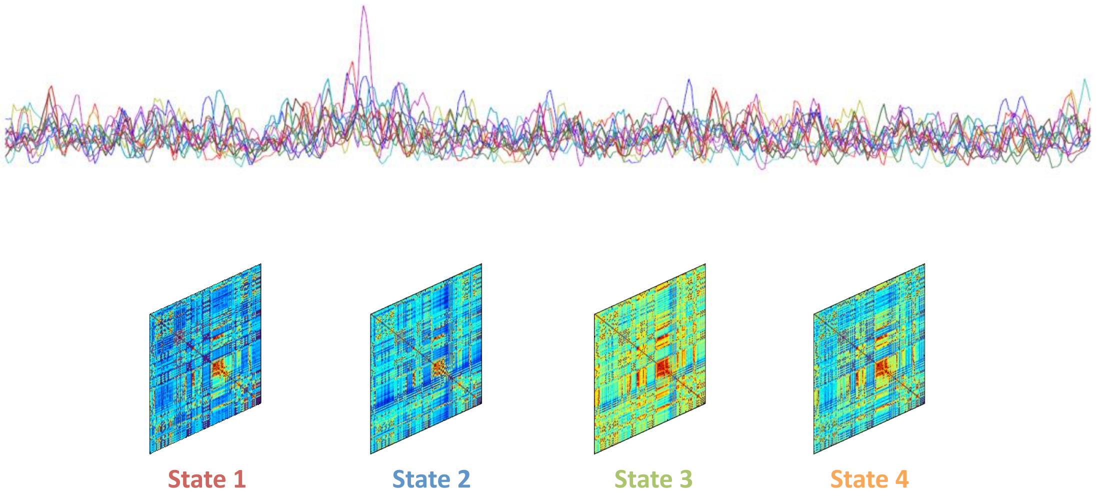
GLHMM workshop

Introductory module

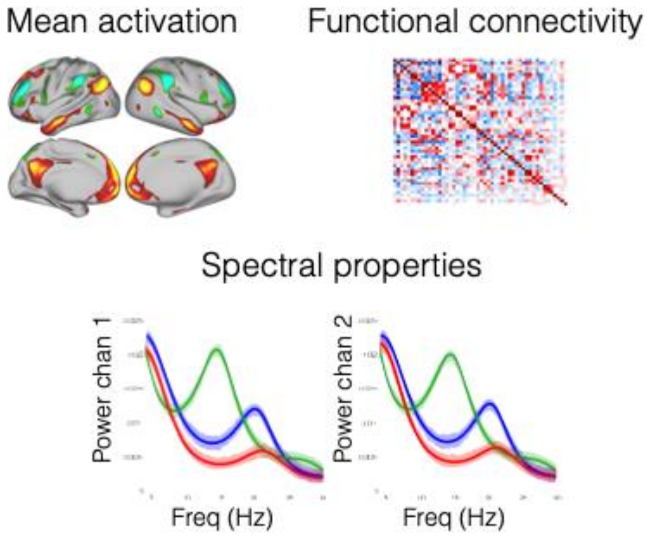
Laura Masaracchia, PhD

HMM recap

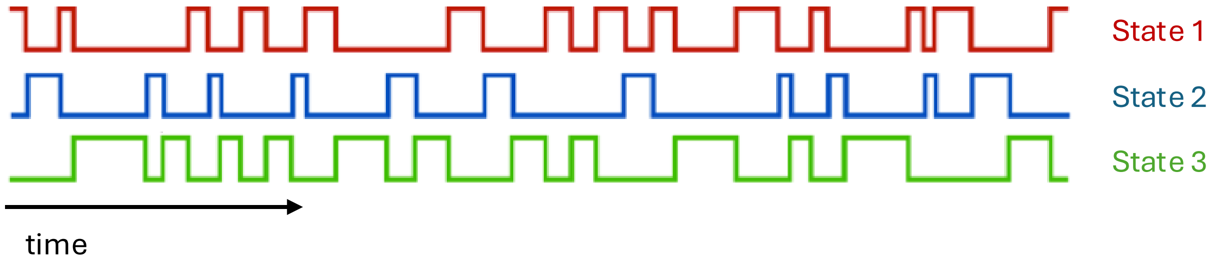
HMM finds recurring patterns in the data (states)



Key elements of HMM states:



State time courses



HMM families: flexible state definitions

Different HMM families: flexible state models

- Gaussian (aka standard) HMM, [Baker et al., 2014, eLife]
- Gaussian Linear (aka decoding) HMM, [Vidaurre et al., 2025, Imaging Neuroscience]
- Multivariate autoregressive (MAR) HMM, [Vidaurre et al., 2016, NeuroImage]
- Time-delay embedded (TDE) HMM, [Vidaurre et al., 2018, Nature Communications]

