Minimizing the distortions in electrophysiological source imaging of cortical oscillatory activity *via* Spectral Structured Sparse Bayesian Learning

Deirel Paz-Linares[1,2†], Eduardo Gonzalez-Moreira[1,3,4,5†], Ariosky Areces-Gonzalez[1,6], Ying Wang[1], Min Li[1], Mayrim Vega-Hernandez[2], Qing Wang[1,7,8], Jorge Bosch-Bayard[1,7,8], Maria L. Bringas-Vega[1,2], Eduardo Martinez-Montes[2], Mitchel J. Valdes-Sosa[1,2] and Pedro A. Valdes-Sosa[1,2]*

[1]MOE Key Lab for Neuroinformation, The Clinical Hospital of Chengdu Brain Science Institute, University of Electronic Science and Technology of China, Chengdu, China, [2]Neuroinformatics Department, Cuban Neuroscience Center, Havana, Cuba, [3]Center for Biomedical Imaging and Neuromodulation, Nathan Kline Institute for Psychiatric Research, Orangeburg, NY, United States, [4]Research Unit for Neurodevelopment, Institute of Neurobiology, Autonomous University of Mexico, Querétaro, Mexico, [5]Faculty of Electrical Engineering, Central University "Marta Abreu" of Las Villas, Santa Clara, Cuba, [6]Faculty of Technical Sciences, University of Pinar del Río "Hermanos Saiz Montes de Oca", Pinar del Rio, Cuba, [7]McGill Centre for Integrative Neurosciences MCIN, Montreal Neurological Institute, McGill University, Montreal, QC, Canada, [8]Ludmer Centre for Mental Health, Montreal Neurological Institute, McGill University, Montreal, QC, Canada

Oscillatory processes at all spatial scales and on all frequencies underpin brain function. Electrophysiological Source Imaging (ESI) is the data-driven brain imaging modality that provides the inverse solutions to the source processes of the EEG, MEG, or ECoG data. This study aimed to carry out an ESI of the source cross-spectrum while controlling common distortions of the estimates. As with all ESI-related problems under realistic settings, the main obstacle we faced is a severely ill-conditioned and high-dimensional inverse problem. Therefore, we opted for Bayesian inverse solutions that posited *a priori* probabilities on the source process. Indeed, rigorously specifying both the likelihoods and *a priori* probabilities of the problem leads to the proper Bayesian inverse problem of cross-spectral matrices. These inverse solutions are our formal definition for cross-spectral ESI (cESI), which requires *a priori* of the source cross-spectrum to counter the severe ill-condition and high-dimensionality of matrices. However, inverse solutions for this problem were NP-hard to tackle or approximated within iterations with bad-conditioned matrices in the standard ESI setup. We introduce cESI with a *joint a priori* probability upon the source cross-spectrum to avoid these problems. cESI inverse solutions are low-dimensional ones for the set of random vector instances and not random matrices. We achieved cESI inverse solutions through the variational approximations *via* our Spectral Structured Sparse Bayesian Learning (ssSBL) algorithm https://github.com/CCC-members/Spectral-Structured-Sparse-Bayesian-Learning. We compared low-density EEG (10−20 system) ssSBL inverse solutions with reference cESIs for two experiments: (a) high-density MEG that were used to simulate EEG and (b) high-density macaque ECoG that were recorded simultaneously with EEG. The ssSBL resulted in two orders of magnitude with less distortion than the state-of-the-art ESI methods.

Introduction

Brain function is embodied in large-scale dynamic networks underlying all behavior and cognition. The natural modes of these networks are oscillations at functionally specific frequencies (Engel et al., 2001; Varela et al., 2001). Direct (invasive) observations of brain oscillations are increasingly more fine-grained and informative (Buzsáki et al., 2013; Frauscher et al., 2018). However, they cannot be feasibly examined in the whole brain *in vivo* human studies. Hence, considerable effort is dedicated to developing indirect (non-invasive) imaging methods to reveal the brain's oscillatory architecture with fine-grained time resolution. Notably, such an imaging modality should reveal the two different but interrelated aspects of networks: (a) the *activity* of each dynamical unit (node) and (b) their *connectivity*.

Electrophysiological Source Imaging (ESI) is a prime candidate for mapping brain network activity and connectivity. ESI is predicated upon the fact that *the Electro-encephalogram* (EEG), *Magneto-encephalogram* (MEG), or *Electrocorticogram* (ECoG) are generated by the electrical currents of macroscopic neural masses (nodes) resulting from the local mean field of post-synaptic potential activity (Jirsa and Haken, 1996, 1997; Jirsa, 2004; Deco et al., 2008; Friston, 2009; Daunizeau et al., 2010, 2011; Moran et al., 2013).

In addition, ESI has been used in attempts to estimate these currents, also known as source activity, from their electrophysiological observables. For a neural mass, the current is directly proportional to the mean field activity (Valdes-Sosa et al., 2009; Rosa et al., 2010). In turn, each neural mass is the collection of neurons to the extent of a few millimeters such that a mean-field observable (Freeman, 1975; Vinck and Perrenoud, 2019) is the "source" of the observations. In this study, we attempted the most general formulation of the problem as a reference for future work while providing concrete examples.

In principle, source activity in a network is modeled as a strictly dynamic random field $\iota(\varrho, \varsigma)$ over a continuous spatiotemporal manifold with $\varrho \in \mathbb{R}^3$ (the parts of the brain that generate observations) and continuous time $\varsigma \in \mathbb{R}$. In practice, the random field $\iota(\varrho, \varsigma)$ must be discretized at spatial points $\varrho_g; g = 1, \cdots, G$, and at discrete time points $\varsigma_t = t \triangle \varsigma; t = -T, \cdots, T$. In this scenario $\triangle \varsigma$ is the sampling period. Thus, we focused on the vector time series $\iota(t)$, defined as the vector function with entries $\iota(g, t) = \iota(\varrho_g, \varsigma_t)$. The (multi-channel) electro-physiological data is the vector time series $\nu(t)$ with entries $\nu(e, t)$ for each sensor, $e = 1 \cdots E$ that arises from the discretization of $\nu(\varrho, \varsigma)$, where the electromagnetic field was produced by $\iota(\varrho, \varsigma)$.

The data $\nu(t)$ from its latent source activity $\iota(t)$ is presented in the following forward model (Eq. 1).

$$\nu(t) = \mathbf{L}_{\nu\iota}\iota(t) + \xi(t) \tag{1}$$

Where $\mathbf{L}_{\nu\iota}$ is the real-valued lead field matrix or forward operator that projects sources $\iota(t)$ to forward the data $\nu(t)$, and $\xi(t)$ is the time series of instrumental noise assumed to be independent of the source activity $\iota(t)$. The forward operator (lead field) $\mathbf{L}_{\nu\iota}$ is linear and stationary by definition, derived from the discretization of a quasi-static electromagnetic forward model (Hämäläinen et al., 1993; Riera and Fuentes, 1998; Hallez et al., 2007; Lei et al., 2011; Piastra et al., 2020). For operators, we used suffixes that indicate the operator's codomain and domain.

ESI can be defined as the generally non-linear inverse solution (Eq. 2) (Nunez, 1974; Hämäläinen and Ilmoniemi, 1994; Nunez et al., 1994; Baillet et al., 2001; Nunez and Srinivasan, 2006; Burle et al., 2015) *via* the optimal inverse operator that we denote with a hat $\hat{\mathbf{W}}_{\iota\nu}$. An inverse operator $\hat{\mathbf{W}}_{\iota\nu}$ projects the data $\nu(t)$ to explain approximately its source and produce its estimator $\hat{\iota}(t)$.

$$\hat{\iota}(t) \leftarrow \hat{\mathbf{W}}_{\iota\nu}(\nu(t)) \tag{2}$$

Optimizing $\mathbf{W}_{\iota\nu}$ from the data solves an inverse problem, Eq. (1), that is not only ill-posed in the sense of Hadamard (Hadamard and Morse, 1953) with degeneracy in a $(G - S)$-dimensional space but also severely ill-conditioned and high dimensional (with $G \gg E$). Overcoming these challenges to obtain acceptable inverse solutions has been the subject of much research in specific optimization methods for $\mathbf{W}_{\iota\nu}$ which was well-summarized in the study of Knösche and Haueisen (2022).

Although we have defined ESI in terms of time domain signals, it is well-known that brain activity is oscillatory at all scales, from the local field potentials at the neuronal level (Freeman, 1975; Vinck and Perrenoud, 2019) to the EEG, MEG, or ECoG (Niedermeyer and da Silva, 2005; Le Van Quyen and Bragin, 2007; Frauscher et al., 2018). In tailoring ESI for oscillatory activity, a natural framework is that of the frequency domain $\iota(f)$, a random vector representing the (discrete) Fourier transform of vector time series $\iota(t)$, and comprising complex-valued entries $\iota(g, f)$ for each source $g = 1, \cdots, G$ and frequency $f = -T, \cdots, T$. Considering that we may compute the physical frequency as $\nu_f = f \triangle \nu$ for a spectral period $\triangle \nu = \frac{1}{(2T + 1) \triangle \varsigma}$ where $(2T + 1) \triangle \varsigma$ is the Nyquist frequency, the corresponding frequency domain data term is $\nu(f)$.

Then, the equivalent expressions to the previous Eqs. (1) and (2) in the frequency domain are Eq. (3), the corresponding inverse problem for the Fourier transform $\iota(f)$, and its inverse solution

leading to the estimator $\hat{\iota}(f)$. Considering that solving an inverse problem in the frequency domain should be (ideally) optimizing the inverse operator $\hat{W}_{\iota v}$ Eq. (3) from the data Fourier transform $v(f)$ (and not from the process $v(t)$).

$$
\begin{aligned}
v(f) &= L_{v\iota}\iota(f) + \xi(f) \\
\hat{\iota}(f) &\leftarrow \hat{W}_{\iota v}(v(f))
\end{aligned} \tag{3}
$$

Frequency domain descriptions of the electrophysiological observations and their inverse solutions (Eqs. 1–3) have been used fruitfully to probe behavior and cognition (Valdés et al., 1992; Engel et al., 2001; Varela et al., 2001; Le Van Quyen and Bragin, 2007; Marzetti et al., 2008; Valdes-Sosa et al., 2009; Brookes et al., 2011a,b; Faes and Nollo, 2011; Faes et al., 2012, 2017; Friston et al., 2012; Hipp et al., 2012; Colclough et al., 2015; Wens et al., 2015; Mahjoory et al., 2017; Vidaurre et al., 2018a,b; Tewarie et al., 2019; Nolte et al., 2020).

Our concern, then, is with frequency domain ESI in Eq. (3), our primary target being the source cross-spectral density matrix or source cross-spectrum $\Sigma_{\iota\iota}(f) = \langle \iota(f) \iota^{\dagger}(f)\rangle$, with the expected value over the sample space of the discrete Fourier transform $\iota(f)$ (Valdés et al., 1992; Engel et al., 2001; Varela et al., 2001; Nunez and Srinivasan, 2006; Hipp et al., 2012; Vidaurre et al., 2018b). The corresponding data cross-spectrum is $\Sigma_{vv}(f) = \langle v(f) v^{\dagger}(f)\rangle$. In practice, this later quantity must be substituted by its estimator $\overline{\Sigma}_{vv}(f)$ Eq. (4).

$$
\overline{\Sigma}_{vv}(f) = \left\langle v_m(f) v_m^{\dagger}(f); M \right\rangle = \frac{1}{M}\sum_{m=1}^{M} v_m(f) v_m^{\dagger}(f) \tag{4}
$$

$\overline{\Sigma}_{vv}(f)$ is denoted with a different hat type since the expectation is for a finite number of instances M with an index $m = 1\cdots M$. Toward this estimation, we followed the standard practice of segmenting the data time series $v(t)$ into segments $(v_m(t); \forall m)$. Thus, we worked with instances $(v_m(f); \forall m)$ of the discrete Fourier transform applied to realizations or observations for these segments.

The frequency domain inverse problem and the inverse solution for the source cross-spectrum $\Sigma_{\iota\iota}(f)$ may be stated as Eq. (5) (He et al., 2019), valid under the condition of independence between the source process $\iota(t)$ and the noise process $\xi(t)$ in the previous forward model (Eq. 1).

$$
\begin{aligned}
\Sigma_{vv}(f) &= L_{v\iota}\Sigma_{\iota\iota}(f) L_{\iota v} + \Sigma_{\xi\xi}(f) \\
\hat{\Sigma}_{\iota\iota}(f) &\leftarrow \hat{W}_{\iota v}(\overline{\Sigma}_{vv}(f))
\end{aligned} \tag{5}
$$

The pursuit of the cross-spectrum $\Sigma_{\iota\iota}(f)$ is rewarding since it completely specifies the multivariate linear properties of any stochastically driven system, be it linear or non-linear (Brillinger, 2001, 2012), though the latter requires additional higher-order kernels for a complete description (Brillinger, 1965; Brillinger and Rosenblatt, 1967). We used the asymptotic stochastic properties of the discrete Fourier transform to introduce these developments (Section Asymptotic probability theory of the Fourier transform and cross-spectrum).

Cross-spectral diagonal elements $\Sigma_{\iota\iota}(g, g, f)$ have intuitive interpretations: the variances $\sigma_{\iota\iota}^2(g, f)$ $(\sigma_{\iota\iota}^2(f) = diag(\Sigma_{\iota\iota}(f)))$ are the spectra of source activity, the *Cortical Spectral Topography*

(CST). Off-diagonal elements $\Sigma_{\iota\iota}(g, g', f)$ are the cross-spectra reflecting functional connectivity. Optimal inverse operators $\hat{W}_{\iota v}$ for the cross-spectrum $\Sigma_{\iota\iota}(f)$ Eq. (5) is a novel form of electrophysiological source imaging: *cross-spectral ESI* (cESI).

From expressions (Eqs. 1–5), the time and frequency domain variants of the inverse problem, it is evident that cESI, the inverse solution $\hat{\Sigma}_{\iota\iota}(f)$ of the source cross-spectrum $\Sigma_{\iota\iota}(f)$, may be obtained from three different types of primary information: $v(t)$, or $v(f)$, or $\overline{\Sigma}_{vv}(f)$. Initially, it might seem that an inverse solution $\hat{\Sigma}_{\iota\iota}(f)$ from any of these data types would be equivalent. However, as shown further into the study, each data type requires optimizing its specific inverse operator $\hat{W}_{\iota v}$, thus defining different estimation "routes" to cross-spectrum $\Sigma_{\iota\iota}(f)$.

We describe in detail the theory, benefits, and problems of each route, in terms of the forward operator $L_{v\iota}$ (Section Forward projection by the lead field linear and stationary operator), and the inverse operator $W_{\iota v}$ (Section Bayesian (MAP) inverse operators. Quasilinear (F-invariant) approximation).

In order to explore these routes in the following sections (Sections Bayesian MAP1 inverse operators. MNE, eLORETA, and LCMV as particular cases and Data used for the validation), we need to select an inverse solution framework, of which there are many potentially valuable approaches (Knösche and Haueisen, 2022). We adopted the Bayesian *maximum a posteriori* (MAP) probability approach (MacKay, 2003). The MAPs (for each route) are derived from finding the latent variables $\mathcal{X}$ that maximize the *posterior* probability $q(\mathcal{X}|\mathcal{Y})$(Eq. 6).

$$
q(\mathcal{X}|\mathcal{Y}) \propto p(\mathcal{Y}|\mathcal{X}) p(\mathcal{X}) \tag{6}
$$

Where $p(\mathcal{Y}|\mathcal{X})$ is the *likelihood* of the data $\mathcal{Y}$ conditional on the latent variable $\mathcal{X}$. In our context, $\mathcal{Y}$ can corresponding to be $v(t)$, or $v(f)$, or $\overline{\Sigma}_{vv}(f)$, and $\mathcal{X}$ can be corresponding to $\iota(t)$, or $\iota(f)$, or $\Sigma_{\iota\iota}(f)$ (Grave de Peralta Menendez et al., 2004; Friston et al., 2006; Mattout et al., 2006; Friston K. J. et al., 2008; Wipf and Nagarajan, 2009). The term $p(\mathcal{X})$ is the respective *a priori* probability upon the latent variable $\iota(t)$, or $\iota(f)$, or $\Sigma_{\iota\iota}(f)$.

As can be observed in the following section (Section Data used for the validation), the third route inverse solution (Eq. 5) has desirable properties for cESI. However, if not impossible in a realistic cESI setup, such inverse solutions are N-P hard and therefore require Approximated Bayesian Computation (ABC) (Csilléry et al., 2010). This issue has been discussed in detail in the Bayesian literature (Dempster et al., 1977; Liu and Rubin, 1994; Daunizeau and Friston, 2007; Friston et al., 2007; Nummenmaa et al., 2007; Friston K. J. et al., 2008; Paz-Linares et al., 2018).

In this study, we took a more practical path. Rather than attempting to use complicated ABCs to obtain more general solutions, the cESI rationale we focused on is to restrict Bayesian inverse operators $W_{\iota v}$ in Eqs. (1–5) within the "quasilinear" class $W_{\iota v} \cong T_{\iota v}$. Quasilinear inverse operators are also known in a more general context as the linear proximal operators or as linear back projectors that solve non-linear optimization or inverse problems (Kaplan and Tichatschke, 1998; Piotrowski and Yamada, 2008; Gramfort et al., 2012; Tirer and Giryes, 2020).

Here, quasilinear inverse operator $T_{v\iota}$, which holds the same linearity attribute as the forward operator $L_{v\iota}$, produced cESI that preserved the amplitude and phase information in

the frequency domain (Mantini et al., 2007; Marzetti et al., 2008; Brookes et al., 2011b; Hipp et al., 2012; Lopes da Silva, 2013; He et al., 2019; Nolte et al., 2020). That is, we ensured that $\mathbf{T}_{\nu}$ possesses the attribute of what we call "F-invariance" for essential properties (Brillinger, 2001, 2012).

An immediate consequence of taking $\mathbf{W}_{\iota\nu} \cong \mathbf{T}_{\iota\nu}$, a quasilinear approximation of the inverse operator in the third cESI route (Eq. 5), is the following representation of source cross-spectrum $\mathbf{\Sigma}_{\iota\iota}(f) \cong \overline{\mathbf{\Sigma}}_{\iota\iota}(f)$ (Eq. 7). Here, we considered $\overline{\mathbf{\Sigma}}_{\iota\iota}(f)$ to be calculated from the set of random instances $(\iota_m(f); \forall m)$ of the latent vector process $\iota(f)$.

$$\overline{\mathbf{\Sigma}}_{\iota\iota}(f) = \left\langle \iota_m(f)\,\iota_m^{\dagger}(f); M \right\rangle = \frac{1}{M}\sum_{m=1}^{M} \iota_m(f)\,\iota_m^{\dagger}(f) \qquad (7)$$

Once $\mathbf{\Sigma}_{\iota\iota}(f) \cong \overline{\mathbf{\Sigma}}_{\iota\iota}(f)$ was assumed, the ABCs were simplified dramatically since a MAP for the matrix $\overline{\mathbf{\Sigma}}_{\iota\iota}(f)$ (Eq. 7) turns into a joint-MAP (Section Validation rationale) for the random vectors $(\iota_m(f); \forall m)$ that are implicit in $\overline{\mathbf{\Sigma}}_{\iota\iota}(f)$ (Hsiao et al., 1998, 2002; Yeredor, 2000; Davis et al., 2001; Auranen et al., 2005; Wipf and Nagarajan, 2009; Chen et al., 2011).

We implemenedt the joint-MAP *via* Spectral Structured Sparse Bayesian Learning (ssSBL), the type of ABC developed in Section Measures of distortion. As shown under realistic inverse problem settings (Section Results), the third cESI route (Eq. 5) implemented *via* the ssSBL approach leads to less distorted estimates than the traditional methods for the first and second cESI routes (Eqs. 1–3). To judge distortions, we employed the well-known ESI methods as a baseline: *Exact Low-Resolution Electromagnetic Tomographic Analysis* (ELORETA) (Pascual-Marqui et al., 2006) and *Linearly Constrained Minimum Variance* (LCMV) (Van Veen et al., 1997).

The theoretical framework allowed one to consider the fundamental problem of ESI distortions. Indeed, significant distortions are expected with any state-of-the-art inverse solutions in a realistic ESI setup. The distortions, which we explore later, are localization error and leakage (blurring). These distortions are pervasive comparing simulated topographic vectors, say $\iota(t)$ or $\iota(f)$, vs. their inverse solution $\hat{\iota}(t)$ or $\hat{\iota}(f)$ (Kobayashi et al., 2003; Grova et al., 2008; Schoffelen and Gross, 2009; Haufe et al., 2013; Burle et al., 2015; Colclough et al., 2015; Bradley et al., 2016; Mahjoory et al., 2017; Stokes and Purdon, 2017; He et al., 2018, 2019; Palva et al., 2018; Haufe and Ewald, 2019; Marinazzo et al., 2019).

As we will show in Section Discussion, the topographic distortions of the inverse solutions for a random vector $\iota(t)$ or $\iota(f)$ can reach unacceptable levels for second-order statistics, such as the sample estimator for the cross-spectrum $\overline{\mathbf{\Sigma}}_{\hat{\iota}\hat{\iota}}(f)$ calculated from an inverse solution $\hat{\iota}(t)$ or $\hat{\iota}(f)$. Minimizing CST distortions (for the estimator of the spectrum $\sigma_{\iota\iota}^2(f)$) benefits the overall cross-spectral estimation (for $\mathbf{\Sigma}_{\iota\iota}(f)$). Our results suggest that that ssSBL significantly reduces these distortions.

