sparse **B**ayesian learning **M**eets cortical **W**avelet**s**

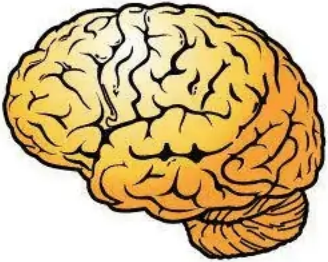
Bruno Torrésani
Institut de Mathématiques de Marseille
Aix-Marseille Université

—

joint work with S. Mokhtari and the BMWs gang

BMWs meeting,
Fréjus, June 2025

THE BRAIN IS THE MOST IMPORTANT ORGAN YOU HAVE



ACCORDING TO THE BRAIN.

Context: M/EEG source localization/inverse problem I

Observation model :

$$Z_0 = G_0 S + B_0$$

- Observations $Z_0[j, t]$: approximately $J \approx 250$ sensors, and T time samples
- Unknown sources $S[k, t]$: cortex discretization on approximately $K \approx 10000$ points k
- Leadfield matrix $G_0 : J \times K$
- Baseline $B_0 : \; J \times T$

Context: M/EEG source localization/inverse problem II

Spatial whitening

- An operation that turn baseline sensor data into samples of independent, zero mean, unit variance random samples (ideally iid, which is true only if time whitening is done too)
- Done by a whitening matrix $\Upsilon \approx \Sigma_B^{-1/2}$, where Σ_B is an estimate of the baseline covariance matrix.

$$Z \longleftarrow \Upsilon Z_0 \ , \qquad G \longleftarrow \Upsilon G_0 \ , \qquad B_0 \longleftarrow \text{spatially white noise}$$

yields spatially whitened observation model

$$Z = GS + B$$

Goals

In the framework of distributed models

- Search for spatially extended brain activity
- Reduce standard biases
- Obtain quantitatively relevant estimates

Sparsity

- Accurately solving such an ill conditioned inverse problem is impossible, unless the number of unknown is controlled
- Insider trading: you know (or decide) in advance that relevant activity is characterized by $K' \ll K$ variables, with $K' \approx J$... and solve the problem in the source subspace generated by these
- Sparsity: you let your algorithm decide which ones of the variables are relevant

Variational formulations: penalized least squares I

Classical variational approaches solve problems of the form

$$S_* = \underset{S\in\mathbb{R}^{K\times T}}{\arg\min}\ \frac{1}{2}\|Z - GS\|_F^2 + \Psi_\theta(S)$$

(remember data are spatially whitened) where

- $\Psi_\theta(S)$ attempts to favour specific behaviors, depends on a (multi)parameter θ
- Equivalent to MAP estimation in a Bayesian perspective, with $p(S) \sim e^{-\Psi_\theta(S)}$.

Quadratic penalizations I

Minimal norm estimates: MNE, wMNE,...

Quadratic penalization (gaussian prior in MAP estimate)

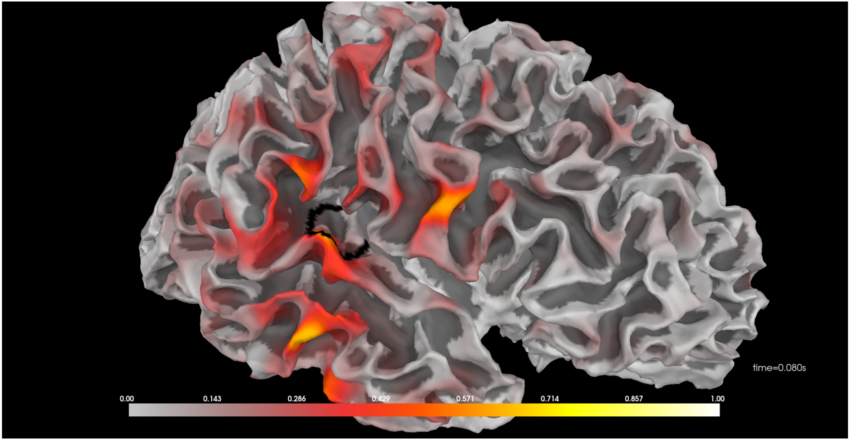
$$\Psi_\theta(S) = \frac{1}{2}\|\Gamma^{-1/2}S\|_F^2$$

Closed form, linear solution

$$S_* = HZ = \Gamma G^\top \left(\mathrm{I}_J + G\Gamma G^\top\right)^{-1} Z$$

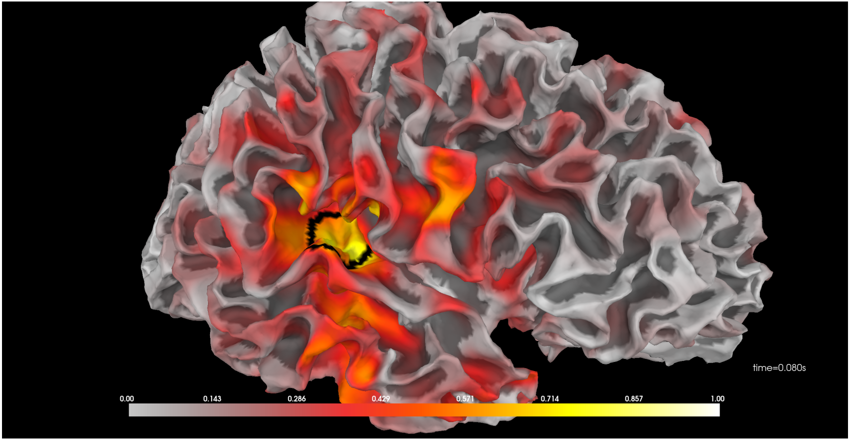
- ☺ Fast computation
- ☹ Important blurring, depth bias
- ☹ No coupling in time
- ☹ Choosing Γ (prior source covariance) is not so easy
  - MNE: $\Gamma = \lambda^{-1}\mathrm{I}$, we use $\lambda_{\mathrm{MNE}} = \|G_0\|_F^2/(\mathrm{SNR}^2 - 1)\mathrm{Tr}(\Sigma_{B_0})$.
  - eLORETA: data driven estimate of a diagonal Γ (provides exact localization with zero error to test point sources).

