Biomag competition: Dementia screening

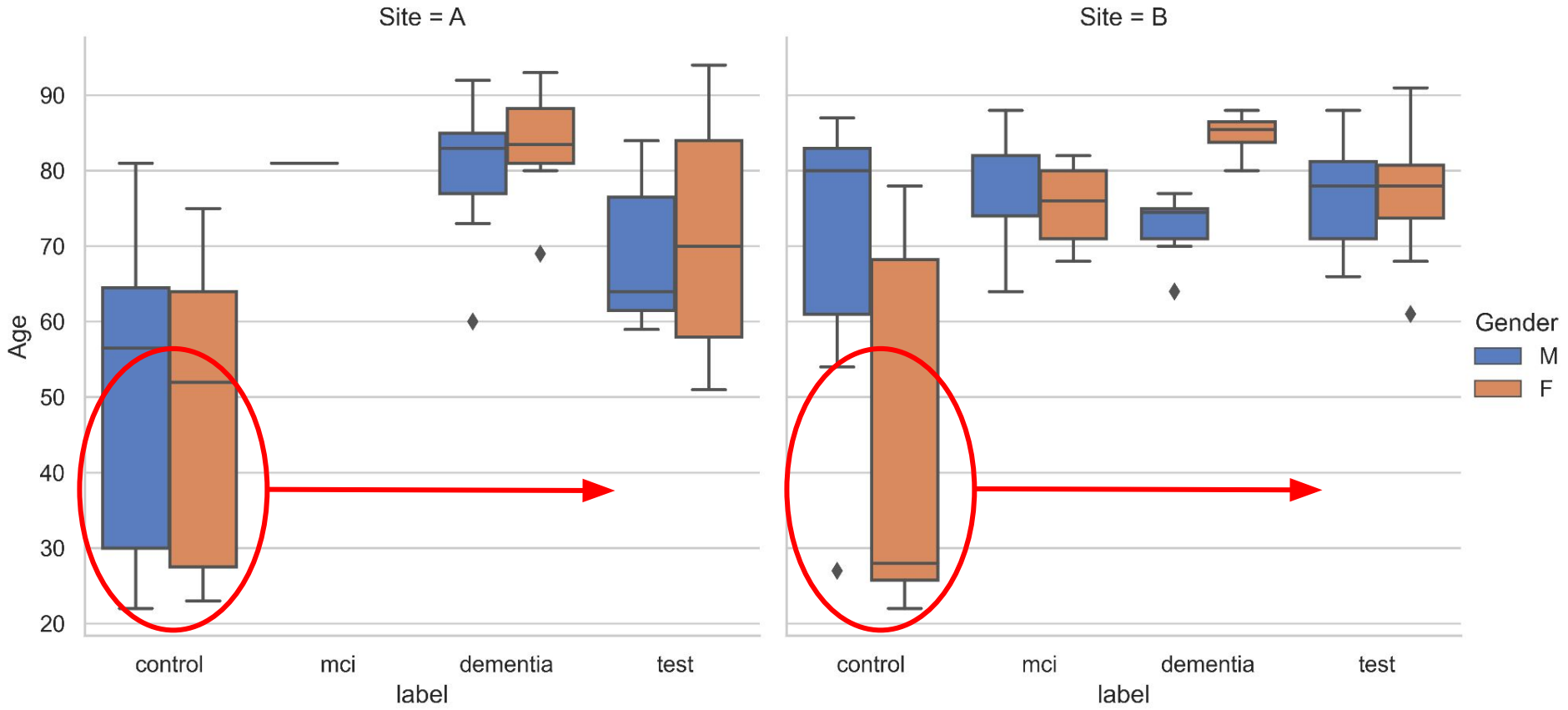
Apolline MELLOT - Benoît MALEZIEUX - Cédric ALLAIN