Standard cESI theory

Asymptotic probability theory of the Fourier transform and cross-spectrum

The material in this section (with minor differences in notation) is described in greater detail in Brillinger (2012). For an exhaustive definition of terms, refer to Supplementary material (SD, Section Introduction). Our primary interest will be in frequency domain quantities. Let $\mathbf{x}(t)$ be an R-dimensional vector time series. We worked with the following definitions:

- The *vector stochastic process* $\mathbf{x}(f)$ (Eq. 8) is defined in the discrete frequency domain $\nu_f = f \triangle \nu$ with $f = -T, \cdots, T$, as the discrete Fourier transform of the vector time series $\mathbf{x}(t)$.

$$\mathbf{x}(f) = \sum_{t=-T}^{T} \mathbf{x}(t)\, e^{-i2\pi(f \triangle \nu)(t \triangle \varsigma)} \qquad (8)$$

- The *cross-spectral density matrix* or *cross-spectrum* $\mathbf{\Sigma}_{\mathbf{xx}}(\nu)$ (Eq. 9) was defined in the frequency domain ν as the Fourier transform of the *auto-covariance matrix* $\mathbf{\Sigma}_{\mathbf{xx}}(\tau)$.

$$\mathbf{\Sigma}_{\mathbf{xx}}(\nu) = \sum_{\tau=-\infty}^{+\infty} \mathbf{\Sigma}_{\mathbf{xx}}(\tau)\, e^{-i2\pi\nu(\tau \triangle \varsigma)} \qquad (9)$$

Here, the *auto-covariance* $\mathbf{\Sigma}_{\mathbf{xx}}(\tau) = \left\langle \mathbf{x}(t)\, \mathbf{x}^{\dagger}(t+\tau) \right\rangle$ depends on the time-lag τ and does not vary with time t, thus being second-order stationary. We will furthermore assume (for technical reasons) that $\mathbf{x}(t)$ is a *strictly stationary vector time series* where all moments are also translation invariants. We also assumed that the strong mixing condition holds. This condition was due to the rapid decrease in the magnitude of the autocovariance $\mathbf{\Sigma}_{\mathbf{xx}}(\tau)$, and all higher-order moments as the time-lag τ increases.

A fundamental result on which we based our work is Theorem 4.1.1 of Brillinger (2001), which can be understood as the equivalent for Fourier coefficients of the central limit theorem under stationarity and strong mixing conditions (Rosenblatt, 1956). In our notation, this theorem states:

"Assume that the number of time points in the discrete Fourier transform (Eq. 8), goes to the infinity ($T \rightarrow +\infty$), and let the sampling period go to zero ($\triangle \varsigma \rightarrow 0+$) so that the spectral resolution $\triangle \nu = \frac{1}{((2\,T+1)\,\triangle\varsigma)} = constant$ holds constant. Then, it holds that $\mathbf{x}(f)$ is asymptotically independent for all f and converges in probability to the *circularly symmetric multivariate complex-valued Gaussian probability density* (Eq. 10). Where the Hermitian covariance matrix $\mathbf{\Sigma}_{\mathbf{xx}}(f)$ is the cross-spectrum (Eq. 9) at the frequencies $\nu_f = f \triangle \nu$ with $f = -T, \cdots, T$."

$$N^{\mathbb{C}}\left(\mathbf{x}(f)\big|0, \mathbf{\Sigma}_{\mathbf{xx}}(f)\right) = \frac{1}{\left|\pi\, \mathbf{\Sigma}_{\mathbf{xx}}(f)\right|} e^{-\mathbf{x}^{\dagger}(f)\mathbf{\Sigma}_{\mathbf{xx}}^{-1}(f)\mathbf{x}(f)} \qquad (10)$$

We emphasize that this Gaussian distribution asymptotic distribution not only holds for $\mathbf{x}(f)$ but also for time-varying

estimators such as the time-windowed discrete Fourier transform or the Hilbert transform $\mathbf{x}(t, f)$ (Bruns, 2004). The latter has been widely used in the literature (Faes and Nollo, 2011; Friston et al., 2012; Nolte et al., 2020).

It is important to note that Gaussianity might not be valid for the original time series $\mathbf{x}(t)$ or even one of its band-filtered versions, as many assume in the literature (Grave de Peralta Menendez et al., 2004; Friston et al., 2006; Mattout et al., 2006; Friston K. J. et al., 2008; Wipf and Nagarajan, 2009; Paz-Linares et al., 2017). In fact, even in the case of non-Gaussian, non-linear time series Brillinger's theorem is valid as long as stationarity and strong mixing hold. This validity does not imply that the spectral density matrix $\mathbf{\Sigma_{xx}}(f)$ completely characterizes the non-linear or non-Gaussian system. Cumulant information of orders higher than two may be necessary for a complete system description (Brillinger, 1965; Brillinger and Rosenblatt, 1967).

Brillinger's theorem leads to the probability density of any sampled estimator of the cross-spectrum. In particular, it applies to $\overline{\mathbf{\Sigma}}_{\mathbf{xx}}(f) = \left\langle \mathbf{x}_m(f) \ \mathbf{x}_m^\dagger(f) \ ; \ M \right\rangle$ the sampled estimator for Fourier transform instances $(\mathbf{x}_m(f) \ ; \ \forall m)$ with sample size M. Here, these instances $(\mathbf{x}_m(f) \ ; \ \forall m)$ represent the discrete Fourier transform applied to sample realizations $(\mathbf{x}_m(t) \ ; \ \forall m)$ obtained from time segments of the observations. Then, it follows that a Hermitian Wishart $W^{\mathbb{C}}$ (Eq. 11), with R-dimensional scale matrix $\mathbf{\Sigma_{xx}}(f)$ (cross-spectrum), and degree of freedom M (with $M \geqslant R$), is the probability density of the estimator $\overline{\mathbf{\Sigma}}_{\mathbf{xx}}(f)$.

$$W^{\mathbb{C}}\left(\overline{\mathbf{\Sigma}}_{\mathbf{xx}}(f)\big|\mathbf{\Sigma_{xx}}(f), M\right) \propto \frac{\left|\overline{\mathbf{\Sigma}}_{\mathbf{xx}}(f)\right|^{M-R}}{\left|\mathbf{\Sigma_{xx}}(f)\right|^M} e^{-Mtr\left(\mathbf{\Sigma_{xx}^{-1}}(f)\overline{\mathbf{\Sigma}}_{\mathbf{xx}}(f)\right)} \quad (11)$$

Since we based our further developments on the assumption of Gaussianity (Eqs. 10, 11), we carried out the statistical test for the distribution of the discrete Fourier transform $\mathbf{x}(f)$ (Eq. 10) with two examples of resting state sensor data: MEG data from the Human Connectome Project (HCP) (Van Essen et al., 2013) and Macaque ECoG data from the Neurotycho project (Nagasaka et al., 2011). The outcome of this test for data did not allow us to reject the hypothesis of Gaussianity, which then was also plausible for the sources.

Forward projection by the lead field linear and stationary operator

We emphasize that linearity and stationarity are essential "conservative" attributes of the operators for our target (cESI). We note, this is the reason why one can state all the forward "routes" in terms of the same operator $\mathbf{L}_{v\iota}$ (Eq. 12). Preserving the Gaussianity of the Fourier transform $\iota(f)$ is only possible under the linear forward operator $\mathbf{L}_{v\iota}$ and subsequent linear or quasilinear inverse operator $\mathbf{T}_{\iota v}$ (Marzetti et al., 2008; He et al., 2019; Nolte et al., 2020).

$$\begin{aligned} (route\ 1)\ \boldsymbol{v}(t) &= \mathbf{L}_{v\iota}\iota(t) + \boldsymbol{\xi}(t) \\ (route\ 2)\ \boldsymbol{v}(f) &= \mathbf{L}_{v\iota}\iota(f) + \boldsymbol{\xi}(f) \\ (route\ 3)\ \mathbf{\Sigma}_{vv}(f) &= \mathbf{L}_{v\iota}\mathbf{\Sigma}_{\iota\iota}(f)\mathbf{L}_{\iota v} + \mathbf{\Sigma}_{\xi\xi}(f) \end{aligned} \quad (12)$$

Furthermore, both attributes are crucial to avoid non-linear warping and delays of the amplitude and phase information in the

frequency domain for cESI (Reid et al., 2019). Therefore, we define operators with these properties for the frequency domain as "F-invariant." Though we restricted our attention later to F-invariant operators $\mathbf{T}_{\iota v}$, non-linear operators $\mathbf{W}_{\iota v}$ have been useful in other contexts (Picton and Hillyard, 1974; Picton et al., 1974; Lopes da Silva et al., 1991; Clark et al., 1994; Makeig et al., 1999, 2004; Makeig, 2002; Eichele et al., 2005; Harrison et al., 2008; Vega-Hernández et al., 2008; Maurer and Dierks, 2012).

Noteworthily, distinguishing forward "routes" leads to variants of the inverse problems or inverse operators $\mathbf{W}_{\iota v}$ for estimating the specific latent variable $\iota(t)$, or $\iota(f)$, or $\mathbf{\Sigma}_{\iota\iota}(f)$, as discussed in the next section (Section Bayesian (MAP) inverse operators. Quasilinear (F-invariant) approximation). Estimation of the cross-spectrum $\mathbf{\Sigma}_{\iota\iota}(f)$, involves the challenging inverse problem of the matrix equation (route 3) detailed in section (Section Bayesian MAP1 inverse operators. MNE, eLORETA, and LCMV as particular cases).

We currently illustrate the effects of the forward operator with the topographic maps: *Cortical Spectral Topography* (CST), a map of the cortical spectrum $\overline{\sigma}_{\iota\iota}^2(f) = diag\left(\overline{\mathbf{\Sigma}}_{\iota\iota}(f)\right)$, and *Sensor Spectral Topography* (SST), a map of the sensor spectrum $\overline{\sigma}_{vv}^2(f) = diag\left(\overline{\mathbf{\Sigma}}_{vv}(f)\right)$ (Figure 1). On the left is a hypothetical CST, and on the right is the corresponding SST. The inverse problem consists in estimating the latent CST from observed SST.

Bayesian (MAP) inverse operators, quasilinear (F-invariant) approximation

From the theory of inverse problems (Tarantola, 2005), one may seek inverse solutions for each of the latent variables in the routes corresponding to $\iota(t)$, or $\iota(f)$, or $\mathbf{\Sigma}_{\iota\iota}(f)$ (Eq. 13) and Figure 2. In each route, the theoretical inverse operator $\hat{\mathbf{W}}_{\iota v}$ represents the symbolic projection to the source space and these inverse solutions. In turn, inverse operators are determined (analytically or numerically) by data-driven methods that depend on the routes from the data variable: $\boldsymbol{v}(t)$, or $\boldsymbol{v}(f)$, or $\overline{\mathbf{\Sigma}}_{vv}(f)$, and the forward operator $\mathbf{L}_{v\iota}$.

$$\begin{aligned} (route\ 1)\ &\hat{\mathbf{W}}_{\iota v}(\boldsymbol{v}(t)) \\ (route\ 2)\ &\hat{\iota}(f) \leftarrow \hat{\mathbf{W}}_{\iota v}(\boldsymbol{v}(f)) \\ (route\ 3)\ &\hat{\mathbf{\Sigma}}_{\iota\iota}(f) \leftarrow \hat{\mathbf{W}}_{\iota v}\left(\overline{\mathbf{\Sigma}}_{vv}(f)\right) \end{aligned} \quad (13)$$

Let any of the data variables be represented generically by $\mathcal{Y}$, and likewise, $\mathcal{X}$ represent all types of latent variables in Eq. (13). A *maximum a posteriori* (*MAP*) Bayesian inverse operator $\hat{\mathbf{W}}_{\mathcal{X}\mathcal{Y}}$ (Eq. 14) achieves the maximum of *a posteriori* probability $q(\mathcal{X}|\mathcal{Y})$. We define the *a posteriori* $q(\mathcal{X}|\mathcal{Y})$ as the conditional probability determined from the *likelihood* $p(\mathcal{Y}|\mathcal{X})$ and *a priori* $p(\mathcal{X})$. Depending on $q(\mathcal{X}|\mathcal{Y})$ an inverse operator $\hat{\mathbf{W}}_{\mathcal{X}\mathcal{Y}}$ can be non-linear or linear, computed numerically or analytically, also intractable or tractable.

$$\begin{aligned} \mathcal{X} &\leftarrow \hat{\mathbf{W}}_{\mathcal{X}\mathcal{Y}}(\mathcal{Y}) \\ \hat{\mathbf{W}}_{\mathcal{X}\mathcal{Y}} &= argmax_{\mathbf{W}_{\mathcal{X}\mathcal{Y}}}\left(q\left(\mathbf{W}_{\mathcal{X}\mathcal{Y}}(\mathcal{Y})\big|\mathcal{Y}\right)\right) \\ q(\mathcal{X}|\mathcal{Y}) &\propto p(\mathcal{Y}|\mathcal{X})\,p(\mathcal{X}) \end{aligned} \quad (14)$$

The essential role of the *a priori* $p(\mathcal{X})$ is to overcome the ill condition of the *likelihood* $p(\mathcal{Y}|\mathcal{X})$, in other words, to provide a
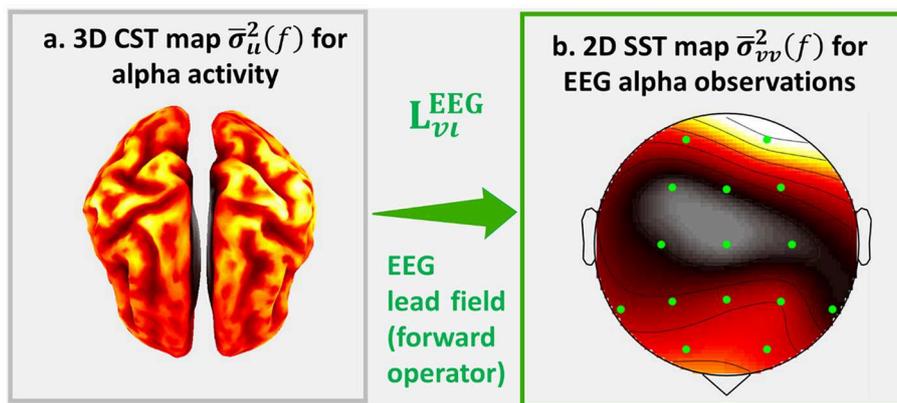
FIGURE 1
An illustration of $\boldsymbol{\Sigma}_{u}\left(f\right)$ the source cross-spectrum of human cortical alpha activity and its EEG cross-spectrum $\boldsymbol{\Sigma}_{vv}\left(f\right)$. For this illustrations we calculate the sampled estimates $\overline{\boldsymbol{\Sigma}}_{u}(f)$ and $\overline{\boldsymbol{\Sigma}}_{vv}\left(f\right)$ of which we represent the topographic projections: **(a)** Cortical Spectral Topography (CST) $\overline{\sigma}_{u}^{2}\left(f\right)=\overline{\boldsymbol{\Sigma}}_{u}\left(f\right)$ and **(b)** Sensor Spectral Topography (SST) $\overline{\sigma}_{vv}^{2}\left(f\right)=\overline{\boldsymbol{\Sigma}}_{vv}\left(f\right)$.
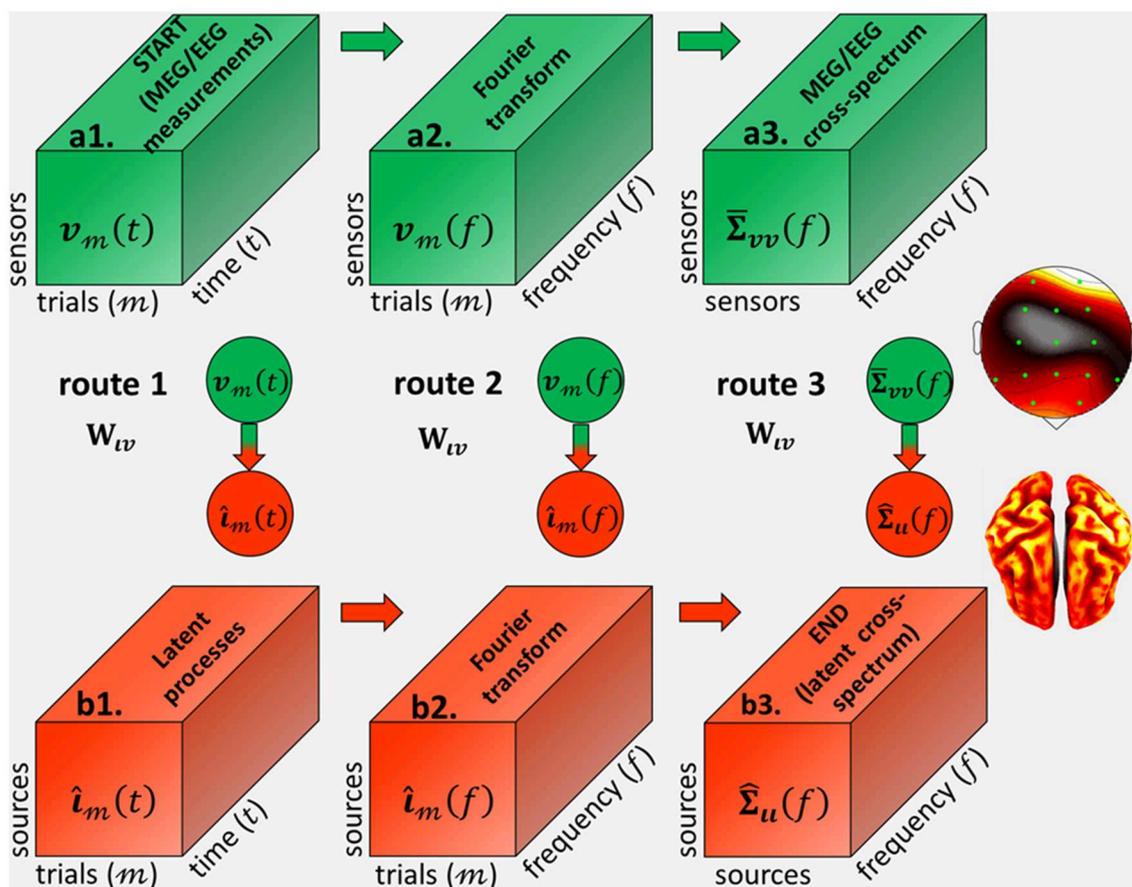
FIGURE 2
Inverse operators from **(a1)** the MEG/EEG/ECoG sensor signal $\boldsymbol{v}_{t}\left(t\right)$, defined as 3D tensor with every trial and time-point observation, to **(b3)** the source cross-spectrum at every frequency $\boldsymbol{\Sigma}_{u}\left(f\right)$, *via* the different *cross-spectral Electrophysiological Source Imaging* (cESI) routes. Route 1 (**a1**→ **b1**→ **b2**→ **b3**) *via* an inverse operator $W_{\iota v}$ in the time-domain first determines **(b1)** the source processes $\iota_{3}(t)$ and then compute **(b2)** their Fourier transform $\iota_{3}(f)$. Route 2 (**a1**→ **a2**→ **b2**→ **b3**) first computes **(a2)** the data Fourier transform $\boldsymbol{v}_{m}\left(f\right)$ and then *via* an inverse operator $W_{\iota v}$ in the frequency domain determines **(b2)** their source Fourier transform $\iota_{m}(f)$. Route 3 (**a1**→ **a2**→ **a3**→ **b3**) first computes **(a2)** the data Fourier transform $\boldsymbol{v}_{3}\left(f\right)$ and then determines **(a3)** the data cross-spectrum $\boldsymbol{\Sigma}_{vv}\left(f\right)$. We revindicate Route 3, which is *via* an inverse operator $W_{\iota v}$ in the spectral-domain with a priori probabilities upon the source cross-spectrum.

unique solution to the inverse problem for the source variable $\mathcal{X}$. Here we formulate this *a priori* $p(\mathcal{X})$ as a Gibbs probability density (Eq. 15).

$$p(\mathcal{X}) \propto \exp(H(\mathcal{X})|\alpha), \qquad (15)$$

Where α is the Gibbs temperature parameter. The Gibbs energy function $H(\mathcal{X})$ is commonly defined using the norms of vectors or matrices (Petersen and Pedersen, 2008; Golub and Van Loan, 2013), usually used to reflect empirical criteria about the structure and density of the source variable $\mathcal{X}$ defined over the spatial, time, or frequency domains. Assimilating these source qualities requires, in addition, the empirically determined (data-driven) scale parameter α, by some method from the data variable $\mathcal{Y}$.

We emphasize that this Bayesian MAP formalism is particularly helpful in developing the cESI routes and our optimal inverse solution in the following sections (Sections Bayesian MAP1 inverse operators. MNE, eLORETA, and LCMV as particular cases and Data used for the validation). In another context, Eqs. (14, 15) may be completely equivalent to the classical Tikhonov regularization (Tikhonov and Arsenin, 1977), taking *a posteriori* $p(\mathcal{X}|\mathcal{Y})$ under the $-\log$ transformation. However, the Bayesian formalism is more general than the Tikhonov regularization since the Gibbs energy $H(\mathcal{X})$ (Eq. 15) describes the statistical properties of any physical system (Landau and Lifshitz, 1980).

As stated before, we constrained all routes (Eq. 13) to the class of quasilinear inverse operators $\mathbf{T}_{\iota v}$ (Eq. 16) (Baillet et al., 2001; Grech et al., 2008). Moreover, the quasilinear class leads to the type of F-invariant inverse solutions for the properties in Eqs. (10, 11), and particularly the tractable cESI route 3.

$$
\begin{aligned}
&(route\ 1)\ \hat{\iota}(t) \leftarrow \hat{\mathbf{W}}_{\iota v}(\mathbf{v}(t)) \cong \hat{\mathbf{T}}_{\iota v}(\mathbf{v}(t))\,\mathbf{v}(t) \\
&(route\ 2)\ \hat{\iota}(f) \leftarrow \hat{\mathbf{W}}_{\iota v}(\mathbf{v}(f)) \cong \hat{\mathbf{T}}_{\iota v}(\mathbf{v}(f))\,\mathbf{v}(f) \\
&(route\ 3)\ \hat{\mathbf{\Sigma}}_{\iota\iota}(f) \leftarrow \hat{\mathbf{W}}_{\iota v}(\overline{\mathbf{\Sigma}}_{vv}(f)) \cong \hat{\mathbf{T}}_{\iota v}(\overline{\mathbf{\Sigma}}_{vv}(f)) \\
&\qquad\qquad \overline{\mathbf{\Sigma}}_{vv}(f)\,\hat{\mathbf{T}}_{v\iota}(\overline{\mathbf{\Sigma}}_{vv}(f))
\end{aligned}
\qquad (16)
$$

The term "quasilinear" $\mathbf{T}_{\iota v}$ is applies to inverse operators defined as non-linear matrix functions of the vector argument $\mathbf{v}(t)$ or $\mathbf{v}(f)$, or the matrix argument $\overline{\mathbf{\Sigma}}_{vv}(f)$. A matrix function comprises entries $\mathbf{T}_{\iota v}(g, s)$ in the $G \times E$ cartesian product of sources $g = 1 \cdots G$ and sensors $e = 1 \cdots E$. These entries are non-linear functions of the vector arguments $\mathbf{T}_{\iota v}(g, s, \mathbf{v}(t))$, or $\mathbf{T}_{\iota v}(g, s, \mathbf{v}(f))$, or the matrix argument $\mathbf{T}_{\iota v}(g, s, \overline{\mathbf{\Sigma}}_{vv}(f))$.

