Modelling and Simulating the Progression of Alzheimer's Disease through the Analysis of Multi-modal Neuroimages and Clinical Data

Clément Abi Nader

UNIVERSITÉ CÔTE D'AZUR

ÉCOLE DOCTORALE
SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

THÈSE DE DOCTORAT

Modélisation et Simulation de l'Évolution de la Maladie d'Alzheimer à partir de l'Analyse d'Images Cérébrales Multimodales et de Données Cliniques

Clément ABI NADER

INRIA, Équipe EPIONE

Thèse dirigée par Nicholas AYACHE et co-dirigée par Philippe ROBERT

Soutenue le 15 Juin 2021

Présentée en vue de l'obtention du grade de DOCTEUR EN AUTOMATIQUE, TRAITEMENT DU SIGNAL ET DES IMAGES de l'UNIVERSITÉ CÔTE D'AZUR.

Devant le jury composé de :

Daniel ALEXANDER	University College London	Rapporteur
Stanley DURRLEMAN	Inria Paris	Rapporteur
Giovanni B. FRISONI	Hôpitaux Universitaires de Genève	Président
Nicholas AYACHE	Inria Sophia Antipolis	Directeur de thèse
Philippe ROBERT	Université Côte d'Azur (EA Cobtek)	Co-directeur de thèse
Marco LORENZI	Inria Sophia Antipolis	Co-encadrant
Adam SCHWARZ	Takeda Pharmaceutical Company	Invité

Inria

UCA JEDI INITIATIVE D'EXCELLENCE

médecine Numérique mNC3 Cerveau Cognition Comportement

Modélisation et Simulation de l'Évolution de la Maladie d'Alzheimer à partir de l'Analyse d'Images Cérébrales Multimodales et de Données Cliniques

Modelling and Simulating the Progression of Alzheimer's Disease through the Analysis of Multi-modal Neuroimages and Clinical Data

Jury

Président du jury

Giovanni B. Frisoni	Professeur	Hôpitaux Universitaires de Genève

Rapporteurs

Daniel Alexander	Professeur	University College London
Stanley Durrleman	Directeur de recherche	INRIA Paris

Examinateurs

Daniel Alexander	Professeur	University College London
Stanley Durrleman	Directeur de recherche	INRIA Paris
Nicholas Ayache	Professeur	INRIA Sophia-Antipolis
Philippe Robert	Professeur	Université Côte d'Azur
Marco Lorenzi	Chargé de recherche	INRIA Sophia-Antipolis

Invités

Adam Schwarz	Docteur	Takeda Pharmaceutical Company

Abstract

Alzheimer's disease (AD) is a neurodegenerative disorder characterized by biological and morphological processes which spread over decades, ultimately leading to cognitive and behavioral decline. These processes can be monitored thanks to biomarkers and imaging measurements. As the mechanisms underlying the evolution of the pathology remain partially unknown, disease progression models have been introduced to describe biomarkers progression in a data-driven manner. These methods hold great clinical interest as they can be used for diagnosis, prognosis and monitoring drug efficacy in clinical trials. Most of currently available approaches are essentially based on the joint analysis of scalar biomarkers data, or on the analysis of multivariate data derived from a single imaging modality only. Scalability of these methods to jointly account for high resolution multi-modal imaging data is currently an issue to develop a comprehensive model of the natural history of AD. Moreover, while current methods provide a descriptive analysis of the biomarkers evolution, they don't explicitly model their dynamical interplay, nor they allow the statistical assessment of the efficacy of hypothetical interventions scenarios. Within this context, the main objective of this thesis is to propose novel methodological frameworks for modeling the natural history of AD through the analysis of high-resolution multi-modal imaging data, while enabling the investigation of biomarkers dynamics and intervention strategies.

First, we introduce a novel theory to identify the spatio-temporal dynamics characterizing the pathological processes at stake during AD, by relying on the analysis of multi-modal volumetric images. To this end, we decompose the spatio-temporal data through matrix factorization between temporal and spatial sources, while constraining the spatio-temporal sources to a set of biologically-inspired constraints in order to estimate a realistic model of disease progression.

Second, we present a novel approach proposing to model AD progression by means of a dynamical system relating clinical and multi-modal imaging markers. The method is inspired by neural ordinary differential equations, to project the data in a low-dimensional space in which we estimate the dynamical system underlying the pathological progression. Thanks to our approach, we can simulate hypothetical scenarios of disease progression. In particular, we investigate the effect of drug intervention on cognitive outcomes.

Third, we evaluate our work on an independent memory clinic cohort. We show that our dynamical model allows to accurately perform individual disease staging for patients from the external cohort. This result highlights the potential of our model as a prognosis tool or for monitoring treatment efficacy.

Overall, the computational methods presented in this thesis coherently account for multi-modal neuroimaging and clinical data to provide a thorough description of AD progression. Our approaches entail great potential for clinical application: whether to simulate the effect of potential treatments on the pathological progression, or to provide complementary information for diagnosis and prognosis through automatic disease staging. The work presented in this thesis sets the basis for several research directions. Our methods can be generalized to investigate other neurodegenerative disorders, such as Huntington's disease and Parkinson's disease. Moreover, our framework can be extended to provide personalized pathological evolution accounting for a set of relevant risk factors which may affect AD course.

Keywords: Alzheimer's disease; clinical trials; Magnetic Resonance Imaging; Positron Emission Tomography; machine learning; Gaussian processes; variational autoencoder; dynamical systems

Résumé

La maladie d'Alzheimer est une maladie neuro dégénérative caractérisée par un ensemble de processus biologiques et morphologiques s'étendant sur plusieurs décennies et induisant un sévère déclin des capacités cognitives et comportementales. Ces processus peuvent être suivis grâce à des biomarqueurs ainsi que des mesures extraites d'images médicales. Alors que les mécanismes sous-tendant l'évolution de la pathologie restent en partie inconnus, des modèles de progression de la maladie ont été présentés afin de décrire l'évolution des biomarqueurs à partir de la seule analyse de données. Ces méthodes comportent un intérêt médical certain pour faciliter le diagnostic et le pronostic, mais également afin de mesurer l'efficacité d'un médicament durant un essai clinique. La plupart des méthodes qui existent actuellement se fondent sur l'analyse conjointe de biomarqueurs scalaires, ou sur l'analyse de données multivariées extraites d'un type d'imagerie donné. Le passage à l'échelle de ces méthodes pour l'analyse de données en grande dimension issues de l'acquisition d'images multimodales est actuellement un défi majeur, qui pourrait permettre de développer un modèle détaillé de l'histoire naturelle de la maladie. De plus, bien que les méthodes proposées jusqu' à maintenant fournissent une analyse descriptive de l'évolution des biomarqueurs, elles ne modélisent pas explicitement les interactions dynamiques entre eux, et ne permettent pas non plus d'entreprendre une évaluation statistique de l'efficacité d'une hypothétique intervention thérapeutique. Par conséquent, l'objectif principal de cette thèse est de proposer de nouvelles méthodes permettant de modéliser l'évolution de la maladie d'Alzheimer à travers l'analyse d'images médicales volumiques multimodales, ainsi que de permettre l'étude des dynamiques des différents biomarqueurs et de l'effet potentiel d'une intervention thérapeutique sur l'évolution de la maladie.

Dans la première partie, nous présentons une méthode dont l'objectif est d'identifier les dynamiques spatio-temporelles qui caractérisent les processus en jeu durant la maladie d'Alzheimer, en s'appuyant sur l'analyse d'images médicales volumiques multimodales. Pour y parvenir, nous proposons de modéliser les données via une factorisation matricielle entre des sources temporelles et spatiales. Nous imposons également un ensemble de contraintes biologiques aux sources spatio-temporelles afin de garantir l'estimation d'un modèle de progression de la maladie réaliste.

Dans la seconde partie, nous proposons une approche originale qui consiste à modéliser la progression de la maladie d'Alzheimer via un système dynamique reliant des évaluations cliniques ainsi que des données extraites d'images multimodales. Notre méthode s'appuie sur un autoencoder variationel afin de projeter les données dans un espace de faible dimension. A l'intérieur de cet espace nous supposons l'existence d'un système dynamique contrôlant la progression de la maladie et dont nous estimons les paramètres. Notre méthode permet de simuler des scénarios de progression hypothétiques de la maladie. Nous étudions en particulier l'effet de traitements potentiels sur les capacités cognitives.

Dans la troisième partie, nous testons notre second modèle sur une cohorte indépendante provenant d'une clinique de la mémoire. Nous montrons que le modèle permet d'évaluer de manière précise le stade clinique des sujets provenant de la cohorte externe. Ces résultats soulignent le potentiel de notre modèle en tant qu'outil clinique, que ce soit pour fournir un pronostic ou bien pour participer au suivi de l'efficacité de traitements.

Dans l'ensemble, les méthodes computationnelles présentées dans cette thèse exploitent la disponibilité d'images multimodales afin d'estimer une progression détaillée de la maladie d'Alzheimer. Outre l'exploration de l'évolution de la maladie, nos approches possèdent un large éventail d'applications cliniques: que ce soit pour simuler l'effet de traitements potentiels sur la progression pathologique, ou pour fournir des informations complémentaires en termes de diagnostic et pronostic grâce à l'estimation automatique du stade de la maladie. Enfin, le travail présenté dans cette thèse ouvre différentes pistes de recherche. Par exemple, nos méthodes pourraient être appliquées à l'étude d'autres maladies neuro dégénératives comme la maladie d'Huntington ou la maladie de Parkinson. Enfin, une piste d'amélioration majeure serait de personnaliser les modèles de progression de la maladie, en prenant en compte un ensemble de facteurs de risque qui caractérisent les sous-types existant de la maladie d'Alzheimer.

Mots-clés: maladie d'Alzheimer; essais cliniques; Imagerie par Résonance Magnétique; Tomographie par Émission de Positons; apprentissage automatique; processus Gaussiens; autoencoder variationel; systèmes dynamiques

Acknowledgement

Je souhaiterais en premier lieu remercier mes encadrants Nicholas, Philippe et Marco qui m'ont offert la chance d'effectuer ma thèse au sein d'Epione, une équipe riche d'histoire et de contributions scientifiques. Nicholas, je te remercie pour tes conseils avisés et la liberté avec laquelle tu m'as laissé mener ma thèse. Je garderai aussi un agréable souvenir de nos pauses déjeuner sous le soleil de la Côte d'Azur. Philippe, je me souviendrai d'une conversation un été à l'ICP durant laquelle tu m'avais fait remarquer les limites du travail que j'enteprenais pour repondre aux problèmes d'un medecin. Ce jour-là tu as fait germer dans mon esprit une des idées les plus importantes de ma thèse et je t'en remercie. Enfin, je tiens à remercier chaleureusement Marco Lorenzi dont les idées pavent ce manuscrit. Marco tu as été l'artisan de tout ce travail dont cette thèse représente l'aboutissement. Je te suis reconnaissant pour m'avoir transmis ta rigueur, ton exigence, et pour m'avoir formé tout au long de ces quatre années.

Je tiens également à remercier les membres du jury et en particulier le Professeur Daniel Alexander et le Docteur Stanley Durrleman pour avoir accepté de prendre le temps de lire ce manuscrit et de partager leurs commentaires et conseils. Je suis aussi reconnaissant envers le Professeur Giovanni Frisoni qui a accepté de faire partie de mon jury de thèse, mais qui a également eu la courtoisie de nous faire part de son expertise à maintes reprises durant ces quatre années. Enfin, je voudrais remercier le Docteur Adam Schwarz qui a accepté notre invitation afin d'assister à la soutenance de thèse.

Ces quatre années ont également été émaillées de rencontres qui me sont chères, au sein de l'équipe et ailleurs. Je ne saurai donc écrire ce manuscrit sans mentionner mes compagnons de route Luigi et Jaume avec qui je partage bien plus que cette thèse mais une belle amitié. Je pense aussi à Selma Souihel et Yann Thanwerdas et nos discussions sans fin dans notre repère antibois. Je ne peux bien évidemment pas oublier Santiago dont le sourire permanent et la joie de vivre auront irradié ces années de thèse. Enfin, je souhaite remercier tous les membres d'Epione, permanents et doctorants, avec une pensée affectueuse pour: Raphael Sivera, Julian Krebs, Tania Bacoyannis, Shuman Jia, Nicolas Guigui, Sara Garbarino, Bastien Manach, Marco Milanesio, Pawel Mlynarski, Benoit Audelan, Nicolas Cedilnik et Zihao Wang.

Enfin, il m'est impossible de conclure ces quelques lignes sans remercier ma famille qui aura été mon plus précieux soutien durant toutes mes années d'études que cette thèse vient sanctionner. Je pense en particulier à ma mère et à mon père. De Toulouse à Sophia-Antipolis en passant par Brest, la route ne fut pas exactement droite et souvent par ma faute. Mais vous étiez toujours derrière moi pour me soutenir et me guider.

Financial support

This work has been supported by the French government, through the UCA^{JEDI} and 3IA Côte d'Azur Investments in the Future project managed by the National Research Agency (ref.n ANR-15-IDEX-01 and ANR-19-P3IA-0002), the grant AAP Santé 06 2017-260 DGA-DSH, and by the Inria Sophia-Antipolis-Méditerranée, "NEF" computation cluster.

Contents

1 Introduction **1**

1.1 Clinical Context . . . 1
1.1.1 Alzheimer's disease . . . 1
1.1.2 Monitoring and understanding Alzheimer's disease progression . 2
1.2 Disease Progression Modelling . . . 4
1.2.1 Basic approaches and limitations . . . 4
1.2.2 Data-driven disease progression models . . . 5
1.2.3 Spatio-temporal approaches . . . 6
1.3 Current Challenges . . . 8
1.4 Objectives and Organization of the Thesis . . . 9
1.5 Publications . . . 10

2 Monotonic Gaussian Process for Spatio-Temporal Disease Progression Modeling **11**

2.1 Introduction . . . 12
2.2 Related Work in Neurodegeneration Modeling . . . 15
2.3 Methods . . . 16
2.3.1 Individual time-shift . . . 16
2.3.2 Data modeling . . . 16
2.3.3 Spatio-temporal processes . . . 18
2.3.4 Sparsity . . . 20
2.3.5 Variational inference . . . 21
2.4 Experiments and Results . . . 23
2.4.1 Synthetic tests on spatio-temporal trajectory separation . . . 23
2.4.2 Synthetic tests on trajectory separation and time reparameterization 24
2.4.3 Application to spatio-temporal brain progression modeling . . . 28
2.5 Discussion . . . 36
2.6 Appendix . . . 39
2.6.1 Lower bound derivation . . . 39
2.6.2 Kullback-Leibler computation . . . 40
2.6.3 Experimental setting . . . 40
2.6.4 Optimization procedure . . . 41
2.6.5 Synthetic experiments using ICA, NMF and PCA . . . 44
2.6.6 Model convergence . . . 45

2.6.7 Application of ICA, NMF and PCA on brain imaging data 45

3 Simulating the Outcome of Amyloid Treatments in Alzheimer's Disease 61
3.1 Introduction 62
3.2 Materials and Methods 64
3.2.1 Study cohort and biomarkers' changes across clinical groups . . . 65
3.2.2 Model overview 67
3.2.3 Data modelling 67
3.2.4 Simulating the long-term progression of AD 69
3.2.5 Simulating intervention 69
3.2.6 Evaluating disease severity 70
3.2.7 Statistical analysis 70
3.3 Results 70
3.3.1 Data acquisition and pre-processing 71
3.3.2 Progression model and latent relationships 71
3.3.3 Simulating clinical intervention 73
3.4 Discussion 75
3.5 Conclusion 81
3.6 Appendix 82
3.6.1 Variational inference 82
3.6.2 Model optimization 83
3.6.3 Lower bound 84
3.6.4 KL divergence 84
3.6.5 Graphical model 85
3.6.6 Lower bound computation 85
3.6.7 Time-shift comparison and validation 86
3.6.8 Simulated clinical endpoints 86

4 SimulAD: A dynamical model for personalized simulation and disease staging 89
4.1 Introduction 90
4.2 Material and Methods 93
4.2.1 Experimental cohort 93
4.2.2 Image preprocessing 95
4.2.3 Modeling framework 95
4.2.4 Estimated model 96
4.2.5 Evaluation strategy 96
4.2.6 Data adjustment 97
4.3 Results 100
4.3.1 Known-groups validity 100
4.3.2 Concurrent validity 102
4.3.3 Model reliability 103
4.4 Discussion 103

4.5 Conclusion 107
4.6 Appendix 108
4.6.1 Clinical scores imputation 108
4.6.2 FDG prediction 108
4.6.3 Tracer correction 108
4.6.4 Robustness to the clinical scores prediction 112
4.6.5 Comparison of the simulated evolution of clinical and imaging measures based on ADNI and GMC cohorts 112

5 Conclusion **115**
5.1 Summary of the Main Contributions 115
5.2 Perspectives and Future Applications 118
5.2.1 Application to other neurodegenerative disorders 118
5.2.2 Learning from multi-centric studies 119
5.2.3 Identifying subtypes of disease progression and accounting for risk factors: towards precision medicine 120

Bibliography **121**

1 Introduction

Contents

1.1 Clinical Context **1**
1.1.1 Alzheimer's disease 1
1.1.2 Monitoring and understanding Alzheimer's disease progression 2
1.2 Disease Progression Modelling **4**
1.2.1 Basic approaches and limitations 4
1.2.2 Data-driven disease progression models 5
1.2.3 Spatio-temporal approaches 6
1.3 Current Challenges **8**
1.4 Objectives and Organization of the Thesis **9**
1.5 Publications **10**

1.1 Clinical Context

Dementia is a neurological disorder whose symptoms include memory loss, aphasia, disorientation as well as further behavioral issues (apathy, self-neglect, mood swings), ultimately precluding patients from living autonomously [Scott, 2007]. Due to the aging of the population, the number of people living with this syndrome has more than doubled in the past 30 years to approximately reach 43.8 million [Nichols, 2019]. This increase in the prevalence of individuals suffering from dementia negatively impacts families and raises growing public-health concerns [Etters, 2008].

1.1.1 Alzheimer's disease

Alzheimer's disease (AD), which was first described by Dr A. Alzheimer in 1906, is the most common cause of dementia and accounts for 60% to 80% of the cases [Alzheimer Association Report, 2020]. The pathology is classified as a proteinopathy defined by two processes: an abnormal processing of amyloid precursor protein leading to the formation in the brain cortex of amyloid-beta neuritic plaques, and accumulation of hyperphosphorylated tau in neurofibrillary tangles [Braak, 1991; Braak, 1997; Hardy, 2002; Long, 2019]. While a definitive diagnosis can only be established by post-mortem

examination showing abnormal deposits of both amyloid and tau, patients are traditionally suspected to suffer from AD based on the presence of a characteristic multidomain amnestic dementia [McKhann, 1984].

Nonetheless, recent studies showed that diagnosis of AD dementia based on a set of neuro-psychological assessments is not necessarily associated with neuropathological changes. In [Nelson, 2011], the authors estimated that 10% to 30% of patients clinically diagnosed with AD dementia did not exhibit the typical neuropathological changes of the disease. Moreover, it has been observed that between 30% to 40% of cognitively unimpaired elderly individuals have abnormal amyloid and tau deposits at autopsy [Knopman, 2003; Bennett, 2006]. These observations show that clinical symptoms are neither specific nor sensitive to the toxic aggregation of amyloid and tau. Therefore, it has been recently proposed to operate a clear distinction between the dementia syndrome and one of its most common cause which is AD. Given this context, evidence of abnormal deposits of amyloid and tau should be sufficient to assign the label "AD" to an individual, independently from the existence of clinical symptoms [Jack, 2018]. Nevertheless, if only post-mortem examination allows to reliably diagnose the pathology, while neuro-psychological assessments indicate a probable presence of AD but not a definitive diagnosis, it is necessary to define specific biological criteria for identifying and following patients during the disease course.

1.1.2 Monitoring and understanding Alzheimer's disease progression

One of the major challenges for understanding AD is that the pathology evolves asymptomatically for a long period (between 15-20 years) before the manifestation of cognitive and behavioral symptoms. Therefore, efforts have focused on finding a set of biomarkers that would allow to detect and monitor pathophysiological changes characteristic of AD. These efforts resulted in the identification of three main biomarkers categories, namely: amyloid, tau and neurodegeneration [Jack, 2016]. This latter process encompasses neuron losses and synaptic dysfunction, which highly correlate with clinical symptoms [Terry, 1991]. Given this context, it was recently proposed to define AD as a "biological construct identified by biomarkers" [Jack, 2018], and whose hallmark is the abnormal deposition of both amyloid and tau in the brain cortex [Montine, 2012; Hyman, 2012]. This means that AD can be tracked via biomarkers which indicate the abnormality of specific physiological processes. For instance, measurements of concentration of Cerebrospinal fluid (CSF) Aβ_{42} and t-tau allow to detect pathological levels of amyloid-beta and tau respectively [Blennow, 2003]. Besides CSF concentration, techniques such as Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) enabled to define non-invasive imaging biomarkers which are suited to monitor many pathophysiological changes involved in AD. For example, abnormal deposition of amyloid-beta

can be measured via (18)F-florbetapir-PET (AV45-) scan [Klunk, 2004; Clark, 2011]. Accumulation of neurofibrillary tangles is quantified thanks to (18)F-flortaucipir-PET (Tau-) scans [Okamura, 2014; Barthel, 2020]. Finally, neurodegeneration is indicated by cerebral atrophy and glucose hypometabolism from MRI scans [Fox, 2004] and (18)F-fluorodeoxyglucose-PET (FDG-) [Herholz, 2012] imaging respectively.

The possibility to collect biomarkers fostered the conduct of longitudinal studies following individuals over years, in order to investigate the set of neuropathological changes occurring during AD, and to understand how they ultimately lead to dementia. Recently, efforts have focused in developing models of disease progression which aimed at providing a global understanding of the large amount of biomarkers acquired [Aisen, 2010; Frisoni, 2010; Jack, 2010; Jack, 2013b]. The objective of such models is to shed light on the natural history of the disease course, which could facilitate early diagnosis of patients suffering from AD and the design of therapeutic intervention. We show in Figure 1.1 the most common model of AD progression, which supports the idea that AD follows a pathological cascade of events. This cascade would be triggered by the abnormal deposition of amyloid-beta in the brain cortex, which occurs decades before the onset of clinical symptoms. Formation of amyloid-beta plaques induces a burden of neurofibrillary tangles detected by pathological accumulation of tau. This latter pathophysiological phenomenon leads to the neurodegeneration process which is the proximate cause of dementia.

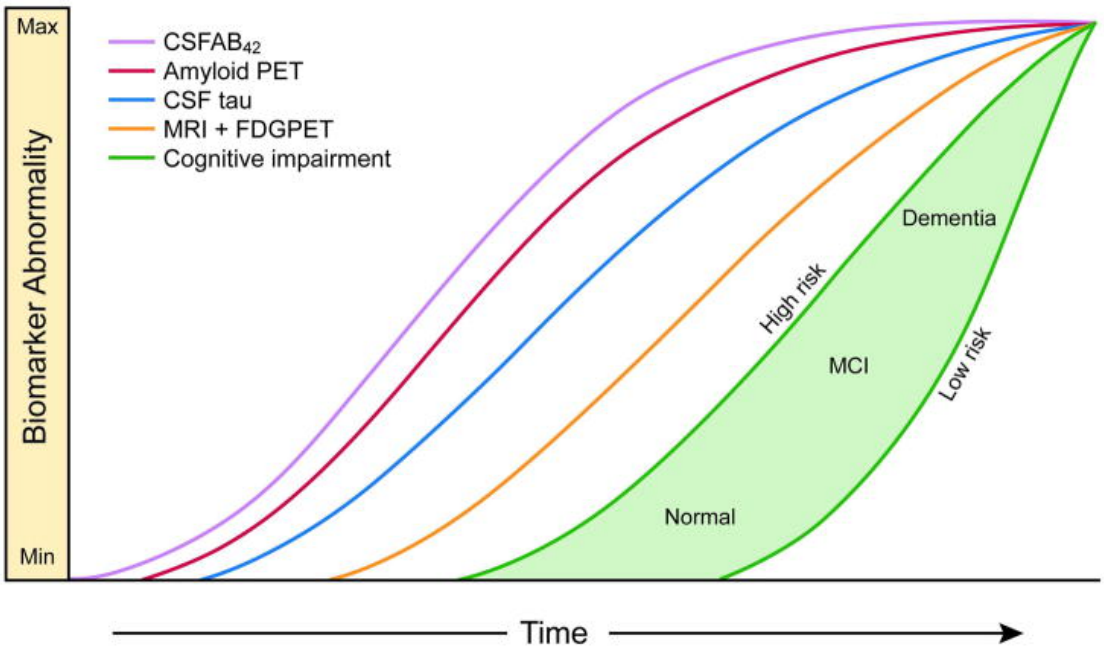

Figure 1.1: Hypothetical model of AD progression. Key biomarkers at stake in AD evolve from a normal to an abnormal stage. Reprinted from The Lancet Neurology, Vol. 12, Jack et al., Tracking pathophysiological processes in AD: an updated hypothetical model of dynamic biomarkers , 207–216, Copyright (2013), with permission from Elsevier.

1.2 Disease Progression Modelling

While models such as the one from Figure 1.1 provide a precious groundwork for understanding AD progression, they remain hypothetical in the sense that they are built by organizing observations extracted from multiple studies in a comprehensive qualitative diagram. Indeed, they lack quantification of key aspects such as the progression speed of the different biomarkers, or the time between the reaching of abnormal levels for these biomarkers and clinical symptoms onset. Moreover, there may exist additional mechanisms driving the pathological progression that might be "missed" by a model solely based on a global analysis of biomarkers. Therefore, the development of data-driven models of disease progression based on the analysis of biomarkers data is fundamental to validate current hypothesis about the evolution of the pathology. Furthermore, these models would also provide a quantitative reference progression against which biomarkers could be compared, thus allowing to stage the individual disease severity by locating a subject along the disease course. This capability entails great potential to provide complimentary insights for diagnosis, prognosis as well as in the context of clinical trials. Indeed, there is currently need of testing pharmacological treatments during the pre-symptomatic phase of the disease [Sperling, 2011; Aisen, 2018], when intervention might be more likely to result in improvement of cognitive endpoints. Within this context, models of disease progression are crucial as they could help identifying subjects during the pre-symptomatic stage, thus reducing heterogeneity of participants. They could also be used to monitor the efficacy of clinical trials.

1.2.1 Basic approaches and limitations

One of the first models of AD progression entirely informed by the analysis of longitudinal data was proposed by [Bateman, 2012]. This study focused on patients suffering from autosomal dominant AD, which is a sub-type of the disease genetically inherited. In this particular case, the estimated years before onset (EYO) can be computed by subtracting the age of the parents at onset with the age of the patient. Measures of amyloid deposition, tau accumulation, glucose metabolism, cerebral atrophy and cognition were collected for all the participants. The authors used a linear mixed model to regress each of these biomarkers against the mutation status and EYO. Thus, they could estimate in a data-driven manner the evolution of each biomarker over 40 years. Their findings are in agreement with the hypothesis of the pathological cascade, as they showed that deposition of amyloid beta starts approximately 20 years before the appearance of clinical symptoms, followed by cerebral atrophy and a decrease of glucose metabolism, ultimately leading to cognitive and behavioral impairment.

However, approaches such as the one proposed in [Bateman, 2012] are not suitable for modeling sporadic forms of AD. This form of AD affects people without a family record of the disease and accounts for 90% of the cases [Bekris, 2010]. As the pathological cascade is initiated at different moments and evolves at heterogeneous speed, the age at onset is generally unpredictable. Therefore, we cannot directly apply a conventional regression model for estimating the evolution of the different biomarkers, due to the lack of a temporal variable describing an absolute time-line of the disease course, such as the EYO. Moreover, it would also be necessary to determine if the estimated progression can be completely transposed to other sub-types of the pathology such as sporadic AD.

So far, there have been attempts trying to circumvent this difficulty through the analysis of longitudinal data. For instance, in [Lo, 2011] the authors compute mean rates of change for many biomarkers measured for hundreds of subjects followed during three years. The participants were classified in three clinical groups: healthy controls, mild cognitive impairment (MCI) and AD dementia. Thus, they could approximate the average evolution of the different biomarkers per clinical group over three years, and derived a "long-term" model of AD evolution by juxtaposing the progression from the healthy, MCI and AD dementia groups. However, by dividing the disease course in three clinical stages this approach provides a coarse temporal resolution of the pathological progression. Another work presented in [Villemagne, 2013] overcomes this problem. The authors propose to find the model best fitting the relationship between the individual rate of change and its baseline value for different biomarkers. Then, the authors integrate the fitted-model over time to obtain long-term trajectories for each biomarker. Their findings suggest that AD is characterized by a long preclinical phase. Indeed, they estimate that deposition of amyloid beta becomes abnormal 17 years before the onset of dementia, while hippocampal atrophy and memory impairment reach pathological levels 4 years and 3 years respectively before AD dementia diagnosis. However, the time-line describing the disease progression is arbitrary, as it depends on abnormality cut-points defined for each biomarkers. Since there exists different principles allowing to establish such cut-points [Bartlett, 2012], this may affect the pathological progression they estimate.

Overall, these works illustrate the difficulty of estimating data-driven models of AD progression, and the need to resort to more advanced mathematical approaches to overcome the lack of an absolute temporal reference describing the disease course.

1.2.2 Data-driven disease progression models

Given this context, a variety of disease progression models aimed at automatically characterizing AD evolution through the sole analysis of biomarkers measurements from large cohort studies. The Event-Based Model (EBM) [Fonteijn, 2012; Young, 2014], is one of the first fully data-driven methods tackling this problem. In the EBM, the

pathological evolution is modelled as a sequence of events representing a collection of biological, morphological and cognitive changes due to AD, and whose ordering needs to be inferred. Each event is statistically defined as the transition of a given biomarker from a normal to an abnormal state. The biomarkers are separately modeled as a mixture of two Gaussians providing distributions of normal and abnormal values respectively. Data is assumed to follow a generative model, and the most probable sequence of events is inferred by maximizing the data likelihood. The EBM allows to estimate an ordering of events spanning the full disease course in agreement with the hypothetical model of [Jack, 2013b], while also assessing the uncertainty of the location of each event within the sequence. Moreover, this approach enables to evaluate the disease stage of a given individual by finding where it lies along the sequence of events. An extension of this approach that improves the accuracy of individual disease staging was proposed in [Venkatraghavan, 2019].

Nonetheless, the EBM is based on the simplistic assumption describing AD progression as a discrete sequence of biomarkers transitions from normal to abnormal states. To overcome this limitation, alternative approaches attempt to model long-term trajectories of key biomarkers over a continuous time-line describing the disease history. Since we lack a temporal axis that could be used as an absolute reference, these methods propose to map the subjects' observations on a common temporal scale associated with a long-term progression from healthy towards pathological states. This is generally done by introducing an individual time-warping function [Jedynak, 2012]. Moreover, these methods also aim to estimate a clinically plausible progression for each biomarkers, by constraining them to follow a steady pathological evolution. To do so, the biomarkers' dynamics can be modelled by sigmoid functions as proposed in [Schiratti, 2015; Marinescu, 2019a]. Other models adopt a non-parametric approach to allow a broader range of temporal profiles, such as [Lorenzi, 2017; Lorenzi, 2018] who model the biomarkers' dynamics thanks to Gaussian Processes associated with a monotonicity constraint. Overall, these approaches enable the estimation of long-term dynamics for a wide variety of scalar biomarkers along a continuous time-line representing the disease course. They can be used to perform individual disease staging, by identifying the time-point of the pathological progression which best reflects the state of a subject with respect to its measurements.

1.2.3 Spatio-temporal approaches

Thus far, we reviewed models of disease progression that were based on the analysis of scalar biomarkers data: amyloid level in the CSF, regional gray matter volumes, clinical scores, etc. However, computational anatomy approaches such as Voxel-Based Morphometry [Ashburner, 2000] or surface-based analysis [Fischl, 2000] allow to work on a variety of data types like volumetric images or cortical shapes. This kind of spatial information provides a richer anatomical description compared to data obtained

by averaging regional imaging measurements. Therefore, building models of disease progression based on the analysis of multivariate spatial data derived from a given imaging modality could bring new insights on the topography and dynamics of the changes affecting the brain during AD.

One of the first methods addressing this challenge was presented in [Koval, 2017; Koval, 2018], where the authors proposed a non-linear mixed-effect model to describe spatio-temporal changes affecting the brain during AD based on cortical thickness data. The fixed-effect accounts for a global progression, while the random-effects are estimated using warping functions that evaluate how a given individual differs in space and time from the average trajectory. Results show that the model is able to capture patterns of cortical atrophy characteristic of AD. It can also be used to evaluate the individual disease severity of a subject based on its cortical thickness measurements. However, although the method accounts for the spatio-temporal shift of an individual with respect to the estimated average pattern of cortical atrophy, the authors assume that atrophy follows a unique spatio-temporal trajectory during AD. Since it was observed that brain regions can be characterized by different atrophy rates [Whitwell, 2010], another approach recently proposed to jointly estimate various spatio-temporal patterns of atrophy [Marinescu, 2019a]. To do so, they assign each vertex of the cortical surface to a cluster associated with a specific sigmoid function describing the pathological progression. Results show that the authors are able to detect differential spatio-temporal patterns of atrophy affecting the brain during AD.

Apart from cortical thickness data, multi-modal volumetric images can provide anatomical information not observable via the analysis of cortical shapes, while bringing insights on additional pathological processes such as amyloid deposition. Therefore, there also exists methods which aim to describe the pathological spatio-temporal changes associated with AD based on the analysis of volumes of brain images. For instance [Bowles, 2018] identifies patterns of regional atrophy due to AD by analyzing MRI images, while [Bilgel, 2016] describes the topography of amyloid deposition during AD based on the analysis of amyloid-PET scans.

Overall, estimating models of disease progression through the analysis of multivariate spatial data is a challenging task, due to the complexity and high-dimensionality of this type of data. Therefore, the spatio-temporal methods presented so far usually face tractability issues, and need to resort to advanced optimization techniques to estimate the models parameters. For instance, in [Bilgel, 2016; Koval, 2018; Marinescu, 2019a] parameters inference is carried out by relying on modified versions of the Expectation-Maximization algorithm. Moreover, it is necessary in some cases to downsample the volumetric images and cortical thickness data, such as in [Bilgel, 2016; Bowles, 2018] and [Koval, 2017; Koval, 2018] respectively, to guarantee tractability of the methods.

While this downsampling step enables these approaches to scale to the analysis of imaging data, it prevents them from fully leveraging the spatial information they provide.

1.3 Current Challenges

In the previous section, we introduced disease progression modelling which aims at identifying and understanding the morphological and physiological processes at stake during AD, based on the analysis of clinical and imaging data. Currently, the more advanced methods rely on mathematical techniques that allow to model plausible long-term trajectories of high-dimensional imaging data. Such models hold great potential to improve our knowledge of the mechanisms driving AD, which still remain partially unknown. They could also be used in clinical practice to provide complimentary information for diagnosis, prognosis and for identifying patients at risk of cognitive decline, which is critical to reduce heterogeneity in clinical trials. However, these approaches present some limitations and disease progression models still face challenges.

One of the main limitation of current disease progression models is related to their scalability to high resolution multi-modal imaging data. Until now, disease progression models have been applied on multivariate data such as cortical thickness or on downsampled volumetric images, but they do not scale to the joint analysis of 3-dimensional (3D) multi-modal images at their native resolution. Yet, estimating long-term trajectories of volumetric images could provide valuable insights about the topography of pathological patterns for each type of imaging modality. Indeed, the analysis of images at the voxel level could enable the identification of pathological spatial changes that cannot be detected at a lower resolution.

A second problem concerns the limited interpretation that can be made from disease progression models. Although current methods enable to describe the evolution of multiple biomarkers during AD, they don't provide any information about the dynamical interplay across them. However, investigating the interactions between the biomarkers dynamics is essential to understand how the different pathological processes affect each other. For instance, this could allow to determine the exact effect of amyloid deposition on the disease course. Moreover, the presented approaches allow to perform individual disease staging, which can prove particularly useful to monitor the effect of a treatment during clinical trials, but they do not enable to predict the effect of a treatment on the disease evolution. Such capability could be used to assess the efficacy of therapeutic intervention *in silico*, which could help planning clinical trials.

Finally, a major challenge for disease progression models in the next years will be their implementation in a clinical setting. Indeed, until now these models were mostly

developed and applied on well-defined research datasets. As they hold promises to assist with diagnosis and prognosis via individual disease staging, it is essential to assess their generalization to independent cohorts in order to ultimately adopt them in clinical routine. However, this assessment of disease progression models on independent cohorts is a challenging task, which will require to deal with differences between datasets such as data acquisition, missing data or data heterogeneity.

1.4 Objectives and Organization of the Thesis

In this thesis, we develop novel computational methods for modeling the natural history of AD through the analysis of 3D multi-modal imaging data, while enabling the investigation of biomarkers dynamics and intervention strategies. We also aim to demonstrate the translational aspect of our approach by applying it on an independent cohort from a memory clinic. The manuscript is organized in the following way in order to address the previously stated limitations and challenges.

First, we present in **Chapter 2** a model of AD progression based on the analysis of clinical data and multi-modal volumetric images. This model aims at identifying the multiple spatio-temporal trajectories characterizing the pathological processes at stake during the disease progression. To do so, data is modelled via matrix factorization between spatio-temporal sources, which are constrained by clinically plausible statistical priors. Thanks to the proposed approach, we are able to disentangle different temporal processes of amyloid deposition, glucose hypometabolism and gray matter atrophy associated to specific regional patterns at a high spatial resolution.

Secondly, we introduce in **Chapter 3** SimulAD, a computational framework allowing to model AD progression through a dynamical system linking clinical and multi-modal imaging-derived data. Similarly to Chapter 2, this approach can be used to estimate a global model of progression for clinical and imaging markers. Most importantly, by formulating AD progression through dynamical systems, we can investigate the causal relationships between the processes driving the disease evolution, and simulate hypothetical progression scenarios. In particular, we study for the first time *in silico* the effect of anti-amyloid treatments on cognitive outcomes depending on intervention time and drug dosage.

Thirdly, we present in **Chapter 4** an application of SimulAD on an independent cohort from the Geneva Memory Center (GMC). We develop a specific pre-processing pipeline allowing to test our model on this external dataset. We propose a set experiments on the GMC cohort to assess the generalization of the model of disease progression, as well as its validity in terms of disease staging.

Finally, we conclude the manuscript in **Chapter 5** by summarizing the main contributions of this work. We also present potential applications of our methods and build upon their limitations to propose future research perspectives.

1.5 Publications

The contributions of the three central chapters of this manuscript led to the following publications and submissions in conferences and peer-reviewed journals.

- *Alzheimer's disease modelling and staging through independent Gaussian process analysis of spatio-temporal brain changes*, Clément Abi Nader, Nicholas Ayache, Philippe Robert, and Marco Lorenzi. *International Workshop on Machine Learning in Clinical Neuroimaging. Springer, 2018, pp. 3-14.*

- *Monotonic Gaussian Process for Spatio-Temporal Disease Progression Modeling in Brain Imaging Data*, Clément Abi Nader, Nicholas Ayache, Philippe Robert, and Marco Lorenzi. *NeuroImage* 2020.

- *Simulating the outcome of amyloid treatments in Alzheimer's disease from imaging and clinical data*, Clément Abi Nader, Nicholas Ayache, Giovanni B. Frisoni, Philippe Robert, and Marco Lorenzi. Accepted for publication at *Brain Communications* 2021.

- *SimulAD: A dynamical model for personalized simulation and disease staging in Alzheimer's disease*, Clément Abi Nader, Federica Ribaldi, Giovanni B. Frisoni, Valentina Garibotto, Philippe Robert, Nicholas Ayache and Marco Lorenzi. Submitted to *Neurobiology of Aging* 2021.

2

Monotonic Gaussian Process for Spatio-Temporal Disease Progression Modeling in Brain Imaging Data

Contents

2.1 Introduction **12**
2.2 Related Work in Neurodegeneration Modeling **15**
2.3 Methods **16**
2.3.1 Individual time-shift 16
2.3.2 Data modeling 16
2.3.3 Spatio-temporal processes 18
2.3.4 Sparsity 20
2.3.5 Variational inference 21
2.4 Experiments and Results **23**
2.4.1 Synthetic tests on spatio-temporal trajectory separation 23
2.4.2 Synthetic tests on trajectory separation and time reparameterization 24
2.4.3 Application to spatio-temporal brain progression modeling . . 28
2.5 Discussion **36**
2.6 Appendix **39**
2.6.1 Lower bound derivation 39
2.6.2 Kullback-Leibler computation 40
2.6.3 Experimental setting 40
2.6.4 Optimization procedure 41
2.6.5 Synthetic experiments using ICA, NMF and PCA 44
2.6.6 Model convergence 45
2.6.7 Application of ICA, NMF and PCA on brain imaging data . . . 45

In this chapter, we present a method to model Alzheimer's disease (AD) progression based on the analysis of high resolution multi-modal images. Our goal is to provide a thorough description of the topographic pathological patterns characterizing AD, as well as their associated dynamics of evolution. We postulate that each pathological process

can be potentially characterized by multiple spatio-temporal trajectories that we aim to identify. To do so, we propose a probabilistic generative framework in which the data is modelled via a matrix factorization between spatio-temporal sources. In order to ensure the plausibility of the estimated pathological progression, we introduce a set of clinically-inspired constraints for the temporal and spatial sources. The chapter is published in NeuroImage [Abi Nader, 2020] and is based on a previous work presented at MLCN 2018 [Abi Nader, 2018].

Abstract. We introduce a probabilistic generative model for disentangling spatio-temporal disease trajectories from collections of high-dimensional brain images. The model is based on spatio-temporal matrix factorization, where inference on the sources is constrained by anatomically plausible statistical priors. To model realistic trajectories, the temporal sources are defined as monotonic and time-reparameterized Gaussian Processes. To account for the non-stationarity of brain images, we model the spatial sources as sparse codes convolved at multiple scales. The method was tested on synthetic data favourably comparing with standard blind source separation approaches. The application on large-scale imaging data from a clinical study allows to disentangle differential temporal progression patterns mapping brain regions key to neurodegeneration, while revealing a disease-specific time scale associated to the clinical diagnosis.

2.1 Introduction

Neurodegenerative disorders such as Alzheimer's disease (AD) are characterized by morphological and molecular changes of the brain, ultimately leading to cognitive and behavioral decline. Clinicians suggested hypothetical models of the disease evolution, showing how different types of biomarkers interact and lead to the final dementia stage [Jack, 2010]. In the past years, efforts have been made in order to collect large databases of imaging and clinical measures, hoping to obtain more insights about the disease progression through data-driven models describing the trajectory of the disease over time. This kind of models are of critical importance for understanding the pathological progression in large scale data, and would represent a valuable reference for improving the individual diagnosis.

Current clinical trials in AD are based on longitudinal monitoring of biomarkers. Disease progression modelling aims at providing an interpretable way of modelling the evolution of biomarkers according to an estimated history of the pathology, as proposed for example in [Donohue, 2014; Fonteijn, 2012; Jedynak, 2012; Lorenzi, 2017; Young, 2014]. Therefore, disease progression models are promising methods for automatically staging

patients, and quantifying their progression with respect to the underlying model of the pathology. These approaches entail a great potential for automatic stratification of individuals based on their estimated stage and progression speed, and for assessment of efficacy of disease modifying drugs. Within this context, we propose a spatio-temporal generative model of disease progression, aimed at disentangling and quantifying the independent dynamics of changes observed in datasets of multi-modal data. With this term we indicate data acquired via different imaging modalities such as Magnetic Resonance Imaging (MRI) or Positron-Emission Tomography (PET), as well as non-imaging data such as clinical scores assessed by physicians. Moreover, we aim at automatically inferring the disease severity of a patient with respect to the estimated trajectory. Defining such a disease progression model raises a number of methodological challenges.

AD spreads over decades with a temporal mismatch between the onset of the disease and the moment where the clinical symptoms appear. Either age of diagnosis, or the chronological age, are therefore not suitable as a temporal reference to describe the disease progression in time. Moreover, as the follow-up of patients doesn't exceed a few years, the development of a model of long-term pathological changes requires to integrate cross-sectional data from different individuals, in order to consider a longer period of time. In virtue of the lack of a well defined temporal reference, observations from different individuals are characterized by large and unknown variability in the onset and speed of the disease. It is therefore necessary to account for a time-reparameterization function, mapping each individuals' observations to a common temporal axis associated to the absolute disease trajectory [Jedynak, 2012; Schiratti, 2015]. This would allow to estimate an absolute time-reference related to the natural history of the pathology.

The analysis of MRI and PET data, requires to account for spatio-temporally correlated features (voxels, i.e. volumetric pixels) defined over arrays of more than a million entries. The development of inference schemes jointly considering these correlation properties thus raises scalability issues, especially when accounting for the non-stationarity of the image signal. Furthermore, the brain regions involved in AD exhibit various dynamics in time, and evolve at different speed [Whitwell, 2010]. From a modeling perspective, accounting for differential trajectories over space and time raises the problem of source identification and separation. This issue has been widely addressed in neuroimaging via Independent Component Analysis (ICA) [Comon, 1994], especially on functional MRI (fMRI) data [Calhoun, 2009]. Nevertheless, while fMRI time-series are usually defined over a few hundreds of time points acquired per subject, our problem consists in jointly analyzing short-term and cross-sectional data observations with respect to an unknown time-line. This problem cannot be tackled with standard ICA, as time is generally an independent variable on which inference is not required. Moreover, ICA retrieves spatial sources based on the assumption of statistical independence. This assumption does not necessarily lead to clinically interpretable findings. Indeed, dependency across temporal

patterns can be still highly relevant to the pathology, for example when modeling temporal delay across similar sources.

The problem of providing a realistic description of the biological processes is critical when analyzing biomedical data, such as medical images. For example, to describe a plausible evolution of AD from normal to pathological stages, smoothness and monotonicity are commonly assumed for the temporal sources. It is also necessary to account for the non-stationarity of changes affecting the brain from global to localized spatio-temporal processes. As a result, spatial sources need to account for different resolutions at which these changes take place. While several multi-scale analysis approaches have been proposed to model spatio-temporal signals [Mallat, 1989; Bullmore, 2004; Hackmack, 2012], extending this type of methods to the high-dimension of medical images is generally not trivial due to scalability issues. Finally, the noisy nature of medical images, along with the large signal variability across observations, requires a modeling framework robust to bias and noise.