Spring 2022

https://github.com/apmellot/Dementia_screening_challenge_biomag_2022.git

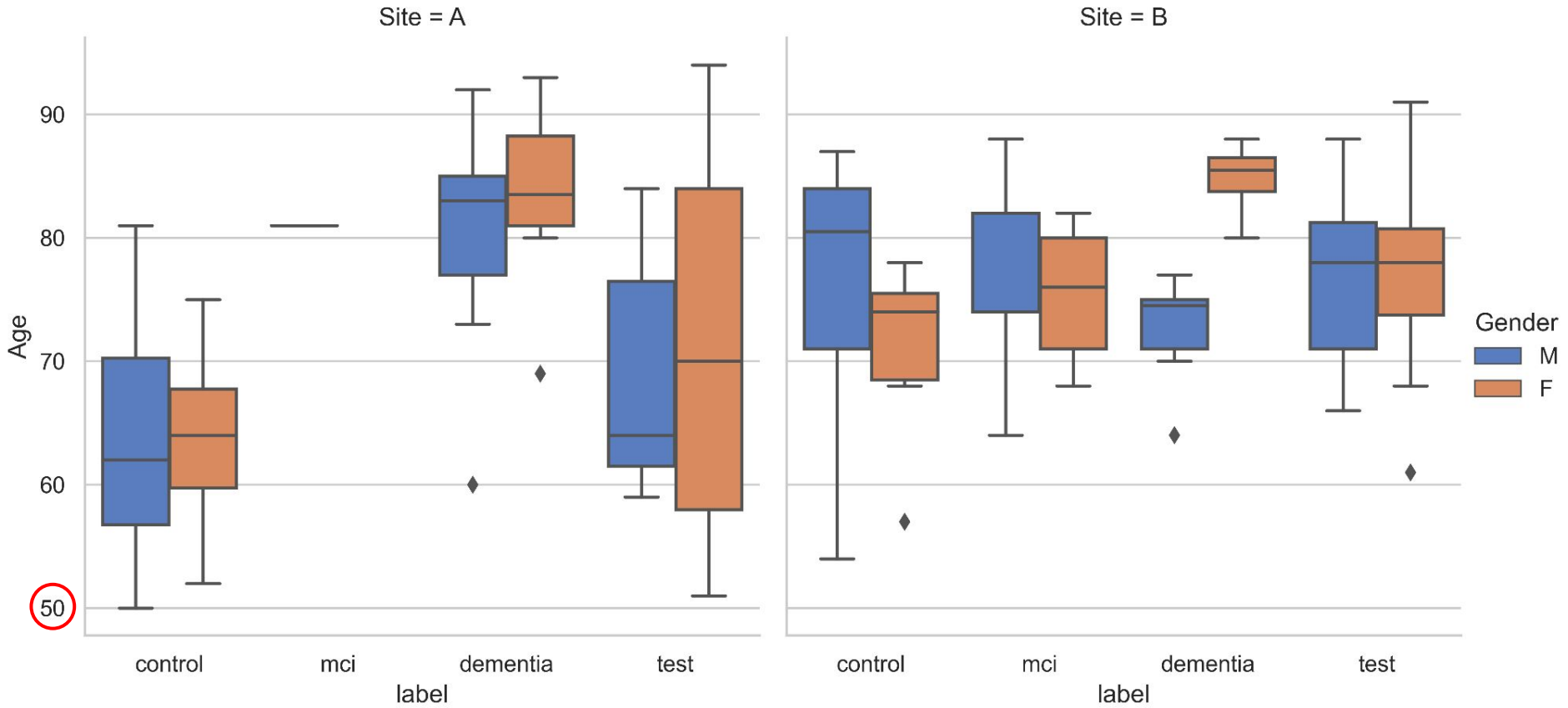
Age distribution across sites:

The train control dataset is deeply imbalanced in terms of age.