Bayesian MAP1 inverse operators. MNE, eLORETA, and LCMV as particular cases

Inverse solutions for routes 1 and 2 (Eq. 13) are the traditional *activation Electrophysiological Source Imaging* (aESI). We obtained these inverse solutions as a *first-type MAP* (MAP1) (Eq. 17), which estimates the source variable $\iota(t)$ in the time domain (Hämäläinen and Ilmoniemi, 1994), or the source variable $\iota(f)$ in the frequency

domain (Salmelin and Hämäläinen, 1995).

$$
\begin{aligned}
&(route\ 1)\ \hat{\iota}(t) \leftarrow \hat{\mathbf{W}}_{\iota v}(\mathbf{v}(t)) \\
&\hat{\mathbf{W}}_{\iota v} = argmax_{\mathbf{W}_{\iota v}}\, q(\mathbf{W}_{\iota v}(\mathbf{v}(t))|\mathbf{v}(t)) \\
&q(\iota(t)|\mathbf{v}(t)) \propto q(\mathbf{v}(t)|\iota(t))\, p(\iota(t)) \\
&(route\ 2)\ \hat{\iota}(f) \leftarrow \hat{\mathbf{W}}_{\iota v}(\mathbf{v}(f)) \\
&\hat{\mathbf{W}}_{\iota v} = argmax_{\mathbf{W}_{\iota v}}\, q(\mathbf{W}_{\iota v}(\mathbf{v}(f))|\mathbf{v}(f)) \\
&q(\iota(f)|\mathbf{v}(f)) \propto p(\mathbf{v}(f)|\iota(f))\, p(\iota(f))
\end{aligned}
\qquad (17)
$$

The *a posteriori* probabilities require definitions of the *likelihood,* and the *a priori* is given below in Eq. (18). For the data $\mathbf{v}(t)$ a real-valued Gaussian is commonly assumed with mean $\mathbf{L}_{v\iota}\iota(t)$ and noise covariance $\mathbf{\Sigma}_{\xi\xi}(t)$. This assumption might not be valid for all types of time domain data. In contrast, and in virtue of Brillinger's theorem cited above, data $\mathbf{v}(f)$ obtained with the discrete Fourier transform will almost certainly have a complex-valued Gaussian *likelihood* with mean $\mathbf{L}_{v\iota}\iota(f)$ and noise cross-spectrum $\mathbf{\Sigma}_{\xi\xi}(f)$ and thus the one we adopted. The *a priori* probabilities in Eq. (18) are placed upon the activation Gibbs energy in the time domain $H(\iota(t))$ or the frequency domain $H(\iota(f))$ (Eq. 18) as expressed by the vector *p-norm* (frequently *p-norm^p*) upon real-valued $\iota(t)$ or complex-valued $\iota(f)$ (Riesz, 1910; Rudin, 1970; Dunford and Schwartz, 1988; Bourbaki, 2013).

$$
\begin{aligned}
&(route\ 1)\ p(\mathbf{v}(t)|\iota(t)) = N^{\mathbb{R}}(\mathbf{v}(t)|\mathbf{L}_{v\iota}\iota(t), \mathbf{\Sigma}_{\xi\xi}(t)) \\
&\qquad p(\iota(t)) \propto \exp(H(\iota(t))|\alpha(t)) \\
&(route\ 2)\ p(\mathbf{v}(f)|\iota(f)) = N^{\mathbb{C}}(\mathbf{v}(f)|\mathbf{L}_{v\iota}\iota(f), \mathbf{\Sigma}_{\xi\xi}(f)) \\
&\qquad p(\iota(f)) \propto \exp(H(\iota(f))|\alpha(f))
\end{aligned}
\qquad (18)
$$

Henceforth, we ruled out route 1 since it is based on the *ad-hoc* Gaussian assumption for $p(\mathbf{v}(t)|\iota(t))$, which, as mentioned before, is not always tenable. Furthermore, our interest was in the frequency domain, which concentrates our attention on route 2, and which bases the *likelihood* of the Fourier transform $p(\mathbf{v}(f)|\iota(f))$ (Eq. 18) on the complex-valued Gaussian probability (Eq. 10).

Toward cESI, selecting *a priori* $p(\iota(f))$ (Eq. 18) follows the standard rationale that oscillatory brain networks and their activity are characterized by a large-scale and dense (or non-sparse) distribution in conditions of resting-state or task in a block design (Mantini et al., 2007; Brookes et al., 2011b; Hipp et al., 2012; Lopes da Silva, 2013). Such activity is the stochastic and stationary process $\iota(t)$ that is composited by oscillations in the frequency domain, as described by the Fourier transform $\iota(f)$ (Engel et al., 2001; Varela et al., 2001; Vidaurre et al., 2018a,b; Tewarie et al., 2019).

A general smooth *a priori* model posits the following Gibbs energy $H_{2,\mathbf{A}(f)}(\iota(f))$ (Eq. 19) which, in addition, specifies $\mathbf{A}(f)$ as some positive definite and symmetric, or Hermitian, matrix. Specifying the matrix $\mathbf{A}(f)$ may follow some types of goodness criteria of the inverse solution, and data-driven methods, which lead to the most common cases of quasilinear (F-invariant) inverse operators (Baillet et al., 2001; Hauk, 2004; Friston K. J. et al., 2008; Grech et al., 2008; Marzetti et al., 2008; Henson et al., 2011; Hindriks, 2020).

$$H_{2,\mathbf{A}(f)}(\iota(f)) = \iota^{\dagger}(f)\,\mathbf{A}^{-1}(f)\,\iota(f) \qquad (19)$$

Representing the most common quasilinear cases the following inverse operator $\hat{\mathbf{T}}_{\iota v}(\mathbf{A}(f), \mathbf{B}(f))$, in brief notation $\hat{\mathbf{T}}_{\iota v}(f)$ (Eq.

20), the constrained generalized inverse of the forward operator $\mathbf{L}_{\iota v}$ that incorporates the regularization matrices $\mathbf{A}(f)$ and $\mathbf{B}(f)$.

$$\hat{\imath}(f) \leftarrow \hat{\mathbf{T}}_{\iota v}(f)\, v(f)$$
$$\hat{\mathbf{T}}_{\iota v}(f) = \hat{\mathbf{\Pi}}_u(f)\, \mathbf{L}_{\iota v} \mathbf{B}^{-1}(f) \;\; with \;\; \hat{\mathbf{\Pi}}_u(f) = \left(\mathbf{L}_{\iota v}\, \mathbf{B}^{-1}(f)\, \mathbf{L}_{v\iota} + \alpha(f)\, \mathbf{A}^{-1}(f)\right)^{-1}$$
$$q\left(\iota(f)\middle| v(f)\right) \propto N^{\mathbb{C}}\left(\iota(f)\middle| \hat{\mathbf{T}}_{\iota v}(f)\, v(f), \hat{\mathbf{\Pi}}_u(f)\right) \tag{20}$$

Where $\hat{\mathbf{T}}_{\iota v}(f)$ arises from the route 2 MAP1 (Eq. 17) that incorporates the Gibbs energy $H_{2,\mathbf{A}(f)}(\iota(f))$ (Eq. 19), and $\mathbf{B}(f)$ any approximation to the noise cross-spectrum $\mathbf{\Sigma}_{\xi\xi}(f)$ in probabilities (Eq. 18); and $q(\iota(f)|v(f))$ is the Gaussian *a posteriori* $N^{\mathbb{C}}$, with mean $\hat{\mathbf{T}}_{\iota v}(f)\, v(f)$ and covariance matrix $\hat{\mathbf{\Pi}}_u(f)$, that follows from the conjugated relation between the *likelihood* and *a priori* probabilities in route 2 MAP1.

The cESI estimator is then $\overline{\mathbf{\Sigma}}_{\hat{u}}(f)$ (Eq. 21) for any set of inverse solution instances $(\hat{\imath}_m(f);\ \forall m)$, or more compactly from an estimate of the data cross-spectrum $\overline{\mathbf{\Sigma}}_{vv}(f)$ (Eq. 22).

$$\overline{\mathbf{\Sigma}}_{\hat{u}}(f) = \left\langle \hat{\imath}_m(f)\, \hat{\imath}_m^{\dagger}(f);\ M \right\rangle \left\langle \hat{\imath}_m(f)\, \hat{\imath}_m^{\dagger}(f);\ M \right\rangle = \left(\frac{1}{M}\right) \sum_{m=1}^{M} \hat{\imath}_m(f)\, \hat{\imath}_m^{\dagger}(f) \tag{21}$$

$$\overline{\mathbf{\Sigma}}_{\hat{u}}(f) = \hat{\mathbf{T}}_{\iota v}(f)\, \overline{\mathbf{\Sigma}}_{vv}(f)\, \hat{\mathbf{T}}_{v\iota}(f) \tag{22}$$

Important examples of cESI that follow route 1 are MNE, eLORETA and LCMV:

- The *Minimum Norm Estimate* (MNE) (Hämäläinen and Ilmoniemi, 1994). The basic smooth model of the source variables, that could either disregard the weight matrix $\mathbf{A}(f) = \mathbf{I}$ or consider it *ad-hoc* $\mathbf{A}(f) = \mathbf{A}^{ac}$ based on anatomical information.
- The *Linearly Constrained Minimum Variance* (LCMV) (Van Veen et al., 1997). The beamformer method that estimates a diagonal weight matrix $\mathbf{A}(f) = diag(\mathbf{a})$. This estimation produces an ideal filter (inverse operator) for each source variable, suppressing the interference from the other source variables and performing ideally under focalized distribution around one or a few sources.
- The *Exact Low-Resolution Electromagnetic Tomographic Analysis* (ELORETA) (Pascual-Marqui et al., 2006). A regression method that estimates $\mathbf{A}(f)$ so that the localization of the maximum for the estimated source variables corresponds exactly to the true maximum. This estimation performs ideally under a unimodal and smooth distribution of source variables.

The variants of these techniques are the one optimized for route 2, ESI in the frequency domain: the Spectral eLORETA (seLORETA) (Nolte et al., 2020) and the Spectral LCMV (sLCMV) (Larson-Prior et al., 2013). An additional solution used in this study as a reference is the spectral MNE (sMNE).

Novel cESI theory leading to the sSSBL approximation

Having described the theory of state-of-the-art cESI methods, we now focus on more sophisticated MAP theory and the sSSBL approximation that allows their practical implementation.

Bayesian MAP2 inverse operators

It is important to stress that posing *ad-hoc priors* and their hyperparameters is always necessary due to the uncertainties involved, by definition of the MAPs (Eqs. 14, 15) [54], [126]. In route 2, the prior was placed upon the Fourier transform $\iota(f)$ (Eq. 19). An alternative is to leverage the asymptotic distribution of the Fourier transform (Eq. 10) and place the *a priori* upon $\mathbf{\Sigma}_u(f)$.

This approach brings us to our main contribution, the theoretically promising cESI, which leads to an optimal inverse operator $\hat{\mathbf{W}}_{\iota v}$ based on the *second-type MAP* (MAP2) (Eq. 23).

$$(route\ 3)\ \hat{\mathbf{\Sigma}}_u(f) \leftarrow \hat{\mathbf{W}}_{\iota v}\left(\overline{\mathbf{\Sigma}}_{vv}(f)\right)$$
$$\hat{\mathbf{W}}_{\iota v} = argmax_{\mathbf{W}_{\iota v}} q\left(\mathbf{W}_{\iota v}\left(\overline{\mathbf{\Sigma}}_{vv}(f)\right)\middle| \overline{\mathbf{\Sigma}}_{vv}(f)\right)$$
$$q\left(\mathbf{\Sigma}_u(f)\middle| \overline{\mathbf{\Sigma}}_{vv}(f)\right) \propto p\left(\overline{\mathbf{\Sigma}}_{vv}(f)\middle| \mathbf{\Sigma}_u(f)\right) p\left(\mathbf{\Sigma}_u(f)\right) \tag{23}$$

This MAP2 above posits the *a priori* probability $p(\mathbf{\Sigma}_u(f))$ upon the source cross-spectrum $\mathbf{\Sigma}_u(f)$, considered a random hyper-parameter matrix. Incorporating an *a priori* to match the *likelihood* upon the sampled estimator $\overline{\mathbf{\Sigma}}_{vv}(f)$ leads to *a posteriori* $q\left(\mathbf{\Sigma}_u(f)\middle| \overline{\mathbf{\Sigma}}_{vv}(f)\right)$. Such a *likelihood* is the Complex Wishart probability density $W^{\mathbb{C}}$ (Eq. 11) now defined for $\overline{\mathbf{\Sigma}}_{vv}(f)$ (Eq. 24) with scale matrix $\mathbf{\Sigma}_{vv}(f)$ and M degrees of freedom. The specific scale matrix $\mathbf{\Sigma}_{vv}(f)$ in this *likelihood* posits the relation to the matrix equation of source cross-spectrum $\mathbf{\Sigma}_u(f)$ (Eq. 12). An *a priori* probability is then upon some cross-spectral Gibbs energy $P(\mathbf{\Sigma}_u(f))$ defined by norms such as the *vectorized (entry-wise) p-norm* (Ding et al., 2006) and *Schatten p-norm* (Fan, 1951; Schatten, 2013) upon the matrix $\mathbf{\Sigma}_u(f)$.

$$p\left(\overline{\mathbf{\Sigma}}_{vv}(f)\middle| \mathbf{\Sigma}_u(f)\right) = W^{\mathbb{C}}\left(\overline{\mathbf{\Sigma}}_{vv}(f)\middle| \mathbf{L}_v \mathbf{\Sigma}_u(f)\, \mathbf{L}_{\iota v} + \mathbf{\Sigma}_{\xi\xi}(f), M\right)$$
$$p\left(\mathbf{\Sigma}_u(f)\right) \propto \exp\left(H\left(\mathbf{\Sigma}_u(f)\right)\middle| \alpha(f)\right) \tag{24}$$

Depending on the *likelihood* and *a priori* probability (Eq. 24), the optimal inverse operator $\hat{\mathbf{W}}_{\iota v}$ (Eq. 23) is often non-linear and numerically intractable. In this case, which we followed in this article, optimizing $\hat{\mathbf{W}}_{\iota v}$ requires Approximated Bayesian Computation (ABC) (Csilléry et al., 2010). The ABC we describe here employed $\hat{\mathbf{W}}_{\iota v}^{(k+1)}$, a non-convex but numerically tractable successive approximation to $\hat{\mathbf{W}}_{\iota v}$. In turn, the tractable $\hat{\mathbf{W}}_{\iota v}^{(k+1)}$ followed from $q^{(k)}\left(\mathbf{\Sigma}_u(f)\middle| \overline{\mathbf{\Sigma}}_{vv}(f)\right)$, the non-convex relaxation of the *a posteriori* $q\left(\mathbf{\Sigma}_u(f)\middle| \overline{\mathbf{\Sigma}}_{vv}(f)\right)$ in Expectation-Maximization (EM) iterations. Obtaining the *a posteriori* relaxation is *via* the *Variational Bayes* (VB) treatment (Dempster et al., 1977; Liu and Rubin, 1994; Daunizeau and Friston, 2007; Friston et al., 2007; Nummenmaa et al., 2007; Friston K. J. et al., 2008). Such a VB treatment could target the separable model for $q\left(\mathbf{\Sigma}_u(f)\middle| \overline{\mathbf{\Sigma}}_{vv}(f)\right)$, or the separable *Hierarchical Bayesian* (HB) model for the *a priori* $p(\mathbf{\Sigma}_u(f))$.

Bayesian (joint-MAP) inverse operators and cross-spectral norms

Considering the MAP2 (Eq. 23) and the latent cross-spectrum matrix $\Sigma_u(f)$ constrained within the vector subspace generated by random instances $(\iota_m(f); \forall m)$, i.e., as if it was defined by its usual estimator $\Sigma_u(f) \cong \overline{\Sigma}_u(f)$ (Eq. 7). Hence, the MAP2 is in probability equivalent to the joint-MAP (Hsiao et al., 1998, 2002; Yeredor, 2000; Davis et al., 2001; Auranen et al., 2005; Chen et al., 2011).

We introduce this joint-MAP (Eq. 25) substituting in the general MAP definition (Eqs. 14, 15) the data $\mathcal{Y}$ and source variable $\mathcal{X}$ by the instances $\mathcal{Y} = (v_m(f); \forall m)$ and $\mathcal{X} = (\iota_m(f); \forall m)$.

$$
\begin{aligned}
(\iota_m\widehat{(f)}; \forall m) &\leftarrow \hat{\mathbf{W}}_{\iota v}(v_m(f); \forall m) \\
\hat{\mathbf{W}}_{\iota v}(v_m(f); \forall m) &= argmax_{\mathbf{W}_{\iota v}} p(\mathbf{W}_{\iota v}(v_m(f); \forall m)|v_m(f); \forall m) \\
p(\iota_m(f); \forall m|v_m(f); \forall m) &\propto p(v_m(f); \forall m|\iota_m(f); \forall m) p(\iota_m(f); \forall m)
\end{aligned}
\quad (25)
$$

In our case, we considered the above $p(v_m(f); \forall m|\iota_m(f); \forall m)$ the factorizable *joint likelihood* (Eq. 26) whose factors are present throughout the complex-valued Gaussian probability density (*likelihood* for route 2) (Eq. 18). In addition, one may assume $\Sigma_{\xi\xi} = diag(\beta(f))$ a univariate (in many cases homogenous) noise model as expressed in terms of the noise spectrum $\beta(f)$.

$$
\begin{aligned}
p(v_m(f); \forall m|\iota_m(f); \forall m) &= \prod_{m=1}^{M} p(v_m(f)|\iota_m(f)) \\
p(v_m(f)|\iota_m(f)) &= N^{\mathbb{C}}(v_m(f)|\mathbf{L}_{v\iota}\iota_m(f), diag(\beta(f)))
\end{aligned}
\quad (26)
$$

We now introduce the *joint a priori* $p(\iota_m(f); \forall m)$ (Eq. 27) upon the Gibbs energy $H(\iota_m(f); \forall m)$ redefining the previous (Eq. 15) over instances $(\iota_m(f); \forall m)$. Toward cESI, this function may be read as cross-spectral Gibbs energy $H(\overline{\Sigma}_u(f))$.

$$
\begin{aligned}
p(\iota_m(f); \forall m) &\propto \exp(H(\iota_m(f); \forall m)|\alpha(f)) = \\
&\exp(H(\overline{\Sigma}_u(f))|\alpha(f))
\end{aligned}
\quad (27)
$$

Our purpose now is to define the type of cross-spectral Gibbs energy that must deal with a severely ill-conditioned and high dimensional cESI setup ($G \gg S$). We shall deal with these problems employing the vector or matrix norms, such as the vector structured p, q-norm (Kowalski and Torrésani, 2009a,b; Gramfort et al., 2012, 2013) and the Schatten matrix p-norm (Fan, 1951; Schatten, 2013).

To recap, our approach toward cESI is then with the joint-MAP inverse operator $\hat{\mathbf{W}}_{\iota v}$ (Eqs. 25–27) [instead of the MAP2 inverse operator (Eqs. 23, 24)]. Toward our target (cESI), we must leverage the class of *joint a priori* probabilities expressed by the cross-spectral Gibbs energy $H(\overline{\Sigma}_u(f))$, not the more general case of Gibbs energy defined upon the instances $H(\iota_m(f); \forall m)$.

We leverage H_1 (Eq. 28) the structured $p = 1, q = 2\text{-norm}$ (square) that performs sparse regularization of the cross-spectrum topographic projection or spectrum $tr^{\frac{1}{2}}(\overline{\Sigma}_u(f))$. H_1 is the well-known *Least Absolute Shrinkage and Selection Operator* (LASSO) (Tibshirani, 1996) for vectors $\iota(f)$ extended or structured over the second dimension index m for the set of instances $(\iota_m(f); \forall m)$. Such a structured norm is a matrix *quasinorm* that does not fulfill the triangle equality over the cross-spectrum $\overline{\Sigma}_u(f)$.

$$
H_1(\overline{\Sigma}_u(f)) = \sum_{g=1}^{G}\left(\sum_{m=1}^{M}|\iota_m(g,f)|^2\right)^{\frac{1}{2}} = tr^{\frac{1}{2}}(\overline{\Sigma}_u(f)) \quad (28)
$$

In addition, we introduce H_2 (Eq. 29) the structured $p = 2, q = 2\text{-norm}$ (square), Shatten $p = 1\text{-norm}$ or nuclear norm $tr(\overline{\Sigma}_u(f))$, to compensate sparse bias of H_1 (Eq. 28) and regularize eigenvalues for the cross-spectrum $\overline{\Sigma}_u(f)$. H_2 is the well-known Ridge operator (Hoerl and Kennard, 1970) for a vector $\iota(f)$ structured over the second dimension index m for the set of instances $(\iota_m(f); \forall m)$.

$$
H_2(\overline{\Sigma}_u(f)) = \sum_{m=1}^{M}\sum_{g=1}^{G}|\iota_m(g,f)|^2 = tr(\overline{\Sigma}_u(f)) \quad (29)
$$

ESI practice, which is based on either the *sparse* or the *smooth* models, may be insufficient, though, in many scenarios where brain activity is patch-wise or not wholly sparse, Elastic Net, the linear combination of *sparse/smooth* models, may be ideal (Zou and Hastie, 2005; Vega-Hernández et al., 2008). Such a regularization style is a spatially structured sparsity due to the linear combination of the *sparse* LASSO and *smooth* Ridge operators upon the vector $\iota(f)$. Hence, for the vector instances $(\iota_m(f); \forall m)$, the linear combination of the *quasinorm* H_1 (Eq. 28) and the *nuclear norm* H_2 (Eq. 29) leads to the following *joint a priori* probability $p(\iota_m(f); \forall m)$ (Eq. 30).

$$
\begin{aligned}
p(\iota_m(f); \forall m) &\propto \exp(H_1(\overline{\Sigma}_u(f))|\alpha_1(f)) \\
&\exp(H_2(\overline{\Sigma}_u(f))|\alpha_2(f))
\end{aligned}
\quad (30)
$$

Motivated by ESI practice, our definition of the Elastic Net (Eq. 30) (nuclear quasinorm) diverges from that of previous studies. From these works, the Elastic Net nuclear norm combines the nuclear norm $tr(\overline{\Sigma}_u(f))$ with the square Schatten $p = 2\text{-norm}$ (Frobenius norm) $tr(\overline{\Sigma}_u^2(f))$ (Sun and Zhang, 2012; Chen et al., 2013; Kim et al., 2015; Zhang et al., 2017). This Elastic Net nuclear norm resolves convexity problems of the nuclear norm in the context of matrix completion (Candes and Recht, 2012; Hu et al., 2012) due to a non-convex problem with the sole nuclear norm. Although it was not our purpose to investigate convexity here, this property holds for our Elastic Net nuclear quasinorm, assuming cross-spectrum $\overline{\Sigma}_u(f)$ upon vector basis (Zou and Hastie, 2005). In addition, we did not consider the vectorized (entry-wise) p-norm (Ding et al., 2006). The latter, which might be necessary to regularize off-diagonal entries in some cases (Paz-Linares et al., 2018), failed to ameliorate ill-condition or distortions.