Quadratic penalizations II



Response to left auditory stimulus
(Black annulus: MNE-Python "Aud-rh" label)
MNE solution: $\Gamma = \lambda I$ (data driven λ)

Quadratic penalizations III



Response to left auditory stimulus
(Black annulus: MNE-Python "Aud-rh" label)
eLORETA solution: data driven diagonal Γ

Quadratic penalizations IV

Post-processing required:

- Smoothing
- Multiple testing
- Validation by expert
- Many ad hoc, more or less legal, manipulations
- ...

Sparse solution I

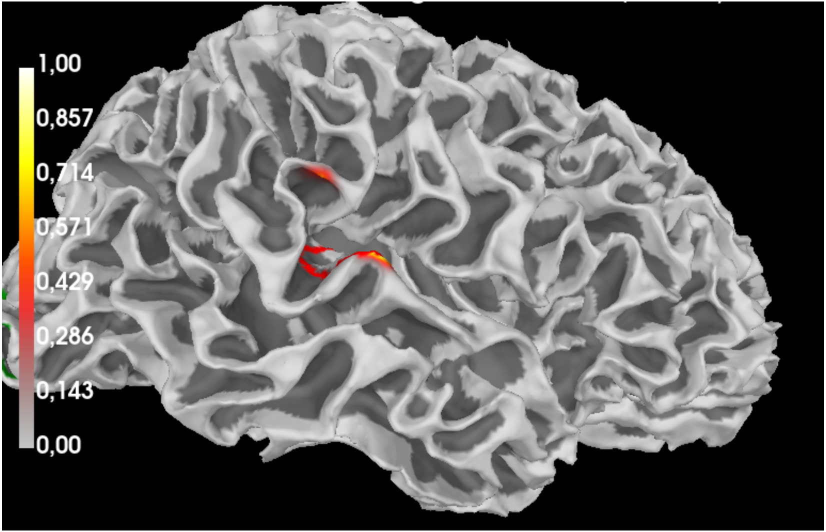
Minimum current: MCE

Introduce sparsity in penalization:

$$\Psi_\lambda(S) = \lambda \|S\|_1 = \lambda \sum_{k,t} |S[k,t]|$$

- ☺ Sparsity (ℓ^1 penalizations induces thresholding)
- ☹ Too sparse: spiky behavior (depends on constraint parameter)
- ☹ Still too superficial
- ☹ No closed form solution, requires optimization algorithm (ISTA, FISTA, ADMM,..), fairly slow
- ☹ λ difficult to tune (there is a heuristics based upon SNR computation)
- ☹ Still no coupling in time

Sparse solution II



Response to left auditory stimulus
MCE solution, with data driven λ

Sparse solution III

Other sparse solvers

MxNE (social sparsity): add time dependence in sources S:

$$\Psi_\lambda(S) = \lambda\|S\|_{21} = \lambda \sum_k \sqrt{\sum_t S[k,t]^2}$$

- ☺ Coupling in time (thresholding on norms $\|S[k,\cdot]\|_2$)
- ☹ Still too sparse, with time persistent spike locations, same optimization algorithms, λ difficult to tune. We use $\lambda_{\text{MCE}} = \sqrt{2\lambda_{\text{MNE}}}$

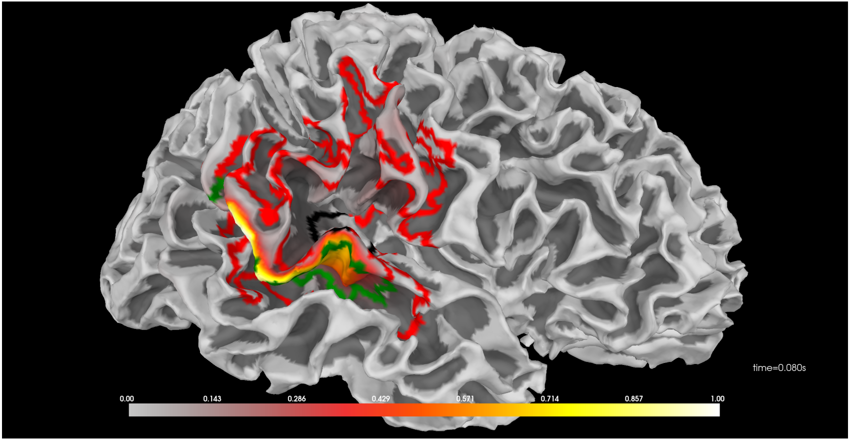
VB-SCCD: total variation based (social) sparsity (using a gradient on the cortex mesh)

$$\Psi_\lambda(S) = \lambda\left(\epsilon\|\nabla S\|_1 + (1-\epsilon)\|S\|_1\right)$$

(the 1-norm may be replaced with 21-norm)

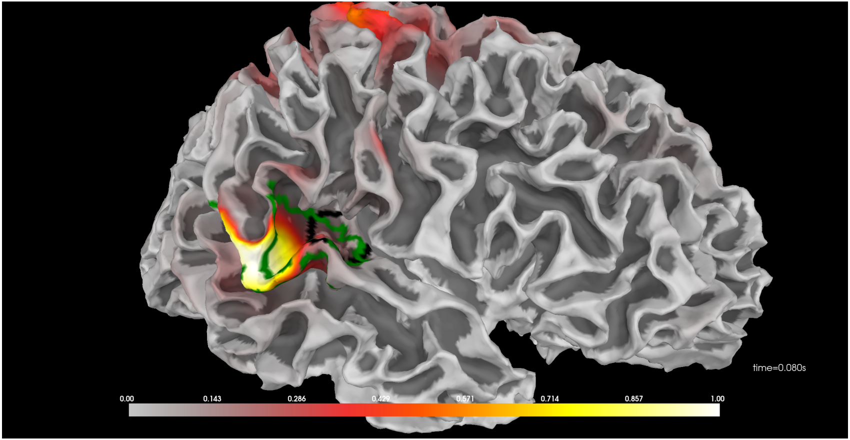
- ☺ Extended solutions
- ☹ Parameters λ, ϵ extremely difficult to tune

Sparse solution IV



Response to left auditory stimulus
VB-SCCD solution, $\lambda = \lambda_{\text{MCE}}$, $\epsilon = .3$
Black annulus: MNE-Python "aud-rh" label
Red and green lines: 1% and 10% level curves

Sparse solution V



Response to left auditory stimulus
VB-SCCD solution, $\lambda = \lambda_{\text{MCE}}$, $\epsilon = .7$
Black annulus: MNE-Python "aud-rh" label
Red and green lines: 1% and 10% level curves

Summary

Problems:

- An extremely ill conditioned inverse problem
- In such cases sparsity is a good paradigm: state that the solution can be written as a linear combination of very few "building blocks"
- Sparsity in the cortex domain yield extremely sparse solutions (for which there exist better methods)
- Parameter tuning is a real issue...

