



Inria

universit
PARIS-SACL

cea

Aligning M/EEG data with covariance-based techniques to enhance predictive regression modeling

Apolline Mellot, Antoine Collas, Pedro Rodrigues, Denis Engemann, Alexandre Gramfort

In preparation for submission to NeuroImage

MIND team meeting - February 21, 2023

Motivations: M/EEG signals and brain health



MEG recording setup



EEG recording setup

Motivations: M/EEG signals and brain health

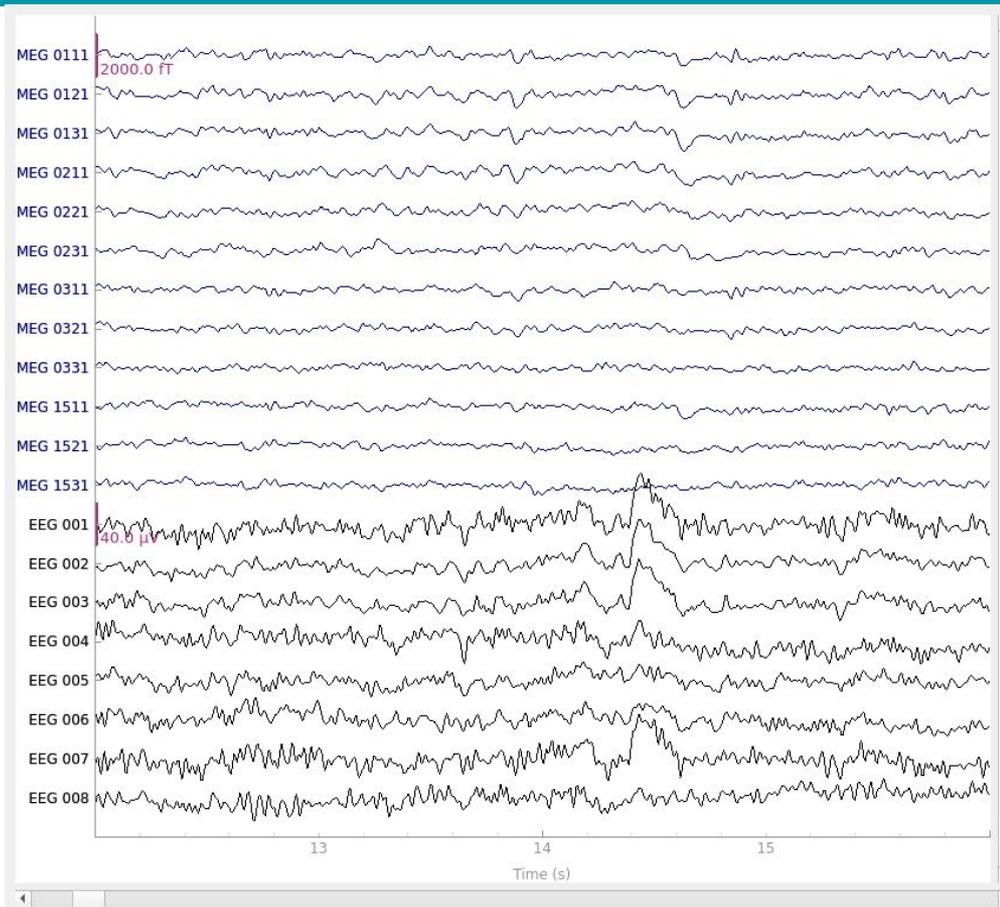
- M/EEG signals:

$$X(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \in \mathbb{R}^{P \times T}$$

P : number of electrodes

T : number of time points

- M/EEG provides relevant info on the brain and related diseases
[Lopes Da Silva *et al.*, 2013]



Motivations: M/EEG signals and brain health

Predictive modeling with M/EEG signals:

- **Classification:** Brain Computer Interface, epileptic seizure detection, sleep staging, etc.
- **Regression:** Risk scores, optimal drug-dosage, brain age, etc.

→ Focus on **regression context**.

Problem: statistical discrepancies when changing population, tasks or acquisition device

[Engemann *et al.*, 2018]

Motivations: Predictive modeling with M/EEG covariances

Main approaches for predictive regression modeling:

- **Spatial filtering** like Common Spatial Filtering (CSP) or Source Power Comodulation (SPoC)...
- **Deep learning**: end-to-end processing
- **Covariance-based techniques**: simple and competitive [Barachant *et al.*, 2012] [Sabbagh *et al.*, 2019] [Engemann *et al.*, 2022]

→ In this work, focus on a **covariance-based technique**.

Motivations: The issue of dataset shift

Predictive model with machine learning:

Increasing the number of data
=
Increasing the performances

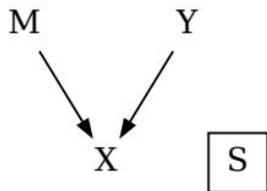
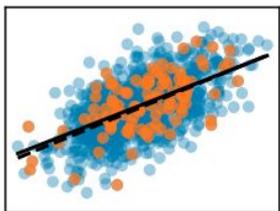
What we would like to do:

Pooling several existing datasets  recent emergence of large databases
high variability

Challenge: Statistical and physiological variabilities between datasets

Motivations: The issue of dataset shift

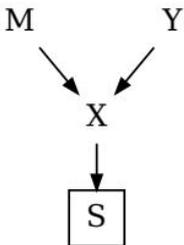
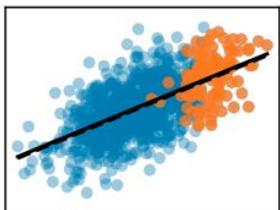
Uniform selection



$S \perp\!\!\!\perp (X, Y)$ No dataset shift
 $P(X, Y)$ unchanged

≠ preprocessing

Selection based on X



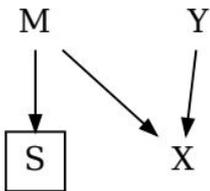
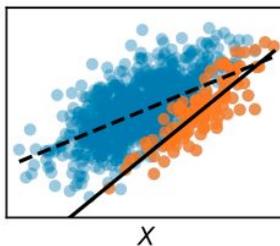
$Y \perp\!\!\!\perp S | X$ Covariate shift
 $P(Y|X)$ unchanged

≠ tasks

≠ populations

≠ recording devices

Selection based on M



$Y \not\perp\!\!\!\perp S | X$ Selection bias
 $P(Y|X)$ modified

≠ number of channels

[Dockes *et al.*, 2021]

Motivations: Domain adaptation to tackle dataset shift

Domain adaptation techniques to reduce dataset shift.

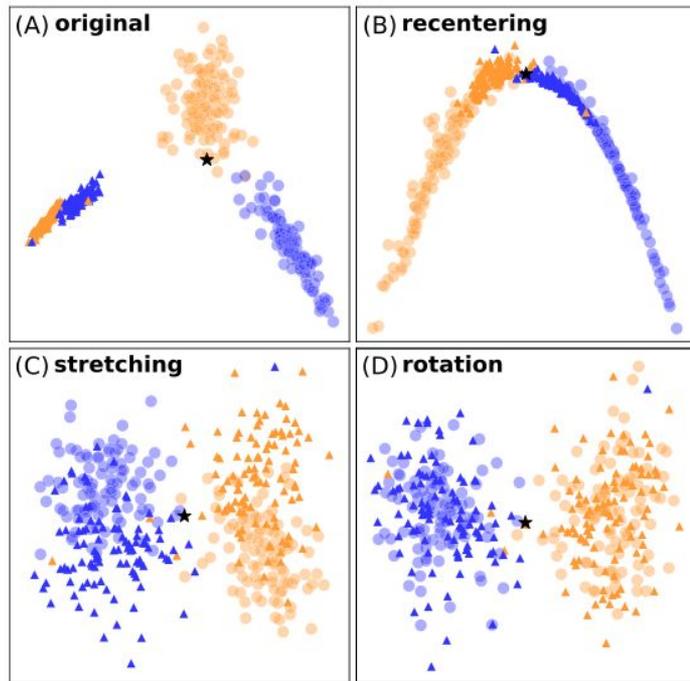
→ Focus on **covariance-based techniques**

for predictive model and for domain adaptation.