In this work, we propose to jointly address these issues within a Bayesian framework for the spatio-temporal analysis of large-scale collections of multi-modal brain data. We show that this framework allows us to naturally encode plausibility constraints through clinically-inspired priors, while accounting for the uncertainty of the temporal profiles and brain structures we wish to estimate. Similarly to the ICA setting, we formulate the problem of trajectory modeling through matrix factorization across temporal and spatial sources. This is done for each modality by inferring their specific spatio-temporal sources. To promote smoothness in time and avoid any unnecessary hypothesis on the temporal trajectories, we rely on non-parametric modeling based on Gaussian Process (GP). We account for a plausible evolution from healthy to pathological stages thanks to a monotonicity constraint applied on the GP. Moreover, individuals' observations are temporally re-aligned on a common scale via a time-warping function. In case of imaging data, to model the non-stationarity of the spatial signal, the spatial sources are defined as sparse activation maps convolved at different scales. We show that our framework can be efficiently optimized through stochastic variational inference, allowing to exploit automatic differentiation and GPU support to speed up computations.

This work is organized as follows: Section 2.2 analyzes related work on spatio-temporal modeling of neurodegeneration, while Section 2.3 details our method. In Section 2.4 we present experiments on synthetic data in which we compare our model to standard blind source separation approaches. We finally provide a demonstration of our method on the modeling of imaging data from a large scale clinical study. Prospects for future work and conclusions are drawn in section 2.5. Additional derivations and experiments are provided in Appendix 2.6.

2.2 Related Work in Neurodegeneration Modeling

To deal with the uncertainty of the time-line of neurodegenerative pathologies, the concept of time-reparameterization of imaging-derived features has been used in several works. The underlying principle consists in estimating an absolute time-scale of disease progression by temporally re-aligning data from different subjects. For instance, in [Young, 2015] the time-evolution was approximated as a sequence of events which need to be re-ordered for each patient. This approach thus considers the evolution of neurodegenerative diseases as a collection of transitions between discrete stages. This hypothesis is however limiting, as it doesn't reflect the continuity of changes affecting the brain along the course of the pathology.

To address this limitation, we rely on a continuous parameterization of the time-axis as in [Lorenzi, 2017; Donohue, 2014]. In particular, individuals' observations are time-realigned on a common temporal scale via a time-warping function. Using a set of relevant scalar biomarkers, this kind of approach allows to learn a time-scale describing the pathology evolution, and to estimate a data-driven time-line markedly correlated with the decline of cognitive abilities. Similarly, in [Bilgel, 2015] a disease progression score was estimated using biomarkers from molecular imaging. These methods are however based on the analysis of low-dimensional measures, such as collections of clinical variables. Therefore, they do not allow to scale to the high dimension of multi-modal medical images. Our work tackles this shortcoming thanks to a scalable inference scheme based on stochastic variational inference.

Concerning the spatio-temporal representation of neurodegeneration, a mixed-effect model was proposed by [Koval, 2017] to learn an average spatio-temporal trajectory of brain evolution on cortical thickness data. The fixed-effect describes the average trajectory, while random effects are estimated through individual spatio-temporal warping functions, modeling how each subject differs from the global progression. Still, the extension of this approach to image volumes raises scalability issues. It has also to be noted that, to allow computational tractability, the brain evolution was assumed to be stationary both in space and time, thus limiting the ability of the model to disentangle the multiple dynamics of the brain structures involved in AD.

An attempt to source separation is proposed in [Marinescu, 2019a], through the decomposition of cortical thickness measurements as a mixture of spatio-temporal processes. This is performed by associating to each cortical vertex a temporal progression modeled by a sigmoid function, which may be however too simplistic to describe the progression of AD temporal processes. We propose to overcome this issue by non-parametric modeling of the temporal sources through GPs. Moreover, the model in [Marinescu, 2019a] is lacking of an explicit vertex-wise correlation model, as it only assumes correlation

between clustering parameters at the resolution of the mesh graph. For this reason, it may still be sensitive to spatial variation at different scales and noise. We address this problem by modeling the spatial sources through convolution of sparse maps at multiple resolutions, allowing to deal with signal non-stationarity and robustness to noise.

2.3 Methods

In the following sections a matrix will be denoted by an uppercase letter $\boldsymbol{X}$, its n-th row will be given by $\boldsymbol{X}_{n:}$ and its n-th column by $\boldsymbol{X}_{:n}$. A column vector will be denoted by a lowercase letter $\boldsymbol{x}$. Subscript indices will be used to index the elements of matrices, vectors or sets of scalars. Superscipt indices will allow to index the blocks of block diagonal matrices.

2.3.1 Individual time-shift

To account for the uncertainty of the time-line of individual measurements, we assume that the observations are defined with respect to an absolute temporal reference τ. This is performed through a time-warping function $t_p = \boldsymbol{f}_p(\tau)$, that models the individual time-reparameterization. We choose an additive parameterization such that:

$$\boldsymbol{f}_p(\tau) = \tau + \delta_p. \tag{2.1}$$

Within this setting the individual time-shift δ_p encodes the temporal position of subject p, which in our application can be interpreted as the disease stage of subject p with respect to the long-term disease trajectory. We denote by $\boldsymbol{\delta} = \{\delta_p\}_{p=0}^{P}$ the set of time-shift parameters.

2.3.2 Data modeling

We represent the spatio-temporal data $\boldsymbol{D}$ by a block diagonal matrix in which we differentiate two main blocks $\boldsymbol{Y}$ and $\boldsymbol{V}$ as illustrated in Figure 2.1. Each sub-block $\boldsymbol{Y}^m$ is a matrix containing the data represented by one of the M imaging modalities we wish to consider. These matrices have dimensions $P \times F_m$, where P denotes the number of subjects and F_m the number of imaging features for modality m, which in our case is the number of voxels. The matrix $\boldsymbol{V}$ accounts for non-imaging or scalar data such as clinical scores and has dimensions $P \times C$, where C is the number of scalar features considered. We postulate a generative model and decompose the data as shown in Figure 2.1.

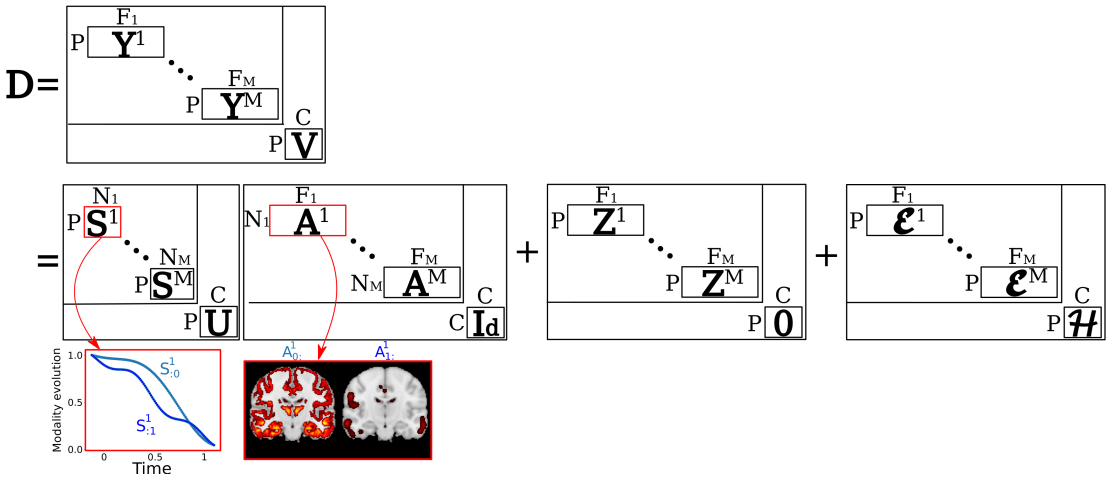

Figure 2.1: Spatio-temporal decomposition of each data block. A data matrix composed by M imaging modalities is decomposed as the product of monotonic temporal sources $\boldsymbol{S}^m$ and corresponding activation maps $\boldsymbol{A}^m$. Monotonic sources are also used to model the scalar biomarkers $\boldsymbol{V}$, while we assume additive constant terms $\boldsymbol{Z}^m$, and noise $\boldsymbol{\mathcal{E}}^m$.

For each sub-block $\boldsymbol{Y}^m$, the data is factorized in a set of N_m spatio-temporal sources $\boldsymbol{Y}^m = \boldsymbol{S}^m\boldsymbol{A}^m$. The columns of the matrix $\boldsymbol{S}^m$ describe the non-linear temporal evolution of the corresponding spatial maps contained in the rows of $\boldsymbol{A}^m$. Therefore, their product represents the voxel-wise linear combination of the spatial maps modulated by the corresponding temporal sources. The subjects share the same set of temporal sources across $\boldsymbol{S}^1, .., \boldsymbol{S}^M$, as these sources describe the temporal evolution of the group-wise images through the regression problem specified in Figure 2.1. The data in matrix $\boldsymbol{V}$ is modelled by a matrix $\boldsymbol{U}$ whose columns depict the temporal trajectories of the different scalar scores. In the case of imaging data, we also consider a constant term modeling brain areas which don't exhibit any intensity changes over time. This is done by including constant matrix terms $\boldsymbol{Z}^m$ that we need to estimate. We assume for a given modality m that the vectors $\boldsymbol{Z}^m_{p:}$ are common to every subjects. Finally, for each modality m, scalar score c, and subject p, we assume Gaussian observational noise $\boldsymbol{\mathcal{E}}^m_{p:} \sim \mathcal{N}(\boldsymbol{0}, \sigma_m^2\boldsymbol{I})$, and $\boldsymbol{\mathcal{H}}_{p,c} \sim \mathcal{N}(0, \nu_c^2)$ for respectively imaging and scalar information.

Therefore, if we consider the data from modality m and scalar c of patient p observed at time $\boldsymbol{f}_p(\tau)$ we have:

$$\begin{aligned}\boldsymbol{Y}^m_{p:}(\boldsymbol{f}_p(\tau), \theta_m, \psi_m) &= \boldsymbol{S}^m_{p:}(\boldsymbol{f}_p(\tau), \theta_m)\boldsymbol{A}^m(\psi_m) + \boldsymbol{Z}^m_{p:} + \boldsymbol{\mathcal{E}}^m_{p:},\\ \boldsymbol{V}_{p,c}(\boldsymbol{f}_p(\tau), \theta_c) &= \boldsymbol{U}_{p,c}(\boldsymbol{f}_p(\tau), \theta_c) + \boldsymbol{\mathcal{H}}_{p,c}.\end{aligned} \tag{2.2}$$

We denote by θ_m and θ_c the temporal parameters related respectively to the modality m and scalar feature c, while ψ_m represents the set of spatial parameters of modality m. We

assume conditional independence across modalities and scalar scores given the time-shift information:

$$p(\boldsymbol{Y}, \boldsymbol{V}|\boldsymbol{A}, \boldsymbol{S}, \boldsymbol{Z}, \boldsymbol{U}, \boldsymbol{\delta}, \sigma, \nu) = \Big(\prod_m p(\boldsymbol{Y}^m|\boldsymbol{A}^m, \boldsymbol{S}^m, \boldsymbol{Z}^m, \boldsymbol{\delta}, \sigma_m)\Big)\Big(\prod_c p(\boldsymbol{V}_{:c}|\boldsymbol{U}_{:c}, \boldsymbol{\delta}, \nu_c)\Big). \tag{2.3}$$

Relying on classical regression formulation, we assume exchangeability across subjects allowing us to derive the data likelihood for a given modality m. According to the generative model we can write:

$$\begin{aligned} p(\boldsymbol{Y}^m|\boldsymbol{A}^m, \boldsymbol{S}^m, \boldsymbol{Z}^m, \boldsymbol{\delta}, \sigma_m) = \Big(\prod_p \frac{1}{(2\pi\sigma_m^2)^{\frac{F_m}{2}}} \exp(-\frac{1}{2\sigma_m^2}||\boldsymbol{Y}^m_{p:}(\boldsymbol{f}_p(\tau), \theta_m, \psi_m) \\ - \boldsymbol{S}^m_{p:}(\boldsymbol{f}_p(\tau), \theta_m)\boldsymbol{A}^m(\psi_m) - \boldsymbol{Z}^m_{p:}||^2)\Big). \end{aligned} \tag{2.4}$$

Naturally, a similar equation holds for $p(\boldsymbol{V}_{:c}|\boldsymbol{U}_{:c}, \boldsymbol{\delta}, \nu_c)$.

Within a Bayesian modeling framework, we wish to maximize the marginal log-likelihood $\log(p(\boldsymbol{Y}, \boldsymbol{V}|\boldsymbol{Z}, \boldsymbol{\delta}, \sigma, \nu))$, to obtain posterior distributions for the spatio-temporal processes. Since the derivation of this quantity in a closed-form is not possible, we tackle this optimization problem through stochastic variational inference. Based on this formulation, in what follows we illustrate our model by detailing the variational approximations imposed on the spatio-temporal sources, along with the priors and constraints we impose to represent the data (Sections 2.3.3 and 2.3.4). Finally, we detail the variational lower bound and optimization strategy in Section 2.3.5. For ease of notation we will drop the m and c indexes in Sections 2.3.3 and 2.3.4. As a result the matrix $\boldsymbol{S}$ will indistinctly refer to either any $\boldsymbol{S}^m$ or $\boldsymbol{U}$, while matrix $\boldsymbol{A}$ will refer to any $\boldsymbol{A}^m$, and $\boldsymbol{Y}$ to any $\boldsymbol{Y}^m$. For a given modality m, the number of patients P will be indexed by p, the number of sources N^m or the number of scalar scores C will be indexed by n, and finally f will index the number of imaging features F^m.

2.3.3 Spatio-temporal processes

Temporal sources

In order to flexibly account for non-linear temporal patterns, the temporal sources are encoded in a matrix $\boldsymbol{S}$ in which each column $\boldsymbol{S}_{:n}$ is a GP representing the evolution of source n and is independent from the other sources. To allow computational tractability within a variational setting, we rely on the GP approximation proposed in [Cutajar, 2017], through kernel approximation via random feature expansion [Rahimi, 2008]. Within this framework, a GP can be approximated as a Bayesian Neural Network with form: $\boldsymbol{S}_{:n}(\boldsymbol{t}) = \phi(\boldsymbol{t}(\boldsymbol{\omega}^n)^T)\boldsymbol{w}^n$. For example, in the case of the Radial Basis Function (RBF) co-

variance, $\boldsymbol{\omega}^n$ is a linear projection in the spectral domain. It is equipped with a Gaussian distributed prior $p(\boldsymbol{\omega}^n) \sim \mathcal{N}(\mathbf{0}, l_n \boldsymbol{I})$ with a zero-mean and a covariance parameterized by a scalar l_n, acting as the length-scale parameter of the RBF covariance. The non-linear basis functions activation is defined by setting $\phi(\cdot) = (\cos(\cdot), \sin(\cdot))$, while the regression parameter $\boldsymbol{w}^n$ is given with a standard normal prior. The GP inference problem can be conveniently performed by estimating approximated variational distributions for all the $\boldsymbol{\omega}^n$ and $\boldsymbol{w}^n$ (Section 2.3.5). We will respectively denote by $\boldsymbol{\Omega}$ and $\boldsymbol{W}$ the block diagonal matrices whose blocks are the $(\boldsymbol{\omega}^n)^T$ and $\boldsymbol{w}^n$. Considering the N temporal sources, we can write $p(\boldsymbol{\Omega}) = \prod_n p(\boldsymbol{\omega}^n)$ and $p(\boldsymbol{W}) = \prod_n p(\boldsymbol{w}^n)$.

We wish also to account for a steady evolution of the temporal processes, hence constraining the temporal sources to monotonicity. This is relevant in the medical case, where one would like to model the steady progression of a disease from normal to pathological stages. In our case, we want to constrain the space of the temporal sources to the set of solutions $\mathcal{C}_n = \{\boldsymbol{S}_{:n}(\boldsymbol{t}) \mid \boldsymbol{S}'_{:n}(\boldsymbol{t}) \geq 0 \quad \forall\, \boldsymbol{t}\}$. This can be done done consistently within the regression setting of [Riihimäki, 2010], and in particular with the GP random feature expansion framework as shown in [Lorenzi, 2018]. In that work, the constraint is introduced as a second likelihood term on the temporal sources dynamics:

$$p(\mathcal{C}|\boldsymbol{S}', \gamma) = \prod_{p,n} (1 + \exp(-\gamma \boldsymbol{S}'_{p,n}(\boldsymbol{t})))^{-1}, \tag{2.5}$$

where $\boldsymbol{S}'$ contains every derivatives $\boldsymbol{S}'_{:n}$, γ controls the magnitude of the monotonicity constraint, and $\mathcal{C} = \bigcap_n \mathcal{C}_n$. According to [Lorenzi, 2018] this constraint can be specified through the parametric form for the derivative of each $\boldsymbol{S}_{:n}$:

$$\boldsymbol{S}'_{:n}(t) = \frac{d\phi(\boldsymbol{t}(\boldsymbol{\omega}^n)^T)}{d\boldsymbol{t}} \boldsymbol{w}^n. \tag{2.6}$$

This setting leads to an efficient scheme for estimating the temporal sources through stochastic variational inference (Section 2.3.5).

Spatial sources.

According to the model introduced in Section 2.3.2, each observation $\boldsymbol{Y}_{p:}$ is obtained as the linear combination at a specific time-point between the temporal and spatial sources. In order to deal with the multi-scale nature of the imaging signal, we propose to represent the spatial sources at multiple resolutions. To this end, we encode the spatial sources in a matrix $\boldsymbol{A}$ whose rows $\boldsymbol{A}_{n:}$ represent a specific source at a given scale. The scale is prescribed by a convolution operator $\boldsymbol{\Sigma}^n$, which is a applied to a map $\boldsymbol{B}_{n:}$ that we wish to infer. This problem can be specified by defining $\boldsymbol{A}_{n:} = \boldsymbol{B}_{n:}\boldsymbol{\Sigma}^n$, where $\boldsymbol{\Sigma}^n$ is an $F \times F$ Gaussian kernel matrix imposing a specific spatial resolution. The length-scale parameter

λ_n of the Gaussian kernel is fixed for each source, to force the model to pick details at that specific scale. Due to the high-dimension of the data we are modeling, performing stochastic variational inference in this setting raises scalability issues. For instance, if we assume a Gaussian distribution $\mathcal{N}(\mu_{\boldsymbol{B}_{n:}}, diag(\boldsymbol{\Lambda}))$ for $\boldsymbol{B}_{n:}$, the distribution of the spatial signal would be $p(\boldsymbol{A}_{n:}) \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{B}_{n:}}\boldsymbol{\Sigma}^n, \boldsymbol{\Sigma}^n diag(\boldsymbol{\Lambda})(\boldsymbol{\Sigma}^n)^T)$. As a result, sampling from $p(\boldsymbol{A}_{n:})$ is not computationally tractable due to the size of the covariance matrix, which prevents the use of standard inference schemes on $\boldsymbol{B}_{n:}$. This can be overcome thanks to the separability of the Gaussian convolution kernel [Marquand, 2014; Lorenzi, 2015b], according to which the 3D convolution matrix $\boldsymbol{\Sigma}^n$ can be decomposed into the Kronecker product of 1D matrices, $\boldsymbol{\Sigma}^n = \boldsymbol{\Sigma}^n_x \otimes \boldsymbol{\Sigma}^n_y \otimes \boldsymbol{\Sigma}^n_z$. This decomposition allows to efficiently perform standard operations such as matrix inversion, or matrix-vector multiplication [Saatçi, 2011]. Thanks to this choice, we recover tractability for the inference of $\boldsymbol{B}_{n:}$ through sampling, as required by stochastic inference methods [Kingma, 2013].

2.3.4 Sparsity

In order to detect specific brain areas involved in neurodegeneration, we propose to introduce a sparsity constraint on the maps (or codes) $\boldsymbol{B}_{n:}$. Consistently with our variational inference scheme, we induce sparsity via *Variational Dropout* as proposed in [Kingma, 2015b]. This approach leverages on an improper log-scale uniform prior $p(|\boldsymbol{B}_{n:}|) \propto \prod_f 1/|\boldsymbol{B}_{n,f}|$, along with an approximate posterior distribution:

$$q_1(\boldsymbol{B}) = \prod_{n=1}^{N} \mathcal{N}(\boldsymbol{M}_{n:}, diag(\alpha_{n,1}\boldsymbol{M}^2_{n,1}...\alpha_{n,F}\boldsymbol{M}^2_{n,F})). \tag{2.7}$$

In this formulation, the dropout parameter $\alpha_{n,f}$ is related to the individual dropout probability $p_{n,f}$ of each weight by $\alpha_{n,f} = p_{n,f}(1-p_{n,f})^{-1}$. When the parameter $\alpha_{n,f}$ exceeds a fixed threshold, the dropout probability $p_{n,f}$ is considered high enough to ignore the corresponding weight $\boldsymbol{M}_{n,f}$ by setting it to zero. However, this framework raises stability issues affecting the inference of the dropout parameters due to large-variance gradients, thus limiting $p_{n,f}$ to values smaller than 0.5. To tackle this problem, we leverage on the extension of *Variational Dropout* proposed in [Molchanov, 2017]. In this setting, the variance parameter is encoded in a new independent variable $\boldsymbol{P}_{n,f} = \alpha_{n,f}\boldsymbol{M}^2_{n,f}$, while the posterior distribution is optimized with respect to $(\boldsymbol{M}, \boldsymbol{P})$. Therefore, in order to minimize the cost function for large variance $\boldsymbol{P}_{n,f} \to \infty$ ($\alpha_{n,f} \to \infty$ i.e $p_{n,f} \to 1$), the value of the weight's magnitude must be controlled by setting to zero the corresponding parameter $\boldsymbol{M}_{n,f}$. As a result, by dropping out weights in the code, we sparsify the estimated spatial maps, thus better isolating relevant spatial sub-structures. Spatial correlations in the images are obtained thanks to the convolution operation detailed in Section 2.3.3.

2.3.5 Variational inference

We detailed in the previous sections the choices of priors and constraints that we apply to the spatio-temporal processes in order to plausibly model the data. To illustrate the overall formulation of the method, we provide in Figure 2.2 the graphical model over the M modalities in the case of imaging data. Naturally, this graph simplifies when we deal with scalar data as we don't need to account for any spatial dependence. To infer

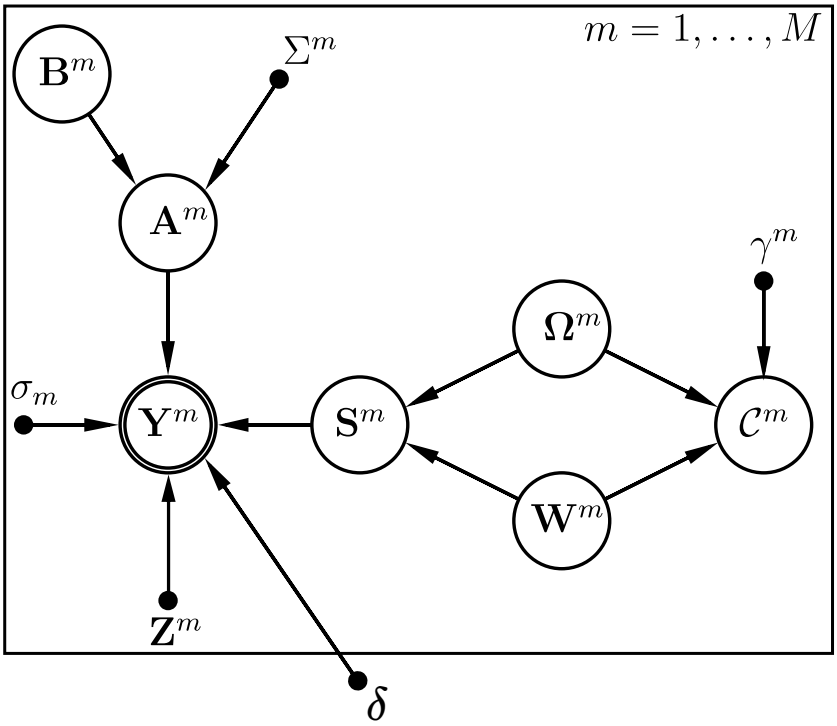

Figure 2.2: Graphical model for imaging data, $\boldsymbol{Y} = \{\boldsymbol{Y}^m\}$.

the time-shift parameter $\boldsymbol{\delta}$, the sets of parameters θ_m, θ_c, and ψ_m, as well as $\boldsymbol{Z}$, σ and ν, we need to jointly optimize the data evidence according to priors and constraints:

$$\log(p(\boldsymbol{Y}, \boldsymbol{V}, \mathcal{C} | \boldsymbol{Z}, \boldsymbol{\delta}, \sigma, \nu, \gamma)) = \sum_m \log(p(\boldsymbol{Y}^m, \mathcal{C}^m | \boldsymbol{Z}^m, \boldsymbol{\delta}, \sigma_m, \gamma_m)) + \sum_c \log(p(\boldsymbol{V}_{:c}, \mathcal{C}^c | \boldsymbol{\delta}, \nu_c, \gamma_c)). \tag{2.8}$$

We tackle the optimization of Equation (2.8) via stochastic variational inference. Following [Cutajar, 2017] and [Lorenzi, 2018] we introduce approximations, $q_2(\boldsymbol{\Omega}^m)$ and $q_3(\boldsymbol{W}^m)$ in addition to $q_1(\boldsymbol{B}^m)$ in order to derive a lower bound $\boldsymbol{\mathcal{L}}_m$ for each modality. We recall that the temporal trajectories $\boldsymbol{S}^m$ and $\boldsymbol{U}$ are treated similarly as described in Section 2.3.3. We also note that the choice of distributions q_1, q_2 and q_3 is the same across modalities, while their parameters will be inferred independently. This leads to:

$$
\begin{aligned}
\log(p(\boldsymbol{Y}^m, \mathcal{C}^m | \boldsymbol{Z}^m, \boldsymbol{\delta}, \sigma_m, \gamma_m)) \geqslant & \mathbb{E}_{q_1,q_2,q_3}[\log(p(\boldsymbol{Y}^m | \boldsymbol{B}^m, \boldsymbol{\Omega}^m, \boldsymbol{W}^m, \boldsymbol{Z}^m, \boldsymbol{\delta}, \sigma_m))] \\
& + \mathbb{E}_{q_2,q_3}[\log(p(\mathcal{C}^m | \boldsymbol{\Omega}^m, \boldsymbol{W}^m, \boldsymbol{\delta}, \gamma_m))] \\
& - \mathcal{D}[q_1(\boldsymbol{B}^m) || p(\boldsymbol{B}^m)] - \mathcal{D}[q_2(\boldsymbol{\Omega}^m) || p(\boldsymbol{\Omega}^m)] \\
& - \mathcal{D}[q_3(\boldsymbol{W}^m) || p(\boldsymbol{W}^m)], \\
\log(p(\boldsymbol{V}_{c:}, \mathcal{C}^c | \boldsymbol{\delta}, \nu_c, \gamma_c)) \geqslant & \mathbb{E}_{q_2,q_3}[\log(p(\boldsymbol{V}_{c:} | \boldsymbol{\Omega}^c, \boldsymbol{W}^c, \boldsymbol{\delta}, \sigma_c))] \\
& + \mathbb{E}_{q_2,q_3}[\log(p(\mathcal{C}^c | \boldsymbol{\Omega}^c, \boldsymbol{W}^c, \boldsymbol{\delta}, \gamma_c))] \\
& - \mathcal{D}[q_2(\boldsymbol{\Omega}^c) || p(\boldsymbol{\Omega}^c)] - \mathcal{D}[q_3(\boldsymbol{W}^c) || p(\boldsymbol{W}^c)]
\end{aligned}
\tag{2.9}
$$

Where $\mathcal{D}$ refers to the Kullback-Leibler (KL) divergence. Combining the lower bounds of the different modalities we obtain:

$$
\log(p(\boldsymbol{Y}, \boldsymbol{V}, \mathcal{C} | \boldsymbol{Z}, \boldsymbol{\delta}, \sigma, \nu, \gamma)) \geqslant \sum_m \boldsymbol{\mathcal{L}}_m + \sum_c \boldsymbol{\mathcal{L}}_c. \tag{2.10}
$$

A detailed derivation of the lower bound is given in Appendix 2.6.1. The approximated distributions $q_2(\boldsymbol{\Omega}^m)$ and $q_3(\boldsymbol{W}^m)$ are factorized across GPs such that:

$$
\begin{aligned}
q_2(\boldsymbol{\Omega}^m) &= \prod_{n=1}^{N_m} q_2(\boldsymbol{\omega}^n)^m = \prod_{n=1}^{N_m} \prod_{j=1}^{N_{rf}} \mathcal{N}(\boldsymbol{R}_{n,j}, \boldsymbol{Q}_{n,j}^2)^m, \\
q_3(\boldsymbol{W}^m) &= \prod_{n=1}^{N_m} q_3(\boldsymbol{w}^n)^m = \prod_{n=1}^{N_m} \prod_{j=1}^{N_{rf}} \mathcal{N}(\boldsymbol{T}_{n,j}, \boldsymbol{V}_{n,j}^2)^m,
\end{aligned}
\tag{2.11}
$$

where N_{rf} is the number of random features used for the projection in the spectral domain. Using Gaussian priors and approximations we introduced above, we can obtain a closed-form formula for the KL divergence. Moreover, the choice of prior and approximate posterior distribution for the maps of $\boldsymbol{B}^m$ leads to an approximation for the divergence $\mathcal{D}[q_1(\boldsymbol{B}^m) || p(\boldsymbol{B}^m)]$ detailed in [Molchanov, 2017]. This allows to analytically compute all the KL terms in our cost function. Formulas for the KL divergences are detailed in Appendix 2.6.2.

Finally, we optimize the individual time-shifts $\boldsymbol{\delta} = \{\delta_p\}_{p=0}^P$, $\boldsymbol{Z}$, $\sigma = \{\sigma_m\}_{m=1}^M$, $\nu = \{\nu_c\}_{c=1}^C$ as well as the overall sets of spatio-temporal parameters $\boldsymbol{\theta} = \{\theta_m\}_{m=1}^M \cup \{\theta_c\}_{c=1}^C$ and $\boldsymbol{\psi} = \{\psi_m\}_{m=1}^M$.

$$
\begin{aligned}
\boldsymbol{\theta} &= \{\boldsymbol{R}_{n:}^m, \boldsymbol{Q}_{n:}^m, \boldsymbol{T}_{n:}^m, \boldsymbol{V}_{n:}^m, l_n, n \in [1, N_m]\}_{m=1}^M \cup \{\boldsymbol{R}_{c:}, \boldsymbol{Q}_{c:}, \boldsymbol{T}_{c:}, \boldsymbol{V}_{c:}^c, l_c\}_{c=1}^C, \\
\boldsymbol{\psi} &= \{\boldsymbol{M}_{n:}^m, \boldsymbol{P}_{n:}^m, n \in [1, N_m]\}_{m=1}^M.
\end{aligned}
\tag{2.12}
$$

Following [Kingma, 2013] and using the reparameterization trick, we can efficiently sample from the approximated distributions q_1, q_2 and q_3 to compute the two expectation terms from Equation (2.9) for each modality. We chose to alternate the optimization

between the spatio-temporal parameters and the time-shift. We set γ_m to the minimum value that gives monotonic sources. This was done through multiple tests on data batches with different numbers of imaging features F_m and sources N_m. We empirically found that monotonicity was enforced when the magnitude of γ_m was in the order of $F_m \times N_m$. The threshold for the dropout probability above which we set a weight $\boldsymbol{B}^m_{n,f}$ to zero was fixed at 95% (i.e $\alpha = 19$), while the σ_m and ν_m were optimized during training along with the spatio-temporal parameters. The model is implemented and trained using the Pytorch library [Paszke, 2019]. The complete experimental setting is detailed in Appendix 2.6.3. We also provide a pseudo-code detailing the optimization procedure in Appendix 2.6.4. In the following sections we will refer to our method as Monotonic Gaussian Process Analysis (MGPA).

2.4 Experiments and Results

In this section we first benchmark MGPA on synthetic data to demonstrate its reconstruction and separation properties while comparing it to standard sources separation methods. We finally apply our model on a large set of medical data from a publicly available clinical study, demonstrating the ability of our method to retrieve spatio-temporal processes relevant to AD, along with a time-scale describing the course of the disease.

2.4.1 Synthetic tests on spatio-temporal trajectory separation

For the synthetic tests we considered the case where the data is associated to a single imaging modality only. We tested MGPA on synthetic data generated as a linear combination of temporal functions and 3D activation maps at prescribed resolutions. The goal was to assess the method's ability to identify the spatio-temporal sources underlying the data. We benchmarked our method with respect to ICA, Non-Negative Matrix Factorization (NMF), and Principal Component Analysis (PCA), which were applied from the standard implementation provided in the Scikit-Learn library [Pedregosa, 2011].

The benchmark was specified by defining a 10-folds validation setting, generating the data at each fold as a linear combination of temporal sources $\widetilde{\boldsymbol{S}}(\boldsymbol{t}) = [\widetilde{\boldsymbol{S}}_{:0}(\boldsymbol{t}), \widetilde{\boldsymbol{S}}_{:1}(\boldsymbol{t})]$, and spatial maps $\widetilde{\boldsymbol{A}} = [\widetilde{\boldsymbol{A}}_{0:}, \widetilde{\boldsymbol{A}}_{1:}]$. The data was defined as $\boldsymbol{Y}_{p:} = \widetilde{\boldsymbol{S}}_{p:}(\boldsymbol{t}_p)\widetilde{\boldsymbol{A}} + \boldsymbol{\mathcal{E}}_{p:}$ over 50 time points $\boldsymbol{t}_p$, where $\boldsymbol{t}_p$ was uniformly distributed in the range $[0, 0.7]$, and $\boldsymbol{\mathcal{E}}_{p:} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$. The temporal sources were specified as sigmoid functions $\widetilde{\boldsymbol{S}}_{p,i}(\boldsymbol{t}_p) = 1/(1+\exp(-\boldsymbol{t}_p+\alpha_i))$, while the spatial structures had dimensions $(30 \times 30 \times 30)$ such that $\widetilde{\boldsymbol{A}}_{i:} = \widetilde{\boldsymbol{B}}_{i:}\widetilde{\boldsymbol{\Sigma}}^i$. The $\widetilde{\boldsymbol{\Sigma}}^i$ were chosen as Gaussian convolution matrices with respective length-scale of $\lambda = 2$ mm and $\lambda = 1$ mm. The $\widetilde{\boldsymbol{B}}_{i:}$ were randomly sampled sparse 3D maps.

Variable selection. We applied our method by specifying an over-complete set of six

sources with respective spatial length-scale of $\lambda = \{2, 2, 1, 1, 0.5, 0.5\ mm\}$. Figure 2.3 shows an example of the sparse maps obtained for a specific fold. The model prunes the signal for most of the maps, while retaining two sparse maps, $\boldsymbol{B_{0:}}$ and $\boldsymbol{B_{4:}}$, whose length-scale are $\lambda = 2$ mm and $\lambda = 1$ mm, thus correctly estimating the right number of sources and their spatial resolution. As it can be qualitatively observed in Figure 2.3, we notice that the estimated sparse code convolved with a Gaussian kernel matrix with $\lambda = 1$ mm is closer to its ground truth than the one convolved with a length-scale $\lambda = 2$ mm. According to our tests, sparse codes associated to high resolution details (low λ) are indeed more identifiable. On the contrary, the identifiability of images obtained via a convolution operator with larger kernels (large λ) is lower, since these maps can be equivalently obtained through the convolution of different sparse codes.

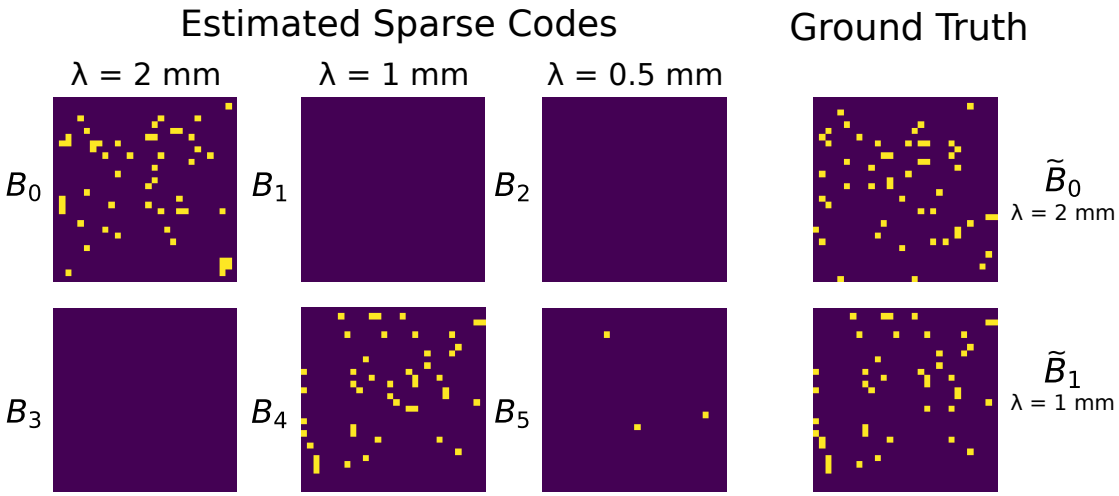

Figure 2.3: Slices extracted from the six sparse codes and the ground truth. Blue: Rejected points. Yellow: Retained points.

Sources separation. We observe in Table 2.1 that the lowest Mean-Squared Error (MSE) for the temporal sources reconstruction is obtained by MGPA, closely followed by ICA. Similarly, our model and ICA show the highest Structural Similarity (SSIM) score [Wang, 2004], which quantifies the image reconstruction accuracy with respect to the ground truth maps, while accounting for the inter-dependencies between neighbouring pixels. An example of image reconstruction from a sample fold is illustrated in Figure 2.4. In this standard benchmark, we note that MGPA leads to comparable results with respect to the state of the art. In the following section, we compare the models in the more challenging setting in which the time-line has to be estimated as well.

2.4.2 Synthetic tests on trajectory separation and time reparameterization

In this test, we modify the experimental benchmark by introducing a further element of variability associated to the time-axis. The temporal and spatial sources were modelled

Table 2.1: MSE and SSIM between respectively the ground truth temporal and spatial sources with respect to the ones estimated by the different methods.

	Temporal (MSE)	Spatial (SSIM)
MGPA	$(\mathbf{8 \pm 4}).\mathbf{10^{-5}}$	$\mathbf{98\% \pm 1}$
ICA	$(6 \pm 3).10^{-4}$	$97\% \pm 2$
NMF	$(3 \pm 2).10^{-2}$	$40\% \pm 17$
PCA	0.44 ± 10^{-3}	$15\% \pm 1$

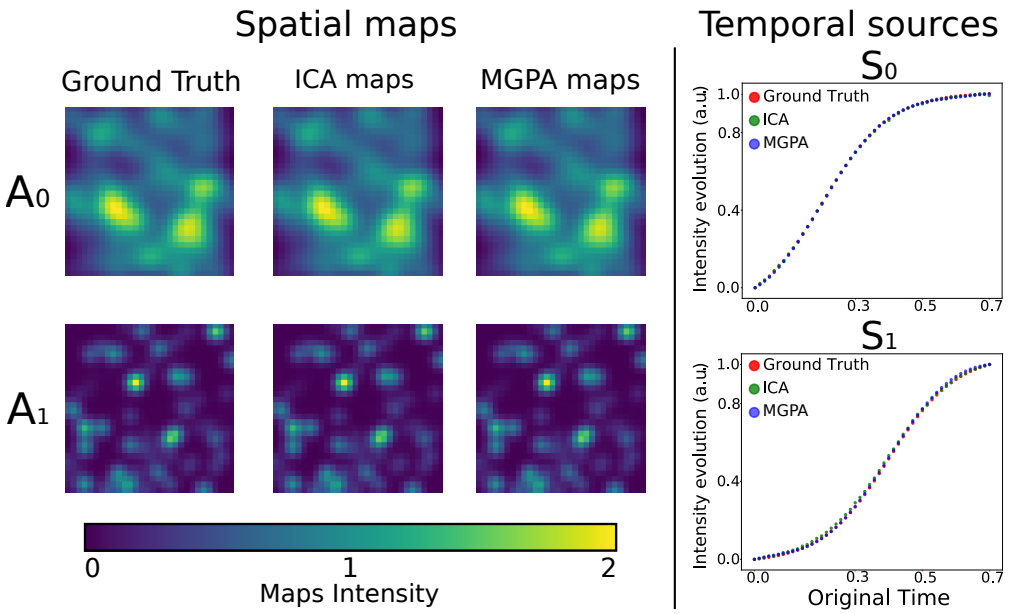

Figure 2.4: Spatio-temporal reconstruction when inference on the time-line is not required. Spatial maps: Sample slice from ground truth images (A_0 $\lambda = 2$ mm, A_1 $\lambda = 1$ mm), the maps estimated by ICA, and the ones estimated by MGPA. Temporal sources: Ground truth temporal sources (red) along with sources estimated by ICA (green) and MGPA (blue).

following the same procedure as in Section 2.4.1, however the observations were mixed along the temporal axis. To do so we generated longitudinal data as $\boldsymbol{Y}_{p,j,:} = \widetilde{\boldsymbol{S}}_{p:}(\boldsymbol{t})\widetilde{\boldsymbol{A}} + \boldsymbol{\mathcal{E}}_{j:}$, by sampling between 1 and 10 images per time-point and randomly re-arranging them along the time-axis (*cf.* time-shift t_p of each observation at initialization in Figures 2.5 and 2.6, panel "Time-Shift"). The goal was to assess the sources separation performances of MGPA when the time-line is unknown. The experiment was run on 10 folds and Figures 2.5 and 2.6 illustrate the sources estimation for two different folds. We present these two figures to demonstrate how the time-shift inference affects the temporal sources reconstruction. Since the model is agnostic of a time-scale, we note that the time-shift may have a different range than the original time-axis. However, its relative ordering should be consistent with the original time points. We fitted a linear regression model over the 10 folds between the original time and the estimated time-shift parameter, and obtained an average R^2 coefficient of 0.98 with a standard deviation of 0.005 (*cf.* Table 2.2). This is illustrated for two different folds in the Time-Shift panel of Figures 2.5 and

Table 2.2: MSE and SSIM between respectively the ground truth temporal and spatial sources with respect to the ones estimated by MGPA. R^2 coefficient of the linear regression between the original time-line and the estimated time-shift.

	Temporal (MSE)	Spatial (SSIM)	R^2
MGPA	$\mathbf{(2 \pm 0.8).10^{-2}}$	$\mathbf{95\% \pm 4}$	$\mathbf{0.98 \pm 0.005}$

2.6, where we observe a strong linear correlation with the original time-line, meaning that the algorithm correctly re-ordered the data with respect to the original time-axis. However, we notice in Table 2.2 that the MSE of the temporal sources significantly increased, due to the additional difficulty brought by the time-shift estimation. Indeed, in order to reconstruct the temporal signal we need to perfectly re-align hundreds of observations. This is the case in Figure 2.5 (optimal reconstruction result), where the time-shift is highly correlated with the original time-line, allowing to distinguish every single observation and reconstruct the original temporal profiles. Whereas in Figure 2.6 (sub-optimal reconstruction result), the estimated time-shift doesn't exhibit a perfect fit, and generally underestimates the time-reparameterization for the later and earlier time points. This is related to the challenging setting of reconstructing the time-line identified by the original temporal sources. Indeed, we observe that $\boldsymbol{S}_{:0}$ reaches a plateau for early time points, while $\boldsymbol{S}_{:1}$ is flat for later ones. This behaviour increases the difficulty of differentiating time points with low signal differences. As a result, it impacts the time-shift optimization and adds variability to the time-shift estimation performances, thus deteriorating the reconstruction of the temporal sources over the 10 folds compared to the previous benchmark. The spatial sources estimation remains comparable to the one without time-shift both quantitatively, with an average SSIM of 95%, and qualitatively, as shown in Figures 2.5 and 2.6. Within this setting, ICA, NMF and PCA poorly perform as they can't reconstruct the time-line. Results obtained using these three methods are provided in Appendix 2.6.5.

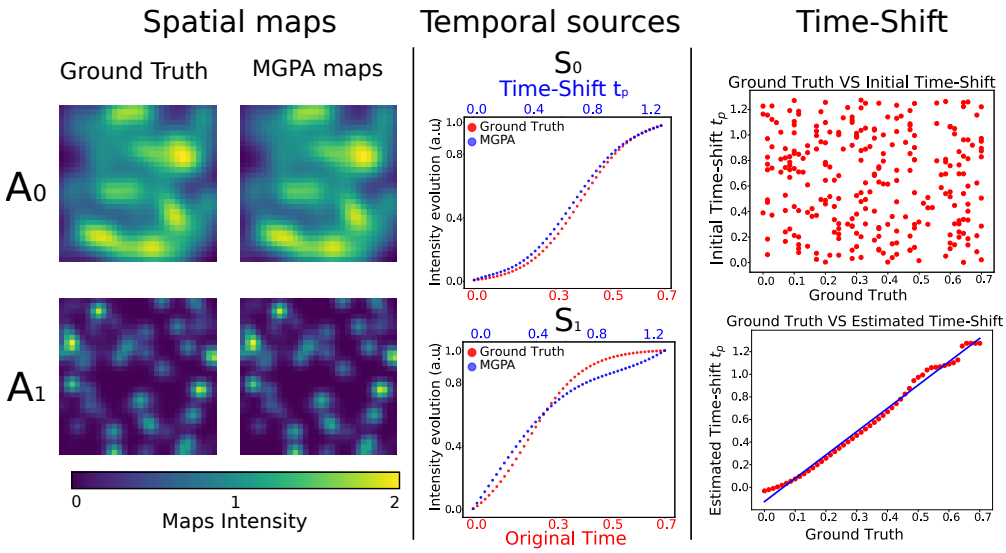

Figure 2.5: Spatio-temporal reconstruction when inference on the time-line is required. Optimal reconstruction result. Spatial maps: Sample slice from ground truth images (A_0 $\lambda = 2$ mm, A_1 $\lambda = 1$ mm) and estimated spatial sources. Temporal sources: In red the original temporal sources, in blue the estimated temporal sources. Time-Shift: Time-shift t_p of each image at initialization (top), and after estimation (bottom). In blue, linear fit with the ground truth.

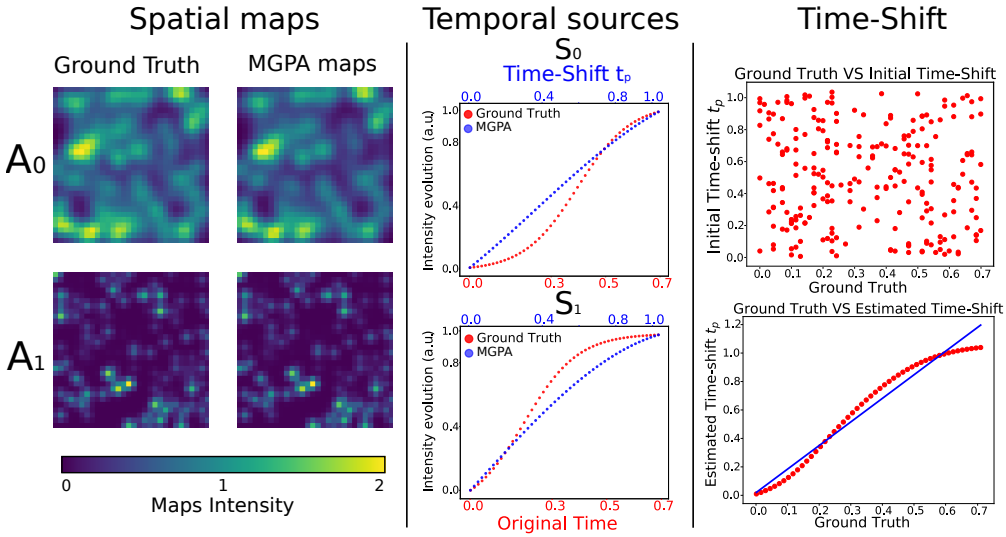

Figure 2.6: Spatio-temporal reconstruction when inference on the time-line is required. Sub-optimal reconstruction result. Spatial maps: Sample slice from ground truth images (A_0 $\lambda = 2$ mm, A_1 $\lambda = 1$ mm) and estimated spatial sources. Temporal sources: In red the original temporal sources, in blue the estimated temporal sources. Time-Shift: Time-shift t_p of each image at initialization (top), and after estimation (bottom). In blue, linear fit with the ground truth.

2.4.3 Application to spatio-temporal brain progression modeling

Data processing

Data used in the preparation of this work were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. For up-to-date information, see www.adni-info.org.

We selected a cohort of 544 amyloid positive subjects of the ADNI database composed of 103 controls (NL), 164 Mild Cognitive Impairment (MCI), 114 AD patients, 34 healthy individuals converted to MCI or to AD (NL converter) and 129 MCI converted to AD (MCI converter). The term amyloid positive refers to subjects whose amyloid level in the cerebrospinal fluid (CSF) is below the nominal cutoff of 192 pg/ml. Conversion to MCI or AD was determined using the last follow-up available information. We provide in Table 2.3 socio-demographic and clinical information across the different groups.

Table 2.3: Baseline socio-demographic and clinical information for study cohort. Average values and standard deviation in parenthesis. NL: normal individuals, NL converter: normal subjects who converted to MCI or to AD, MCI: mild cognitive impairment, MCI converter: MCI subjects who converted to AD, AD: Alzheimer's patients. ADAS13: Alzheimer's Disease Assessment Scale-cognitive subscale, 13 items. FAQ: Functional Assessment Questionnaire. FDG: (18)F-fluorodeoxyglucose Positron Emission Tomography (PET) imaging. AV45: (18)F-florbetapir Amyloid PET imaging.

Group	NL	NL converter	MCI	MCI converter	AD
N	103	34	164	129	114
Age	73 (6)	78 (5)	73 (7)	73 (7)	74 (8)
Education (yrs)	16.3 (3)	16 (3)	15.7 (3)	16 (3)	15.6 (3)
ADAS13	9.1 (4.4)	11.4 (4.3)	14.6 (5.5)	20.4 (6.5)	31.6 (8.5)
FAQ	0.3 (0.7)	0.2 (0.6)	1.9 (2.8)	5.0 (4.6)	13.5 (6.9)
Entorhinal (cm^3)	3.8 (0.5)	3.5 (0.5)	3.6 (0.6)	3.2 (0.7)	2.8 (0.6)
Hippocampus (cm^3)	7.4 (0.9)	6.9 (0.7)	6.9 (0.9)	6.4 (0.9)	5.9 (0.8)
Ventricles (cm^3)	31 (16)	42 (21)	39 (23)	40 (19)	48 (23)
Whole brain (cm^3)	1033 (104)	1019 (91)	1058 (103)	1037 (102)	1005 (115)
FDG	1.3 (0.1)	1.3 (0.1)	1.2 (0.1)	1.1 (0.1)	1.0 (0.1)
AV45	1.3 (0.2)	1.3 (0.1)	1.3 (0.2)	1.4 (0.2)	1.5 (0.2)

MRI, FDG-PET and AV45-PET of each individual were processed in order to obtain respectively, volumes of gray matter density, glucose uptake, and amyloid load in a standard anatomical space.

MRI processing protocol. Baseline MRI images were analyzed according to the SPM12

processing pipeline [Ashburner, 2000]. Each image was initially segmented into grey, white matter and CSF probabilistic maps. Grey matter images were used for the following analysis, normalized to a group-wise reference space via DARTEL [Ashburner, 2007], and modulated using the Jacobian determinant of the subject-to-template transformation. The subsequent modeling was carried out on the normalised images at the original spatial resolution.

PET processing protocol. Individuals' baseline PET images were initially affinely aligned to the corresponding MRI. After scaling the intensities to the cerebellum, the images were normalized to the grey matter template obtained with DARTEL and smoothed with a FWHM parameter of 4.55.

Images have dimension $102 \times 130 \times 107$ before vectorization, leading to $1,418,820$ spatial features per patient. These spatial features represent for each voxel their gray matter concentration in the case of MRI images, their glucose metabolism for FDG-PET images, or their amyloid concentration for AV45-PET images. To exploit the ability of our model to automatically adapt to different spatial scales, we chose to keep the MRI images at their native resolution for the analysis, and thus do not perform additional smoohting to equalize to the PET FWHM. In addition to the imaging data of each patient, we also integrate the ADAS13 score assessed by clinicians. High values of this score indicate a decline of cognitive abilities. We consider three matrices $\boldsymbol{Y}^{MRI}$, $\boldsymbol{Y}^{FDG}$, and $\boldsymbol{Y}^{AV45}$ of dimension $(543 \times 1,418,820)$ containing the images of all the subjects, and a matrix $\boldsymbol{V}$ of dimension (543×1) containing their ADAS13 score. From now on we will refer to the data as the block diagonal matrix containing the four matrices $\boldsymbol{Y}^{MRI}$, $\boldsymbol{Y}^{FDG}$, $\boldsymbol{Y}^{AV45}$, and $\boldsymbol{V}$ as described in Section 2.3.2. We note that the analysis is performed by only considering a single scan per imaging modality and ADAS13 score for each patient. Therefore, the temporal evolution has to be inferred solely through the analysis of relative differences between the brain morphologies, glucose metabolisms, amyloid concentrations and cognitive abilities across individuals.

Model specification

We aim at showing how MGPA applied on the data extracted from the ADNI cohort is able to temporally re-align patients in order to describe AD progression in a plausible way, while detecting relevant spatio-temporal processes at stake in AD. The model estimates AD progression by relying on MR, FDG-PET, AV45-PET scans and ADAS13 score of each patient. The temporal sources $\boldsymbol{S}^{MRI}$ and $\boldsymbol{S}^{FDG}$ associated respectively to the loss of gray matter, and to the decrease of glucose uptake, are enforced to be monotonically decreasing. On the contrary, the temporal sources $\boldsymbol{S}^{AV45}$ and $\boldsymbol{U}_{:ADAS13}$, modeling respectively the evolution of amyloid concentration, and ADAS13 score, are enforced

to be monotonically increasing. Since we don't consider any information about the disease stage of each individual before applying our method, all the observations are initialized at the same time reference $\tau = 0$. Therefore, as for the tests in Section 2.4.2, the time-shift reparameterization describes a relative re-ordering of the subjects not related to a specific time-unit. To decompose the imaging data we apply our model by specifying an over-complete basis of six sources with $\lambda = \{8, 8, 4, 4, 2, 2\ mm\}$, to cover both different scales and the associated variety of temporal evolution. Due to the high-dimension of the data matrix, the computations were parallelized over six GPUs, and the model required eighteen hours to complete the training. Details on the model convergence during training are provided in Appendix 2.6.6.

Estimated spatio-temporal brain dynamics

In Figure 2.7 we show the spatio-temporal processes retained by the model for each imaging modality. Interestingly, the model adapts to the spatial resolution of MRI and PET images. Indeed, we notice that the model accounts for the high-resolution of MRI images by retaining a source associated to the lowest length-scale ($\lambda = 2$ mm). Concerning PET data, we observe that the induced sparsity discards the highest resolution codes ($\lambda = 2$ mm) for both FDG and AV45, highlighting the ability of the model to adapt to the coarser resolution of the PET signal.

In the case of MRI data, two sources were retained at two different resolutions ($\lambda = 4$ mm and $\lambda = 2$ mm). Source $\boldsymbol{S}_4^{MRI}$ describes gray matter loss encompassing a large extent of the brain with a focus on cortical areas (see $\boldsymbol{A}_4^{MRI}$). We note that this map also targets subcortical areas such as the hippocampi, which are key regions of AD. Source $\boldsymbol{S}_2^{MRI}$ ($\lambda = 4$ mm) indicates a mild decrease of gray matter which accelerates in the latest stages of the disease, and targets the temporal poles (see $\boldsymbol{A}_2^{MRI}$). It is interesting to notice that this differential pattern of gray matter loss also affects the parahippocampal region, whose atrophy is known to be prominent in AD [Echavarri, 2011]. These results underline the complex evolution of brain atrophy, and the ability of the model to disentangle spatio-temporal processes mapping different regions involved in the pathology [Frisoni, 2010; Bateman, 2012]. Concerning the spatio-temporal processes extracted from the FDG-PET data, we see on Figure 2.7 that the model retained two sources at the coarsest resolutions ($\lambda = 8$ mm). Source $\boldsymbol{S}_1^{FDG}$ indicates a pattern of hypometabolism that tends to plateau and which involves most of the brain regions, thus describing a global effect of the pathology on the glucose uptake. Source $\boldsymbol{S}_0^{FDG}$ describes a linear pattern of hypometabolism targeting areas such as the precuneus and the parietal lobe, which are known to be strongly affected during the evolution of the disease [Brown, 2014]. Finally, the model extracted two spatio-temporal sources from the AV45-PET data at two different resolutions ($\lambda = 8$ mm and $\lambda = 4$ mm). We observe that source $\boldsymbol{S}_2^{AV45}$

highlights an increase of amyloid deposition mapping a large extent of the brain, such as the parietal and frontal lobes as well as temporal areas, thus concurring with clinical evidence [Rodrigue, 2009]. Similarly to the FDG-PET processes, we have a source $\boldsymbol{S}_0^{AV45}$ exhibiting a differential pattern of amyloid deposition targeting mostly frontal, temporal, occipital areas and precuneus.

The estimated spatio-temporal processes can be combined to obtain an estimated evolution $\boldsymbol{S}^m\boldsymbol{A}^m$ of the brain along the time-shift axis for each modality. In Figure 2.8, we show the ratio $|\boldsymbol{S}^m_{p:}\boldsymbol{A}^m - \boldsymbol{S}^m_{0:}\boldsymbol{A}^m|/\boldsymbol{S}^m_{0:}\boldsymbol{A}^m$ between the image predicted at four time-points t_p and the image predicted at t_0 for the three imaging modalities. This allows us to visualize the trajectory of a brain going from a healthy to a pathological state in terms of atrophy, glucose metabolism and amyloid load according to our model.

Finally, we also applied ICA, NMF and PCA on the ADNI data, showing that the associated results are characterized by poor interpretability and high variability. The complete experimental setting and results are detailed in Appendix 2.6.7.

Model Consistency

To verify the plausibility of the fitted model, we compare in Figure 2.9 the concentration predicted by the model and the raw concentration measures in different brain areas for the three imaging modalities. We observe a decrease of gray matter and glucose metabolism as we progress along the estimated time-line, allowing to relate large time-shift values to lower gray matter density and glucose uptake. Moreover, we notice the agreement between the predictions made by the model (in blue) and the raw concentration measures (in red). In the case of AV45 data there is only a mild increase of amyloid load according to the model, probably due to the fact that the subjects selected in the cohort are already amyloid positive. As a result, they already show a high baseline amyloid level concentration, close to plateau levels.

In Figure 2.10, we show the estimated GP $\boldsymbol{U}_{:ADAS13}$. We observe that the model is able to plausibly describe the evolution of this cognitive score, while demonstrating a larger variability than in the case of imaging modalities.

Plausibility with respect to clinical evidence

We assessed the clinical relevance of the estimated time-shift by relating it to independent medical information which were not included in the model during training. To this end, we compared the estimated time-shift to ADAS11, MMSE and FAQ scores. High values of

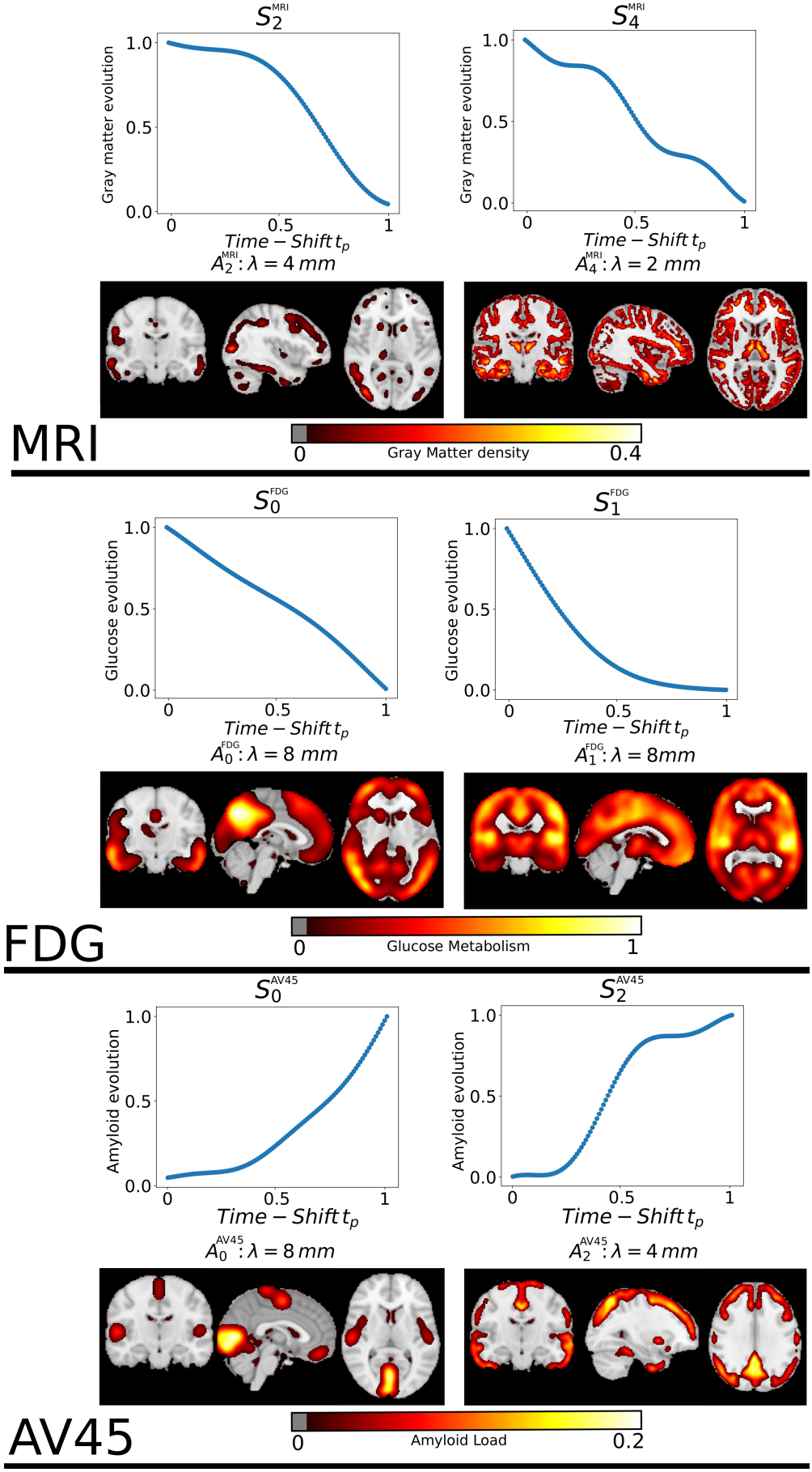

Figure 2.7: Estimated spatio-temporal processes for the three imaging modalities. The time-scale was re-scaled to the arbitrary range [0, 1].

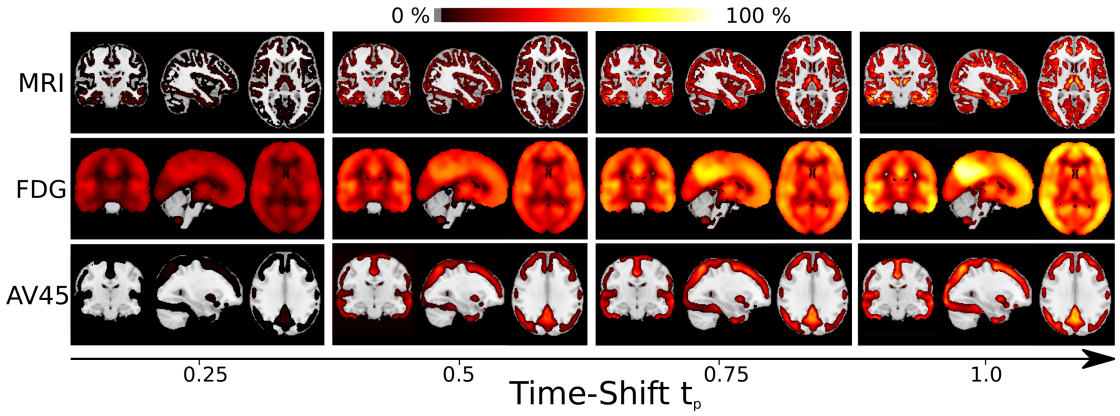

Figure 2.8: Ratio between the model prediction at time t_p and the prediction at t_0 for the three imaging modalities. The time-scale was re-scaled to the arbitrary range [0, 1].

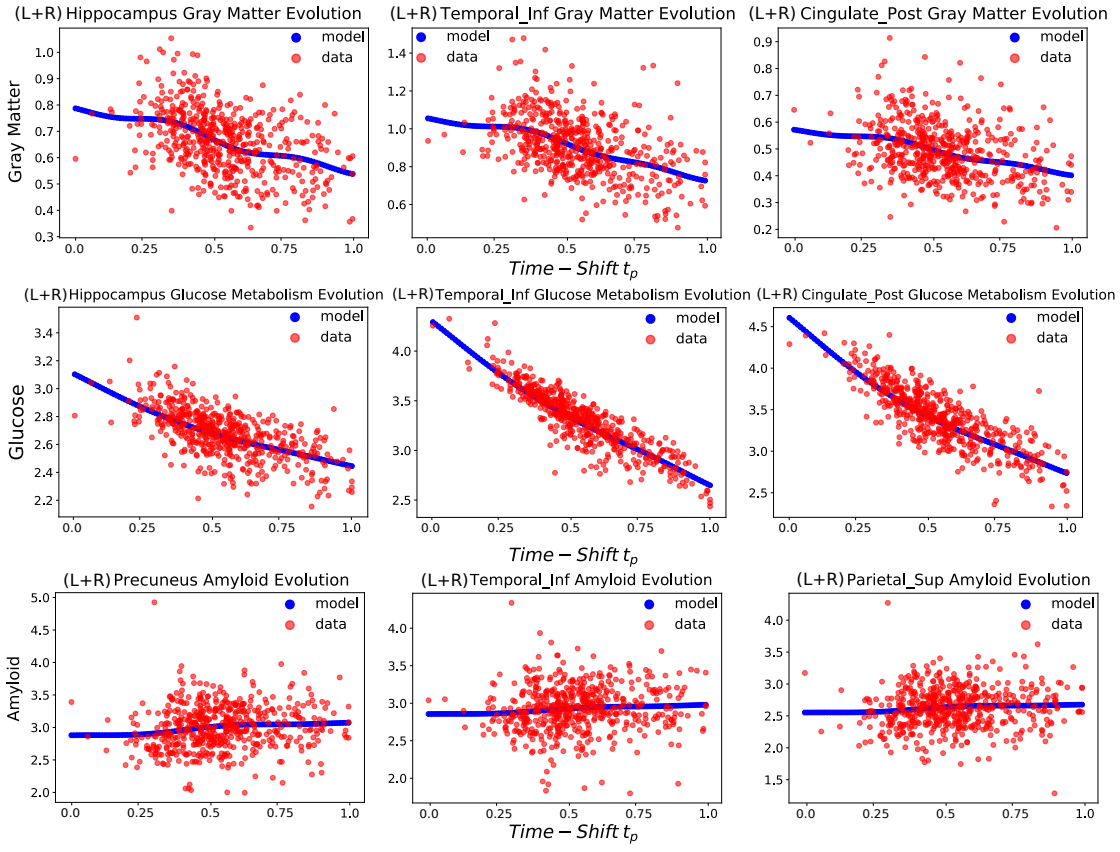

Figure 2.9: Model prediction averaged on specific brain areas (blue line), and observed values (red dots), along the estimated time-line for the three imaging modalities. L and R respectively stand for left and right. The time-scale was re-scaled to the arbitrary range [0, 1].

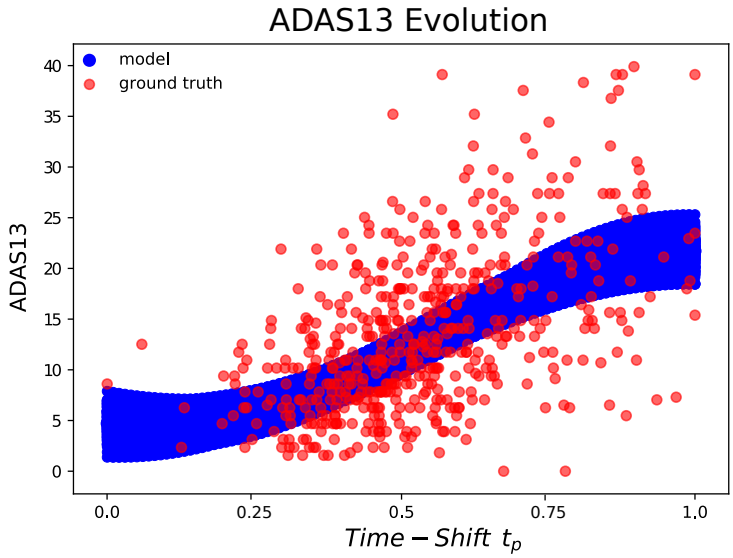

Figure 2.10: Model prediction of the ADAS13 score (blue line), and observed values (red dots) along the estimated time-line. The time-scale was re-scaled to the arbitrary range [0, 1].

ADAS11 and FAQ or low values of MMSE indicate a decline of performances. We show in Figure 2.11 that the estimated time-shift correlates with a decrease of cognitive and functional abilities. In particular, a cubic model slightly better describes the relationship between ADAS11 and the time-shift (according to BIC and AIC), with a significance for the cubic coefficient of $p = 0.04$. Concerning MMSE and FAQ, quadratic and linear models were almost equivalent; the significance of the linear coefficients was $p < 0.01$, while the quadratic coefficient was never significant. Pearson correlation coefficients for ADAS11, FAQ and MMSE were respectively of 0.49, 0.41, and -0.45, with corresponding p-values $p < 0.01$.

The box-plot of Figure 2.12 shows the time-shift distribution across clinical groups. We observe an increase of the estimated time-shift when going from healthy to pathological stages. The high uncertainty associated to the MCI group is due to the broad definition of this clinical category, which includes subjects not necessarily affected by dementia. We note that MCI subjects subsequently converted to AD (MCI converter) exhibit higher time-shift than the clinically stable MCI group, highlighting the ability of the model to differentiate between conversion status. A similar distinction can be noticed between NL and NL converter groups. We found significant differences between median time-shift for NL-NL converter, MCI-MCI converter and MCI converter-AD (comparisons $p < 0.01$, Figure 2.12). It is also important to recall that this result is obtained from the analysis of a single scan per imaging modality and ADAS13 score for each patient.

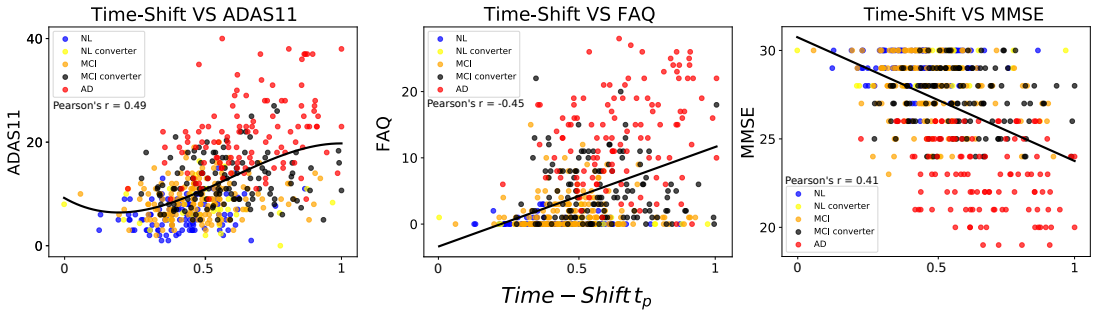

Figure 2.11: Evolution of the ADAS11 (left), FAQ (middle) and MMSE (right) along the estimated time-line. The time-scale was re-scaled to the arbitrary range [0, 1].

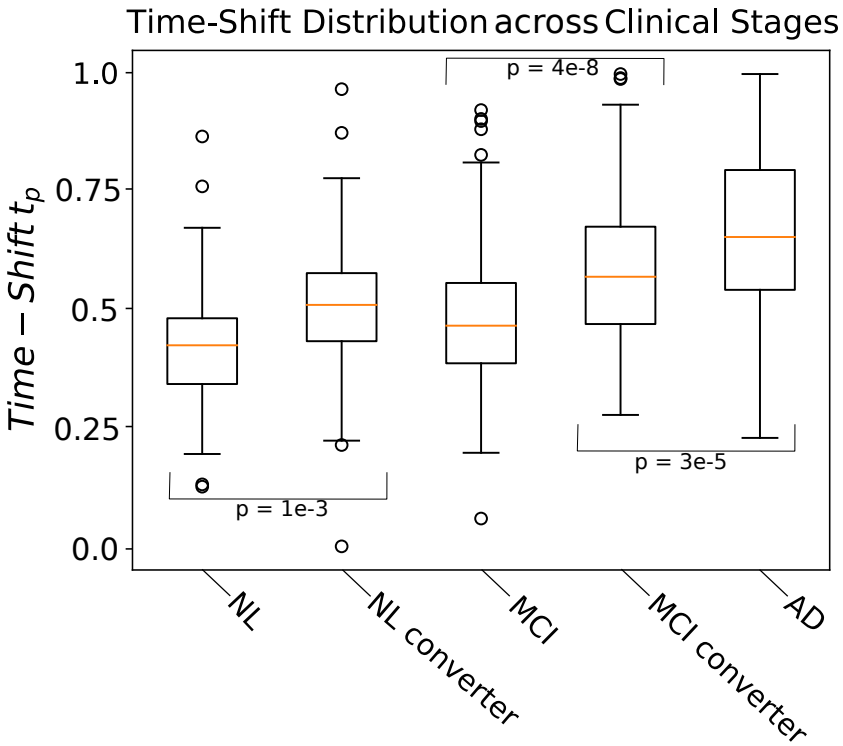

Figure 2.12: Distribution of the time-shift values over the different clinical stages. The time-scale was re-scaled to the arbitrary range [0, 1].

2.5 Discussion

We presented a generative approach to spatio-temporal disease progression modeling based on matrix factorization across temporal and spatial sources. The proposed application on a large set of medical images shows the ability of the model to disentangle relevant spatio-temporal processes at stake in AD, along with an estimated time-scale related to the disease evolution.

The model was compared to standard methods such as ICA, NMF and PCA since they perform blind source separation similarly to our method. This allowed us to demonstrate the advantages of building more complex approaches such as MGPA for the problem we tackle in this work. Concerning the comparison with the state of the art in disease progression modelling, to the best of our knowledge the two closest approaches are [Marinescu, 2019a] and [Koval, 2017]. However, these two methods are specifically designed for modelling data defined on brain surfaces. On the contrary, our method aims at progression modeling using full 3D volumetric information. The data dimension we tackle is thus an order of magnitude greater than the one of [Marinescu, 2019a] and [Koval, 2017], preventing these methods to scale to the spatial geometry of our data.

There are several avenues of improvement for the proposed approach. We found that the optimization is highly sensitive to the initialization of the spatial sources. This is typical of such complex non-convex problems, and requires further investigations to better control the algorithm convergence. More generally, the problem of source separation tackled in this work is intrinsically ill-posed, as the given data can be explained by several solutions. This was illustrated for example in our tests on synthetic data (Section 2.4.2), where the identification of the sources was more challenging in the case of coarse resolution codes and of flat temporal sources. We note however that this issue is general, and intrinsic to the problem of disease progression modeling.

Indeed, identifiability ultimately remains a critical issue when training the model. Concerning the spatio-temporal parameters, their number is extremely high due to the fact that we scale our method to 3D volumetric images. Estimating a single spatial source from a single modality requires to estimate the mean and variance of its sparse code, i.e $1,418,820 \times 2 = 2,837,640$ parameters. In practice, hypotheses are explicitly introduced to reduce the number of effective parameters. For instance, the convolution of the spatial maps using Gaussian kernels allows to enforce smoothness, and thus reduces the number of effective degrees of freedom via spatial correlation across the related parameters. This is equivalent to the regularization applied to image registration problems, in which the number of parameters is of the same order of magnitude than in our setting. Moreover, our sparsity constraint allows to sensibly reduce the number of parameters at test time. Indeed, after training, the sparse codes of the MRI sources have $2,213,359$ non-zero

elements instead of $17,025,840$, which amounts in 87% reduction in the number of parameters. In the case of the FDG-PET and AV45-PET sparse codes, the number of non-zero elements at test time is respectively of $9,023,695$ and $1,362,067$, which is equivalent to a reduction in the number of parameters of 53% and 92%. Nonetheless, this high number of parameters still remains a factor of potential convergence issues during the parameters estimation procedure. We present graphs in Appendix F showing the evolution of the different terms composing the cost function during training. These figures show convergence profiles typical of those obtained with stochastic variational inference schemes, such as with Variational Autoencoders or Bayesian Neural Networks. Moreover, the stability of the solution has been ensured through multiple runs of the model. Finally, as mentioned in Section 2.3.4, the *Variational Dropout* framework leads to stability issues affecting inference, which are mostly due to the use of an improper prior. This problem may motivate the identification of alternative ways to induce sparsity on the spatial maps.

In this work, we modeled the time-shift of each subject as a translation with respect to a common temporal reference. However, since pathological trajectories are different across individuals, it would be valuable to account for individual speed of progressions by introducing a scaling effect, as it has been proposed for example in [Koval, 2017; Schiratti, 2015]. This was not in the scope of the current study, as we focused on the analysis of cross-sectional data, thus having only one data point per subject. Therefore, one of the main extensions of this model will be the integration of longitudinal data for each individual, which will allow a more specific time-reparameterization.

Our noise model for the reconstruction problem of Equation 2.2 is homoscedastic and i.i.d. Gaussian with zero mean. For this reason, data variability for the entire image is encoded by the variance parameter of the Gaussian noise. Similarly as in standard regression problems, this modelling choice has been motivated to promote simplicity of the model and computational efficiency. However, around 40% of the values in the brain images do not provide relevant information as they represent zero and constant background areas. For this reason, during training, the model can perfectly fit this background and increases its confidence on the overall regression solution, thus lowering the value of the noise variance σ_m (*cf* Figure 2.9). This is in contrast to what we observe with the ADAS13 data (*cf* Figure 2.10), where the problem corresponds to standard univariate regression. A potential way to fix this issue could be to train the model only on non-zero image areas, or by implementing an heteroscedastic noise model. However, this latter solution may further increase the number of model parameters.

The modeling results are also sensitive to the specification of the spatio-temporal processes priors. In our case, the monotonicity constraint imposed to the GPs may be too restrictive to completely capture the complexity of the progression of neurodegeneration. From a clinical point of view, the model could also benefit from the integration of data

measuring the concentration of Tau protein via PET imaging, in order to quantify key neurobiological processes associated to AD [Kametani, 2018].

In order to guarantee that all the subjects belong to the same pathological trajectory due to AD, the model has only been applied to a cohort of amyloid positive subjects. However, this choice restricts the dynamics of evolution that we could estimate. Indeed, only considering these subjects narrows down the time-line of the pathology, as we study patients at potentially advanced disease stages. Therefore, it would be interesting in a future work to apply the model on a cohort including amyloid negative subjects, to model the brain dynamics over the whole disease natural history. This extension would require to define a proper methodology for disentangling sub-trajectories associated, for example with normal ageing and different pathological subtypes [Lorenzi, 2015a; Sivera, 2019; Young, 2018]. Moreover, we know that many patients diagnosed with AD can be associated to mixed pathologies such as vascular disease or Lewy bodies. Therefore, a potential clinical application of our method could be to investigate if the spatio-temporal dynamics estimated by MGPA are able to disentangle the contribution of each comorbidity.

Assessment of clinical plausibility of MGPA on the ADNI must be corroborated by further validation on independent datasets. Therefore, in a future work, we wish to validate the model on different cohorts to demonstrate its generalization properties. The validation step for each subject would be done by estimating the time-point minimizing the cost between the images of each tested individual, and the image progression model previously estimated on ADNI. The estimated time-shift would provide a measure of the pathological stage of the individual with respect to the modelled trajectory, and could be then compared with the clinical diagnosis of the subject, allowing to test the reliability of our model. This additional validation step could ultimately allow to use the model as a diagnostic instrument of AD. This validation would require an important effort in terms of data harmonisation across multiple cohorts, as well as in terms of clinical interpretation. For this reason, this work will be part of a subsequent publication.

We planned to release the source-code along with instructions in order for the model to be used by a large audience. It will be available as a complementary tool on the platform `http://gpprogressionmodel.inria.fr/`, which already offers a simple front-end to Gaussian Process Progression model.

2.6 Appendix

2.6.1 Lower bound derivation

In this Appendix, we detail the complete derivation of the lower bound.

$$\begin{aligned}\log(p(\boldsymbol{Y}^m, \mathcal{C}^m | \boldsymbol{Z}^m, \boldsymbol{\delta}, \sigma_m, \gamma_m)) &= \log\Big[\int p(\boldsymbol{Y}^m|\boldsymbol{B}^m, \boldsymbol{S}^m, \boldsymbol{Z}^m, \boldsymbol{\delta}, \sigma_m) p(\mathcal{C}|\frac{d\boldsymbol{S}^m}{d\boldsymbol{t}}, \boldsymbol{\delta}, \gamma_m) p(\boldsymbol{B}^m) \\ &\qquad p(\boldsymbol{S}^m, \frac{d\boldsymbol{S}^m}{d\boldsymbol{t}}|\boldsymbol{\delta}, \gamma) d\boldsymbol{B}^m d\boldsymbol{S}^m\Big] \\ &= \log\Big[\int p(\boldsymbol{Y}^m|\boldsymbol{B}^m, \boldsymbol{S}^m, \boldsymbol{Z}^m, \boldsymbol{\delta}, \sigma_m) p(\mathcal{C}|\frac{d\boldsymbol{S}^m}{d\boldsymbol{t}}, \boldsymbol{\delta}, \gamma_m) p(\boldsymbol{B}^m) \\ &\qquad p(\frac{d\boldsymbol{S}^m}{d\boldsymbol{t}}|\boldsymbol{S}^m, \boldsymbol{\delta}, \gamma) p(\boldsymbol{S}^m) d\boldsymbol{B}^m d\boldsymbol{S}^m\Big].\end{aligned} \tag{2.13}$$

By observing that $\frac{d\boldsymbol{S}^m}{d\boldsymbol{t}}$ is completely identified by $\boldsymbol{S}^m$, the equation can be written as:

$$\begin{aligned}\log(p(\boldsymbol{Y}^m, \mathcal{C}^m | \boldsymbol{Z}^m, \boldsymbol{\delta}, \sigma_m, \gamma_m)) &= \log\Big[\int p(\boldsymbol{Y}^m|\boldsymbol{B}^m, \boldsymbol{S}^m, \boldsymbol{Z}^m, \boldsymbol{\delta}, \sigma_m) p(\mathcal{C}|\frac{d\boldsymbol{S}^m}{d\boldsymbol{t}}, \boldsymbol{\delta}, \gamma_m) p(\boldsymbol{B}^m) \\ &\qquad p(\boldsymbol{S}^m) d\boldsymbol{B}^m d\boldsymbol{S}^m\Big].\end{aligned} \tag{2.14}$$

Similarly this derivation can be applied to $\log(p(\boldsymbol{V}_{:c}, \mathcal{C}^c | \boldsymbol{\delta}, \nu_c, \gamma_c))$.

$$\begin{aligned}\log(p(\boldsymbol{Y}^m, \mathcal{C}^m | \boldsymbol{Z}^m, \boldsymbol{\delta}, \sigma_m, \gamma_m)) &= \log\Big[\int p(\boldsymbol{Y}^m|\boldsymbol{B}^m, \boldsymbol{S}^m, \boldsymbol{Z}^m, \boldsymbol{\delta}, \sigma_m) p(\mathcal{C}|\frac{d\boldsymbol{S}^m}{d\boldsymbol{t}}, \boldsymbol{\delta}, \gamma_m) p(\boldsymbol{B}^m) \\ &\qquad p(\boldsymbol{S}^m) d\boldsymbol{B}^m d\boldsymbol{S}^m\Big] \\ &= \log\Big[\int p(\boldsymbol{Y}^m|\boldsymbol{B}^m, \boldsymbol{\Omega}^m, \boldsymbol{W}^m, \boldsymbol{Z}^m, \boldsymbol{\delta}, \sigma_m) p(\mathcal{C}|\boldsymbol{\Omega}^m, \boldsymbol{W}^m, \boldsymbol{\delta}, \gamma_m) p(\boldsymbol{B}^m) \\ &\qquad p(\boldsymbol{\Omega}^m) p(\boldsymbol{W}^m) d\boldsymbol{B}^m d\boldsymbol{\Omega}^m d\boldsymbol{W}^m\Big] \\ &= \log\Big[\int p(\boldsymbol{Y}^m|\boldsymbol{B}^m, \boldsymbol{\Omega}^m, \boldsymbol{W}^m, \boldsymbol{Z}^m, \boldsymbol{\delta}, \sigma_m) p(\mathcal{C}|\boldsymbol{\Omega}^m, \boldsymbol{W}^m, \boldsymbol{\delta}, \gamma_m) p(\boldsymbol{B}^m) \\ &\qquad p(\boldsymbol{\Omega}^m) p(\boldsymbol{W}^m) \frac{q_1(\boldsymbol{B}^m) q_2(\boldsymbol{\Omega}^m) q_3(\boldsymbol{W}^m)}{q_1(\boldsymbol{B}^m) q_2(\boldsymbol{\Omega}^m) q_3(\boldsymbol{W}^m)} d\boldsymbol{B}^m d\boldsymbol{\Omega}^m d\boldsymbol{W}^m\Big] \\ &= \log\Big[\mathbb{E}_{q_1,q_2,q_3} \frac{p(\boldsymbol{Y}^m|\boldsymbol{B}^m, \boldsymbol{\Omega}^m, \boldsymbol{W}^m, \boldsymbol{Z}^m, \boldsymbol{\delta}, \sigma_m) p(\mathcal{C}|\boldsymbol{\Omega}^m, \boldsymbol{W}^m, \boldsymbol{\delta}, \gamma_m)}{q_1(\boldsymbol{B}^m) q_2(\boldsymbol{\Omega}^m) q_3(\boldsymbol{W}^m)} \\ &\qquad \frac{p(\boldsymbol{B}^m) p(\boldsymbol{\Omega}^m) p(\boldsymbol{W}^m)}{q_1(\boldsymbol{B}^m) q_2(\boldsymbol{\Omega}^m) q_3(\boldsymbol{W}^m)}\Big] \\ &\geq \mathbb{E}_{q_1,q_2,q_3}\Big(\log\Big[\frac{p(\boldsymbol{Y}^m|\boldsymbol{B}^m, \boldsymbol{\Omega}^m, \boldsymbol{W}^m, \boldsymbol{Z}^m, \boldsymbol{\delta}, \sigma_m) p(\mathcal{C}|\boldsymbol{\Omega}^m, \boldsymbol{W}^m, \boldsymbol{\delta}, \gamma_m)}{q_1(\boldsymbol{B}^m) q_2(\boldsymbol{\Omega}^m) q_3(\boldsymbol{W}^m)} \\ &\qquad \frac{p(\boldsymbol{B}^m) p(\boldsymbol{\Omega}^m) p(\boldsymbol{W}^m)}{q_1(\boldsymbol{B}^m) q_2(\boldsymbol{\Omega}^m) q_3(\boldsymbol{W}^m)}\Big]\Big) \\ &= \mathbb{E}_{q_1,q_2,q_3}[\log(p(\boldsymbol{Y}^m|\boldsymbol{B}^m, \boldsymbol{\Omega}^m, \boldsymbol{W}^m, \boldsymbol{Z}^m, \boldsymbol{\delta}, \sigma_m))]\end{aligned}$$

$$
\begin{aligned}
&+ \mathbb{E}_{q_2,q_3}[\log(p(\mathcal{C}^m|\boldsymbol{\Omega}^m, \boldsymbol{W}^m, \boldsymbol{\delta}, \gamma_m))] \\
&- \mathcal{D}[q_1(\boldsymbol{B}^m)||p(\boldsymbol{B}^m)] - \mathcal{D}[q_2(\boldsymbol{\Omega}^m)||p(\boldsymbol{\Omega}^m)] \\
&- \mathcal{D}[q_3(\boldsymbol{W}^m)||p(\boldsymbol{W}^m)]. \qquad (2.15)
\end{aligned}
$$

This derivation gives us the lower bound $\boldsymbol{\mathcal{L}}_m$ of a given modality m. The same technique can be used to derive a lower bound for $\log(p(\boldsymbol{V}_{c:}, \mathcal{C}^c|\boldsymbol{\delta}, \nu_c, \gamma_c))$, and by summation over m and c we obtain the lower bound of Equation 2.10 for $\log(p(\boldsymbol{Y}, \boldsymbol{V}, \mathcal{C}|\boldsymbol{Z}, \boldsymbol{\delta}, \sigma, \nu, \gamma))$.

2.6.2 Kullback-Leibler computation

In this section we provide formulas for computing the three KL terms of the lower bound. The total KL divergences are:

$$
\begin{aligned}
\mathcal{D}[q_1(\boldsymbol{B})||p(\boldsymbol{B})] &= \sum_m \mathcal{D}[q_1(\boldsymbol{B}^m)||p(\boldsymbol{B}^m)], \\
\mathcal{D}[q_2(\boldsymbol{\Omega})||p(\boldsymbol{\Omega})] &= \sum_m \mathcal{D}[q_1(\boldsymbol{\Omega}^m)||p(\boldsymbol{\Omega}^m)] + \sum_c \mathcal{D}[q_1(\boldsymbol{\Omega}^c)||p(\boldsymbol{\Omega}^c)], \\
\mathcal{D}[q_3(\boldsymbol{W})||p(\boldsymbol{W})] &= \sum_m \mathcal{D}[q_3(\boldsymbol{W}^m)||p(\boldsymbol{W}^m)] + \sum_c \mathcal{D}[q_3(\boldsymbol{W}^c)||p(\boldsymbol{W}^c)].
\end{aligned} \qquad (2.16)
$$

For ease of notation we will drop the m and c indices and will give formulas for a single modality. In [Molchanov, 2017], authors provide an approximation of the KL for the maps $\boldsymbol{B}$:

$$
-\mathcal{D}[q_1(\boldsymbol{B})||p(\boldsymbol{B})] = \sum_{n,f} k_1 h(k_2 + k_3 \log(\alpha_{n,f})) - 0.5 \log(1 + \alpha_{n,f}^{-1}) - k_1, \qquad (2.17)
$$

where h is the sigmoid function and $k_1 = 0.63576,\ k_2 = 1.87320,\ k_3 = 1.48695$.
In the case of $\boldsymbol{\Omega}$ and $\boldsymbol{W}$, we've seen that they have Gaussian priors and approximations which are detailed in Sections 2.3.3 and 2.3.5. As a result we can obtain closed-form formulas for their KL, leading to:

$$
\begin{aligned}
\mathcal{D}[q_2(\boldsymbol{\Omega})|p(\boldsymbol{\Omega})] &= \frac{1}{2} \sum_{n,j} \boldsymbol{Q}_{n,j}^2 l_n + \boldsymbol{R}_{n,j}^2 l_n - 1 - \log(\boldsymbol{Q}_{n,j}^2 l_n), \\
\mathcal{D}[q_3(\boldsymbol{W})|p(\boldsymbol{W})] &= \frac{1}{2} \sum_{n,j} \boldsymbol{V}_{n,j}^2 + \boldsymbol{T}_{n,j}^2 - 1 - \log(\boldsymbol{V}_{n,j}^2).
\end{aligned} \qquad (2.18)
$$

By summation over the different modalities we finally obtain the total KL divergences.

2.6.3 Experimental setting

We provide in this Appendix details for the experiments on real data.

- The number of random features for the GP estimation was set to 10, as it was enough to recover the temporal sources in the synthetic experiments.

- The γ parameter controlling monotonicity was set to $\gamma_m = 10^7$ for each imaging modality ($F_m = 1,418,820$ imaging features and $N_m = 6$ sources) and $\gamma_c = 1$ for ADAS13 ($C_c = 1$ scalar feature).

- The lower bound was optimized using the ADAM optimizer [Kingma, 2015a].

- We used an alternate optimization scheme between the spatio-temporal parameters and the time-shift of [2000, 1000] iterations repeated 20 times, followed by 30000 iterations in which we only optimized the spatio-temporal parameters.

- The expectation terms in the lower bound were approximated using only one Monte-Carlo sample as proposed in [Kingma, 2013].

- The table below gives the learning rates (LR) of all the parameters of the model.

Table 2.4: Learning rates (LR) of the different parameters of the model.

	$\boldsymbol{\theta}$	$\boldsymbol{M}$	$\boldsymbol{P}$	$\boldsymbol{Z}$	$\sigma,\ \nu$	$\boldsymbol{\delta}$
lr	10^{-2}	10^{-3}	10^{-1}	10^{-1}	10^{-2}	10^{-4}

2.6.4 Optimization procedure

In this Appendix, we first provide a pseudo-code for sampling from a normal distribution using the reparameterization trick (see Algorithm 1). The second pseudo-code (Algorithm 2) details the steps to compute the lower bound $\boldsymbol{\mathcal{L}}_m$ for a given imaging modality m. We recall that we want to optimize the following sets of parameters (see Section 2.3.5): $\boldsymbol{\delta} = \{\delta_p\}_{p=0}^{P}$, $\boldsymbol{Z}$, $\sigma = \{\sigma_m\}_{m=1}^{M}$, $\nu = \{\nu_c\}_{c=1}^{C}$, $\boldsymbol{\theta} = \{\theta_m\}_{m=1}^{M} \cup \{\theta_c\}_{c=1}^{C}$, and $\boldsymbol{\psi} = \{\psi_m\}_{m=1}^{M}$. Where P is the number of subjects, M the number of imaging modalities, C the number of scalar features, and N_m the number of spatio-temporal sources for a given modality m.

$$\begin{aligned}\boldsymbol{\theta} &= \{\boldsymbol{R}^m_{n:}, \boldsymbol{Q}^m_{n:}, \boldsymbol{T}^m_{n:}, \boldsymbol{V}^m_{n:}, l_n, n \in [1, N_m]\}_{m=1}^{M} \cup \{\boldsymbol{R}_{c:}, \boldsymbol{Q}_{c:}, \boldsymbol{T}_{c:}, \boldsymbol{V}_{c:}, l_c, \}_{c=1}^{C}, \\ \boldsymbol{\psi} &= \{\boldsymbol{M}^m_{n:}, \boldsymbol{P}^m_{n:}, n \in [1, N_m]\}_{m=1}^{M}.\end{aligned} \tag{2.19}$$

Similarly to Algorithm 2, we can derive a function LOSS_SCALAR when dealing with scalar scores by removing the computations on the spatial sources. Finally the last

pseudo-code (Algorithm 3) details the model optimization. For sake of clarity we denote by $\mathbf{\Pi}$, the set of all the spatio-temporal parameters of the model.

$$\begin{aligned}\boldsymbol{\theta} &= \{\boldsymbol{R}^m_{n:}, \boldsymbol{Q}^m_{n:}, \boldsymbol{T}^m_{n:}, \boldsymbol{V}^m_{n:}, l_n, n \in [1, N_m]\}_{m=1}^M \cup \{\boldsymbol{R}_{c:}, \boldsymbol{Q}_{c:}, \boldsymbol{T}_{c:}, \boldsymbol{V}_{c:}, l_c, \}_{c=1}^C, \\ \boldsymbol{\psi} &= \{\boldsymbol{M}^m_{n:}, \boldsymbol{P}^m_{n:}, n \in [1, N_m]\}_{m=1}^M.\end{aligned} \tag{2.20}$$

Similarly to Algorithm 2, we can derive a function LOSS_SCALAR when dealing with scalar scores by removing the computations on the spatial sources. Finally the last pseudo-code (Algorithm 3) details the model optimization. For sake of clarity we denote by $\mathbf{\Pi}$, the set of all the spatio-temporal parameters of the model.

Algorithm 1 Sampling from $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ using the reparameterization trick.

function RT($\boldsymbol{\mu}, \boldsymbol{\Sigma}$)
  $\boldsymbol{\epsilon} \leftarrow$ random sample from $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$
  $\boldsymbol{z} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\epsilon}$ ▷ Gives one sample from $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
  Return $\boldsymbol{z}$
end function

Algorithm 2 Compute loss for a given imaging modality m.

function LOSS_IMAGE($\boldsymbol{Y}^m, \theta_m, \psi_m, \boldsymbol{Z}^m, \sigma_m, \boldsymbol{\delta}, \gamma_m, N_m, F_m, P$)
For ease of notation we drop the m index in the pseudo-code.
  for n=1 to N **do** ▷ For each source
    $\boldsymbol{B}_{n:} =$ RT($\boldsymbol{M}_{n:}, diag(\boldsymbol{P}_{n,:})$) ▷ Sampling from q_1
    $\boldsymbol{\omega}^n =$ RT($\boldsymbol{R}_{n:}, diag(\boldsymbol{Q}^2_{n:})$) ▷ Sampling from q_2
    $\boldsymbol{w}^n =$ RT($\boldsymbol{T}_{n:}, diag(\boldsymbol{V}^2_{n:})$) ▷ Sampling from q_3
    $\boldsymbol{A}_{n:} = \boldsymbol{B}_{n:}\boldsymbol{\Sigma}^n$ ▷ Convolution of the sparse code of source n at a given spatial resolution
    $\boldsymbol{S}_{:n}(\boldsymbol{\delta}) = \phi(\boldsymbol{\delta}(\boldsymbol{\omega}^n)^T)\boldsymbol{w}^n$ ▷ Compute temporal trajectory of source n
    $\boldsymbol{S}'_{:n}(\boldsymbol{\delta}) = \frac{d\phi(\boldsymbol{\delta}(\boldsymbol{\omega}^n)^T)}{d\boldsymbol{\delta}}\boldsymbol{w}^n$ ▷ Compute derivative of temporal trajectory of source n
  end for
  $\boldsymbol{\Omega} \leftarrow$ block diagonal matrix containing all the set of $(\boldsymbol{\omega}^n)^T$
  $\boldsymbol{W} \leftarrow$ block diagonal matrix containing all the set of $\boldsymbol{w}^n$
  $\mathbb{E}_{q_1,q_2,q_3}[\log(p(\boldsymbol{Y}|\boldsymbol{B}, \boldsymbol{\Omega}, \boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{\delta}, \sigma))] \approx \sum_p -\frac{F}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}||\boldsymbol{Y}_{p:} - \boldsymbol{S}_{p:}\boldsymbol{A} - \boldsymbol{Z}_{p:}||^2$
  $\mathbb{E}_{q_2,q_3}[\log(p(\mathcal{C}|\boldsymbol{\Omega}, \boldsymbol{W}, \boldsymbol{\delta}, \gamma))] \approx -\sum_{p,n}\log((1 + \exp(-\gamma\boldsymbol{S}'_{p,n}(\boldsymbol{\delta})))$
    ▷ The two expectations terms are approximated using only one Monte-Carlo sample as proposed in [Kingma, 2013].
  KL $= \mathcal{D}[q_1(\boldsymbol{B})||p(\boldsymbol{B})] + \mathcal{D}[q_2(\boldsymbol{\Omega})||p(\boldsymbol{\Omega})] + \mathcal{D}[q_3(\boldsymbol{W})||p(\boldsymbol{W})]$ ▷ This tern is computed using approximations and formulas of Appendix 2.6.2.
  $\boldsymbol{\mathcal{L}} = \mathbb{E}_{q_1,q_2,q_3}[\log(p(\boldsymbol{Y}|\boldsymbol{B}, \boldsymbol{\Omega}, \boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{\delta}, \sigma))] + \mathbb{E}_{q_2,q_3}[\log(p(\mathcal{C}|\boldsymbol{\Omega}, \boldsymbol{W}, \boldsymbol{\delta}, \gamma))] -$ KL
  Return $\boldsymbol{\mathcal{L}}$
end function

Algorithm 3 Model optimization.

function OPTIMIZE($\boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{\Pi}, \boldsymbol{\delta}$, n_iter0, n_iter1, n_iter2)
  Initialize $\boldsymbol{\Pi}^{(0)}$, $\boldsymbol{\delta}^{(0)}$
  $i, j, k = 0$
  while i $\leq$ n_iter0 **do**
    for l=1 to n_iter1 **do** ▷ Optimizing spatio-temporal parameters only
      $\boldsymbol{\mathcal{L}} = 0$
      for m=1 to M **do** ▷ For each modality
        $\boldsymbol{\mathcal{L}}$ += LOSS_IMAGE($\boldsymbol{Y}^m, \theta_m, \psi_m, \boldsymbol{Z}^m, \sigma_m, \boldsymbol{\delta}, \gamma_m, N_m, F_m, P$)
      end for
      for c=1 to C **do** ▷ For each scalar feature
        $\boldsymbol{\mathcal{L}}$ += LOSS_SCALAR($\boldsymbol{V}_{:c}, \theta_c, \nu_c, \boldsymbol{\delta}, \gamma_m, P$)
      end for
      Compute $\frac{d\boldsymbol{\mathcal{L}}}{d\boldsymbol{\Pi}^{(j)}}$ through backpropagation
      $\boldsymbol{\Pi}^{(j+1)} = \text{ADAM}(\frac{d\boldsymbol{\mathcal{L}}}{d\boldsymbol{\Pi}^{(j)}}, \boldsymbol{\Pi}^{(j)}, \text{LR}(\boldsymbol{\Pi}))$ ▷ The spatio-temporal parameters are optimized by gradient descent using the ADAM optimizer. LR refers to the overall set of learning rates (cf Appendix 2.6.3)
      j += 1
    end for
    for l=1 to n_iter2 **do** ▷ Optimizing time-shift only
      $\boldsymbol{\mathcal{L}} = 0$
      for m=1 to M **do**
        $\boldsymbol{\mathcal{L}}$ += LOSS_IMAGE($\boldsymbol{Y}^m, \theta_m, \psi_m, \boldsymbol{Z}^m, \sigma_m, \boldsymbol{\delta}, \gamma_c, N_m, F_m, P$)
      end for
      for c=1 to C **do**
        $\boldsymbol{\mathcal{L}}$ += LOSS_SCALAR($\boldsymbol{V}_{:c}, \theta_c, \nu_c, \boldsymbol{\delta}, \gamma_c, P$)
      end for
      Compute $\frac{d\boldsymbol{\mathcal{L}}}{d\boldsymbol{\delta}^{(k)}}$ through backpropagation
      $\boldsymbol{\delta}^{(k+1)} = \text{ADAM}(\frac{d\boldsymbol{\mathcal{L}}}{d\boldsymbol{\delta}^{(k)}}, \boldsymbol{\delta}^{(k)}, \text{LR}(\boldsymbol{\delta}))$
      k += 1
    end for
    i += 1
  end while
end function

2.6.5 Synthetic experiments using ICA, NMF and PCA

In this Appendix, we show results obtained with standard methods (ICA, NMF, PCA) when applied within the experimental setting of Section 2.4.2. We recall that for these experiments observations were randomly aligned along the time-axis. The goal was to assess the ability of the different methods to reconstruct the spatio-temporal sources underlying the data when the time-axis is unknown. Results obtained in Table 2.5 show a substantial decrease of performances for the MSE and SSIM compared to MGPA (*cf* Table 2.2 in Section 2.4.2). Indeed, these methods do not consider time as a variable on which inference is required, thus preventing them from reconstructing correctly the temporal sources. Figure 2.13 shows an example of reconstruction when using ICA. We observe that even though the spatial reconstruction remains acceptable, the estimated temporal sources are not interpretable as ICA reconstructs the data using the time-axis on which observations have been mixed.

Table 2.5: MSE and SSIM between respectively the ground truth temporal and spatial sources with respect to the ones estimated by the different standard methods.

	Temporal (MSE)	Spatial (SSIM)
ICA	0.24 ± 0.08	$54\% \pm 2$
NMF	0.25 ± 0.03	$22\% \pm 14$
PCA	0.66 ± 0.05	$9\% \pm 3$

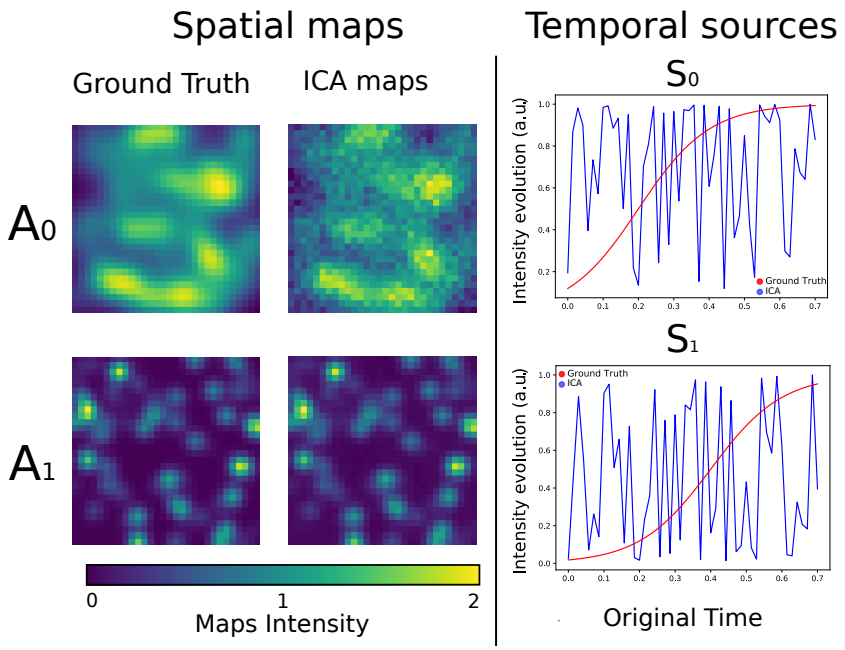

Figure 2.13: Spatial maps: Sample slice from ground truth images (A_0 $\lambda = 2$ mm, A_1 $\lambda = 1$ mm), the maps estimated by ICA. Temporal sources: Ground truth temporal sources (red) along with sources estimated by ICA (blue).

2.6.6 Model convergence

We provide in this Appendix details on the model convergence when applied on the ADNI data. The training was divided in three iterations of 30000 epochs each. During the two first iterations the spatio-temporal parameters and the time-shift are trained alternatively following a scheme of [2000,1000] epochs ten times. The third iteration only optimizes the spatio-temporal parameters. In Figure 2.14, we show the evolution of the total loss and the different terms composing it during training. The term reconstruction cost stands for $\sum_m \mathbb{E}_{q_1,q_2,q_3}[\log(p(\boldsymbol{Y}^m|\boldsymbol{B}^m,\boldsymbol{\Omega}^m,\boldsymbol{W}^m,\boldsymbol{Z}^m,\boldsymbol{\delta},\sigma_m))]$, monotonicity cost for $\sum_m \mathbb{E}_{q_2,q_3}[\log(p(\mathcal{C}^m|\boldsymbol{\Omega}^m,\boldsymbol{W}^m,\boldsymbol{\delta},\gamma_m))]$ and KL for $\sum_m \mathcal{D}[q_1(\boldsymbol{B}^m)||p(\boldsymbol{B}^m)] + \mathcal{D}[q_2(\boldsymbol{\Omega}^m)||p(\boldsymbol{\Omega}^m)] + \mathcal{D}[q_3(\boldsymbol{W}^m)||p(\boldsymbol{W}^m)]$. We observe that through the first two iterations the reconstruction and monotonicity costs decrease, and become stable during the last iteration. Differently, the KL cost increases during the first iteration as the model is driven by the reconstruction and monotonicity constraints. The KL term decreases during the second iteration, thus regularizing the model, before becoming stable during the third iteration. We also note that the graphs in Figure 2.14 show convergence profiles typical of those obtained with stochastic variational inference schemes, such as with Variational Autoencoders or Bayesian Neural Networks.

2.6.7 Application of ICA, NMF and PCA on brain imaging data

In this Appendix, we provide the results obtained when applying ICA, NMF and PCA on the ADNI data of Section 2.4.3. We used the three imaging modalities for each subject and concatenated these images in a (544×4256460) matrix. Our goal was to compare the spatio-temporal processes extracted using these standard methods with the ones from MGPA. We recall that in the case of MGPA the model automatically re-aligns the observations following monotonic assumptions for each biomarker, while these standard methods don't perform any inference on the time variable. Therefore, we created three experimental settings in which we changed the observations' alignment. In the first one, subjects were aligned by their chronological age (Figures 2.15, 2.16 and 2.17), in the second one by ADAS13 (Figures 2.18, 2.19 and 2.20) and in the last one time was randomly initialized like in the experiments of Section 2.4.3 (Figures 2.21, 2.22 and 2.23). We extracted six spatio-temporal sources for each method and each time-alignment, like in 2.4.3.

We observe that the temporal profiles are generally noisy and hard to interpret due to the lack of constraints on the temporal evolution. This motivates the need of smooth and monotonic constraints as in MGPA. Moreover, due to the concatenation of all the modalities they all share the same temporal patterns. This is an important difference with the modality-specific modelling of MGPA. Finally, we note that the spatial patterns associated with these methods are very similar, independently from the time-initialization,

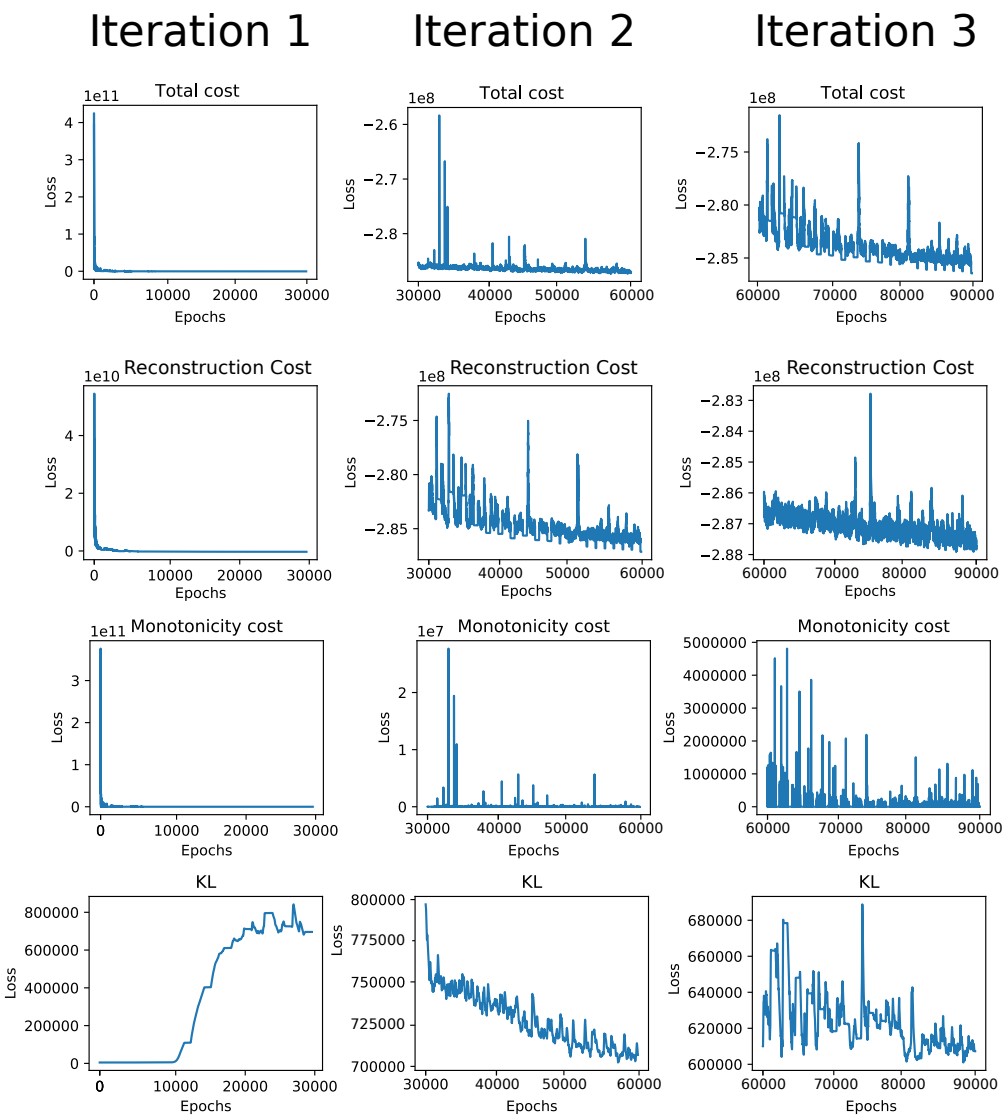

Figure 2.14: Evolution of the total loss, reconstruction cost, monotonicity cost and KL during training. Each iteration corresponds to 30000 epochs.

while the temporal sources substantially differ. This is also true when time is randomly initialized. These observations point to the challenge of giving a clinical interpretation of the results obtained with these approaches, and therefore to the need of plausible spatio-temporal constraints as provided in MGPA.

Subjects aligned by age.

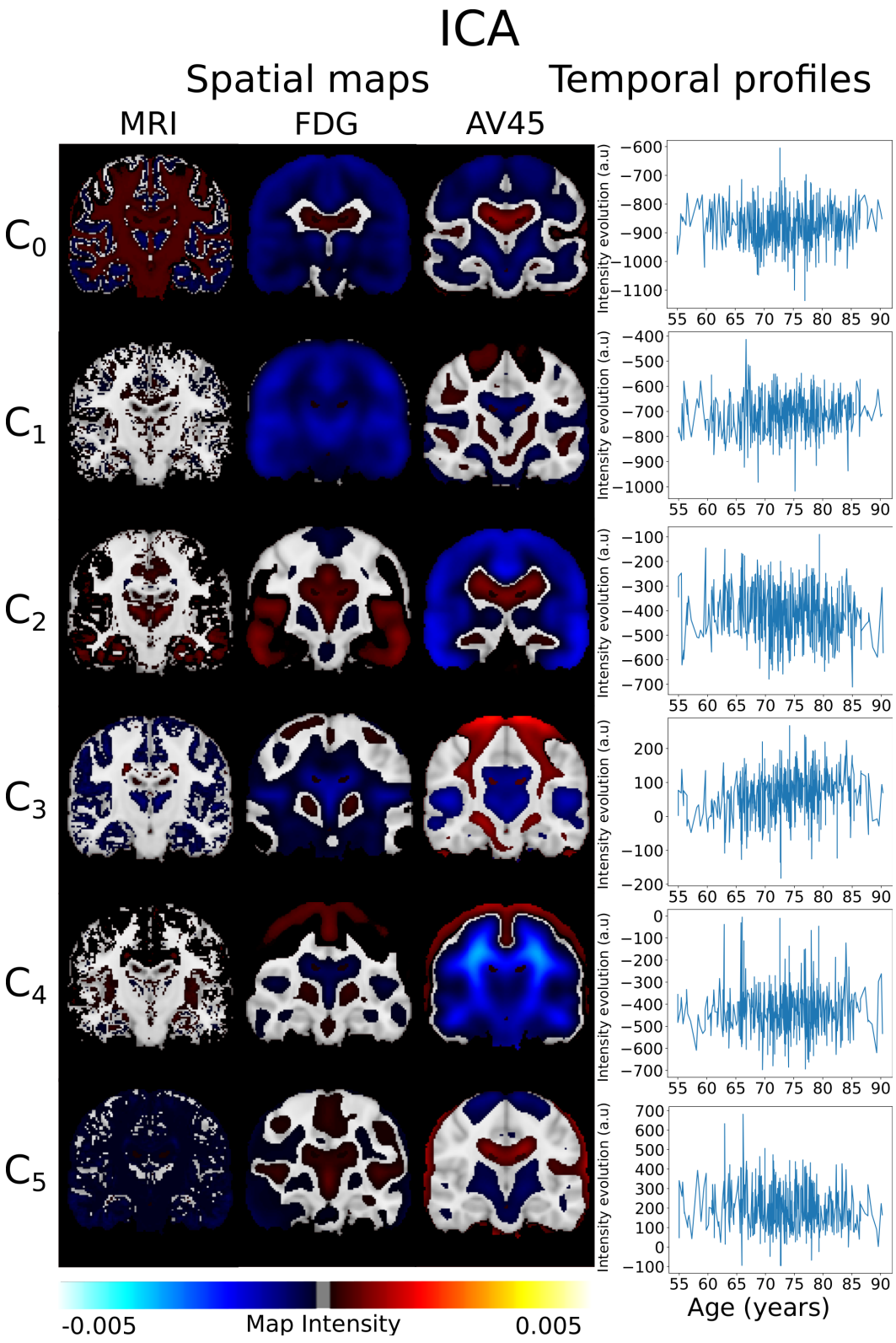

Figure 2.15: Spatio-temporal processes extracted by ICA with subjects aligned by age.

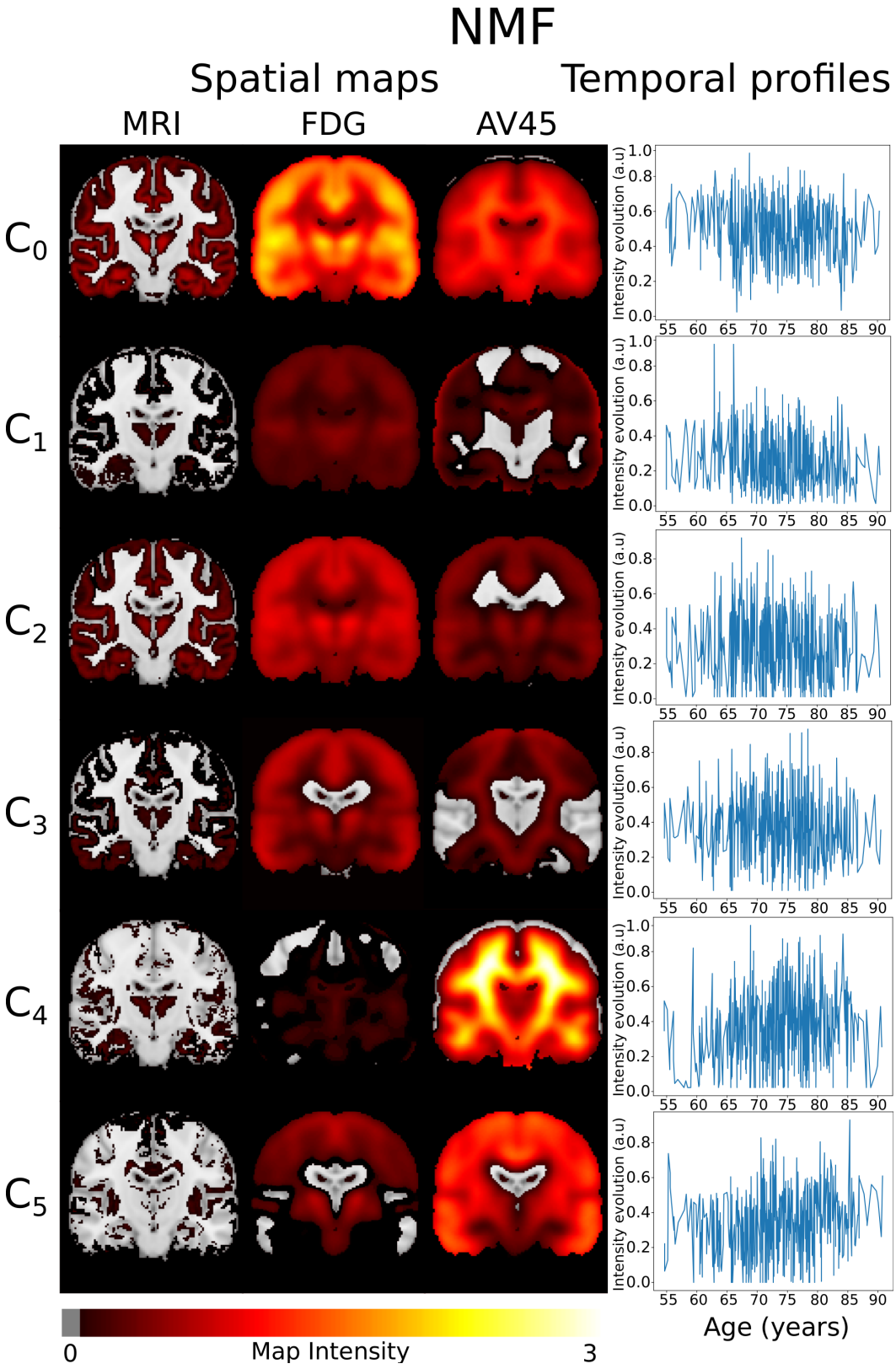

Figure 2.16: Spatio-temporal processes extracted by NMF with subjects aligned by age.

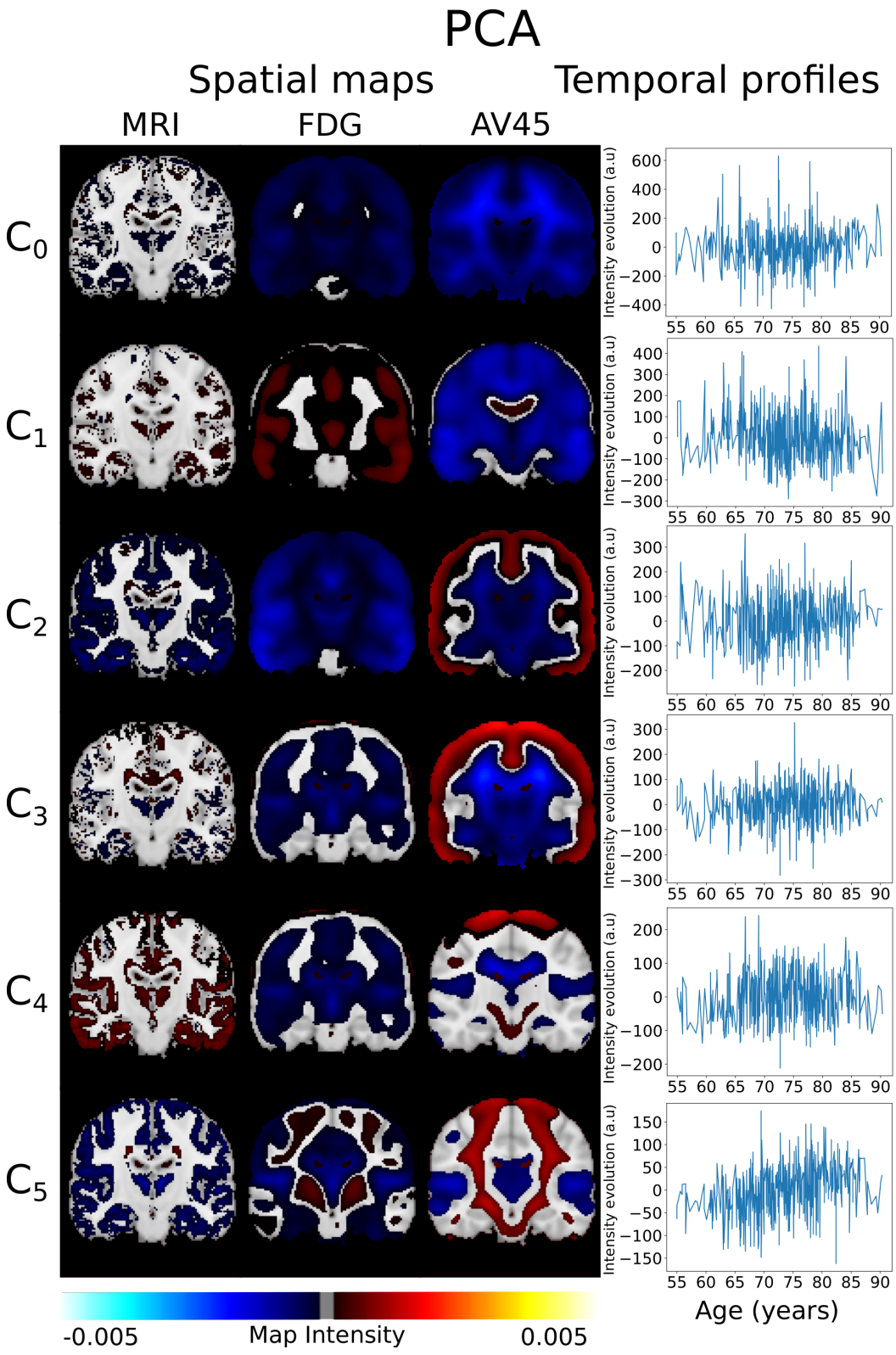

Figure 2.17: Spatio-temporal processes extracted by PCA with subjects aligned by age.

Subjects aligned by ADAS13.

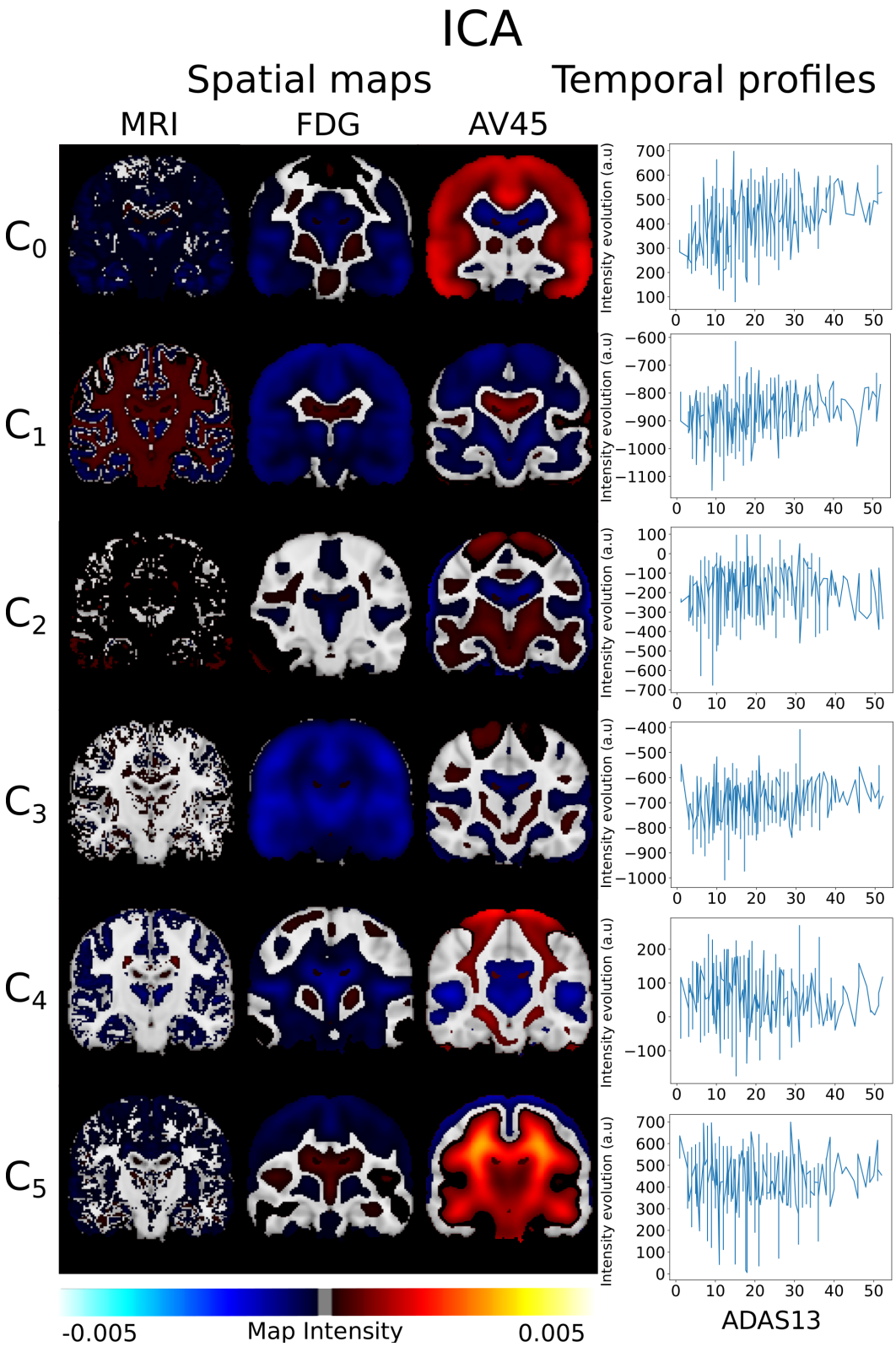

Figure 2.18: Spatio-temporal processes extracted by ICA with subjects aligned by ADAS13.

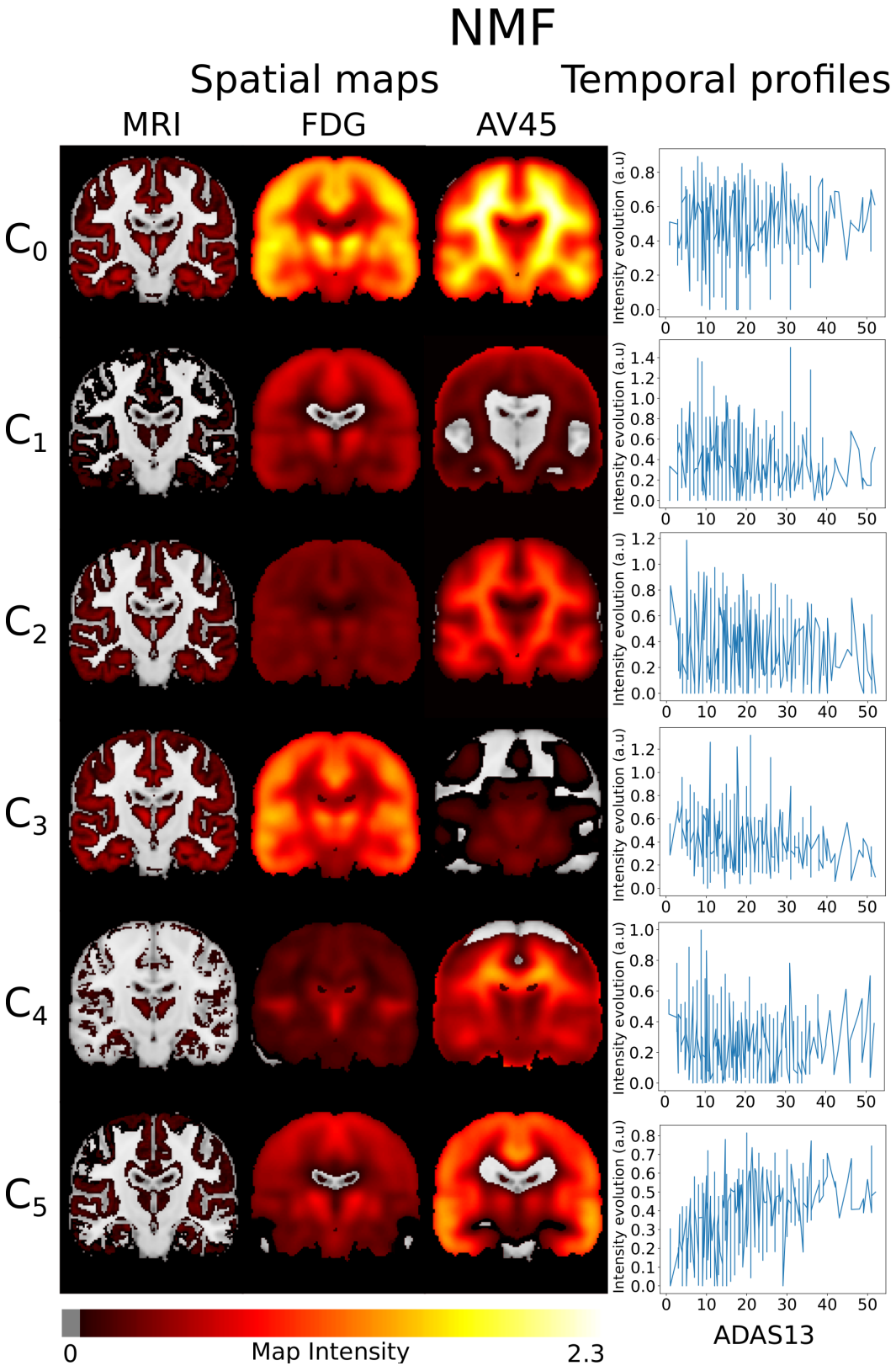

Figure 2.19: Spatio-temporal processes extracted by NMF with subjects aligned by ADAS13.

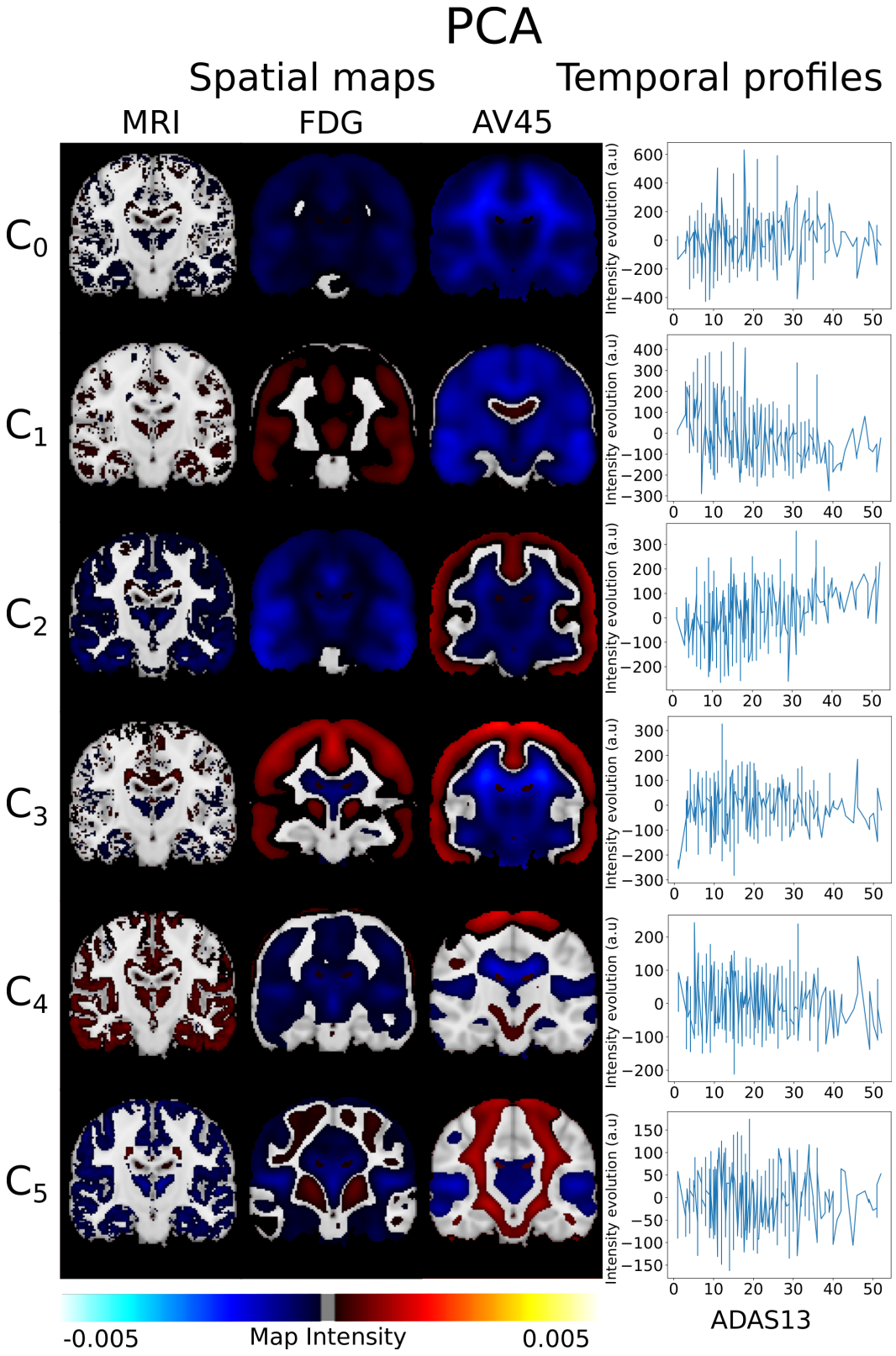

Figure 2.20: Spatio-temporal processes extracted by PCA with subjects aligned by ADAS13.

Subjects randomly aligned.

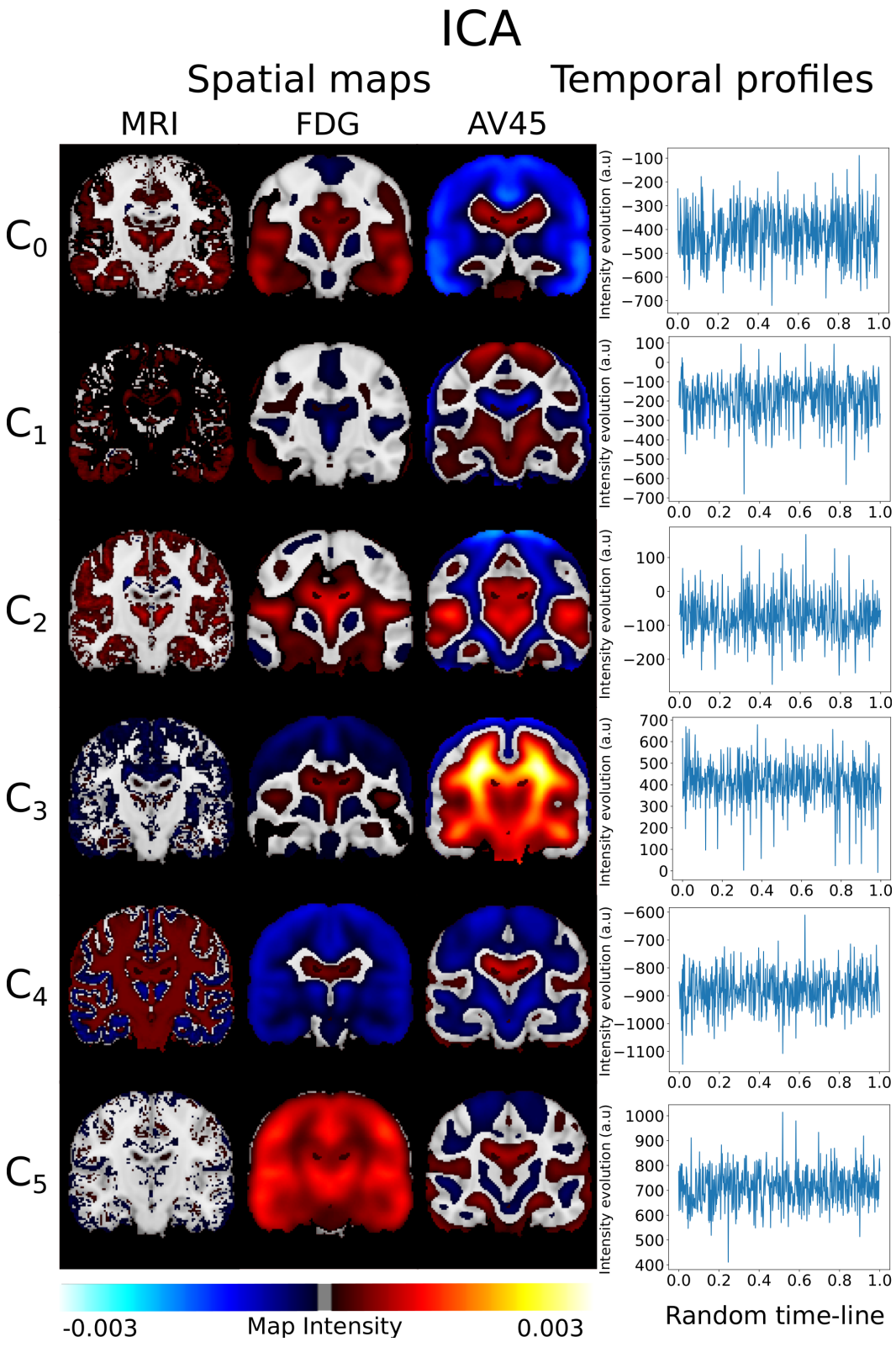

Figure 2.21: Spatio-temporal processes extracted by ICA with subjects randomly aligned.

