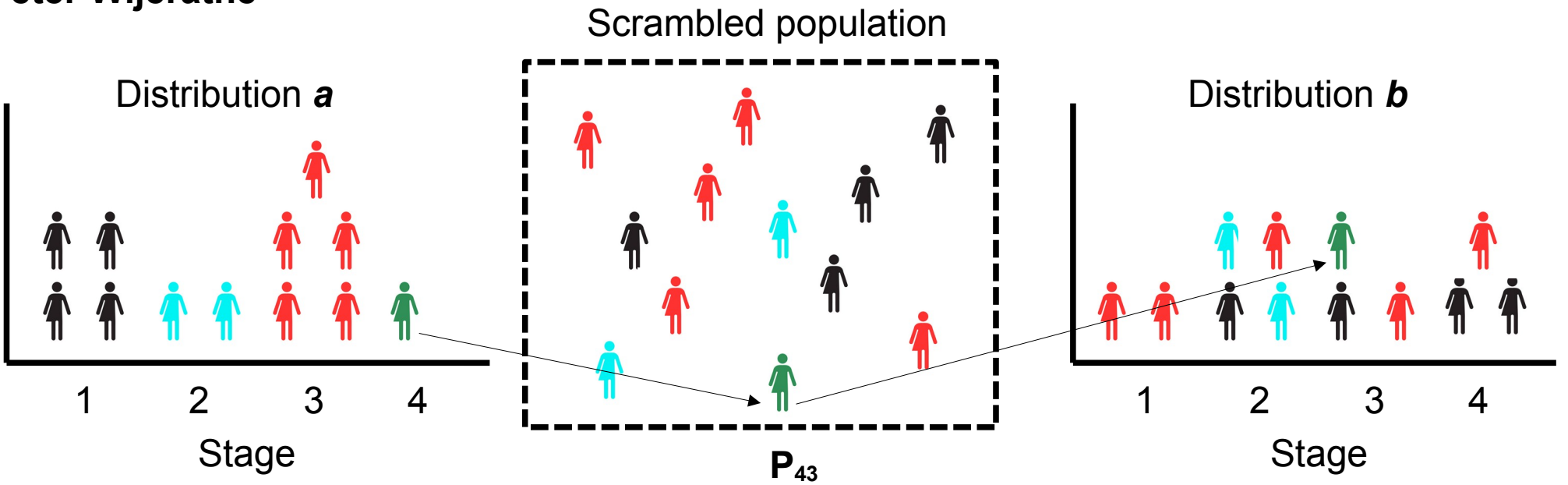
A probabilistic perspective on modelling disease progression (with some optimal transport)

Sussex Maths seminar
20/03/25

Peter Wijeratne

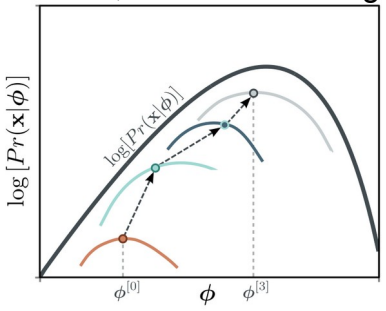


Our research: bridging computer and life sciences

UNIVERSITY OF SUSSEX

Informatics

* Prince, S. Understanding Deep Learning. MIT Press



Machine learning & statistics

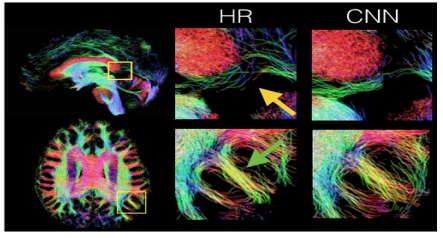
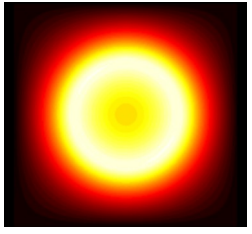


Image processing



Physics-guided modelling

Sussex AI Centre

Life sciences

sussex NEUROSCIENCE
open neuroscience

Neuroscience

Clinical imaging sciences

brighton and sussex medical school

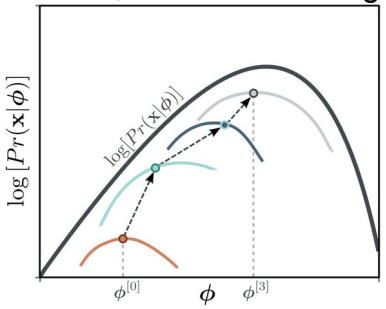
Clinical and translational research

Our research: bridging computer and life sciences

UNIVERSITY OF SUSSEX

Informatics

* Prince, S. Understanding Deep Learning. MIT Press



Machine learning & statistics

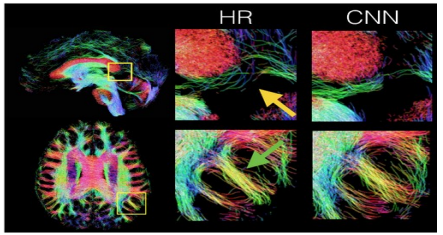
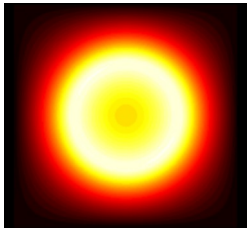


Image processing



Physics-guided modelling

Learners of Latents

Aisha Shuaibu
Paula Seidler
Kieran Gibb
Ivor Simpson
Sanduni Pinnawala
Giuseppe Castiglione
Luca Trautmann
Annie Hartanto

Life sciences

sussex NEUROSCIENCE
open neuroscience

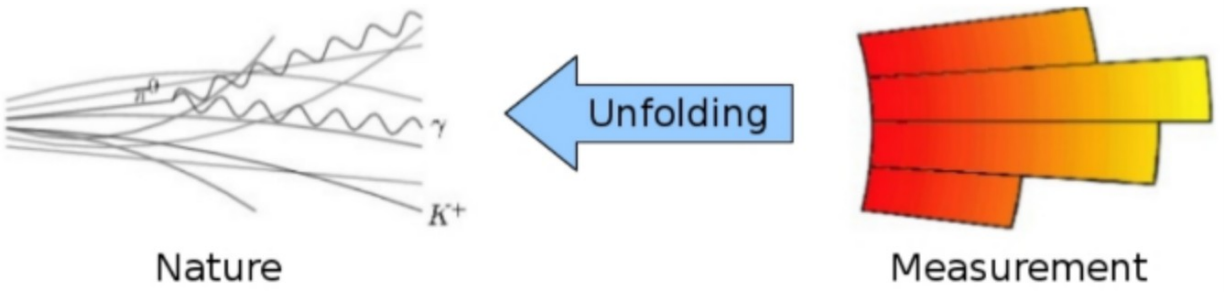
Neuroscience

Clinical imaging sciences

brighton and sussex medical school

Clinical and translational research

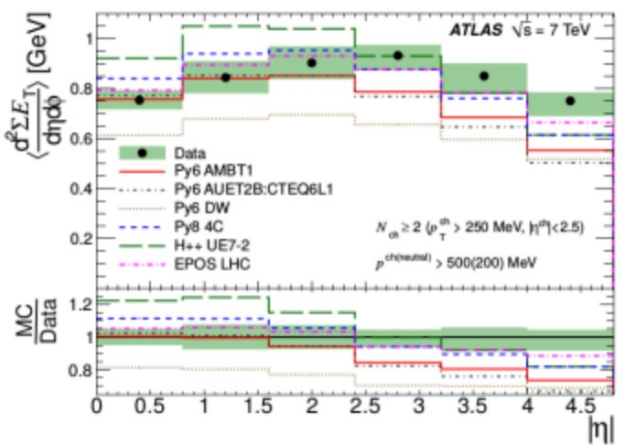
Background: PhD in particle physics



$$n(C_i^{data}) = \frac{1}{\epsilon_i} \sum_j P(T_i^{MC} \mid R_j^{MC}) n(R_j^{data})$$

- Real data are dependent on the detector used to measure them

Bring data back to their natural state by applying hypothesis-driven corrections derived from simulation



Postdoc 1: Biophysical modelling

UNIVERSITY OF SUSSEX

The Chemical Basis of Morphogenesis

A. M. Turing

Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, Vol. 237, No. 641. (Aug. 14, 1952), pp. 37-72.

Computational Modeling

$$\frac{\partial u}{\partial t} = F(u,v) - d_u v + D_u \Delta u$$

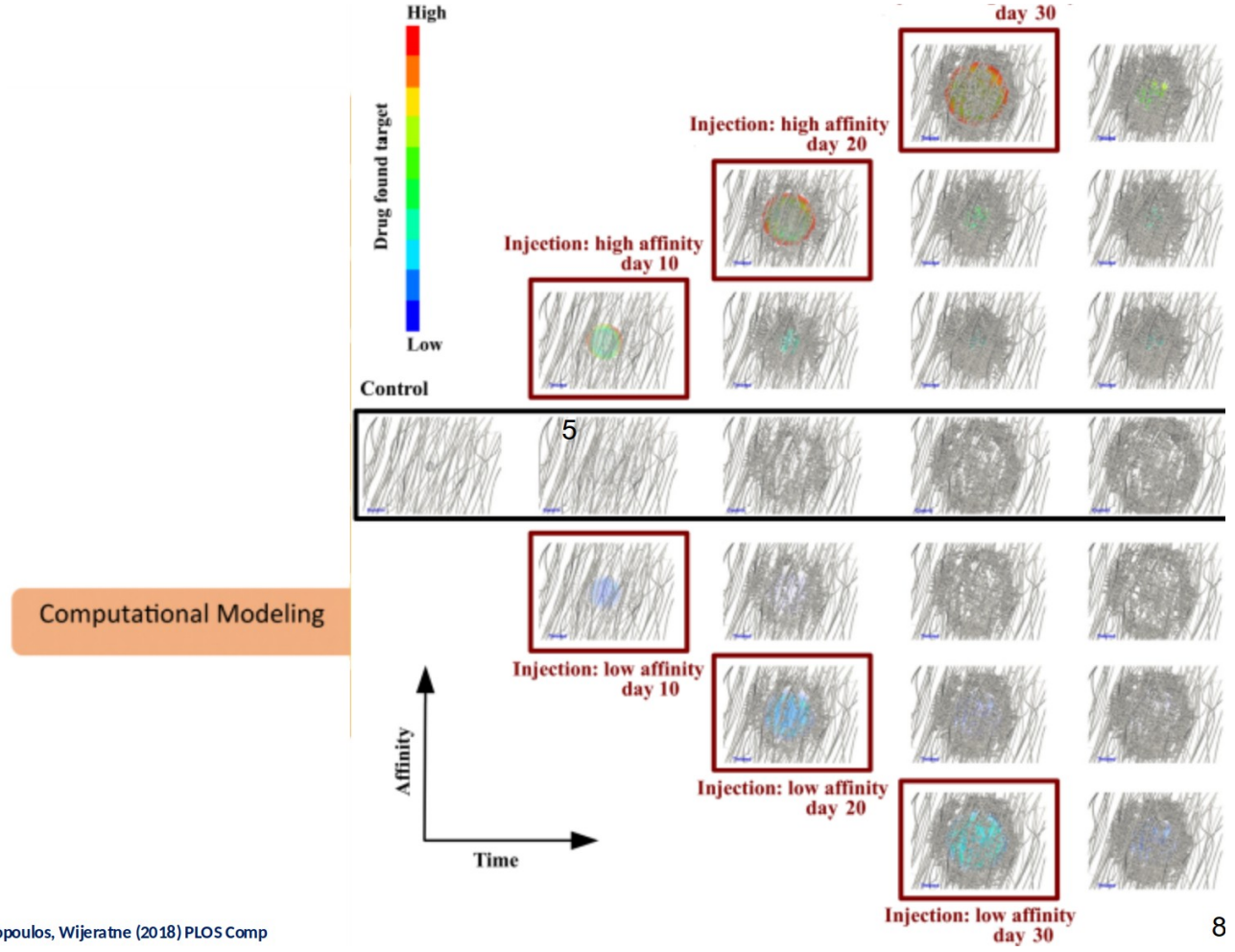
$$\frac{\partial v}{\partial t} = G(u,v) - d_v v + D_v \Delta v$$

Rate of concentration change | Production | Degradation | Diffusion

Reaction

FIGURE 2. An example of a 'dappled' pattern as resulting from a type (*a*) morphogen system. A marker of unit length is shown. See text, §9, 11.

Postdoc 1: Biophysical modelling



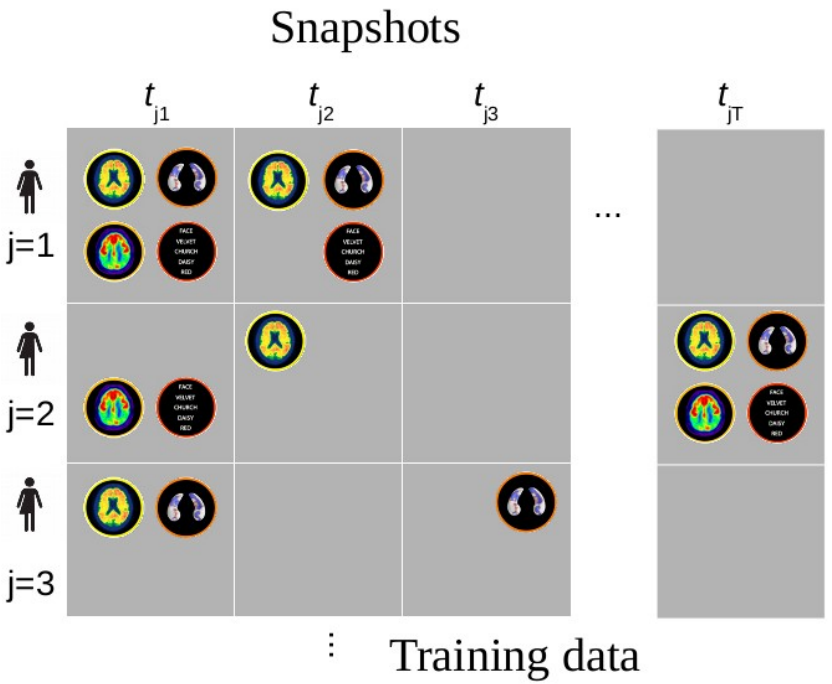
Vavourakis, Stylianopoulos, Wijeratne (2018) PLOS Comp

