
Disease Progression Analysis of typical Alzheimer's Disease and Posterior Cortical Atrophy using Data-Driven Models

Author:

Razvan V. MARINESCU

Supervisors:

Prof. Daniel C. ALEXANDER

Dr. Sebastian CRUTCH

Abstract

In this work we study disease progression models, which are mathematical models used to study the evolution of neurodegenerative diseases such as Alzheimer’s disease and to stage patients along the disease time course. We improve existing models such as the Event-Based Model (EBM) or the Differential Equation Model (DEM) and evaluate their performance using a few evaluation metrics. We also develop a new disease progression model that works with vertexwise data such as cortical thickness. This vertexwise model reveals patterns of atrophy to a high spatial detail, without being constrained to a-priori defined regions-of-interest (ROIs). We apply these models to study the progression of two distinct neurodegenerative disorders: typical Alzheimer’s disease (AD) and Posterior Cortical Atrophy (PCA).

Contents

1	Introduction	9
1.1	Alzheimer's Disease	9
1.1.1	Posterior Cortical Atrophy	9
1.2	Disease progression models	9
1.3	Problem Statement	10
1.4	Project goals	10
1.5	Contributions	11
1.6	Report Structure	11
2	Background	13
2.1	Alzheimer's disease	13
2.1.1	Progression of AD	14
2.2	Posterior Cortical Atrophy	14
2.3	Disease Progression Models	15
2.4	The Event-Based Model	16
2.4.1	Theory	18
2.4.2	Event Sequence Estimation	19
2.4.3	MCMC sampling - Metropolis Hastings	20
2.4.4	Mixture Models For Data Likelihood	21
2.4.5	Patient Staging and Diagnosis Prediction	21
2.4.6	Advantages and Limitations	21
2.5	Differential Equation Model	22
2.5.1	Advantages and Limitations	22
2.6	The Disease Progression Score model	23
2.6.1	Model fitting	23
2.6.2	Advantages and Limitations	24
2.7	The Self-Modelling Regression model	24
2.7.1	Parameter fitting	25
2.7.2	Advantages and limitations	25
2.8	The Manifold-based mixed effects model	25
2.8.1	The straight lines model	26
2.8.2	The logistic curve model	26
2.8.3	Advantages and limitations	26
2.9	The voxelwise disease progression model	26
2.9.1	Model fitting	27
2.9.2	Advantages and limitations	28
2.10	The Network Diffusion Model	29
2.10.1	Advantages and Limitations	30
2.11	Discriminative models	30
2.11.1	Advantages and Limitations	30

3 Project Data	33
3.1 ADNI dataset	33
3.1.1 Image preprocessing	33
3.2 DRC dataset	34
4 Performance Evaluation	35
4.1 Introduction	35
4.2 EBM improvements	35
4.2.1 EBM - Simultaneous Sampling	35
4.2.2 EBM - Expectation Maximisation	36
4.3 DEM improvements	39
4.3.1 Trajectory alignment	39
4.3.2 Optimised trajectory alignment	39
4.3.3 Avoiding Singularities	40
4.4 Performance Evaluation Methods	40
4.4.1 Staging consistency	40
4.4.2 Time-lapse prediction	41
4.4.3 Diagnosis and conversion prediction	41
4.4.4 Differential diagnosis	41
4.5 Experiments	42
4.6 Results	42
4.6.1 DRC	42
4.6.2 ADNI	43
4.7 Discussion	44
4.7.1 DRC	44
4.7.2 ADNI	44
4.7.3 Staging-based performance	44
4.7.4 Diagnosis prediction performance	45
4.8 Summary	45
4.8.1 Limitations	45
4.9 Conclusion	46
5 Event-based Model	47
5.1 PCA Analysis	47
5.1.1 Hypothesis	47
5.1.2 Method	47
5.1.3 Results	47
5.1.4 Discussion	49
5.2 AD Analysis	49
5.2.1 Hypothesis	49
5.2.2 Method	49
5.2.3 Results	49
5.2.4 Discussion	49
5.3 PCA subgroups	51
5.3.1 Background	51
5.3.2 Hypotheses	51
5.3.3 Method	52
5.3.4 Results	52
5.3.5 Discussion	52
5.4 Conclusion	57

6 Differential Equation Model	59
6.1 Clinical questions	59
6.2 Method	59
6.3 Results	60
6.4 Discussion	60
6.5 Conclusion	62
7 Voxelwise Temporal Clustering Model	65
7.1 Introduction	65
7.2 Methods	66
7.2.1 Model	66
7.2.2 Fitting the Model using EM	67
7.2.3 Initialisation and Implementation	68
7.3 Simulation results	70
7.4 Experimental results	71
7.4.1 Data acquisition and preprocessing	71
7.4.2 Results with ADNI and DRC datasets	71
7.4.3 Model Validation	73
7.5 Discussion	74
8 Project Plan	75
8.1 Voxelwise Temporal Clustering Model	75
8.2 Performance Evaluation	76
8.2.1 Metrics	76
8.2.2 Models	76
8.2.3 Dataset	77
8.3 PCA vs tAD Analysis	77
9 Bibliography	79
A	91
A.1 ADNI extra figures	91
A.2 Performance evaluation - Synthetic results	92

Chapter 1

Introduction

1.1 Alzheimer's Disease

Alzheimer's disease (AD) is a chronic progressive neurodegenerative disorder that accounts for 60% to 70% of all cases of dementia worldwide [1, 2]. Its symptoms include cognitive dysfunction such as memory loss and language difficulties and psychiatric symptoms such as depression, hallucinations, delusions and agitation. People suffering from Alzheimer's disease also have difficulty performing daily tasks such as driving or shopping. The symptoms of Alzheimer's disease progress from mild symptoms such as memory loss to very severe dementia [1]. Diagnosis is usually based on the person's medical history, information from relatives and behavioural observations. Medical imaging modalities such as computed tomography (CT) or magnetic resonance imaging (MRI) can be used to aid diagnosis and exclude other cerebral pathologies.

The causes of AD are currently unknown, although genetic and environmental risk factors have been found [1]. Amyloid plaques and neurofibrillary tangles are the main histological features of Alzheimer's disease. Genetic risk factors include a specific isoform of apolipoprotein (APOE4). So far there exists no effective treatment that stops or reverses neurodegeneration.

In 2010 it was estimated that up to 35 million people worldwide suffered from AD [2]. Bullock et al. [3] estimated that around 5% of the population older than 65 years is affected by AD. The prevalence doubles approximately every 5 years beyond age 65 [4, 5] and some studies suggest that more than half of the population older than 85 years might be suffering from AD [4, 6].

1.1.1 Posterior Cortical Atrophy

Posterior cortical atrophy (PCA), also called Benson's syndrome [7] is a subtype of Alzheimer's disease that causes atrophy in the posterior part of the cortex, resulting in disruptions of the visual and motor systems. Early symptoms include blurred vision, inability to read, difficulty with depth perception and navigating through space [8, 9]. More severe symptoms emerge as neurodegeneration spreads, including inability to recognise familiar faces and objects and visual hallucinations. If the anterior part of the brain is also affected, symptoms similar to those from Alzheimer's disease may occur, such as memory loss. As with typical Alzheimer's disease, the cause of PCA is still unknown and there is no fully accepted diagnostic criteria [9].

1.2 Disease progression models

During the progression of Alzheimer's disease, many biomarkers based on Magnetic Resonance Imaging (MRI) such as cortical thickness become abnormal at different points in the progression. Finding out the precise temporal evolution of these biomarkers is crucial for 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 heterogeneity in the cohort analysed.

A hypothetical model of disease progression has been proposed by [10], 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 [11] or [12] 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.

A multitude of data-driven disease progression models (DPMs) have been proposed in recent years. On such model is the Event-Based Model [13], which models the progression of disease as a sequence of discrete events, representing underlying biomarkers switching from a normal to abnormal state. Another model, the Differential Equation Model (DEM), reconstructs a continuous trajectory of biomarker measurements from change in short-term follow-up data, which represent samples of the slope at different points along the trajectory. Other models such as the Disease Progression Score (DPS) [14] or self-modelling regression approaches [15] 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 [16].

1.3 Problem Statement

The data-driven disease progression models have several aims. First of all, they try to accurately reconstruct multimodal biomarker trajectories across the temporal domain. Secondly, they optimally place subject visits along this temporal axis and can even model progression speeds or higher order moments for every subject individually. Other things they might try to model include correlations between biomarker measurements or distinct trajectories for different subgroups in the population, which accounts for the disease heterogeneity.

There are several reasons for using data-driven DPMs. First of all, they can provide a better understanding of the disease process by constructing a very detailed picture of the temporal dynamics of biomarkers. Compared to previous methods, they don't rely on *a-priori* staging (e.g. "mild", "moderate", "severe") of patients, which are usually crude, inaccurate and biased. Secondly, DPMs are valuable for clinical trials, where we need tools that can accurately stage subjects. Without accurate staging, drug trials might find no differences between the placebo group and the drug group. Third, these tools can be deployed in the clinic to provide personalised prognostic information to patients.

1.4 Project goals

Our project has three main aims:

1. Develop better disease progression models that can be used for accurate and robust modelling of the evolution of dementia and staging of subjects in clinical trials. Apart from improving existing models, we are also interested in developing novel methods that can uncover new insights into the disease process.
2. Evaluate the performance of these models on real and synthetic datasets. This also includes finding new performance metrics that can be used to evaluate the models.

3. Use these models for understanding the differences in the progression of two distinct AD subtypes: typical AD and Posterior Cortical Atrophy. This will help with differential diagnosis, clinical management of PCA patients and the design of research studies. For this analysis, we use a dataset from the Dementia Research Center (DRC), UK which contains Controls, PCA and tAD subjects.

1.5 Contributions

The key contributions of our work are summarised below:

1. Improved the fitting of the Event-Based Model by simultaneously estimating the abnormality sequence and event distribution parameters using two approaches: a blocked MCMC approach and an Expectation-Maximisation approach. This work contributes to aim 1 and is presented in section 4.2.
2. Improved the fitting of the Differential Equation Model by devising ways to align the biomarker trajectories along the temporal axis. This work contributes to aim 1 and is presented in section 4.3.
3. Evaluated the performance of two main classes of models: the Event-Based Model and the Differential Equation Model. We also proposed new metrics for evaluating the performance of these disease progression models. This work contributes towards aim 2 and is presented in chapter 4.
4. Developed a disease progression model that uses vertex-wise data such as cortical thickness as input. The model combines 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 contributes towards aim 1 and is presented in chapter 7.
5. Used the Event-Based Model, Differential Equation Model and the vertexwise approach from contribution 2 to compare the differences in the progression of two distinct AD subtypes: typical AD and PCA. These results contribute towards aim 3. The EBM results are presented in chapter 5, the differential equation results are presented in chapter 6 while the results with the vertexwise model are presented in chapter 7.

1.6 Report Structure

The report has the following structure:

- Chapter 2 contains background information on Alzheimer’s disease, Posterior Cortical Atrophy and disease progression models.
- Chapter 3 contains information about the datasets analysed and the subject demographics.
- Chapter 4 presents the performance evaluation of the EBM and DEM, as well as the improved EBM fitting procedures.
- Chapter 5 presents results with the Event-Based Model on modelling the progression of PCA and tAD as well as PCA subgroups.
- Chapter 6 presents results but using the Differential Equation Model instead, which can give extra information such as rate and extent of atrophy.
- Chapter 7 presents the temporal clustering model that used vertexwise thickness data.

- Chapter 8 we propose the future work plan for the rest of the PhD period.

Chapter 2

Background

2.1 Alzheimer's disease

Neurodegenerative diseases such as Alzheimer's disease (AD) result in progressive degeneration and death of neuron cells and affect more than 35 million people worldwide [17]. Other neurodegenerative diseases include Parkinson's disease, Huntington's disease, Amyotrophic lateral sclerosis and Lewy bodies dementia. However, Alzheimer's disease is by far the most common type of dementia, accounting for 60% to 70% of the total cases of dementia [1, 2].

Alzheimer's disease is a progressive neurodegenerative disease that usually affects people over 65 years of age [1, 2]. Its early symptoms include short-term memory loss, but as the disease advances other symptoms appear such as problems with language, disorientation, mood swings, loss of motivation and behavioural issues [1, 2]. The average life expectancy after clinical diagnosis is between three to nine years, although the speed of progression can vary [17]. The principal risk factor of AD is age, with incidence rates doubling every 5 years after 65 years of age [17, 18].

Its causes are poorly understood and around 70% of the risk is thought to be genetic, with many risk genes involved such as APOE, GSK3 β , DYRK1A and many others [19]. Other risk factors include head injuries, depression and hypertension [1]. Pathologically, Alzheimer's disease is characterised by the presence of amyloid plaques and neurofibrillary tangles in the brain tissue [20]. In 1991, Hardy et al. postulated that amyloid- β deposits are a central cause in the development of AD, followed by tau phosphorylation and tangle formation and neuronal death [21]. This was supported by the discovery of a pathogenic mutation in the amyloid precursor protein (APP) gene on chromosome 21 and by the occurrence of AD in Down syndrome. On the other hand, the tau hypothesis suggests that abnormalities related to the tau proteins initiate the disease cascade [22]. In this case, tau binding to microtubules is disrupted by phosphorylation which results in free tau that aggregates into neurofibrillary tangles. This ends up destroying the cell's cytoskeleton which collapses the neuron's transport system, ultimately resulting in neuronal death.

There are no effective treatments available to stop or reverse neuronal degeneration in AD [1]. Nevertheless, there exists some medication that can temporarily reduce some symptoms or slow down the progression in some people. Among these, most of them are acetylcholinesterase inhibitors: tacrine, rivastigmine, galantamine and donepezil [23]. Psychosocial interventions are used to complement pharmaceutical treatment, but they don't usually focus on AD but on dementia in general.

Over the last decades, various biomarkers have been developed that monitor the extent of neuronal atrophy of the brain in vivo: CSF amyloid- β_{1-42} [24] and amyloid PET imaging using the Pittsburgh compound B [25] measure amyloid plaque deposits; CSF total tau and phospho-rylated tau [24] measure neurofibrillary tangles and neuroaxonal damage; fluorodeoxyglucose (FDG) PET [26] measures metabolic activity in the brain; MRI cortical volumes, thickness and

atrophy rate measured using the Boundary Shift Integral [27] detect shrinkage of individual brain areas that is caused by neurodegeneration. Lastly, cognitive test scores such as the Mini-Mental State Examination (MMSE) [28] are used to assess memory and cognitive performance. Characterising the trajectories of these biomarkers can help us better understand the dynamics of the disease, explain its causes and perform better differential diagnosis.

2.1.1 Progression of AD

Several studies have been done so far on Alzheimer's disease progression [10, 29, 30, 31, 20, 32]. It is currently believed that in AD amyloid-beta and tau proteins become abnormal first, long before symptoms occur, followed by cognitive decline such as memory loss and brain structural atrophy [10, 33]. Neuropathological staging of AD brains showed that the earliest change in brain structure are in the medial temporal lobe, particularly in the entorhinal cortex and hippocampus [20]. Afterwards, in medium stages amyloid deposits are found in almost all isocortical areas apart from primary sensory areas and motor field [20]. In late stages, atrophy gets worse and spreads even further to other areas including the sensory and motor fields [20]. As a result of early hippocampal and entorhinal atrophy, many volumetric MRI studies focus on these regions and show that even in mild AD the entorhinal area and hippocampus shrink by 20-25% compared to controls [34, 35, 36, 37]. Results by Schott et al. [32] and Ridha et al. [29] show that atrophy of the medial temporal lobe precedes the clinical onset of AD by approximately 3.5 years [32].

2.2 Posterior Cortical Atrophy

Posterior cortical atrophy (PCA) is an early-onset subtype of Alzheimer's disease that affects the posterior part of the brain, resulting in disruption of the visual cortex. The syndrome was first reported by Benson et al. [7] in 1988 to describe five patients with fairly homogeneous but otherwise unclassified symptoms. The most common symptoms include inability to read, blurred vision, light sensitivity, trouble navigating through space and issues with depth perception [8, 9]. Additional symptoms also include apraxia (disorder of movement planning), visual agnosia (object recognition deficit) and agraphia (loss of writing ability) [7, 38]. These symptoms get worse as the disease progresses, with patients becoming unable to recognise familiar people and objects and difficulty navigating familiar places. Some studies [39, 40, 41] reported visual hemineglect (difficulty seeing one half of the visual field) to be frequent in PCA patients, especially if asymmetrical atrophy takes place in the occipital areas.

PCA patients face difficulties in diagnosis due to the young age at onset and the fact that there are no fully accepted diagnostic criteria. Patients are sometimes misdiagnosed with depression, anxiety or even malingering in early stages of the disease [8]. They are often initially referred to opticians and ophthalmologists in the belief that ocular abnormalities are causing their visual deficits, often leading to unnecessary medical procedures such as cataract surgery. Neuroimaging modalities such as magnetic resonance imaging (MRI), positron emission tomography (PET) or single photon emission computed tomography (SPECT) can aid diagnosis of PCA [42].

The causes of PCA are still unknown, due to the rarity of the disease, gradual onset of symptoms and no fully accepted diagnostic criteria [9, 8]. The progressive neurodegeneration that characterises PCA is often attributed to Alzheimer's disease, but alternative causes including dementia with Lewy bodies, corticobasal degeneration and prion disease have also been identified [8]. Genetic factors that underlie PCA are also not well understood [8, 9]. Empirical findings suggest that there are no significant differences in the number of patients with a positive family history of PCA and typical AD [8]. Some studies also report no differences in apolipoprotein E (APOE) genotypes between PCA and typical AD [43, 44, 45].

There is no known efficient treatment of PCA that will reverse or stop neurodegeneration [9]. Patients with PCA are usually treated with the same medication as for AD, namely cholinesterase inhibitors: tacrine, rivastigmine, galantamine and donepezil [9]. However, there are no studies analysing the efficiency of these drugs in PCA patients [9]. A few non-pharmacological therapies have also been attempted recently in some patients that included psycho-educative programs [46] or a combination of speech therapy, occupational therapy and physiotherapy [47].

As opposed to AD, PCA has a distinct progression pattern. Studies by Hof et al. [48] and Tang-Wei et al. [44] show a greater concentration of senile plaques and neurofibrillary tangles in the occipital and parietal lobes and at the occipito-temporal junction. In later stages of the disease neurodegeneration spreads to anterior areas of the brain, and as a result symptoms similar to Alzheimer's disease may occur, such as memory loss.

2.3 Disease Progression Models

A disease progression model describes the evolution of a neurodegenerative disease, which is characterised by a set of pathological changes and symptoms that occur in the brain. Although the underlying molecular mechanisms that cause these changes are not fully understood, biomarker measurements can be used to track the state of the patient and of various anatomical structures. A disease progression model therefore characterises the temporal trajectory of these biomarkers as they go from a normal to an abnormal state. Most models will be able to provide information on the ordering of the biomarkers as they become abnormal and the biomarker rate of decline.

A hypothetical model of disease progression has been proposed by Jack et al. [10, 33], which describes the trajectory of several key biomarkers during the progression of Alzheimer's disease (fig. 2.1). The model suggests that amyloid- β and tau protein biomarkers become abnormal long before the onset of any dementia symptoms. Afterwards, during the mild cognitive impairment (MCI) phase, cognitive functions such as memory become abnormal along with brain structure measured using MRI. These biomarkers continue to be affected in the dementia stage, while A β and tau seem to reach a plateau at this point. This hypothesised model of disease progression is shaping the current field of AD research [15].

Some of the simplest disease progression models are based on symptomatic staging of patients into a small number of groups, e.g. "pre-symptomatic", "mild", "moderate" and "severe" [13]. They then describe the differences in biomarker measurements among these groups. Scahill et al. [31] devised such a method that finds changes in brain structure using voxel-based analysis of serial nonlinear-registered MRI images. Other models based on symptomatic staging are those of Dickerson et al., 2009 [49] and Thomson et al. 2001 [50], 2003 [51]. One key limitation of these methods is that the clinical assessment is usually subjective and biased, which limits the temporal resolution of the disease progression models. This limitation is one of the main factors motivating the development of data-driven models described below. Another method by Bateman et al. [11] uses estimated years from onset to conduct cross-sectional analyses of baseline data in order to determine the relative order and magnitude of pathophysiological changes. However, this method can only be applied to dominantly inherited AD, due to the need to estimate the years from onset using data from parents.

A range of data-driven models of disease progression have also been proposed, which do not need prior knowledge of the disease stages of subjects. One such model is the *event-based model* [13], which describes disease progression as a sequence of discrete events, where each event corresponds to a biomarker becoming abnormal. The model finds the optimal ordering in which biomarkers become abnormal given the data. It has so far been used to model neurodegeneration in familial AD [13], Huntington's disease [13] and sporadic AD [52]. Other discrete event models include those based on Hidden Markov Models [53, 54]. The advantage of the EBM is the computational simplicity of the formulation. However, it is unable to model the time between

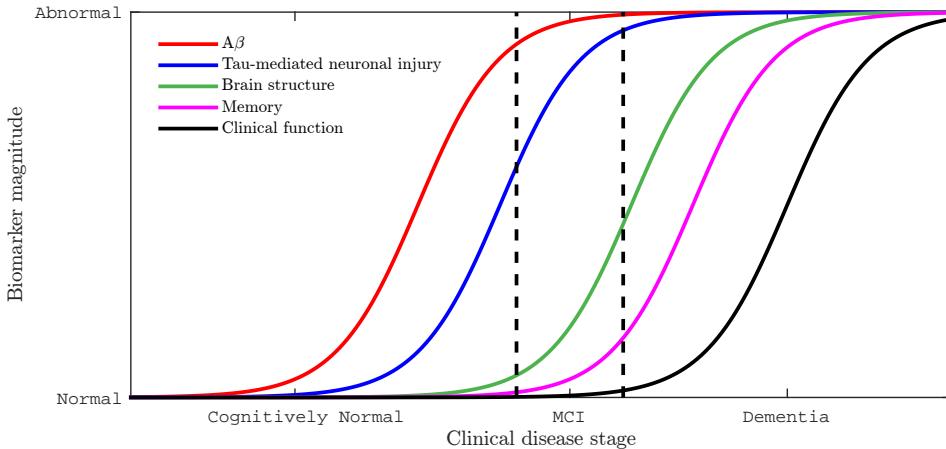


Figure 2.1: Dynamic biomarkers of the AD cascade as hypothesised by Jack et al. [10]. $A\beta$ and tau are thought to become abnormal before the onset of any dementia symptoms, while brain structure, memory and clinical function are thought to become abnormal later, during MCI and dementia stages. Reprinted from [10].

events or the rate of biomarker decline, which limits its utility for prognosis.

Another class of disease progression models are the *differential equation models* [55, 56, 57, 58, 59, 60]. These can be used to reconstruct an average biomarker trajectory which is continuous in contrast with the discrete event-based model trajectories. The models use follow-up biomarker measurements to provide sample gradients of the biomarker trajectory and integrate a differential equation to determine the average trajectory for the entire cohort. These models have been used in a variety of settings: finding the time it takes for amyloid accumulation to become abnormal in amyloid-PET images [60], calculate the rates of amyloid β deposition, cerebral atrophy and cognitive decline [55]. Moreover, *stochastic differential equation models* can also express the deviation from the average trajectory and have been used to analyse MRI brain volumes [56]. The advantage of these models is that they can compute the rate of biomarker decline, which is very useful for prognosis at an individual level. On the other hand, there is no guarantee of correspondence across disease stage and prognosis estimates between different biomarkers [61].

Self-modelling regression approaches can also estimate biomarker trajectories by fitting sets of curves under the assumption of a common shape throughout the population [15]. These approaches have been used to determine the dynamics of regional brain volumes, cognitive tests, PET-based biomarkers and CSF amyloid-beta and tau values [15]. Jedynak et al. [62] used similar biomarkers to derive a *disease progression score* and applied it to the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. The main advantage of these models is that they provide a complete picture of the disease which is useful for clinical understanding [61]. Disadvantages include risk of overfitting due to the large number of parameters that need to be estimated.

2.4 The Event-Based Model

The event-based model was introduced by Fonteijn et al. [13] in 2012 and describes the disease as a sequence of discrete events. Each event represents a change in the patient state, such as the onset of a new symptom (e.g. 'patient shows a drop in cognitive performance') or measurement of tissue pathology (e.g. 'lumbar puncture shows reduced amyloid beta'). The aim of the model is to find the ordering in which events occur given a dataset of biomarker measurements. In section 2.4.1 we present the theory behind the EBM, in section 2.4.2 we present the MCMC

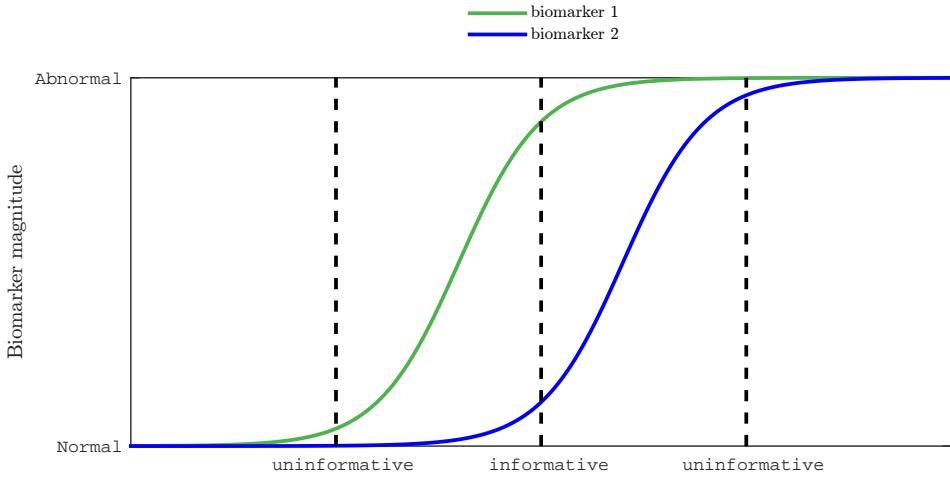


Figure 2.2: Two hypothetical biomarker progressions along with three different snapshots, that may come from the same subject or from different subjects. The left-most and the right-most snapshots have measurements that are both either low or high, so this does not give any information regarding which biomarker becomes abnormal first. However, the middle snapshot is informative because the measurement for biomarker 1 is high while the measurement for biomarker 2 is low, which suggests that biomarker 1 becomes abnormal first.

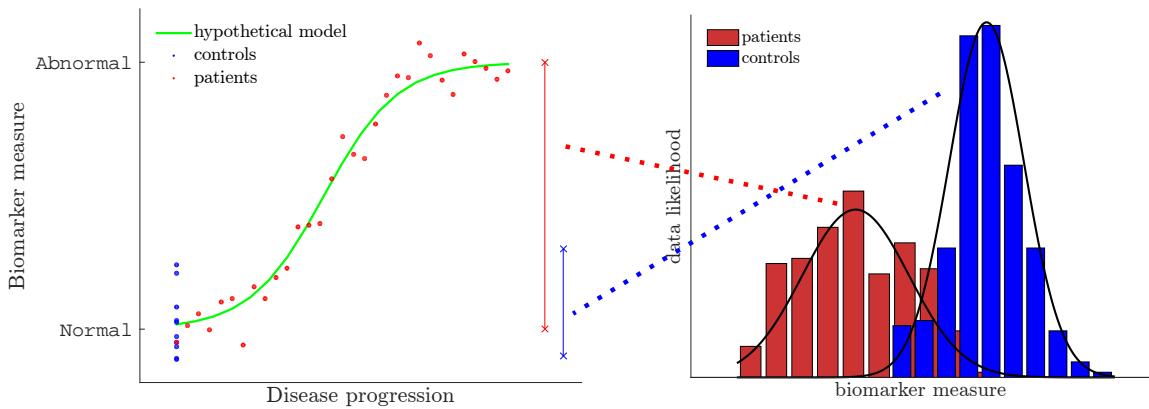


Figure 2.3: The left figure shows biomarker measurements for controls and patients along with the underlying hypothetical (true) progression curve. The right figure shows the mixture model for the two groups of subjects. Note that there is overlap between the biomarker values. The Gaussian fit for the control groups represents $p(x|\neg E)$, while the fit for the patients group represents $p(x|E)$.

and greedy ascent methods that are used to estimate the abnormality sequence, in section 2.4.4 we show how to fit the mixture model parameters and in section 2.4.5 we show how to stage the subjects using the EBM.