Solutions: to estimate extended brain activity

- Cortical wavelets
- Sparse Bayesian Learning

Wavelets on the cortex I

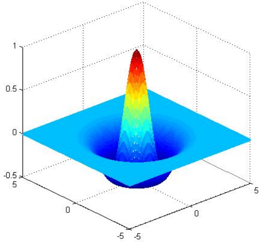
The case of the plane: Oscillatory functions ψ such that a suitable family of translates and dilates span $L^2(\mathbb{R}^2)$. Possible constructions:

- Tensor products of 1D wavelets vs 2D radial functions
- Wavelet bases vs wavelet frames

$$\psi_{ju}(x) = s^{-j}\psi\left(s^{-j}(x-u)\right) \ , \quad j \in \mathbb{Z}\, , u \in \mathbb{Z}^2$$

In the Fourier domain

$$\widehat{\psi_{ju}}(\xi) = e^{2i\pi\xi u}\hat{\psi}(s\xi)$$



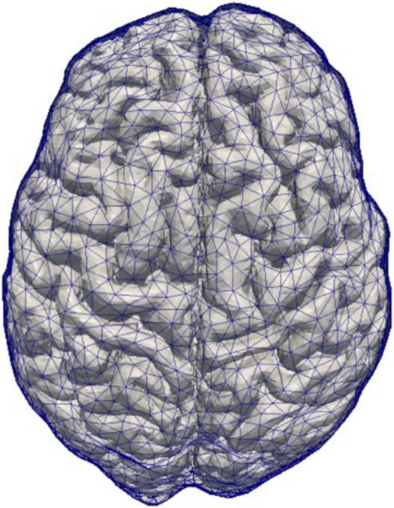
Mexican hat radial 2D wavelet

The case of the discretized cortex surface:

- No natural translation or dilation structure
- Fourier transform: possible as soon as a Laplacian is available: graph Laplacian

Wavelets on the cortex II

- Discretized surface:
  - Topologically equivalent to two spheres
  - Triangulated (Delaunay-type triangulation)
  - Post-processings: topological corrections, smoothing, mesh refinement...
- Eventually: surface of interest represented by a graph, with edges and vertices (mainly topological information, unless edge lengths are taken into account).



Wavelets on the cortex III

Consider a symmetric graph $G(\mathcal{V}, \mathcal{W})$ with vertex set $\mathcal{V}$ ($\sharp\mathcal{V} = K$), and $K \times K$ weight matrix $\mathcal{W}$. The graph Laplacian is

$$\mathcal{L} = \mathcal{D} - \mathcal{W} \qquad \text{with} \qquad \mathcal{D} = \mathrm{diag}(d_1, \dots d_K) \ , \qquad d_k = \sum_i w_{ki} \ .$$

Consider the graph Laplacian eigenvalue decomposition $\mathcal{L} = F\,\Xi F^\top$, $\Xi = \mathrm{diag}(\xi_1, \dots \xi_K)$.

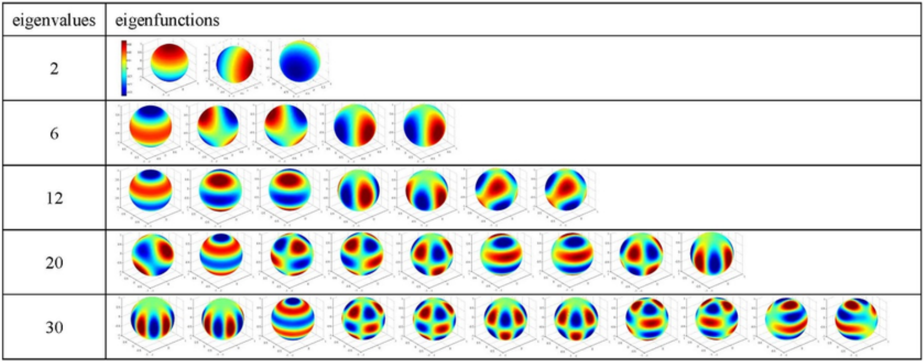
- Eigenvalues $0 = \xi_1 \leq \xi_2 \leq \cdots \leq \xi_K$ interpreted as spatial frequencies
- Corresponding eigenvectors $\{\phi_1, \dots \phi_K\}$ interpreted as Fourier modes
- The graph Fourier transform (GFT) of $f \in \mathbb{C}^K$ is defined by

$$f \longmapsto \widehat{f} = F^\top f : \qquad \widehat{f}(k) = \langle f, \phi_k \rangle$$

GFT is an orthogonal transform, hence invertible, norm preserving,...

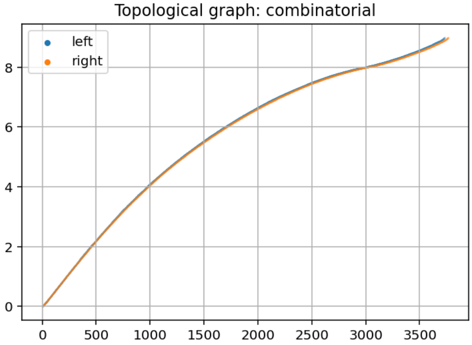
Wavelets on the cortex IV

Laplacian eigenvectors: the case of the sphere



Wavelets on the cortex V

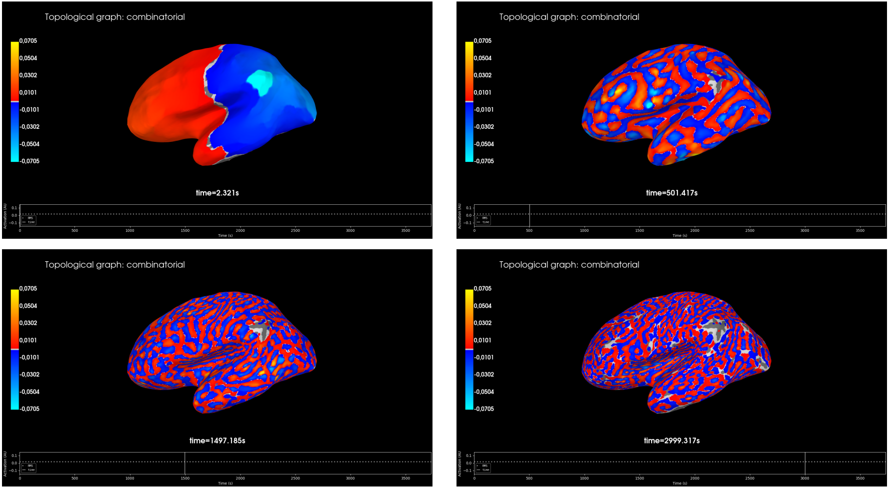
Graph filtering: if we have frequencies, we can create low-pass, band-pass, high-pass filters... and wavelet filter banks



Eigenvalues (cortical frequencies) as a function of rank

Question: Do Laplacian eigenvalues qualify as frequencies ?