8

Postdoc 2 - present: latent variable modelling of dementia

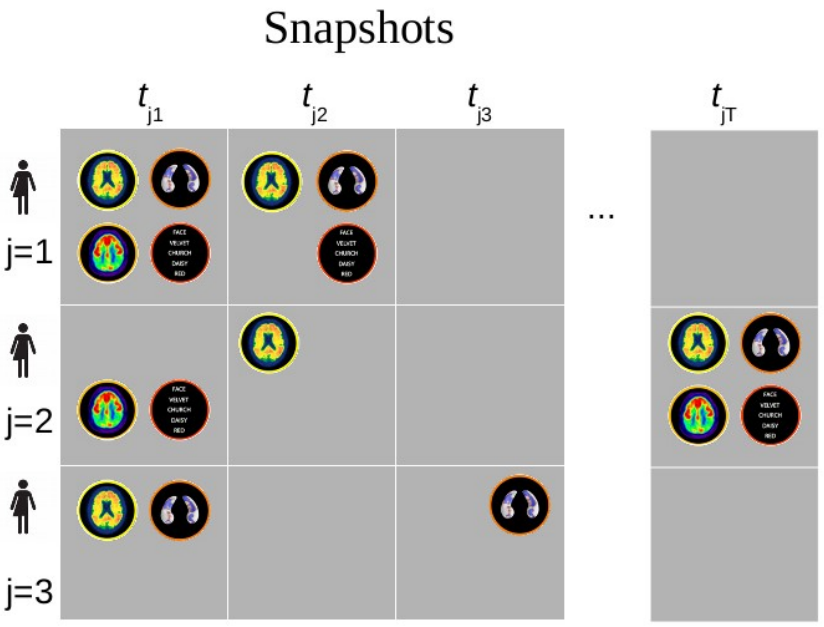
UNIVERSITY OF SUSSEX

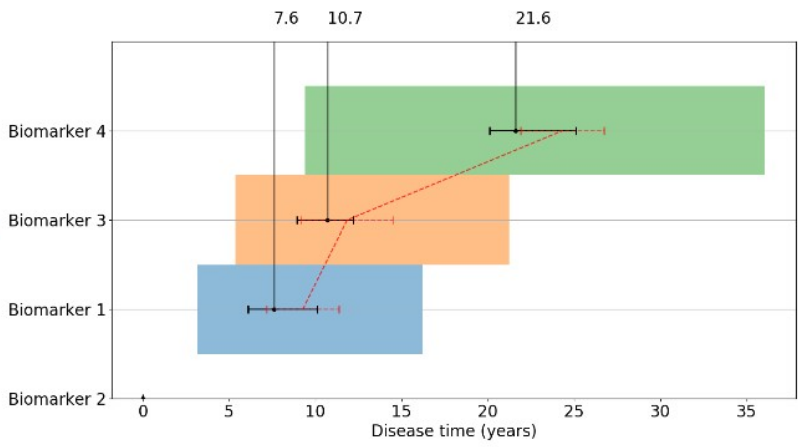
Build latent variable models (LVMs) that can leverage multi-modal data to characterise and predict progression



The big idea: LVM in neuro-degeneration and development

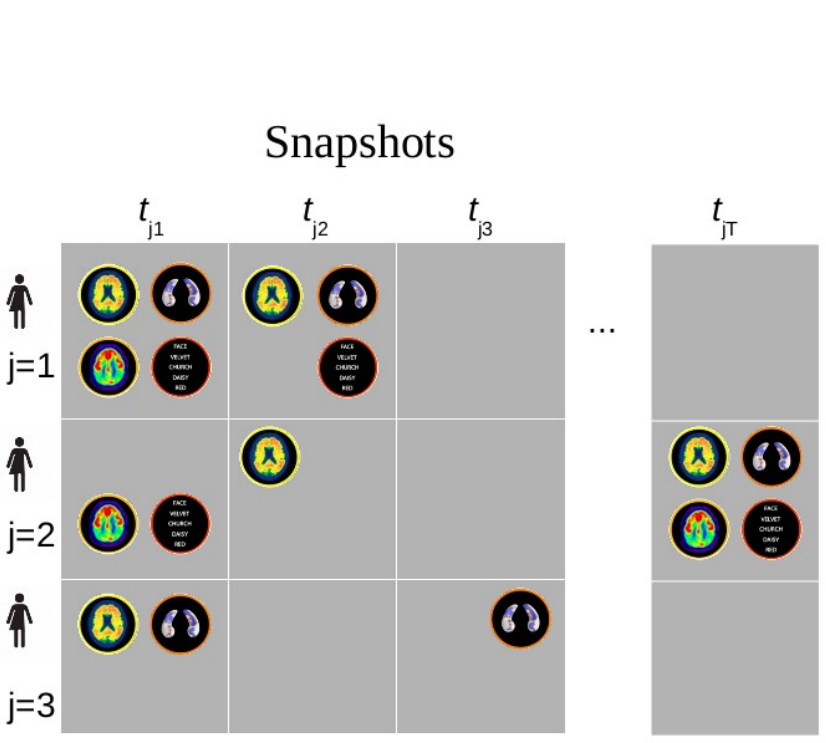
Build latent variable models (LVMs) that can leverage multi-modal data to characterise and predict progression

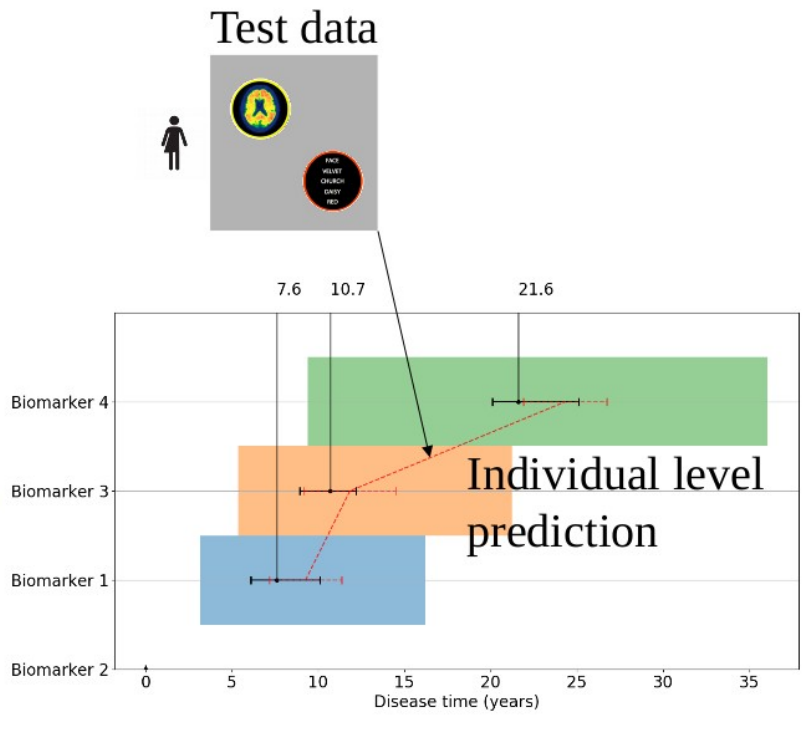




The big idea: ML in neuro-degeneration and development

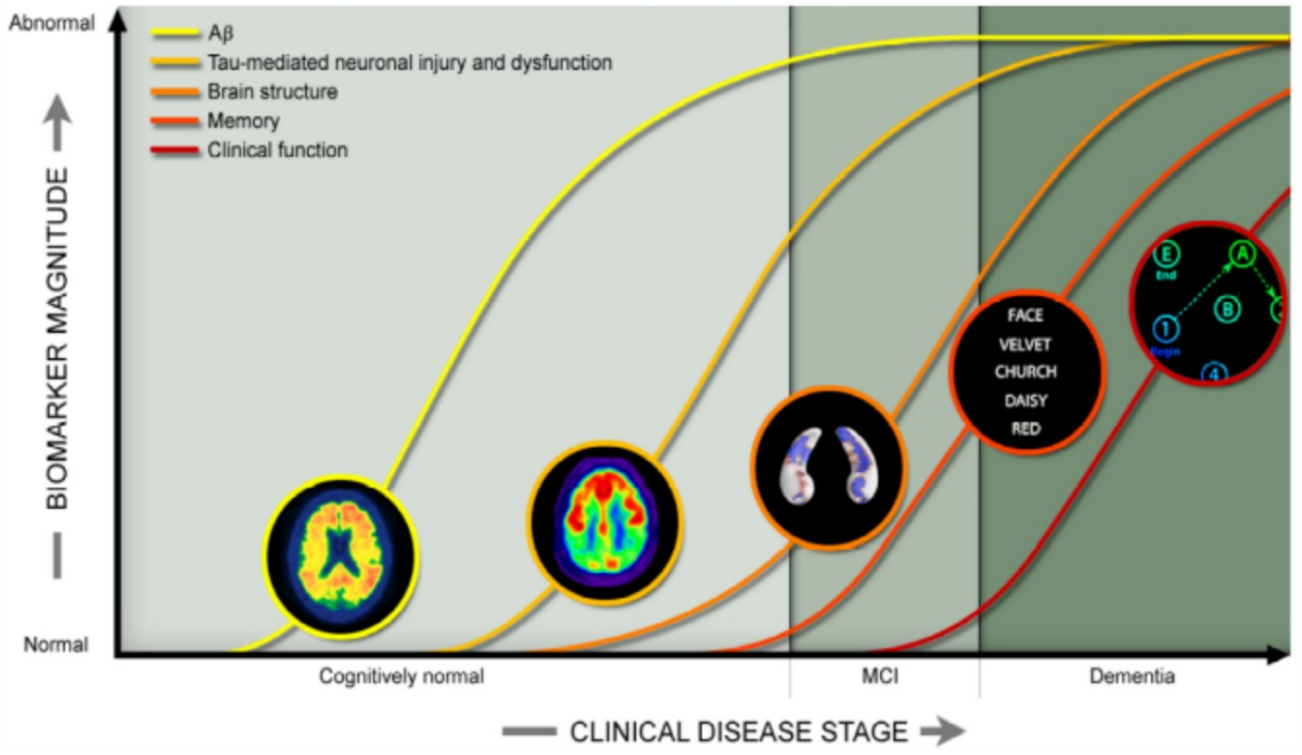
Build latent variable models (LVMs) that can leverage multi-modal data to characterise and predict progression





High level: disease progression modelling

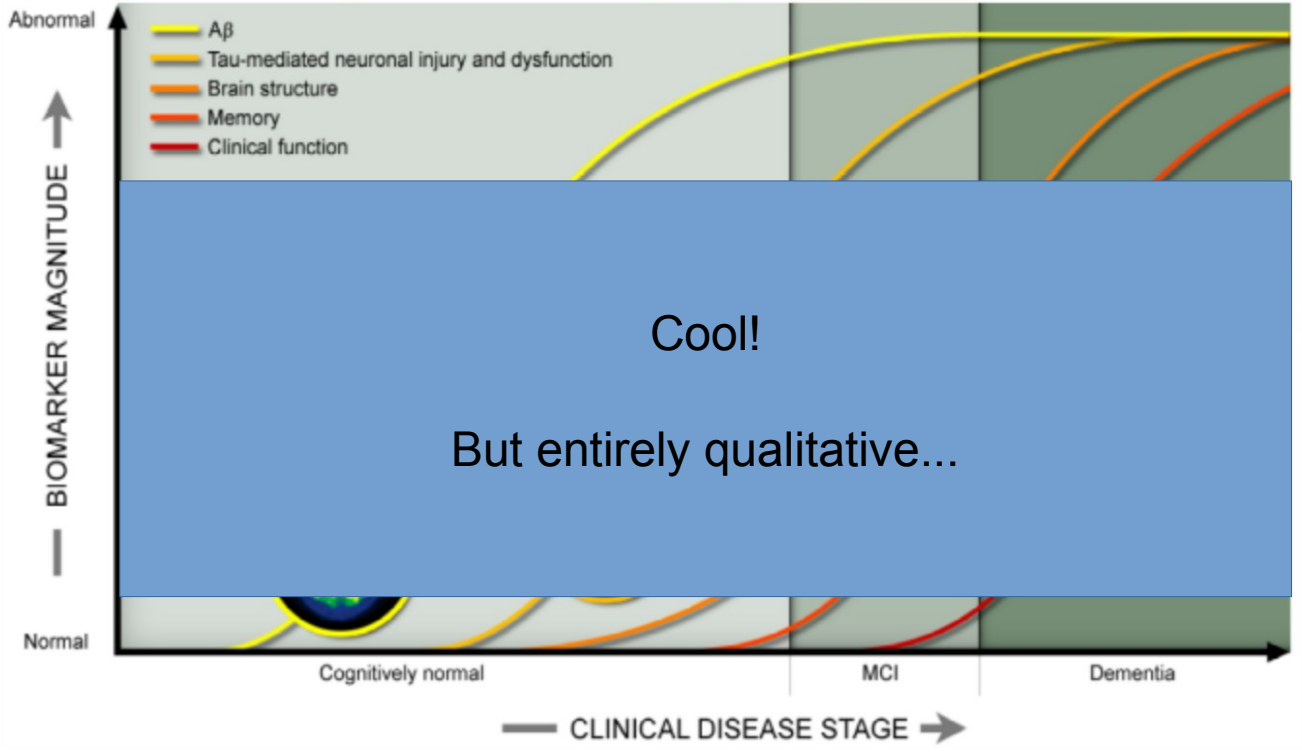
http://adni.loni.usc.edu/study-design/#background-container



A picture of how components of a disease progresses over time

High level: disease progression modelling

http://adni.loni.usc.edu/study-design/#background-container



A picture of how components of a disease progresses over time

Data-driven disease progression modelling

UNIVERSITY OF SUSSEX

Phenomenological – model disease progression in terms of observable changes in markers

e.g., how does the brain change over the course of Alzheimer's?

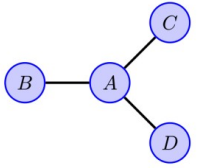
Mechanistic – model disease progression in terms of underlying mechanisms

e.g., why does the brain change over the course of Alzheimer's?

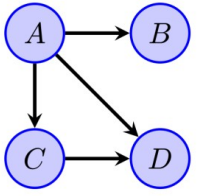
Low level: disease progression modelling

UNIVERSITY OF SUSSEX

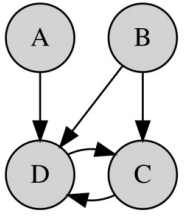
- Undirected graphical models (a.k.a. Markov networks)



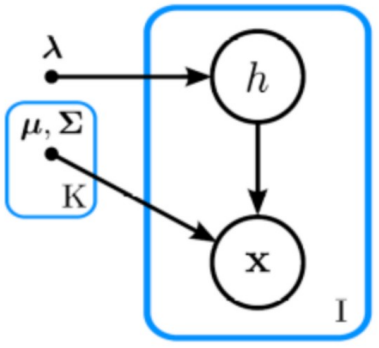
- Directed graphical models (a.k.a. Bayesian networks)
  - Directed acyclic graph (DAG)



  - Directed cyclic graph



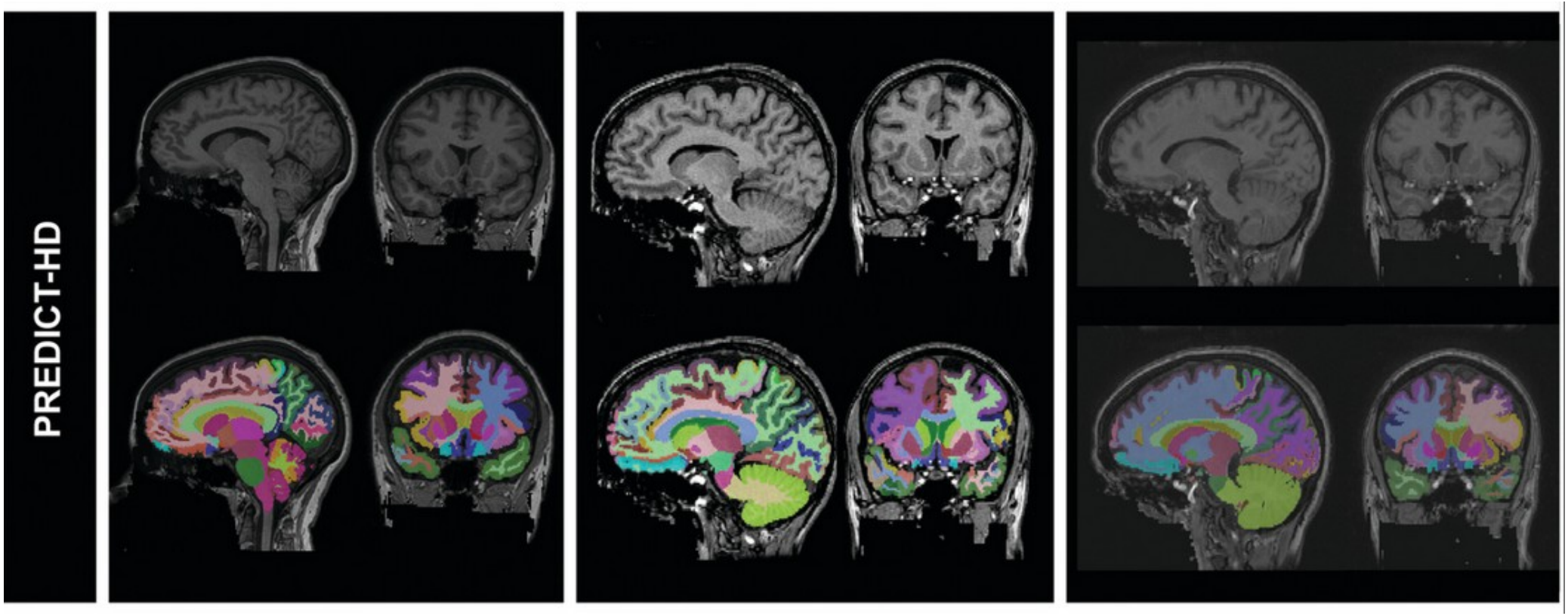
Low level: disease progression modelling



Mixture of Gaussians

- Circles: random variables.
- Blue boxes: "plates", signifies that the contents of the box should be repeated number of times in bottom corner.
- Bullet: variables that are not treated as uncertain.