2.4.1 Theory

The event-based model consists of a series of events E_1, E_2, \dots, E_N and an ordering $S = [s(1), \dots, s(N)]$ which is a permutation of the integers $1, \dots, N$ creating the event ordering $E_{s(1)}, E_{s(2)}, \dots, E_{s(N)}$. The set of events is specified a-priori. Moreover, the model uses a dataset X which contains a set of X_j measurements for each subject j . Each set $X_j = \{x_{1j}, x_{2j}, \dots, x_{Nj}\}$ where x_{ij} represents the value of biomarker i in subject j and is informative of event E_i in subject j .

The event-based model makes two key assumptions: first, measurements decrease monotonically as the disease progresses and second, the event ordering is the same across all patients. The first assumption fits with the hypothetical model presented by Jack et al. [10] in fig. 2.1. Therefore, a patient for whom event E_i has occurred cannot revert to a state where event E_i did not occur. This assumption is essential because it ensures snapshots are informative about the event ordering [13]. The second assumption is necessary to be able to aggregate information about the event ordering from the entire set of subjects.

The aim of the event-based model is to find the probability density function $p(S|X)$ of an event ordering given the biomarker data. One starts by fitting a model for the likelihood function $p(x_{ij}|E_i)$ the likelihood of measuring x_{ij} given event E_i occurred. A similar fit is obtained for $p(x_{ij}|\neg E_i)$, the likelihood of measuring x_{ij} given event E_i has not occurred. More information about mixture model fitting can be found in section 2.4.4. If a subject j is at stage k in the disease progression, events $E_{s(1)}, \dots, E_{s(k)}$ have occurred while events $E_{s(k+1)}, \dots, E_{s(N)}$ have not occurred. We can therefore define the likelihood of the data from subject j given ordering S as:

$$p(X_j|S, k) = \prod_{i=1}^k p(x_{s(i),j}|E_{s(i)}) \prod_{i=k+1}^N p(x_{s(i),j}|\neg E_{s(i)}) \quad (2.1)$$

where measurements x_{ij} are assumed to be independent. Since the subject could potentially be at any stage k in the progression, we integrate over k :

$$p(X_j|S) = \sum_{k=0}^N p(k)p(X_j|S, k) \quad (2.2)$$

where $p(k)$ is the prior probability of the subject being at position k in the sequence. A uniform prior is usually assumed here. Further assuming independence of measurements across patients we get:

$$p(X|S) = \prod_{j=1}^J p(X_j|S) \quad (2.3)$$

Combining equations 2.1, 2.2, 2.3 we get the total likelihood:

$$p(X|S) = \prod_{j=1}^J \left[\sum_{k=0}^N p(k) \left(\prod_{i=1}^k p(x_{s(i),j}|E_{s(i)}) \prod_{i=k+1}^N p(x_{s(i),j}|\neg E_{s(i)}) \right) \right] \quad (2.4)$$

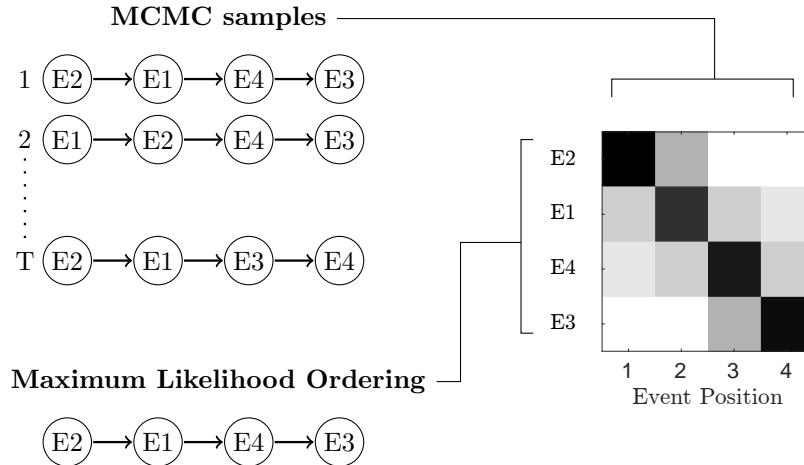


Figure 2.4: MCMC sampling and positional variance computation. MCMC sampling finds a series of T samples, which are then used to derive the *characteristic ordering*, where events are ordered according to their average position in the MCMC samples. Entries $M(i, j)$ in the positional variance matrix stores the relative number of times each event appeared in each position in the sequence. The events in the positional variance matrix are ordered according to the characteristic ordering.

2.4.2 Event Sequence Estimation

Applying Bayes' theorem we can get the posterior on the sequence:

$$p(S|X) = \frac{p(S)p(X|S)}{p(X)} \quad (2.5)$$

As the marginal distribution $p(X)$ is analytically intractable, one can use a Markov-chain Monte Carlo (MCMC) algorithm to sample from the posterior $p(S|X)$. A detailed description on how the MCMC algorithm works is given in section 2.4.3. One assumes flat priors on the sequence S as any sequence could be equally likely.

In the MCMC phase, at each iteration the sequence S can be perturbed by swapping two randomly chosen events. This perturbation rule has been used by Fonteijn et al. [13]. However, another perturbation method used by Young et al. [52] randomly selects a source and target event and places the source event after the target event, sliding the other biomarkers accordingly. See figure 2.5 for a representation of the two perturbation rules. The resulting sequence S^{new} is accepted with probability $p = \min(1, a)$ where $a = p(X|S^{new})/p(X|S)$. Otherwise the old sequence is stored and the process is repeated. As MCMC depends on accurate initialisation, one also runs a greedy ascent algorithm in order to find the sequence with the highest likelihood. The greedy ascent is very similar to the MCMC phase, the only difference being that a is set to 1 if $p(X|S^{new}) > p(X|S)$ and to zero otherwise. Depending on the number of biomarkers, the greedy ascent is run for a few thousand iterations and repeated 10 times, with different random permutations of integers $1, \dots, N$ as the starting position. The maximum likelihood sequence obtained from greedy ascent is then used to initialise MCMC sampling, which usually runs for at least 100,000 iterations, again depending on problem size.

The resulting MCMC-sampled sequences are usually plotted in a positional variance matrix M , which is a compact way to represent uncertainty in the event ordering. Each element $M(i, j)$ represents the proportion of times event $E_{s(i)}$ appeared on position j in the sampled sequences, given some *master sequence* S . S is usually set to be the maximum likelihood sequence or the *characteristic ordering*, which is given by the average position of the events in the MCMC samples [13]. The characteristic sequence is particularly useful in plotting positional variance matrices from cross-validation or bootstrapping.

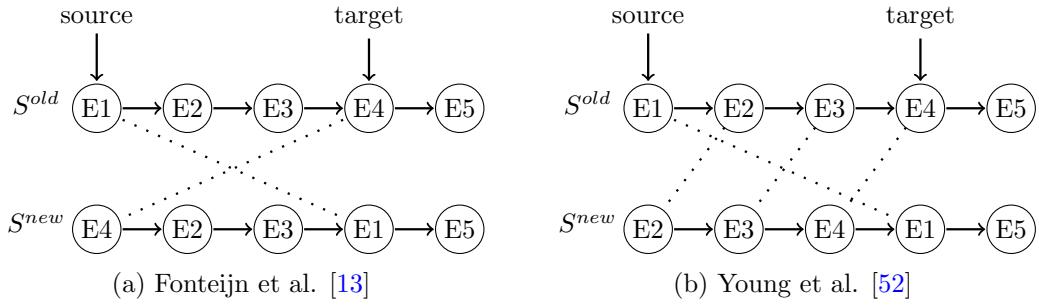


Figure 2.5: MCMC perturbation rules used by (a) Fonteijn et al. [13] and (b) Young et al. [52]. Both methods assume randomly selected source and target events. The method by Fonteijn et al. only swaps the source event (E1) with the target event (E4). On the other hand, the perturbation used by Young et al. moves a source event after a target event and slides the other biomarkers accordingly.

2.4.3 MCMC sampling - Metropolis Hastings

In order to sample the event ordering, one can use the Metropolis Hastings algorithm, which is a Markov chain Monte Carlo sampling method. The original Metropolis algorithm was proposed by Metropolis et al. in 1953 [63]. W. K. Hastings extended it to a more general case in 1970 [64].

Suppose we have a function $f(x)$ which is proportional to a probability distribution $p(x)$, which we need to take samples from. It is often the case in practice that the normalisation factor is not known because it is extremely difficult to compute it. The Metropolis Hastings algorithm works by producing a sequence of samples X_0, X_1, \dots, X_n where each sample X_{t+1} is generated from X_t using a proposal distribution $Q(X_{t+1}|X_t)$. At each iteration, a candidate sample X_{cand} is generated from the proposal distribution $Q(X_{cand}|X_t)$. A probability ratio a is calculated as follows:

$$a = \frac{f(X_{cand})}{f(X_t)} \frac{Q(X_t|X_{cand})}{Q(X_{cand}|X_t)} \quad (2.6)$$

Finally, the sample X_{cand} is accepted with probability $\min(1, a)$. Note that in most cases the *proposal distribution* is symmetric, i.e. $Q(X_t|X_{cand}) = Q(X_{cand}|X_t)$, in which case the ratio a becomes:

$$a = \frac{f(X_{cand})}{f(X_t)}$$

Metropolis-Hastings algorithm

1. Choose a starting point X_1
2. for $t = 2$ to T
 - 2.1. Draw a candidate sample X^{cand} from the proposal distribution $Q(X|X^t)$
 - 2.2. Compute the acceptance ratio a as follows:

$$a = \frac{f(X_{cand})}{f(X_t)} \frac{Q(X_t|X_{cand})}{Q(X_{cand}|X_t)}$$

- 2.3. Draw a random number r from the unit interval $[0, 1]$

2.4. Accept the candidate sample X^{cand} only if $a \geq r$. That is:

$$X^{t+1} = \begin{cases} X^{cand} & \text{if } a \geq r \\ X^t & \text{otherwise} \end{cases}$$

2.4.4 Mixture Models For Data Likelihood

In equation 2.1 we need to model the distributions $p(x_{i,j}|E_i)$ and $p(x_{i,j}|\neg E_i)$ using the measurements in X . However, the fact that a subject is a control or patient does not give any information whether event E_i has occurred. Therefore, it is necessary to use a mixture model, whose components will correspond to $p(x_{i,j}|E_i)$ and $p(x_{i,j}|\neg E_i)$ respectively. The particular choice of the model depends on the data that is being modelled. Fonteijn et al. [13] used a Gaussian distribution for $p(x_{i,j}|\neg E_i)$ and a uniform distribution for $p(x_{i,j}|E_i)$. On the other hand, Young et al. [52] used two Gaussian distributions for both distributions and applied some constraints on their mean and variance. The models are fit using numerical optimisation or Expectation-Maximisation (EM).

2.4.5 Patient Staging and Diagnosis Prediction

After the maximum likelihood sequence has been found using the greedy ascent method described in section 2.4.2, each subject can be assigned a stage k as follows:

$$k = \arg \max_k p(k)p(X_j|S, k) = p(k) \prod_{i=1}^k p(x_{s(i),j}|E_{s(i)}) \prod_{i=k+1}^N p(x_{s(i),j}|\neg E_{s(i)}) \quad (2.7)$$

As before, the prior $p(k)$ is assumed to be uniform. It should be noted that stages range from zero to N , the number of events. If a subject is at stage k it means that all events up to and including k have occurred while the events after k have not occurred.

One can use patient stages to classify subjects into controls and AD, or any other available subgroups, using a simple method by Young et al. [52]. Given a threshold stage t , one can predict all subjects having a stage less than or equal to t to be controls and all subjects with stages greater than t to be patients. The optimal threshold is the one which maximises the balanced accuracy, defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (2.8)$$

where TP , FP , FN , TN represent the number of true positive, false positive, false negative and true negative subjects respectively. After the optimal threshold is found, one can also compute the sensitivity and specificity defined as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (2.9)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (2.10)$$

2.4.6 Advantages and Limitations

The event-based model by Fonteijn et al. [13] has several advantages. It is a data-driven progression model which does not use a-priori defined clinical stages, which can often be biased and limit the temporal resolution of the model. Moreover, it does not require longitudinal data, which makes it very useful for analysing rare types of dementia for which comprehensive

longitudinal datasets do not exist. The Bayesian framework in which it is formulated also allows it to estimate uncertainty in the abnormality sequence. The model can also easily combine data from different modalities.

The current model has several limitations. The trajectory parameters are modelled as step functions, which is a strong assumption given the continuous nature of almost every MRI and CSF biomarker commonly used in the field. Secondly, in the fitting process the parameters and sequence are assumed independent and therefore fitted separately. Third, the model assumes all biomarker measurements are independent with each other and also across different visits of the same subject. Finally, the model also assumes that all subjects follow the same progression sequence, which is not the case in heterogeneous datasets such as ADNI.

2.5 Differential Equation Model

The differential equation model (DEM) [57, 65, 59, 55] constructs the biomarker trajectories from the change in biomarker values between different visits. In many datasets such as ADNI, the biomarker scores s are observed for each subject over a few visits. By determining how these scores change (Δs) over time (t) during a specified time interval Δt , the temporal rate of progression ($\Delta s/\Delta t$) can be modelled as a function of the mean biomarker value $f(s)$ [57]:

$$\frac{\Delta s}{\Delta t} \approx f(s) \quad (2.11)$$

The model given by $f(s)$ can be for example linear, polynomial or non-parametric models such as Gaussian Processes (GP). We then perform a line integral along $f(s)$ to recover $s(t)$. More explicitly, if we take the limit as $\Delta \rightarrow 0$ from Eq. 2.11, we get that:

$$\lim_{\Delta t \rightarrow 0} \frac{\Delta s}{\Delta t} = \frac{\delta s}{\delta t} = f(s) \quad (2.12)$$

Solving this numerically is done using the Euler method. We set an initial (t_0, s_0) and small increment step δt and finding the next pair (t_1, s_1) as follows:

$$\begin{aligned} t_1 &= t_0 + \delta t \\ s_1 &= s_0 + f(s_0)\delta t \end{aligned} \quad (2.13)$$

This is the repeated until the full curve defined by $(t_0, s_0), (t_1, s_1), \dots, (t_n, s_n)$ is reconstructed. The process is repeated independently for the other biomarkers.

2.5.1 Advantages and Limitations

The differential equation model has several advantages. It is a fully data-driven method that does not require a-priori defined clinical categories. In contrast to the event-based model, it can estimate non-parametric biomarker trajectories which makes minimal assumptions on the shape of the biomarker trajectories, with the only real constraint being monotonicity.

The model has several limitations. First of all, the biomarker trajectories are fit independently, which requires alignment on the temporal axis after they are recovered. Secondly, in this formulation the model does not allow to directly estimate uncertainty in the trajectory values along the y-axis. The only option for estimating uncertainty is by integrating posterior samples of the differential model and then aligning them all on the temporal axis. A third limitation is that the model is very sensitive to noise in the biomarker measurements.

2.6 The Disease Progression Score model

The disease progression score (DPS) model was proposed by Jedynak et al. [62]. It is based on three main assumptions:

- Subjects follow a common disease progression but they have a different age on onset and progression speed.
- Each biomarker trajectory is a monotonic curve that follows a sigmoidal shape
- The speed of progression of each subject is the same across the entire disease time-course.

The model estimates the optimal shape of the biomarker trajectories while estimating a disease progression score for each subject, which is the stage along the disease time course. The disease progression score s_{ij} for subject i at visit j is defined as a linear transformation of age t_{ij} :

$$s_{ij} = \alpha_i t_{ij} + \beta_i \quad (2.14)$$

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

The DPS model assumes that biomarker measurements are independent and follow a sigmoidal trajectory $f(s)$ given the disease progression score s . The sigmoidal function for biomarker k is parametrised as $\theta_k = [a_k, b_k, c_k, d_k]$ where:

$$f(s; \theta_k) = \frac{a_k}{1 + \exp(-b_k(s - c_k))} + d_k \quad (2.15)$$

where d_k is the minimum value, $d_k + a_k$ is the maximum value, $a_k b_k / 4$ is the maximum slope and c_k is the inflexion point. Previous work has shown that sigmoidal curves provide a better fit for biomarker data compared to linear models [59, 66]. The value y_{ijk} of biomarker k from subject i at visit j is a normally distributed random variable:

$$p(y_{ijk} | \alpha_i, \beta_i, \theta_k, \sigma_k) = N(y_{ijk} | f(\alpha_i t_{ij} + \beta_i; \theta_k), \sigma_k) \quad (2.16)$$

We further define I to be the set of all triplets (i, j, k) for which measurements are available. Assuming independence across all measurements we get the following model likelihood:

$$p(y | \alpha, \beta, \theta, \sigma) = \prod_{(i,j,k) \in I} p(y_{ijk} | \alpha_i, \beta_i, \theta_k, \sigma_k) \quad (2.17)$$

where $y = [y_{ijk}]$ for $(i, j, k) \in I$. 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 biomarkers, represent the stacked parameters for the sigmoidal trajectories and measurement noise specific to each biomarker.

The parameters of the model are therefore $\Theta = [\alpha, \beta, \theta, \sigma]$ and the log-likelihood function associated with it is:

$$l(\alpha, \beta, \theta, \sigma) = \sum_{(i,j,k) \in I} \log \sigma_k + \frac{1}{2\sigma_k^2} (y_{ijk} - f(\alpha_i t_{ij})) \quad (2.18)$$

2.6.1 Model fitting

Model fitting is done by alternating between optimising the sigmoidal parameters σ and the subject specific parameters α, β using the method of least squares. Figure 1 show the fitting procedure. In line 1, we initialise $\alpha_i = 1$, $\beta_i = 0$ for every i . On lines 4 and 5 the optimal

parameters for every biomarker trajectory are computed. On line 8 the subject specific shifts and progression speeds are found. Sometimes unfeasible parameters can be retrieved, so on line 14 the parameters are mapped into the feasible space. In order to make the model identifiable, parameters α_i and β_i are rescaled for every subject on line 18 so that the disease progression scores of healthy controls have a mean μ_N of 0 and a standard deviation σ_N of 1.

```

1 Initialise  $\alpha^{(0)}, \beta^{(0)}$ ;
2 for  $l = 1$  to  $L$  do
3   for  $k = 1$  to  $K$  do
4      $\theta_k^{(1)} = \arg \min_{\theta_k} \sum_{(i,j) \in I_k} (y_{ijk} - f(\alpha_i^{(0)} t_{ij} + \beta_i^{(0)} | \theta_k))^2$ 
5      $\sigma_k^{(1)2} = \frac{1}{|I_k - 2I - 4|} \sum_{(i,j) \in I_k} (y_{ijk} - f(\alpha_i^{(0)} t_{ij} + \beta_i^{(0)} | \theta_k))^2$ 
6   end
7   for  $i = 1$  to  $I$  do
8      $\alpha_i^{(1)}, \beta_i^{(1)} = \arg \min_{\alpha_i, \beta_i} \sum_{(j,k) \in I_i} \frac{1}{\sigma_k^{(1)2}} (y_{ijk} - f(\alpha_i t_{ij} + \beta_i | \theta_k^{(1)}))^2$ 
9   end
10   $\alpha^{(0)} = \alpha^{(1)}, \beta^{(0)} = \beta^{(1)}$ 
11 end
12 for  $k = 1$  to  $K$  do
13   if  $b_k < 0$  then
14      $a_k^{(1)} = -a_k^{(1)}, b_k^{(1)} = -b_k^{(1)}, d_k^{(1)} = d_k^{(1)} + a_k^{(1)}$ 
15   end
16 end
17 for  $i = 1$  to  $I$  do
18    $\alpha_i = \frac{\alpha_i}{\sigma_N}, \beta_i = \frac{\beta_i - \mu_N}{\sigma_N}$ 
19 end
```

Algorithm 1: The optimisation procedure for the disease progression score by [62].

2.6.2 Advantages and Limitations

The model by Jedynak et al. [62] has several advantages. As opposed to the differential equation model, the model automatically aligns the biomarker trajectories on the temporal axis. Furthermore, compared to the event-based model by [13], the biomarker trajectories are modelled as sigmoidal trajectories instead of step functions and each subject has an associated time shift and progression speed.

The model has several limitations. The main limitation of this model is that the trajectories are assumed to be sigmoidal, which is not necessarily the case for many biomarkers in ADNI (e.g. MRI volumes or cortical thickness) that don't show any signs of plateauing. Furthermore, each subject is assumed to follow the same progression pattern, which is not true in many heterogeneous datasets such as ADNI.

2.7 The Self-Modelling Regression model

Self-modelling regression (SEMR) is a method that fits several curves under the assumption of a common shape [15]. This approach has been used by Donohue et al. [15] to estimate non-parametric biomarker trajectories with linear subject-specific effects. We assume that Y_{ij} is the measurement of biomarker j in subject i , g_j is a continuously differentiable monotone function, $\gamma_i \sim N(0, \sigma_\gamma^2)$ is the time shift for subject i . The model is defined as follows:

$$Y_{ij}(t) = g_j(t + \gamma_i) + \alpha_{0ij} + \alpha_{1ij}t + \epsilon_{ij}(t) \quad (2.19)$$

where $\alpha_{0ij}, \alpha_{1ij} \sim N(0, \Sigma_j)$ model a linear perturbation of the non-parametric trajectory g_j for subject i and biomarker j , t is the time and $\epsilon_{ij}(t) \sim N(0, \sigma_j)$ is the measurement noise.

2.7.1 Parameter fitting

Fitting the model is done by iteratively estimating each set of parameters (g_j, γ_i, α) until convergence of the Residual Sum of Squares (RSS). The algorithm makes use of the following residuals:

$$\begin{cases} R_{ij}^g(t) = Y_{ij}(t) - \alpha_{0ij} - \alpha_{1ij}t & E[R_{ij}^g(t)|g_j, t, \gamma_i] = g_j(t + \gamma_i) \\ R_{ij}^\alpha(t) = Y_{ij}(t) - g_j(t + \gamma_i) & E[R_{ij}^\alpha(t)|\alpha_{0ij}, \alpha_{1ij}, t] = \alpha_{0ij} + \alpha_{1ij}t \\ R_{ij}^\gamma(t) = t - g_j^{-1}(Y_{ij}(t)) & E[R_{ij}^\gamma(t)|\gamma_i] \approx g_j^{-1}(g_j(t + \gamma_i)) - t = \gamma_i \end{cases} \quad (2.20)$$

Using the above residuals, the model is fit by initialising γ_i and iterating the following steps[15]:

1. Given γ_i , estimate g_j by setting $\alpha_{0ij} = \alpha_{1ij} = 0$ and iterating the following subroutine:
 - (a) Estimate g_j by a monotone curve fit on $R_{ij}^g(t)$
 - (b) Estimate $\alpha_{0ij}, \alpha_{1ij}$ using the linear mixed model of $R_{ij}^\alpha(t)$. Repeat steps a and b until convergence of each $RSS_j = \sum_{it} [Y_{ij}(t) - g_j(t + \gamma_i) - \alpha_{0ij} - \alpha_{1ij}t]^2$
2. Given the estimated g_j , set $\alpha_{0ij} = \alpha_{1ij} = \epsilon_{ij}(t) = 0$ and estimate each γ_i with the average of R_{ij} over all j and t . Steps 1 and 2 are repeated until convergence of the total $RSS = \sum_{ijt} [Y_{ij}(t) - g_j(t + \gamma_i) - \alpha_{0ij} - \alpha_{1ij}t]^2$

2.7.2 Advantages and limitations

The model have many advantages. It robustly models the average trajectory using a non-parametric model, in contrast with previous approaches [62]. Moreover, it also models unique trajectories for each subject as linear deviations from the average trajectories using a mixed effects model. Each subject also has it's own temporal shift.

One limitation of the model is that it does not model different progression speeds for each individual. Also, estimating all the subject-specific parameters $(\alpha_{0ij}, \alpha_{1ij}, \gamma_i)$ requires suitable priors and many longitudinal measurements, which might not be available in some datasets. Moreover, it also assumes that biomarker measurements are independent, which makes it unsuitable for modelling large sets of correlated biomarkers such as vertex-wise cortical thickness measures or voxel-wise amyloid burden.

2.8 The Manifold-based mixed effects model

The manifold-based mixed effects model was introduced by Schiratti et al. in 2015 [16]. The model generalises a previous linear mixed effects model by [67] to account for time shifts and describes it in a Riemannian manifold setting. Let us assume we observe p individuals, each having n_i observations obtained at times $t_{i,1} < \dots < t_{i,n_i}$, each having $y_{i,1}, \dots, y_{i,n_i}$ biomarker measurements. For a point $p_0 \in M$, $v_0 \in T_{p_0}M$ we define $\gamma_{p_0, t_0, v_0} = Exp_{t_0, p_0}(\alpha_i v_0)(.)$ as the geodesic which passes through point p_0 at time t_0 with velocity v_0 . We also consider $t_{i,j}$ and $y_{i,j}$ as the age and measurement for one biomarker in subject i at visit j . This gives us the following model:

$$y_{i,j} = Exp_{t_0 + \tau_i, p_0}(\alpha_i v_0)(t_{i,j}) + \epsilon_{i,j} \quad (2.21)$$

where

$$\begin{cases} \alpha_i = \exp(\nu_i) \\ \nu_i \sim \bigotimes_{i=1}^p N(0, \sigma_\nu^2) \\ \tau_i \sim \bigotimes_{i=1}^p N(0, \sigma_\tau^2) \\ \epsilon_{i,j} \sim \bigotimes_{i,j} N(0, \sigma^2) \end{cases} \quad (2.22)$$

The model assumes that each η_i and τ_i are independent. The parameters of the model are $\theta = [p_0, t_0, v_0, \sigma_\eta, \sigma_\tau, \sigma]$. The model above can be re-written as:

$$\begin{aligned} y_{i,j} &= \gamma_{p_0, t_0, v_0}(\alpha_i v_0(t_{i,j} - t_0 - \tau_i)) + \epsilon_{i,j} \\ &= \gamma_i(t_{i,j}) + \epsilon_{i,j} \end{aligned} \quad (2.23)$$

where γ_i is the subject specific trajectory that is modelled as an affine reparametrisation of the average trajectory γ_{p_0, t_0, v_0} . The model described here is univariate. However, an extension of the model has been published by Schiratti et al. [68] which extends this framework to a multivariate analysis.

2.8.1 The straight lines model

If we assign the canonical (i.e. Euclidean) metric to the manifold $M = \mathbb{R}$, we get the "straight lines" model:

$$y_{i,j} = p_0 + \alpha_i v_0(t_{i,j} - t_0 - \tau_i) + \epsilon_{i,j} \quad (2.24)$$

2.8.2 The logistic curve model

Some biomarkers such as Mini-Mental State Examination or other cognitive tests are bounded in a specific interval that can be mapped to (0,1). In order to give the (0,1) interval a Riemannian manifold structure we need to equip it with a Riemannian metric. For all points $p_0 \in M$ we define g_0 to be the inner product $g_{p_0}(u, v) = uM(p_0)v$ where $M(p_0) = \frac{1}{p_0^2(1-p_0)^2}$. This results in the geodesics of M to be logistic curves of the form $t \rightarrow (1 + a \exp(-rt))^{-1}$ [16]. Therefore, the model from Eq. 2.21 writes:

$$y_{i,j} = \left[1 + \left(\frac{1}{p_0} - 1 \right) \exp \left(-\frac{\alpha_i v_0}{p_0(1-p_0)} (t_{i,j} - t_0 - \tau_i) \right) \right]^{-1} + \epsilon_{i,j} \quad (2.25)$$

2.8.3 Advantages and limitations

The model has several strengths and weaknesses. The main strength lies in the flexible Riemannian manifold framework, that allows one to create different models depending on how the inner product is defined. Moreover, the model estimates subject specific trajectories γ_i , time shifts τ_i and progression speeds α_i . However, one of the limitations of the model is that it assumes a parametric form of the biomarker trajectories (i.e. sigmoidal).

2.9 The voxelwise disease progression model

A voxelwise disease progression model has been introduced in 2016 by Bilgel et al. [69]. This model allows the discovery of patterns of atrophy that are not confined to a given region of interest (ROI). Since the input data is represented by voxel-wise measurements such as amyloid burden, a spatial correlation function is used to model correlation between voxel values. The model is built on the framework of the disease progression score by Jedynak et al. [62].

