Statistical machine learning methods for neuroscientific data: Analysing brain activity, psychopathology and cognitive function in adolescents

Final Seminar
Owen Forbes
May 5 2023

Distinguished Professor Kerrie Mengersen (Principal Supervisor)
Dr Paul Wu (Associate Supervisor)
Dr Edgar Santos Fernandez (Associate Supervisor)

Queensland University of Technology

Mental Health in Adolescence

Mental Health in Adolescence

- Mental health issues are a significant and growing problem in youth populations (Sawyer et al., 2018; McGorry & Mei, 2018)

- Many mental health problems & disorders are thought to emerge in adolescence; period of rapid brain development and change (Paus et al., 2008; McGorry et al., 2011)

- Mental health risk in childhood & adolescence is understood to be **pluripotential**
  - Non-specific risk overlaps across a number of clinical diagnostic categories (Raballo & Poletti, 2020; Hartmann et al., 2019)

👉 We need **improved data-driven statistical modeling** to better understand the **complex relationships between neurophysiology and mental health in adolescence** 🧠

Case-control and Data-driven Approaches

Case-control studies

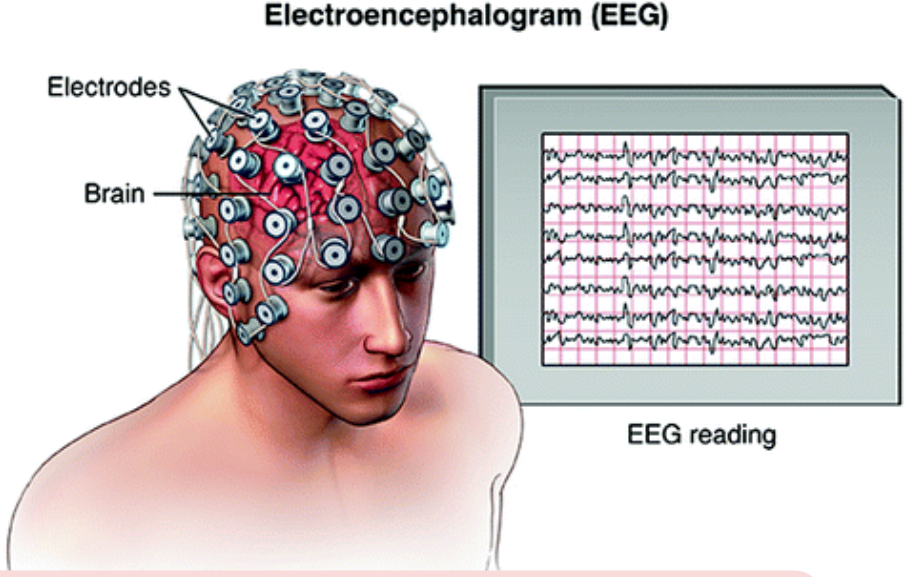
- Identifies differences between diagnosed and undiagnosed groups
- Limited to find information within diagnostic categories based on clinical judgement
- Limited generalisability / replicability outside of sample (Latzman & DeYoung, 2020)

Data-driven empirical methods

- Identifies subgroups of symptoms or individuals based on patterns in data, rather than diagnostic categories / clinician judgement
- Prioritises population-based cohort studies (Latzman & DeYoung, 2020)
- Growing emphasis on data-driven electroencephalography (EEG) research to find new discoveries out of scope from traditional diagnostic categories (Loo et al., 2016; Keizer, 2019)

Electroencephalography (EEG) Data

- Records electrical brain activity using electrodes on the scalp
- Signals represent cumulative activity of rhythmic firing across networks of many neurons



Strengths

- High temporal resolution for measuring oscillations & dynamic brain states (Burle et al., 2015)
- Low cost, widely available (Liedorp et al., 2009)
- Non-invasive, low participant burden

Challenges

- Low signal-to-noise ratio (Wu et al., 2015)
  - Eye and muscle movements
  - Background electrical noise
  - Electrode contact quality
- Non-stationary (Sun and Zhou, 2014)
  - Significant variation in underlying processes
- High dimensionality
  - Large number of channels & features

Growing need for increased large scale research studies collecting data on brain, psychopathology and cognitive function

Longitudinal Adolescent Brain Study (LABS)

- Longitudinal cohort study of youth in the Sunshine Coast
- Follows participants from age 12 - 17
- Recruitment goal: 500 participants

Thesis data scope:

- 59 individuals aged 12 years (T1)
- Resting state EEG, psychological distress, sleep quality, cognitive function
- Chapters 3 & 4

Healthy Brain Network (HBN)

- Large cross-sectional study in New York region
- Participants aged 5-21
- Goal: Biobank of 10,000 participants

Thesis data scope:

- 503 participants aged 9-15
- Pre-processed resting state EEG, psychopathology and cognitive function
- Chapters 5 & 6

Aims & Objectives

Applied Motivation

Develop novel statistical methods to identify data-driven EEG phenotypes in adolescents

Investigate relationships of phenotypes to psychopathology and cognitive function, with potential applications in personalised healthcare, risk prediction and early intervention

(1) Develop an end-to-end pipeline for EEG-based clusters in adolescents **(Ch 3)**

(1a) Using LABS data, examine associations with health measures. **(Ch 3 & 6)**

(2) Build Bayesian Model Averaging framework to combine clusters with probabilistic allocation & model-based uncertainty **(Ch 4)**

(2a) Assess clinical & applied utility of BMA features with case study using LABS EEG clusters **(Ch 4 & 6)**

(3) Develop nested functional analysis framework for time series data including functional characteristics in latent states **(Ch 5)**

(3a) Investigate associations between functional analysis outputs & health measures using HBN data **(Ch 5 & 6)**

(4) Replicate and compare feature extraction and clustering pipelines (1) and (3) across LABS and HBN data **(Ch 6)**

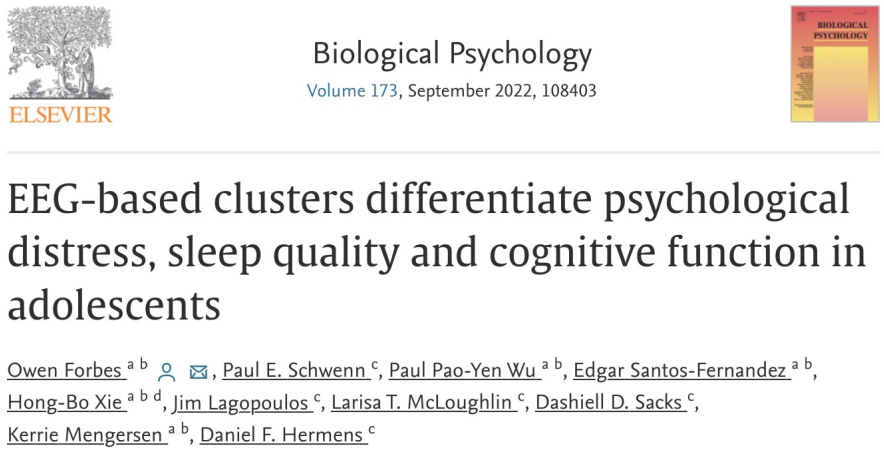
(4a) Explore cross-method, cross-sample replication of robust EEG phenotypes in adolescents using methods in (1) and (3) **(Ch 6)**

Chapter 3

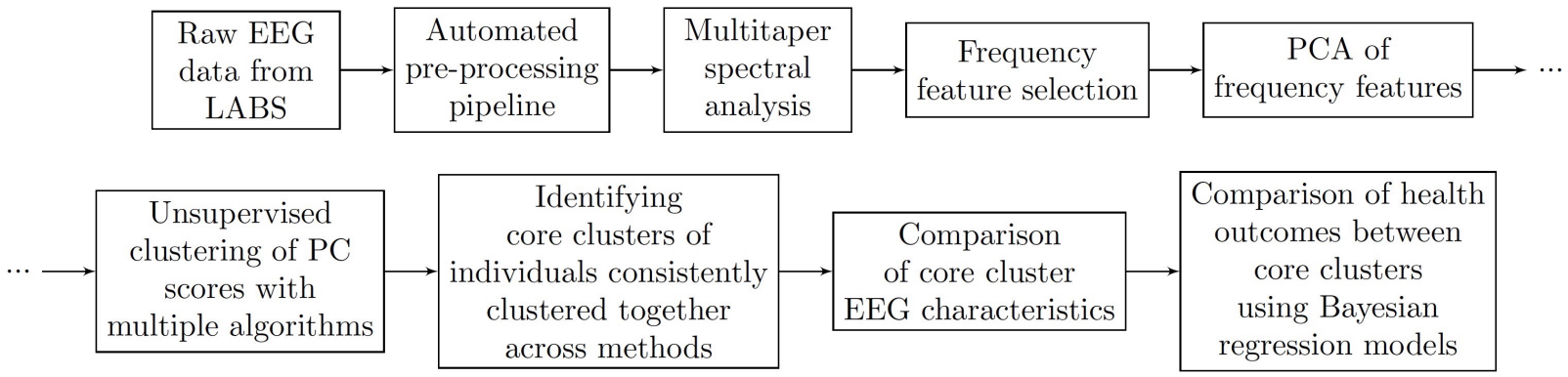
EEG-based clusters differentiate psychological distress, sleep quality and cognitive function in adolescents

Chapter 3: Novel Contributions

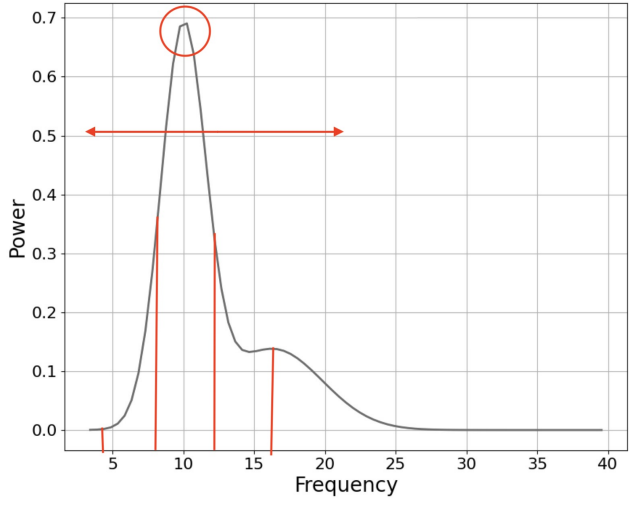
- First work to investigate phenotypes using unsupervised clustering of EEG data from a population sample of adolescents
- New insights about the presence of resting state EEG phenotypes, and clear associations with psychological distress, sleep quality and cognitive function
- Novel statistical pipeline implemented in **freely available open source software** (R and MATLAB) for clinicians and researchers to apply in their own work
- SAGE Award for Best HDR Student Publication (Faculty of Science) 2022
- Published in *Biological Psychology*



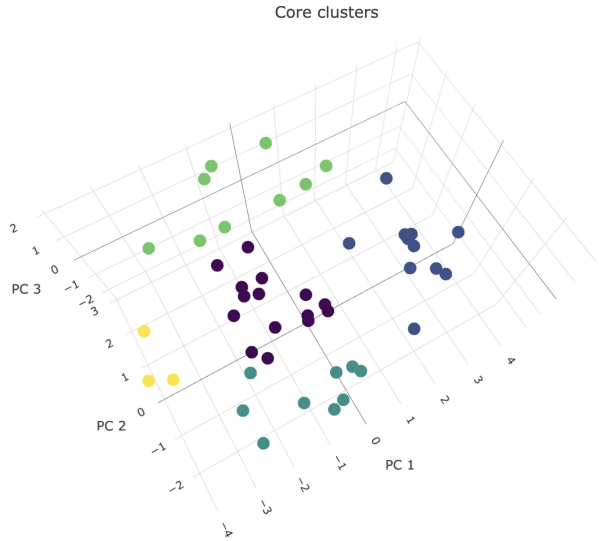
Chapter 3: Methods

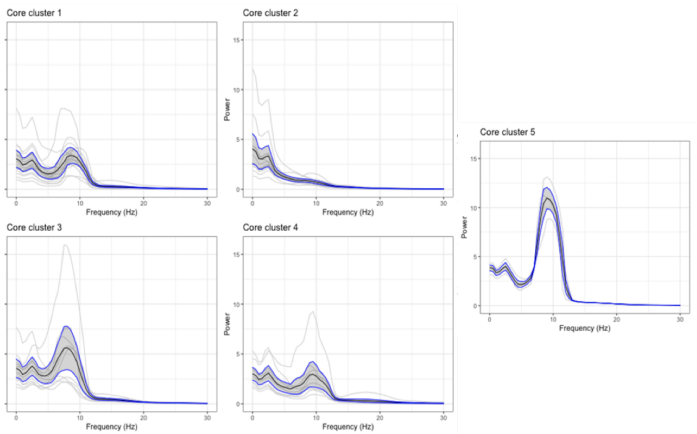


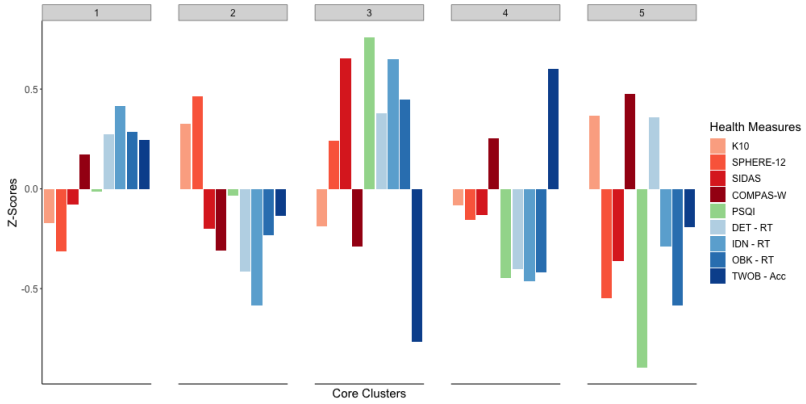
- Time-averaged summary features selected *a priori*
- Dimension Reduction with PCA
- 3 clustering algorithms
- Overlapping 'core' clusters

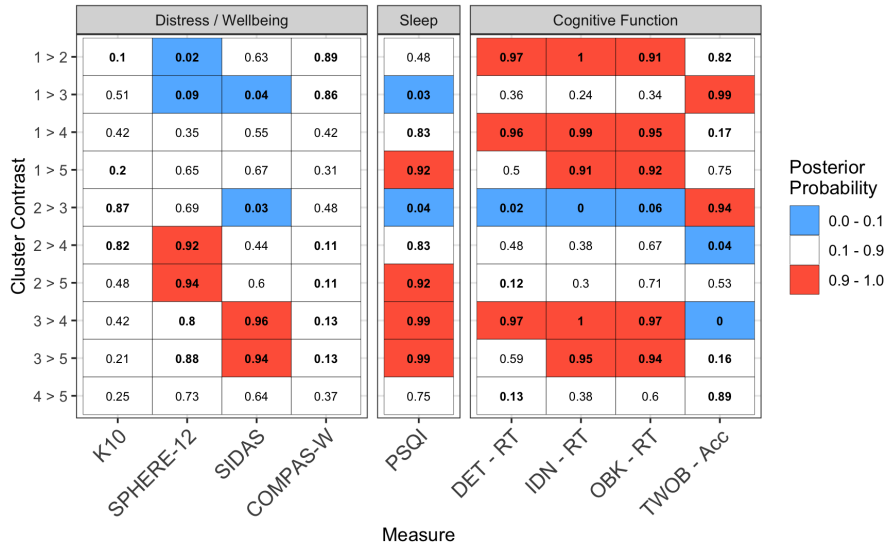


Chapter 3: Results









Exciting preliminary results

- Clusters have distinct EEG characteristics and substantial differences in mental health and cognition
- Indicate resting state EEG phenotypes in adolescents, potential *risk* and *protective* biomarkers