The type of ABC introduced here is known as "gamma-MAP," the standard VB treatment applied to *a priori* probabilities in the joint-MAP. In turn, the gamma-MAP leads to the quasilinear successive approximations $\hat{\mathbf{T}}_{\iota v}^{(k)}$. Solving the gamma-MAP under cross-spectral Gibbs energy $H(\overline{\Sigma}_u(f))$ defined by the Elastic Net nuclear quasinorm is the Structured Sparse Bayesian learning algorithm (ssSBL) described in the next section.

Gamma-MAP and implementation of SSBL

Similarly to the MAP2 (Eq. 23), the joint-MAP inverse operator $\hat{\mathbf{W}}_{\iota v}$ (Eq. 25) could be non-linear by nature, depending on the *joint a priori* probability $p(\iota_m(f); \forall m)$. We now introduce the gamma-MAP that achieves the quasilinear joint-MAP version leveraging the idea of mean-field approximation (Kadanoff, 2009). Such a mean-field approximation is the main idea behind the

VB treatment applied to the latent variables of neuroimaging data (MacKay, 1998; Roweis and Ghahramani, 1999; Ghahramani and Beal, 2000, 2001; Eyink et al., 2004; Friston et al., 2007; Nummenmaa et al., 2007; Friston K. J. et al., 2008; Babacan et al., 2009).

By assuming the definition for this *joint a priori* $p\left(\iota_m\left(f\right);\,\forall m\right)$ upon Gibbs energy $H\left(\iota_m\left(f\right);\,\forall m\right)$ (Eq. 27), the mean-field approximation is then to search for the self-consistent energy form $\tilde{H}\left(\iota_m\left(f\right);\,\forall m\right)$ (Eq. 31) as represented by separable (in most cases also indistinguishable) energy terms $H_g\left(\iota_m\left(g,f\right);\,\forall m\right)$ for each source. A self-consistent energy form $H_g\left(\iota_m\left(g,f\right);\,\forall m\right)$ also summarizes the interaction field $g\;\leftrightarrow\;\forall g^{'}$, as a property of the original Gibbs energy $H\left(\iota_m\left(f\right);\,\forall m\right)$.

$$H\left(\iota_m\left(f\right);\,\forall m\right)\cong\tilde{H}\left(\iota_m\left(f\right);\,\forall m\right)=\sum_g H_g\left(\iota_m\left(g,f\right);\,\forall m\right) \tag{31}$$

Obtaining the self-consistent form Eq. (31) was always plausible under the Gibbs energy $H\left(\iota_m\left(f\right);\,\forall m\right)$ that fulfills pair-wise separability in identical functions $H\left(\iota_m\left(g,f\right),\,\iota_m\left(g^{'},f\right);\,\forall m\right)$ (Eq. 32). Note that this property is known for ensuring the convergence of solutions for the self-consistency field equations in the literature of magnetism (Weiss, 1907, 2001; Le Boudec et al., 2007; Kadanoff, 2009; Zheng et al., 2015). The cross-spectral Gibbs energy $H\left(\overline{\boldsymbol{\Sigma}}_\iota\left(f\right)\right)$, as defined by the norms considered here (Elastic nuclear quasinorm (Eq. 30) or others), fulfills such a property.

$$H\left(\overline{\boldsymbol{\Sigma}}_\iota\left(f\right)\right)=\sum_{g=1}^{G}\sum_{g^{'}=1}^{G}H\left(\left\langle\iota_m\left(g,f\right)\iota_m^{\dagger}\left(g^{'},f\right);\,M\right\rangle\right) \tag{32}$$

Then, if the *joint a priori* probability $p\left(\iota_m\left(f\right);\,\forall m\right)$ (Eq. 27) is written in terms of the separable Gibbs energy $H\left(\overline{\boldsymbol{\Sigma}}_\iota\left(f\right)\right)$ (Eq. 32) it turns out the Markov random field property (Eq. 33) that is summarized by $p\left(\iota_m\left(g,f\right);\,\forall m\big|\iota_m\left(g^{'},f\right);\,\forall m\right)$ the pair-wise and identical probability factors (Kindermann et al., 1980; Lafferty et al., 2001). Note that such probability factors may approximate but do not strictly represent conditional probabilities.

$$p\left(\iota_m\left(f\right);\,\forall m\right)=\prod_{g=1}^{G}\prod_{g^{'}=1}^{G}p\left(\iota_m\left(g,f\right);\,\forall m\big|\iota_m\left(g^{'},f\right);\,\forall m\right)$$
$$p\left(\iota_m\left(g,f\right);\,\forall m\big|\iota_m\left(g^{'},f\right);\,\forall m\right)\propto$$
$$\exp\left(P\left(\left\langle\iota_m\left(g,f\right)\iota_m^{\dagger}\left(g^{'},f\right);\,M\right\rangle\right)\big|\alpha\left(f\right)\right) \tag{33}$$

Hence, we target VB—the factorizable (separable) approximation $h^q\left(\iota_m\left(f\right);\,\forall m\right)$ (Eq. 34) of the *joint a priori* $p\left(\iota_m\left(f\right);\,\forall m\right)$ (Eq. 33). Essential to obtain $h_g^q\left(\iota_m\left(g,f\right);\,\forall m\right)$ is the *Hierarchical Bayes* (HB) or *probability mixture* bellow, identical *a priori h* and *a posteriori q* probabilities that are upon some type of "variational" hyper-parameters $\boldsymbol{\gamma}\left(g,f\right)$. Such an *a posteriori q* is also known as the proxy for belief propagation or message passing, denominated iterated conditional mode in the general literature of Markov random fields (Pearl, 1988, 2022; Weiss, 2001; Yedidia et al., 2003; Zheng et al., 2015).

$$p\left(\iota_m\left(f\right);\,\forall m\right)\cong h^q\left(\iota_m\left(f\right);\,\forall m\right)\propto\prod_{g=1}^{G}h_g^q\left(\iota_m\left(g,f\right);\,\forall m\right)$$
$$h_g^q\left(\iota_m\left(g,f\right);\,\forall m\right)=$$
$$\int h\left(\iota_m\left(g,f\right);\,\forall m\big|\boldsymbol{\gamma}\left(g,f\right)\right)q\left(\boldsymbol{\gamma}\left(g,f\right)\big|\iota_m\left(f\right);\,\forall m\right)d\boldsymbol{\gamma}\left(g,f\right) \tag{34}$$

Where the hyper-parameters $\boldsymbol{\gamma}\left(g,f\right)$ condense field interactions $H\left(\left\langle\iota\left(g,f\right)\iota^{\dagger}\left(g^{'},f\right);\,M\right\rangle\right)$ (Eq. 33) by the definition of an *a posteriori* probability q upon $\left(\iota_m\left(f\right);\,\forall m\right)$.

After specifying an *a priori h*, the VB treatment is applied to search for estimators of the *a posteriori* $\hat{q}$ (Eq. 35). That is to minimize the Kullback-Leibler divergence $D_{KL}\left(p,\,h^q\right)$ of the approximation h^q regarding to the *joint a priori p* (Eq. 34).

$$\hat{q}=argmin_q D_{KL}\left(p\,||\,h^q\right)$$
$$D_{KL}\left(p\,||\,h^q\right)=$$
$$\int p\left(\iota_m\left(f\right);\,\forall m\right)\log\left(p\left(\iota_m\left(f\right);\,\forall m\right)/h^q\left(\iota_m\left(f\right);\,\forall m\right)\right)$$
$$d\left(\iota_m\left(f\right);\,\forall m\right) \tag{35}$$

Minimizing the $D_{KL}\left(p,\,h^q\right)$, although theoretically achievable, turned out to be a difficult variational calculus problem for non-parametric HB (mixture) models (Blei and Jordan, 2006; Bryant and Sudderth, 2012; Gershman and Blei, 2012; Gershman et al., 2012; Nguyen and Bonilla, 2013; Duvenaud et al., 2016). Thus, the D_{KL} approach is common with parametric solutions constraining models (*a priori h* and *a posteriori q*) to the exponential family of probabilities (Andersen, 1970; Casella and Berger, 2021).

A simplifying assumption that could bypass the Kullback-Leibler divergence $D_{KL}\left(p,\,h^q\right)$ (Eq. 35) is to regard the *a priori h* and *a posteriori q* (34) within a family of conjugate HB models of the *likelihood*. Furthermore, we considered *a priori h* factorizable over the set of instances since the pair-wise interactions are additive by definition of $\left\langle\iota\left(g,f\right)\iota^{\dagger}\left(g^{'},f\right);\,M\right\rangle$ the cross-spectrum entries in Eq. (7).

Henceforth, a complex-valued Gaussian probability $N^{\mathbb{C}}$ defines the *a priori* $h\left(\iota_m\left(f\right)\big|\boldsymbol{\gamma}\left(f\right)\right)$, and a probability in the Gamma family Γ defines the *a posteriori* $q\left(\boldsymbol{\gamma}\left(f\right)\big|\iota_m\left(f\right);\,\forall m\right)$ which we incorporate into the following HB model (Eq. 36).

$$h^q\left(\iota_m\left(f\right);\,\forall m\right)\propto\int\left(\prod_{m=1}^{M}h\left(\iota_m\left(f\right)\big|\boldsymbol{\gamma}\left(f\right)\right)\right)q\left(\boldsymbol{\gamma}\left(f\right)\big|\iota_m\left(f\right);\,\forall m\right)$$
$$d\boldsymbol{\gamma}\left(f\right)$$
$$h\left(\iota_m\left(f\right)\big|\boldsymbol{\gamma}\left(f\right)\right)=N^{\mathbb{C}}\left(\iota_m\left(f\right)\big|0,diag\left(\boldsymbol{\gamma}\left(f\right)\right)\right)$$
$$q\left(\boldsymbol{\gamma}\left(f\right)\big|\iota_m\left(f\right);\,\forall m\right)=\prod_{g=1}^{G}\Gamma\left(\boldsymbol{\gamma}\left(g,f\right)\big|\boldsymbol{\delta}\left(g,\iota_m\left(f\right);\,\forall m\right)\right) \tag{36}$$

The previous HB model is the *Generalized Gaussian Scale Mixture Model* (GGSMM) prescribed by the *Andrews and Mallows lemma* (Andrews, 1974), where the particular and "Gamma" probability Γ depends on *joint a priori* to be approximated *via* $D_{KL}\left(p,\,h^q\right)$ (Eq. 35) and falls within the class of scale mixture Gamma models (McLachlan and Basford, 1988; Lindsay, 1995; Böhning and Seidel, 2003; Hancock and Samuelsen, 2007).

In such a mixture model, the hyper-parameter vector $\boldsymbol{\gamma}\left(f\right)$ may then be interpreted as the "variational spectrum," which specifies a Complex Gaussian *a priori* probability $N^{\mathbb{C}}$ (Eq. 36). The *a posteriori* for the variational spectrum entries $\boldsymbol{\gamma}\left(g,f\right)$ is a form of probability that belongs to the Gamma Γ family, with a parameterization $\boldsymbol{\delta}\left(g,\iota_m\left(f\right);\,\forall m\right)$ that condenses field interactions $H\left(\left\langle\iota\left(g,f\right)\iota^{\dagger}\left(g^{'},f\right);\,M\right\rangle\right)$ (Eq. 32).

The joint-MAP (Eq. 25) can be approximated sequentially within iterations for the set $\left(\iota_m^{(k)}\left(f\right);\,\forall m\right)$, or directly the cross-spectrum $\hat{\boldsymbol{\Sigma}}_\iota^{(k)}\left(f\right)$, due to the quasilinear inverse operator $\mathbf{T}_{\iota v}^{(k)}$ (37). Quasilinear $\mathbf{T}_{\iota v}^{(k)}$ was then equivalent to iterated

MAP1s for the Fourier transform (Eq. 20), positing complex-valued Gaussian *likelihood* $p\left(v_m\left(f\right)\big|\iota_m\left(f\right)\right)$ (Eq. 26) and *a priori* $h\left(\iota_m\left(f\right)\big|\gamma\left(f\right)\right)$ (36).

$$\left(\widehat{\iota_m^{(k)}\left(f\right)};\ \forall\ m\right) \leftarrow \mathbf{T}_{\iota v}^{(k)}\left(f\right)\left(v_m\left(f\right);\ \forall\ m\right)$$
$$\hat{\overline{\boldsymbol{\Sigma}}}_{\iota\iota}^{(k)}\left(f\right) = \mathbf{T}_{\iota v}^{(k)}\left(f\right)\overline{\boldsymbol{\Sigma}}_{vv}\left(f\right)\mathbf{T}_{v\iota}^{(k)}\left(f\right) \quad (37)$$

These iterated MAP1s are for an equivalent *a posteriori* $q^{(k)}\left(\iota_m\left(f\right)\big|v_m\left(f\right)\right)$ (Eq. 38), specified by the mean $\mathbf{T}_{\iota v}^{(k)}\left(f\right)v\left(f\right)$ and the covariance $\boldsymbol{\Pi}_{\iota\iota}^{(k)}\left(f\right)$, which are functions of the variational spectrum $diag\left(\gamma^{(k)}\left(f\right)\right)$ and the noise spectrum $diag\left(\boldsymbol{\beta}^{(k)}\left(f\right)\right)$. These spectrums specified the type of univariate approximations for the source cross-spectrum $\boldsymbol{\Sigma}_{\iota\iota}$, and the noise cross-spectrum $\boldsymbol{\Sigma}_{\xi\xi}$ in the quasilinear MAP1 of the Fourier transform (Eq. 20). Henceforth, we place the emphasis in the variational spectrum iterations $\gamma^{(k)}\left(f\right)$ and defer this noise spectrum $\boldsymbol{\beta}^{(k)}\left(f\right)$ which is not essential for our main exposition.

$$q^{(k)}\left(\iota_m\left(f\right)\big|v_m\left(f\right)\right) = N^{\mathbb{C}}\left(\iota_m\left(f\right)\big|\mathbf{T}_{\iota v}^{(k)}\left(f\right)v_m\left(f\right),\boldsymbol{\Pi}_{\iota\iota}^{(k)}\left(f\right)\right)$$
$$\mathbf{T}_{\iota v}^{(k)}\left(f\right) = \hat{\boldsymbol{\Pi}}_{\iota\iota}^{(k)}\left(f\right)\mathbf{L}_{\iota v}diag^{-1}\left(\boldsymbol{\beta}^{(k)}\left(f\right)\right) \quad (38)$$
$$\boldsymbol{\Pi}_{\iota\iota}^{(k)}\left(f\right) = \left(\mathbf{L}_{\iota v}\,diag^{-1}\left(\boldsymbol{\beta}^{(k)}\left(f\right)\right)\mathbf{L}_{v\iota} + diag^{-1}\left(\gamma^{(k)}\left(f\right)\right)\right)^{-1}$$

Targeting the variational vector $\gamma\left(f\right)$ was under the marginal or hyper-parametrized *likelihood* for data $p\left(v_m\left(f\right)\big|\gamma\left(f\right)\right)$ (Eq. 39). This *likelihood* was due to integration (expectation) of $p\left(v_m\left(f\right)\big|\iota_m\left(f\right)\right)$ under the *a priori* probability for parameters $h\left(\iota_m\left(f\right)\big|\gamma\left(f\right)\right)$, which is upon the variational hyper-parameters (spectrum) $\gamma\left(f\right)$ (Eq. 30) (MacKay, 1999).

$$p\left(v_m\left(f\right)\big|\gamma\left(f\right)\right) = \int p\left(v_m\left(f\right)\big|\iota_m\left(f\right)\right)h\left(\iota_m\left(f\right)\big|\gamma\left(f\right)\right)d\iota_m\left(f\right)$$
$$p\left(v_m\left(f\right)\big|\iota_m\left(f\right)\right) = N^{\mathbb{C}}\left(v_m\left(f\right)\big|\mathbf{L}_{v\iota}\iota_m\left(f\right),diag\left(\boldsymbol{\beta}\left(f\right)\right)\right) \quad (39)$$
$$h\left(\iota_m\left(f\right)\big|\gamma\left(f\right)\right) = N^{\mathbb{C}}\left(\iota_m\left(f\right)\big|0,diag\left(\gamma\left(f\right)\right)\right)$$

However, the actual marginal *likelihood* $p\left(v_m\left(f\right)\big|\gamma\left(f\right)\right)$ (Eq. 39) was intractable, with approximations *via* the expectation of the *log joint* probability $p\left(v_m\left(f\right),\iota_m\left(f\right)\big|\gamma\left(f\right)\right)$ (Dempster et al., 1977; Liu and Rubin, 1994) under the iterated *a posteriori* $q^{(k)}\left(\iota_m\left(f\right)\big|v_m\left(f\right)\right)$ (Eq. 40). Then an approximated marginal *likelihood* $L\left(v_m\left(f\right)\big|\gamma^{(k)}\left(f\right),\gamma\left(f\right)\right)$ depended on the variational hyper-parameters (spectrum) $\gamma^{(k)}\left(f\right)$ in the current iteration.

$$\log\left(p\left(v_m\left(f\right)\big|\gamma\left(f\right)\right)\right) \cong log\left(L\left(v_m\left(f\right)\big|\gamma^{(k)}\left(f\right),\gamma\left(f\right)\right)\right) =$$
$$\int q^{(k)}\left(\iota_m\left(f\right)\big|v_m\left(f\right)\right)\log\left(p\left(v_m\left(f\right)\big|\mathbf{L}_{v\iota}\iota_m\left(f\right)\right)h\left(\iota_m\left(f\right)\big|\gamma\left(f\right)\right)\right)$$
$$d\iota_m\left(f\right) \quad (40)$$

An inverse operator $\mathbf{W}_{\iota v}$ of the variational spectrum $\gamma\left(f\right)$ is theoretically equivalent to the previous joint-MAP (Eq. 25), which is inverse operator of the samples $\left(\iota_m\left(f\right);\ \forall\ m\right)$. This is known from the literature as "gamma-MAP" (Hsiao et al., 1998, 2002; Wipf and Nagarajan, 2009). Within the loop (Eq. 41) such an inverse operator $\mathbf{W}_{\iota v}^{(k)}$ was the tractable successive approximations,

with the iterated *a posteriori* $q\left(\gamma\left(f\right)\big|\iota_m^{(k)}\left(f\right);\forall m\right)$ and *likelihood* $L\left(v_m\left(f\right)\big|\gamma^{(k)}\left(f\right),\gamma\left(f\right)\right)$.

$$\gamma^{(k+1)}\left(f\right) \leftarrow \hat{\mathbf{W}}_{\iota v}^{(k)}\left(v_m\left(f\right);\ \forall m\right)$$
$$\hat{\mathbf{W}}_{\iota v}\left(v_m\left(f\right);\ \forall m\right) =$$
$$argmax_{\mathbf{W}_{\iota v}}p^{(k)}\left(\mathbf{W}_{\iota v}\left(v_m\left(f\right);\ \forall m\right)\big|v_m\left(f\right);\forall m\right)$$
$$p^{(k)}\left(\gamma\left(f\right)\big|v_m\left(f\right);\forall m\right) \propto \left(\prod_{m=1}^{M}L\left(v_m\left(f\right)\big|\gamma^{(k)}\left(f\right),\gamma\left(f\right)\right)\right) \quad (41)$$
$$q\left(\gamma\left(f\right)\big|\iota_m^{(k)}\left(f\right);\forall m\right)$$

Here we implemented the Bayesian learning algorithm that effectuates the gamma-MAP $\gamma^{(k+1)}\left(f\right)$ (41) and computes the quasilinear $\hat{\mathbf{T}}_{\iota v}^{(k)}$ (Eq. 38), for a *joint a priori* based on the Elastic Net nuclear quasinorm (Eq. 30). For this *joint a priori*, the solution to the HB (mixture) model (Eq. 36) is exact, following an extension of the *Andrews and Mallows* lemma to the structured sparsity norms which are here upon the spectrum $\overline{\sigma}_\iota^2\left(f\right)$.

Henceforth, we refer to this algorithm as Spectral Structured Sparse Bayesian Learning (ssSBL). The full derivation of the gamma-MAP and the ssSBL algorithm are developed in Supplementary material (Section Standard cESI theory).

Comparison of the distortions produced by seLORETA, sLCMV, and sSSBL

Data used for the validation

We used two different sets of EEG data to calculate estimated CST, each with their corresponding gold standard CST:

(1) Realistically simulated low-density EEG obtained considering as sources a CST obtained from a MEG recording with a very high sensor density (simulated-EEGvsMEG). The gold standard here is the sMNE CST of the MEG recording due to the well-known advantages of MEG over EEG and the very high sensor density.

(2) Real low-density macaque EEG (EEGvsECoG). The gold standard here is the sMNE CST of the simultaneously recorded EcoG.