Data: neuroimaging in humans



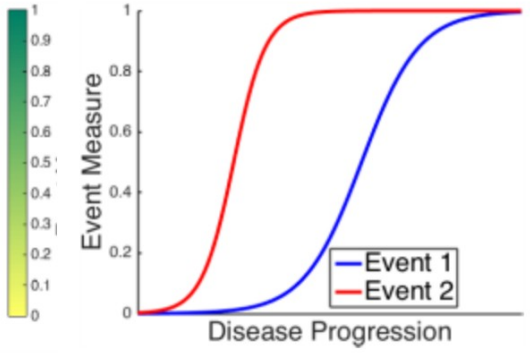
Typically use magnetic resonance imaging (MRI) data of various types as inputs to models

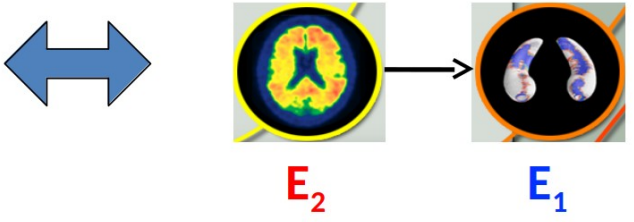
Classic example: the Event-Based Model (EBM)

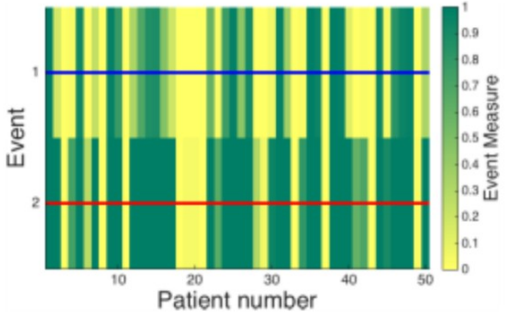
EBM describes disease progression as a sequence of abnormality **events**

Events represent the change of a biomarker from a healthy to abnormal state

Learns from cross-sectional data of any type (imaging, clinical, biofluid...)







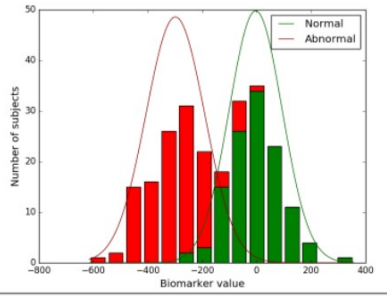
Simple example: 2 event measures

More patients have greater abnormality in Event 2 than Event 1

→Event 2 **measurably abnormal** before Event 1

Classic example: the Event-Based Model (EBM)

More formally: EBM is a generative model of observed data from a hidden sequence



data

uniform prior

$$P(X|S)=\prod_{j=1}^{N}\left[\sum_{k=0}^{Z}\left(P(k)\prod_{i=1}^{k}P(x_{ij}|E_i)\prod_{i=k+1}^{Z}P(x_{ij}|\neg E_i)\right)\right]$$

sequence

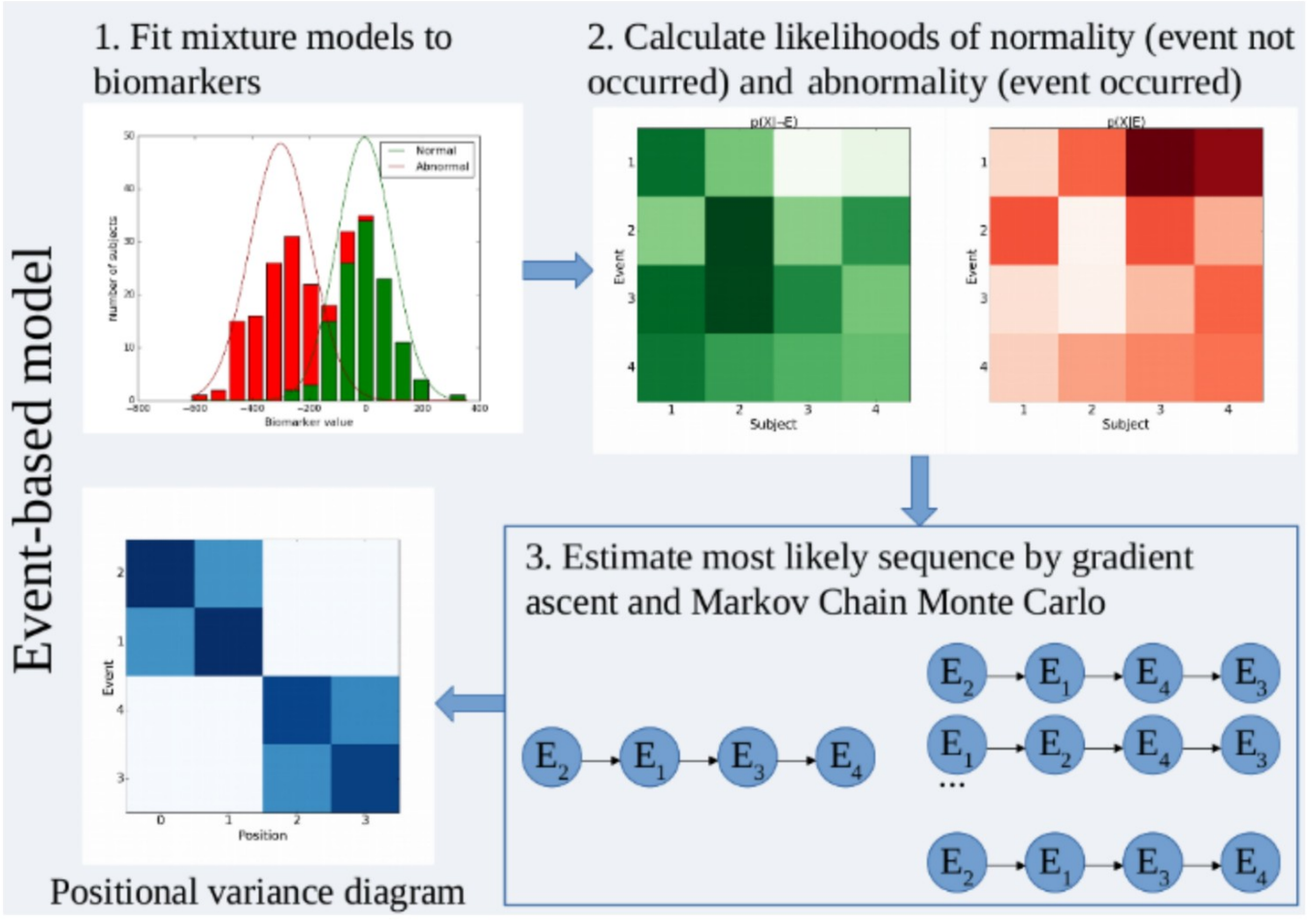
prob. Abnormal

prob. Healthy

- The EBM needs likelihood distributions for healthy and abnormal individuals

→Learn directly from data

Classic example: the Event-Based Model (EBM)

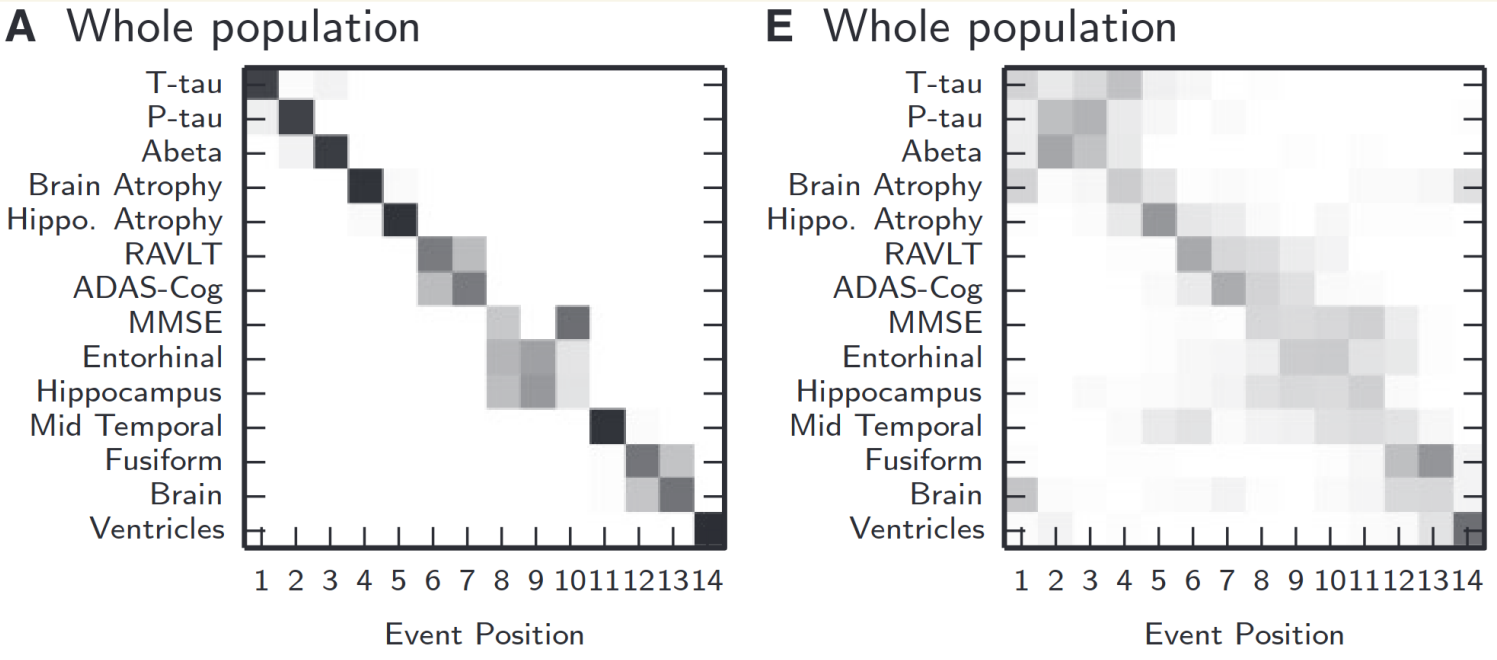


Fonteijn et al. NeuroImage (2012) DOI: 10.1016/j.neuroimage.2012.01.062

Classic example: the Event-Based Model (EBM)

UNIVERSITY OF SUSSEX

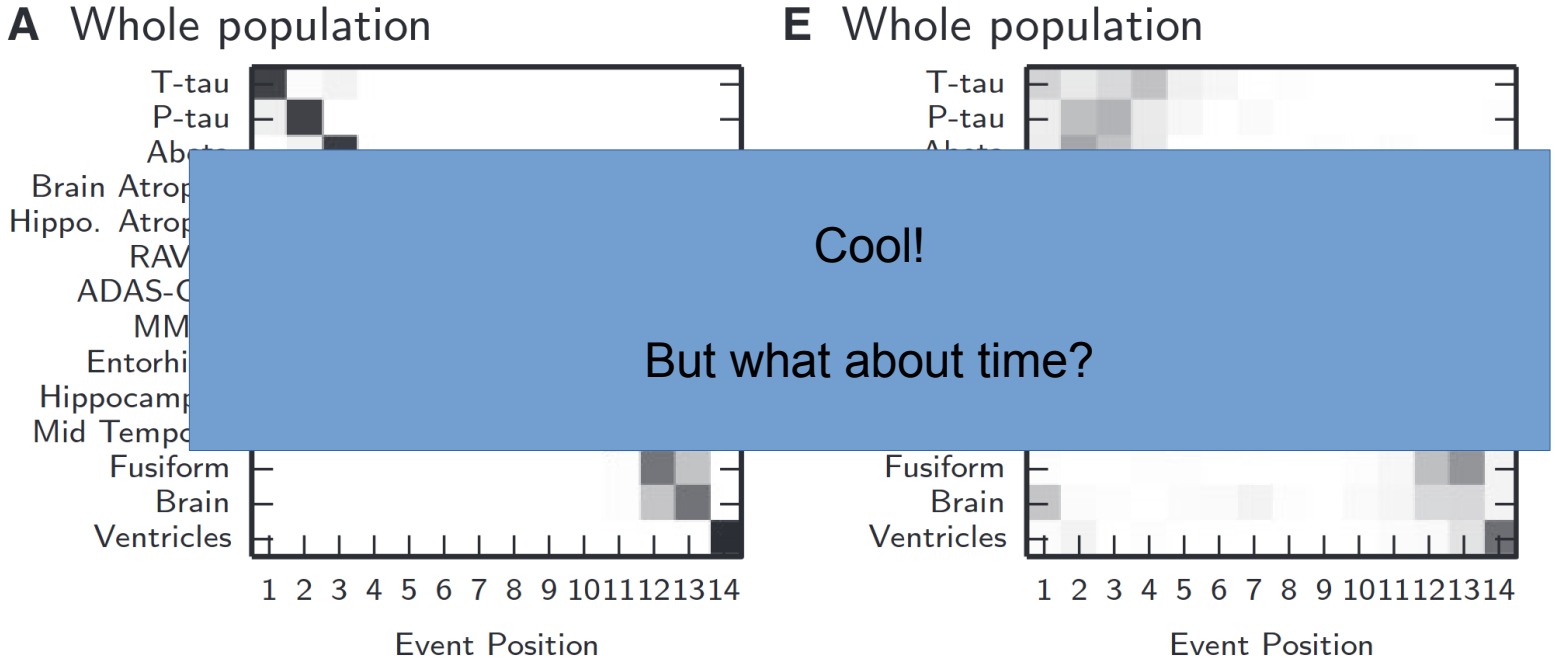
Brain 2014: 137; 2564–2577 A. L. Young *et a*



Classic example: the Event-Based Model (EBM)

UNIVERSITY OF SUSSEX

Brain 2014: 137; 2564–2577 A. L. Young *et*



Cool!

But what about time?