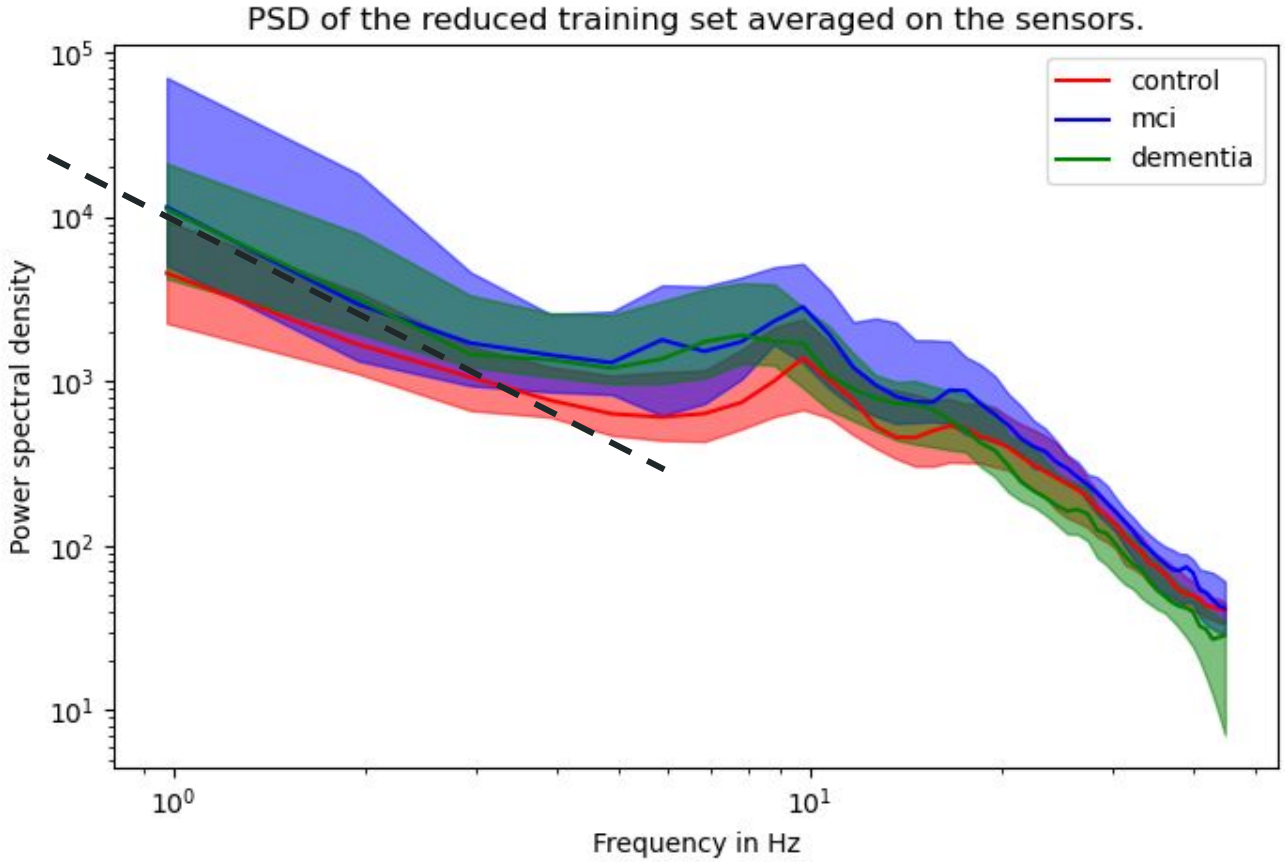
Age distribution across sites:

Only subjects over 50 years of age are kept
(43 subjects removed)



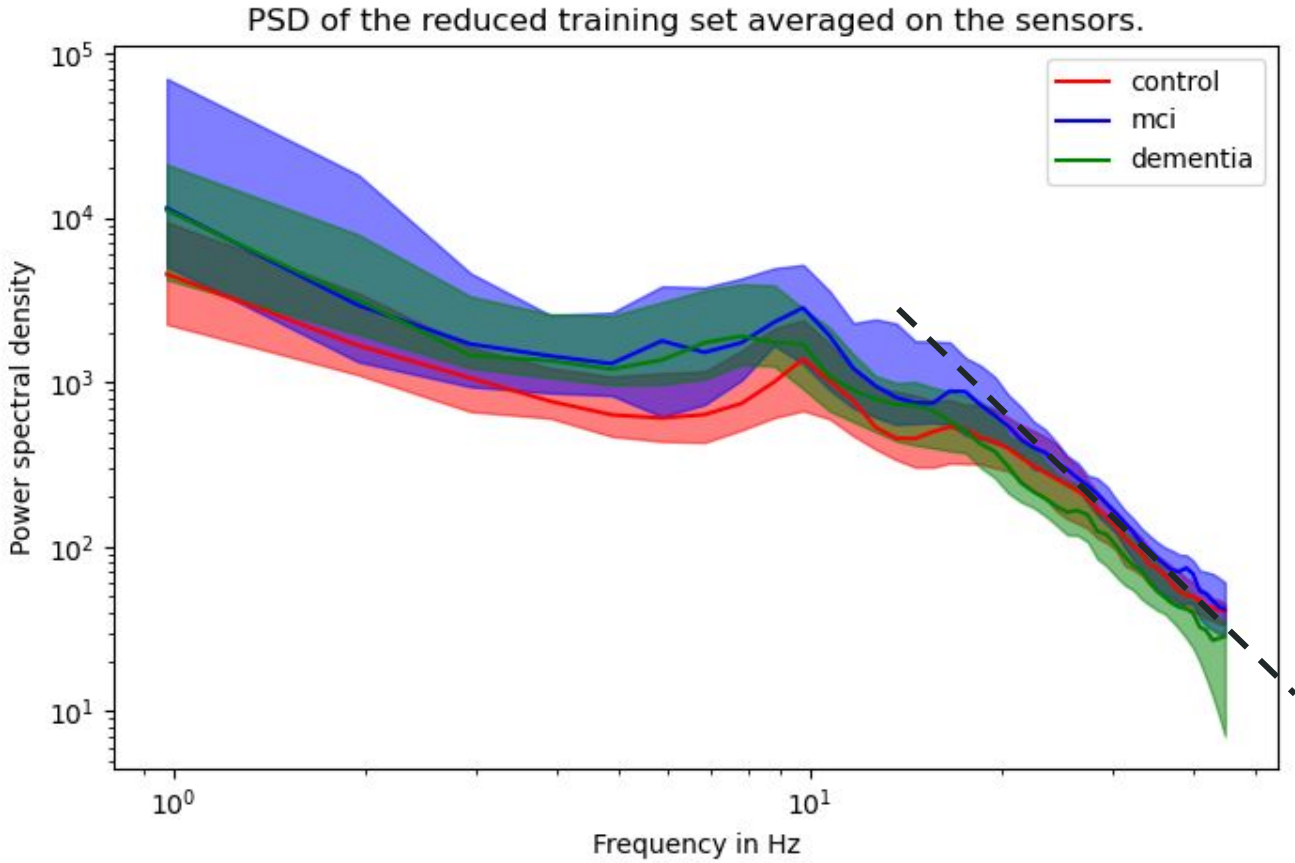
Features extracted from the Power Spectral Densities:

➔ 1/f slope between 0.1 to 1.5 Hz



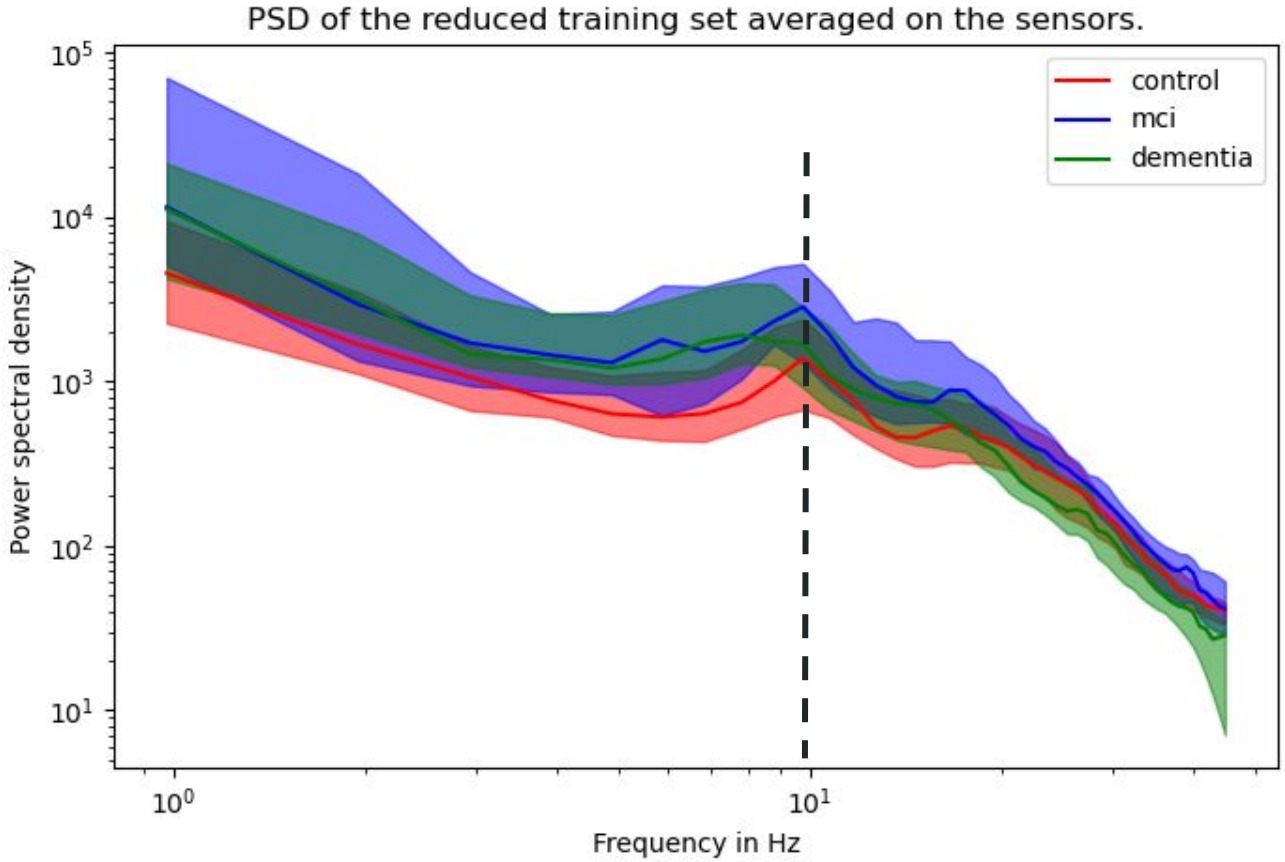
Features extracted from the Power Spectral Densities:

- ➔ 1/f slope between 0.1 to 1.5 Hz
- ➔ 1/f slope between 35 to 49 Hz



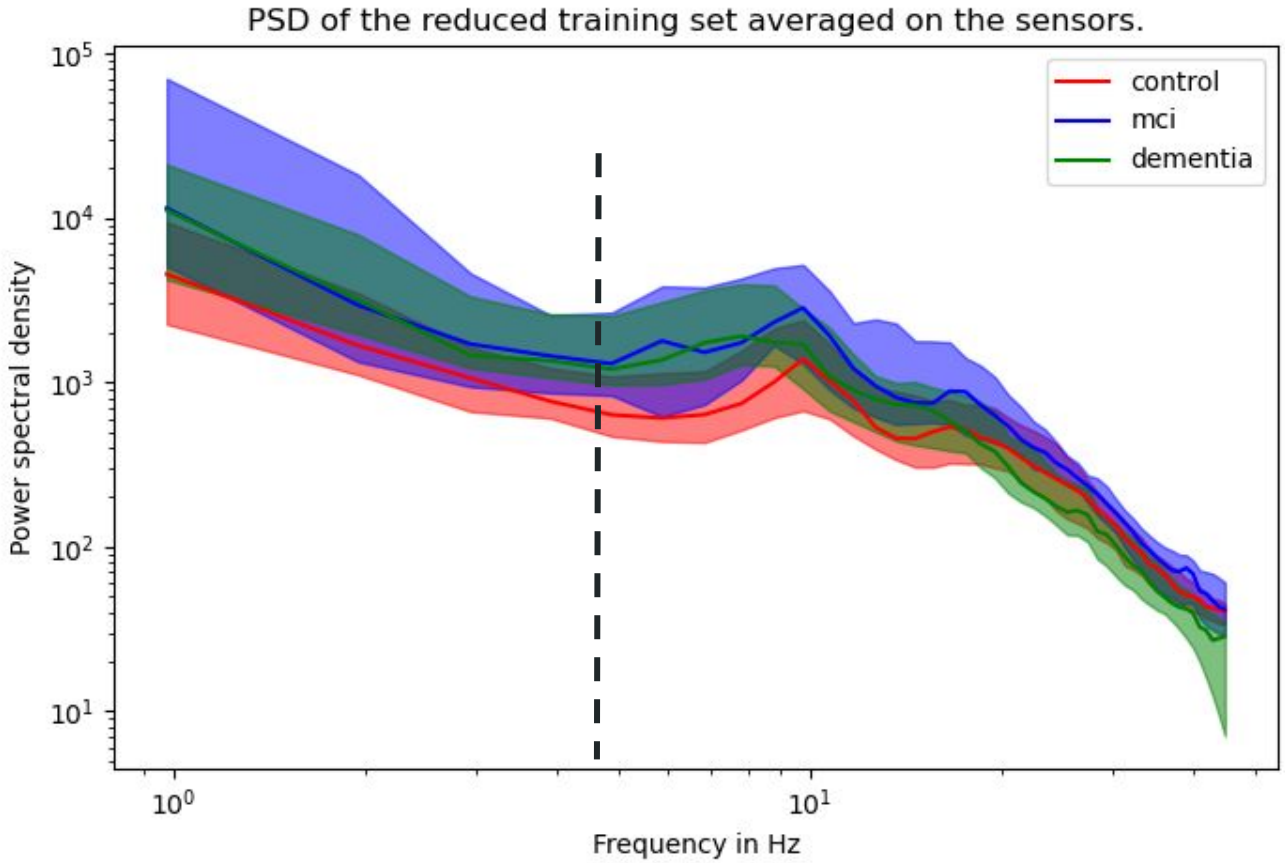
Features extracted from the Power Spectral Densities:

- ➔ 1/f slope between 0.1 to 1.5 Hz
- ➔ 1/f slope between 35 to 49 Hz
- ➔ Alpha peak frequency



Features extracted from the Power Spectral Densities:

- ➔ 1/f slope between 0.1 to 1.5 Hz
- ➔ 1/f slope between 35 to 49 Hz
- ➔ Alpha peak frequency
- ➔ Median power frequency



Feature vectors extracted from covariances:

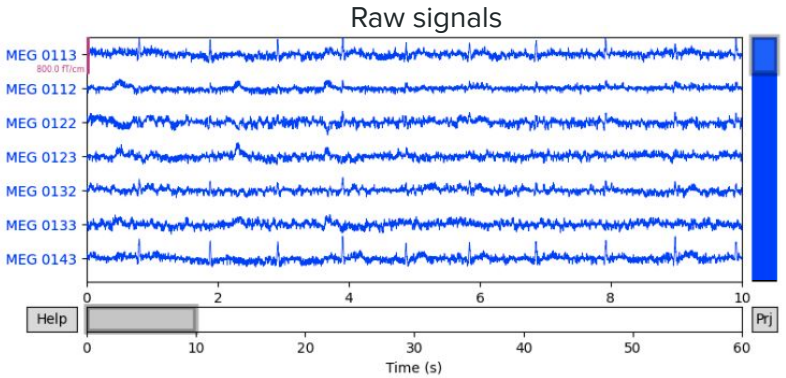
M/EEG signal:

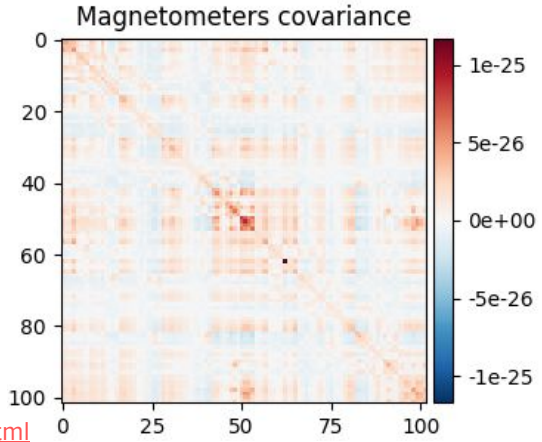
$$X(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \in \mathbb{R}^{n \times T_S}$$

Covariances:

$$C = \frac{1}{T_s - 1} X(t) X(t)^T \in \mathbb{R}^{n \times n}$$

where n is the number electrodes and T_s the number of sampled time points





https://mne.tools/stable/index.html

Feature vectors extracted from covariances:

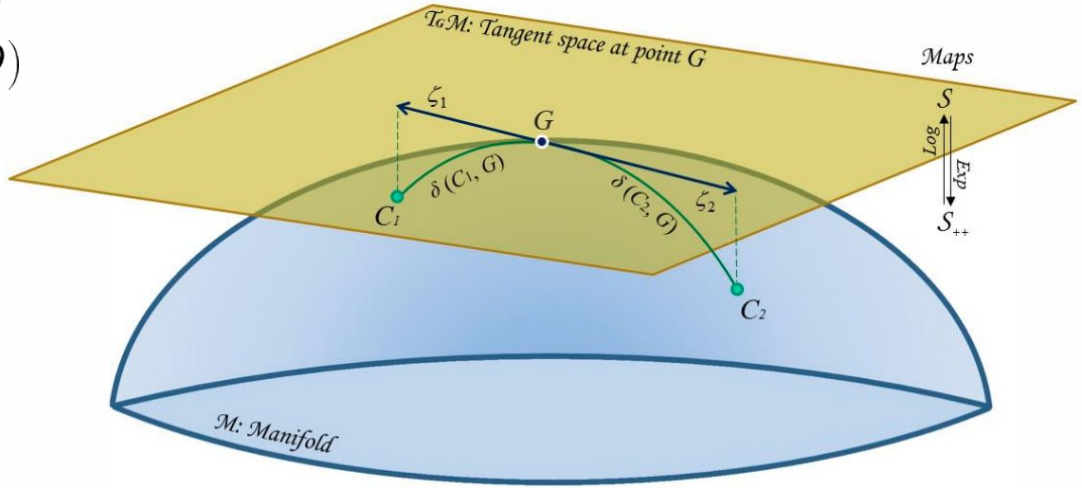
1. Covariance matrices (here of size 160x160) lie in a Riemannian manifold.
2. Dimension reduction from 160x160 to 120x120.
3. Tangent space projection to get an Euclidean framework:

$$\zeta_1 = Log_G(C_1) = G^{\frac{1}{2}}\, log(G^{-\frac{1}{2}}\, C_1\, G^{-\frac{1}{2}})\, G^{\frac{1}{2}}$$

4. Vectorization: $z_1 = uvec(\zeta_1\ o\ Q)$

with Q a matrix holding 1 on the diagonal elements and $\sqrt{2}$ elsewhere.