Simulated–EEG vs. MEG

Figure 3 was based on a high-quality resting state MEG recording for subject 175,237 from the HCP database. The MEG signal selected for this purpose (Figure 3a2) was the 246-channel preprocessed data file. The electrical and magnetic lead fields were calculated with the subject's cortical and head geometry. With the magnetic lead field, Spectral MNE (sMNE) was used to calculate the MEG sources, which were taken as sources for the EEG. These sources were passed through the electrical Lead Field (Figure 3a3) to simulate a low-density 19-channel EEG recording (Figure 3a1). The simulation design is standard, essentially the same as those based on more straightforward configurations, where dipoles or patches are taken as the "ground truth" (Haufe and Ewald, 2019). Here, we used a much
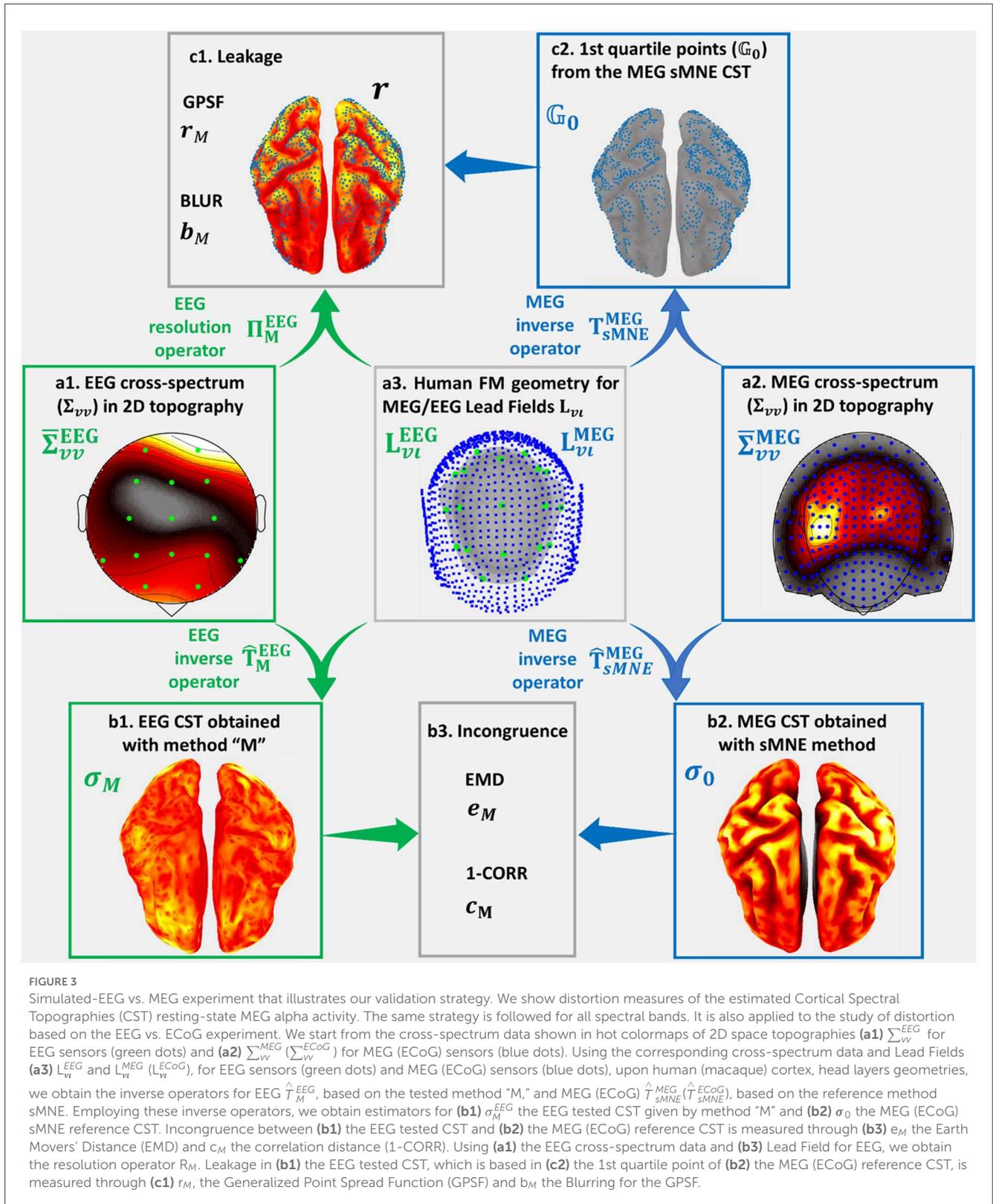
FIGURE 3
Simulated-EEG vs. MEG experiment that illustrates our validation strategy. We show distortion measures of the estimated Cortical Spectral Topographies (CST) resting-state MEG alpha activity. The same strategy is followed for all spectral bands. It is also applied to the study of distortion based on the EEG vs. ECoG experiment. We start from the cross-spectrum data shown in hot colormaps of 2D space topographies **(a1)** $\sum_{vv}^{EEG}$ for EEG sensors (green dots) and **(a2)** $\sum_{vv}^{MEG}$ ($\sum_{vv}^{ECoG}$) for MEG (ECoG) sensors (blue dots). Using the corresponding cross-spectrum data and Lead Fields **(a3)** $L_{v\iota}^{EEG}$ and $L_{v\iota}^{MEG}$ ($L_{v\iota}^{ECoG}$), for EEG sensors (green dots) and MEG (ECoG) sensors (blue dots), upon human (macaque) cortex, head layers geometries, we obtain the inverse operators for EEG $\hat{T}_{M}^{EEG}$, based on the tested method "M," and MEG (ECoG) $\hat{T}_{sMNE}^{MEG}$ ($\hat{T}_{sMNE}^{ECoG}$), based on the reference method sMNE. Employing these inverse operators, we obtain estimators for **(b1)** σ_{M}^{EEG} the EEG tested CST given by method "M" and **(b2)** σ_0 the MEG (ECoG) sMNE reference CST. Incongruence between **(b1)** the EEG tested CST and **(b2)** the MEG (ECoG) reference CST is measured through **(b3)** e_M the Earth Movers' Distance (EMD) and c_M the correlation distance (1-CORR). Using **(a1)** the EEG cross-spectrum data and **(b3)** Lead Field for EEG, we obtain the resolution operator R_M. Leakage in **(b1)** the EEG tested CST, which is based in **(c2)** the 1st quartile point of **(b2)** the MEG (ECoG) reference CST, is measured through **(c1)** r_M, the Generalized Point Spread Function (GPSF) and b_M the Blurring for the GPSF.

more realistic set of sources, determined from the MEG. Code availability: https://github.com/CCC-members/MEGvsSimulated-EEG.

EEG vs. ECoG

Figure 4 was based on high-density macaque ECoG recordings acquired in 128 sensors, concurrently with a low-density EEG

FIGURE 4
Confirmation of Cortical Spectral Topographies (CST) based in EEG recorded simultaneously with **(a1)** ECoG implanted in the macaque. An X-ray image shows **(a2)** the high-density ECoG array placed onto the macaque cortical surface. ECoG recordings and **(a3)** their Lead Fields provide a more fine-grained reference for confirming CST estimators and measures of distortions for the EEG. The validation here includes elements analogous to the MEGvsSimulated-EEG experiment of Figure 3 **(a3, b1–b3, c1, c2)**. Incongruence between **(b1)** σ_M the EEG tested CST and **(b2)** σ_o the ECoG reference CST is measured through **(b3)** $e_M^{ECoG\rightarrow EEG}$ the Earth Movers' Distance (EMD) and $C_M^{ECoG\rightarrow EEG}$ the correlation distance (1-CORR). Leakage in **(b1)** σ_M the EEG tested CST, which is based in **(c2)** on the first quartile point of **(b2)** σ_o the ECoG reference CST is measured through **(c1)** r_M, the Generalized Point Spread Function (GPSF) and b_M the Blurring for the GPSF. Elements **(a1, a2)** of this figure, are freely available in http://www.www.neurotycho.org/.

acquired at 19 sensors simultaneously with the ECoG in the resting state (Nagasaka et al., 2011). This macaque preparation allowed the realistic measurement of distortions in resting-state connectivity estimated from low-density EEG using different aESI solutions (Wang et al., 2019). A sensor array placed surgically on the left macaque's cortical surface (Figure 4a1) allowed dense recordings of

ECoG (Figure 4a2). The lead fields for EEG and ECoG (Figure 4a3) were obtained using the macaque's cortical and head geometry. The ECoG lead field was used to generate a reference or ground truth CST using sMNE. It was unnecessary to simulate data since the concurrent low-density EEG was available. Code availability: https://github.com/CCC-members/ECoGvsEEG.

Validation rationale

We evaluated the low-density EEG recordings using ssSBL, sELORETA, and sLCMV to produce the CSTs σ_{sSBL}, σ_{sELOR}, and σ_{sLCMV}, respectively. We denote these CST generically as σ_M, with M specifying the method. These calculations were carried out for the previously mentioned experiments: simulated EEG vs. MEG and EEG vs. ECOG. These CSTs were compared with reference CSTs (σ_0) considered the "gold standards." This reference was varied according to the type of measurement and will be detailed in the context below. Finally, each method's distortion of its reference was assessed using the measures described next.

Measures of distortion

Measures of distortion to compare σ_M with σ_0 fall into two groups, leakage and incongruence indices that we illustrate in the Simulated-EEG vs. MEG experiment (Figure 3). The measures of the macaque EEG vs. ECoG experiment in figure elements (Figures 4b1–b3, c1, c2) are analogous to those of the Simulated-EEG vs. MEG experiment.

Leakage (Palva et al., 2018; Van de Steen et al., 2019), or spread, is quantified using the Generalized Point Spread Function and a blurring measure. These are based on the concept of the resolution operator $\mathbf{R}_M = \mathbf{T}_{\iota\nu,M}\mathbf{L}_{\nu\iota}$. Here $\mathbf{T}_{\iota\nu,M}$ is the inverse operator for the method M, and $\mathbf{L}_{\nu\iota}$ the lead field. Consider a point source $\mathbf{u}_o$ indexed by g_0, any column $\mathbf{R}_M(:, g_0)$ is then the voltages $\mathbf{v}_0$ produced by this source.

- The Generalized Point Spread Function (Grova et al., 2006b; Haufe et al., 2008) (GPSF) $\mathbf{r}_M$ in Figure 3c1, depicted with a hot colormap, represents the leakage (spread) of the low-density EEG CST estimators for a given set of cortical points. These cortical points, shown as blue dots (Figure 3c2), are selected as the most active 25% of the reference σ_0 and are the set $\mathbb{G}_0$. The GPSF, for any point $g \in \mathbb{G}$ in the set of cortical points, is then:

$$\mathbf{r}_M(g) = \sqrt{\frac{1}{|\mathbb{G}_0|}\sum_{g_0 \in \mathbb{G}_0}\left|\mathbf{R}_M(g, g_0)\right|^2} \qquad (42)$$

Where $|\mathbb{G}_0|$ denotes the number of elements in that set.

- The blurring measure for images (BLUR) b_M is defined as the Spatial Dispersion (SD) of the GPSF (Grova et al., 2006a; Haufe et al., 2008). It is worth clarifying that for a perfect cESI solution, with zero $b_M = 0$ in Figure 3c1, there would be an exact coincidence between the GPSF $\mathbf{r}_M$ non-zero values in

the colormap and the blue dots in Figure 3c2. Formally, b_M is defined as:

$$b_M = \sqrt{\frac{1}{|\mathbb{G}_0|}\sum_{g_0 \in \mathbb{G}_0}\vartheta_M^2(g_0)}, \qquad (43)$$

Where $\vartheta_M^2(g_0)$ is the spatial dispersion around the reference point g_0 and

$$\vartheta_M^2(g_0) = \frac{1}{|\mathbb{G}|}\sum_{g \in \mathbb{G}}\left|\mathbf{R}_M(g, g_0)\right|d_{gg_0}^2, \qquad (44)$$

Where $d_{gg_0}^2$ is the square of the geodesic distance between those points.

Incongruence (Wang et al., 2019) quantifies the global level of distortion (leakage and localization error):

(1) The Earth Movers' Distance (EMD) e_M in Figure 3b3, which measures the effort to deform the CST spatial density determined from EEG (Figure 3b1) into the reference (Figure 3b2) (Grova et al., 2006b; Haufe et al., 2008; Paz-Linares et al., 2017).

(2) The correlation distance (1-CORR) c_M which measures deformations from the expected collinearity between pairs of spatially distributed CSTs.

These incongruence measures (EMD) e_M correlation distance (1-CORR) c_M combine sensitivity to leakage and localization error.

Results

Simulated-EEG vs. MEG inverse solutions

Both sELORETA and sLCMV CST estimators were seriously affected by leakage (Figure 5), judging by the mismatch between their estimated GPSF $\mathbf{r}_{sELOR}$ and $\mathbf{r}_{sLCMV}$ compared to the reference points (blue dots). They were centered at incorrect sites (opposite the blue dots) and with a much larger spread. For ssSBL, the set of local maxima for $\mathbf{r}_{ssSBL}$ was correctly centered around reference points and did not extend significantly beyond these. In other words, as can be appreciated qualitatively, $\mathbf{r}_{ssSBL}$ minimized leakage compared to sELORETA and LCMV.

Initially, these results differed from those reported by other authors for $\mathbf{r}_{sELOR}$ and $\mathbf{r}_{sLCMV}$ (Haufe and Ewald, 2019). The explanation may be due to their using idealized EEG simulations (point generators or discrete patches) that generate fewer distortions, thus reducing the apparent differences in the performance of different aESI methods. We verified this in recent aESI studies (Paz-Linares et al., 2017), concluding that simple SSBL (not the cESI ssSBL implemented in this paper) outperforms eLORETA and other methods only by a narrow margin.

An aESI solution computed with LCMV was usually sparser than an ELORETA solution, which was expected due to the thresholding strategy implemented in the original LCMV (not sLCMV) (Van Veen et al., 1997). However, this did not lead to any improvement in terms of the leakage observed in the GPSFs $\mathbf{r}_{sELOR}$ and $\mathbf{r}_{sLCMV}$. The "zero localization error" of eLORETA has
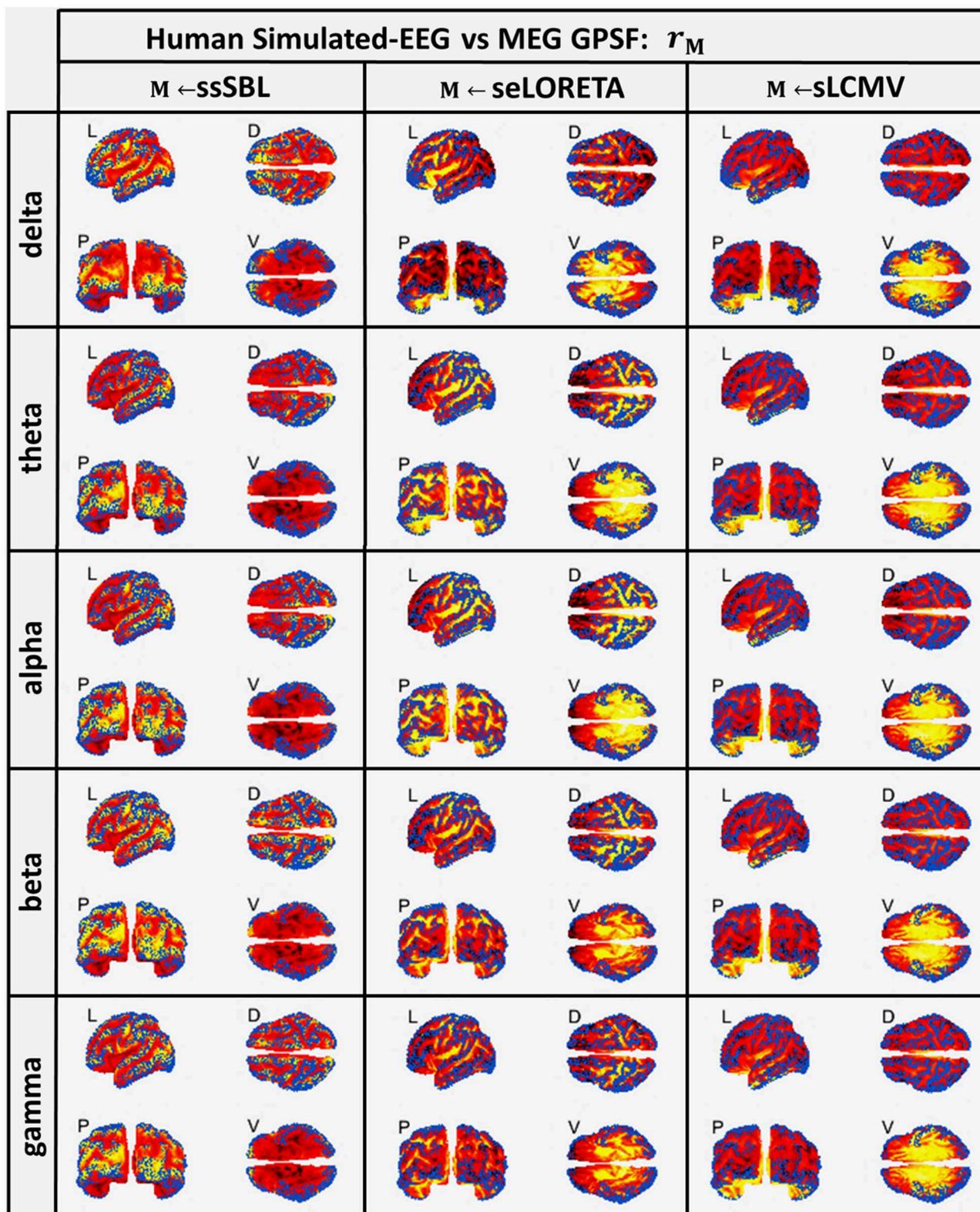
FIGURE 5
In hot colormaps, measurements of "Leakage" in the EEG-based Cortical Spectral Topographies (CST) were obtained with all tested methods "M" (ssSBL, seLORETA, and sLCMV). Leakage is measured by employing r_M, the Generalized Point Spread Function (GPSF) regarding to the MEG sMNE reference CST, shown in the cortical views Left (L), Dorsal (D), Posterior (P), and Ventral (V) and for five characteristic spectral bands (delta, theta, alpha, beta, and gamma). Leakage distortions are proportional to mismatch and spread in the GPSF values (hot colormaps) regarding reference points (blue dots). Calculations of the GPSF follow the procedure described in Figure 3 for the MEGvsSimulated-EEG experiment.

been claimed to be the unique advantage in its favor (Pascual-Marqui et al., 2006). However, this property has been theoretically demonstrated only for the peak (maximum) activity and not for the numerous local maxima of secondary activations common in real-life scenarios. Therefore, unsurprisingly, eLORETA can produce much better results in idealized simulations that use activity modeled as a focalized concentrated unimodal distribution (e.g., Gaussian) or only a few simple point sources (Kobayashi et al., 2003; Grova et al., 2008; Schoffelen and Gross, 2009; Haufe et al., 2013; Burle et al., 2015; Colclough et al., 2015; Bradley et al., 2016; Mahjoory et al., 2017; Stokes and Purdon, 2017; Palva et al., 2018; Haufe and Ewald, 2019; Marinazzo et al., 2019).

A striking observation was that despite σ_0 being highly frequency dependent, the GPSF pattern was $\mathbf{r}_{seLORETA}$ and $\mathbf{r}_{sLCMV}$ was relatively invariant for all frequencies across the spectrum. This invariance suggests that the pattern was mainly due to Lead Field properties rather than physiologically (Lopes da Silva et al., 1974; Niedermeyer and da Silva, 2005; Lopes da Silva, 2013). In contrast, the GPSF $\mathbf{r}_{ssSBL}$ pattern was closely tied to the physiological fluctuations across frequencies: from activity in the slow delta band to the faster alpha band. These fluctuations were inherited from the reference MEG data.

The quantitative analysis of the leakage measure b_M (BLUR) (Figure 6, left-column) confirms our qualitative impressions based on the GPSF $\mathbf{r}_M$. The radar graphs showed a decrease in leakage of ssSBL compared to sELORETA and sLCMV ($b_{sELOR} > b_{ssSBL}$, $b_{sLCMV} > b_{ssSBL}$). This improvement was valid for all spectral bands (delta, theta, alpha, gamma 1, and gamma 2).

EMD values e_M are intuitively the amount of work to deform the EEG-based CST estimator to the reference CST estimator ($\sigma_M \rightarrow \sigma_0$). They were one to two orders larger for sELORETA and sLCMV than SSBL ($e_{sELOR} >> e_{ssSBL}$, and $e_{sLCMV} >> e_{ssSBL}$).

A linear model adjusted to every source of the EEG and MEG CSTs estimated data pairs shows a clear linear tendency (Figure 7), with correlation distances larger for sELORETA and sLCMV. Some correlations are even negative correlations. By contrast, c_{sELOR} was in the range of around 0.7, for all frequency bands. This behavior was congruent with that observed in the colormaps for GPSF ($\mathbf{r}_{sELOR}$ and $\mathbf{r}_{sLCMV}$) (Figure 5), where the maximum values appear in opposite areas to the blue dots (reference estimation).

These results suggest that commonly used ESI validation procedures, limited to idealized simulations of local neural currents, may not accurately assess the actual distortions (Haufe et al., 2013; Haufe and Ewald, 2019). Our simulation of EEG that incorporates realistic local neural currents, estimated from high-density MEG, shows that the leakage and localization error in ESI applied to actual data might be much more severe than expected (Palva et al., 2018; He et al., 2019; Van de Steen et al., 2019). Therefore, future efforts should consider validation benchmarks based on realistic simulations like those described here. When using this benchmark, the ssSBL approach appears to considerably control the effect of distortions with respect to other techniques.

EEG vs. ECoG inverse solutions

The GPSF colormaps (Figure 8) for each method showed a consistent behavior as those of the Simulated-EEG vs. MEG

experiment (Figure 5). As evident in the GPSF maps $\mathbf{r}_{ssSBL}$ and measured in b_{ssSBL} (Figure 6, right-column), the performance of ssSBL in reducing leakage was superior compared to sELORETA ($b_{sELOR} > b_{ssSBL}$) and sLCMV ($b_{sLCMV} > b_{ssSBL}$).

The measurements of incongruence by the EMD e_M (Figure 6, right-column) were consistent with those of the previous experiment (left-column), confirming the improvement of ssSBL regarding sELORETA ($e_{sELOR} > e_{ssSBL}$ by a considerably narrow margin) and LCMV ($e_{sLCMV} >> e_{ssSBL}$ by a wider margin of three orders of magnitude) for all spectral bands.

The correlation distance c_M for all methods showed that linear regression described the relation of EEG CST to ECoG CST well. The correlation was positive for ssSBL but negative for sELORETA and sLCMV as seen in Figure 6 (right-column). As a consequence ($c_{sLCMV} > c_{LORETA} > c_{ssSBL}$) for all spectral bands. This linear tendency in Figure 9, similarly to Figure 7 confirms the feasibility of cESI, even with low-density EEG, given its close relationship to a more direct observation modality like ECoG. The results of sELORETA and sLCMV did not reveal a clear linear positive tendency. These results confirm and extend the results of previous studies (Marinazzo et al., 2019; Papadopoulou et al., 2019; Wang et al., 2019), suggesting that some types of ESI should be interpreted with extreme care.

Discussion

We now summarize and evaluate our results from a theoretical point of view. We introduce a general Bayesian framework for cESI, the estimation of source cross-spectral matrices. This approach allowed us to address the high level of topographic distortions, which arise from the severely ill-conditioned nature of the underlying inverse problem (Kobayashi et al., 2003; Grova et al., 2008; Schoffelen and Gross, 2009; Haufe et al., 2013; Burle et al., 2015; Colclough et al., 2015; Bradley et al., 2016; Mahjoory et al., 2017; Stokes and Purdon, 2017; He et al., 2018, 2019; Palva et al., 2018; Haufe and Ewald, 2019; Marinazzo et al., 2019). We indicated that the problems originating from the ill-posedness were compounded by the habitual implementation of linear and non-sparse (smooth) types of inverse solutions (Mantini et al., 2007; Marzetti et al., 2008; Brookes et al., 2011b, 2012; Hipp et al., 2012; Lopes da Silva, 2013; Colclough et al., 2015; Marinazzo et al., 2019; Nolte et al., 2020).

Non-linear and sparse inverse solutions might ameliorate distortions (Tibshirani, 1996; Zou and Hastie, 2005; Yuan and Lin, 2006; Haufe et al., 2008; Vega-Hernández et al., 2008; Kowalski and Torrésani, 2009a,b; Li and Tian, 2011; Gramfort et al., 2012, 2013). Unfortunately, they may also introduce warping and other biases of the cross-spectral (cESI) estimator. ESI practice with these inverse solutions was most beneficial when addressing the deterministic time/frequency waveforms of event-related brain processes within the framework of spatially short-scale distributed event-related brain networks (Picton and Hillyard, 1974; Picton et al., 1974; Lopes da Silva et al., 1991; Clark et al., 1994; Makeig et al., 1999, 2004; Makeig, 2002; Eichele et al., 2005; Harrison et al., 2008; Vega-Hernández et al., 2008; Maurer and Dierks, 2012). This approach to ESI empiricism suggests that the best results were obtained
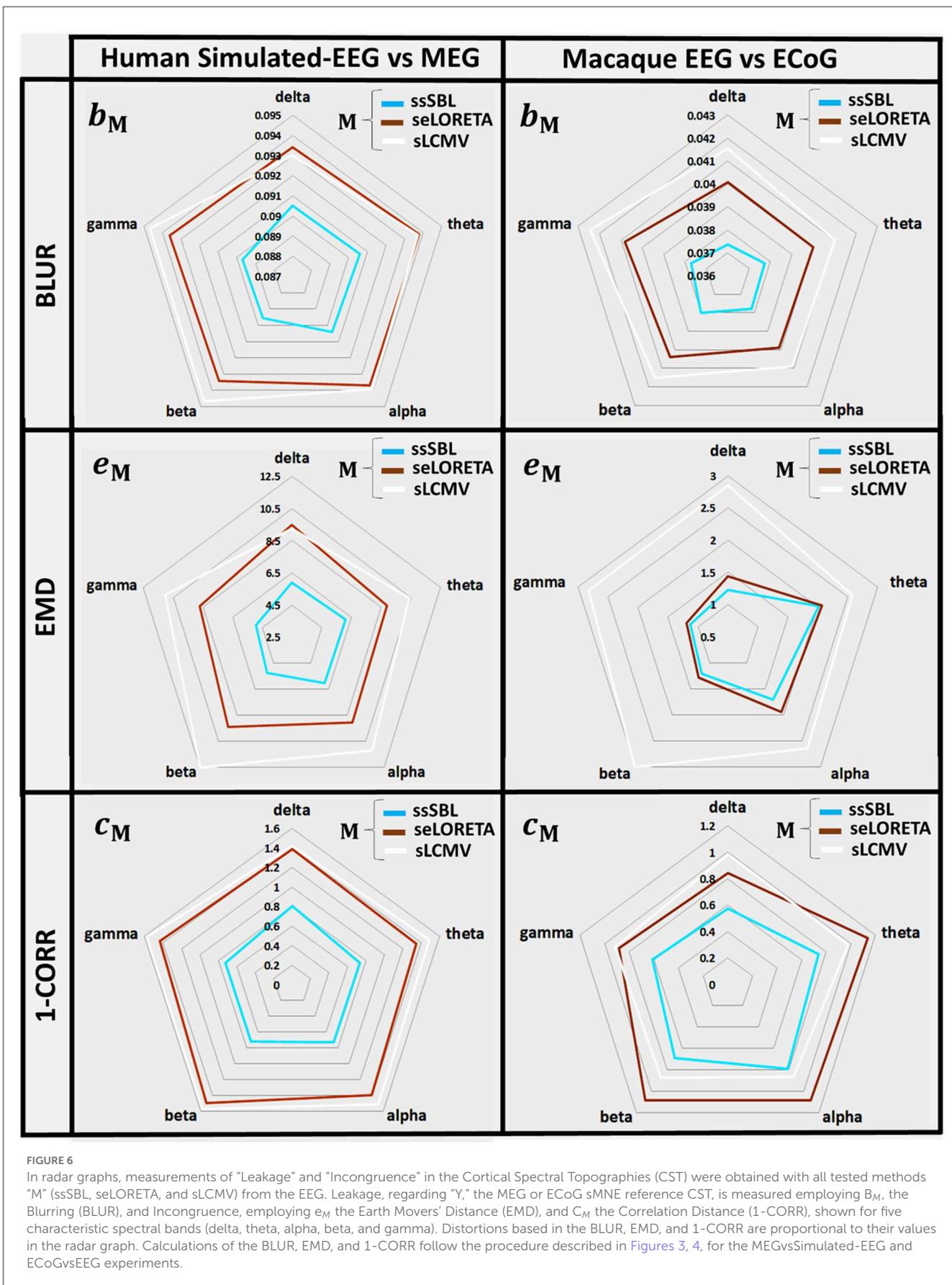
FIGURE 6
In radar graphs, measurements of "Leakage" and "Incongruence" in the Cortical Spectral Topographies (CST) were obtained with all tested methods "M" (ssSBL, seLORETA, and sLCMV) from the EEG. Leakage, regarding "Y," the MEG or ECoG sMNE reference CST, is measured employing B_M, the Blurring (BLUR), and Incongruence, employing e_M the Earth Movers' Distance (EMD), and C_M the Correlation Distance (1-CORR), shown for five characteristic spectral bands (delta, theta, alpha, beta, and gamma). Distortions based in the BLUR, EMD, and 1-CORR are proportional to their values in the radar graph. Calculations of the BLUR, EMD, and 1-CORR follow the procedure described in Figures 3, 4, for the MEGvsSimulated-EEG and ECoGvsEEG experiments.
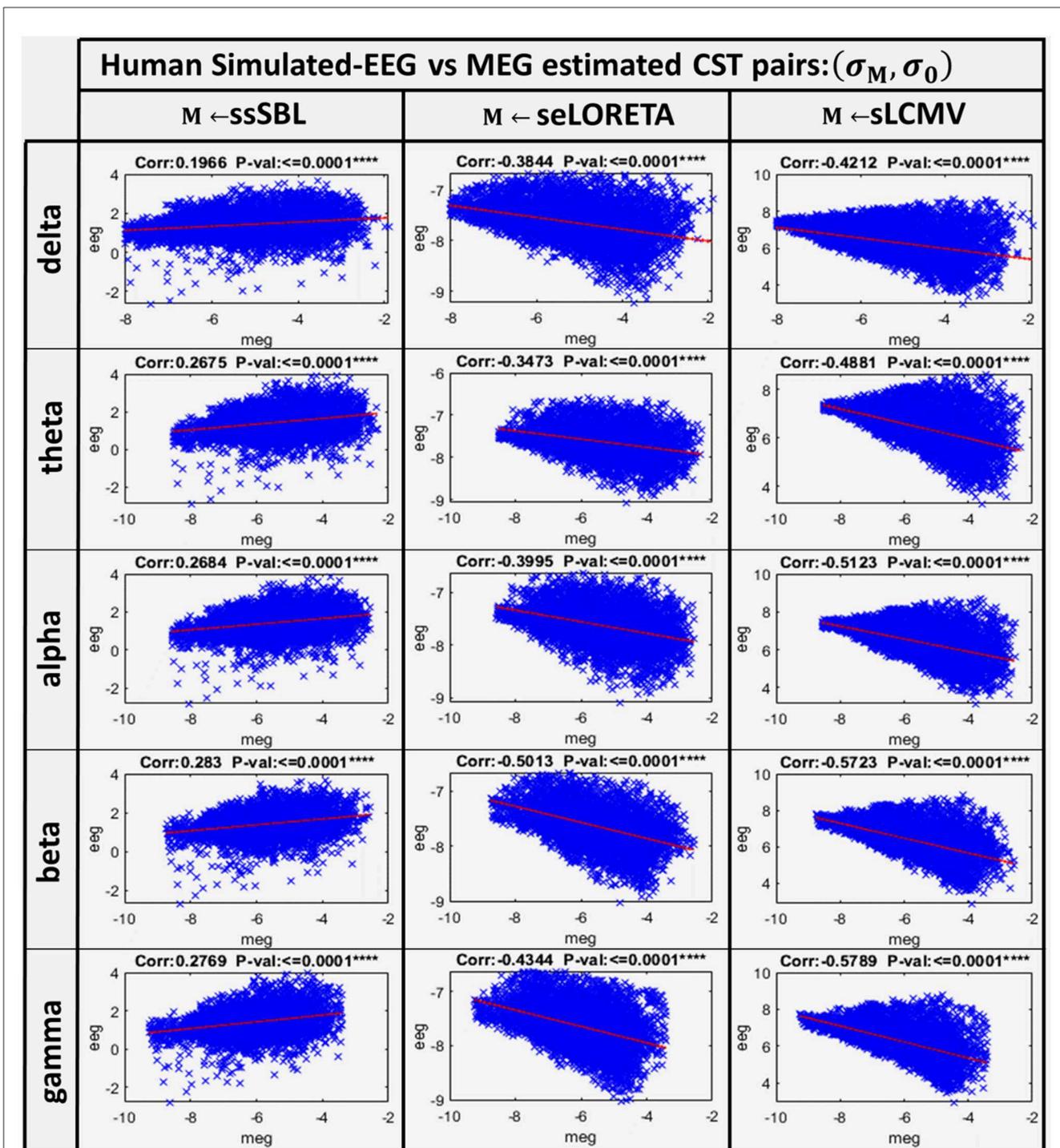
FIGURE 7
Linear model and correlations for the Cortical Spectral Topographies (CST) obtained from the MEGvsSimulated-EEG experiment in the human. These were adjusted from the CST data pairs (σ_M, σ_0) for all tested methods "M" (ssSBL, seLORETA, and sLCMV), and in five characteristic spectral bands (delta, theta, alpha, beta, and gamma).

with flexible smooth/sparse a priori models (Zou and Hastie, 2005; Vega-Hernández et al., 2008; Li and Lin, 2010). Quasilinear inverse solutions of these smooth/sparse models can provide good data-driven approximations.

We formalized the three possible *routes* toward cESI (our target) *via* MAP1 inverse solutions, for the vector processes or their

Fourier transform and *via* MAP2 or joint-MAP inverse solutions for their cross-spectrum. The cross-spectral MAP2 or joint-MAP are plausible inverse solutions, which target the quantity of interest and theretofore posit the *a priori* upon the cross-spectrum $\Sigma_u\left(f\right)$.

In contrast, most cESI have been previously limited to route 1 for the processes or route 2 for the Fourier transform
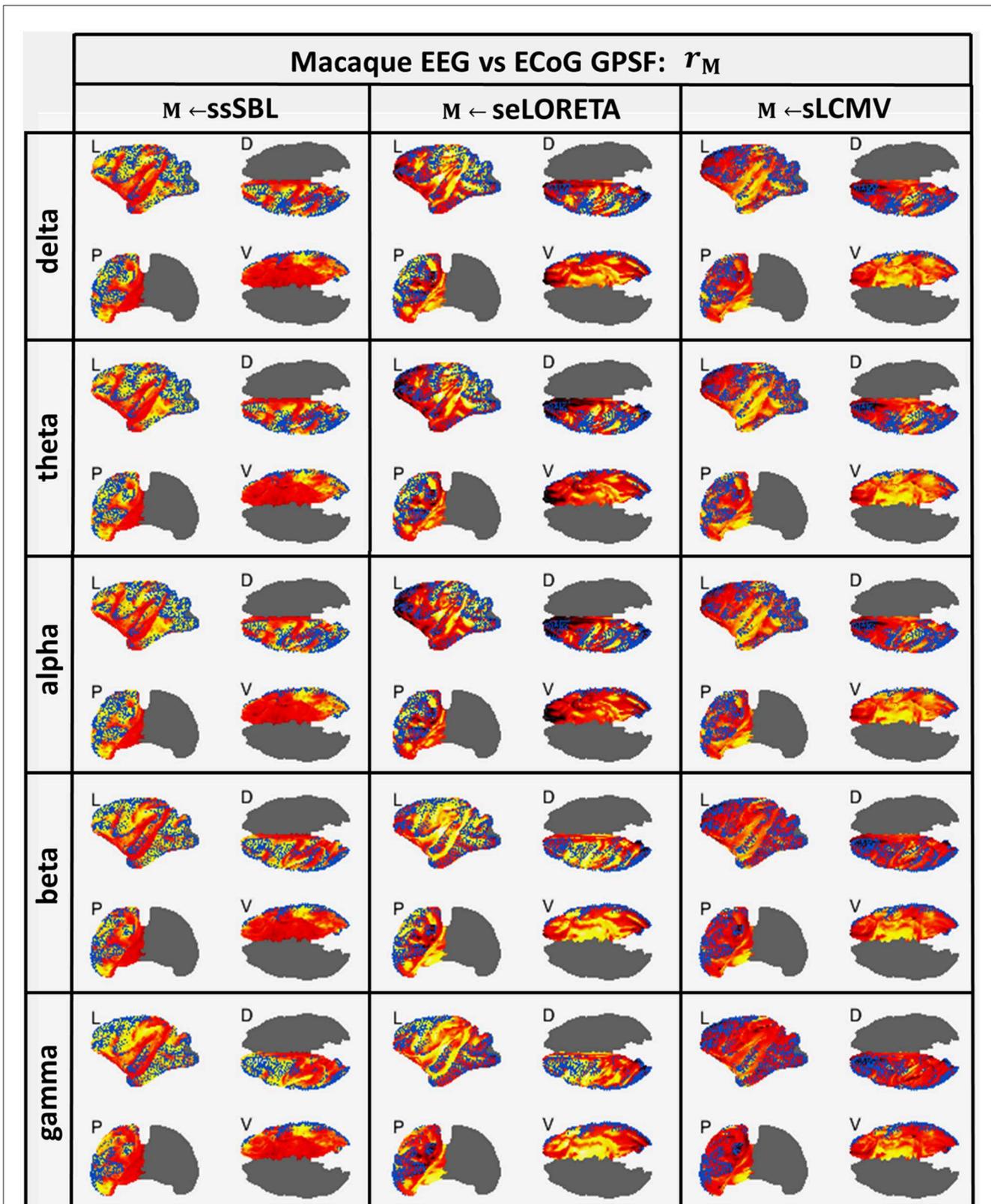
FIGURE 8
In hot colormaps, measurements of "Leakage" in the EEG-based Cortical Spectral Topographies (CST) were obtained with all tested methods "M" (ssSBL, seLORETA, and sLCMV). Leakage is measured by employing r_M, the Generalized Point Spread Function (GPSF) regarding to the ECoG sMNE reference CST, shown in the cortical views Left (L), Dorsal (D), Posterior (P), and Ventral (V) and for five characteristic spectral bands (delta, theta, alpha, beta, and gamma). Leakage distortions are proportional to mismatch and spread in the GPSF values (hot colormaps) regarding reference points (blue dots). Calculations of the GPSF follow the procedure described in Figure 4 for the ECoGvsEEG experiments.
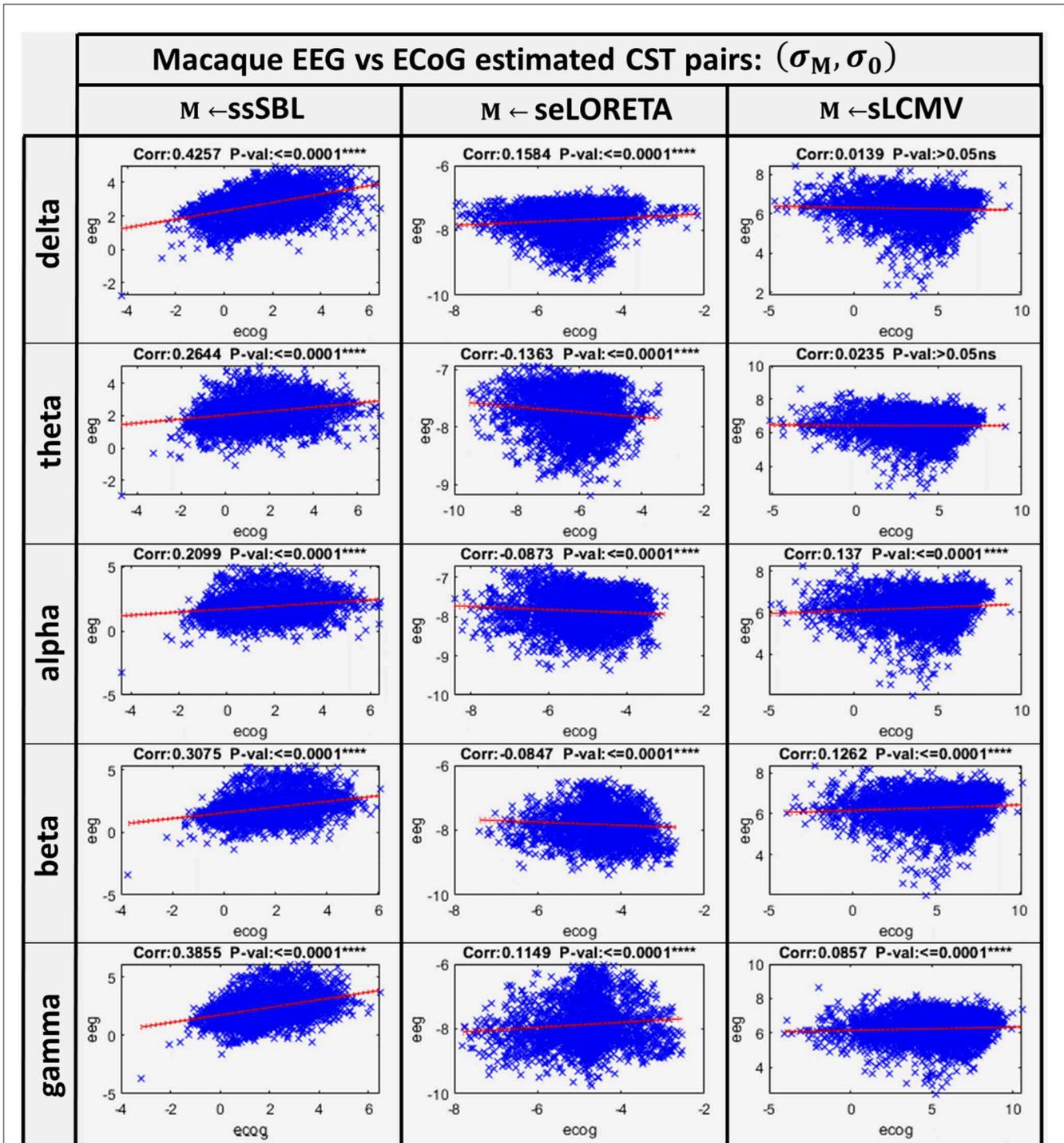
FIGURE 9
Linear model and correlations for the Cortical Spectral Topographies (CST) obtained from the ECoGvsEEG experiment in the macaque. These were adjusted from the CST data pairs (σ_M, σ_M) for all tested methods "M" (ssSBL, seLORETA, and sLCMV), and in five characteristic spectral bands (delta, theta, alpha, beta, and gamma).

(aESI) wherein the *a priori* is placed upon $\iota(t)$ or $\iota(f)$. Therefore, essential statistics like cross-spectrum were incorrectly addressed as the subsequent step to aESI. As we demonstrated in simulations and real data, the aESI procedure is not suitable, with cESI amplifying the distortions previously produced during aESI. Indeed, a way to achieve statistical guarantees in cESI is

through *route 3* associated with the source cross-spectrum for the data.

We have deferred for now to the complete implementation of the MAP2 inverse solution for the cross-spectrum. This MAP2, which targets a source matrix, was not straightforward, given a cESI setup that is severely ill-conditioned and of high dimensionality.

Hence, we adopted the joint-MAP, as an approximation to the MAP2, which targets $\overline{\Sigma}_{\iota\iota}(f)$ of the sampled source cross-spectrum.

An essential concept introducing any MAP inverse solution was the Gibbs energy, a concept first formulated as the Gibbs free energy of any physical system (Landau and Lifshitz, 1980). From the point of view of the inverse problem theory, the Gibbs free energy describes the forward energy exchange of a system (the source variable) to the media (the data) (Ghahramani and Beal, 2001; Grave de Peralta Menendez et al., 2004; Friston et al., 2006; Mattout et al., 2006; Friston K. J. et al., 2008; Friston K. L. et al., 2008; Wipf and Nagarajan, 2009). Thus, this is an important generalization of the Tikhonov regularization and ESI for source cross-spectrum.

We employed this cross-spectral Gibbs energy to model the *joint a priori* probabilities and second-order multivariate properties of the Fourier transform. The Gibbs energy in cESI must be a function of the cross-spectrum entries. This assumption follows from the Gaussianity of Fourier transform (Brillinger, 1965; Brillinger and Rosenblatt, 1967), or statistical sufficiency of the cross-spectrum. This assumption is valid for ESI under a variety of experimental conditions, as it can be demonstrated with high-quality MEG and ECoG data (Nagasaka et al., 2011; Van Essen et al., 2013).

Quasilinear inverse solutions preserved the F-invariance for Gaussianity, avoiding warping of cross-spectral amplitude and phase information (Marzetti et al., 2008; He et al., 2019; Nolte et al., 2020). F-invariance is also valid for a MAP1 assuming Gaussianity of the Fourier transform, and streamlining the dimensionality reduction for MAP2, leading to our joint-MAP interpretation (Hsiao et al., 1998, 2002; Yeredor, 2000; Davis et al., 2001; Auranen et al., 2005; Chen et al., 2011).

We indicate the importance of the nuclear norm (trace) and nuclear quasinorm (square root trace) for matrices, from the context of matrix completion inverse problems (Fan, 1951; Ding et al., 2006; Sun and Zhang, 2012; Chen et al., 2013; Schatten, 2013; Kim et al., 2015; Zhang et al., 2017) which translated into a sparse/smooth spectrum model. Per the definition of cross-spectrum upon vector basis, this model indeed could be unified with the structured vector norms (Kowalski and Torrésani, 2009a,b; Gramfort et al., 2012, 2013). Furthermore, this has been very common in aESI with sparse (LASSO) (Tibshirani, 1996) and smooth (MNE) (Hoerl and Kennard, 1970) model. Our sparse-smooth model was, therefore, an Elastic Net nuclear quasinorm.