Continuous time hidden Markov model (CTHMM)

UNIVERSITY OF SUSSEX

What's a Continuous Time Hidden Markov Model (CTHMM)?

Markov Model: a stochastic description of a sequence of observable events, where the probability of each event depends only on the previous state (ie. the probability is *conditional* on the previous state; cf. Poisson process)

Andrey Andreyevich Markov (14th June 1856 – 20th July 1922)

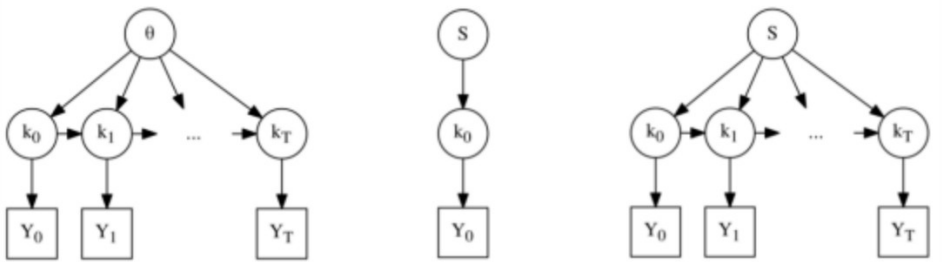
Continuous time hidden Markov model (CTHMM)

CTHMMs can be used to estimate state durations

Problem with Event-Based Model (EBM) – no time between events

Realised EBM is essentially a state-space model

→Reformulate EBM as a special CTHMM: Temporal EBM (TEBM)



(a) CTHMM (b) EBM (c) TEBM

$$P(Y|\theta,S)=\prod_{j=1}^{J}\left[\sum_{k=0}^{N}P(k_{j,t=0})\prod_{t=1}^{T_j}P(k_{j,t}|k_{j,t-1})\prod_{t=0}^{T_j}\prod_{i=1}^{k_{j,t}}P(Y_{i,j,t}|k_{j,t},\theta_i^p,S)\prod_{i=k_{j,t}+1}^{I}P(Y_{i,j,t}|k_{j,t},\theta_i^c,S)\right]$$

Wijeratne et al. Imaging Neuroscience 2023

Temporal Event-Based Model: provenance

UNIVERSITY OF SUSSEX

- Guided by this 2014 paper on a continuous-time hidden Markov disease progression model

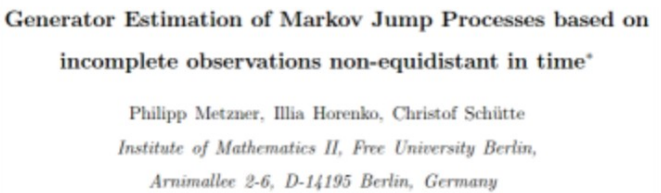
Unsupervised Learning of Disease Progression Models

Xiang Wang
IBM Research
Yorktown Heights, NY
wangxi@us.ibm.com

David Sontag
New York University
New York, NY
dsontag@cs.nyu.edu

Fei Wang
IBM Research
Yorktown Heights, NY
fwang@us.ibm.com

- Found this 2007 paper deriving a variable-interval continuous-time hidden Markov model



Generator Estimation of Markov Jump Processes based on incomplete observations non-equidistant in time*

Philipp Metzner, Illia Horenko, Christof Schütte
Institute of Mathematics II, Free University Berlin, Arnimallee 2-6, D-14195 Berlin, Germany

- Complemented by various algorithmic implementations in this 2016 paper

Efficient Learning of Continuous-Time Hidden Markov Models for Disease Progression

Yu-Ying Liu, Shuang Li, Fuxin Li, Le Song, and James M. Rehg

- Can calculate Q(t) using eigendecomposition (fast, requires Q diagonalisable) or directly using Padé method for matrix exponential (slow)

Example: the Temporal Event-Based Model (TEBM)

UNIVERSITY OF SUSSEX

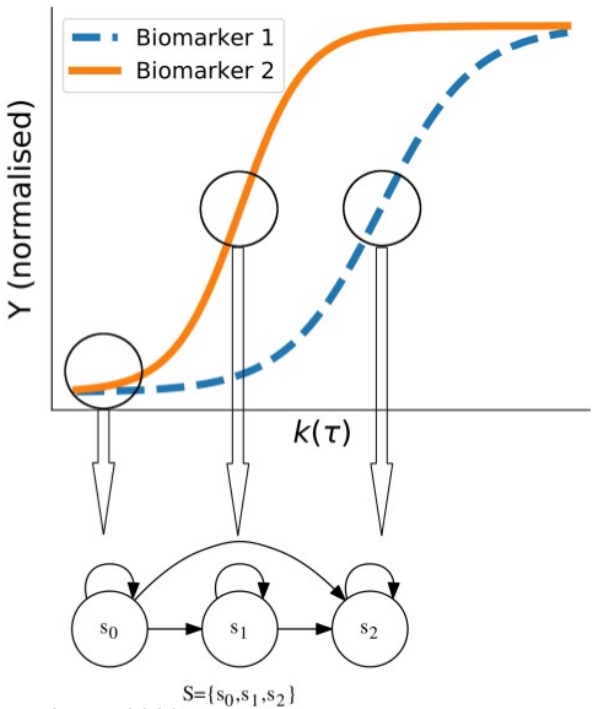
The TEBM is a generative model of observations (e.g., biomarkers) conditional on latent variables (e.g., disease states)

(2021). To formulate the TEBM, we make three main assumptions: *i*) monotonic biomarker dynamics at the group level; *ii*) a consistent event sequence across the whole population; and *iii*) Markov stage transitions at the individual level. We can write the TEBM joint distribution as a hierarchical Bayesian model using the chain rule:

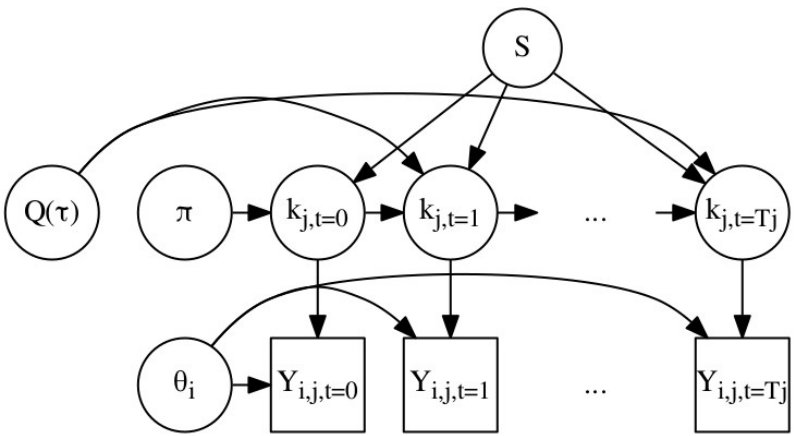
$$\begin{aligned} P(S, \Theta_i, k_{j,t}, Y_{i,j,t}) &= P(S, \Theta_i) \cdot P(Y_{i,j,t}, k_{j,t} | \Theta_i, S) \\ &= P(S) \cdot P(\Theta_i) \cdot P(k_{j,t} | S, \pi, Q) \cdot P(Y_{i,j,t} | k_{j,t}, \theta_i, S). \end{aligned} \quad (1)$$

Example: the Temporal Event-Based Model (TEBM)

The TEBM is a generative model of observations (e.g., biomarkers) conditional on latent variables (e.g., disease states)

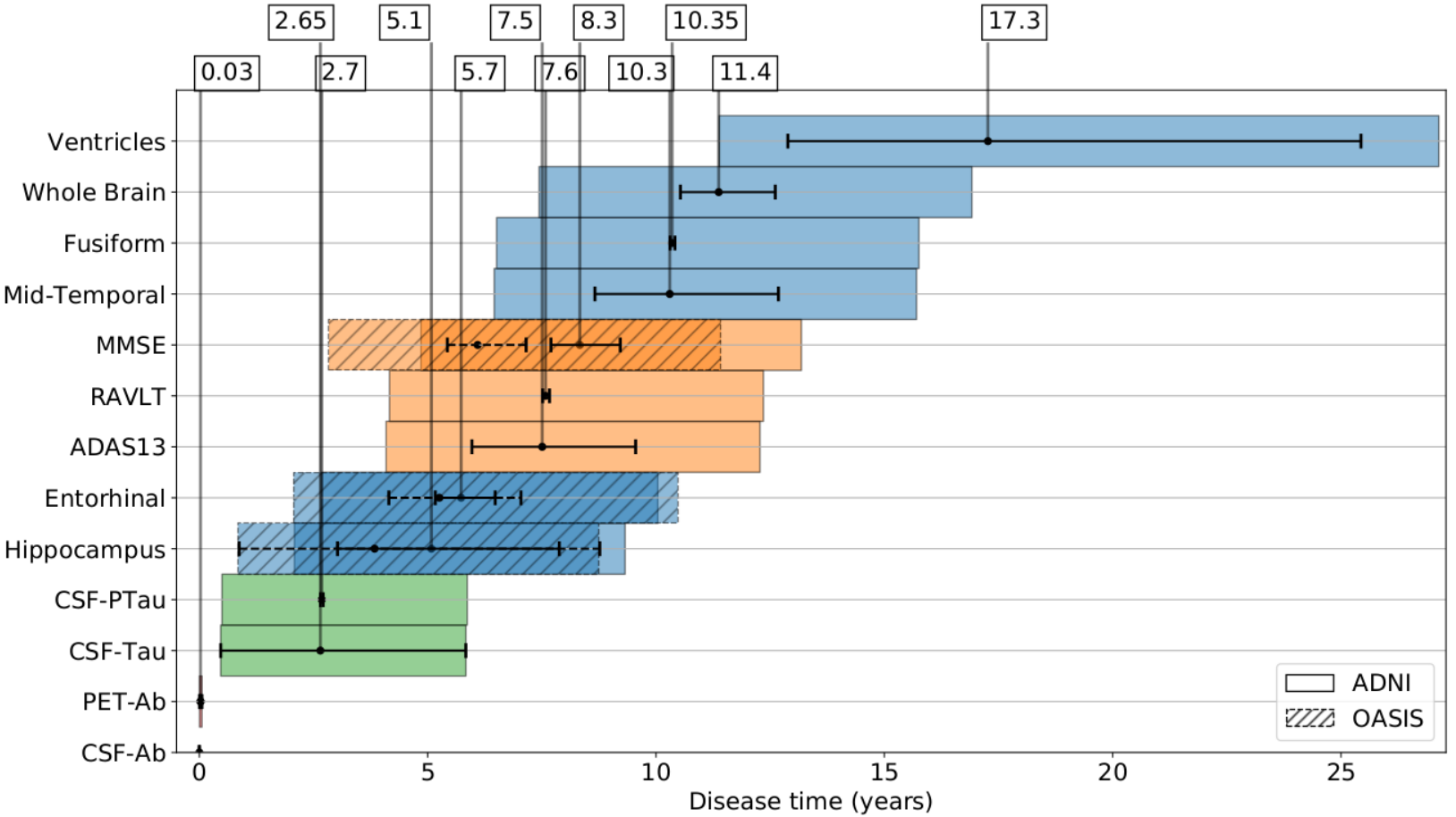


Graphical model



Wijeratne et al. Imaging Neuroscience 2023

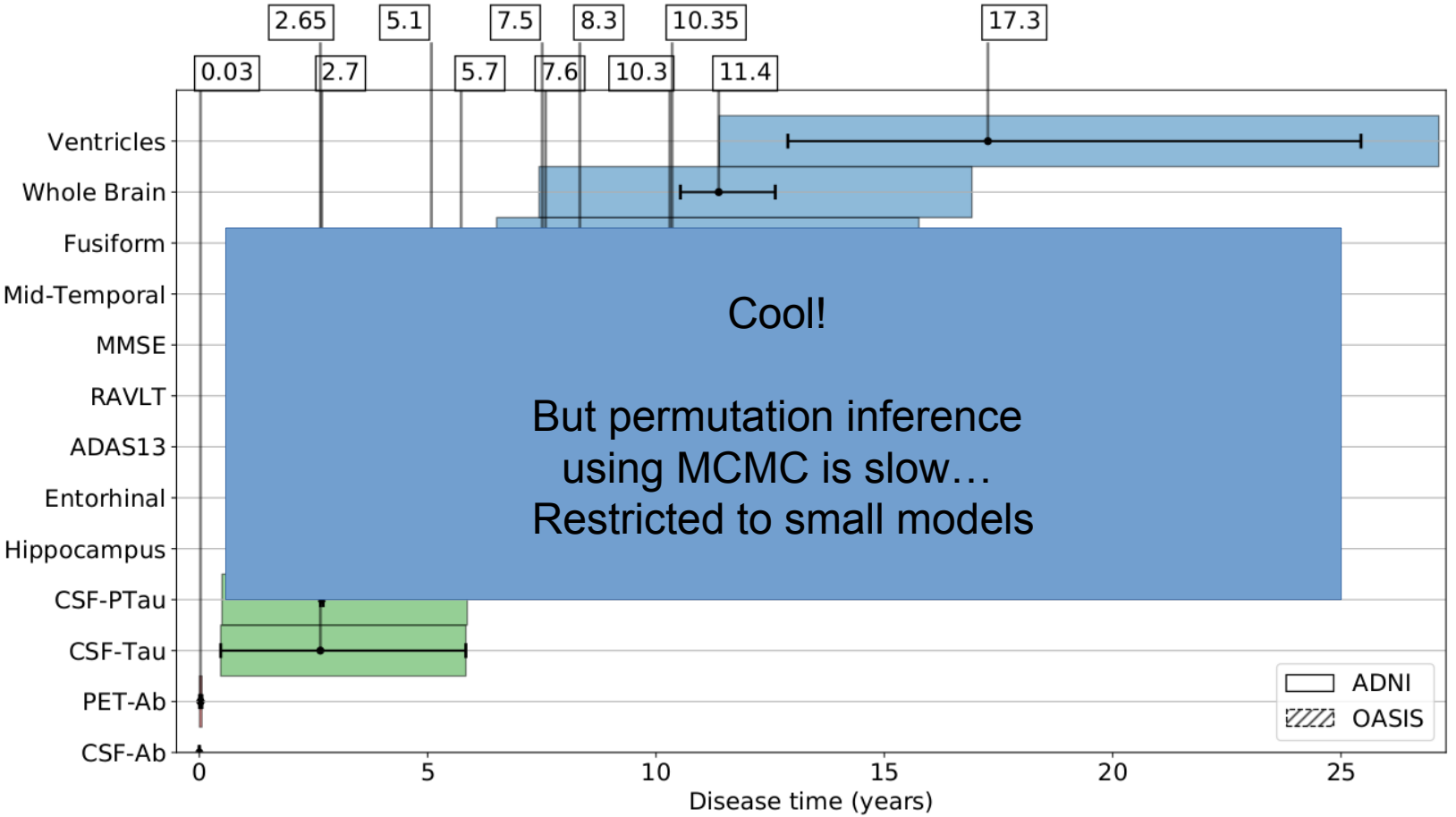
Application: learning Alzheimer's disease timeline



