DIVE: A spatiotemporal progression model of brain pathology in neurodegenerative disorders

Răzvan V. Marinescu [a,*], Arman Eshaghi [b], Marco Lorenzi [a,c], Alexandra L. Young [a], Neil P. Oxtoby [a], Sara Garbarino [a,c], Sebastian J. Crutch [d], Daniel C. Alexander [a], for the Alzheimer's Disease Neuroimaging Initiative

[a] Centre for Medical Image Computing, University College London, Gower Street, London, United Kingdom
[b] Queen Square Multiple Sclerosis Centre, UCL Institute of Neurology, Faculty of Brain Sciences, University College London, United Kingdom
[c] Epione Research-Project, Inria, France and Université Côte D'Azur, France
[d] Dementia Research Centre, University College London Institute of Neurology, London, United Kingdom

ARTICLE INFO

ABSTRACT

Current models of progression in neurodegenerative diseases use neuroimaging measures that are averaged across pre-defined regions of interest (ROIs). Such models are unable to recover fine details of atrophy patterns; they tend to impose an assumption of strong spatial correlation within each ROI and no correlation among ROIs. Such assumptions may be violated by the influence of underlying brain network connectivity on pathology propagation – a strong hypothesis e.g. in Alzheimer's Disease. Here we present DIVE: **D**ata-driven **I**nference of **V**ertexwise **E**volution. DIVE is an image-based disease progression model with single-vertex resolution, designed to reconstruct long-term patterns of brain pathology from short-term longitudinal data sets. DIVE clusters vertex-wise (i.e. point-wise) biomarker measurements on the cortical surface that have similar temporal dynamics across a patient population, and concurrently estimates an average trajectory of vertex measurements in each cluster. DIVE uniquely outputs a parcellation of the cortex into areas with common progression patterns, leading to a new signature for individual diseases. DIVE further estimates the disease stage and progression speed for every visit of every subject, potentially enhancing stratification for clinical trials or management. On simulated data, DIVE can recover ground truth clusters and their underlying trajectory, provided the average trajectories are sufficiently different between clusters. We demonstrate DIVE on data from two cohorts: the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Dementia Research Centre (DRC), UK. The DRC cohort contains patients with Posterior Cortical Atrophy (PCA) as well as typical Alzheimer's disease (tAD). DIVE finds similar spatial patterns of atrophy for tAD subjects in the two independent datasets (ADNI and DRC), and further reveals distinct patterns of pathology in different diseases (tAD vs PCA) and for distinct types of biomarker data – cortical thickness from Magnetic Resonance Imaging (MRI) vs amyloid load from Positron Emission Tomography (PET). We demonstrate that DIVE stages have potential clinical relevance, despite being based only on imaging data, by showing that the stages correlate with cognitive test scores. Finally, DIVE can be used to estimate a fine-grained spatial distribution of pathology in the brain using any kind of voxelwise or vertexwise measures including Jacobian compression maps, fractional anisotropy (FA) maps from diffusion tensor imaging (DTI) or other PET measures.

1. Introduction

Many biomarkers exist that can be used to track the severity of neurodegenerative diseases such as Alzheimer's disease (AD). Clinical function can be measured using cognitive assessments performed by an expert clinician and brain atrophy can be measured using Magnetic Resonance Imaging (MRI). Other measures include molecular markers such as aggregation of misfolded amyloid-beta or tau measured using Positron Emission Tomography (PET), and measures of white-matter degradation such as fractional anisotropy (FA) from Diffusion Tensor Imaging (DTI). The evolution of these biomarkers across the disease time-course creates a unique signature of the disease that can be used to stage

---

patients, which may potentially help stratification in clinical trials (Holland et al., 2012).

A hypothetical model of disease progression has been proposed by (Jack Jr et al., 2010), describing the order of abnormality of key biomarkers along the progression of AD. 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 hypothetical model, several data-driven disease progression models have been proposed in recent years, which aggregate information from multiple biomarkers into a single time frame representing disease progression. One such model is the Event-Based Model (EBM) (Fonteijn et al., 2012; Young et al., 2014), which models the progression of disease as a sequence of discrete events, representing underlying biomarkers switching from a normal to abnormal state. Other types of disease progression models (Donohue et al., 2014; Jedynak et al., 2012; Li et al., 2017; Lorenzi et al., 2017; Schiratti et al., 2015) have been developed, that build continuous trajectories by "stitching" together short-term follow-up data from individual subjects. In contrast to the discrete disease stages that are estimated by the EBM, these models also compute a continuous measure of disease stage for every individual by estimating individual time shifts and progression speeds.

Current image-based disease progression models estimate the evolution of the disease using a small set of biomarkers corresponding to predefined regions-of-interest (ROI). This ROI parcellation is usually coarse, doesn't allow one to find spatially dispersed patterns of atrophy. While spatiotemporal longitudinal models have already been demonstrated (Derado et al., 2010; Hyun et al., 2016; Lorenzi et al., 2015), these models regress against pre-defined sets of covariates such as age, time since baseline or clinical markers. This is problematic because, age-based alignment of subjects assumes all subjects have the same age of disease onset, while for time since baseline, its relationship with disease onset is unknown. Similarly, clinical markers are noisy, biased, suffer from floor/ceiling and training effects, are not sensitive in pre-symptomatic phases, and have low test-retest reliability (Johnson et al., 2012). Recently, some spatiotemporal models that estimate subject-specific time shifts have been developed (Bilgel et al., 2016; Koval et al., 2017). However, these models generally cannot recover dispersed and disconnected pathological patterns, because they assume correlation of voxels based on spatial distance, either through a distance function or distance from control points. However, spatially dispersed pathological patterns have been observed in AD and related dementias and are hypothesized to appear due to interactions of pathology and brain networks (Seeley et al., 2009). Discovering such fine-grained patterns could provide new insight into underlying mechanisms of pathology propagation along these brain networks (Seeley et al., 2009; Zhou et al., 2012). However, a spatiotemporal disease progression model that allows recovery of dispersed atrophy patterns present in AD, is not currently available.

In this work, we present DIVE: **D**ata-driven **I**nference of **V**ertexwise **E**volution. DIVE is a novel disease progression model with single vertex resolution that makes only weak assumptions on spatial correlation. In contrast to approaches which model temporal trajectories for a small set of biomarker measures based on *a priori* defined ROIs, DIVE models temporal trajectories for each vertex on the cortical surface. DIVE combines unsupervised learning and disease progression modelling to identify clusters of vertices on the cortical surface that show a similar trajectory of brain pathology over a particular patient cohort. This formulation enables us to estimate a fine-grained spatial distribution of pathology and also provides a novel parcellation of the brain based on temporal change. In this exploratory study, we aim to demonstrate in Alzheimer's disease DIVE's ability to identify new detail in pathology patterns, as well as evaluate the potential benefit of estimating fine-grained patterns of pathology for disease staging and prognosis. We first test DIVE on synthetic data and show that the model can recover known biomarker trajectories and disease progression scores (DPS). We

then demonstrate the model on both MRI and PET data from two cohorts: the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Dementia Research Centre (DRC), UK, containing patients with Posterior Cortical Atrophy. We use the model to reveal spatiotemporal patterns of pathology to a much finer resolution than previous models and demonstrate the ability to assign subjects to stages that predict progression. Finally, we validate DIVE in terms of how robust are the estimated pathology patterns and how well the disease progression scores correlate with cognitive tests. Source code for DIVE is available online: https://github.com/mrazvan22/dive.

2. Method

In this section we describe the mathematical formulation of DIVE (section 2.1), then we show how to fit the model using Expectation Maximisation (section 2.2) and we describe further implementation details of the algorithm (section 2.3). Afterwards, we outline the synthetic data-generation process (section 2.4) for testing the model in the presence of ground truth, as well as the pipeline for pre-processing the ADNI and DRC datasets (section 2.5).

2.1. The DIVE model

Fig. 1 illustrates the DIVE aims and implementation. As opposed to previous models (Donohue et al., 2014; Jedynak et al., 2012) that used extracted features as input, DIVE input measures are vertexwise or voxelwise biomarker measures in the brain (Fig. 1A), such as cortical thickness or amyloid load. A vertex is a location on the cortical surface at which a biomarker of pathology is quantifiable (e.g. cortical thickness). For each vertex on the cortical surface (or voxel in the 3D brain volume), we estimate a unique trajectory along the disease progression timeline (Fig. 1B), while also estimating subject/visit-specific disease progression scores (i.e. disease stages). We do that by grouping vertices with similar biomarker trajectories into clusters (Fig. 1C), and we estimate a representative trajectory for every cluster (Fig. 1D). Each trajectory is a function of subject-/visit-specific disease progression scores (DPS) (Fig. 1E). The DPS depends linearly on the time since baseline visit, but with subject-specific slope and intercept.

2.1.1. Modelling subject-specific time shifts and progression speeds

The disease progression score s_{ij} for subject i at visit j is a latent variable denoting the current disease stage of the subject at this visit. It is defined as a linear transformation of time since baseline measurement t_{ij} (in years):

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

where α_i and β_i represent the speed of progression and time shift (i.e. disease onset) of subject i. The range of DPS is $(-\infty, +\infty)$, where large positive values are indicative of advanced stages of pathology. For interpretability, we later standardise the values so that the DPS of controls follow a $N(0,1)$ distribution, so any DPS above 1 can be considered abnormal.

2.1.2. Modelling biomarker trajectory for a single vertex

DIVE assumes that the biomarker measure at each vertex on the cortical surface follows a sigmoidal trajectory $f(.; \theta)$ over the disease progression score s and with parameters θ. We choose a parametric sigmoid function because it is a parsimonious parametric model that offers better fit compared to linear models, is monotonic, and can account for floor and ceiling effects (Caroli and Frisoni, 2010; Sabuncu et al., 2011). We also assume that vertices are grouped into K clusters and we model a unique trajectory for each cluster $k \in [1, ..., K]$, which will be referred to as cluster trajectories. The trajectory function $f(s; \theta_k)$ for cluster k for is defined as:
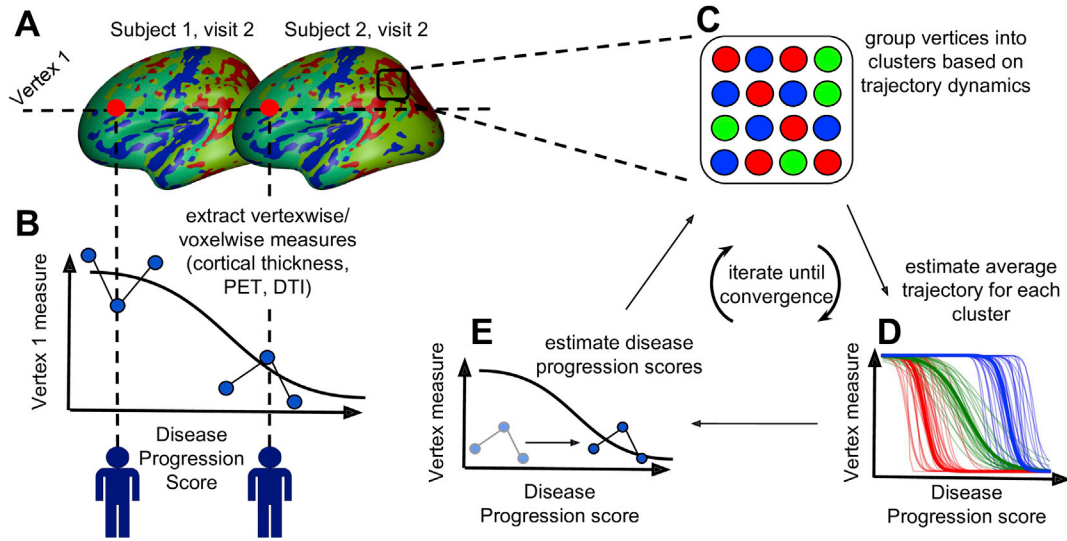
Fig. 1. Diagram of the proposed DIVE model. DIVE assumes that biomarkers of pathology (e.g. cortical thinning) can be measured at many vertices (i.e. locations) on the cortical surface (**A**), where each vertex has a distinct trajectory of change during disease progression (**B**). In (**B**), each individual has measurements for vertex 1 at three visits. DIVE assigns to every cortical vertex one of a small set of temporal trajectories describing the change in some image-based measurement (e.g. cortical thickness, amyloid PET, DTI fractional anisotropy measures) from beginning to end of the disease progression. The estimation process simultaneously estimates the set of clusters, the trajectory defining each cluster, and the disease progression scores, representing the position of each subject along the latent disease space. The process iterates assignment of each vertex to clusters (red, green and blue in this diagram) (**C**), estimation of the trajectory in each cluster (**D**) and estimation of the disease progression score for each subject (**E**), all within an expectation-maximisation framework, until convergence. In particular, (**E**) shows how the disease progression score, which is initially set to the individual's age, converges to the disease stage of the subject.

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

where s is the disease progression score from Eq. (1) and $\theta_k = [a_k, b_k, c_k, d_k]$ are parameters controlling the shape of the trajectory – d_k and $d_k + a_k$ represent the lower and upper limits of the sigmoidal function, c_k represents the inflection point and $a_k b_k/4$ represents the slope at the inflection point.

For parsimony, we model the all vertices within a cluster with the same trajectory. For a given subject i at visit j, the value V_l^{ij} of its biomarker measurement at vertex l is a random variable that has an associated discrete latent variable $Z_l \in [1, ..., K]$ 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:

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

where $N(V_l^{ij} \, |f(\alpha_i \, t_{ij} + \beta_i; \, \theta_{Zl}), \sigma_{Zl})$ represents the probability density function (pdf) of the normal distribution that models the measurement noise along the sigmoidal trajectory of cluster Z_l, having variance σ_{Zl}.

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 Eq. (1). Moreover, we also assume a uniform prior on cluster membership Z_l. This gives the following model:

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

where $I = [(i, j)]$ represents the set of all the subjects i and their corresponding visits j. Furthermore, $V_l = [V_l^{ij} \mid (i,j) \in I]$ is the 1D array of all the values for vertex l across every visit of every subject. Vectors $\alpha = [\alpha_1, ..., \alpha_S]$ and $\beta = [\beta_1, ..., \beta_S]$, where S is the number of subjects, denote the stacked parameters for the subject shifts. If a subject i has multiple visits, these visits share the same parameters α_i and β_i. Vectors $\theta = [\theta_1, ..., \theta_K]$

and $\sigma = [\sigma_1, ..., \sigma_K]$, with K being the number of clusters, represent the stacked parameters for the sigmoidal trajectories and measurement noise specific to each cluster.

2.1.3. Modelling biomarker trajectories for all vertices

So far we have a model for only one vertex on the brain surface. We therefore extend the formulation to all the vertices by assuming all these vertex measurements are spatially independent, giving the complete data likelihood:

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

where $V = [V_1, ..., V_L], Z = [Z_1, ..., Z_L]$, L being the total number of vertices on the cortical surface. The formulation assumes spatial independence between measurements in different vertices, but in section 2.1.4 the model is extended to capture spatial correlations. We get the final model log likelihood for incomplete data by marginalising over the latent variables Z:

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

Throughout the article, we will use the shorthand $z_{lk} = p(Z_l = k)$.

2.1.4. Modelling spatial correlation

The model so far assumes spatial independence between hidden Z_l variables, conditional on the parameters $[\alpha, \beta, \theta, \sigma]$. However, the regional organisation of the cortex suggests we would expect spatial correlation of the vertex measurements. More precisely, measures of cortical thickness or other modalities are often similar in neighbouring vertices on the cortical surface and likely belong to the same cluster. The model is now extended to include mild spatial constraints on the correlation of vertex measurements via a Markov Random Field (MRF),

which encourages neighbouring vertices to have the same corresponding cluster (Bishop, 2006). We hypothesise that incorporating such constraints should reduce the effects of noise and produce a more stable clustering with "filled holes". However, this does not model correlation between the actual vertex values, but only between the latent variables Z_l, i.e. the cluster membership of each vertex. The MRF thus has the advantage of not requiring the use of huge covariance matrices, which are otherwise needed if we want to model correlation of vertex values directly. More precisely, the dimensionality of Z is L (number of vertices, here around 160,000) x K (number of clusters) instead of L x L for covariance matrices. Moreover, in contrast to previous methods that use correlation based on spatial distance (Bilgel et al., 2016; Koval et al., 2017), we use neighbourhood correlations, which allow us to estimate fine-grained spatial patterns of pathology. With the MRF, the full-data likelihood function of the model now becomes:

$$p(V, Z | \alpha, \beta, \theta, \sigma, \lambda) = \prod_{l}^{L} \left[\prod_{(i,j) \in I} N\big(V_l^{ij} | f(\alpha_i t_{ij} + \beta_i; \theta_{Z_i}), \sigma_{Z_i}\big) \prod_{l_2 \in N_l} \Psi(Z_l, Z_{l_2}) \right]$$

(7)

where $\Psi(Z_l, Z_{l_2})$ is a clique term representing the likelihood of a neighbouring vertex l_2 to have similar label with vertex l. The formula for the clique term is:

$$\Psi(Z_l = k, Z_{l_2} = k_2) = \begin{cases} exp(g(\lambda)) & \text{if } k = k_2 \\ exp(-h(\lambda)) & \text{otherwise} \end{cases}$$

(8)

where λ is a parameter controlling how much to penalise neighbouring vertices that belong to distinct clusters, and g and h are positive, monotonic functions over the $\lambda > 0$ range. We choose $g(\lambda) = \lambda$ and $h(\lambda) = \lambda^2$, which results in a concave objective function for λ, ensuring that it can later be optimised (see M-step).

Therefore, the model parameters that need to be estimated are $M = [\alpha, \beta, \theta, \sigma, \lambda]$ where α and β are the subject specific shifting parameters, θ and σ are the cluster specific trajectory and noise parameters and λ controls the penalisation of vertices belonging to different clusters.

2.2. Fitting the model using generalised expectation-maximisation

We choose to fit our model using Expectation-Maximisation (EM), because it offers a fast convergence given the large number of parameters that need to be estimated and the huge dimensionality of relevant datasets (e.g. 1973 subjects x 163,842 vertices in ADNI). This is because an approximate form of the incomplete-data log-likelihood is used that can be evaluated (i.e. is computationally tractable). Moreover, highly-coupled parameters are optimised jointly, while low-coupled parameters are optimised independently. In the next two sections we outline the E-step and M-step. While both of these steps have no closed-form solution, we will solve them using numerical optimisation, which only results in an increase in the objective function at each iteration. However, the EM algorithm is still guaranteed to converge, and this approach is called Generalised EM (Bishop, 2006).

Algorithm 1 shows the model fitting procedure using the EM algorithm. The procedure first initialises (line 1) some parameters required to start the EM algorithm: the subject-specific parameters α and β and the latent parameters z_{lk} which represent the assignment of vertices to clusters. In the M-step, the method updates the trajectories of each cluster (lines 4–6), the subjects-specific parameters (line 9) and the clique penalty term λ (line 17). In the E-step, the method computes z_{lk} (line 18) using previously defined functions that compute z_{lk} given a fixed λ (line 14).

Algorithm 1. The optimisation procedure for fitting DIVE using the Expectation-Maximisation algorithm.

```
1  Initialise α^(0), β^(0), z_lk^(0)
2  while θ, σ, α, β or z_lk not converged do
     ; // M-step 1:  For each cluster, optimise its trajectory
3      for k = 1 to K do
4          θ_k^(u) = arg min_θk Σ_l=1^L z_lk^(u-1) Σ_(i,j)∈I (V_l^ij - f(α_i^(u-1) t_ij + β_i^(u-1) | θ_k))^2 - log p(θ_k)
5          θ_k^(u) = make_identifiable(θ_k^(u))
6          (σ_k^(u))^2 = 1/|I| Σ_l=1^L z_lk^(u-1) Σ_(i,j)∈I (V_l^ij - f(α_i^(u-1) t_ij + β_i^(u-1) | θ_k^(u)))^2 - log p(σ_k)
7      end
     ; // M-step 2:  For each subject, optimise its time shift α_i and progression speed β_i
8      for i = 1 to S do
9          α_i^(u), β_i^(u) = arg min_αi,βi [ Σ_l=1^L Σ_k=1^K z_lk^(u-1)/(2(σ_k^(u-1))^2) Σ_j∈I_i (V_l^ij - f(α_i t_ij + β_i | θ_k^(u)))^2 ) - log p(α_i, β_i) ]
10     end
     ; // E-step 1:  Define functions ζ_lk(λ) computing z_lk, the probability of vertex l being assigned
     // to cluster k, given fixed λ
11     for l = 1 to L do
12         for k = 1 to K do
             ; // Pre-compute data fit terms D_lk
13             D_lk = -1/2 log (2π(σ_k^(u))^2)|I| - 1/(2(σ_k^(u))^2) Σ_i,j∈I (V_l^ij - f(α_i^(u) t_ij + β_i^(u) | θ_k^(u)))^2
14             ζ_lk(λ) ≈ exp ( D_lk + Σ_l2∈N_l log [ exp(-λ^2) + z_l2k^(u-1)(exp(λ) - exp(-λ^2)) ] )
15         end
16     end
     ; // M-step 3:  optimise clique term λ using above definitions in E-step 1
17     λ^(u) = arg max_λ Σ_l=1^L Σ_k=1^K ζ_lk(λ) [ D_lk + λ Σ_l2∈N_l ζ_l2k(λ) - λ^2 Σ_l2∈N_l (1 - ζ_l2k(λ)) ]
     ; // E-step 2:  Compute next z_lk using the best λ
18     z_lk^(u) = ζ_lk(λ^(u))
19 end
20 α_i^(u) = α_i^(u)/σ_N, β_i^(u) = β_i^(u) - μ_N / σ_N ; // Re-scale subject shifts
```

2.2.1. E-step

In the Expectation step, at iteration u we seek an estimate of $p(Z | V, M^{(u-1)})$, given the current estimates of the parameters $M^{(u-1)} = [\theta_k^{(u-1)}, \sigma_k^{(u-1)}, \alpha_i^{(u-1)}, \beta_i^{(u-1)}, \lambda_i^{(u-1)}]$. We perform this using Iterated Conditional Modes (Bishop, 2006), which performs coordinate-wise gradient ascent. This works by conditioning the clique terms Z on the values of Z from the previous iterations. This approximation gives the following factorizable likelihood:

$$p\big(Z | V, M^{(u-1)}\big) \approx p\big(Z | V, M^{(u-1)}, Z^{(u-1)}\big) = \prod_{l=1}^{L} p\Big(Z_l | V_l, M^{(u-1)}, Z_{N_l}^{(u-1)}\Big)$$

(9)

The factorised form allows for tractable computation and memory storage of $p(Z)$.

Let $z_{lk}^{(u)} = p(Z_l = k \mid V_l, M^{(u-1)}, Z^{(u-1)})$. After simplifications we reach the following update rule:

$$\log z_{lk}^{(u)} \propto \left[\frac{-\log (2\pi\sigma_k^2)|I|}{2} - \sum_{i,j\in I} \frac{1}{2\sigma_k^2} \big(V_l^{ij} - f(\alpha_i t_{ij} + \beta_i; \theta_k)\big)^2 \right]$$

$$+ \left[\sum_{l_2 \in N_l} \log \big[exp(-\lambda^2) + z_{l_2 k}^{(u-1)}(exp(\lambda) - exp(-\lambda^2)) \big] \right]$$

(10)

The full derivation is given in the Supplementary material. In order to enable optimisation over λ, a final modification of this step is performed, by considering z_{lk} to be functions $\zeta_{lk}(\lambda)$ over λ. This results in the update equation from Alg. 1, line 18 which is based on pre-defined terms on lines 13–14.

2.2.2. M-step

In the Maximisation step we try to estimate the model parameters $M = (\alpha, \beta, \theta, \sigma, \lambda)$ that maximise $E_{Z|V,M}^{(u-1)}[log \, p(V,Z|M)]$. We cannot simultaneously optimise all 5 sets of parameters, so we optimise them independently. 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. The key observation here is that if we assume fixed α, β and Z, then the trajectory parameters θ_k for every cluster k are conditionally independent, i.e. $\theta_k \perp \theta_m \mid (Z, \alpha, \beta, \sigma) \, \forall \, (k, m), \, k \neq m$. This allows us to maximise every θ_k independently using the following equation:

$$\theta_k = \arg\max_{\theta_k} \sum_{z_1,\dots,z_L}^{K} p\big(Z$$

$$= (z_1,\dots,z_L)|V,M^{(u-1)}\big) \log\left[\prod_{l=1}^{L} \prod_{(i,j)\,\in\,I} N\big(V_l^{ij}\big|f\big(\alpha_i t_{ij} + \beta_i; \theta_{z_l}\big), \sigma_{z_l}\big)\right]$$

$$+ \log p(\theta_k) \tag{11}$$

A similar observation of conditional independence can also be observed for the latent variables Z. This allows us to decompose the joint distribution over Z, and after expanding the noise model we reach the optimisation problem from Alg. 1, line 4. See Supplementary material for full derivation. This does not have a closed-form solution, so we use numerical optimisation for finding θ_k that maximises the equation from Alg. 1, line 4.

A similar equation, yet in closed form, is also obtained for σ_k (Alg. 1, line 6). After estimating θ and σ for every cluster, we use the new values to estimate the subject specific parameters α and β. For every subject i, we maximise the expected log likelihood with respect to α_i, β_i independently, and after simplifications we obtain the update rule from Alg. 1, line 9, which is again solved using numerical optimisation. For the numerical optimisation of θ we used the Nelder-Mead method for its robustness, while for α and β we used the second-order Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm due to fast convergence. Finally, we achieved a significant speed-up in the evaluation of objective functions by computing a z_{lk}-weighted average of vertex measurements within each cluster (see Supplementary section 5). This resulted in a convergence time of around 6 h for the larger datasets (ADNI).

For optimising λ, we again try to optimise in the M-step the expected full data likelihood under the Z estimates from the previous iteration:

$$\lambda^{(u)} = \arg\max_{\lambda} E_{p(Z|V,M^{(u-1)},\lambda,Z^{(u-1)})}\left[\log p\big(V,Z|M^{(u-1)}\big)\right] \tag{12}$$

We simplify the above equation by expanding the likelihood model and approximating the joint over Z with the product of the marginals z_{lk} over all vertices l. This results in the update equation from Alg. 1 line 17 – see supplementary material for full derivation. In this final equation we also replaced z_{lk} with a function $\zeta_{lk}(\lambda)$ over λ, which updates z_{lk} based on the current value of λ being evaluated. This is done to increase convergence, as latent variables z_{lk} are highly coupled with the value of λ being evaluated.

2.3. Implementation details

2.3.1. Parameter initialisation and priors

Before starting the fitting process, we need to initialise α, β and the clustering probabilities z_{lk} (Alg. 1, line 1). We set α_i and β_i to be 1 and 0 respectively for each subject, which sets the initial disease progression score to the time since baseline of the subject at the clinical visit. In line with previous approaches that used models of increasing complexity for parameter initialisation (Bilgel et al., 2016), we initialise z_{lk} using a simpler clustering method, k-means clustering, of the vectors V_l. We also tried other informed starting points for z_{lk} based on splitting the histogram of V into K parts, one for each cluster, as well as atlas based initialisation. However, we found that the initialisation using k-means always converged to the solution with highest likelihood. For initialisation of θ, we set each sigmoid parameter (minimum, maximum, centre, slope) based on the biomarker (i.e. vertex) values and the DPS range of the subjects; we also initialise λ to 1. At each EM loop, to ensure the algorithm doesn't get stuck in a local minimum, we also use 15 perturbed starting points for each θ_k, 3 starting points for α_i and β_i and 20 for λ. We take the best solution every time.

To ensure model identifiability, we need to account for extra degrees of freedom in DIVE. First, the DPS scores have two extra degrees of freedom (scale and shift), which we account for by setting informative gamma and Gaussian priors on parameters α_i and β_i respectively ($\alpha_i \sim \Gamma(16e4, 16e4)$, $\beta_i \sim N(0, 0.1)$). These informative priors further aid convergence and work well in practice as they result in realistic ranges for α_i and β_i of around [0.3, 3] and [-15,15] respectively. Such informative priors on α_i and β_i and help deal with singularities in the objective functions of α_i and β_i when the biomarker trajectories are flat. Secondly, as already explained in (Jedynak et al., 2012), the sigmoid parameters θ_k are not identifiable, so we need to apply the following transformation on line 5 of Alg. 1: if $b_k^{(u)} < 0$ then $a_k^{(u)} = -a_k^{(u)}$; $b_k^{(u)} = -b_k^{(u)}$; $d_k^{(u)} = d_k^{(u)} - a_k^{(u)}$.

To make the DPS scores more interpretable, we normalise them so that the DPS scores of controls have a mean of 0 and standard deviation of 1 (Alg. 1, line 20). This means that a DPS score above 1 will reflect the presence of pathology. The DPS scores will also be comparable between different cohorts if the same control population is used (e.g. DRC tAD vs DRC PCA, see section 2.5). However, DPS scores are not comparable between different datasets (ADNI vs DRC) or different modalities (MRI vs PET).

As a Bayesian generative model, DIVE can automatically deal with missing data by restricting the objective function (Alg 1 lines 4,9 and 12) to only use available data.

2.3.2. Estimating the optimal number of clusters

The EM procedure needs to specify a-priori the number of clusters to fit on the data. We optimise the number of clusters K using Akaike Information Criterion (AIC), which we found to better agree with ground truth in simulations than other information criteria such as the Bayesian Information Criteria (BIC). The number of parameters of the fitted model is $5K+2S + 1$, where S is the number of subjects. Note that z_{lk} are not included as parameters of the model because they are latent variables that are marginalised (see Eq. (6)). We repeat the fitting procedure for each K from 2 to 100 clusters and select the K that minimises the AIC.

2.4. Simulation experiments

2.4.1. Motivation

Initial assessment of DIVE performance uses synthetic data, where we know the ground truth. The aim is to explore how accurately we can recover ground truth parameters as the problem becomes harder in three different scenarios:

Scenario 1: as the number of clusters increases, evaluate how well DIVE can estimate the correct number of clusters using AIC and BIC.

Scenario 2: as the trajectories become more similar, test how well we can recover the assignment of vertices to clusters and the DIVE parameters.

Scenario 3: same as Scenario 2, but for decreasing number of subjects.

2.4.2. Synthetic data generation

We first designed a basic simulation, which the model should be able to fit well since the trajectories were designed to be well separated and enough subject data was generated along the disease time course. The data in the basic simulation was generated as follows:

1. Sample age a_{i1} and subject-specific parameters α_i, β_i for 300 subjects with 4 timepoints (each timepoint 1 year apart), with $a_{i1} \sim U(40,80)$, $\alpha_i \sim \Gamma(6.25, 6.25)$, $\beta_i \sim N(0, 10)$. Time since baseline has been obtained for every visit j of subject i as follows: $t_{ij} = a_{ij} - a_{i1}$
2. Generate three sigmoids with different (slope, centre) parameters: [(-0.1, −15), (−0.1, 2.5), (−0.1, 20)] (Fig. 2A, red lines). Upper and lower limits have been set to 1 and 0 respectively.
3. Randomly assign every vertex $l \in [1, \dots, L]$ where $L = 1000$, to a cluster $a[l] \in [1,2,3]$
4. Sample a set of L perturbed trajectories θ_l from each of the original trajectories, one for each vertex (Fig. 2A, gray lines) using covariance matrix $C_\theta = diag([0, 2b_k/15, 11.6, 0])$.
5. Sample subject data for every vertex l from its corresponding perturbed trajectory θ_l with noise standard deviation $\sigma_l = 1$.

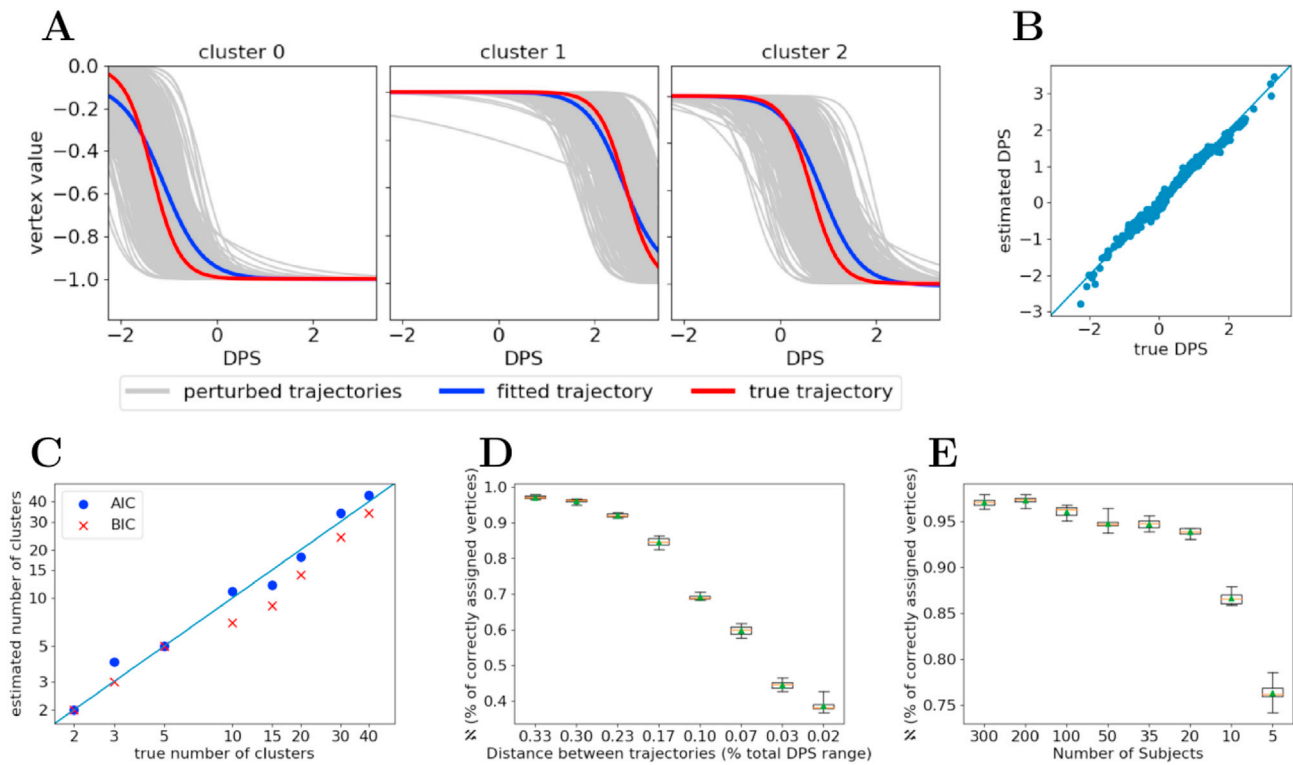From the basic simulation, we generated synthetic data for each of the

Fig. 2. (A–B) Results for the basic simulation, where trajectories are relatively well separated. (A) Reconstructed temporal trajectories (blue) plotted against the true trajectories (red). The x-axis shows the disease progression score (DPS), while the y-axis shows the biomarker values of the vertices. (B) Estimated subject-specific DPS scores compared to the true scores. (C–E) Simulation results for the three scenarios: (C) increasing number of clusters, (D) trajectories becoming similar and (E) decreasing number of subjects. On the x-axes we show the variable that was changing within each scenario: true number of clusters, distance between trajectory centres and the number of subjects. X-axis tick marks are equally spaced to improve visualisation, with increasing problem complexity from left to right. On the y-axis we show the agreement measure $\varkappa$, representing the percentage of vertices that were assigned to the correct cluster.

three scenarios by varying one parameter at a time and kept the other parameters constant, having the same values as in the basic simulation. We varied the following parameters:

Scenario 1: number of clusters - 2, 3, 5, 10, 15, 20, 30 and 40. The cluster centres were spread evenly across a fixed total DPS range where data was available.

Scenario 2: distance δ between trajectory centres $\delta = c_2 - c_1 = c_3 - c_2$ (as proportion of total DPS range sampled) – 0.33, 0.30, 0.23, 0.17, 0.10, 0.07, 0.03 and 0.02.

Scenario 3: number of subjects - 300, 200, 100, 50, 35, 20, 10 and 5.

2.4.3. Model fitting and evaluation

Since there was no spatial information in the data generation procedure, we used DIVE without the MRF extension. For Scenario 1, we estimated using AIC and BIC the optimal number of clusters. For Scenarios 2 and 3, after fitting the parameters of DIVE, we calculated the agreement between the final clustering probabilities $p(Z_l)$ and the true clustering assignments $a[l]$. This agreement, which we will call the *clustering agreement,* is defined as $= \varkappa max_\tau (1/L)\Sigma^L_{l=1} p(Z_l = \tau(a[l]))$, *where τ is any permutation of cluster labels.* We also computed the error in the DPS estimation (sum of squared differences, SSD) and trajectory estimation (SSD between predicted trajectory and true trajectory at DPS points of every subject visit).

2.5. Data acquisition and pre-processing

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 MR

images that have undergone gradient warping, intensity correction, and scaling for gradient drift. We included subjects that had at least 3 scans, to ensure we get a robust estimate of the subject specific parameters. This resulted in 138 healthy controls, 235 subjects with mild cognitive impairment (MCI) and 81 subjects with Alzheimer's disease.

We also downloaded all AV45 PET images from ADNI that were fully pre-processed, having the tag *'Co-reg, Avg, Std Img and Vox Siz, Uniform Resolution'*. This meant that the images were co-registered, averaged across the 6 five-minute frames, standardised with respect to the orientation and voxel size and smoothed to produce a uniform resolution of 8 mm full-width/half-max (FWHM).

The DRC dataset consisted of T1 MRI scans from 26 healthy controls, 23 PCA and 20 typical AD subjects with at least 3 scans each. All PCA patients fulfilled the (Tang-Wai et al., 2004) criteria and (Mendez et al., 2002) criteria based on clinical review. The typical AD patients all met the criteria for probable Alzheimer's disease (Dubois et al., 2007, 2010).

Given that the ADNI and DRC datasets contained subjects with different modalities or diseases, we ran DIVE independently on the following four cohorts (see Table 1 for demographics):

1) *ADNI MRI*: controls, MCI and tAD subjects from ADNI (cortical thickness data)
2) *DRC tAD*: tAD subjects and controls from the DRC dataset (cortical thickness data)
3) *DRC PCA*: PCA subjects and controls from the DRC dataset (cortical thickness data)
4) *ADNI PET*: AV45 scans from ADNI containing subjects with following diagnoses: healthy controls, subjective memory complaints, early MCI, late MCI and Alzheimer's disease.

Table 1
Demographics of the three cohorts used in our analysis. ADNI MRI and the DRC cohorts were used for the cortical thickness analysis, while ADNI PET was used for the PET AV45 analysis. MCI – mild cognitive impairment, SMC - subjective memory complaints, EMCI – early MCI, LMCI – late MCI.

Diagnosis	Number of Subjects	Number of Scans	Age (baseline)	Gender (% male)
ADNI MRI				
CTL	138	4.4 ± 1.1	76.3 ± 5.2	52%
MCI	235	4.6 ± 1.3	74.8 ± 7.4	39%
AD	81	3.5 ± 0.6	75.8 ± 7.7	56%
DRC tAD				
CTL	26	5.0 ± 1.7	66.3 ± 5.6	46%
AD	20	5.5 ± 1.6	71.2 ± 7.1	50%
DRC PCA				
CTL	26	5.0 ± 1.7	66.3 ± 5.6	46%
PCA	23	4.2 ± 1.3	62.7 ± 6.5	30%
ADNI PET				
CTL	141	2.4 ± 0.5	85.6 ± 3.7	46%
SMC	27	2.0 ± 0.0	86.1 ± 4.0	59%
EMCI	149	2.4 ± 0.5	85.6 ± 3.6	38%
LMCI	104	2.4 ± 0.5	86.0 ± 3.6	55%
AD	12	2.0 ± 0.0	87.3 ± 3.8	25%

2.5.1. Image pre-processing

2.5.1.1. MR images. On both datasets, in order to extract reliable cortical thickness measures, we ran the Freesurfer longitudinal pipeline (Reuter et al., 2012), which first registers the MR scans to an unbiased within-subject template space using inverse-consistent registration. The longitudinally registered images were then registered to the average Freesurfer template. No further smoothing was performed on these images (FWHM level of zero mm). From these template-registered volumetric images, cortical thickness measurements were computed at each vertex (i.e. point) on an average 2D cortical surface manifold. 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.

2.5.1.2. PET AV45 images. We computed amyloid SUVR levels using the PetSurfer pipeline (Greve et al., 2014, 2016), which is available with Freesurfer version 6. The PetSurfer pipeline first registers the PET image with the corresponding MRI scan, then applies Partial Volume Correction, and finally resamples the voxelwise SUVR values onto the cortical surface. While the final images also had a resolution of 163,842 vertices, the PET data we obtained from ADNI was inherently more smooth than the MRI cortical thickness data (8 mm FWHM).

2.5.2. The MRF neighbourhood graph

We estimated the MRF neighbourhood graph based on a Freesurfer triangular mesh for the *fsaverage* template. Each vertex was a triangle on the brain surface estimated with Freesurfer, and we connected the vertices if the corresponding triangles had a shared edge. For the MRF neighbourhood graph, we used a 3rd degree neighbourhood structure, meaning that two vertices were considered neighbours if the shortest path between them was not higher than 3.

3. Results

We first present results on synthetic data (section 3.1), then on ADNI and DRC datasets (section 3.2), followed by model evaluation (section 3.3) using cross-validation and correlation with cognitive markers.

3.1. Results on synthetic data

In the basic simulation, we obtained a clustering agreement $\varkappa$ of *0.97*,

which suggests that almost all vertices were assigned to the correct cluster. Fig. 2A 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. In Fig. 2B we plotted the recovered DPS of each subject along with the true DPS. The results for the three scenarios are shown in Fig. 2C–E. In Fig. 2C, we show for Scenario 1 the estimated number of clusters against the true number of clusters using both AIC and BIC criteria. In Fig. 2D and E we show the distributions for $\varkappa$ in Scenarios 2 and 3 as the problem becomes harder in each successive step.

The results show that, in a simple experiment where the trajectories are well separated, DIVE can very accurately estimate which clusters generated each vertex. Moreover, the recovered trajectories and DPS scores are close to the true values. The results of Scenario 1 also suggest that both AIC and BIC are effective at estimating the correct number of known clusters, with AIC having slightly better performance than BIC for larger numbers of clusters. On the other hand, the results of the stress test scenarios 2 and 3 show that performance measure $\varkappa$ drops when the trajectories become very similar with each other or when the number of subjects decreases. This happens because small differences in trajectories are hard to detect in the presence of measurement noise, while a small number of subjects doesn't provide enough data to accurately estimate the parameters. Similar decreases in performance for scenarios 2 and 3 are observed also for other measures, such as the error in recovered trajectories or DPS scores (Supplementary Fig. 1).

3.2. Results on ADNI and DRC datasets

3.2.1. Initial hypotheses

Using ADNI and DRC datasets, we aim to recover the spatial distribution of cortical atrophy and amyloid pathology, as well as the rate and timing of these pathological processes. In particular, we hypothesise that these spatial patterns of pathology and their evolution will be:

1. Similar on two independent typical AD datasets: ADNI and DRC
2. Different on distinct diseases: tAD vs PCA
3. Different in distinct modalities: cortical thickness from MRI vs amyloid load from AV45 PET.

3.2.2. Results

The optimal number of clusters, as estimated with AIC, was three for the ADNI MRI dataset, three for the DRC tAD dataset, five for the DRC PCA dataset and eighteen for the ADNI PET dataset. Fig. 3A-left shows the results from the ADNI MRI dataset, where in the left image we coloured the vertices on the cortical surface according to the cluster they most likely belong to. We assigned a colour for each cluster (both the brain figures on the left and the trajectory figures on the right) according to the extent of pathology of its corresponding trajectory at a DPS score of 1. The cluster colours range from red (severe pathology) to blue (moderate pathology). In Fig. 3A-right, we show the resulting cluster trajectories with samples from the posterior distribution of each θ_k. Similar results are shown for the other three datasets: the DRC tAD dataset (Fig. 3B), DRC PCA dataset (Fig. 3C) and the ADNI PET dataset (Fig. 3D). To better visualise the continuous progression of pathology, we show movies of the pathology progression for all four datasets (Movie 1, online version only).

Supplementary video related to this article can be found at https://doi.org/10.1016/j.neuroimage.2019.02.053.

We notice that in tAD subjects using the ADNI datasets (Fig. 3A), there is more severe cortical thinning mainly in the inferior temporal lobe (red cluster), with disperse atrophy also in parietal and frontal regions (green cluster), with relative sparing of the inferior frontal and occipital lobes. In tAD subjects from the DRC dataset, we see a relatively similar pattern, however with more pronounced atrophy in the supramarginal cortex (red cluster) compared to ADNI. The spatial distribution of cortical thinning found with DIVE resembles results from previous longitudinal studies
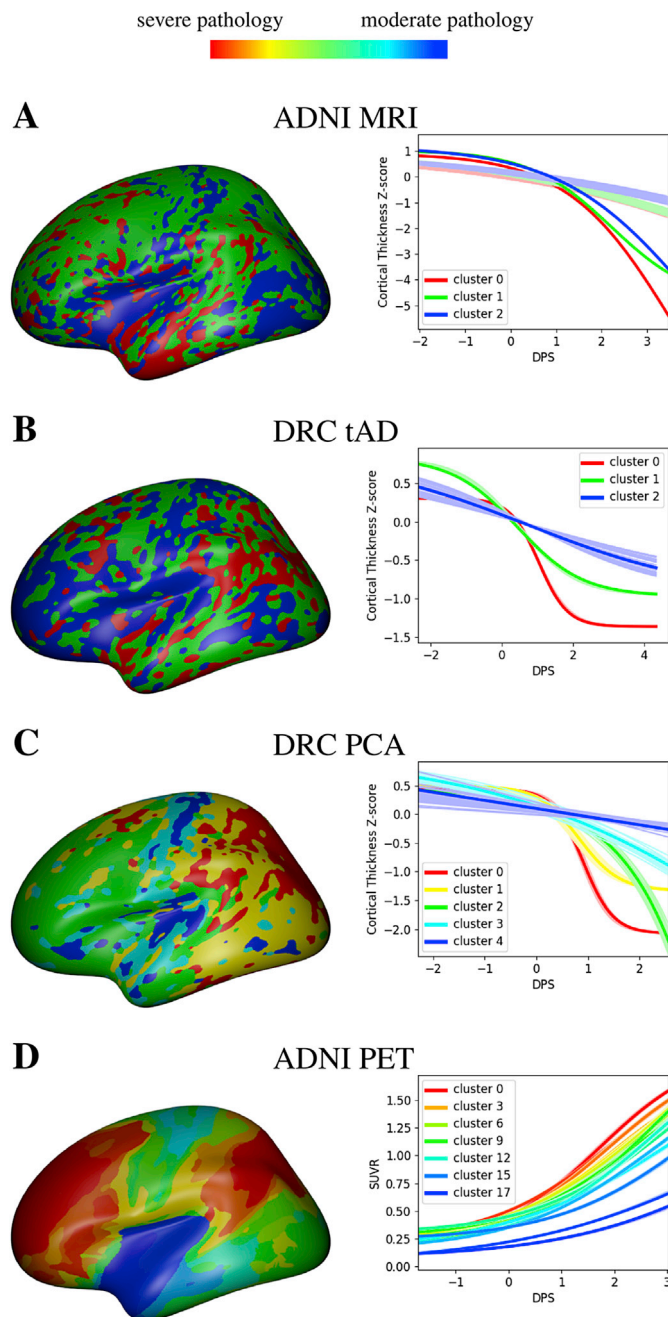
Fig. 3. (left column) DIVE estimated clusters (left column) and corresponding disease progression trajectories (right column) on four datasets: (A) ADNI MRI (B) DRC tAD (C) DRC PCA and (D) ADNI PET. We coloured each cluster according to the extent of pathology (cortical thickness or amyloid uptake) at DPS = 1.

such as (Dickerson et al., 2008; Singh et al., 2006). However, in contrast to these approaches, our model gives insight into the timing and rate of atrophy and is also able to stage subjects across the disease time course. We also find that the cluster trajectories in the DRC tAD dataset have similar dynamics to the ADNI MRI dataset, although they show a clearer separation between each other.

In the PCA subjects (Fig. 3C), we find that atrophy is mainly focused on the posterior part of the brain, with limited spread in the motor cortex, anterior temporal and frontal areas. This posterior-focused pattern of atrophy is different from the one found in the tAD datasets, and agrees with previous findings in the literature (Crutch et al., 2012; Lehmann et al., 2011). However, as opposed to the results from (Lehmann et al.,

2011), we notice that there are two clusters within the posterior region with different pathology dynamics, with the superior parietal and supramarginal areas affected more that the remaining posterior regions. This might be attributable to DIVE's ability to model the disease onset and progression speed for every individual and non-linear cortical thinning dynamics. However, other differences in the datasets and pre-processing pipelines might also influence the results.

In ADNI PET (Fig. 3D) we see that the regions with the highest amyloid uptake are more spatially continuous, comprising the precuneus and anterior frontal areas. On the other hand, the anterior-superior temporal gyrus shows the least uptake of amyloid. This result closely matches the result by (Bilgel et al., 2016), which used a completely different dataset and modelling technique. The "layers of clusters" starting from the precuneus and frontal lobes, which range from severe to less severe atrophy, suggest a continuum of variation in vertex trajectories in the case of the PET dataset (Fig. 3D-right). These trajectories all start with a low amyloid SUVR, between 0 and 0.25, but in late stages the trajectories for some clusters such as cluster 0 can reach an SUVR of 1.5. The reason for seeing this continuum might be because the PET images have a much lower resolution than MR images and were smoothed by ADNI during the pre-processing steps.

3.3. Model evaluation

3.3.1. Motivation
We further tested the robustness and validity of the model as follows:

1. *Robustness in parameter estimation*: test whether similar spatial clustering is estimated for different subsets of the data
2. *Clinical validity of DPS scores*: test whether the subject disease progression scores, based purely on MRI or PET data, correlate 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).
4. *Comparison with other models*: to evaluate the benefit of estimating fine-grained patterns of pathology in DIVE, as well as latent time shifting of subjects, we compared the performance of DIVE with two standard approaches: (1) an *ROI-based method,* which groups vertices based on an a-priori atlas like the approach of (Jedynak et al., 2012) – this is a special case of DIVE where latent variables z_{lk} are fixed instead of being marginalised – and (2) a *no-staging model* that doesn't estimate subject time shifts and progression speeds (i.e. fixes $\alpha_i = 1$, $\beta_i = 0$).

3.3.2. Evaluation procedure
For all scenarios, we ran 10-fold cross-validation (CV) on the ADNI MRI dataset. At each fold we fit the model using 3 clusters, since this was the optimal number of clusters found previously on the entire dataset. The trained model was then used to estimate the DPS of the test subjects.

For the performance comparison of DIVE with other models, we compute two performance metrics: (1) between-subject correlation of the models' estimated DPS values with cognitive tests, which evaluates the models' ability to map brain-behaviour relationships; here we estimated a unique DPS for every subject and every visit, which we then matched with the corresponding cognitive tests at that subject's visit and (2) prediction root mean squared error (RMSE) between the predicted vertex-wise values and actual measurements, averaged over all subjects and all locations on the brain; to evaluate these predictions, for every subject we use the first *n-1* scans for training and the last scan for testing the prediction.

3.3.3. Evaluation results
Fig. 4 shows the brain clusters and corresponding trajectories, estimated for all the cross-validation folds after fitting the model on the training data. The clusters have been coloured using a similar colour
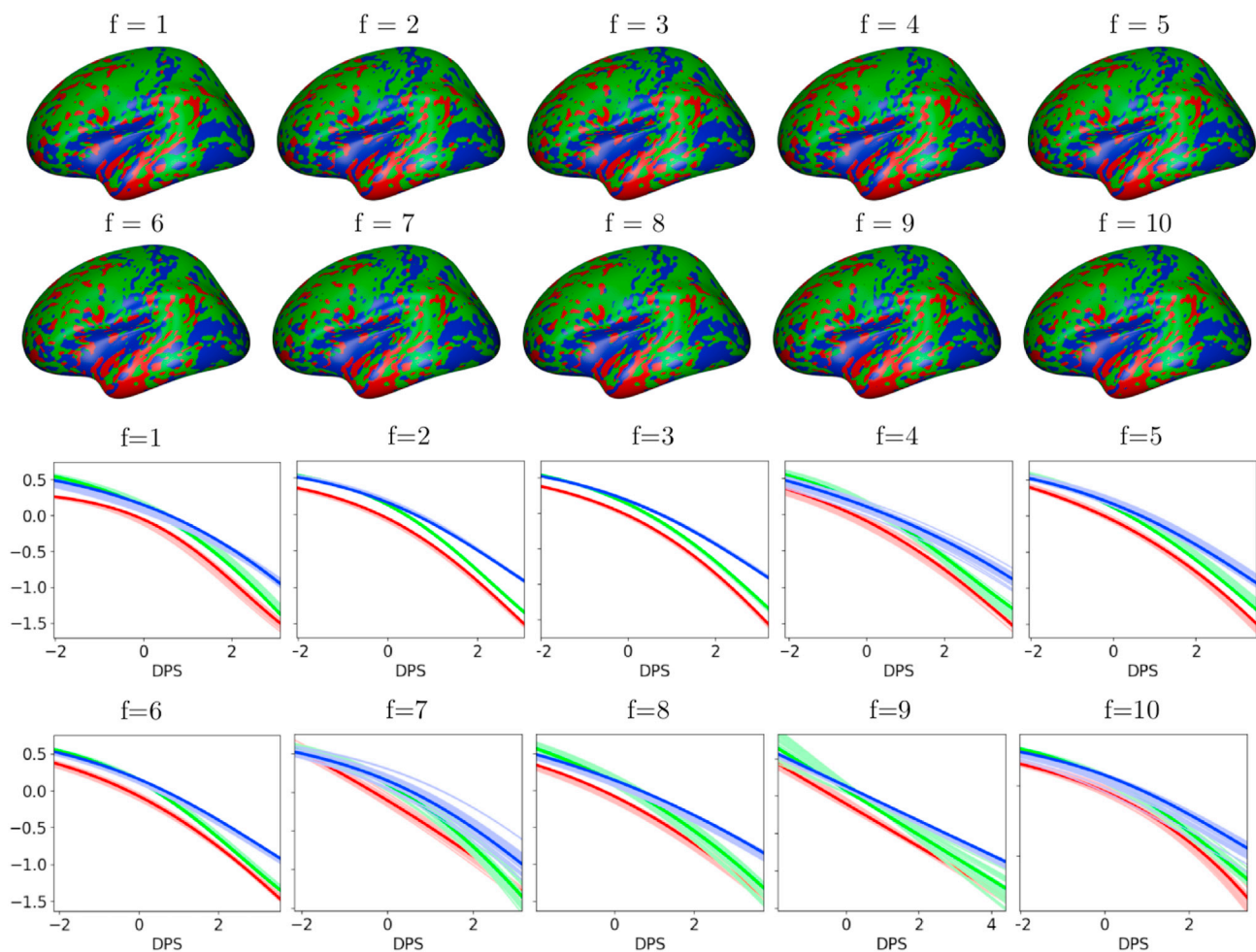
Fig. 4. (top) Clusters estimated from 10-fold cross-validation training sets on the ADNI MRI dataset. (bottom) Estimated trajectories for each fold.

scheme as in Fig. 3. In Fig. 5 we show scatter plots of the DPS scores with clinical measures such as CDRSOB, ADAS-COG, MMSE and RAVLT.

The results in Fig. 4 demonstrate that DIVE is robust in cross-validation, as the estimated clusters and trajectory parameters are all similar across folds. The average Dice score overlap across the 10-folds range were 0.77, 0.76 and 0.90 for clusters 0, 1 and 2 respectively. The DIVE-derived DPS scores, which were estimated purely based on MRI data, are also clinically relevant as they correlate with cognitive tests (Figs. 5, *0.32 < ρ 0.38, p 1e-38*). These results are also comparable with correlations between whole brain cortical thickness measures and cognitive tests: CDRSOB ($\rho = -0.40$, $p < 1e-76$), ADAS-Cog ($\rho = -0.40$, $p < 1e-76$), MMSE ($\rho = 0.40$, $p < 1e-76$), RAVLT ($\rho = 0.34$, $p < 1e-56$).

Moreover, all the subject-specific parameters (time shifts, progression speeds) and the DPS scores show significant differences ($p < 0.013$) between the APOE-positive and -negative groups (Supplementary Fig. 2), suggesting they correlate also with genetic features.

When evaluating the DIVE's performance in terms of subject staging and biomarker prediction (Table 2), it performs better than the *no-staging* model due to its ability to estimate subject-specific parameters, and has similar performance to the *ROI-based* model. However, as opposed to the more parsimonious *ROI-based* model, DIVE offers qualitative insight into the fine-grained spatial patterns of pathology and their temporal progression.
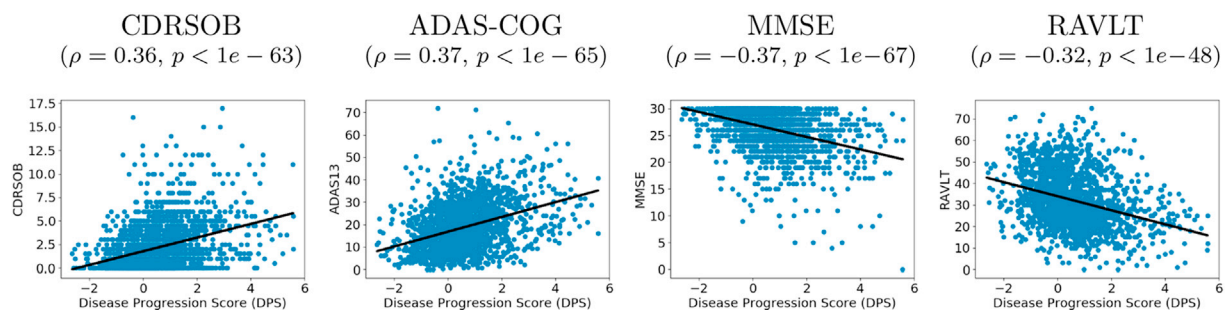

Fig. 5. Scatter plots of the DPS scores estimated from the ADNI MRI dataset, plotted against four cognitive tests: CDRSOB, ADAS-COG, MMSE and RAVLT. For each cognitive test we also report the correlation coefficient and p-value. The disease progression scores, computed only based on MRI cortical thickness data, correlate with these cognitive measures, suggesting that the DPS scores are clinically meaningful.

Table 2

Performance evaluation of DIVE and two other models on the ADNI MRI dataset using 10-fold cross-validation. In the middle four columns, we show between-subject correlations between the DPS scores and several cognitive tests: CDRSOB, ADAS-Cog 13, MMSE and RAVLT. The last column shows the prediction error (RMSE) of cortical thickness values for follow-up scans. (*) Statistically significant differences between the model and DIVE, Bonferroni corrected for multiple comparisons.

Model	CDRSOB (ρ)	ADAS13 (ρ)	MMSE (ρ)	RAVLT (ρ)	Prediction (RMSE)
DIVE	0.37±0.09	0.37±0.10	0.36±0.11	0.32±0.12	1.021±0.008
ROI-based model	0.36±0.10	0.35±0.11	0.34±0.13	0.30±0.13	1.019±0.010
No-staging model	*0.09±0.06	*0.03±0.09	*0.05±0.06	*0.02±0.06	*1.062±0.024

4. Discussion

4.1. Summary and key findings

We presented DIVE, a spatiotemporal model of disease progression that clusters vertex- or voxel-wise measures of pathology in the brain 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. The model also estimates the temporal shift and progression speed for every subject. We applied the model on cortical thickness vertex-wise data from three MRI datasets (ADNI, DRC tAD and DRC PCA), as well as an amyloid PET dataset (ADNI). Our model found qualitatively similar patterns of cortical thinning in tAD subjects using the two independent datasets (ADNI and DRC). Moreover, it also found different patterns of pathology dynamics on two distinct diseases (tAD and PCA) and on different types of data (PET and MRI-derived cortical thickness). Finally, DIVE also provides a new way to parcellate the brain that is specific to the temporal trajectory of a particular disease, and enables staging of individuals at risk of disease, which may potentially help stratification in clinical trials.

4.2. Effect of training data on DIVE results

The characteristics of the subjects used for training can affect the DIVE parameters. For example, while we could not include MCI subjects in the DRC dataset due to unavailability, such subjects in middle stages should be included to accurately estimate the DIVE parameters. Moreover, in cortical thinning analyses we standardised the data with respect to controls, which might have already shown cortical thinning due to early pathology. This can be mitigated through enrichment of the control population to amyloid-negative individuals. To reliably estimate the subject-specific parameters, multiple follow-up scans are required. We mitigated this by using only subjects with at least three scans, and further placing informative priors on these parameters. However, DIVE can still be used on cross-sectional datasets due to the informative priors placed on subject-specific para meters; however, we expect higher parameter uncertainty in these regimes.

4.3. Interpretation of spatial patterns and trajectories

The DIVE-estimated spatial patterns are patchier in MRI compared to PET scans, which had lower resolution and were smoothed a-priori. However, this does not suggest the MRI should be smoothed a-priori, as the spatial correlation mechanism within DIVE enables it to remove high-frequency patterns that are not meaningful automatically. More precisely, as the "hole-filling" parameter λ is automatically optimised based on the input data, this enables DIVE to identify the optimal smoothness level specific to the types and size of input data analysed. Moreover, such a-priori smoothing could potentially loose dispersed patterns of pathology that arise due to underlying disruption of brain networks.

While the DIVE-estimated trajectories in the DRC tAD dataset are similar to the ones in ADNI MRI, they show less variance and higher separation between clusters. We attribute this to both the inclusion of MCI patients in ADNI and to the smaller sample size of the DRC dataset. In particular, MCI subjects are highly heterogenous and some have non-AD underlying pathologies, so including them can increase the trajectory variance. Furthermore, the small sample size can also result in lower variance due to less datapoints available to estimate the trajectories.

As less assumptions are made in DIVE about disease progression compared to previous modelling approaches, there could be differences in the resulting spatial patterns of pathology and its evolution. As compared to voxel-based morphometry studies where spatial patterns are highlighted only if significant differences from controls exist, DIVE can highlight early patterns of mild pathology that are also present in controls. There are some implications to this; for example, DIVE could be used to study the normal ageing process. Moreover, as clusters of vertices with similar evolution are optimally derived from the data, this can result in more sensitive biomarkers that can be used for earlier prediction of dementias.

4.4. Model evaluation

In some cortical thickness prediction tasks (Table 2), DIVE showed some improvements over standard approaches, although the margin of improvement is small. This is because the vertex-wise prediction problem is itself very challenging due to high spatial noise and the lack of a-priori smoothing. Moreover, DIVE was not optimised specifically for this prediction task – our key aim was to improve interpretability of disease progression modelling. Similarly, the correlations between DPS scores and cognitive tests are comparable, yet don't show significant improvements over direct correlations of whole brain cortical thickness with cognitive tests. This can be attributed to cognitive tests becoming affected later than MRI atrophy, along with other characteristics of the analysed data (inherent noise and limited number of scans).

4.5. Limitations and future work

DIVE has some limitations that can be addressed. First, we assumed that cluster trajectories follow sigmoidal shapes, which is not the case for many types of biomarkers in ADNI which do not plateau in later stages. The assumption of sigmoidal trajectories can be avoided using non-parametric curves such as Gaussian Processes (Lorenzi et al., 2017), which would be straightforward to incorporate into the DIVE framework. DIVE also assumes a single trajectory within one cluster to avoid complexity, although in reality there is within-cluster trajectory variability. However, simulations have shown that the single-trajectory simplification is sensible as DIVE can still accurately estimate the mean trajectory. Moreover, DIVE inherently minimises the within-cluster variance as it minimises the sum of squared differences under the Gaussian noise model. However, non-Gaussian variability can exist, e.g. due to registration working better in some regions, so future work can take registration uncertainty into account.

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 pathology that occurs in a small set of subjects. However, DIVE can be extended to account for heterogeneity in the datasets by modelling different progression dynamics for distinct subgroups, using random effects like in (Donohue et al., 2014) or unsupervised learning methods like the SuStaIn model (Young et al., 2018). While SuStaIn, just like DIVE, estimates clusters and trajectories within the dataset, the clusters in SuStaIn are made of *subjects* with similar

disease progression, while the clusters in DIVE are made of *vertices* with similar progression. Future work could combine clustering along both subjects and vertices simultaneously to estimate disease subtypes with distinct spatiotemporal dynamics at the vertexwise level.

Our work also has some limitations regarding parameter estimation. Firstly, the EM framework can sometimes converge to a local minimum. While we accounted for this using informed starting points and setting informative priors on the parameters, future work can explore other approaches such as stochastic EM or reversible-jump Markov Chain Monte Carlo (MCMC); moreover, reversible-jump MCMC would allow us to also estimate the optimal number of clusters simultaneously with the other parameters. Secondly, to get a reliable estimate of the subject-specific parameters, we ran DIVE only on balanced datasets, where subjects had at least three scans. However, DIVE can also be applied to less balanced datasets, by setting stronger priors on these parameters or even fixing the progression speed for every subject to 1.

There are several potential future applications of DIVE. One of the advantages of DIVE is that it can be used to study the link between disconnected patterns of brain pathology and connectomes extracted from diffusion tractography or functional MRI (fMRI). Such an analysis would enable further understanding of the exact underlying mechanisms by which the brain is affected by the disease. Our model, which can estimate fine-grained spatial patterns of pathology, is more suitable than standard ROI-based methods for studying the link between pathology and these structural or functional connectomes, because white matter or functional connections have a fine-grained and spatially-varying distribution of endpoints on the cortex.

Apart from studying the link with brain connectomes, there are other potential applications for DIVE. While we only applied it to vertexwise data, the model can also be applied to study voxelwise data. Moreover, DIVE can be applied to other modalities or types of data, including FDG PET, tau PET, DTI or Jacobian compression maps from MRI. Moreover, the model can also be extended to cluster points on the brain surface according to a more complex disease signature, that can be made of two or more biomarkers. For example, using our cortical thickness and amyloid PET datasets from ADNI, we could have clustered points on the brain based on both modalities simultaneously. Such complex disease signatures can offer important insights into the relationships between different modalities and underlying disease mechanisms.

5. Conclusion

DIVE is a spatiotemporal model that can be used for accurately predicting and staging patients across the progression timeline of neurodegenerative diseases. The spatial patterns of pathology can also be used to test mechanistic hypotheses which consider AD as a network vulnerability disorder. All these avenues can help towards disease understanding, patient prognosis, as well as clinical-trials for assessing efficacy of a putative treatment for slowing down cognitive decline.

DIVE availability

DIVE source code, written in Python 3, is available at https://github .com/mrazvan22/dive and can be easily applied on any registered voxelwise or vertexwise images. ADNI data can be downloaded from the Laboratory of NeuroImaging at the University of Southern California.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.02.053.

References

Bilgel, M., Prince, J.L., Wong, D.F., Resnick, S.M., Jedynak, B.M., 2016. A multivariate nonlinear mixed effects model for longitudinal image analysis: application to amyloid imaging. Neuroimage 134, 658–670.

Bishop, M.C., 2006. Pattern Recognition and Machine Learning. Springer-Verlag, New York.

Caroli, A., Frisoni, G., 2010. The dynamics of Alzheimer's disease biomarkers in the Alzheimer's Disease Neuroimaging Initiative cohort. Neurobiol. Aging 31, 1263–1274.

Crutch, S.J., Lehmann, M., Schott, J.M., Rabinovici, G.D., Rossor, M.N., Fox, N.C., 2012. Posterior cortical atrophy. Lancet Neurol. 11, 170–178.

Derado, G., Bowman, F.D., Kilts, C.D., 2010. Modeling the spatial and temporal dependence in fMRI data. Biometrics 66, 949–957.

Dickerson, B.C., Bakkour, A., Salat, D.H., Feczko, E., Pacheco, J., Greve, D.N., Grodstein, F., Wright, C.I., Blacker, D., Rosas, H.D., 2008. 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. Cerebr. Cortex 19, 497–510.

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., 2014. Estimating long-term multivariate progression from short-term data. Alzheimer's Dementia: J. Alzheimer's Assoc. 10, S400–S410.

Dubois, B., Feldman, H.H., Jacova, C., Cummings, J.L., Dekosky, S.T., Barberger-Gateau, P., Delacourte, A., Frisoni, G., Fox, N.C., Galasko, D., Gauthier, S., Hampel, H., Jicha, G.A., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Sarazin, M., de Souza, L.C., Stern, Y., Visser, P.J., Scheltens, P., 2010. Revising the definition of Alzheimer's disease: a new lexicon. Lancet Neurol. 9, 1118–1127.

Dubois, B., Feldman, H.H., Jacova, C., Dekosky, S.T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P.J., Scheltens, P., 2007. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol. 6, 734–746.

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., 2012. An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. Neuroimage 60, 1880–1889.

Greve, D.N., Salat, D.H., Bowen, S.L., Izquierdo-Garcia, D., Schultz, A.P., Catana, C., Becker, J.A., Svarer, C., Knudsen, G.M., Sperling, R.A., 2016. Different partial volume correction methods lead to different conclusions: an 18F-FDG-PET study of aging. Neuroimage 132, 334–343.

Greve, D.N., Svarer, C., Fisher, P.M., Feng, L., Hansen, A.E., Baare, W., Rosen, B., Fischl, B., Knudsen, G.M., 2014. Cortical surface-based analysis reduces bias and variance in kinetic modeling of brain PET data. Neuroimage 92, 225–236.

Holland, D., McEvoy, L.K., Desikan, R.S., Dale, A.M., Alzheimer's Disease Neuroimaging, I, 2012. Enrichment and stratification for predementia Alzheimer disease clinical trials. PLoS One 7, e47739.

Hyun, J.W., Li, Y., Huang, C., Styner, M., Lin, W., Zhu, H., Initiative, A.s.D.N., 2016. STGP: spatio-temporal Gaussian process models for longitudinal neuroimaging data. Neuroimage 134, 550–562.

Jack Jr, C.R., Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Trojanowski, J.Q., 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol. 9, 119–128.

Jedynak, B.M., Lang, A., Liu, B., Katz, E., Zhang, Y., Wyman, B.T., Raunig, D., Jedynak, C.P., Caffo, B., Prince, J.L., 2012. A computational neurodegenerative disease progression score: method and results with the Alzheimer's disease neuroimaging initiative cohort. Neuroimage 63, 1478–1486.

Johnson, K.A., Fox, N.C., Sperling, R.A., Klunk, W.E., 2012. Brain imaging in Alzheimer disease. Cold Spring Harb. Perspect. Med. 2, a006213.

Koval, I., Schiratti, J.-B., Routier, A., Bacci, M., Colliot, O., Allassonnière, S., Durrleman, S., Initiative, A.s.D.N., 2017. Statistical learning of spatiotemporal patterns from longitudinal manifold-valued networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 451–459.

Lehmann, M., Crutch, S.J., Ridgway, G.R., Ridha, B.H., Barnes, J., Warrington, E.K., Rossor, M.N., Fox, N.C., 2011. Cortical thickness and voxel-based morphometry in posterior cortical atrophy and typical Alzheimer's disease. Neurobiol. Aging 32, 1466–1476.

Li, D., Iddi, S., Thompson, W.K., Donohue, C.M., 2017. Bayesian latent time joint mixed effect models for multicohort longitudinal data. Stat. Methods Med. Res. 00, 1–11. https://doi.org/10.1177/0962280217737566.

Lorenzi, M., Filippone, M., Frisoni, G.B., Alexander, D.C., Ourselin, S., Initiative, A.s.D.N., 2017. Probabilistic disease progression modeling to characterize diagnostic uncertainty: application to staging and prediction in Alzheimer's disease. Neuroimage. https://doi.org/10.1016/j.neuroimage.2017.08.059.

Lorenzi, M., Ziegler, G., Alexander, D.C., Ourselin, S., 2015. Efficient Gaussian process-based modelling and prediction of image time series. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 626–637.

Mendez, M.F., Ghajarania, M., Perryman, K.M., 2002. Posterior cortical atrophy: clinical characteristics and differences compared to Alzheimer's disease. Dement. Geriatr. Cognit. Disord. 14, 33–40.

Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 61, 1402–1418.

Sabuncu, M.R., Desikan, R.S., Sepulcre, J., Yeo, B.T.T., Liu, H., Schmansky, N.J., Reuter, M., Weiner, M.W., Buckner, R.L., Sperling, R.A., 2011. The dynamics of cortical and hippocampal atrophy in Alzheimer disease. Arch. Neurol. 68, 1040–1048.

Schiratti, J.-B., Allassonniere, S., Colliot, O., Durrleman, S., 2015. Learning spatiotemporal trajectories from manifold-valued longitudinal data. Adv. Neural Inf. Process. Syst. 2404–2412.

Seeley, W.W., Crawford, R.K., Zhou, J., Miller, B.L., Greicius, M.D., 2009. Neurodegenerative diseases target large-scale human brain networks. Neuron 62, 42–52.

Singh, V., Chertkow, H., Lerch, J.P., Evans, A.C., Dorr, A.E., Kabani, N.J., 2006. Spatial patterns of cortical thinning in mild cognitive impairment and Alzheimer's disease. Brain 129, 2885–2893.

Tang-Wai, D.F., Graff-Radford, N.R., Boeve, B.F., Dickson, D.W., Parisi, J.E., Crook, R., Caselli, R.J., Knopman, D.S., Petersen, R.C., 2004. Clinical, genetic, and neuropathologic characteristics of posterior cortical atrophy. Neurology 63, 1168–1174.

Young, A.L., Marinescu, R.V., Oxtoby, N.P., Bocchetta, M., Yong, K., Firth, N.C., Cash, D.M., Thomas, D.L., Dick, K.M., Cardoso, J., van Swieten, J., 2018. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. Nat. Commun. 9 (1), 4273.

Young, A.L., Oxtoby, N.P., Daga, P., Cash, D.M., Fox, N.C., Ourselin, S., Schott, J.M., Alexander, D.C., 2014. A data-driven model of biomarker changes in sporadic Alzheimer's disease. Brain 137, 2564–2577.

Zhou, J., Gennatas, E.D., Kramer, J.H., Miller, B.L., Seeley, W.W., 2012. Predicting regional neurodegeneration from the healthy brain functional connectome. Neuron 73, 1216–1227.