Our approach has an important connection to Bayesian learning (Tipping, 2001; Wipf et al., 2006; Park and Casella, 2008; Wipf and Nagarajan, 2009; Casella et al., 2010; Li and Lin, 2010; Li et al., 2011; Paz-Linares et al., 2017) and provided the link between the joint-MAP and quasilinear inverse solutions. We introduced a type of variational approximation to the *joint a priori* probability upon cross-spectral Gibbs energy. This variational approximation is similar to the previous Bayesian Learning methods with extended applicability to high dimensional inverse problems that are solved in iterated conditional mode (Pearl, 1988, 2022; Weiss, 2001; Yedidia et al., 2003; Zheng et al., 2015).

Here, we restricted ourselves to the Gaussian Mixture Model approach (Blei and Jordan, 2006; Bryant and Sudderth, 2012;

Gershman and Blei, 2012; Gershman et al., 2012; Nguyen and Bonilla, 2013; Duvenaud et al., 2016) applied to the specific *a priori* per the definition of the Elastic Net nuclear quasinorm (Candes and Recht, 2012; Hu et al., 2012). Our Bayesian learning method (ssSBL) was a generalization of the SBL and sSBL approaches.

Studying the distortions in cESI required a validation based on a reference estimation (ground truth) closer to actual source distributions in the brain rather than simulations of brain electrical activity based on ideal source configurations (Kobayashi et al., 2003; Grova et al., 2008; Schoffelen and Gross, 2009; Haufe et al., 2013; Burle et al., 2015; Colclough et al., 2015; Bradley et al., 2016; Mahjoory et al., 2017; Stokes and Purdon, 2017; Palva et al., 2018; Haufe and Ewald, 2019; Marinazzo et al., 2019).

We demonstrated with several quality measures (Grova et al., 2006b; Haufe et al., 2008; Paz-Linares et al., 2017; Van de Steen et al., 2019; Wang et al., 2019) that cESI estimator distortions in actual data are more larger than expected for some state-of-the-art methods, such as sELORETA and LCMV. Our Simulated-EEG vs. MEG and EEG vs. ECoG validation method benchmarked the cESI distortions in the 10–20 EEG system (19 channels), which is considered the lower bound for all ESI. Therefore, we can infer that the behavior of our methods in EEG or other techniques that provide denser recordings can only improve.

The human Simulated-EEG vs. MEG validation method produces measurements of the cESI estimator distortions very similar to those of the EEG vs. ECoG simultaneously recorded in the macaque (Nagasaka et al., 2011; Wang et al., 2019). This similarity strongly supports the possibility of effectively measuring the distortions expected in many real-life scenarios with our Simulated-EEG vs. MEG design, which is easily replicable for other MEG acquisitions or different experimental conditions (not only resting state).

Our results indicate that ssSBL produces less cESI distortions than sELORETA and sLCMV, according to all leakage measures computed for the Simulated-EEG vs. MEG. These results were supported by cESI obtained from low-density EEG recordings in macaque compared to more fine-grained cESI obtained from simultaneously recorded high-density ECoG.

Conclusion

In this manuscript, we leveraged Bayesian theory to investigate the benefits and shortcomings of facing cESI with a large family of methods. We show that cESI must be taken care of in severely ill-conditioned and high-dimensional settings such as the ones dealt with here. In such settings, achieving cESI *via* exact Bayesian MAP2 methods is unfeasible and requires approximations. We have introduced quasilinearity, a reasonable cESI assumption, leading to the joint MAP that is feasible *via* the variational approximation to MAP. Our implementation, ssSBL specifies a priori to diminish CST distortions and exhibits good properties, outperforming state-of-the-art methods. The Bayesian theory and methods presented here could potentially be applied to signal processing and imaging other biological phenomena described by the cross-spectrum.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

DP-L and EG-M contributed in developing the theory, mathematical demonstrations, design of methods, implementation, and paper preparation. AA-G, ML, and YW prepared part of the data, code, and results used in this work. MV-H, QW, and EM-M prepared the macaque data for experimental confirmation and contributed to the theory and methods. MB-V, JB-B, and MV-S as senior authors analyzed the results. PV-S as senior author, introduced the theoretical background, CNS program of and guided this work. All authors contributed to the article and approved the submitted version.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2023.978527/full#supplementary-material

References

Andersen, E. B. (1970). Sufficiency and exponential families for discrete sample spaces. *J. Am. Stat. Assoc.* 65, 1248–1255. doi: 10.1080/01621459.1970.10481160

Andrews, D. F. (1974). Scale mixtures of normal distributions. *J. R. Stat. Soc. Ser. B* 148, 148–162.

Auranen, T., Nummenmaa, A., Hämäläinen, M. S., Jääskeläinen, I. P., Lampinen, J., Vehtari, A., et al. (2005). Bayesian analysis of the neuromagnetic inverse problem with l(p)-norm priors. *Neuroimage* 26, 870–884. doi: 10.1016/j.neuroimage.2005.02.046

Babacan, S. D., Mancera, L., Molina, R., and Katsaggelos, A. K. (2009). Bayesian compressive sensing using non-convex priors. *Eur. Signal Process. Conf.* 19, 53–63. doi: 10.1109/TIP.2009.2032894

Baillet, S., Mosher, J. C., and Leahy, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Process. Mag.* 18, 14–30. doi: 10.1109/79.962275

Blei, D. M., and Jordan, M. I. (2006). Variational inference for dirichlet process mixtures. *Bayesian Anal.* 1, 121–143. doi: 10.1214/06-BA104

Böhning, D., and Seidel, W. (2003). Recent developments in mixture models. *Comput. Stat. Data Anal.* 41, 349–357. doi: 10.1016/S0167-9473(02)00161-5

Bourbaki, N. (2013). *Topological Vector Spaces: Chapters 1–5*. Berlin: Springer Science & Business Media.

Bradley, A., Yao, J., Dewald, J., and Richter, C.-P. (2016). Evaluation of electroencephalography source localization algorithms with multiple cortical sources. *PLoS ONE* 11, e0147266. doi: 10.1371/journal.pone.0147266

Brillinger, D. R. (1965). An introduction to polyspectra. *Ann. Math. Stat.* 1351–1374. doi: 10.1214/aoms/1177699896

Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory*. Philadelphia, PA: SIAM.

Brillinger, D. R. (2012). "Asymptotic normality of finite fourier transforms of stationary generalized processes," in *Selected Works of David Brillinger* (Springer), 65–72. doi: 10.1007/978-1-4614-1344-8_6

Brillinger, D. R., and Rosenblatt, M. (1967). Asymptotic theory of estimates of kth-order spectra. *Proc. Nat. Acad. Sci. U. S. A.* 57, 206–210. doi: 10.1073/pnas.57.2.206

Brookes, M. J., Hale, J. R., Zumer, J. M., Stevenson, C. M., Francis, S. T., Barnes, G. R., et al. (2011a). Measuring functional connectivity using meg: methodology and comparison with FcMRI. *Neuroimage* 56, 1082–1104. doi: 10.1016/j.neuroimage.2011.02.054

Brookes, M. J., Woolrich, M., Luckhoo, H., Price, D., Hale, J. R., Stephenson, M. C., et al. (2011b). Investigating the electrophysiological basis of resting state networks using magnetoencephalography. *Proc. Natl. Acad. Sci. U. S. A.* 108, 16783–16788. doi: 10.1073/pnas.1112685108

Brookes, M. J., Woolrich, M. W., and Barnes, G. R. (2012). Measuring functional connectivity in meg: a multivariate approach insensitive to linear source leakage. *Neuroimage* 63, 910–912. doi: 10.1016/j.neuroimage.2012.03.048

Bruns, A. (2004). Fourier-, Hilbert- and Wavelet-based signal analysis: are they really different approaches? *J. Neurosci. Methods* 137, 237. doi: 10.1016/j.jneumeth.2004.03.002

Bryant, M., and Sudderth, E. (2012). Truly nonparametric online variational inference for hierarchical dirichlet processes. *Adv. Neural Inf. Process. Syst.* 25, 2699–2707. doi: 10.5555/2999325.2999436

Burle, B., Spieser, L., Roger, C., Casini, L., Hasbroucq, T., and Vidal, F. (2015). Spatial and temporal resolutions of EEG: is it really black and white? A scalp current density view. *Int. J. Psychophysiol.* 97, 210–220. doi: 10.1016/j.ijpsycho.2015.05.004

Buzsáki, G., Logothetis, N., and Singer, W. (2013). Scaling brain size, keeping timing: evolutionary preservation of brain rhythms. *Neuron* 80, 751–764. doi: 10.1016/j.neuron.2013.10.002

Candes, E., and Recht, B. (2012). Exact matrix completion via convex optimization. *Commun. ACM* 55, 111–119. doi: 10.1145/2184319.2184343

Casella, G., and Berger, R. L. (2021). *Statistical Inference*. Boston, MA: Cengage Learning.

Casella, G., Ghosh, M., Gill, J., and Kyung, M. (2010). Penalized regression, standard errors, and bayesian lassos. *Bayesian Anal.* 5, 369–411. doi: 10.1214/10-BA607

Chen, K., Dong, H., and Chan, K.-S. (2013). Reduced rank regression via adaptive nuclear norm penalization. *Biometrika* 100, 901–920. doi: 10.1093/biomet/ast036

Chen, Y., Li, Y., Yu, W., Luo, L., Chen, W., and Toumoulin, C. (2011). Joint-MAP tomographic reconstruction with patch similarity based mixture prior model. *Multiscale Model. Simul.* 9, 1399–1419. doi: 10.1137/100814184

Clark, V. P., Fan, S., and Hillyard, S. A. (1994). Identification of early visual evoked potential generators by retinotopic and topographic analyses. *Hum. Brain Mapp.* 2, 170–187. doi: 10.1002/hbm.460020306

Colclough, G. L., Brookes, M. J., Smith, S. M., and Woolrich, M. W. (2015). A symmetric multivariate leakage correction for MEG connectomes. *Neuroimage* 117, 439–448. doi: 10.1016/j.neuroimage.2015.03.071

Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., and François, O. (2010). Approximate Bayesian computation (ABC) in practice. *Trends Ecol. Evol.* 25, 410–418. doi: 10.1016/j.tree.2010.04.001

Daunizeau, J., David, O., and Stephan, K. E. (2011). Dynamic causal modelling: a critical review of the biophysical and statistical foundations. *Neuroimage* 58, 312–322. doi: 10.1016/j.neuroimage.2009.11.062

Daunizeau, J., and Friston, K. J. (2007). A mesostate-space model for EEG and MEG. *Neuroimage* 38, 67–81. doi: 10.1016/j.neuroimage.2007.06.034

Daunizeau, J., Laufs, H., and Friston, K. J. (2010). EEG-FMRI information fusion: biophysics and data analysis. *EEGFMRI Physiol. Basis Techn. Applic.* 511–26. doi: 10.1007/978-3-540-87919-0_25

Davis, L. M., Collings, I. B., and Hoeher, P. (2001). Joint MAP equalization and channel estimation for frequency-selective and frequency-flat fast-fading channels. *IEEE Trans. Commun.* 49, 2106–2114. doi: 10.1109/26.974257

Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M., and Friston, K. (2008). The dynamic brain: from spiking neurons to neural masses and cortical fields. *PLoS Comput. Biol.* 4, e1000092. doi: 10.1371/journal.pcbi.1000092

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* 39, 1–22. doi: 10.1111/j.2517-6161.1977.tb01600.x

Ding, C., Zhou, D., He, X., and Zha, H. (2006). "R 1-Pca: rotational invariant l 1-norm principal component analysis for robust subspace factorization," in *Proceedings of the 23rd International Conference on Machine Learning* (Pittsburgh, PA), 281–88. doi: 10.1145/1143844.1143880

Dunford, N., and Schwartz, J. T. (1988). *Linear Operators, Part 1: General Theory.* Vol. 10. Hoboken, NJ: John Wiley & Sons.

Duvenaud, D., Maclaurin, D., and Adams, R. (2016). "Early stopping as nonparametric variational inference," in *Artificial Intelligence and Statistics* (PMLR), 1070–1077.

Eichele, T., Specht, K., Moosmann, M., Jongsma, M. L. A., Quiroga, R. Q., Nordby, H., et al. (2005). Assessing the spatiotemporal evolution of neuronal activation with single-trial event-related potentials and functional MRI. *Proc. Nat. Acad. Sci. U. S. A.* 102, 17798–17803. doi: 10.1073/pnas.0505508102

Engel, A. K., Fries, P., and Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top–down processing. *Nat. Rev. Neurosci.* 2, 704–716. doi: 10.1038/35094565

Eyink, G. L., Restrepo, J. M., and Alexander, F. J. (2004). A mean field approximation in data assimilation for nonlinear dynamics. *Phys. D Nonlinear Phenomena* 195, 347–368. doi: 10.1016/j.physd.2004.04.003

Faes, L., Erla, S., and Nollo, G. (2012). Measuring connectivity in linear multivariate processes: definitions, interpretation, and practical analysis. *Comput. Math. Methods Med.* 2012, 140513. doi: 10.1155/2012/140513

Faes, L., and Nollo, G. (2011). Multivariate frequency domain analysis of causal interactions in physiological time series. *Biomed. Eng. Trends Electron. Commun. Softw.* 8, 403–428. doi: 10.5772/13065

Faes, L., Stramaglia, S., and Marinazzo, D. (2017). On the interpretability and computational reliability of frequency-domain Granger causality. *F1000 Res.* 6, 1710. doi: 10.12688/f1000research.12694.1

Fan, K. (1951). Maximum properties and inequalities for the eigenvalues of completely continuous operators. *Proc. Nat. Acad. Sci. U. S. A.* 37, 760–766. doi: 10.1073/pnas.37.11.760

Frauscher, B., Ellenrieder, N. V., Zelmann, R., Doležalová, I., Minotti, L., Olivier, A., et al. (2018). Atlas of the normal intracranial electroencephalogram: neurophysiological awake activity in different cortical areas. *Brain* 141, 1130–1144. doi: 10.1093/brain/awy035

Freeman, W. J. (1975). *Mass Action in the Nervous System : Examination of the Neurophysiological Basis of Adaptive Behavior through the EEG.* New York, NY; London: Academic Press. Available online at: https://librarysearch.cardiff.ac.uk/openurl/44WHELF_CAR/44WHELF_CAR:44WHELF_CAR_VU1??u.ignore_date_coverage=true&rft.mms_id=9911670637102420 (accessed February 01, 2023).

Friston, K., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., et al. (2008). Multiple sparse priors for the M/EEG inverse problem. *Neuroimage* 39, 1104–1120. doi: 10.1016/j.neuroimage.2007.09.048

Friston, K., Henson, R., Phillips, C., and Mattout, J. (2006). Bayesian estimation of evoked and induced responses. *Hum. Brain Mapp.* 27, 722–735. doi: 10.1002/hbm.20214

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., and Penny, W. (2007). Variational free energy and the laplace approximation. *Neuroimage* 34, 220–234. doi: 10.1016/j.neuroimage.2006.08.035

Friston, K. J. (2009). Modalities, modes, and models in functional neuroimaging. *Science* 326, 399–403. doi: 10.1126/science.1174521

Friston, K. J., Bastos, A., Litvak, V., Stephan, K. E., Fries, P., and Moran, R. J. (2012). DCM for complex-valued data: cross-spectra, coherence and phase-delays. *Neuroimage* 59, 439–455. doi: 10.1016/j.neuroimage.2011.07.048

Friston, K. J., Trujillo-Barreto, N., and Daunizeau, J. (2008). DEM: a variational treatment of dynamic systems. *Neuroimage* 41, 849–885. doi: 10.1016/j.neuroimage.2008.02.054

Gershman, S., Hoffman, M., and Blei, D. (2012). Nonparametric variational inference. *ArXiv Preprint ArXiv:1206.4665.* 53, 1–12.

Gershman, S. J., and Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *J. Math. Psychol.* 56, 1–12. doi: 10.1016/j.jmp.2011.08.004

Ghahramani, Z., and Beal, M. J. (2001). Propagation algorithms for variational bayesian learning. *Adv. Neural Inform. Process. Syst.* 3, 507–513.

Ghahramani, Z., and Beal, M. M. J. (2000). Variational inference for Bayesian mixtures of factor analysers. *Adv. Neural Inform. Process. Syst.* 12, 449–455.

Golub, G. H., and Van Loan, C. F. (2013). *Matrix Computations.* Charles Village, MA: JHU Press. doi: 10.56021/9781421407944

Gramfort, A., Kowalski, M., and Hämäläinen, M. (2012). Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. *Phys. Med. Biol.* 57, 1937. doi: 10.1088/0031-9155/57/7/1937

Gramfort, A., Strohmeier, D., Haueisen, J., Hämäläinen, M. S., and Kowalski, M. (2013). Time-frequency mixed-norm estimates: sparse M/EEG imaging with non-stationary source activations. *Neuroimage* 70, 410–422. doi: 10.1016/j.neuroimage.2012.12.051

Grave de Peralta Menendez, R., Murray, M. M., Michel, C. M., Martuzzi, R., and Andino, S. L. G. (2004). Electrical neuroimaging based on biophysical constraints. *Neuroimage* 21, 527–539. doi: 10.1016/j.neuroimage.2003.09.051

Grech, R., Cassar, T., Muscat, J., Camilleri, K. P., Fabri, S. G., Zervakis, M., et al. (2008). Review on solving the inverse problem in EEG source analysis. *J. Neuroeng. Rehabil.* 5, 1–33. doi: 10.1186/1743-0003-5-25

Grova, C., Daunizeau, J., Kobayashi, E., Bagshaw, A. P., Lina, J. M., Dubeau, F., et al. (2008). Concordance between distributed EEG source localization and simultaneous EEG-FMRI studies of epileptic spikes. *Neuroimage* 39, 755–774. doi: 10.1016/j.neuroimage.2007.08.020

Grova, C., Daunizeau, J., Lina, J. M., Bénar, C. G., Benali, H., and Gotman, J. (2006a). Evaluation of EEG localization methods using realistic simulations of interictal spikes. *Neuroimage* 29, 734–753. doi: 10.1016/j.neuroimage.2005.08.053

Grova, C., Makni, S., Flandin, G., Ciuciu, P., Gotman, J., and Poline, J. B. (2006b). Anatomically informed interpolation of FMRI data on the cortical surface. *Neuroimage* 31, 1475–1486. doi: 10.1016/j.neuroimage.2006.02.049

Hadamard, J., and Morse, P. M. (1953). Lectures on Cauchy's problem in linear partial differential equations. *Phys. Today* 6, 18–18. doi: 10.1063/1.3061337

Hallez, H., Vanrumste, B., Grech, R., Muscat, J., Clercq, W. D., Vergult, A., et al. (2007). Review on solving the forward problem in EEG source analysis. *J. Neuroeng. Rehabil.* 4, 46. doi: 10.1186/1743-0003-4-46

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., and Lounasmaa, O. V. (1993). Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Mod. Phys.* 65, 413. doi: 10.1103/RevModPhys.65.413

Hämäläinen, M. S., and Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* 32, 35–42. doi: 10.1007/BF02512476

Hancock, G. R., and Samuelsen, K. M. (2007). *Advances in Latent Variable Mixture Models.* Charlotte, NC: IAP.

Harrison, L. M., Penny, W., Daunizeau, J., and Friston, K. J. (2008). Diffusion-based spatial priors for functional magnetic resonance images. *Neuroimage* 41, 408–423. doi: 10.1016/j.neuroimage.2008.02.005

Haufe, S., and Ewald, A. (2019). A simulation framework for benchmarking EEG-based brain connectivity estimation methodologies. *Brain Topogr.* 32, 625–642. doi: 10.1007/s10548-016-0498-y

Haufe, S., Nikulin, V. V., Müller, K. R., and Nolte, G. (2013). A critical assessment of connectivity measures for EEG data: a simulation study. *Neuroimage* 64, 120–133. doi: 10.1016/j.neuroimage.2012.09.036

Haufe, S., Nikulin, V. V., Ziehe, A., Müller, K. R., and Nolte, G. (2008). Combining sparsity and rotational invariance in EEG/MEG source reconstruction. *Neuroimage* 42, 726–738. doi: 10.1016/j.neuroimage.2008.04.246

Hauk, O. (2004). Keep it simple: a case for using classical minimum norm estimation in the analysis of EEG and MEG data. *Neuroimage* 21, 1612–1621. doi: 10.1016/j.neuroimage.2003.12.018

He, B., Astolfi, L., Valdes-Sosa, P. A., Marinazzo, D., Palva, S. O., Benar, C.-G., et al. (2019). Electrophysiological brain connectivity: theory and implementation. *IEEE Trans. Biomed. Eng.* 66, 2115–2137. doi: 10.1109/TBME.2019.2913928

He, B., Sohrabpour, A., Brown, E., and Liu, Z. (2018). Electrophysiological source imaging: a noninvasive window to brain dynamics. *Annu. Rev. Biomed. Eng.* 20:171–196. doi: 10.1146/annurev-bioeng-062117-120853

Henson, R. N., Wakeman, D. G., Litvak, V., and Friston, K. J. (2011). A parametric empirical bayesian framework for the EEG/MEG inverse problem: generative models for multi-subject and multi-modal integration. *Front. Hum. Neurosci.* 5, 76. doi: 10.3389/fnhum.2011.00076

Hindriks, R. (2020). A methodological framework for inverse-modeling of propagating cortical activity using MEG/EEG. *Neuroimage* 223:117345. doi: 10.1016/j.neuroimage.2020.117345

Hipp, J. F., Hawellek, D. J., Corbetta, M., Siegel, M., and Engel, A. K. (2012). Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nat. Neurosci.* 15, 884–890. doi: 10.1038/nn.3101

Hoerl, A. E., and Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12, 55–67. doi: 10.1080/00401706.1970.10488634

Hsiao, T., Rangarajan, A., and Gindi, G. (1998). "Joint-MAP reconstruction/segmentation for transmission tomography using mixture-models as priors," in *1998 IEEE Nuclear Science Symposium Conference Record* 1998. IEEE Nuclear Science Symposium and Medical Imaging Conference (Cat. No. 98CH36255), Vol. 3 (Toronto, ON: IEEE), 1689–1693.