First timeline of biomarker events in Alzheimer's disease

Wijeratne et al. Imaging Neuroscience 2023

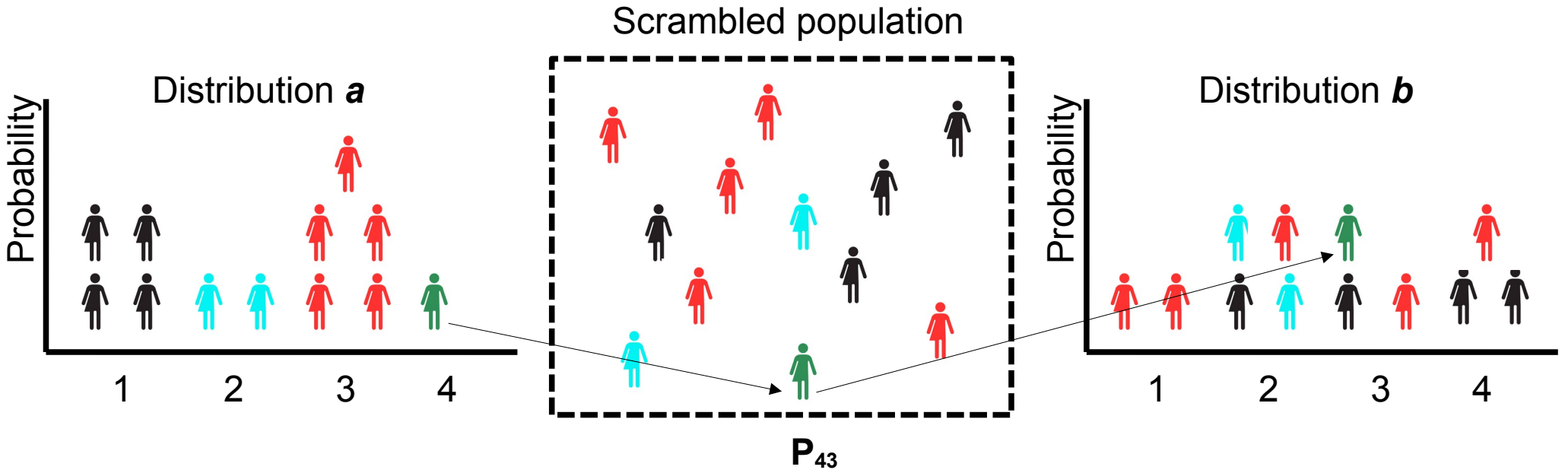
Application: learning Alzheimer's disease timeline



First timeline of biomarker events in Alzheimer's disease

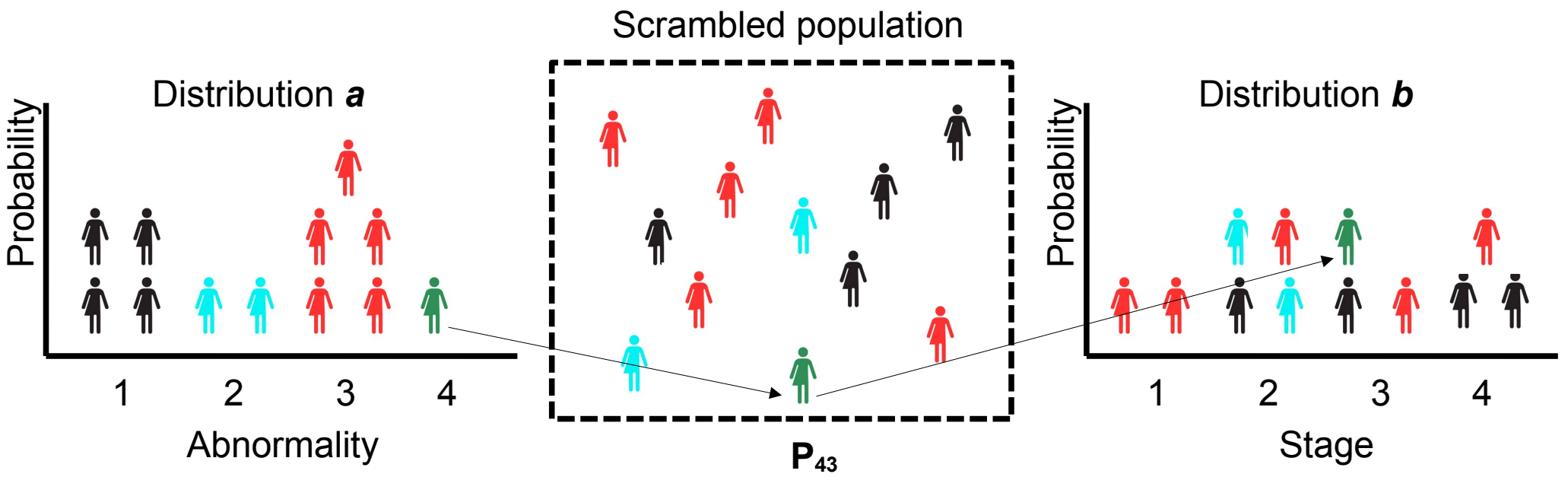
Wijeratne et al. Imaging Neuroscience 2023

Optimal transport



Find the optimal mapping ***P*** that minimises the cost of transporting probability distribution ***a*** to ***b***

Idea: optimal transport for disease trajectory modelling



Transport people to their optimal (latent) stage along an event-based disease trajectory

Variational event-based model: key idea

$$\mathcal{B}_N = \Big\{ X : \quad X_{m,n} \geq 0 \quad \forall\, m, n \in 1, \ldots, N;$$
$$\sum_{n=1}^{N} X_{m,n} = 1 \quad \forall\, m \in 1, \ldots, N;$$
$$\sum_{m=1}^{N} X_{m,n} = 1 \quad \forall\, n \in 1, \ldots, N \Big\}.$$

These linear row- and column-normalization constraints restrict $\mathcal{B}_N$ to a $(N-1)^2$ dimensional subset of $\mathbb{R}^{N \times N}$.

Reframe the discrete sequence *s* as a permutation matrix *S* belonging to the Birkhoff polytope

Wijeratne & Alexander, NeurIPS 2024, doi: arXiv:2410.14388

Aside: Birkhoff polytopes

What's a Birkhoff polytope?

https://en.wikipedia.org/wiki/Birkhoff_polytope
Paffenholz, 2013. arXiv:1304.3948

"The Birkhoff polytope, B, is the convex polytope in R^N (where N = n^2) whose points are the doubly stochastic matrices, X, i.e., the n×n matrices whose entries are non-negative real numbers and whose rows and columns each add up to 1."

→Maps permutation matrices, X, to geometric objects, B

→The top vertex in Figure 2.1 would have a permutation matrix like:

|1, 0, 0|
|0, 1, 0| = X
|0, 0, 1|

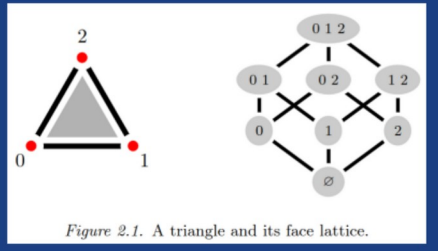


Figure 2.1. A triangle and its face lattice.

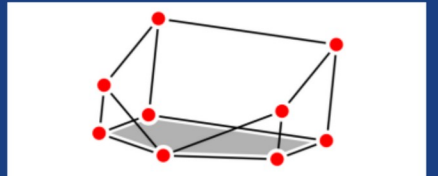
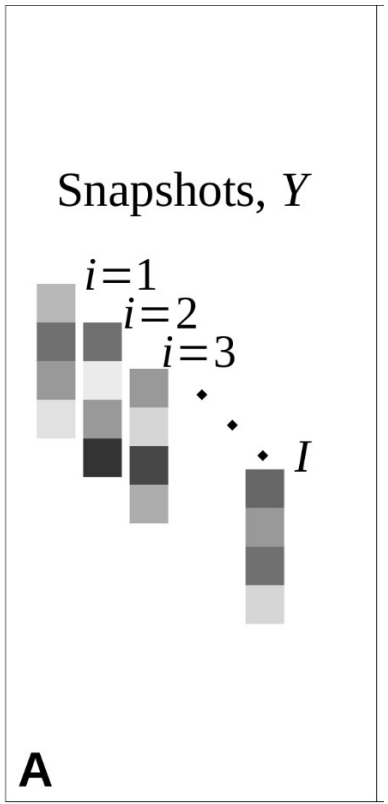


Figure 2.2. The wedge over a vertex of a pentagon.

Reframe the discrete sequence *s* as a permutation matrix *S* belonging to the Birkhoff polytope

Wijeratne & Alexander, NeurIPS 2024, doi: arXiv:2410.14388

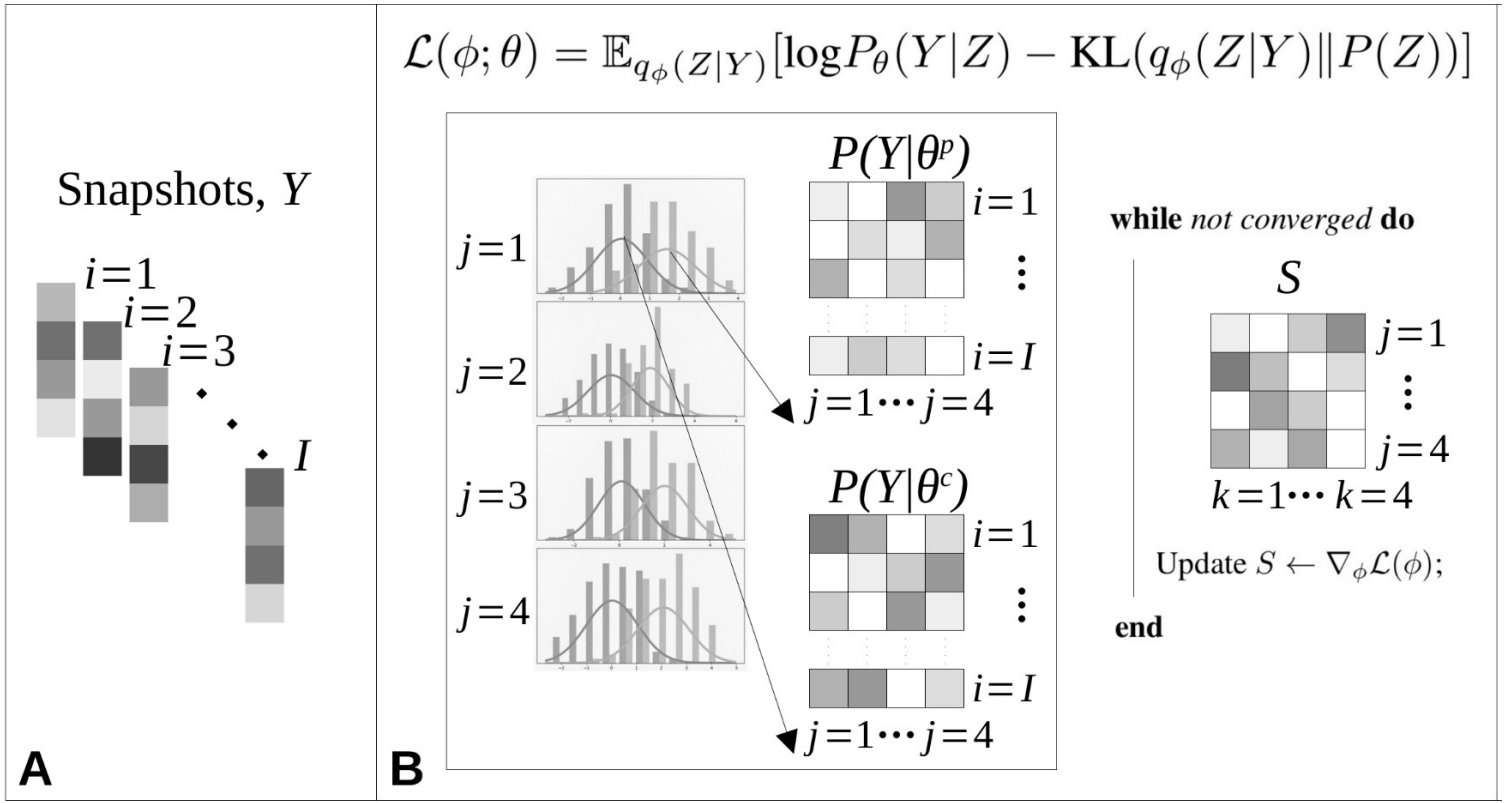
Variational event-based model



Input “snapshots” (single observations) from different individuals

Wijeratne & Alexander, NeurIPS 2024, doi: arXiv:2410.14388

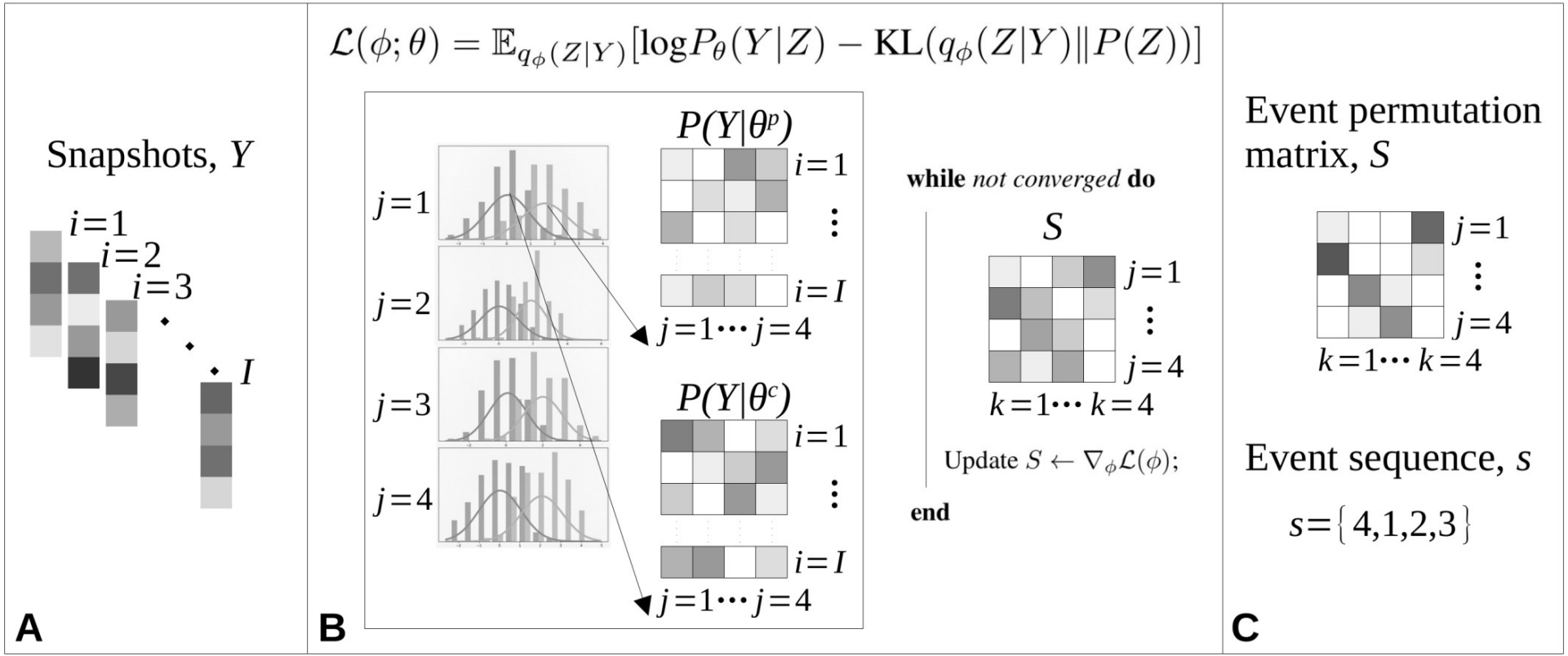
Variational event-based model



Model likelihood specified by distributions “normality” and “abnormality”

Wijeratne & Alexander, NeurIPS 2024, doi: arXiv:2410.14388

Variational event-based model



Learn optimal permutation of hidden events and corresponding event sequence

Wijeratne & Alexander, NeurIPS 2024, doi: arXiv:2410.14388

Variational event-based model: key ingredients

$$K(X/\tau) = \underset{S \in \mathcal{B}_{\mathcal{N}}}{\operatorname{argmax}} \langle S, X \rangle_F + \tau H(S).$$

Entropy-regularised optimal transport problem

Cuturi. NeurIPS 2013, doi: arXiv.1306.0895

$$M(X) = \begin{bmatrix} e_{s(0)} \\ \vdots \\ e_{s(N)} \end{bmatrix} = \lim_{\tau \to 0} K(X/\tau).$$

Hard permutation from Sinkhorn-Knopp operator

Sinkhorn & Knopp. Pacific Journal of Mathematics, 1967

$$G(X, \tau) \sim K((X + \epsilon)/\tau).$$

Gumbel-Sinkhorn distribution

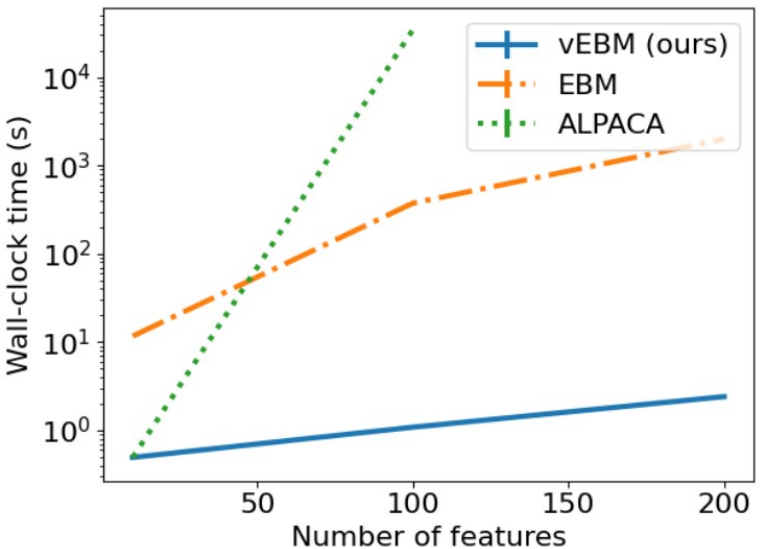
Mena et al. ICLR 2018, doi: arXiv.1802.08665

$$\begin{aligned} \log P(Y) \geq \mathcal{L}(\phi; \theta) &= \mathbb{E}_{q_\phi(Z|Y)}[\log P_\theta(Y|Z) - \mathrm{KL}(q_\phi(Z|Y) \| P(Z))] \\ &= \mathbb{E}_{q_\phi(Z|Y)}[\log P_\theta(Y|Z) - \mathrm{KL}(G_\phi(X, \tau) \| G(X = 0, \tau_{\text{prior}}))]. \end{aligned}$$

Enable differentiability by parametrising prior & posterior using Gumbel-Sinkhorn distribution

Wijeratne & Alexander, NeurIPS 2024, doi: arXiv:2410.14388

Fast inference

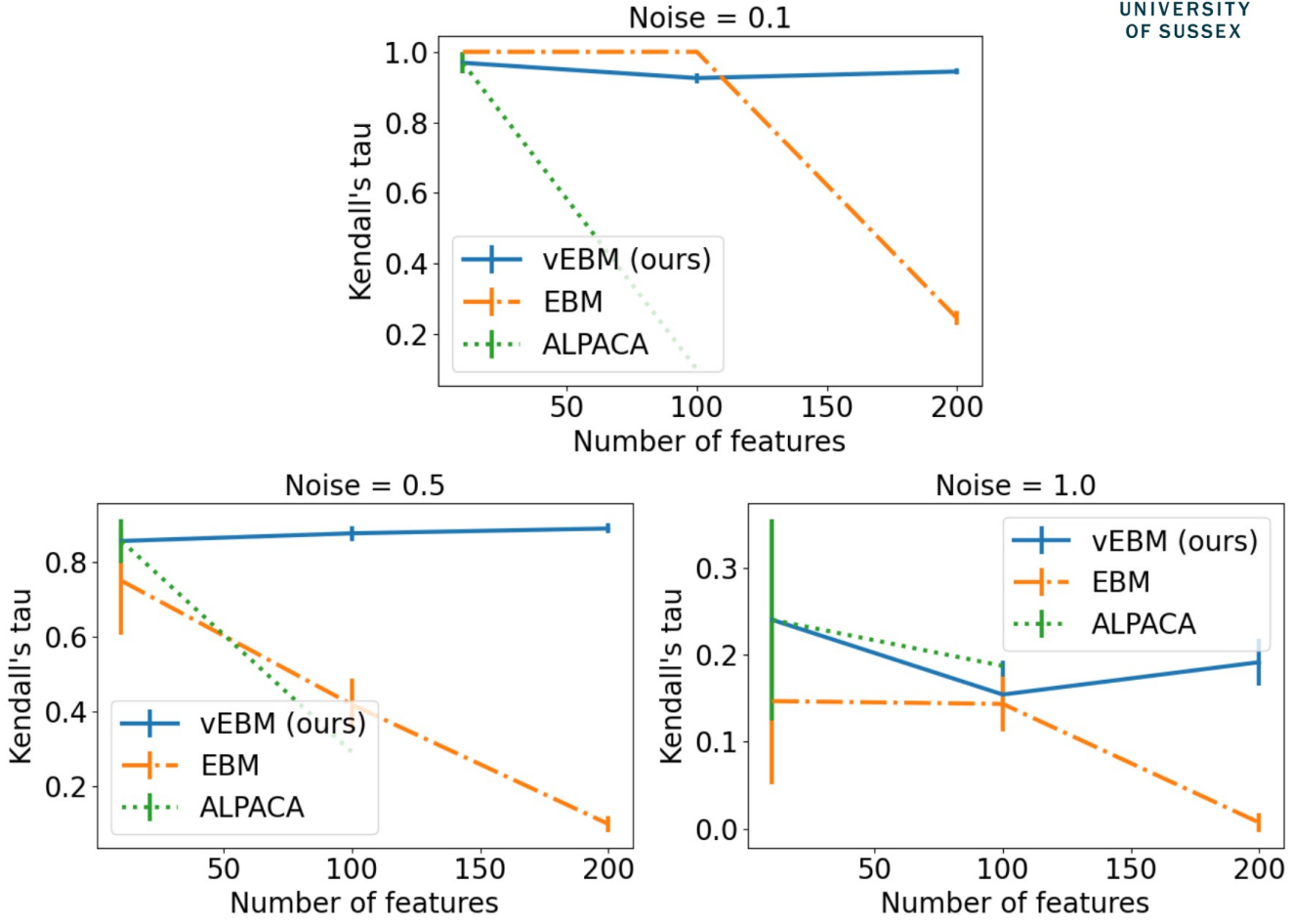


>1000 x faster than baselines,
which use maximum likelihood

Wijeratne & Alexander, NeurIPS 2024, doi: arXiv:2410.14388

Robust to noise

Maintains inference accuracy with increasing noise



Wijeratne & Alexander, NeurIPS 2024, doi: arXiv:2410.14388

Real data: pixel-level atrophy using tensor-based morphometry data

UNIVERSITY OF SUSSEX

Use ADNI dataset “TBM Jacobian Maps MDT-SC”

Cross-sectional TBM maps from 816 individuals (299 controls, 399 mild cognitive impairment, 188 AD)

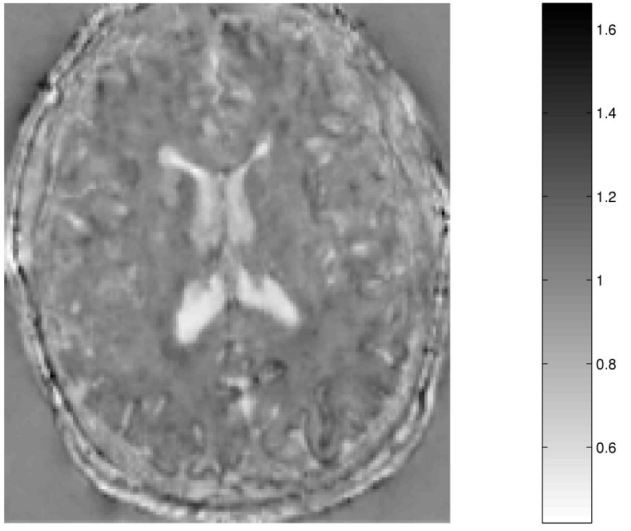
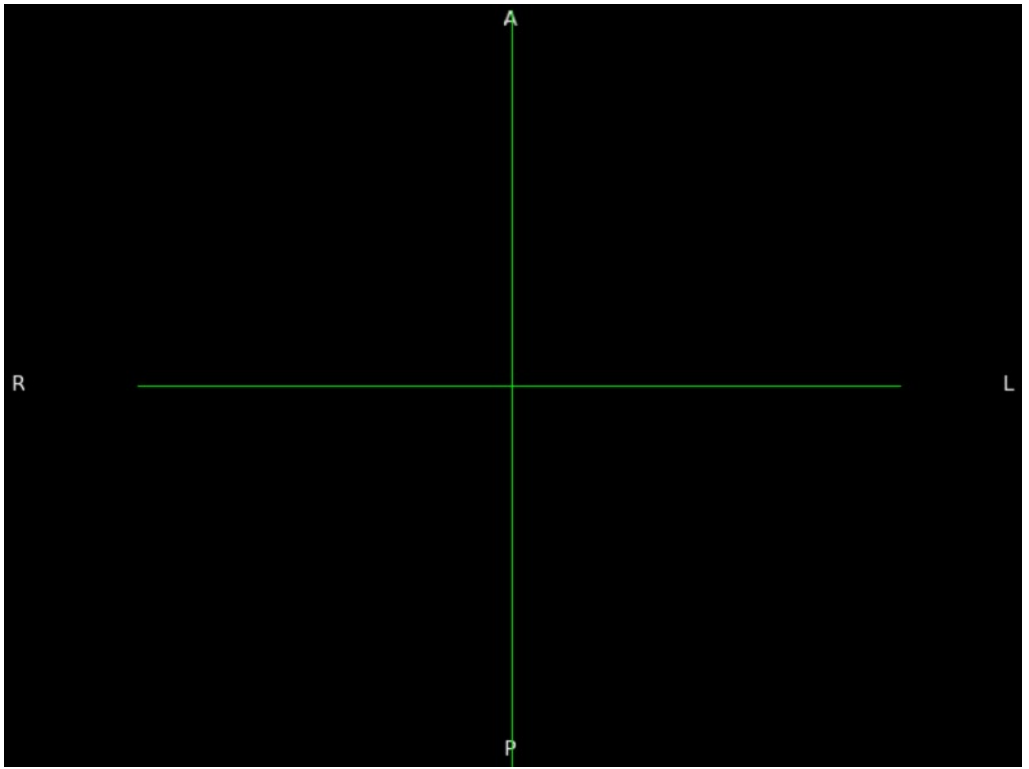


Figure 6.4: This figure illustrates the volume changes estimated by warping together the images shown in Figure 6.3. The relative volumes are the Jacobian determinants of the deformation field. Smaller determinants are obtained when a region of the template maps to a smaller region in the source image. In this example, they represent regions that have expanded between the early and late scans. Regions where there are no measurable volume changes have Jacobian determinants with a value of one.

https://www.fil.ion.ucl.ac.uk/spm/doc/books/hbf2/pdfs/Ch6.pdf

*Jacobian Map == Deformation tensor

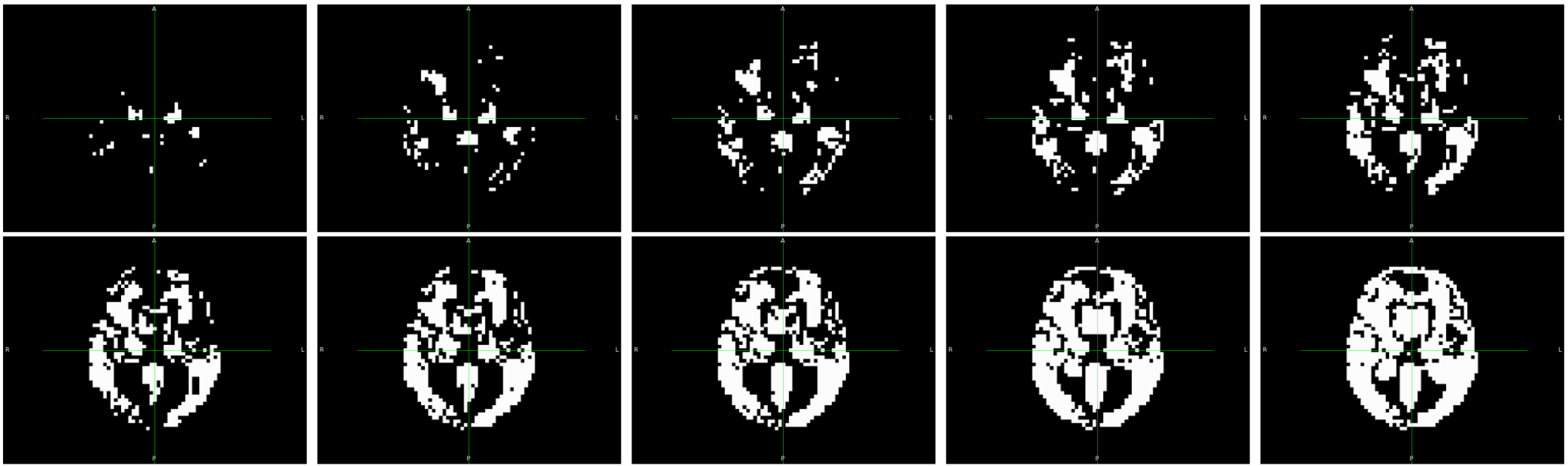
Pixel-level progression Alzheimer's disease (AD)



New insights into tissue-level AD progression

Pixel-level progression Alzheimer's disease (AD)