Wavelets on the cortex VI



Some eigenvectors (cortical Fourier modes ?) with increasing rank

Wavelets on the cortex VII

Graph wavelet design (Hammond, Gribonval, Vandergheynst 2011): Given functions $g = h_0, h_1, \dots h_{N_s}$ on the continuous extension $[0, \xi_K]$ of the spectrum of L set

$$W_n = F\,\mathrm{diag}(h_n(\xi_1), \dots h_n(\xi_K))F^\top \ .$$

The columns of W_n are the wavelets at scale n. The concatenation W of the matrices W_n is the wavelet matrix, with dimension $K \times K(N_s + 1)$

$$W = \begin{bmatrix} W_0 & W_1 & \dots & W_{N_s} \end{bmatrix}$$

Proposition

The columns of W form a frame of $\mathbb{R}^K$ if there are constants $0 < C_1 \leq C_2$ s.t.

$$C_1 \leq \sum_{n=0}^{N_s} |h_n(\xi_k)|^2 \leq C_2 \ , \qquad k = 1, \dots K \ .$$

Then, for all $S \in \mathbb{R}^K$, there exists $A \in \mathbb{R}^{(J+1)N_s}$ such that $S = WA$.

Wavelets on the cortex VIII

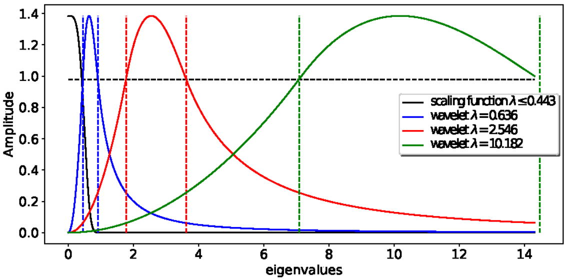
Remarks

- If $N_s > 0$, there are infinitely many $A \in \mathbb{R}^{(N_s+1)K}$ such that $S = WA$, including the OLS solution $A = (WW^\top)^{-1}W^\top S$, ... far from being the most interesting.
- More interesting: find the "optimal" A, in a sense to be defined.
- The ratio C_2/C_1 is the condition number of the frame. For the construction given next slide (3 wavelet scales + 1 low-pass scale, see (Hammond et al 2011))), the condition number is about 1.5.
- The construction can be modified so that $C_1 = C_2 = 1$, yields a Parseval formula.

Wavelets on the cortex IX

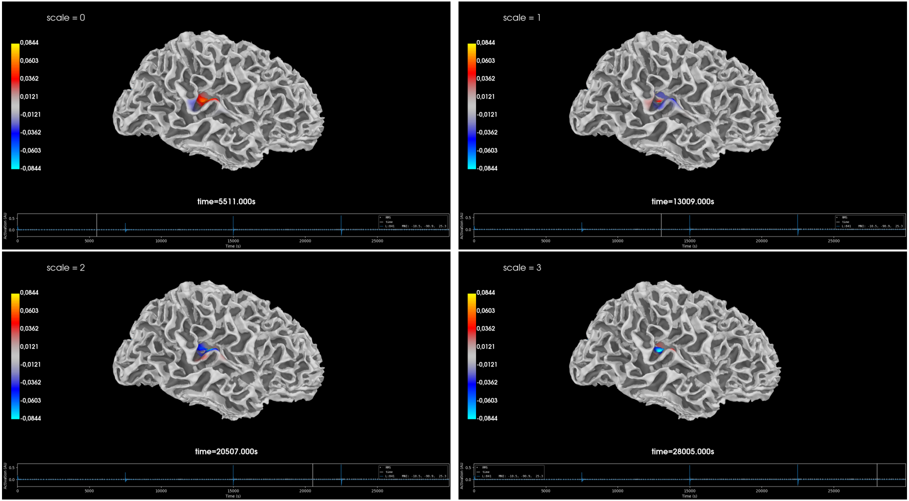
Graph wavelet design (cont'd): choose $g = h_0$ as the frequency response of a low-pass filter, and the h_n as rescaled copies of the frequency response of a reference band-pass filter:

$$h_n(\xi) = h(s^n\xi) \ , \qquad \text{for some} \qquad s > 0 \ .$$



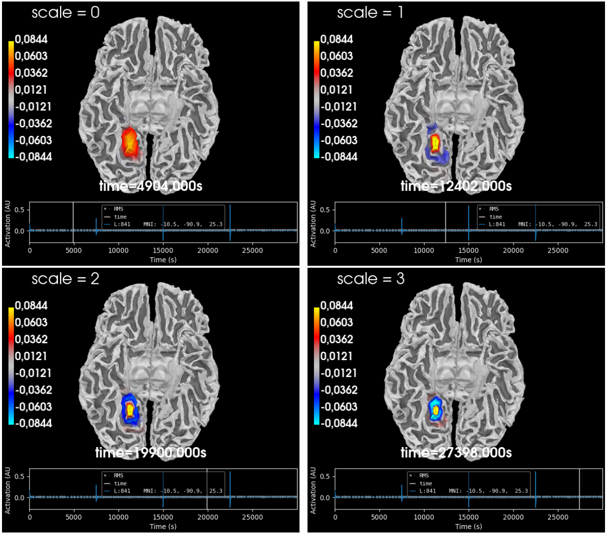
Frequency responses of graph wavelets (frequency = Laplacian spectrum).
Here: 1 low-pass function $g = h_0$, and 3 band-pass functions h_1, h_2, h_3.

