
Sommariva, S., Sorrentino, A., 2014. Sequential monte carlo samplers for semi-linear inverse problems and application to magnetoencephalography. Inverse Probl. 30 (11), 114020.
Sperli, F., Spinelli, L., Seeck, M., Kurian, M., Michel, C.M., Lantz, G., 2006. Eeg source imaging in pediatric epilepsy surgery: a new perspective in presurgical workup. Epilepsia 47 (6), 981–990.
Strohmeier, D., Haueisen, J., Gramfort, A., 2014. Improved MEG/EEG source localization with reweighted mixed-norms. In: 2014 International Workshop on Pattern Recognition in Neuroimaging, pp. 1–4. doi:10.1109/PRNI.2014.6858545.
Sun, R., Sohrabpour, A., Worrell, G.A., He, B., 2022. Deep neural networks constrained by neural mass models improve electrophysiological source imaging of spatiotemporal brain dynamics. Proc. Natl. Acad. Sci. 119 (31). e2201128119
Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M., 2011. Brainstorm: a user-friendly application for MEG/EEG analysis. Comput. Intell. Neurosci. 2011.
Tenney, J.R., Fujiwara, H., Horn, P.S., Rose, D.F., 2014. Comparison of magnetic source estimation to intracranial EEG, resection area, and seizure outcome. Epilepsia 55 (11), 1854–1863.
Unnwongse, K., Rampp, S., Wehner, T., Kowoll, A., Parpaley, Y., von Lehe, M., Lanfer, B., Rusiniak, M., Wolters, C., Wellmer, J., 2023. Validating EEG source imaging using intracranial electrical stimulation. Brain Commun. 5 (1), fcad023.
Valentin, A., Anderson, M., Alarcon, G., Seoane, J.J.G., Selway, R., Binnie, C.D., Polkey, C.E., 2002. Responses to single pulse electrical stimulation identify epileptogenesis in the human brain in vivo. Brain 125 (8), 1709–1718.
Van Veen, B.D., Van Drongelen, W., 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.
Viani, A., Luria, G., Bornfleth, H., Sorrentino, A., 2020. Where Bayes tweaks Gauss: conditionally gaussian priors for stable multi-dipole estimation. arXiv preprint arXiv:2006.04141.
Wipf, D., Nagarajan, S., 2009. A unified bayesian framework for MEG/EEG source imaging. Neuroimage 44 (3), 947–966.
Yao, J., Dewald, J.P.A., 2005. Evaluation of different cortical source localization methods using simulated and experimental EEG data. Neuroimage 25 (2), 369–382.