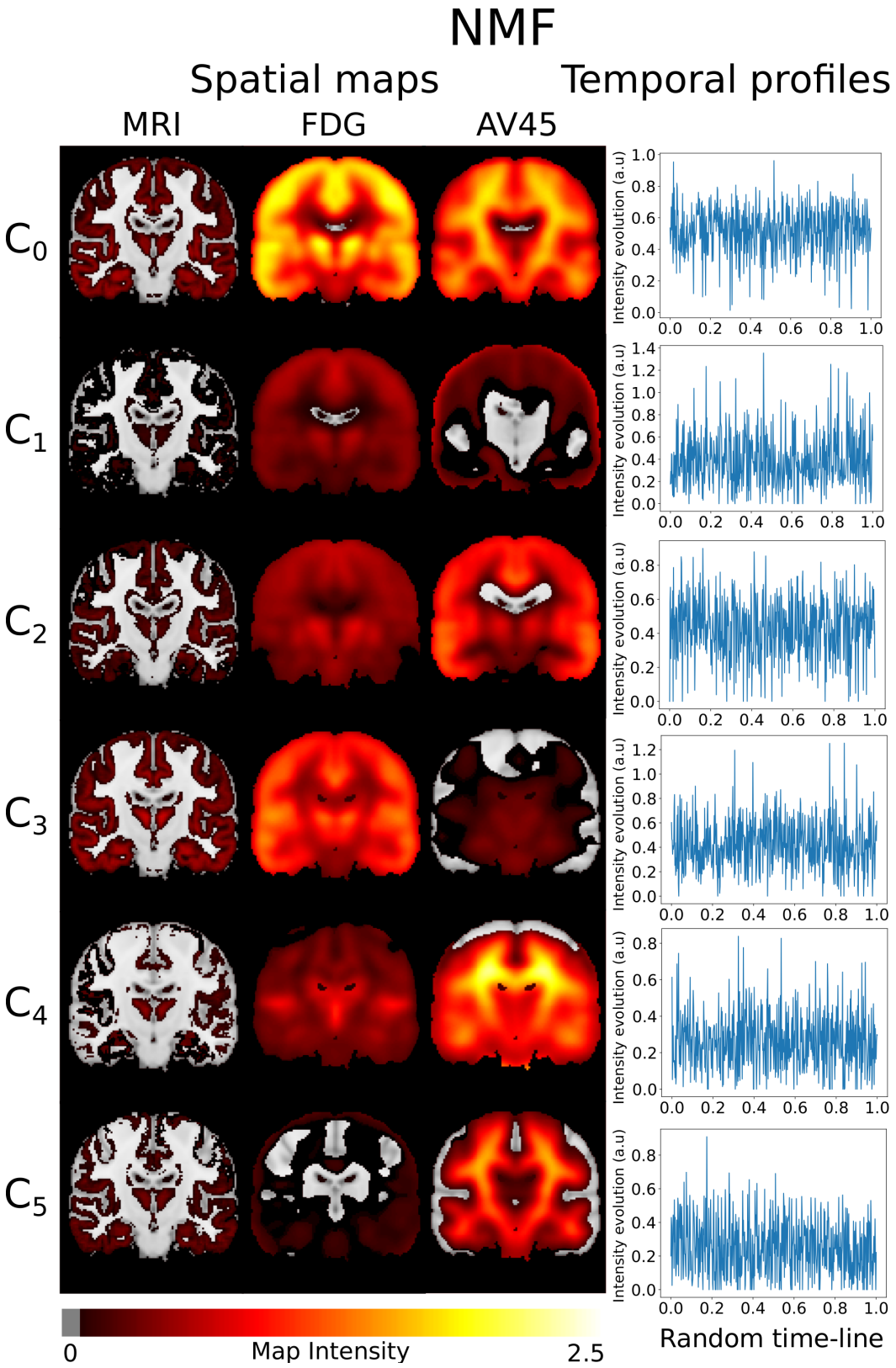

Figure 2.22: Spatio-temporal processes extracted by NMF with subjects randomly aligned.

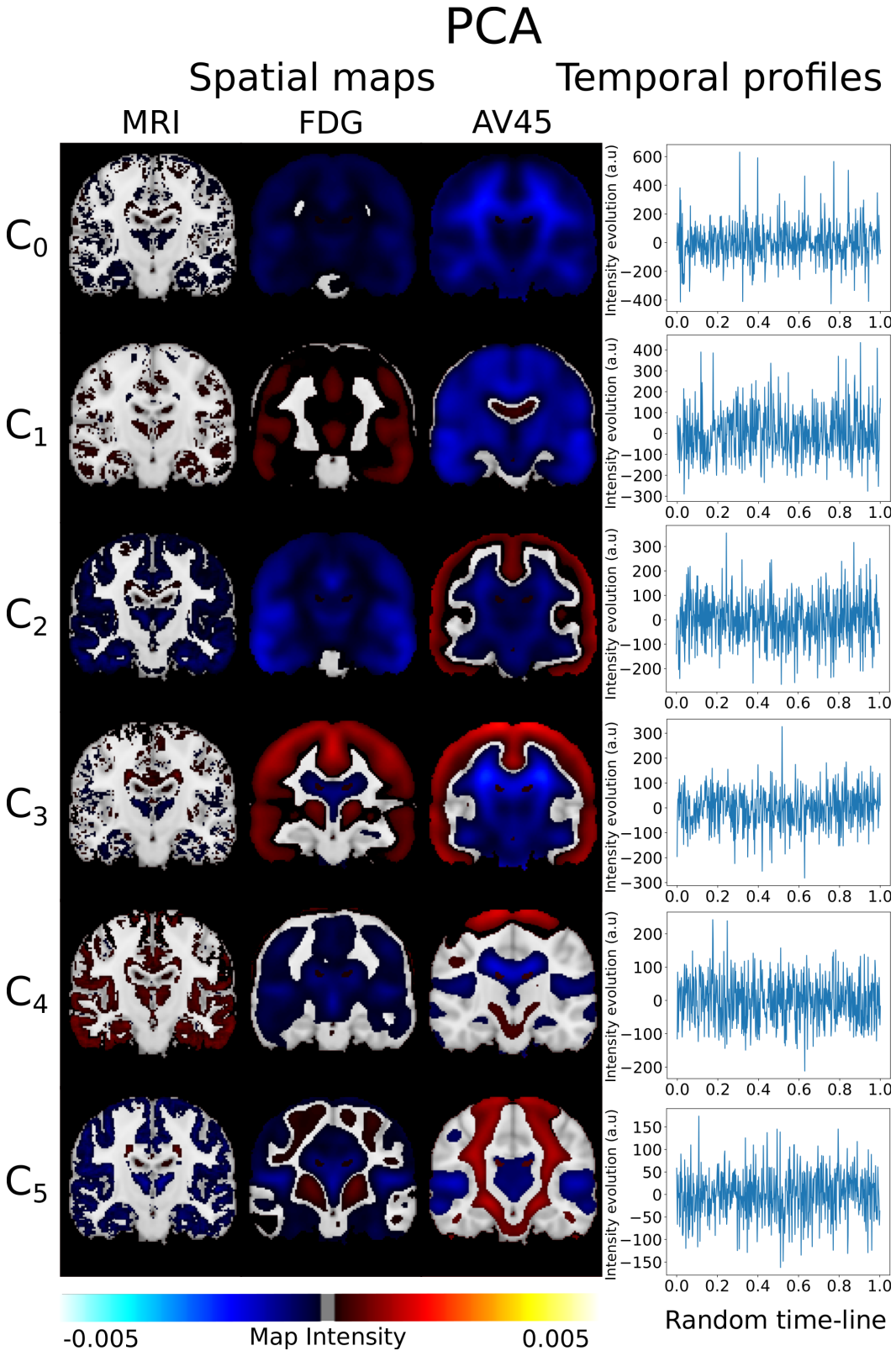

Figure 2.23: Spatio-temporal processes extracted by PCA with subjects randomly aligned.

3 Simulating the Outcome of Amyloid Treatments in Alzheimer's Disease from Imaging and Clinical data

Contents

3.1 Introduction **62**
3.2 Materials and Methods **64**
3.2.1 Study cohort and biomarkers' changes across clinical groups . 65
3.2.2 Model overview 67
3.2.3 Data modelling 67
3.2.4 Simulating the long-term progression of AD 69
3.2.5 Simulating intervention 69
3.2.6 Evaluating disease severity 70
3.2.7 Statistical analysis 70
3.3 Results **70**
3.3.1 Data acquisition and pre-processing 71
3.3.2 Progression model and latent relationships 71
3.3.3 Simulating clinical intervention 73
3.4 Discussion **75**
3.5 Conclusion **81**
3.6 Appendix **82**
3.6.1 Variational inference 82
3.6.2 Model optimization 83
3.6.3 Lower bound 84
3.6.4 KL divergence 84
3.6.5 Graphical model 85
3.6.6 Lower bound computation 85
3.6.7 Time-shift comparison and validation 86
3.6.8 Simulated clinical endpoints 86

In the previous chapter, we presented a method allowing to identify differential temporal processes of amyloid deposition, glucose hypometabolism and gray matter atrophy as well as the brain regions targeted by these mechanisms. While this approach provides valuable insights about Alzheimer's disease (AD) evolution, it remains purely descriptive

as it cannot explain the causal relationships between the pathological processes at stake. However, investigating such relationships entails great interest, in particular for simulating hypothetical scenarios of intervention, which could help planning clinical trials. To address this challenge we present in this chapter a novel method, SimulAD, which models AD progression through a dynamical system relating clinical and multi-modal imaging data. Thanks to this approach, we can investigate the effect of drug intervention on the disease progression, and particularly the impact of anti-amyloid treatments on cognitive outcomes. This chapter has been accepted for publication in Brain Communications [Abi Nader, 2021].

Abstract. In this study we investigate SimulAD, a novel quantitative instrument for the development of intervention strategies for disease modifying drugs in Alzheimer's disease. SimulAD is based on the modeling of the spatio-temporal dynamics governing the joint evolution of imaging and clinical biomarkers along the history of the disease, and allows the simulation of the effect of intervention time and drug dosage on the biomarkers' progression. When applied to multi-modal imaging and clinical data from the Alzheimer's Disease Neuroimaging Initiative the method enables to generate hypothetical scenarios of amyloid lowering interventions. The results quantify the crucial role of intervention time, and provide a theoretical justification for testing amyloid modifying drugs in the pre-clinical stage. Our experimental simulations are compatible with the outcomes observed in past clinical trials, and suggest that anti-amyloid treatments should be administered at least seven years earlier than what is currently being done in order to obtain statistically powered improvement of clinical endpoints.

3.1 Introduction

The number of people affected by Alzheimer's disease (AD) has recently exceeded 46 millions and is expected to double every 20 years [Prince, 2015], thus posing significant healthcare challenges. Yet, while the disease mechanisms remain in large part unknown, there are still no effective pharmacological treatments leading to tangible improvements of patients' clinical progression. One of the main challenges in understanding AD is that its progression goes through a silent asymptomatic phase that can stretch over decades before a clinical diagnosis can be established based on cognitive and behavioral symptoms. To help designing appropriate intervention strategies, hypothetical models of the disease history have been proposed, characterizing the progression by a cascade of morphological and molecular changes affecting the brain, ultimately leading to cognitive impairment [Jack, 2013a; Jack, 2013b]. The dominant hypothesis is that disease dynamics along the asymptomatic period are driven by the deposition in the brain of the amyloid peptide,

triggering the so-called “amyloid cascade” [Bateman, 2012; Braak, 1991; Delacourte, 1999; Murphy, 2010; Villemagne, 2013]. Based on this rationale, clinical trials have been focusing on the development and testing of disease modifiers targeting amyloid aggregates [Cummings, 2019b], for example by increasing its clearance or blocking its accumulation. Although the amyloid hypothesis has been recently invigorated by a post-hoc analysis of the aducanumab trial [Howard, 2020], clinical trials failed so far to show efficacy of this kind of treatments [Schwarz, 2019], as the clinical primary endpoints were not met [Egan, 2019; Wessels, 2019; Honig, 2018], or because of unacceptable adverse effects [Henley, 2019]. In the past years, growing consensus emerged about the critical importance of intervention time, and about the need of starting anti-amyloid treatments during the pre-symptomatic stages of the disease [Aisen, 2018]. Nevertheless, the design of optimal intervention strategies is currently not supported by quantitative analysis methods allowing to model and assess the effect of intervention time and dosing [Klein, 2019]. The availability of models of the pathophysiology of AD would entail great potential to test and analyze clinical hypothesis characterizing AD mechanisms, progression, and intervention scenarios.

Within this context, quantitative models of disease progression, Disease progression Models referred to as DPMs, have been proposed [Fonteijn, 2012; Jedynak, 2012; Abi Nader, 2020; Oxtoby, 2017; Schiratti, 2015], to quantify the dynamics of the changes affecting the brain during the whole disease span. These models rely on the statistical analysis of large datasets of different data modalities, such as clinical scores, or brain imaging measures derived from MRI, Amyloid- and Fluorodeoxyglucose- PET [Bilgel, 2015; Burnham, 2020; Donohue, 2014; IturriaMedina, 2016; Koval, 2018]. In general, DPMs estimate a long-term disease evolution from the joint analysis of multivariate time-series acquired on a short-term time-scale. Due to the temporal delay between the disease onset and the appearance of the first symptoms, DPMs rely on the identification of an appropriate temporal reference to describe the long-term disease evolution [Lorenzi, 2017; Marinescu, 2019a]. These tools are promising approaches for the analysis of clinical trials data, as they allow to represent the longitudinal evolution of multiple biomarkers through a global model of disease progression. Such a model can be subsequently used as a reference in order to stage subjects and quantify their relative progression speed [Insel, 2020; Li, 2019; Oxtoby, 2018; Young, 2014]. However, these approaches remain purely descriptive as they don’t account for causal relationships among biomarkers. Therefore, they generally don’t allow to simulate progression scenarios based on hypothetical intervention strategies, thus providing a limited interpretation of the pathological dynamics. This latter capability is of utmost importance for planning and assessment of disease modifying treatments.

To fill this gap, recent works such as [Hao, 2016; Petrella, 2019] proposed to model AD progression based on specific assumptions on the biochemical processes of pathological protein propagation. These approaches explicitly define biomarkers interactions through

the specification of sets of Ordinary Differential Equations (ODEs), and are ideally suited to simulate the effect of drug interventions [IturriaMedina, 2017]. However, these methods are mostly based on the arbitrary choices of pre-defined evolution models, which are not inferred from data. This issue was recently addressed by [Garbarino, 2019], where the authors proposed an hybrid modeling method combining traditional DPMs with dynamical models of AD progression. Still, since this approach requires to design suitable models of protein propagation across brain regions, extending this method to jointly account for spatio-temporal interactions between several processes, such as amyloid propagation, glucose metabolism, and brain atrophy, is considerably more complex. Finally, these methods are usually designed to account for imaging data only, which prevents to jointly simulate heterogeneous measures [Antelmi, 2019], such as image-based biomarkers and clinical outcomes, the latter remaining the reference markers for patients and clinicians.

In this work we present SimulAD, a novel computational model of AD progression allowing to simulate intervention strategies across the history of the disease. The model is here used to quantify the potential effect of amyloid modifiers on the progression of brain atrophy, glucose metabolism, and ultimately on the clinical outcomes for different scenarios of intervention. To this end, we model the joint spatio-temporal variation of different modalities along the history of AD by identifying a system of ODEs governing the pathological progression. This latent ODEs system is specified within an interpretable low-dimensional space relating multi-modal information, and combines clinically-inspired constraints with unknown interactions that we wish to estimate. The interpretability of the relationships in the latent space is ensured by mapping each data modality to a specific latent coordinate. The model is formulated within a Bayesian framework, where the latent representation and dynamics are efficiently estimated through stochastic variational inference. To generate hypothetical scenarios of amyloid lowering interventions, we apply SimulAD to multi-modal imaging and clinical data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Our results provide a meaningful quantification of different intervention strategies, compatible with findings previously reported in clinical studies. For example, we estimate that in a study with 100 individuals per arm, statistically powered improvement of clinical endpoints can be obtained by completely arresting amyloid accumulation at least 11 years before Alzheimer's dementia. The minimum intervention time decreases to 7 years for studies based on 1000 individuals per arm.

3.2 Materials and Methods

In the following sections, healthy individuals will be denoted as NL stable, subjects with mild cognitive impairment as MCI stable, subjects diagnosed with Alzheimer's

disease dementia as AD dementia. We define conversion as the change of diagnosis towards a more pathological state. Therefore, NL converters are subjects who were diagnosed as cognitively normal at baseline and whose diagnosis changed either in MCI or AD dementia during their follow-up visits. MCI converters are subjects who were diagnosed as MCI at baseline and subsequently progressed to AD dementia. Diagnosis was established using the DX column from the ADNIMERGE file (`https://adni.bitbucket.io/index.html`), which reflects the standard ADNI clinical assessment based on Wechsler Memory Scale, Mini-Mental State Examination, and Clinical Dementia Rating. Amyloid concentration and glucose metabolism are respectively measured by (18)F-florbetapir Amyloid (AV45)-PET and (18)F-fluorodeoxyglucose (FDG)-PET imaging. Cognitive and functional abilities are assessed by the following neuro-psychological tests: Alzheimer's Disease Assessment Scale (ADAS11), Mini-Mental State Examination (MMSE), Functional Assessment Questionnaire (FAQ), Rey Auditory Verbal Learning Test (RAVLT) immediate, RAVLT learning, RAVLT forgetting, and Clinical Dementia Rating Scale Sum of Boxes (CDRSB).

3.2.1 Study cohort and biomarkers' changes across clinical groups

Our study is based on a cohort of 442 amyloid positive individuals composed of 71 NL stable subjects, 33 NL converters subjects, 131 subjects diagnosed with MCI, 105 MCI converters subjects, and 102 AD dementia patients. Among the 131 MCI subjects, 78 were early MCI and 53 were late MCI. Concerning the group of MCI converters, 80 subjects were late MCI at baseline and 25 were early MCI. The term "amyloid positive" refers to subjects whose amyloid level in the CSF was below the nominal cutoff of 192 pg/ml [Shaw, 2009] either at baseline, or during any follow-up visit, and conversion to AD dementia was determined using the last available follow-up information. This preliminary selection of patients aims at constituting a cohort of subjects for whom it is more likely to observe "Alzheimer's pathological changes" [Jack, 2018]. The length of follow-up varies between 0 and 16 years. Further information about the data are available on `https://adni.bitbucket.io/reference/`, while details on data acquisition and processing are provided in Section 3.3.1.

We show in Table 3.1A socio-demographic information for the training cohort across the different clinical groups. Table 3.1B shows baseline values and annual rates of change across clinical groups for amyloid burden (average normalized AV45 uptake in frontal cortex, anterior cingulate, precuneus and parietal cortex), glucose metabolism (average normalized FDG uptake in frontal cortex, anterior cingulate, precuneus and parietal cortex), for hippocampal and medial temporal lobe volumes, and for the cognitive ability as measured by ADAS11. Compatibly with previously reported results [Schuff, 2009; Cash, 2015], we observe that while regional atrophy, glucose metabolism and cognition

show increasing rate of change when moving from healthy to pathological conditions, the change of AV45 is maximum in NL stable, NL converters and MCI stable subjects. We also notice the increased magnitude of ADAS11 in AD dementia as compared to the other clinical groups. Finally, we note that glucose metabolism and regional atrophy show comparable magnitudes of change.

The observations presented in Table 3.1 provide us with a coarse representation of the biomarkers' trajectories characterizing AD. The complexity of the dynamical changes we may infer is limited, as the clinical stages roughly approximate a temporal scale describing the disease history, while very little insights can be obtained about the biomarkers' interactions. Within this context, our model allows the quantification of the fine-grained dynamical relationships across biomarkers at stake during the history of the disease. Investigation of intervention scenarios can be subsequently carried out by opportunely modulating the estimated dynamics parameters according to specific intervention hypothesis (e.g. amyloid lowering at a certain time).

Table 3.1: A: Baseline socio-demographic information for training cohort (442 subjects for 2781 data points, follow-up from 0 to 16 years depending on subjects). Average values, standard deviation in parenthesis. B: Baseline values (bl) and annual rates of change (% change / year) of amyloid burden (average normalized AV45 uptake in frontal cortex, anterior cingulate, precuneus and parietal cortex), glucose hypometabolism (average normalized FDG uptake in frontal cortex, anterior cingulate, precuneus and parietal cortex), hippocampus volume, medial temporal lobe volume, and ADAS11 score for the different clinical groups. Median values, interquartile range below. The volumes of the hippocampus and the medial temporal lobe are averaged across left and right hemispheres. NL: healthy individuals, MCI: individuals with mild cognitive impairment. AD dementia: Alzheimer's disease dementia. APOE4: apolipoprotein E ϵ4. FDG: (18)F-fluorodeoxyglucose Positron Emission Tomography (PET) imaging. AV45: (18)F-florbetapir Amyloid PET imaging. SUVR: Standardized Uptake Value Ratio. MTL: Medial Temporal Lobe. ADAS11: Alzheimer's Disease Assessment Scale-cognitive subscale, 11 items.

A: Socio-demographics

	NL stable	NL converters	MCI stable	MCI converters	AD dementia
N	71	33	131	105	102
Age (yrs)	74 (6)	76 (4)	72 (8)	73 (7)	74 (8)
Education (yrs)	16 (2)	17 (2)	16 (3)	16 (3)	16 (2)
APOE4-carrier (%)	41)	51)	61	75	71

B: Biomarkers and rates of change

	NL stable		NL converters		MCI stable		MCI converters		AD dementia	
	bl	% change / year	bl	% change / year	bl	% change / year	bl	% change / year	bl	% change / year
Global AV45 (SUVR)	1.21 [1.06 ; 1.37]	0.8 [0.1 ; 2.2]	1.36 [1.28 ; 1.55]	1.24 [0.43 ; 2.2]	1.27 [1.12 ; 1.44]	1.21 [0.1 ; 2.5]	1.41 [1.29 ; 1.53]	0.03 [-1.5 ; 1.4]	1.45 [1.34 ; 1.57]	0.06 [-1.9 ; 3.3]
Global FDG (SUVR)	1.27 [1.19 ; 1.34]	-0.47 [-1.8 ; 0.9]	1.22 [1.16 ; 1.33]	-1.6 [-2.2 ; -1.0]	1.28 [1.19 ; 1.36]	-0.92 [-2.9 ; 0.0]	1.16 [1.05 ; 1.25]	-3.1 [-4.8 ; -1.4]	1.04 [0.97 ; 1.14]	-5.0 [-7.9 ; -2.0]
Hippocampus (ml)	3.7 [3.4 ; 4.0]	-1.6 [-2.2 ; -0.4]	3.5 [3.1 ; 3.8]	-1.8 [-3.3 ; -2.2]	3.5 [3.1 ; 3.8]	-1.5 [-3.3 ; -0.7]	3.1 [2.7 ; 3.4]	-3.8 [-5.1 ; -2.3]	2.8 [2.5 ; 3.2]	-4.5 [-6.8 ; -2.0]
MTL (ml)	10.0 [9.3 ; 10.5]	-0.8 [-2.0 ; 0.1]	9.8 [8.5 ; 10.5]	-1.3 [-2.3 ; -0.7]	10.4 [9.8 ; 11.2]	-1.0 [-2.2 ; 0.4]	9.1 [8.2 ; 10.1]	-2.9 [-4.7 ; -1.5]	8.5 [7.6 ; 9.3]	-5.0 [-7.9 ; -1.9]
ADAS11	5.0 [3.1 ; 7.0]	0.1 [-0.2 ; 0.8]	8.0 [5.0 ; 12.2]	1.7 [-0.6 ; 2.8]	9.0 [6.0 ; 11.6]	1.2 [0.3 ; 2.8]	14.3 [11.0 ; 20.0]	5.0 [2.3 ; 8.4]	22.0 [17.0 ; 28.0]	10.3 [4.1 ; 21.0]

3.2.2 Model overview

We provide in Figure 3.1 an overview of SimulAD. Baseline multi-modal imaging and clinical information for a given subject are transformed into a latent variable composed of four z-scores quantifying respectively the overall severity of atrophy, glucose metabolism, amyloid burden, and cognitive and functional assessment. The model estimates the dynamical relationships across these z-scores to optimally describe the temporal transitions between follow-up observations. These transition rules are here mathematically defined by the parameters of a system of ODEs, which is estimated from the data. This dynamical system allows to compute the evolution of the z-scores over time from any baseline observation, and to predict the associated multi-modal imaging and clinical measures. It is important to note that this modelling choice requires to have at least one visit per patient for which all the measures are available, in order to compute the z-scores temporal evolution.

SimulAD thus enables to simulate the pathological progression of biomarkers across the entire history of the disease. Once the model is estimated, we can modify the ODEs parameters to simulate different evolution scenarios according to specific hypothesis. For example, by reducing the parameters associated with the progression rate of amyloid, we can investigate the relative change in the evolution of the other biomarkers. This setup thus provides us with a data-driven system enabling the exploration of hypothetical intervention strategies, and their effect on the pathological cascade.

3.2.3 Data modelling

We consider observations $\mathbf{X}_i(t) = [\mathbf{x}_i^1(t), \mathbf{x}_i^2(t), ..., \mathbf{x}_i^M(t)]^T$, which correspond to multivariate measures derived from *M* different modalities (e.g clinical scores, MRI, AV45, or FDG measures) at time *t* for subject *i*. Each vector $\mathbf{x}_i^m(t)$ has dimension D_m. We postulate the following generative model, in which the modalities are assumed to be independently generated by a common latent representation of the data $\mathbf{z}_i(t)$:

$$\begin{aligned} p(\mathbf{X}_i(t)|\mathbf{z}_i(t), \boldsymbol{\sigma^2}, \boldsymbol{\psi}) &= \prod_m p(\mathbf{x}_i^m(t)|\mathbf{z}_i(t), \sigma_m^2, \psi_m) \\ &= \prod_m \mathcal{N}(\mu_m(\mathbf{z}_i(t), \psi_m), \sigma_m^2), \\ \mathbf{z_i}(t) &= \Lambda(\mathbf{z}_i(t_0), t), \\ \mathbf{z}_i(t_0) &\sim p(\mathbf{z}_i(t_0)), \end{aligned} \tag{3.1}$$

where σ_m^2 is measurement noise, while ψ_m are the parameters of the function μ_m which maps the latent state to the data space for the modality *m*. For simplicity of notation we denote $\mathbf{z}_i(t)$ by $\mathbf{z}(t)$. We assume that each coordinate of $\mathbf{z}$ is associated to a specific modality *m*, leading to an *M*-dimensional latent space. The Λ operator which gives the

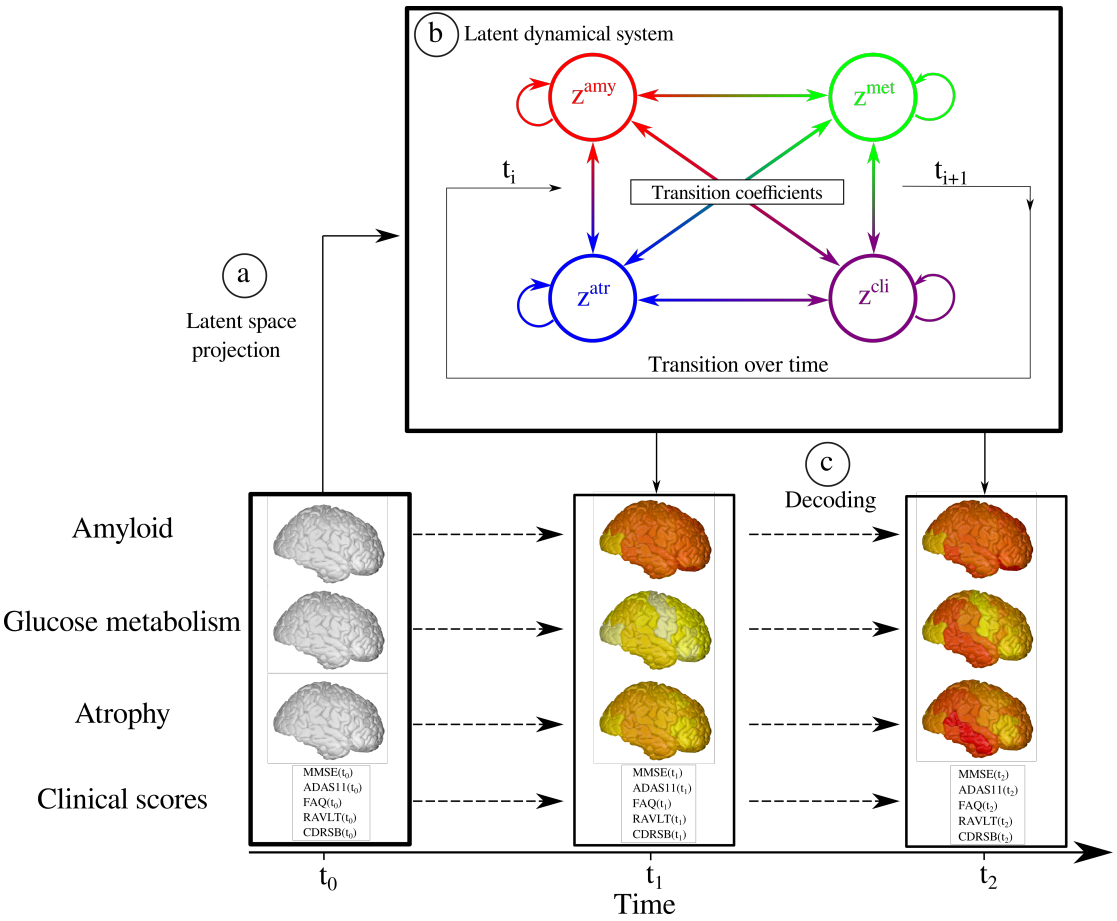

Figure 3.1: Overview of SimulAD. a) High-dimensional multi-modal measures are projected into a 4-dimensional latent space. Each data modality is transformed in a corresponding z-score z^{amy}, z^{met}, z^{atr}, z^{cli}. b) The dynamical system describing the relationships between the z-scores allows to compute their transition across the evolution of the disease. c) Given the latent space and the estimated dynamics, the follow-up measurements can be reconstructed to match the observed data.

value of the latent representation at a given time t, is defined by the solution of the following system of ODEs:

$$\frac{dz^m(t)}{dt} = k_m z^m(t)(1 - z^m(t)) + \sum_{j \neq m} \alpha_{m,j} z^j(t), \quad m=1,...,M. \tag{3.2}$$

For each coordinate, the first term of the equation enforces a sigmoidal evolution with a progression rate k_m, while the second term accounts for the relationship between modalities *m* and *j* through the parameters $\alpha_{m,j}$. This system can be rewritten as:

$$\begin{aligned} &\frac{d\mathbf{z}(t)}{dt} = \mathbf{W}\mathbf{z}(t) - \mathbf{V}\mathbf{z}^2(t) = g(\mathbf{z}(t), \theta_{ODE}) \text{ where,} \\ &\left(\mathbf{W}_{i,j}\right) = \begin{cases} k_i \text{ if i=j,} \\ \alpha_{i,j} \text{ otherwise;} \end{cases} \text{ and } \left(\mathbf{V}_{i,j}\right) = \begin{cases} k_i \text{ if i=j} \\ 0 \text{ otherwise,} \end{cases} \end{aligned} \tag{3.3}$$

θ_{ODE} denotes the parameters of the system of ODEs, which correspond to the entries of the matrices $\mathbf{W}$ and $\mathbf{V}$. According to Equation 3.3, for each initial condition $\mathbf{z}(0)$, the latent state at time *t* can be computed through integration, $\mathbf{z}(t) = \mathbf{z}(0) + \int_0^t g(\mathbf{z}(x), \theta_{ODE})dx$.

We resort to variational inference and stochastic gradient descent in order to optimize the parameters of the model. The procedure is detailed in Appendices 3.6.1 and 3.6.2.

3.2.4 Simulating the long-term progression of AD

To simulate the long-term progression of AD we first project the AD dementia patients in the latent space via the encoding functions. We can subsequently follow the trajectories of these subjects backward and forward in time, in order to estimate the associated trajectory from the healthy to their respective pathological condition. In practice, a Gaussian Mixture Model is used to fit the empirical distribution of the AD dementia subjects' latent projection. The number of components and covariance type of the GMM is selected by relying on the Akaike information criterion [Akaike, 1998]. The fitted GMM allows us to sample pathological latent representations $\mathbf{z}_i(t_0)$, that can be integrated forward and backward in time thanks to the estimated set of latent ODEs, to finally obtain a collection of latent trajectories $\mathbf{Z}(t) = [\mathbf{z}_1(t), ..., \mathbf{z}_N(t)]$ summarising the distribution of the long-term AD evolution.

3.2.5 Simulating intervention

In this section we assume that we computed the average latent progression of the disease $\mathbf{z}(t)$. Thanks to the modality-wise encoding (*cf.* Appendix 3.6.1) each coordinate of the latent representation can be interpreted as representing a single data modality.

Therefore, we propose to simulate the effect of an hypothetical intervention on the disease progression, by modulating the vector $\frac{d\mathbf{z}(t)}{dt}$ after each integration step such that:

$$\left(\frac{d\mathbf{z}(t)}{dt}\right)^* = \mathbf{\Gamma}\frac{d\mathbf{z}(t)}{dt} \text{ where, } \mathbf{\Gamma} = \begin{pmatrix} \gamma_1 & & \\ & \ddots & \\ & & \gamma_m \end{pmatrix}. \tag{3.4}$$

The values γ_m are fixed between 0 and 1, allowing to control the influence of the corresponding modalities on the system evolution, and to create hypothetical scenarios of evolution. For example, for a 100% (resp. 50%) amyloid lowering intervention we set $\gamma_{amy} = 0$ (resp. $\gamma_{amy} = 0.5$).

3.2.6 Evaluating disease severity

Given an evolution $\mathbf{z}(t)$ describing the disease progression in the latent space, we propose to consider this trajectory as a reference and to use it in order to quantify the individual disease severity of a subject $\mathbf{X}$. This is done by estimating a time-shift τ defined as:

$$\begin{aligned} \tau &= \arg\min_t ||f(\mathbf{X}, \boldsymbol{\phi}^1) - \mathbf{z}(t)||_1 \\ &= \sum_m |f(\mathbf{x}^m, \phi^1) - z^m(t)|. \end{aligned} \tag{3.5}$$

This time-shift allows to quantify the pathological stage of a subject with respect to the disease progression along the reference trajectory $\mathbf{z}(t)$. Moreover, the time-shift can still be estimated even in the case of missing data modalities, by only encoding the available measures of the observed subject.

3.2.7 Statistical analysis

The model was implemented using the Pytorch library [Paszke, 2019]. The estimated disease severity was compared group-wise via two-sided Wilcoxon-Mann-Whitney test ($P < 0.01$). Differences between the clinical outcomes distribution after simulation of intervention were compared via two-sided Student's T-test ($P < 0.01$). Shadowed areas in the different figures show ± standard deviation of the mean.

3.3 Results

In the following, MRI, FDG-PET, and AV45-PET images are processed in order to respectively extract regional gray matter density, glucose metabolism and amyloid load from

a brain parcellation. The z-scores of gray matter atrophy (z^{atr}), glucose metabolism (z^{met}), and amyloid burden (z^{amy}), are computed using the measures obtained by this pre-processing step. The clinical z-score z^{cli} is derived from neuro-psychological scores: ADAS11, MMSE, FAQ, RAVLT immediate, RAVLT learning, RAVLT forgetting and CDRSB. This panel of scores was chosen to provide a comprehensive representation of cognitive, memory and functional abilities.

3.3.1 Data acquisition and pre-processing

Data used in the preparation of this work were obtained from the ADNI database. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. For up-to-date information, see `www.adni-info.org`.

We considered four types of biomarkers, related to clinical scores, gray matter atrophy, amyloid load and glucose metabolism, and respectively denoted by *cli*, *atr*, *amy* and *met*. MRI images were processed following the longitudinal pipeline of Freesurfer [Reuter, 2012], to obtain gray matter volumes in a standard anatomical space. AV45-PET and FDG-PET images were aligned to the closest MRI in time and normalized to the cerebellum uptake. Regional gray matter density, amyloid load and glucose metabolism were extracted from the Desikan-Killiany parcellation [Desikan, 2006]. We discarded white-matter, ventricular, and cerebellar regions, thus obtaining 82 regions that were averaged across hemispheres. Therefore, for a given subject, $\mathbf{x}^{atr}, \mathbf{x}^{amy}$ and $\mathbf{x}^{met}$ are respectively 41-dimensional vectors. The variable $\mathbf{x}^{cli}$ is composed of the neuro-psychological scores ADAS11, MMSE, RAVLT immediate, RAVLT learning, RAVLT forgetting, FAQ, and CDRSB. The total number of measures is of 2781 longitudinal data points. We recall that the model estimation requires a visit for which all the measures are available in order to obtain the z-scores evolution of a given subject, but can handle missing data in the follow-up by finding the parameters that best match the available measures.

3.3.2 Progression model and latent relationships

We show in Figure 3.2 panel I) the dynamical relationships across the different z-scores estimated by SimulAD, where direction and intensity of the arrows quantify the estimated increase of one variable with respect to the other. Being the scores adimensional, they have been conveniently rescaled to the range [0,1] indicating increasing pathological levels. These relationships extend the summary statistics reported in Table 3.1 to a much finer temporal scale and wider range of possible biomarkers' values. We observe in Figure 3.2A, 3.2B and 3.2C that large values of the amyloid score z^{amy} trigger the increase of the remaining ones: z^{met}, z^{atr}, and z^{cli}. Figure 3.2D shows that large increase of the atrophy score z^{atr} is associated to pathological glucose metabolism indicated by large

values of z^{met}. Moreover, we note that high z^{met} values also contribute to an increase of z^{cli} (Figure 3.2E). Finally, Figure 3.2F shows that high atrophy values lead to an increase mostly along the clinical dimension z^{cli}. This chain of relationships is in agreement with the cascade hypothesis of AD [Jack, 2013a; Jack, 2013b].

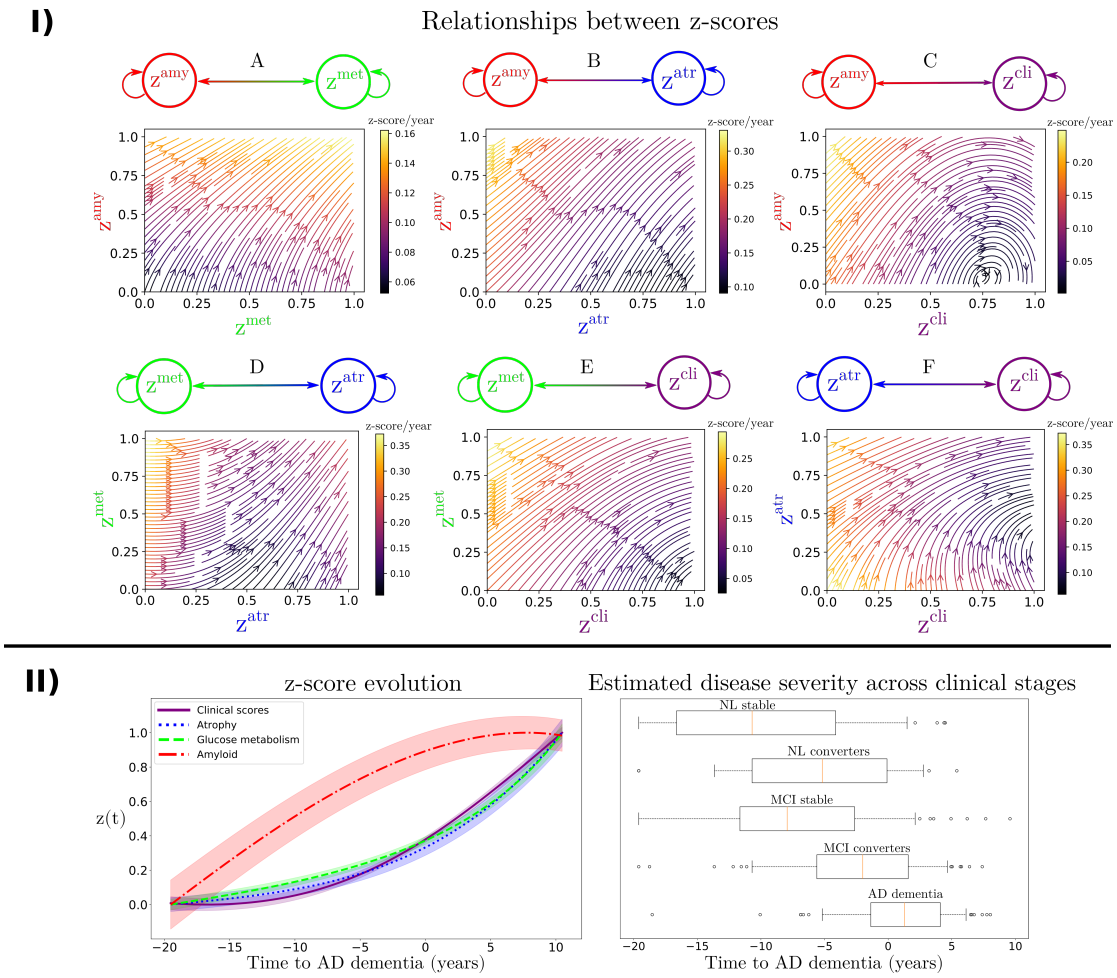

Figure 3.2: Panel I: Estimated dynamical relationships across the different z-scores (A to F). Given the values of two z-scores, the arrow at the corresponding coordinates indicates how one score evolves with respect to the other. The intensity of the arrow gives the strength of the relationship between the two scores. Panel II, left: Estimated long-term latent dynamics (time is relative to conversion to AD dementia). Shadowed areas represent the standard deviation of the average trajectory. Panel II, right: Distribution of the estimated disease severity across clinical stages, relatively to the long-term dynamics on the left. NL: normal individuals, MCI: mild cognitive impairment, AD dementia: Alzheimer's disease dementia.

Relying on the dynamical relationships shown in Figure 3.2 panel I), starting from any initial set of biomarkers values we can estimate the relative trajectories over time. Figure 3.2 panel II) (left), shows the evolution obtained by extrapolating backward and forward in time the trajectory associated to the z-scores of the AD dementia group. The x-axis represents the years from conversion to AD dementia, where the instant t=0 corresponds to the average time of diagnosis estimated for the group of MCI progressing to dementia.

As observed in Figure 3.2 panel I) and Table 3.1, the amyloid score z^{amy} increases and saturates first, followed by z^{met} and z^{atr} scores whose progression slows down when reaching clinical conversion, while the clinical score exhibits strong acceleration in the latest progression stages. Figure 3.2 panel II) (right) shows the group-wise distribution of the disease severity estimated for each subject relatively to the modelled long-term latent trajectories. The group-wise difference of disease severity across groups is statistically significant and increases when going from healthy to pathological stages (Wilcoxon-Mann-Whitney test $p < 0.01$ for each comparisons). The reliability of the estimation of disease severity was further assessed through testing on an independent cohort, and by comparison with a previously proposed disease progression modeling method from the state-of-the-art [Lorenzi, 2017]. The results are provided in Appendix 3.6.7 and show positive generalization results as well as a favorable comparison with the benchmark method.

From the z-score trajectories of Figure 3.2 panel II) (left) we predict the progression of imaging and clinical measures shown in Figure 3.3. We observe that amyloid load globally increases and saturates early, compatibly with the positive amyloid condition of the study cohort. Abnormal glucose metabolism and gray matter atrophy are delayed with respect to amyloid, and tend to map prevalently temporal and parietal regions. Finally, the clinical measures exhibit a non-linear pattern of change, accelerating during the latest progression stages. These dynamics are compatible with the summary measures on the raw data reported in Table 3.1.

3.3.3 Simulating clinical intervention

This experimental section is based on two intervention scenarios: a first one in which amyloid is lowered by 100%, and a second one in which it is reduced by 50% with respect to the estimated natural progression. In Figure 3.4 we show the latent z-scores evolution resulting from either 100% or 50% amyloid lowering performed at the time t=-20 years. According to these scenarios, intervention results in a sensitive reduction of the pathological progression for atrophy, glucose metabolism and clinical scores, albeit with a stronger effect in case of total blockage.

We further estimated the resulting clinical endpoints associated with the two amyloid lowering scenarios, at increasing time points and for different sample sizes. Clinical endpoints consisted in the simulated ADAS11, MMSE, FAQ, RAVLT immediate, RAVLT learning, RAVLT forgetting and CDRSB scores at the reference conversion time (t=0). The case placebo indicates the scenario where clinical values were computed at conversion time from the estimated natural progression shown in Figure 3.2 panel II) (left). Figure 3.5 shows the change in statistical power depending on intervention time and sample sizes. For large sample sizes (1000 subjects per arm) a power greater than 0.8 can be

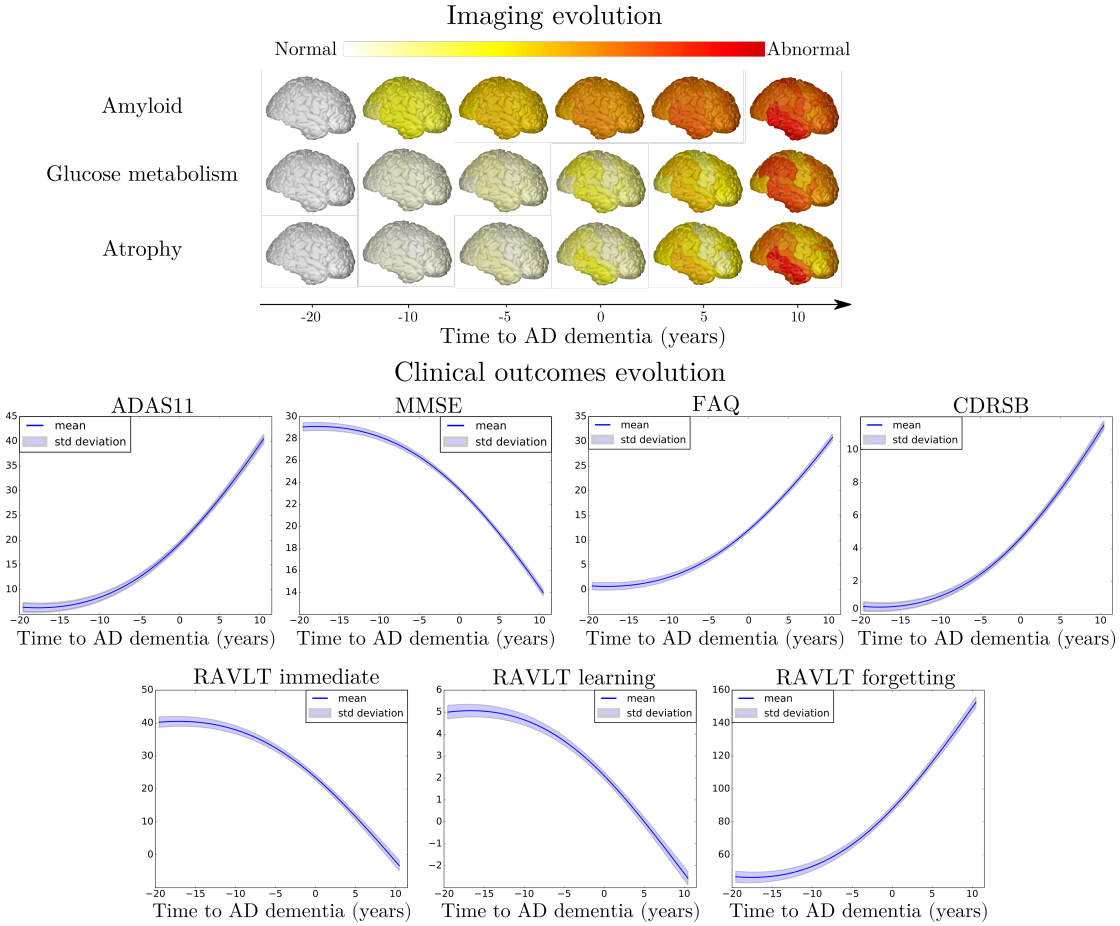

Figure 3.3: Estimated long-term evolution of cortical measurements for the different types of imaging markers, and clinical scores. Shadowed areas represent the standard deviation of the average trajectory. Brain images were generated using the software provided in [Marinescu, 2019b].

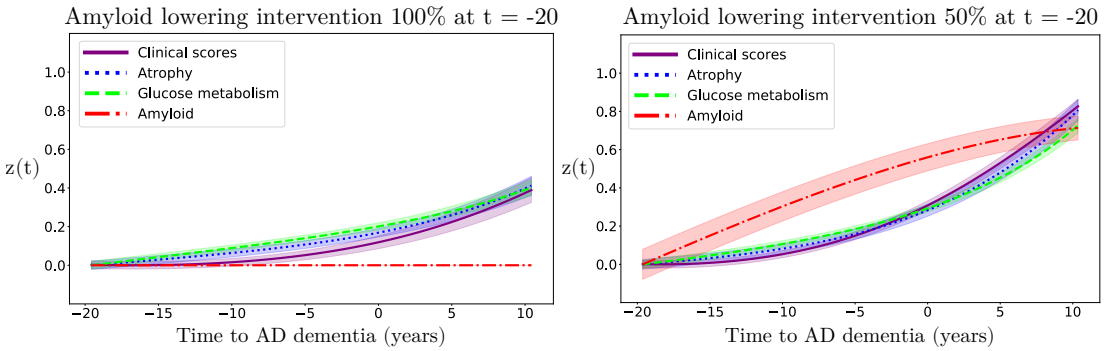

Figure 3.4: Hypothetical scenarios of irreversible amyloid lowering interventions at t=-20 years from Alzheimer's disease dementia diagnosis, with a rate of 100% (left) or 50% (right). Shadowed areas represent the standard deviation of the average trajectory.

obtained around 7 years before conversion, depending on the outcome score, where in general we observe that RAVLT forgetting exhibits a higher power than the other scores. When sample size is lower than 100 subjects per arm, a power greater than 0.8 is reached if intervention is performed at the latest 11 years before conversion, with a mild variability depending on the considered clinical score. We notice that in the case of 50% amyloid lowering, in order to reach the same power intervention needs to be consistently performed earlier compared to the scenario of 100% amyloid lowering for the same sample size and clinical score. For instance, if we consider ADAS11 with a sample size of 100 subjects per arm, a power of 0.8 is obtained for a 100% amyloid lowering intervention performed 11.5 years before conversion, while in case of a 50% amyloid lowering the equivalent effect would be obtained by intervening 15 years before conversion.

We provide in Table 3.2 the estimated improvement for each clinical score at conversion with a sample size of 100 subjects per arm for both 100% and 50% amyloid lowering depending on the intervention time. We observe that for the same intervention time, 100% amyloid lowering always results in a larger improvement of clinical endpoints compared to 50% amyloid lowering. We also note that in the case of 100% lowering, clinical endpoints obtained for intervention at t=-15 years correspond to typical cutoff values for inclusion into AD trials (ADAS11 $= 13.7 \pm 5.9$, MMSE $= 25.7 \pm 2.5$, see Table 3.4 in Appendix 3.6.8) [Kochhann, 2010; Gamberger, 2017].

3.4 Discussion

We presented SimulAD, a framework to jointly model the progression of multi-modal imaging and clinical data, based on the estimation of latent biomarkers' relationships governing AD progression. The model is designed to simulate intervention scenarios in clinical trials, and in this study we focused on assessing the effect of anti-amyloid drugs on biomarkers' evolution, by quantifying the effect of intervention time and drug efficacy on clinical outcomes. Our results underline the critical importance of intervention time, which should be performed sensibly early during the pathological history to effectively appreciate the effectiveness of disease modifiers.

The results obtained with our model are compatible with findings reported in recent clinical studies [Honig, 2018; Egan, 2019; Wessels, 2019]. For example, if we consider 500 patients per arm and perform a 100% amyloid lowering intervention for 2 years to reproduce the conditions of the recent trial of Verubecestat [Egan, 2019], the average improvement of MMSE predicted by our model is of 0.02, falling in the 95% confidence interval measured during that study ([-0.5 ; 0.8]). While recent anti-amyloid trials such as [Honig, 2018; Egan, 2019; Wessels, 2019] included between 500 and 1000 mild AD

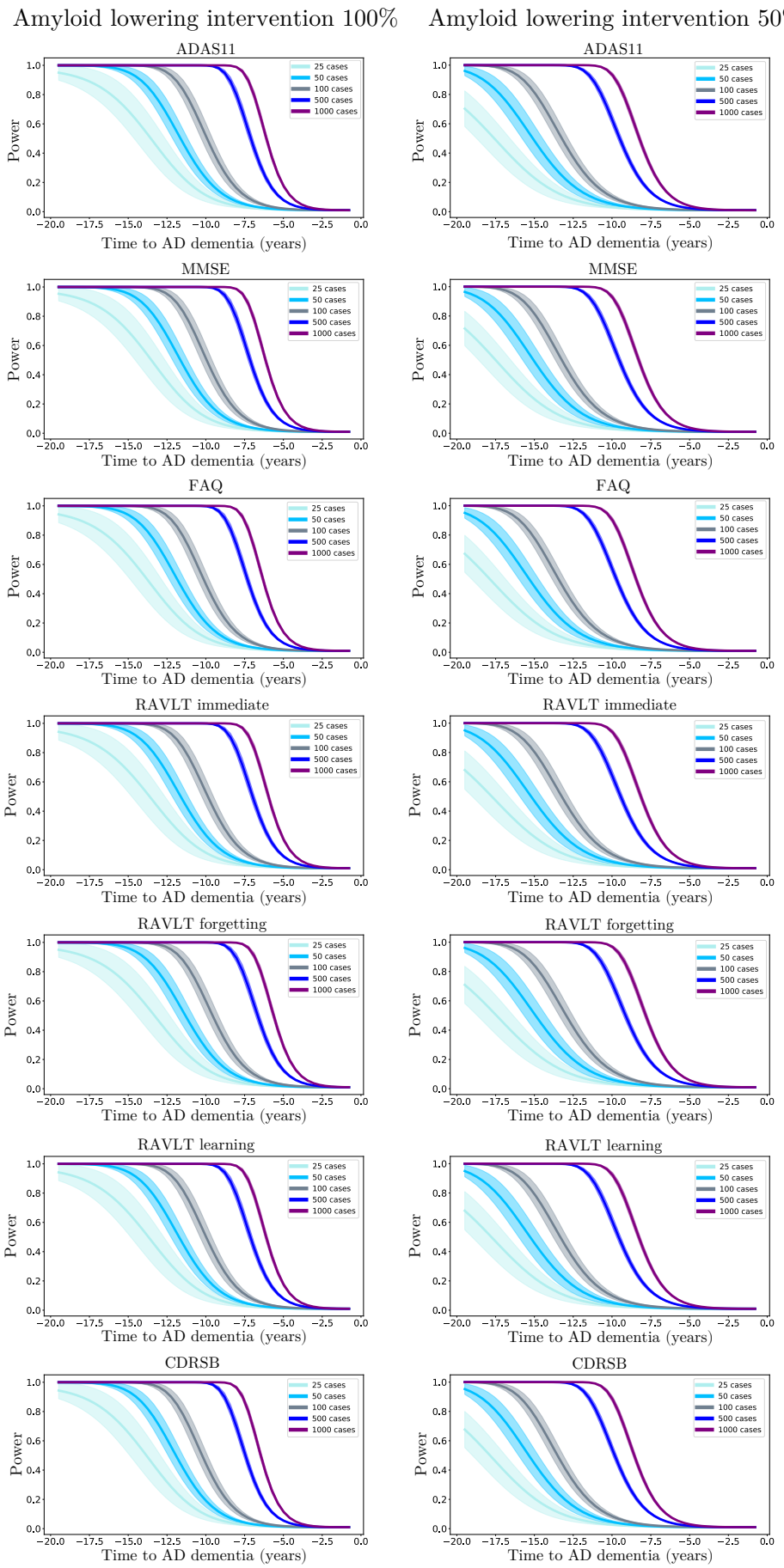

Figure 3.5: Statistical power of the Student t-test comparing the estimated clinical outcomes at conversion time between placebo and treated scenarios, according to the year of simulated intervention, lowering rate and sample size.

Table 3.2: Estimated mean (standard deviation) improvement of clinical outcomes at predicted conversion time for the normal progression case by year of simulated intervention (100% and 50% amyloid lowering interventions). Results in bold indicate a statistically significant difference between placebo and treated scenarios ($p<0.01$, two-sided t-test, 100 cases per arm). AD: Alzheimer's disease dementia, ADAS11: Alzheimer's Disease Assessment Scale, MMSE: Mini-Mental State Examination, FAQ: Functional Assessment Questionnaire, RAVLT: Rey Auditory Verbal Learning Test, CDRSB: Clinical Dementia rating Scale Sum of Boxes.

Amyloid lowering intervention 100%
Point improvement per intervention time

Score \ Years to AD	−20	−15	−12.5	−10	−5	−3	−2	−1
ADAS11	**11.1 (6.4)**	**5.2 (2.9)**	**3.0 (1.7)**	**1.6 (1.0)**	0.3 (0.2)	0.1 (0.1)	0.0 (0.0)	0.0 (0.0)
MMSE	**4.9 (2.8)**	**2.3 (1.3)**	**1.3 (0.8)**	**0.7 (0.4)**	0.1 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
FAQ	**9.6 (5.6)**	**4.5 (2.5)**	**2.6 (1.5)**	**1.4 (0.8)**	0.2 (0.2)	0.1 (0.1)	0.0 (0.0)	0.0 (0.0)
RAVLT immediate	**15.3 (8.9)**	**7.2 (4.1)**	**4.2 (2.4)**	**2.3 (1.4)**	0.5 (0.3)	0.2 (0.1)	0.1 (0.1)	0.0 (0.0)
RAVLT learning	**2.7 (1.6)**	**1.3 (0.7)**	**0.7 (0.4)**	**0.4 (0.2)**	0.1 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
RAVLT forgetting	**37.2 (21.5)**	**17.7 (9.9)**	**10.5 (6.0)**	**5.8 (3.5)**	1.3 (0.9)	0.5 (0.4)	0.2 (0.2)	0.1 (0.1)
CDRSB	**3.5 (2.0)**	**1.6 (0.9)**	**0.9 (0.5)**	**0.5 (0.3)**	0.1 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)

Amyloid lowering intervention 50%
Point improvement per intervention time

Score \ Years to AD	−20	−15	−12.5	−10	−5	−3	−2	−1
ADAS11	**5.0 (2.8)**	**2.4 (1.3)**	**1.4 (0.8)**	0.8 (0.5)	0.2 (0.1)	0.1 (0.0)	0.0 (0.0)	0.0 (0.0)
MMSE	**2.2 (1.2)**	**1.1 (0.6)**	**0.6 (0.4)**	0.3 (0.2)	0.1 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
FAQ	**4.3 (2.4)**	**2.1 (1.1)**	**1.2 (0.7)**	0.7 (0.4)	0.1 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
RAVLT immediate	**6.9 (3.8)**	**3.4 (1.9)**	**2.0 (1.1)**	1.1 (0.7)	0.2 (0.2)	0.1 (0.1)	0.0 (0.0)	0.0 (0.0)
RAVLT learning	**1.2 (0.7)**	**0.6 (0.3)**	**0.3 (0.2)**	0.2 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
RAVLT forgetting	**16.7 (9.2)**	**8.3 (4.6)**	**5.0 (2.8)**	2.8 (1.7)	0.6 (0.4)	0.2 (0.2)	0.1 (0.1)	0.0 (0.0)
CDRSB	**1.6 (0.9)**	**0.8 (0.4)**	**0.4 (0.2)**	0.2 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)

dementia subjects per arm and were conducted over a period of two years at most, our analysis suggests that clinical trials performed with less than 1000 subjects with mild AD dementia may be consistently under-powered. Indeed, we see in Figure 3.5 that with a sample size of 1000 subjects per arm and a total blockage of amyloid production, a power of 0.8 can be obtained only if intervention is performed at least 7 years before conversion.

These results allow to quantify the crucial role of intervention time, and provide a theoretical justification for testing amyloid modifying drugs in the pre-clinical stage [Sperling, 2011; Aisen, 2018]. This is for example illustrated in Table 3.2, in which we notice that clinical endpoints are close to placebo even when the simulated intervention takes place 10 years before conversion, while stronger cognitive and functional changes happen when amyloid is lowered by 100% or 50% earlier. These findings may be explained by considering that amyloid accumulates over more than a decade, and that when amyloid clearance occurs the pathological cascade is already entrenched [Rowe, 2010]. Our results are thus supporting the need to identify subjects at the pre-clinical stage, that is to say still cognitively normal, which is a challenging task. Currently, one of the main criteria to enroll subjects into clinical trials is the presence of amyloid in the brain, and blood-based markers are considered as potential candidates for identifying patients at risk for AD [Zetterberg, 2019]. Moreover, recent works such as [Blennow, 2010; Westwood, 2016] have proposed more complex entry criteria to constitute cohorts based on multi-modal measurements. Within this context, our model could also be used as an enrichment tool by quantifying the disease severity based on multi-modal data as shown in Figure 3.2 panel II) (right). Similarly, SimulAD could be applied to predict the evolution of single patient given its current available measurements.

An additional critical aspect of anti-amyloid trials is the effect of dose exposure on the production of amyloid [Klein, 2019]. Currently, β-site amyloid precursor protein cleaving enzyme (BACE) inhibitors allow to suppress amyloid production from 50% to 90%. In this study we showed that lowering amyloid by 50% consistently decreases the treatment effect compared to a 100% lowering at the same time. For instance, if we consider a sample size of 1000 subjects per arm in the case of a 50% amyloid lowering intervention, 80% power can be reached only 10 years before conversion instead of 7 years for a 100% amyloid lowering intervention. This ability of SimulAD to control the rate of amyloid progression is fundamental in order to provide realistic simulations of anti-amyloid trials.

In Figure 3.2 panel I) we showed that amyloid triggers the pathological cascade affecting the other markers, thus confirming its dominating role on disease progression. Assuming that the data used to estimate the model is sufficient to completely depict the history of the pathology, our model can be interpreted from a causal perspective. However, we cannot exclude the existence of other mechanisms driving amyloid accumulation, which our

model cannot infer from the existing data. Therefore, our findings should be considered with care, while the integration of additional biomarkers of interest will be necessary to account for multiple drivers of the disease. It is worth noting that recent works ventured the idea to combine drugs targeting multiple mechanisms at the same time [Gauthier, 2019]. For instance, pathologists have shown tau deposition in brainstem nuclei in adolescents and children [Kaufman, 2018], and clinicians are currently investigating the pathological effect of early tau spreading on AD progression [Pontecorvo, 2019], raising crucial questions about its relationship with amyloid accumulation, and the impact on cognitive impairment [Cummings, 2019a]. In this study, 190 subjects underwent at least one Tau-PET scan. However, when considering the subjects for whom there exists one visit in which all the data modalities were available, the number of patients in the study cohort decreased to 33. This low sample size prevented us from estimating reliable trajectories for this biomarker. It is also important to note that among the 190 subjects with at least one Tau-PET scan, only 19 of them had one follow-up visit. This means that tau markers dynamics cannot be reliably estimated. Including tau data will require studies on larger cohorts with complete sets of PET imaging acquisitions. This could be part of future extensions of this work, where the inclusion of tau markers will allow to simulate scenarios of production blockage of both amyloid and tau at different rates or intervention time.

Lately, disappointing results of clinical studies led to hypothesize specific treatments targeting AD dementia sub-populations based on their genotype [Safieh, 2019]. While in our work we describe a global progression of AD, in the future we will account for sub-trajectories due to genetic factors, such as the presence of 4 allele of apolipoprotein (APOE4), which is a major risk for developing AD influencing both disease onset and progression [Kim, 2009]. This could be done by estimating dynamical systems specific to the genetic condition of each patient. This was not possible in this study due to a strong imbalance between the number of carriers and non-carriers across the different clinical groups (cf. Table 3.1). Indeed, we observe that the number of ADNI non-carriers is much lower than the number of carriers, especially in the latest stages of the disease (MCI converters and AD dementia). On the contrary, the majority of NL stable subjects are non-carriers. Therefore, applying the model in such conditions would lead to a bias towards more represented groups during the different stages of the disease progression (APOE4- at early stages and APOE4+ at late ones), thus preventing us from differentiating the biomarkers dynamics based on the genetic status. Yet, simulating dynamical relationships specific to genetic factors is a crucial avenue of improvement of our approach, as it would allow to evaluate the effect of APOE4 on intervention time or drug dosage. In addition to this example, there exist numerous non-genetic aggravating factors that may also affect disease evolution, such as diabetes, obesity or smoking. Extending our model to account for panels of risk factors would ultimately allow to test in silico personalized intervention strategies. Moreover, a key aspect of clinical trials is their economic cost. SimulAD could be extended to help designing clinical trials by optimizing intervention with respect to

the available funding. Given a budget, we could simulate scenarios based on different sample size, and trials duration, while estimating the expected cognitive outcome.

Results presented in this work are based on a model estimated by relying solely on a subset of subjects and measures from the ADNI cohort, and therefore they may not be fully representative of the general AD progression. Indeed, subjects included in this cohort were either amyloid-positive at baseline, or became amyloid-positive during their follow-up visits. This was motivated by the consideration that evidence of pathological amyloid levels is a necessary condition for diagnosing AD as it puts subjects within the "Alzheimer's disease continuum" [Jack, 2018]. By narrowing the list of subjects to a subgroup of amyloid positive we increase the chances of selecting a set of patients likely to develop the disease. Moreover, the inclusion of subjects at various clinical stages allows to span the entire spectrum of morphological and physiological changes affecting the brain. Through the joint analysis of markers of amyloid, neurodegeneration and cognition, SimulAD estimates the average trajectory that best describes the progression of the observed measures when going from NL individuals towards AD dementia patients. The selection of amyloid positive patients aims at increasing the signal of Alzheimer's pathological changes within this cohort, in order to estimate long-term dynamics for the biomarkers that can be associated to the disease. We believe that this modeling choice is based on a clinically plausible rationale, and allows us to perform our study on a sufficiently large cohort enabling the estimation of our model.

Bearing this in mind, we acknowledge the potential presence of bias towards the specific inclusion criterion adopted in this work. Indeed, the present results may provide a limited representation of the pathological temporal window captured by the model. For example, applying SimulAD on a cohort containing amyloid-negative subjects may provide additional insights on the overall disease history. However, this is a challenging task as it would require to identify sub-trajectories dissociated from normal ageing [Lorenzi, 2015a; Sivera, 2020]. Another potential bias affecting the results may come from the choice of the clinical scores used to estimate our model. In this study, we relied on a panel of 7 neuro-psychological assessments providing a comprehensive representation of cognitive, memory and functional abilities: ADAS11, MMSE, RAVLT immediate, RAVLT learning, RAVLT forgetting, FAQ, and CDRSB. The choice of these particular scores is consistent with previous literature on DPM [Donohue, 2014; Lorenzi, 2017]. However, it is important to note that SimulAD can handle any type of clinical assessment. Therefore, investigating the effect of adding supplementary clinical scores on the model's findings would be an interesting future application of our approach, and could be done without any modification of its current formulation. Finally, in addition to these specific characteristics of the cohort, there exists additional biases impacting the model estimation. For instance, the fact that gray matter atrophy and glucose metabolism become abnormal approximately at the same time in Figure 3.3 can be explained by the high atrophy rate of change in some key regions in normal elders, such as in the

hippocampus, compared to the rate of change of FDG (see Table 3.1). We note that this stronger change of atrophy with respect to glucose metabolism can already be appreciated in the clinically healthy group.

3.5 Conclusion

In this study we introduced SimulAD, a novel quantitative instrument for the development of intervention strategies for disease modifying drugs in AD. Our framework enables the simulation of the effect of intervention time and drug dosage on the evolution of imaging and clinical biomarkers in clinical trials. The proposed data-driven approach is based on the modeling of the spatio-temporal dynamics governing the joint evolution of imaging and clinical measurements throughout the disease. The model is formulated within a Bayesian framework, where the latent representation and dynamics are efficiently estimated through stochastic variational inference. To generate hypothetical scenarios of amyloid lowering interventions, we applied SimulAD to multi-modal imaging and clinical data from ADNI. The results quantify the crucial role of intervention time, and provide a theoretical justification for testing amyloid modifying drugs in the pre-clinical stage. Our experimental simulations are compatible with the outcomes observed in past clinical trials and suggest that anti-amyloid treatments should be administered at least 7 years earlier than what is currently being done in order to obtain statistically powered improvement of clinical endpoints.

3.6 Appendix

3.6.1 Variational inference

We rewrite $p(\mathbf{X}_i(t)|\mathbf{z}_i(t), \boldsymbol{\sigma^2}, \boldsymbol{\psi})$ as $p(\mathbf{X}_i(t)|\mathbf{z}_i(t_0), \theta_{ODE}, \boldsymbol{\sigma^2}, \boldsymbol{\psi})$. Assuming independence between subjects, the marginal log-likelihood writes as:

$$\begin{aligned}\mathcal{L} &= \sum_{i}^{N} \log\left[p(\mathbf{X}_i(t)|\theta_{ODE}, \boldsymbol{\sigma}^2, \boldsymbol{\psi})\right] \\ &= \sum_{i}^{N} \log\left[\int p(\mathbf{X}_i(t)|\mathbf{z}_i(t_0), \theta_{ODE}, \boldsymbol{\sigma}^2, \boldsymbol{\psi}) p(\mathbf{z}_i(t_0)) d\mathbf{z}_i(t_0)\right].\end{aligned} \tag{3.6}$$

For ease of notation, we drop the *i* index, and dependence on *t* and t_0 is made implicit. Within a Bayesian framework, we wish to maximize $\mathcal{L}$ in order to obtain a posterior distribution for the latent variable $\mathbf{z}$. Since derivation of this quantity is generally not tractable, we resort to stochastic variational inference to tackle the optimization problem. We assume a $\mathcal{N}(\mathbf{0}, \mathbf{I})$ prior for $p(\mathbf{z})$, and introduce an approximate posterior distribution $q(\mathbf{z}|\mathbf{X})$ [Ghahramani, 2001], in order to derive a lower-bound (ELBO) $\mathcal{E}$ for the marginal log-likelihood:

$$\begin{aligned}\log p(\mathbf{X}|\theta_{ODE}, \boldsymbol{\sigma}^2, \boldsymbol{\psi}) &\geq \mathbb{E}_{q(\mathbf{z}|\mathbf{X})}\Big[\log p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \boldsymbol{\sigma^2}, \boldsymbol{\psi})\Big] \\ &\quad - \mathcal{D}\Big[q(\mathbf{z}|\mathbf{X})|p(\mathbf{z})\Big] \\ &= \mathcal{E},\end{aligned} \tag{3.7}$$

where $\mathcal{D}$ refers to the Kullback-Leibler (KL) divergence. We propose to factorize the distribution $q(\mathbf{z}|\mathbf{X})$ across modalities such that, $q(\mathbf{z}|\mathbf{X}) = \prod_m q(z^m|\mathbf{x}^m)$, where $q(z^m|\mathbf{x}^m) = \mathcal{N}(f(\mathbf{x}^m, \phi_m^1), h(\mathbf{x}^m, \phi_m^2))$, is a variational Gaussian approximation with moments parameterized by the functions f and h. This modality-wise encoding of the data enables to interpret each coordinate of $\mathbf{z}$ as a compressed representation of the corresponding modality. Moreover, the lower-bound simplifies as:

$$\mathcal{E} = \sum_m \mathbb{E}_{q(\mathbf{z}|\mathbf{X})}\Big[\log p(\mathbf{x}^m|\mathbf{z}, \theta_{ODE}, \sigma_m^2, \psi_m)\Big] - \mathcal{D}\Big[q(z^m|\mathbf{x}^m)|p(z^m)\Big]. \tag{3.8}$$

Details about the ELBO derivation and the computation of the KL divergence are given in Appendices 3.6.3 and 3.6.4. A graphical model of the method is also provided in Figure 3.6, while Algorithm 4 details the steps to compute the ELBO.

3.6.2 Model optimization

Using the reparameterization trick [Kingma, 2013], we can efficiently sample from the posterior distribution $q(\mathbf{z}(t_0)|\mathbf{X}(t_0))$ to approximate the expectation terms. Moreover, thanks to our choices of priors and approximations the KL terms can be computed in closed-form. In practice, we sample from $q(\mathbf{z}(t_0)|\mathbf{X}(t_0))$ to obtain a latent representation $\mathbf{z}(t_0)$ at baseline, while the follow-up points are estimated by decoding the latent time-series obtained through the integration of the ODEs of Equation 3.3. The model is trained by computing the total ELBO for all the subjects at all the available time points. The parameters $\psi, \phi^1, \phi^2, \theta_{ODE}, \sigma$ are optimized using gradient descent, which requires to backpropagate through the integration operation.

In order to enable backpropagation through the ODEs integration we need to numerically solve the differential equation using only operations that can be differentiated. In this work, we used the Midpoint method which follows a second order Runge-Kutta scheme. The method consists in evaluating the derivative of the solution at $(t_{i+1} + t_i)/2$, which is the midpoint between t_i at which the correct $\mathbf{z}(t)$ is evaluated, and the following t_{i+1}:

$$\begin{aligned}\int_{t_i}^{t_{i+1}} g(\mathbf{z}(x))dx &\approx h \cdot g\Big(\mathbf{z}(\frac{t_i + t_{i+1}}{2})\Big) \\ &\approx h \cdot g\Big(\mathbf{z}(t_i) + \frac{h}{2} g(\mathbf{z}(t_i))\Big), \;\; h = t_{i+1} - t_i.\end{aligned} \tag{3.9}$$

Therefore, solving the system of Equation 3.3 on the interval $[t_0, ..., t]$ only requires operations that can be differentiated, allowing to compute the derivatives of the ELBO with respect to all the parameters, and to optimize them by gradient descent. Moreover, in order to control the variability of the estimated latent trajectory $\mathbf{z}(t)$ due to the error propagation during integration, we initialized the weights of ϕ^1 and ϕ^2 such that the approximate posterior of the latent representation for each modality *m* at baseline was following a $\mathcal{N}(0, 0.01)$ distribution. Finally, we also tested other ODE solvers such as Runge-Kutta 4, which gave similar results than the Midpoint method with a slower execution time due its more expensive approximation scheme.

Concerning the implementation, we trained the model using the ADAM optimizer [Kingma, 2015a] with a learning rate of 0.01. The functions f, h and μ_m were parameterized as linear transformations. The model was implemented in Pytorch [Paszke, 2019], and we used the *torchdiffeq* package developed in [Chen, 2018] to backpropagate through the ODE solver.

3.6.3 Lower bound

We provide here the detailed derivation to obtain the ELBO of Equation 3.8.

$$\begin{aligned}
\log p(\mathbf{X}|\boldsymbol{\sigma}^2, \boldsymbol{\psi}) &= \log \Big[\int p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \boldsymbol{\sigma^2}, \boldsymbol{\psi}) p(\mathbf{z}) d\mathbf{z} \Big] \\
&= \log \Big[\int p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \boldsymbol{\sigma^2}, \boldsymbol{\psi}) p(\mathbf{z}) \frac{q(\mathbf{z}|\mathbf{X})}{q(\mathbf{z}|\mathbf{X})} d\mathbf{z} \Big] \\
&= \log \Big[\mathbb{E}_{q(\mathbf{z}|\mathbf{X})} \frac{p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \boldsymbol{\sigma^2}, \boldsymbol{\psi}) p(\mathbf{z})}{q(\mathbf{z}|\mathbf{X})} \Big] \\
&\overset{\text{Jensen}}{\geq} \mathbb{E}_{q(\mathbf{z}|\mathbf{X})} \Big[\log \frac{p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \boldsymbol{\sigma^2}, \boldsymbol{\psi}) p(\mathbf{z})}{q(\mathbf{z}|\mathbf{X})} \Big] \\
&= \mathbb{E}_{q(\mathbf{z}|\mathbf{X})} \Big[\log p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \boldsymbol{\sigma^2}, \boldsymbol{\psi}) \Big] - \mathcal{D}\Big[q(\mathbf{z}|\mathbf{X}) | p(\mathbf{z}) \Big] \\
&= \mathcal{E}.
\end{aligned} \tag{3.10}$$

Given that:

$$p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \boldsymbol{\sigma^2}, \boldsymbol{\psi}) = \prod_m p(\mathbf{x}^m|\mathbf{z}, \theta_{ODE}, \sigma_m^2, \psi_m),\ q(\mathbf{z}|\mathbf{X}) = \prod_m q(z^m|\mathbf{X}),\ \text{and}\ p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{3.11}$$

We obtain:

$$\mathcal{E} = \sum_m E_{q(\mathbf{z}|\mathbf{X})} \Big[\log p(\mathbf{x}^m|\mathbf{z}, \theta_{ODE}, \sigma_m^2, \psi_m) \Big] - \mathcal{D}\Big[q(z^m|\mathbf{x}^m) | p(z^m) \Big]. \tag{3.12}$$

3.6.4 KL divergence

We have that:

$$\begin{aligned}
q(z^m|\mathbf{X}) &= \mathcal{N}(f(\mathbf{x}^m, \phi_m^1), h(\mathbf{x}^m, \phi_m^2)), \\
p(z^m) &= \mathcal{N}(0, 1).
\end{aligned} \tag{3.13}$$

We use the closed-form formula to calculate the KL divergence between two normal distributions:

$$\begin{aligned}
\mathcal{D}\Big[q(\mathbf{z}|\mathbf{X}) | p(\mathbf{z}) \Big] &= \sum_m \mathcal{D}\Big[q(z^m|\mathbf{x}^m) | p(z^m) \Big] \\
&= \frac{1}{2} \sum_m \Big[-\log(h(\mathbf{x}^m, \phi_m^2)) - 1 + h(\mathbf{x}^m, \phi_m^2) + f(\mathbf{x}^m, \phi_m^1)^2 \Big].
\end{aligned} \tag{3.14}$$

3.6.5 Graphical model

Figure 3.6 below provides the graphical model illustrating the method presented in Section 3.2.3.

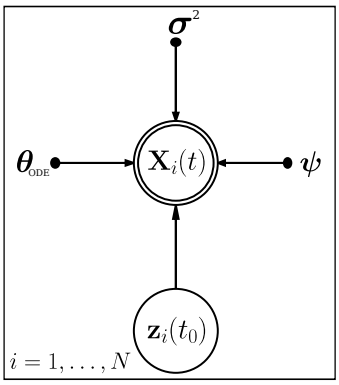

Figure 3.6: Graphical model of the proposed method.

3.6.6 Lower bound computation

Algorithm 4 below details the steps to compute the lower-bound for a given subject i at time t.

Algorithm 4 Forward pass to compute the lower-bound for a given subject i at time t.

function COMPUTE_ELBO($\mathbf{X}(t), \mathbf{X}(t_0), \theta_{ODE}, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\sigma}^2$)
For ease of notation we drop the i index in the pseudo-code.
    Sample $\mathbf{z}(t_0) \sim q(\mathbf{z}(t_0)|\mathbf{X}(t_0)) = \prod_m \mathcal{N}(f(\mathbf{x}^m(t_0), \phi_m^1), h(\mathbf{x}^m(t_0), \phi_m^2))$ ▷ Baseline latent representation (reparameterization trick).
    Compute $\mathbf{z}(t) = MIDPOINT(\mathbf{z}(t_0), g, \theta_{ODE}, t)$ ▷ Predict latent representation at time t by numerically solving the ODEs system.
    Compute $\mathbb{E}_{q(\mathbf{z}(t_0)|\mathbf{X}(t_0))}\Big[\log p(\mathbf{x}^m(t)|\mathbf{z}(t), \theta_{ODE}, \sigma_m^2, \psi_m)\Big] \approx -\frac{D_m}{2}\log(2\pi\sigma_m^2) - \frac{1}{2\sigma_m^2}||\mathbf{x}^m(t) - \mu_m(\mathbf{z}(t))||^2$ ▷ Expectation term Equation 3.8.
    Compute $\mathcal{D}\Big[q(z^m(t_0)|\mathbf{x}^m(t_0))|p(z^m(t_0))\Big] = \frac{1}{2}\Big[-\log(h(\mathbf{x}^m(t_0), \phi_m^2)) - 1 + h(\mathbf{x}^m(t_0), \phi_m^2) + f(\mathbf{x}^m(t_0), \phi_m^1)^2\Big]$ ▷ KL divergence Equation 3.8.
    Compute $\mathcal{E} = \sum_m \mathbb{E}_{q(\mathbf{z}(t_0)|\mathbf{X}(t_0))}\Big[\log p(\mathbf{x}^m(t)|\mathbf{z}(t), \theta_{ODE}, \sigma_m^2, \psi_m)\Big] - \mathcal{D}\Big[q(z^m(t_0)|\mathbf{x}^m(t_0))|p(z^m(t_0))\Big]$.
    Return $\mathcal{E}$
end function

3.6.7 Time-shift comparison and validation

We compared our estimated disease severity (Figure 3.2 panel II right) with the one obtained applying the monotonic Gaussian Process (GP) model of [Lorenzi, 2017] from the state-of-the-art (Figure 3.7A). While both methods estimate significant time differences when going from healthy to pathological stages, SimulAD captures a larger temporal variability for both earlier and later stages of the disease, as shown in Figure 3.7B, highlighting a stronger separability across clinical stages.

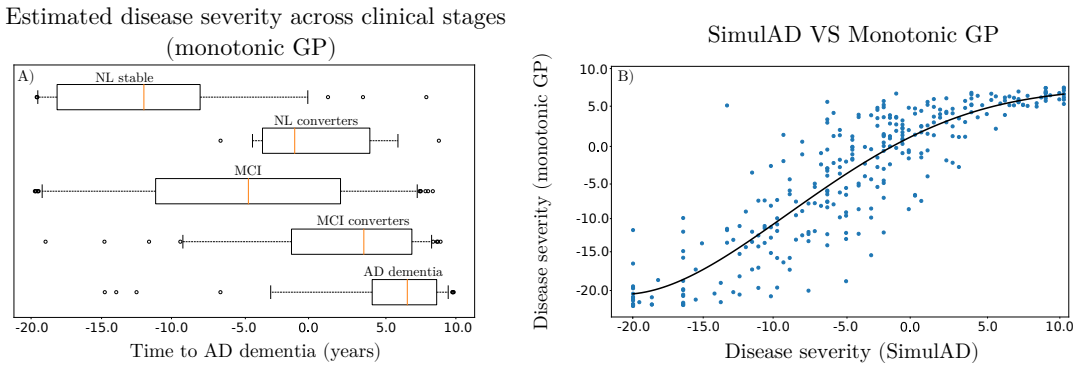

Figure 3.7: A: Distribution of the disease severity estimated by the monotonic GP method [Lorenzi, 2017] on the training set. NL: normal individuals, MCI: mild cognitive impairment, AD dementia: Alzheimer's disease dementia. B: Comparison of the disease severity estimated by SimulAD with respect to the one estimated by the monotonic GP.

We also tested SimulAD on an independent testing cohort from the ADNI composed of 130 NL stable, 10 NL converters, 125 MCI stable, 7 MCI converters, and 12 AD dementia subjects which were not necessarily amyloid positive. It is important to note that no PET-FDG data was available for these subjects. We provide in Table 3.3 socio-demographic and clinical information for the testing cohort across the different clinical groups. Despite the fact that no FDG data was used to estimate the disease severity, we observe in Figure 3.8 that the method still exhibits good separating performances between clinical stages, coherently with the clinical status of the testing individuals.

3.6.8 Simulated clinical endpoints

We provide in Table 3.4 the estimated values for each clinical score at predicted conversion time for the normal progression case when performing the simulations presented in Section 3.3.3.

Table 3.3: Baseline socio-demographic and clinical information for testing cohort (284 subjects for 2116 data points). Average values, standard deviation in parenthesis. NL: normal individuals, MCI: mild cognitive impairment, AD dementia: Alzheimer's disease dementia. ADAS11: Alzheimer's Disease Assessment Scale-cognitive subscale, 11 items. AV45: (18)F-florbetapir Amyloid PET imaging. SUVR: Standardized Uptake Value Ratio.

	NL stable	NL converters	MCI stable	MCI converters	AD dementia
N	130	10	125	7	12
Age (yrs)	72 (6)	74 (8)	71 (8)	73 (9)	78 (6)
Education (yrs)	17 (2)	16 (2)	16 (3)	14 (3)	17 (2)
ADAS11	5.4 (2.8)	7.7 (4.1)	7.8 (3.3)	14.3 (5.2)	15.0 (6.7)
WholeBrain (cm^3)	1063 (103)	1104 (98)	1054 (97)	966 (104)	1010 (108)
AV45 (SUVR)	0.9 (0.1)	1.0 (0.1)	1.0 (0.1)	1.1 (0.2)	1.2 (0.3)

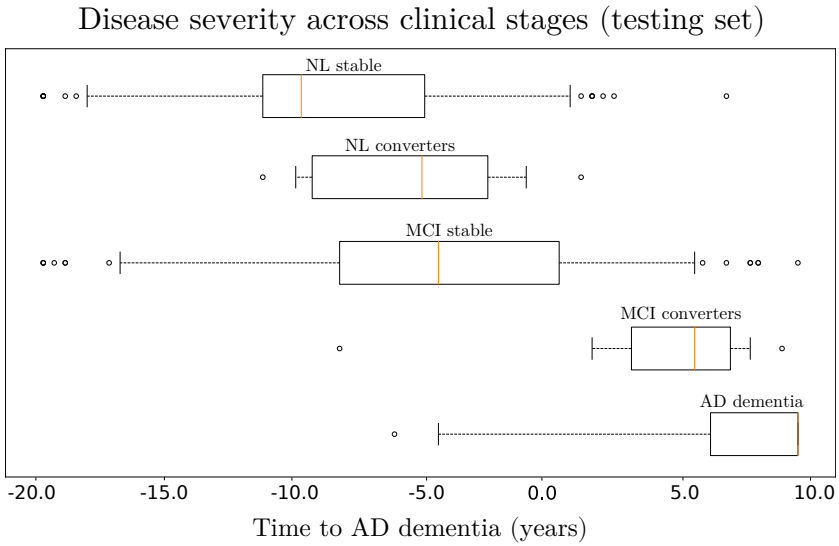

Figure 3.8: Distribution of the disease severity estimated for the subjects of the testing set, relatively to the long-term dynamics of Figure 3.2 panel II) (left) in the manuscript. NL: normal individuals, MCI: mild cognitive impairment, AD dementia: Alzheimer's disease dementia.

Table 3.4: Estimated mean (standard deviation) of the clinical outcomes at predicted conversion time for the normal progression case by year of simulated intervention (100% and 50% amyloid lowering interventions). Results in bold indicate a statistically significant difference between placebo and treated scenarios ($p<0.01$, two-sided t-test, 100 cases per arm). AD: Alzheimer's disease dementia, ADAS11: Alzheimer's Disease Assessment Scale, MMSE: Mini-Mental State Examination, FAQ: Functional Assessment Questionnaire, RAVLT: Rey Auditory Verbal Learning Test, CDRSB: Clinical Dementia Rating Scale Sum of Boxes.

Amyloid lowering intervention 100%

Score per intervention time

Score \ Years to AD	−20	−15	−12.5	−10	−5	−3	−2	−1
ADAS11	**7.8 (8.9)**	**13.7 (5.9)**	**15.9 (5.0)**	**17.3 (4.5)**	18.6 (4.2)	18.8 (4.1)	18.8 (4.1)	18.9 (4.1)
MMSE	**28.3 (3.8)**	**25.7 (2.5)**	**24.8 (2.1)**	**24.2 (2.0)**	23.6 (1.8)	23.5 (1.8)	23.5 (1.8)	23.5 (1.8)
FAQ	**2.3 (7.9)**	**7.5 (5.2)**	**9.3 (4.5)**	**10.5 (4.0)**	11.7 (3.7)	11.8 (3.7)	11.9 (3.7)	11.9 (3.7)
RAVLT immediate	**39.2 (12.7)**	**31.0 (8.4)**	**28.1 (7.1)**	**26.1 (6.4)**	24.3 (5.8)	24.0 (5.7)	23.9 (5.7)	23.9 (5.7)
RAVLT learning	**4.9 (2.2)**	**3.4 (1.5)**	**2.9 (1.2)**	**2.6 (1.1)**	2.2 (1.0)	2.2 (1.0)	2.2 (1.0)	2.2 (1.0)
RAVLT forgetting	**49.9 (29.8)**	**69.4 (19.7)**	**76.6 (16.8)**	**81.3 (15.1)**	85.9 (13.7)	86.6 (13.5)	86.9 (13.4)	87.0 (13.4)
CDRSB	**1.0 (2.8)**	**2.9 (1.9)**	**3.6 (1.6)**	**4.0 (1.5)**	4.4 (1.4)	4.5 (1.3)	4.5 (1.3)	4.5 (1.3)

Amyloid lowering intervention 50%

Score per intervention time

Score \ Years to AD	−20	−15	−12.5	−10	−5	−3	−2	−1
ADAS11	**14.0 (5.9)**	**16.6 (4.8)**	**17.6 (4.5)**	18.3 (4.4)	18.9 (4.2)	19.0 (4.2)	19.0 (4.2)	19.0 (4.2)
MMSE	**25.6 (2.5)**	**24.5 (2.1)**	**24.0 (2.0)**	23.7 (1.9)	23.5 (1.8)	23.4 (1.8)	23.4 (1.8)	23.4 (1.8)
FAQ	**7.7 (5.2)**	**10.0 (4.3)**	**10.8 (4.1)**	11.4 (3.9)	11.9 (3.7)	12.0 (3.7)	12.0 (3.7)	12.0 (3.7)
RAVLT immediate	**30.5 (8.3)**	**27.0 (6.8)**	**25.6 (6.4)**	24.7 (6.1)	23.8 (5.8)	23.7 (5.8)	23.6 (5.8)	23.6 (5.8)
RAVLT learning	**3.3 (1.5)**	**2.7 (1.2)**	**2.5 (1.1)**	2.3 (1.1)	2.2 (1.0)	2.1 (1.0)	2.1 (1.0)	2.1 (1.0)
RAVLT forgetting	**70.9 (19.5)**	**79.4 (16.1)**	**82.7 (15.0)**	84.9 (14.3)	87.1 (13.7)	87.4 (13.6)	87.6 (13.6)	87.6 (13.6)
CDRSB	**3.0 (1.9)**	**3.8 (1.6)**	**4.1 (1.5)**	4.3 (1.4)	4.5 (1.4)	4.5 (1.4)	4.6 (1.4)	4.6 (1.4)

4

SimulAD: A dynamical model for personalized simulation and disease staging in Alzheimer's disease

Contents

4.1 Introduction **90**
4.2 Material and Methods **93**
4.2.1 Experimental cohort 93
4.2.2 Image preprocessing 95
4.2.3 Modeling framework 95
4.2.4 Estimated model 96
4.2.5 Evaluation strategy 96
4.2.6 Data adjustment 97
4.3 Results **100**
4.3.1 Known-groups validity 100
4.3.2 Concurrent validity 102
4.3.3 Model reliability 103
4.4 Discussion **103**
4.5 Conclusion **107**
4.6 Appendix **108**
4.6.1 Clinical scores imputation 108
4.6.2 FDG prediction 108
4.6.3 Tracer correction 108
4.6.4 Robustness to the clinical scores prediction 112
4.6.5 Comparison of the simulated evolution of clinical and imaging measures based on ADNI and GMC cohorts 112

In the previous chapters we introduced two different approaches allowing to estimate a model of Alzheimer's disease (AD) progression. In both cases, the methods were developed on the Alzheimer's Disease Neuroimaging Initiative cohort, but their generalization to independent datasets was not evaluated. Yet, assessing the generalization properties of disease progression models is essential in order to finally deploy them in clinical routine. Given this context, we propose in this chapter to evaluate the generalization capabilities of SimulAD by applying it on an independent cohort from the Geneva Memory Center.

The goal of this preliminary validation is to demonstrate the reliability of the model of disease progression provided by our approach, as well as its validity for performing individual disease staging. This work has been submitted to the journal Neurobiology of Aging.

Abstract. SimulAD is a disease progression model initially developed on the ADNI database to simulate the evolution of clinical and imaging markers characteristic of AD. Based on multi-modal and clinical imaging data, SimulAD estimates the disease severity of a subject with respect to the simulated progression. In this work, we assessed the generalization capabilities of SimulAD when applied to an independent cohort from the Geneva Memory Center (GMC). The GMC cohort included 93 subjects who underwent MRI, amyloid-PET, FDG-PET, tau-PET scans and neuropsychological evaluations. The distribution of the disease severity estimated by SimulAD was compared between clinical groups using Student's *t*-test and Cohen's *d*. The relationship between SimulAD disease severity, and clinical scores and imaging biomarkers was assessed using Spearman correlation coefficient. We also simulated the evolution of clinical and imaging markers based on the GMC cohort, and computed the average error between this model of evolution and the one previously estimated on the ADNI database. The difference between SimulAD disease severity of healthy, mild cognitive impairment and AD dementia groups was statistically significant (p-values < 0.05; $d \geq 0.8$). The disease severity correlated with MMSE ($\rho = -0.55$), hippocampal atrophy ($\rho = -0.62$), glucose hypometabolism ($\rho = -0.67$), amyloid burden ($\rho = 0.31$) and tau deposition ($\rho = 0.62$) (p-values < 0.01). The average error between the evolutions estimated by SimulAD on the ADNI and the GMC cohort were of 7%, 8%, 9%, 6% for MRI, FDG-PET, amyloid-PET derived markers and clinical scores respectively. The evaluation of SimulAD on the GMC cohort illustrates its robustness and generalization properties, highlighting the interest of this model to support diagnosis and to identify individuals at risk of cognitive decline over time.