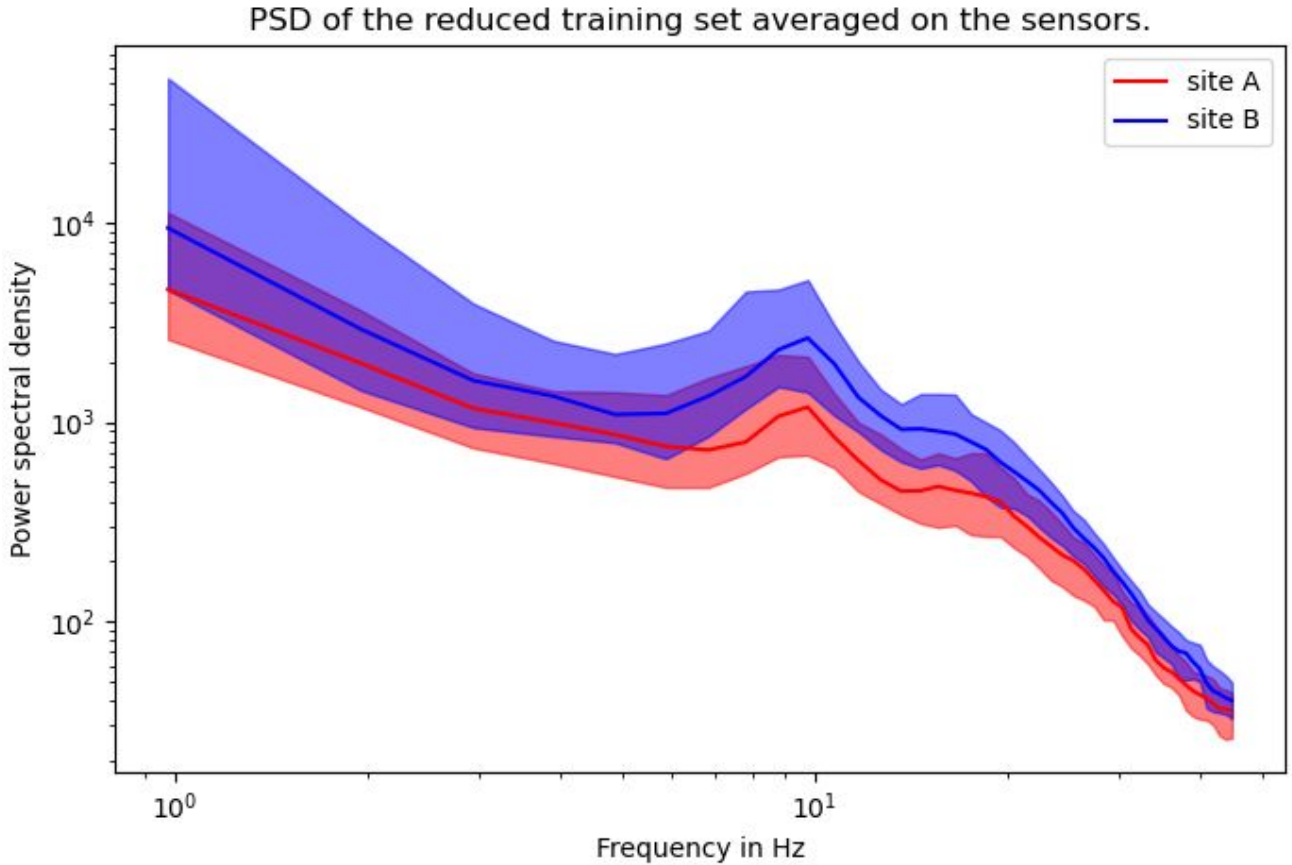
[Sabbagh et al. 2019, 2020]



[Congedo et al. 2017]