Wavelets on the cortex X



Wavelets on the primary auditory cortex:
Top left: scaling function; others: wavelets at various scales

Wavelets on the cortex XI



Wavelets on the ventral cortex:

Top left: scaling function; others: wavelets at various scales

SGW-based solvers I

Synthesis-based approaches: denote by A synthesis multiscale coefficients, and solve

$$A_* = \arg\min_A \left[\frac{1}{2}\|Z - \widetilde{G}A\|^2 + \psi_\theta(A)\right]$$
$$= \arg\min_A \left[\frac{1}{2}\|Z - \widetilde{G}A\|^2 + \psi_\theta(A)\right] , \quad \text{with} \quad \widetilde{G} = GW \ .$$

The estimate is then

$$S_* = WA_*$$

Wavelets used only twice:

- pre-computation of combined wavelet leadfield matrix $\widetilde{G} = GW$
- Synthesis of the solution WA_* after evaluation of A_*.

Choices for ψ_λ:

- Quadratic: $\psi_\gamma(A) = \|\Gamma^{-1/2}A\|_2^2$, with $\Gamma = \text{diag}(\gamma)$ diagonal
- Sparsity promoting: $\psi(A) = \lambda\|A\|_1$, group sparsity $\psi_\lambda(A) = \lambda\|A\|_{21}$
- ...

SGW-based solvers II

Quadratic penalty: $\psi(A) = \|\Gamma^{-1/2}A\|_2^2$, yields a closed form solution

$$\begin{aligned} A_* &= \Gamma\widetilde{G}^\top\left(\mathrm{I}_J + \widetilde{G}\Gamma\widetilde{G}^\top\right)^{-1} Z \\ &= \Gamma\widetilde{G}^\top\Sigma_Z(\Gamma)^{-1}Z \end{aligned}$$

Sparsity promoting penalty: $\psi(A) = \lambda\|A\|_1$, requires numerical minimization: ISTA (or accelerated version FISTA), ADMM, ...

$$A^{(i+1)} = \mathbb{S}_{\lambda\tau}\left[A^{(i)} - \tau\widetilde{G}^\top(A^{(i)} - \widetilde{G}Z)\right]$$

which converges as soon as $\tau \leq \|\widetilde{G}\|^{-2}$.

$\mathbb{S}_{\lambda\tau}$ is the soft thresholding operator with threshold $\lambda\tau$.

In all situations, the enemy is the parameter θ (can be a number λ, several numbers, a matrix Γ,...)

Type II Bayesian learning

Assume the T columns of Z (resp. A) are i.i.d. samples from J-dimensional (resp. (N-dimensional, with $N = K(N_s + 1)$) distribution.

Previous solutions correspond to MAP estimates, also called Type I estimates

$$A_* = \arg\max_A p(Z|A)p_\theta(A)$$

Consider a families of distributions p_θ parameterized by a set of hyper-parameters θ. These hyper-parameters can be learned from the data along with the model parameters using a hierarchical empirical Bayesian approach, called Type II Bayesian learning:

$$\theta_* = \arg\max_\theta p_\theta(Z) = \arg\max_\theta \int p(Z|A)p_\theta(A)\, dA$$

Once θ_* has been learned, it can be plugged in the MAP estimate.

Sparse Bayesian Learning I

SBL corresponds to the case $A \sim \mathcal{N}(0, \Gamma)$, with $\Gamma = \text{diag}(\gamma_1, \ldots \gamma_K)$.

Maximization of $p_\gamma(Z)$ boils down to minimizing a Bregman's `log det` divergence

$$\begin{aligned}
\gamma_* &= \arg\min_{\gamma \in \mathbb{R}_+^N} \left[\frac{1}{T} \sum_{t=1}^{T} \text{Tr}\Big(Z(t)^\top \Sigma_Z(\gamma)^{-1} Z(t) \Big) + \ln\det(\Sigma_Z(\gamma)) \right] \\
&= \arg\min_{\gamma \in \mathbb{R}_+^N} \left[\text{Tr}\Big(C_Z \Sigma_Z(\gamma)^{-1} \Big) + \ln\det(\Sigma_Z(\gamma)) \right]
\end{aligned}$$

with C_Z the sample covariance matrix of observations, and $\Sigma_Z(\gamma)$ is the posterior covariance matrix

$$\Sigma_Z(\gamma) = \mathrm{I}_J + \widetilde{G} \Gamma \widetilde{G}^\top$$

Again the estimate can be plugged in the expression of S to get

$$A_* = \Gamma_* \widetilde{G}^\top \Sigma_Z(\gamma_*)^{-1} Z$$

Sparse Bayesian Learning II

The minimization problem to be solve is not so easy:

$$\underbrace{\mathrm{Tr}\Big(C_Z\Sigma_Z(\gamma)^{-1}\Big)}_{\text{convex}} + \underbrace{\ln\det(\Sigma_Z(\gamma))}_{\text{concave}}$$

Most algorithms rely on duality arguments: in CHAMPAGNE (Wipf et al 2011), write

$$\mathrm{Tr}\Big(C_Z\Sigma_Z(\gamma)^{-1}\Big) = \min_{X\in\mathbb{R}^{N\times T}} \frac{1}{T}\left[\sum_{t=1}^{T}\|Z(t)-\widetilde{G}X(t)\|^2 + \sum_{t=1}^{T}\|\Gamma^{-1/2}X(t)\|^2\right]$$

and

$$\ln\det(\Sigma_Z(\gamma)) = \min_Y\left[Y^\top\gamma - h^*(Y)\right]\ , \qquad h^*(Y) = \nabla_\gamma \ln\det(\Sigma_Z(\gamma))$$

Sparse Bayesian Learning III

Hence, we end up with the following problem

$$(\gamma_*, X_*, Y_*) = \underset{\gamma\in\mathbb{R}_+^N, X\in\mathbb{R}^{N\times T}, Y\in\mathbb{R}^N}{\arg\min} \frac{1}{T}\left[\sum_{t=1}^{T} \|Z(t) - \widetilde{G}X(t)\|^2\right] + \mathcal{R}(\gamma, Y)$$

with

$$\mathcal{R}(\gamma, Y) = \frac{1}{T}\sum_{t=1}^{T}\sum_{n=1}^{N}\frac{X_n(t)^2}{\gamma_n} + \gamma^\top Y - h^*(Y)\ .$$