4.1 Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder whose evolution has been hypothesized to follow a cascade of events [Jack, 2013b]. Deposition of the beta-amyloid protein in the brain cortex is believed to initiate this cascade, and to subsequently cause the aggregation of hyperphosphorylated tau protein in neurofibrillary tangles. This is followed by a process of neurodegeneration (i.e glucose hypometabolism and gray matter atrophy) ultimately leading to dementia. An inherent difficulty in diagnosing AD is that patients go through a long asymptomatic phase spanning approximately 10 to 20 years

[Sperling, 2011] before showing clinical symptoms. To provide a biological assessment of the disease, AD has been recently defined as a pathology characterized by three main biomarkers categories, namely: amyloid, tau and neurodegeneration [Jack, 2018]. These three biomarkers can be measured thanks to imaging techniques, such as Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET), or in the case of amyloid and tau, also by lumbar puncture and blood collection. Monitoring these biomarkers is paramount in order to track the disease progression [Risacher, 2017], and to potentially facilitate prevention or assessment of drug efficacy [Schwarz, 2019].

In the past years, the proliferation of studies collecting large amounts of biomarkers, combined with the growth of machine learning, fostered the development of computational models for automated AD diagnosis. For instance, many studies focused on the development of data-driven approaches for automatic assessment of clinical diagnosis [Davatzikos, 2009; Falahati, 2014; Arbabshirani, 2017]. Based on the sole analysis of imaging-derived data, these methods showed that it is possible to automatically identify healthy controls, subjects with mild cognitive impairment, and patients suffering from AD dementia, some of them reporting results comparable to diagnosis rates obtained by expert physicians [Klöppel, 2008]. However, most of these approaches have been exclusively developed to solve a predictive task, and generally don't allow to understand the mechanisms relating the different biomarkers throughout AD evolution. As these mechanisms still remain partially unknown, different methods known as disease progression models were therefore introduced in order to estimate, in a data-driven fashion, the long-term progression of biomarkers [Jedynak, 2012]. Due to the lack of an absolute time-line describing AD evolution, these models usually assume that the disease is characterized by monotonic changes, such that the modelled biomarkers steadily evolve from normal to pathological values. This assumption allows to reconstruct a time-line on which we can track the disease progression [Lorenzi, 2017]. Moreover, these methods can be applied to a variety of data types, such as cortical and subcortical shapes [Marinescu, 2019a] or volumetric images [Khanal, 2016; Khanal, 2017], thus offering a fine-grained spatial description of the changes affecting the brain. These models also allow to automatically assess the individual disease severity by comparing the clinical and imaging measurements of a given subject to the estimated disease progression. This latter capability of disease progression models is usually referred as disease staging. Ultimately, these approaches could potentially be used for identifying individuals at risk of cognitive decline, or for assessing drug efficacy in clinical trials.

Since these statistical models have been mostly developed on publicly available research datasets, such as the one provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI), their generalization to external cohorts from memory clinics still requires additional testing and validation [Mendelson, 2017]. As clinical cohorts may fundamentally differ from the ADNI one, whether it be in terms of data acquisition or study population, automated diagnosis pipelines usually show a prominent decrease in performances.

Regarding disease progression models, it is conceivable that biomarkers' trajectories estimated solely via the analysis of a single cohort may not be fully representative of the natural disease course. This aspect would question the generalization of the progression model when tested on subjects from independent clinical cohorts. It is therefore essential to assess the generalization of this kind of models on independent datasets, with respect to their specific biases which can encompass a broad range of differences between cohorts such as data acquisition, missing data or data heterogeneity. These differences need to be addressed to finally deploy disease progression models in a practical clinical setting [Castro, 2020].

Recently, the event-based-model (EBM) [Fonteijn, 2012] and the Discriminative EBM (DEBM) [Venkatraghavan, 2019] underwent an extensive evaluation effort. These approaches model AD progression as a sequence of events representing the transition of a set of biomarkers from a normal to an abnormal state. Both EBM and DEBM have been applied on subjects from independent cohorts, providing accurate patients staging [Archetti, 2019]. However, the kind of progression model estimated by these methods presents important limitations. First, both approaches are based on the simplistic assumption describing the pathological progression as a discrete sequence of biomarkers transitions from normal to abnormal states, which doesn't reflect the continuous nature of the changes affecting the brain during the disease. Second, they allow the analysis of summary measures only, such as regional brain uptake of grey matter density values, and thus don't enable the fine-grained quantification of the spatial patterns of changes associated with the disease. Third, while these two models inform us about the sequence of events characterizing AD, they don't provide insights about the dynamical interplay between biomarkers. Investigating such interactions would allow a deeper understanding of how the pathological processes at stake during the disease affect each other. Fourth, the EBM and the DEBM don't allow to simulate hypothetical scenarios of disease progression. Yet, such capability could be used to assess the effect of drug intervention on the disease evolution *in silico*, which could help planning and monitoring clinical trials.

To address these limitations, more refined approaches to disease progression modeling have been proposed, to enable the fine-grained description of the pathological evolution in space and time [Bilgel, 2016; Koval, 2018; Marinescu, 2019a]. Within this context, SimulAD [Abi Nader, 2021] is a recent method allowing the analysis of clinical scores and multivariate imaging data extracted from MRI and PET scans to estimate a continuous spatio-temporal model of disease progression. Compared to the EBM and the DEBM, this approach offers a higher resolution for the imaging biomarkers, allowing to track the evolution of regional changes affecting the brain along a continuous temporal scale describing the disease course over 30 years. The trajectories estimated for clinical and imaging markers can subsequently be used as a reference to assess the individual disease severity, by locating subjects along the temporal scale describing the disease evolution. Moreover, SimulAD estimates the dynamical relationships between key biomarkers at

stake during AD progression, namely: amyloid deposition, glucose hypometabolism, cerebral atrophy, cognitive and behavioural decline. Based on these relationships, the model can be applied to simulate the personalized evolution of any patient or group of patients only from the knowledge of their baseline clinical and imaging measurements. Finally, SimulAD enables to assess the impact of therapeutic intervention, such as anti-amyloid treatment, on cognitive outcomes depending on the intervention time (*cf.* Chapter 3).

While SimulAD was previously trained on the ADNI cohort, its generalization to independent datasets has not been evaluated. As it generally holds for statistical and machine learning models, one of the main difficulty for the generalization of SimulAD to independent cohorts lies in the challenge of accounting for potential missing data, heterogeneity due to different acquisition protocols or even data incompatibility between cohorts. Yet, evaluating the generalization of SimulAD to independent cohorts is essential to demonstrate the reliability of the dynamics allowing to personalize the models of disease progression, as well as the validity of the approach for providing accurate disease staging.

Within this context, we assess in this study the generalization capabilities of SimulAD. To this end, we test the robustness and reliability of this approach when applied to an independent dataset from a memory clinic, namely the Geneva Memory Center (GMC). This cohort includes patients with cognitive complaints, who underwent a baseline clinical and neuropsychological evaluation, MRI, amyloid-PET, (18)F-fluorodeoxyglucose-PET (FDG-) and (18)F-flortaucipir-PET (tau-) scans. Our evaluation procedure relies on three key aspects: (i) Development of a pre-processing pipeline allowing to apply SimulAD on the GMC cohort; (ii) Assessment of SimulAD validity for individual disease staging. (iii) Evaluation of the reliability of the progression of imaging and clinical markers estimated by SimulAD.

4.2 Material and Methods

In this work subjects were divided in five clinical groups: cognitively healthy (NL stable), individuals diagnosed with mild cognitive impairment (MCI stable), patient suffering from Alzheimer's disease dementia (AD dementia), subjects progressing from NL to MCI or AD dementia (NL converters), and finally subjects progressing from MCI to AD dementia (MCI converters).

4.2.1 Experimental cohort

The GMC cohort included 93 subjects: 23 NL stable, 28 MCI stable, 25 MCI converters, 17 AD dementia. In Table 4.1, we provide socio-demographic information across clinical

groups for this dataset. The clinical spectrum of the cohort spans a broad range of cognitive severity, from healthy to moderate and severe dementia. Conversion to AD dementia was determined using the last available follow-up information. All the participants were amyloid positive at baseline, and "amyloid positivity" was evaluated using visual assessment performed by an expert nuclear medicine physician and following the tracer manufacturers guidelines. Multi-modal biomarkers consisting of neuropsychological tests and imaging-derived measures were collected. Each participant underwent the Mini-Mental State Examination (MMSE). Moreover, imaging data coming from MRI, FDG-PET, and amyloid-PET were available. In addition, a tau-PET scan was acquired for 50 subjects. Summary statistics about clinical and imaging-derived information across clinical groups are reported in Table 4.1.

The cohort used to train SimulAD was composed by a subset of individuals from the ADNI database. Further information on this cohort is available in Section 3.2.1.

Table 4.1: Baseline characteristic of the GMC cohort. Average values, standard deviation in parenthesis. Acronyms: GMC: Geneva Memory Center; NL: cognitively healthy; MCI: mild cognitive impairment; AD dementia: Alzheimer's disease dementia; FDG: (18)F-fluorodeoxyglucose Positron Emission Tomography (PET) imaging; SUVR: Standardized Uptake Value Ratio; MMSE: Mini Mental State Examination; Tau: (18)F-flortaucipir PET imaging. Converters are NL patients progressing to MCI or AD dementia, or MCI individuals progressing to AD dementia. Hippocampal volume: extracted with Freesurfer. Amyloid burden: voxel-number weighted average of the amyloid uptake in the frontal, anterior/posterior cingulate, lateral parietal, and lateral temporal regions normalized to the cerebellum. Early amyloid: voxel-number weighted average of the uptake extracted from the early-phase (6 min) of amyloid-PET in the frontal, anterior/posterior cingulate, lateral parietal, and lateral temporal regions normalized to the cerebellum. Glucose metabolism: voxel-number weighted average of the FDG uptake in the angular, temporal, and posterior cingulate cortex normalized to the cerebellum. Tau burden: voxel-number weighted average of the tau uptake in the entorhinal, amygdala, parahippocampal, fusiform, inferior temporal, and middle temporal regions normalized to the cerebellum. Missing data for 36[a] and 43[b] subjects.

Group	NL stable	MCI stable	MCI converters	AD dementia
	GMC cohort			
N	23	28	25	17
Female (%)	61	61	65	53
Age (years)	69.1 (7.5)	74.3 (6.5)	73.8 (4.6)	70.4 (11.1)
Education (years)	17.3 (3.9)	14.0 (3.1)	12.5 (4.7)	11.6 (3.8)
MMSE	28.5 (1.0)	25.2 (3.1)	23.7 (4.3)	18.2 (6.5)
Hippocampus (mm^3)	4271 (435)	3621 (534)	3634 (448)	3409 (436)
Amyloid (SUVR)	0.70 (0.12)	0.93 (0.13)	0.91 (0.13)	0.89 (0.11)
Early amyloid (SUVR)	0.53 (0.03)	0.50 (0.03)	0.47 (0.02)	0.45 (0.03)
FDG[a] (SUVR)	0.57 (0.06)	0.54 (0.04)	0.52 (0.04)	0.49 (0.06)
Tau[b] (SUVR)	0.60 (0.06)	0.81 (0.20)	0.94 (0.27)	1.21 (0.34)

4.2.2 Image preprocessing

We derived volumes of gray matter density in a standard anatomical space by relying on the longitudinal pipeline of Freesurfer [Reuter, 2012]. Regional gray matter density was extracted from the Desikan-Killiany parcellation [Desikan, 2006]. amyloid-PET, FDG-PET and tau-PET images were registered to their corresponding T1-MRI acquisition, and normalized to the cerebellum uptake. Regional amyloid load, glucose metabolism and tau burden were computed thanks to the PetSurfer software [Greve, 2014]. For every imaging modality we discarded white-matter, ventricular, and cerebellar regions, thus obtaining 82 regions that were averaged across hemispheres.

4.2.3 Modeling framework

SimulAD is based on the hypothesis that AD evolution can be mathematically modelled by a set of key biomarkers following a dynamical system. These biomarkers are namely *clinical scores*, *gray matter atrophy*, *amyloid load* and *glucose metabolism*. This assumption has two consequences: the first one is that, at any given time, AD severity is uniquely associated with the values of these biomarkers. The second one is that past and futures states of the disease can be computed from the current ones thanks to mathematical relationships linking the biomarkers evolutions.

To estimate the complex relationships between high-dimensional imaging and clinical measures, the model first transforms baseline neuropsychological assessments and measures derived from MRI, amyloid-PET and FDG-PET data in a set of four corresponding z-scores. The transformation consists in a weighted average of the measurements derived from each type of data modality (i.e regional grey matter measurements in the case of *atrophy*). The obtained z-scores are respectively denoted z^{cli}, z^{atr}, z^{amy} z^{met}, and describe the overall pathological status of an individual. We hypothesize that these four z-scores are related by a set of relationships driving the disease progression, which are mathematically modelled by a system of Ordinary Differential Equations (ODEs). This system of ODEs provides us with an interaction rule that describes how the z-scores jointly evolve over time. The parameters controlling the system of ODEs are optimized such that the predicted evolution of the z-scores best matches the available follow-up clinical and imaging measurements of each individual. Further details about the mathematical formulation of the model are provided in Section 3.2.

Trajectory modelling. Thanks to this mathematical formulation, SimulAD can be used to simulate the progression of changes characterizing AD by considering the subjects diagnosed with AD dementia, and for whom we compute corresponding z-scores based on their baseline measures. Relying on the estimated set of relationships between z-scores we follow their evolution forward and backward in time, thus simulating the

subjects' evolution from their original healthy condition to their current pathological state. We obtain z-scores trajectories summarizing the overall progression of AD, and from which we can estimate the long-term evolution of the corresponding clinical and imaging measurements.

Disease severity quantification. Relying on the reference trajectory estimated for the four z-scores summarizing AD evolution, we can subsequently perform individual disease staging. Based on the multi-modal imaging and clinical data of a given subject collected at any visit, we compute z-scores for each type of marker, and find the time-point τ jointly minimizing the distance between the individual z-scores and the reference trajectory. In the rest of the paper, we will refer to this time-point as the disease severity. The estimated disease severity τ locates a subject on the reference trajectory, thus quantifying its pathological state. It is also important to note that the disease severity can still be estimated even in the case of missing data, by only computing the z-scores of the available measures of the observed subject. We provide mathematical details on the disease severity estimation in Section 3.2.6.

4.2.4 Estimated model

The parameters of the resulting model, presented in Chapter 3, were estimated through the analysis of multi-modal longitudinal data from the ADNI cohort. The clinical scores consisted in the Clinical Dementia Rating Scale Sum of Boxes (CDRSB), Alzheimer's Disease Assessment Scale (ADAS11), Functional Assessment Questionaire (FAQ), Rey Auditory Verbal Learning Test (RAVLT) learning, RAVLT immediate, RAVLT forgetting and MMSE. Regional gray matter density, amyloid load and glucose metabolism were derived following the procedure detailed in Section 4.2.2. No tau-PET data was included in the model. Additional baseline socio-demographic information and summary statistics about clinical and imaging data for the subjects from the ADNI cohort are provided in Section 3.2.1. The disease progression previously estimated by SimulAD on the ADNI database is illustrated in Figures 3.2 and 3.3 in which we show the evolution of the z-scores and their associated imaging and clinical measures.

4.2.5 Evaluation strategy

We considered the model of evolution estimated by SimulAD on the ADNI cohort as the reference progression for AD. We evaluated SimulAD by conducting a series of experiments on both ADNI and GMC cohorts which aimed at demonstrating respectively the known-groups validity, the concurrent validity and the reliability of the model.

Known-groups validity. We evaluated how the estimated disease severity discriminates subjects across clinical groups within each cohort. It is expected that values of disease severity should increase along with the severity of the clinical status. We further assessed the group-wise consistency of the disease severity distribution, by comparing its values for similar clinical groups between ADNI and GMC cohorts. Differences between groups were assessed using Student's *t*-test and Cohen's *d* effect size.

Concurrent validity. We assessed the extent to which the estimated disease severity correlates with validated clinical and imaging assessments. For each subject MMSE score was available and imaging-biomarkers assessment as follows: hippocampal atrophy evaluated with Freesurfer, glucose metabolism and amyloid burden computed by extracting standardized uptake value ratio in a composite mask of regions of interest (MetaROI approach [Jagust, 2009; Landau, 2010; Landau, 2012]). In the case of GMC, we also had 50 subjects who underwent a tau-PET scan. We compared their estimated disease severity with respect to their tau burden computed in a composite mask of relevant regions (MetaROI approach [Jack, 2017]). Correlation between the estimated disease severity and the different variables was assessed using Spearman rank correlation (ρ).

Reliability. The two previous experiments aimed to quantify the validity of SimulAD in terms of disease staging based on the reference progression previously simulated on the ADNI database. We verified the consistency and robustness of the dynamics estimated by SimulAD by simulating the evolution of clinical and imaging-derived markers based on the GMC data. To this end, we applied the procedure described in Section 4.2.3 on the AD dementia subjects from the GMC cohort, thus providing us with a new model of progression for clinical and imaging measurements, as well as specific z-scores trajectories personalized to the GMC cohort. We compared the disease progression models obtained on ADNI and GMC cohorts by computing the average error between their z-scores trajectories over time.

4.2.6 Data adjustment

In order to implement the aforementioned assessment strategy, a number of additional pre-processing steps had to be carried out.

Missing measures imputation. We recall that SimulAD relies on 7 neuropsychological tests (CDRSB, MMSE, ADAS11, FAQ, RAVLT learning, RAVLT immediate and RAVLT forgetting) to compute the score z^{cli}, and that the only common clinical test between the ADNI and GMC cohorts is the MMSE. However, relying only on the MMSE would bias the computation of z^{cli} for the subjects from the GMC cohort. To overcome this issue, we imputed the 6 missing clinical scores for all the subjects from the GMC cohort. Imputation was carried out through k-means regressions trained on the ADNI database

to predict each clinical score based on the MMSE and the measures of regional grey matter volume. The trained models were subsequently applied on the GMC cohort to estimate the associated clinical scores. We performed a 10-fold cross validation on the ADNI database to evaluate the prediction performances of the models. We show in Table 4.2 the average and the 95% confidence interval of the relative error between the ground truth and predicted score. The average relative error remains below 10% for the CDRSB, ADAS11, RAVLT immediate and FAQ, while not exceeding 20% for the RAVLT learning and RAVLT forgetting. We provide additional information about the estimation error in Appendix 4.6.1.

FDG data harmonization. We can observe in Table 4.1 that in the GMC cohort 36 subjects out of 93 are missing an FDG-PET scan, thus preventing the computation of their score z^{met}. For these subjects their disease severity can therefore be estimated only based on three z-scores (z^{cli}, z^{atr}, z^{cli}), leading to potential bias and mis-estimation compared to the rest of the cohort. To prevent this issue we computed the regional FDG uptake for every subjects of the GMC cohort based on their corresponding early-phase (6 min) of amyloid-PET scan [Daerr, 2017]. This was done by fitting a linear regression between the early-amyloid and FDG uptake of all the subjects for each brain region. Figure 4.1 illustrates the linear fit between early-amyloid and FDG regional uptake for three brain regions. We show similar relationships for additional brain regions in Appendix 4.6.2.

Tracer bias correction. The amyloid-PET scan of 17 subjects from the GMC cohort was acquired using the flutemetamol tracer, while the amyloid-PET scans of the remaining subjects from both ADNI and GMC cohorts were acquired using florbetapir. To compensate the effect of the tracer on the regional uptake for these subjects, we converted the extracted amyloid burden on the centiloid scale and back to a florbetapir scale [Klunk, 2015; Battle, 2018; Navitsky, 2018]. We observe in Figure 4.2 that the correction reduces the variability of the amyloid uptake values for the flutemetamol group, increasing the overlapping with the florbetapir one. Appendix 4.6.3 provides histograms illustrating the same effect in other brain regions.

Table 4.2: Relative error between the ground truth and k-means prediction of the different clinical scores in the ADNI cohort. Average values and 95% confidence interval. CDRSB: Clinic Dementia Rating Scale Sum of Boxes; ADAS11: Alzheimer's Disease Assessment Scale; FAQ: Functional Assessment Questionnaire; RAVLT: Rey Auditory Verbal Learning Test. ADNI: Alzheimer's Disease Neuroimaging Initiative. CI: Confidence interval.

Score	CDRSB	ADAS11	RAVLT immediate	RAVLT learning	RAVLT forgetting	FAQ
Relative error (%)	5.2	6.0	7.5	14.2	17.8	9.1
95% CI	[1.4 ; 10.3]	[3.2 ; 10.6]	[5.2 ; 10.2]	[10.7 ; 18.8]	[10.3 ; 25.8]	[1.5 ; 18.8]

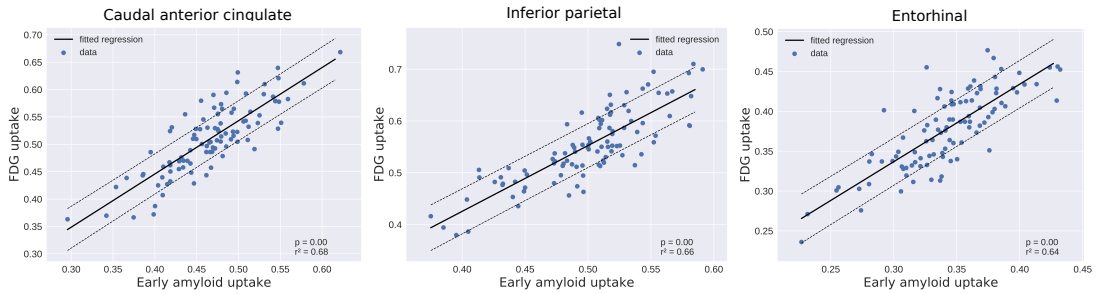

Figure 4.1: Scatter plot between the regional early-amyloid uptake and the corresponding FDG uptake for 57 patients of the GMC cohort. Solid black lines show the fitted linear model between regional FDG and early-amyloid. The dashed-lines represent 95% confidence interval. GMC: Geneva Memory Center. FDG: (18)F-fluorodeoxyglucose.

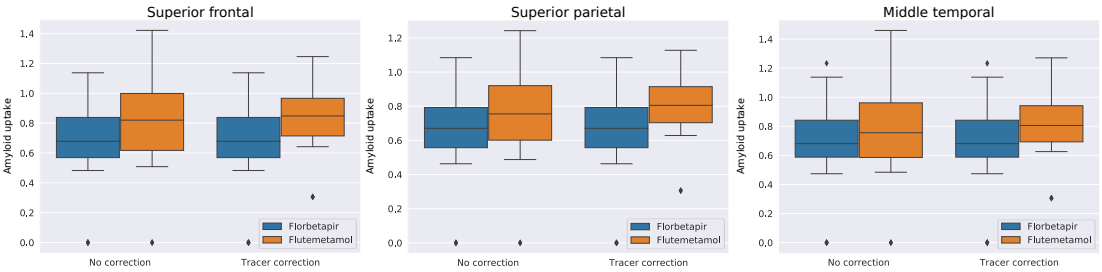

Figure 4.2: Distribution of the regional amyloid uptake depending on the tracer used during acquisition for the subjects from the GMC cohort. 76 amyloid-PET scans were acquired using florbetapir and 17 using flutemetamol. Tracer correction indicates that the regional amyloid uptake of subjects whose PET scan was acquired using flutemetamol was converted to a florbetapir scale. GMC: Geneva Memory Center; SUVR: Standardized Uptake Value Ratio; PET: Positron Emission Tomography.

4.3 Results

4.3.1 Known-groups validity

In this section, we considered the pathological progression previously estimated in Chapter 3 by SimulAD as a reference trajectory, and computed the disease severity of the individuals from both cohorts based on the procedure of Section 4.2.3. We show in Figure 4.3 the group-wise distribution of the disease severity estimated by SimulAD for each subject in the ADNI and GMC datasets. We observe that for both cohorts the disease severity increases when going from healthy to pathological stages. The group-wise difference of disease severity across clinical groups is statistically significant for each comparison (Student's *t*-test $p < 0.05$) except in the case of MCI stable vs MCI converters for the GMC cohort (*cf.* Table 4.3a). We also notice rather large differences between clinical groups ($d > 0.7$, *cf.* Table 4.3a) for both cohorts except in the case of NL stable vs MCI stable for the ADNI cohort and MCI stable vs MCI converters for the GMC database. We also evaluated the consistency of the disease severity by comparing its distribution for similar clinical groups across cohorts. We observe in Table 4.3b that the estimated disease severity of similar clinical groups is not significantly different ($p > 0.05$), and differences between cohorts are rather small ($d \leq 0.1$), except in the case of MCI stable. We recall that, apart from the MMSE, the clinical scores of the subjects from the GMC cohort were imputed based on the procedure detailed in Section 4.2.6. We show in Figure 4.10 in Appendix 4.6.4 that the estimation of the individual disease severity is robust to this approximation.

Table 4.3: Comparison of the estimated disease severity distribution between clinical groups within each cohorts (a) and between similar clinical groups across cohorts (b); We report p-values of Student's *t*-test as well as the associated effect size (Cohen's *d*).

(a)

Within cohorts disease severity comparison

	NL stable vs NL converters		NL stable vs MCI stable		MCI stable vs MCI converters		MCI stable vs AD dementia		MCI converters vs AD dementia	
Cohort	ADNI	GMC	ADNI	GMC	ADNI	GMC	ADNI	GMC	ADNI	GMC
p-value	$7.5 \cdot 10^{-4}$	/	$1.8 \cdot 10^{-2}$	$1.0 \cdot 10^{-4}$	$1.2 \cdot 10^{-11}$	$4.4 \cdot 10^{-1}$	$1.1 \cdot 10^{-26}$	$3.8 \cdot 10^{-3}$	$3.1 \cdot 10^{-7}$	$1.6 \cdot 10^{-2}$
Cohen's d	0.75	/	0.35	1.1	0.91	0.21	1.5	0.80	0.74	0.81

(b)

Between cohorts disease severity comparison

	NL stable	MCI stable	MCI converters	AD dementia
p-value	$8.2 \cdot 10^{-1}$	$7.0 \cdot 10^{-3}$	$9.5 \cdot 10^{-1}$	$7.3 \cdot 10^{-1}$
Cohen's d	0.03	0.57	0.10	0.09

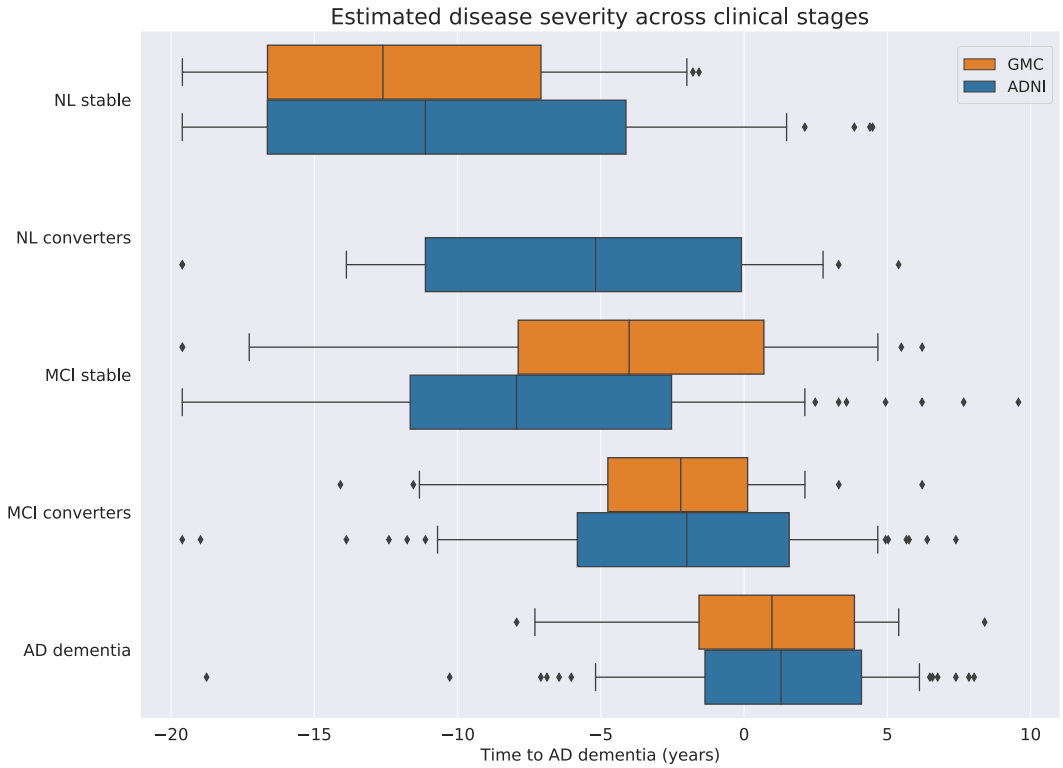

Figure 4.3: Distribution of the disease severity estimated by SimulAD across clinical stages for the ADNI and GMC cohorts relatively to the estimated model of disease progression (Figure 3.2). ADNI: Alzheimer's Disease Neuroimaging Initiative; GMC: Geneva Memory Center; NL: cognitively healthy; MCI: mild cognitive impairment; AD dementia: Alzheimer's disease dementia. Converters are NL patients progressing to MCI or AD dementia, or MCI individuals progressing to AD dementia.

4.3.2 Concurrent validity

We show in Figure 4.4 the progression of standard clinical and imaging markers with respect to the disease severity estimated by SimulAD for subjects from the ADNI and GMC databases. For both cohorts, the disease severity significantly correlates with the MMSE score (ADNI: $\rho = -0.58$, $p < 0.01$; GMC: $\rho = -0.55$, $p < 0.01$). Regarding imaging-biomarkers, in both ADNI and GMC datasets the estimated disease severity correlates with hippocampal volume (ADNI: $\rho = -0.57$, $p < 0.01$; GMC: $\rho = -0.62$, $p < 0.01$), glucose metabolism (ADNI: $\rho = -0.80$, $p < 0.01$; GMC: $\rho = -0.67$, $p < 0.01$) and amyloid burden (ADNI: $\rho = 0.44$, $p < 0.01$; GMC: $\rho = 0.31$, $p < 0.01$). Since 50 individuals from the GMC cohort underwent a tau-PET scan, we also compare their estimated disease severity with their tau burden and show a significant correlation between them in Figure 4.4 ($\rho = 0.62$, $p < 0.01$). This latter correlation of the disease severity with a typical biomarker of AD that was not used for building the model supports the reliability of such a measure to summarize the overall severity of AD.

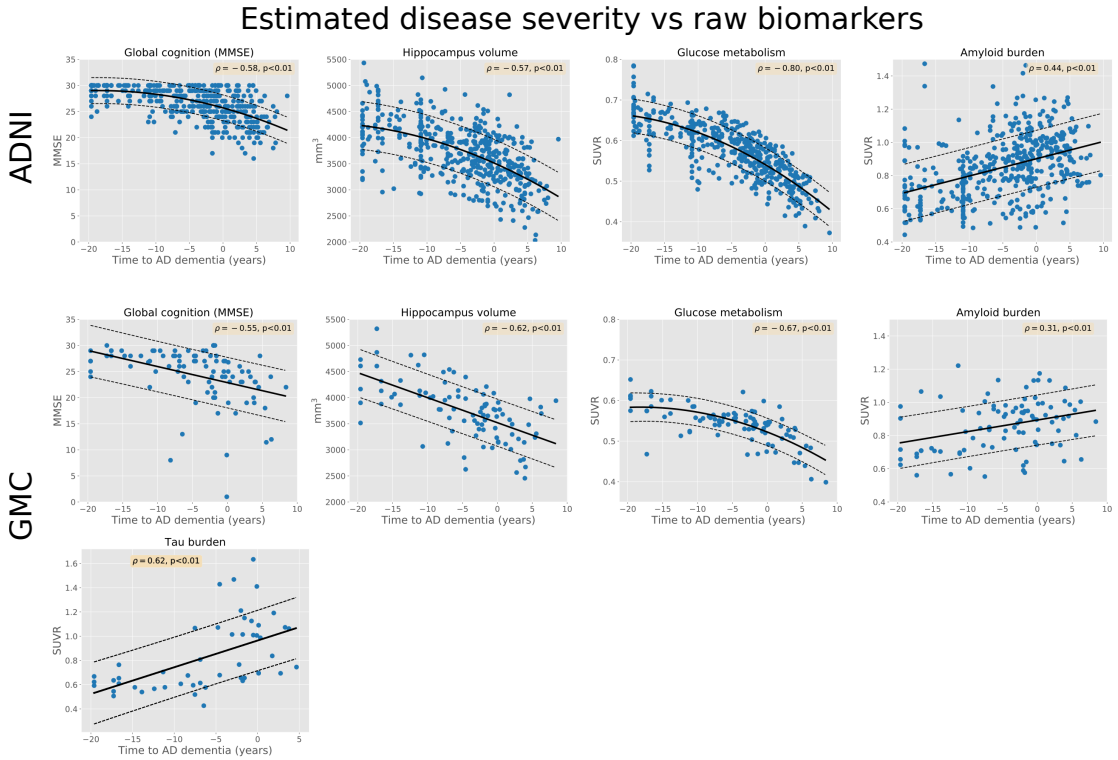

Figure 4.4: Relationships between the estimated disease severity and global cognition (MMSE), hippocampal volume, glucose metabolism, amyloid burden and tau burden (only for the GMC cohort). For each subject the estimated disease severity quantifies their position with respect to the model of disease progression (Figure 3.2). The dashed-lines represent 95% confidence interval. ADNI: Alzheimer's Disease Neuroimaging Initiative; GMC: Geneva Memory Center; SUVR: Standardized Uptake Value Ratio.

4.3.3 Model reliability

In the previous sections, we assessed the disease severity of the subjects based on the model of progression simulated on the ADNI cohort. In order to demonstrate the reliability of the dynamics estimated by SimulAD, we simulated a new model of disease progression personalized to the GMC cohort. We show in Figure 4.5 the predicted evolution of imaging and clinical measurements based on this dataset. Similarly to what has been observed on the model previously trained on the ADNI database (Figure 3.3), amyloid load increases and saturates early while following a uniform spatial pattern. Amyloid deposition is followed by a delayed process of neurodegeneration, more specifically a decrease of glucose metabolism and gray matter atrophy, mostly affecting temporal and parietal regions. Finally, clinical scores such as the MMSE or the CDRSB show a non-linear evolution accelerating during the latest stages of the disease.

We also compared the models of progression obtained on the ADNI and GMC cohorts. Figure 4.6 shows the evolution of the different z-scores depending on the cohort used to estimate the disease progression. These z-scores indicate the overall evolution of clinical scores, cerebral atrophy, amyloid deposition and glucose metabolism during AD. Given that the z-scores are not related to a physical unit but rather quantify the abnormality of a particular process, they were re-scaled between 0 and 1 to illustrate the progression from healthy towards pathological stages. We observe that the four z-scores exhibit similar evolution patterns, whether they have been estimated on the ADNI or the GMC cohort. When averaged across time, the error between the z-scores of the two cohorts is of 6%, 7%, 8% and 9% for z^{cli}, z^{atr}, z^{met} and z^{amy} respectively. We provide in Appendix 4.6.5 the evolution of the error between the evolutions of clinical scores and imaging regional measurements estimated based on the ADNI and the GMC cohort. When averaged over time, brain regions and clinical scores, the error is of 3%, 6%, 7% and 12% for MRI, FDG-PET, amyloid-PET derived regional measurements and clinical scores respectively. Finally, in spite of the fact that most of the clinical scores were imputed in the GMC cohort, Appendix 4.6.4 shows that the resulting z-scores trajectories are robust to this estimation. Indeed, we observe that adding an additional error when imputing the clinical scores in the GMC cohort leads to rather small changes for the estimated z-scores trajectories compared to the results presented in Figure 4.6.

4.4 Discussion

In this study, we presented a thorough assessment of SimulAD on the independent GMC clinical cohort. The model was initially estimated based on the analysis of longitudinal imaging and clinical data from a subset of the ADNI database (*cf.* Chapter 3). Due to a mismatch between cohorts in terms of missing measurements and data acquisition, we

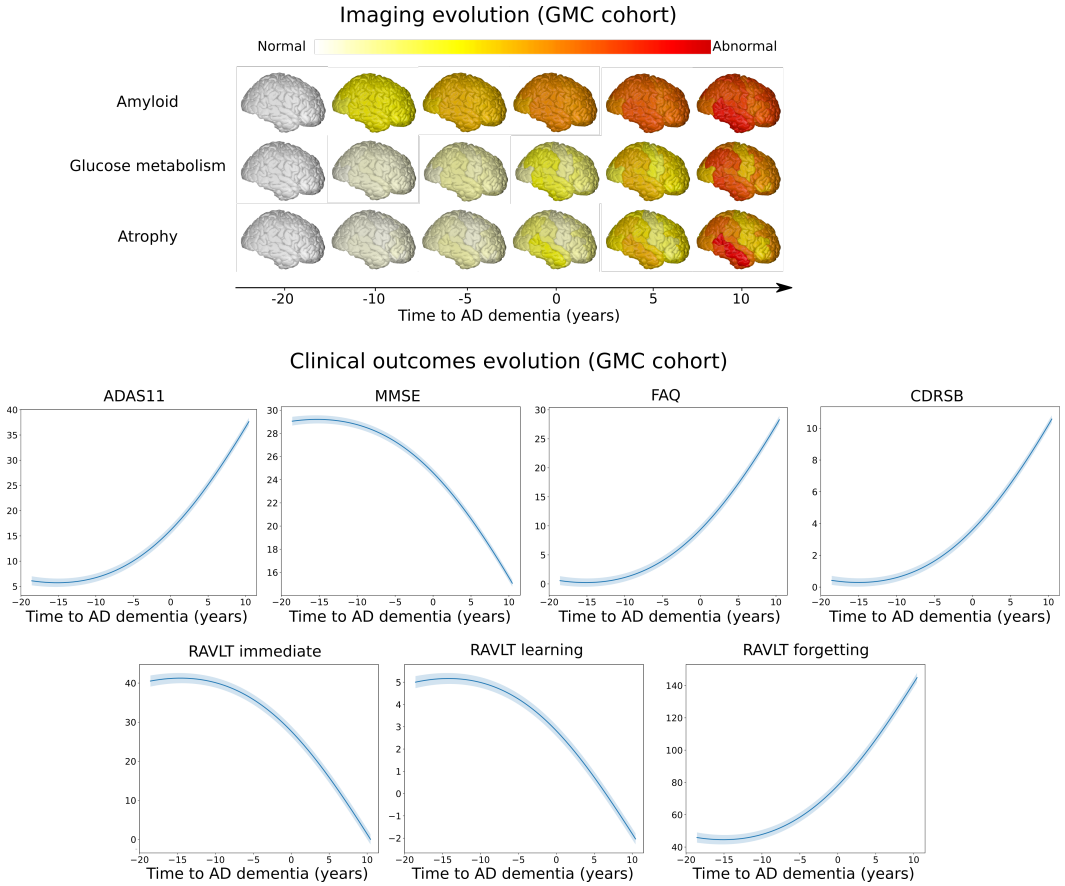

Figure 4.5: Simulated long-term evolution of cortical measurements for the different types of imaging markers and clinical scores based on the GMC cohort. Shadowed areas represent the standard deviation of the average trajectory. GMC: Geneva Memory Center. Images were generated thanks to the BrainPainter software [Marinescu, 2019b].

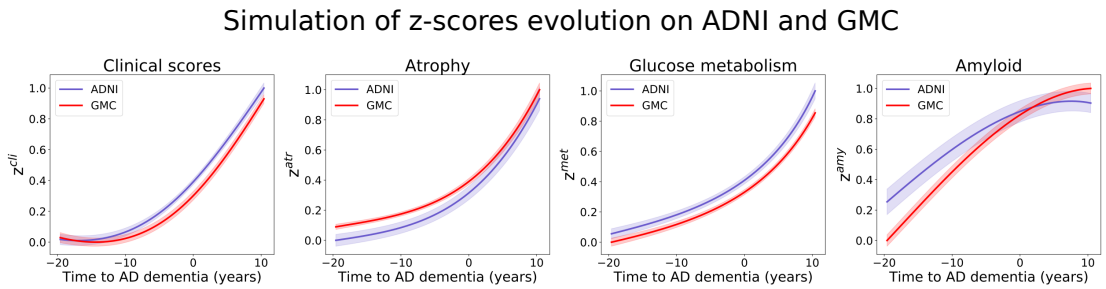

Figure 4.6: Estimated long-term dynamics depending on the cohort used for simulating the z-scores trajectories (time is relative to conversion to Alzheimer's dementia). The z-scores have been re-scaled between 0 and 1 to illustrate the progression of each process from normal to pathological stages. ADNI: Alzheimer's Disease Neuroimaging Initiative; GMC: Geneva Memory Center.

designed a pre-processing pipeline in order to enable the application of the model. We subsequently carried out a set of experiments to evaluate the validity of our approach and showed that the results obtained on the ADNI database were reproducible on the GMC cohort, outlining the robustness and generalization properties of SimulAD.

The present results underlined the feasibility of applying our model in a clinical context. We showed that SimulAD was able to differentiate subjects across clinical groups for both cohorts (Figure 4.3, Table 4.3a). Known-groups validity was established by assessing the similarity between cohorts of the estimated disease severity distribution for the NL stable, MCI converters and AD dementia groups (Table 4.3b). Moreover, the estimated disease severity correlated with clinical and imaging assessments in both cohorts, and especially with tau burden on the GMC cohort. We recall that AD is defined by evidence of an abnormal load of both amyloid and tau [Jack, 2018]. Therefore, given that the model was estimated based on a subset of the ADNI cohort without accounting for tau-PET data, the correlation between the disease severity and tau burden on an external dataset is an appreciable demonstration of the concurrent validity of SimulAD. Finally, we observed that, independently from the cohort used to simulate the disease progression, the long-term evolution of the z-scores were similar between cohorts (Figure 4.6), thus outlining the reliability of the proposed model of AD progression. Overall, these results indicate general robustness of SimulAD when applied to independent cohorts characterized by different imaging acquisition protocols.

Some results highlighted in the proposed analysis deserve further discussion. There is no statistically significant difference between the estimated severity of the MCI stable and converters in the GMC cohort, while there is a significant difference of the estimated disease severity between the MCI stable of the ADNI and GMC cohorts. Concerning the first remark, it is important to note that among the 28 MCI stable subjects of the GMC cohort, a single visit only was available for 16 of them. It is therefore likely that this sub-group contains potential converters who may bias the MCI stable group with non-representative measurements. Regarding the second remark, we note in Figure 4.3 that the median disease severity of the MCI stable group of the ADNI dataset is approximately of -8 years, while it is close to -4 years for the GMC cohort. This means that, according to our model, the pathological condition of the group of MCI stable from the GMC cohort is more severe than the one of the ADNI subset considered in the study. This is in agreement with the fact that compared to the ADNI database, the MCI stable subjects from the GMC cohort exhibit a statistically significant lower MMSE score, lower volume of the hippocampus, lower glucose metabolism and higher amyloid burden (p-values$<$0.05, Student's *t*-test). Moreover, we note that the distribution of the disease severity is conserved between both cohorts for NL stable, MCI converters and AD dementia groups. This result may point to the generalization capabilities of the disease severity estimation, which seems to be solely influenced by the clinical status. It also shows the ability of the model to identify specific clinical sub-types.

To our knowledge, the EBM and the DEBM are the only data-driven models of disease progression which have been evaluated on several external cohorts [Archetti, 2019]. The type of study here presented is therefore of relevant experimental value to enable the future application of disease progression models. We note that SimulAD presents certain advantages compared to standard event-based models, as it provides a finer-grained description of the disease for both spatial and temporal scales, in which regional changes affecting the brain are modelled on a continuous long-term time span. In addition, the method allows to personalize the evolution of clinical and imaging measurements for any patient or group of patients. Finally, the proposed mathematical formulation of AD progression as a system of ODEs allows to go beyond a simple description of the dynamics at stake during the disease evolution. In particular, we can simulate the impact of an anti-amyloid treatment on cognitive outcomes.

As data-driven models are becoming more popular in healthcare thanks to their ability to leverage large scale clinical data, it is of utmost importance to facilitate their transfer from a research context to clinical practice. A growing number of regulatory institutions provided guidelines to help designing machine learning models that could be applied in clinical practice [Health, 2019]. The major problem revolves around the generalization of the models beyond the dataset used to develop them, the main obstacle being their robustness to biases [Ghassemi, 2019]. In this study, we developed a pipeline to mitigate the biases due to data heterogeneity and missing measurements. This work included data imputation for clinical data, and data standardization for PET imaging scans. Even though SimulAD proved robust to these approximations, one of the main challenge highlighted in this study points to the complexity of data integration across cohorts and studies. In our particular case, neuropsychological tests could be standardized on a common scale, thus allowing to replace a score by another if they assess similar functions, while PET data could be systematically converted to the centiloid scale.

Finally, this study motivates further extensions of SimulAD that would foster its adoption in clinical practice. For instance, a limitation of our model is the underlying hypothesis that there exists a unique progression of AD which is common across individuals, while the disease is in fact highly heterogeneous. SimulAD could be extended in the future to account for multiple risk factors, such as the presence of APOE4 [Kim, 2009], thus leading to a higher level of personalization of the predictions. Another avenue of improvement would be to account for a larger panel of biomarkers, such as tau [Pontecorvo, 2019], in order to better comprehend the disease progression. Currently, SimulAD is still a research software and future efforts should focus on the development of a user-friendly platform that could be deployed in clinical routine.

4.5 Conclusion

We presented a preliminary validation of SimulAD on the clinical cohort of the GMC. The results highlighted the reliability of the dynamics simulated by SimulAD for the disease key biomarkers, and showed encouraging performances in terms of disease staging on both cohorts. SimulAD is a promising modeling tool that may enable in the future the identification of subjects for enrollment in clinical trials, or the monitoring of the efficacy of disease modifying drugs.