Choice of HMM family depends on

- Data characteristics
- Scope of research

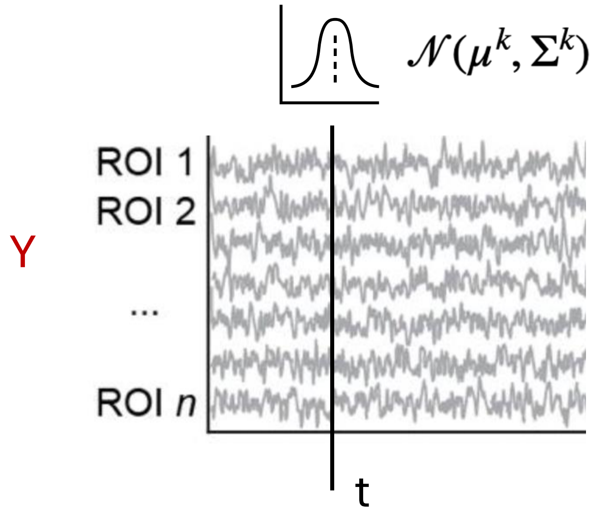
Gaussian (aka standard) HMM

Baker et al., 2014, eLife

Definition:

- For each time point t in data **Y**: computes probability of $\mathbf{Y_t}$ coming from a **Gaussian** distribution.
- Focus on instantaneous changes in the mean of the signal

$$Y_t \mid s_t = k \sim \mathcal{N}(\mu^k, \Sigma^k)$$



Application:

- Most appropriate for whole-brain fMRI data
- States interpreted as functional connectivity

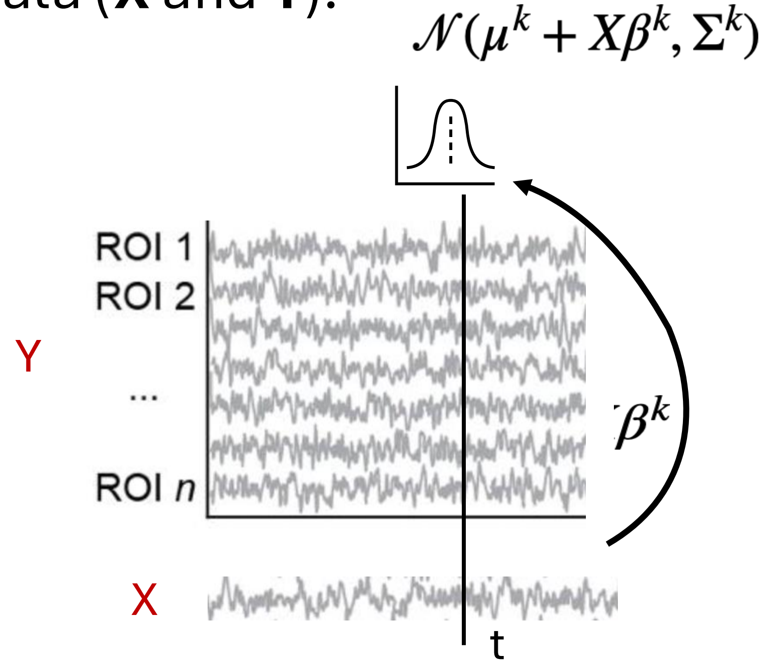
Gaussian-Linear (aka Decoding) HMM

[Vidaurre et al., 2025, Imaging Neuroscience]

Definition:

- Models the relationship between two sets of timeseries data (**X** and **Y**).
- Models $\mathbf{Y_t}$ like the standard HMM + predicting $\mathbf{Y_t}$ from **X**

$$Y_t \mid s_t = k \sim \mathcal{N}(\mu^k + X\beta^k, \Sigma^k)$$



Application:

- Useful to link behavioural and fMRI timeseries, or
- neural data across different brain regions / spatial scales

Multivariate Autoregressive (MAR) HMM

[Vidaurre et al., 2016, NeuroImage](#)

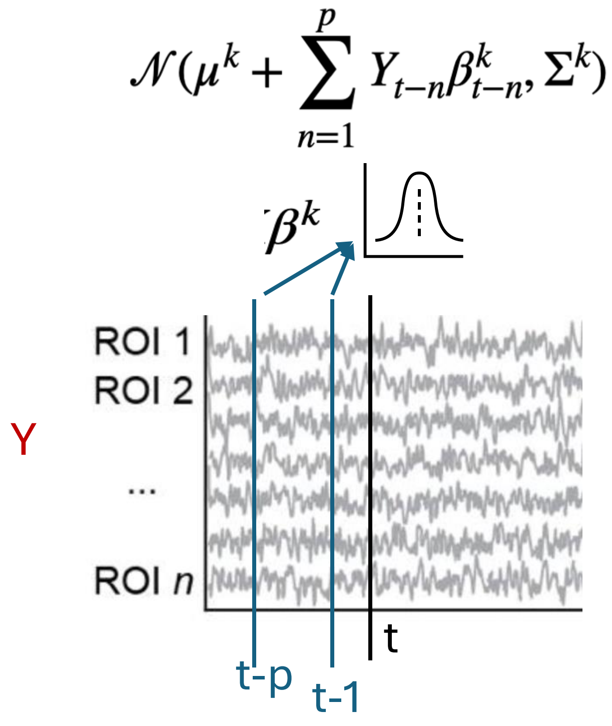
Definition:

- Predicts $\mathbf{Y_t}$ from *p* previous time points
- Focus on temporal changes in the signal

$$Y_t \mid s_t = k \sim \mathcal{N}(\mu^k + \sum_{n=1}^{p} Y_{t-n}\beta^k_{t-n}, \Sigma^k)$$

Application:

- Best to use low-dimensional data (MEG, LFP, EEG) for computational reasons
- Most appropriate to detect frequency changes in the data



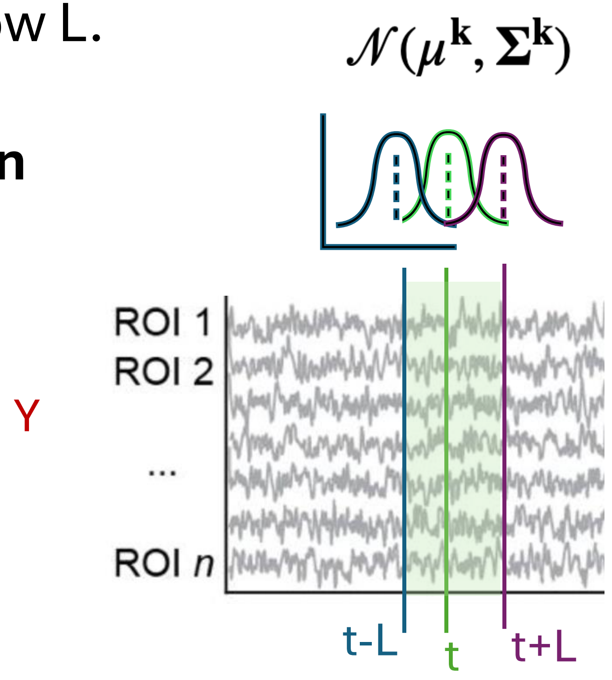
Time-delay embedded (TDE) HMM

Vidaurre et al., 2018, Nature Communications

Definition:

- Models the autocovariance of the signal within a time window L.
- For each t in the data **Y**: computes the probability of the **autocovariance** of $\mathbf{Y_t}$ to come from a **multivariate Gaussian** distribution.

$$Y_{t-L}, .. Y_t, .., Y_{t+L} \,|\, s_t = k \sim \mathcal{N}(\mu^{\mathbf{k}}, \mathbf{\Sigma}^{\mathbf{k}})$$