Let us assume that t_{ij} represent the age for subject i at visit j . The progression score s_{ij} for subject i at visit j is an affine transformation of age t_{ij} :

$$\begin{cases} s_{ij} = \alpha_i t_{ij} + \beta_i = \mathbf{q}_{ij}^T \mathbf{u}_i \\ \mathbf{u}_i \sim N_2(\mathbf{m}, V(\boldsymbol{\nu})) \end{cases} \quad (2.26)$$

where α_i, β_i are the progression speed and time shift of subject i , $\mathbf{q}_{ij} = [t_{ij}, 1]^T$ and $\mathbf{u}_i = [\alpha_i, \beta_i]^T$. The prior covariance matrix V is modelled as a 2×2 positive definite matrix that has a Log-Cholesky parametrisation given by $\boldsymbol{\nu}$. More precisely, if we consider a matrix U such that $V = UU^T$ and $U = \begin{bmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix}$, then $\boldsymbol{\nu} = [\log U_{11}, \log U_{12}, \log U_{22}]$

Furthermore, let us denote by \mathbf{y}_{ij} the $K \times 1$ vector of biomarker measurements for subject i at visit j . The longitudinal trajectories corresponding to these measurements are modelled as follows:

$$\begin{cases} \mathbf{y}_{ij} = \mathbf{a}s_{ij} + \mathbf{b} + \epsilon_{ij} \\ \epsilon_{ij} \sim N_K(0, R(\boldsymbol{\lambda}, \boldsymbol{\rho})) \end{cases} \quad (2.27)$$

where $\mathbf{a} = [a_1, \dots, a_K]^T$, $\mathbf{b} = [b_1, \dots, b_K]^T$ are the coefficients of the linear model and ϵ_{ij} is the measurement noise that is independent and identically distributed across different subjects and visits. The matrix $R(\boldsymbol{\lambda}, \boldsymbol{\rho})$ is the spatial covariance that is assumed to have the form $R = \Lambda C \Lambda$, where Λ is a diagonal matrix with diagonal elements $\boldsymbol{\lambda}$ and C is a correlation matrix that is parametrised by $\boldsymbol{\rho}$ [69]. This ensures that the matrix $R(\boldsymbol{\lambda}, \boldsymbol{\rho})$ is positive definite. In order to model correlation among voxel measurements, the elements $C_{kk'}$ of matrix C must be a function of the distance $d \equiv d(k, k')$ between voxels k and k' . Several such options exist:

- Exponential: $C_{kk'} = \exp(-d/\rho)$
- Gaussian: $C_{kk'} = \exp(-(d/\rho)^2)$
- Exponential: $C_{kk'} = (1 + (d/\rho)^2)^{-1}$
- Spherical: $C_{kk'} = \left(1 - \frac{3}{2} \frac{d}{\rho} + \frac{1}{2} \left(\frac{d}{\rho}\right)^3\right)$ if $d < \rho$

The model parameters are therefore $\boldsymbol{\theta} = [\mathbf{m}, \boldsymbol{\nu}, \mathbf{a}, \mathbf{b}, \boldsymbol{\lambda}, \boldsymbol{\rho}]$. The model is a mixed effects model where \mathbf{a}, \mathbf{b} are the fixed effects and \mathbf{u}_i are the random effects.

2.9.1 Model fitting

The model is fit using the Expectation-Maximisation (EM), described below. The complete data log-likelihood is:

$$\begin{aligned} l(\mathbf{y}, \mathbf{u}; \boldsymbol{\theta}) = \sum_i l(\mathbf{y}_i, \mathbf{u}_i; \boldsymbol{\theta}) = \\ -\frac{1}{2} \sum_{i,j} \log|2\pi R| - \frac{1}{2} \sum_{i,j} (\mathbf{y}_{ij} - Z_{ij} \mathbf{u}_i - \mathbf{b})^T \\ - \frac{1}{2} \sum_i \log|2\pi V| - \frac{1}{2} \sum_i (\mathbf{u}_i - \mathbf{m})^T V^{-1} (\mathbf{u}_i - \mathbf{m}) \end{aligned} \quad (2.28)$$

E-step

Let (\mathbf{y}, \mathbf{u}) be the complete data and $\boldsymbol{\theta}' = [\mathbf{m}', \boldsymbol{\nu}', \mathbf{a}', \mathbf{b}', \boldsymbol{\lambda}', \boldsymbol{\rho}']$ be the parameters estimated at the previous EM iteration. Bilgel et al. [69] show that the E-step integral $Q(\boldsymbol{\theta}, \boldsymbol{\theta}')$ is proportional to $\sum_i \int \Phi(\tilde{\mathbf{u}}_i; \hat{\mathbf{u}}'_i, \Sigma'_i) l(\mathbf{y}_i, \tilde{\mathbf{u}}_i; \boldsymbol{\theta}) d\tilde{\mathbf{u}}_i$, where Φ is a multivariate normal probability density function with mean:

$$\hat{\mathbf{u}}'_i = \left(\sum_j Z_{ij}^T R'^{-1} Z_{ij}' + V'^{-1} \right)^{-1} \left(\sum_j Z_{ij}^T R'^{-1} (\mathbf{y}_{ij} - \mathbf{b}') + V'^{-1} \mathbf{m}' \right) \quad (2.29)$$

and covariance matrix $\Sigma'_i = \left(\sum_j Z_{ij}^T R'^{-1} Z_{ij}' + V'^{-1} \right)^{-1}$. Evaluating the integral gives the following final form:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}') = -\frac{1}{2} \sum_{ij} \log|R| - \frac{1}{2} \sum_{ij} (\mathbf{y}_{ij} - Z_{ij} \hat{\mathbf{u}}'_i - \mathbf{b}) - \frac{1}{2} \sum_{ij} \text{Tr}(Z_{ij}^T R^{-1} Z_{ij} \Sigma'_i) - \frac{1}{2} \sum_i \log|V| - \frac{1}{2} \sum_i (\hat{\mathbf{u}}'_i - \mathbf{m})^T V^{-1} (\hat{\mathbf{u}}'_i - \mathbf{m}) - \frac{1}{2} \sum_i \text{Tr}(V^{-1} \Sigma'_i) \quad (2.30)$$

M-step

At the M-step we need to find $\boldsymbol{\theta} = \arg \max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}')$. The full derivations are given in [69], yielding the following updates:

$$\mathbf{a} = \frac{(\sum_i \nu_i) (\sum_{ij} \mathbf{y}_{ij} s'_{ij}) - (\sum_{ij} \mathbf{y}_{ij}) (\sum_{ij} s'_{ij})}{(\sum_i \nu_i) (\sum_{ij} \mathbf{q}_{ij}^T \Sigma'_i \mathbf{q}_{ij} + s'^2_{ij}) - (\sum_{ij} s'_{ij})^2} \quad (2.31)$$

$$\mathbf{b} = \frac{(\sum_{ij} \mathbf{y}_{ij}) (\sum_{ij} \mathbf{q}_{ij}^T \Sigma'_i \mathbf{q}_{ij} + s'^2_{ij}) - (\sum_{ij} \mathbf{y}_{ij} s'_{ij}) (\sum_{ij} s'_{ij})}{(\sum_i \nu_i) (\sum_{ij} \mathbf{q}_{ij}^T \Sigma'_i \mathbf{q}_{ij} + s'^2_{ij}) - (\sum_{ij} s'_{ij})^2} \quad (2.32)$$

$$\mathbf{m} = \frac{1}{n} \sum_i \hat{\mathbf{u}}'_i \quad (2.33)$$

$$\boldsymbol{\nu} = \arg \max_{\boldsymbol{\nu}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}') \quad (2.34)$$

$$\boldsymbol{\lambda}, \boldsymbol{\rho} = \arg \max_{\boldsymbol{\lambda}, \boldsymbol{\rho}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}') \quad (2.35)$$

2.9.2 Advantages and limitations

The model by Bilgel et al. [69] has several advantages. First of all, it is specifically tailored for dealing with voxelwise measurements such as amyloid load by modelling the spatial correlations. Secondly, like the disease progression score by Jedynak et al. [62], it estimates subject specific temporal shifts and progression speeds.

The model has several limitations. First of all, the biomarker trajectories are assumed to be linear, which is a strong assumption especially for early many biomarkers such as amyloid, which start to plateau when subjects are in the MCI stages. Moreover, while it is necessary to incorporate a spatial correlation function due to the nature of the data, this has the effect of "blurring out" the image and is not suitable for some voxels that are close-by and undergo different trajectories.

2.10 The Network Diffusion Model

The network diffusion model was introduced by Raj et al. in 2012 [70] and later extended in 2015 [71]. The model is inspired by evidence that Alzheimer's disease atrophy spreading process occurs along vulnerable pathways in a prion-like manner rather than by spatial proximity [72, 73, 74]. The model works by simulating a diffusion process driven by concentration gradients along a connectivity graph that is obtained from tractography data of healthy controls.

The diffusion process is modelled on a hypothetical brain network $G = \{V, E\}$ whose nodes $v_i \in V$ represent the gray matter structure of region i and edges (i, j) having weight c_{ij} represent fibre connections between regions i and j . Structures v_i are obtained from parcellated regions-of-interest while the connection strength c_{ij} is measured using tractography [75]. If we denote the disease factor in region i as x_i , then the flow of the disease from a region i to another region j in a short time interval δt is $\beta(x_i - x_j)c_{ij}\delta t$, where β is a diffusivity constant controlling the diffusion speed. As $\delta t \rightarrow 0$, this results in the following first-order differential equation:

$$\frac{dx_j}{dt} = \beta c_{ij}(x_i - x_j) \quad (2.36)$$

Now let us denote by $\mathbf{x}(t) = \{x(v, t), v \in V\}$ the disease factor at time t in every node of the network. Eq. 2.36 will then translate into the "network heat equation" [76]:

$$\frac{d\mathbf{x}(t)}{dt} = \beta H\mathbf{x}(t) \quad (2.37)$$

where H is the Laplacian matrix of G defined as:

$$H(i, j) = \begin{cases} \sum_{j' \neq i} c_{ij'} & \text{for } i = j \\ -c_{ij} & \text{otherwise} \end{cases} \quad (2.38)$$

We model the cortical atrophy in region k as the accumulation of the disease process:

$$\phi_k(t) = \int_0^t x_k(\tau) d\tau \quad (2.39)$$

Extending this to the whole brain gives

$$\Phi(t) = \int_0^t \mathbf{x}(\tau) d\tau \quad (2.40)$$

The solution to the above equation is given by:

$$\mathbf{x}(t) = \exp(-\beta Ht)\mathbf{x}_0 \quad (2.41)$$

where \mathbf{x}_0 is the initial disease concentration, where the term $\exp(-\beta Ht)$ acts as a smoothing operator. Performing eigenvalue decomposition on $H = U\Lambda U^\dagger$, where $U = [\mathbf{u}_1, \dots, \mathbf{u}_N]$ is the matrix of eigenmodes, allows to express $\mathbf{x}(t)$ as:

$$\mathbf{x}(t) = U\exp(-\Lambda\beta t)U^\dagger\mathbf{x}_0 = \sum_{i=1}^N \left(\exp(-\beta\lambda_i t) \mathbf{u}_i^\dagger \mathbf{x}_0 \right) \mathbf{u}_i \quad (2.42)$$

The time evolution of atrophy can then be described as:

$$\Phi(t) = \int_0^t \sum_{i=1}^N \left(\exp(-\beta\lambda_i t) \mathbf{u}_i^\dagger \mathbf{x}_0 \right) \mathbf{u}_i dt = \sum_{i=1}^N \frac{1}{\beta\lambda_i} (1 - \exp(-\beta\lambda_i t)) \mathbf{u}_i^\dagger \mathbf{x}_0 \mathbf{u}_i \quad (2.43)$$

Raj et al. [70] present evidence to suggest that the eigenmodes i with the highest corresponding eigenvalues λ_i represent the areas that are normally affected by normal ageing, AD and behavioural variant frontotemporal dementia (bvFTD) respectively. They suggest that these areas are selectively vulnerable to these types of dementia, in line with previous findings in the literature [77, 78].

2.10.1 Advantages and Limitations

The diffusion model by Raj et al. [70] has several advantages. In contrast with the models presented above, it is able to model the propagation of atrophy along white matter tracts, which can be used to test the prion hypothesis. Secondly, this approach allows one to test for hypothesis of network-based atrophy spread such as nodal stress, transneuronal spread, trophic failure, and shared vulnerability [79].

The model has several limitations. The model assumes static networks, even though the network dynamically evolves during the time course of the disease. The model also assumes a parametric form of the biomarker trajectories, either exponential or sigmoidal.

2.11 Discriminative models

Discriminative models such as Support Vector Machines (SVM) can also be used to model the progression of the disease. A Support Vector Machine is a non-probabilistic binary linear classifier that was originally developed by Vladimir N. Vapnik and Alexey Ya. Chervonenkis [80]. They can perform non-linear classification by mapping the input data into higher-dimensional feature spaces using the *kernel trick* and finding the hyperplane that optimally separates the data. They have been successfully used for differential diagnosis of different types of dementia using post-mortem confirmed subjects [81]. Other uses of SVMs on medical images and other high-dimensional data have also been shown [82, 83, 84, 85].

A different type of discriminative model is the random forest [86, 87]. These work by building a multitude of decision trees on training data, using them to make predictions on the test data and performing majority voting on the predictions of individual trees to select the final labels. Random forests have been demonstrated in medical imaging for diagnosis classification [88], image quality transfer [89] or myocardium segmentation [90].

2.11.1 Advantages and Limitations

Discriminative models such as SVMs or random forests have several advantages. They work for a wide variety of problems, datasets and underlying models and are generally robust to high-dimensional data. Some of them such as random forests are also resistant to overfitting.

These discriminative models also have several fundamental limitations. First of all, they require labelled data, in the form of a-priori defined clinical categories or stages, which are usually coarse, inaccurate and biased. These limit the temporal resolution of the model. Moreover, it is also harder to interpret what these models learn from the data, which limits their use for understanding the disease process.

Model	Trajectory shape	Subject Staging		Handles Missing Data	Subject-specific trajectories	Models Biomarker Correlation	Main Limitation
		onset	speed				
Discriminative Models	N/A	✗	✗	✓	✗	✗	biased categories
Event-based Model	step-function	✓	✗	✗	✗	✗	no notion of time
Differential Equation Model	non-parametric	✗*	✗*	✓	✗	✗	trajectories not aligned
Disease Progression Score	sigmoid	✓	✓	✓	✗	✗	sigmoidal trajectory assumption
Self-Modelling Regression	non-parametric	✓	✓	✗	✓	✗	requires large datasets
Manifold Model	linear, sigmoid	✓	✓	✗	✓	✗	sigmoidal trajectory assumption
Voxelwise Model	linear	✓	✓	✓	✗	✓	linear trajectory assumption
Network diffusion Model	exponential, sigmoidal	✗*	✗*	✗	✗	✓	assumes static connectome

Table 2.1: Comparison of features of various disease progression models. Models with * cannot be used for subject staging in their basic formulation, but extensions to the model can enable them to estimate subject-specific disease onset and progression speed.

Chapter 3

Project Data

For this project we used two main datasets: (1) the Alzheimer’s Disease Neuroimaging Initiative (ADNI) and (2) the Dementia Research Center, UK (DRC) dataset. In sections 3.1 and 3.2 we present the subject demographics and how the data has been acquired and pre-processed.

3.1 ADNI dataset

The ADNI dataset is a multisite, longitudinal study that assesses clinical, imaging, genetic and biospecimen biomarkers through the process of normal ageing to mild cognitive impairment (MCI) and dementia (AD) [91]. Magnetic resonance imaging (MRI), (¹⁸F)-fluorodeoxyglucose positron emission tomography (FDG PET), urine serum and cerebrospinal fluid (CSF) data is acquired along with cognitive test information at multiple time points. At 55 participating sites in North America, the data was collected from 200 elderly cognitively normal, 400 MCI and 200 AD subjects. All subjects were scanned with 1.5T MRI at each time point and half of these were also scanned with FDG PET [91]. The main aims of the ADNI project was to make the data openly available to the scientific community, develop technical standards for imaging in longitudinal studies, find out optimal methods for image acquisition and analysis and validate imaging biomarkers by correlating them with cognitive tests.

The ADNI dataset used for this project is the same one used by Young et al. in 2014 [52]. It was downloaded from LONI (www.loni.ucla.edu/ADNI/) on 5 February 2013 and included 285 subjects that had 1.5T MRI scans at baseline and 1 year, a CSF examination at baseline and cognitive tests at baseline which included the Mini-Mental State Examination (MMSE) [28], the Alzheimer’s Disease Assessment Scale-Cognitive Subscale (ADAS-COG) [92] which omits item 13 (ADAS13) and the Rey Auditory Verbal Learning Test (RAVLT) [93]. FDG-PET and amyloid PET were not included in the dataset used by Young et al. [52]. CSF measures of amyloid beta, tau and total tau were performed centrally, and CSF tau and total tau were log transformed, to improve normality [52]. For the longitudinal analysis, imaging we used clinical and CSF data at 12- and 24-months follow-up periods. We also used the clinical diagnosis at all available timepoints up to 72 months. Demographics of the population analysed is included in table 3.1.

3.1.1 Image preprocessing

The MRI acquisition protocols have already been detailed by Jack et al. in 2008 [91]. In the dataset used for this project, cross-sectional brain regions of interest (ROIs) such as the hippocampus, entorhinal cortex, middle temporal gyrus, fusiform gyrus, ventricles and the whole brain, as well as total intracranial volume (TIV) were calculated at baseline and followup by Young et al. [52] using Freesurfer 4.3, a freely available brain imaging software (<http://www.freesurfer.net/>). All regional volumes were normalised by the total intracranial volume

	Number of Subjects	Gender M/F	Age (years, mean \pm SD)	Education (years, mean \pm SD)	APOE +/-
Controls	92	48/44 (52%)	75 \pm 5	15.6 \pm 2.9	22/70 (24%)
MCI	129	82/47 (64%)	73 \pm 7	15.9 \pm 3	72/57 (56%)
AD	64	34/30 (53%)	75 \pm 8	15 \pm 3	45/19 (70%)

Table 3.1: Baseline population demographics for the ADNI cohort.

of each subject. The Boundary Shift Integral (BSI) [27] was used to measure volume change between 0 and 12 months. The KN-BSI method by Leung et al. [94] was used to measure change in brain volume while the MAPS-HBSI method [95] was used to measure the hippocampal volume change.

3.2 DRC dataset

The Dementia Research Center, UK provided us with a dataset of imaging data and cognitive test scores which contains healthy subjects along with subjects diagnosed with Alzheimer's disease (AD) and Posterior Cortical Atrophy (PCA). A total of 89 controls, 70 PCA and 65 AD subjects have been scanned with 1.5T/3T MRI at baseline. Some of them had follow-up scans for up to ten different timepoints. A total of 510 MRI scans have been acquired from these subjects at different timepoints. Demographic information about the DRC cohort is given in table 3.2.

In order to perform tissue segmentation and parcellation, we used the Geodesic Information Flows (GIF) algorithm by Cardoso et al. [96] which is available for download at <http://cmictig.cs.ucl.ac.uk/niftyweb/>. The atlas that has been used for segmentation is the Neuromorphometrics atlas, which produced 146 different brain ROIs across the left and right hemisphere. The brain volumes have been corrected for total intracranial volume (TIV), age and gender using a linear regression approach. Left and right brain regions were averaged into one region.

	Number of Subjects	Gender M/F	Age at baseline (years, mean \pm SD)	Years from onset (years, mean \pm SD)	Number of scans (mean \pm SD)
Controls	89	33/56	60.5 \pm 11	-	2.2 \pm 1.3
PCA	70	27/43	63.0 \pm 7	4.4 \pm 2.8	2.4 \pm 1.4
AD	65	34/31	66.3 \pm 8	4.8 \pm 2.6	2.1 \pm 1.3

Table 3.2: Baseline population demographics for the DRC cohort.

Chapter 4

Performance Evaluation

4.1 Introduction

Data-driven disease progression models such as the Event-Based Model (section 2.4), the Differential Equation Model (section 2.5) or the Disease Progression Score (DPS) (section 2.6) make several assumptions about the biomarker data and have inherent limitations, so it is therefore important to evaluate their performance. However, this is complicated by the fact that for most datasets analysed we don't know the ground truth on the shape and alignment of the biomarker trajectories or on the stages of subjects. Such ground truth information would only be available in synthetic datasets, where it is straightforward to validate the models as shown in the work of [97]. The ability to evaluate disease progression models will help find models and types of fitting algorithms that work best on different biomarker datasets, offer a correct picture of how the disease evolves and can accurately stage patients, which is vital in clinical trials. We are not aware of any previous work that has been done on performance evaluation of data-driven disease progression models. We searched on Google Scholar for the following keywords: "disease", "progression", "model", "performance", "evaluation" but we did not find anything relevant.

In this chapter we suggest ways to evaluate the performance of disease progression models on real datasets such as ADNI or DRC where we don't have accurate ground truth information about biomarker trajectories or subject stages. We seek to evaluate two classes of models: the event-based model and the differential equation model. For each of these models, we devise improvements in the fitting procedures and compare them against the standard fitting methods. In section 4.2 we present improvements in the Event-Based Model while in section 4.3 we present improvements in the Differential Equation Model. Afterwards, in section 4.4 we propose some performance metrics that can be used to evaluate the performance of these disease progression models.

4.2 EBM improvements

In this section we outline two improvements in the fitting of the event-based model: a simultaneous blocked MCMC sampling of the distribution parameters and event ordering (section 4.2.1), an Expectation-Maximisation approach (section 4.2.2). Furthermore, we also present an improvement in temporal alignment of the DEM trajectories in section 4.3.2).

4.2.1 EBM - Simultaneous Sampling

In this section we present an improvement in the fitting of the event-based model that uses simultaneous MCMC sampling of the distribution parameters and event sequence. So far in the work of [13] or [52], the event based model has been fit by first estimating the parameters of

the $p(x|E)$ and $p(x|\neg E)$ distributions, and then estimating the maximum likelihood sequence. This was under the assumption at the sequence and the parameters of the normal and abnormal distributions are independent. However, this is not the case because choosing different distributions (which define the "soft" abnormality threshold) for a single biomarker might change its position in the sequence and potentially the position of other biomarkers too. We therefore propose a new *Simultaneous Sampling* method that optimises both the distribution parameters and the ordering at the same time by sampling them using a blocked Markov Chain Monte Carlo (MCMC) method.

Let us define a set of events E_1, E_2, \dots, E_N and an ordering $S = [S(1), \dots, S(N)]$ which is a permutation of the integers $1, \dots, N$ creating the event ordering $E_{S(1)}, E_{S(2)}, \dots, E_{S(N)}$. The likelihood of dataset X given the ordering S is given by:

$$p(X|S) = \prod_{i=1}^P \left[\sum_{k=0}^N p(k) \left(\prod_{j=1}^k p(x_{i,S(j)}|E_{S(j)}) \prod_{j=k+1}^N p(x_{i,S(j)}|\neg E_{S(j)}) \right) \right] \quad (4.1)$$

where x_{ij} represents the value from subject i for biomarker j and is informative of event E_j in subject i , P is the number of subjects and N is the number of biomarkers. Let us further assume the biomarker distributions are normally distributed, i.e. $p(x|E_j) \sim N(\mu_j^a, \sigma_j^a)$ and $p(x|\neg E_j) \sim N(\mu_j^n, \sigma_j^n)$ where μ_j^a and σ_j^a model the distribution of abnormal values for biomarker j (i.e. event E_j occurred), while μ_j^n and σ_j^n model the distribution of normal values for biomarker j (event E_j did not occur). Thus, the full set of parameters that need to be modelled is $\theta = [\mu_j^n, \sigma_j^n, \mu_j^a, \sigma_j^a]^{j=1\dots N}, S]$. Therefore, the likelihood in equation 4.1 can be explicitly written as:

$$p(X|S, [\mu_j^n, \sigma_j^n, \mu_j^a, \sigma_j^a]^{j=1\dots N}) = \prod_{i=1}^P \left[\sum_{k=0}^N p(k) \left(\prod_{j=1}^k p(x_{i,S(j)}|\mu_{S(j)}^a, \sigma_{S(j)}^a) \prod_{j=k+1}^N p(x_{i,S(j)}|\mu_{S(j)}^n, \sigma_{S(j)}^n) \right) \right] \quad (4.2)$$

In order to maximise this likelihood, one can perform MCMC or Gibbs sampling and get the sample θ with the highest likelihood. However, this can take a very long time to converge due to the large dimensionality and the dependence between parameters. We therefore propose a blocked MCMC sampling approach, where at each step we only propose parameters for biomarker j , i.e. $[\mu_j^n, \sigma_j^n, \mu_j^a, \sigma_j^a]$ along with a new sequence S_j^{new} where only event E_j changed its position. The distribution parameters for the other biomarkers and the ordering of the other events $i \neq j$ are kept the same. The proposed sample θ^{new} is then accepted or rejected according to the normal Metropolis rule. At next iteration, we sample parameters for the next biomarker and so on until convergence. This blocked approach works well because there is strong dependence between parameters corresponding to the same biomarker and between the position of the corresponding event in the sequence.

The covariance matrix of the proposal distribution is estimated by taking 100 bootstraps of the dataset and computing the covariance of $[\mu^c, \sigma^c, \mu^p, \sigma^p]$, where μ^c, σ^c are the mean and standard deviation of the control group while μ^p, σ^p are the mean and standard deviation of the patient group. While we have used Gaussian distributions for $p(x|E)$ and $p(x|\neg E)$, one can actually use any other mixture of parametric distributions.

4.2.2 EBM - Expectation Maximisation

In this section we present how the event-based model can be fit using Expectation Maximisation (EM). EM works well in this scenario because each stage of the subjects is a latent variable that can be estimated in the E-step, while the distribution parameters and the optimal sequence can be estimated in the M-step. The motivation for using EM is the same as for the simultaneous

sampling approach: we can fit the data better by not assuming that the distribution parameters and the event sequence are independent.

Let us assume that $p(x|E)$ and $p(x|\neg E)$ follow normal distributions, where $p(x|E_k) \sim N(\mu_k^a, \sigma_k^a)$ and $p(x|\neg E_k) \sim N(\mu_k^n, \sigma_k^n)$. Our vector of parameters is then $\theta = [[\mu_k^n, \sigma_k^n, \mu_k^a, \sigma_k^a]^{k=1..N}, S]$ where S is the event ordering. Moreover, we define $Z = [Z_1, Z_2, \dots, Z_P]$ a vector of (latent) discrete random variables representing the stage of each subject which can take values $0 \dots N$, where N is the number of biomarkers. The dataset is denoted by X where x_{ij} represents the data from subject i for biomarker j while $X_i = [x_{i1}, \dots, x_{iN}]$ is a vector of biomarker data for subject i .

E-step

In the E-step of the EM framework we aim to compute the following quantity:

$$Q(\theta|\theta^{old}) = \mathbb{E}_{Z|X,\theta^{old}}[\log p(X, Z|\theta)] =$$

Assuming a uniform prior on Z , i.e. $\log p(Z = z) = C$ and expanding Z we get:

$$Q(\theta|\theta^{old}) = C + \sum_{z_1=0}^N \cdots \sum_{z_P=0}^N p(Z = z|X, \theta^{old}) \log p(X|Z_1 = z_1, \dots, Z_P = z_P, \theta)$$

Further assuming that the data from each subject is conditionally independent, i.e. $X_i \perp\!\!\!\perp X_j | \theta$ and $X_i \perp\!\!\!\perp Z_j | \theta$ for $i \neq j$ we get:

$$Q(\theta|\theta^{old}) = C + \sum_{z_1, \dots, z_P} p(Z = z_1, \dots, z_P | X, \theta^{old}) \log \left[\prod_{i=1}^P p(X_i | Z_i = z_i, \theta) \right]$$

where P is the number of patients. After further expansion of $p(X_i | Z_i = z_i, \theta)$, moving the log inside the products and removing the constant C we get:

$$\sum_{z_1, \dots, z_P} p(Z = z_1, \dots, z_P | X, \theta^{old}) \sum_{i=1}^P \left[\sum_{j=1}^{z_i} \log p(x_{ij} | E_{S(j)}) + \sum_{j=z_i+1}^N \log p(x_{ij} | \neg E_{S(j)}) \right]$$

Replacing $p(x|E)$ and $p(x|\neg E)$ with the pdf of a Gaussian distribution we get:

$$\sum_{z_1, \dots, z_P} p(Z = z_1, \dots, z_P | X, \theta^{old}) \sum_{i=1}^P \left[\sum_{j=1}^{z_i} \log N(x_{ij} | \mu_{S(j)}^a, \sigma_{S(j)}^a) + \sum_{j=z_i+1}^N \log N(x_{ij} | \mu_{S(j)}^n, \sigma_{S(j)}^n) \right] \quad (4.3)$$

M-step

In the M-step we aim to find the arguments θ^* that maximise the expected log-likelihood of the complete data $\theta^* = \arg \max_{\theta} Q(\theta|\theta^{old})$. As the function is differentiable with respect to all parameters apart from S (which is discrete), we can find the maximum by solving $\nabla_{\theta} Q(\theta|\theta^{old}) = 0$. We next show the derivation for parameter μ_k^n , which is the solution of $\frac{d}{d\mu_k^n} Q(\theta|\theta^{old}) = 0$. Using the result from equation 4.3 and moving the derivation operator inside the sums we get:

$$\sum_{z_1, \dots, z_P} p(Z = z_1, \dots, z_P | X, \theta^{old}) \sum_{i=1}^P \left[\sum_{j=1}^{z_i} \frac{d}{d\mu_k^n} \log N(x_{ij} | \mu_{S(j)}^a, \sigma_{S(j)}^a) + \sum_{j=z_i+1}^N \frac{d}{d\mu_k^n} \log N(x_{ij} | \mu_{S(j)}^n, \sigma_{S(j)}^n) \right] = 0$$

The derivative term cancels all likelihood terms apart from the one where $S(j) = k$:

$$\sum_{z_1=0}^N \cdots \sum_{z_P=0}^N p(Z_1 = z_1, \dots, Z_P = z_P | X, \theta^{old}) \sum_{i=1}^P \left[\sum_{j=z_i+1}^N \mathbb{I}[S(j) = k] \frac{d}{d\mu_k^n} \log N(x_{ik} | \mu_k^n, \sigma_k^n) \right] = 0$$

which can be rewritten as:

$$\sum_{z_1=0}^N \cdots \sum_{z_P=0}^N p(Z_1 = z_1, \dots, Z_P = z_P | X, \theta^{old}) \sum_{i=1}^P \left[\frac{d}{d\mu_k^n} \log N(x_{ik} | \mu_k^n, \sigma_k^n) \sum_{j=z_i+1}^N \mathbb{I}[j = S^{-1}(k)] \right] = 0$$

$$\sum_{z_1=0}^N \cdots \sum_{z_P=0}^N p(Z_1 = z_1, \dots, Z_P = z_P | X, \theta^{old}) \sum_{i=1}^P \left[\frac{d}{d\mu_k^n} \log N(x_{ik} | \mu_k^n, \sigma_k^n) \mathbb{I}[S^{-1}(k) > z_i] \right] = 0$$

Further rearranging the sum terms we get:

$$\begin{aligned} \sum_{i=1}^P \frac{d}{d\mu_k^n} \log N(x_{ik} | \mu_k^n, \sigma_k^n) \sum_{z_i=0}^N \mathbb{I}[S^{-1}(k) > z_i] \sum_{z_j \neq z_i} p(Z_1 = z_1, \dots, Z_P = z_P | X, \theta^{old}) &= 0 \\ \sum_{i=1}^P \frac{d}{d\mu_k^n} \log N(x_{ik} | \mu_k^n, \sigma_k^n) \sum_{z_i=0}^N \mathbb{I}[S^{-1}(k) > z_i] p(Z_i = z_i | X, \theta^{old}) &= 0 \\ \sum_{i=1}^P \frac{d}{d\mu_k^n} \log N(x_{ik} | \mu_k^n, \sigma_k^n) p(S^{-1}(k) > Z_i | X, \theta^{old}) &= 0 \end{aligned}$$

Inserting the formula for the Gaussian pdf we get:

$$\sum_{i=1}^P \frac{d}{d\mu_k^n} \frac{(x_{ik} - \mu_k^n)^2}{2(\sigma_k^n)^2} p(S^{-1}(k) > Z_i | X, \theta^{old}) = 0$$

which results in the update rule for μ_k^n , the mean of $p(x | \neg E_k)$

$$\mu_k^n = \sum_{i=1}^P x_{ik} w_i^n$$

where

$$w_i^n = \frac{p(S^{-1}(k) > Z_i | X, \theta^{old})}{\sum_{i=1}^P p(S^{-1}(k) > Z_i | X, \theta^{old})}$$

Using a similar approach we get the update rules for σ_k^n , μ_k^a , σ_k^a :

$$\sigma_k^n = \sqrt{\sum_{i=1}^P w_i^n (x_{ik} - \mu_k^n)^2}$$

$$\mu_k^a = \sum_{i=1}^P x_{ik} w_i^a$$

$$\sigma_k^a = \sqrt{\sum_{i=1}^P w_i^a (x_{ik} - \mu_k^a)^2}$$

where

$$w_i^a = \frac{p(S^{-1}(k) \leq Z_i | X, \theta^{old})}{\sum_{i=1}^P p(S^{-1}(k) \leq Z_i | X, \theta^{old})}$$

Solving for S in the M-step is unfortunately intractable, so we use MCMC sampling where at each step of the sampling process we propose a new sequence S^{new} , find the optimal distribution parameters for each biomarker given S^{new} using the EM update rules and then evaluate the likelihood $Q(\theta|\theta^{old})$. The sequence and parameters that maximise the likelihood are chosen and the EM proceeds to a new iteration. Although this approach might not guarantee that we truly find the optimal parameters, it still results in an increase of $Q(\theta|\theta^{old})$. This approach, called generalised EM, guarantees that the method will still converge to a local maxima. We also use the mean and standard deviation of the control and patient populations as a starting point for EM.

4.3 DEM improvements

4.3.1 Trajectory alignment

The DEM model presented in background section 2.5 fits a model for each biomarker independently. However, the reconstructed trajectories are not aligned on the temporal axis. One simple way to align them is to set, for each biomarker, time $t = 0$ correspond to the average biomarker value of one of the diagnostic groups. More precisely, for each biomarker we set $f_b(0) = x_b$ where f_b is the trajectory of biomarker b and x_b is the average value of biomarker b for one diagnostic group (e.g. AD, MCI). There are two limitations with this approach: (1) the disease onset can be much earlier than the baseline visit and (2) patients are at different stages in the disease at the baseline visit. An improved alignment method is presented in section 4.3.2.

4.3.2 Optimised trajectory alignment

In this section we present a data-driven Bayesian model that takes trajectories obtained from the differential equation model (DEM) and aligns them on the same temporal axis. Let us denote by X our dataset where x_{pb} is the measurement of biomarker b in patient p , Z_p is the discrete stage of patient p and f_b is the shape of the trajectory for biomarker b . For every biomarker b , we aim to estimate a temporal shift t_b of the trajectory and a measurement noise σ_b . The log likelihood for the data X_p from patient p who is at stage z_p can be expressed as:

$$p(X_p | f_1, \dots, f_B, t_1, \dots, t_B, \sigma_1, \dots, \sigma_B, z_p) = \prod_{b=1}^B N(x_{pb} | f_b(z_p - t_b), \sigma_b)$$

Multiplying by the prior on z_p and summing over all the possible values of z_p we get the marginal:

$$p(X_p | f_1, \dots, f_B, t_1, \dots, t_B, \sigma_1, \dots, \sigma_B) = \sum_{z_p} p(Z_p = z_p) \prod_{b=1}^B N(x_{pb} | f_b(z_p - t_b), \sigma_b)$$

Assuming the data from each subject is independent, we get the full likelihood:

$$p(X | f_1, \dots, f_B, t_1, \dots, t_B, \sigma_1, \dots, \sigma_B) = \prod_{p=1}^P \sum_{z_p} p(Z_p = z_p) \prod_{b=1}^B N(x_{pb} | f_b(z_p - t_b), \sigma_b)$$

This likelihood can be optimised with any method of choice such as MCMC sampling or gradient methods. In this work, we chose an iterative approach where, given an initial estimate of the trajectory shifts t we find the optimal levels of measurement noise σ , then optimise for the trajectory shifts and so on. For the datasets analysed, this approach converges after around 30-40 iterations.

4.3.3 Avoiding Singularities

Due to noise in the biomarker measurements, the mean of $f(s)$ can cross the $Y = 0$ axis which results in the integrated trajectory having a horizontal asymptote and then changing its direction. In order to avoid this we limit the integration of the line only on the segment that is in the $Y < 0$ half-space, where Y is the rate of biomarker change. This is the same as constraining the biomarker trajectories to be monotonically decreasing, an assumption that is consistent with the nature of most neurodegenerative diseases. If there exist two or more segments in the $Y < 0$ half-space, we choose the largest one that overlaps with the range of biomarker values of the patient data.

4.4 Performance Evaluation Methods

In this section we present a few ways that we used to evaluate the performance of disease progression models. In section 4.4.1 we present metrics which test staging consistency, i.e. that follow-up stages are greater or equal to baseline stages. We then generalise this concept in section 4.4.2 to test the accuracy of models in predicting the time lapse between two visits of a subject. Finally, in section 4.4.3 we present metrics that test the performance on predicting clinical diagnosis or conversion status.

4.4.1 Staging consistency

The staging consistency metrics test whether stages at a follow-up visit are greater than or equal to the stages at baseline. Let us consider a set of random variables z_t^i representing the stage of subject i at timepoint t , where $i \in [1 \dots N], t \in [1 \dots T_i]$, N being the number of subjects and T_i the number of time points for subject i . For most disease progression models we can find the posterior $p(z_t^i | X_i, \theta)$, which we will call the staging probabilities. Moreover, let $M_t^i = \arg \max_s p(z_t^i = s)$ be the maximum likelihood stage for subject i at time point t . This allows us to introduce the first performance metric, which we will call the *hard staging consistency* C_h , as follows:

$$C_h = \frac{1}{-N + \sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=2}^{T_i} \mathbb{I}[M_t^i > M_{t-1}^i] \quad (4.4)$$

C_h simply counts the proportion of stages from consecutive visits of every subject where the stage at the later visit must be greater than the stage at the earlier visit. The element $-N + \sum_{i=1}^N T_i$ is a normalising constant that represents the number of pairs of consecutive visits from all subjects and time points in the dataset. The C_h metric ranges from 0 (no consistent pairs of stages) to 1 (all pairs of stages are consistent).

The main limitation of the *hard staging consistency* is that it only uses the maximum likelihood stages. We are however interested in defining a soft version of this metric that would use all the values in the distribution of z_t^i . Therefore, we define the *soft staging consistency* $C_s^i(t_1, t_2)$ for subject i given two time points t_1 and t_2 as:

$$C_s^i(t_1, t_2) = p(z_{t_1}^i \leq z_{t_2}^i) = \sum_{s \in S} p(z_{t_2}^i = s) p(z_{t_1}^i \leq s) \quad (4.5)$$

where S is the set of possible stages in the disease progression model. We then define the *soft staging consistency* for the whole population as the mean of subject-specific consistencies for consecutive timepoints:

$$C_s = \frac{1}{-N + \sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=2}^{T_i} C_s^i(t_1, t_2) \quad (4.6)$$

It should be noted that the soft staging consistency would yield the same result as the hard staging consistency if the staging probabilities were expressed as Dirac delta functions, with all the likelihood being assigned to the maximum likelihood stage. However, most of the time there is uncertainty in the staging, and this uncertainty is penalised automatically by this metric. C_s also ranges from 0 (inconsistency between pairs of predicted stages) to 1 (consistency and usually very small uncertainty in predicted stages).

4.4.2 Time-lapse prediction

Time-lapse prediction is a generalisation of the staging consistency, where the disease progression model needs to predict how much time passed between two visits of the same subject, which is compared against the true time elapsed. As previously described, the soft and hard staging consistency metrics penalise if a follow-up stage is smaller than a baseline stage. However, we are also interested in a *time-lapse* metric that tests how well a model predicts the interval of time that passed between a baseline and follow-up visit. We can define the *hard time-lapse* metric D_h as follows:

$$D_h = \frac{1}{-N + \sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=2}^{T_i} |\tau(M_t^i) - \tau(M_{t-1}^i) - (a_t^i - a_{t-1}^i)| \quad (4.7)$$

where a_t^i is the age of subject i at timepoint t , M_t^i is the maximum likelihood stage for subject i at timepoint t and $\tau(M_t^i)$ is the estimated time from onset associated with stage M_t^i . The equivalent *soft time-lapse* metric D_s is defined as:

$$D_s = \frac{1}{-N + \sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=2}^{T_i} |\mathbb{E}[\tau(z_t^i) - \tau(z_{t-1}^i)] - (a_t^i - a_{t-1}^i)| \quad (4.8)$$

This metric therefore requires the model to be able to estimate a time from onset for every stage. This is not directly available for the event-based model, as we don't explicitly model the function τ . However, for the differential equation model τ is readily available. We have therefore only applied these consistency metrics to the differential equation model.

4.4.3 Diagnosis and conversion prediction

The accuracy of prediction of clinical diagnosis or conversion prediction can also be used to evaluate the performance of disease progression models. In order to avoid circularity, the model should not use cognitive tests that are used for setting the diagnosis. Diagnosis prediction can be done by finding the optimal threshold stage which separates two classes of subjects (e.g. Controls vs AD or MCI vs AD) using their predicted stages. For a new subject, one can then stage the subject and then assign a diagnosis depending on which side of the threshold it is. The threshold is usually chosen as the one that maximised a chosen metric such as balanced accuracy. The method above can also be used to predict conversion from one symptomatic group to another by finding the optimal staging threshold that separates converters from stable subjects. In order to avoid overfitting, N-fold cross-validation should be used, where the training data is used for model fitting and for finding the optimal threshold, while the testing data is used for measuring the predictive accuracy.

4.4.4 Differential diagnosis

Disease progression models such as the EBM or the DEM can also be evaluated on the accuracy in differential diagnosis, which requires building a different disease progression model for each disease. We propose to perform differential diagnosis in the following manner: for every disease

we train a disease progression model M_d on their corresponding dataset. Then, given the biomarker data X_j of a new patient, we compute the likelihood of the data given the model as:

$$p(X_j|M_d) = \sum_{s \in S} p(X_j|M_d, \theta, z_j = s)p(z_j = s) \quad (4.9)$$

where z^j is the stage of subject j and S is the set of all possible stages. We also set the prior on z_j to be uniform. Finally, we assign a new disease d_j to subject j as $d_j = \arg \max_d p(X_j|M_d)$.

4.5 Experiments

We evaluated two classes of disease progression models: the event-based model and the differential equation model. We tested three methods of fitting the EBM: (a) the *standard* EBM fitting method described in background section 2.4, (b) the *simultaneous MCMC sampling* approach described in section 4.2.1 and (c) the *Expectation-Maximisation* (EM) approach (section 4.2.2). For the DEM model we used two methods of aligning the obtained trajectories: (a) the standard method that aligns the trajectories using the average biomarker value of patients at baseline and (b) the optimised method described in section 4.3.2.

Each model has been tested on the staging consistency, time-lapse prediction on the DRC and ADNI datasets. On the DRC dataset, we also evaluated the models with respect to diagnosis prediction, while on ADNI we tested them on conversion prediction. We also tested a Support Vector Machine (SVM) on the differential diagnosis tasks using the same biomarker data. We performed 10-fold cross-validation on all experiments.

4.6 Results

4.6.1 DRC

In table 4.1 we show the staging-based metrics for the PCA cohort. Each entry shows the mean and standard deviation of the metric calculated over the 10 folds. Similar results are shown for the AD cohort in table 4.2. In table 4.3 we show the results of differential diagnosis in the DRC cohort. In each entry we show the mean and standard deviation of the balanced accuracy across the cross-validation folds.

Model	Staging Consistency		Time-lapse	
	Hard	Soft	Hard	Soft
EBM - Standard	0.88 ± 0.12	0.66 ± 0.09	-	-
EBM - Simultaneous Sampling	0.96 ± 0.06	0.70 ± 0.06	-	-
EBM - EM	0.95 ± 0.10	0.68 ± 0.11	-	-
DEM - Standard Alignment	0.94 ± 0.06	0.95 ± 0.05	0.54 ± 0.31	0.52 ± 0.29
DEM - Optimised Alignment	0.95 ± 0.05	0.95 ± 0.04	0.56 ± 0.28	0.52 ± 0.27

Table 4.1: Staging metrics on the DRC data using the PCA cohort. The mean and standard deviations are calculated for each testing set in 10-fold cross-validation.

Model	Staging Consistency		Time-lapse	
	Hard	Soft	Hard	Soft
EBM - Standard	0.91 ± 0.16	0.71 ± 0.07	-	-
EBM - Simultaneous Sampling	0.96 ± 0.07	0.76 ± 0.10	-	-
EBM - EM	0.99 ± 0.01	0.72 ± 0.07	-	-
DEM - Standard Alignment	0.87 ± 0.10	0.88 ± 0.08	0.72 ± 0.91	0.67 ± 0.92
DEM - Optimised Alignment	0.87 ± 0.10	0.88 ± 0.08	0.74 ± 0.92	0.69 ± 0.92

Table 4.2: Staging metrics on the DRC data using the AD cohort. The mean and standard deviations are calculated for each testing set in 10-fold cross-validation.

Model	PCA vs AD Controls vs PCA Controls vs AD		
	PCA vs AD	Controls vs PCA	Controls vs AD
EBM - Standard	0.72 ± 0.13	0.95 ± 0.05	0.90 ± 0.06
EBM - Simultaneous Sampling	0.79 ± 0.09	0.94 ± 0.06	0.90 ± 0.05
EBM - EM	0.80 ± 0.07	0.95 ± 0.05	0.87 ± 0.05
DEM - Standard Alignment	0.81 ± 0.07	0.95 ± 0.05	0.90 ± 0.11
DEM - Optimised Alignment	0.82 ± 0.09	0.93 ± 0.06	0.88 ± 0.14
Support Vector Machine	0.79 ± 0.14	0.91 ± 0.06	0.88 ± 0.07

Table 4.3: Diagnosis prediction on DRC data. Each entry shows the mean and standard deviation of the balanced accuracy across the cross-validation folds.

4.6.2 ADNI

In table 4.4 we show the staging-based performance results of the progression models on the ADNI dataset. As with the DRC results, for each metric we show its mean and standard deviation over the 10 cross-validation folds. In table 4.5 we also evaluated the models on how well they predict conversion from MCI to AD at 12-months, 24-months and 36-months from baseline visit. We did not compute results for prediction of conversion status in controls due to small and very imbalanced datasets.

Model	Staging Consistency		Time-lapse	
	Hard	Soft	Hard	Soft
EBM - Standard	0.83 ± 0.07	0.76 ± 0.05	-	-
EBM - Simultaneous Sampling	0.84 ± 0.05	0.76 ± 0.06	-	-
EBM - EM	0.84 ± 0.08	0.74 ± 0.06	-	-
DEM - Standard Alignment	0.87 ± 0.05	0.83 ± 0.08	0.85 ± 0.17	0.85 ± 0.16
DEM - Optimised Alignment	0.87 ± 0.05	0.84 ± 0.07	0.86 ± 0.15	0.86 ± 0.16

Table 4.4: Staging metrics on ADNI data

Model	Duration between baseline and follow-up		
	12 months	24 months	36 months
EBM - Standard	0.69 ± 0.14	0.64 ± 0.11	0.72 ± 0.15
EBM - Simultaneous Sampling	0.66 ± 0.14	0.63 ± 0.10	0.74 ± 0.14
EBM - EM	0.69 ± 0.15	0.63 ± 0.10	0.76 ± 0.15
DEM - Standard Alignment	0.73 ± 0.13	0.72 ± 0.14	0.70 ± 0.13
DEM - Optimised Alignment	0.64 ± 0.11	0.69 ± 0.12	0.75 ± 0.14
SVM	0.68 ± 0.15	0.70 ± 0.10	0.77 ± 0.08

Table 4.5: Prediction of conversion from MCI to AD on ADNI data.

4.7 Discussion

4.7.1 DRC

In the PCA cohort, most EBM models perform as good as the DEM models in terms of *hard staging consistency* but relatively worse in *soft staging consistency*. There is however some improvement in the novel EBM fitting methods compared to the standard EBM. Moreover, the DEM models perform very well in both hard and soft staging consistency, with a significantly better *soft staging consistency* compared to the EBM models. We also tested the DEM models using the time-lapse metrics, and found that there is no significant difference between the optimised and standard alignment procedures for biomarker trajectories. We could not test the EBMs using the time-lapse metrics because there is no notion of time in the EBM, which is a discrete model.

In the AD cohort we notice that the EBM models actually perform better than the DEM models in terms of *hard staging consistency*, but worse in *soft staging consistency*. These "contradictory" results could be explained by the fact that the soft staging consistency penalises uncertainty in the staging probabilities, and the EBM might have a lot of uncertainty due to not fitting the data well. The novel EBM methods again show some improvements compared to the Standard EBM method. In terms of the time-lapse, there is again no statistically significant difference between the two DEM methods.

In the diagnosis prediction tasks, most disease progression models have similar performance, with only the Standard EBM having a low performance in the PCA vs AD test. Surprisingly, for some tests such as the Controls vs PCA prediction, the disease progression models generally have similar or higher balanced accuracy than the SVM, which is a discriminative model optimised for this classification task.

4.7.2 ADNI

In terms of staging consistency the DEM models perform better than the EBM models, especially in terms of *soft staging consistency*. As with the DRC results, this might be explained by the fact that the soft staging consistency penalises uncertainty in the subject staging. This uncertainty might be higher in the EBM due to a bad fit of the data, as the EBM assumes biomarker trajectories are step-functions. However, in the case of both the EBM and DEM, the staging performance of the novel and standard methods is very similar, suggesting that the standard methods already model the data well enough. Also in terms of time-lapse, the two DEM methods perform equally well.

The results on conversion prediction in ADNI show that all models have a broadly similar performance at this task. However, a few clear differences can be noticed in some models. The DEM with standard trajectory alignment performs best at 12-months and 24-months conversion prediction, while the SVM and the novel fitting methods for EBMs and DEMs perform best at 36-month conversion prediction. This inconsistency might be due to some models having a better fit on certain parts of the disease time course.

4.7.3 Staging-based performance

The staging-based performance measures can pick up differences in the performance of the models in both DRC and ADNI datasets, in particular between different classes of models such as EBM vs DEM, where we see that DEM models usually perform better than EBM models. This might be the case because the DEM constructs continuous non-parametric biomarker trajectories which might give a better fit than the step-wise trajectories of the EBM. However, the disadvantage of non-parametric trajectories in the DEM is that they can overfit the data due to the large number of parameters that need to be estimated.

The staging consistency metrics can also pick up improvements in different fitting procedures, such as in the case of the EBM models. On the other hand, there are generally no statistically significant differences in the DEM between the standard versus optimised alignment, probably due to the fact that the standard alignment is already good enough for the two datasets that we tested it on.

There are some differences between the *hard* and *soft staging consistencies*. Fundamentally, the soft staging consistency penalises for uncertainty in the subject staging while the *hard* version does not. Nevertheless, in most of the experiments the *hard* and *soft staging consistency* metrics had consistent changes in values across different models. One exception was in the DRC data using the AD cohort, where the EBM methods had a higher score than DEM methods in terms of the *hard staging consistency*, but lower score in terms of the *soft staging consistency*. This can be explained by the fact that the hard staging consistency tends to favour models with a smaller set of possible stages such as the EBM due to floor and ceiling effects as well the fact that subject stages are forced to be snapped to the closest (discrete) stage. For example, a model with only one stage will always have 100% accuracy in the *staging consistency* metrics, due to all subjects being assigned the same stage.

4.7.4 Diagnosis prediction performance

The models have also been tested on predicting the clinical diagnosis of subjects or predicting converter status across different time periods. We generally found these metrics to discriminate less between different types of models. For example, in many experiments there were no consistent differences between most models in terms of diagnosis performance, suggesting that the models were all equally good or that the diagnosis was itself noisy. Also, in terms of prediction of conversion in ADNI, there was a lot of variability in the metric values across folds and also in the model rankings across experiments, making it hard to identify an overall winner model. For example, the DEM with standard alignment performed very well at prediction of conversion status after 12-months and 24-months from baseline, but relatively worse at 36-months from baseline. This suggests this model had a better estimation of biomarker trajectories in earlier stages compared to later stages. The variability of the diagnosis prediction metrics can also be attributed to inaccuracies and biases in the diagnosis process.

4.8 Summary

In this work we presented several extensions of the EBM and the DEM. We further devised performance metrics that measure the accuracy of the predicted subject stages and of predicted clinical diagnosis. We tested these on two independent datasets: ADNI and DRC. Our results show that in many situations the novel EBM and DEM fitting methods show improvements with respect to our performance metrics compared to the standard versions.

4.8.1 Limitations

The performance metrics we used for evaluation have certain inherent limitations which might limit their use for some disease progression models. For example, the staging consistency metrics are ultimately a consistency metric, which means that a specially crafted model can get perfect accuracy if it simply assigns the same stage to all subjects. If we knew the true underlying stage of each subject, we would be able to compare the predicted stages against these true stages without having to evaluate staging consistency. However, this is not the case with most datasets such as ADNI, where we only have a rough idea of where the subjects lie on the disease time-course based on their symptoms. However, the time-difference metric does not suffer from this problem, due to the fact that the model needs to predict precisely the time that passed between different visits to the clinic.

Another limitation of the *hard staging consistency* metric is that it tends to favour models with a small number of states, in our case the EBM methods, because of floor and ceiling effects. For example, if a subject might ideally lie in between two discrete stages it is instead snapped to the nearest stage. For instance, if a follow-up visit comes shortly after the baseline visit, then they are both likely to be assigned the same stage even if the biomarkers might be sensitive enough to pick up the subtle changes. Depending on the chosen biomarkers, it might take several years in order to progress from one stage of the EBM to the next one. This is not the case with continuous time models such as the DEM, where the stage of subjects can be measured more precisely given enough accuracy in the biomarker measurements and that at least one biomarker is "active" (i.e. not plateaued) during each stage of the disease.

The staging consistency metrics are based on the idea that the biomarkers evolve monotonically as the disease progresses. This is however not the case with some neurological disorders such as Multiple Sclerosis (MC), where a patient can have an attack (relapse) followed by a period of steady recovery (remission).

One limitation of the time difference metrics is that it requires the disease progression model to estimate the time from onset for every stage. This is not normally modelled in discrete models such as the event-based model or other methods based on Markov chains (e.g. [98]). However, it might be possible to extend these discrete models in order to estimate time since onset for each of the states.

The main limitation of the performance metrics based on prediction of diagnosis or conversion status is that these labels are usually biased and inaccurate, especially if they are not confirmed post-mortem by histology. This is a problem especially in ADNI, where a large number of controls already have abnormal CSF biomarkers and will develop MCI status within a few years.

4.9 Conclusion

Given a lack of ground truth in the evolution of biomarker trajectories and accurate subject stages, these performance metrics will allow us to test and validate any types disease progression models, identify limitations in these models and find corresponding improvements. The performance of some models can vary depending on the dataset used, the way the data was pre-processed, cohort size or how well-defined diagnostic groups are. While we have so far demonstrated these performance metrics only on event-based models and differential equation models, they could be used to test any other disease progression models such as the self-modelling regression approach by [15] or the disease progression score method by [62].

Future work will focus on evaluating other types of disease progression models such as the ones previously mentioned or on devising other performance metrics. A different avenue of future research can focus on devising validation techniques for diseases such as MS where biomarkers do not monotonically decrease.

Chapter 5

Event-based Model

In this chapter we present analysis that has been done with the event-based model on the DRC dataset. In section 5.1 we present the EBM analysis on the progression of Posterior Cortical Atrophy, while in section 5.2 we present the similar results on the progression of typical AD. Finally, in section 5.3 we present the predicted disease progression on three PCA subgroups.

5.1 PCA Analysis

5.1.1 Hypothesis

We are interested to find out in which order are brain regions affected along the time course of PCA. Previous studies that compared PCA vs control groups have shown that the posterior parts of the brain, namely the parietal, precuneus and occipital areas are mostly affected in PCA compared to controls [99, 100]. Therefore, we hypothesise that these are the earliest regions that become abnormal in PCA.

5.1.2 Method

We ran the standard Event-Based Model as formulated by Fontejin et al. [13] (section 2.4) on healthy controls and PCA data from the DRC dataset, using 48 biomarkers representing ROI volumes that have been corrected for TIV, age and gender. See section 3.2 for more details on how the images were acquired and segmented using the GIF algorithm. For the EBM event distributions, we used a mixture of a gaussian and a uniform distribution. We assumed that the control population are well-defined in the DRC dataset, so we fit the gaussian distribution directly on the control data. For the uniform distribution, we set the minimum and maximum values so as to cover all observed values of the biomarkers. For finding the optimal sequence, we used 25 different starting points and performed greedy ascent for 10,000 sequences. We chose the sequence with the maximum likelihood as the final sequence.

5.1.3 Results

The results for the PCA progression are shown in Fig. 5.1. In subfigure 5.1a we show the order in which cortical brain regions are affected at several stages of the EBM, according to our model. We show stages 8,16,24,32 and 40 out of 48, which is the total number of biomarkers. White regions are not yet affected while red regions are affected at the corresponding stage. The model assumes that all regions will eventually be affected by the disease. Moreover, in subfigure 5.1b we show the histogram of EBM stages for controls and PCA subjects.

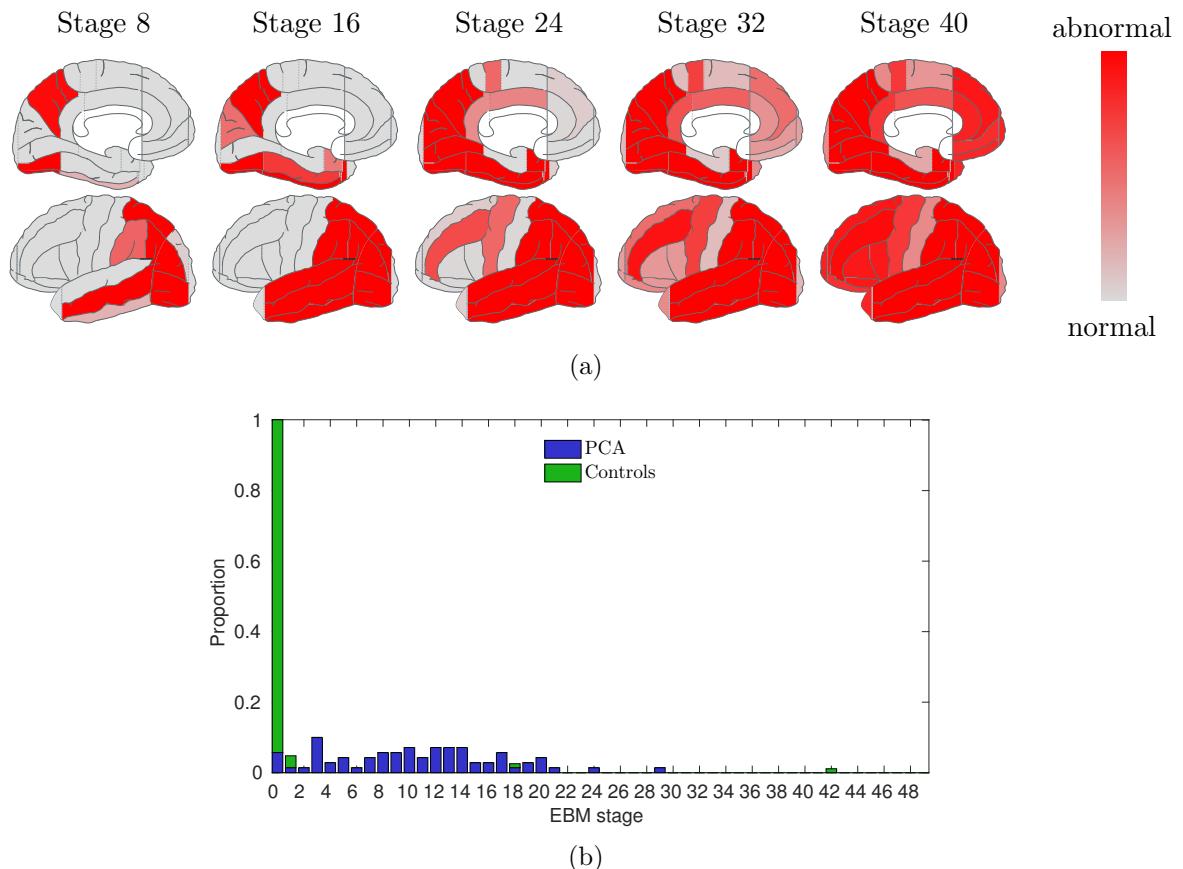


Figure 5.1: (a) Timing of atrophy in PCA subjects according to the Event-based Model. White regions have not been affected, while red regions have been affected by the corresponding stage. (b) The histogram of EBM stages for controls and PCA subjects.

5.1.4 Discussion

The EBM abnormality ordering confirms our original hypothesis, where showing that posterior regions such as occipital and superior parietal areas are the earliest to become abnormal, followed by temporal areas and inferior parietal. Moreover, in the staging histogram most controls are placed at stage zero, suggesting they have no abnormalities while PCA subjects are spread across the other stages.

5.2 AD Analysis

5.2.1 Hypothesis

In a similar way to 5.1, we are interested to find out the progression of tAD using the event-based model. Previous literature has shown that the hippocampus, entorhinal cortex and other temporal areas are the first brain regions to become affected in AD [20, 101, 102]. We are interested to test this hypothesis and further find out a fine-grained picture of progression in typical AD.

5.2.2 Method

For doing this analysis, we used the EBM on the healthy control and tAD data from the DRC dataset. The method is exactly the same as the PCA analysis described in section 5.1.2.

5.2.3 Results

The results for the tAD progression are shown in Fig. 5.2. In subfigure 5.2a we show the order in which cortical brain regions are affected at several stages of the EBM, according to our model. White regions are not yet affected while red regions are affected at the corresponding stage. The model assumes that all regions will eventually be affected by the disease. Moreover, in subfigure 5.2b we show the histogram of EBM stages for controls and tAD subjects.

5.2.4 Discussion

The tAD progression obtained by our data-driven model confirms previous patterns of atrophy, where the hippocampus, entorhinal, amygdala (not shown) are the earliest regions to become abnormal on in the disease time course. This is followed by other regions in this order: inferior and middle temporal regions, occipital regions, angular gyrus, precuneus and ventricles. The patient staging histogram also shows that controls are mostly at stage zero (no abnormal regions), while patients are spread across the disease stages.

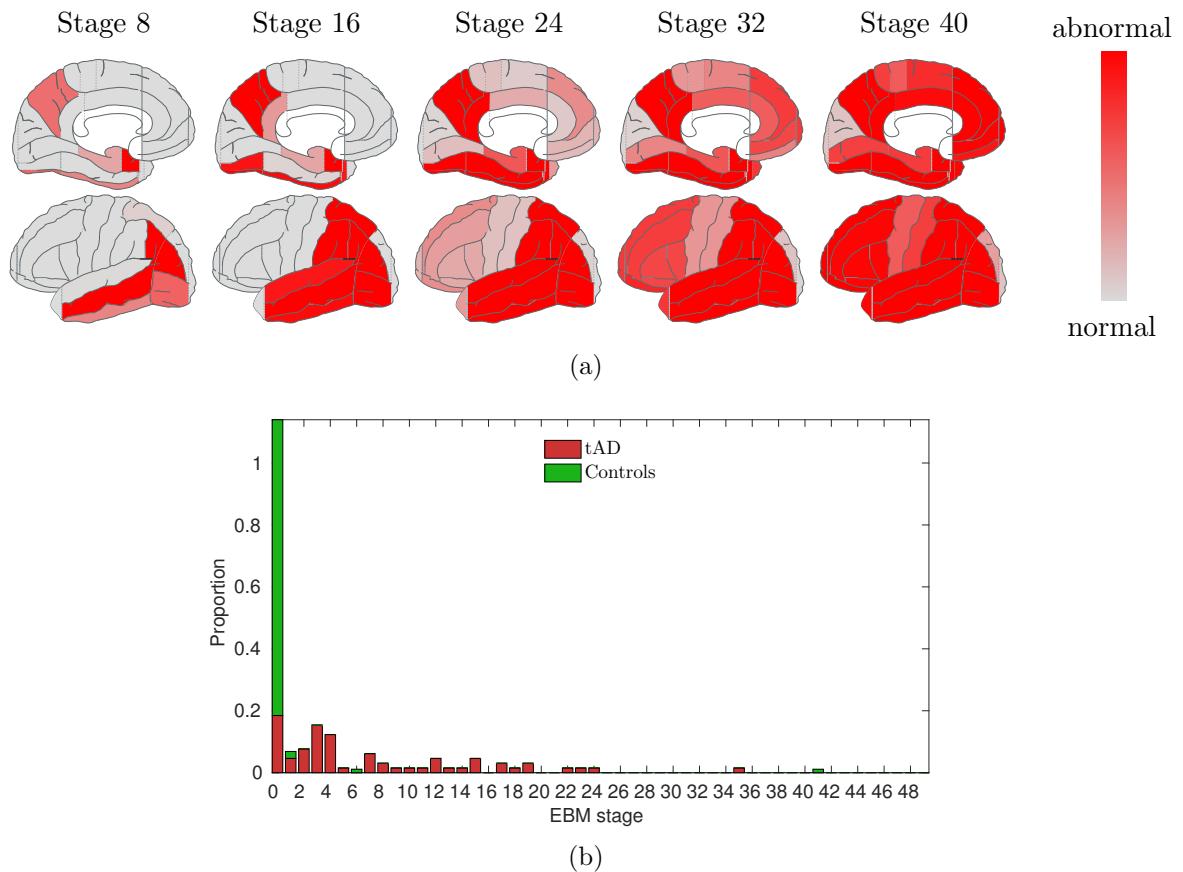


Figure 5.2: (a) Timing of atrophy in AD subjects according to the Event-based Model. White regions have not been affected, while red regions have been affected by the corresponding stage. (b) The histogram of EBM stages for controls and AD subjects.

5.3 PCA subgroups

5.3.1 Background

Although PCA is considered a homogeneous disease, there have been claims in past literature that there exist three main subgroups[103]: parietal (dorsal) and occipitotemporal (ventral) forms of PCA have been reported by Ross et al. [104], while a third primary visual (striate cortex, caudal) form of PCA has been proposed by Galton et al. [105]. In this section we are interested to analyse and compare the progression of these different subgroups.

We split our dataset into three groups based on performance on a suite of cognitive tests. For each subject we computed the following composite tests by summing up scores from individual tests:

1. Early visual processing: shape discrimination, figure-ground discrimination and simple crowding
2. Visuoperceptual processing: object decision, fragmented letters and usual and unusual views.
3. Visuospatial processing: number location, dot counting and a cancellation time (cut off at 90s)
4. Episodic memory: short recognition memory test (sRMT) for words and sRMT for faces

We then classified the subjects into three groups. The worst 1/3 of subjects on the early visual processing tests as compared to the memory tests (i.e. difference between early visual and memory tests) were assigned to the *vision subgroup*. The remaining 2/3 of participants were split into two groups according to a worse performance on the visuoperceptual or visuospatial tests. More precisely, the difference between visuoperceptual and visuospatial was computed and the top half of participants was assigned to the *spatial subgroup*, while the bottom half was assigned to the *object subgroup*. Detailed demographics for the three PCA subgroups are shown in table 5.1.

5.3.2 Hypotheses

We formulate the following hypotheses, based on evidence from previous research[104, 105]:

- *Vision subgroup*: atrophy starts in the occipital lobe
- *Space subgroup*: atrophy starts in the superior parietal lobe
- *Object subgroup*: atrophy starts in the inferior temporal lobe

Subgroup	Number of Subjects	Gender M/F	Age at baseline (years, mean \pm std)	Years from onset (years, mean \pm std)	Number of scans (mean \pm std)
Vision	23	6/17	64.4 \pm 7.7	4.5 \pm 2.8	2.2 \pm 1.1
Space	21	10/11	62.0 \pm 6.8	4.4 \pm 2.8	2.4 \pm 1.4
Object	18	7/11	62.3 \pm 8.2	4.5 \pm 2.7	2.7 \pm 1.6

Table 5.1: Baseline population demographics for PCA subgroups.

5.3.3 Method

We ran the EBM on data from each of the three subgroups separately, along with the shared healthy controls population. We used the same method as in the PCA analysis described in section 5.1.2. The only difference is that we increased the range of the uniform distribution to make it span a wider interval. This was done because, due to the use of a gaussian and uniform mixture of distributions, some controls with "very good" biomarker values would be assigned to the abnormal group and subsequently later stages. Therefore, for each of the groups (*vision*, *space*, *object*), we multiplied the initial range of the uniform distribution by a factor of 2,5 and 1 respectively.

5.3.4 Results

The positional variance matrix for the three subgroups (*vision*, *space*, *object*) are shown in figures 5.3, 5.5 and 5.7 respectively. On the y-axis, we show the biomarkers representing ROI volumes ordered according to the maximum likelihood sequence S from Eq. 2.4, while on the x-axis we show the position at which they become abnormal in the sequence. Each gray square at position (i, j) gives the probability that biomarker i is placed on position j in the sequence, ranging from 0 (white cells) and to 1 (black cells).

5.3.5 Discussion

For the *vision subgroup*, our hypothesis is confirmed, as the EBM suggests that the superior occipital, occipital fusiform and inferior occipital areas are the first to be affected (in no particular order). This is followed by the whole brain and parietal areas (superior parietal and precuneus). This is further followed by angular, middle occipital and middle temporal.

In the *space subgroup*, we find that the superior parietal lobe and the inferior occipital gyrus become abnormal first, again confirming our initial hypothesis. This is then followed by atrophy in the middle occipital, middle temporal, occipital fusiform and angular regions. Moreover, atrophy in the whole brain becomes detectable at this stage.

In the *object subgroup*, we find that the first two regions to become affected are the inferior occipital gyrus and the ventricles, which does not match our initial hypothesis that the inferior temporal region is affected at the very beginning. However, the inferior temporal region is affected right after, which is still early on in the sequence. The fact that the occipital regions are also affected in this subgroup was also confirmed by previous studies [104, 106].

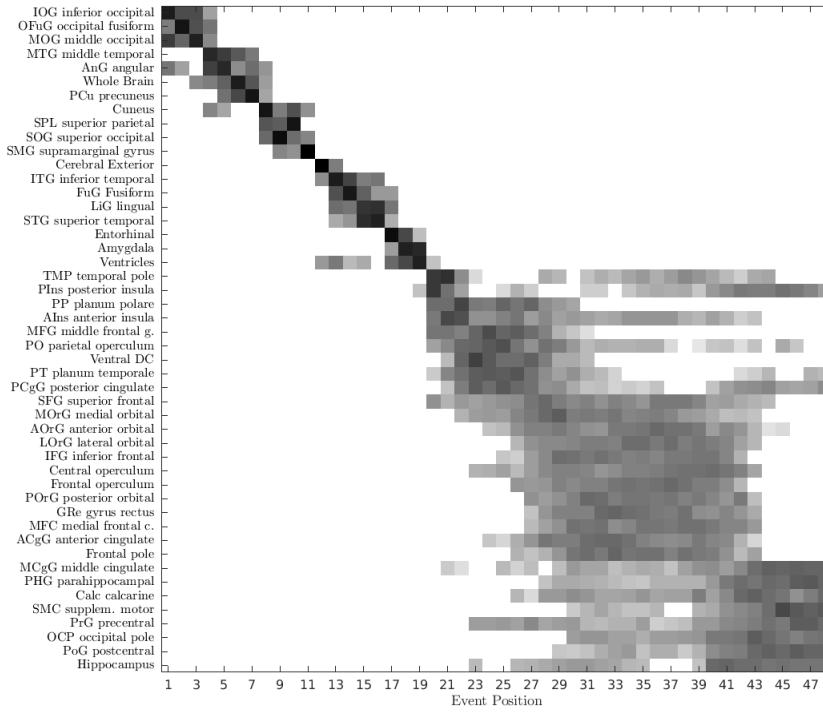


Figure 5.3: Positional variance matrix for the *vision subgroup*. On the y-axis, we show the biomarkers representing ROI volumes ordered according to the maximum likelihood sequence, while on the x-axis we show the position at which they become abnormal in the sequence. Each gray square at position (i, j) gives the probability that biomarker i is placed on position j in the sequence, ranging from 0 (white cells) and to 1 (black cells). Our initial hypothesis was that atrophy starts in the occipital lobe, which is confirmed by the EBM analysis.

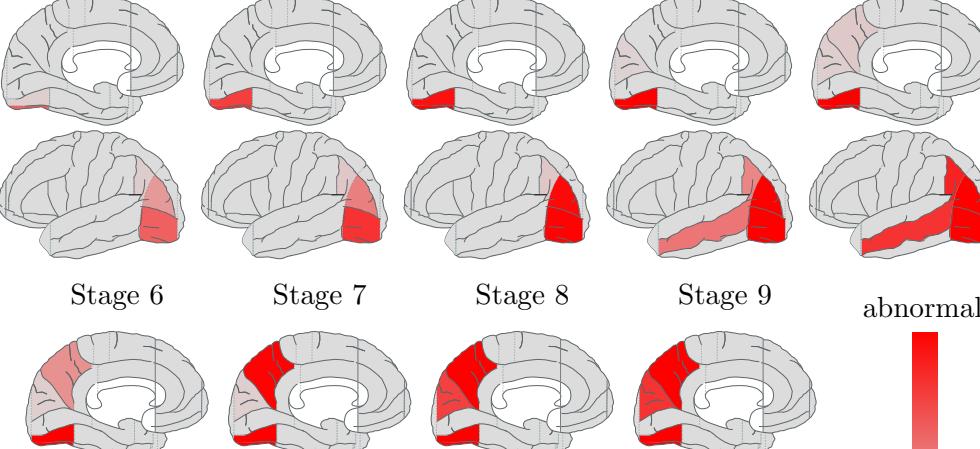


Figure 5.4: (a) Timing of atrophy for the *vision subgroup*. White regions have not been affected, while red regions have been affected by the corresponding stage.

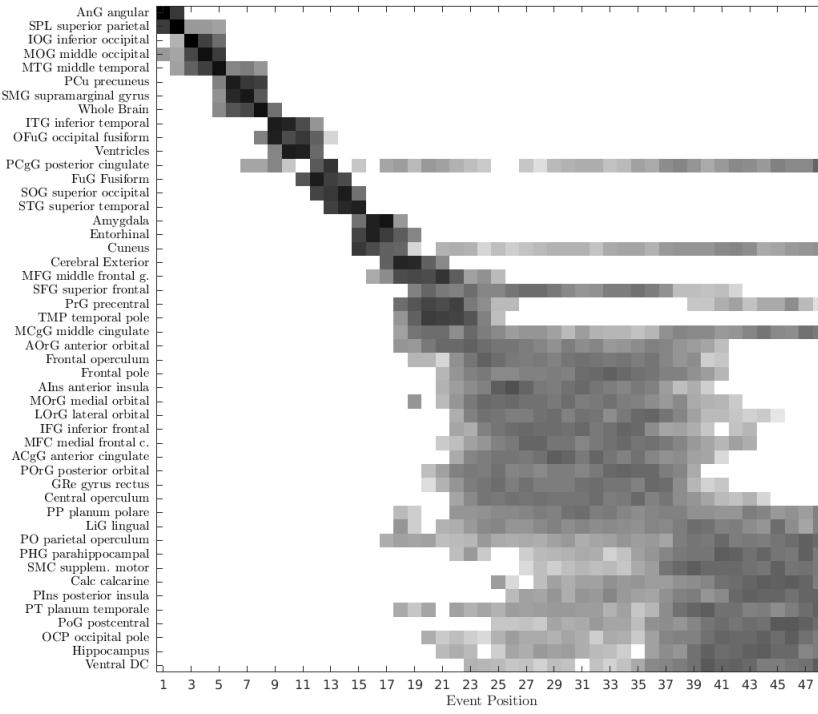


Figure 5.5: Positional variance matrix for the *space subgroup*. On the y-axis, we show the biomarkers representing ROI volumes ordered according to the maximum likelihood sequence, while on the x-axis we show the position at which they become abnormal in the sequence. Each gray square at position (i, j) gives the probability that biomarker i is placed on position j in the sequence, ranging from 0 (white cells) and to 1 (black cells). Our initial hypothesis was that atrophy starts in the superior parietal lobe. This is confirmed by the EBM analysis, which suggests atrophy starts in the angular and superior parietal regions.

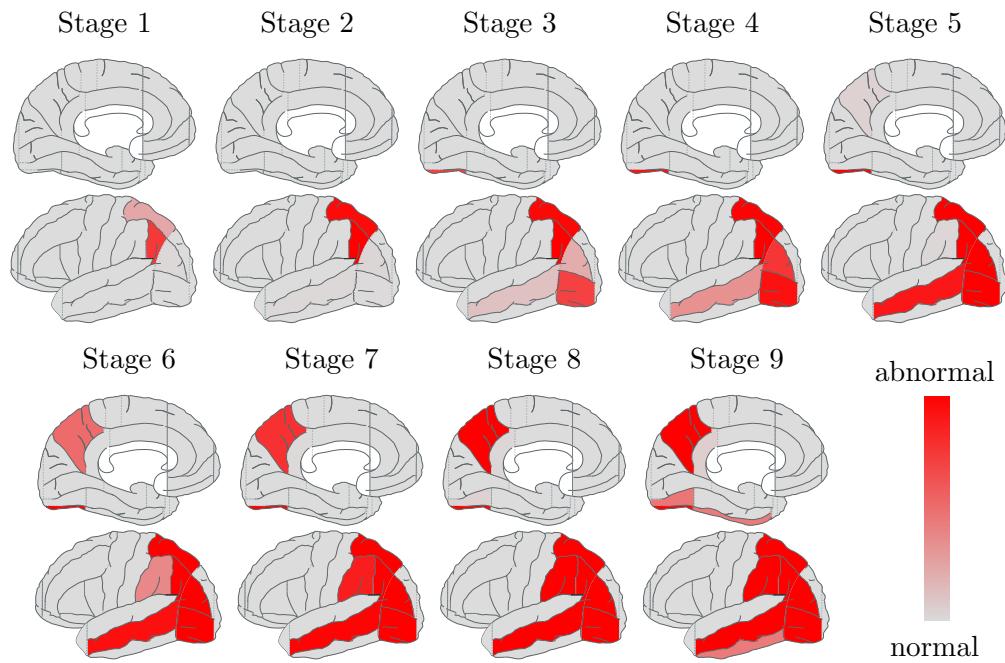


Figure 5.6: (a) Timing of atrophy for the *space* subgroup. White regions have not been affected, while red regions have been affected by the corresponding stage.

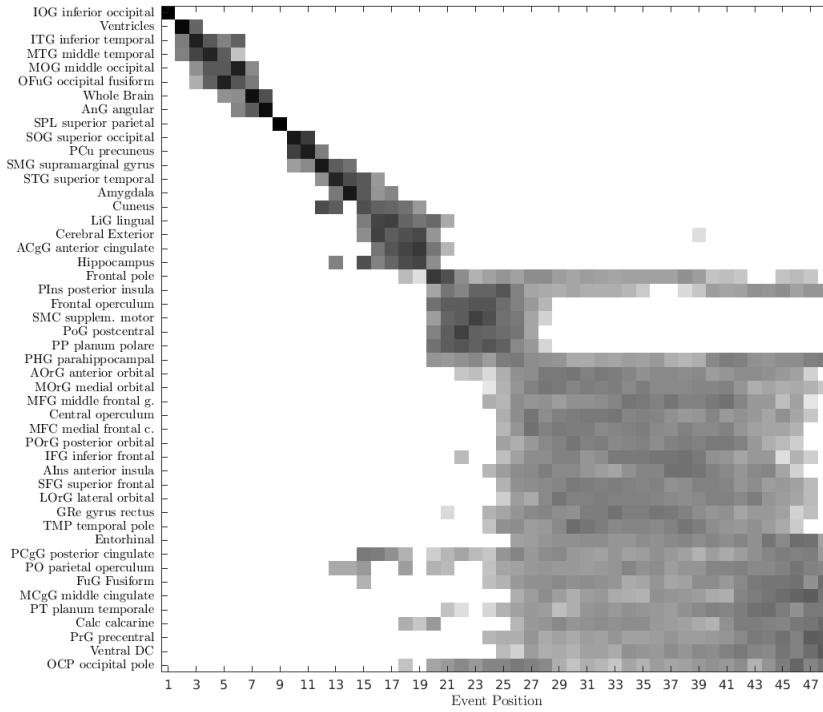


Figure 5.7: Positional variance matrix for the *object subgroup*. On the y-axis, we show the biomarkers representing ROI volumes ordered according to the maximum likelihood sequence, while on the x-axis we show the position at which they become abnormal in the sequence. Each gray square at position (i, j) gives the probability that biomarker i is placed on position j in the sequence, ranging from 0 (white cells) and to 1 (black cells). Our initial hypothesis was that atrophy starts in the inferior temporal lobe, which is not confirmed by the EBM analysis as the inferior temporal region becomes abnormal only after the inferior occipital and ventricles.

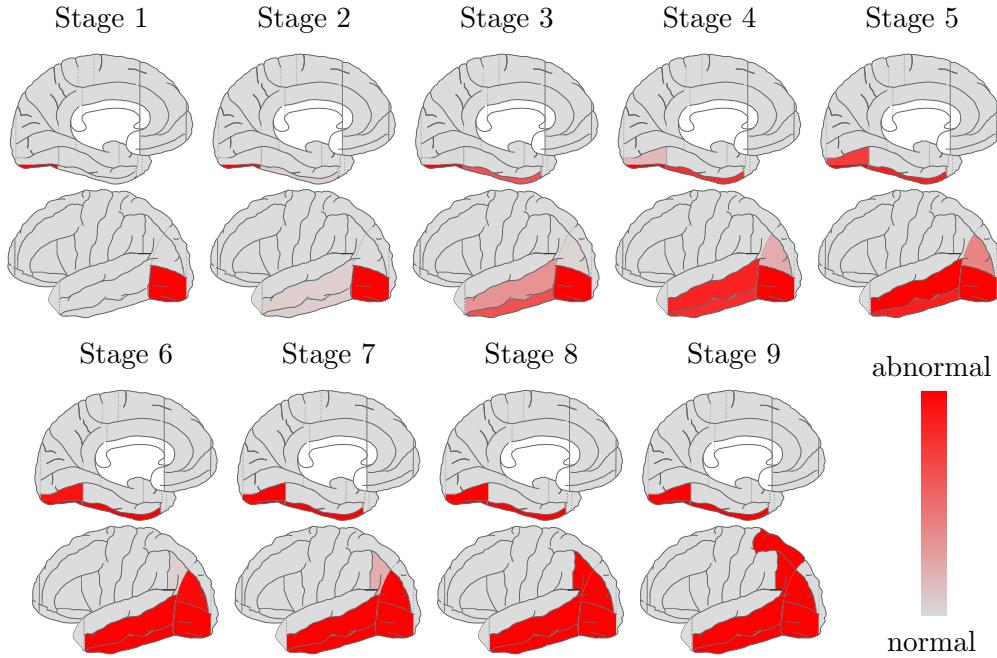


Figure 5.8: (a) Timing of atrophy for the *object* subgroup. White regions have not been affected, while red regions have been affected by the corresponding stage.

5.4 Conclusion

In this chapter we showed that the EBM is a powerful tool for studying progression of typical Alzheimer disease and other subtypes. The main advantages of the model is that it does not require longitudinal data, making it especially useful for analysing rare dementias where no large, longitudinal datasets currently exist. We have shown that the EBM results confirm patterns of atrophy previously reported in the literature, but gives a much more fine-grained picture of the temporal progression than these studies.

The current analysis has several limitations. First of all, we assumed that the control population was well defined in order to fit the distribution of normal biomarker values directly on control data. However, the controls might show some abnormalities in CSF biomarkers, but no such data was available. Secondly, we only explored brain volumes, but other biomarkers such as cortical thickness can be further analysed using our data.

There are several avenues of future research. For fitting the distributions of normal and abnormal biomarker values we could use data-driven methods such as mixture model optimisation. Another option is to use the optimised EBM fitting procedures such as simultaneous sampling (section 4.2.1) or Expectation-Maximisation (section 4.2.2). Further analysis also needs to be done in order to answer questions related to the extent and rate of atrophy in PCA and the respective subgroups. This analysis has been done with the differential equation model in chapter 6, but not at the PCA subgroup level due to the lack of sufficient longitudinal data.

Chapter 6

Differential Equation Model

In this chapter we present a longitudinal analysis performed on the DRC data using the differential equation model (see section 2.5). In section 6.1 we present the clinical questions that we want to answer, in section 6.2 we present the methodological details, in 6.3 we present the results and in 6.4 we discuss our findings.

6.1 Clinical questions

The main questions we want to answer in this longitudinal analysis are:

1. In PCA, what are the differences in timing of atrophy and rates of atrophy between different brain regions?
2. In tAD, what are the differences in timing of atrophy and rates of atrophy between different brain regions?
3. Do people with PCA and AD lose brain volume at same speed?
4. What are the differences in rates of atrophy across brain regions in PCA compared to AD?

6.2 Method

We used the differential equation model as described in section 2.5, where we fit a Gaussian Process (GP) to the change in biomarker scores $\Delta s/\Delta t$. The GP was fit independently for each biomarker. We used a non-parametric model such as GPs in order not to impose any parametric shape of the biomarker trajectory. Furthermore, given that the DRC dataset has a well-defined control population, we do not include the controls in the fitting of the GP as there is very little change in their biomarker values across timepoints which is not due to measurement noise. Moreover, in order to get a good fit of the GP we normalise the average biomarker values to z-scores and standardise the rates of change by dividing them with the average rate of change of all patients. The prior of the GP is kept at $Y = 0$, which ensures that the trajectory has upper and lower asymptotes.

After fitting the GP process we extracted the mean of the GP and sampled 20 trajectories from the posterior distribution of the GP in order to measure uncertainty in the model fit. The mean trajectory and trajectory samples were integrated using the numerical approach from Eq. 2.13. The resulting trajectories were however not aligned on the temporal axis, since we fit the model for each biomarker independently. We therefore aligned the $t = 0$ point so that $f(t = 0)$ is the average biomarker value for every patient at baseline visit. More information about the alignment is give in section 4.3.1. This process was repeated independently for every mean

trajectory in every biomarker. We also aligned the sampled trajectories in a similar fashion, with the only distinction that at the alignment point we added an amount of noise on the y-axis. This noise was computed as the standard deviation of the biomarker measurements for each patient at baseline visit. In order to discuss the differences in rates of atrophy across different biomarkers of diseases, we qualitatively compare the rates of atrophy at the point where the maximum rate of atrophy is attained. Moreover, a meaningful comparison of rates of atrophy across biomarkers requires them to be in a similar scale, so we converted the biomarker values representing ROI volumes to z-scores relative to controls.

6.3 Results

In Fig. 6.1a we show the temporally aligned average trajectories for PCA subjects across several ROIs, while in Fig. 6.1b we show the trajectories for AD subjects. On the x-axis we show the number of years since baseline visit, while on the y-axis we show the z-score of the ROI volume relative to controls. The direction of the "ventricles" biomarker has been manually flipped for consistency with the other biomarkers. In Fig. 6.2 we show the trajectories of ROI volumes aligned for PCA and tAD along with integrated samples from the GP posterior that have been integrated and temporally aligned. The integrated posterior samples give an idea of the confidence in the average biomarker trajectory. Results from figure 6.1 can help us answer questions 1 and 2, while results from Fig. 6.2 can help us answer questions 3 and 4.

6.4 Discussion

In the PCA progression from 6.1a, in terms of timing of atrophy we find that the occipital and parietal lobes are the regions most affected early in the disease time course, before baseline visit. However, soon after baseline visit other biomarkers such as whole brain, temporal lobe and ventricles become abnormal. This matches our previous results on the same dataset using a fundamentally different model (the EBM, see section 5.1), but provides more detail on the exact timing of atrophy. In terms of rates of atrophy, we see that the whole brain, ventricles and the temporal lobe are the ones showing the highest drop, which happens between baseline visit and 5 years later. While these findings confirm previous results in the literature [99, 100], they provide a more detail regarding the exact timing and rate of atrophy decline over a timespan of 30 years.

In the tAD progression from 6.1a, in terms of timing of atrophy we find that the hippocampus is the first to become abnormal in the early stages, which is something that has been consistently confirmed across many other studies. Our model also predicts that in early stages the hippocampus is more sensitive than the entorhinal cortex and that the hippocampal atrophy precedes entorhinal atrophy. This has been confirmed by some previous studies [107, 13] but rejected by others [108, 109]. However, in middle and late stages we see a much more global patterns of atrophy affecting most brain areas. In terms of rate of atrophy, we find that in early stages many regions have a similar rate of decline, with only the hippocampus and ventricles showing a slightly slower rate. This is in contrast with previous findings in the literature such as [31] that find the hippocampus to have the highest rate of decline. We also find very similar rates especially right after the baseline visit, with slightly higher drops in volume for temporal and whole brain biomarkers.

When comparing PCA versus tAD, we find several differences in atrophy timings, rates and extent. In terms of whole brain, there is a comparable timing and rate of atrophy in both PCA and tAD in early stages, but in later stages at around baseline we find that there is more atrophy in PCA as compared to tAD. For the ROIs, we find that in PCA there is a significant amount of early and extensive atrophy in occipital and parietal areas as compared to tAD. On the other hand, in tAD there is significantly more atrophy in the hippocampus and entorhinal

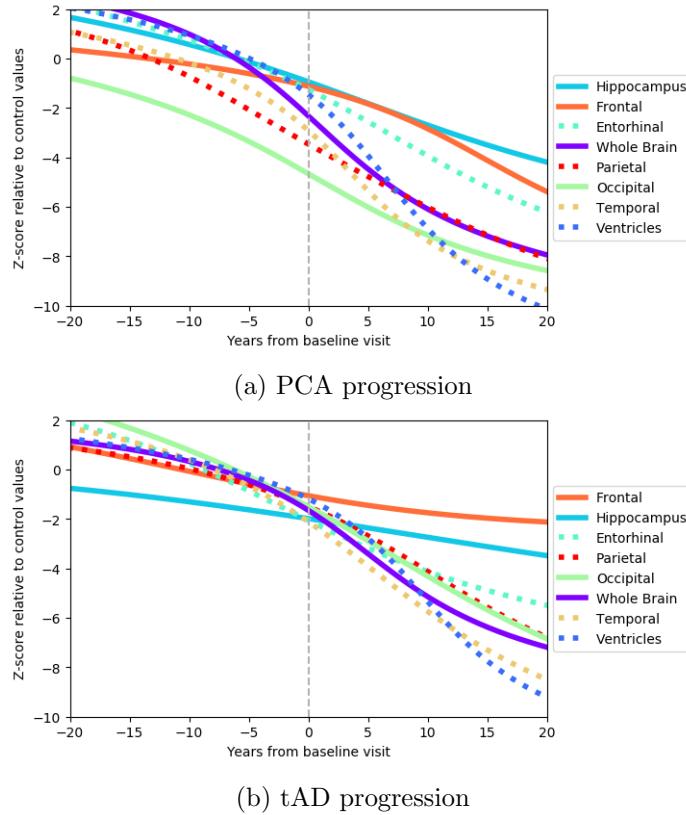


Figure 6.1: Trajectories of different ROI volumes from the differential equation model on the (a) PCA progression and (b) tAD progression. On the x-axis we show the number of years since baseline visit, while on the y-axis we show the z-score of the ROI volume relative to controls.

areas at any point in the disease, with more extensive atrophy in the hippocampus. Moreover, the hippocampus also reaches abnormal levels (-1 standard deviation) much earlier (15 to 10 years before baseline scan) in tAD compared with PCA (2 to 3 years after baseline scan). We notice that for many biomarkers such as temporal, entorhinal and frontal there are reasonably similar rates of atrophy at many stages of the two diseases.

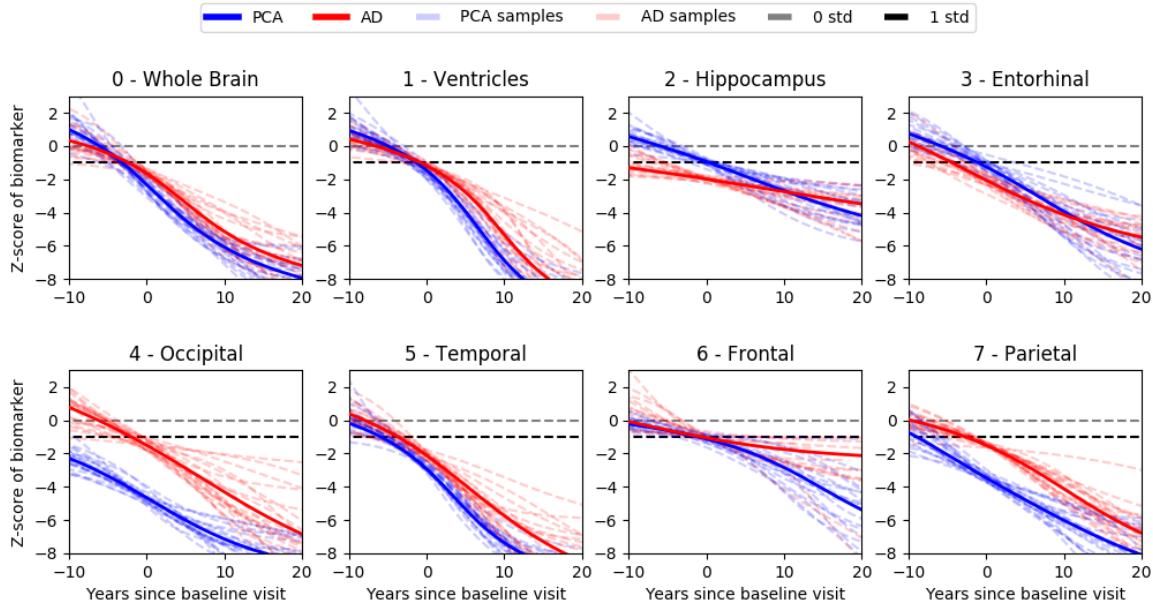


Figure 6.2: Mean trajectories of ROI volumes for PCA and tAD aligned on the same temporal scale along with samples from the posterior distribution showing the confidence in the mean trajectory. On the x-axis we show the number of years since baseline visit, while on the y-axis we show the z-score of the ROI volume relative to controls.

6.5 Conclusion

In this section we presented a longitudinal analysis of volume changes in PCA and tAD using the differential equation model. We compared the timing, rates and extent of atrophy across different regions in both PCA and tAD. Moreover, we also compared the evolution of each of these biomarkers across the two AD subtypes: tAD and PCA. While the patterns of tAD progression have been extensively studied using MRI biomarkers, there is significantly less literature on the progression of PCA. This analysis is the first to use data-driven disease progression models for studying atrophy evolution in PCA. Our analysis is also the first to directly compare MRI biomarkers in PCA vs tAD in terms of timing, rate and atrophy extent.

There are several limitations with our analysis. First of all, we aligned the trajectories using the average biomarker value at patient baseline visit. This makes the assumption that the biomarker values corresponding to the stages of each patient at baseline only change linearly in that region. Secondly, the model fits each biomarker trajectory independently from the others and doesn't model biomarker correlations. Third, the confidence given by the posterior samples is not a true confidence interval, because these samples also have to be temporally aligned. This can bias the confidence interval, in making it smaller near the alignment point and larger at the ends. Moreover, the time $t = 0$ is set as the time since baseline scan instead of disease onset. This becomes problematic when comparing across different diseases (PCA vs tAD) as symptoms might develop at different stages in the disease and this will influence when the subjects will come to the clinic for a baseline scan and assessment.

There are several possibilities for future work. The model can be extended to a multimodal framework by fitting an N-dimensional GP on the data and then integrating along all the biomarker values at the same time. This extension will remove the need to align the trajectories from different biomarker on the same temporal axis, but one still needs to set a $t = 0$ starting point. Other clinical questions should also be explored, such as analysing subgroups within PCA, performing time to symptom onset analyses or asymmetry analyses. Other types of

MRI biomarker measurements such as cortical thickness should also be analysed and linked to cognitive measures.

Chapter 7

Voxelwise Temporal Clustering Model

7.1 Introduction

The main limitation of data-driven models such as the Event-Based Model (section 2.4), the Differential Equation Model (section 2.5) or the Disease Progression Score (DPS) (section 2.6) 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 would give new insights into the disease process and enable more precise staging. A voxel-wise disease progression model by Bilgel et al. [69] (see section 2.9) has been recently proposed to mitigate this problem, that uses amyloid measures in each voxel as its input data.

The voxel-wise model by [69] has two limitations: (1) it does not show which areas of the brain have a similar progression dynamics; (2) the biomarker trajectories are assumed to be linear, which are not suitable given that many biomarkers such as amyloid or tau plateau after a certain point. Moreover, 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 [77].

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 [69].

We first show using simulated data that our model is able to recover the underlying clusters, trajectory parameters and subject stages. We then apply our model to cortical thickness vertex-wise measures using ADNI and DRC datasets 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.

7.2 Methods

7.2.1 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 [14, 15, 16], 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 [14, 15, 16], 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 each trajectory 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} :

$$s_{ij} = \alpha_i t_{ij} + \beta_i \quad (7.1)$$

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

$$f(s; \theta_k) = \frac{a_k}{1 + \exp(-b_k(s - c_k))} + d_k \quad (7.2)$$

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:

$$p(V_l^{ij} | \alpha_i, \beta_i, \theta_{Z_l}, \sigma_{Z_l}, Z_l) = N(V_l^{ij} | f(\alpha_i t_{ij} + \beta_i | \theta_{Z_l}), \sigma_{Z_l}) \quad (7.3)$$

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 7.1, 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:

$$p(V_l, Z_l | \alpha, \beta, \theta, \sigma) = \prod_{(i,j) \in I} N(V_l^{ij} | f(\alpha_i t_{ij} + \beta_i | \theta_{Z_l}), \sigma_{Z_l}) \quad (7.4)$$

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:

$$p(V, Z | \alpha, \beta, \theta, \sigma) = \prod_l^L \prod_{(i,j) \in I} N(V_l^{ij} | f(\alpha_i t_{ij} + \beta_i | \theta_{Z_l}), \sigma_{Z_l}) \quad (7.5)$$

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 :

$$p(V | \alpha, \beta, \theta, \sigma) = \prod_{l=1}^L \sum_{k=1}^K p(Z_l = k) \prod_{(i,j) \in I} N(V_l^{ij} | f(\alpha_i t_{ij} + \beta_i | \theta_k), \sigma_k) \quad (7.6)$$

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.

7.2.2 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:

$$z_{lk} = \frac{\prod_{(i,j) \in I} N(V_l^{ij} | f(\alpha_i^{old} t_{ij} + \beta_i^{old} | \theta_k^{old}), \sigma_k^{old})}{\sum_{m=1}^K \prod_{(i,j) \in I} N(V_l^{ij} | f(\alpha_i^{old} t_{ij} + \beta_i^{old} | \theta_m^{old}), \sigma_m^{old})} \quad (7.7)$$

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

$$\log z_{lk} \propto -\frac{1}{2} \log (2\pi (\sigma_k^{old})^2) |I| - \frac{1}{2(\sigma_k^{old})^2} \sum_{(i,j) \in I} (V_l^{ij} - f(\alpha_i^{old} t_{ij} + \beta_i^{old} | \theta_k^{old}))^2 \quad (7.8)$$

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:

$$\theta_k = \arg \min_{\theta_k} \left[\sum_{l=1}^L z_{lk} \sum_{(i,j) \in I} (V_l^{ij} - f(\alpha_i t_{ij} + \beta_i | \theta_k))^2 \right] - \log p(\theta_k) \quad (7.9)$$

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 further simplifications we obtain the following problem:

$$\alpha_i, \beta_i = \arg \min_{\alpha_i, \beta_i} \left[\sum_{l=1}^L \sum_{k=1}^K z_{lk} \frac{1}{2\sigma_k^2} \sum_{j \in I_i} (V_l^{ij} - f(\alpha_i t_{ij} + \beta_i | \theta_k))^2 \right] - \log p(\alpha_i, \beta_i) \quad (7.10)$$

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 it has been shown that EM still works if we only find an increase in the log-likelihood [110]. This approach that involves a partial M-step is called Generalised EM.

7.2.3 Initialisation and Implementation

The optimisation algorithm is given in figure 7.1. 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 [14], 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 also perform successive refinements of parameters θ, α and β using equations 7.9, 7.10 until convergence. For the numerical optimisation we used 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.

```

1 Initialise  $\alpha, \beta, z_{lk}$ 
; // Outer loop
2 while  $z_{lk}$  not converged do
    ; // M-step
3     while  $\theta, \sigma, \alpha, \beta$  not converged do
        ; // Optimise trajectories
        for  $k = 1$  to  $K$  do
             $\theta_k = \arg \min_{\theta_k} \sum_{l=1}^L z_{lk} \sum_{(i,j) \in I} (V_l^{ij} - f(\alpha_i t_{ij} + \beta_i | \theta_k))^2 - \log p(\theta_k)$ 
             $\sigma_k^2 = \frac{1}{|I|} \sum_{l=1}^L z_{lk} \sum_{(i,j) \in I} (V_l^{ij} - f(\alpha_i t_{ij} + \beta_i | \theta_k))^2 - \log p(\sigma_k)$ 
        end
        ; // Optimise subject shifts
        for  $i = 1$  to  $S$  do
             $\alpha_i, \beta_i = \arg \min_{\alpha_i, \beta_i} \left[ \sum_{l=1}^L \sum_{k=1}^K z_{lk} \frac{1}{2\sigma_k^2} \sum_{j \in I_i} (V_l^{ij} - f(\alpha_i t_{ij} + \beta_i | \theta_k))^2 \right] - \log p(\alpha_i, \beta_i)$ 
        end
        ; // Re-scale subject shifts
        for  $i = 1$  to  $S$  do
             $\alpha_i = \frac{\alpha_i}{\sigma_N}, \beta_i = \frac{\beta_i - \mu_N}{\sigma_N}$ 
        end
        ; // Make trajectory parameters identifiable
        for  $k = 1$  to  $K$  do
             $\theta_k = \text{make\_identifiable}(\theta_k)$ 
             $\theta_k = \text{rescale}(\theta_k)$ 
        end
    end
    ; // E-step
    for  $l = 1$  to  $L$  do
        for  $k = 1$  to  $K$  do
             $z_{lk} \propto -\frac{1}{2} \log (2\pi\sigma_k^2 |I| - \frac{1}{2\sigma_k^2} \sum_{(i,j) \in I} (V_l^{ij} - f(\alpha_i t_{ij} + \beta_i | \theta_k))^2)$ 
        end
    end
24 end

```

Figure 7.1: The model fitting algorithm.

7.3 Simulation results

We first tested the model in a toy scenario using synthetic data to see how well the cluster probabilities, trajectories and subject shifts are recovered. We generated our data 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. 7.2a, 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. 7.2a, gray lines)
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 $A = \frac{1}{L} \sum_{l=1}^L p(Z_l = a[l])$. However, the recovered clusters can be found in a permuted order, so we used the permutation that yielded the highest agreement A . The method also requires us to set the number of clusters *a priori*, so in order to find the optimal number of clusters we used model selection with the Bayesian Information Criterion (BIC).

The BIC analysis correctly predicted 3 clusters for this synthetic experiment. Using three clusters, we also obtained a clustering agreement of $A = 1.0$. Figure 7.2a 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. Moreover, in Fig. 7.2b 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.

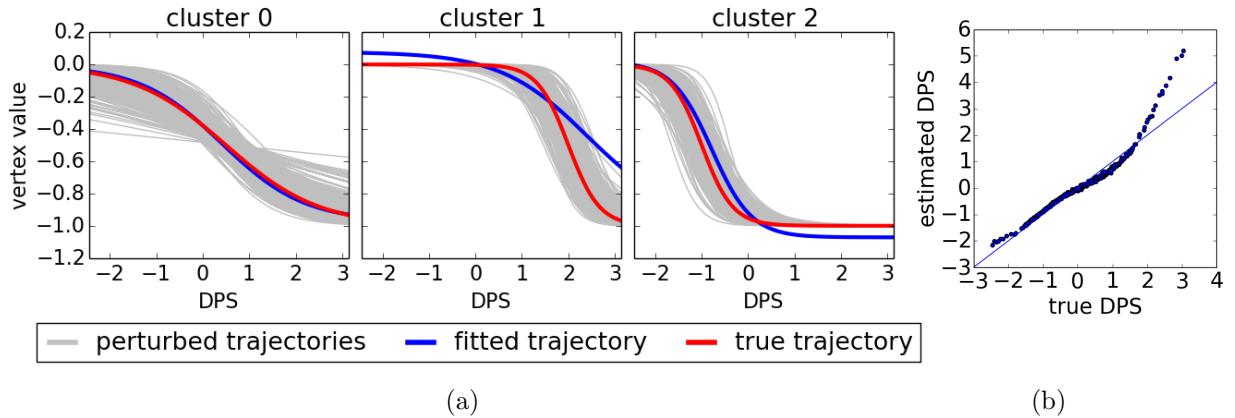


Figure 7.2: (a) Reconstructed temporal trajectories (blue) from the synthetic data along with the true trajectories (red). The data was generated from the perturbed trajectories, which in turn were generated from the true trajectories. (b) Estimated subject-specific disease progression scores compared to the true scores.

7.4 Experimental results

7.4.1 Data acquisition and preprocessing

We analysed cortical thickness measures from two datasets: ADNI and DRC dataset. For ADNI, we downloaded from <http://adni.loni.usc.edu/> 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 thickness measures we ran the Freesurfer longitudinal pipeline [111], 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 Freesurfer template and smoothed at various full-width/half-max (FWHM) values. We used the FWHM zero level and 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.

7.4.2 Results with ADNI and DRC datasets

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 would like to find out: (1) if we get similar results using our model on two independent tAD datasets: ADNI and DRC datasets 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. 7.3a 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.

[7.3d](#) 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. [7.3b](#) and [7.3e](#)) and PCA subjects (Fig. [7.3c](#) and [7.3f](#)).

We notice that in tAD subjects using both ADNI and the DRC dataset (Fig. [7.3](#)), 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 [49, 50]. However, in contrast to those approaches, our model gives insight into the timing, rate and extent of atrophy.

In the PCA subjects (Fig. [7.3c](#)), 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 [8]. 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. [7.3d](#), [7.3e](#), [7.3f](#)). 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.

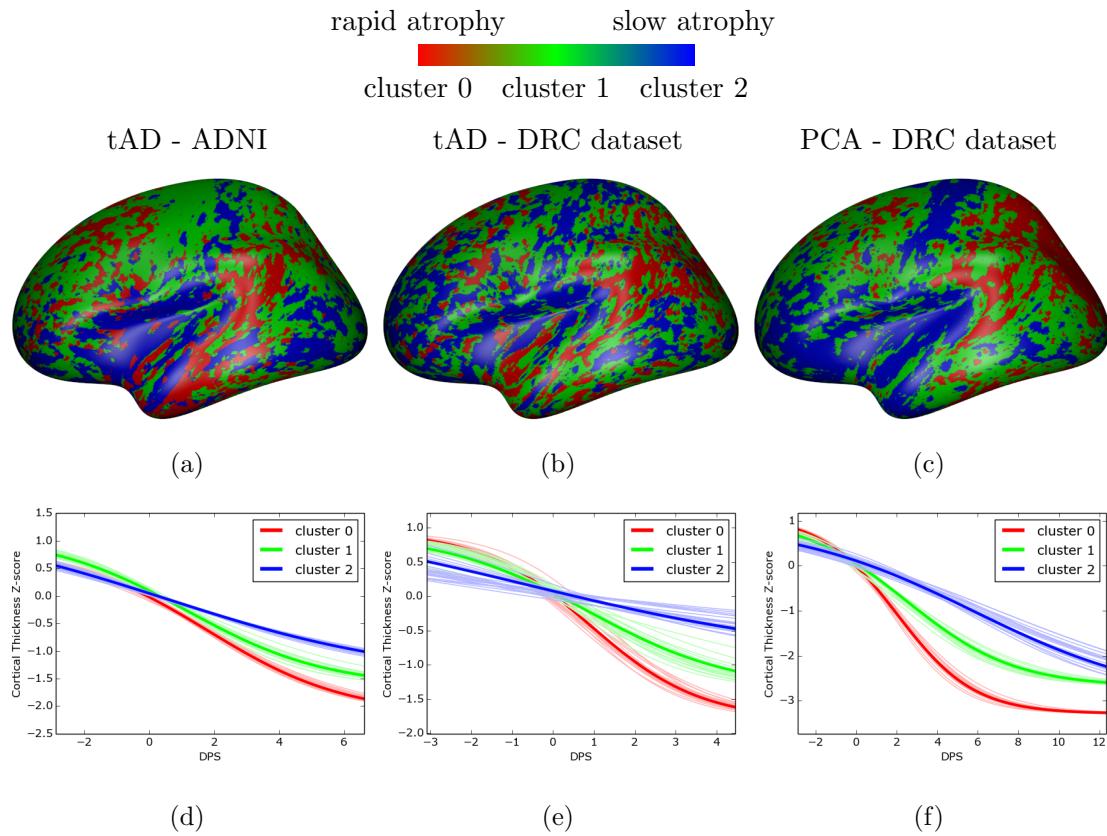


Figure 7.3: (a) Clustering results on the ADNI data using our model, where each cluster is coloured according to the slope of its corresponding trajectory, from red (high slope suggesting very affected areas) to blue (low slope suggesting less affected areas). (d) The corresponding trajectories and samples 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.

7.4.3 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 7.4 shows the clusters that were estimated at each fold from the training data only. Moreover, in Fig. 7.5 we plot the estimated DPS (i.e. disease stage) of each subject from the test set against their age.

The results in Fig. 7.4 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.

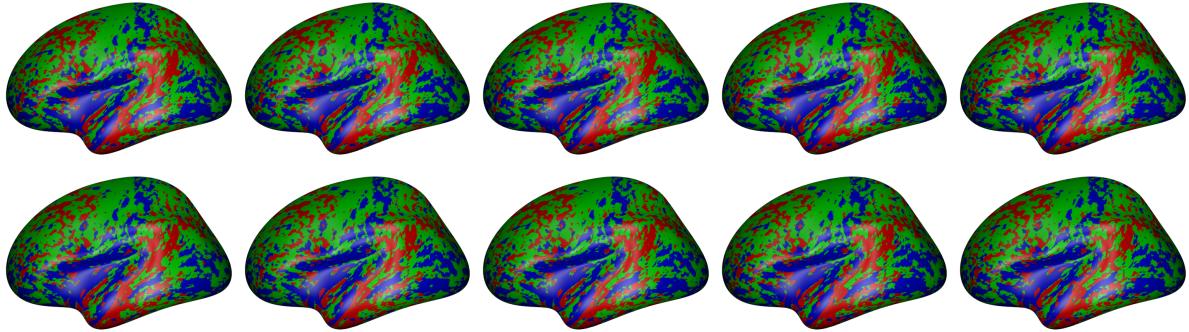


Figure 7.4: Clusters estimated for each of the 10 cross-validation folds in ADNI. As before, each cluster is coloured according to the slope of its corresponding trajectory, from red (high rate of atrophy) to blue (low rate of atrophy).

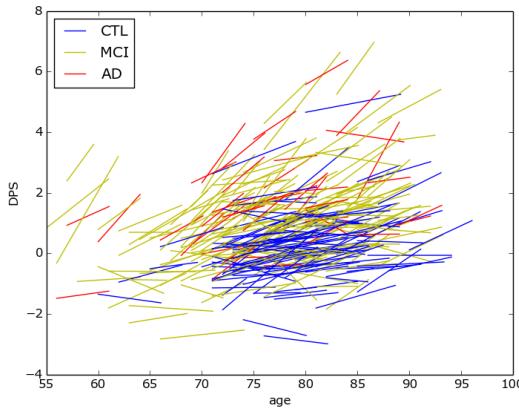


Figure 7.5: The disease progression score for each subject from the ADNI dataset estimated during 10-fold cross validation. Each line represents an individual i with different visits j . Later visits generally have a higher corresponding stage.

7.5 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 have applied it to cortical thickness vertex-wise data from ADNI and DRC cohorts. Our model found similar patterns of atrophy dynamics in the tAD subjects using two independent datasets. Moreover, it also found different patterns of atrophy dynamics on two distinct diseases: typical AD and PCA.

The model has some limitations. First of all, we have 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 the DRC cohort. Furthermore, the data we analysed has been standardised with respect to controls, which can introduce biases due to the fact that controls might already show some abnormalities in cortical thickness.

There are several potential avenues of future research. While we have 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 or self-modelling regression techniques [15]. One could also model different progression dynamics for distinct groups in the population by clustering subjects together like the approach of [112]. The analysis could also be improved by using only amyloid negative controls for data standardisation.

Our approach can be used for accurately predicting and staging neurodegenerative disease progression. 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.

Chapter 8

Project Plan

In this section we present our future work plan until the end of the PhD. In section 8.1 we suggest further modelling that can be done using the voxelwise temporal clustering model presented in chapter 7. In section 8.2 we present further analysis that should be done on performance evaluation of disease progression models 8.2. In section 8.3 we present further applications for our disease progression models, more precisely on the PCA vs tAD comparison. Finally, in Fig. 8.1 we present a Gantt chart showing the proposed work plan.

8.1 Voxelwise Temporal Clustering Model

We plan to extend the work done on the voxelwise temporal clustering in several ways and do further analysis. These are summarised in the list below:

1. Must do:
 - (a) Run model on amyloid PET imaging from ADNI. This would complement the work currently done using cortical thickness and demonstrate the different types of data which can be analysed by the model.
 - (b) Perform more validation on ADNI and DRC datasets. This can include: correlation of stages with cognitive measures, more staging consistency metrics and other goodness of fit measures.
 - (c) Test a similar model that fits trajectories according to the mean value of vertex measurements in every cluster. We believe that this model gives very similar results but it can run much faster, as the sum over every vertex in equation 7.6 will disappear.
2. Should do:
 - (a) Extensive testing on synthetic data. This can include experiments for many different types of trajectories and number of clusters. For each experiment we plan to plot error-bars showing the error on subject staging and recovered trajectory parameters.
3. Might do:
 - (a) Extend model to non-parametric curves such as Gaussian Processes (if time permits). This would involve changing the fundamentals of the model and can be difficult to fine-tune the parameters for the model on each dataset, otherwise the model can easily overfit.
 - (b) Extend the model to analyse multimodal data. In order to construct an accurate temporal axis of the disease progression, it can additionally use cognitive tests, CSF biomarkers or combine different voxelwise imaging modalities (cortical thickness +

amyloid images). Combining multiple biomarkers is important, as they become abnormal at different stages of the disease.

- (c) Explore other types of voxel dependence, such as spatial correlation functions.
- (d) Extend the model to account fit subgroups of subjects with different clusters of voxels. This will account for disease heterogeneity.

8.2 Performance Evaluation

We plan to extend the work on performance evaluation in order to turn this into a comprehensive evaluation of many different metrics (section 8.2.1) and disease progression models (section 8.2.2). Based on this work, we also consider getting involved in a prediction challenge using the ADNI dataset. This challenge will involve training some models on current ADNI data and using a subsequent release of ADNI data as the testing set.

8.2.1 Metrics

We plan to further evaluate the following performance metrics or scenarios:

1. Must do:
 - (a) Correlation of stages with cognitive measures
 - (b) Goodness of fit: AIC, BIC
 - (c) Synthetic data - error in estimating the true parameters
2. Should do:
 - (a) Resampling methods: cross-validation, bootstrapping (measure std of estimated parameters)
 - (b) Robustness to added noise levels in real data
3. Might do:
 - (a) Experiment design metrics: these will test which biomarkers are the most informative. This can help make the clinical studies more economical by removing scans or cognitive tests that do not bring in new information.

8.2.2 Models

We plan to incorporate other fundamentally different models such as:

1. Must do:
 - (a) Disease Progression Score (DPS) (Jedynak et al. [62])
 - (b) Self-Modelling Regression (SEMOR) (Donohue et al. [15])
2. Might do:
 - (a) Riemannian manifold model (Schiratti et al. [68])

Testing other models such as these will help us understand if the assumptions they make are reasonable for the ADNI biomarkers. Moreover, based on the results we will also be able to point out improvements in these models.

8.2.3 Dataset

Our plan is to do the performance analysis only on the ADNI dataset. We will carefully choose a set of biomarkers derived from MRI (ROI volumes, cortical thickness, atrophy rates), CSF (amyloid-beta, tau), PET (FDG, PIB) and cognitive tests. We need to be careful so that the choice of these biomarkers will not favour certain models.

8.3 PCA vs tAD Analysis

We plan to continue the application of our disease progression models to PCA and tAD analysis on the DRC dataset. We suggest to perform the following analysis¹:

1. Must do:

- (a) Study rates of atrophy for different PCA subtypes. This analysis will be similar to the one in section 5.3, but will study atrophy rates instead of timing. This can be done with a simple linear mixed effects model or, if there is enough data, with the differential equation model. We would like to test the following hypotheses:
 - i. Vision subgroup: greater occipital lobe atrophy compared to other subgroups?
 - ii. Space subgroup: greater superior parietal atrophy compared to other subgroups?
 - iii. Object subgroup: greater inferior temporal atrophy compared to other subgroups?
- (b) Study the continuum of the three subgroups: build EBM with each subgroup separately, and then try to assign subjects to different models. Plot the normalised likelihoods of belonging to each model on a triangle.

2. Should do:

- (a) Predict progression of cognitive tests from imaging biomarkers at baseline. This task would comprise generating cross-sectional occipital-hippocampal volume discrepancy values at baseline and then test whether different subgroups on this occipital-hippocampal continuum predict different longitudinal patterns of cognitive impairment. More precisely, we hypothesise that patients that have a lower occipital volume at baseline (PCA-like) will have early visual loss with memory spared until later. On the other hand, patients who have a lower hippocampal volume at baseline (tAD-like) will have earlier multi-domain cognitive impairment.

3. Might do:

- (a) Repeat the same analysis as in 1 (a) but for cortical thickness measures.
- (b) Study the relationship between posterior and anterior grey matter loss.
 - i. Based on anatomical subgroups, test if greater inferior posterior atrophy predict greater inferior anterior atrophy and vice versa.
 - ii. Test if dorsolateral prefrontal lobe atrophy and cortical thinning differ significantly between the three clinical subgroups as follows: (Highest atrophy) Space > Object > Vision (Lowest atrophy).
 - iii. Test if inferior prefrontal atrophy and cortical thinning differ significantly between the three subgroups as follows: (Highest atrophy) Object > Space > Vision (Lowest atrophy).
- (c) Asymmetry analyses:

¹ideas proposed by Dr. Sebastian Crutch

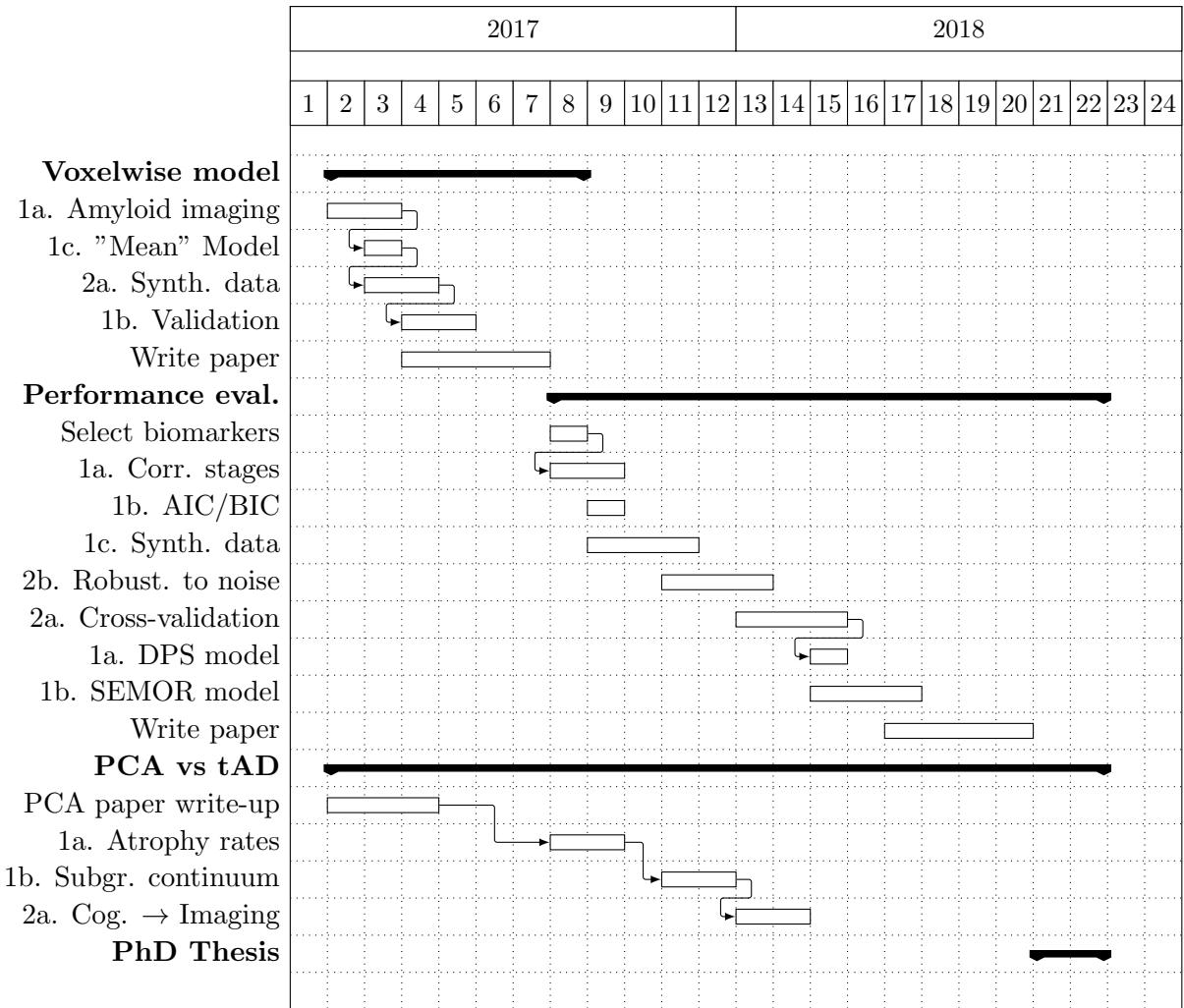


Figure 8.1: Gantt chart showing the work schedule for the PhD duration.

- i. Measure timing of atrophy: Run the EBM on PCA and tAD using separate biomarkers for left and right hemisphere regions.
- ii. Measure rate and extent of atrophy: Run the DEM as above. In addition to that, also plot each left vs right ROI trajectories and measure if there is any statistically significant different in timing, rate and amount of atrophy between regions in the left and right hemisphere.

Chapter 9

Bibliography

- [1] Alistair Burns and Steve Iliffe. Alzheimer's disease. *BMJ*, 338:467–471 ST – Alzheimer's disease, 2009.
- [2] World Health Organization (WHO) et al. Dementia fact sheet n. 362 2012.
- [3] Roger Bullock. Future directions in the treatment of alzheimer's disease. *Expert opinion on investigational drugs*, 13(4):303–314, 2004.
- [4] Hans-Wolfgang Klafki, Matthias Staufenbiel, Johannes Kornhuber, and Jens Wiltfang. Therapeutic approaches to alzheimer's disease. *Brain*, 129(11):2840–2855, 2006.
- [5] Jeffrey L. Cummings. Alzheimers disease. *New England Journal of Medicine*, 351(1):56–67, 2004. PMID: 15229308.
- [6] Elizabeth Forsyth and Pamela D Ritzline. An overview of the etiology, diagnosis, and treatment of alzheimer disease. *Physical therapy*, 78(12):1325–1331, 1998.
- [7] D Frank Benson, R Jeffrey Davis, and Bruce D Snyder. Posterior cortical atrophy. *Archives of neurology*, 45(7):789–793, 1988.
- [8] Sebastian J Crutch, Manja Lehmann, Jonathan M Schott, Gil D Rabinovici, Martin N Rossor, and Nick C Fox. Posterior cortical atrophy. *The Lancet Neurology*, 11(2):170–178, 2012.
- [9] François-Xavier Borruat. Posterior cortical atrophy: review of the recent literature. *Current neurology and neuroscience reports*, 13(12):1–8, 2013.
- [10] Clifford R Jack, David S Knopman, William J Jagust, Leslie M Shaw, Paul S Aisen, Michael W Weiner, Ronald C Petersen, and John Q Trojanowski. Hypothetical model of dynamic biomarkers of the alzheimer's pathological cascade. *The Lancet Neurology*, 9(1):119–128, 2010.
- [11] Randall J Bateman, Chengjie Xiong, Tammie LS Benzinger, Anne M Fagan, Alison Goate, Nick C Fox, Daniel S Marcus, Nigel J Cairns, Xianyun Xie, Tyler M Blazey, et al. Clinical and biomarker changes in dominantly inherited alzheimer's disease. *New England Journal of Medicine*, 367(9):795–804, 2012.
- [12] Alexander Schmidt-Richberg, Ricardo Guerrero, Christian Ledig, Helena Molina-Abril, Alejandro F Frangi, Daniel Rueckert, Alzheimers Disease Neuroimaging Initiative, et al. Multi-stage biomarker models for progression estimation in alzheimers disease. In *International Conference on Information Processing in Medical Imaging*, pages 387–398. Springer, 2015.

- [13] Hubert M Fonteijn, Marc Modat, Matthew J Clarkson, Josephine Barnes, Manja Lehmann, Nicola Z Hobbs, Rachael I Scahill, Sarah J Tabrizi, Sébastien Ourselin, Nick C Fox, et al. An event-based model for disease progression and its application in familial alzheimer’s disease and huntington’s disease. *NeuroImage*, 60(3):1880–1889, 2012.
- [14] Bruno M. Jedynak, Andrew Lang, Bo Liu, Elyse Katz, Yanwei Zhang, Bradley T. Wyman, David Raunig, C. Pierre Jedynak, Brian Caffo, and Jerry L. Prince. A computational neurodegenerative disease progression score: Method and results with the Alzheimer’s disease neuroimaging initiative cohort. *NeuroImage*, 63(3):1478–1486, 2012.
- [15] Michael C Donohue, Hélène Jacqmin-Gadda, Mélanie Le Goff, Ronald G Thomas, Rema Raman, Anthony C Gamst, Laurel A Beckett, Clifford R Jack, Michael W Weiner, Jean-François Dartigues, et al. Estimating long-term multivariate progression from short-term data. *Alzheimer’s & Dementia*, 10(5):S400–S410, 2014.
- [16] J-B Schiratti, Stéphanie Allassonniere, Alexandre Routier, Olivier Colliot, Stanley Durleman, Alzheimers Disease Neuroimaging Initiative, et al. A mixed-effects model with time reparametrization for longitudinal univariate manifold-valued data. In *International Conference on Information Processing in Medical Imaging*, pages 564–575. Springer, 2015.
- [17] Henry W Querfurth and Frank M LaFerla. Mechanisms of disease. *N Engl J Med*, 362(4):329–344, 2010.
- [18] Stephen Todd, Stephen Barr, Mark Roberts, and A Peter Passmore. Survival in dementia and predictors of mortality: a review. *International journal of geriatric psychiatry*, 28(11):1109–1124, 2013.
- [19] Clive Ballard, Serge Gauthier, Anne Corbett, Carol Brayne, Dag Aarsland, and Emma Jones. Alzheimer’s disease. *The Lancet*, 377(9770):1019–1031, 2011.
- [20] H Braak and E Braak. Neuropathological stageing of alzheimer-related changes. *Acta neuropathologica*, 82(4):239–259, 1991.
- [21] John Hardy and David Allsop. Amyloid deposition as the central event in the aetiology of alzheimer’s disease. *Trends in pharmacological sciences*, 12:383–388, 1991.
- [22] Amritpal Mudher and Simon Lovestone. Alzheimer’s disease—do tauists and baptists finally shake hands? *Trends in neurosciences*, 25(1):22–26, 2002.
- [23] George T Grossberg and Abhilash K Desai. Management of alzheimer’s disease. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 58(4):M331–M353, 2003.
- [24] Kaj Blennow and Harald Hampel. Csf markers for incipient alzheimer’s disease. *The Lancet Neurology*, 2(10):605–613, 2003.
- [25] William E Klunk, Henry Engler, Agneta Nordberg, Yanming Wang, Gunnar Blomqvist, Daniel P Holt, Mats Bergström, Irina Savitcheva, Guo-Feng Huang, Sergio Estrada, et al. Imaging brain amyloid in alzheimer’s disease with pittsburgh compound-b. *Annals of neurology*, 55(3):306–319, 2004.
- [26] Karl Herholz. Use of fdg pet as an imaging biomarker in clinical trials of alzheimer’s disease. *Biomarkers in medicine*, 6(4):431–439, 2012.
- [27] Peter Freeborough, Nick C Fox, et al. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat mri. *Medical Imaging, IEEE Transactions on*, 16(5):623–629, 1997.

- [28] Guy McKhann, David Drachman, Marshall Folstein, Robert Katzman, Donald Price, and Emanuel M Stadlan. Clinical diagnosis of alzheimer's disease report of the nincds-adrda work group* under the auspices of department of health and human services task force on alzheimer's disease. *Neurology*, 34(7):939–939, 1984.
- [29] Basil H Ridha, Josephine Barnes, Jonathan W Bartlett, Alison Godbolt, Tracey Pepple, Martin N Rossor, and Nick C Fox. Tracking atrophy progression in familial alzheimer's disease: a serial mri study. *The Lancet Neurology*, 5(10):828–834, 2006.
- [30] NC Fox, RI Scahill, WR Crum, and MN Rossor. Correlation between rates of brain atrophy and cognitive decline in ad. *Neurology*, 52(8):1687–1687, 1999.
- [31] Rachael I Scahill, Jonathan M Schott, John M Stevens, Martin N Rossor, and Nick C Fox. Mapping the evolution of regional atrophy in alzheimer's disease: unbiased analysis of fluid-registered serial mri. *Proceedings of the National Academy of Sciences*, 99(7):4703–4707, 2002.
- [32] Jonathan M Schott, Nick C Fox, Chris Frost, Rachael I Scahill, John C Janssen, Dennis Chan, Rhian Jenkins, and Martin N Rossor. Assessing the onset of structural change in familial alzheimer's disease. *Annals of neurology*, 53(2):181–188, 2003.
- [33] Clifford R Jack, David S Knopman, William J Jagust, Ronald C Petersen, Michael W Weiner, Paul S Aisen, Leslie M Shaw, Prashanthi Vemuri, Heather J Wiste, Stephen D Weigand, et al. Tracking pathophysiological processes in alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *The Lancet Neurology*, 12(2):207–216, 2013.
- [34] Clifford R Jack, Ronald C Petersen, Yue Cheng Xu, Stephen C Waring, Peter C O'Brien, Eric G Tangalos, Glenn E Smith, Robert J Ivnik, and Emre Kokmen. Medial temporal atrophy on mri in normal aging and very mild alzheimer's disease. *Neurology*, 49(3):786–794, 1997.
- [35] Stephane Lehericy, Michel Baulac, Jacques Chiras, Laurent Pierot, Nadine Martin, Bernard Pillon, Bernard Deweer, Bruno Dubois, and Claude Marsault. Amygdalohippocampal mr volume measurements in the early stages of alzheimer disease. *American Journal of Neuroradiology*, 15(5):929–937, 1994.
- [36] Kirsi Juottonen, Mikko P Laakso, Kaarina Partanen, and Hilkka Soininen. Comparative mr analysis of the entorhinal cortex and hippocampus in diagnosing alzheimer disease. *American Journal of Neuroradiology*, 20(1):139–144, 1999.
- [37] M Bobinski, MJ De Leon, J Wegiel, S Desanti, A Convit, LA Saint Louis, H Rusinek, and HM Wisniewski. The histological validation of post mortem magnetic resonance imaging-determined hippocampal volume in alzheimer's disease. *Neuroscience*, 95(3):721–725, 1999.
- [38] Maarten Goethals and Patrick Santens. Posterior cortical atrophy. two case reports and a review of the literature. *Clinical neurology and neurosurgery*, 103(2):115–119, 2001.
- [39] Katia Andrade, Dalila Samri, Marie Sarazin, Leonardo C de Souza, Laurent Cohen, Michel T de Schotten, Bruno Dubois, and Paolo Bartolomeo. Visual neglect in posterior cortical atrophy. *BMC neurology*, 10(1):68, 2010.
- [40] Katia Andrade, Aurélie Kas, Romain Valabregue, Dalila Samri, Marie Sarazin, Marie-Odile Habert, Bruno Dubois, and Paolo Bartolomeo. Visuospatial deficits in posterior cortical atrophy: structural and functional correlates. *Journal of Neurology, Neurosurgery & Psychiatry*, pages jnnp–2012, 2012.

- [41] Katia Andrade, Aurélie Kas, Dalila Samri, Marie Sarazin, Bruno Dubois, Marie-Odile Habert, and Paolo Bartolomeo. Visuospatial deficits and hemispheric perfusion asymmetries in posterior cortical atrophy. *Cortex*, 49(4):940–947, 2013.
- [42] Martin A Goldstein, Iliyan Ivanov, and Michael E Silverman. Posterior cortical atrophy: an exemplar for renovating diagnostic formulation in neuropsychiatry. *Comprehensive psychiatry*, 52(3):326–333, 2011.
- [43] Mario F Mendez, Mehdi Ghajarania, and Kent M Perryman. Posterior cortical atrophy: clinical characteristics and differences compared to alzheimers disease. *Dementia and geriatric cognitive disorders*, 14(1):33–40, 2002.
- [44] David F Tang-Wai, NR Graff-Radford, BF Boeve, DW Dickson, JE Parisi, R Crook, RJ Caselli, DS Knopman, and RC Petersen. Clinical, genetic, and neuropathologic characteristics of posterior cortical atrophy. *Neurology*, 63(7):1168–1174, 2004.
- [45] MH Rosenbloom, A Alkalay, N Agarwal, SL Baker, JP O’Neil, M Janabi, IV Yen, M Growdon, J Jang, C Madison, et al. Distinct clinical and metabolic deficits in pca and ad are not related to amyloid distribution. *Neurology*, 76(21):1789–1796, 2011.
- [46] Hélène Videaud, Frédéric Torny, Leslie Cartz-Piver, N Deschamps-Vergara, and Philippe Couratier. [impact of drug-free care in posterior cortical atrophy: Preliminary experience with a psycho-educative program]. *Revue neurologique*, 168(11):861–867, 2012.
- [47] A Weill-Choulamountry, F Poncet, S Crop, N Hesly, A Mouton, D Samri, M Sarazin, and P Pradat-Diehl. Physical medicine and rehabilitation multidisciplinary approach in a case of posterior cortical atrophy. *Annals of physical and rehabilitation medicine*, 55(6):430–439, 2012.
- [48] Patrick R Hof, Brent A Vogt, Constantin Bouras, and John H Morrison. Atypical form of alzheimer’s disease with prominent posterior cortical atrophy: a review of lesion distribution and circuit disconnection in cortical visual pathways. *Vision research*, 37(24):3609–3625, 1997.
- [49] Bradford C Dickerson, Akram Bakkour, David H Salat, Eric Feczkó, Jenni Pacheco, Douglas N Greve, Fran Grodstein, Christopher I Wright, Deborah Blacker, H Diana Rosas, et al. 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):497–510, 2009.
- [50] Paul M Thompson, Michael S Mega, Roger P Woods, Chris I Zoumalan, Chris J Lindshield, Rebecca E Blanton, Jacob Moussai, Colin J Holmes, Jeffrey L Cummings, and Arthur W Toga. Cortical change in alzheimer’s disease detected with a disease-specific population-based brain atlas. *Cerebral Cortex*, 11(1):1–16, 2001.
- [51] Paul M Thompson, Kiralee M Hayashi, Greig De Zubicaray, Andrew L Janke, Stephen E Rose, James Semple, David Herman, Michael S Hong, Stephanie S Dittmer, David M Doddrell, et al. Dynamics of gray matter loss in alzheimer’s disease. *The Journal of Neuroscience*, 23(3):994–1005, 2003.
- [52] Alexandra L Young, Neil P Oxtoby, Pankaj Daga, David M Cash, Nick C Fox, Sébastien Ourselin, Jonathan M Schott, and Daniel C Alexander. A data-driven model of biomarker changes in sporadic alzheimer’s disease. *Brain*, 137(9):2564–2577, 2014.

- [53] Christopher H Jackson, Linda D Sharples, Simon G Thompson, Stephen W Duffy, and Elisabeth Couto. Multistate markov models for disease progression with classification error. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(2):193–209, 2003.
- [54] Chantal Guienneuc-Jouyaux, Sylvia Richardson, and Ira M Longini. Modeling markers of disease progression by a hidden markov process: application to characterizing cd4 cell decline. *Biometrics*, 56(3):733–741, 2000.
- [55] Victor L Villemagne, Samantha Burnham, Pierrick Bourgeat, Belinda Brown, Kathryn A Ellis, Olivier Salvado, Cassandra Szoek, S Lance Macaulay, Ralph Martins, Paul Maruff, et al. Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic alzheimer’s disease: a prospective cohort study. *The Lancet Neurology*, 12(4):357–367, 2013.
- [56] Neil P Oxtoby, Alexandra L Young, Nick C Fox, Pankaj Daga, David M Cash, Sebastien Ourselin, Jonathan M Schott, Daniel C Alexander, Alzheimers Disease Neuroimaging Initiative, et al. Learning imaging biomarker trajectories from noisy alzheimers disease data using a bayesian multilevel model. In *Bayesian and grAphical Models for Biomedical Imaging*, pages 85–94. Springer, 2014.
- [57] J Wesson Ashford and Frederick A Schmitt. Modeling the time-course of alzheimer dementia. *Current psychiatry reports*, 3(1):20–28, 2001.
- [58] Eric Yang, Michael Farnum, Victor Lobanov, Tim Schultz, R Verbeeck, N Raghavan, MN Samtani, G Novak, V Narayan, and A DiBernardo. Quantifying the pathophysiological timeline of alzheimer’s disease. *Journal of Alzheimer’s disease: JAD*, 26(4):745–753, 2010.
- [59] Mert R Sabuncu, Rahul S Desikan, Jorge Sepulcre, Boon Thye T Yeo, Hesheng Liu, Nicholas J Schmansky, Martin Reuter, Michael W Weiner, Randy L Buckner, Reisa A Sperling, et al. The dynamics of cortical and hippocampal atrophy in alzheimer disease. *Archives of neurology*, 68(8):1040–1048, 2011.
- [60] Clifford R Jack, Heather J Wiste, Timothy G Lesnick, Stephen D Weigand, David S Knopman, Prashanthi Vemuri, Vernon S Pankratz, Matthew L Senjem, Jeffrey L Gunter, Michelle M Mielke, et al. Brain β -amyloid load approaches a plateau. *Neurology*, 80(10):890–896, 2013.
- [61] Alexandra L Young, Neil P Oxtoby, Jonathan M Schott, and Daniel C Alexander. Data-driven models of neurodegenerative disease — ACNR — Online Neurology Journal.
- [62] Bruno M Jedynak, Andrew Lang, Bo Liu, Elyse Katz, Yanwei Zhang, Bradley T Wyman, David Raunig, C Pierre Jedynak, Brian Caffo, Jerry L Prince, et al. A computational neurodegenerative disease progression score: method and results with the alzheimer’s disease neuroimaging initiative cohort. *Neuroimage*, 63(3):1478–1486, 2012.
- [63] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.
- [64] W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

- [65] Eric Yang, Michael Farnum, Victor Lobanov, Tim Schultz, Nandini Raghavan, Mahesh N Samtani, Gerald Novak, Vaibhav Narayan, and Allitia DiBernardo. Quantifying the pathophysiological timeline of alzheimer’s disease. *Journal of Alzheimer’s Disease*, 26(4):745–753, 2011.
- [66] A Caroli, GB Frisoni, Alzheimer’s Disease Neuroimaging Initiative, et al. The dynamics of alzheimer’s disease biomarkers in the alzheimer’s disease neuroimaging initiative cohort. *Neurobiology of aging*, 31(8):1263–1274, 2010.
- [67] Manasi Datar, Prasanna Muralidharan, Abhishek Kumar, Sylvain Gouttard, Joseph Piven, Guido Gerig, Ross Whitaker, and P Thomas Fletcher. Mixed-effects shape models for estimating longitudinal changes in anatomy. In *International Workshop on Spatio-temporal Image Analysis for Longitudinal and Time-Series Image Data*, pages 76–87. Springer, 2012.
- [68] Jean-Baptiste Schiratti, Stéphanie Allassonniere, Olivier Colliot, and Stanley Durrleman. Learning spatiotemporal trajectories from manifold-valued longitudinal data. In *Advances in Neural Information Processing Systems*, pages 2404–2412, 2015.
- [69] Murat Bilgel, Jerry L Prince, Dean F Wong, Susan M Resnick, and Bruno M Jedynak. A multivariate nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging. *NeuroImage*, 134:658–670, 2016.
- [70] Ashish Raj, Amy Kuceyeski, and Michael Weiner. A network diffusion model of disease progression in dementia. *Neuron*, 73(6):1204–1215, 2012.
- [71] Ashish Raj, Eve LoCastro, Amy Kuceyeski, Duygu Tosun, Norman Relkin, Michael Weiner, Alzheimers Disease Neuroimaging Initiative (ADNI, et al. Network diffusion model of progression predicts longitudinal patterns of atrophy and metabolism in alzheimers disease. *Cell reports*, 10(3):359–369, 2015.
- [72] Nicolas Villain, Marine Fouquet, Jean-Claude Baron, Florence Mézenge, Brigitte Landeau, Vincent de La Sayette, Fausto Viader, Francis Eustache, Béatrice Desgranges, and Gaël Chételat. Sequential relationships between grey matter and white matter atrophy and brain metabolic abnormalities in early alzheimer’s disease. *Brain*, 133(11):3301–3314, 2010.
- [73] E Englund, A Brun, and C Alling. White matter changes in dementia of alzheimer’s type. *Brain*, 111(6):1425–1439, 1988.
- [74] Beth Kuczynski, Elizabeth Targan, Cindee Madison, Michael Weiner, Yu Zhang, Bruce Reed, Helena C Chui, and William Jagust. White matter integrity and cortical metabolic associations in aging and dementia. *Alzheimer’s & dementia*, 6(1):54–62, 2010.
- [75] TEJ Behrens, H Johansen Berg, Saad Jbabdi, MFS Rushworth, and MW Woolrich. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *Neuroimage*, 34(1):144–155, 2007.
- [76] Risi Imre Kondor and John Lafferty. Diffusion kernels on graphs and other discrete input spaces. In *ICML*, volume 2, pages 315–322, 2002.
- [77] William W Seeley, Richard K Crawford, Juan Zhou, Bruce L Miller, and Michael D Greicius. Neurodegenerative diseases target large-scale human brain networks. *Neuron*, 62(1):42–52, 2009.

- [78] Juan Zhou, Michael D Greicius, Efstathios D Gennatas, Matthew E Growdon, Jung Y Jang, Gil D Rabinovici, Joel H Kramer, Michael Weiner, Bruce L Miller, and William W Seeley. Divergent network connectivity changes in behavioural variant frontotemporal dementia and alzheimers disease. *Brain*, 133(5):1352–1367, 2010.
- [79] Juan Zhou, Efstathios D Gennatas, Joel H Kramer, Bruce L Miller, and William W Seeley. Predicting regional neurodegeneration from the healthy brain functional connectome. *Neuron*, 73(6):1216–1227, 2012.
- [80] Vladimir Vapnik. *Estimation of dependences based on empirical data*. Springer Science & Business Media, 2006.
- [81] Stefan Klöppel, Cynthia M Stonnington, Carlton Chu, Bogdan Draganski, Rachael I Scahill, Jonathan D Rohrer, Nick C Fox, Clifford R Jack, John Ashburner, and Richard SJ Frackowiak. Automatic classification of mr scans in alzheimer’s disease. *Brain*, 131(3):681–689, 2008.
- [82] Zhiqiang Lao, Dinggang Shen, Zhong Xue, Bilge Karacali, Susan M Resnick, and Christos Davatzikos. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *Neuroimage*, 21(1):46–57, 2004.
- [83] Yong Fan, Dinggang Shen, and Christos Davatzikos. Classification of structural images via high-dimensional image warping, robust feature extraction, and svm. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 1–8. Springer, 2005.
- [84] Janaina Mourão-Miranda, Arun LW Bokde, Christine Born, Harald Hampel, and Martin Stetter. Classifying brain states and determining the discriminating activation patterns: support vector machine on functional mri data. *NeuroImage*, 28(4):980–995, 2005.
- [85] Yasuhiro Kawasaki, Michio Suzuki, Ferath Kherif, Tsutomu Takahashi, Shi-Yu Zhou, Kazue Nakamura, Mie Matsui, Tomiki Sumiyoshi, Hikaru Seto, and Masayoshi Kurachi. Multivariate voxel-based morphometry successfully differentiates schizophrenia patients from healthy controls. *Neuroimage*, 34(1):235–242, 2007.
- [86] Tin Kam Ho. Random decision forests. In *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, volume 1, pages 278–282. IEEE, 1995.
- [87] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [88] Katherine R Gray, Paul Aljabar, Rolf A Heckemann, Alexander Hammers, Daniel Rueckert, Alzheimer’s Disease Neuroimaging Initiative, et al. Random forest-based similarity measures for multi-modal classification of alzheimer’s disease. *NeuroImage*, 65:167–175, 2013.
- [89] Daniel C Alexander, Darko Zikic, Jiaying Zhang, Hui Zhang, and Antonio Criminisi. Image quality transfer via random forest regression: applications in diffusion mri. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 225–232. Springer, 2014.
- [90] Victor Lempitsky, Michael Verhoek, J Alison Noble, and Andrew Blake. Random forest classification for automatic delineation of myocardium in real-time 3d echocardiography. In *International Conference on Functional Imaging and Modeling of the Heart*, pages 447–456. Springer, 2009.

- [91] Clifford R Jack, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L Whitwell, Chadwick Ward, et al. The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging*, 27(4):685–691, 2008.
- [92] Wilma G Rosen, Richard C Mohs, and Kenneth L Davis. A new rating scale for alzheimer's disease. *The American journal of psychiatry*, 1984.
- [93] André Rey. L'examen clinique en psychologie. 1958.
- [94] Kelvin K Leung, Matthew J Clarkson, Jonathan W Bartlett, Shona Clegg, Clifford R Jack, Michael W Weiner, Nick C Fox, Sébastien Ourselin, Alzheimer's Disease Neuroimaging Initiative, et al. Robust atrophy rate measurement in alzheimer's disease using multi-site serial mri: tissue-specific intensity normalization and parameter selection. *Neuroimage*, 50(2):516–523, 2010.
- [95] Kelvin K Leung, Josephine Barnes, Gerard R Ridgway, Jonathan W Bartlett, Matthew J Clarkson, Kate Macdonald, Norbert Schuff, Nick C Fox, Sébastien Ourselin, Alzheimer's Disease Neuroimaging Initiative, et al. Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and alzheimer's disease. *Neuroimage*, 51(4):1345–1359, 2010.
- [96] M Jorge Cardoso, Marc Modat, Robin Wolz, Andrew Melbourn, David Cash, Daniel Rueckert, and Sébastien Ourselin. Geodesic information flows: Spatially-variant graphs and their application to segmentation and fusion. 2015.
- [97] Alexandra L Young, Neil P Oxtoby, Sébastien Ourselin, Jonathan M Schott, Daniel C Alexander, Alzheimer's Disease Neuroimaging Initiative, et al. A simulation system for biomarker evolution in neurodegenerative disease. *Medical image analysis*, 26(1):47–56, 2015.
- [98] Rafid Sukkar, Elyse Katz, Yanwei Zhang, David Raunig, and Bradley T Wyman. Disease progression modeling using hidden markov models. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2845–2848. IEEE, 2012.
- [99] Manja Lehmann, Sebastian J Crutch, Gerard R Ridgway, Basil H Ridha, Josephine Barnes, Elizabeth K Warrington, Martin N Rossor, and Nick C Fox. Cortical thickness and voxel-based morphometry in posterior cortical atrophy and typical alzheimer's disease. *Neurobiology of aging*, 32(8):1466–1476, 2011.
- [100] Jennifer L Whitwell, Clifford R Jack, Kejal Kantarci, Stephen D Weigand, Bradley F Boeve, David S Knopman, Daniel A Drubach, David F Tang-Wai, Ronald C Petersen, and Keith A Josephs. Imaging correlates of posterior cortical atrophy. *Neurobiology of aging*, 28(7):1051–1061, 2007.
- [101] Antonio Convit, MonyJ de Leon, Chaim Tarshish, Susan De Santi, Alan Kluger, Henry Rusinek, and AjaxE George. Hippocampal volume losses in minimally impaired elderly. *The Lancet*, 345(8944):266, 1995.
- [102] Y Xu, CR Jack, PC Obrien, E Kokmen, Glenn E Smith, Robert J Ivnik, Bradley F Boeve, RG Tangalos, and Ronald Carl Petersen. Usefulness of mri measures of entorhinal cortex versus hippocampus in ad. *Neurology*, 54(9):1760–1767, 2000.
- [103] Manja Lehmann, Josephine Barnes, Gerard R Ridgway, John Wattam-Bell, Elizabeth K Warrington, Nick C Fox, and Sebastian J Crutch. Basic visual function and cortical thickness patterns in posterior cortical atrophy. *Cerebral cortex*, 21(9):2122–2132, 2011.

- [104] SJ Ross, Naida Graham, Lindsay Stuart-Green, Miriam Prins, John Xuereb, Karalyn Patterson, and John R Hodges. Progressive biparietal atrophy: an atypical presentation of alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 61(4):388–395, 1996.
- [105] Clare J Galton, Karalyn Patterson, John H Xuereb, and John R Hodges. Atypical and typical presentations of alzheimer's disease: a clinical, neuropsychological, neuroimaging and pathological study of 13 cases. *Brain*, 123(3):484–498, 2000.
- [106] Motohiro Kiyosawa, Thomas M Bosley, John Chawluk, Dara Jamieson, Norman J Schatz, Peter J Savino, Robert C Sergott, Martin Reivich, and Abass Alavi. Alzheimer's disease with prominent visual symptoms: clinical and metabolic evaluation. *Ophthalmology*, 96(7):1077–1086, 1989.
- [107] Mikko P Laakso, Giovanni B Frisoni, Mervi Könönen, Mia Mikkonen, Alberto Beltramello, Claudia Geroldi, Angelo Bianchetti, Marco Trabucchi, Hilkka Soininen, and Hannu J Aronen. Hippocampus and entorhinal cortex in frontotemporal dementia and alzheimers disease: a morphometric mri study. *Biological psychiatry*, 47(12):1056–1063, 2000.
- [108] Heiko Braak and Eva Braak. Morphological criteria for the recognition of alzheimer's disease and the distribution pattern of cortical changes related to this disorder. *Neurobiology of aging*, 15(3):355–356, 1994.
- [109] Corina Pennanen, Miia Kivipelto, Susanna Tuomainen, Päivi Hartikainen, Tuomo Hänninen, Mikko P Laakso, Merja Hallikainen, Matti Vanhanen, Aulikki Nissinen, Eeva-Liisa Helkala, et al. Hippocampus and entorhinal cortex in mild cognitive impairment and early ad. *Neurobiology of aging*, 25(3):303–310, 2004.
- [110] Christopher M Bishop. Pattern recognition. *Machine Learning*, 128, 2006.
- [111] Martin Reuter, Nicholas J Schmansky, H Diana Rosas, and Bruce Fischl. Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage*, 61(4):1402–1418, 2012.
- [112] Alexandra L Young, Neil P Oxtoby, Jonathan Huang, Razvan V Marinescu, Pankaj Daga, David M Cash, Nick C Fox, Sebastien Ourselin, Jonathan M Schott, Daniel C Alexander, et al. Multiple orderings of events in disease progression. In *Information Processing in Medical Imaging*, pages 711–722. Springer, 2015.
- [113] C Bishop. Pattern recognition and machine learning (information science and statistics), 1st edn. 2006. corr. 2nd printing edn, 2007.
- [114] Stephen M Stigler. Francis galton's account of the invention of correlation. *Statistical Science*, 4(2):73–79, 1989.
- [115] Joseph Lee Rodgers and W Alan Nicewander. Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59–66, 1988.
- [116] Gabor J Szekely and Maria L Rizzo. Hierarchical clustering via joint between-within distances: Extending ward's minimum variance method. *Journal of classification*, 22(2):151–183, 2005.
- [117] Robert R Sokal. A statistical method for evaluating systematic relationships. *Univ Kans Sci Bull*, 38:1409–1438, 1958.

- [118] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.
- [119] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
- [120] D Luenberger and Y Ye. Linear and nonlinear optimization. *Linear and Nonlinear Optimization*, 1984.
- [121] A Aizerman, Emmanuel M Braverman, and LI Rozoner. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and remote control*, 25:821–837, 1964.
- [122] John Moody and Christian J Darken. Fast learning in networks of locally-tuned processing units. *Neural computation*, 1(2):281–294, 1989.
- [123] Vikas Dhikav and Kuljeet Anand. Potential predictors of hippocampal atrophy in alzheimers disease. *Drugs & aging*, 28(1):1–11, 2011.
- [124] Sebastian J Crutch, Jonathan M Schott, Gil D Rabinovici, Bradley F Boeve, Stefano F Cappa, Bradford C Dickerson, Bruno Dubois, Neill R Graff-Radford, Pierre Krolak-Salmon, Manja Lehmann, et al. Shining a light on posterior cortical atrophy. *Alzheimer's & Dementia*, 9(4):463–465, 2013.
- [125] Elizabeth K Warrington and Merle James. *The visual object and space perception battery*. Thames Valley Test Company Bury St Edmunds, 1991.
- [126] Elizabeth K Warrington, Tim Shallice, et al. Category specific semantic impairments. *Brain*, 107(3):829–853, 1984.
- [127] David Wechsler. Wechsler adult intelligence scale–fourth edition (waits–iv). *San Antonio, TX: NCS Pearson*, 2008.
- [128] John TE Richardson. Measures of short-term memory: a historical review. *Cortex*, 43(5):635–650, 2007.
- [129] Jonathan Young, Marc Modat, Manuel J Cardoso, Alex Mendelson, Dave Cash, Sebastien Ourselin, Alzheimer's Disease Neuroimaging Initiative, et al. Accurate multimodal probabilistic prediction of conversion to alzheimer's disease in patients with mild cognitive impairment. *NeuroImage: Clinical*, 2:735–745, 2013.
- [130] Nick C Fox and Peter A Freeborough. Brain atrophy progression measured from registered serial mri: validation and application to alzheimer's disease. *Journal of Magnetic Resonance Imaging*, 7(6):1069–1075, 1997.
- [131] PJ Nestor, D Caine, TD Fryer, J Clarke, and JR Hodges. The topography of metabolic deficits in posterior cortical atrophy (the visual variant of alzheimers disease) with fdg-pet. *Journal of Neurology, Neurosurgery & Psychiatry*, 74(11):1521–1529, 2003.
- [132] Klaus Schmidtke, PD Dr Michael Hüll, and Jochen Talazko. Posterior cortical atrophy: variant of alzheimers disease? *Journal of neurology*, 252(1):27–35, 2005.
- [133] Eunhee Kim, Yunsoo Lee, Jongkeol Lee, and Seol-Heui Han. A case with cholinesterase inhibitor responsive asymmetric posterior cortical atrophy. *Clinical neurology and neurosurgery*, 108(1):97–101, 2005.

- [134] M Wakai, H Honda, A Takahashi, T Kato, K Ito, and T Hamanaka. Unusual findings on pet study of a patient with posterior cortical atrophy. *Acta neurologica scandinavica*, 89(6):458–461, 1994.
- [135] João Maroco, Dina Silva, Ana Rodrigues, Manuela Guerreiro, Isabel Santana, and Alexandre de Mendonça. Data mining methods in the prediction of dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC research notes*, 4(1):299, 2011.

Appendix A

A.1 ADNI extra figures

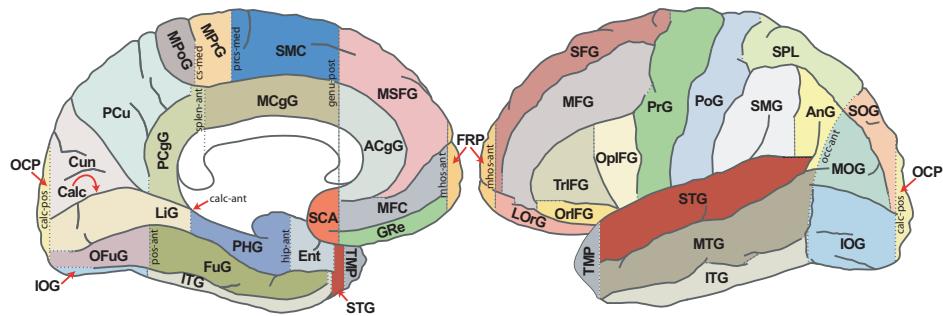


Figure A.1: Labels of the different areas analysed in the EBM progression snapshots from chapter 5.

A.2 Performance evaluation - Synthetic results

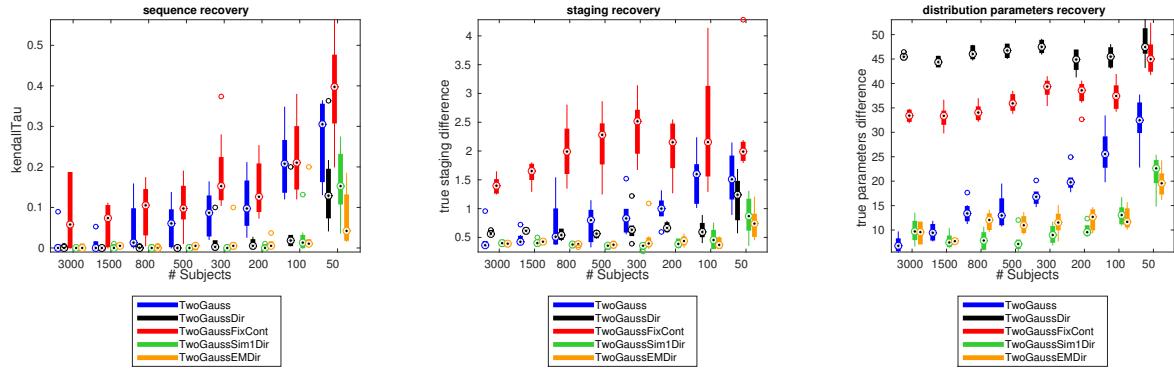


Figure A.2: Simulation results comparing different EBM model fitting methods. In blue we have the standard EBM method which first fits the distribution parameters and then the sequence, while in black and red we have the EBM Expectation-Maximisation method with two different starting points. Each of these three methods have been fit on eight synthetic datasets where we used a decreasing number of subjects, from 3000 down to 50. The left figure shows the Kendall-tau distance between the sequence recovered by the model and the true sequence. The middle figure shows the L1 norm of the difference between predicted subject stages and true stages. The right figure shows the L1 norm of the difference between the recovered distribution parameters and the true parameters.

TwoGauss	fit distribution parameters by optimising a mixture model, like Fontejin et al. [13], Young et al. [52]
TwoGaussDir	fitting each distribution directly on the control and patient data, optimise sequence independently
TwoGaussFixCont	fit the control distribution directly on control data, then fit patient distribution using mixture model optimisation.
TwoGaussSim1	simultaneous sampling of parameters and sequence, initialise the method from the solution found by TwoGauss
TwoGaussSim1Dir	same as TwoGaussSim1, initialise the method from the solution found by TwoGaussDir

Table A.1: Legend description for Fig. A.2

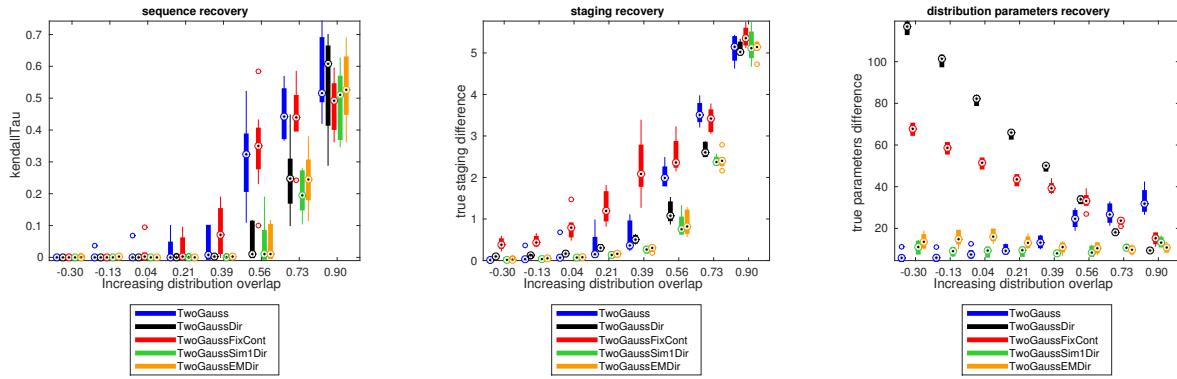


Figure A.3: Simulation results comparing different EBM model fitting methods. Same setup as in Fig. A.2, but this time we keep the number of subjects constant and instead increase the overlap of distributions $p(x|E)$ and $p(x|\neg E)$. The distribution overlap ranges from -0.3 (low overlap) to 0.9 (high overlap).

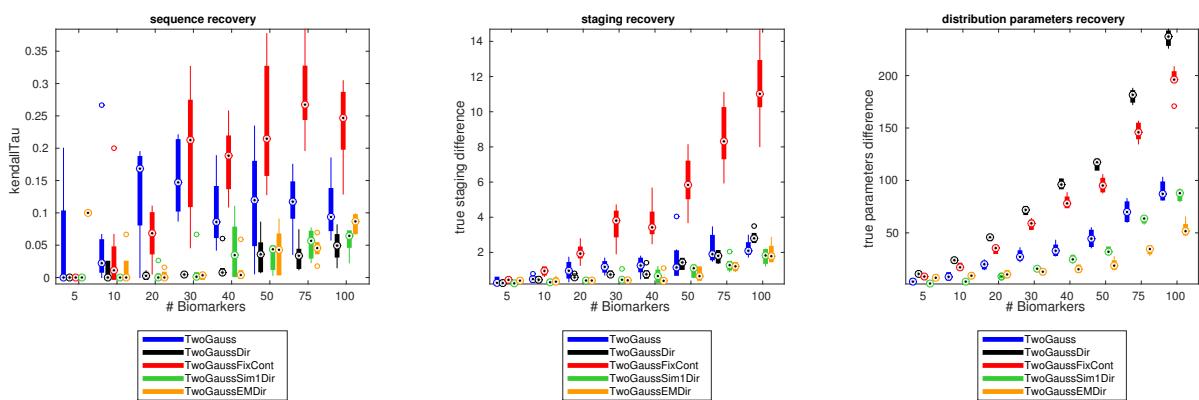


Figure A.4: Simulation results comparing different EBM model fitting methods. Same setup as in Fig. A.2, but this time we keep the number of biomarkers used, from 0 to 100.

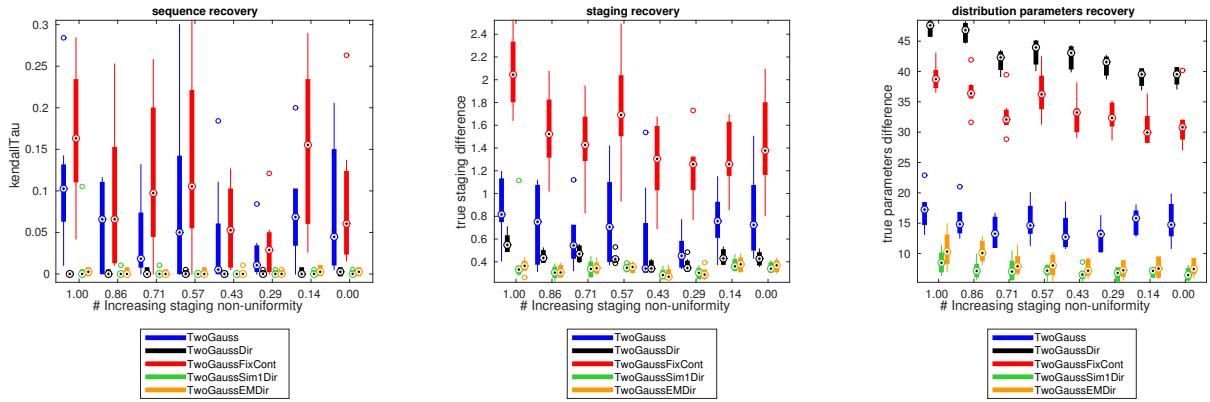


Figure A.5: Simulation results comparing different EBM model fitting methods. Same setup as in Fig. A.2, but this time we vary the uniformity of the subject stages, ranging from 1 (uniform probability across the disease timecourse) to 0 (subjects are placed more at beginning or end stages). The pdfs of the stage distributions for each experiment are shown in Fig. A.6.

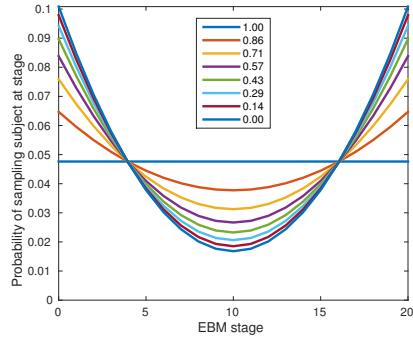


Figure A.6: Pdf of the staging probability functions used to generate the synthetic datasets in Fig. A.5. The shapes of the pdf distributions are quadratic curves for which we vary the second-order factor.

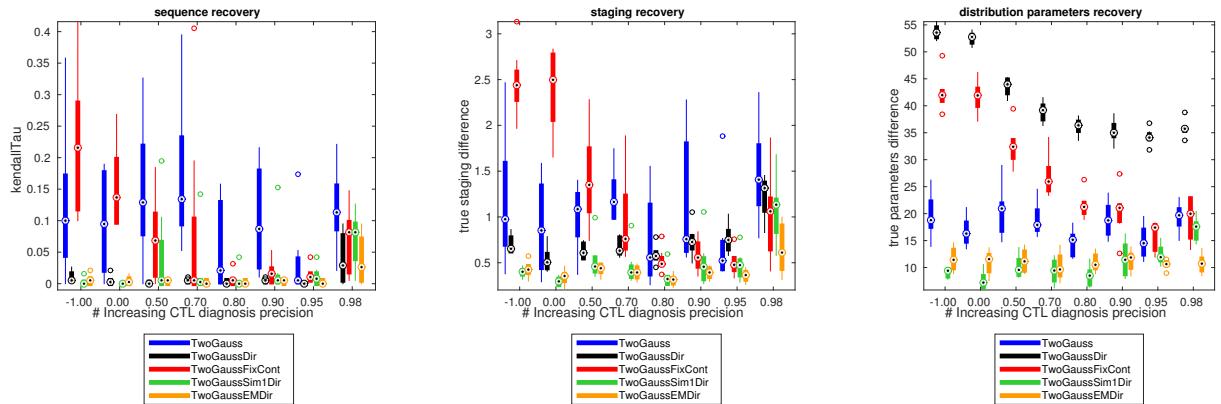


Figure A.7: Simulation results comparing different EBM model fitting methods. Same setup as in Fig. A.2, but this time we vary the precision of the control distribution, ranging from -1 (well-defined control population) to 1 (not so well defined control population). The pdfs of the stage distribution for control subjects in each experiment are shown in Fig. A.8.

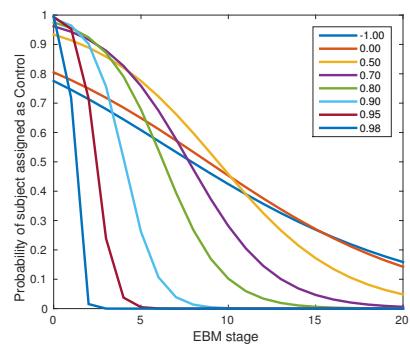


Figure A.8: Probability density functions for the sampling distribution of the stages of controls. This ranges from -1 (well-defined control population) to 1 (not so well defined control population)