Approach: aligning data distributions leveraging the geometry of the data.

- Re-center
- Equalize dispersion / Re-scale
- Rotation correction

→ **But** developed and evaluated for classification



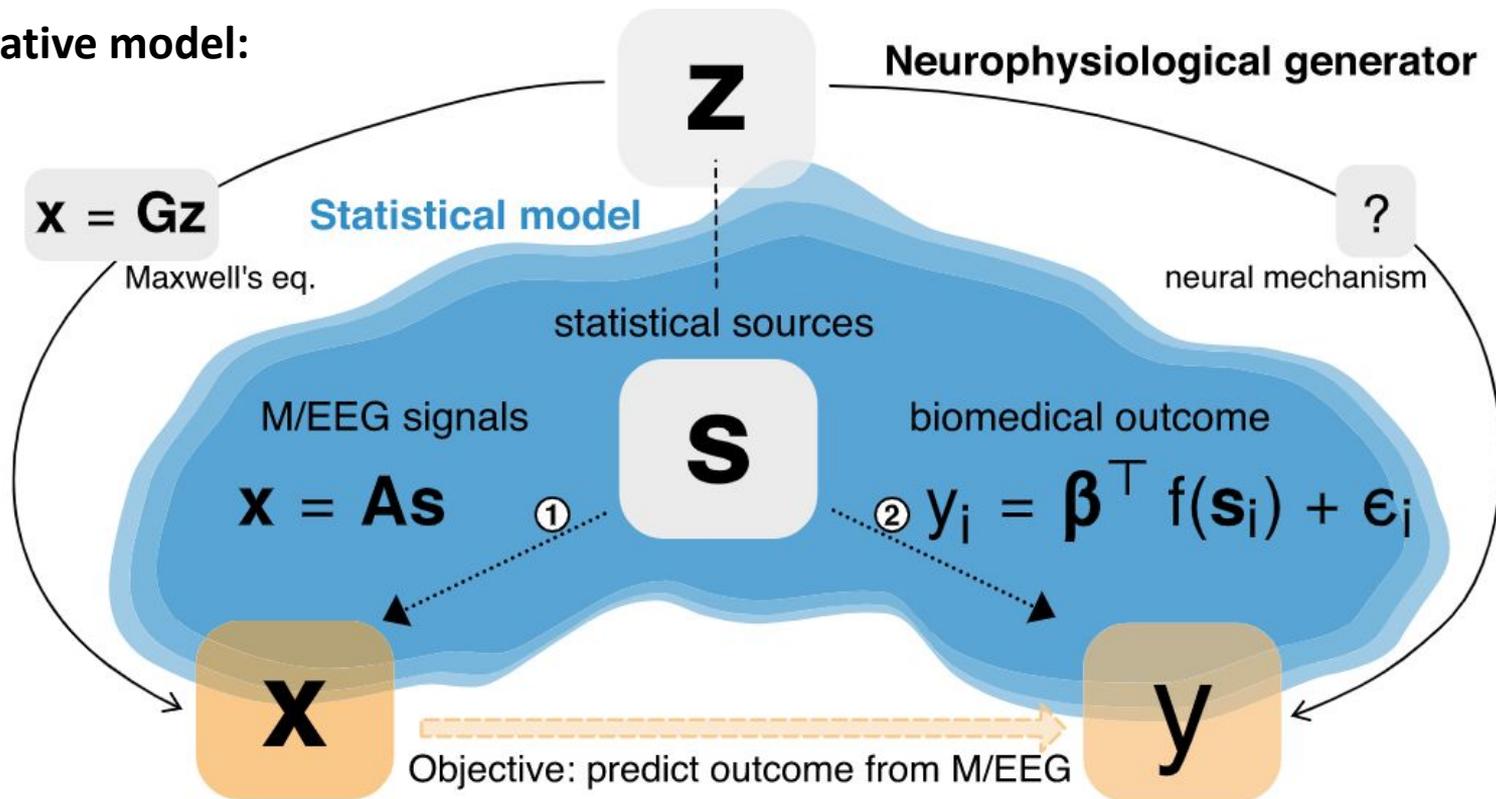
[Rodrigues *et al.*, 2019]

What this work is about:

- Extension of the covariance-based alignment steps to **regression context**.
- Investigate dataset shifts scenarios based on **M/EEG generative mechanisms**.
- **Simulation study** exploring these scenarios and validation of the alignment methods.
- Alignment methods **relevance shown one MEG dataset and two EEG datasets**.

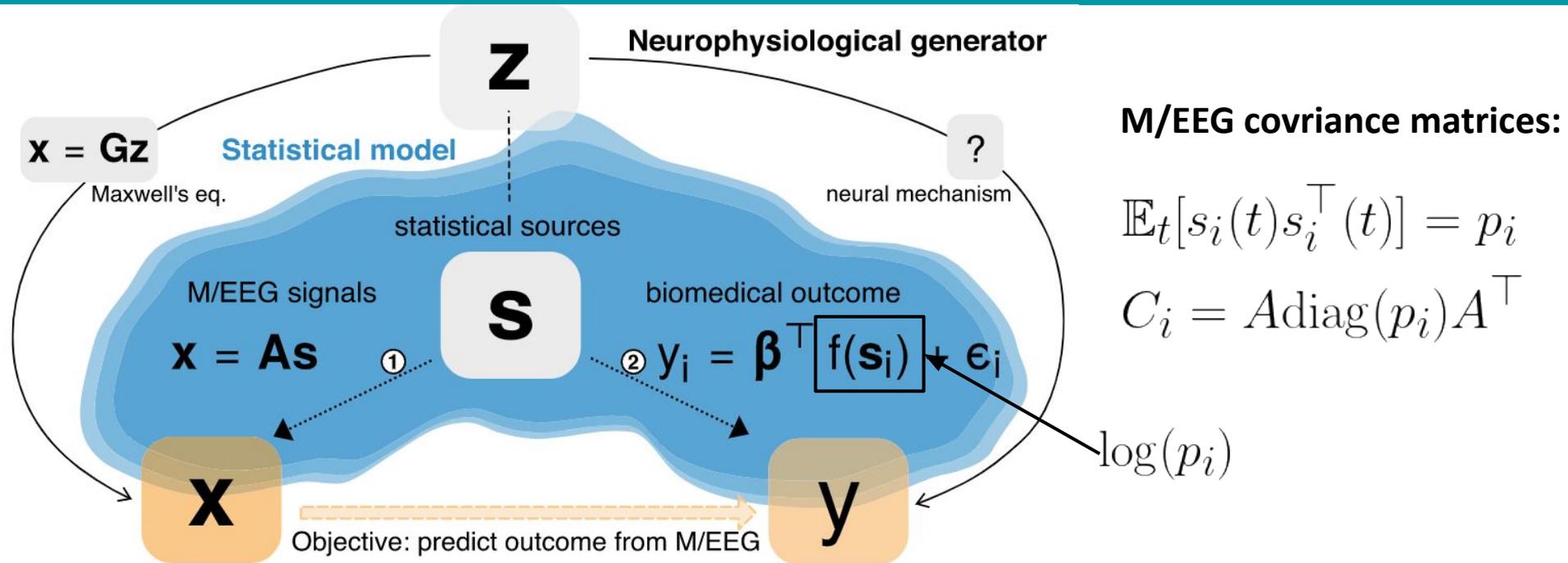
Methods: Generative model for regression with M/EEG

Generative model:



[Sabbagh et al., 2020]