Block coordinate descent

$$\begin{aligned}
Y_n^{(i)} &= \widetilde{G}_n^\top \Sigma_Z\Big(\gamma^{(i)}\Big)^{-1} \widetilde{G}_n\ , & n = 1, \ldots N \\
X^{(i)}(t) &= \Gamma^{(i)}\widetilde{G}^\top\Sigma_Z\Big(\gamma^{(i)}\Big)^{-1} Z(t)\ , & t = 1, \ldots T \\
\gamma_n^{(i+1)} &= \sqrt{\frac{1}{TY_n^{(i)}}\sum_{t=1}^{T} X_n^{(i)}(t)^2}\ , & n = 1, \ldots N
\end{aligned}$$

Sparse Bayesian Learning IV

Remarks

1. A main result (Wipf et al 2011) states that the number of nonzero diagonal entries in γ_* is smaller than or equal to the rank of G_W, and therefore the number of sensors J. Since $\gamma_k = 0 \implies A_k = 0$, this yields sparsity !
2. Several similar algorithms have been derived. For example McKay updates exploits duality with respect to $\ln(\gamma)$ instead of γ. EM can also be used, but is extremely slow here.
3. These algorithms have been put into the common language of MM (Majorization-Minimization) algorithms by (Hashemi et al 2021), where proofs of convergence to a stationary point have been given (exploiting earlier results by Hunter & Lange 2004)

At convergence,

$$A_*(t) = \lim X^{(i)} \ , \qquad S_*(t) = WA_*(t) \ .$$

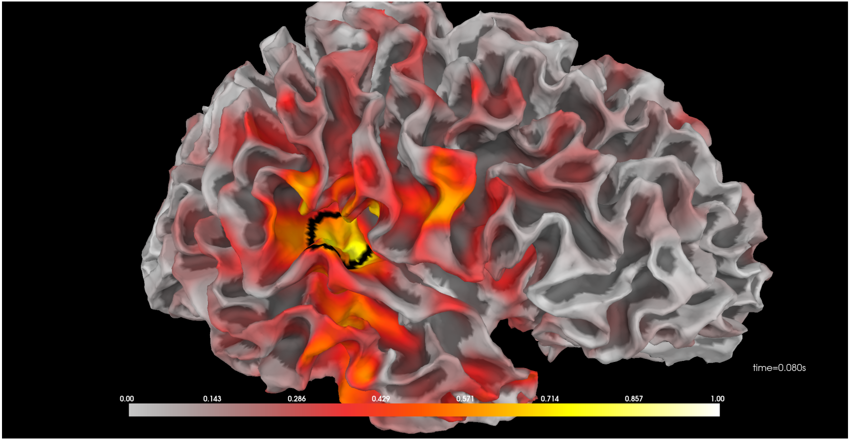
Summary

- SBL approach is indeed able to obtain sparse solutions, i.e. solutions that involve a very small number of nonzero coordinates
- In the context of MEG source reconstruction, results are acceptable if the brain activity is very focal
- For extended activity, sparsity in wavelet domain does the job

Auditory Evoked Potential dataset I

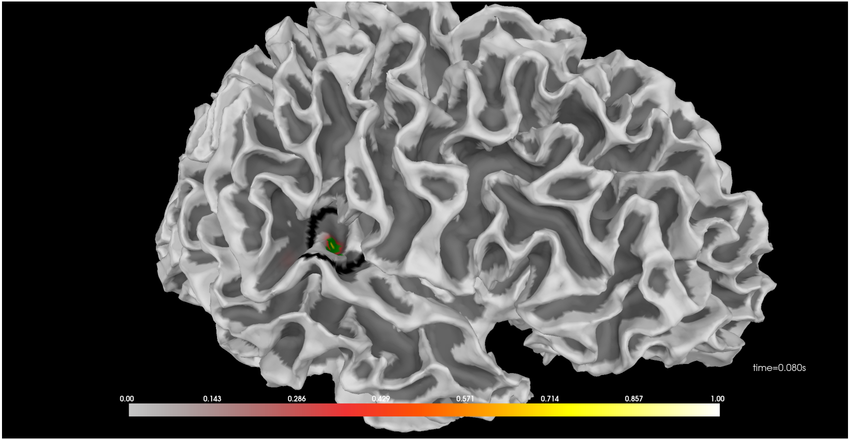
- Data : reference dataset from the MNE-Python software suite.
- Quote from mne-python web site : *checkerboard patterns were presented to the subject into the left and right visual field, interspersed by tones to the left or right ear. The interval between the stimuli was 750 ms. Occasionally a smiley face was presented at the center of the visual field...*
- Available data for auditory stimulation : ~ 50 trials for each side, time-locked to the stimulus, and averaged to increase SNR.
- Whitening : spatial whitening using a baseline covariance matrix estimated by tools from mne-python.

Auditory Evoked Potential dataset II



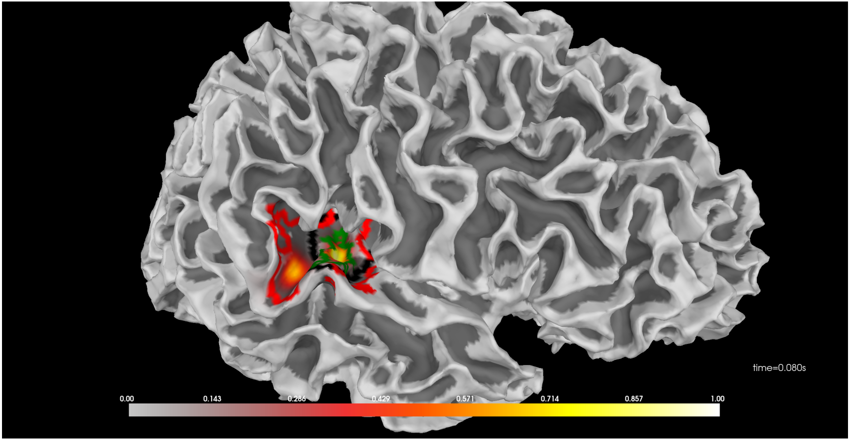
Response to left auditory stimulus
Black annulus: MNE-Python "aud-rh" label
Red and green lines: 1% and 10% level curves
eLORETA solution: data driven diagonal Γ

Auditory Evoked Potential dataset III



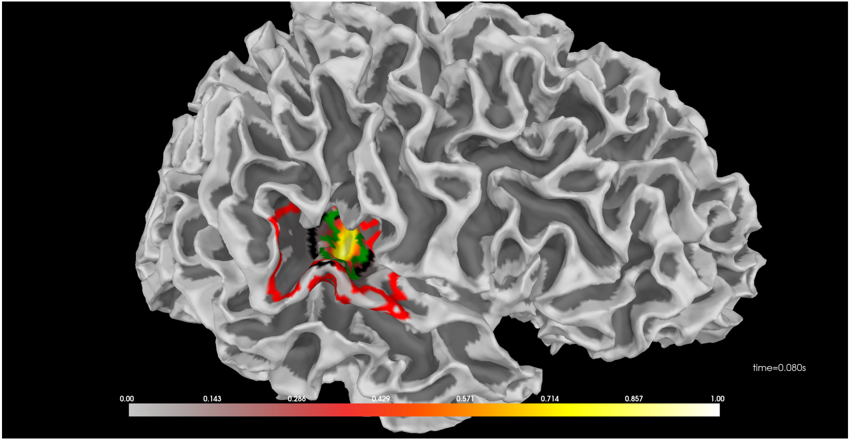
Response to left auditory stimulus
Black annulus: MNE-Python "aud-rh" label
Red and green lines: 1% and 10% level curves
SBL (in spatial domain) solution, with data driven λ: extremely sparse

Auditory Evoked Potential dataset IV



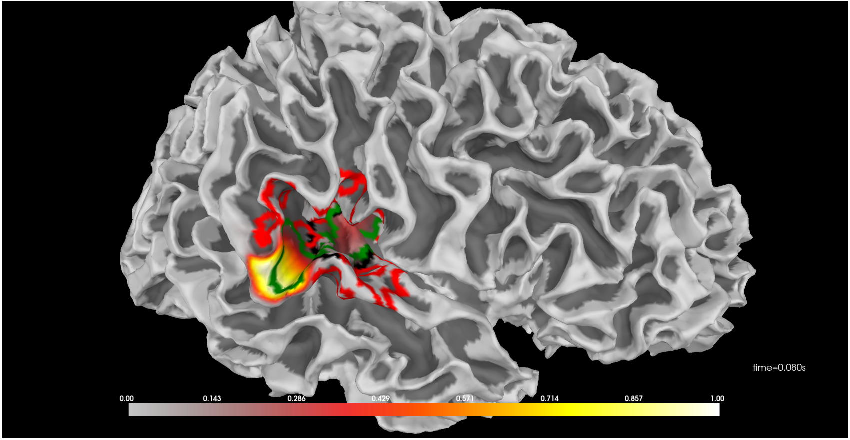
Response to left auditory stimulus
Black annulus: MNE-Python "aud-rh" label
Red and green lines: 1% and 10% level curves
Wavelet-SBL solution: $N_s = 1$

Auditory Evoked Potential dataset V



Response to left auditory stimulus
Black annulus: MNE-Python "aud-rh" label
Red and green lines: 1% and 10% level curves
Wavelet-SBL solution: $N_s = 2$

Auditory Evoked Potential dataset VI



Response to left auditory stimulus
Black annulus: MNE-Python "aud-rh" label
Red and green lines: 1% and 10% level curves
Wavelet-SBL solution: $N_s = 3$

Evaluation metrics I

Amplitude map:

$$\mathcal{A}[k] = \sqrt{\langle S[k,\cdot]^2\rangle_{\rm soi}}\ , \qquad \widetilde{\mathcal{A}} = \mathcal{A}/\sum_{p=1}^{N}\mathcal{A}_p\ .$$

Absolute metrics: characterize properties of a given amplitude map $\mathcal{A}$

- Spatial dispersion (equivalent radius)
- Center of mass depth
- Coifman-Wickerhauser entropy (measures sparsity)

Comparison metrics: measure discrepancies between two amplitude maps

- Region Localization Error
- Wasserstein distance
- Distance between centers of mass
- Ratio of norms (to test the amplitude of the reconstruction)

Evaluation metrics II

Comparison with "aud-rh" label

	Ref.	**MNE**	**MCE**	**SBL**	**L1-TV(1)**	**L1-TV(2)**	**w-SBL1**	**w-SBL2**
Size	13	111	19	8	73	76	**14**	**18**
Depth	15	0.72	3.2	10.5	1.6	11.2	10.5	**17.3**
Entropy	4.75	12.27	1.86	1.21	10.18	11.79	**5.17**	**6.57**

Results on simulated data I

Simulated data

Input data:

- Cortical surface: vertex locations (K vertices), triangles and vertex connectivity (adjacency matrix)
- Leadfield matrix G ($J \times K$)
- Baseline covariance matrix Σ ($J \times J$)
- Input SNR ρ_{in}, defined in sensor space

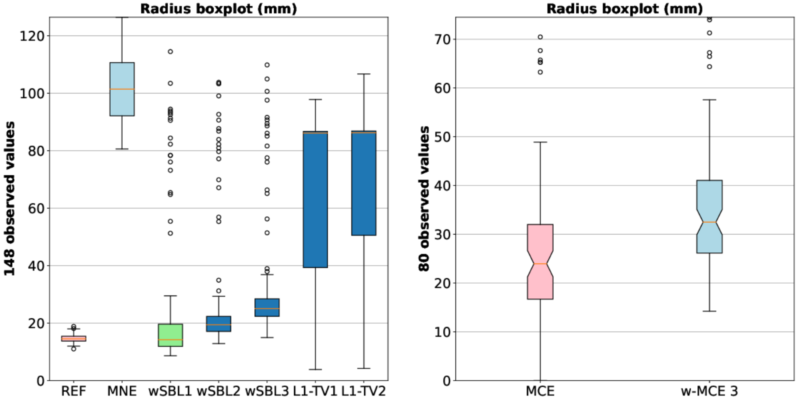
Simulation protocole:

- Randomly located patches, of variable depth and two different sizes.
- Realistic time course
- Propagate to sensors and add (realistic) baseline

Comparisons:

- Run various solvers
- Compute a set of validation metrics, that (hopefully) cover most relevant information.

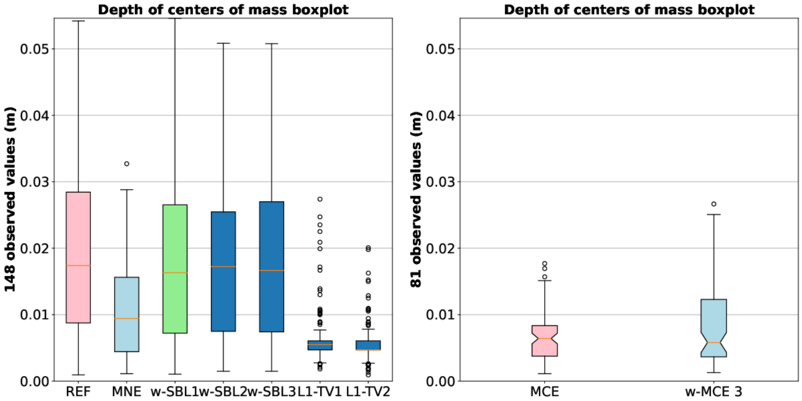
Results on simulated data II



Absolute metrics: equivalent radius Δ_{SD}
20mm patch

- The number of scales in wavelet decomposition influences the spatial extension
- Other solvers: far too extended solutions

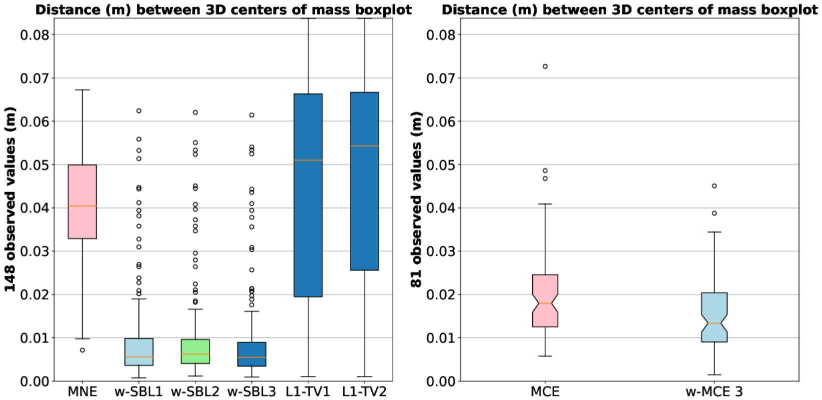
Results on simulated data III



Absolute metrics: depth, ground truth and estimated activity
20mm patch

- w-SBL: depths compatible with reference.
- Other solvers: too superficial

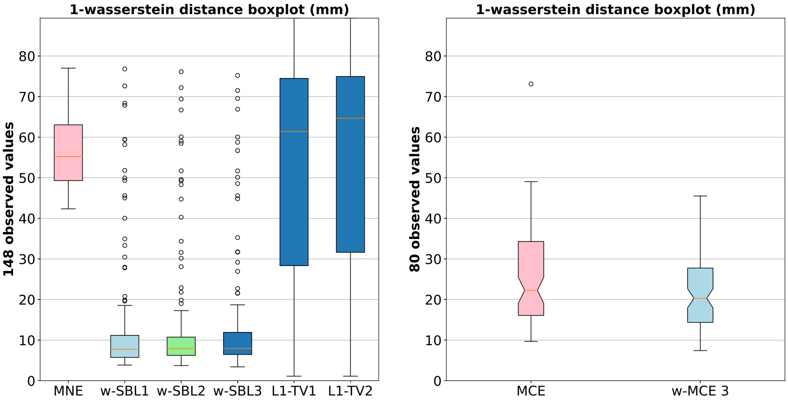
Results on simulated data IV



Comparison metrics: 3D distance between centers of mass, ground truth and estimated activity 20mm patch

- No comment !

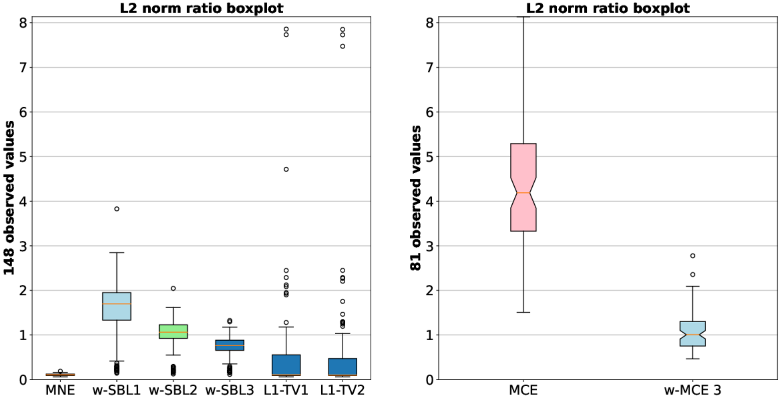
Results on simulated data V



Comparison metrics: Wasserstein distances W_1 between ground truth and estimated 20mm patch

- No comment !

Results on simulated data VI



footnotesize Comparison metrics: Ratio of L2 norms of ground truth and estimated brain activity
20mm patch

- w-SBL does quite a good job in terms of amplitudes

Outline

1. The combination of spectral graph wavelets with SBL appears to be a good choice for the reconstruction of extended sources.
2. Compares very favorably with concurrent approaches, with all tested metrics (similar resuls obtained for 10mm patches)
3. No difficult parameter tuning, but...
4. The influence of wavelet parameters remain to be better understood
   - Most notably the number of decomposition levels, which influences the solution
   - The choice of wavelet parameters (relative bandwidth, Laplacian,...)

Thanks for your attention

References

- S. Baillet, Nature Neuroscience, 20:3, pp. 327–339, 03 2017.
- H. Becker, L. Albera, P. Comon, R. Gribonval & I. Merlet, EUSIPCO 2014, pp. 41–45.
- L. Ding, Physics in Medicine and Biology, 54:9, pp. 2683–2697, 2009.
- D.K. Hammond, P. Vandergheynst & R. Gribonval, Appl. Comput. Harmon. Anal., 30(2):129 – 150, 2011.
- A. Hashemi et al, *NeuroImage* 239 (2021), pp. 1-20.
- D.P. Wipf, B. Rao & S.S. Nagarajan, *IEEE Trans Info Th* 57 (2011), pp. 6236-6255.
- S. Mokhtari, J.-M. Badier, C.-G. Bénar and B. Torrésani. *IEEE Transactions Biomedical Imaging*, under revision
- S. Mokhtari, B. Torrésani, J.M. Badier & C. Bénar, Proceedings of SampTA'23 (Sampling Theory and Applications), (YALE, New Haven, USA, july 2023).
- S. Mokhtari, J.M. Badier, C. Bénar & B. Torrésani, Proceedings of GRETSI (French Signal Processing COnference, 2023).