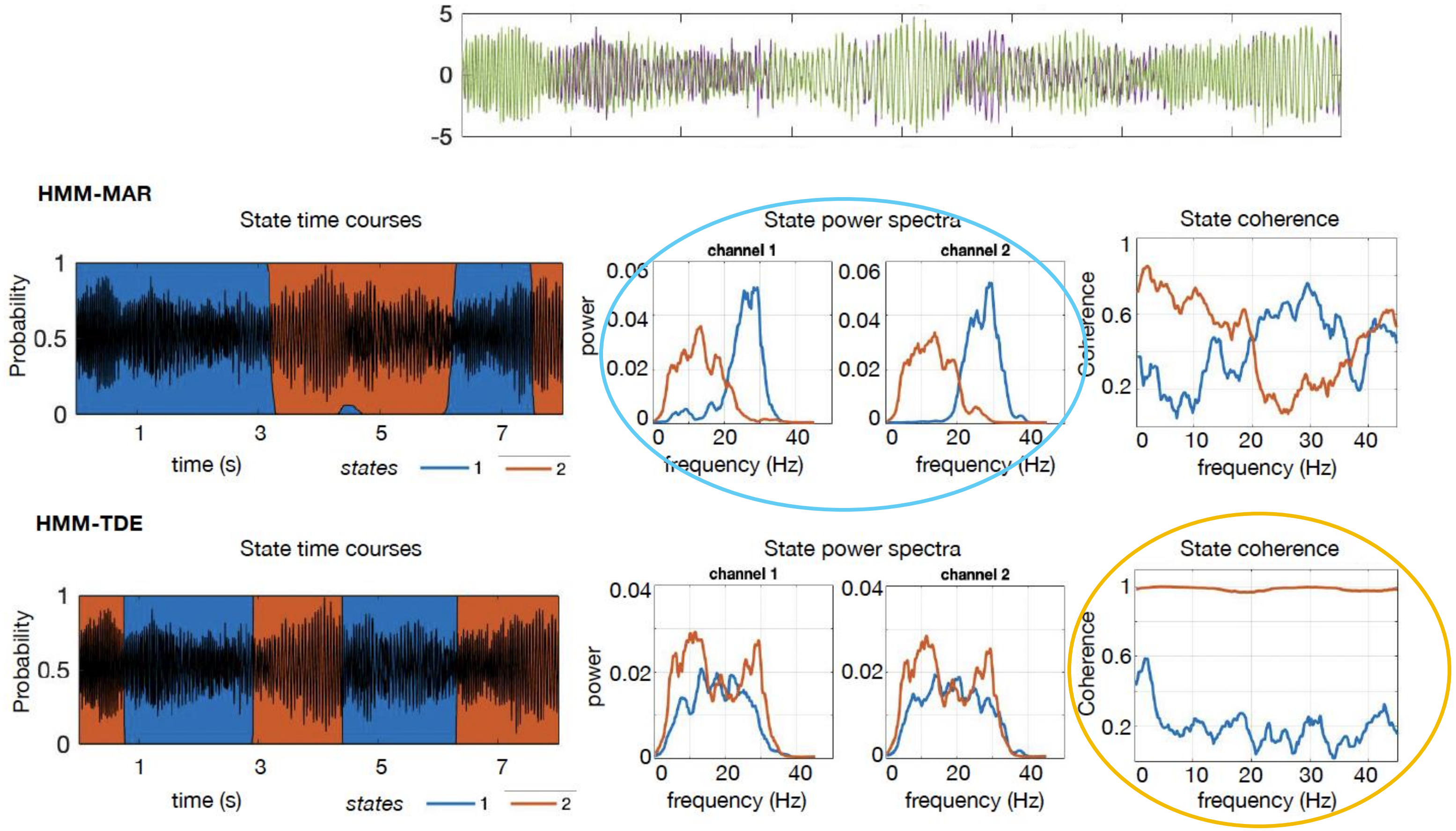


Application:

- Most appropriate for MEG multi-channel (whole-brain) data
- Most sensitive to frequency changes and cross-channel coherence in the data.

Practical example: HMM-MAR vs HMM-TDE

Two-channel synthetic signal



HMM-MAR : states focus on frequency

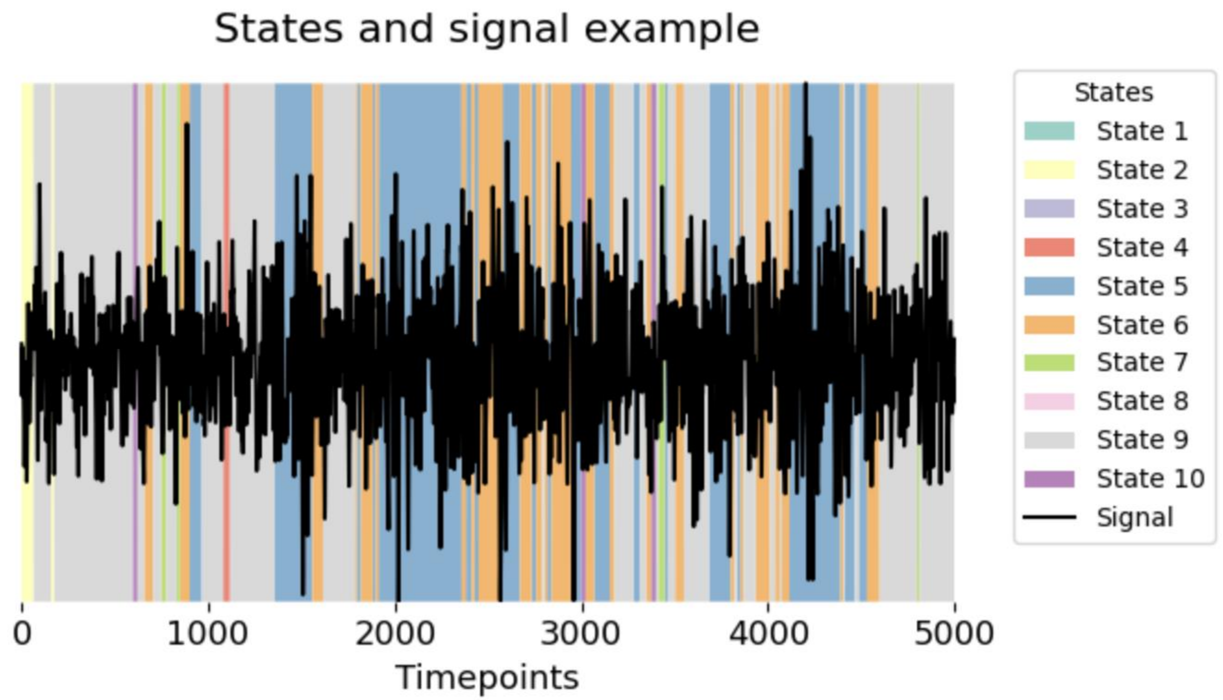
HMM-TDE: states focus on cross-channel coherence

Masaracchia *et al.*, 2023. *Journal of Neurophysiology*

The HMM output: State time courses

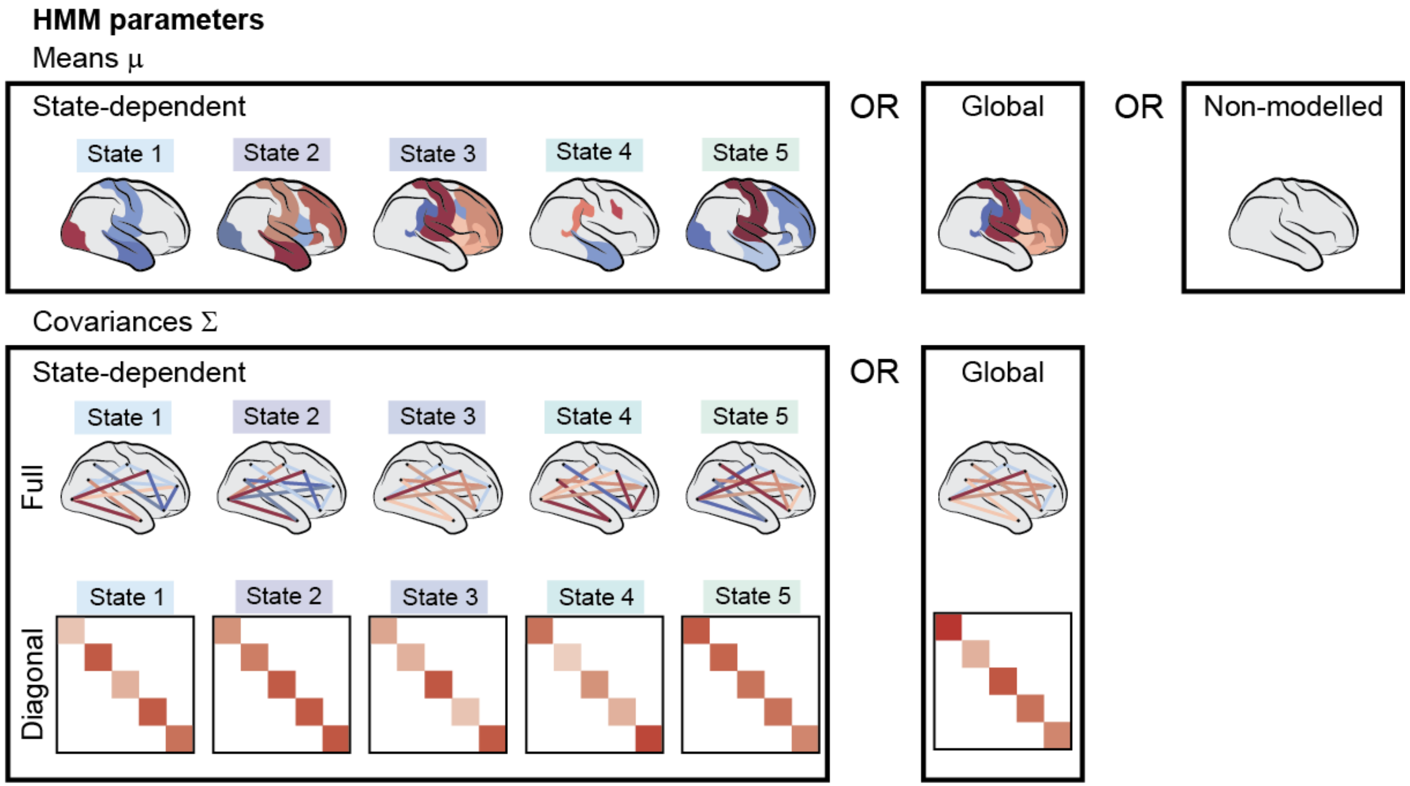
- State time courses (Gamma): the probability of each time point to belong to a state

- Viterbi path (vpath): a categorical version of the Gamma



The HMM output: States

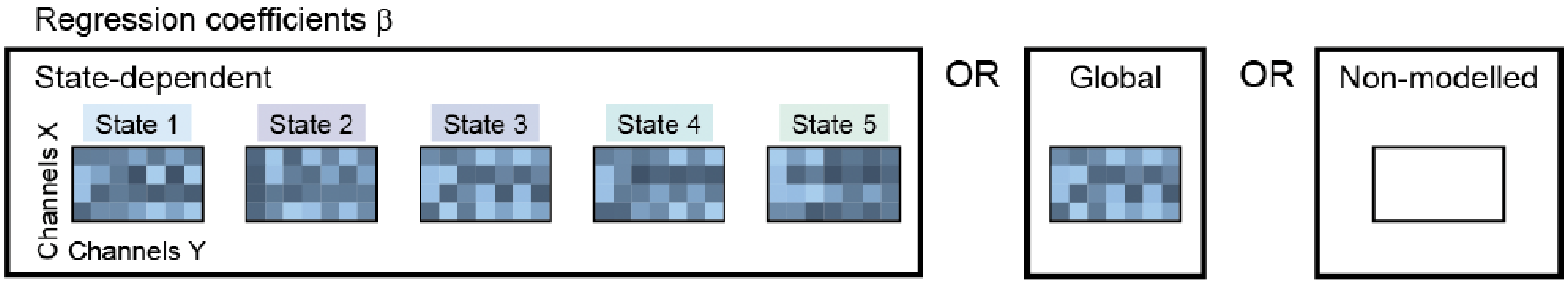
- States:
  - Mean (μ / mu), when included : amplitude of the state
  - Covariance (Σ / Sigma) : Functional Connectivity – patterns of activity across channels defining a state



Source: [Vidaurre et al., 2025, Imaging Neuroscience]

The HMM output: Regression Coefficients

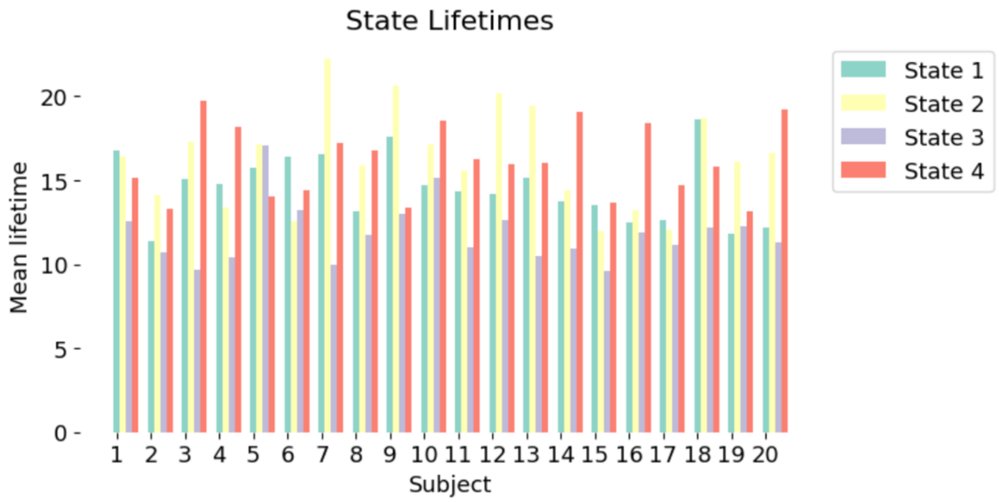
- Beta (β), when included: link to previous time points or to another time series



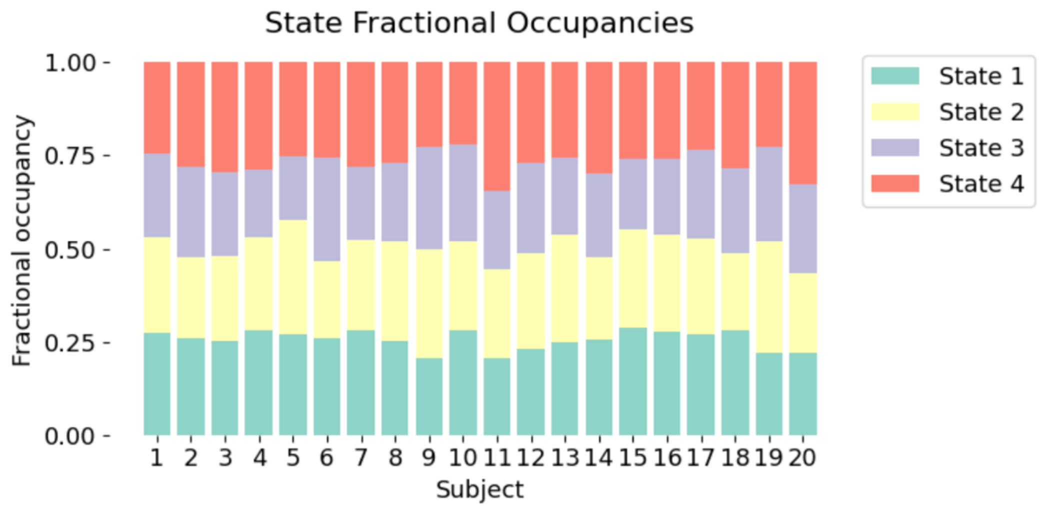
Source: [Vidaurre et al., 2025, Imaging Neuroscience]

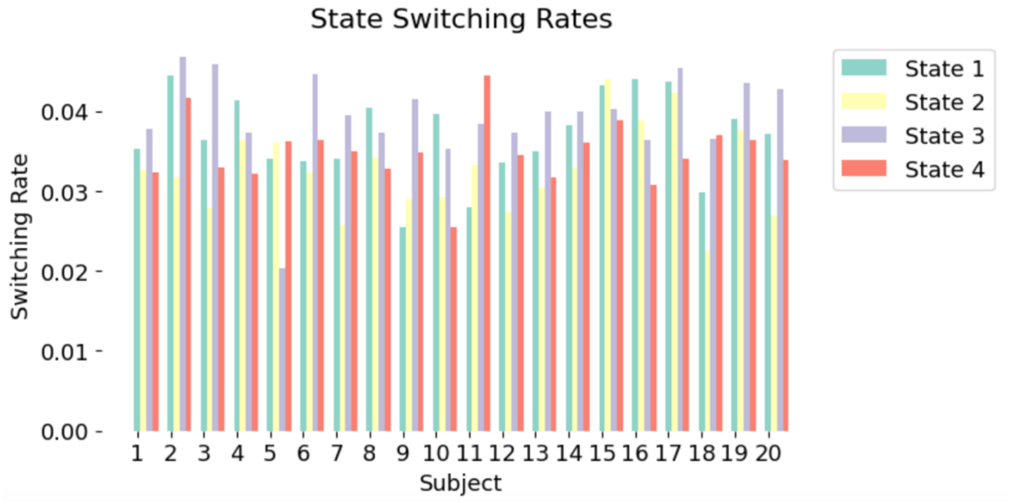
Summary metrics

- States Fractional Occupancy (FO) – how much on average each state is active
- States Lifetime (LT) – average time each state is active



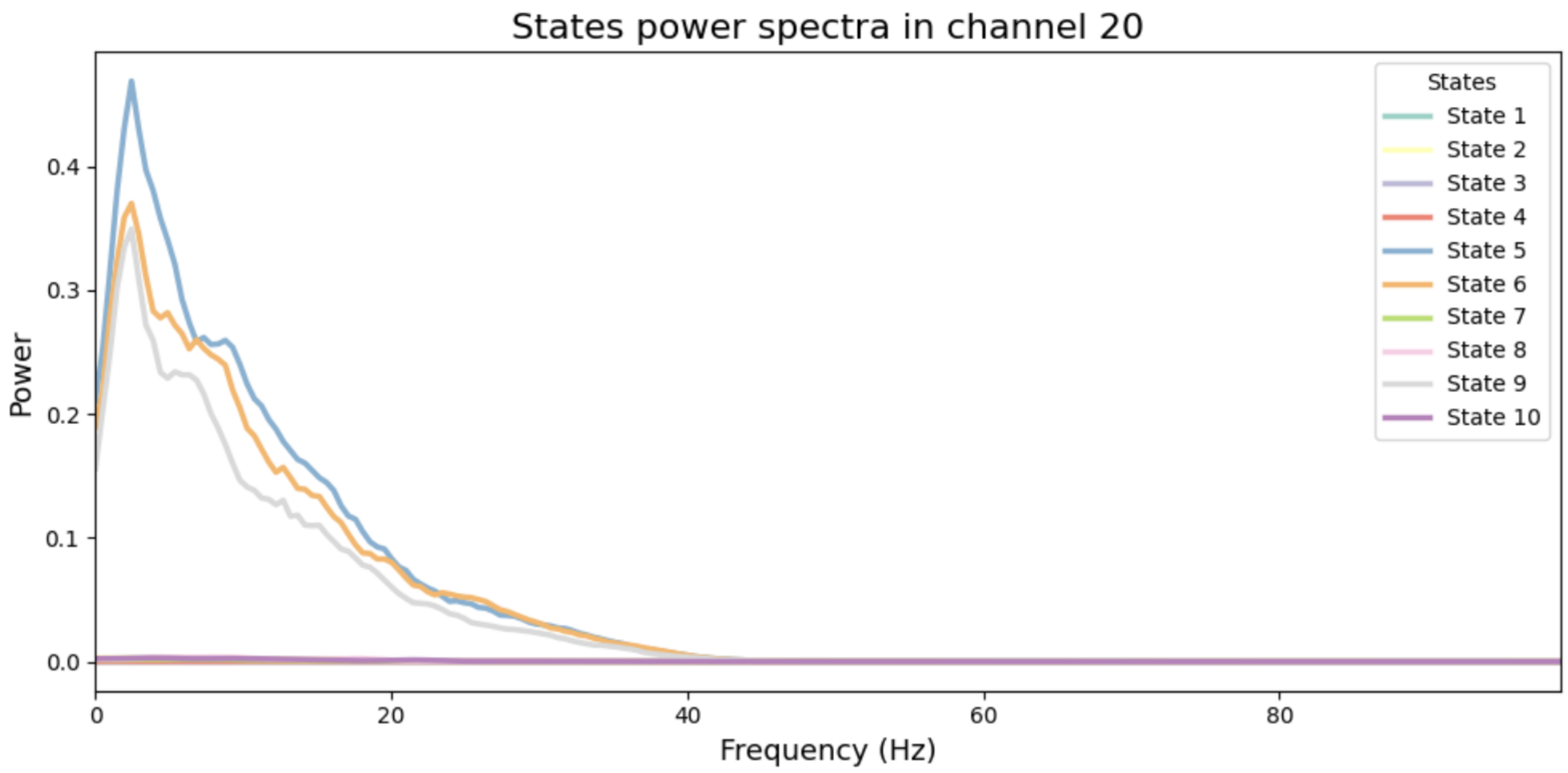
- States Switching Rate (SR) – how often a state switch occurs





Spectral characteristics

Computing the power spectrum and cross-channel coherence of the states

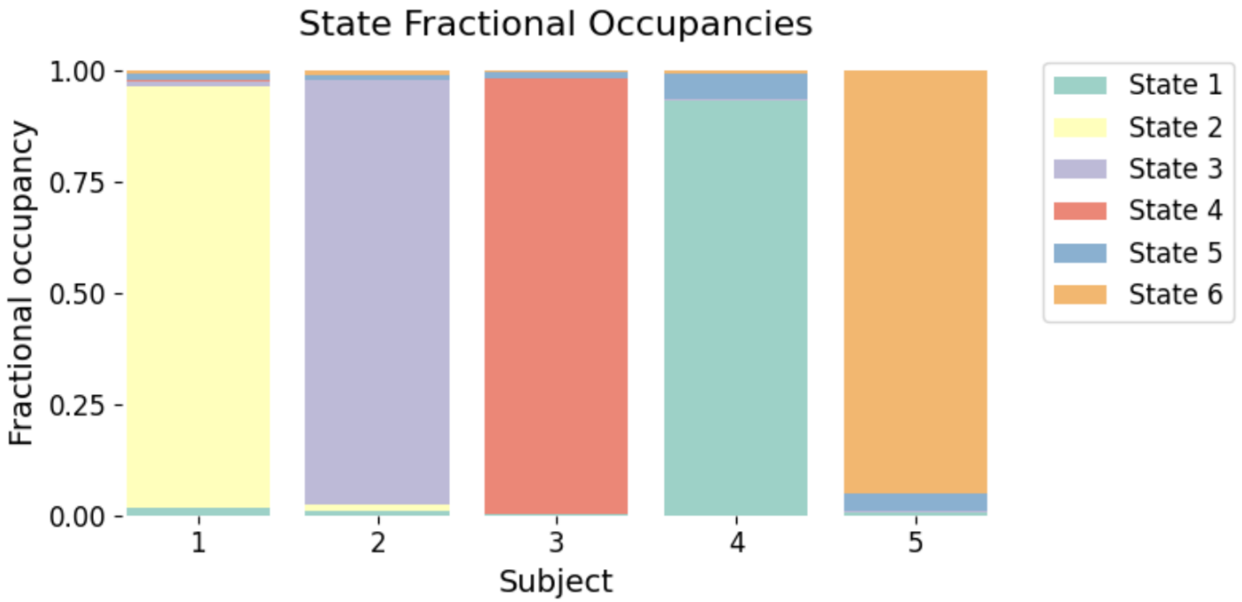


Sanity checks and visualisations

!!! Inspect your HMM after training

Useful things to plot:

- Some examples of state time courses (vpath) with signal
- per-subject fractional occupancy, states lifetime and switching rate
- overall transition probability of the states



See Ahrends et al., 2022, NeuroImage for considerations on data and HMM parameters for good estimations

Dealing with big data: stochastic learning

HMM can deal with large amount of data, performing *stochastic training* (Vidaurre et al., 2018, NeuroImage)

When to use stochastic training?

- Too big dataset to be loaded in memory
- Too long training time for the complete dataset

How does it work?

- The HMM loads randomly selected chunks of data
- Learns recursively from the data that handles step by step

GLHMM paper & python toolbox

The Gaussian Linear Hidden Markov Model (GLHMM) is a technique designed to characterise brain dynamics across different modalities (*Vidaurre et al., 2025, Imaging Neuroscience*). With the paper, we introduce a **python toolbox.**

The toolbox implements various HMMs and has options for:

- data preprocessing,
- analysis and visualization of the HMM outputs,
- prediction and statistical testing

Read the paper at https://doi.org/10.1162/imag_a_00460
Find the toolbox at https://github.com/vidaurre/glhmm
Find the documentation at: https://glhmm.readthedocs.io/en/latest/index.html

The GLHMM toolbox: tutorials

The GLHMM toolbox includes **tutorials** sketching typical neuroscience projects, for example:

- Hypothetical fMRI study, with decoding HMM and prediction of phenotypes
- Hypothetical MEG study, with TDE-HMM and states spectral analysis
- Various types of statistical testing on HMM outputs
- … and more!

Find the tutorials at: https://github.com/vidaurre/glhmm/tree/main/docs/notebooks

The remaining time

- Go through the first tutorial at your own pace

  Tutorial's content:

  - Download example data
  - Basic preprocessing of the data
  - Instantiate and train appropriate HMM
  - Plot summary metrics and basic sanity checks
  - Spectral analysis, plot states power spectra
- Ask questions
- Get Help