Methods: Generative model for regression with M/EEG



M/EEG covariance matrices:

$$\mathbb{E}_t[s_i(t)s_i^\top(t)] = p_i$$

$$C_i = A \text{diag}(p_i) A^\top$$

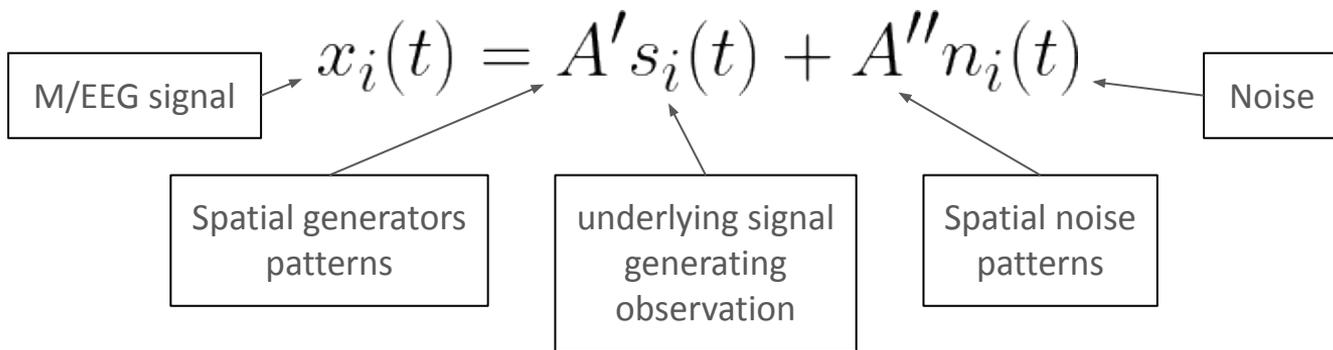
→ Riemannian geometry: **linear** relationship between covariance vector and y

[Sabbagh et al., 2019, 2020]

Methods: Statistical generative model of M/EEG signals

Generative model: $i \in [1, \dots, N]$ ^{samples} $t \in [1, \dots, T]$ ^{time samples} $j \in [1, \dots, Q]$ ^{generators} $Q \leq P$ ^{sensors}

M/EEG signals:



$$\begin{aligned}
 x_i(t) &\in \mathbb{R}^P \\
 s_i(t) &\in \mathbb{R}^Q \\
 n_i(t) &\in \mathbb{R}^{P-Q} \\
 A' &\in \mathbb{R}^{P \times Q} \\
 A'' &\in \mathbb{R}^{P \times P-Q} \\
 \eta_i(t) &\in \mathbb{R}^P \\
 A &\in \mathbb{R}^{P \times P}
 \end{aligned}$$

or more conveniently: $x_i(t) = A \eta_i(t)$

Mixing matrix $\left(\begin{array}{c|c} A' & 0 \\ \hline 0 & A'' \end{array} \right)$ $\left(\begin{array}{c} s_i(t) \\ n_i(t) \end{array} \right)$

Methods: Statistical generative model of M/EEG signals

Generative model: $i \in [1, \dots, N]$ ^{samples} $t \in [1, \dots, T]$ ^{time samples} $j \in [1, \dots, Q]$ ^{generators} $Q \leq P$ ^{sensors}

M/EEG signals: $x_i(t) = A\eta_i(t)$

M/EEG covariance matrices: $C_i = AH_iA^\top$ (or $C_i = A\text{diag}(p_i)A^\top$ if $Q = P$)

we assume $\mathbb{E}_t[s_i(t)s_i^\top(t)] = p_i$ and $\mathbb{E}_t[s_i(t)n_i^\top(t)] = 0$

$$\Rightarrow H_i = \mathbb{E}_t[\eta_i(t)\eta_i^\top(t)] = \left(\begin{array}{c|c} \text{diag}(p_i) & \mathbf{0} \\ \hline \mathbf{0} & \mathbb{E}_t[n_i(t)n_i^\top(t)] \end{array} \right)$$

Methods: Statistical generative model of M/EEG signals

Generative model: $i \in [1, \dots, N]$ ^{samples} $t \in [1, \dots, T]$ ^{time samples} $j \in [1, \dots, Q]$ ^{generators} $Q \leq P$ ^{sensors}

Label modeling: $y_i = \sum_{j=1}^Q \beta_j \log \left(\frac{1}{T} \sum_{t=1}^T s_{i,j}^2(t) \right) + \epsilon_i$

Outcome to predict \rightarrow y_i

Regression coefficient \rightarrow β_j

Generators' powers \rightarrow $s_{i,j}^2(t)$

Noise \rightarrow ϵ_i

Model violation: $A_i = A + E_i$

Reminder: $x_i(t) = A\eta_i(t)$ and $C_i = AH_iA^\top$

Methods: Dataset shifts associated with this model

Experimental setting:

we want to fit a regression algorithm to a source dataset $X^{(s)} = \{x_i^{(s)}, i = 1, \dots, N^{(s)}\}$

and predict on a different target dataset $X^{(t)} = \{x_i^{(t)}, i = 1, \dots, N^{(t)}\}$

Possible shifts:

Different SNR

Different population

$$\begin{cases} x_i^{(s)}(t) = A_i^{\prime(s)} s_i^{(s)}(t) + A_i^{\prime\prime(s)} n_i^{(s)}(t) = A_i^{(s)} \eta_i^{(s)}(t) \\ y_i^{(s)} = \sum_{j=1}^Q \beta_j \log \left(\frac{1}{T} \sum_{t=1}^T (s_{i,j}^{(s)}(t))^2 \right) + \epsilon_i^{(s)} \end{cases}$$

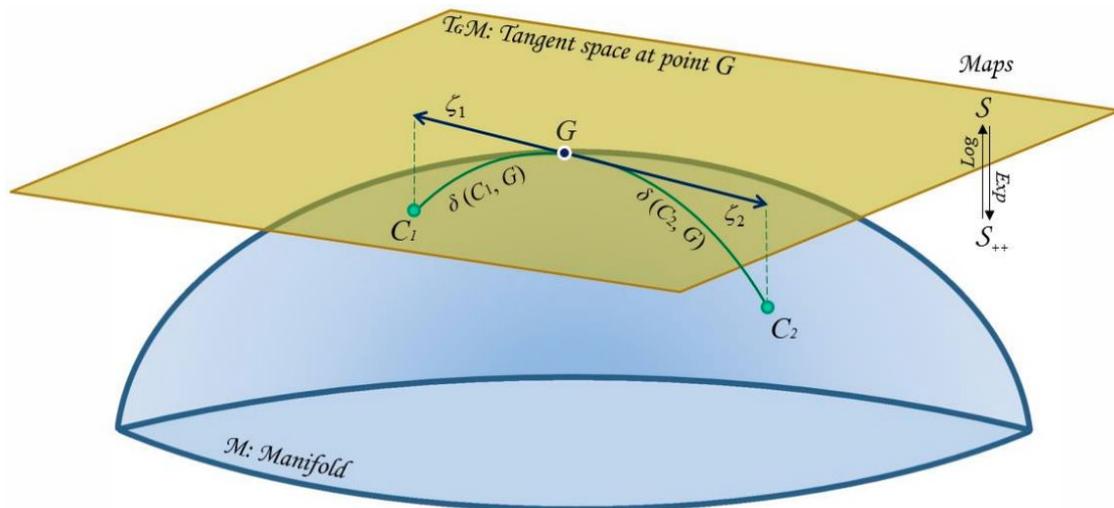
$A_i = A + E_i$

Different label noise

Methods: Riemannian geometry basics

→ Covariance matrices are symmetric positive-definite (SPD) matrices

log of a matrix: $M = U\Delta U^\top \mapsto \log(M) = U\log(\Delta)U^\top$ (same for exp or power)



Logarithmic map:

$$\zeta_1 = \text{Log}_G(C_1) = G^{\frac{1}{2}} \log(G^{-\frac{1}{2}} C_1 G^{-\frac{1}{2}}) G^{\frac{1}{2}}$$

Exponential map:

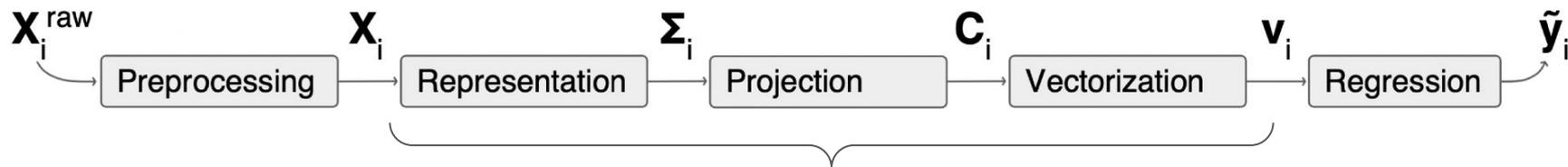
$$C_1 = \text{Exp}_G(\zeta_1) = G^{\frac{1}{2}} \exp(G^{-\frac{1}{2}} \zeta_1 G^{-\frac{1}{2}}) G^{\frac{1}{2}}$$

Vector representation:

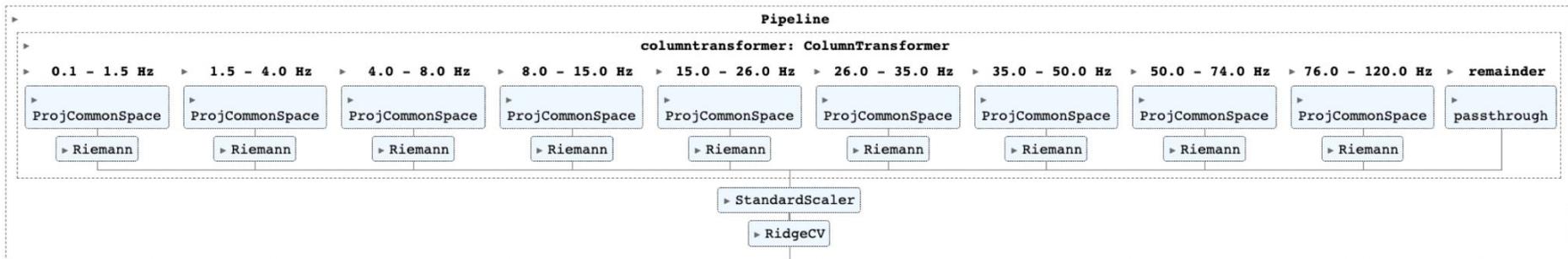
$$z_1 = \text{uvec}(S \odot \zeta_1)$$

[Barachant et al., 2011] → **Generative model:** tangent vectors are linear combination of $\log(p_i)$
[Congedo et al., 2017]

Methods: Filterbank regression pipeline

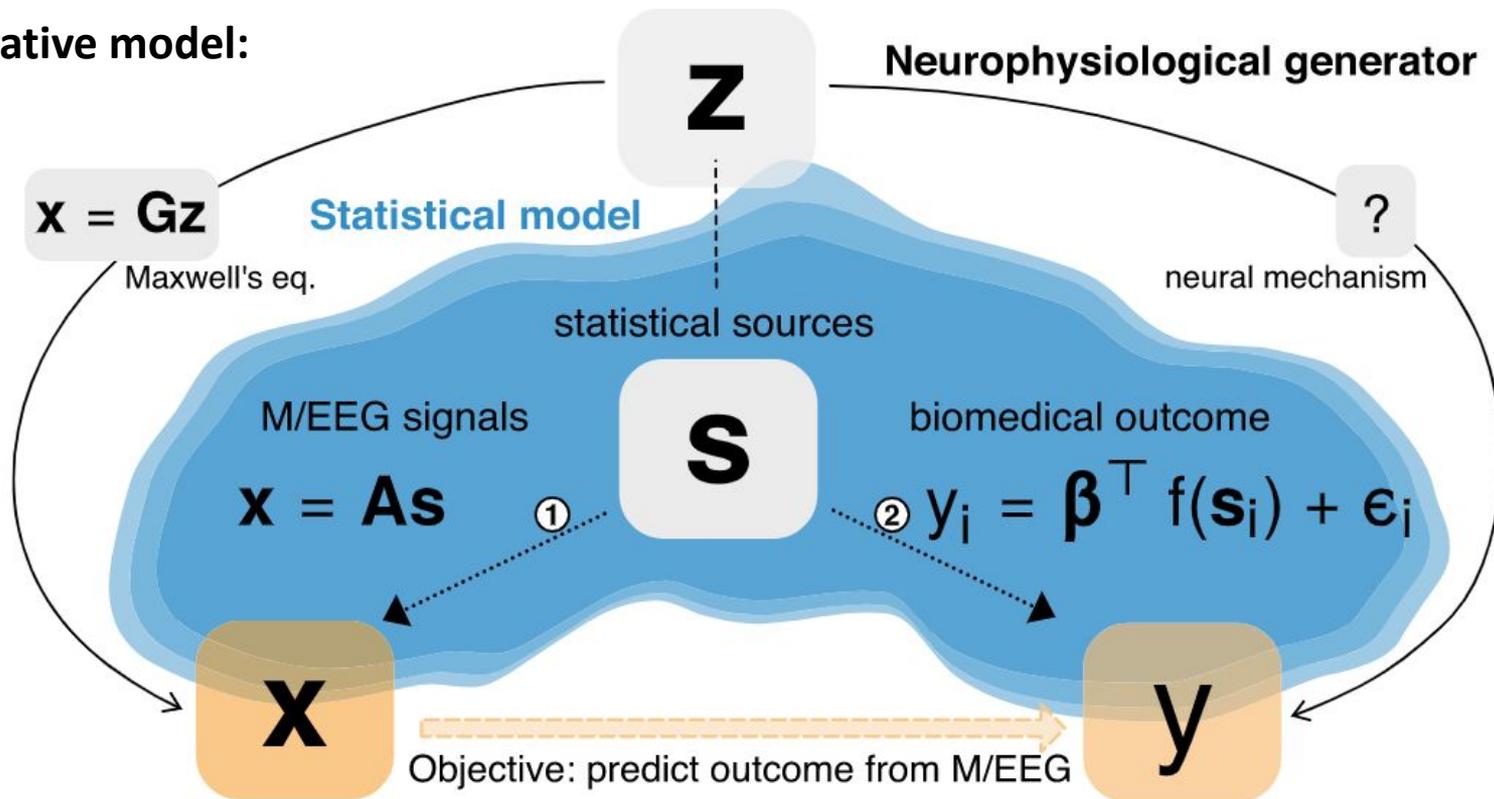


applied separately for each frequency band



Methods: Dataset shifts associated with this model

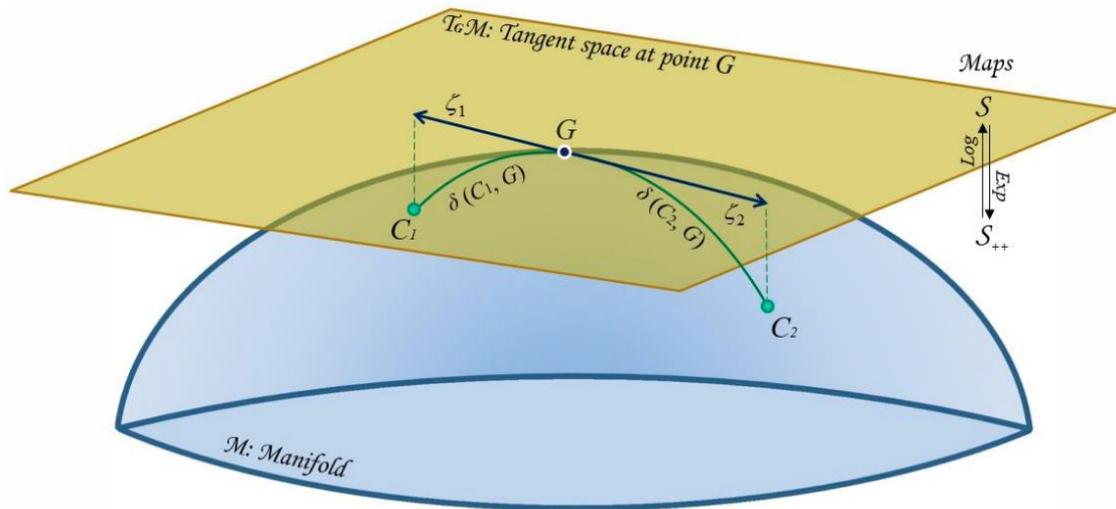
Generative model:



Methods: Riemannian geometry basics

→ Covariance matrices are symmetric positive-definite (SPD) matrices

log of a matrix: $M = U\Delta U^\top \mapsto \log(M) = U\log(\Delta)U^\top$ (same for exp or power)



Riemannian distance:

$$\delta_R(C_1, C_2) = \|\log(C_1^{-1}C_2)\|_F$$

Geometric mean:

$$\bar{C} = \arg \min_{C \in \mathcal{S}_P^{++}} \sum_{i=1}^N \delta_R^2(C, C_i)$$

Dispersion:

$$d = \frac{1}{N} \sum_{i=1}^N \delta_R^2(C_i, \bar{C})$$

[Barachant et al., 2011]

[Congedo et al., 2017]

Methods: Alignment steps

Step 1: re-center to Identity

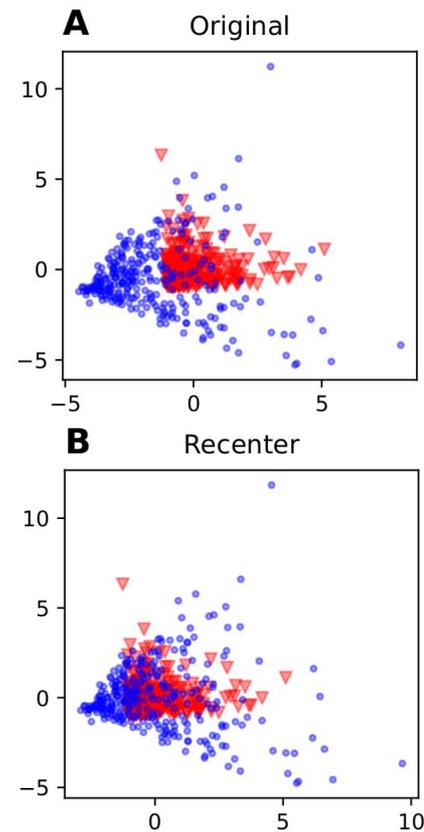
1. Compute geometric means

Reminder: $\bar{C} = \arg \min_{C \in \mathcal{S}_P^{++}} \sum_{i=1}^N \delta_R^2(C, C_i)$

2. Separate whitening

source: $h_{\bar{C}(s)}^{\text{rct}}(C_i^{(s)}) = \bar{C}(s) - \frac{1}{2} C_i^{(s)} \bar{C}(s) - \frac{1}{2}$

target: $h_{\bar{C}(t)}^{\text{rct}}(C_i^{(t)}) = \bar{C}(t) - \frac{1}{2} C_i^{(t)} \bar{C}(t) - \frac{1}{2}$



[Zanini et al., 2018] [Yair et al., 2019]

Methods: Alignment steps

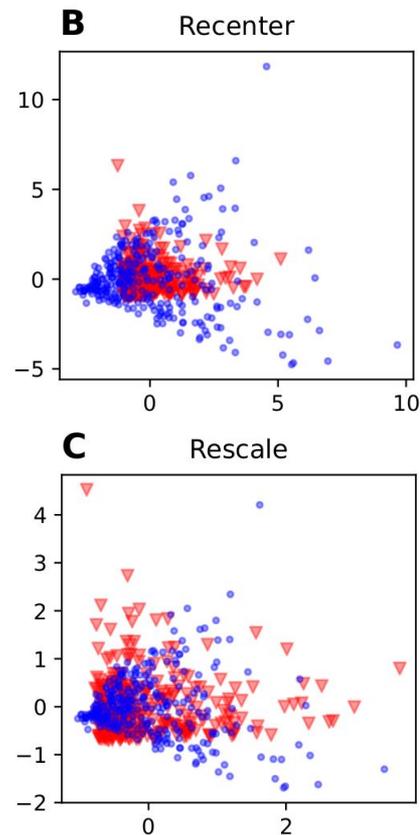
Step 2: equalize the dispersion

1. Compute dispersions

Reminder:
$$d = \frac{1}{N} \sum_{i=1}^N \delta_R^2(C_i, \bar{C})$$

2. Separate re-scaling

source: $h_{d^{(s)}}^{\text{str}}(C_i^{(s)}) = C_i^{(s)} \sqrt{\frac{1}{d^{(s)}}}$ target: $h_{d^{(t)}}^{\text{str}}(C_i^{(t)}) = C_i^{(t)} \sqrt{\frac{1}{d^{(t)}}}$



[Rodrigues et al., 2019]

Methods: Alignment steps

Step 3: Correct the rotation - Method unpaired

1. Singular value decomposition (SVD):

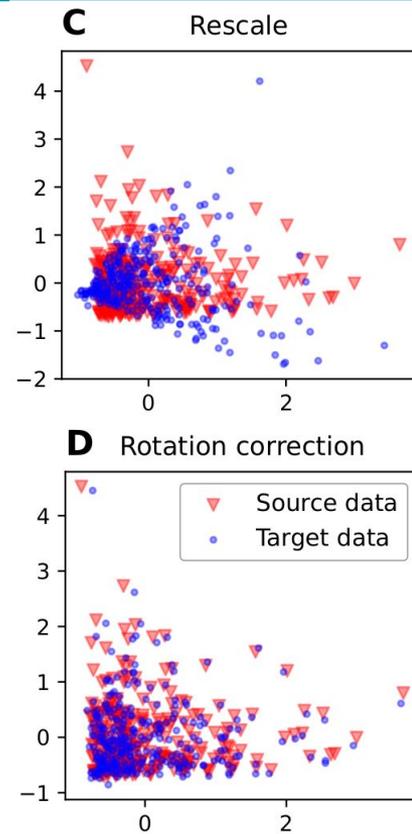
source: $Z^{(s)} = \{z_i^{(s)}, i = 1 \dots N^{(s)}\}$ target: $Z^{(t)} = \{z_i^{(t)}, i = 1 \dots N^{(t)}\}$

$$Z^{(s)} = U^{(s)\top} S^{(s)} V^{(s)}$$

$$Z^{(t)} = U^{(t)\top} S^{(t)} V^{(t)}$$

2. Separate rotation correction

source: $h_{U^{(s)}}^{\text{rot}}(Z^{(s)}) = U^{(s)\top} Z^{(s)}$ target: $h_{U^{(t)}}^{\text{rot}}(Z^{(t)}) = U^{(t)\top} Z^{(t)}$



[Maman et al., 2019]

Methods: Alignment steps

Step 3: Correct the rotation - Method paired

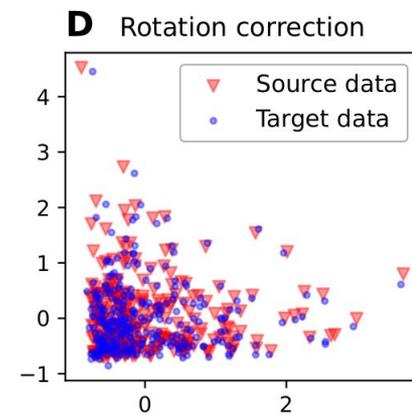
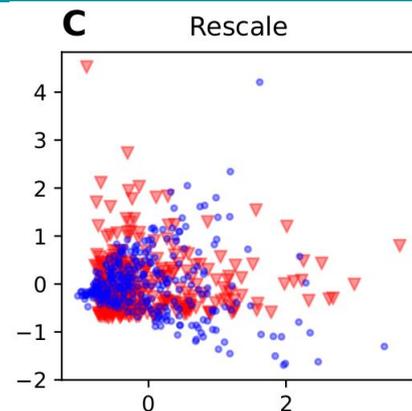
1. SVD:

$$Z^{(t)} Z^{(s)\top} = U^\top S V \implies P = V U^\top$$

2. Rotation correction

$$h_P^{\text{rot}}(Z^{(t)}) = P Z^{(t)}$$

→ Same number of matching observations



[Bleuze et al., 2021]

Methods: Alignment methods

4 functions:

Re-center to Identity: $h_{\bar{C}}^{\text{rect}}$

Normalize dispersion: h_d^{str}

Rotation correction unpaired: h_U^{rot}

Rotation correction paired: h_P^{rot}

5 methods:

No alignment

Recenter: $h_{\bar{C}}^{\text{rect}}$

Rescale: $h_d^{\text{str}} \circ h_{\bar{C}}^{\text{rect}}$

Procrustes unpaired: $h_U^{\text{rot}} \circ h_d^{\text{str}} \circ h_{\bar{C}}^{\text{rect}}$

Procrustes paired: $h_P^{\text{rot}} \circ h_d^{\text{str}} \circ h_{\bar{C}}^{\text{rect}}$

+ comparison with a **z-score** method

Experiments: Simulated data and shift scenarios

Simulation study: generative model with $Q = P = 20$ and $N=300$

Reminder: $C_i = A \text{diag}(p_i) A^\top$ and $y_i = \sum_{j=1}^Q \beta_j \log(p_{i,j})$

Experiments: Simulated data and shift scenarios

Simulation study: generative model with $Q = P = 20$ and $N=300$

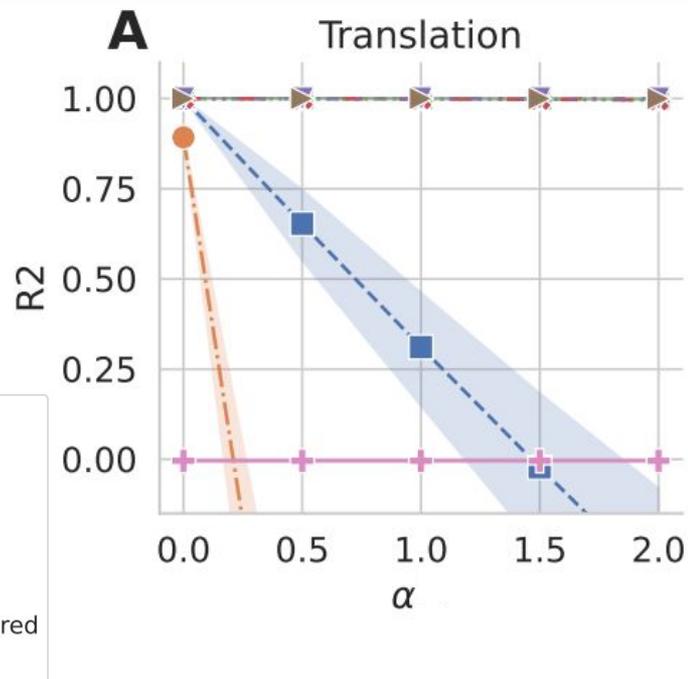
Reminder: $C_i = A \text{diag}(p_i) A^\top$ and $y_i = \sum_{j=1}^Q \beta_j \log(p_{i,j})$

Scenario 1: Translation

$$A^{(t)} = (B)^\alpha A^{(s)} \quad \alpha \in \mathbb{R}_*^+$$

Results:

- Re-center compensates the shift
- z-score performs worse than using Riemannian framework without alignment



Experiments: Simulated data and shift scenarios

Simulation study: generative model with $Q = P = 20$

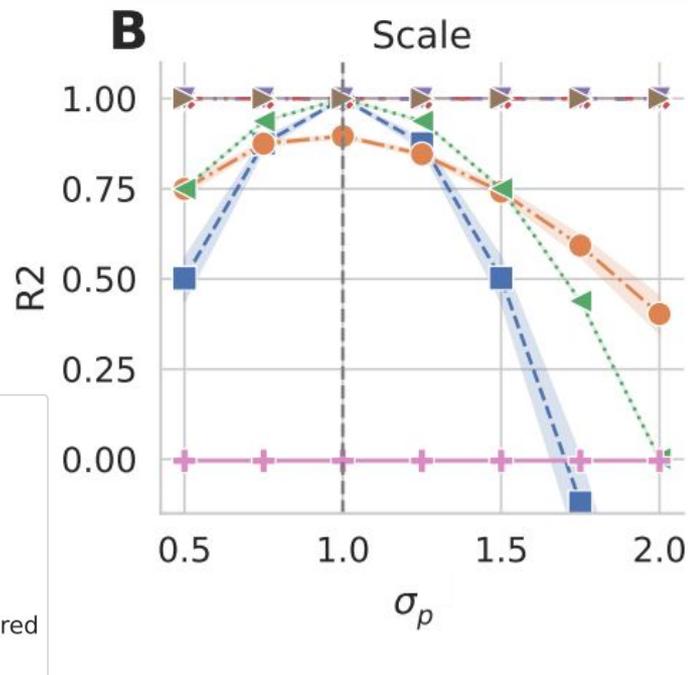
Reminder: $C_i = A \text{diag}(p_i) A^\top$ and $y_i = \sum_{j=1}^Q \beta_j \log(p_{i,j})$

Scenario 2: Scale

$$p_i^{(t)} = (p_i^{(s)})^{\sigma_p} \quad \sigma_p \in \mathbb{R}_*^+$$

Results:

- Re-scale compensates the shift
- Re-center is better than no alignment
- z-score improves over re-center when the shift increases



Experiments: Simulated data and shift scenarios

Simulation study: generative model with $Q = P = 20$

Reminder: $C_i = A \text{diag}(p_i) A^\top$ and $y_i = \sum_{j=1}^Q \beta_j \log(p_{i,j})$

Scenario 3: Translation and rotation

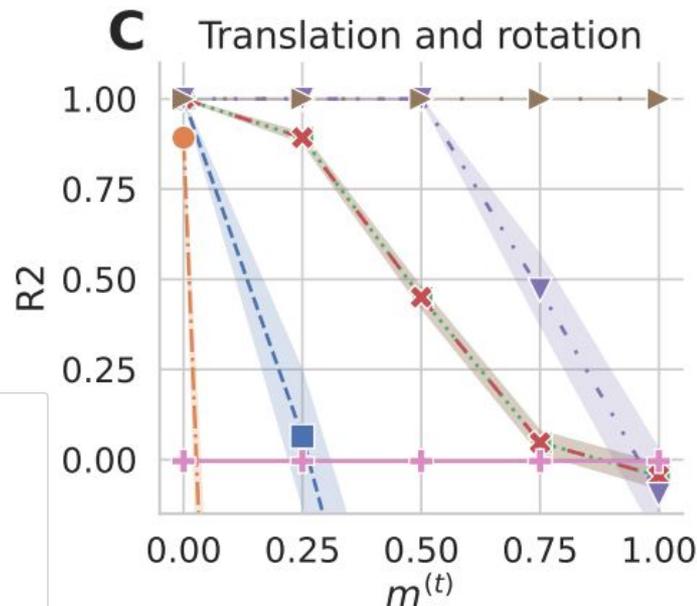
Gaussian random matrix A_t

$$A^{(t)} = mA_t + (1 - m)A^{(s)} \quad m \in [0, 1]$$

Results:

→ Procrustes paired reach perfect performance and Procrustes unpaired breaks when the shift increases.

→ Re-center is not enough



Experiments: Simulated data and shift scenarios

Simulation study: data are simulated following the generative model with $Q = P = 20$

Reminder: $C_i = A \text{diag}(p_i) A^\top$ and $y_i = \sum_{j=1}^Q \beta_j \log(p_{i,j})$

Scenario 4: Noise on mixing matrix

$$A_i^{(s)} = A + E_i^{(s)} \quad \text{with} \quad E_i^{(s)} \sim \mathcal{N}(0, (\sigma_A^{(s)})^2)$$

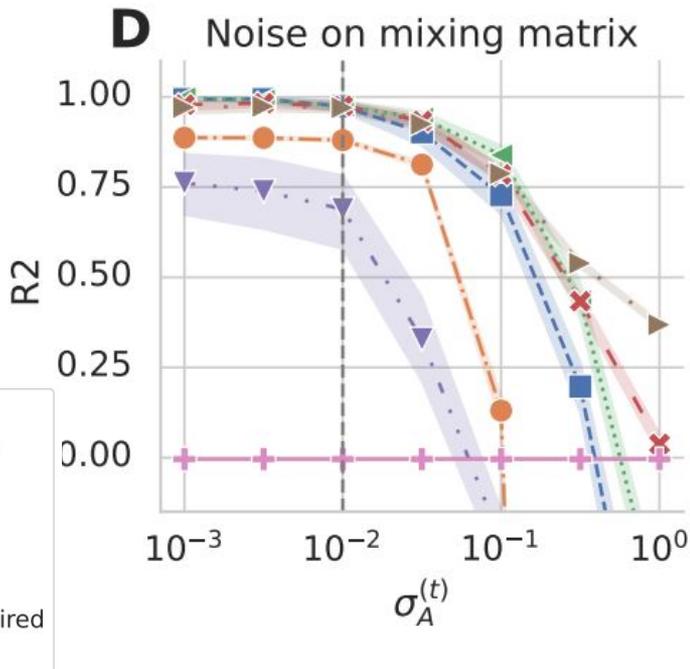
and similarly for target.

Results:

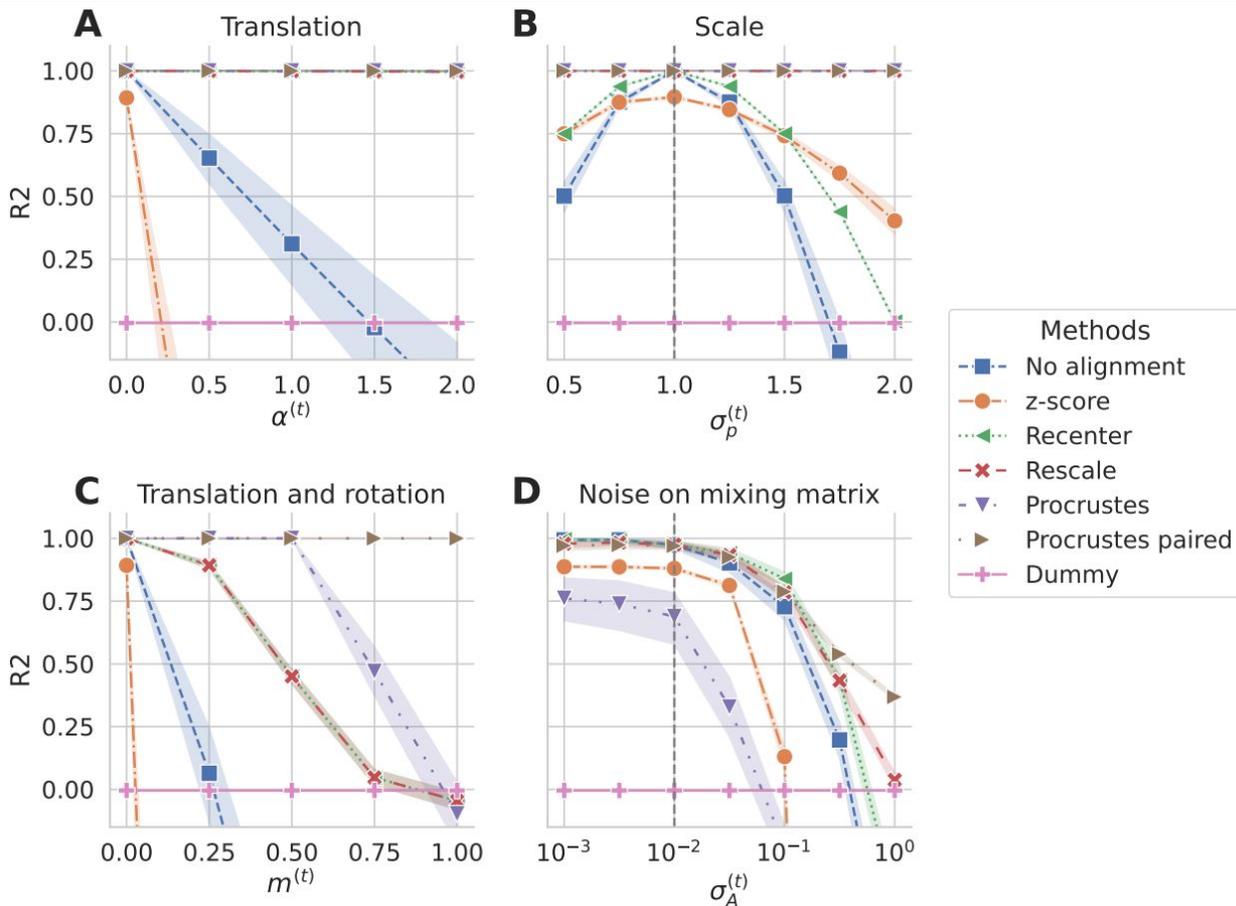
→ Procrustes unpaired performs the worst followed by z-score

→ All other methods perform similarly

→ Procrustes paired is slightly better



Experiments: Simulated data and shift scenarios



- Procrustes paired performs best in all scenarios but requires the source and target sets to be the same size and have corresponding points
- When the points are unpaired re-center and re-scale is the best solution
- Procrustes unpaired is unstable when the shift is too big or when there is noise

Experiments: Cam-CAN

MEG experiments: 646 healthy subjects from the Cam-CAN dataset, 18 to 89 years old.

3 tasks: resting-state, passive task, and somatosensory task → **domains**

Processing:

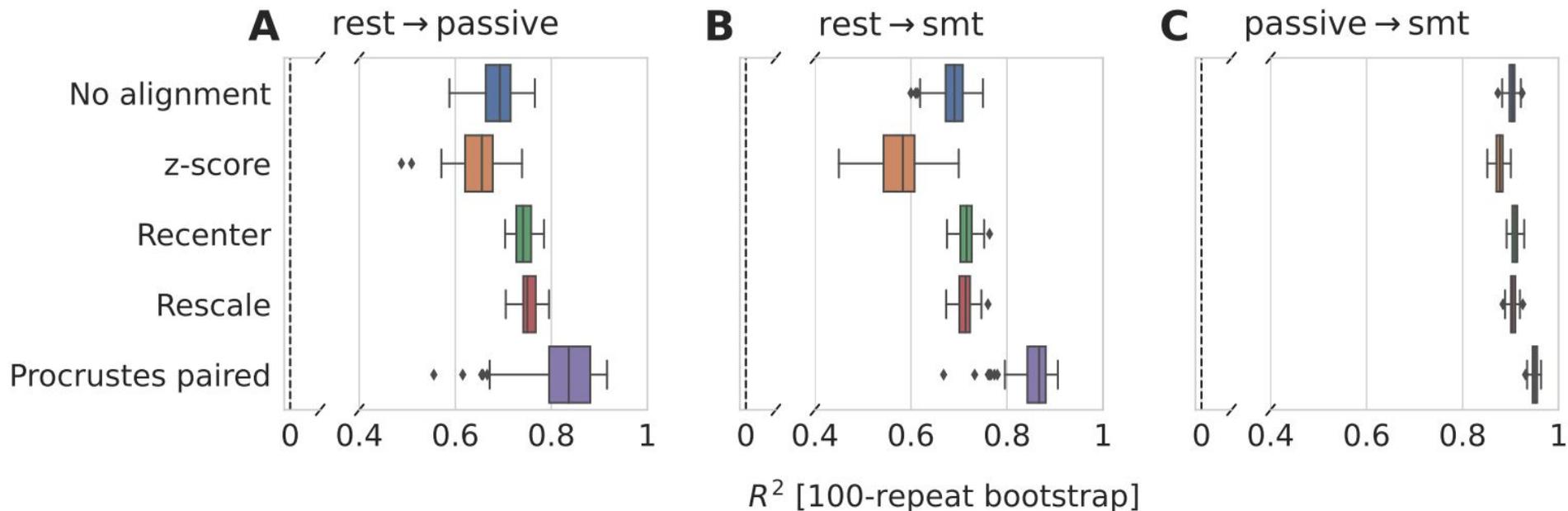
- Band-pass filter between 0.1 and 49 Hz
- Resampling at 200 Hz
- temporal signal-space-separation (tSSS)
- 10-second epochs filtered in 7 frequency bands and averaged
- Computation of the averaged epochs' covariance matrices

→ 7 covariance matrices per subject

We evaluate the alignment methods on **age** prediction with Ridge regression.

Experiments: Cam-CAN - different tasks, same subjects

1st experiment: same subjects in source and target



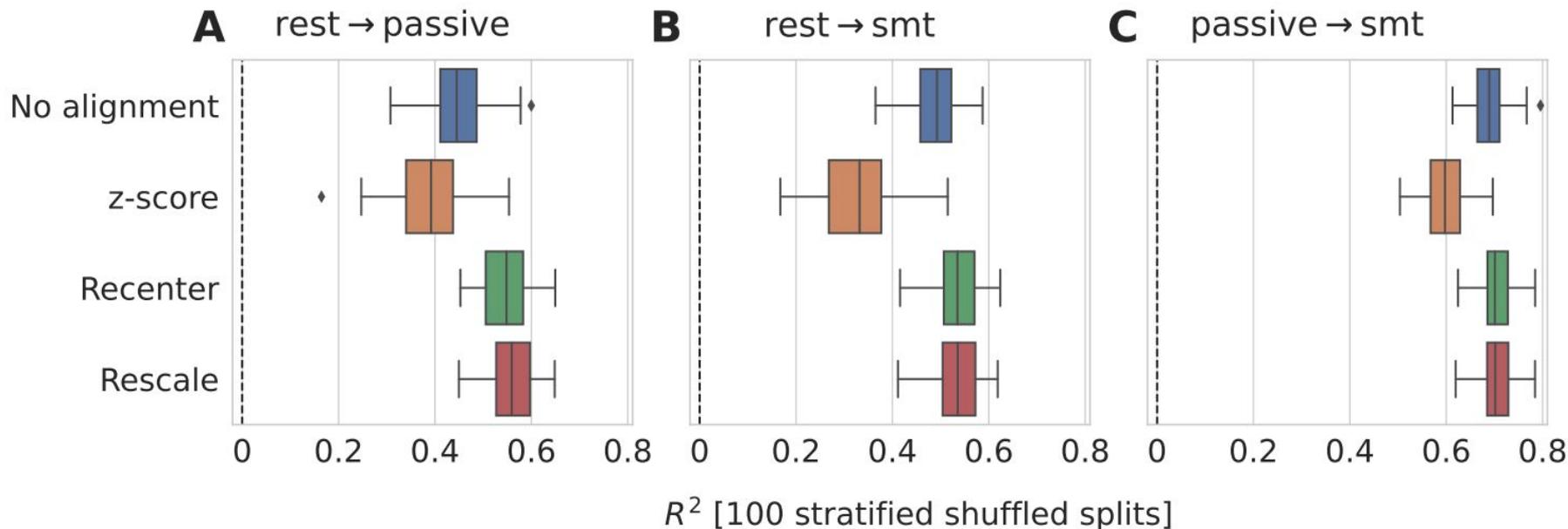
→ Re-center leads to a clear gain in performance

→ Re-scale do not affect the predictions

→ Procrustes paired improves the score of 83 splits compared to re-center

Experiments: Cam-CAN - different tasks, different subjects

2nd experiment: different subjects in source and target



- Re-center leads to a clear gain in performance
- Re-scale do not affect the predictions
- Not possible to apply Procrustes Paired in this setting

Experiments: EEG datasets - TUAB → LEMON

EEG experiments: 1385 healthy subjects from the TUAB dataset and 213 from LEMON

TUAB (source): 0 to 95 years old

LEMON (target): 20 to 35 and 55 to 77 years old

Processing:

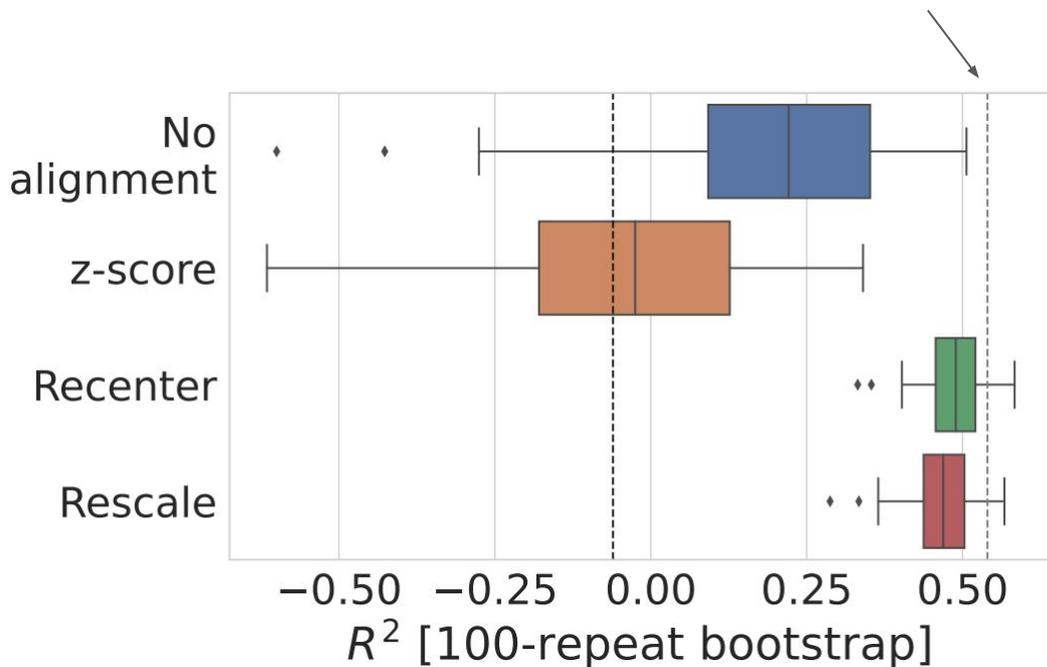
- Band-pass filter between 0.1 and 49 Hz
- Resampling at 200 Hz
- 10-second epochs filtered in 7 frequency bands and averaged
- Computation of the averaged epochs' covariance matrices

→ 7 covariance matrices per subject

We evaluate the alignment methods on **age** prediction with Ridge regression.

Experiments: EEG datasets - TUAB → LEMON

Score obtained with a 10-fold cross-validation on LEMON



→ z-score at the level of dummy

→ Re-center improves the performance while re-scale has still no impact

→ With re-center we reach scores similar to what is obtained when the training is done on LEMON data

Conclusion:

- The z-score method systematically performed worse than all the others.
- Recenter leads to a clear improvement in performance in all experiments, especially in the EEG experiment.
- Rescale is not helping in our M/EEG experiments.
- Rotation correction is useful when it is possible to pair subjects

Limitations:

- No rotation correction methods in the two last experiments.
- The methods only works with the same sensors in source and target domains.
- We only performed age prediction in healthy subjects
- Need for more investigation: other populations / datasets, other prediction context.