Bottom-up control of leakage in spectral electrophysiological source imaging via structured sparse bayesian learning

Eduardo Gonzalez-Moreira [a,b,c,#], Deirel Paz-Linares [a,d,#], Ariosky Areces-Gonzalez [a,e], Ying Wang [a], Min Li [a], Thalia Harmony [c], and Pedro A. Valdes-Sosa [a,d,*]

[a] The Clinical Hospital of Chengdu Brain Science fadsInstitute, MOE Key Lab for Neuroinformation, UESTC, Chengdu, China.[b] Central University "Marta Abreu" of Las Villas, Santa Clara, Cuba.[c] Institute of Neurobiology, UNAM, Querétaro, México.[d] Cuban Neuroscience Center, La Habana, Cuba.[e] University of Pinar del Río "Hermanos Saiz Montes de Oca", Pinar del Río, Cuba. [f] Montreal Neurological Institute
[#] contributed equally to this work
[*] corresponding authors:PAVS (pedro.valdes@neuroinformatics-collaboratory.org)

Abstract

Brain electrical activity in different spectral bands has been associated with diverse mechanisms underlying Brain function. Deeper reconnoitering of these mechanisms entails mapping in grayordinates (Gray Matter coordinates), the spectral features of electrophysiological Brain signals. Such mapping is possible through MEG/EEG signals, due to their wide Brain coverage and excellent temporal resolution in reflecting neural -electrical- activity. This process -coined Electrophysiological Source Imaging (ESI)- can only produce approximated images of Brain activity, which are severely distorted by leakage: a pervasive effect in almost any imaging technique. It has been proposed that leakage control to tolerable levels can be achived through using priors or regularization within ESI, but their implementation commonly yields meager statistical guaranties. We introduce bottom-up control of leakage: defined as maximum Bayesian evidence search braced with priors precisely on the spectral responses. This is feasible due to an instance of Bayesian learning of complex valued data: spectral Structured Sparse Bayesian Learning (sSSBL). "Spectral" refers to specific spatial topologies that are reflected by the MEG/EEG spectra. We also present a new validation benchmark based on the concurrency between high density MEG and its associated pseudo-EEG of lower density. This reveals that prevealing methods like eLORETA and LCMV can fall short of expectations whereas sSSBL exibits an exellent performance. A final qualitative assesment reveals that sSSBL can outline brain lessions using just low density EEG, according to the T2 MRI shine through of the affected areas.

[#] contributed equally as first author to this research

[*] corresponding author at The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Lab for Neuroinformation, University of Electronic Science and Technology of China, Chengdu, China. Tel.: +86-(028)- 61830972.

Email address: pedro.valdes@neuroinformatics-collaboratory.org

Keywords: MEG/EEG, leakage, spectral, SSBL, priors

Introduction

Band specific synchronized -spectral- activity is the underlying mechanism for large scale Brain integration or networks of functionally specialized regions from which coherent behavior and cognition emerge [Engel et al., 2001; Varela et al. 2001; Valdes-Sosa et al 1992; Nunez and Srinivasan 2006; Vidaurre et al 2018]. Its cause dwells mostly microscopically in the neural architecture and interactions within layers at the cortical columns [Freeman 1975; Jirsa 1997], but still macroscopically its topological organization -spatial distribution and connectivity pattern- at the macroscopic level possesses a tremendous descriptive power in behavioral and clinical aspects of healthy, developing, aging, and diseased human brains [Bullmore and Sporns 2009; Rubinov and Sporns 2010; Sporns 2011].

An important part in describing this spectral activity involves mapping its spatial distribution alone, by the localization of responsive areas at the observational level of experimental techniques like Functional Magnetic Resonance Imaging (fMRI) and Electro/Magnetoencephalogram (EEG/MEG) [Mantini 2007]. For either technique, mapping these responses cannot be tackled directly from the observed data: the spectral composition of fMRI signals is severely distorted (spectral leakage) by the slow metabolic-hemodynamic cascade of processes following the actual neural activations [Buxton 1998; Logothetis 2001], whereas EEG/MEG signals are affected in a different manner, their low resolution and blurring effect of head volume conduction (activation leakage) [Haufe et al 2013, Stokes et al 2017, Van de Steen et al 2019].

However, MEG/EEG signals have purely electrophysiological causes. Synchronic neural activity, at both microscopic and macroscopic Brain scales, elicit observable currents that are sensed through EEG electrodes or MEG gradiometers. Therefore, these signals are a sharper reflection of neural activity $\boldsymbol{\iota}(t)$, converted directly into voltage $\boldsymbol{v}(t)$ or magnetic field $\boldsymbol{b}(t)$ by the laws of the electromagnetic field in media alone. In the neuroimaging context this is denominated forward equation, a linear model whose operator is coined with the term Lead Field (LF) $\mathbf{L}_{\boldsymbol{v\iota}}$ [Grech et al 2008; Nunez et al 1994, Burle et al 2015], see Figure1. Full details on the notation in Section 1 of the "Supporting Information" (SI) SI-1. Thus, EEG/MEG carry on rich information of the neural mechanisms underpinned by purely electrical activity, that is registered with excellent temporal resolution and without spectral distortion. This is the reason why they have -initially EEG and later

MEG- provided most of the existent knowledge on the healthy brain dynamics [Freeman 1975, Lopes da Silva 2013, Larson-Prior 2013].

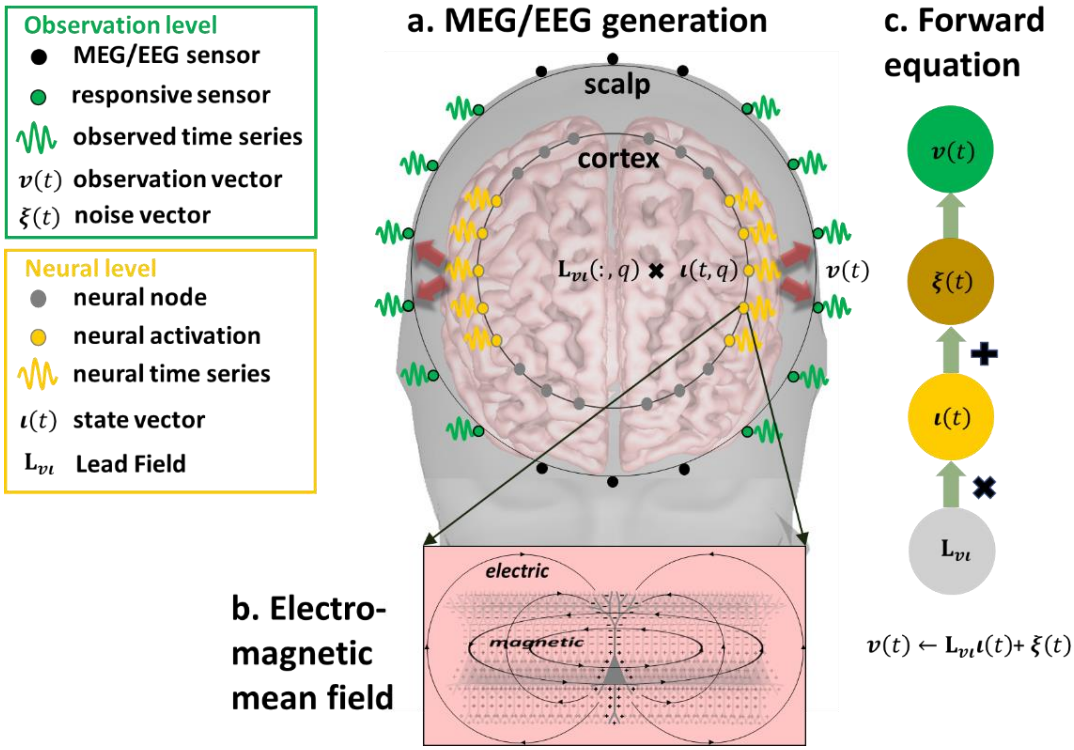


***Figure 1**: Illustration of the mechanisms underlying the generation of MEG/EEG signals. Their actual causes sojourn in the time variable (t) pattern of macroscopic currents at gray mater points (q) represented as $\boldsymbol{\iota}(t,q)$ (a). These currents, originated from a homogeneous polarization of the pyramidal layers, sustain a local electromagnetic mean field (b) that can be observed externally through MEG/EEG sensors (electrodes or gradiometers) $\boldsymbol{v}(t)$ (a). Whereas the spatial distribution of the external signal attributed to a single gray mater generator is explained in terms of a stationary linear equation for the electromagnetic field in media $\mathbf{L}_{v\iota}(:,q)$ (a). From the principle of superposition stems the linear generative model of MEG/EEG signals explaining the composited effect of currents distributed across the whole gray mater $\boldsymbol{v}(t) \leftarrow \boldsymbol{L}_{v\iota}\boldsymbol{\iota}(t) + \boldsymbol{\xi}(t)$ (c). Where the process $\boldsymbol{\xi}(t)$ represents the effect of instrumental or environmental noise.*

Leakage of electrophysiological source imaging

The basic flaw is the pervasive activation leakage of routine Electrophysiological Source Imaging (ESI) solutions, pledged with the coarse spatial undertone of sensed EEG/MEG responses with Gray Matter (in grayordinates or Gray Matter coordinate system) activations [Hämäläinen et al 1994; Pascual-Marqui et al 1994; Pascual-Marqui 1999; Srinivasan 1999, Babiloni et al 2001, Pascual-Marqui 2002, He et al 2019]. Where ESI is defined as the pseudo-inversion of the operator $\mathbf{L}_{v\iota}$ of the Forward equation. This inversion process or ESI is formally expressed through a transfer operator $\mathbf{T}_{\iota v}$

that attempts to explain in grayordinates the neural causes of signals [Nunez et al 1994, Burle et al 2015], see Figure 2.

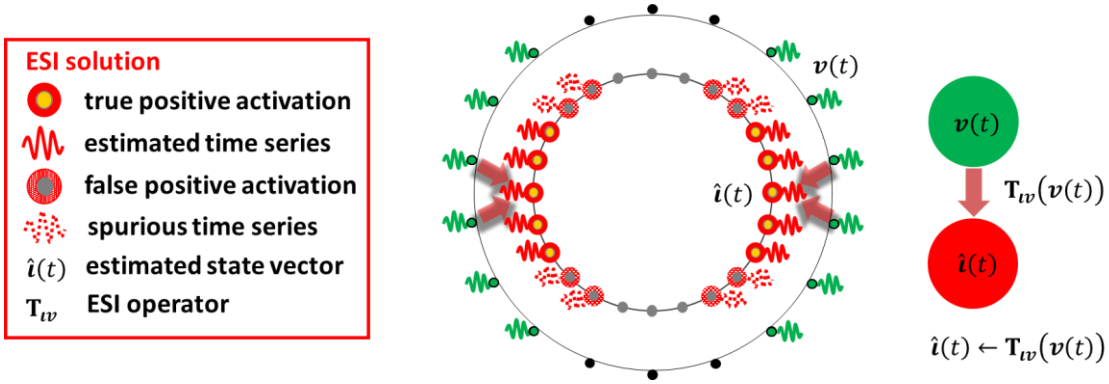


***Figure 2**: Basic leakage problem with ESI estimation. Attempting to retrieve the source activity via certain operator $\boldsymbol{T}_{\imath v}$ (this can be linear/nonlinear and encode different prior information) typical methods produce a blurred estimation $\hat{\boldsymbol{\imath}}(t)$. This carries on spurious (false positive) activations that extend beyond the truly active areas.*

The pitfall of leakage has been followed by the deployment of several "generations" of ESI methods within just two decades [Gonzalez-Moreira et al 2018b]. The term "generation" coins basic differences of the statistical learning method towards the construction of the EEG/MEG source transfer operator -via Linear, Nonlinear Univariate, Nonlinear Multivariate models- and the inclusion of prior information in order to ameliorate Leakage, we defer details on this to SI-2.

More sophisticated models (Nonlinear Multivariate) leveraged by the 3rd ESI generation (SI-2) can lead to very inaccurate results without adequate prior variable selection of 2nd generation methods, that could possibly constrain the grayordinate active subspace. This flaw is caused by inaccurate high dimensional computations and ill-conditioning of the coactivation matrix (which regard multivariate patterns) used in their nonlinear algorithms [Friston et al 2007; Paz-Linares et al 2018; Gonzalez-Moreira et al 2018a; 2018b; 2018c]. Likewise, this is critical for EEG/MEG connectivity methods preceded by any ESI solution [Coulclogh, 2016; Marinazzo et al 2019; Haufe et al 2013, Stokes et al 2017, Van de Steen et al 2019]. Learning the truly responsive sources in the spectral domain remains an uneasy enterprise underscoring any further source space analysis. This is connected first, to the selection of priors for controlling leakage, second, statistical guarantees of the inference framework, and third, appropriate experimental confirmation.

Bottom-up control of leakage

We aknowledge a major dificulty with leakage control in ESI arises from a very general methodological unconsitency, this is the top-down approach to a composited problem (Figure 3) in

two steps, first a procedure targeting the solution to an ill posed problem at the top a) but which affects ireversibly derived statistical quantities intended to gauge a given effect at the bottom b). We are concerned with the case in which a) is stated as the discovery of hidden source activity from MEG/EEG observations in time or spectral domain via ESI (Figure 3a); and b) as the statistical relevance of spectral responses upon which the Leakage effect is ultimately observed. The latter are irrversibly affected by ESI inacurracy and therefore are a common target of Leakage corrections (Figure 3b).

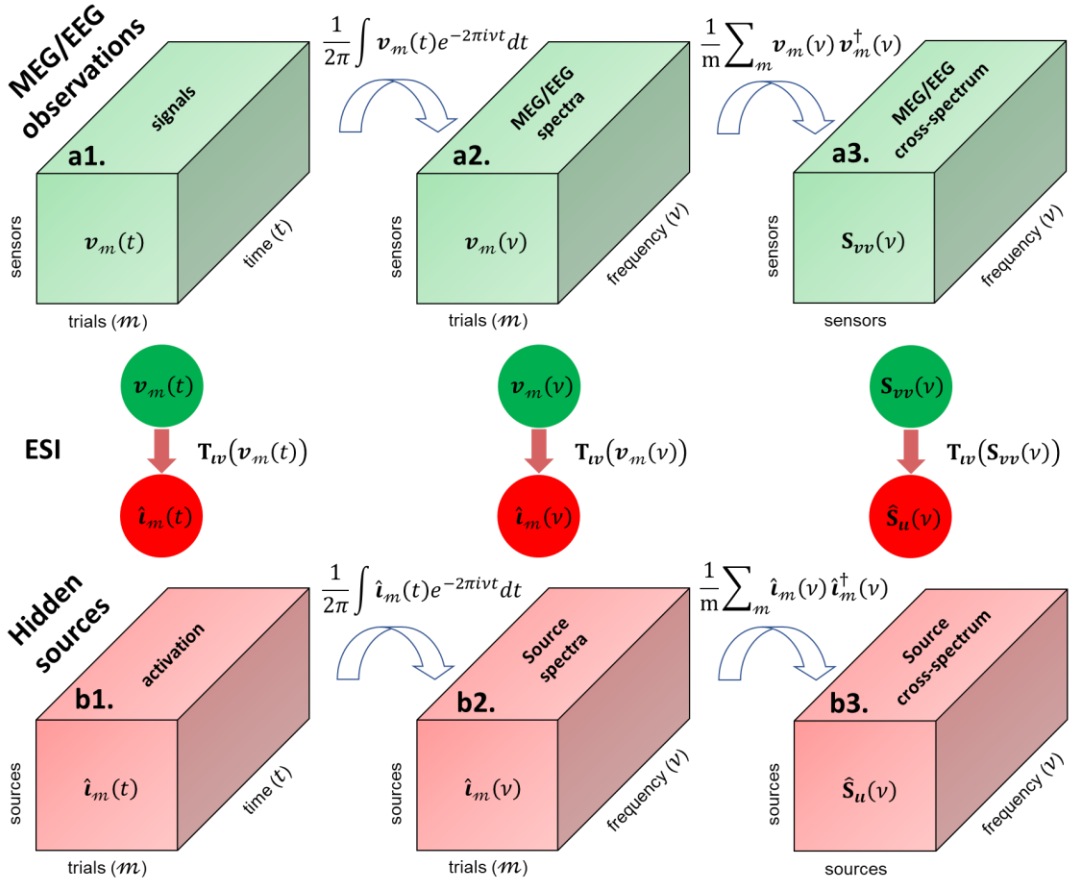


Figure 3*: Transit from observations (a) to hidden source activation (b) via three possible ESI routes. The starting point is a tensor structure containing MEG/EEG signals (a1) that are recorded through time (**t**) and for different trials (**m**). These are used to discover spectral properties encoded by the cross-spectrum tensor (b3), i.e. the activations (tensor diagonal entries) and coactivation (tensor off-diagonal entries). From left to right, the ESI routes to define the source transfer operator stride across qualitatively different learning paths: via independent instances in time a1=>b1 and frequency a2=>b2 domains or unifidely via the cross-spectrum a3=>b3.*

As Figure 3 illustrates, the alternatives for solving leakage stride across three different ESI routes $a_k \rightarrow b_k$, k = 1,2,3, in the transit from observations a) to hidden source activation b). Note that the

segregation of ESI routes refres to the form in which the data is used for statistical learning of the transfer operator $\mathbf{T}_{\boldsymbol{\iota v}}$ and in disregar to the type of solution which can be either Linear, Nonlinear univariated or Nonlinear multivariated. We enphasize that the top-down inconsistency relies on the lack of statistical guaranties in leakage control upon b_3 via ESI routes $a_1 \rightarrow b_1$ and $a_2 \rightarrow b_2$, in other words, by means of priors on independent instances of source activity. Therefore, the only available alternative is bottom-up Leakage control by placing priors directly on the cross-spectrum b_3 whose estimation is carried out via ESI route $a_3 \rightarrow b_3$.

Indeed this explains why some implementions of 2nd and 3rd generation methods are far superior to the linear methods in the first generation. The reason relies on the idea that, within the nonlinear learning process, parameters underlaying the estimation across samples are prone to perform leakage control. An importamt limitation hitched to direct spectral ESI solutions is the unavailability of learning models for complex valued data. Aknowledging these facts, we present a spectral modality of ESI provided with the property of botom-up control of leakage. The target of our method is to directly estimate the cross-spectrum tensor b_3, with complex valued structured sparsity priors via Bayesian learning. We refer to this method as Spectral Structured Sparse Bayesian Learning (sSSBL).

Spectral Structured Sparse Bayesian Learning (sSSBL) and extensions (sSSBL++)

The basic principle stems behind the idea of EEG/MEG scalp-spectral features are caused by focalized cortical topologies and are therefore a better target for sparse ESI inversion (Figure 4a). sSSBL targets individual topologies by means of a group penalization model upon the sample space of spectral source reponses $(\boldsymbol{\iota}_1(\nu) \cdots \boldsymbol{\iota}_m(\nu))$ that may possibly explain the observed cros-sspectum $(\mathbf{S}_{\boldsymbol{vv}}(\nu))$ (Figure 4b). This translates into a sparse univariate model for the selection of rows/colums corresponding to diagonal elements on the source cross-spectrum $(\mathbf{S}_{\boldsymbol{\iota\iota}}(\nu))$, whose statistical relevance is controled locally through the variances $\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2$ of the Hierarchical Elastic Net model. The global level of sparsity is controlled via a flexible twofolded regularization model with parameters (α, ρ) (Figure 4c). See the details on the formulation of the model underlying the sSSBL model in SI-3 and SI-4.

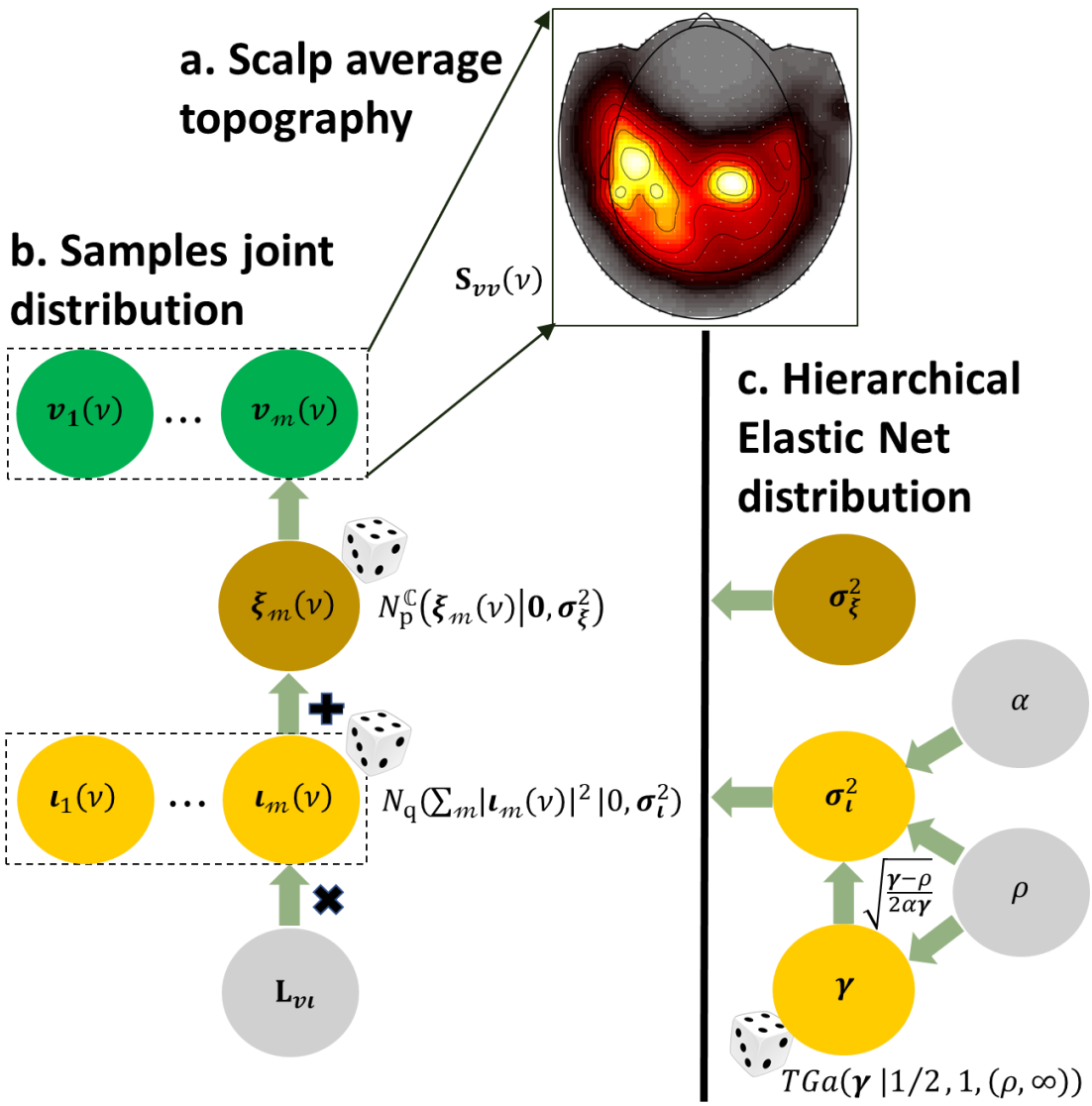


Figure 4*: Model of the Spectral Structured Sparse Bayesian Learning. Priors are used directly on the spectral domain of the MEG/EEG generative model, to target source activity that explains a statistical tendency (a) at a given frequency of the observed cross-spectrum $\boldsymbol{S}_{vv}(\nu)$. Gaussianity is assumed at two levels (b): First, on the spectral noise process $\boldsymbol{\xi}_m(\nu)$, as many typical ESI methods do in the time domain, which yields in this case a hermitian Gaussian model of the MEG/EEG cross-spectrum with variance spectral noise $\boldsymbol{\sigma}_{\xi}^2$. Second, on the spectral activations $\boldsymbol{\iota}_m(\nu)$, but through a joint model of the different instances m. The statistical relevance of a given activation pattern common for all instances m is determined by the variances $\boldsymbol{\sigma}_{\iota}^2$. The estimation of these variances is based on the Hierarchical Elastic Net (c), with sparse variable selection performed through the parameters $\boldsymbol{\gamma}$ of Truncated Gamma distribution. The global level of sparsity or shrinkage is controlled by α and ρ.*

The statistical gurantees in part rely on the maximum evidence towards determination of the variances that control leakage. Within the inference process these variances directly affect the source transfer

operator $\mathbf{T}_{\iota\nu}$, whose estimation is carried out here suboptimally, but highly eficientlly, via the Expectation-Maximization (E-M) algorithm. See the graphical sSSBL algorithm in the section 1 of "Materials and Methods" (MM) MM-1. With "suboptimal" we highlight the E-M limited capacity to only provide a local optima of the model evidence, but which can be higly accurate with the further inclusion of priors. We denominate this expanded sSSBL (sSSBL++). A critical set of five priors is introduced onto the model of Figure 4 serving to the purpose of not only constraining the optimal evidence search process but providing the estimated source activation with relevant physical features.

We use a joint activation field model, which introduces two different type of links (Figure 5a). First, a link onto its three spatial directions, via probabilistic 2D rotational invariance or isotropy (Figure 5b) reducing the degree of freedom, but still regarding surface normals as the field preferential direction. This diverges from the use of rigid constrains in assuming that surface normals are the only plausible orientation for the field of cortical currents. Second, a link onto the fields of neighbour cortical points via the Laplacian, which builds on assumptions of a graph structure underlaying the acitve sources spatial distribution (Figure 5c).

We aknowledge that ESI solutions based on cortical space can be wrapped at or biased towards surface giri due to atenuation of fields in the deeper sulci areas. Therefore we introduce a compensation factor based on curvature coeficients to unwrapping the Lead Field (Figure5d). To avoid the reverse bias towards sulci the final estimator is computed as the average of two preliminar solutions, one favoring the sulci activation and a second one doing so onto the giri.

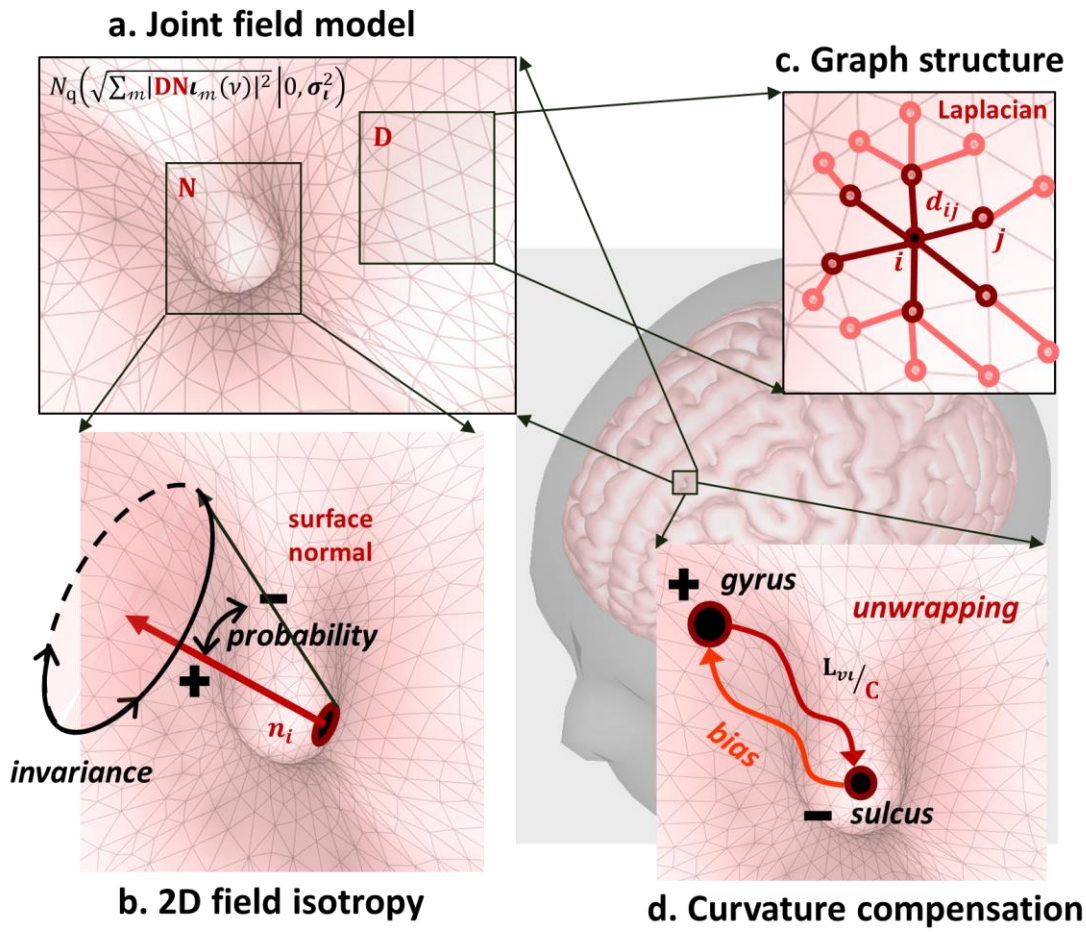


Figure 5*: Links introduced in the Gaussian distribution of the spectral activations via the composited effect of two operators (a). First, reducing the field components 3D degeneracy through the normal projection operator (**N**) and second, introducing dependencies between neighbor activations through the Laplacian operator (**D**). **N** yields rotational invariance of the activations Gaussian probability model, with pivot axis on the surface normal ($\boldsymbol{n_i}$) at every cortical point (**i**). The probability decreases as the orientation of activity deviates from the normal direction. Whereas with **D** the probabilities are not assigned individually to cortical points (**i**), but to its linear weighted combination with neighbors (**j**). With weights of the Laplacian matrix elements ($\boldsymbol{d_{ij}}$). Compensation of ESI curvature wrap or bias towards gyri (d). The operator (**C**) represents the curvature of the cortical surface to unwrapping the Lead Field and therefore the ESI solution.*

We also extended the sample-space group penalization to a 3D cartesian space, with the inclusion of two aditional spaces (Figure 6a). First, the space of frequencies within a spectral band carrying on common MEG/EEG scalp topologies (Figure 6b) as typicaly observed in real data. Second, the space of generators within cortical parcels, asuming the coactivation of generators within areas that are defined upon a structural or functional atlas (Figure 6c).

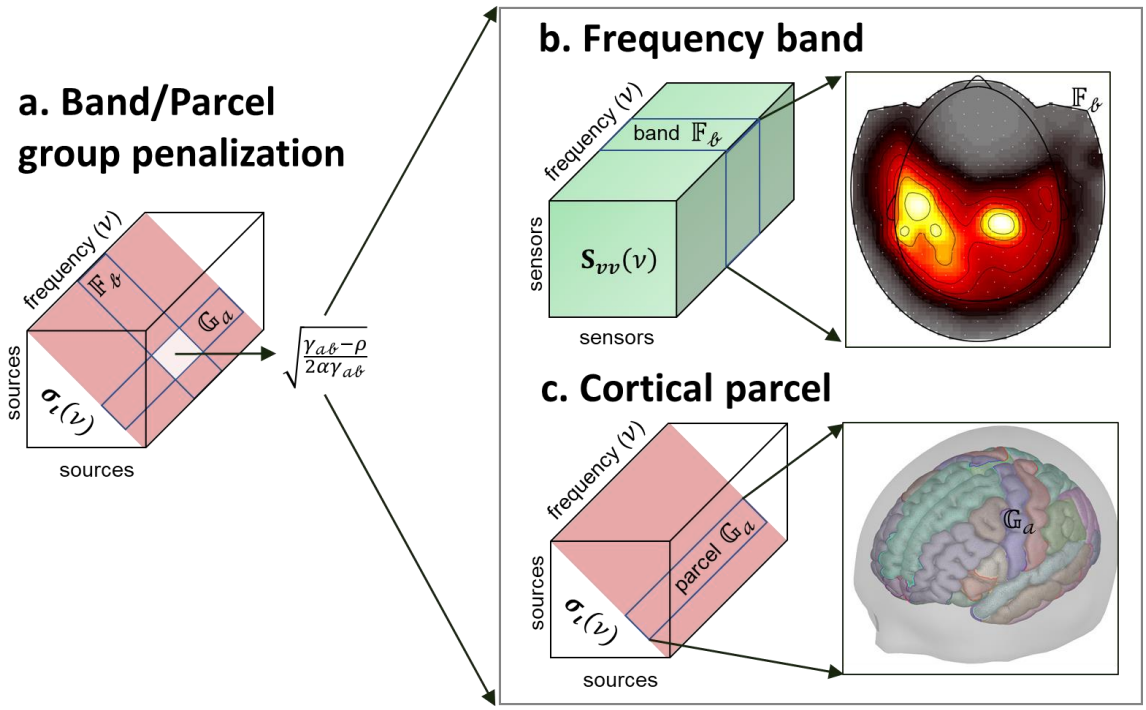


***Figure 6**: Group penalization of the Hierarchical Elastic Net with implications in the estimation of the variances $\boldsymbol{\sigma}_\iota$ (a), which determine the statistical relevance of activation patterns. As a consequence of this penalization the variances are computed from a single parameter γ_{ab} that performs a variable selection of groups of parameters. These are representing certain activations belonging to the interception of two groups $\mathbb{F}_b$ (band) and $\mathbb{G}_a$ (parcel). First, for a spectral range $\mathbb{F}_b$ corresponding to an electrophysiological band (b), with common observable spatial patterns for all the frequency components within this band. Second, activity corresponding to generators within a cortical area $\mathbb{G}_a$ (c), defined into a given neuroanatomical Atlas, which acknowledges activations within delimited areas are highly integrated and thus prone to appear simultaneously.*

Results and Discussion

Concurrency of sSSBL++ in the low-density pseudo-EEG of a high-density MEG

In evaluating the sSSBL++ performance we diverge from the use of synthetic MEG/EEG signal simulations, typically stablished upon idealization of the spatial distribution and spectral composition of underlaying neural activity [Pascual-Marqui et al 2014; Haufe and Arne 2016]. This is a major flaw of the state-of-the-art validation methods that deserves special attention here. We aim to resolve this by leveraging the concept of concurrency between neuroimaging techniques (Wang, 2019), through introducing a completely new validation benchmark based on the comparison between ESI solutions for the MEG and EEG modalities (Figure 7). One from a high-density real MEG signal (Figure 7a) to establish a landmark for source activation (Figure 7b), that is later used to simulate the pseudo signal of a low-density EEG (Figure 7c). The other, attempting to retrieve the original source activation from the pseudo EEG signal alone (Figure 7d).

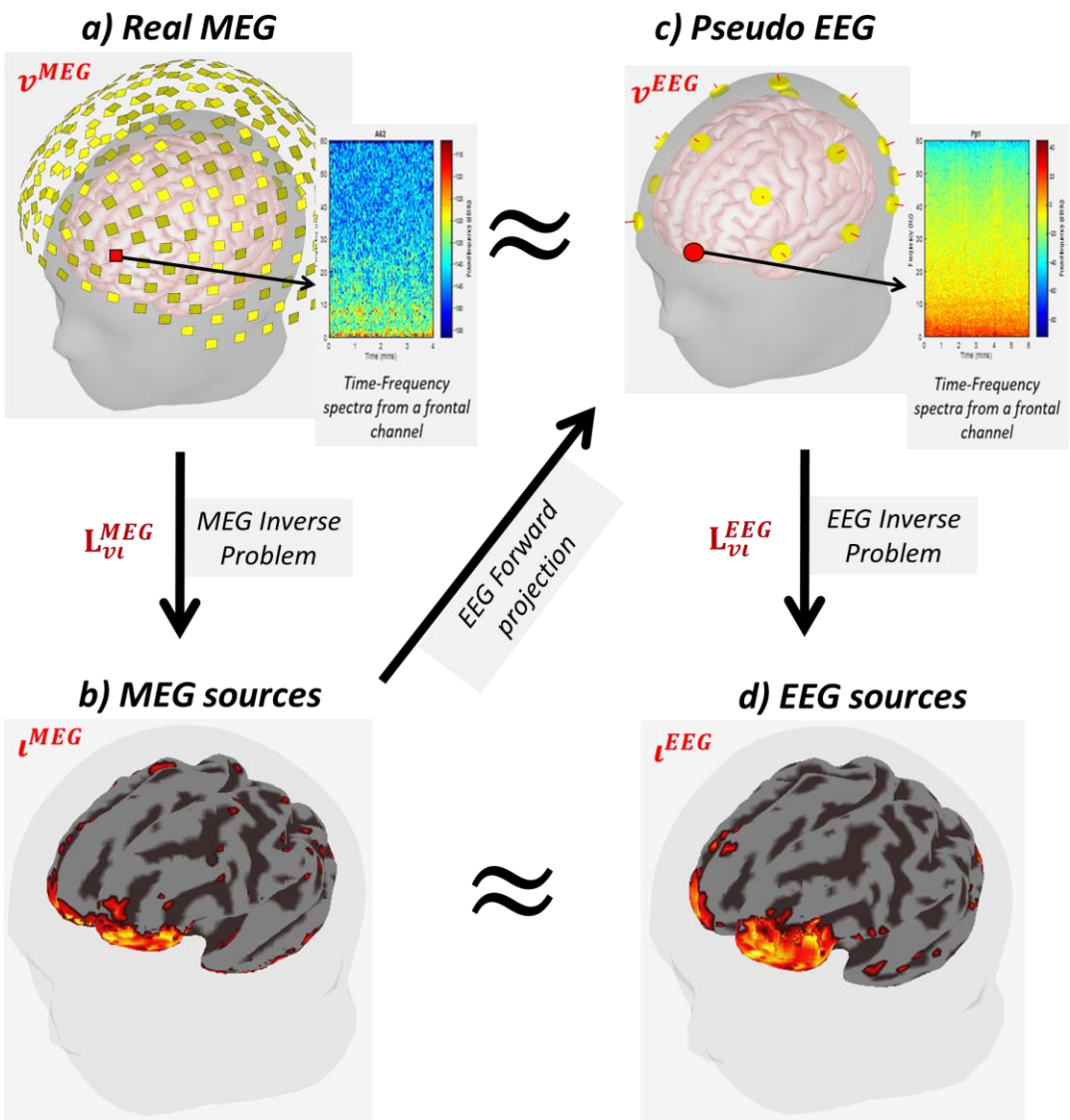


Figure 7*: Novel methodology for ESI validation, based on the concurrency between MEG and its pseudo EEG. a) Actual distribution of MEG sensors in the subject's native space and time frequency composition a frontal sensor signal. b) A landmark for alpha band spectral activity construed through sSSBL++ for the MEG signals, with Brainstorm head model of the three overlapping spheres. c) 10-20 EEG system adjusted to the native subject's space, used to produce pseudo-EEG from the projected MEG alpha source landmark, with a Brainstorm head model of the Boundary Element Method. With sSSBL++ we achieve a concurrent pseudo-EEG to judge from the time-frequency composition of the signal for the frontal sensor analogous to that of the MEG. Concurrent alpha source activity retrieved from the pseudo-EEG with the sSSBL++.*

Following the scheme of Figure 7 we studied the sSSBL++ concurrency on real MEG -pseudo EEG-data in five bands of frequency (delta 0-4 Hz, theta 4-8 Hz, alpha 8-12 Hz, beta 12-16 Hz and gamma 16-32 Hz). The real MEG was selected from the Human Connectome Project (HCP) database [Van Essen et al 2013], opportunely after its selection by Brainstorm team [Tadel et al 2011] to illustrate the MEG database usage, see details in MM-2.

The sSSBL++ exhibited a good performance in the concurrency test (Figure 8a), to judge roughly for its similarity in performance with typical examples of idealized simulations [Haufe et al 2013, Stokes et al 2017, Van de Steen et al 2019]. In SI-7 we offer an example of the results for sSSBL++ and other methods in such simulations. Meaning that major activations retrieved from the pseudo-EEG across areas coincided with the real-MEG landmark with only minor leakage distortion. This was also supported by the concurrency of fluctuations across frequency from activity in the slow band delta to the faster beta band. For both pseudo-EEG and real-MEG, these transited smoothly from frontal areas (delta) to occipital areas (alpha) as reported by multiple studies [da Silva 2013]. The same effect was observed for the transition bands theta (delta → alpha) and beta (alpha → gamma). Leakage of the pseudo-EEG persisted but relevantly did not overpass specific anatomical areas where major activations were detected.

The evaluation was extended to a prevalent instance of ESI: The Exact Low-Resolution Tomographic Analysis (eLORETA) [Pascual-Marqui et al 2011] (Figure 8b). See details on methods used for comparison in MM-2. Roughly, only few activations were expressed concurrently in both the pseudo-EEG and real-MEG eLORETA solutions, which were also persistent across the different frequency bands. Interestingly, their results did defer greatly from those in idealized simulations (Section 7 of SI), meaning in this realistic concurrency scenario its performance was severely affected by leakage.

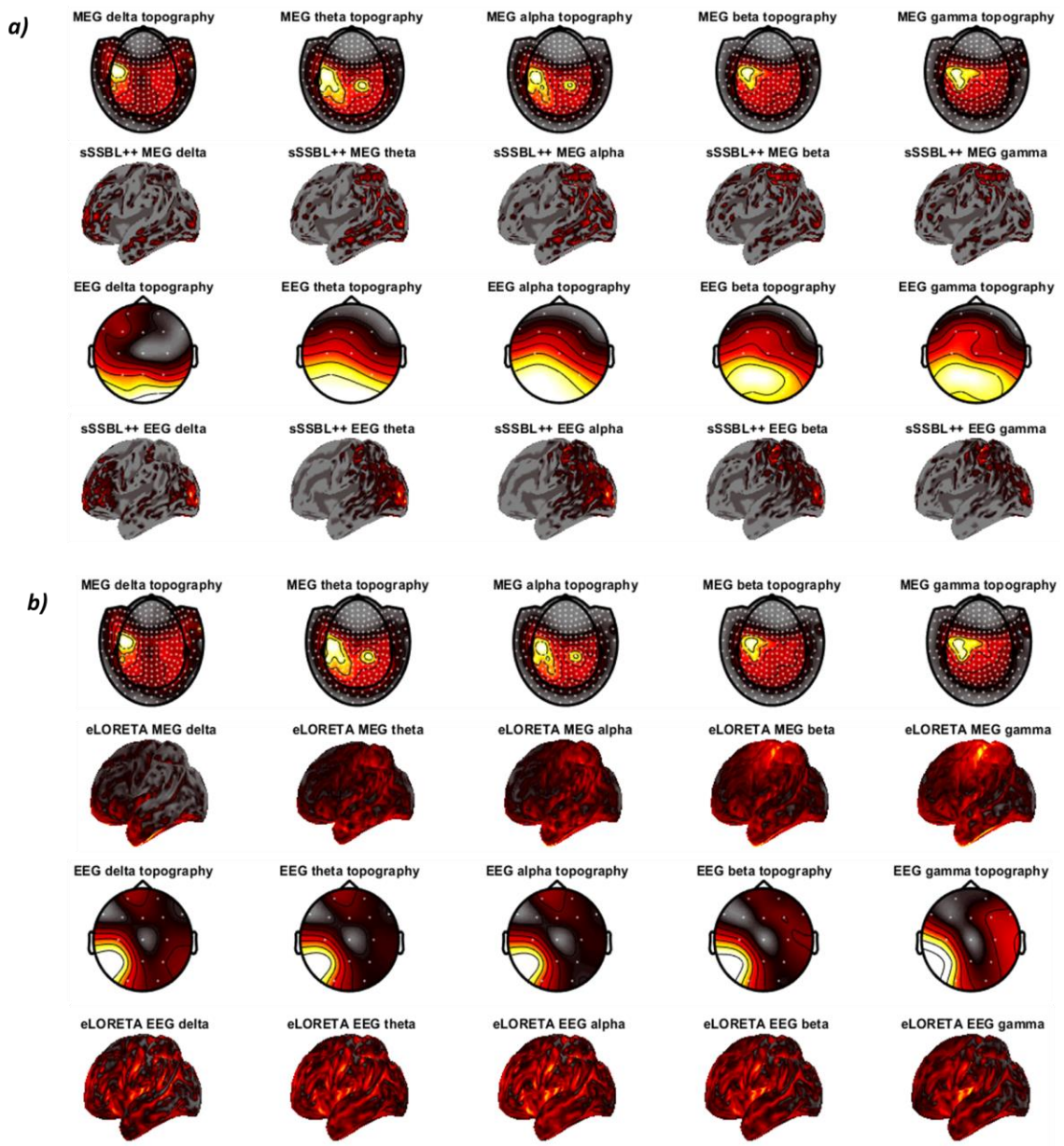


Figure 8*: Hot colormaps with results of the concurrency evaluation for (a) sSSBL++ and (b) eLORETA on 4 spectral bands (delta, theta, alpha and beta). The topographies at the sensor level represent the cross-spectrum slices for different spectral bands used for the ESI estimation with MEG (top) and EEG (bottom). The performance can be judged qualitative by comparing the source activity (cortical maps) estimated from the MEG (top) and its pseudo-EEG (bottom).*

A quantitative analysis of concurrency, through the surface based Earth Mover's Distance (EMD) metric [Paz-Linares et al., 2017], confirms this dramatic effect of leakage in eLORETA for all frequencies (Figure9). The EMD quantifies effort to draw leaked activity, produced in the estimation with the pseudo-EEG, back to the MEG landmark. In this situation we use EMD values rated by the ESI differences in the spectral domain, due to the lack of gold standard, in order to discriminate

concurrency from bias undergoing ESI estimation, see MM2. The evaluation was extended to the Linearly Constrained Minimum Variance (LCMV), the second generation ESI method selected by HCP for the study of brain connectome based electrophysiology [Van Essen et al 2013; Larson-Prior et al 2013], see MM-2 and SI-2. The leakage for the LCMV solution measured in terms of EMD reaches far beyond that of eLORETA.

The values of EMD for sSSBL++ were representative of the results in Figure 8, lower to those of eLORETA and therefor much lower in comparison to those of LCMV. This quantitative analysis was also extended to several instances of sSSBL++, adding on each of the proposed priors to confirm their possible effect on leakage in isolation. See the summary of these in Figures 5 and 6. Importantly, overall the performance of sSSBL++ was far more superior than that of any instance of sSSBL, only outperformed by the 2D field isotropy in some spectral bands. Also, overall all sSSBL instances performed better than eLORETA and LCMV only falling behind for the gamma band the spectral smoothness (sSSBL) and parcel smoothness (sSSBLparcelled), which performed very similar.

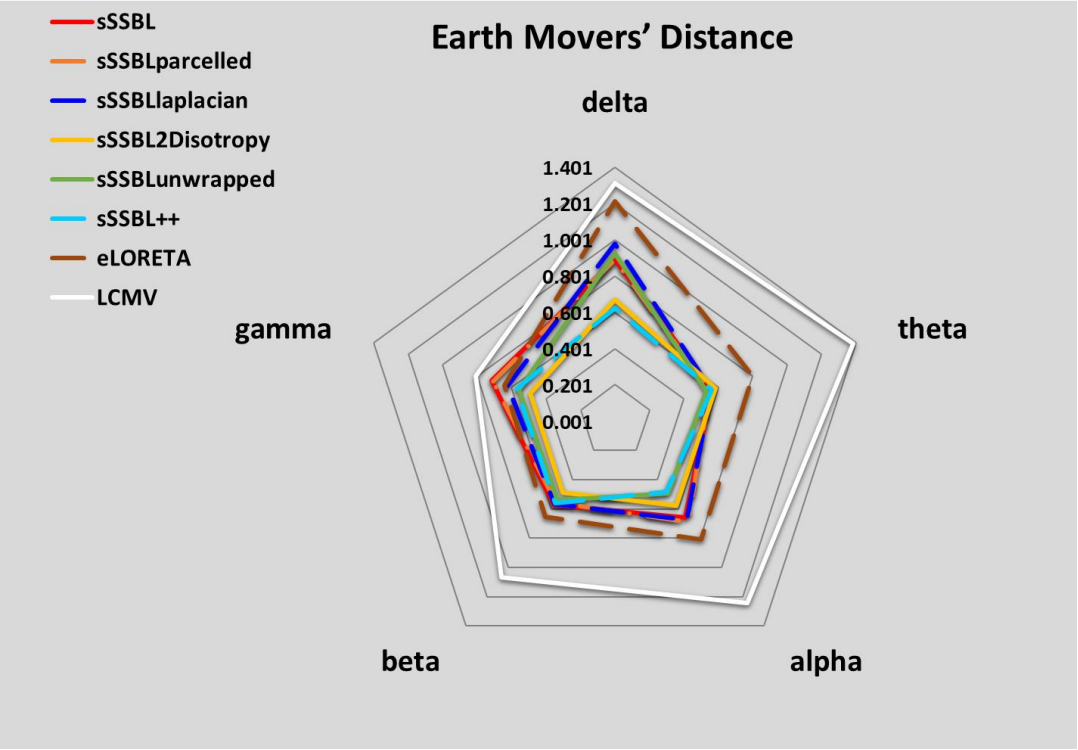


Figure 9*: Radar graph with values of Earth Movers' Distance (EMD) for sSSBL with several instances of priors and state-of-the-art-methods eLORETA and LCMV, in 5 spectral bands (delta, theta, alpha, beta, gamma). The priors introduced produce the following effects: spectral smoothness that we refer simply with sSSBL (Figure6b), parcel smoothness sSSBLparcelled (Figure6c), neighbors link sSSBLlaplacian (Figure5c), field rotational invariance sSSBL2Disotrpy (Figure5b), curvature compensation sSSBLunwrapped (Figure5d). These were added in isolation to evaluate their effect on leakage independently and altogether denominated sSSBL++. Here the EMD is quantifying the present amount of*

leakage produced by any method in the concurrency evaluation of Figure7, rated by the EMD for the deviation of these methods across the spectral bands, see MM-2.

These results suggest, first a universal lack of rigor in ESI validation procedures, which are limited to evaluate the performance only in idealized circumstances [Haufe and Ewald 2016], and second that the effect of leakage in real data for typical ESI methods might be much more severe than usually expected [Van de Steen et al 2019]. Therefore, it would be required for future efforts to consider a validation benchmark like the one proposed here or at least fair simulations of brain activity. Relevantly, in this scenario the sSSBL++ bottom-up approach allows curtailing of this effect considerably.

We acknowledge the concurrency test based on real-MEG/pseudo-EEG, although superior to previous validation methods, is lacking an actual landmark for brain activity. But for different reasons this is not readily accessible by means of noninvasive techniques, e.g. fMRI. Previous works on MEG/fMRI concurrency have provided maps with positively correlated features, but with no account for those expressed exclusively in either technique, which precisely explain their flaws manifested in form of leakage.

Confirmation of sSSBL++ with the MRI of brain lesions

We propose an alternative solution to discriminate on the effect of leakage in ESI methods based on the comparison of EEG sources with the MRI shine through of hemorrhagic brain lesions. This is possible due to the good definition of this type of lesion in the MRI image and their abnormal electrophysiological responses that could possibly be observed through EEG in any spectral band. In this scenario we define leakage as the responses estimated ESI on the affected areas that could be classified as normal. See details in MM-3.

The MRI and EEG collected belong to an infant born at 36 gestational weeks without any noticeable obstetric or clinical complication (Apgar 9/9, birth weight of 2,370 grams). However, at the age of 2 months an intracranial hemorrhage was detected, caused by hemorrhagic disease of the newborn (HDN). HDN is a common cause of the “acquired homeostatic disorder of newborn infants”, which can take place without the presence of any underlying disorder [Bör et al 2000]. The HDN diagnosis was confirmed through a T2 MRI study, see the three views at Figure10top. This shows clearly an extended subdural hematoma and right hemispheric stroke located in the area surrounding the middle and posterior cerebral artery.

EEG data was recorded a month later (3 months of age) in expectancy of a higher degree of maturity -in terms of electrophysiological brain age- in which not only delta spectral features but the faster

band theta may possibly be appreciated [Niedermeyer and da Silva 2005; André et al 2010]. Aiming to reconstruct the source of all neurophysiological activity, that may possibly be expressed in the EEG, we computed the sSSBL++ solution for a broad spectral band (0-7 Hz) unifying the delta and theta bands. See in Figure10middle the results in red color, where the red emphasis corresponds to activity classified as abnormal for the broad range of electrophysiological frequencies.

These are in high correspondence with the lesions mapped through the T2 MRI, provided all the abnormal areas detected by sSSBL++ were in the left hemisphere. Also, their distribution across different lobes and extension resemble that of the hemorrhage, which involved the occipital, temporal, tempo-parietal, frontal and fronto-parietal areas. Importantly, no perceptible normal activity infiltrating inwards to affected areas could be observed in the sSSBL++ solution. Rather, the areas detected as abnormal were extended far beyond the lesions.

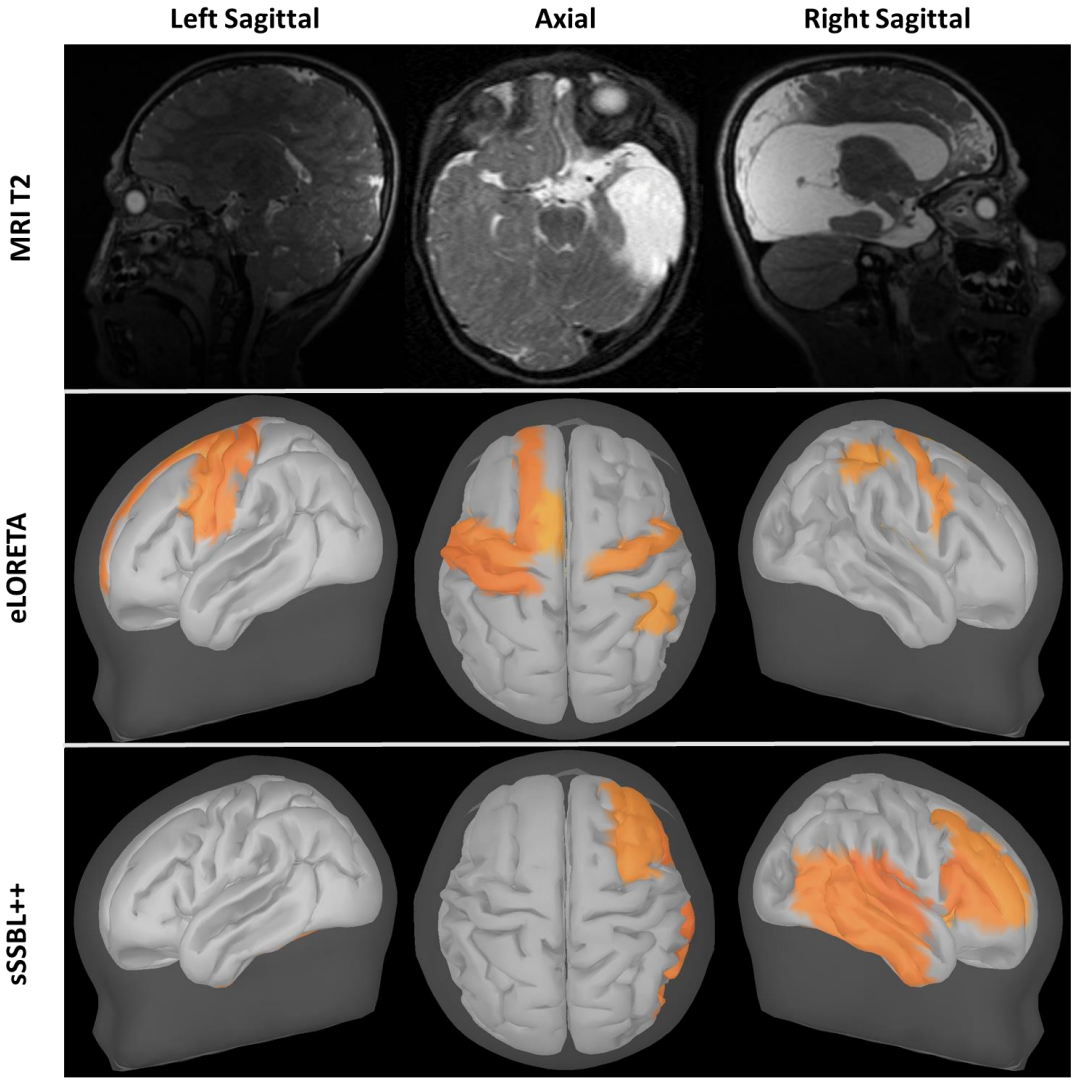


Figure 10*: Detection of abnormal source electrophysiological activity caused by hemorrhagic brain lesions that were observed in the MRI T2 of the subject (top). The results of ESI with eLORETA (middle) and sSSBL++ were obtained from the signal of a 19 channel EEG in the electrophysiological range of frequencies. The performance can be judged qualitative in terms of the undertone of the areas detected as abnormal by the methods with their MRI T2 shine-through.*

On the contrary the abnormal activity detected with eLORETA exhibited great differences with the MRI shine through of the lesions. Most of the abnormal activity was detected in the left hemisphere which cannot be justified to judge for the normal T2 contrast of this. Also, in the right hemisphere there was a wide mismatch between the shine through of lesions and the abnormal areas detected by eLORETA.

Notice that the ESI solutions were computed with an average head model and therefore in different space as the MRI. A T1 MRI -adequate structural space to define the individual head model- was not registered anticipating the typically deficient tissue contrast observed in T1 relaxation of newborns. An adequate quantification of the present amount of leakage would require projection of both the T2 MRI and EEG to a common structural space. Up to now we lacked a processing pipeline that could perform with acceptable precision the registration, segmentation and head model construction based on the T2 MRI of infants. This process, if performed by means of the standardized tools, would produce a roughly approximated head model, therefore requiring further developments.

Materials and Methods (MM)

MM-1 sSSBL algorithm

sSSBL pursues estimation of the cross-spectrum through a maximum "evidence" search via the Expectation-Maximization algorithm. Where evidence is defined as the conditional probabilities of two groups of parameters, given the available data samples or MEG/EEG cross-spectrum $\mathbf{S}_{\boldsymbol{vv}}(\nu)$: these are $\boldsymbol{\sigma_\iota}$, that controls the statistical relevance the source cross-spectral components $\mathbf{S}_{\boldsymbol{\iota\iota}}(\nu)$, and $\sigma_{\boldsymbol{\xi}}$, that controls the level of noise of the observations. This is done through an iterated scheme that produces an approximated representation of the evidence (expectation) followed by its maximization, that guarantees convergence to a local maximum. See in Figure 11 the algorithm for the expectation step illustrated in graphical form. In SI-5 and SI-6 we provide the derivation of the later, extended with all priors included in the sSSBL++.

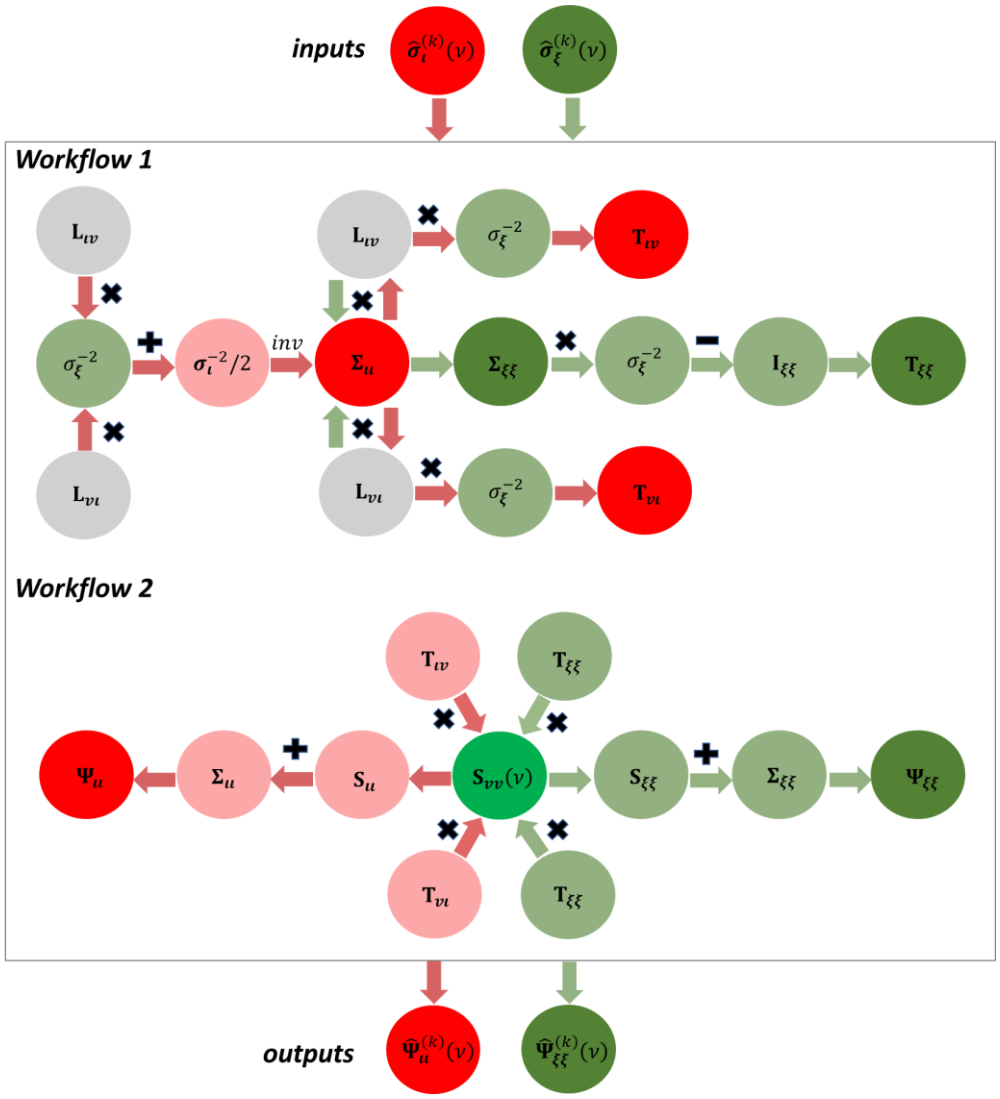


Figure 11*: Graphical representation of the sSSBL Expectation algorithm that precedes the Maximization. The red color identifies quantities (circles) and mathematical operations (arrows) associated to source activity, whereas the green identifies those for observation noise. The constant* $\boldsymbol{L}_{\iota v}$ *(Lead Field) and its transconjugated* $\boldsymbol{L}_{\iota v}$ *is identified with a gray circle. Workflow 1 uses the variances of noise* $\hat{\boldsymbol{\sigma}}_{\xi}^{(k)}$ *and activations* $\hat{\boldsymbol{\sigma}}_{\iota}^{(k)}$ *to produce two locally linear transfer operators: one asymmetric for the re-estimation of ESI solution (*$\boldsymbol{T}_{\iota v}$ *and its transconjugated* $\boldsymbol{T}_{v\iota}$*) and another one symmetric for the noise re-estimation* $\boldsymbol{T}_{\xi\xi}$*. Workflow 2 produces the effective empirical covariances of source activity* $\hat{\boldsymbol{\Psi}}_{u}^{(k)}$ *and noise* $\hat{\boldsymbol{\Psi}}_{\xi\xi}^{(k)}$*, used in the Maximization step.*

The maximization step is carried out via estimation formulas that we refer to here as the group Hierarchical Elastic Net, see Figure 12. This resembles the estimation formulas of the vector regression Elastic Net [Paz-Linares et al., 2017] and the Sparse Bayesian Learning [Wipf et al., 2009], but in this case through the arithmetic mean $\boldsymbol{\psi}_{\iota}$ of typical vector regression inputs corresponding to the samples. See the derivation of these formulas in SI-7 also extended to the sSSBL++. Although

not shown here, we perform control of the global sparsity level through estimating the parameters α and ρ, in completely analogous form to the procedure in [Paz-Linares et al., 2017], see SI-7.

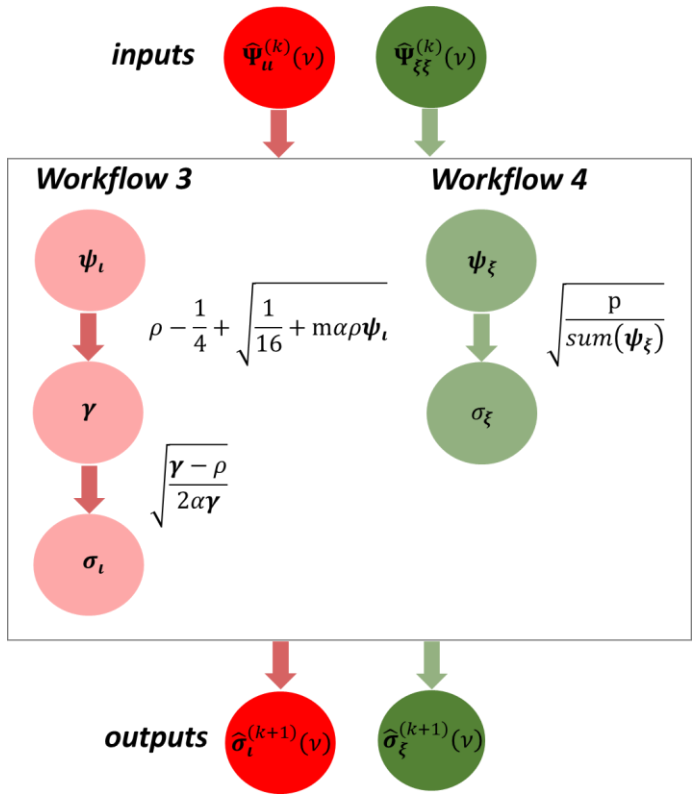


***Figure 12**: Graphical representation of the sSSBL Maximization algorithm, following the Expectation. The red color identifies quantities (circles) and mathematical operations (arrows) associated to source activity, whereas the green identifies those for observation noise. Workflow 3 uses the effective empirical covariance of the activations $\widehat{\boldsymbol{\Psi}}_{\boldsymbol{u}}^{(k)}$ to produce the source variance $\widehat{\boldsymbol{\sigma}}_{\iota}^{(k+1)}$. Workflow 4 uses the effective empirical covariance of the noise $\widehat{\boldsymbol{\Psi}}_{\xi\xi}^{(k)}$ to produce the noise variances $\widehat{\boldsymbol{\sigma}}_{\xi}^{(k+1)}$. The variances are used in the Expectation step of the next iteration.*

To help with the reproducibility we provide in SI-8 a standard pseudocode (nongraphical) and all formulas in a compact way. The code is freely available in GitHub as part of a more general toolbox for MEG/EEG source analysis (BC-VARETA) [Paz-Linares et al 2018; Gonzalez-Moreira et al 2018a; 2018b; 2018c].

MM-2 MEG/EEG concurrency study

The MEG corresponded to 246 channel preprocessed resting state data of the subject 175237 from the HCP database [Van Essen et al 2013; Larson-Prior et al 2013]. We construed the MEG head model following the Brainstorm pipeline [Tadel et al 2011], designed specifically to utilize HCP structural and functional data, and which illustrates the head model processing with the subject cited

above (175237). The cross-spectral tensor used in ESI (Figure3a3) was computed at a frequency resolution of about 0.5 Hz from 172 Fourier coefficients samples (trials), extracted through the Fourier transformation of the recordings spanning about 2 seconds at a sampling rate of 508Hz for each trial. The EEG head model corresponded to 19 channels in the 10-20 system, computed also following Brainstorm recommendations via Boundary Element Method upon 3 tissue layers (inner skull, outer skull, scalp) extracted with FSL [Smith et al 2004; Jenkinson et al 2012] and on 6K generators of a discretization of the midthickness.

We performed the concurrency evaluation, according to the illustration in Figure8, following three steps. First, the inversion of real MEG spectral data upon the MEG head model, via ESI with a given method. Second, the generation through the EEG head model of the pseudo EEG spectral data, corresponding to spectral activity determined in the first step. Third, the inversion of the pseudo EEG spectral data upon the EEG head model, via ESI with the same method as in the first step. The code and data to reproduce these results is freely available in GitHub (sSSBL-Concurrency).

One of the two methods proposed for comparison (eLORETA) belongs to the class of linear methods therefore not provided of any capacity towards bottom-up leakage control (see SI-2) [Pascual-Marqui et al., 2006]. The other method (LCMV) was a Beamformer that belongs to the second ESI generation, which can provide bottom-up control of leakage. But it does partially, due to its lack of a specific spectral model. Nevertheless, we leverage a more efficient implementation provided in fieldtrip that provides eLORETA the capacity to perform the direct computation of the source cross-spectrum via a linear transfer function and with regularization parameters that can be adjusted with complex variable cross-validation [Oostenveld et al. 2011].

To judge the quality of methods or leakage in the concurrency evaluation we used an implementation of Earth Movers' Distance on surface [Paz-Linares et al., 2017]. In this case their values for the difference between the MEG and EEG solutions in the concurrency test were rated by their values for the difference between delta and alpha bands of the MEG. This turns the validation metrics sensitive to detecting possible bias of the methods, towards certain cortical areas, as with the case of eLORETA and LCMV. Bias makes it difficult to judge, in the absence of the "gold standard" as is the case of our concurrency evaluation, whether the present similarities can be explained with good quality or with low sensitivity to changes in the spectral domain.

MM-3 MRI/EEG confirmatory study

Acquisitions like the EEG and T2 MRIs used in the confirmatory study here are part of the routine examination following the admission of cases in the neurotherapeutic area of the Neurodevelopment

Research Unit located in Queretaro campus of UNAM. The data used for the confirmatory study will be available upon request of the readers.

EEG was recorded in resting state condition with eyes closed on 19 channels defined in the 10-20 system using a Medicid™ IV (*Neuronic Mexicana, S.A.; Mexico*). The contact impedance for all electrodes was tuned down to a level of 5 kΩ or below and the bandwidth of the amplifiers was set within the limits of 0.3 Hz to 30 Hz. The preprocessing was performed off-line and through manual artifact rejection guided by expert hands on the standards of the International Federation of Clinical Neurophysiology (IFCN). The preprocessing yielded an amount of 16 artifact-free trails each spanning over 2.56 seconds. To construct the head model for ESI we used the Brainstorm default anatomy ICBM152, via Boundary Element Method upon 3 tissue layers (inner skull, outer skull, scalp) and on 6K generators of a discretization of the midthickness.

Structural T2 MRIs were acquired using a 3T scanner (General Electric Healthcare, Milwaukee, Wisconsin, US) with 16-channels of a head neurovascular coil (HDNV). The infant was awake, laying back and wearing earplugs for protection against the MR room noises. Figure11c shows structural images acquired with two different pulse sequences: A T2-weigthed spin echo sequence (Left and Right), TR/TE 2500/68 ms, flip angle 90°, slices 196, slice thickness 1 mm, matrix 224 x 224, FoV 220 x 220 mm2, voxel sizes of 0.8 x 1.0 x 0.8 mm3. A T2-weigthed turbo spin echo sequence (Center), TR/TE 7000/100 ms, flip angle 111°, slices 70, slice thickness 2 mm, matrix 448 x 352, FoV 220 x 220 mm2, voxel sizes of 0.4 x 0.4 x 2.0 mm3.

The classification of abnormal activity was carried out in statistical comparison with norms for EEG source activity in infants of the same range of age [Otero et al 2011]. For each method (eLORETA and sSSBL) the norm of source activity was computed using the EEG of 179 newborns collected on the basis discussed at the beginning of this section. Then, the estimated activity for the pathological patient for every method was compared to the mean and standard deviation of its norm. The criteria for the selection of abnormal areas was according to the difference, across all frequencies, between the norm mean and the estimated activity for the patient. The threshold used for classification was twice the standard deviation of the norm.

Conclusions

We demonstrated that the impact of leakage in common Electrophysiological Source Imaging methods can be much more severe than expected. Shedding light on the actual extent of leakage required completelly new validation procecedures slanted towards a landmark based on the real data

rather than in simulations of Brian electrical activity. This was furnished here via two different instances, that according to our knowledge are lacking of any precedent and may else serve to future efforts: First, concurrency of the ESI solution in the low-density pseudo-EEG of a high-denstity MEG. Second, comparison against the MRI shine through of brain lesions which turn the affected areas into a source of abnormal electrophysiological activity.

We proposed ESI via direct sparse Bayesian learning of spectral responses -coined spectral Structured Sparse Bayesian Learning (sSSBL) and its extension with multiple priors (sSSBL++). This procedure is hitched to technical difficulties for the statistical learning of complex valued data models, that we have solved here. The results with sSSBL++ display far more control of leakage benchmarked in our validation framework against commonly used methods eLORETA and LCMV. This strongly supports our working hypothesis of bottom-up control on leakage, where a possible suggested route is through using ESI priors and learning process directly to the ultimate estimators targeted by leakage.

Acknowledgments

This study was supported by Grant 61673090 from the National Nature Science Foundation of China, and Grants DGAPA-IN200917, CONACYT 4971 and CONACYT 251309 from the National Council on Science and Technology of Mexico.

Contributions

EGM and DPL designed the method and wrote the paper. AAG, YW and ML performed validation and implementation of software platforms. JBB guided MEG/EEG analysis and confirmation of the method. TH senior author who designed the experimental confirmation of the method and collected the data. PAVS senior author who introduced the theoretical background.

References

[1] André, Monique, M-D. Lamblin, Anne-Marie d'Allest, Lilia Curzi-Dascalova, F. T. S. N. T. Moussalli-Salefranque, S. Nguyen The Tich, M-F. Vecchierini-Blineau, F. Wallois, E. Walls-Esquivel, and P. Plouin. "Electroencephalography in premature and full-term infants. Developmental features and glossary." *Neurophysiologie clinique/Clinical neurophysiology* 40, no. 2 (2010): 59-124.

[2] Andrews, David F., and Colin L. Mallows. "Scale mixtures of normal distributions." *Journal of the Royal Statistical Society: Series B (*Methodological) 36.1 (1974): 99-102.

[3] Artemenko, Christina, et al. "The neural correlates of mental arithmetic in adolescents: a longitudinal fNIRS study." *Behavioral and Brain Functions* 14.1 (2018): 5.

[4] Attias, Hagai. "A variational baysian framework for graphical models." *Advances in neural information processing systems*. (2000).

[5] Avena-Koenigsberger, Andrea, Bratislav Misic, and Olaf Sporns. "Communication dynamics in complex brain networks*." Nature Reviews Neuroscience* 19.1 (2018): 17.

[6] Babiloni, Fabio, et al. "Spatial enhancement of EEG data by surface Laplacian estimation: the use of magnetic resonance imaging-based head models." (2001): 724-727.

[7] Baillet, Sylvain, and Line Garnero. "A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem." *IEEE transactions on Biomedical Engineering* 44.5 (1997): 374-385.

[8] Baillet, Sylvain. "Electromagnetic brain mapping using MEG and EEG." The Oxford handbook of social neuroscience. *Oxford Library of Psychology*, (2011): 97.

[9] Barry, Robert J., Adam R. Clarke, and Stuart J. Johnstone. "A review of electrophysiology in attention-deficit/hyperactivity disorder: I. Qualitative and quantitative electroencephalography." *Clinical neurophysiology* 114.2 (2003): 171-183.

[10] Barry, Robert J., Adam R. Clarke, Stuart J. Johnstone, Christopher A. Magee, and Jacqueline A. Rushby. "EEG differences between eyes-closed and eyes-open resting conditions." *Clinical Neurophysiology* 118, no. 12 (2007): 2765-2773.

[11] Becerra, Judith, et al. "Neurofeedback in healthy elderly human subjects with electroencephalographic risk for cognitive disorder." *Journal of Alzheimer's Disease* 28.2 (2012): 357-367.

[12] Belardinelli, Paolo, et al. "Source reconstruction accuracy of MEG and EEG Bayesian inversion approaches." *PloS one* 7.12 (2012): e51985.

[13] Blinowska, Katarzyna J., Rafał Kuś, and Maciej Kamiński. "Granger causality and information flow in multivariate processes." *Physical Review E* 70.5 (2004): 050902.

[14] Blinowska, Katarzyna J. "Review of the methods of determination of directed connectivity from multichannel data." *Medical & biological engineering & computing* 49.5 (2011): 521-529.

[15] Bör, Özcan, Nejat Akgün, Ayten Yakut, Fatih Sarhuş, and Suna Köse. "Late hemorrhagic disease of the newborn." *Pediatrics International* 42, no. 1 (2000): 64-66.

[16] Bosch-Bayard, J., et al. "3D statistical parametric mapping of EEG source spectra by means of variable resolution electromagnetic tomography (VARETA)." *Clinical Electroencephalography* 32.2 (2001): 47-61.

[17] Budzynski, Thomas H., et al., eds. Introduction to quantitative EEG and neurofeedback: Advanced theory and applications. *Academic Press*, (2009).

[18] Burle, Boris, et al. "Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view." *International Journal of Psychophysiology* 97.3 (2015): 210-220.

[19] Butterworth, Brian, Sashank Varma, and Diana Laurillard. "Dyscalculia: from brain to education." *Science* 332.6033 (2011): 1049-1053.

[20] Chabot, Robert J., Flavia di Michele, and Leslie Prichep. "The role of quantitative electroencephalography in child and adolescent psychiatric disorders." *Child and Adolescent Psychiatric Clinics* 14.1 (2005): 21-53.

[21] da Silva, Fernando Lopes. "EEG and MEG: relevance to neuroscience." *Neuron* 80.5 (2013): 1112-1128.

[22] Dehaene, Stanislas, et al. "Three parietal circuits for number processing." *Cognitive neuropsychology* 20.3-6 (2003): 487-506.

[23] Dehaene, Stanislas, et al. "Arithmetic and the brain." *Current opinion in neurobiology* 14.2 (2004): 218-224.

[24] Desikan, Rahul S., et al. "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest." *Neuroimage* 31.3 (2006): 968-980.

[25] Dupuy, Franca E., et al. "Sex differences between the combined and inattentive types of attention-deficit/hyperactivity disorder: an EEG perspective." *International Journal of Psychophysiology* 89.3 (2013): 320-327.

[26] Fernández, T, et al. "Sources of EEG activity in learning disabled children." *Clinical Electroencephalography* 33.4 (2002): 160-164.

[27] Fernandez, T., et al. "EEG and behavioral changes following neurofeedback treatment in learning disabled children." *Clinical Electroencephalography* 34.3 (2003): 145-152.

[28] Freeman, Walter J. "Mass action i78u 4gn the nervous system" Vol. 2004. *Academic Press*, New York, (1975).

[29] Freeman, Walter J., et al. "Fine spatiotemporal structure of phase in human intracranial EEG." *Clinical Neurophysiology* 117.6 (2006): 1228-1243.

[30] Friston, Karl, et al. "Variational free energy and the Laplace approximation." *Neuroimage* 34.1 (2007): 220-234.

[31] Friston, Karl, et al. "Multiple sparse priors for the M/EEG inverse problem." *NeuroImage* 39.3 (2008): 1104-1120.

[32] Geary, David C. "Mathematical disabilities: Reflections on cognitive, neuropsychological, and genetic components." *Learning and individual differences* 20.2 (2010): 130-133.

[33] Gonzalez-Moreira, Eduardo, et al. "Populational Super-Resolution Sparse M/EEG Sources and Connectivity Estimation." *bioRxiv* (2018a): 346569.

[34] Gonzalez-Moreira, Eduardo, et al. "Caulking the Leakage Effect in MEEG Source Connectivity Analysis." *arXiv preprint arXiv*:1810.00786 (2018b).

[35] Gonzalez-Moreira, Eduardo, et al. "Third generation MEEG source connectivity analysis toolbox (BC-VARETA 1.0) and validation benchmark." *arXiv preprint arXiv*:1810.11212 (2018c).

[36] Grech, Roberta, et al. "Review on solving the inverse problem in EEG source analysis." *Journal of neuroengineering and rehabilitation* 5.1 (2008): 25

[37] Grefkes, Christian, and Gereon R. Fink. "The functional organization of the intraparietal sulcus in humans and monkeys." *Journal of anatomy* 207.1 (2005): 3-17.

[38] Groetsch, Charles W., and C. W. Groetsch. Inverse problems in the mathematical sciences. Vol. 52. *Braunschweig*: Vieweg, (1993).

[39] Hämäläinen, Matti S., and Risto J. Ilmoniemi. "Interpreting magnetic fields of the brain: minimum norm estimates." *Medical & biological engineering & computing* 32.1 (1994): 35-42.

[40] Harmony, Thalia, et al. "Correlation between EEG spectral parameters and an educational evaluation." *International Journal of Neuroscience* 54.1-2 (1990): 147-155.

[41] Harville, David A. "Maximum likelihood approaches to variance component estimation and to related problems." *Journal of the American statistical association* 72.358 (1977): 320-338.

[42] Haufe, Stefan, et al. "Combining sparsity and rotational invariance in EEG/MEG source reconstruction." *NeuroImage* 42.2 (2008): 726-738.

[43] Haufe, Stefan, et al. "A critical assessment of connectivity measures for EEG data: a simulation study." *Neuroimage* 64 (2013): 120-133.

[44] Haufe, Stefan, and Arne Ewald. "A simulation framework for benchmarking EEG-based brain connectivity estimation methodologies." *Brain topography* (2016): 1-18.

[45] He, Bin, et al. "Electrophysiological Brain Connectivity: Theory and Implementation." *IEEE Transactions on Biomedical Engineering* (2019).

[46] Jenkinson, Mark, Christian F. Beckmann, Timothy EJ Behrens, Mark W. Woolrich, and Stephen M. Smith. "Fsl." *Neuroimage* 62, no. 2 (2012): 782-790.

[47] John, E. R., et al. "Neurometric evaluation of cognitive dysfunctions and neurological disorders in children." *Progress in Neurobiology* 21.4 (1983): 239-290.

[48] Kaye, Herbert, et al. "Neurometric evaluation of learning disabled children." *International Journal of Neuroscience* 13.1 (1981): 15-25.

[49] Krishnaswamy, Pavitra, et al. "Sparsity enables estimation of both subcortical and cortical activity from MEG and EEG." *Proceedings of the National Academy of Sciences* 114.48 (2017): E10465-E10474.

[50] Kus, Rafal, Maciej Kaminski, and Katarzyna J. Blinowska. "Determination of EEG activity propagation: pair-wise versus multichannel estimate." *IEEE transactions on Biomedical Engineering* 51.9 (2004): 1501-1510.

[51] Landerl, Karin, Anna Bevan, and Brian Butterworth. "Developmental dyscalculia and basic numerical capacities: A study of 8–9-year-old students." *Cognition* 93.2 (2004): 99-125.

[52] Larson-Prior, Linda J., Robert Oostenveld, Stefania Della Penna, G. Michalareas, F. Prior, Abbas Babajani-Feremi, J-M. Schoffelen et al. "Adding dynamics to the Human Connectome Project with MEG." *Neuroimage* 80 (2013): 190-201.

[53] Marinazzo, Daniele, et al. "Controversies in EEG Source Imaging and Connectivity: Modeling, Validation, Benchmarking." (2019): 1-3.

[54] Matejko, Anna A., and Daniel Ansari. "How do individual differences in children's domain specific and domain general abilities relate to brain activity within the intraparietal sulcus during arithmetic? An fMRI study." *Human brain mapping* 38.8 (2017): 3941-3956.

[55] Mattout, Jérémie, et al. "MEG source localization under multiple constraints: an extended Bayesian framework." *NeuroImage* 30.3 (2006): 753-767.

[56] Monastra, Vincent J. "Quantitative electroencephalography and attention-deficit/hyperactivity disorder: implications for clinical practice." Current psychiatry reports 10.5 (2008): 432-438.

[57] Niedermeyer, Ernst, and FH Lopes da Silva, eds. Electroencephalography: basic principles, clinical applications, and related fields. *Lippincott Williams & Wilkins*, (2005).

[58] Nunez, Paul L., and Ramesh Srinivasan. Electric fields of the brain: the neurophysics of EEG. Oxford University Press, USA, (2006).

[59] Nunez, P. L., et al. "A theoretical and experimental study of high resolution EEG based on surface Laplacians and cortical imaging." Electroencephalography and clinical neurophysiology 90.1 (1994): 40-57.

[60] Ogrim, Geir, Juri Kropotov, and Knut Hestad. "The quantitative EEG theta/beta ratio in attention deficit/hyperactivity disorder and normal controls: sensitivity, specificity, and behavioral correlates." Psychiatry research 198.3 (2012): 482-488.

[61] Otero, G. A., T. Harmony, F. B. Pliego-Rivero, J. Ricardo-Garcell, J. Bosch-Bayard, R. Porcayo-Mercado, A. Fernandez-Bouzas et al. "QEEG norms for the first year of life." *Early human development* 87, no. 10 (2011): 691-703.

[62] Pascual-Marqui, Roberto Domingo. "Review of methods for solving the EEG inverse problem." International journal of bioelectromagnetism 1.1 (1999): 75-86.

[63] Pascual-Marqui, Roberto D., Christoph M. Michel, and Dietrich Lehmann. "Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain." International Journal of psychophysiology 18.1 (1994): 49-65.

[64] Pascual-Marqui, Roberto Domingo. "Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details." Methods Find Exp Clin Pharmacol 24. Suppl D (2002): 5-12.

[65] Pascual-Marqui, Roberto D., et al. "Exact low resolution brain electromagnetic tomography (eLORETA)." Neuroimage 31. Suppl 1 (2006).

[66] Pascual-Marqui, Roberto D. "Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization." arXiv preprint arXiv:0710.3341 (2007).

[67] Pascual-Marqui, Roberto D., Dietrich Lehmann, Martha Koukkou, Kieko Kochi, Peter Anderer, Bernd Saletu, Hideaki Tanaka et al. "Assessing interactions in the brain with exact low-resolution electromagnetic tomography." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 369, no. 1952 (2011): 3768-3784.

[68] Pascual-Marqui, R. D., R. J. Biscay, J. Bosch-Bayard, D. Lehmann, K. Kochi, N. Yamada, T. Kinoshita, and N. Sadato. "Isolated effective coherence (iCoh): causal information flow excluding indirect paths." *arXiv preprint arXiv*:1402.4887 (2014).

[69] Patterson, H. Desmond, and Robin Thompson. "Recovery of inter-block information when block sizes are unequal." *Biometrika* 58.3 (1971): 545-554.

[70] Paz-Linares, Deirel, et al. "Spatio temporal EEG source imaging with the hierarchical bayesian elastic net and elitist lasso models." *Frontiers in neuroscience* 11 (2017): 635.

[71] Paz-Linares, Deirel, et al. "Neural Connectivity with Hidden Gaussian Graphical State-Model." *arXiv preprint arXiv*: 1810.01174 (2018).

[72] Larson-Prior, L.J., Oostenveld, R., Della Penna, S., Michalareas, G., Prior, F., Babajani-Feremi, A., Schoffelen, J.M., Marzetti, L., de Pasquale, F., Di Pompeo, F. and Stout, J., 2013. Adding dynamics to the Human Connectome Project with MEG. Neuroimage, 80, pp.190-201.

[73] Robert Oostenveld, Pascal Fries, Eric Maris, and Jan-Mathijs Schoffelen. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Computational Intelligence and Neuroscience, vol. 2011, Article ID 156869, 9 pages, 2011. doi:10.1155/2011/156869.

[74] Roca-Stappung, Milene, et al. "Electroencephalographic characterization of subgroups of children with learning disorders." PloS one 12.7 (2017): e0179556.

[75] Rubner, Yossi, Carlo Tomasi, and Leonidas J. Guibas. "A metric for distributions with applications to image databases." *Sixth International Conference on Computer Vision* (IEEE Cat. No. 98CH36271). IEEE, (1998).

[76] Rubner, Yossi, Carlo Tomasi, and Leonidas J. Guibas. "The earth mover's distance as a metric for image retrieval." *International journal of computer vision* 40.2 (2000): 99-121.

[77] Sakkalis, Vangelis. "Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG." *Computers in biology and medicine* 41.12 (2011): 1110-1117.

[78] Schomer, Donald L., and Fernando Lopes Da Silva. "Niedermeyer's electroencephalography: basic principles, clinical applications, and related fields. " *Lippincott Williams & Wilkins*, (2012).

[79] Smith, Stephen M., Mark Jenkinson, Mark W. Woolrich, Christian F. Beckmann, Timothy EJ Behrens, Heidi Johansen-Berg, Peter R. Bannister et al. "Advances in functional and structural MR image analysis and implementation as FSL." *Neuroimage* 23 (2004): S208-S219.

[80] Snyder, Steven M., et al. "Blinded, multi-center validation of EEG and rating scales in identifying ADHD within a clinical sample." Psychiatry research 159.3 (2008): 346-358.

[81] Srinivasan, Ramesh. "Methods to improve the spatial resolution of EEG." *International Journal of Bioelectromagnetism* 1.1 (1999): 102-111.

[82] Stokes, Patrick A., and Patrick L. Purdon. "A study of problems encountered in Granger causality analysis from a neuroscience perspective." *Proceedings of the National Academy of Sciences* 114.34 (2017): E7063-E7072.

[83] Tadel, François, Sylvain Baillet, John C. Mosher, Dimitrios Pantazis, and Richard M. Leahy. "Brainstorm: a user-friendly application for MEG/EEG analysis." *Computational intelligence and neuroscience* 2011 (2011).

[84] Valdés, Pedro, et al. "High resolution spectral EEG norms for topography. Brain" *Brain topography* 3 (1990): 281-282.

[85] Valdés, Pedro, et al. "Frequency domain models of the EEG." *Brain topography* 4.4 (1992): 309-319.

[86] Valdes-Sosa, P., et al. "Variable resolution electric-magnetic tomography." *Biomag* 96. Springer, New York, NY, 2000. 373-376.

[87] Van Essen, David C., Stephen M. Smith, Deanna M. Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, and Wu-Minn HCP Consortium. "The WU-Minn human connectome project: an overview." *Neuroimage* 80 (2013): 62-79.

[88] Van de Steen, Frederik, et al. "Critical comments on EEG sensor space dynamical connectivity analysis." *Brain topography* 32.4 (2019): 643-654.

[89] Van Veen et al. "Localization of brain electrical activity via linearly constrained minimum variance spatial filtering." IEEE Transactions on biomedical engineering 44, no. 9 (1997): 867-880.

[90] Vega-Hernández, Mayrim, et al. "Penalized least squares methods for solving the EEG inverse problem." *StatisticaSinica* (2008): 1535-1551.

[91] Vidaurre, Diego, et al. "Spontaneous cortical activity transiently organises into frequency specific phase-coupling networks." *Nature communications* 9.1 (2018): 2987.

[92] Wang et al. "EECoG-Comp: An Open Source Platform for Concurrent EEG/ECoG Comparisons—Applications to Connectivity Studies." Brain topography 32, no. 4 (2019): 550-568.

[93] Wipf, David, and Srikantan Nagarajan. "A unified Bayesian framework for MEG/EEG source imaging." *NeuroImage* 44.3 (2009): 947-966.

Supporting Information (SI)

SI-1 Mathematical notation

$\mathbb{E}$	**Sensor space for EEG/MEG.**
$\mathbb{G}$	Generator space from a discretization of the Gray Matter.
$\mathbb{F}$	Frequency domain for EEG/MEG signals.
$\mathbb{M}$	Random sample space of Fourier coefficients from EEG/MEG signals.
$\mathbb{A}$	Pacellation space from a given functional or structural atlas of the Gray Matter.
p	Number of sensors $\mathrm{p} = \lvert\mathbb{E}\rvert$.
q	Number of generators $\mathrm{q} = \lvert\mathbb{G}\rvert$.
f	Number of frequency components $\mathrm{f} = \lvert\mathbb{F}\rvert$.
m	Number of random samples $\mathrm{m} = \lvert\mathbb{M}\rvert$.
$\boldsymbol{l}$	Number of parcells $l = \lvert\mathbb{A}\rvert$.
$\boldsymbol{e} \in \mathbb{E}$	Indices for sensors $e = 1 \ldots \mathrm{p}$.
$\boldsymbol{g} \in \mathbb{G}$	Indices for generators $g = 1 \ldots \mathrm{q}$.
$\boldsymbol{\nu} \in \mathbb{F}$	Indices for frequencies $\nu = 1 \ldots \mathrm{f}$.
$\boldsymbol{m} \in \mathbb{M}$	Indices for samples $m = 1 \ldots \mathrm{m}$.
$\boldsymbol{a} \in \mathbb{A}$	Indices for parcels $a = 1 \ldots l$.
$\boldsymbol{v_m}(\boldsymbol{\nu})$	Sampled complex vector of the Fourier coefficients for EEG/MEG data coreresponding to a frequency component (observed variables or *data*).
$\boldsymbol{\iota_m}(\boldsymbol{\nu})$	Sampled complex vector of the Fourier coefficients for EEG/MEG source activity coreresponding to a frequency component (hidden variables or spectral activations).
$\mathbf{L}_{\boldsymbol{\nu\iota}}$	Lead Field matrix of $\mathrm{p} \times \mathrm{q}$ size.

$\boldsymbol{\xi_m(\nu)}$	Sampled complex vector of the Fourier coefficients for EEG/MEG noise coreresponding to a frequency component (observation noise process or spectral noise).
$\boldsymbol{\Sigma_{\iota\iota}}$	Hermitian and positive semidefinite Posterior Covariance matrix of spectral EEG/MEG source activity.
$\boldsymbol{\Sigma_{\xi\xi}}$	Hermitian and positive semidefinite Posterior Covariance matrix of spectral EEG/MEG noise.
$\boldsymbol{\sigma_\iota}$	Variances of spectral EEG/MEG source activity.
$\boldsymbol{\sigma_\xi}$	Variances of spectral EEG/MEG noise.
$\boldsymbol{\lambda}$	Regularization parameters or tuning hyperparameters vector of the general penalty function.
$\mathbf{T}_{\iota\nu}^{(k)}$	EEG/MEG assymetric Data to Source Transfer operator.
$\mathbf{T}_{\xi\xi}^{(k)}$	EEG/MEG Hermitian Data to noise Transfer operator.
$\mathbf{S}_{\iota\iota}^{(k)}$	Hermitian and positive semidefinite empirical Covariance matrix of spectral EEG/MEG source activity.
$\mathbf{S}_{\xi\xi}^{(k)}$	Hermitian and positive semidefinite empirical Covariance matrix of spectral EEG/MEG source activity.
$\boldsymbol{\Psi}_{\iota\iota}^{(k)}$	Hermitian and positive semidefinite effective empirical Covariance matrix of spectral EEG/MEG source activity.
$\boldsymbol{\Psi}_{\xi\xi}^{(k)}$	Hermitian and positive semidefinite effective empirical Covariance matrix of spectral EEG/MEG source activity.
$\mathbf{S}_{\nu\nu}$	Complex Hermitian matrix EEG/MEG data Fourier coefficients Covariance matrix.

SI-2 ESI concepts and generations

The derived inverse problem of ESI contains three main elements: (i) EEG/MEG data observed in p-sensors; (ii) anatomically constrained EEG/MEG space of q-generators from a brain tissue segmentation of cortical mid-thickness surface or volume; (iii) px3q Lead Field matrix (an electromagnetic vector field weighting the contribution q generators) given a volume conductor

model that entails nonbrain tissue segmentation, sensor positioning and nature of observed data (electric or magnetic fields) [Grech et al 2008].

Estimation of the neural causes can be done by inversion of the EEG forward model [Hämäläinen et al 1994; Pascual-Marqui et al 1994; Pascual-Marqui 1999]. In this case we propose a space-frequency analysis by directly inverting its spectral equivalent in virtue of the linearity of the Discrete Fourier Transform (DFT):

$$\boldsymbol{v}_m(\nu) = \mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\boldsymbol{\iota}_m(\nu) + \boldsymbol{\xi}_m(\nu) \qquad [2.1]$$

where ν represents frequencies in the spectral domain, m time segments used to compute the DFT, $\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}$ is the $\mathrm{p} \times 3\mathrm{q}$ Lead Field matrix. The spectral vectors in the equation represent the EEG $\boldsymbol{v}_m(\nu)$ of size $\mathrm{p} \times 1$, source activation $\boldsymbol{\iota}_m(\nu)$ of size $3\mathrm{q} \times 1$ and a sensor noise process $\boldsymbol{\xi}_m(\nu)$ of size $\mathrm{p} \times 1$

Usually the vector fields ($\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}(i,:)$ and $\boldsymbol{\iota}$) are projected by the scalar product with another vector field $\boldsymbol{d}$ (structural not electromagnetic) of Gray Matter tissue directions yielding only q-p degeneracy ($\langle \mathbf{L}'_{\boldsymbol{v}\boldsymbol{\iota}}(i,:), \boldsymbol{d}\rangle$ and $\langle \boldsymbol{\iota}, \boldsymbol{d}\rangle$).

ESI methods target solving the 3q-p degeneracy EEG/MEG inverse problem kernel (space defining all possible solutions that could explain the recordings) in analogy to the general inverse problem [Groetsch et al 1993]. Reducing the degeneracy beyond further anatomical constraints (Gray Matter ROIs or tissue directionality) is only achievable through regularization or inclusion of some mathematical a priori information [Hämäläinen et al 1994, Baillet et al 1997, Friston et al 2008], which produces a new system with stable unique solution.

The pitfall of leakage has been followed by the deployment of several "generations" of ESI methods within just two decades [Gonzalez-Moreira et al 2018b]. The term "generation" coins basic differences of the statistical learning method towards the construction of the EEG/MEG source transfer operator -via Linear, Nonlinear Univariate, Nonlinear Multivariate models- and the inclusion of prior information in order to ameliorate Leakage. See in Table1 a deeper reconnoitering of such models.

Table 1: Family of Linear/Nonlinear univariate/Nonlinear multivariate ESI methods

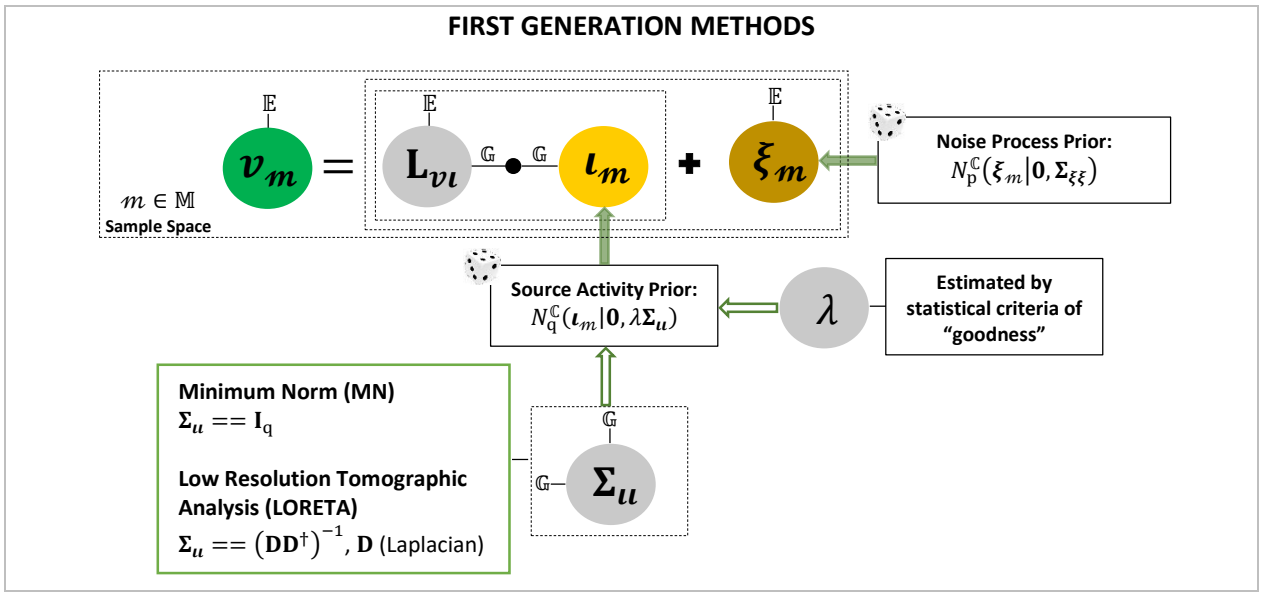
	Parameters Covariance	**Data Conditional Covariance**	**Algorithm**	**Linear/ Nonlinear/ Nonlinear Multivariate**
Minimum Norm Estimator (MME) (Hämäläinen and Ilmoniemi, 1994)	Constrained to Scaled Identity Matrix. Fixed value.	Constrained to Scaled Identity	Explicit ***Parameters*** Posterior Mean estimator.	Linear

	Connectivity-Postprocessing.	Matrix. Fixed value.		
Low Resolution Electromagnetic Tomography (LORETA) (Pascual-Marqui et al., 1994).	Fixed Laplacian Operator. Connectivity-Postprocessing.	Full Matrix. Prior Free.	Iterated Explicit ***Parameters*** Posterior Mean estimator and empirical formulas of the Conditional ***Data*** Covariance.	Linear
Exact Low Resolution Electromagnetic Tomography (eLORETA) (Pascual-Marqui et al., 2006).	Constrained to a Diagonal Matrix. Prior Free. Connectivity-Postprocessing.	Full Matrix. Prior Free.	Iterated Explicit ***Parameters*** Posterior Mean estimator and empirical formulas of the Conditional ***Data*** Covariance and ***Parameters*** Variances. Provides Zero Localization Error in case of a single Source reconstruction.	Nonlinear
Standardized Low Resolution Electromagnetic Tomography (sLORETA/eLORETA) (Pascual-Marqui, 2002).	Full Matrix (Connectivity). Prior Free.	Full Matrix. Prior Free.	Iterated Explicit ***Parameters*** Posterior Mean estimator and empirical formulas of the Conditional ***Data*** Covariance and ***Parameters*** Covariances.	Nonlinear
Variable Resolution Tomographic Analysis (VARETA) (Valdes-Sosa, 1996, Bosch-Bayard, et al., 2001).	Full Matrix (Connectivity). Prior Free.	Constrained to Scaled Identity Matrix. Prior free.	Iterated Explicit ***Parameters*** Posterior Mean estimator and Expectation Maximization (EM) formulas of the Conditional ***Data*** Variance and ***Parameters*** Covariance.	Nonlinear Multivariate
Automatic Relevance Determination (ARD) (Neal, 1998; Tipping, 2001; Sato et al., 2004; Wipf et al., 2006; Wipf and Rao, 2007; Daunizeau and Friston, 2007).	Constrained to a Diagonal Matrix. Jeffrey Improper Priors. Connectivity-Postprocessing.	Constrained to Scaled Identity Matrix. Prior free.	Iterated Explicit ***Parameters*** Posterior Mean estimator and Empirical Bayes (EB) formulas of the Conditional ***Data*** Variance and ***Parameters*** Variances. Induces Sparsity by pruning the ***Parameters*** Variances estimates.	Nonlinear
Restricted Maximum Likelihood (ReLM) (Patterson and Thompson, 1971, Harville, 1977; Friston et al., 2007; Wipf et al., 2009; Belardinelli et al., 2012; Wu et al., 2016)	Full Matrix (Connectivity) hyper-parametrized on Function Basis. Sparse Priors on the Function Basis ***Hyperparameters*** .	Constrained to Scaled Identity Matrix. Prior free.	Iterated Explicit ***Parameters*** Posterior Mean estimator and Maximum Likelihood plus Restrictions formulas of the Conditional ***Data*** Variance and Function Basis ***Hyperparameters***.	Nonlinear Multivariate
Structured Sparse Bayesian Learning (SSBL) (Wipf et al., 2010; Zhang and Rao, 2011; Babacan et al., 2012; Wan et al., 2014; Balkan et al. 2014;	Constrained to a Diagonal Matrix (or Block Diagonal). Sparse Gamma Priors.	Constrained to Scaled Identity Matrix. Jeffrey	Iterated Explicit ***Parameters*** Posterior Mean estimator and Empirical Bayes (EB) formulas of the Conditional	Nonlinear

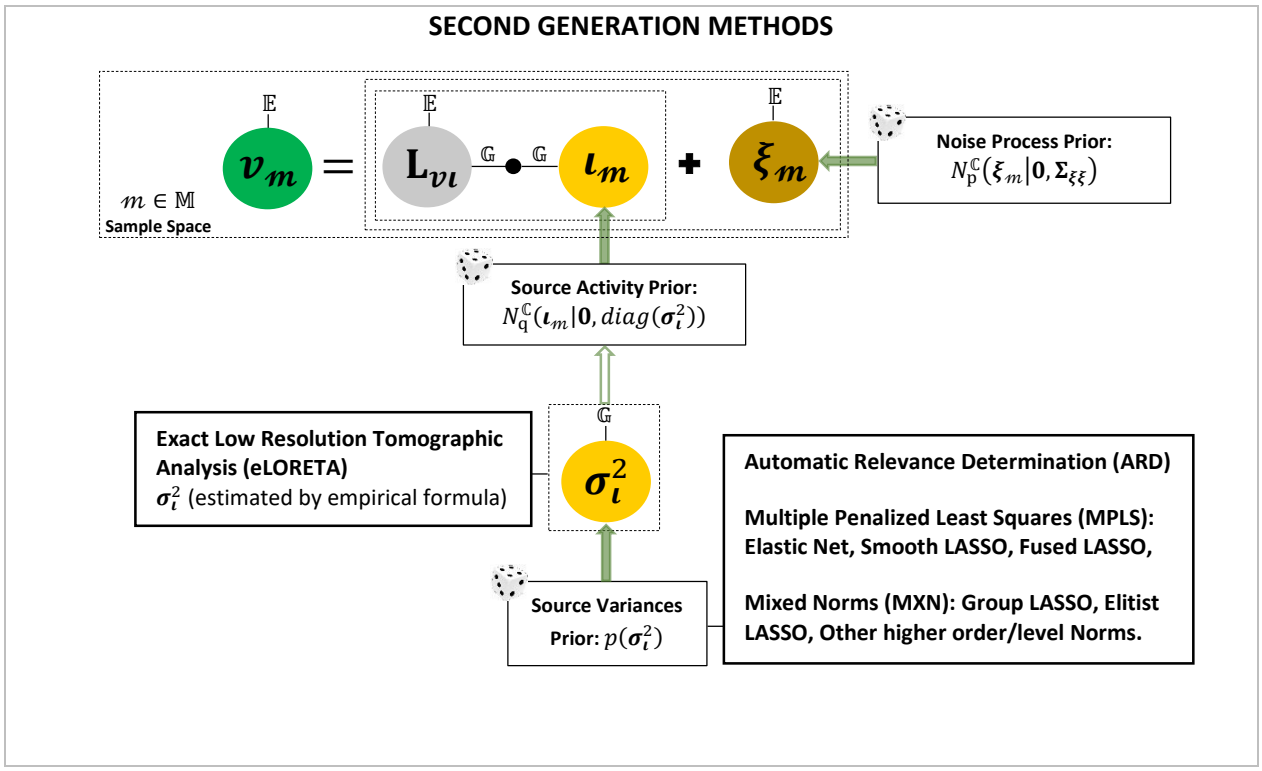
Paz-Linares et al., 2017; Zhang et al., 2016; Zhang et al., 2017).	Connectivity-Postprocessing.	Improper Prior.	***Data*** Variance and ***Parameters*** Variances.	

Summarizing ESI methodologies, we may refer as **Zero-generation** those endeavors of the pre-ESI period, pledged with the search of association between sensor level information and processes at the Gray Matter level [Blinowska et al 2004, Kus et al 2004, Blinowska 2011, Sakkalis 2011], and which has been fairly haunted by critics [Haufe et al 2013, Stokes et al 2017, Van de Steen et al 2019].

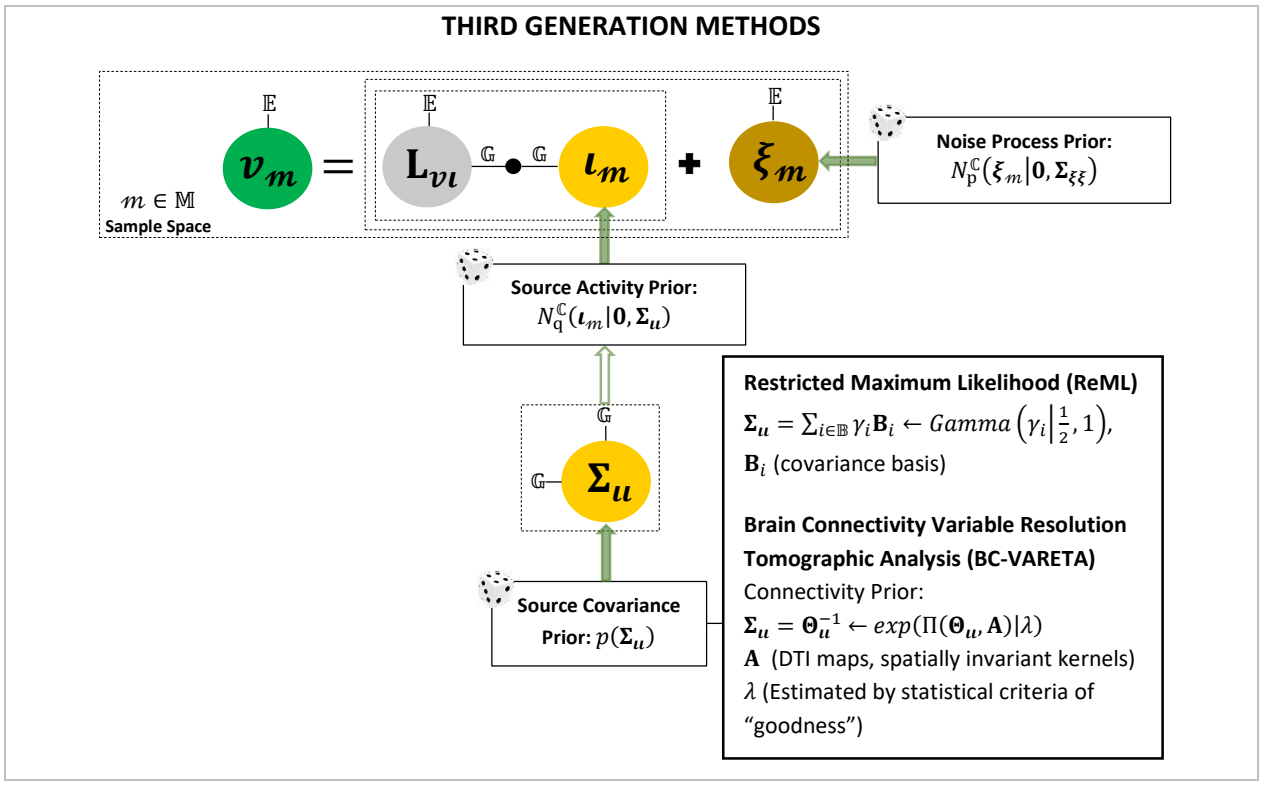
First-generation, linear nature ESI solution which imposes regularization on the LF ill-conditioning via static priors and which reflects -due it independence of the sensed activity and linearity- similar properties of the Zero-generation with exception of improved anatomical association, e.g. Minimum-Norm Estimates (MNE) [Hämäläinen et al 1994] and LORETAs [Pascual-Marqui et al 1994, Pascual-Marqui 1999].



Second-generation, nonlinear nature ESI solution for which the transfer function depends on the sensed activity and that usually pursues a sparsity pattern of activation, introduced in form of flexible and univariate priors that improve variable selection, e.g. Exact LORETA (eLORETA) [Pascual-Marqui et al 2006, Pascual-Marqui, 2007], Multiple Penalized Least Squares (MPLS) [Vega-Hernández et al., 2008], and Structured Sparse Bayesian Learning (SSBL) [Paz-Linares et al., 2017].



Third-generation: nonlinear and multivariate nature ESI solution which acknowledges possible coactivation due to the link between active sources, and therefore leads in some cases to improved determination of activity and its statistical dependencies, e.g. Variable Resolution Tomographic Analysis (VARETA) [Valdes-Sosa et al 1996, Bosch-Bayard et al 2001], Restricted Likelihood Maximization (ReLM) [Patterson and Thompson 1971, Harville 1977, Mattout et al 2006, Friston et al 2007, Friston et al 2008, Wipf et al 2009, Belardinelli et al, 2012] and Brain Connectivity VARETA (BC-VARETA) [Paz-Linares et al 2018, Gonzalez-Moreira et al 2018a, 2018b, 2018c]. *(ESI generations)*



SI-3 Andrews and Mallows Lemma for the complex variable Hierarchical Elastic Net

The results hierarchical Elastic Net can be extended to complex domain by modifying the corollary of Andrews and Mallows Lemma for Real ENET Gibbs pdf (Gaussian-Laplace) in [Andrews and Mallows, 1974]. For the complex case, the integral representation holds.

$$e^{-\alpha_1|z|-\alpha_2|z|^2} \propto \int N_1\big(|z|\big|0, f(\gamma)\big)p(\gamma)d\tau \qquad [3.1]$$

Where the variances are defined as $f(\gamma) = \frac{1}{2\alpha_2}\left(1 - \frac{\alpha_1^2}{4\alpha_2\gamma}\right)$. The measurable space in which the variable $z|\gamma$ is defined has a unnormalized density function given by the Gaussian pdf $p(z|\gamma) = N_1\big(|z|\big|0, f(\gamma)\big)$ and its variance $f(\gamma)$ is dependent on the random variable γ which has Truncated Gamma pdf.

$$p(\gamma) = TGa\left(\gamma\middle|\frac{1}{2}, 1, \left(\frac{\alpha_1^2}{4\alpha_2}, \infty\right)\right) \qquad [3.2]$$

So, the measure in the space product of z and $p(\gamma)$ is had density represented as an unnormalized product of Gaussian and Gamma densities.

$$p(z,\gamma) \propto N_1\big(|z|\big|0, f(\gamma)\big)TGa\left(\gamma\middle|\frac{1}{2}, 1, \left(\frac{\alpha_1^2}{4\alpha_2}, \infty\right)\right) \qquad [3.3]$$

SI-4 Tensor spatio-frequency prior probabilities of the group Hierarchical Elastic Net.

We introduce tensor group penalization on the 3D cartesian space of generators, samples and frequencies with:

$$|z| = \sqrt{\sum_{m\in\mathbb{M}}\sum_{\nu\in\mathbb{F}_b}\sum_{g\in\mathbb{G}_a}|\tilde{\boldsymbol{\iota}}_m(\nu;g)|^2} \qquad [4.1]$$

where

$\tilde{\boldsymbol{\iota}}_m(\nu) = \mathbf{DN}\boldsymbol{\iota}_m(\nu)$, $m \in \mathbb{M}$, $\nu \in \mathbb{F}$

a refers to a specific Gray matter area $a \in \mathbb{A}$

b refers to a specific frequency band $b \in \mathbb{B}$

Then the transformed prior of the parameters is described analytically by the following distribution

$$\begin{cases} \{\tilde{\boldsymbol{\iota}}_m(\nu;g), m \in \mathbb{M}, \nu \in \mathbb{F}_b, g \in \mathbb{G}_a\} \sim \dfrac{e^{-\frac{1}{2}f(\gamma_a)^{-1}\sum_{m\in\mathbb{M}}\sum_{\nu\in\mathbb{F}_b}\sum_{g\in\mathbb{G}_a}|\tilde{\iota}_m(\nu;g)|^2}}{|2\pi f(\gamma_a)|^{\frac{1}{2}}} \\ \qquad or \\ \tilde{\boldsymbol{\iota}}_m(\nu;g) \sim \dfrac{e^{-\frac{1}{2}\tilde{\iota}_m^{\dagger}(\nu;g)f(\gamma_a)^{-1}\tilde{\iota}_m(\nu;g)}}{|2\pi f(\gamma_a)|^{\frac{1}{2\mathrm{q}_a\mathrm{f}_b\mathrm{m}}}}, m \in \mathbb{M}, \nu \in \mathbb{F}_b, g \in \mathbb{G}_a \end{cases} \qquad [4.2]$$

where $f(\gamma_a) = \frac{1}{2a_2}\left(1 - \frac{\alpha_1^2}{4\alpha_2\gamma_a}\right)$ and $\gamma_a \sim TGa\left(\gamma_a \left|\frac{1}{2}, 1, \left(\frac{a_1^2}{4a_2}, \infty\right)\right.\right)$

In vector form they are expressed as

$$\begin{cases} \tilde{\boldsymbol{\iota}}_m(\nu) \sim N_{\mathrm{q}}^{unn}\left(|\tilde{\boldsymbol{\iota}}_m(\nu)| \middle| 0, diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2)\right), m \in \mathbb{M}, \nu \in \mathbb{F}_b \\ \qquad or \\ \boldsymbol{\iota}_m(\nu) \sim N_{\mathrm{q}}^{unn}\left(|\mathbf{DN}\boldsymbol{\iota}_m(\nu)| \middle| 0, diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2)\right), m \in \mathbb{M}, \nu \in \mathbb{F}_b \end{cases} \qquad [4.3]$$

where

N_{q}^{unn} is the unnormalized distribution $e^{-\frac{1}{2}\tilde{\boldsymbol{\iota}}_m^{\dagger}(\nu)diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2)^{-1}\tilde{\boldsymbol{\iota}}_m(\nu)} / |2\pi diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2)|^{\frac{1}{2\,(\prod_{a\in\mathbb{A}}\mathrm{q}_a)\mathrm{f}_b\mathrm{m}}}$

$\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2(g) = f(\gamma_a), g \in \mathbb{G}_a$, $a \in \mathbb{A}$ are the variances

The full vector Bayesian model is as follows:

$\boldsymbol{\nu}_m(\nu) \sim N_{\mathrm{p}}^{\mathbb{C}}\left(\boldsymbol{\nu}_m(\nu) \middle| \mathbf{L}_{\boldsymbol{\nu}\boldsymbol{\iota}}\boldsymbol{\iota}_m(\nu), \sigma_{\boldsymbol{\xi}}^2\mathbf{I}\right)$, $m \in \mathbb{M}$, $\nu \in \mathbb{F}_b$ "Likelihood" [4.4]

$\boldsymbol{\iota}_m(\nu) \sim N_{\mathrm{q}}^{unn}\left(|\mathbf{DN}\boldsymbol{\iota}_m(\nu)| \middle| 0, diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2)\right)$, $m \in \mathbb{M}$, $\nu \in \mathbb{F}_b$ "Parameters prior" [4.5]

$\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2(g) = f(\gamma_a), g \in \mathbb{G}_a$, $a \in \mathbb{A}$ "Variance parametrization" [4.6]

$$\boldsymbol{\gamma} \sim \prod_{a\in\mathbb{A}} TGa\left(\gamma_a \left|\frac{1}{2}, 1, \left(\frac{\alpha_1^2}{4\alpha_2}, \infty\right)\right.\right) \qquad \text{“Hyperparameters prior”} \qquad [4.7]$$

SI-5 Parameters posterior probability

For the joint distribution of data and parameters $p\left(\boldsymbol{v}_m(\nu), \boldsymbol{\iota}_m(\nu)\right)$ the following factorization holds, for simplicity we avoid the use of argument for frequency ν and samples m:

$$N_{\mathrm{p}}^{\mathbb{C}}\left(\boldsymbol{v}\middle|\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\boldsymbol{\iota}, \sigma_{\xi}^2\mathbf{I}\right) N_{\mathrm{q}}^{unn}\left(|\mathbf{DN}\boldsymbol{\iota}|\middle|0, diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2)\right) =$$
$$\left|\pi\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}\boldsymbol{\iota}}\right| N_{\mathrm{q}}^{\mathbb{C}}\left(\boldsymbol{\iota}\middle|\hat{\boldsymbol{\iota}}, \breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}\boldsymbol{\iota}}\right) N_{\mathrm{p}}^{\mathbb{C}}\left(\boldsymbol{v}\middle|\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\hat{\boldsymbol{\iota}}, \sigma_{\xi}^2\mathbf{I}\right) N_{\mathrm{q}}\left(|\mathbf{DN}\hat{\boldsymbol{\iota}}|\middle|0, diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2)\right) \qquad [5.1]$$

The quantities $\hat{\boldsymbol{\iota}}$ (posterior mean) and $\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}\boldsymbol{\iota}}$ (posterior covariance) are defined as follows

$$\hat{\boldsymbol{\iota}} = \breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}\boldsymbol{\iota}}\mathbf{L}_{\boldsymbol{\iota}\boldsymbol{v}}\left(\sigma_{\xi}^2\mathbf{I}\right)^{-1}\boldsymbol{v} \qquad \text{“Posterior mean”} \qquad [5.2]$$

$$\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}\boldsymbol{\iota}} = \left(\mathbf{L}_{\boldsymbol{\iota}\boldsymbol{v}}\left(\sigma_{\xi}^2\mathbf{I}\right)^{-1}\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}} + \frac{1}{2}\boldsymbol{\Sigma}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{-1}\right)^{-1} \qquad \text{“Posterior Covariance”} \qquad [5.3]$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{-1} = (\mathbf{DN})^{\mathcal{T}}\left(diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2)\right)^{-1}\mathbf{DN}$

This can be demonstrated by writing their distributions explicitly.

$$N_{\mathrm{p}}^{\mathbb{C}}\left(\boldsymbol{v}\middle|\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\boldsymbol{\iota}, \sigma_{\xi}^2\mathbf{I}\right) = \frac{e^{-(\boldsymbol{v}-\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\boldsymbol{\iota})^{\dagger}\left(\sigma_{\xi}^2\mathbf{I}\right)^{-1}(\boldsymbol{v}-\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\boldsymbol{\iota})}}{\left|\pi\sigma_{\xi}^2\mathbf{I}\right|} \qquad \text{“Likelihood”}$$

$$N_{\mathrm{q}}\left(|\mathbf{DN}\boldsymbol{\iota}|\middle|0, diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2)\right) = \frac{e^{-\frac{1}{2}\boldsymbol{\iota}^{\dagger}\boldsymbol{\Sigma}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{-1}\boldsymbol{\iota}}}{\left|2\pi diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2)\right|^{\frac{1}{2\left(\prod_{a\in\mathbb{A}}\mathrm{q}_a\right)\mathrm{f}_{\mathcal{B}}\mathrm{m}}}} \qquad \text{“Parameters prior”}$$

The form of the resultant distribution can be found by analyzing the terms that depend on the parameters (exponential argument) in the formula above

$$-\boldsymbol{v}^{\dagger}\left(\sigma_{\xi}^2\mathbf{I}\right)^{-1}\boldsymbol{v} - \boldsymbol{\iota}^{\dagger}\mathbf{L}_{\boldsymbol{\iota}\boldsymbol{v}}\left(\sigma_{\xi}^2\mathbf{I}\right)^{-1}\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\boldsymbol{\iota} + \boldsymbol{v}^{\dagger}\left(\sigma_{\xi}^2\mathbf{I}\right)^{-1}\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\boldsymbol{\iota} + \boldsymbol{\iota}^{\dagger}\mathbf{L}_{\boldsymbol{\iota}\boldsymbol{v}}\left(\sigma_{\xi}^2\mathbf{I}\right)^{-1}\boldsymbol{v} - \frac{1}{2}\boldsymbol{\iota}^{\dagger}\boldsymbol{\Sigma}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{-1}\boldsymbol{\iota}$$

Reorganizing $\mathbf{L}_{\boldsymbol{\iota}\boldsymbol{v}}\left(\sigma_{\xi}^2\mathbf{I}\right)^{-1}\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}} + \frac{1}{2}\boldsymbol{\Sigma}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{-1}$ of terms 2 and 5 to render $\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{-1}$ and $\mathbf{L}_{\boldsymbol{\iota}\boldsymbol{v}}\left(\sigma_{\xi}^2\mathbf{I}\right)^{-1}\boldsymbol{v}$ in terms 3 and 4 to render $\hat{\boldsymbol{\iota}}$

$$-\boldsymbol{v}^{\dagger}\left(\sigma_{\xi}^2\mathbf{I}\right)^{-1}\boldsymbol{v} + \left(\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}\boldsymbol{\iota}}\mathbf{L}_{\boldsymbol{\iota}\boldsymbol{v}}\left(\sigma_{\xi}^2\mathbf{I}\right)^{-1}\boldsymbol{v}\right)^{\dagger}\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{-1}\boldsymbol{\iota} + \boldsymbol{\iota}^{\dagger}\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{-1}\left(\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}\boldsymbol{\iota}}\mathbf{L}_{\boldsymbol{\iota}\boldsymbol{v}}\left(\sigma_{\xi}^2\mathbf{I}\right)^{-1}\boldsymbol{v}\right)$$
$$-\boldsymbol{\iota}^{\dagger}\left(\mathbf{L}_{\boldsymbol{\iota}\boldsymbol{v}}\left(\sigma_{\xi}^2\mathbf{I}\right)^{-1}\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}} + \frac{1}{2}\boldsymbol{\Sigma}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{-1}\right)\boldsymbol{\iota} \qquad [5.4]$$

$$-\boldsymbol{v}^{\dagger}\left(\sigma_{\xi}^2\mathbf{I}\right)^{-1}\boldsymbol{v} + \hat{\boldsymbol{\iota}}^{\dagger}\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{-1}\boldsymbol{\iota} + \boldsymbol{\iota}^{\dagger}\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{-1}\hat{\boldsymbol{\iota}} - \boldsymbol{\iota}^{\dagger}\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{-1}\boldsymbol{\iota} \qquad [5.5]$$

Completing with the term $+\hat{\boldsymbol{\iota}}^{\dagger}\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}}^{-1}\hat{\boldsymbol{\iota}}$

$$-(\boldsymbol{\iota}-\hat{\boldsymbol{\iota}})^{\dagger}\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}}^{-1}(\boldsymbol{\iota}-\hat{\boldsymbol{\iota}})-\boldsymbol{v}^{\dagger}\left(\sigma_{\xi}^{2}\mathbf{I}\right)^{-1}\boldsymbol{v}+\hat{\boldsymbol{\iota}}^{\dagger}\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}}^{-1}\hat{\boldsymbol{\iota}} \qquad [5.6]$$

Completing with the terms $-\boldsymbol{v}^{\dagger}\left(\sigma_{\xi}^{2}\mathbf{I}\right)^{-1}\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\hat{\boldsymbol{\iota}}$, $-\hat{\boldsymbol{\iota}}^{\dagger}\mathbf{L}_{\boldsymbol{\iota}\boldsymbol{v}}\left(\sigma_{\xi}^{2}\mathbf{I}\right)^{-1}\boldsymbol{v}$ and $+\hat{\boldsymbol{\iota}}^{\dagger}\mathbf{L}_{\boldsymbol{\iota}\boldsymbol{v}}\left(\sigma_{\xi}^{2}\mathbf{I}\right)^{-1}\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\hat{\boldsymbol{\iota}}$ we obtain:

$$-(\boldsymbol{\iota}-\hat{\boldsymbol{\iota}})^{\dagger}\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}}^{-1}(\boldsymbol{\iota}-\hat{\boldsymbol{\iota}})-(\boldsymbol{v}-\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\hat{\boldsymbol{\iota}})^{\dagger}\left(\sigma_{\xi}^{2}\mathbf{I}\right)^{-1}(\boldsymbol{v}-\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\hat{\boldsymbol{\iota}})-\boldsymbol{v}^{\dagger}\left(\sigma_{\xi}^{2}\mathbf{I}\right)^{-1}\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\hat{\boldsymbol{\iota}}$$
$$-\hat{\boldsymbol{\iota}}^{\dagger}\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}^{\mathcal{T}}\left(\sigma_{\xi}^{2}\mathbf{I}\right)^{-1}\boldsymbol{v}+\hat{\boldsymbol{\iota}}^{\dagger}\mathbf{L}_{\boldsymbol{\iota}\boldsymbol{v}}\left(\sigma_{\xi}^{2}\mathbf{I}\right)^{-1}\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\hat{\boldsymbol{\iota}}+\hat{\boldsymbol{\iota}}^{\dagger}\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}}^{-1}\hat{\boldsymbol{\iota}} \qquad [5.7]$$

Then terms 3, 4 and 6 can be reorganized into $-\hat{\boldsymbol{\iota}}^{\dagger}\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}}^{-1}\hat{\boldsymbol{\iota}}$ since $\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}^{\mathcal{T}}\left(\sigma_{\xi}^{2}\mathbf{I}\right)^{-1}\boldsymbol{v}=\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}}^{-1}\hat{\boldsymbol{\iota}}^{\dagger}$ to obtain:

$$-(\boldsymbol{\iota}_{m}-\hat{\boldsymbol{\iota}})^{\dagger}\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}}^{-1}(\boldsymbol{\iota}_{m}-\hat{\boldsymbol{\iota}})-(\boldsymbol{v}_{m}-\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\hat{\boldsymbol{\iota}})^{\dagger}\left(\sigma_{\xi}^{2}\mathbf{I}\right)^{-1}(\boldsymbol{v}_{m}-\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\hat{\boldsymbol{\iota}})$$
$$-\hat{\boldsymbol{\iota}}^{\dagger}\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}}^{-1}\hat{\boldsymbol{\iota}}+\hat{\boldsymbol{\iota}}^{\dagger}\mathbf{L}_{\boldsymbol{\iota}\boldsymbol{v}}\left(\sigma_{\xi}^{2}\mathbf{I}\right)^{-1}\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\hat{\boldsymbol{\iota}} \qquad [5.8]$$

But combining terms 3 and 4 yields $-\frac{1}{2}\hat{\boldsymbol{\iota}}^{\dagger}\boldsymbol{\Sigma}_{\boldsymbol{\iota}}^{-1}\hat{\boldsymbol{\iota}}$

$$-(\boldsymbol{\iota}-\hat{\boldsymbol{\iota}})^{\dagger}\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}}^{-1}(\boldsymbol{\iota}-\hat{\boldsymbol{\iota}})-(\boldsymbol{v}-\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\hat{\boldsymbol{\iota}})^{\dagger}\left(\sigma_{\xi}^{2}\mathbf{I}\right)^{-1}(\boldsymbol{v}-\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\hat{\boldsymbol{\iota}})-\frac{1}{2}\hat{\boldsymbol{\iota}}^{\dagger}\boldsymbol{\Sigma}_{\boldsymbol{\iota}}^{-1}\hat{\boldsymbol{\iota}} \qquad [5.9]$$

From this, it holds that:

$$N_{\mathrm{p}}^{\mathbb{C}}\left(\boldsymbol{v}\middle|\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\boldsymbol{\iota},\sigma_{\xi}^{2}\mathbf{I}\right)N_{\mathrm{q}}^{unn}\left(|\mathbf{D}\mathbf{N}\boldsymbol{\iota}|\middle|0,diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^{2})\right)=$$
$$\left|\pi\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}}\right|N_{\mathrm{q}}^{\mathbb{C}}\left(\boldsymbol{\iota}\middle|\hat{\boldsymbol{\iota}},\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}}\right)N_{\mathrm{p}}^{\mathbb{C}}\left(\boldsymbol{v}\middle|\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\hat{\boldsymbol{\iota}},\sigma_{\xi}^{2}\mathbf{I}\right)N_{\mathrm{q}}^{unn}\left(|\mathbf{D}\mathbf{N}\hat{\boldsymbol{\iota}}|\middle|0,diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^{2})\right) \qquad [5.10]$$

SI-6 Empirical Bayes and hyperparameters posterior probability

From the joint distribution $p\left(\left\{\left(\boldsymbol{v}_{m}(\nu),\boldsymbol{\iota}_{m}(\nu)\right),m\in\mathbb{M},\nu\in\mathbb{F}_{\mathcal{b}}\right\}\right)$ we can derive iteratively and approximated representation the Type II-Likelihood $p^{(k)}(\{\boldsymbol{v}_{m}(\nu),m\in\mathbb{M},\nu\in\mathbb{F}_{\mathcal{b}}\})$:

$$p(\{\boldsymbol{v}_{m}(\nu),m\in\mathbb{M},\nu\in\mathbb{F}_{\mathcal{b}}\})=\prod_{\nu\in\mathbb{F}_{\mathcal{b}}}\prod_{m\in\mathbb{M}}\int p\left(\boldsymbol{v}_{m}(\nu),\boldsymbol{\iota}_{m}(\nu)\right)d\boldsymbol{\iota}_{m}(\nu) \qquad [6.1]$$

$$p(\{\boldsymbol{v}_{m}(\nu),m\in\mathbb{M},\nu\in\mathbb{F}_{\mathcal{b}}\})\approx p^{(k)}(\{\boldsymbol{v}_{m}(\nu),m\in\mathbb{M},\nu\in\mathbb{F}_{\mathcal{b}}\}) \qquad [6.2]$$

where

$$p^{(k)}(\{\boldsymbol{v}_{m}(\nu),m\in\mathbb{M},\nu\in\mathbb{F}_{\mathcal{b}}\})=\prod_{\nu\in\mathbb{F}_{\mathcal{b}}}\prod_{m\in\mathbb{M}}\int N_{\mathrm{p}}^{\mathbb{C}}\left(\boldsymbol{v}_{m}(\nu)\middle|\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\boldsymbol{\iota}_{m}(\nu),\sigma_{\xi}^{2}\mathbf{I}\right)\times$$
$$N_{\mathrm{q}}^{unn}\left(|\mathbf{D}\mathbf{N}\boldsymbol{\iota}_{m}(\nu)|\middle|0,diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^{2})\right)d\boldsymbol{\iota}_{m}(\nu) \qquad [6.3]$$

The analysis of section 5 yields

$$\int N_{\mathrm{p}}^{\mathbb{C}}\left(\boldsymbol{v}_{m}(\nu)\middle|\mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\boldsymbol{\iota}_{m}(\nu),\sigma_{\xi}^{2}\mathbf{I}\right)N_{\mathrm{q}}^{unn}\left(|\mathbf{D}\mathbf{N}\boldsymbol{\iota}_{m}(\nu)|\middle|0,diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^{2})\right)d\boldsymbol{\iota}_{m}(\nu)=$$

$$\left|\pi\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota\iota}}\right| N_{\mathrm{p}}^{\mathbb{C}}\left(\boldsymbol{v}_m(\nu)\middle|\mathbf{L}_{\boldsymbol{v\iota}}\hat{\boldsymbol{\iota}}_m(\nu),\sigma_{\boldsymbol{\xi}}^2\mathbf{I}\right) N_{\mathrm{q}}^{unn}\left(\left|\mathbf{DN}\hat{\boldsymbol{\iota}}_m(\nu)\right|\middle|0, diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2)\right) \qquad [6.4]$$

Then iteratively upon fixed values of the parameters $\hat{\boldsymbol{\iota}}_m^{(k)}(\nu)$ the Type II-Likelihood is expressed as:

$$p^{(k)}(\{\boldsymbol{v}_m(\nu), m \in \mathbb{M}, \nu \in \mathbb{F}_{\mathcal{b}}\}) =$$

$$\begin{cases} \prod_{\nu\in\mathbb{F}_{\mathcal{b}}}\prod_{m\in\mathbb{M}}\left|\pi\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota\iota}}\right| N_{\mathrm{p}}^{\mathbb{C}}\left(\boldsymbol{v}_m(\nu)\middle|\mathbf{L}_{\boldsymbol{v\iota}}\hat{\boldsymbol{\iota}}_m^{(k)}(\nu),\sigma_{\boldsymbol{\xi}}^2\mathbf{I}\right) \times \\ \qquad N_{\mathrm{q}}^{unn}\left(\left|\mathbf{DN}\hat{\boldsymbol{\iota}}_m^{(k)}(\nu)\right|\middle|0, diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2)\right) \\ \qquad or \\ \dfrac{\left|\pi\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota\iota}}\right|^{\mathrm{f}_{\mathcal{b}}\mathrm{m}}}{\left|\pi\sigma_{\boldsymbol{\xi}}^2\mathbf{I}\right|^{\mathrm{f}_{\mathcal{b}}\mathrm{m}}\left|2\pi diag(f(\boldsymbol{\gamma}))\right|^{\frac{1}{2}}} e^{-\frac{\mathrm{f}_{\mathcal{b}}\mathrm{m}}{2}tr\left(diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^{-2})\tilde{\mathbf{S}}_{\boldsymbol{\iota\iota}}^{(k)}\right)} e^{-\mathrm{f}_{\mathcal{b}}\mathrm{m}\sigma_{\boldsymbol{\xi}}^{-2}tr\left(\mathbf{S}_{\boldsymbol{\xi\xi}}^{(k)}\right)} \end{cases} \qquad [6.5]$$

Where the covariances $\tilde{\mathbf{S}}_{\boldsymbol{\iota\iota}}^{(k)}$ and $\mathbf{S}_{\boldsymbol{\xi\xi}}^{(k)}$ are

$$\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota\iota}} = \left(\mathbf{L}_{\boldsymbol{\iota v}}\left(\sigma_{\boldsymbol{\xi}}^2\mathbf{I}\right)^{-1}\mathbf{L}_{\boldsymbol{v\iota}} + \frac{1}{2}\boldsymbol{\Sigma}_{\boldsymbol{\iota\iota}}^{-1}\right)^{-1} \text{ with } \boldsymbol{\Sigma}_{\boldsymbol{\iota\iota}}^{-1} = (\mathbf{DN})^{\mathcal{T}}\left(diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2)\right)^{-1}\mathbf{DN}$$

$$\tilde{\mathbf{S}}_{\boldsymbol{\iota\iota}}^{(k)} = \mathbf{DNS}_{\boldsymbol{\iota\iota}}^{(k)}(\mathbf{DN})^{\mathcal{T}} \text{ with } \mathbf{S}_{\boldsymbol{\iota\iota}}^{(k)} = \frac{1}{\mathrm{f}_{\mathcal{b}}\mathrm{m}}\sum_{\nu\in\mathbb{F}_{\mathcal{b}}}\sum_{m\in\mathbb{M}}\hat{\boldsymbol{\iota}}_m^{(k)}(\nu)\left(\hat{\boldsymbol{\iota}}_m^{(k)}(\nu)\right)^{\dagger}$$

$$\mathbf{S}_{\boldsymbol{\xi\xi}}^{(k)} = \frac{1}{\mathrm{f}_{\mathcal{b}}\mathrm{m}}\sum_{\nu\in\mathbb{F}_{\mathcal{b}}}\sum_{m\in\mathbb{M}}\left(\boldsymbol{v}_m(\nu) - \mathbf{L}_{\boldsymbol{v\iota}}\hat{\boldsymbol{\iota}}_m^{(k)}(\nu)\right)\left(\boldsymbol{v}_m(\nu) - \mathbf{L}_{\boldsymbol{v\iota}}\hat{\boldsymbol{\iota}}_m^{(k)}(\nu)\right)^{\dagger}$$

Parameter estimators

The parameters $\hat{\boldsymbol{\iota}}_m^{(k)}(\nu)$ are determined in the previous iteration in terms of the iteratively linear source transfer operator $\breve{\mathbf{T}}_{\boldsymbol{\iota v}}^{(k)}$

$$\hat{\boldsymbol{\iota}}_m^{(k)}(\nu) = \breve{\mathbf{T}}_{\boldsymbol{\iota v}}^{(k)}\boldsymbol{v}_m(\nu) \qquad [6.6]$$

The source transfer operator is defined upon fixed values of the hyperparameters:

$$\breve{\mathbf{T}}_{\boldsymbol{\iota v}}^{(k)} = \breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota\iota}}^{(k)}\mathbf{L}_{\boldsymbol{\iota v}}\left(\sigma_{\boldsymbol{\xi}}^{2^{(k)}}\mathbf{I}\right)^{-1} \qquad [6.7]$$

where

$$\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota\iota}}^{(k)} = \left(\mathbf{L}_{\boldsymbol{\iota v}}\left(\sigma_{\boldsymbol{\xi}}^{2^{(k)}}\mathbf{I}\right)^{-1}\mathbf{L}_{\boldsymbol{v\iota}} + \frac{1}{2}\boldsymbol{\Sigma}_{\boldsymbol{\iota\iota}}^{(k)^{-1}}\right)^{-1} \text{ with } \boldsymbol{\Sigma}_{\boldsymbol{\iota\iota}}^{(k)^{-1}} = (\mathbf{DN})^{\mathcal{T}}\left(diag\left(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^{2^{(k)}}\right)\right)^{-1}\mathbf{DN}$$

This simplifies the computations of $\mathbf{S}_{\boldsymbol{\iota\iota}}^{(k)}$ expressed as:

$$\mathbf{S}_{\boldsymbol{\iota\iota}}^{(k)} = \breve{\mathbf{T}}_{\boldsymbol{\iota v}}^{(k)}\mathbf{S}_{\boldsymbol{vv}}\breve{\mathbf{T}}_{\boldsymbol{v\iota}}^{(k)} \qquad [6.8]$$

where

$$\mathbf{S}_{\boldsymbol{vv}} = \frac{1}{\mathrm{f}_{\mathscr{b}}\mathrm{m}} \sum_{\nu \in \mathbb{F}_{\mathscr{b}}} \sum_{m \in \mathbb{M}} \boldsymbol{v}_m(\nu) \boldsymbol{v}_m^{\dagger}$$

SI-7 Hyperparameter estimators

The estimation formulas can be derived by applying maximum posterior of the combined Type II Likelihood and priors. To do so we reformulate the targeted hyperparameters as follows:

$$\rho \leftarrow \frac{\alpha_1^2}{4\alpha_2} \qquad [7.1]$$

$$\alpha \leftarrow \alpha_2 \qquad [7.2]$$

$$\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a) \leftarrow 2a\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2(\mathscr{g}), \mathscr{g} \in \mathbb{G}_a, a \in \mathbb{A} \qquad [7.3]$$

Variances

First the estimator of variances $\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)$ can be computed from the stationary values of the expression:

$$\begin{cases} \frac{\partial}{\partial \tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)} \log\left(p^{(k)}(\{\boldsymbol{v}_m(\nu), m \in \mathbb{M}, \nu \in \mathbb{F}_{\mathscr{b}}\}) p(\gamma_a)\right) = 0 \\ or \\ \frac{\partial}{\partial \tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)} \left(\mathrm{f}_{\mathscr{b}}\mathrm{m} \log\left|\pi \breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota\iota}}\right| - \frac{\mathrm{f}_{\mathscr{b}}\mathrm{m}}{2} tr\left(diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^{-2}) \tilde{\mathbf{S}}_{\boldsymbol{\iota\iota}}^{(k)}\right)\right) \\ -\frac{1}{2} \frac{\partial}{\partial \tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)} \log f(\gamma_a) + \frac{\partial}{\partial \tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)} \log p(\gamma_a) = 0 \end{cases} \qquad [7.4]$$

Due to the chain rule of matrix derivatives the first term can be expressed in a close form, see preposition 2.3.2b in [Paz-Linares, 2017].

$$\frac{\partial}{\partial \tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)} \left(\mathrm{f}_{\mathscr{b}}\mathrm{m} \log\left|\pi \breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota\iota}}\right| - \frac{\mathrm{f}_{\mathscr{b}}\mathrm{m}}{2} tr\left(diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^{-2}) \tilde{\mathbf{S}}_{\boldsymbol{\iota\iota}}^{(k)}\right)\right) =$$

$$\begin{cases} \sum_{\mathscr{g} \in \mathbb{G}_a} \frac{\partial}{\partial \boldsymbol{\sigma}_{\boldsymbol{\iota}}^2(\mathscr{g})} \left(\mathrm{f}_{\mathscr{b}}\mathrm{m} \log\left|\pi \breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota\iota}}\right| - \frac{\mathrm{f}_{\mathscr{b}}\mathrm{m}}{2} tr\left(diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^{-2}) \tilde{\mathbf{S}}_{\boldsymbol{\iota\iota}}^{(k)}\right)\right) \frac{\partial \boldsymbol{\sigma}_{\boldsymbol{\iota}}^2(\mathscr{g})}{\partial \tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)} \\ or \\ \frac{\mathrm{f}_{\mathscr{b}}\mathrm{m}}{2} \boldsymbol{\sigma}_{\boldsymbol{\iota}}^{-4}(\mathscr{g}) \left(\tilde{\breve{\boldsymbol{\Sigma}}}_{\boldsymbol{\iota\iota}}^{(k)}(\mathscr{g}, \mathscr{g}) + \tilde{\mathbf{S}}_{\boldsymbol{\iota\iota}}^{(k)}(\mathscr{g}, \mathscr{g})\right) \end{cases} \qquad [7.5]$$

where

$$\tilde{\mathbf{S}}_{\boldsymbol{\iota\iota}}^{(k)} = \mathbf{DNS}_{\boldsymbol{\iota\iota}}^{(k)} (\mathbf{DN})^{\mathcal{T}} \text{ with } \mathbf{S}_{\boldsymbol{\iota\iota}}^{(k)} = \breve{\mathbf{T}}_{\boldsymbol{\iota v}}^{(k)} \mathbf{S}_{\boldsymbol{vv}} \breve{\mathbf{T}}_{\boldsymbol{v\iota}}^{(k)}$$

$$\tilde{\breve{\boldsymbol{\Sigma}}}_{\boldsymbol{\iota\iota}}^{(k)} = \mathbf{DN} \breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota\iota}}^{(k)} (\mathbf{DN})^{\mathcal{T}}$$

$$\frac{\partial \boldsymbol{\sigma}_{\boldsymbol{\iota}}^2(\mathscr{g})}{\partial \tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)} = \begin{cases} \frac{1}{2\alpha^{(k)}} & , \mathscr{g} \in \mathbb{G}_a \\ 0 & , otherwise \end{cases}$$

Yielding:

$$\frac{\partial}{\partial \tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)}\left(\mathrm{f}_{\mathscr{b}}\mathrm{m}\log\left|\pi\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}\boldsymbol{\iota}}\right| - \frac{\mathrm{f}_{\mathscr{b}}\mathrm{m}}{2} tr\left(diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^{-2})\tilde{\mathbf{S}}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{(k)}\right)\right) =$$

$$\mathrm{f}_{\mathscr{b}}\mathrm{m}\frac{\alpha^{(k)}}{\left(\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)\right)^2}\sum_{\mathscr{g}\in\mathbb{G}_a}\left(\tilde{\tilde{\boldsymbol{\Sigma}}}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{(k)}(\mathscr{g},\mathscr{g}) + \tilde{\mathbf{S}}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{(k)}(\mathscr{g},\mathscr{g})\right) \qquad [7.6]$$

The second and third term derivatives are

$$\frac{\partial}{\partial \tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)}\log f(\gamma_a) = \frac{1}{f(\gamma_a)}\frac{\partial}{\partial \eta_a} f(\gamma_a) = \frac{1}{\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)} \qquad [7.7]$$

where

$f(\gamma_a) = \frac{1}{2\alpha^{(k)}}\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)$ and $\frac{\partial}{\partial \tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)} f(\gamma_a) = \frac{1}{2\alpha^{(k)}}$

$$\frac{\partial}{\partial \tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)}\log p(\gamma_a) = \frac{\partial}{\partial \tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)}\log TGa\left(\frac{\rho^{(k)}}{1-\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)}\middle|\frac{1}{2}, 1, (r,\infty)\right) = -\frac{1}{2\left(1-\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)\right)} - \frac{\rho^{(k)}}{\left(1-\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)\right)^2} \qquad [7.8]$$

Substituting the derivatives, we obtain:

$$\frac{\partial}{\partial \tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)}\log\left(p^{(k)}(\{\boldsymbol{v}_m(\nu), m \in \mathbb{M}, \nu \in \mathbb{F}_{\mathscr{b}}\})p(\gamma_a)\right) =$$

$$\mathrm{f}_{\mathscr{b}}\mathrm{m}\frac{\alpha^{(k)}}{\left(\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)\right)^2}\sum_{\mathscr{g}\in\mathbb{G}_a}\left(\tilde{\tilde{\boldsymbol{\Sigma}}}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{(k)}(\mathscr{g},\mathscr{g}) + \tilde{\mathbf{S}}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{(k)}(\mathscr{g},\mathscr{g})\right) - \frac{1}{2}\frac{1}{\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)} - \frac{1}{2\left(1-\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)\right)} - \frac{\rho^{(k)}}{\left(1-\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)\right)^2} \qquad [7.9]$$

But since it holds that $-\frac{1}{2}\frac{1}{\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)} - \frac{1}{2\left(1-\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)\right)} = -\frac{1}{2\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)\left(1-\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)\right)}$ the expression is much more compact:

$$\frac{\partial}{\partial \tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)}\log\left(p^{(k)}(\{\boldsymbol{v}_m(\nu), m \in \mathbb{M}, \nu \in \mathbb{F}_{\mathscr{b}}\})p(\gamma_a)\right) =$$

$$\mathrm{f}_{\mathscr{b}}\mathrm{m}\frac{\alpha^{(k)}}{\left(\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)\right)^2}\sum_{\mathscr{g}\in\mathbb{G}_a}\left(\tilde{\tilde{\boldsymbol{\Sigma}}}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{(k)}(\mathscr{g},\mathscr{g}) + \tilde{\mathbf{S}}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{(k)}(\mathscr{g},\mathscr{g})\right) - \frac{1}{2\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)\left(1-\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)\right)} - \frac{\rho^{(k)}}{\left(1-\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)\right)^2} \qquad [7.10]$$

Using the auxiliary quantity η_a so that $\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a) = \frac{\eta_a}{\left(\rho^{(k)}+\eta_a\right)}$ we obtain:

$$\mathrm{f}_{\mathscr{b}}\mathrm{m}\,\alpha^{(k)}\left(\rho^{(k)}\right)^2\sum_{\mathscr{g}\in\mathbb{G}_a}\left(\tilde{\tilde{\boldsymbol{\Sigma}}}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{(k)}(\mathscr{g},\mathscr{g}) + \tilde{\mathbf{S}}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{(k)}(\mathscr{g},\mathscr{g})\right) - \frac{1}{2}\rho^{(k)}\eta_a - \rho^{(k)}(\eta_a)^2 = 0 \qquad [7.11]$$

Therefore, the only possible solution for the conditions set by the problem statement and estimator can be obtained from η_a with:

$$\eta_a^{(k+1)} = -\frac{1}{4} + \sqrt{\frac{1}{16} + \mathrm{f}_{\mathscr{b}}\mathrm{m}\,\alpha^{(k)}\rho^{(k)}\sum_{\mathscr{g}\in\mathbb{G}_a}\left(\tilde{\tilde{\boldsymbol{\Sigma}}}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{(k)}(\mathscr{g},\mathscr{g}) + \tilde{\mathbf{S}}_{\boldsymbol{\iota}\boldsymbol{\iota}}^{(k)}(\mathscr{g},\mathscr{g})\right)} \qquad [7.12]$$

$$\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^{2\,(k+1)}(a) = \frac{\eta_a^{(k+1)}}{\left(\rho^{(k)}+\eta_a^{(k+1)}\right)} \qquad [7.13]$$

Regularization parameters ρ and α

The estimator of the regularization parameters r and a can be computed from the stationary values of the expression below, following the same steps as for the variances, see preposition 2.3.2c in [Paz-Linares, 2017]:

$$\begin{cases} \frac{\partial}{\partial\rho}\log p(\boldsymbol{\gamma}) + \frac{\partial}{\partial r}\log p(\rho) = 0 \\ and \\ \frac{\partial}{\partial\alpha}\left(\mathrm{f}_{\mathscr{b}}\mathrm{m}\log\left|\pi\breve{\boldsymbol{\Sigma}}_{\boldsymbol{u}}\right| - \frac{\mathrm{f}_{\mathscr{b}}\mathrm{m}}{2}tr\left(diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^{-2})\tilde{\mathbf{S}}_{\boldsymbol{u}}^{(k)}\right)\right) \\ -\frac{1}{2}\frac{\partial}{\partial\alpha}\log\left|2\pi diag\left(f(\boldsymbol{\gamma})\right)\right| + \frac{\partial}{\partial\alpha}\log p(\alpha) = 0 \end{cases} \qquad [7.14]$$

Writing explicitly the TGa distribution and using the chain rule in the derivative $\frac{\partial}{\partial r}\log p(\boldsymbol{\gamma})$:

$$\frac{\partial}{\partial r}\log p(\boldsymbol{\gamma}) = \sum_{a\in\mathbb{A}}\left(\frac{\partial}{\partial\gamma_a}\log Ga\left(\gamma_a\left|\frac{1}{2},1,(\rho,\infty)\right.\right)\frac{\partial\gamma_a}{\partial r} - \frac{\partial}{\partial r}\log\int_\rho^\infty Ga\left(x\left|\frac{1}{2},1\right.\right)dx\right) \qquad [7.15]$$

where

$$\frac{\partial}{\partial\gamma_a}\log Ga\left(\gamma_a\left|\frac{1}{2},1,(\rho,\infty)\right.\right) = -\frac{1-\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^2(a)}{2\rho} - 1$$

$$\frac{\partial\gamma_a}{\partial\rho} = \frac{1}{1-\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^{2(k)}(a)}$$

$$\frac{\partial}{\partial\rho}\log\int_\rho^\infty Ga\left(x\left|\frac{1}{2},1\right.\right)dx = -Ga\left(\rho\left|\frac{1}{2},1\right.\right)/\int_\rho^\infty Ga\left(x\left|\frac{1}{2},1\right.\right)dx$$

The derivative $\frac{\partial}{\partial\rho}\log p(\rho)$ is given by:

$$\frac{\partial}{\partial\rho}\log p(\rho) = \frac{s_\rho}{\rho} - r_\rho \qquad [7.16]$$

where $p(\rho) \propto Ga(\rho|s_\rho+1, r_\rho)$ with shape $(s_\rho+1)$ and rate r_ρ

From this we obtain the equation for ρ:

$$\rho^{(k+1)} = zero_\rho\left(-\sum_{a\in\mathbb{A}}\frac{1}{1-\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^{2(k)}(a)} + l\frac{Ga\left(\rho\left|\frac{1}{2},1\right.\right)}{\int_\rho^\infty Ga\left(x\left|\frac{1}{2},1\right.\right)dx} + \left(s_\rho - \frac{l}{2}\right)\frac{1}{\rho} - r_\rho\right) \qquad [7.17]$$

Due to the chain rule of matrix derivatives the first term for the parameter α can be expressed as:

$$\frac{\partial}{\partial\alpha}\left(\mathrm{f}_{\mathscr{b}}\mathrm{m}\log\left|\pi\breve{\boldsymbol{\Sigma}}_{\boldsymbol{u}}\right| - \frac{\mathrm{f}_{\mathscr{b}}\mathrm{m}}{2}tr\left(diag(\boldsymbol{\sigma}_{\boldsymbol{\iota}}^{-2})\tilde{\mathbf{S}}_{\boldsymbol{u}}^{(k)}\right)\right) =$$

$$\sum_{a\in\mathbb{A}}\sum_{\mathscr{g}\in\mathbb{G}_a}\frac{\mathrm{f}_{\mathscr{b}}\mathrm{m}}{2}\boldsymbol{\sigma}_{\boldsymbol{\iota}}^{-4}(\mathscr{g})\left(\tilde{\breve{\boldsymbol{\Sigma}}}_{\boldsymbol{u}}^{(k)}(\mathscr{g},\mathscr{g}) + \tilde{\mathbf{S}}_{\boldsymbol{u}}^{(k)}(\mathscr{g},\mathscr{g})\right)\frac{\partial\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2(\mathscr{g})}{\partial\alpha} \qquad [7.18]$$

where

$\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2(g) = \frac{1}{2a}\widetilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^{2\,(k)}(a),\ g \in \mathbb{G}_a,\ a \in \mathbb{A}$

$\frac{\partial \boldsymbol{\sigma}_{\boldsymbol{\iota}}^2(g)}{\partial \alpha} = -\frac{1}{2a^2}\widetilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^{2\,(k)}(a),\ g \in \mathbb{G}_a,\ a \in \mathbb{A}$

The derivative $\frac{\partial}{\partial \alpha}\log\left|2\pi diag\left(f(\boldsymbol{\gamma})\right)\right|$ is given by:

$$\frac{\partial}{\partial \alpha}\log\left|2\pi diag\left(f(\boldsymbol{\gamma})\right)\right| = \sum_{a\in\mathbb{A}}\frac{1}{f(\gamma_a)}\frac{\partial}{\partial \alpha}f(\gamma_a) \qquad [7.19]$$

where

$f(\gamma_a) = \frac{1}{2a}\widetilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^{2\,(k)}(a)$

$\frac{\partial}{\partial \alpha}f(\gamma_a) = -\frac{1}{2a^2}\widetilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^{2\,(k)}(a)$

The derivative $\frac{\partial}{\partial \rho}\log p(\rho)$ is given by:

$$\frac{\partial}{\partial \alpha}\log p(\alpha) = \frac{s_\alpha}{\alpha} - r_\alpha \qquad [7.20]$$

where $p(\alpha) \propto Ga(\alpha|(s_\alpha + 1), r_\alpha)$ with shape $(s_\alpha + 1)$ and rate r_α

From this we obtain the equation for α:

$$-\mathrm{f}_{\mathcal{B}}\mathrm{m}\sum_{a\in\mathbb{A}}\frac{\sum_{g\in\mathbb{G}_a}\left(\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}}^{(k)}(g,g) + \tilde{\mathbf{S}}_{\boldsymbol{\iota}}^{(k)}(g,g)\right)}{\widetilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^{2\,(k)}(a)} + \frac{l}{2\alpha} + \frac{s_\alpha}{\alpha} - r_\alpha = 0 \qquad [7.21]$$

or

$$\alpha^{(k+1)} = \left(s_\alpha + \frac{l}{2}\right)\Bigg/\left(\mathrm{f}_{\mathcal{B}}\mathrm{m}\sum_{a\in\mathbb{A}}\frac{\sum_{g\in\mathbb{G}_a}\left(\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}}^{(k)}(g,g) + \tilde{\mathbf{S}}_{\boldsymbol{\iota}}^{(k)}(g,g)\right)}{\widetilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^{2\,(k)}(a)} + r_\alpha\right) \qquad [7.22]$$

Noise parameter $\boldsymbol{\sigma}_{\boldsymbol{\xi}}^2$

The estimator of the noise parameter $\sigma_{\boldsymbol{\xi}}^2$ can be computed from the stationary values of the expression below, following the same steps as for the variances, see preposition 2.3.2d in [Paz-Linares, 2017]:

$$\frac{\partial}{\partial \sigma_{\xi}^2}\left(\mathrm{f}_{\mathcal{B}}\mathrm{m}\log\left|\pi\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}}\right| - \mathrm{f}_{\mathcal{B}}\mathrm{m}\sigma_{\xi}^{-2}tr\left(\mathbf{S}_{\xi\xi}^{(k)}\right)\right) - \mathrm{f}_{\mathcal{B}}\mathrm{m}\frac{\partial}{\partial \sigma_{\xi}^2}\log\left|\pi\sigma_{\xi}^2\mathbf{I}\right| + \frac{\partial}{\partial \sigma_{\xi}^2}\log p\left(\sigma_{\xi}^2\right) = 0 \qquad [7.23]$$

Due to the chain rule of matrix derivatives the first term for the parameter α can be expressed as:

$$\frac{\partial}{\partial \sigma_\xi^2}\left(\mathrm{f}_{\mathscr{b}}\mathrm{m}\log\left|\pi\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}}\right| - \mathrm{f}_{\mathscr{b}}\mathrm{m}\sigma_\xi^{-2} tr\left(\mathbf{S}_{\xi\xi}^{(k)}\right)\right) = \mathrm{f}_{\mathscr{b}}\mathrm{m}\,\sigma_\xi^{-4}\sum_{e\in\mathbb{E}}\left(\breve{\boldsymbol{\Sigma}}_{\xi\xi}^{(k)}(e,e) + \mathbf{S}_{\xi\xi}^{(k)}(e,e)\right) \qquad [7.24]$$

where $\breve{\boldsymbol{\Sigma}}_{\xi\xi}^{(k)} = \mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}}^{(k)}\mathbf{L}_{\boldsymbol{\iota}\boldsymbol{v}}$ and $\mathbf{S}_{\xi\xi}^{(k)} = \breve{\mathbf{T}}_{\xi\boldsymbol{v}}^{(k)}\mathbf{S}_{\boldsymbol{v}\boldsymbol{v}}\breve{\mathbf{T}}_{\boldsymbol{v}\xi}^{(k)}$ with $\breve{\mathbf{T}}_{\xi\boldsymbol{v}}^{(k)} = \mathbf{I} - \mathbf{L}_{\boldsymbol{v}\boldsymbol{\iota}}\breve{\mathbf{T}}_{\boldsymbol{\iota}\boldsymbol{v}}^{(k)}$

The derivative $\frac{\partial}{\partial \sigma_\xi^2}\log\left|\pi\sigma_\xi^2\mathbf{I}\right|$ is given by:

$$\frac{\partial}{\partial \sigma_\xi^2}\log\left|\pi\sigma_\xi^2\mathbf{I}\right| = \frac{\mathrm{p}}{\sigma_\xi^2} \qquad [7.25]$$

The derivative $\frac{\partial}{\partial \sigma_\xi^2}\log p\left(\sigma_\xi^2\right)$ is given by:

$$\frac{\partial}{\partial \sigma_\xi^2}\log p\left(\sigma_\xi^2\right) = -\frac{s_\xi}{\sigma_\xi^2} + \frac{r_\xi}{\sigma_\xi^{-4}} \qquad [7.26]$$

where $p\left(\sigma_\xi^2\right) \propto Ga\left(1/\sigma_\xi^2\,\middle|\left(s_\xi+1\right), r_\xi\right)$ with shape $\left(s_\xi+1\right)$ and rate r_ξ

From this we obtain the equation for α:

$$\mathrm{f}_{\mathscr{b}}\mathrm{m}\,\sigma_\xi^{-4}\sum_{e\in\mathbb{E}}\left(\breve{\boldsymbol{\Sigma}}_{\xi\xi}^{(k)}(e,e) + \mathbf{S}_{\xi\xi}^{(k)}(e,e)\right) - \mathrm{f}_{\mathscr{b}}\mathrm{m}\frac{\mathrm{p}}{\sigma_\xi^2} - \frac{s_\xi}{\sigma_\xi^2} + \frac{r_\xi}{\sigma_\xi^{-4}} = 0 \qquad [7.27]$$

or

$$\sigma_\xi^{2^{(k+1)}} = \left(s_\xi + \mathrm{f}_{\mathscr{b}}\mathrm{mp}\right)/\left(\mathrm{f}_{\mathscr{b}}\mathrm{m}\sum_{e\in\mathbb{E}}\left(\breve{\boldsymbol{\Sigma}}_{\xi\xi}^{(k)}(e,e) + \mathbf{S}_{\xi\xi}^{(k)}(e,e)\right) + r_\xi\right) \qquad [7.28]$$

SI-8 Spectral Structured Sparse Bayesian Learning with multiple priors (sSSBL++)

Inputs	**$\mathbf{L}_{\boldsymbol{\iota}\boldsymbol{v}}$ (Lead Field), $\boldsymbol{scale}$L (scale factor for $\mathbf{L}_{\boldsymbol{\iota}\boldsymbol{v}}$), $\{\mathbf{S}_{\boldsymbol{v}\boldsymbol{v}}(\boldsymbol{v}), \boldsymbol{v} \in \mathbb{F}_{\mathscr{b}}\}$ (Observed cross-spectrum), $\boldsymbol{scale}$S (scale factor for $\mathbf{S}_{\boldsymbol{v}\boldsymbol{v}}$), p (Number of sensors), $\mathrm{f}_{\mathscr{b}}$ (Band size), m (Sample size), D (Laplacian) N (Normal directions), $\{\mathbb{G}_{\boldsymbol{a}}, \boldsymbol{a} \in \mathbb{A}\}$ (parcels), $\boldsymbol{l}$ (Parcellation size), $\boldsymbol{s_\xi}$ (noise shape), $\boldsymbol{r_\xi}$ (noise ratio), $\boldsymbol{s_\alpha}$ (alpha shape), $\boldsymbol{r_\alpha}$ (alpha ratio), $\boldsymbol{s_\rho}$ (rho shape), $\boldsymbol{r_\rho}$ (rho shape), *maxiter* (Maximum number of iterations)**
Outputs	$\mathbf{T}_{\boldsymbol{\iota}\boldsymbol{v}}$ (Source transfer operator), $\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota}}$ (Source posterior covariance), $\{\mathbf{S}_{\boldsymbol{\iota}}(\nu), \nu \in \mathbb{F}_{\mathscr{b}}\}$ (Source covariance or cross-spectrum),
Initial values	$\alpha^{(0)}$, $\rho^{(0)}$, $\sigma_\xi^{2^{(0)}}$ and $\boldsymbol{\sigma}_{\boldsymbol{\iota}}^{2^{(0)}}$ $\mathbf{S}_{\boldsymbol{v}\boldsymbol{v}} = \frac{1}{\mathrm{f}_{\mathscr{b}}}\sum_{\nu\in\mathbb{F}_{\mathscr{b}}}\mathbf{S}_{\boldsymbol{v}\boldsymbol{v}}(\nu)$

	$\mathbf{L}_{\iota v} = \mathbf{L}_{\iota v}/scale\mathbf{L}$ $\mathbf{S}_{vv} = \mathbf{S}_{vv}/scale\mathbf{S}$
START	
$\boldsymbol{FOR\ k = 1 \cdots maxiter}$	
Parameters	
step 1	$\boldsymbol{\Sigma}_{\iota\iota}^{(k)^{-1}} = (\mathbf{DN})^{\mathcal{T}} \left(diag \left(\boldsymbol{\sigma}_{\iota}^{2^{(k)}} \right) \right)^{-1} \mathbf{DN}$
step 2	$\breve{\boldsymbol{\Sigma}}_{\iota\iota}^{(k)} = \left(\mathbf{L}_{\iota v} \left(\sigma_{\xi}^{2^{(k)}} \mathbf{I} \right)^{-1} \mathbf{L}_{v\iota} + \frac{1}{2} \boldsymbol{\Sigma}_{\iota\iota}^{(k)^{-1}} \right)^{-1}$
step 3	$\breve{\tilde{\boldsymbol{\Sigma}}}_{\iota\iota}^{(k)} = \mathbf{DN} \breve{\boldsymbol{\Sigma}}_{\iota\iota}^{(k)} (\mathbf{DN})^{\mathcal{T}}$
step 4	$\breve{\boldsymbol{\Sigma}}_{\xi\xi}^{(k)} = \mathbf{L}_{v\iota} \breve{\boldsymbol{\Sigma}}_{\iota\iota}^{(k)} \mathbf{L}_{\iota v}$
step 5	$\breve{\mathbf{T}}_{\iota v}^{(k)} = \breve{\boldsymbol{\Sigma}}_{\iota\iota}^{(k)} \mathbf{L}_{\iota v} \left(\sigma_{\xi}^{2^{(k)}} \mathbf{I} \right)^{-1}$
step 6	$\breve{\mathbf{T}}_{\xi v}^{(k)} = \mathbf{I} - \mathbf{L}_{v\iota} \breve{\mathbf{T}}_{\iota v}^{(k)}$
step 7	$\mathbf{S}_{\iota\iota}^{(k)} = \breve{\mathbf{T}}_{\iota v}^{(k)} \mathbf{S}_{vv} \breve{\mathbf{T}}_{v\iota}^{(k)}$
step 8	$\mathbf{S}_{\xi\xi}^{(k)} = \breve{\mathbf{T}}_{\xi v}^{(k)} \mathbf{S}_{vv} \breve{\mathbf{T}}_{v\xi}^{(k)}$
Hyperparameters	
$\boldsymbol{FOR\ a = 1 \cdots l}$	
step 9	$\eta_{a}^{(k+1)} = -\frac{1}{4} + \sqrt{\frac{1}{16} + \mathrm{f}_{\mathcal{b}} \mathrm{m}\ \alpha^{(k)} \rho^{(k)} \sum_{\mathcal{g} \in \mathbb{G}_a} \left(\breve{\tilde{\boldsymbol{\Sigma}}}_{\iota\iota}^{(k)} (\mathcal{g}, \mathcal{g}) + \tilde{\mathbf{S}}_{\iota\iota}^{(k)} (\mathcal{g}, \mathcal{g}) \right)}$
step 10	$\tilde{\boldsymbol{\sigma}}_{\iota}^{2^{(k+1)}}(a) = \frac{\eta_{a}^{(k+1)}}{\left(\rho^{(k)} + \eta_{a}^{(k+1)} \right)}$
step 11	$\boldsymbol{\sigma}_{\iota}^{2^{(k+1)}}(\mathbb{G}_a) = \frac{1}{2\alpha^{(k)}} \tilde{\boldsymbol{\sigma}}_{\iota}^{2^{(k)}}(a)$
END	

step 12	$\rho^{(k+1)} = zero_{\rho}\left(-\sum_{a=1}^{l}\frac{1}{1-\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^{2(k)}(a)} + l\,\frac{Ga\left(\rho\middle\vert\frac{1}{2},1\right)}{\int_{\rho}^{\infty}Ga\left(x\middle\vert\frac{1}{2},1\right)dx} + \left(s_{\rho}-\frac{l}{2}\right)\frac{1}{\rho} - r_{\rho}\right)$
step 13	$\alpha^{(k+1)} = \left(s_{\alpha}+\frac{l}{2}\right)\Big/\left(\mathrm{f}_{\mathcal{b}}\mathrm{m}\sum_{a=1}^{l}\frac{\sum_{\mathcal{g}\in\mathbb{G}_a}\left(\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\iota\iota}}^{(k)}(\mathcal{g},\mathcal{g})+\tilde{\mathbf{S}}_{\boldsymbol{\iota\iota}}^{(k)}(\mathcal{g},\mathcal{g})\right)}{\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\iota}}^{2(k)}(a)} + r_{\alpha}\right)$
step 14	$\sigma_{\xi}^{2(k+1)} = \left(s_{\xi}+\mathrm{f}_{\mathcal{b}}\mathrm{mp}\right)\Big/\left(\mathrm{f}_{\mathcal{b}}\mathrm{m}\sum_{e=1}^{\mathrm{p}}\left(\tilde{\boldsymbol{\Sigma}}_{\xi\xi}^{(k)}(e,e)+\mathbf{S}_{\xi\xi}^{(k)}(e,e)\right)+r_{\xi}\right)$
END	
Converged estimators	
step 15	$\mathbf{T}_{\boldsymbol{\iota\nu}} = \tilde{\mathbf{T}}_{\boldsymbol{\iota\nu}}^{(maxiter)}$
step 16	$\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\iota\iota}} = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\iota\iota}}^{(maxiter)}$
step 17	$\mathbf{S}_{\boldsymbol{\iota\iota}}(\nu) = \mathbf{T}_{\boldsymbol{\iota\nu}}\mathbf{S}_{\boldsymbol{\nu\nu}}(\nu)\mathbf{T}_{\boldsymbol{\nu\iota}}$ for $\nu \in \mathbb{F}_{\mathcal{b}}$
END	

Algorithm interpretation

The estimation procedure is based on maximum posterior analysis extended from [Paz-Linares et al 2017]. The parameters ρ (step 12) and α (step 13) tune the global level of sparsity by balancing the contribution of the group Hierarchical Elastic Net of step 9. The selection of specific neural causes is controlled by $\boldsymbol{\sigma}_{\boldsymbol{\iota}}^2$ (step 11), which does it for the grouped frequency components $\nu \in \mathbb{F}_{\mathcal{b}}$, $\mathbf{m}$ samples and generators into an area $\mathcal{g} \in \mathbb{G}_a$. Thus, allowing to identify the spatial signature of the band robustly and area robustly with the information of all samples.

Algorithm statistics

The sSSBL or sSSBL++ allows screening out the neural space by thresholding the posterior distribution statistic: the ratio of the posterior mean and posterior variances. After convergence the estimated source activity can be thresholded by means of an unbiased statistic: this is due to the posterior distribution of source activity $N_{\mathrm{q}}^{\mathbb{C}}\left(\boldsymbol{\iota}\middle\vert\hat{\boldsymbol{\iota}},\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\iota\iota}}\right)$ of formula [5.1], where the quantities $\hat{\boldsymbol{\iota}}$ (posterior mean) and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\iota\iota}}$ (posterior covariance) are defined as follows:

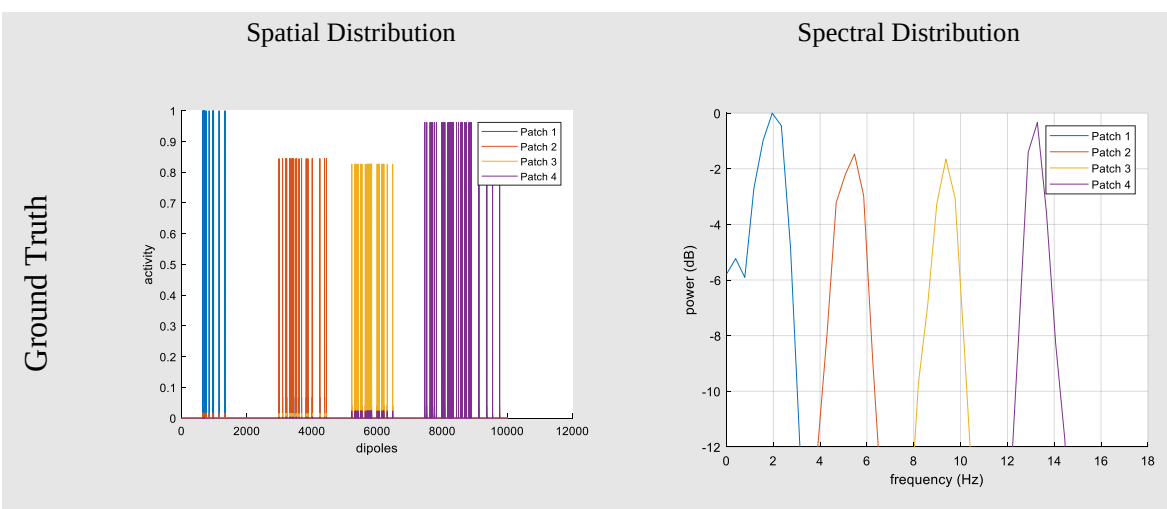
$$\hat{\boldsymbol{\iota}} = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\iota\iota}}\mathbf{L}_{\boldsymbol{\iota\nu}}\left(\sigma_{\xi}^2\mathbf{I}\right)^{-1}\boldsymbol{\nu} \qquad \text{“Posterior mean”} \qquad [5.2]$$

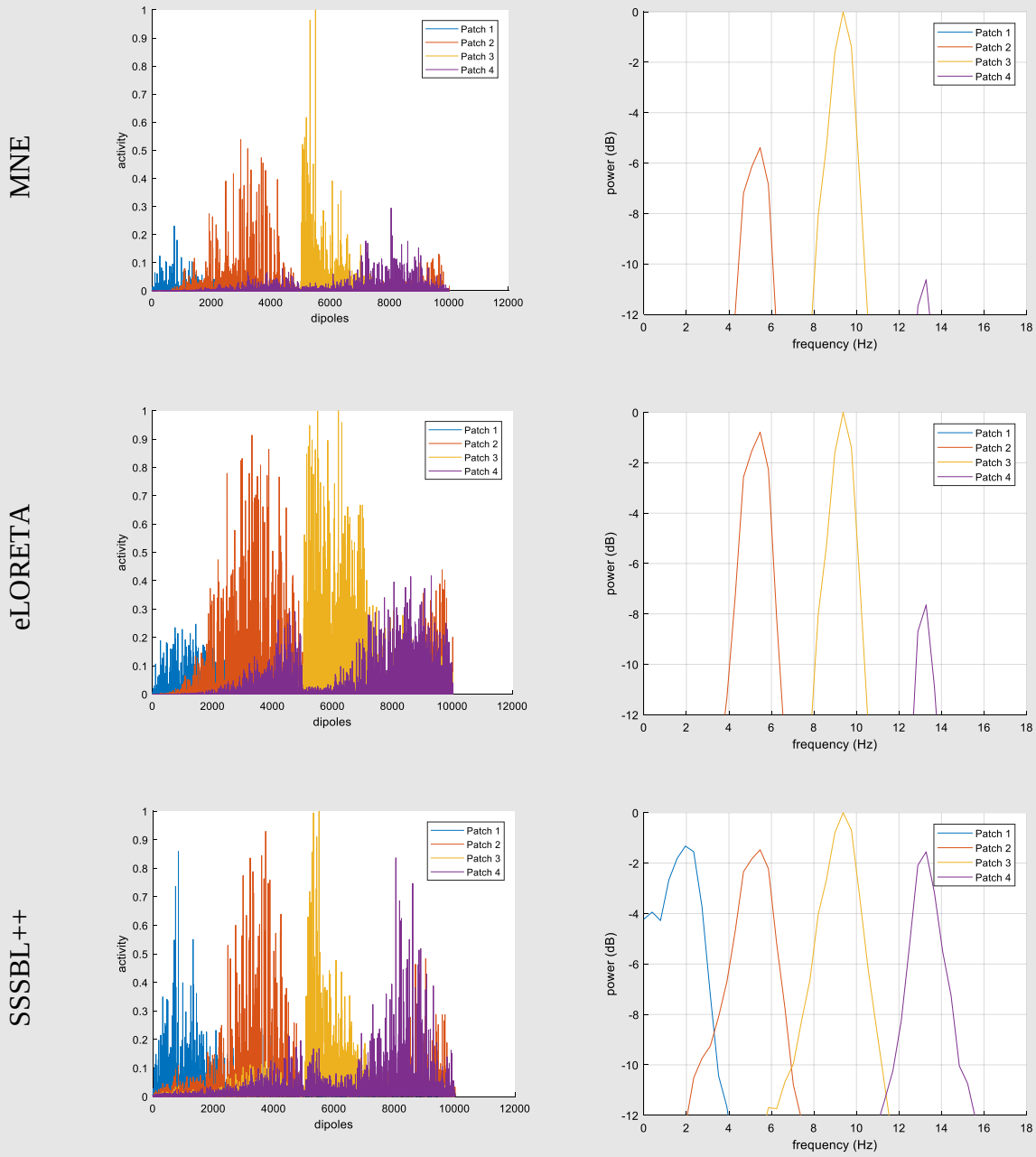
$$\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota\iota}} = \left(\mathbf{L}_{\boldsymbol{\iota v}}\left(\sigma_{\boldsymbol{\xi}}^2\mathbf{I}\right)^{-1}\mathbf{L}_{\boldsymbol{v\iota}} + \frac{1}{2}\boldsymbol{\Sigma}_{\boldsymbol{\iota\iota}}^{-1}\right)^{-1} \qquad \text{“Posterior Covariance”} \qquad [5.3]$$

In this distribution, $diag(\mathbf{S}_{\boldsymbol{\iota\iota}})$ is the posterior mean and $\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota\iota}}$ the posterior covariance. The z-statistic for the analysis of variance has the following form: $z_{stat} = sqrt\left(diag(\mathbf{S}_{\boldsymbol{\iota\iota}})./diag\left(\breve{\boldsymbol{\Sigma}}_{\boldsymbol{\iota\iota}}\right)\right)$. A plausible way to screen out the active sources is to extract the set of nodes $\boldsymbol{\mathcal{J}}$ that return a value of the z-statistic greater than 1: $\mathbb{G}_0 = \{\boldsymbol{\iota}: z_stat_{\iota} \geq 1\}$.

SI-9 Synthetic EEG

To evaluate the performance of the sSSBL we also use typical synthetic EEG data simulations ($\boldsymbol{v}_m$, m = 1…m) corresponding distributed four sources patches with four different spectral components, see an instance of the “Ground True” in next Figure. Source activity is defined as random complex vectors ($\boldsymbol{\iota}_m$, m = 1…m) multivariate Gaussian engine with random Hermitic covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\iota\iota}}$. For more realism the covariance is defined from its inverse or precision matrix $\boldsymbol{\Theta}_{\boldsymbol{\iota\iota}}$ a graphical connectivity model. Synthetic data samples ($\boldsymbol{v}_m$, m = 1…m) and its corresponding noisy empirical covariance matrix $\mathbf{S}_{\boldsymbol{vv}}$ were used to evaluate the ESI method performances. To avoid the inverse crime, we used different Lead Field for the inversion to the one used for simulations purpose. The data was corrupted with noise at both levels: (i) sensor (instrumental) noise ($\boldsymbol{\xi}_m^{(eeg)}$, m = 1…m) sampled from an univariate Gaussian engine (ii) source (biological) noise ($\boldsymbol{\xi}_m^{(biol)}$,$m$ = 1…m) obtained from the Lead Field of an univariate Gaussian engine defined as above. Both types of noise were energy normalized and scaled to a fraction of the EEG signal to achieve the specified signal-to-noise ratio coefficient. See pseudocode for generating synthetic data in next session. For comparison purpose, we use the well-stablished ESI methods MNE and eLORETA.





Results for three ESI methods (MNE, eLORETA and sSSBL++). Simulation framework was based on four patches distributed at cortical surface with different spectral activity.

For an evaluation of sSSBL++ inference framework, one-hundred random trials of patches centroids were generated. Every random trial included three patches with spatially distributed activity over thirty nodes correspondingly. The synthetic data empirical covariance matrix was obtained from four-hundred samples of simulated source activity and noise. The signal to noise ratio was adjusted to four different levels of noise: none (∞dB), low (19dB), medium (7dB), and high (0dB), as is showed in the next Table.

EMD MEASURE IN SOURCE LOCALIZATION PERFORMANCE FOR THREE ESI METHODS: MNE, ELORETA, AND SSSBL++. MEAN AND STANDARD DEVIATION ARE SHOWED FOR ONE HUNDRED TRIALS WITH FOUR DIFFERENT LEVELS OF NOISE: NONE (∞DB), LOW (19DB), MEDIUM (7DB) AND HIGH (0DB).

NOISE LEVEL \ ESI	MNE	eLORETA	SSSBL++
NONE	8.13 ± 0.64	12.01 ± 2.14	5.99 ± 0.66
LOW	14.92 ± 0.58	20.39 ± 3.37	11.22 ± 0.85
MEDIUM	14.86 ± 0.57	41.37 ± 4.81	10.91 ± 0.58
HIGH	14.89 ± 0.64	51.65 ± 2.18	10.92 ± 0.62

SI-10 Pseudocode for synthetic data generation

Inputs	**L** (Lead Field), $\mathbb{G}$ (Generator space from a discretization of the Gray Matter), $\mathbb{E}$ (Sensor space for EEG/MEG), **p** (Number of sensors), **q** (Number of sources), **m** (Sample size), $\boldsymbol{\sigma_\xi}$ (Sensors' noise)
Outputs	$\mathbf{S}_{\boldsymbol{u}}$ (Synthetic source empirical covariance matrix), $\mathbf{S}_{\boldsymbol{vv}}$ (Synthetic noisy data empirical covariance matrix)
START	
Activity	
step 1	$\Re(\boldsymbol{\Theta}_{\boldsymbol{u}}) \leftarrow randn(\mathrm{q},\mathrm{q})$ Real part in subspace $\mathbb{G}$
step 2	$\Re(\boldsymbol{\Theta}_{\boldsymbol{u}}) \leftarrow \left(\Re(\boldsymbol{\Theta}_{\boldsymbol{u}}) + \Re(\boldsymbol{\Theta}_{\boldsymbol{u}})^{\mathcal{T}}\right)/2$ Symmetric property
step 3	$\Im(\boldsymbol{\Theta}_{\boldsymbol{u}}) \leftarrow randn(\mathrm{q},\mathrm{q})$ Imaginary part at subspace $\mathbb{G}$
step 4	$\Im(\boldsymbol{\Theta}_{\boldsymbol{u}}) \leftarrow \left(\Im(\boldsymbol{\Theta}_{\boldsymbol{u}}) - \Im(\boldsymbol{\Theta}_{\boldsymbol{u}})^{\mathcal{T}}\right)/2$ Antisymmetric property
step 5	$\boldsymbol{\Theta}_{\boldsymbol{u}} \leftarrow complex\left(\Re(\boldsymbol{\Theta}_{\boldsymbol{u}}), \Im(\boldsymbol{\Theta}_{\boldsymbol{u}})\right)$ Complex compositing
step 6	Set to zero elements in $\boldsymbol{\Theta}_{\boldsymbol{u}}$ with unconnected indexes
step 7	$If\ min\left(svds(\boldsymbol{\Theta}_{\boldsymbol{u}})\right) \leq 0$ then correct singular values to be greater than zero
step 8	$(\boldsymbol{\iota}_m) \sim N_{\mathrm{q}}^{C}\left(\boldsymbol{\iota}_m \middle\vert \mathbf{0}, \boldsymbol{\Theta}_{\boldsymbol{u}}^{-1}\right);\ m = 1 \ldots \mathbf{m}$

step 9	$\mathbf{S}_{\boldsymbol{u}} \leftarrow \frac{1}{\mathrm{m}} \sum_{m=1}^{\mathrm{m}} \boldsymbol{\iota}_m \boldsymbol{\iota}_m{}^{\dagger}$
Noise	
step 10	$\boldsymbol{\xi}_m^{(inst)} \sim randn(\mathrm{p}, 1);\ m = 1 \ldots \mathbf{m}$ Instrumental noise
step 11	$\mathbf{S}_{\xi\xi}^{(inst)} \leftarrow \frac{1}{\mathrm{m}} \sum_{m=1}^{\mathrm{m}} \boldsymbol{\xi}_m^{(inst)} \boldsymbol{\xi}_m^{(inst)\dagger}$ Instrumental noise empirical covariance matrix
step 12	$\boldsymbol{\xi}_m^{(biol)} \sim randn(\mathrm{q}, 1);\ m = 1 \ldots \mathbf{m}$ Biological noise at subspace $\mathbb{G}$
step 13	$\boldsymbol{\xi}_m^{(biol)} \leftarrow \mathbf{L}\boldsymbol{\xi}_m^{(biol)};\ m = 1 \ldots \mathbf{m}$ Biological noise at subspace $\mathbb{E}$
step 14	$\mathbf{S}_{\xi\xi}^{(biol)} \leftarrow \frac{1}{\mathrm{m}} \sum_{m=1}^{\mathrm{m}} \boldsymbol{\xi}_m^{(biol)} \boldsymbol{\xi}_m^{(biol)\dagger}$ Biological noise empirical covariance matrix
Signal to noise ratio	
step 15	$\mathbf{S}_{\boldsymbol{vv}} \leftarrow \mathbf{L}\mathbf{S}_{\boldsymbol{u}}\mathbf{L}^{\mathcal{T}}$ Synthetic ideal data empirical covariance matrix
step 16	$\boldsymbol{\xi}_m^{(inst)} \leftarrow \boldsymbol{\xi}_m^{(inst)} tr(\mathbf{S}_{\boldsymbol{vv}}) / \left(\boldsymbol{\sigma}_{\xi} tr\left(\mathbf{S}_{\xi\xi}^{(inst)}\right)\right);\ m = 1 \ldots \mathbf{m}$ Instrumental noise
step 17	$\boldsymbol{\xi}_m^{(biol)} \leftarrow \boldsymbol{\xi}_m^{(biol)} tr(\mathbf{S}_{\boldsymbol{vv}}) / \left(\boldsymbol{\sigma}_{\xi} tr\left(\mathbf{S}_{\xi\xi}^{(biol)}\right)\right);\ m = 1 \ldots \mathbf{m}$ Biological noise
step 18	$\boldsymbol{\xi}_m \leftarrow \boldsymbol{\xi}_m^{(inst)} + \boldsymbol{\xi}_m^{(biol)};\ m = 1 \ldots \mathbf{m}$ Total Noise
Data	
step 19	$\boldsymbol{v}_m = \mathbf{L}\boldsymbol{\iota}_m + \boldsymbol{\xi}_m,\ m = 1 \ldots \mathbf{m}$ Synthetic Noisy Data
step 20	$\mathbf{S}_{\boldsymbol{vv}} = \frac{1}{\mathrm{m}} \sum_{m=1}^{\mathrm{m}} \boldsymbol{v}_m \boldsymbol{v}_m{}^{\dagger}$
END	