Hsiao, T., Rangarajan, A., and Gindi, G. (2002). Joint-MAP Bayesian tomographic reconstruction with a gamma-mixture prior. *IEEE Trans. Image Process.* 11, 1466–1477. doi: 10.1109/TIP.2002.806254

Hu, Y., Zhang, D., Ye, J., Li, X., and He, X. (2012). Fast and accurate matrix completion via truncated nuclear norm regularization. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 2117–2130. doi: 10.1109/TPAMI.2012.271

Jirsa, V., and Haken, H. (1996). Field theory of electromagnetic brain activity. *Phys. Rev. Lett.* 77, 960–963. doi: 10.1103/PhysRevLett.77.960

Jirsa, V. K. (2004). Connectivity and dynamics of neural information processing. *Neuroinformatics* 2, 183–204. doi: 10.1385/NI:2:2:183

Jirsa, V. K., and Haken, H. (1997). A derivation of a macroscopic field theory of the brain from the quasi-microscopic neural dynamics. *Phys. D Nonlinear Phenomena* 99, 503–526. doi: 10.1016/S0167-2789(96)00166-2

Kadanoff, L. P. (2009). More is the same; phase transitions and mean field theories. *J. Stat. Phys.* 137, 777–797. doi: 10.1007/s10955-009-9814-1

Kaplan, A., and Tichatschke, R. (1998). Proximal point methods and nonconvex optimization. *J. Glob. Optimiz.* 13, 389–406. doi: 10.1023/A:1008321423879

Kim, E., Lee, M., and Oh, S. (2015). "Elastic-net regularization of singular values for robust subspace learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Boston, MA), 915–923. doi: 10.1109/CVPR.2015.7298693

Kindermann, R., Snell, J. L., Society, A. M., and Collection, K. M. R. (1980). *Markov Random Fields and Their Applications. Contemporary Mathematics - American Mathematical Society.* American Mathematical Society. Available online at: https://books.google.co.uk/books?id=NeVQAAAAMAAJ (accessed February 01, 2023). doi: 10.1090/conm/001

Knösche, T. R., and Haueisen, J. (2022). *EEG/MEG Source Reconstruction: Textbook for Electro-and Magnetoencephalography.* Midtown Manhattan, NY: Springer. doi: 10.1007/978-3-030-74918-7

Kobayashi, K., Yoshinaga, H., Oka, M., Ohtsuka, Y., and Gotman, J. (2003). A simulation study of the error in dipole source localization for EEG spikes with a realistic head model. *Clin. Neurophysiol.* 114, 1069–1078. doi: 10.1016/S1388-2457(03)00064-6

Kowalski, M., and Torrésani, B. (2009a). "Structured sparsity: from mixed norms to structured shrinkage," in *SPARS'09-Signal Processing with Adaptive Sparse Structured Representations* (Saint-Malo).

Kowalski, M., and Torrésani, B. (2009b). Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients. *Signal Image Video Process.* 3, 251–264. doi: 10.1007/s11760-008-0076-1

Lafferty, J., McCallum, A., and Pereira, F. C. N. (2001). "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *ICML '01: Proceedings of the Eighteenth International Conference on Machine Learning* (San Francisco, CA), 282–289. doi: 10.5555/645530.655813

Landau, L. D., and Lifshitz, E. M. (1980). "Chapter III - the gibbs distribution," in *Statistical Physics, 3rd Edn.*, eds L. D. Landau and E. M. Lifshitz (Oxford: Butterworth-Heinemann), 79–110. doi: 10.1016/B978-0-08-057046-4.50010-5

Larson-Prior, L. J., Oostenveld, R., Penna, S. D., Michalareas, G., Prior, F., Babajani-Feremi, A., et al. (2013). Adding dynamics to the human connectome project with MEG. *Neuroimage* 80:190–201. doi: 10.1016/j.neuroimage.2013.05.056

Le Boudec, J.-Y., McDonald, D., and Mundinger, J. (2007). "A generic mean field convergence result for systems of interacting objects," in *Fourth International Conference on the Quantitative Evaluation of Systems (QEST 2007)* (Edinburgh: IEEE), 3–18. doi: 10.1109/QEST.2007.8

Le Van Quyen, M., and Bragin, A. (2007). Analysis of dynamic brain oscillations: methodological advances. *Trends Neurosci.* 30, 365–373. doi: 10.1016/j.tins.2007.05.006

Lei, X., Ostwald, D., Hu, J., Qiu, C., Porcaro, C., Bagshaw, A. P., et al. (2011). Multimodal functional network connectivity: an EEG-FMRI fusion in network space. *PLoS ONE* 6, e24642. doi: 10.1371/journal.pone.0024642

Li, J., Li, R., and Wu, R. (2011). Bayesian group LASSO for nonparametric varying-coefficient. *Group* 16802, 1–19.

Li, Q., and Lin, N. (2010). The Bayesian elastic net. *Bayesian Anal.* 5, 151–170. doi: 10.1214/10-BA506

Li, Z., and Tian, T. S. (2011). A spatio-temporal solution for the EEG/MEG inverse problem using group penalization methods. *Stat. Interface* 4, 521–533. doi: 10.4310/SII.2011.v4.n4.a10

Lindsay, B. G. (1995). Mixture models: theory, geometry, and applications. *NSF-CBMS Regional Conf. Ser. Probab. Statist.* 5, 163. doi: 10.1214/cbms/1462106013

Liu, C., and Rubin, D. B. (1994). The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. *Biometrika* 81, 633–648. doi: 10.1093/biomet/81.4.633

Lopes da Silva, F. (2013). EEG and MEG: relevance to neuroscience. *Neuron* 80, 1112–1128. doi: 10.1016/j.neuron.2013.10.017

Lopes da Silva, F. H., Hoeks, A., Smits, H., and Zetterberg, L. H. (1974). Model of brain rhythmic activity - the alpha-rhythm of the thalamus. *Kybernetik* 15, 27–37. doi: 10.1007/BF00270757

Lopes da Silva, F. H., Wieringa, H. J., and Peters, M. J. (1991). Source localization of EEG versus MEG: empirical comparison using visually evoked responses and theoretical considerations. *Brain Topogr.* 4, 133–142. doi: 10.1007/BF01132770

MacKay, D. J. C. (1999). Comparison of approximate methods for handling hyperparameters. *Neural Comput.* 11, 1035–1068. doi: 10.1162/089976699300016331

MacKay, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms.* Cambridge: Cambridge University Press.

MacKay, D. J. C. J. C. (1998). Choice of basis for laplace approximation. *Mach. Learn.* 33, 77–86. doi: 10.1023/A:1007558615313

Mahjoory, K., Nikulin, V. V., Botrel, L., Linkenkaer-Hansen, K., Fato, M. M., and Haufe, S. (2017). Consistency of EEG source localization and connectivity estimates. *Neuroimage* 152:590–601. doi: 10.1016/j.neuroimage.2017.02.076

Makeig, S. (2002). Response: event-related brain dynamics - unifying brain electrophysiology. *Trends Neurosci.* 25, 390. doi: 10.1016/S0166-2236(02)02198-7

Makeig, S., Delorme, A., Westerfield, M., Jung, T. P., Townsend, J., Courchesne, E., et al. (2004). Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biol.* 2, e176. doi: 10.1371/journal.pbio.0020176

Makeig, S., Westerfield, M., Jung, T. P., Covington, J., Townsend, J., Sejnowski, T. J., et al. (1999). Functionally independent components of the late positive event-related potential during visual spatial attention. *J. Neurosci.* 19, 2665–2680. doi: 10.1523/JNEUROSCI.19-07-02665.1999

Mantini, D., Perrucci, M. G., Gratta, C. D., Romani, G. L., and Corbetta, M. (2007). Electrophysiological signatures of resting state networks in the human brain. *Proc. Nat. Acad. Sci. U. S. A.* 104, 13170–13175. doi: 10.1073/pnas.0700668104

Marinazzo, D., Riera, J. J., Marzetti, L., Astolfi, L., Yao, D., and Sosa, P. A. V. (2019). Controversies in EEG source imaging and connectivity: modeling, validation, benchmarking. *Brain Topogr.* 32, 527–529. doi: 10.1007/s10548-019-00709-9

Marzetti, L., Gratta, C. D., and Nolte, G. (2008). Understanding brain connectivity from EEG data by identifying systems composed of interacting sources. *Neuroimage* 42, 87–98. doi: 10.1016/j.neuroimage.2008.04.250

Mattout, J., Phillips, C., Penny, W. D., Rugg, M. D., and Friston, K. J. (2006). MEG source localization under multiple constraints: an extended bayesian framework. *Neuroimage* 30, 753–767. doi: 10.1016/j.neuroimage.2005.10.037

Maurer, K., and Dierks, T. (2012). *Atlas of Brain Mapping: Topographic Mapping of EEG and Evoked Potentials.* Berlin; Heidelberg: Springer Science & Business Media.

McLachlan, G. J., and Basford, K. E. (1988). *Mixture Models: Inference and Applications to Clustering.* Vol. 38. New York, NY: M. Dekker New York. doi: 10.2307/2348072

Moran, R., Pinotsis, D., a., and Friston, K. (2013). Neural masses and fields in dynamic causal modeling. *Front. Comput. Neurosci.* 7, 57. doi: 10.3389/fncom.2013.00057

Nagasaka, Y., Shimoda, K., and Fujii, N. (2011). Multidimensional recording (MDR) and data sharing: an ecological open research and educational platform for neuroscience. *PLoS ONE* 6, e22561. doi: 10.1371/journal.pone.0022561

Nguyen, T., and Bonilla, E. (2013). "Efficient variational inference for gaussian process regression networks," in *Artificial Intelligence and Statistics* (Scottsdale, AZ: PMLR), 472–480.

Niedermeyer, E., and da Silva, F. H. L. (2005). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields.* LWW Doody's All Reviewed

Collection. Lippincott Williams & Wilkins. Available online at: https://books.google.com.sg/books?id=tndqYGPHQdEC (accessed February 01, 2023).

Nolte, G., Galindo-Leon, E., Li, Z., Liu, X., and Engel, A. K. (2020). Mathematical relations between measures of brain connectivity estimated from electrophysiological recordings for gaussian distributed data. *Front. Neurosci.* 14, 577574. doi: 10.3389/fnins.2020.577574

Nummenmaa, A., Auranen, T., Hämäläinen, M. S., Jääskeläinen, I. P., Lampinen, J., Sams, M., et al. (2007). Hierarchical Bayesian estimates of distributed MEG sources: theoretical aspects and comparison of variational and MCMC methods. *Neuroimage* 35, 669–685. doi: 10.1016/j.neuroimage.2006.05.001

Nunez, P. L. (1974). The brain wave equation: a model for the EEG. *Math. Biosci.* 21, 279–97. doi: 10.1016/0025-5564(74)90020-0

Nunez, P. L., Silberstein, R. B., Cadusch, P. J., Wijesinghe, R. S., Westdorp, A. F., and Srinivasan, R. (1994). A theoretical and experimental study of high resolution eeg based on surface laplacians and cortical imaging. *Electroencephalogr. Clin. Neurophysiol.* 90, 40–57. doi: 10.1016/0013-4694(94)90112-0

Nunez, P. L., and Srinivasan, R. (2006). *Electric Fields of the Brain: The Neurophysics of EEG, 2nd Edn.* New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195050387.001.0001

Palva, J. M., Wang, S. H., Palva, S., Zhigalov, A., Monto, S., Brookes, M. J., et al. (2018). Ghost interactions in MEG/EEG source space: a note of caution on inter-areal coupling measures. *Neuroimage* 173, 632–643. doi: 10.1016/j.neuroimage.2018.02.032

Papadopoulou, M., Friston, K., and Marinazzo, D. (2019). Estimating directed connectivity from cortical recordings and reconstructed sources. *Brain Topogr.* 32, 741–752. doi: 10.1007/s10548-015-0450-6

Park, T., and Casella, G. (2008). The Bayesian lasso. *J. Am. Stat. Assoc.* 103, 681–686. doi: 10.1198/016214508000000337

Pascual-Marqui, R. D., Pascual-Montano, A. D., Lehmann, D., Kochi, K., Esslen, M., Jancke, L., et al. (2006). Exact low resolution brain electromagnetic tomography (ELORETA). *Neuroimage* 31 (Suppl. 1):S86. doi: 10.1016/S1053-8119(08)70001-6

Paz-Linares, D., Gonzalez-Moreira, E., Areces-Gonzalez, A., Wang, Y., Li, M., Martinez-Montes, E., et al. (2018). Identification of oscillatory brain networks with hidden gaussian graphical spectral models of EEG/MEG. *ArXiv [Preprint].* arXiv:1810.01174. doi: 10.48550/arXiv.1810.01174

Paz-Linares, D., Vega-Hernández, M., Rojas-López, P. A., Valdés-Hernández, P. A., Martínez-Montes, E., and Valdés-Sosa, P. A. (2017). Spatio temporal EEG source imaging with the hierarchical bayesian elastic net and elitist lasso models. *Front. Neurosci.* 11, 635. doi: 10.3389/fnins.2017.00635

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* New York, NY: ACM. doi: 10.1016/B978-0-08-051489-5.50008-4

Pearl, J. (2022). "Reverend Bayes on inference engines: a distributed hierarchical approach," in *Probabilistic and Causal Inference: The Works of Judea Pearl* (San Rafael, CA: Morgan & Claypool), 129–38. doi: 10.1145/3501714.3501727

Petersen, K. B., and Pedersen, M. S. (2008). *The Matrix Cookbook.* Lyngby: Technical University of Denmark. Vol. 7, 510.

Piastra, M. C., Nüßing, A., Vorwerk, J., Clerc, M., Engwer, C., and Wolters, C. H. (2020). A comprehensive study on electroencephalography and magnetoencephalography sensitivity to cortical and subcortical sources. *Hum. Brain Mapp.* 42, 978–992. doi: 10.1002/hbm.25272

Picton, T. W., and Hillyard, S. A. (1974). Human auditory evoked potentials. II: effects of attention. *Electroencephal. Clin. Neurophysiol.* 36, 191–200. doi: 10.1016/0013-4694(74)90156-4

Picton, T. W., Hillyard, S. A., Krausz, H. I., and Galambos, R. (1974). Human auditory evoked potentials. I: evaluation of components. *Electroencephalogr. Clin. Neurophysiol.* 36:179–190. doi: 10.1016/0013-4694(74)90155-2

Piotrowski, T., and Yamada, I. (2008). MV-PURE estimator: minimum-variance pseudo-unbiased reduced-rank estimator for linearly constrained ill-conditioned inverse problems. *IEEE Trans. Signal Process.* 56, 3408–3423. doi: 10.1109/TSP.2008.921716

Reid, A. T., Headley, D. B., Mill, R. D., Sanchez-Romero, R., Uddin, L. Q., Marinazzo, D., et al. (2019). Advancing functional connectivity research from association to causation. *Nat. Neurosci.* 22, 1751–1760. doi: 10.1038/s41593-019-0510-4

Riera, J. J., and Fuentes, M. E. (1998). Electric lead field for a piecewise homogeneous volume conductor model of the head. *IEEE Trans. Biomed. Eng.* 45, 746–753. doi: 10.1109/10.678609

Riesz, F. (1910). Untersuchungen über systeme integrierbarer funktionen. *Math. Annalen* 69, 449–497. doi: 10.1007/BF01457637

Rosa, M. J., Daunizeau, J., and Friston, K. J. (2010). EEG-FMRI integration: a critical review of biophysical modeling and data analysis approaches. *J. Integr. Neurosci.* 9, 453–476. doi: 10.1142/S0219635210002512

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. U. S. A.* 42, 43. doi: 10.1073/pnas.42.1.43

Roweis, S., and Ghahramani, Z. (1999). A unifying review of linear gaussian models. *Neural Comput.* 11, 305–345. doi: 10.1162/089976699300016674

Rudin, W. (1970). *Real and Complex Analysis P. 2.* New York, NY: McGraw-Hill.

Salmelin, R. H., and Hämäläinen, M. S. (1995). Dipole modelling of MEG rhythms in time and frequency domains. *Brain Topogr.* 7, 251–257. doi: 10.1007/BF01202384

Schatten, R. (2013). *Norm Ideals of Completely Continuous Operators.* Vol. 27. Midtown Manhattan, NY: Springer-Verlag.

Schoffelen, J., and Gross, J. (2009). Source connectivity analysis with MEG and EEG. *Hum. Brain Mapp.* 30, 1857–1865. doi: 10.1002/hbm.20745

Stokes, P. A., and Purdon, P. L. (2017). A study of problems encountered in granger causality analysis from a neuroscience perspective. *Proc. Nat. Acad. Sci. U. S. A.* 201704663. doi: 10.1073/pnas.1704663114

Sun, T., and Zhang, C.-H. (2012). Calibrated elastic regularization in matrix completion. *Adv. Neural Inf. Process. Syst.* 25, 863–871.

Tarantola, A. (2005). "Inverse problem theory and methods for model parameter estimation," in *Inverse Problem Theory and Methods for Model Parameter Estimation* (Philadelphia, PA: SIAM). doi: 10.1137/1.9780898717921

Tewarie, P., Liuzzi, L., O'Neill, G. C., Quinn, A. J., Griffa, A., Woolrich, M. W., et al. (2019). Tracking dynamic brain networks using high temporal resolution MEG measures of functional connectivity. *Neuroimage* 200, 38–50. doi: 10.1016/j.neuroimage.2019.06.006

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B* 58, 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x

Tikhonov, A. N., and Arsenin, V. I. A. (1977). *Solutions of Ill-Posed Problems.* Washington; New York, NY: Winston; Distributed solely by Halsted Press. Available online at: http://books.google.com/books?id=ECrvAAAAMAAJ (accessed February 01, 2023).

Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* 1, 211–244. doi: 10.1162/15324430152748236

Tirer, T., and Giryes, R. (2020). Back-projection based fidelity term for ill-posed linear inverse problems. *IEEE Trans. Image Process.* 29, 6164–6179. doi: 10.1109/TIP.2020.2988779

Valdés, P., Bosch, J., Grave, R., Hernandez, J., Riera, J., Pascual, R., et al. (1992). Frequency domain models of the EEG. *Brain Topogr.* 4, 309–319. doi: 10.1007/BF01135568

Valdes-Sosa, P. A., Sanchez-Bornot, J. M., Sotero, R. C., Iturria-Medina, Y., Aleman-Gomez, Y., Bosch-Bayard, J., et al. (2009). Model driven EEG/FMRI fusion of brain oscillations. *Hum. Brain Mapp.* 30, 2701–2721. doi: 10.1002/hbm.20704

Van de Steen, F., Faes, L., Karahan, E., Songsiri, J., Valdes-Sosa, P. A., and Marinazzo, D. (2019). Critical comments on EEG sensor space dynamical connectivity analysis. *Brain Topogr.* 32, 643–654. doi: 10.1007/s10548-016-0538-7

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., and Ugurbil, K. (2013). The WU-Minn human connectome project: an overview. *Neuroimage* 80:62–79. doi: 10.1016/j.neuroimage.2013.05.041

Van Veen, B. D., Drongelen, W. V., Yuchtman, M., and Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans. Biomed. Eng.* 44, 867–880. doi: 10.1109/10.623056

Varela, F., Lachaux, J.-P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2, 229–239. doi: 10.1038/35067550

Vega-Hernández, M., Martínez-Montes, E., Sánchez-Bornot, J. M., Lage-Castellanos, A., and Valdés-Sosa, P. A. (2008). Penalized least squares methods for solving the EEG inverse problem. *Stat. Sin.* 18, 1535–1551. Available online at: https://www3.stat.sinica.edu.tw/statistica/J18N4/j18n416/j18n416.html

Vidaurre, D., Abeysuriya, R., Becker, R., Quinn, A. J., Alfaro-Almagro, F., Smith, S. M., et al. (2018a). Discovering dynamic brain networks from big data in rest and task. *Neuroimage* 180, 646–656. doi: 10.1016/j.neuroimage.2017.06.077

Vidaurre, D., Hunt, L. T., Quinn, A. J., Hunt, B. A. E. E., Brookes, M. J., Nobre, A. C., et al. (2018b). Spontaneous cortical activity transiently organises into frequency specific phase-coupling networks. *Nat. Commun.* 9, 2987. doi: 10.1038/s41467-018-05316-z

Vinck, M., and Perrenoud, Q. (2019). Layers of rhythms—from cortical anatomy to dynamics. *Neuron* 101, 358–360. doi: 10.1016/j.neuron.2019.01.028

Wang, Q., Valdés-Hernández, P. A., Paz-Linares, D., Bosch-Bayard, J., Oosugi, N., Komatsu, M., et al. (2019). EECoG-comp: an open source platform for concurrent EEG/ECoG comparisons—applications to connectivity studies. *Brain Topogr.* 32, 550–568. doi: 10.1007/s10548-019-00708-w

Weiss, P. (1907). L'hypothèse Du Champ Moléculaire et La Propriété Ferromagnétique. *J. Phys. Theor. Appl.* 6, 661–690. doi: 10.1051/jphystap:019070060066100

Weiss, Y. (2001). *Comparing the Mean Field Method and Belief Propagation for Approximate Inference in MRFs.* Burlington, MA: Morgan Kaufmann.

Wens, V., Marty, B., Mary, A., Bourguignon, M., Beeck, M. O., d., et al. (2015). A geometric correction scheme for spatial leakage effects in MEG/EEG

seed-based functional connectivity mapping. *Hum. Brain Mapp.* 36, 4604–4621. doi: 10.1002/hbm.22943

Wipf, D., and Nagarajan, S. (2009). A unified Bayesian framework for MEG/EEG source imaging. *Neuroimage* 44, 947–966. doi: 10.1016/j.neuroimage.2008.02.059

Wipf, D. P., Ramirez, R. R., Palmer, J. A., Makeig, S., and Rao, B. D. (2006). *Automatic Relevance Determination for Source Localization with MEG and EEG Data.* Technical Report. San Diego, CA: University of California.

Yedidia, J. S., Freeman, W. T., and Weiss, Y. (2003). Understanding belief propagation and its generalizations. *Exploring Art. Intell. New Millennium* 8, 18–9448. doi: 10.5555/779343.779352

Yeredor, A. (2000). The joint MAP-ml criterion and its relation to ml and to extended least-squares. *IEEE Trans. Signal Process.* 48, 3484–3492. doi: 10.1109/78.887041

Yuan, M., and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B* 68, 49–67. doi: 10.1111/j.1467-9868.2005.00532.x

Zhang, Z., Lai, Z., Xu, Y., Shao, L., Wu, J., and Xie, G.-S. (2017). Discriminative elastic-net regularized linear regression. *IEEE Trans. Image Process.* 26, 1466–1481. doi: 10.1109/TIP.2017.2651396

Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., et al. (2015). "Conditional random fields as recurrent neural networks," in *Proceedings of the IEEE International Conference on Computer Vision* (Santiago), 1529–1537. doi: 10.1109/ICCV.2015.179

Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* 67, 301–320. doi: 10.1111/j.1467-9868.2005.00503.x