Can we develop a better way to combine multiple clustering algorithms, incorporating model-based uncertainty? **(Chapter 4)**

Can we use more sophisticated methods to characterise EEG, include temporal information, and avoid coarse canonical summary features? **(Chapter 5)**

How robust are the phenotypes associated with these clusters? Do we find similar results using different samples, and different methods? **(Chapter 6)**

Chapter 4

clusterBMA: Bayesian Model Averaging for Clustering

Chapter 4: Novel Contributions

- First implementation of Bayesian model averaging to combine solutions across multiple clustering algorithms
- Offers unique features for ensemble clustering including **quantification of model-based uncertainty** and probabilistic allocation to combined clusters
- Substantially better accuracy in benchmarking performance compared to other ensemble algorithms, particularly for high-dimensional data with low separation between clusters
- Implemented in open source & freely available **R package** for data scientists & practitioners to apply across various fields
- Preprint has been cited
- Trending on DeepAI
- Used by researchers at UWA and US government agency

Under review in *PLOS ONE*

Which clustering algorithm?

k-means, K = 5

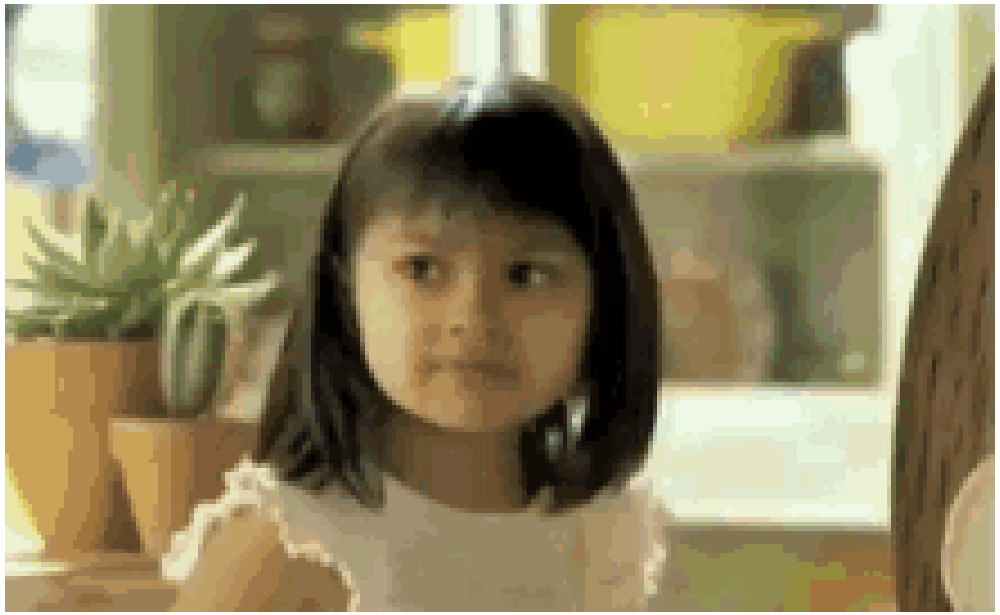
Hierarchical clustering (Ward), K = 5

Gaussian mixture model, K = 5

Inconsistent clustering across algorithms

Which clustering algorithm?

- Different algorithms will emphasise different aspects of clustering structure
- Choosing one 'best' model often arbitrary, unclear choice
  - ➡️ **Inference not calibrated for model-based uncertainty**
- Locking into one method loses insights offered by other methods about plausible clustering structure



BMA offers a nice framework for combining clustering solutions

- simple
- flexible
- intuitive

Combining Clustering Results with Bayesian Model Averaging

- Limited development for Clustering
  - Finite mixture models (Russell et al., 2015)
  - Naive Bayes classifiers (Santafe & Lozano, 2006)
  - Lacks implementation across multiple clustering algorithms

Advantages

- 🆒 Weighted averaging of results incorporating model quality / goodness of fit
- 💯 Intuitive framework for **probabilistic** inferences combining results from **different clustering algorithms**
- 😮 Each input solution can have a different number of clusters K
- 😎 **Quantify model-based uncertainty and enable more robust inferences** calibrated accordingly

Bayesian Model Averaging: Basics

$$P(\Delta|Y) = \sum_{l=1}^{L}(\Delta|Y, M_l)P(M_l|Y)$$

$$P(M_l|Y) = \frac{P(Y|M_l)P(M_l)}{\sum_{l=1}^{L} P(Y|M_l)P(M_l)}$$

BMA for Mixture Models - BIC weighting

- $P(Y|M_l)$ typically involves a difficult/intractable integral and is often approximated for many applications (Fragoso et al., 2018)
- **Russell et al. (2015)** weight results from multiple GMMs according to BIC

$$P(M_l|Y) \approx \frac{exp(\frac{1}{2}BIC_l)}{\sum_{l=1}^{L} exp(\frac{1}{2}BIC_l)}$$

- BIC definition for GMM

$$BIC_l = 2\log(\mathcal{L}) - \kappa_m \log(N)$$

- GMM likelihood

$$\mathcal{L}(\Theta) = \sum_{n=1}^{N}\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)$$

- Multivariate Gaussian density

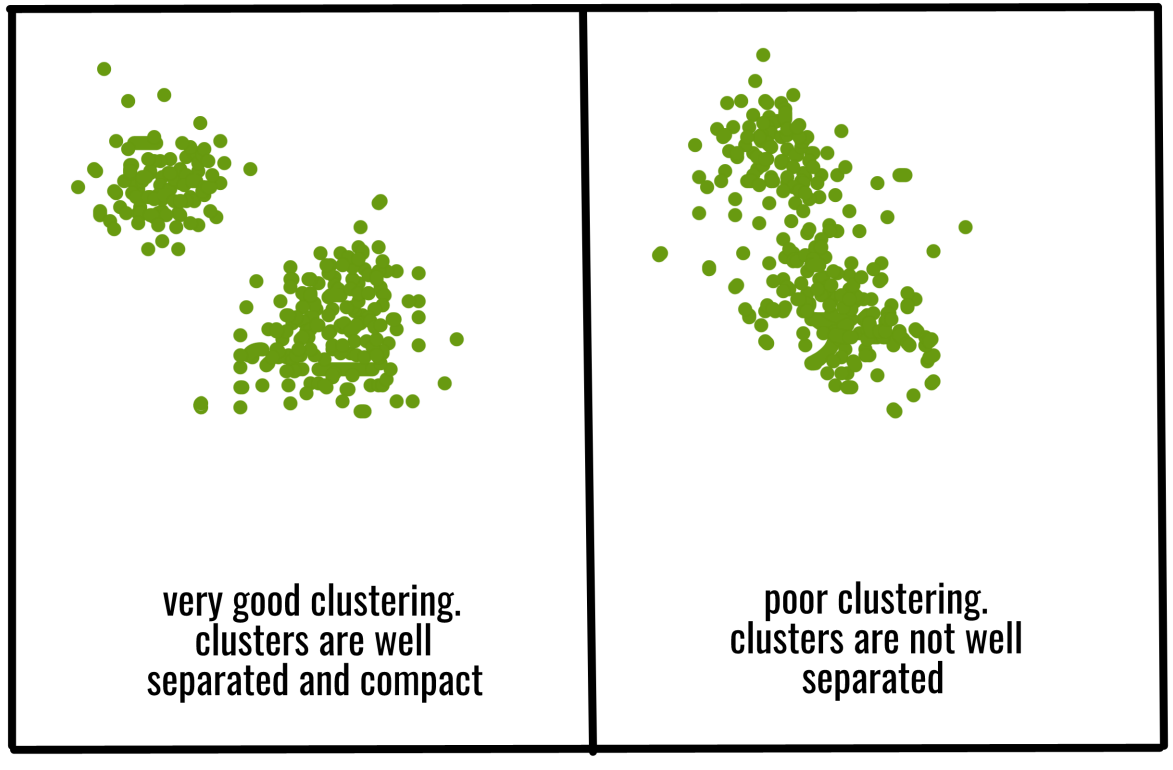
$$\mathcal{N}(x|\mu, \Sigma) = \frac{\exp\{-\frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)\}}{|\Sigma|^{\frac{1}{2}}(2\pi)^{\frac{D}{2}}}$$

Aside: Cluster internal validation indices

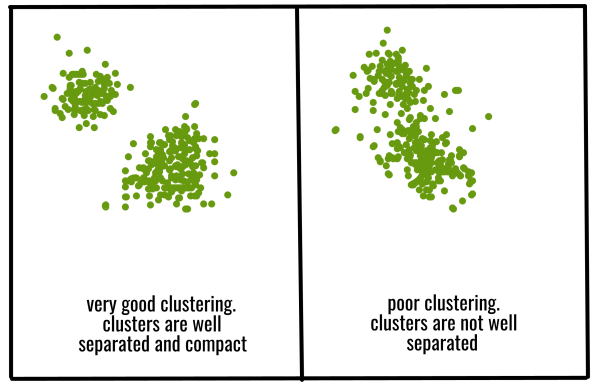
- Often used as a proxy for model quality in clustering
- Choose between candidate models with different numbers of clusters k
- Interpreted similarly to marginal likelihood/model evidence
- Typically measure compactness and/or separation of clusters

Compared to BIC…

- Agnostic to clustering algorithm
- Typically do not require likelihood term



New proposed weighting / approximation for posterior model probability



BIC for GMM driven by **Multivariate Gaussian density:**

$$\mathcal{N}(x|\mu,\Sigma) = \frac{\exp\{-\frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)\}}{|\Sigma|^{\frac{1}{2}}(2\pi)^{\frac{D}{2}}}$$

Calinski-Harabasz Index

- Ratio of separation to compactness (maximise)

$$CH = \frac{\sum_i n_i d^2(c_i,c)/(NC-1)}{\sum_i \sum_{x\in C_i} d^2(x,c_i)/(n-NC)}$$

S_Dbw Index

- Sum of intra-cluster scattering and inter-cluster density (minimise)

$$S\_Dbw = Scat(K) + Dens\_bw(K)$$

- CH and S_Dbw indices are both well validated in the context of comparing model quality across different clustering algorithms

New proposed weighting / approximation for posterior model probability

CH and S_Dbw indices

- conceptually and mathematically similar to BIC
- Unlike BIC, can be calculated + directly compared across different clustering algorithms

New proposed weight:

$$P(Y|\mathcal{M}_m) \approx \hat{\mathcal{W}}_m := \begin{cases} \frac{\mathcal{W}_m}{\sum_{m'=1}^{M} \mathcal{W}_{m'}} & \text{if } \mathcal{W}_m \text{ is to be maximised} \\ \\ \frac{\frac{1}{\mathcal{W}_m}}{\sum_{m'=1}^{M} \frac{1}{\mathcal{W}_{m'}}} & \text{if } \mathcal{W}_m \text{ is to be minimised,} \end{cases}$$

Approximate posterior model probability for weighted averaging:

$$P(\mathcal{M}_m|Y) \approx \frac{\hat{\mathcal{W}}_m\left(\frac{1}{M}\right)}{\sum_{m'=1}^{M} \hat{\mathcal{W}}_{m'}\left(\frac{1}{M}\right)} = \hat{\mathcal{W}}_m.$$

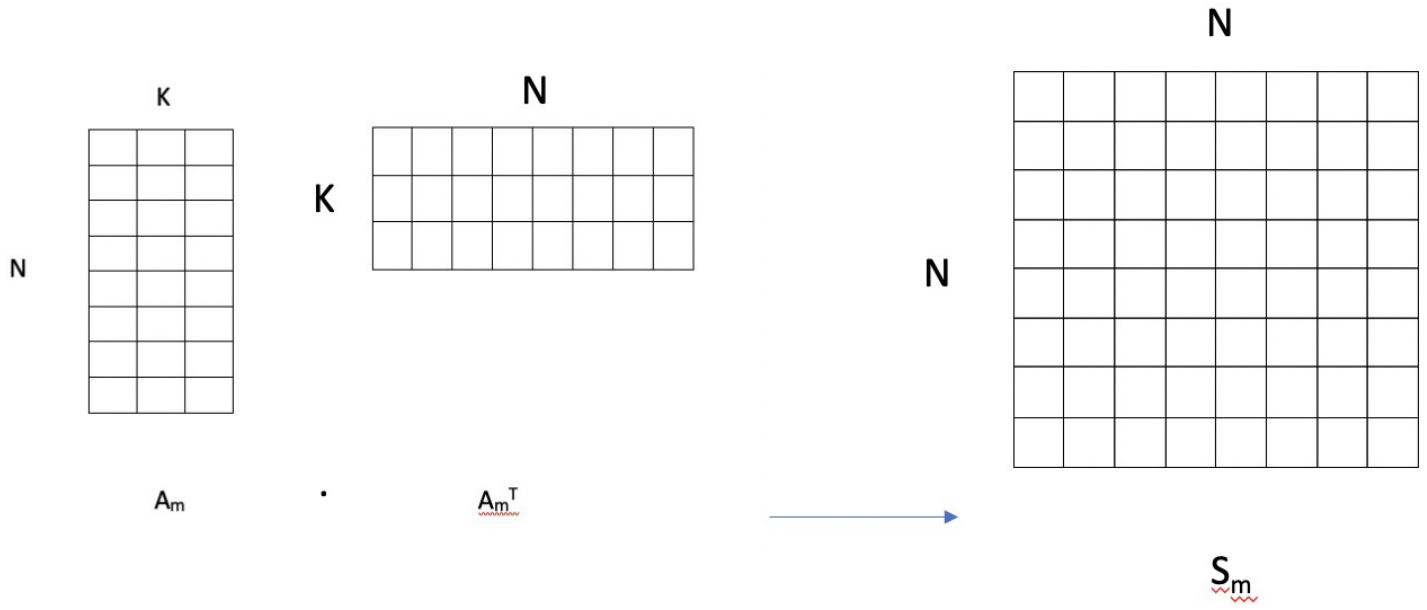
Consistent quantity Δ - Similarity matrices

Previous work (Russell et al., 2015) has used pairwise similarity matrices as Δ for each model

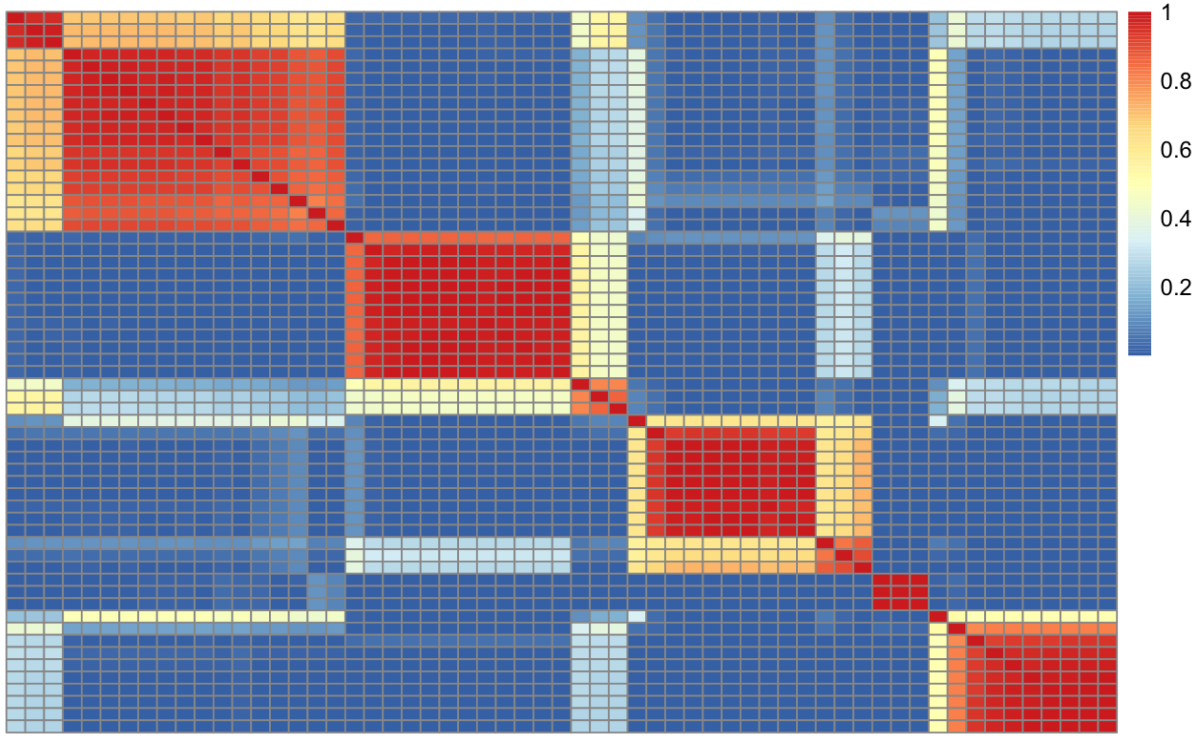
- To get similarity matrix, multiply allocation matrix by its transpose:

$$S_m = A_m A_m^T$$

- **invariant to number and labelling of clusters across solutions**



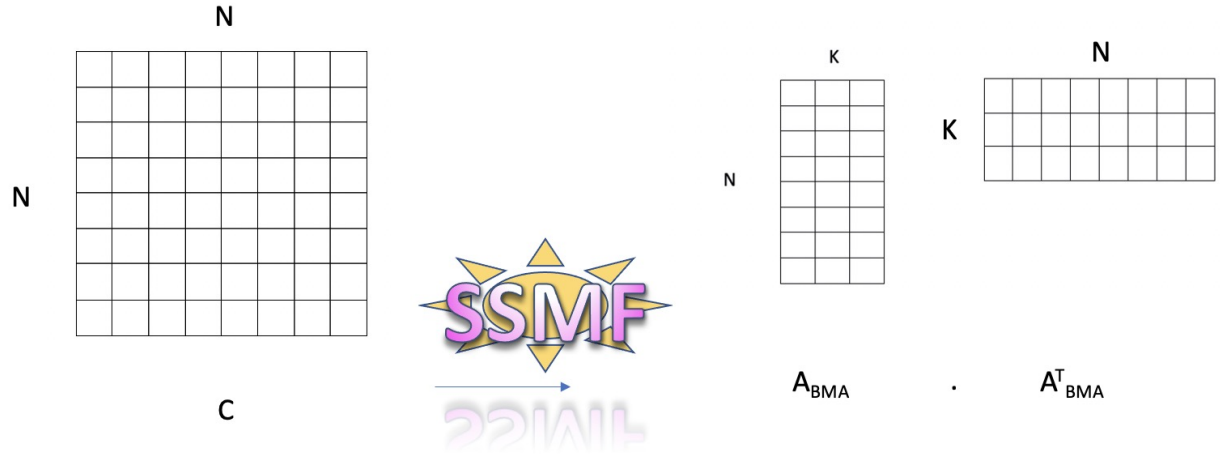
Consensus matrix



$$C = \sum_{m=1}^{M} \hat{\mathcal{W}}_m S_m.$$

Consensus matrix ➡️ Matrix factorisation ➡️ Cluster allocation probabilities

- Symmetric Simplex Matrix Factorisation (SSMF; Duan, 2020) to get $N \times K$ allocation matrix A_m from $N \times N$ consensus matrix C
- Generates probabilistic cluster allocations from pairwise probabilities

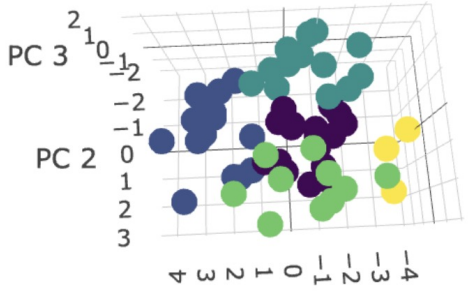


- Includes L2 regularisation step to reduce overfitting & redundant clusters

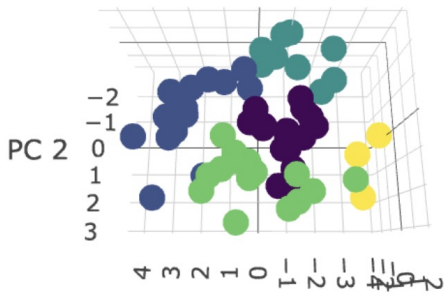
Case study: Clustering adolescents based on resting state EEG recordings

Model results

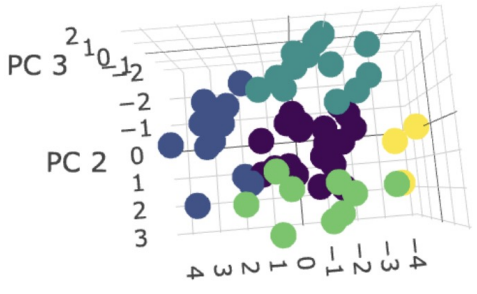
k-means, K = 5



Hierarchical clustering (Ward), K = 5

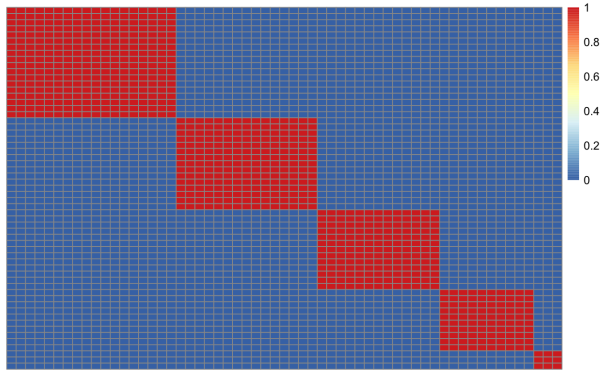


Gaussian mixture model, K = 5

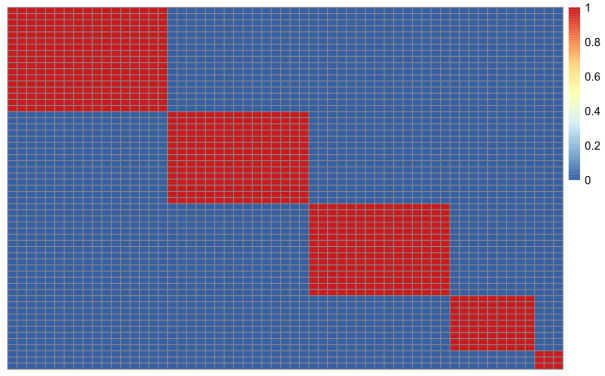


Model results ➡️ Similarity matrices

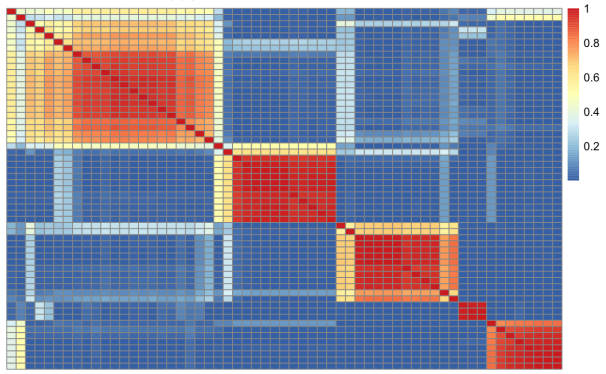
k-means $\hat{\mathcal{W}}_m = 0.36$



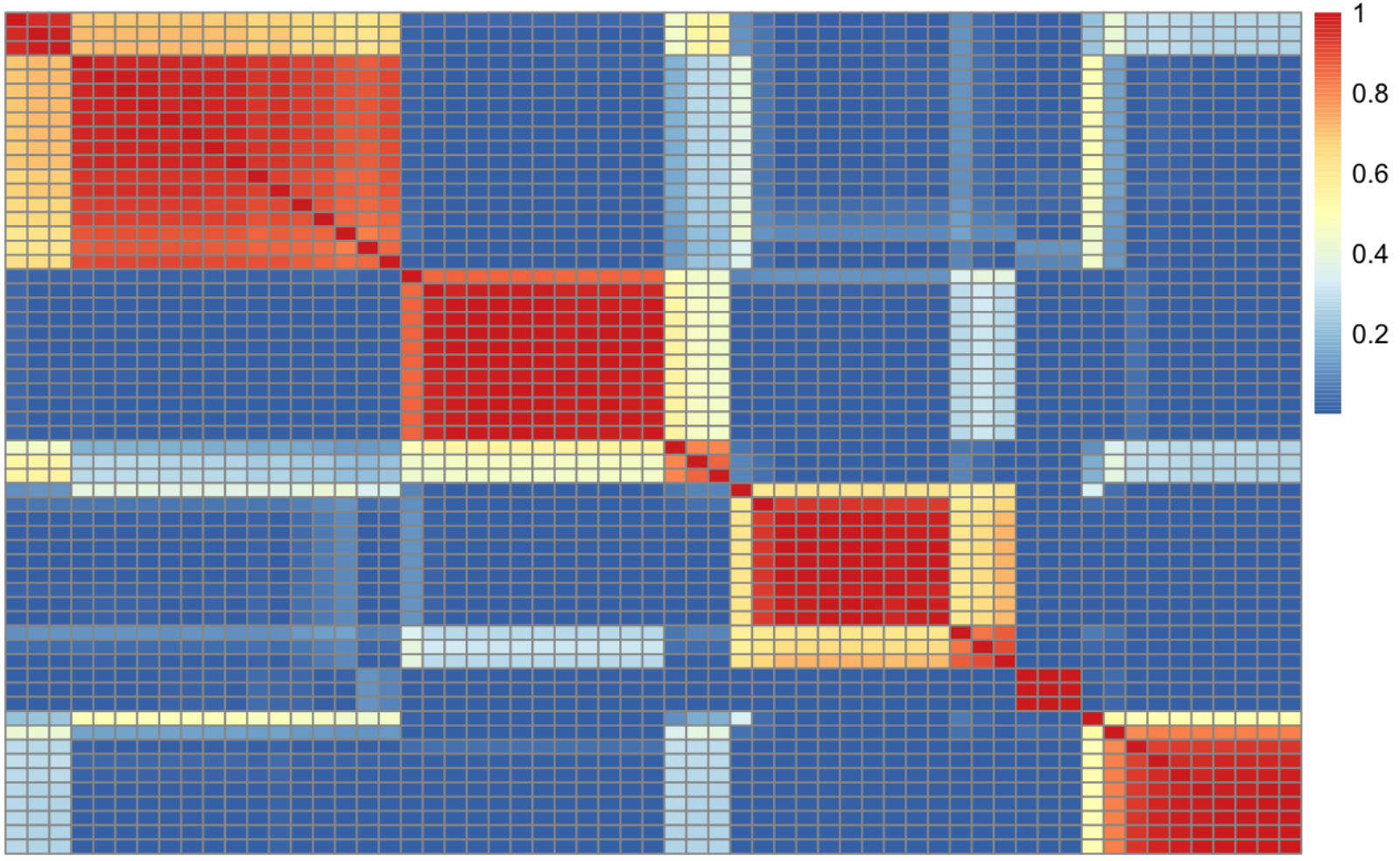
HC $\hat{\mathcal{W}}_m = 0.27$



GMM $\hat{\mathcal{W}}_m = 0.37$



Model results ➡️ Similarity matrices ➡️ Consensus matrix



BMA Clusters with allocation uncertainty

- **Uncertainty can be propagated forward for further analysis in a Bayesian framework**

Benchmarking study: Feature comparison with other ensemble approaches

Method	Combine solutions with different numbers of clusters	Combine solutions from different algorithms	Weight each input solution by model quality	Combine allocation probabilities from 'soft' and 'hard' clustering algorithms [1]	Probabilistic allocation to averaged clusters	Measure model-based uncertainty in allocations to averaged clusters
BMA for GMM [2]	✓	×	✓	×	×	×
CSPA	✓	✓	*	×	×	×
LCE	✓	✓	*	×	×	×
K modes	✓	✓	*	×	×	×
Majority voting	✓	✓	*	×	×	×
clusterBMA	✓	✓	✓	✓	✓	✓

Table 1. Feature comparison between *clusterBMA* and five other ensemble clustering methods.

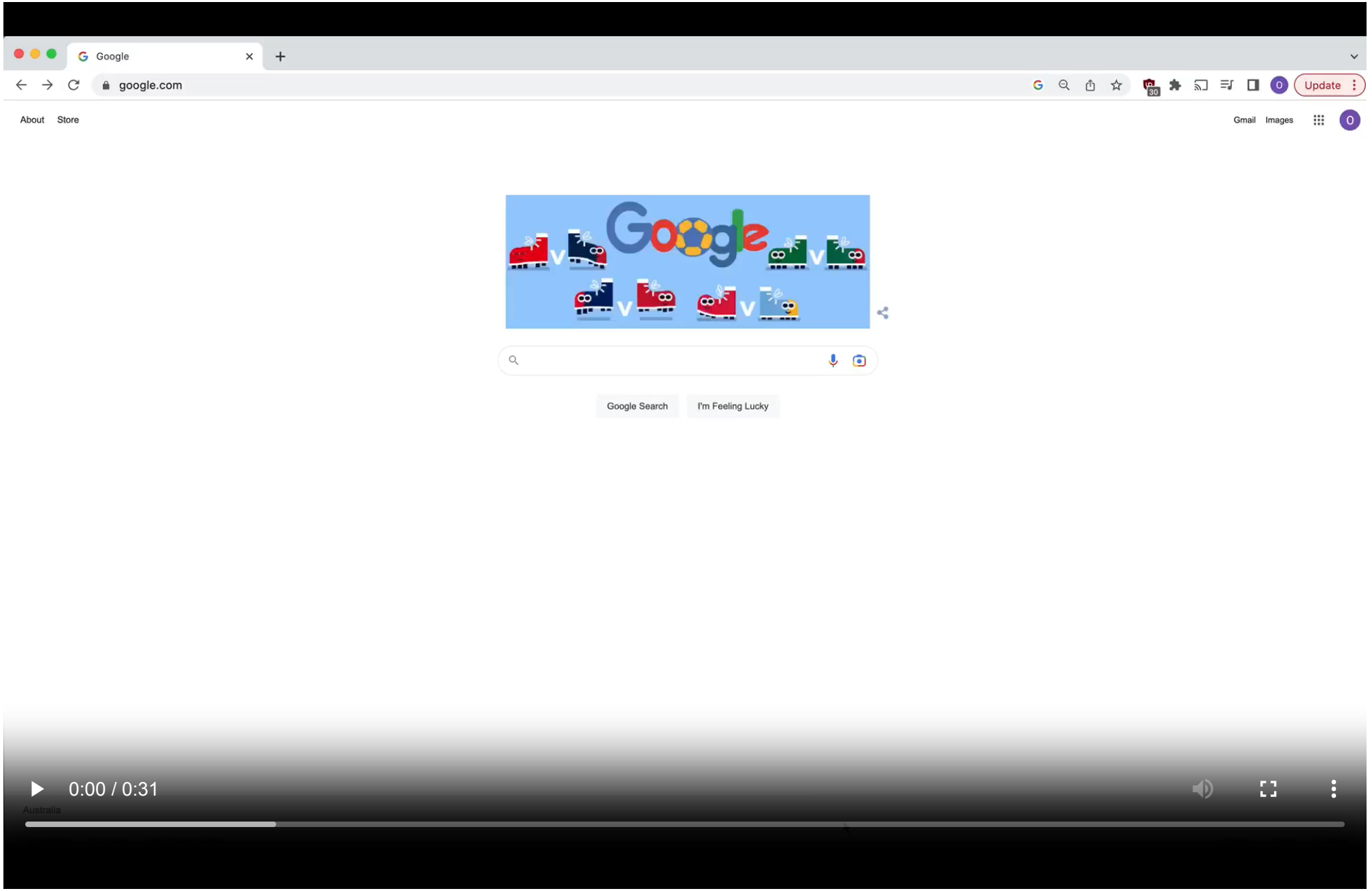
Benchmarking study: Accuracy comparison

- 9 clustering algorithms
- Simulated data with differing dimensions (2, 10, 50) and separation (high, medium, low)
- 10 simulated datasets for each combination of dimensions X separation

Cluster separation	Method	2 Dimensions	10 Dimensions	50 Dimensions
High	CSPA	0.93 (0.03)	0.95 (0.01)	0.94 (0.02)
	LCE	0.89 (0.13)	0.91 (0.13)	0.74 (0.22)
	K-modes	0.93 (0.03)	0.89 (0.18)	0.93 (0.01)
	Majority voting	0.93 (0.03)	0.89 (0.15)	0.38 (0.24)
	clusterBMA	0.93 (0.03)	0.95 (0.01)	0.94 (0.02)
	clusterBMA – high certainty	0.97 (0.01)	0.98 (0.01)	0.97 (0.01)
Medium	CSPA	0.81 (0.05)	0.79 (0.03)	0.69 (0.16)
	LCE	0.81 (0.05)	0.76 (0.09)	0.42 (0.25)
	K-modes	0.81 (0.04)	0.80 (0.02)	0.68 (0.17)
	Majority voting	0.81 (0.04)	0.68 (0.16)	0.11 (0.09)
	clusterBMA	0.81 (0.04)	0.80 (0.02)	0.76 (0.02)
	clusterBMA – high certainty	0.86 (0.04)	0.91 (0.02)	0.86 (0.04)
Low	CSPA	0.63 (0.08)	0.62 (0.05)	0.49 (0.16)
	LCE	0.60 (0.11)	0.61 (0.05)	0.32 (0.17)
	K-modes	0.63 (0.07)	0.58 (0.11)	0.49 (0.18)
	Majority voting	0.62 (0.09)	0.42 (0.15)	0.08 (0.10)
	clusterBMA	0.63 (0.08)	0.61 (0.04)	0.57 (0.13)
	clusterBMA – high certainty	0.70 (0.06)	0.78 (0.04)	0.69 (0.11)

Table 2. Simulation study results - ARI mean (standard deviation) across 10 simulated datasets, comparing clusterBMA with four other ensemble clustering methods.

A quick demo

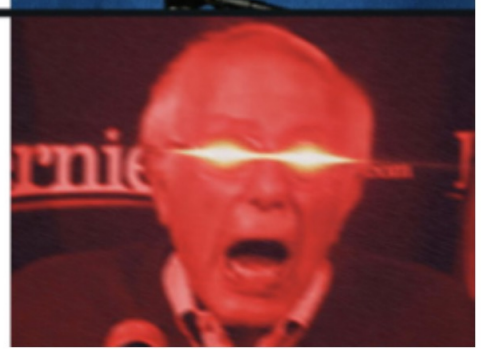

Google
google.com
Update
About
Store
Gmail
Images
Google Search
I'm Feeling Lucky
0:00 / 0:31
Australia

Yet another method for combining clustering solutions...

It's agnostic to the number of clusters and the algorithms used

Intuitive & flexible framework to combine solutions weighted by quality

Calibrate cluster-based inferences for model based uncertainty



Chapter 5

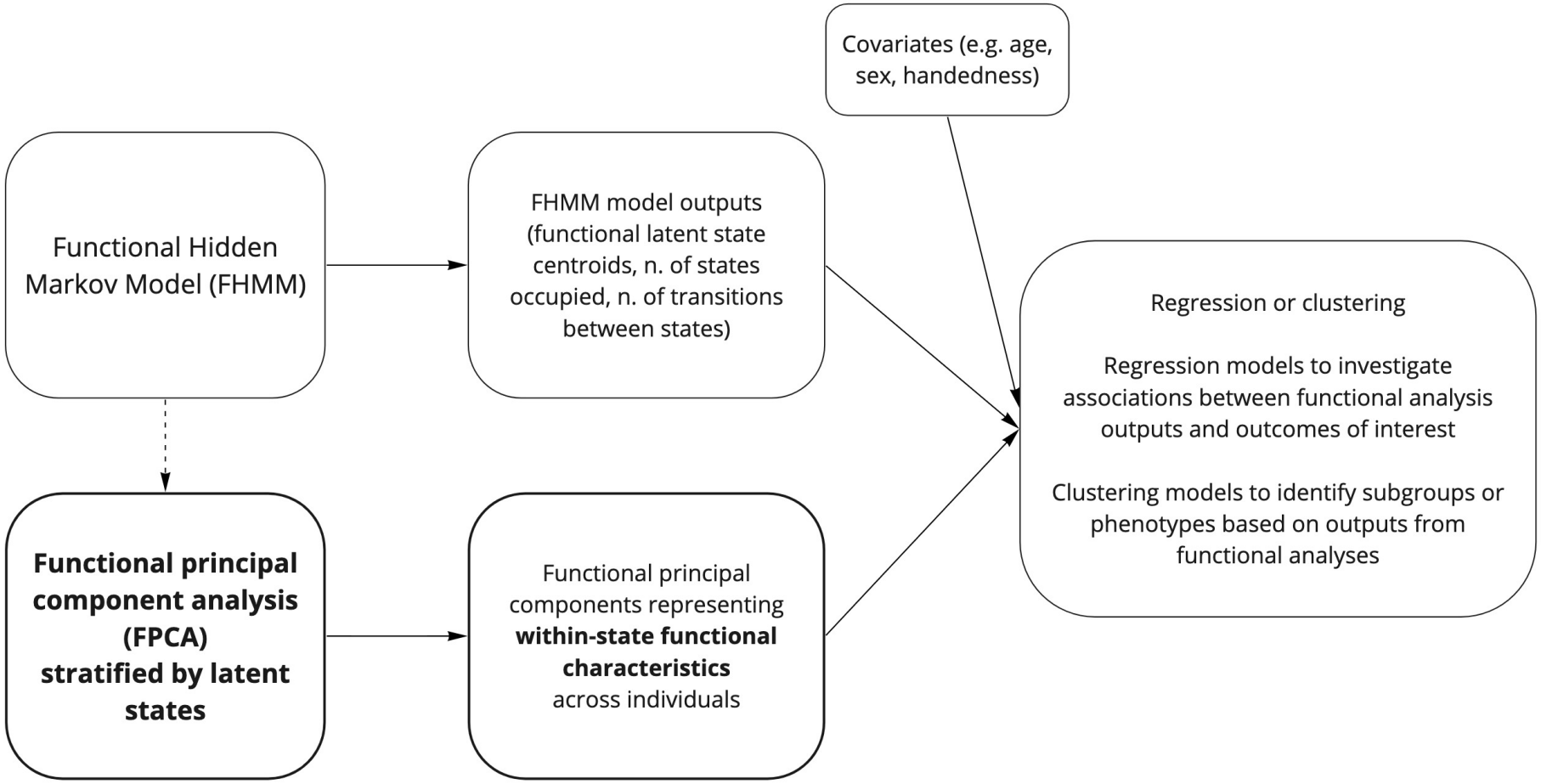
Functional analysis within latent states for neuroscientific data: Linking resting state brain activity, psychopathology and cognitive function in young people

Chapter 5: Novel Contributions

- Novel nested framework for functional data analysis (FDA) of functional time series data
- Integrated analysis of functional latent states using FHMM, and functional characteristics within states using FPCA
- More nuanced insights into variation in resting state EEG, less reliant on flawed assumptions and canonical features
- In HBN data, Bayesian regression models show substantial assocations of temporal dynamics and functional characteristics within states, with measures of psychopathology and cognitive function
- Lays a foundation for more sophisticated phenotype identification including temporal and functional characteristics (Chapter 6)
- Implemented in open source & freely available software in R for practitioners to apply across various fields

In preparation for submission to *Statistical Methods in Medical Research*

Functional analysis within latent states - Overview



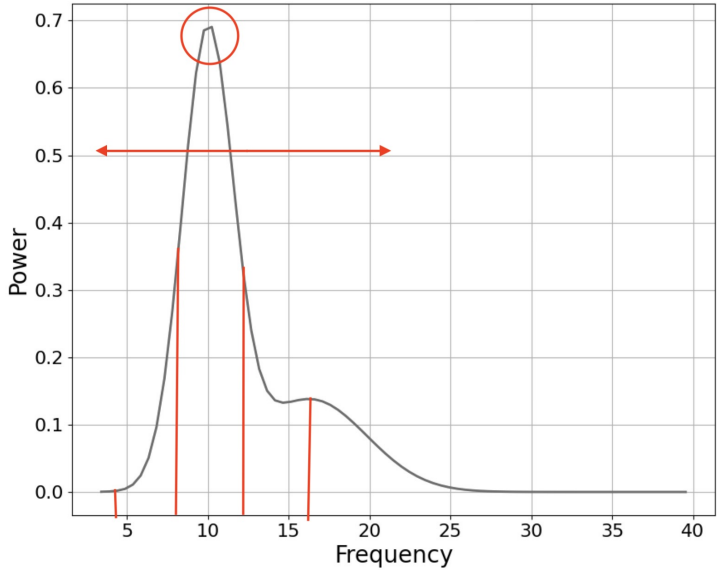
Functional analysis within latent states

Functiona**L** **A**nalysis **W**ithin **L**at**E**nt **S**tate**S**

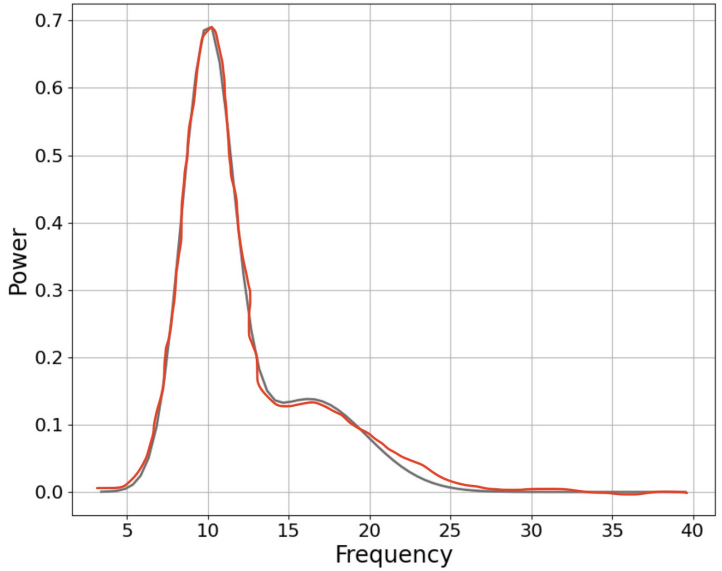
(***flawless***)

Functional Data Analysis (FDA)

- Data represented as functions -- e.g. curves over a continuous functional domain like time or frequency
- Analyse influential characteristics across the whole functional domain of interest
- Copes well with high dimensionality of neuroscience data

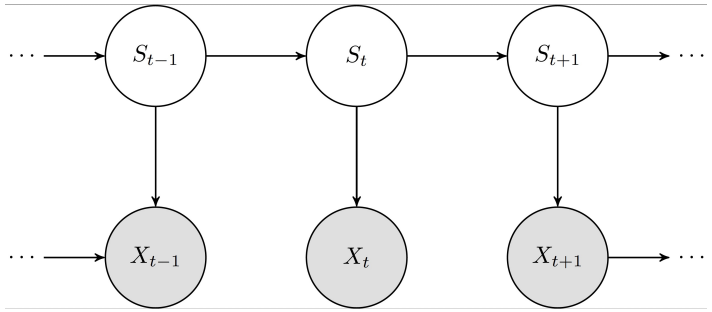


Summary features selected *a priori*



FDA - information across whole curve

Functional Hidden Markov Models



Objective Function

$$\log(\mathcal{L}(\lambda \mid \mathbf{x})) = \sum_{i=1}^{N} \gamma_1(i) \log \nu_i + \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\sum_{k=1}^{K-1} \xi_k(i,j) \right) \log a_{ij} + \sum_{i=1}^{N} \sum_{k=1}^{K} \gamma_k(i) \log b_i \left(\mathbf{x}_k; \boldsymbol{\theta}_i \right).$$

- Similar to regular HMM, but emission function is not a PDF --> function based on L2 distance measure between each curve and mean curve for each state (Martino 2020)

$$b_{\mathbf{X}_k \mid Q_k = s_i} \left(\mathbf{x}_k; \boldsymbol{\mu}_i \right) = h \left(d \left(\mathbf{x}_k, \boldsymbol{\mu}_i \right) \right), \quad i = 1, \ldots, N$$

where $h : \mathbb{R} \to \mathbb{R}$ is a function that transforms the distance into a similarity measure. In particular, the implementation by Martino et al. (2020) uses the function $h(y) = 1/y^2$ and the L^2 distance.

Functional Principal Component Analysis

Captures the essential patterns of variation in functional data in terms of scores on **eigenfunctions** (Ramsay et al., 2009)

FPCA expansion

For a set of functional data $X_p(t)$ where t is the functional domain, the FPCA expension is as follows:

$$X_p(t) = \mu(t) + \sum_{c=1}^{\infty} A_{pc}\phi_c(t),$$

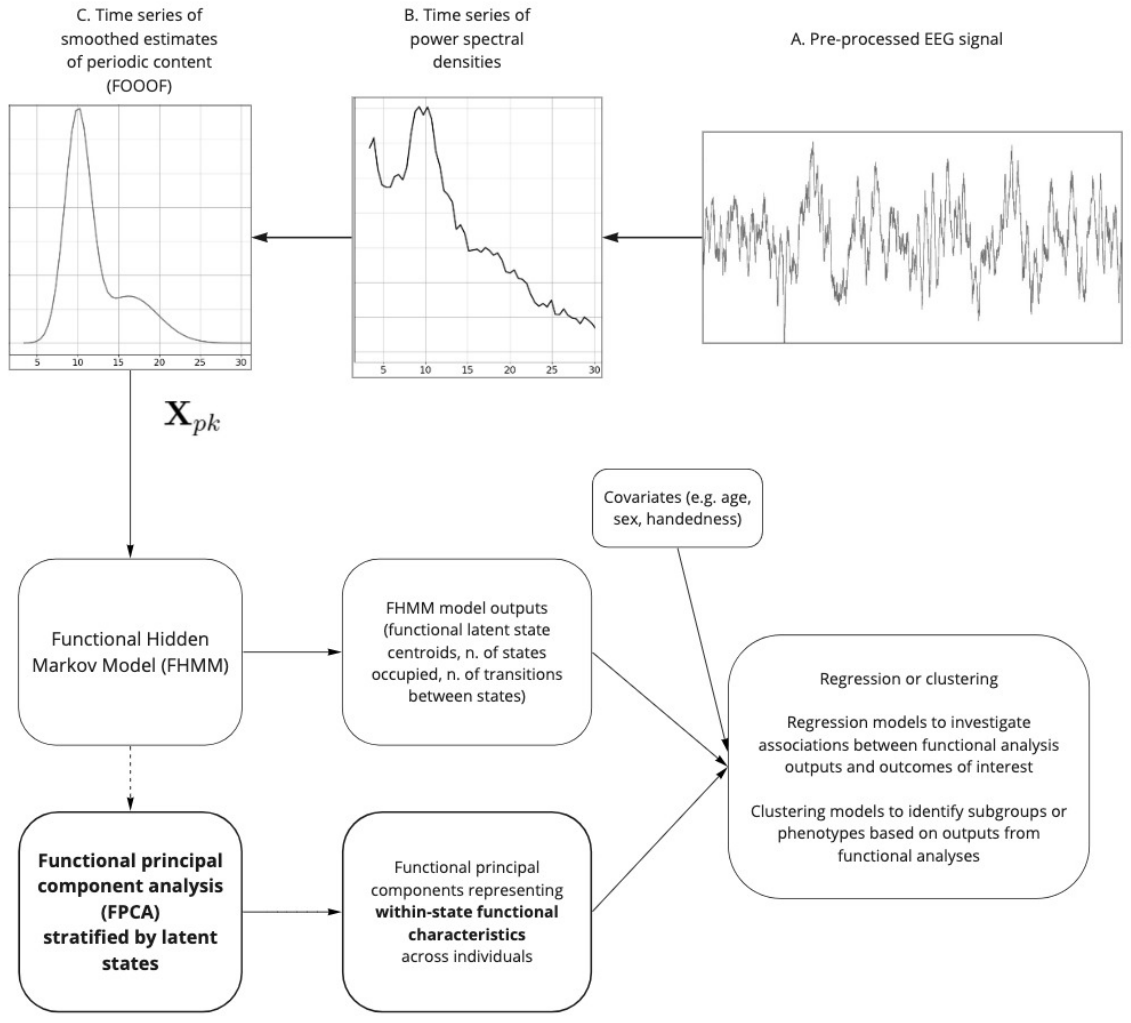
where $\mu(t)$ is the mean function of $X_p(t)$, and $A_{pc} = \int_P (X_p(t) - \mu(t))\,\phi_c(t)\mathrm{d}t$ are the functional principal components of X_p.

Finite n. of functional principal components, `$C$`

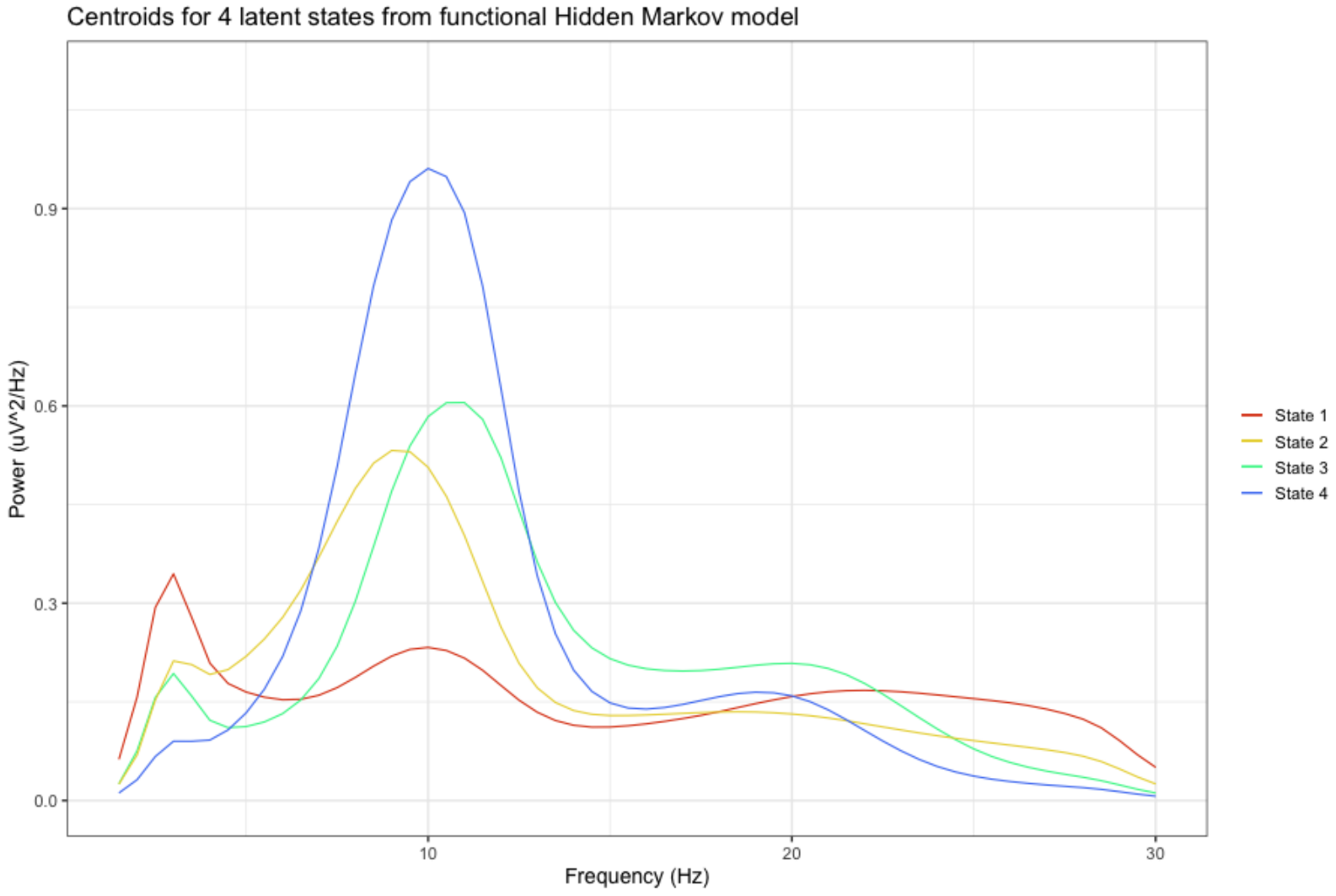
the information contained in X_p is largely contained in the C-dimensional vector of eigenvalues $\mathbf{A}_p = (A_{p1}, \ldots, A_{pC})$ and the approximated processes

$$X_{pC}(t) = \mu(t) + \sum_{c=1}^{C} A_{pc}\phi_c(t).$$

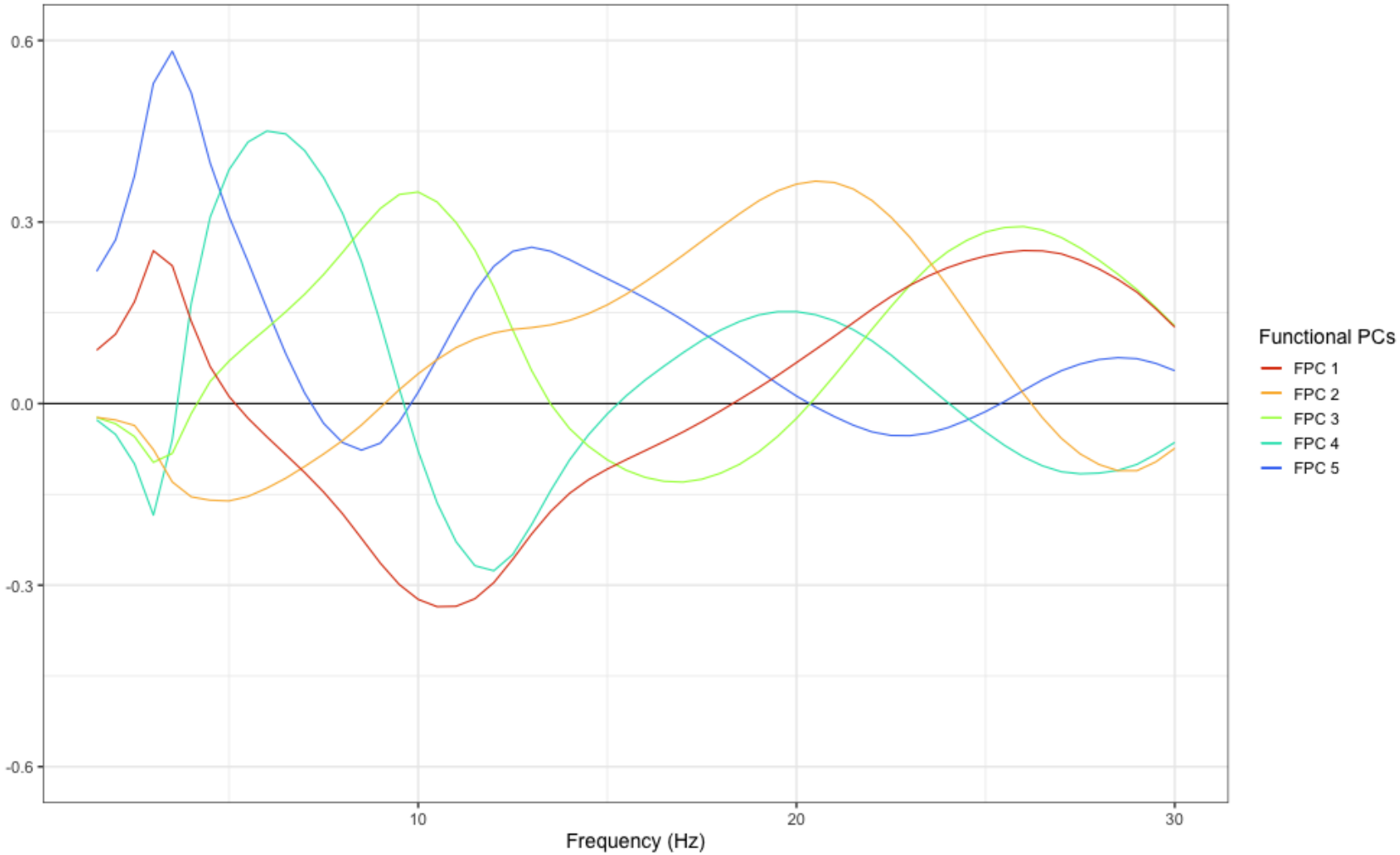
flawless analysis - Overview



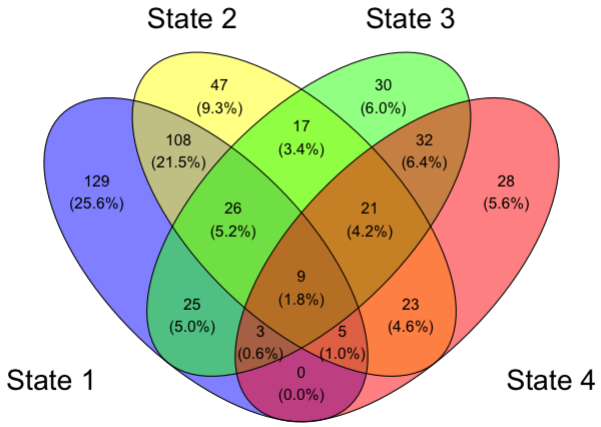
FHMM - Latent State Centroids



Latent State 1 - Functional Principal Components

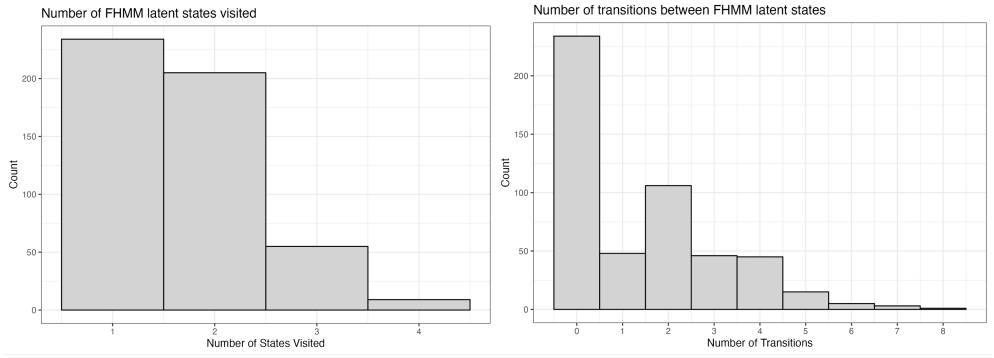


FHMM - Latent State Occupancy



Most common combinations:

- State 1 only (25.6%)
- States 1 and 2 (21.5%)
- State 2 only (9.3%)
- States 3 and 4 (6.4%)
- State 3 only (6.0%)
- State 4 only (5.6%).



- Most visit 1 (n = 234; 46.5%) or 2 (n = 205; 40.8%) latent states
- A majority (n = 479; 95.2%) make between 0 and 4 transitions between states over the recording period.

- Rich & nuanced empirical insights: Functional latent states; Temporal dynamics; Functional variation within states
- Bayesian Regression models: Substantial associations between FHMM & FPCA-within-state outputs, psychopathology and cognitive function

Relative to traditional methods: Less information loss; Fewer flawed assumptions from canonical summary features; Broadens scope beyond traditional features of interest

Relative to either method alone: Avoids compressing over distinct latent states (vs. FPCA alone); Identifies functional variation within states (vs. FHMM alone)

Relative to multi-dimensional functional analyses for EEG: Integrated insights + relationships across multiple levels of analysis; Better interpretability at each level

Chapter 6

Data-driven phenotypes of resting state brain activity in adolescents: Cross-method, cross-sample replication and extension using functional analyses

Chapter 6: Novel Contributions

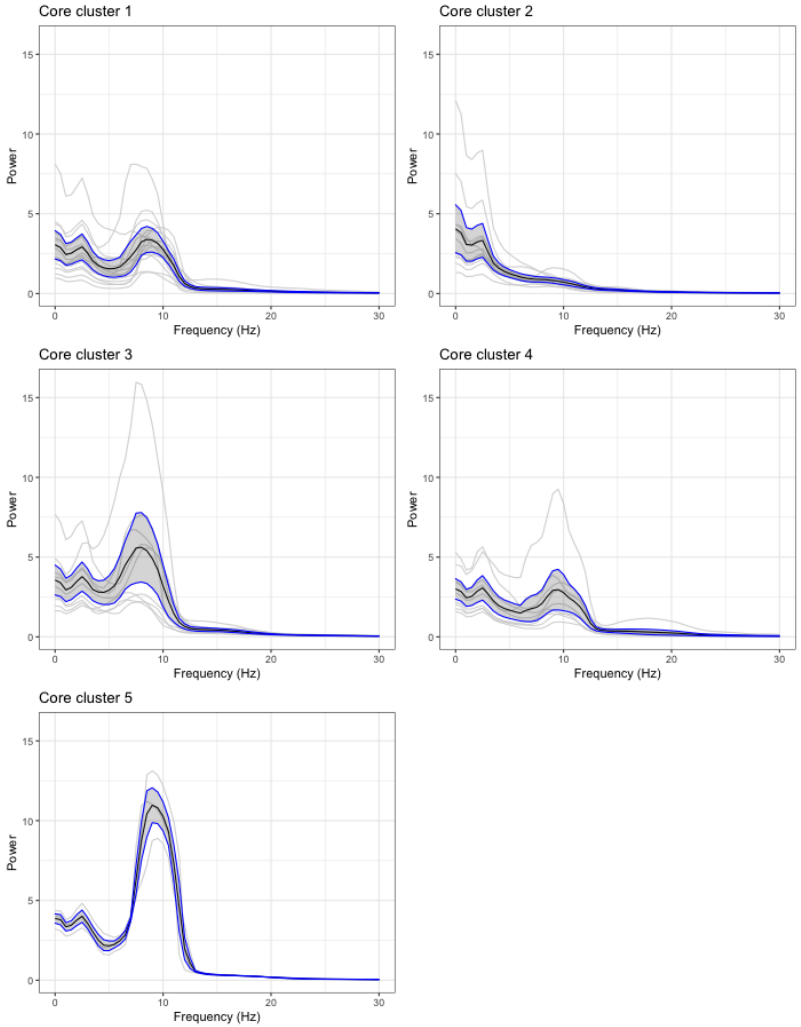
(1) **Cross-sample replication** of phenotypes from Chapter 3 clustering (Summary Features) across LABS and HBN datasets

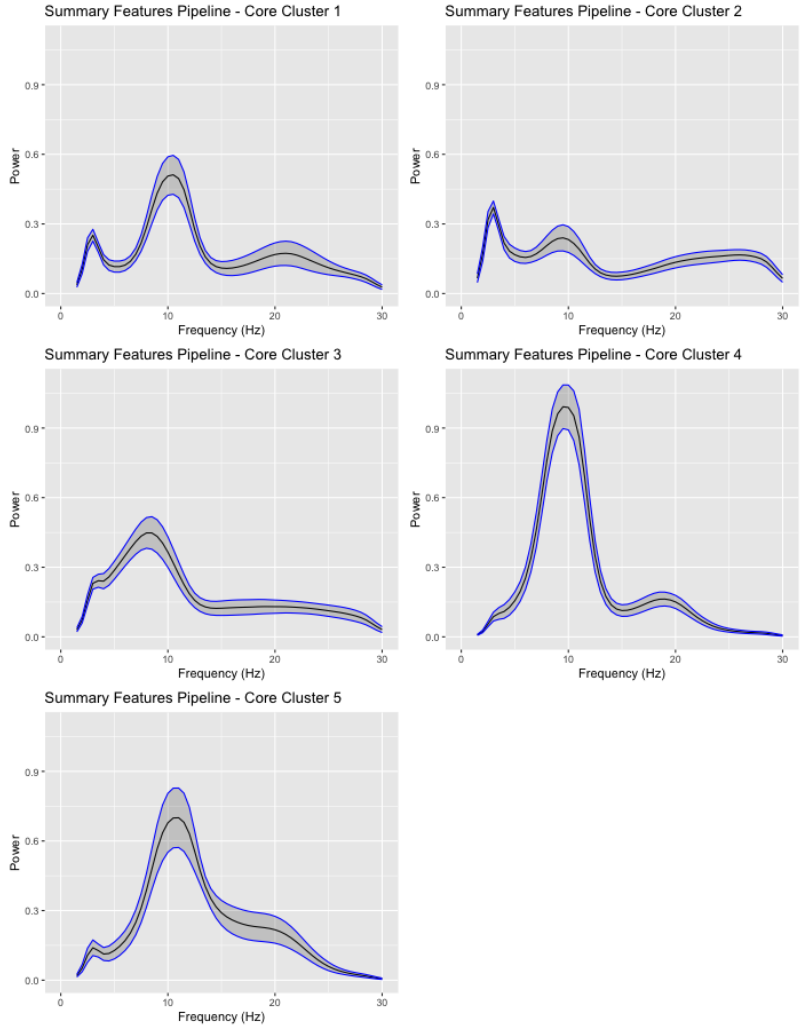
(2) **Cross-method replication** of phenotypes from Chapter 3 (Summary Features) and Chapter 5 (FHMM States) with HBN data

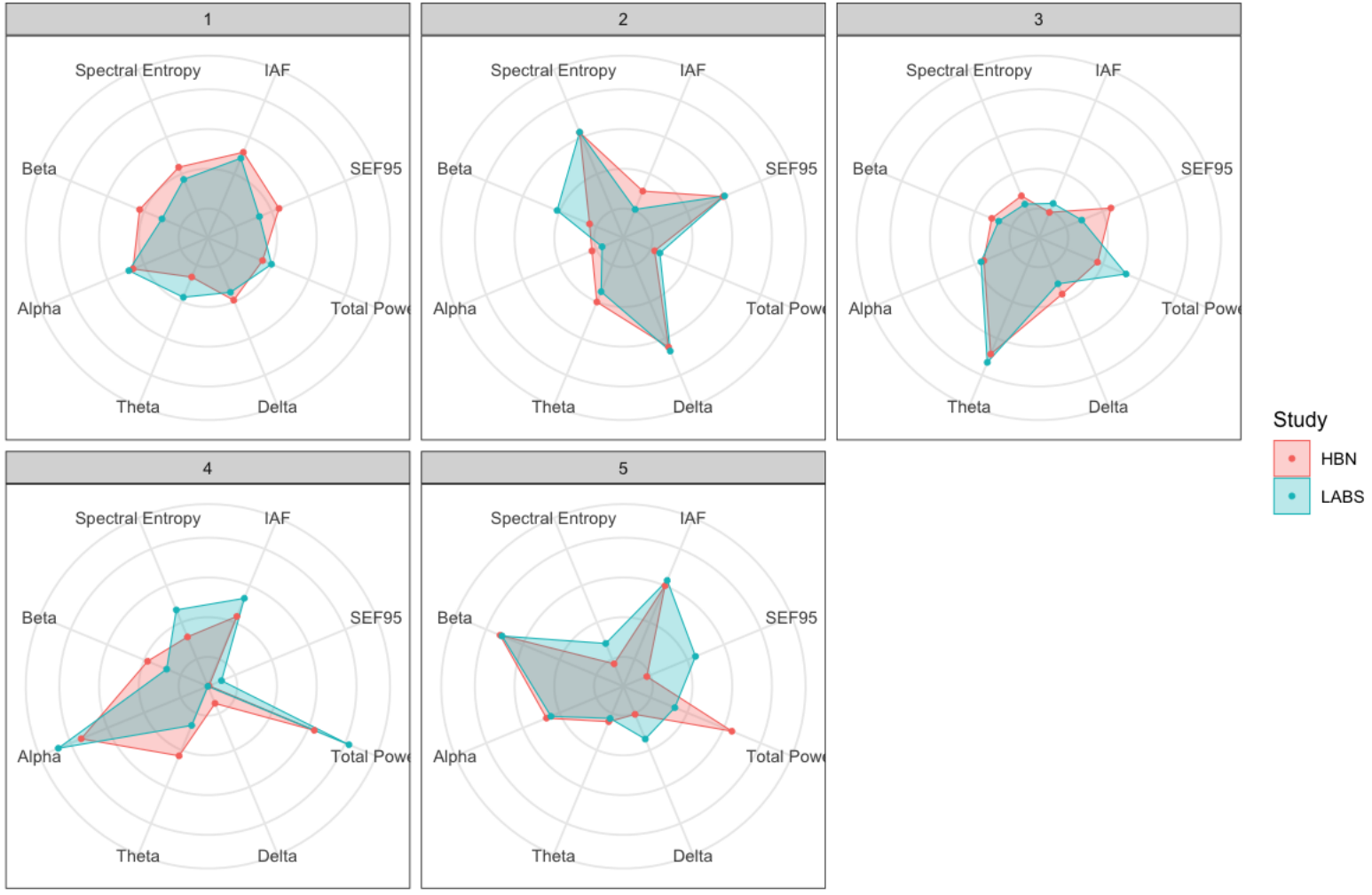
(3) **Extension of phenotypes** by clustering *flawless* analysis outputs for latent state dynamics and FPC scores with HBN data

- Consistent patterns of association with psychopathology and cognitive function
- Replication indicates more substantial evidence & lays foundation for clinical utility in the context of personalised healthcare and risk prediction

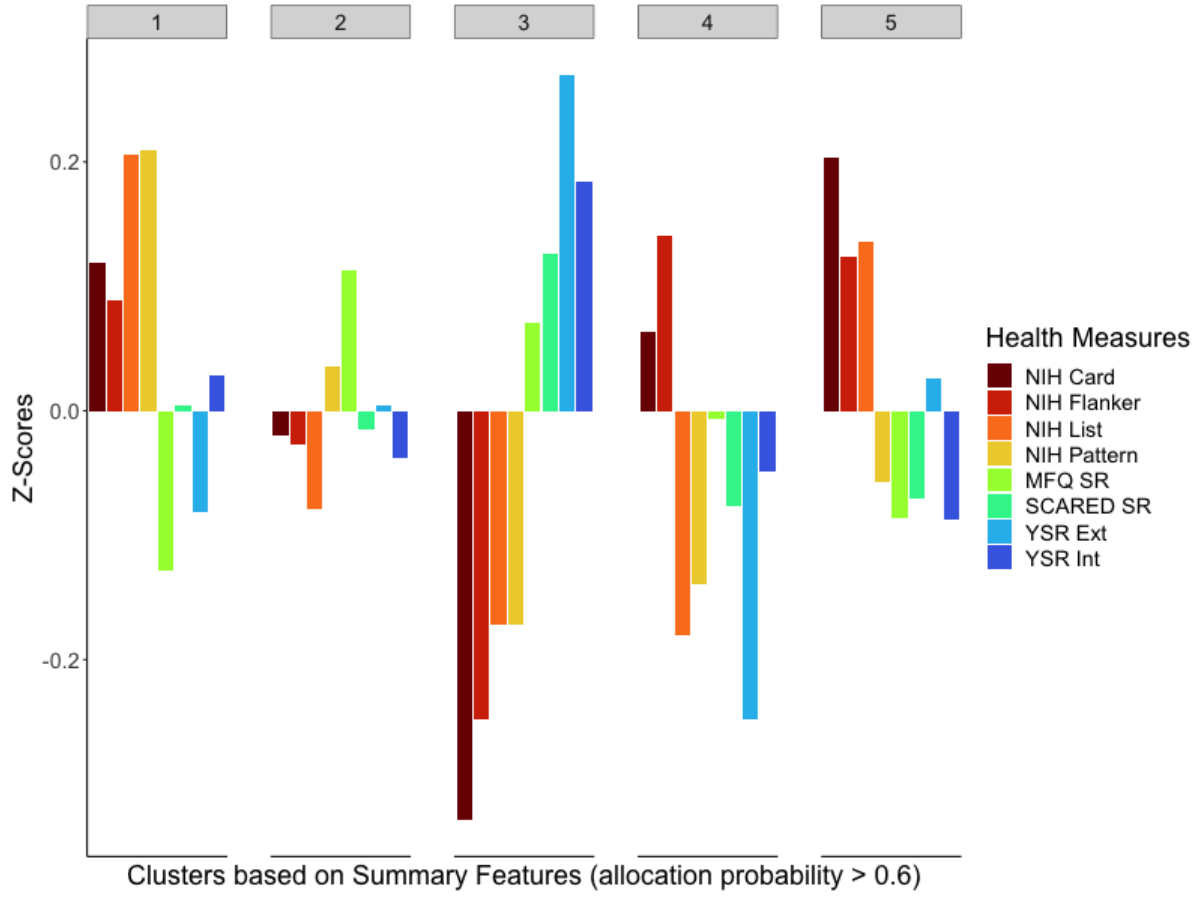
(1) Cross-sample replication of Chapter 3 clustering across LABS and HBN datasets



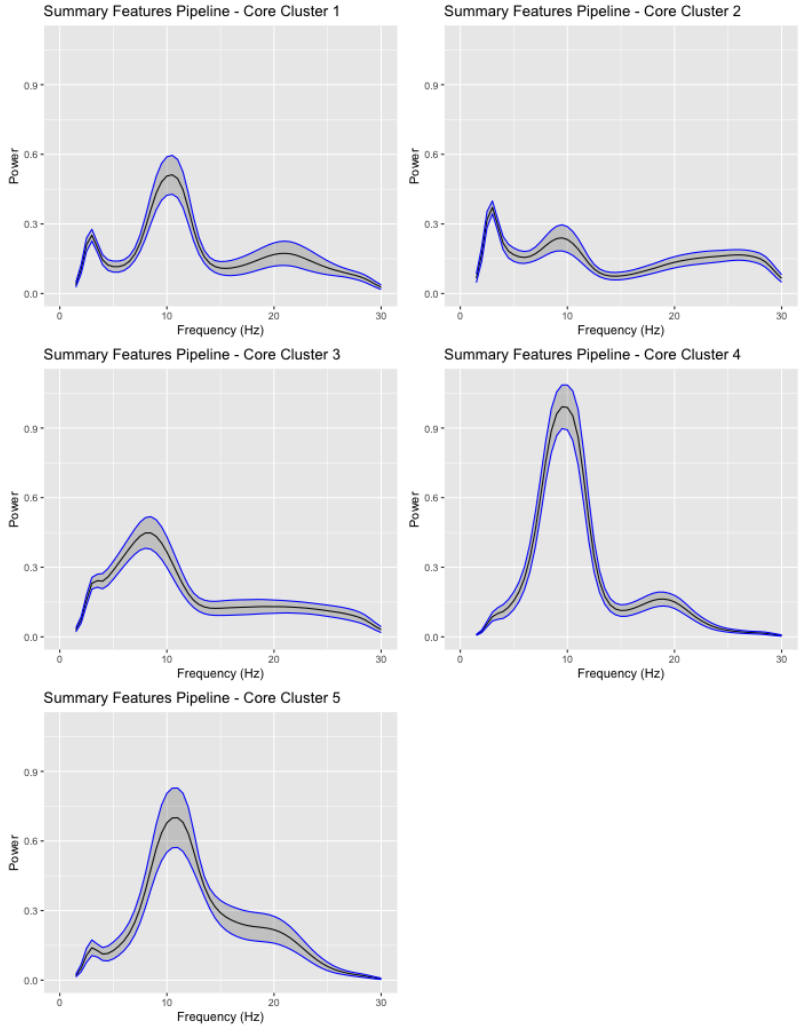


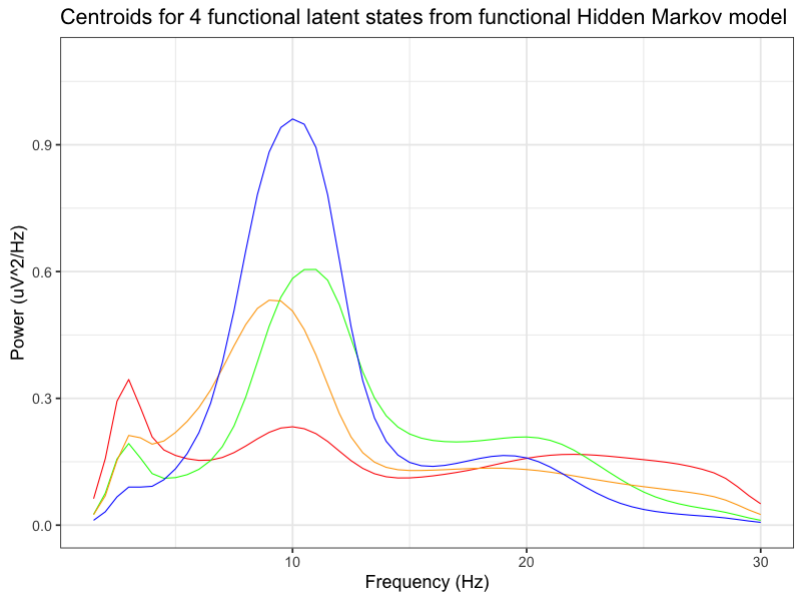

1
2
3
4
5
Spectral Entropy
IAF
SEF95
Total Pow
Delta
Theta
Alpha
Beta
Study
HBN
LABS

Health characteristics across Pipeline 1 clusters (time-averaged summary features)

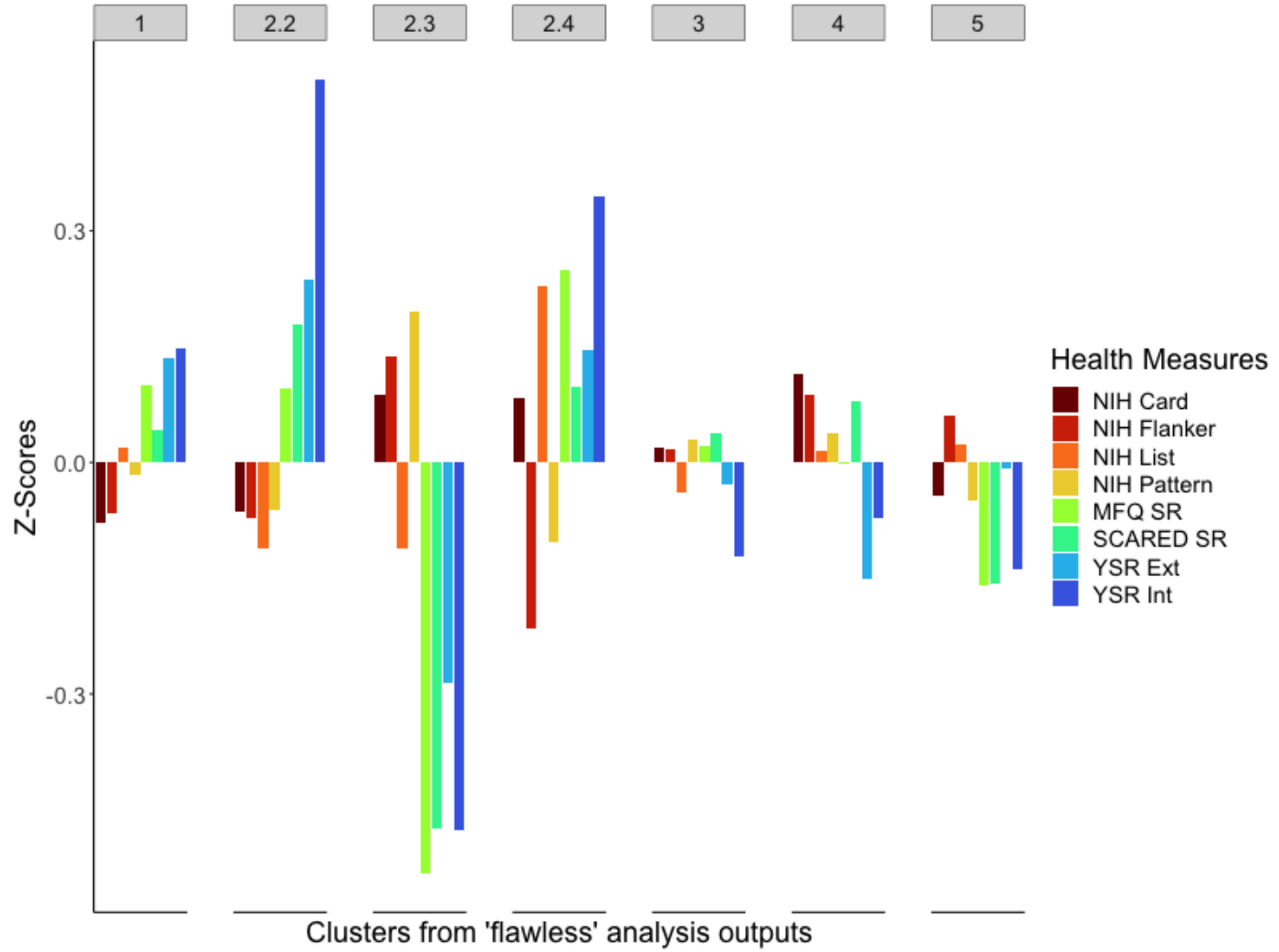


(2) Cross-method replication of clustering from Chapter 3 (Summary Features) and Chapter 5 (FHMM States) with HBN data





(3) Extension of phenotypes by clustering (*flawless*) outputs for latent state dynamics and FPC scores with HBN data



🧠 ⭐ Cross-method, cross-sample replication of resting state EEG phenotypes in adolescents, and extension of more sophisticated phenotypes based on latent states, temporal dynamics, and functional characteristics within states

🔥 🆒 😎 Strong foundation for potential clinical applications - risk/protective biomarkers, support early intervention and care planning personalised medicine 💯 💯 💯

Limitations

Cross-sectional data - limits understanding of phenotypes and health measures changing within individuals over time

Single electrode data - limits understanding of spatial patterns of brain activity and functional connectivity

EEG only - we did not look at other neuroimaging modalities

Future Directions

- Longitudinal data & time-series analyses - understand longer term patterns of change and health risk
- Multi-electrode data - investigate spatial patterns & functional connectivity
- Combine EEG with other neuroimaging modalities (e.g. Diffusion Tensor Imaging, Functional MRI)

- Fully Bayesian implementations of *clusterBMA* and *flawless*

- Development of user-friendly software with GUIs for testing in clinical settings

Acknowledgements

- Supervisors: Kerrie Mengersen, Paul Wu and Edgar Santos-Fernandez
- Panel members: David Lovell, Kate Helmstedt
- PhD stipend and travel funding: Kerrie, QUT, CDS, ACEMS, SSA, IBS-AR
- Researchers, Participants & Caregivers for LABS and HBN studies
- Collaborators: Paul Schwenn, Hongbo Xie
- Friends, family, my partner Satomi

***clusterBMA* R package: bit.ly/clusterBMA**

- 🦪 Twitter: **@oforbes22**

References

Aggarwal, C. C. and Reddy, C. K. (2014). Data clustering: Algorithms and applications. Chapman Hall/CRC Data mining and Knowledge Discovery Series. London.

Babadi, B., & Brown, E. N. (2014). A review of multitaper spectral analysis. IEEE Transactions on Biomedical Engineering, 61(5), 1555-1564.

Bernardo, J. M. and Smith, A. F. (2009). Bayesian theory, volume 405. John Wiley Sons.

Bruckers, L., Molenberghs, G., Drinkenburg, P., & Geys, H. (2016). A clustering algorithm for multivariate longitudinal data. Journal of biopharmaceutical statistics, 26(4), 725-741.

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. International Journal of Psychophysiology, 97(3):210-220.

Cohen, M. X. (2014). Analyzing neural time series data: theory and practice. MIT press.

Connell, A. M., & Frye, A. A. (2006). Growth mixture modelling in developmental psychology: Overview and demonstration of heterogeneity in developmental trajectories of adolescent antisocial behaviour. Infant and Child Development: An International Journal of Research and Practice, 15(6), 609-621.

References

de Lijster, J. M., van den Dries, M. A., van der Ende, J., Utens, E. M., Jaddoe, V. W., Dieleman, G. C., ... & Legerstee, J. S. (2019). Developmental trajectories of anxiety and depression symptoms from early to middle childhood: A population-based cohort study in the Netherlands. Journal of abnormal child psychology, 47(11), 1785-1798.

Duffy, F. H., McAnulty, G. B., and Albert, M. S. (1996). Effects of age upon interhemispheric EEG coherence in normal adults. Neurobiol. Aging 17, 587– 599. doi: 10.1016/0197-4580(96)00007-3

Feng, J., Xu, H., & Yan, S. (2012). Robust PCA in high-dimension: A deterministic approach. arXiv preprint arXiv:1206.4628.

Hartmann, J. A., Nelson, B., Spooner, R., Paul Amminger, G., Chanen, A., Davey, C. G., McHugh, M., Ratheesh, A., Treen, D., and Yuen, H. P. (2019). Broad clinical high-risk mental state (CHARMS): methodology of a cohort study validating criteria for pluripotent risk. Early Intervention in Psychiatry, 13(3):379-386.

Jin, X., Liang, X., & Gong, G. (2020). Functional integration between the two brain hemispheres: evidence from the homotopic functional connectivity under resting state. Frontiers in Neuroscience.

Kaiser, A. K., Gnjezda, M. T., Knasmüller, S., & Aichhorn, W. (2018). Electroencephalogram alpha asymmetry in patients with depressive disorders: current perspectives. Neuropsychiatric disease and treatment.

References

Keizer, A. W. (2019). Standardization and Personalized Medicine Using Quantitative EEG in Clinical Settings. Clinical EEG and neuroscience, page 1550059419874945.

Ke, L., & Li, R. (2009, November). Classification of EEG signals by multi-scale filtering and PCA. In 2009 IEEE International Conference on Intelligent Computing and Intelligent Systems (Vol. 1, pp. 362-366). IEEE.

Latzman, R. D. and DeYoung, C. G. (2020). Using empirically-derived dimensional phenotypes to accelerate clinical neuroscience: The Hierarchical Taxonomy of Psychopathology (HiTOP) framework. Neuropsychopharmacology, 45(7):1083-1085.

Liedorp, M., Van Der Flier, W., Hoogervorst, E., Scheltens, P., and Stam, C. (2009). Associations between patterns of EEG abnormalities and diagnosis in a large memory clinic cohort. Dementia and geriatric cognitive disorders, 27(1):18-23.

Loo, S. K., Lenartowicz, A., and Makeig, S. (2016). Research review: Use of EEG biomarkers in child psychiatry research{current state and future directions. Journal of Child Psychology and Psychiatry, 57(1):4-17.

Margaritella, N., Inácio, V., & King, R. (2021). Parameter clustering in Bayesian functional principal component analysis of neuroscientific data. Statistics in Medicine, 40(1), 167-184.

McGorry, P. D. and Mei, C. (2018). Early intervention in youth mental health: progress and future directions. Evidence-based mental health, 21(4):182-184.

References

McGorry, P. D., Purcell, R., Goldstone, S., and Amminger, G. P. (2011). Age of onset and timing of treatment for mental and substance use disorders: implications for preventive intervention strategies and models of care. Current opinion in psychiatry, 24(4):301-306.

Newson, J. J., & Thiagarajan, T. C. (2019). EEG frequency bands in psychiatric disorders: a review of resting state studies. Frontiers in human neuroscience, 12, 521.

Paus, T., Keshavan, M., and Giedd, J. N. (2008). Why do many psychiatric disorders emerge during adolescence? Nature reviews neuroscience, 9(12):947-957.

Prerau, M. J., Brown, R. E., Bianchi, M. T., Ellenbogen, J. M., & Purdon, P. L. (2017). Sleep neurophysiological dynamics through the lens of multitaper spectral analysis. Physiology, 32(1), 60-92.

Sawyer, M. G., Reece, C. E., Sawyer, A. C., Johnson, S. E., and Lawrence, D. (2018). Has the prevalence of child and adolescent mental disorders in Australia changed between 1998 and 2013 to 2014? Journal of the American Academy of Child Adolescent Psychiatry, 57(5):343-350. e5.

Sun, S. and Zhou, J. (2014). A review of adaptive feature extraction and classification methods for EEG-based brain-computer interfaces. In 2014 International Joint Conference on Neural Networks (IJCNN), pages 1746-1753. IEEE.

References

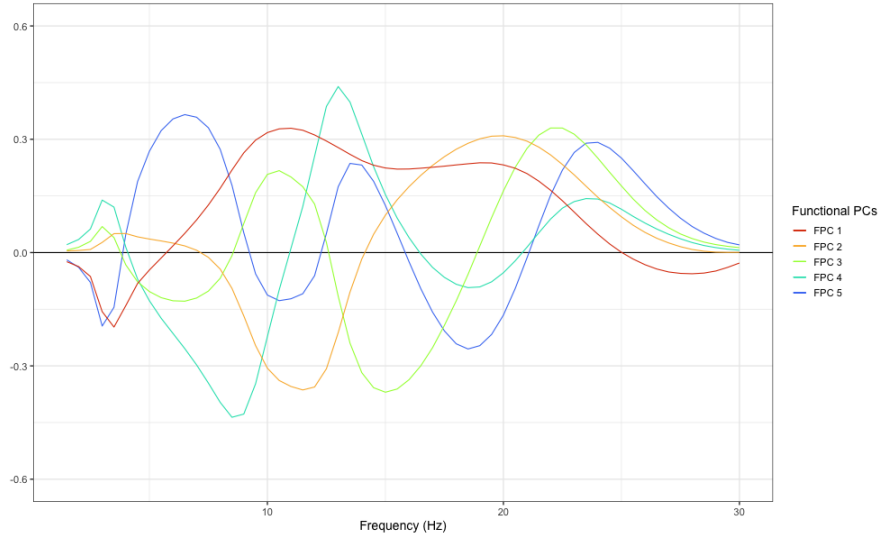
Raballo, A. and Poletti, M. (2020). Advances in early identification of children and adolescents at risk for psychiatric illness. Current Opinion in Psychiatry, 33(6):611-617.

Ramsay, J. O., Hooker, G., & Graves, S. (2009). Functional data analysis with R and MATLAB: Springer Science & Business Media.

Tenke, C. E., & Kayser, J. (2005). Reference-free quantification of EEG spectra: combining current source density (CSD) and frequency principal components analysis (fPCA). Clinical Neurophysiology, 116(12), 2826-2846. Chicago

Wu, W., Nagarajan, S., and Chen, Z. (2015). Bayesian Machine Learning: EEGMEG signal processing measurements. IEEE Signal Processing Magazine, 33(1):14-36.

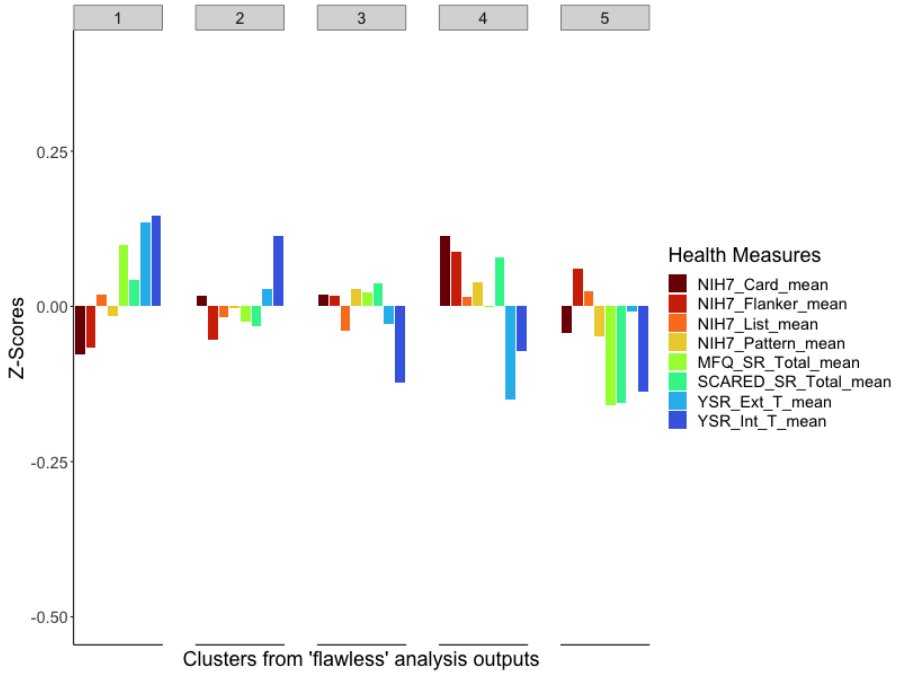
FPCA & Bayesian Regression



- Bayesian regression models indicated substantial associations between:
- Latent state temporal characteristics
- **Functional frequency characteristics within states (FPCA scores)**
- Outcome: Health measures relating to psychopathology & cognitive function

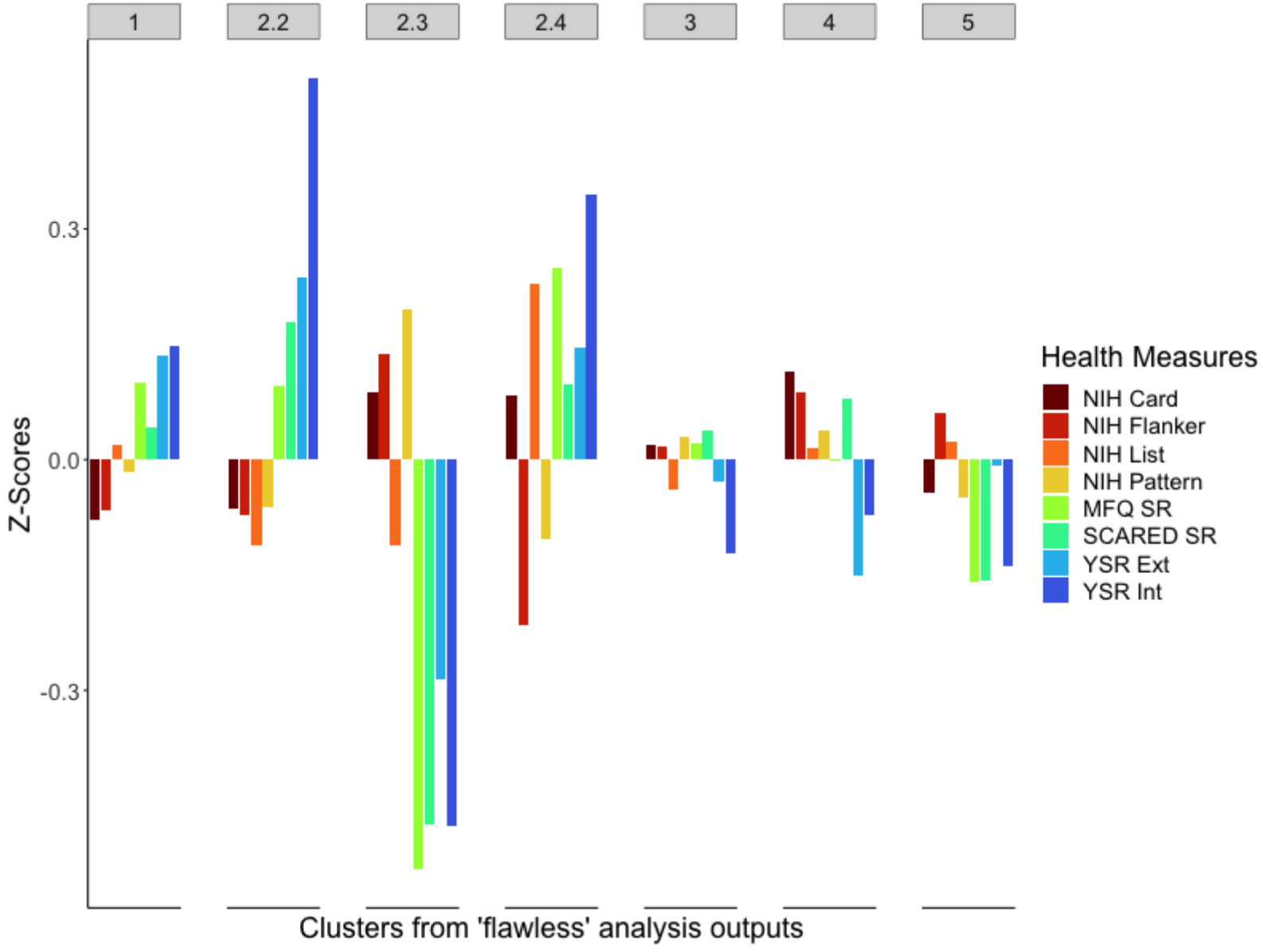
Independent Variable	Regression Coefficient	95% Credible Interval	Dependent Variable
N. States	4.1	-3.1, 11	MFQ SR
N. States	3.5	-6.5, 14	SCARED SR
N. States	7.3	0.11, 15	YSR Ext
N. States	2.4	-5.1, 9.8	YSR Int
N. States	-6.5	-21, 7.6	NIH Card
N. States	-3	-12, 6.2	NIH Flanker
N. States	1.8	-13, 16	NIH List
N. States	9.1	-9.8, 28	NIH Pattern
N. Transitions	-2.8	-5.8, 0.27	MFQ SR
N. Transitions	-3.9	-8.1, 0.34	SCARED SR
N. Transitions	-2.5	-5.5, 0.45	YSR Ext
N. Transitions	-4.5	-7.7, -1.3	YSR Int
N. Transitions	-2.9	-8.9, 3.1	NIH Card
N. Transitions	-1.9	-5.8, 2.1	NIH Flanker
N. Transitions	-2	-8.2, 4.2	NIH List
N. Transitions	3.3	-4.6, 11	NIH Pattern
Dominant State %	-16	-47, 16	MFQ SR
Dominant State %	-45	-89, -1.1	SCARED SR
Dominant State %	-7.2	-38, 24	YSR Ext
Dominant State%	-33	-66, 0.03	YSR Int
Dominant State %	-42	-105, 19	NIH Card
Dominant State %	-30	-70, 11	NIH Flanker
Dominant State %	-17	-81, 47	NIH List
Dominant State %	101	15, 185	NIH Pattern
S3 FPC1	1.4	-14, 17	MFQ SR
S3 FPC1	8.4	-13, 30	SCARED SR
S3 FPC1	1	-14, 16	YSR Ext
S3 FPC1	0.23	-16, 16	YSR Int
S3 FPC1	11	-18, 41	NIH Card
S3 FPC1	0.94	-19, 20	NIH Flanker
S3 FPC1	-27	-58, 4.0	NIH List
S3 FPC1	-19	-60, 21	NIH Pattern
S3 FPC2	-3	-14, 8.2	MFQ SR
S3 FPC2	-5.3	-21, 10	SCARED SR
S3 FPC2	-0.3	-12, 11	YSR Ext
S3 FPC2	-0.6	-12, 11	YSR Int
S3 FPC2	3	-19, 25	NIH Card
S3 FPC2	3	-11, 17	NIH Flanker
S3 FPC2	-4	-26, 18	NIH List
S3 FPC2	-21	-51, 8.2	NIH Pattern
S3 FPC3	13	-0.76, 27	MFQ SR
S3 FPC3	5.9	-13, 24	SCARED SR
S3 FPC3	17	3.9, 31	YSR Ext
S3 FPC3	24	9.5, 38	YSR Int
S3 FPC3	-19	-46, 7.8	NIH Card
S3 FPC3	-12	-30, 5.0	NIH Flanker
S3 FPC3	-0.71	-28, 27	NIH List
S3 FPC3	-3	-39, 34	NIH Pattern
S3 FPC4	16	-1.3, 34	MFQ SR
S3 FPC4	25	0.22, 50	SCARED SR
S3 FPC4	17	-1.1, 35	YSR Ext
S3 FPC4	8.8	-10, 27	YSR Int
S3 FPC4	11	-25, 46	NIH Card
S3 FPC4	24	1.8, 48	NIH Flanker
S3 FPC4	-2.1	-38, 34	NIH List
S3 FPC4	-40	-88, 8.1	NIH Pattern
S3 FPC5	-1.8	-37, 33	MFQ SR
S3 FPC5	12	-36, 60	SCARED SR
S3 FPC5	2	-34, 37	YSR Ext
S3 FPC5	-0.29	-37, 36	YSR Int

Answer to Conor's Question - Exciting Findings from Chapter 6 😎 (extended phenotypes)



- Clusters based on FHMM and FPCA outputs from all latent states

Pipeline 2 (*flawless*) clusters	1	2	3	4	5
N. states	1.02	1.00	2.00	2.11	2.79
N. transitions	0.02	0.00	1.90	2.08	4.21
S1 proportion	1.00	0.00	0.53	0.03	0.28
S2 proportion	0.00	0.45	0.47	0.27	0.31
S3 proportion	0.00	0.29	0.01	0.37	0.23
S4 proportion	0.00	0.27	0.00	0.33	0.18


1
2.2
2.3
2.4
3
4
5
0.3
0.0
-0.3
Z-Scores
Clusters from 'flawless' analysis outputs
Health Measures
NIH Card
NIH Flanker
NIH List
NIH Pattern
MFQ SR
SCARED SR
YSR Ext
YSR Int

Pipeline 2 (*flawless*) clusters	1	2.2	2.3	2.4	3	4	5
N. states	1.015	1.000	1.000	1.000	2.000	2.107	2.790
N. transitions	0.015	0.000	0.000	0.000	1.902	2.083	4.210
S1 proportion	0.995	0.000	0.000	0.000	0.527	0.034	0.280
S2 proportion	0.000	1.000	0.000	0.000	0.465	0.268	0.313
S3 proportion	0.005	0.000	1.000	0.000	0.008	0.371	0.226
S4 proportion	0.000	0.000	0.000	1.000	0.000	0.327	0.181
S1 FPC1	0.072				-0.098	-0.176	-0.133
S1 FPC2	0.010				-0.051	0.239	0.001
S1 FPC3	-0.041				0.072	-0.056	0.081
S1 FPC4	-0.013				0.061	-0.064	-0.064
S1 FPC5	0.004				-0.021	0.084	0.016
S2 FPC1		0.048			-0.134	0.148	-0.037
S2 FPC2		-0.048			-0.006	0.098	0.049
S2 FPC3		0.074			0.022	-0.067	-0.153
S2 FPC4		0.046			-0.019	-0.054	-0.023
S2 FPC5		-0.038			-0.016	0.062	0.043
S3 FPC1	-0.154		0.084		-0.161	0.013	-0.198
S3 FPC2	0.298		0.073		0.056	-0.055	-0.065
S3 FPC3	0.017		-0.013		-0.104	-0.030	0.056
S3 FPC4	-0.006		0.041		0.113	-0.043	-0.065
S3 FPC5	-0.002		-0.005		-0.071	0.009	0.010
S4 FPC1				0.212	-0.043	-0.136	-0.223
S4 FPC2				-0.007	0.008	0.067	-0.062
S4 FPC3				0.017	-0.047	0.012	-0.054
S4 FPC4				-0.001	0.131	0.004	-0.030
S4 FPC5				-0.012	-0.057	0.002	0.005

Table 4: FHMM latent state temporal dynamics and FPC scores between clusters from pipeline 2 based on *flawless* analysis outputs.

FHMM Details

A Hidden Markov Model is a bivariate process $\{(Q_k, \boldsymbol{X}_k)\}_{k\geq 1}$ defined on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that:

- $\{Q_k\}_{k\geq 1}$ is a Markov chain with a discrete and finite state space $\{s_1, \ldots, s_N\}$, with $N \geq 1$, transition matrix $A = \{a_{ij}\} = \mathbb{P}(Q_k = s_j \mid Q_{k-1} = s_i)$ and initial distribution $\boldsymbol{\nu}$, where $\nu_i = \mathbb{P}(Q_1 = s_i)$;
- For each time k, the observation $\boldsymbol{X}_k$ is a d-dimensional random array. In particular, given the state process $\{Q_k\}_{k\geq 1}$, $\boldsymbol{X}_k$ is a sequence of conditionally independent random arrays (vectors or matrices, depending on the type of data) (Cappé et al., 2009; Martino et al., 2020).

In the general case, the objective function for a HMM can be written as:

$$\begin{aligned}\log(\mathcal{L}(\lambda \mid \mathbf{x})) = &\sum_{i=1}^{N} \gamma_1(i) \log \nu_i + \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\sum_{k=1}^{K-1} \xi_k(i,j)\right) \log a_{ij} \\ &+ \sum_{i=1}^{N}\sum_{k=1}^{K} \gamma_k(i) \log b_i(\mathbf{x}_k; \boldsymbol{\theta}_i)\,.\end{aligned} \tag{1}$$

S_Dbw Details

The S_Dbw index is calculated as the sum of an intra-cluster variance term $Scat(K)$ that measures cluster compactness, and a density term $Dens\_bw(K)$ that measures inter-cluster density:

$$S\_Dbw = Scat(K) + Dens\_bw(K). \tag{10}$$

The intra-cluster variance term $Scat(K)$ is defined as:

$$Scat(K) = \frac{1}{K} \sum_{k=1}^{K} \frac{||\sigma(C_k)||}{||\sigma(D)||}, \tag{11}$$

where $\sigma(C_k)$ is the variance of cluster C_k and $\sigma(D)$ is the variance of the dataset. The inter-cluster density term $Dens\_bw(K)$ is defined as:

$$\begin{aligned} Dens\_bw(K) &= \frac{1}{K(K-1)} \sum_{k=1}^{K} \left(\sum_{j \neq k}^{K} \frac{\sum_{x \in C_k \cup C_j} f(x, u_{kj})}{max\left(\sum_{x \in C_k} f(x, c_k), \sum_{x \in C_j} f(x, c_j) \right)} \right) \\ f(x, y) &= \begin{cases} 0 & \text{if } d(x, y) > \frac{1}{K} \sqrt{\sum_{k=1}^{K} ||\sigma(C_k)||} \\ 1 & \text{otherwise,} \end{cases} \end{aligned} \tag{12}$$

where u_{kj} is the mid-point between c_k and c_j. $Dens\_bw$ represents a ratio of inter-cluster density to within cluster density, with lower values indicating better

k-means

- 'Hard' clustering
- Minimises within-cluster sums of squares

k-means objective function

$$J = \sum_{i=1}^{K}(\sum_{k} ||x_k - c_i||^2)$$

Hierarchical Clustering (Ward's Method)

- 'Hard' clustering
- Each observation starts out in its own cluster
- Repeated pairwise fusion of clusters that minimises change in within-cluster sums of squares (Ward)

Ward's objective function

$$D(c_1, c_2) = \delta^2(c_1, c_2) = \frac{|c_1||c_2|}{|c_1| + |c_2|}||c_1 - c_2||^2$$

Gaussian Mixture Model

- 'Soft' clustering
- Models data as coming from a mixture of Gaussian distributions
- Restrictive distribution ('thin tails')
    - Tends to prioritise cluster compactness, may tend to 'overcluster'

Mixture of multivariate Gaussians

$$p(x_n|\mu, \Sigma, \pi, K) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)$$

Chapter 3 details - Summary Features

Features

- Power in canonical bands -- Delta (0-4 Hz), ..., Beta (12-16 Hz)
- Total Power
- Individual Alpha Frequency
- Spectral Edge Frequency
- Spectral Entropy

Principal Component Analysis

- 3 PCs explained 80.6% of the variance

Clustering

- k-means
- Hierarchical Clustering
- Gaussian Mixture Model