

## A vertex clustering model of disease progression

R. V. Marinescu<sup>1</sup>, A. Eshaghi<sup>1,2</sup>, M. Lorenzi<sup>1,4</sup>, A. L. Young<sup>1</sup>, N. P. Oxtoby<sup>1</sup>,  
S. Garbarino<sup>1</sup>, T. J. Shakespeare<sup>3</sup>, S. J. Crutch<sup>3</sup> and D. C. Alexander<sup>1</sup>

<sup>1</sup>Centre for Medical Image Computing, University College London, UK

<sup>2</sup>Queen Square MS Centre, UCL Institute of Neurology, London

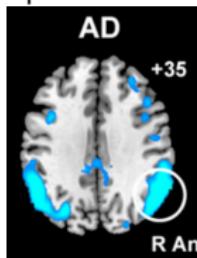
<sup>3</sup>Dementia Research Centre, UCL Institute of Neurology, University College London, UK

<sup>4</sup>University of Côte d'Azur, Inria Sophia Antipolis, Asclepios Research Project

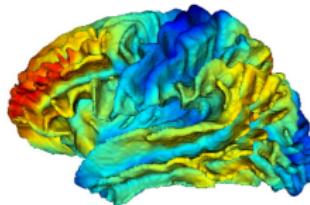


**Aim:** Build a disease progression model of pathology over the brain that:

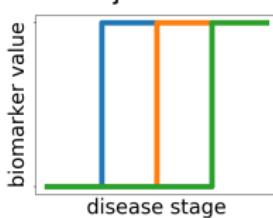
Avoids pre-defined ROI  
parcellation



Avoids simplistic spatial  
correlation structure



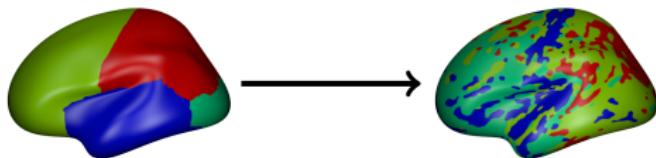
Avoids simplistic biomarker  
trajectories



This leads to a technique that simultaneously:

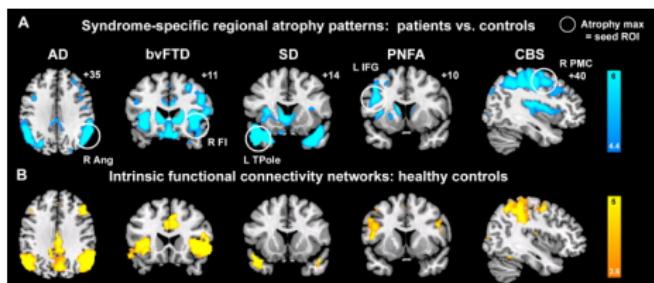
- parcellates the brain into disconnected components that undergo similar progression
- estimates biomarker trajectories

**Aim:** Move from ROI-based analysis to vertexwise:



### Motivation:

1. Atrophy correlates with functional networks, which are not spatially connected (Seeley et al., Neuron, 2009)
2. Better biomarker prediction and disease staging

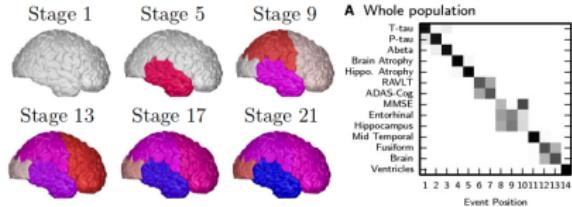


(a) Seeley et al., Neuron, 2009

- Modelling the progression of Alzheimer's disease can potentially help drug development
- Several data-driven disease progression models have been recently formulated:

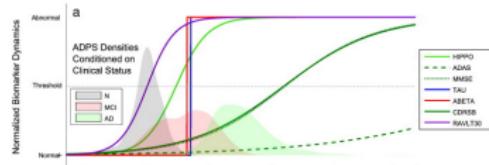
### Event-Based Model

(Fontejin et al., Neuroimage, 2012, Young et al., Brain, 2014)



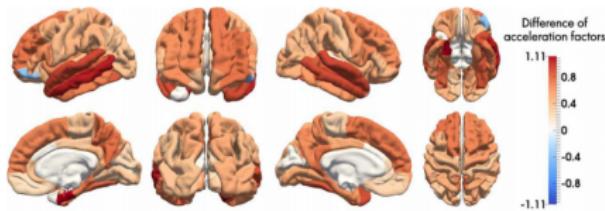
### Disease Progression Score

(Jedynak et al., Neuroimage, 2012)



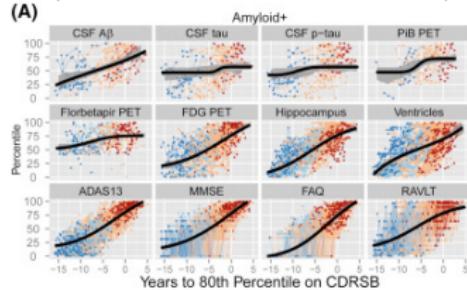
### Manifold-based model

(Schiratti et al., IPMI, 2015)



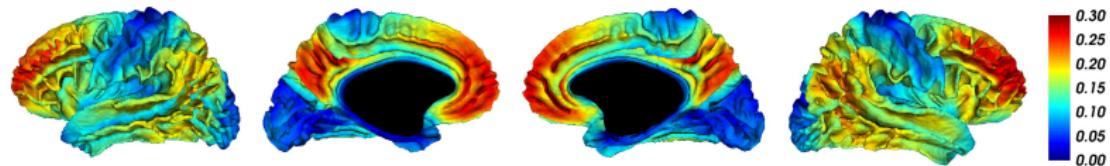
### Self-Modelling Regression

(Donohue et al., Alz. & Dementia, 2014)



## Voxelwise disease progression model (Bilgel et al., IPMI, 2015)

- Built on PET data measuring amyloid load at each voxel
- Estimates a unique trajectory for each voxel
- However, it uses a spatial correlation function



- We aim to avoid imposing spatial correlation

## Idea

- Combine two techniques:
  - unsupervised learning (clustering)
  - disease progression modelling
- Estimate trajectories for each vertex on the cortical surface
- Vertex measures cortical thickness at that location

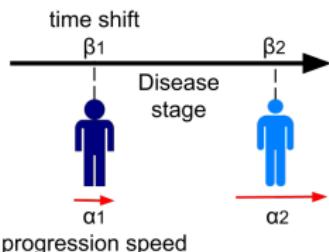
## Method outline:

1. Each subject has a *disease progression score* (DPS):

$$s_{ij} = \alpha_i t_{ij} + \beta_i$$

where:

- $s_{ij}$  - disease progression score of subject  $i$  at timepoint  $j$
- $t_{ij}$  - age of subject  $i$  at timepoint  $j$
- $\alpha_i$  - progression speed of subject  $i$
- $\beta_i$  - time shift of subject  $i$



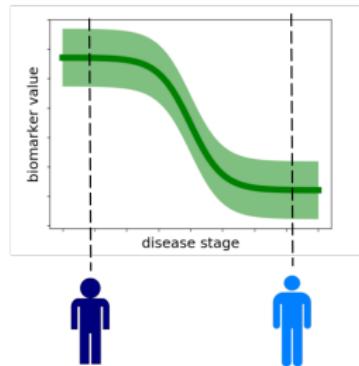
**Method outline - continued:**

2. Each cortical thickness measurement  $V_l^{ij}$  follows a sigmoidal curve  $f(\cdot; \theta)$  along the disease progression:

$$V_l^{ij} \approx f(s_{ij}; \theta_k) = \frac{a_k}{1 + \exp(-b_k(s - c_k))} + d_k$$

where

- $V_l^{ij}$  - cortical thickness at location  $l$  for subject  $i$ , timepoint  $j$
- $\theta_k = [a_k, b_k, c_k, d_k]$  - parameters of  $k$ -th sigmoid curve



3. We assume Gaussian noise along the  $k$ -th trajectory:

$$p(V_l^{ij} | \alpha_i, \beta_i, \theta_k, \sigma_k) \sim N(f(\alpha_i t_{ij} + \beta_i; \theta_k), \sigma_k)$$

where:

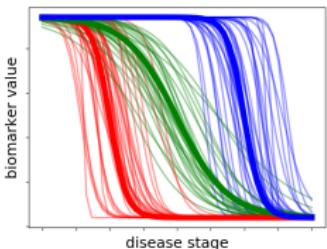
- $N$  - pdf of the Gaussian distribution
- $\sigma_k$  - noise level

**Idea:** Group vertices with similar progression dynamics into clusters

### Method outline - continued:

4. Define  $Z_l$  as the cluster that generated vertex  $l$ :

$$p(V_l^{ij} | \alpha_i, \beta_i, \theta_{Z_l}, \sigma_{Z_l}, Z_l) \sim N(f(\alpha_i t_{ij} + \beta_i; \theta_{Z_l}), \sigma_{Z_l})$$



where

- $Z_l$  - discrete latent variable assigning vertex  $l$  to a cluster  $k \in [1 \dots K]$

5. Extend to all subjects and vertices:

$$p(V, Z | \alpha, \beta, \theta, \sigma) = \prod_l^L \prod_{(i,j) \in I} N(V_l^{ij} | f(\alpha_i t_{ij} + \beta_i; \theta_{Z_l}), \sigma_{Z_l})$$

where

- $L$  - the total number of vertices on the cortical surface
- $I = (i, j)$  - set of available timepoints for each subject  $i$  and timepoint  $j$
- we assume independence across subjects and voxels in different clusters

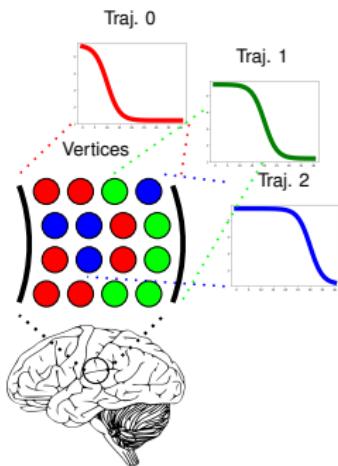
**Method outline - continued:**

6. Marginalise over the hidden variables  $Z_l$  (cluster assignments):

$$p(V|\alpha, \beta, \theta, \sigma) = \prod_{l=1}^L \sum_{k=1}^K p(Z_l = k) \prod_{(i,j) \in l} N(V_l^{ij} | f(\alpha_i t_{ij} + \beta_i; \theta_k), \sigma_k)$$

**Summary:**

1. Model clusters vertices on the brain surface according to disease progression dynamics
  - No assumptions on spatial correlation
2. Model estimates one trajectory for each cluster
3. Model estimates subject progression scores
4. Parameters to estimate:  $[\alpha, \beta, \theta, \sigma]$



## Model fitting with Expectation-Maximisation (EM):

- **E-step:**

- Estimate vertex assignment to clusters:

$$z_{lk} = p(Z_l = k | V_l, \Theta^{old}) = \frac{\prod_{(i,j) \in I} N(V_l^{ij} | f(\alpha_i^{old} t_{ij} + \beta_i^{old} | \theta_k^{old}), \sigma_k^{old})}{\sum_{m=1}^K \prod_{(i,j) \in I} N(V_l^{ij} | f(\alpha_i^{old} t_{ij} + \beta_i^{old} | \theta_m^{old}), \sigma_m^{old})} \quad (1)$$

- **M-step:**

- Update trajectories:

$$\theta_k = \arg \min_{\theta_k} \left[ \sum_{l=1}^L z_{lk} \sum_{(i,j) \in I} (V_l^{ij} - f(\alpha_i t_{ij} + \beta_i | \theta_k))^2 \right] - \log p(\theta_k) \quad (2)$$

- Update subject progression scores:

$$\alpha_i, \beta_i = \arg \min_{\alpha_i, \beta_i} \left[ \sum_{l=1}^L \sum_{k=1}^K z_{lk} \frac{1}{2\sigma_k^2} \sum_{j \in I_l} (V_l^{ij} - f(\alpha_i t_{ij} + \beta_i | \theta_k))^2 \right] - \log p(\alpha_i, \beta_i) \quad (3)$$

## M-step - Numerical optimisation

- M-step has no analytical solution
- Perform numerical optimisation with Broyden-Fletcher-Goldfarb-Shanno (BFGS):
  - fast convergence
  - uses first derivative of objective function
- EM still converges with partial M-step

## Initialisation

- We set  $\alpha_i = 1$  and  $\beta_i = 0, \forall i$
- We initialise  $z_{lk} = p(Z_l = k | V_l, \Theta^{old})$  using k-means clustering
  - feature vector for vertex  $l$ :  $[V_l^{ij} | (i, j) \in l]$  (measurements for all subjects at that location)
- Estimate the optimal number of clusters with the Bayesian Information Criterion (BIC)
  - Number of parameters:  $5K + 2S$

## 1. Simulation tests using synthetic data

## 2. Results on two datasets:

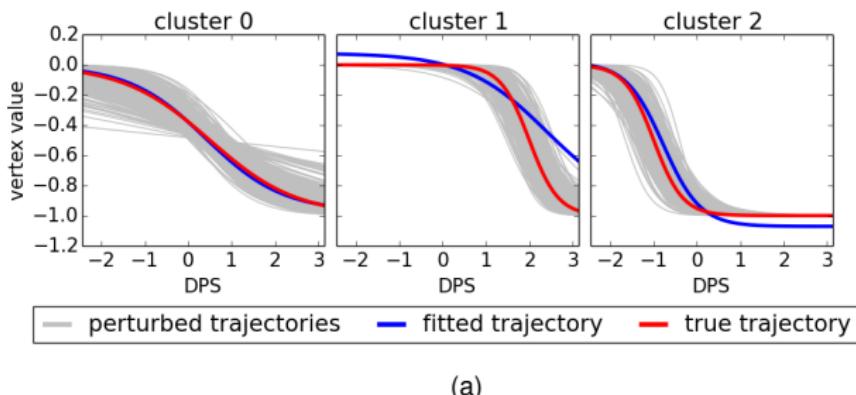
- Alzheimer's Disease Neuroimaging Initiative (ADNI)
- Dementia Research Center, University College London (UCL DRC)

## 3. Model validation

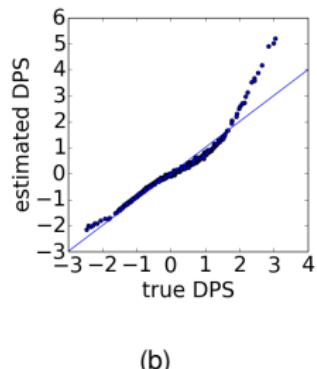
- Robustness
- Staging consistency

## 4. Recent results (not included in paper)

- Simulated data from 1,000 vertices generated from 3 clusters:
  - sampled age and shift parameters from 300 subjects with 4 timepoints (each timepoint 1 year apart), with  $t_{i1} \sim U(40, 80)$ ,  $\alpha_i \sim N(1, 0.05)$ ,  $\beta_i \sim N(0, 10)$
  - generated three sigmoidal trajectories with different center points and slopes (red lines)
  - generated a random cluster assignment for  $L = 1,000$  vertices
  - sampled  $L$  perturbed trajectories  $\theta_l$  from the original trajectories, one for each vertex (gray lines)
  - sampled subject data for every vertex  $l$  from its corresponding perturbed trajectory  $\theta_l$  with  $\sigma_l = 0.5$
- Model was able to recover the:
  - underlying trajectories, Fig. (a)
  - subject progression scores, Fig. (b)



(a)



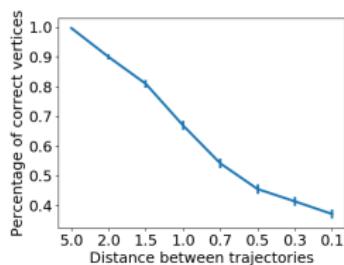
(b)

**Hypothesis:** Model will perform worse when:

- the trajectories are very similar to each other
- number of clusters is large
- number of subjects is decreasing

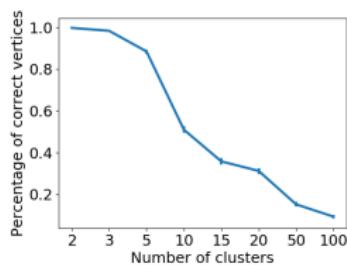
**Method:** Performed several "stress tests" for each scenario:

Trajectories becoming similar



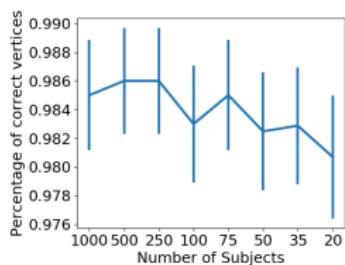
(a)

Increasing number of clusters



(b)

Decreasing number of subjects



(c)

## Conclusion:

- Model performance decreases when:
  - the trajectories become more similar
  - number of clusters increases
- However, performance is not affected by the number of subjects

**Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset:**

	Number of Subjects	Number of Scans	Age at baseline (years)
Controls	138	4.3	76.3
MCI	235	4.6	74.8
AD	81	3.5	75.8

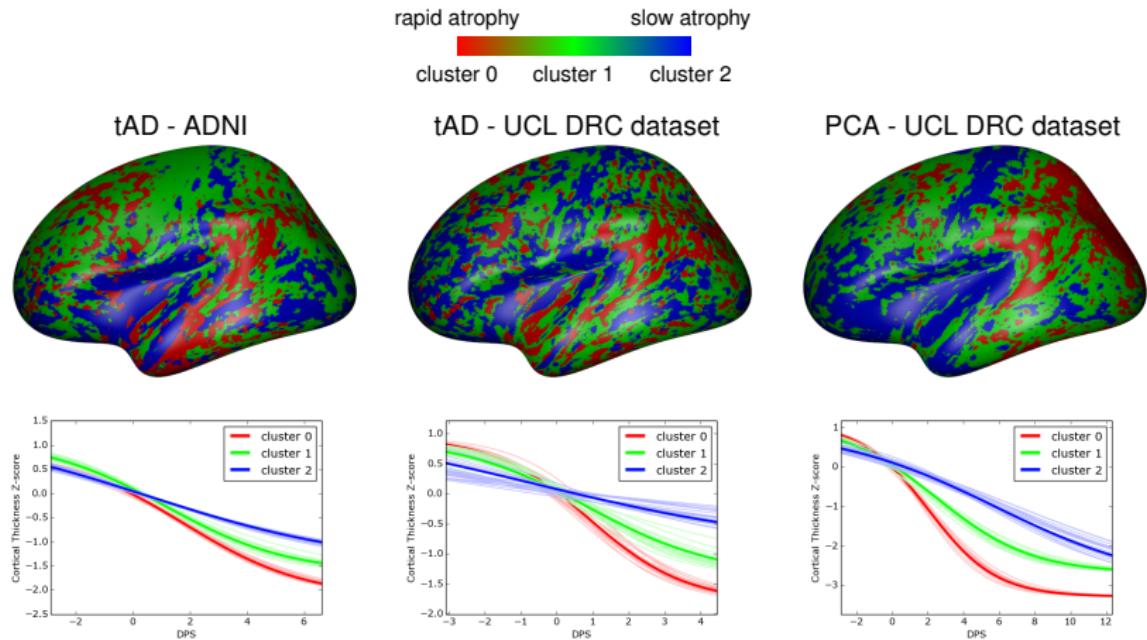
**Dementia Research Center, University College London (UCL DRC) dataset:**

	Number of Subjects	Number of Scans	Age at baseline (years)
Controls	31	5.0	66.3
PCA	32	4.1	62.6
AD	24	5.4	71.2

**Preprocessing:**

- Logitudinally registered all MRI images to a common template using Freesurfer
- Extracted vertexwise cortical thickness measurements
- Cortical thickness values at each vertex were standardised with respect to controls

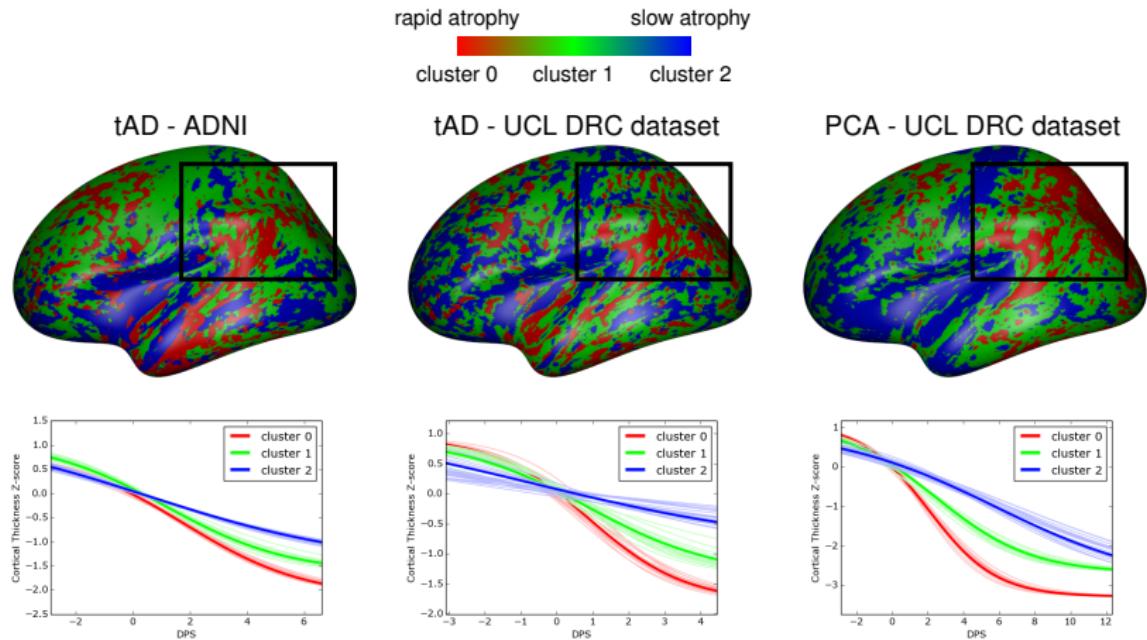
# Results - ADNI and UCL DRC Cohorts



## Conclusions:

- Similar patterns of atrophy in tAD in two independent datasets: ADNI and UCL DRC
- Distinct patterns of atrophy in two different diseases: typical AD and Posterior Cortical Atrophy

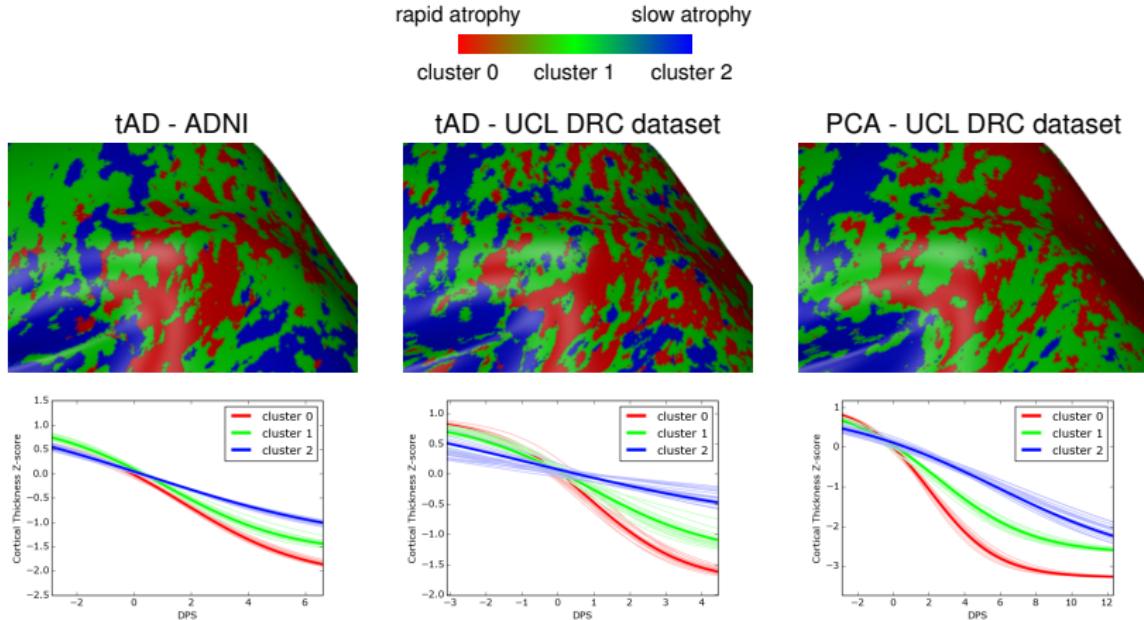
# Results - ADNI and UCL DRC Cohorts



## Conclusions:

- Similar patterns of atrophy in tAD in two independent datasets: ADNI and UCL DRC
- Distinct patterns of atrophy in two different diseases: typical AD and Posterior Cortical Atrophy

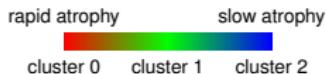
# Results - ADNI and UCL DRC Cohorts



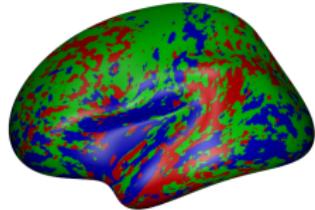
## Conclusions:

- Similar patterns of atrophy in tAD in two independent datasets: ADNI and UCL DRC
- Distinct patterns of atrophy in two different diseases: typical AD and Posterior Cortical Atrophy

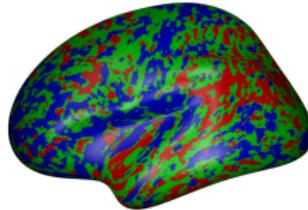
Results resemble previous findings in the literature



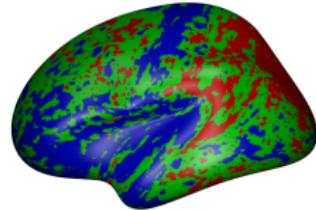
tAD - ADNI



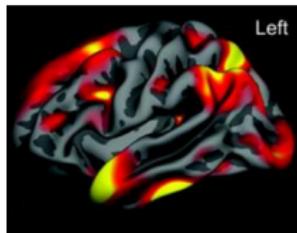
tAD - UCL DRC dataset



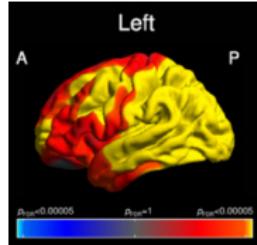
PCA - UCL DRC dataset



tAD Cortical Thinning  
Dickerson et al., Cereb. Cortex, 2009



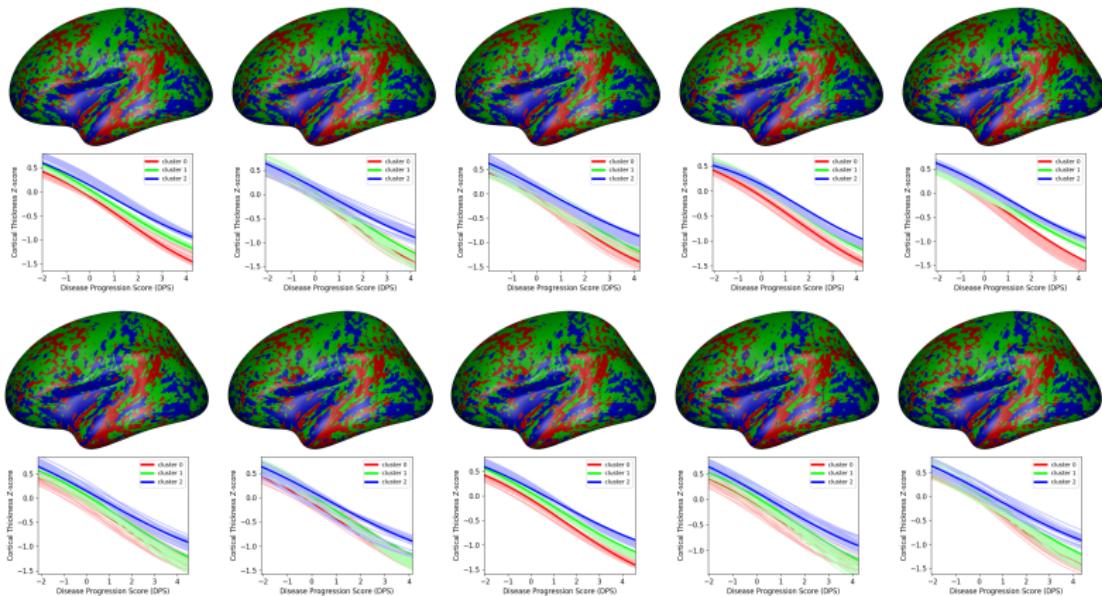
PCA Cortical Thinning  
Lehmann et al., Neurobiol. of Aging, 2011



# Results - Validation of Estimated Atrophy Patterns

## Cross-validation

- Tested the consistency of the spatial clustering in ADNI using 10-fold CV
- Results show good agreement in terms of spatial distribution
- Average dice score overlap for all cluster pairs: 0.89



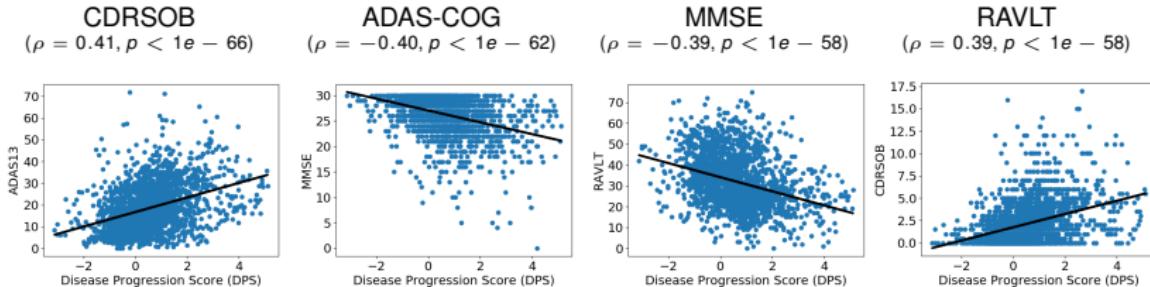
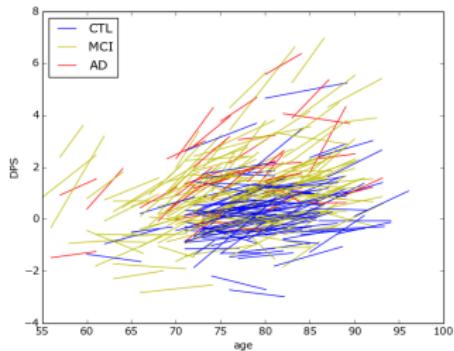
## Hypotheses:

1. Subjects show higher disease progression score (DPS) in later visits
2. DPS correlates with other markers of disease progression

**Method:** Ran our model on ADNI using 10-fold cross-validation

## Results

1. 84% of subjects analysed show increasing DPS
2. Progression scores correlate well with cognitive tests:



## Motivation

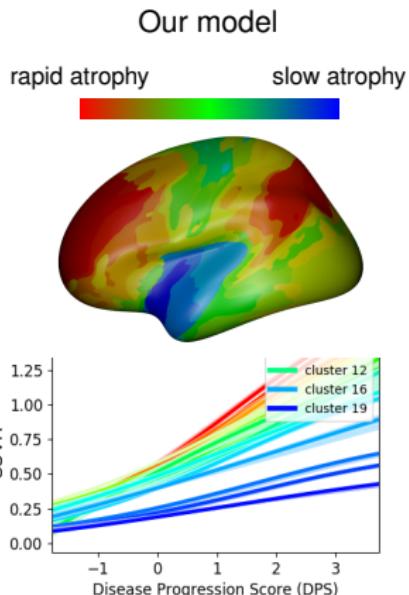
- Prove the model can be used on other types of data
- Hypothesis: PET clusters will be different from cortical thickness ones

## Methods

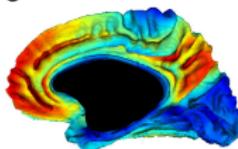
- Used 433 subjects from ADNI with PET AV45 scans
- Preprocessing with the PETSurfer pipeline

## Conclusion

- PET deposition patterns are focused on precuneus and frontal areas
- Spatial patterns match with previous findings (Bilgel et al., IPMI, 2015)
- Model still works with PET data, which is more spatially correlated



Bilgel et al, IPMI, 2015



Developed a method that:

- finds disconnected regions which undergo similar disease progression dynamics
- simultaneously estimates biomarker trajectories
- completely new way of parcellating the brain based on disease progression
- applicable to any kind of vertexwise/voxelwise data (e.g. MRI, PET, DTI)

Findings:

- Model finds plausible patterns in four different datasets
- Estimated disease progression scores are clinically relevant

## Limitations

1. The model assumes all patients follow the same progression pattern
2. Biomarker trajectories are assumed sigmoidal
3. Model relies on a well-defined control population
4. Model doesn't allow variability in cluster trajectories

## Future work

1. Model distinct progression patterns for population subgroups (e.g. Young et al., IPMI, 2015)
2. Our model is easily extensible to non-parametric trajectories
3. Define control population using data-driven techniques (e.g. mixture models).
4. Fit parameters using variational methods (e.g. Variational Bayes)

## Work in progress:

- Correlate the cluster patterns with a structural connectome from the HCP dataset
- Hypothesis: disconnected areas cluster together due to underlying WM connections

# Acknowledgements



## Collaborators

1. Leon Aksman
2. Maura Bellio
3. Arman Eshaghi
4. Nicholas Firth
5. Sara Garbarino
6. Kyriaki Mengoudi
7. Marco Lorenzi
8. Neil Oxtoby
9. Peter Wijeratne
10. Alexandra Young

## Project supervisors

Daniel Alexander



Sebastian Crutch



## Funders

UCL EPSRC CDT in  
 Medical Imaging

The logo for the UCL EPSRC CDT in Medical Imaging. It features the text "UCL EPSRC CDT in" above "Medical Imaging". The word "Medical" is written in a larger, stylized font where the letters "M" and "I" overlap. To the left of the text is a circular icon containing a silhouette of a person standing next to a medical scanner.

  
EPSRC  
Pioneering research  
and skills

The EPSRC logo, consisting of the acronym "EPSRC" in a large, bold, blue font with a green horizontal bar underneath it, followed by the tagline "Pioneering research and skills" in a smaller, purple font.

Join the TADPOLE challenge!



- Challenge: use any model of choice to predict future AD biomarkers
- URL: <https://tadpole.grand-challenge.org/>
- See me for flyers or more details!



[Home](#)   [Background](#)   [Details](#)   [Metrics](#)   [Data](#)   [Submit](#)   [Forum](#)   [Contacts](#)

## Tadpole Challenge

Welcome to The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) challenge.

Brought to you by the EuroPOND consortium in collaboration with the Alzheimer's Disease Neuroimaging Initiative (ADNI).

## Hypotheses

- Dynamic clustering improves staging and biomarker prediction
- Subject staging improves biomarker prediction

**Methods** Compared the performance of our model to two simplified models:

1. Predefined ROI model (i.e. Jedynak et al., Neuroimage, 2012)
2. Dynamic clustering only model, without subject staging: fixes  $\alpha_i = 1, \beta_i = 0, \forall i$

## Results

- Correlation  $\rho$  between progression scores and cognitive tests
- Cortical thickness prediction error (RMSE)

Model	CDRSOB ( $\rho$ )	ADAS13 ( $\rho$ )	MMSE ( $\rho$ )	RAVLT ( $\rho$ )	Prediction (RMSE)
Dynamic clusters + staging (our model)	0.35	0.36	0.35	0.31	1.007 +/- 0.011
Predefined ROI (Jedynak et al., Nimg, 2012)	0.39	0.40	0.40	0.34	1.008 +/- 0.012
Dynamic clusters, no staging	*0.02	*-0.02	*-0.02	*0.01	*1.058 +/- 0.021

## Conclusion

- Dynamic clustering doesn't improve performance vs Predefined ROI model
- Disease staging of subjects improves performance vs model without staging

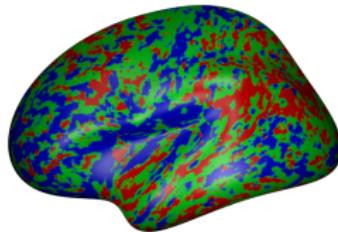
**Motivation**

- measurements from neighbouring vertices are inherently correlated
- could enable better staging and prediction

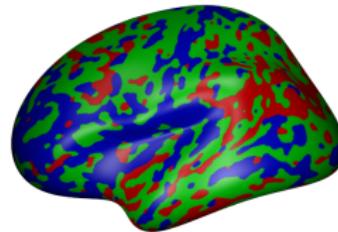
**Method:**

- Original formulation assumes independence between vertices
- Modelled spatial correlation between vertices with a *Markov Random Field* (MRF):

$$p(V, Z | \alpha, \beta, \theta, \sigma) = \prod_l^L \prod_{(i,j) \in I} N(V_l^{ij} | f(\alpha_i t_{ij} + \beta_i | \theta_{Z_l}), \sigma_{Z_l}) \prod_{l_1 \sim l_2} \Psi(Z_{l_1}, Z_{l_2})$$



(a) Without MRF

(b) With MRF,  $\alpha = 5$ .

- Work still in progress (not included in paper)