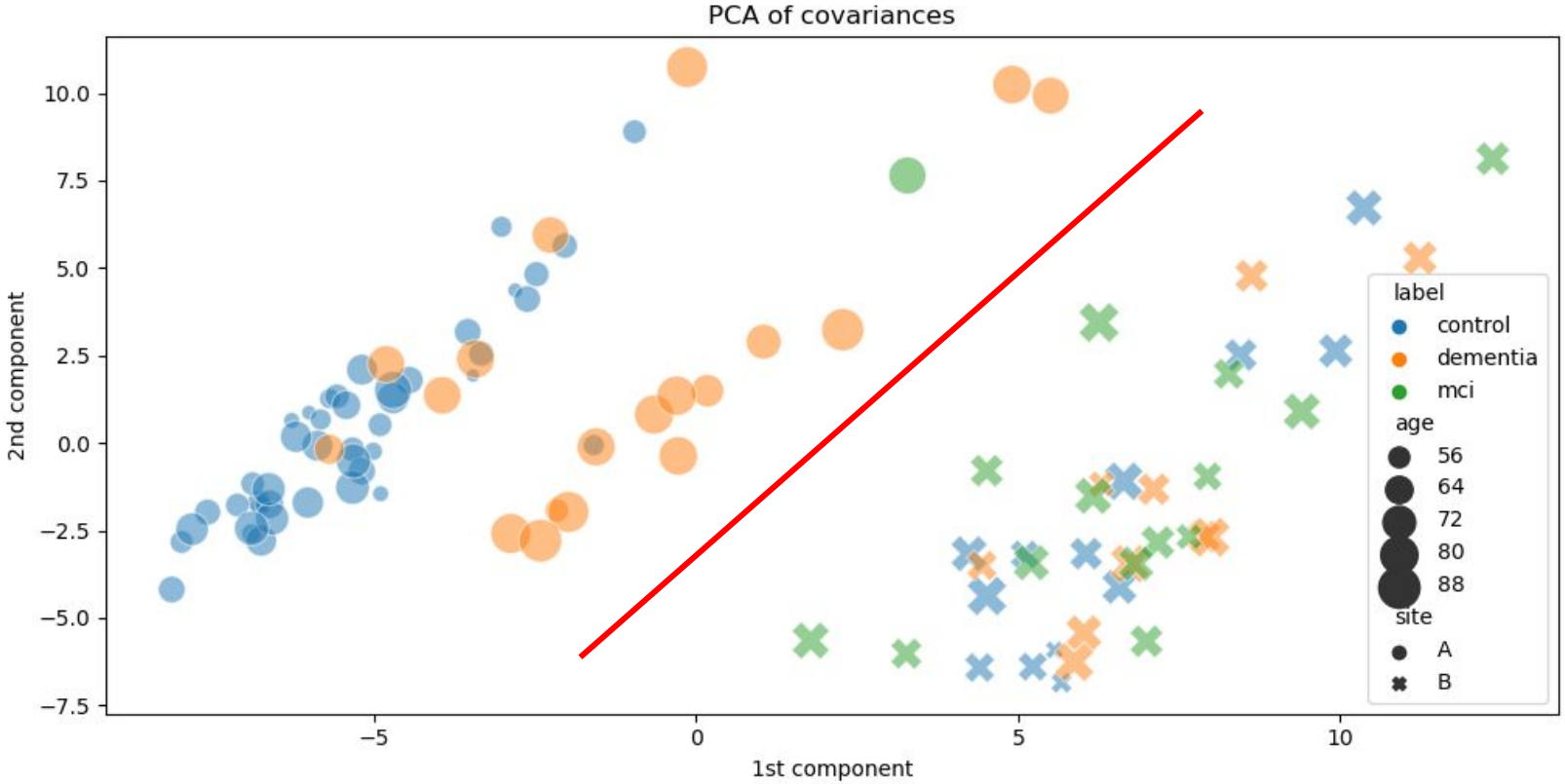
Remaining issue: recording site

PSD intensity offset between the 2 sites.

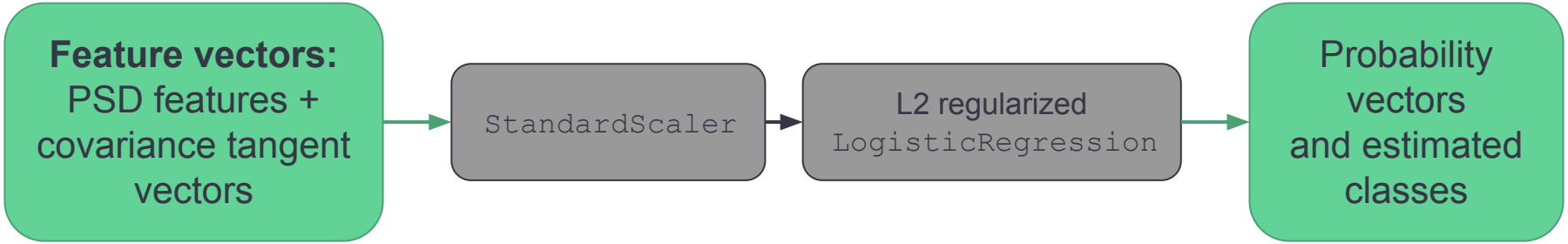


Remaining issue: recording site

The 2 sites are fully separable: prevent generalization



Classification pipeline:



Results on the training set:

➔ `StratifiedShuffleSplit` cross-validation from scikit-learn with 20 splits and `test_size=0.2`

Mean accuracy: 71.9% ± 8.3%

Mean balanced accuracy: 64.3% ± 9.6%

Final results on the test set: