

A vertex clustering model for disease progression: Application to cortical thickness images

Răzvan V. Marinescu¹, Arman Eshaghi^{1,2}, Marco Lorenzi^{1,4}, Alexandra L. Young¹, Neil P. Oxtoby¹, Sara Garbarino¹, Timothy J. Shakespeare³, Sebastian J. Crutch³, and Daniel C. Alexander¹, for the Alzheimers Disease Neuroimaging Initiative*

¹ Centre for Medical Image Computing, Computer Science Department, University College London, UK

² Queen Square MS Centre, UCL Institute of Neurology, London

³ Dementia Research Centre, UCL Institute of Neurology, University College London, UK

⁴ University of Côte d'Azur, Inria Sophia Antipolis, Asclepios Research Project

Abstract. We present a disease progression model with single vertex resolution that we apply to cortical thickness data. Our model works by clustering together vertices on the cortex that have similar temporal dynamics and building a common trajectory for vertices in the same cluster. The model estimates optimal stages and progression speeds for every subject. Simulated data show that it is able to accurately recover the vertex clusters and the underlying parameters. Moreover, our clustering model finds similar patterns of atrophy for typical Alzheimer's disease (tAD) subjects on two independent datasets: the Alzheimer's Disease Neuroimaging Initiative (ADNI) and a cohort from the Dementia Research Centre (DRC), UK. Using a separate set of subjects with Posterior Cortical Atrophy (PCA) from the DRC dataset, we also show that the model finds different patterns of atrophy in PCA compared to tAD. Finally, our model also provides a novel way to parcellate the brain based on disease dynamics.

Keywords: Disease Progression Model, Cortical Thickness, Vertex-wise Measures, Alzheimer's Disease, Posterior Cortical Atrophy

1 Introduction

During the progression of Alzheimer's disease, many biomarkers based on Magnetic Resonance Imaging (MRI) such as cortical thickness become abnormal at

* Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

different points in the progression. Finding out the precise temporal evolution of these biomarkers facilitates patient staging in clinical trials. However, the analysis of disease progression is limited by several factors: short number of follow-up visits available, different disease onset and progression speed for every subject and cohort heterogeneity.

A hypothetical model of disease progression has been proposed by [1], describing the trajectory of key biomarkers along the progression of Alzheimer's disease. The model suggests that amyloid-beta and tau biomarkers become abnormal long before symptoms appear, followed by brain atrophy measures and cognitive decline. Motivated by this idea, several models such as [2] or [3] have been proposed that reconstruct biomarker trajectories and can be used to stage subjects. However, these models make use of *a priori* clinical categories, which are noisy, biased and can limit the temporal resolution of the model. This motivates the use of fully data-driven approaches that do not use *a priori* clinical stages.

Various data-driven disease progression modelling techniques have been proposed in recent years. One such model is the Event-Based Model [4], which models the progression of disease as a sequence of discrete events, representing underlying biomarkers switching from a normal to abnormal state. Other models such as the Disease Progression Score (DPS) [5] or self-modelling regression approaches [6] have been developed, that build continuous trajectories by "stitching" together short-term follow-up data. Models estimating linear or logistic trajectories by means of Riemannian manifold techniques have also been recently shown [7].

The main limitation of these data-driven models is that they use a small set of biomarkers that are obtained by averaging MRI or PET measures across all voxels or vertices in a Region of Interest (ROI). This can be problematic, especially if different parts of the ROI, say the hippocampus, are affected at different speeds or timepoints in the disease process. Therefore, moving to a voxel-wise approach would allow one to estimate the fine-grained spatial distribution of atrophy, which could give new insights into the disease process and potentially enable more precise staging. A voxel-wise disease progression model [8] has been recently proposed to mitigate this problem, that uses amyloid measures in each voxel as its input data. However, the model by [8] has two limitations: (1) the biomarker trajectories are assumed to be linear, so cannot capture the plateau effect observed in amyloid-beta or tau and (2) the model uses a spatial correlation function for modelling correlation between voxels; while this is necessary, due to the nature of the imaging data, it has been shown that in different types of dementia atrophy patterns match functional networks, which are not spatially connected [9].

In this work, we present a new disease progression model with single vertex resolution that avoids assumptions on spatial correlation. We combine unsupervised learning and disease progression modelling to identify clusters of vertices on the cortical surface, with no spatial constraints, that show a similar trajectory of atrophy over a particular patient cohort. This formulation enables us to gain

new insights into the spatial structure of atrophy in different diseases and also provides a novel parcellation of the brain based on temporal change. Moreover, each cluster of vertices has a corresponding sigmoidal trajectory, which avoids the limitation of linear trajectories in [8].

We first show using simulated data that our model is able to recover known underlying clusters, trajectory parameters and subject stages. We then apply our model to cortical thickness vertex-wise measures using ADNI and the DRC dataset and highlight the new insights the model can give. Finally, we validate our model using cross-validation and by correlating the subject stages with cognitive measures.

2 Methods

2.1 Model

We seek to identify groups of image vertices that show a common trajectory during the disease course, while simultaneously placing each visit from each subject within that disease course. In a similar way to [5,6,7], we estimate a time shift and speed (or rate) of progression for each subject. We relate these time shifts and progression speeds by assigning each subject a disease stage which we will refer to as the Disease Progression Score (DPS). In contrast to [5,6,7], which model temporal trajectories for a small set of biomarker measures based on *a priori* defined ROIs, we model temporal trajectories for each vertex on the cortical surface. Each trajectory is a function of the disease progression score (i.e. disease stage) of a subject. We estimate each subjects' time shift, progression speed and trajectory parameters from the data. The disease progression score s_{ij} for subject i at visit j is defined as a linear transformation of age t_{ij} :

$$s_{ij} = \alpha_i t_{ij} + \beta_i \quad (1)$$

where α_i and β_i represent the speed of progression and time shift (i.e. disease onset) of subject i .

Our model assumes that the cortical thickness at each vertex on the cortical surface follows a sigmoidal trajectory $f(s)$ given the disease progression score s . We also assume that vertices are grouped into K clusters and we model a unique trajectory for each cluster $k \in [1, \dots, K]$, which will be referred to as cluster trajectories. The sigmoidal function for cluster k is parametrised as $\theta_k = [a_k, b_k, c_k, d_k]$ where

$$f(s; \theta_k) = \frac{a_k}{1 + \exp(-b_k(s - c_k))} + d_k \quad (2)$$

For a given subject i at visit j , the value V_l^{ij} of its cortical thickness at vertex l is a random variable that has an associated discrete latent variable $Z_l \in [1, \dots, K]$ denoting the cluster it was generated from. The value of V_l^{ij} given that it was generated from cluster Z_l can be modelled as:

$$p(V_l^{ij} | \alpha_i, \beta_i, \theta_{Z_l}, \sigma_{Z_l}, Z_l) = N(V_l^{ij} | f(\alpha_i t_{ij} + \beta_i | \theta_{Z_l}), \sigma_{Z_l}) \quad (3)$$

where $N(V_l^{ij} | f(\alpha_i t_{ij} + \beta_i | \theta_{Z_l}), \sigma_{Z_l})$ represents the pdf of the normal distribution that models the measurement noise along the sigmoidal trajectory of cluster Z_l , having variance σ_{Z_l} . Next, we assume the measurements from different subjects are independent, while the measurements from the same subject i at different visits j are linked using the disease progression score from equation 1, because we estimate only two parameters (α_i and β_i) using the data from all visits j . Moreover, we also assume a uniform prior on Z_l . This gives the following model:

$$p(V_l, Z_l | \alpha, \beta, \theta, \sigma) = \prod_{(i,j) \in I} N(V_l^{ij} | f(\alpha_i t_{ij} + \beta_i | \theta_{Z_l}), \sigma_{Z_l}) \quad (4)$$

where $I = (i, j)$ represents the set of all the subjects i and their corresponding visits j . Furthermore, $V_l = [V_l^{ij} | (i, j) \in I]$ is the 1D array of all the values for vertex l across every subject and corresponding visit. Vectors $\alpha = [\alpha_1, \dots, \alpha_S]$ and $\beta = [\beta_1, \dots, \beta_S]$, where S is the number of subjects, denote the stacked parameters for the subject shifts. Vectors $\theta = [\theta_1, \dots, \theta_K]$ and $\sigma = [\sigma_1, \dots, \sigma_K]$, with K being the number of clusters, represent the stacked parameters for the sigmoidal trajectories and measurement noise specific to each cluster.

We further assume all vertex measurements to be spatially independent, giving the complete data likelihood:

$$p(V, Z | \alpha, \beta, \theta, \sigma) = \prod_l^L \prod_{(i,j) \in I} N(V_l^{ij} | f(\alpha_i t_{ij} + \beta_i | \theta_{Z_l}), \sigma_{Z_l}) \quad (5)$$

where $V = [V_1, \dots, V_L]$, $Z = [Z_1, \dots, Z_L]$, L being the total number of vertices on the cortical surface. We recall that we don't want to enforce spatial correlation between vertices as we are interested to see if vertices from distinct areas of the brain are grouped together in the same cluster. Our assumption is also justified by the fact that we smoothed the cortical thickness images in the preprocessing steps. We get the final model log likelihood for incomplete data by marginalising over the latent variables Z :

$$p(V | \alpha, \beta, \theta, \sigma) = \prod_{l=1}^L \sum_{k=1}^K p(Z_l = k) \prod_{(i,j) \in I} N(V_l^{ij} | f(\alpha_i t_{ij} + \beta_i | \theta_k), \sigma_k) \quad (6)$$

Therefore, the parameters that need to be estimated are $\Theta = [\alpha, \beta, \theta, \sigma]$ where α and β are the subject specific shifting parameters while θ and σ are the cluster specific trajectory and noise parameters.

2.2 Fitting the Model using EM

Due to the summing over the latent variables Z , it is not possible to find a closed form solution to the maximum likelihood. Therefore, we fit our model using Expectation-Maximisation, which is suitable given the large number of data points and parameters that need to be estimated.

E-step In the Expectation step we seek to estimate which cluster has generated each of the L vertices, given the current estimates of the cluster parameters $\theta_k^{old}, \sigma_k^{old}$ as well as the subject specific shift parameters $\alpha_i^{old}, \beta_i^{old}$. More formally, we seek to find $p(Z|V, \Theta^{old}) = \prod_l^L p(Z_l|V_l, \Theta^{old})$, given our independence assumption between vertices. Let us denote by $z_{lk} = p(Z_l = k|V_l, \Theta^{old})$. We then have:

$$z_{lk} = \frac{\prod_{(i,j) \in I} N(V_l^{ij} | f(\alpha_i^{old} t_{ij} + \beta_i^{old} | \theta_k^{old}), \sigma_k^{old})}{\sum_{m=1}^K \prod_{(i,j) \in I} N(V_l^{ij} | f(\alpha_i^{old} t_{ij} + \beta_i^{old} | \theta_m^{old}), \sigma_m^{old})} \quad (7)$$

Ignoring the normalisation factor, we perform a log transformation and expand the pdfs of the normal distributions. This results in the following update equation for the E-step:

$$\log z_{lk} \propto -\frac{1}{2} \log (2\pi (\sigma_k^{old})^2) |I| - \frac{1}{2(\sigma_k^{old})^2} \sum_{(i,j) \in I} (V_l^{ij} - f(\alpha_i^{old} t_{ij} + \beta_i^{old} | \theta_k^{old}))^2 \quad (8)$$

The original probabilities can be easily recovered by exponentiating them and then normalising with respect to their sum.

M-step In the Maximisation step we try to find $\Theta = (\alpha, \beta, \theta, \sigma)$ that maximise $E_{Z|V, \Theta^{old}} [\log p(V, Z|\Theta)]$. Since there is no closed-form solution, we perform successive refinements of θ_k for each cluster k and α_i, β_i for each subject i until convergence.

In order to get the update rule for the trajectory parameters θ_k corresponding to cluster k we need to maximise the expected log likelihood with respect to θ_k . We get the following simplified optimisation problem:

$$\theta_k = \arg \min_{\theta_k} \left[\sum_{l=1}^L z_{lk} \sum_{(i,j) \in I} (V_l^{ij} - f(\alpha_i t_{ij} + \beta_i | \theta_k))^2 \right] - \log p(\theta_k) \quad (9)$$

A similar equation is also obtained for σ_k . After estimating θ and σ for every cluster, we use the new values to estimate the subject specific parameters α and β . Let S be the number of subjects and α_i, β_i be the rate and shift for subject $i \in S$. We again maximise the expected log likelihood with respect to α_i, β_i independently, and after simplifications we obtain the following problem:

$$\alpha_i, \beta_i = \arg \min_{\alpha_i, \beta_i} \left[\sum_{l=1}^L \sum_{k=1}^K z_{lk} \frac{1}{2\sigma_k^2} \sum_{j \in I_i} (V_l^{ij} - f(\alpha_i t_{ij} + \beta_i | \theta_k))^2 \right] - \log p(\alpha_i, \beta_i) \quad (10)$$

In summary, at every single M-step, we iterate between solving for θ, σ and solving for α, β using numerical optimisation until convergence. Due to the use

of numerical optimisation, we are not guaranteed to find the global maxima for the expected log likelihood, but EM still works if we only find an increase in the log-likelihood. This approach that involves a partial M-step is called Generalised EM.

2.3 Initialisation and Implementation

Before starting the fitting process, we need to initialise α , β and the clustering probabilities z_{lk} . We set α_i and β_i to be 1 and 0 respectively for each subject. We initialise z_{lk} using k-means clustering using vectors V_l having $|I|$ number of samples and L features. Furthermore, as already explained in [5], the scale of the DPS is arbitrary so we standardise the scores at each EM iteration such that the DPS of controls have a mean μ_N of zero and a standard deviation σ_N of 1. This also requires a rescaling of the cluster parameters θ_k .

In our implementation, we run the main EM loop until convergence of the clustering probabilities z_{lk} . At each M-step we perform numerical optimisation using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm that makes use of the first derivative of the objective function. In all datasets analysed, the method converges after maximum 25 EM iterations.

3 Simulation Results

We first tested the model in a toy scenario using synthetic data, which we generated as follows: (1) sampled age and shift parameters from 300 subjects with 4 timepoints (each timepoint 1 year apart), with $t_{i1} \sim U(40, 80)$, $\alpha_i \sim N(1, 0.05)$, $\beta_i \sim N(0, 10)$ (2) generated three sigmoids with different center points and slopes (Fig. 1a, red lines) (3) generated a random cluster assignment for $L = 1,000$ vertices, i.e. every vertex l was assigned to a cluster $a[l] \in \{1, 2, 3\}$ (4) sampled a set of L perturbed trajectories θ_l from each of the original trajectories, one for each vertex (Fig. 1a, gray lines) and (5) sampled subject data for every vertex l from its corresponding perturbed trajectory θ_l with $\sigma_l = 0.5$

To fit the data we used a uniform prior on the parameters θ and σ and an informative prior on α and β , with $p(\alpha_i) \sim \Gamma(49, 70)$ and $\beta_i \sim N(0, 1)$. We also normalised age to have a mean of 0 and standard deviation of 1 and rescaled the DPS and cluster trajectories accordingly. After convergence, we calculated the agreement between the final clustering probabilities and the true clustering assignments as $A = \frac{1}{L} \sum_{l=1}^L p(Z_l = a[l])$. The method also requires us to set the number of clusters *a priori*, so we optimised the number of clusters using the Bayesian Information Criterion (BIC).

The BIC analysis correctly predicted three clusters for this synthetic experiment. Using three clusters, we also obtained a clustering agreement of $A = 1.0$. Figure 1a shows the original trajectories and the recovered trajectories using our model, plotted against the disease progression score on the x-axis and the vertex value on the y-axis. Moreover, in Fig. 1b we plotted the recovered DPS of each subject along with the true DPS.

The results show that the model accurately estimated which clusters generated each vertex. Moreover, the recovered trajectories are close to the true trajectories, with some errors for the trajectories corresponding to clusters 1 and 2. The recovered DPS also shows good agreement with the true DPS, with the exception of a few subjects with DPS greater than 2.5. This is explained by the fact that there is not enough signal in that DPS range in terms of trajectory dynamics (i.e. trajectories are mostly flat). The simulation confirms that our model is able to recover the hidden clusters, trajectory parameters and the subject specific parameters. However, more realistic simulations with varying noise levels and numbers of clusters are required to understand the limitations of the model and find out when it fails to recover the true parameters.

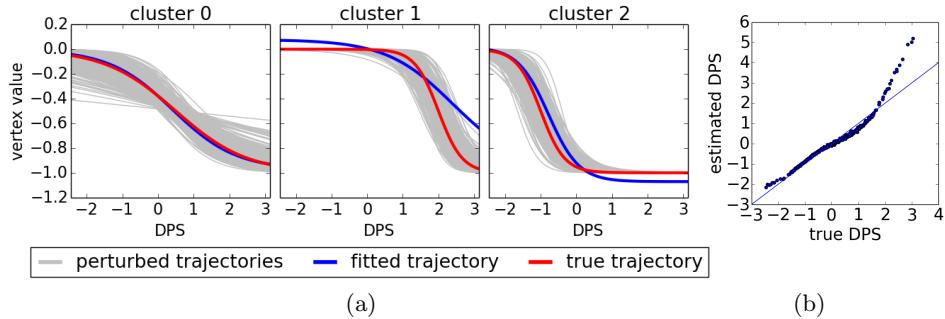


Fig. 1: (a) Reconstructed temporal trajectories (blue) from the synthetic data along with the true trajectories (red). The data was generated from the perturbed trajectories, which in turn were generated from the true trajectories. (b) Estimated subject-specific disease progression scores compared to the true scores.

4 Experimental Results

4.1 Data Acquisition and Preprocessing

Data used in this work were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) and from the Dementia Research Centre, UK. For ADNI, we downloaded all T1 MRI images that have undergone gradwarping, intensity correction and scaling for gradient drift. We included subjects that had at least 4 scans, in order to ensure we get a robust estimate of the subject specific parameters. This resulted in 328 subjects with an average number of 4.95 scans each. The DRC dataset consisted of T1 MRI scans from 31 healthy controls, 32 PCA and 23 typical typical AD subjects with at least 3 scans each and an average of 5.26 scans per subject.

On both datasets, in order to extract reliable cortical thickness measures we ran the Freesurfer longitudinal pipeline [10], which first registers the MRI images

to an unbiased within-subject template space using inverse-consistent registration. The longitudinally registered images were then registered to the average Freesurfer template and smoothed at a full-width/half-max (FWHM) level of zero. For each vertex we averaged the thickness levels from both hemispheres. Finally, we standardised the data from each vertex with respect to the values of that vertex in the control population. Each of the final images had a resolution of 163,842 vertices on the cortical surface.

4.2 ADNI and DRC Results

Using ADNI and DRC datasets, we were interested to find out the spatial distribution of cortical atrophy, as well as the rate and timing of this atrophy process. In particular, we wanted to find out: (1) if we get similar results using our model on two independent tAD datasets: ADNI and DRC and (2) if we get different patterns of atrophy on distinct diseases (tAD and PCA) that match previous studies.

BIC analysis predicted that the optimal number of clusters is two for the ADNI cohort and three for both tAD and PCA subjects from the DRC cohort. In order to make the results easily comparable across the different datasets, we ran all experiments using 3 clusters. Fig. 2a shows the results of our model using all ADNI subjects, where we coloured points on the cortical surface according to the cluster they most likely belong to. We assigned a colour to each cluster according to the slope of its corresponding trajectory, ranging from red (high slope suggesting a fast rate of atrophy) to blue (low slope suggesting a slow rate of atrophy). In Fig. 2d we also show the resulting cluster trajectories with samples from the posterior distribution of each θ_k . We repeated the same analysis on the DRC cohort, separately for the tAD subjects (Fig. 2b and 2e) and PCA subjects (Fig. 2c and 2f).

We notice that in tAD subjects using both ADNI and DRC datasets (Fig. 2), there is widespread atrophy in most temporal, parietal and frontal areas (red cluster), with the notable exception of the motor cortex and the occipital lobe. These patterns of atrophy are similar across the two different datasets. Moreover, the spatial distribution of cortical thinning found with our technique resembles results from previous longitudinal studies such as [11]. However, in contrast to these approaches, our model gives insight into the timing, rate and extent of atrophy and is also able to stage subjects across the disease timecourse.

In the PCA subjects (Fig. 2c), we find that the atrophy is more focused on the posterior part of the brain, mostly the posterior parietal and occipital, with more limited spread in the superior temporal and inferior frontal. This is in contrast with the tAD patterns in the other datasets, that lacks the focus on posterior parietal and occipital regions. This posterior pattern of atrophy also matches previous findings in the literature [12]. For all datasets, we find that the cluster trajectories differ less in timing and more in the slope and minima/maxima values at which they plateau (Fig. 2d, 2e, 2f). Our model therefore predicts that regions on the cortical surface are all affected roughly at the same time, but the rate and extent to which they are affected is different.

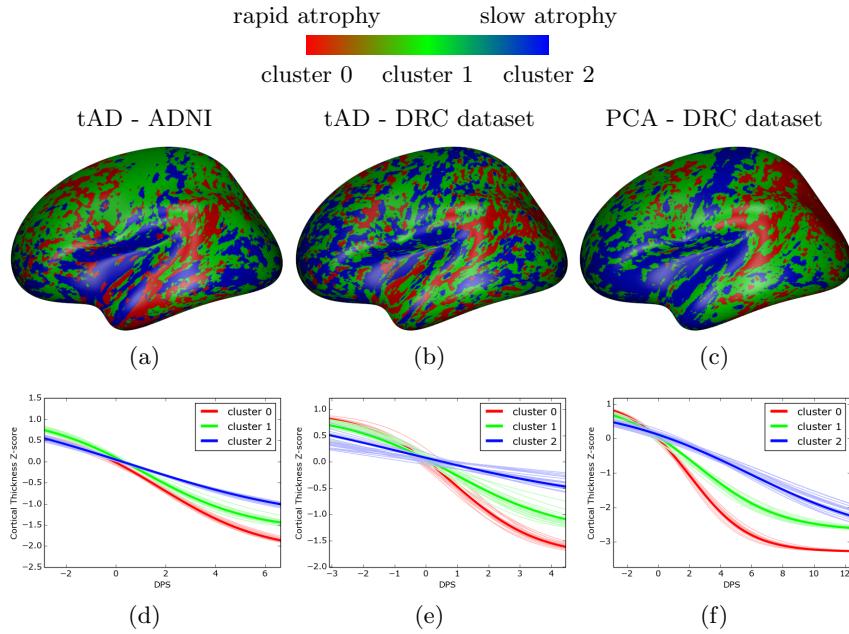


Fig. 2: (a) Clustering results on the ADNI data using our model, where each cluster is coloured according to the slope of its corresponding trajectory, from red (high slope suggesting very affected areas) to blue (low slope suggesting less affected areas). (d) The corresponding trajectories and samples from the posterior distribution of the trajectory parameters for the three clusters in ADNI. The same analysis is shown also for (b, e) tAD subjects from the DRC cohort and (c, f) PCA subjects from the DRC cohort.

4.3 Model Validation

We tested robustness of the model by performing 10-fold cross validation (CV) on ADNI. Our motivation was to test the following: (1) if similar spatial clustering is estimated at each fold, as quantified by Dice score overlap, (2) if the stages of the test subjects were consistent (i.e. were increasing for follow-up visits) and (3) if the stages of test subjects are clinically meaningful, by correlating them with cognitive tests such as Clinical Dementia Rating Scale - Sum of Boxes (CDRSOB), Alzheimer's Disease Assessment Scale - Cognitive (ADAS-COG), Mini-Mental State Examination (MMSE) and Rey Auditory and Verbal Learning Test (RAVLT).

Fig 3 shows the clusters that were estimated at each fold from the training data only. Moreover, in Fig. 4 we plot the estimated DPS (i.e. disease stage) of each subject from the test set against their age.

The results in Fig. 3 prove that the model is robust in cross-validation, as the estimated clusters are all very similar across folds. The average Dice scores we obtained across all pairs of folds were 0.89, 0.89 and 0.90 for clusters 0, 1 and 2 respectively. Furthermore, 84% of the subjects analysed show increased

stages across their follow-up visits, proving that the estimated stages are mostly consistent. Finally, the stages of test subject correlate with clinical measures such as CDRSOB ($\rho = 0.41, p < 1e-66$), ADAS-COG ($\rho = 0.40, p < 1e-62$), MMSE ($\rho = 0.39, p < 1e-58$) and RAVLT ($\rho = 0.35, p < 1e-46$), demonstrating that the stages have clinical validity.

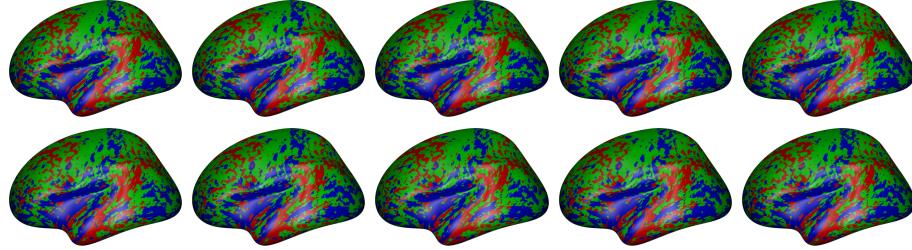


Fig. 3: Clusters estimated for each of the 10 cross-validation folds in ADNI. As before, each cluster is coloured according to the slope of its corresponding trajectory, from red (high rate of atrophy) to blue (low rate of atrophy).

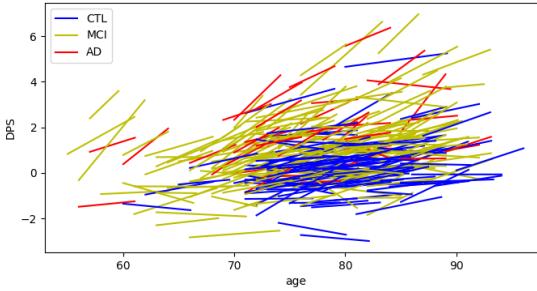


Fig. 4: The disease progression score for each subject from the ADNI dataset estimated during 10-fold cross validation. Each line represents an individual i with different visits j . Later visits generally have a higher corresponding stage.

5 Discussion

We presented a model of disease progression that clusters vertex-wise measures of cortical thickness based on similar temporal dynamics. The model highlights, for the first time, groups of cortical vertices that exhibit a similar temporal trajectory over the population. This provides a new way to parcellate the brain that is specific to the temporal trajectory of a particular disease. The model also finds the optimal temporal shift and progression speed for every subject. We applied the model to cortical thickness vertex-wise data from the ADNI and DRC cohorts. Our model found similar patterns of atrophy dynamics in the tAD subjects using the two independent datasets. Moreover, it also found different patterns of atrophy dynamics on two distinct diseases: tAD and PCA.

The model has some limitations. First of all, we assumed that cluster trajectories follow sigmoidal shapes, which might not be the case for many types of biomarkers such as cortical thickness. Another limitation of the model is that it assumes all subjects follow the same disease progression pattern, which might not be the case in heterogeneous datasets such as ADNI or DRC. This can be a concern, as there might be a pattern of atrophy that occurs in a small set of subjects. Moreover, our cluster-based model might miss atrophy patterns that occur in very small regions. Furthermore, the data we analysed has been standardised with respect to controls, which assumes controls don't show any biomarker abnormalities.

There are several potential avenues of future research. While we only used the model for studying cortical thickness, one can also apply it to other types of data such as amyloid images or Jacobian compression maps. On the methodological side, the assumption of sigmoidal trajectories can be avoided using non-parametric curves such as Gaussian Processes. Another extension is to model different progression dynamics for distinct subgroups using unsupervised learning methods like the approach of [13], or incorporate subject-specific deviations from the standard pattern of atrophy using a mixed-effects model.

Our approach can be used for accurately predicting and staging patients across the progression timeline of neurodegenerative diseases. This is promising for patient prognosis, as well as in clinical-trials for assessing efficacy of a putative treatment for slowing down the degeneration process.

6 Acknowledgements

This work was supported by the EPSRC Centre For Doctoral Training in Medical Imaging with grant EP/L016478/1. AE received a McDonald Fellowship from the Multiple Sclerosis International Federation (MSIF, www.msif.org), and the ECTRIMS - MAGNIMS Fellowship. ALY was supported through EPSRC grant EP/J020990/01. NPO and SG received funding from the EU Horizon 2020 research and innovation programme under grant agreement No 666992. SJC was supported by an Alzheimers Research UK Senior Research Fellowship and ESR-C/NIHR (ES/L001810/1) and EPSRC (EP/M006093/1) grants. DCA's work on this topic has funding from the EU Horizon 2020 research and innovation programme under grant agreement No 666992, as well as EPSRC grants J020990, M006093 and M020533. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). The Dementia Research Centre is an ARUK co-ordination center.

References

1. Jack, C.R., Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C. and Trojanowski, J.Q., 2010. Hypothetical model of dynamic

- biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*, 9(1), pp.119-128.
2. Bateman, R.J., Xiong, C., Benzinger, T.L., Fagan, A.M., Goate, A., Fox, N.C., Marcus, D.S., Cairns, N.J., Xie, X., Blazey, T.M. and Holtzman, D.M., 2012. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *New England Journal of Medicine*, 367(9), pp.795-804.
 3. Schmidt-Richberg, A., Guerrero, R., Ledig, C., Molina-Abril, H., Frangi, A.F., Rueckert, D. and Alzheimers Disease Neuroimaging Initiative, 2015, June. Multi-stage Biomarker Models for Progression Estimation in Alzheimers Disease. In International Conference on Information Processing in Medical Imaging (pp. 387-398). Springer International Publishing.
 4. Fonteijn, H.M., Modat, M., Clarkson, M.J., Barnes, J., Lehmann, M., Hobbs, N.Z., Scahill, R.I., Tabrizi, S.J., Ourselin, S., Fox, N.C. and Alexander, D.C., 2012. An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. *NeuroImage*, 60(3), pp.1880-1889.
 5. Jedynak, B.M., Lang, A., Liu, B., Katz, E., Zhang, Y., Wyman, B.T., Raunig, D., Jedynak, C.P., Caffo, B., Prince, J.L. and Alzheimer's Disease Neuroimaging Initiative, 2012. A computational neurodegenerative disease progression score: method and results with the Alzheimer's Disease Neuroimaging Initiative cohort. *Neuroimage*, 63(3), pp.1478-1486.
 6. Donohue, M.C., Jacqmin-Gadda, H., Le Goff, M., Thomas, R.G., Raman, R., Gamst, A.C., Beckett, L.A., Jack, C.R., Weiner, M.W., Dartigues, J.F. and Aisen, P.S., 2014. Estimating long-term multivariate progression from short-term data. *Alzheimer's & Dementia*, 10(5), pp.S400-S410.
 7. Schiratti, J.B., Alllassonnier, S., Routier, A., Colliot, O., Durleman, S. and Alzheimers Disease Neuroimaging Initiative, 2015, June. A mixed-effects model with time reparametrization for longitudinal univariate manifold-valued data. In International Conference on Information Processing in Medical Imaging (pp. 564-575). Springer International Publishing.
 8. Bilgel, M., Prince, J.L., Wong, D.F., Resnick, S.M. and Jedynak, B.M., 2016. A multivariate nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging. *NeuroImage*, 134, pp.658-670.
 9. Seeley, W.W., Crawford, R.K., Zhou, J., Miller, B.L. and Greicius, M.D., 2009. Neurodegenerative diseases target large-scale human brain networks. *Neuron*, 62(1), pp.42-52.
 10. Reuter, M., Schmansky, N.J., Rosas, H.D. and Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage*, 61(4), pp.1402-1418.
 11. Dickerson, B.C., Bakkour, A., Salat, D.H., Fecenko, E., Pacheco, J., Greve, D.N., Grodstein, F., Wright, C.I., Blacker, D., Rosas, H.D. and Sperling, R.A., 2009. The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. *Cerebral cortex*, 19(3), pp.497-510.
 12. Crutch, S.J., Lehmann, M., Schott, J.M., Rabinovici, G.D., Rossor, M.N. and Fox, N.C., 2012. Posterior cortical atrophy. *The Lancet Neurology*, 11(2), pp.170-178.
 13. Young, A.L., Oxtoby, N.P., Huang, J., Marinescu, R.V., Daga, P., Cash, D.M., Fox, N.C., Ourselin, S., Schott, J.M., Alexander, D.C. and Alzheimers Disease Neuroimaging Initiative, 2015, June. Multiple Orderings of Events in Disease Progression. In International Conference on Information Processing in Medical Imaging (pp. 711-722). Springer International Publishing.