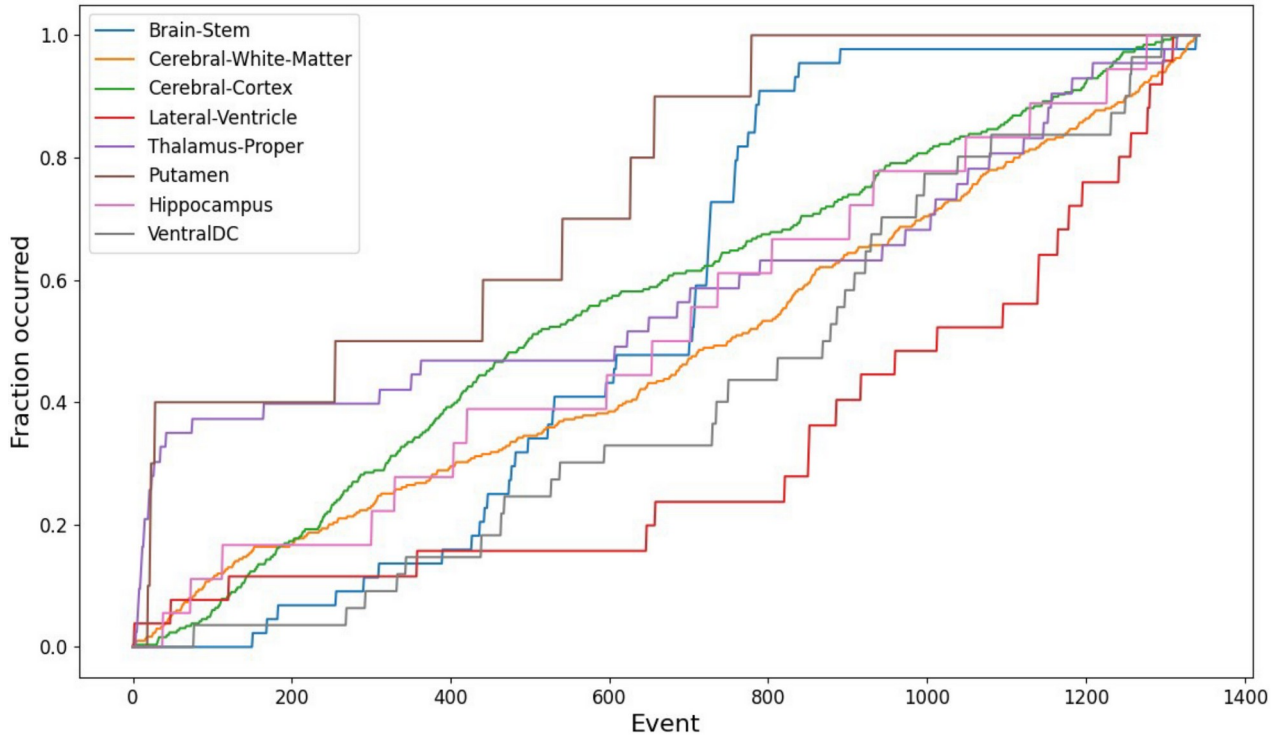
UNIVERSITY OF SUSSEX



New insights into tissue-level AD progression

Wijeratne & Alexander, NeurIPS 2024, doi: arXiv:2410.14388

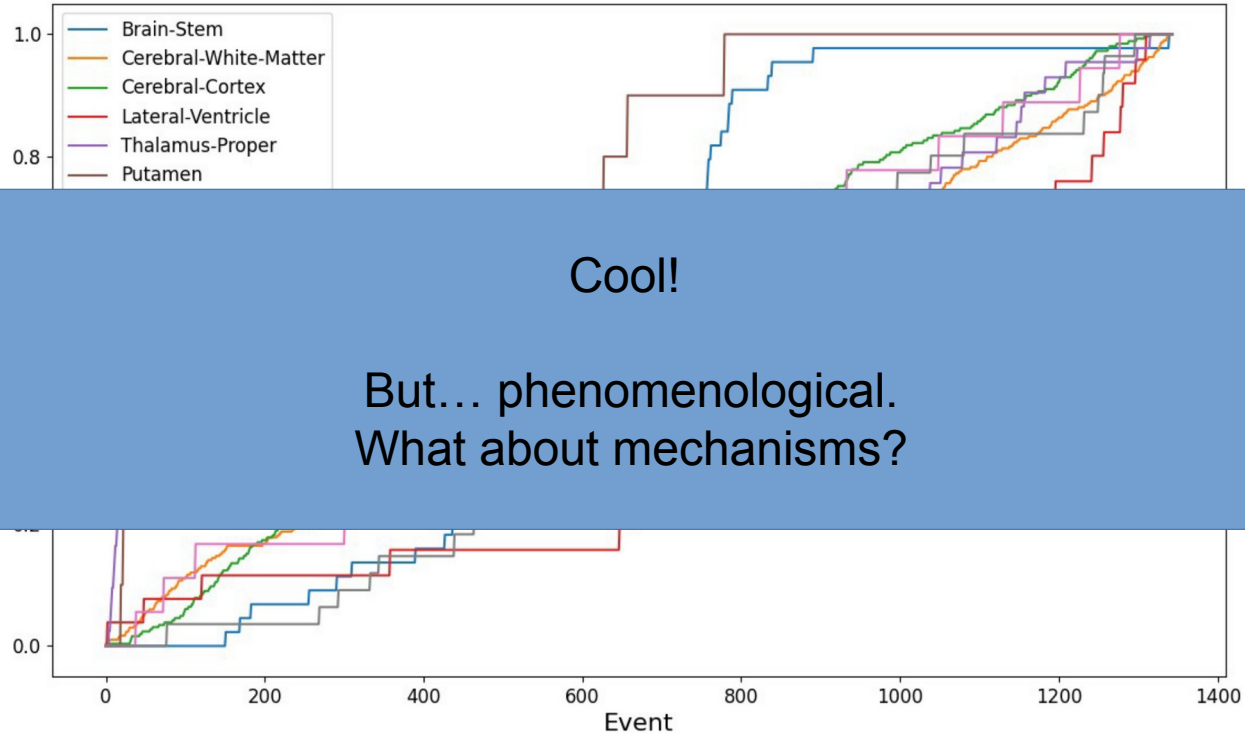
Pixel-level progression Alzheimer's disease (AD)



New insights into tissue-level AD progression

Wijeratne & Alexander, NeurIPS 2024, doi: arXiv:2410.14388

Pixel-level progression Alzheimer's disease (AD)



Cool!

But… phenomenological.
What about mechanisms?

New insights into tissue-level AD progression

Wijeratne & Alexander, NeurIPS 2024, doi: arXiv:2410.14388

Mechanistic modelling of disease propagation

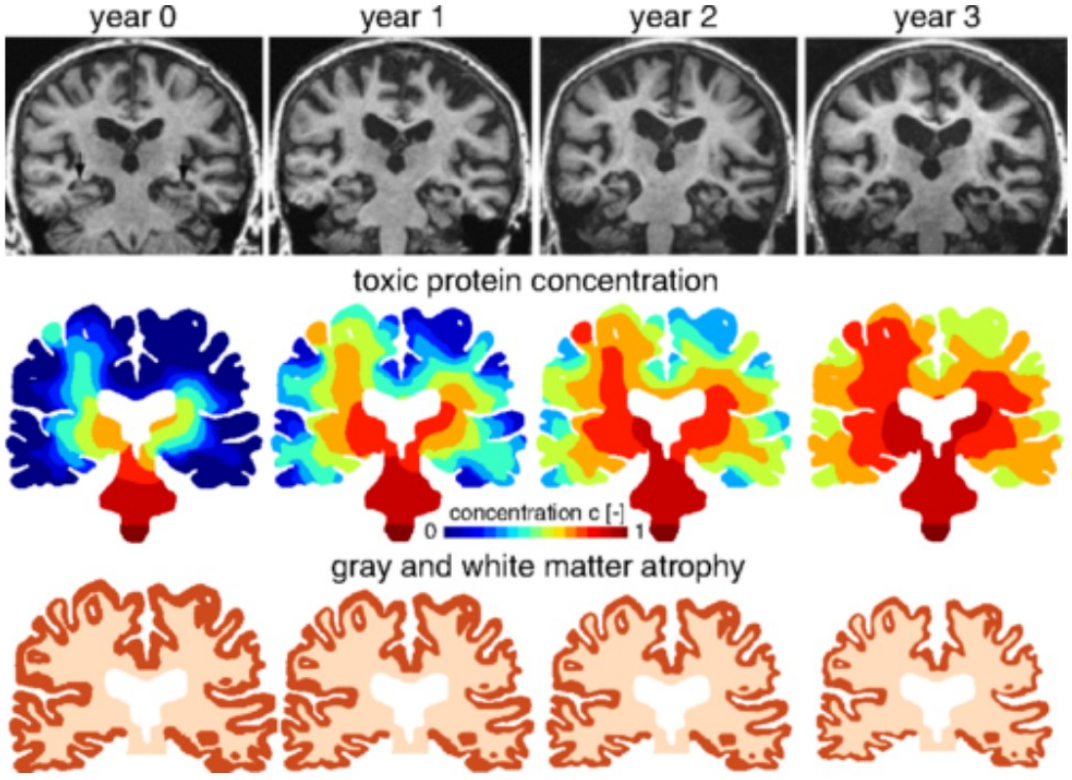


Figure 1: **Top Image:** MRI scans of an Alzheimer's patient showing yearly progression, highlighting increased hippocampal atrophy, enlarged ventricles, and widening cortical sulci. **Middle Image:** Annual increase in toxic protein concentration, starting from the brain stem. **Bottom Image:** Simulated annual atrophy patterns, correlating activation time and toxic protein concentration. Weickenmeier, J., Kuhl, E., & Goriely, A. (2018). Multiphysics of prionlike diseases: Progression and atrophy. Physical review letters , 121 (15), 158101. (https://creativecommons.org/licenses/by/4.0/)

Courtesy of Sanduni Pinnawala

Mechanistic modelling of disease propagation

UNIVERSITY OF SUSSEX

BrainPhys

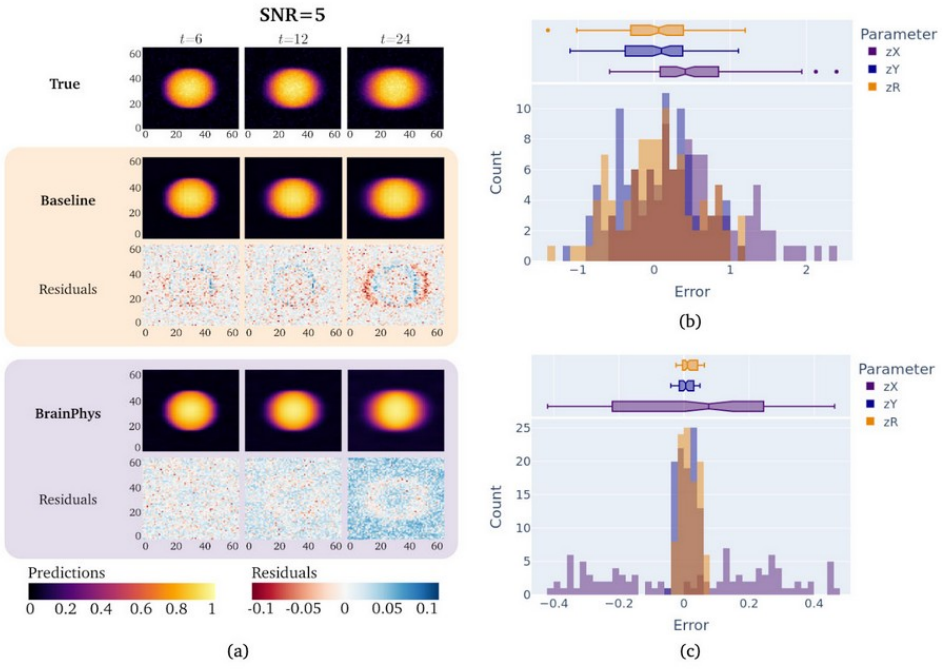
Challenges in studying mechanisms of neurodegeneration

- ➔ Incomplete knowledge of disease biology
- ➔ Scarcity of longitudinal imaging data

Limitations of conventional deep learning

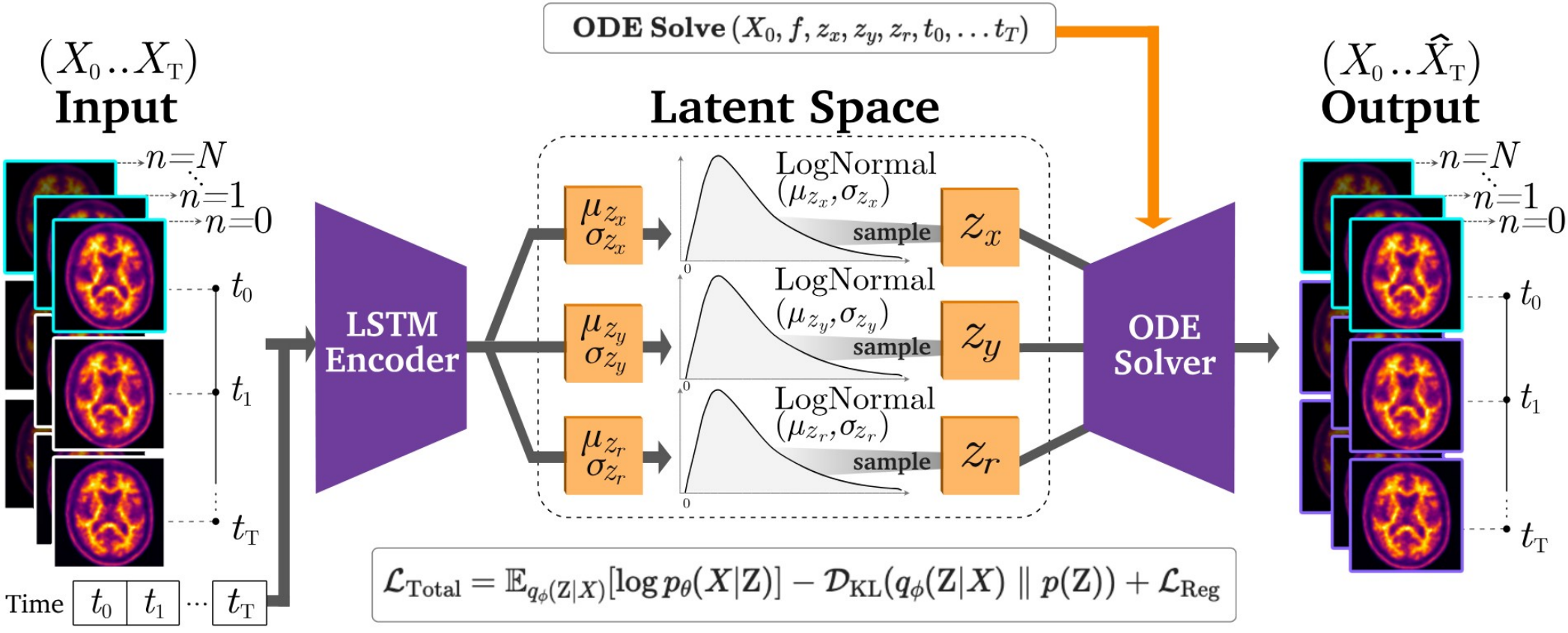
- ➔ Lack interpretability (act as black boxes)
- ➔ Demand large datasets

Physics informed machine learning combines data-driven methods with physical constraints to **improve interpretability**



Courtesy of Sanduni Pinnawala

Mechanistic modelling of disease propagation



Courtesy of Sanduni Pinnawala

Mechanistic modelling of disease propagation

UNIVERSITY OF SUSSEX

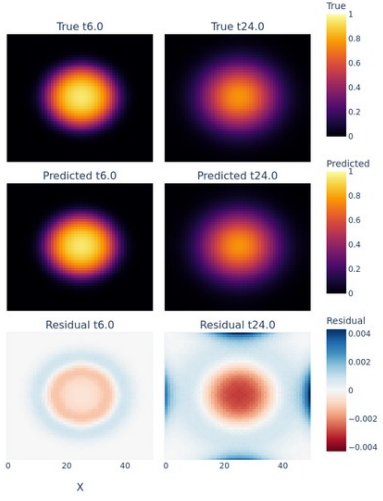
BrainPhys

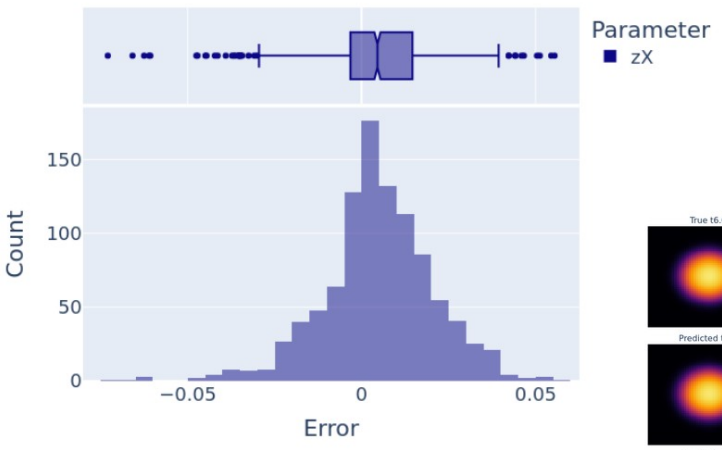
$$\frac{\partial c}{\partial t} = z_X \nabla^2 c$$

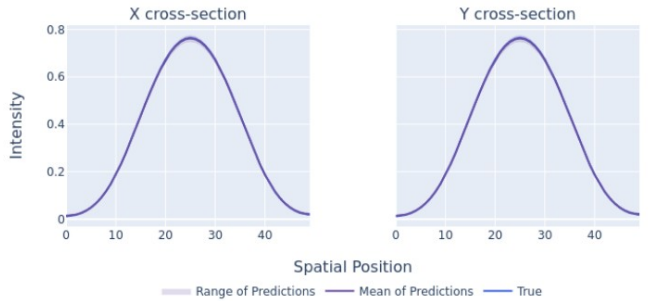
$$\frac{\partial c}{\partial t} = D\nabla^2 c$$

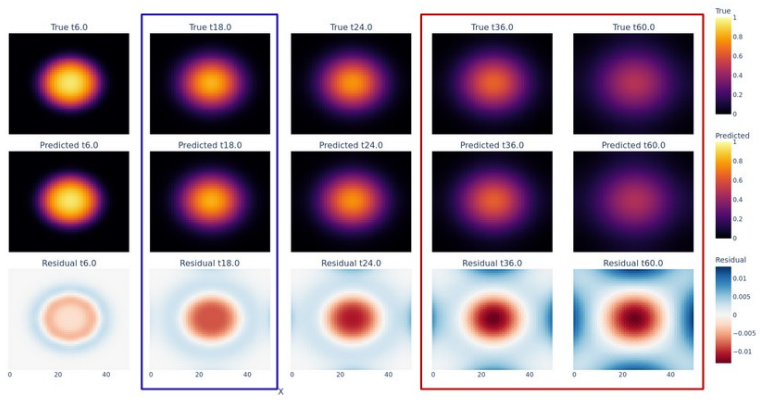
where:

- $\frac{\partial c}{\partial t}$ is the time derivative of c, describing the accumulation and propagation of misfolded protein concentration.
- $D\nabla^2 c$ represents diffusion









Courtesy of Sanduni Pinnawala

Mechanistic modelling of disease propagation

UNIVERSITY OF SUSSEX

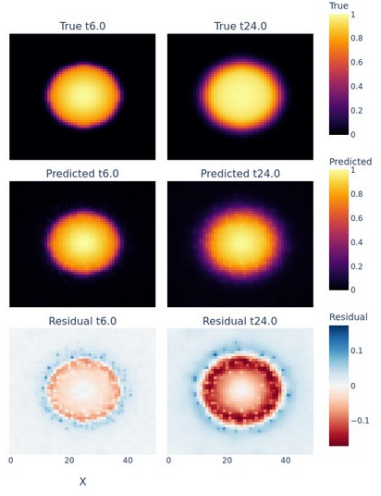
BrainPhys

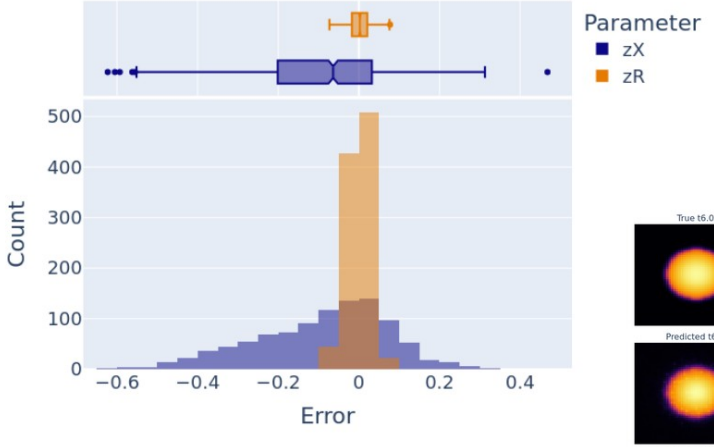
$$\frac{\partial c}{\partial t} = z_X \nabla^2 c + z_R f(c)$$

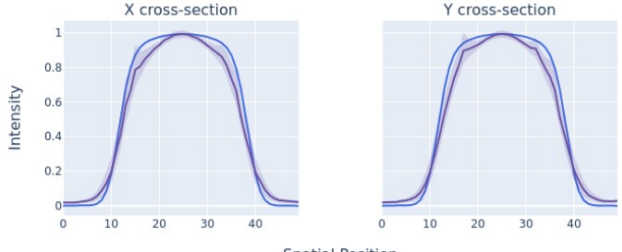
$$\frac{\partial c}{\partial t} = D\nabla^2 c + rc(1-c)$$

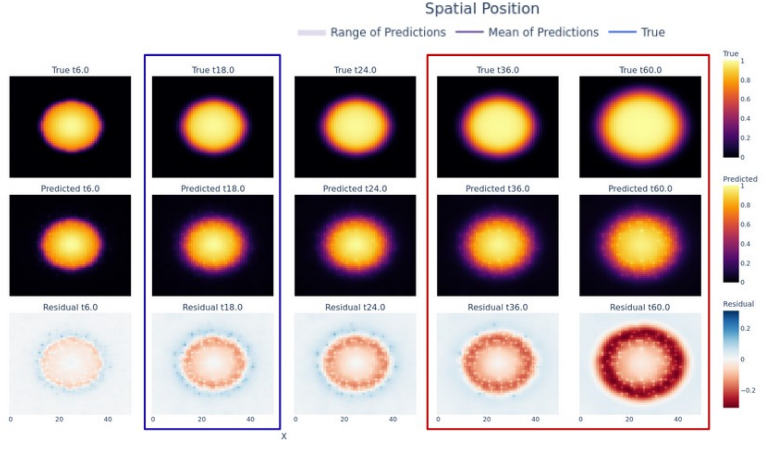
where:

- $\frac{\partial c}{\partial t}$ is the time derivative of c, describing the accumulation and propagation of misfolded protein concentration.
- $D\nabla^2 c$ represents diffusion
- $rc(1-c)$ represents the reaction dynamics









Courtesy of Sanduni Pinnawala

Detecting model mis-specification

UNIVERSITY OF SUSSEX

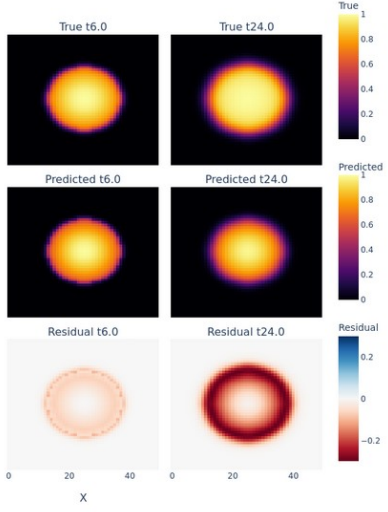
BrainPhys

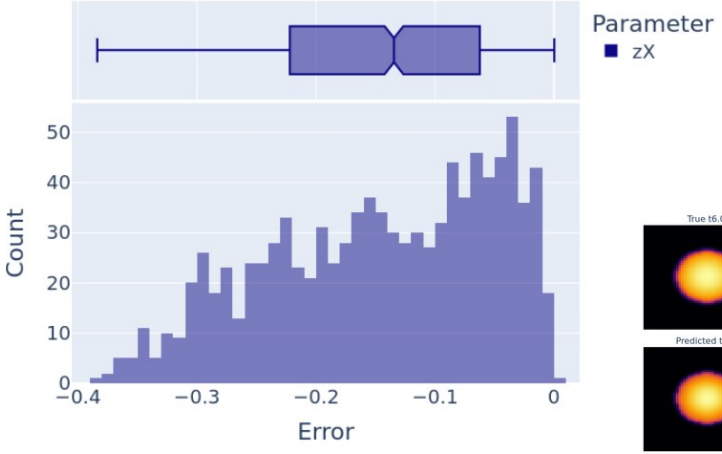
$$\frac{\partial c}{\partial t} = z_X \nabla^2 c$$

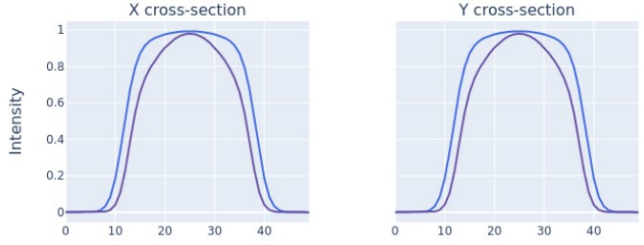
$$\frac{\partial c}{\partial t} = D\nabla^2 c + rc(1-c)$$

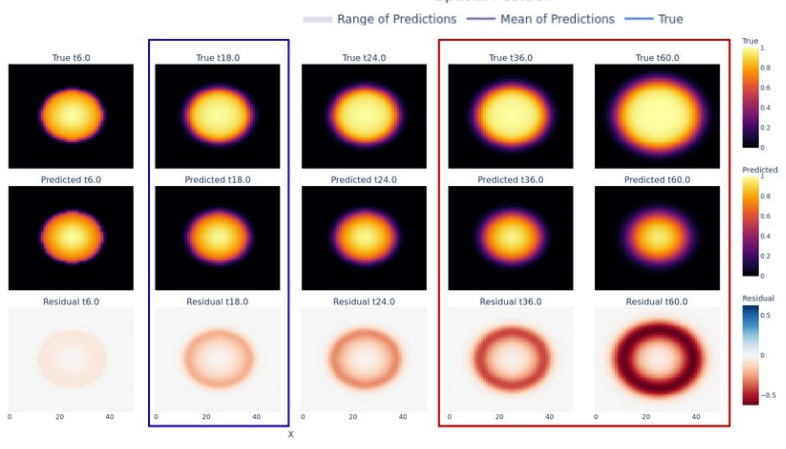
where:

- $\frac{\partial c}{\partial t}$ is the time derivative of c, describing the accumulation and propagation of misfolded protein concentration.
- $D\nabla^2 c$ represents diffusion
- $rc(1-c)$ represents the reaction dynamics









Courtesy of Sanduni Pinnawala

Detecting model mis-specification

UNIVERSITY OF SUSSEX

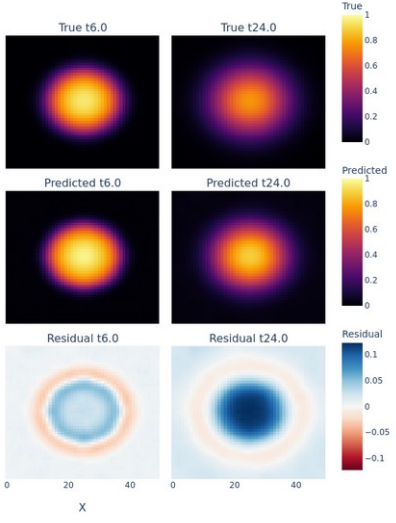
BrainPhys

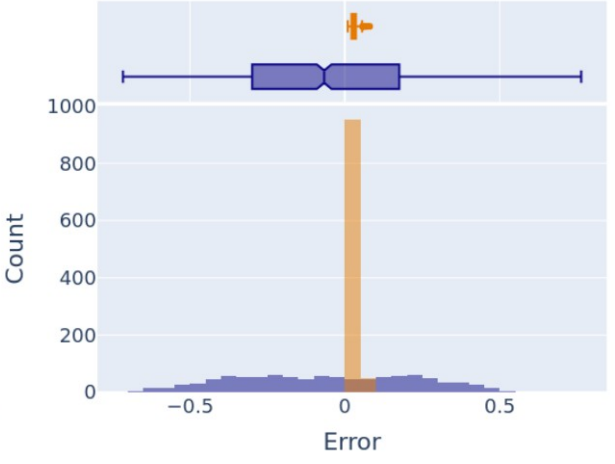
$$\frac{\partial c}{\partial t} = z_X \nabla^2 c + z_R f(c)$$

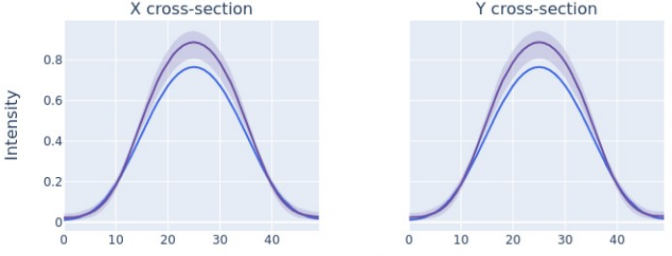
$$\frac{\partial c}{\partial t} = D\nabla^2 c$$

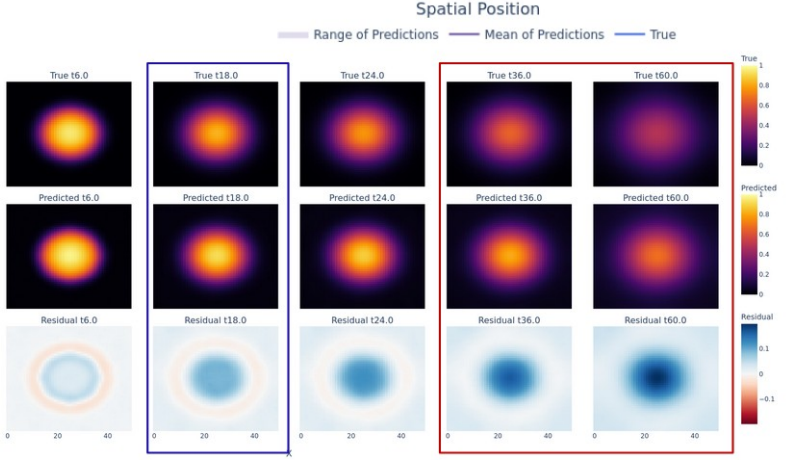
where:

- $\frac{\partial c}{\partial t}$ is the time derivative of c, describing the accumulation and propagation of misfolded protein concentration.
- $D\nabla^2 c$ represents diffusion









Courtesy of Sanduni Pinnawala

Summary and next steps

UNIVERSITY OF SUSSEX

- Probabilistic models can be used to learn hidden information in diseases
- Optimal transport formulation offers many benefits over standard maximum likelihood approaches
- Physics-guided/integrated/informed… machine learning gives flexibility
  - But… identifiability, validation (ground truth?), assumptions, ...

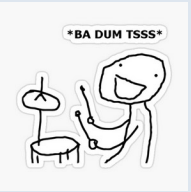
Many interesting theoretical avenues…

Summary and next steps

- Probabilistic models can be used to learn hidden information in diseases
- Optimal transport formulation offers many benefits over standard maximum likelihood approaches
- Physics-guided/integrated/informed… machine learning gives flexibility
  - But… identifiability, validation (ground truth?), assumptions, ...

Many interesting theoretical avenues…

Just need to pick the optimal route!



Backup: uncertainty