4.6 Appendix

4.6.1 Clinical scores imputation

To complete what was presented in Section 4.2.6 concerning the imputation of the neuropsychological tests for the subjects of the GMC cohort, we show in Table 4.4 the Krippendorff coefficient for each clinical score. This coefficient compares the inter-rater agreement, which would be here the agreement between the ground truth score and the estimated one. We observe that the value 0.8 is contained within the confidence interval of the CDRSB, ADAS11, RAVLT immediate and FAQ, which makes the associated k-means models reliable predictors according to the thresholds defined by Krippendorff. The value 0.67 is contained within the confidence interval of the RAVLT learning and RAVLT forgetting, which makes the corresponding k-means models acceptable for drawing prudent conclusions. We also provide the distribution of the absolute prediction error for all the scores across clinical groups in Figure 4.7.

Table 4.4: Krippendorff coefficient assessing the agreement between the ground truth and the k-means prediction for all the clinical scores in the ADNI cohort. Average values and 95% confidence interval. CDRSB: Clinic Dementia Rating Scale Sum of Boxes; ADAS11: Alzheimer's Disease Assessment Scale; FAQ: Functional Assessment Questionnaire; RAVLT: Rey Auditory Verbal Learning Test. ADNI: Alzheimer's Disease Neuroimaging Initiative. CI: Confidence interval.

Score	CDRSB	ADAS11	RAVLT immediate	RAVLT learning	RAVLT forgetting	FAQ
Krippendorff	0.77	0.77	0.80	0.61	0.69	0.78
95% CI	[0.69 ; 0.83]	[0.70 ; 0.83]	[0.74 ; 0.85]	[0.54 ; 0.68]	[0.62 ; 0.75]	[0.72 ; 0.84]

4.6.2 FDG prediction

We show in Figure 4.8 the fitted linear models between the FDG and early-AV45 uptake of 57 subjects from the GMC cohort for additional brain regions, as well as their raw values.

4.6.3 Tracer correction

As mentioned in Section 4.2.6, we provide in Figure 4.9 additional histograms showing the effect of the tracer correction in different brain regions. Consistently with what has been observed, we notice a reduction of the variability of the values for the flutemetamol group.

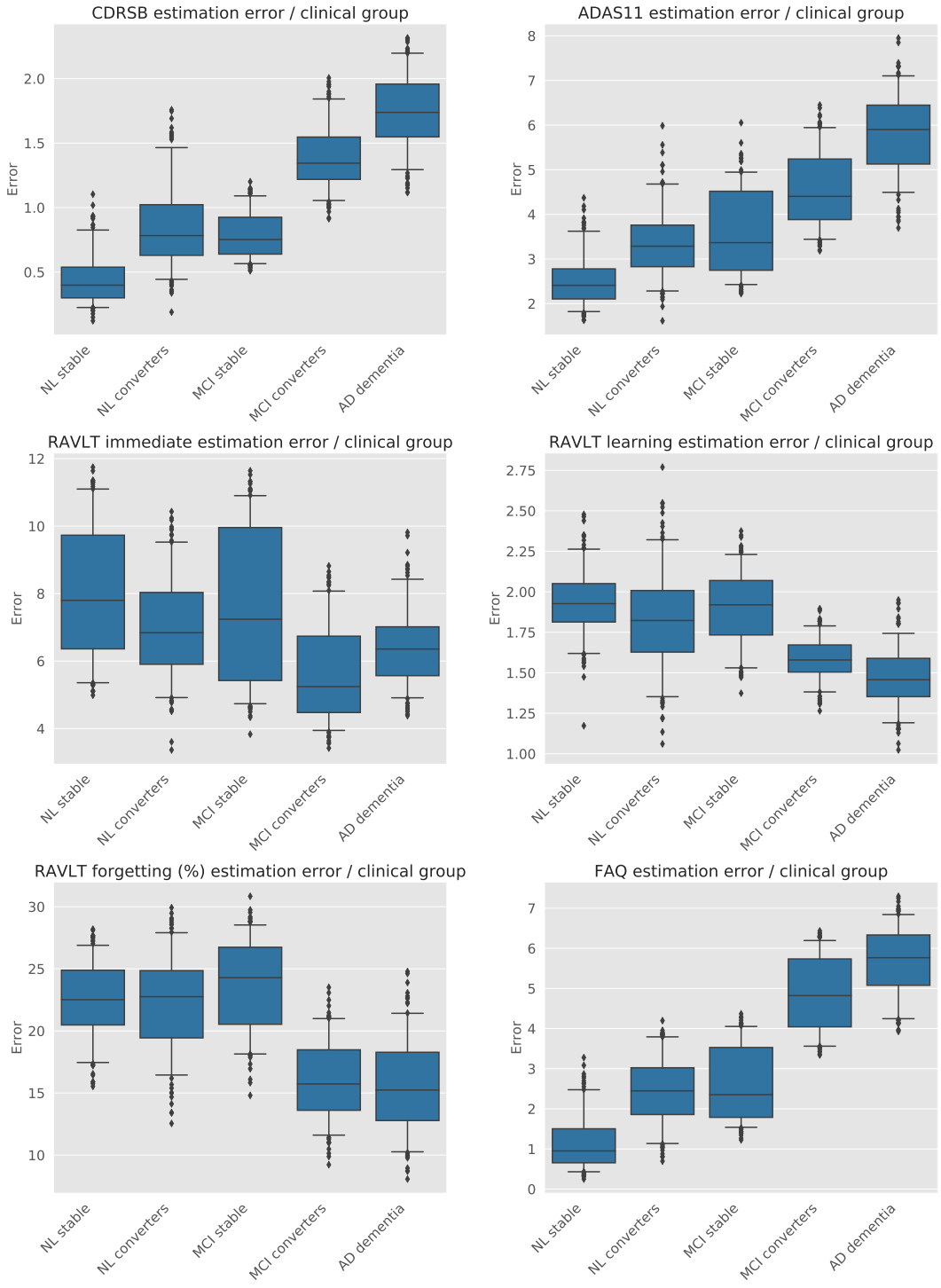

Figure 4.7: Distribution of the absolute error between the ground truth and k-means prediction for all the neuro-psychological tests across clinical groups in the ADNI cohort. NL: cognitively healthy; MCI: mild cognitive impairment; AD dementia: Alzheimer's disease dementia; CDRSB: Clinic Dementia Rating Sum of Boxes; ADAS11: Alzheimer's Disease Assessment Scale; MMSE: Mini-Mental State Examination; FAQ: Functional Assessment Questionnaire; RAVLT: Rey Auditory Verbal Learning Test; ADNI: Alzheimer's Disease Neuroimaging Initiative.

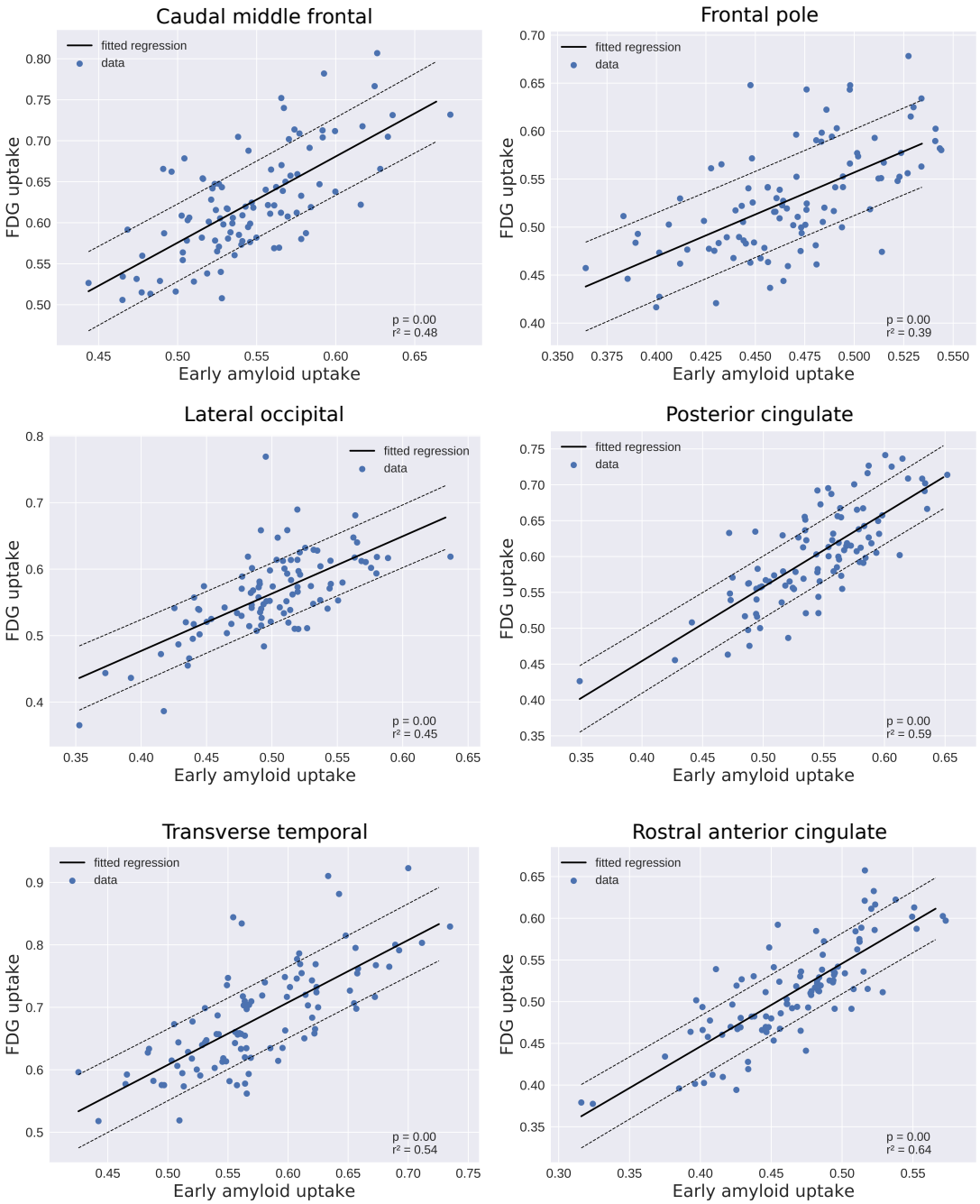

Figure 4.8: Scatter plot between the regional early-amyloid uptake and the corresponding FDG uptake for 57 patients of the GMC cohort. Solid black lines show the fitted linear model between regional FDG and early-amyloid. The dashed-lines represent 95% confidence interval. GMC: Geneva Memory Center FDG: (18)F-fluorodeoxyglucose Positron Emission Tomography (PET) imaging.

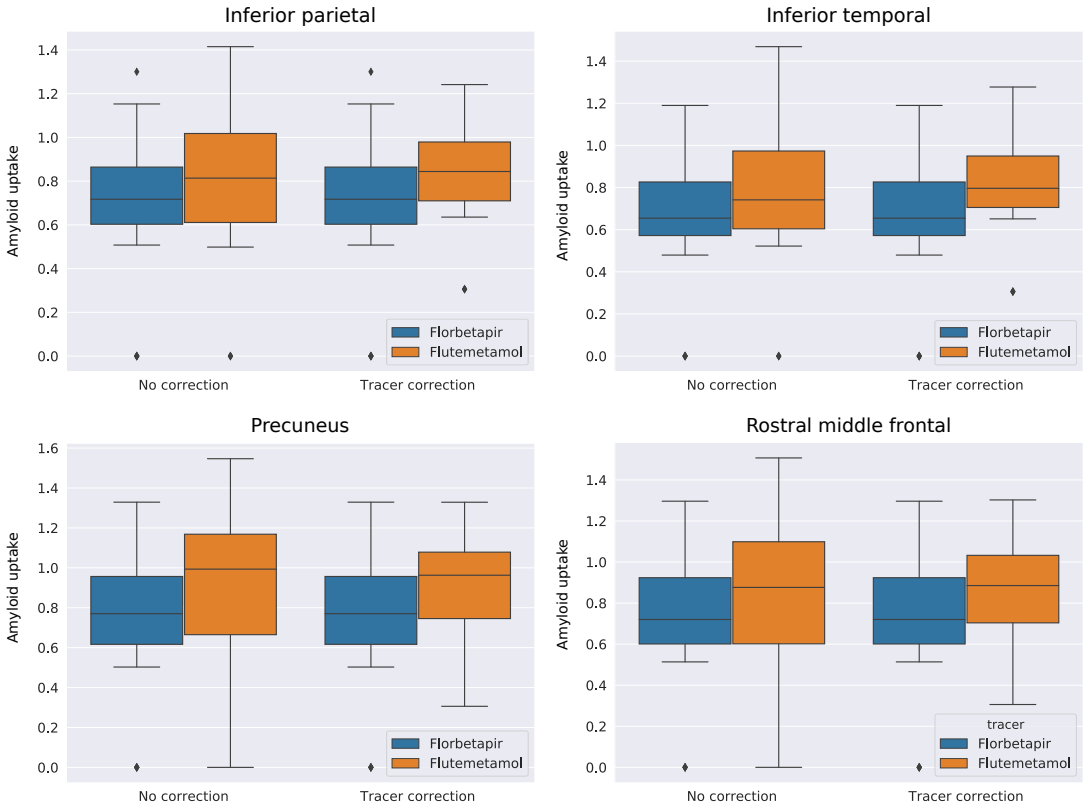

Figure 4.9: Comparison of the distribution of the regional amyloid uptake before and after correction for the subjects from the GMC cohort, depending on the tracer used during acquisition. 76 Amyloid-PET scans were acquired using florbetapir and 17 using flutemetamol. GMC: Geneva Memory Center; SUVR: Standardized Uptake Value Ratio.

4.6.4 Robustness to the clinical scores prediction

In the case of the GMC cohort six clinical scores over seven were imputed based on the procedure detailed in Section 4.2.6. These approximated neuropsychological assessments are used to compute the z-scores z^{cli} of the individuals from the GMC cohort, thus affecting their estimated disease severity shown in Figure 4.3, but also the progression of the z-scores for the GMC cohort in Figure 4.6. We evaluate the impact of this approximation by computing the group-wise disease severity and the progression of the z-scores in three scenarios: (i) clinical scores are imputed following the procedure described in Section 4.2.6; (ii) clinical scores are imputed with an additional approximation by adding twice their respective average prediction error; (iii) clinical scores are imputed with an additional approximation by subtracting twice their average prediction error. We observe in Figures 4.10 and 4.11 that introducing this additional approximation error when imputing the clinical scores leads to rather small changes compared with the results obtained with the regular imputation, for both the group-wise disease severity and the z-scores progression. This shows that SimulAD is robust to the imputation of the clinical scores for the GMC cohort, and that despite the induced margin of error the estimated disease severity and z-scores progression are reliable.

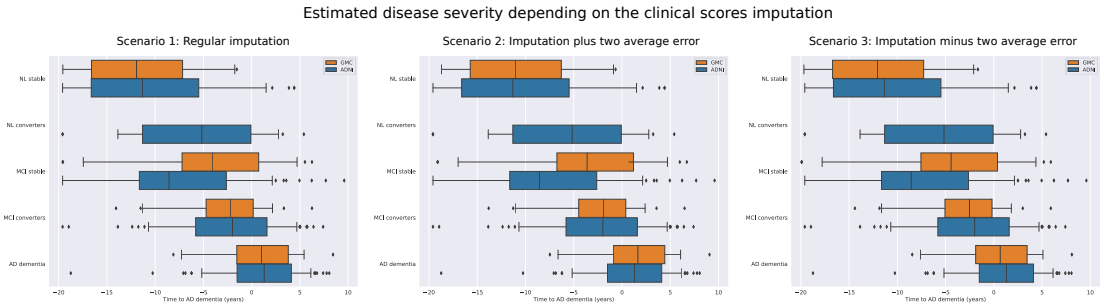

Figure 4.10: Effect of the clinical scores imputation on the estimated disease severity for the subjects from the GMC cohort. Scenario 1: Clinical scores are imputed following the procedure described in Section 4.2.6. Scenario 2: Clinical scores are imputed with an additional approximation by adding twice their respective average prediction error. Scenario 3: Clinical scores are imputed with an additional approximation by subtracting twice their average prediction error. ADNI: Alzheimer's Disease Neuroimaging Initiative; GMC: Geneva Memory Center; NL: cognitively healthy; MCI: mild cognitive impairment; AD dementia: Alzheimer's disease dementia. Converters are cognitively unimpaired and MCI subjects whose clinical diagnosis change during follow-up.

4.6.5 Comparison of the simulated evolution of clinical and imaging measures based on ADNI and GMC cohorts

The similarity between the two models of disease progression based on the ADNI and GMC cohorts respectively (*cf.* Section 4.3.3) is confirmed in Figure 4.12. In this figure, we show the temporal evolution of the error between the clinical and regional imaging

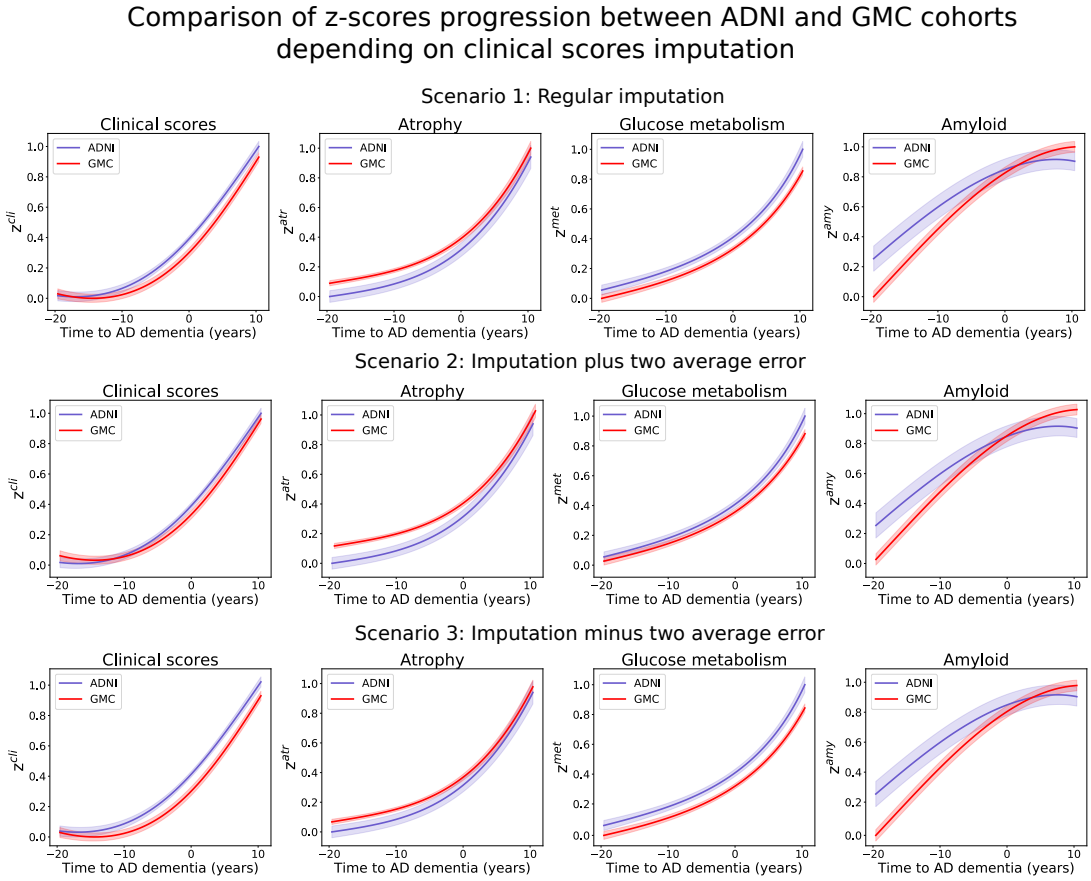

Figure 4.11: Effect of the clinical scores imputation on the long-term dynamics for the subjects from the GMC cohort. Shadowed areas represent the standard deviation of the average trajectory. Scenario 1: Clinical scores are imputed following the procedure described in Section 4.2.6. Scenario 2: Clinical scores are imputed with an additional approximation by adding twice their respective average prediction error. Scenario 3: Clinical scores are imputed with an additional approximation by subtracting twice their average prediction error. ADNI: Alzheimer's Disease Neuroimaging Initiative; GMC: Geneva Memory Center.

measurements simulated based on the ADNI cohort and the ones based on the GMC cohort. We observe that for imaging markers the regional error remains below 10% for most regions across the different modalities, and reaches a maximum of approximately 20% in the case of amyloid deposition at the earliest stages of the disease. Concerning clinical scores, the error remains below 10% and 20% for MMSE and RAVLT immediate respectively. When averaged across time, brain regions and clinical scores, the error is of 3%, 6%, 7% and 12% for MRI, FDG-PET, amyloid-PET data and clinical scores respectively.

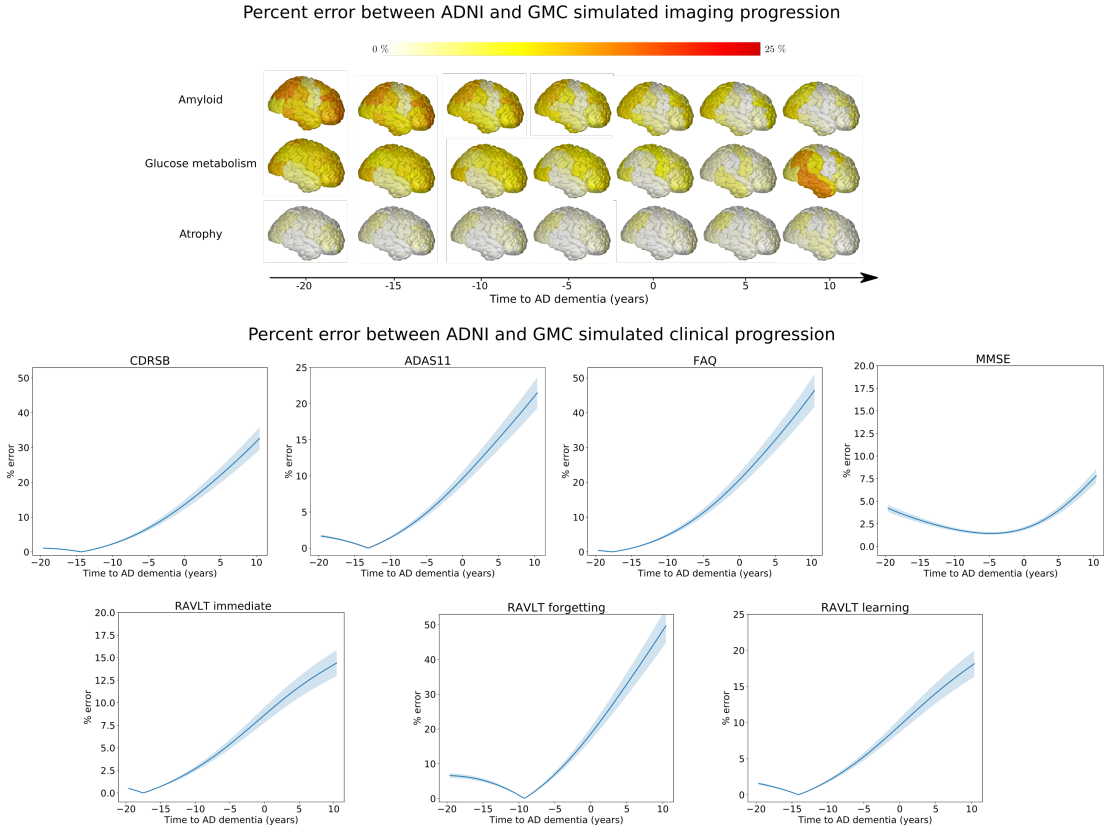

Figure 4.12: Modelled long-term evolution of the error between the progressions simulated based on the ADNI and GMC cohorts respectively in terms of cortical measurements and clinical scores. Shadowed areas represent the standard deviation of the average trajectory. ADNI: Alzheimer's Disease Neuroimaging Initiative; GMC: Geneva Memory Center; CDRSB: Clinic Dementia Rating Sum of Boxes; ADAS11: Alzheimer's Disease Assessment Scale; FAQ: Functional Assessment Questionnaire; RAVLT: Rey Auditory Verbal Learning Test; MMSE: Mini-Mental State Examination.

5 Conclusion

Disease progression models allow to uncover the mechanisms underlying Alzheimer's disease (AD) evolution in a data-driven manner. In addition to providing valuable insights about the pathological progression, they also offer a broad range of potential clinical applications. Indeed, these approaches are perfectly suited for diagnosis, prognosis and for evaluating treatment efficacy. In this thesis, we developed two novel disease progression models built upon state-of-the-art machine learning methods aiming at improving our comprehension of the disease evolution. More specifically, our objective was threefold: (i) propose a computational method able to leverage the availability of high-dimensional multi-modal imaging data; (ii) investigate the causal relationships between the biomarkers evolution and enable the simulation of intervention scenarios; (iii) demonstrate the generalization of our model and its validity in terms of disease staging in a clinical context. In the following sections, we summarize the methodological contributions of this thesis as well as the obtained results. We also propose new applications for our methods and build upon their limitations to propose research perspectives for the field of disease progression modelling.

5.1 Summary of the Main Contributions

Monotonic Gaussian Process for Spatio-Temporal Disease Progression Modeling in Brain Imaging Data.

In Chapter 2, we introduced a novel method to estimate a model of AD progression which leverages multi-modal volumetric images. We assumed that each pathological process can be characterized by multiple spatio-temporal trajectories, and proposed to disentangle them by modelling the data through a matrix factorization between temporal and spatial sources. In order to enforce the estimation of a realistic pathological progression, we introduced a set of clinically-inspired constraints for the temporal and spatial sources. More precisely, the temporal sources were defined as monotonic Gaussian Processes to ensure a steady pathological progression, while the spatial sources were modelled as sparse maps convolved at different scales to deal with the non-stationarity of medical images. The method enabled to disentangle different temporal profiles for amyloid deposition, glucose hypometabolism and gray matter atrophy targeting brain regions associated with neuropathological changes typical of AD. Moreover, these multiple spatio-

temporal trajectories were following an estimated time-line representing the disease progression and which was related to clinical diagnosis.

Main contributions:

- We proposed a novel methodology in which AD progression was formulated as a sources separation problem. We introduced a set of constraints on the spatio-temporal sources to ensure the estimation of a realistic disease progression. Our approach was scalable to the analysis of volumetric multi-modal images, thus allowing to describe the changes affecting the brain during AD at the highest spatial resolution.

- The model identified differential temporal patterns of amyloid deposition, glucose hypometabolism and gray matter atrophy mapping brain regions key to AD.

- The estimated time-line describing the disease course was associated to clinical diagnosis and correlated with neuro-psychological scores.

Simulating the Outcome of Amyloid Treatments in Alzheimer's Disease from Imaging and Clinical data

Our goal in Chapter 3 was to provide a computational framework to study the interactions between the biomarkers dynamics, while enabling the simulation of intervention strategies. To do so we introduced SimulAD, a method modelling AD progression by means of a dynamical system between clinical and multi-modal imaging-derived data. Relying on a variational autoencoder framework, we encoded clinical and imaging data in a lower-dimensional space, within which each dimension was associated to a given process (i.e cognitive decline or gray matter atrophy). We postulated the existence of a system of Ordinary Differential Equations (ODEs) in the latent space relating each coordinate, and inferred its parameters through the analysis of longitudinal clinical and imaging-derived data. Once the model was estimated, this formulation as a latent system of ODEs allowed to directly modify the dynamics of certain processes, such as amyloid deposition, and to investigate the impact of this modification on the pathological progression. We showed that in order to result in statistically significant improvement of cognitive outcomes, anti-amyloid treatment should be administered much earlier than what is currently done. More precisely, we estimated that amyloid deposition should be completely blocked at least seven years before a clinical diagnosis of dementia is made, in order to observe a significantly powered improvement of neuro-psychological evaluations.

Main contributions:

- From the methodological point of view this approach proposed an original formulation, by modelling AD progression as a dynamical system relating clinical and multi-modal imaging-derived measurements. Thanks to this modelling choice, we could study the causal relationships between different pathological processes characteristic of AD.

- SimulAD allowed to estimate the long-term evolution of clinical and imaging-derived measurements on a time-line spanning the whole disease course. This time-line was used as a reference for individual disease staging with performances comparable to a state-of-the-art approach [Lorenzi, 2017; Lorenzi, 2018].

- SimulAD enabled the creation of hypothetical scenarios of disease progression. In particular, we simulated the effect of an anti-amyloid treatment on cognitive endpoints depending on both intervention time and drug dosage. We showed that it is critical to intervene during the pre-symptomatic phase in order to significantly improve cognitive outcomes. This aspect of the model entails great potential for providing quantitative guidelines when planning clinical trials.

SimulAD: A Dynamical Model for Personalized Simulation and Disease Staging in Alzheimer's Disease

In Chapter 4, we presented a preliminary validation of SimulAD on an independent cohort from the Geneva Memory Center (GMC). Our objective was to assess the generalization of SimulAD on an independent dataset, in order to demonstrate its interest in a clinical context. The GMC study used to evaluate the model included subjects who underwent clinical evaluations, MRI, Amyloid-PET, FDG-PET and Tau-PET scans. We developed a specific pre-processing pipeline to correct the various biases between the ADNI cohort (the training dataset) and the GMC cohort (the testing dataset). We showed that SimulAD was able to differentiate subjects across clinical groups and that the disease severity estimated by the method correlated with clinical evaluations and standard imaging biomarkers. Finally, the progression of clinical and imaging markers simulated by our approach was similar on both ADNI and GMC cohorts, highlighting the reliability of the method.

Main contributions:

- We developed a pre-processing pipeline allowing to apply SimulAD on a cohort from a memory clinic.

- We showed that the progression of clinical and imaging-derived measurements estimated by SimulAD on ADNI was reproducible based on the data from the GMC. This result highlights the generalization of the pathological progression estimated by our method.

- The disease severity estimated by SimulAD was associated with clinical diagnosis for subjects from the GMC cohort. We also showed that it correlated with clinical scores and well-established biomarkers of AD. These results support the validity of SimulAD for staging patients.

5.2 Perspectives and Future Applications

5.2.1 Application to other neurodegenerative disorders

In this thesis, the proposed disease progression models were applied to uncover the long-term trajectories of clinical and imaging markers characteristic of AD. Yet, our approaches could be used to study other neurodegenerative disorders leading to dementia and that share common features with AD.

For instance, Huntington's disease is a neurodegenerative disorder whose symptoms principally include motor and cognitive decline, causing devastating effects on patients and greatly reducing their life expectancy [Ross, 2014]. The disease is characterized by a pre-symptomatic period comparable to the one of AD, during which intervention is likely to result in effective results. There exists a broad range of biomarkers allowing to track the disease: motor and cognitive assessments, regional gray matter atrophy, FDG-PET imaging data, but none of them is sufficient alone to accurately identify the individual disease stage during the whole pathological progression. Parkinson's disease is another neurodegenerative disorder which also affects the motor system, ultimately leading to dementia. While recent studies have shown that imaging techniques such as MRI and PET might provide biomarkers that could help identifying and monitoring the disease evolution [Poewe, 2017], the mechanisms underlying the pathological progression still remain unclear.

Based on these observations, disease progression models seem to be perfectly suited to provide valuable insights on the evolution of these neurodegenerative disorders. Lately, the Event-Based Model presented in Chapter 1 was applied in order to reconstruct the sequence of events of Huntington's [Wijeratne, 2018] and Parkinson's diseases [Oxtoby, 2021], from the analysis of clinical and imaging markers relevant to each pathology. Similarly, we believe that the methods presented in Chapter 2 and 3 entail potential to uncover the dynamics characterizing both Huntington's disease and Parkinson's disease.

For instance, the model presented in Chapter 2 would enable the identification of differential spatio-temporal patterns characteristic of these diseases, thus allowing to obtain a fine-grained description of their topographic pathological patterns and associated temporal dynamics. Moreover, the disease-specific time-line estimated by our approaches could allow to identify patients at-risk to develop the disease or to monitor treatment efficacy during clinical trials. Finally, by formulating the disease progression through a system of latent ODEs as proposed in Chapter 3, we could simulate scenarios of drug intervention that could help understanding the relationships between the pathological processes at stake. This could prove particularly useful in the case of Parkinson's disease, where the interactions between the many pathways and mechanisms characterizing the disease remain partially unknown.

5.2.2 Learning from multi-centric studies

In Chapter 4, we emphasized the importance of evaluating the generalization of disease progression models on independent datasets, in order to support their adoption in clinical routine. However, we could also consider the problem of generalization from the training step. To this end, we could develop disease progression models based not only on a single database, such as the one from the Alzheimer's Disease Neuroimaging Initiative (ADNI), but by relying on many cohorts. Moreover, this would allow to consider additional biomarkers, which could provide essential information to better understand AD progression.

Currently, there exists many studies composed of thousands of subjects for which multi-modal imaging data was acquired. Obviously, the ADNI is the first one to come to mind, as it has been used to develop hundreds of machine learning methods. However, studies such as the Australian Imaging Biomarkers and Lifestyle (AIBL), Open Access Series of Imaging Studies (OASIS) or the Amyloid Imaging to Prevent Alzheimer's Disease (AMYPAD) [Ellis, 2009; Marcus, 2010; Lopes Alves, 2020] also provide access to a large variety of multi-modal imaging data for large cohorts. Jointly learning from these databases could enable to estimate disease progression models with better generalization properties, while including a larger panel of biomarkers.

Given this context, multi-task learning [Caruana, 1993; Baxter, 1997] is a relevant solution to leverage the availability of large and complementary multi-centric studies. For instance, in the case of SimulAD we could assume that the encoding and decoding functions are specific to each cohort, while the parameters of the latent ODE system are shared across them. Then, the model would be trained by minimizing the joint data likelihood between the different datasets. In this way, the model would capture the distribution of each dataset, while estimating a unique mechanism driving AD progression. Thanks to this joint analysis of the multiple cohorts, we would therefore obtain a more

generalized and reliable model of disease progression. Moreover, this could improve the performances in terms of disease staging on independent cohorts. Finally, we could consider additional biomarkers, such as tau, which is available in some studies but still remains scarce compared to other biomarkers like amyloid.

5.2.3 Identifying subtypes of disease progression and accounting for risk factors: towards precision medicine

We presented in Chapters 2 and 3 two methods allowing to estimate models of AD evolution. While our approaches provide valuable insights about the pathological progression, they don't account for important aspects of the disease.

For instance, a problem that was not addressed in this thesis is the existence of various spatial patterns that can be associated with AD. Indeed, it has been shown that there exists subtypes of AD [Murray, 2011] that are atypical in the sense that they differ from the expected neuropathological changes. More specifically, it is usually considered that there are three variants of AD based on the distribution of neurofibrillary tangles (NFTs). Typical AD presents a rather uniform distribution of NFTs in the hippocampus, hippocampal-sparing AD is characterized by lower counts of NFTs in the hippocampus, while limbic-predominant AD shows a higher spread of NFTs in the hippocampus. These subtypes of AD are associated with specific atrophy patterns [Whitwell, 2012], and can therefore be identified thanks to structural MRI. Another issue that was not investigated in this thesis is the effect of risk factors on the pathological progression. In particular, it has been observed that genetic factors such as the presence of the allele $\epsilon 4$ of the apolipoprotein (APOE4) strongly impacts the disease onset and progression [Kim, 2009]. Moreover, many studies have shown that cardiovascular risk factors and life-style seem to play an important role in AD [Baumgart, 2015].

Overall, the aforementioned problems point to the direction of precision medicine. Indeed, whether we try to identify a subtype of disease progression or to account for specific risk factors, the underlying objective is to personalize the pathological evolution. Within this context, the methods presented in this manuscript represent a first step but need to be further extended. They allow to estimate a global progression of AD but don't account in their current formulation for the possibility to deviate from this trajectory. A promising extension would be to associate our models with a clustering method, such as Gaussian Mixture Models, in order to automatically identify various pathological trajectories. This could be done in the case of SimulAD by clustering subjects within the latent space and estimate a system of ODEs specific to each cluster. This modelling approach might allow to identify multiple trajectories of AD progression, which would better represent the heterogeneous nature of the pathology.

Bibliography

[Abi Nader, 2018] Clement Abi Nader, Nicholas Ayache, Philippe Robert, and Marco Lorenzi. “Alzheimer’s Disease Modelling and Staging Through Independent Gaussian Process Analysis of Spatio-Temporal Brain Changes”. In: *Understanding and Interpreting Machine Learning in Medical Image Computing Applications*. Cham: Springer International Publishing, 2018, pp. 3–14 (cit. on p. 12).

[Abi Nader, 2020] Clément Abi Nader, Nicholas Ayache, Philippe Robert, and Marco Lorenzi. “Monotonic Gaussian Process for spatio-temporal disease progression modeling in brain imaging data”. In: *NeuroImage* 205 (2020) (cit. on pp. 12, 63).

[Abi Nader, 2021] Clément Abi Nader, Nicholas Ayache, Giovanni B Frisoni, Philippe Robert, and Marco Lorenzi. “Simulating the outcome of amyloid treatments in Alzheimer’s Disease from multi-modal imaging and clinical data”. In: *Brain communications* (Feb. 2021) (cit. on pp. 62, 92).

[Aisen, 2010] P. S. Aisen, R. C. Petersen, M. C. Donohue, A. Gamst, R. Raman, R. G. Thomas, S. Walter, J. Q. Trojanowski, L. M. Shaw, L. A. Beckett, C. R. Jack, W. Jagust, A. W. Toga, A. J. Saykin, J. C. Morris, R. C. Green, and M. W. Weiner. “Clinical Core of the Alzheimer’s Disease Neuroimaging Initiative: progress and plans”. In: *Alzheimers Dement* 6.3 (May 2010), pp. 239–246 (cit. on p. 3).

[Aisen, 2018] P. S. Aisen, E. Siemers, D. Michelson, S. Salloway, C. Sampaio, M. C. Carrillo, R. Sperling, R. Doody, P. Scheltens, R. Bateman, M. Weiner, and B. Vellas. “What Have We Learned from Expedition III and EPOCH Trials? Perspective of the CTAD Task Force”. In: *J Prev Alzheimers Dis* 5.3 (2018), pp. 171–174 (cit. on pp. 4, 63, 78).

[Akaike, 1998] Hirotogu Akaike. “Information Theory and an Extension of the Maximum Likelihood Principle”. In: *Selected Papers of Hirotugu Akaike*. New York, NY: Springer New York, 1998, pp. 199–213 (cit. on p. 69).

[Alzheimer Association Report, 2020] Alzheimer Association Report. “2020 Alzheimer’s disease facts and figures”. In: *Alzheimer’s & Dementia* 16.3 (2020), pp. 391–460. eprint: https://alz-journals.onlinelibrary.wiley.com/doi/pdf/10.1002/alz.12068 (cit. on p. 1).

[Antelmi, 2019] Luigi Antelmi, Nicholas Ayache, Philippe Robert, and Marco Lorenzi. “Sparse Multi-Channel Variational Autoencoder for the Joint Analysis of Heterogeneous Data”. In: *ICML 2019 - 36th International Conference on Machine Learning*. Long Beach, United States, June 2019 (cit. on p. 64).

[Arbabshirani, 2017] Mohammad R. Arbabshirani, Sergey Plis, Jing Sui, and Vince D. Calhoun. “Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls”. In: *NeuroImage* 145 (2017). Individual Subject Prediction, pp. 137–165 (cit. on p. 91).

[Archetti, 2019] D. Archetti, S. Ingala, V. Venkatraghavan, V. Wottschel, A. L. Young, M. Bellio, E. E. Bron, S. Klein, F. Barkhof, D. C. Alexander, N. P. Oxtoby, G. B. Frisoni, and A. Redolfi. “Multi-study validation of data-driven disease progression models to characterize evolution of biomarkers in Alzheimer’s disease”. In: *Neuroimage Clin* 24 (2019), p. 101954 (cit. on pp. 92, 106).

[Ashburner, 2000] John Ashburner and K. J. Friston. “Voxel-based morphometry–the methods”. In: *NeuroImage* 11.6 Pt 1 (June 2000), pp. 805–821 (cit. on pp. 6, 29).

[Ashburner, 2007] John Ashburner. “A fast diffeomorphic image registration algorithm”. In: *NeuroImage* 38.1 (2007), pp. 95–113 (cit. on p. 29).

[Barthel, 2020] Henryk Barthel. “First Tau PET Tracer Approved: Toward Accurate In Vivo Diagnosis of Alzheimer Disease”. In: *Journal of Nuclear Medicine* 61.10 (2020), pp. 1409–1410. eprint: https://jnm.snmjournals.org/content/61/10/1409.full.pdf (cit. on p. 3).

[Bartlett, 2012] J. W. Bartlett, C. Frost, N. Mattsson, T. Skillb?ck, K. Blennow, H. Zetterberg, and J. M. Schott. “Determining cut-points for Alzheimer’s disease biomarkers: statistical issues, methods and challenges”. In: *Biomark Med* 6.4 (Aug. 2012), pp. 391–400 (cit. on p. 5).

[Bateman, 2012] Randall J. Bateman, Chengjie Xiong, Tammie L.S. Benzinger, Anne M. Fagan, Alison Goate, Nick C. Fox, Daniel S. Marcus, Nigel J. Cairns, Xianyun Xie, Tyler M. Blazey, David M. Holtzman, Anna Santacruz, Virginia Buckles, Angela Oliver, Krista Moulder, Paul S. Aisen, Bernardino Ghetti, William E. Klunk, Eric McDade, Ralph N. Martins, Colin L. Masters, Richard Mayeux, John M. Ringman, Martin N. Rossor, Peter R. Schofield, Reisa A. Sperling, Stephen Salloway, and John C. Morris. "Clinical and Biomarker Changes in Dominantly Inherited Alzheimer's Disease". In: *New England Journal of Medicine* 367.9 (2012). PMID: 22784036, pp. 795–804 (cit. on pp. 4, 5, 30, 63).

[Battle, 2018] Mark R. Battle, Lovena Chedumbarum Pillay, Val J. Lowe, David Knopman, Bradley Kemp, Christopher C. Rowe, Vincent Doré, Victor L. Villemagne, and Christopher J. Buckley. "Centiloid scaling for quantification of brain amyloid with [18F]flutemetamol using multiple processing methods". In: *EJNMMI Research* 8.1 (Dec. 2018), p. 107 (cit. on p. 98).

[Baumgart, 2015] M. Baumgart, H. M. Snyder, M. C. Carrillo, S. Fazio, H. Kim, and H. Johns. "Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective". In: *Alzheimers Dement* 11.6 (June 2015), pp. 718–726 (cit. on p. 120).

[Baxter, 1997] Jonathan Baxter. "A Bayesian/Information Theoretic Model of Learning to Learn via Multiple Task Sampling". In: *Machine Learning* 28.1 (July 1997), pp. 7–39 (cit. on p. 119).

[Bekris, 2010] L. M. Bekris, C. E. Yu, T. D. Bird, and D. W. Tsuang. "Genetics of Alzheimer disease". In: *J Geriatr Psychiatry Neurol* 23.4 (Dec. 2010), pp. 213–227 (cit. on p. 5).

[Bennett, 2006] D. A. Bennett, J. A. Schneider, Z. Arvanitakis, J. F. Kelly, N. T. Aggarwal, R. C. Shah, and R. S. Wilson. "Neuropathology of older persons without cognitive impairment from two community-based studies". In: *Neurology* 66.12 (June 2006), pp. 1837–1844 (cit. on p. 2).

[Bilgel, 2015] M. Bilgel, B. Jedynak, D. F. Wong, S. M. Resnick, and J. L. Prince. "Temporal Trajectory and Progression Score Estimation from Voxelwise Longitudinal Imaging Measures: Application to Amyloid Imaging". In: *Inf Process Med Imaging* 24 (2015), pp. 424–436 (cit. on pp. 15, 63).

[Bilgel, 2016] M. Bilgel, J. L. Prince, D. F. Wong, S. M. Resnick, and B. M. Jedynak. “A multivariate nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging”. In: *Neuroimage* 134 (July 2016), pp. 658–670 (cit. on pp. 7, 92).

[Blennow, 2003] K. Blennow and H. Hampel. “CSF markers for incipient Alzheimer’s disease”. In: *Lancet Neurol* 2.10 (Oct. 2003), pp. 605–613 (cit. on p. 2).

[Blennow, 2010] K. Blennow, H. Hampel, M. Weiner, and H. Zetterberg. “Cerebrospinal fluid and plasma biomarkers in Alzheimer disease”. In: *Nat Rev Neurol* 6.3 (Mar. 2010), pp. 131–144 (cit. on p. 78).

[Bowles, 2018] Christopher Bowles, Roger Gunn, Alexander Hammers, and Daniel Rueckert. “Modelling the progression of Alzheimer’s disease in MRI using generative adversarial networks”. In: *Medical Imaging 2018: Image Processing*. Ed. by Elsa D. Angelini and Bennett A. Landman. Vol. 10574. International Society for Optics and Photonics. SPIE, 2018, pp. 397–407 (cit. on p. 7).

[Braak, 1991] H. Braak and E. Braak. “Neuropathological stageing of Alzheimer-related changes”. In: *Acta Neuropathol.* 82.4 (1991), pp. 239–259 (cit. on pp. 1, 63).

[Braak, 1997] H. Braak and E. Braak. “Frequency of stages of Alzheimer-related lesions in different age categories”. In: *Neurobiol Aging* 18.4 (1997), pp. 351–357 (cit. on p. 1).

[Brown, 2014] R. K. Brown, N. I. Bohnen, K. K. Wong, S. Minoshima, and K. A. Frey. “Brain PET in suspected dementia: patterns of altered FDG metabolism”. In: *Radiographics* 34.3 (2014), pp. 684–701 (cit. on p. 30).

[Bullmore, 2004] E. Bullmore, J. Fadili, V. Maxim, L. Sendur, B. Whitcher, J. Suckling, M. Brammer, and M. Breakspear. “Wavelets and functional magnetic resonance imaging of the human brain”. In: *NeuroImage* 23 Suppl 1 (2004), S234–249 (cit. on p. 14).

[Burnham, 2020] Samantha C Burnham, Noelia Fandos, Christopher Fowler, Virginia Pérez-Grijalba, Vincent Dore, James D Doecke, Rosita Shishegar, Timothy Cox, Jurgen Fripp, Christopher Rowe, Manuel Sarasa, Colin L Masters, Pedro Pesini, and Victor L Villemagne. “Longitudinal evaluation of the natural history of amyloid-β in plasma and brain”. In: *Brain Communications* 2.1 (Apr. 2020). fcaa041. eprint: `https://academic.oup.com/braincomms/article-pdf/2/1/fcaa041/33639843/fcaa041.pdf` (cit. on p. 63).

[Calhoun, 2009] V. D. Calhoun, J. Liu, and T. Adali. “A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data”. In: *NeuroImage* 45.1 Suppl (Mar. 2009), S163–172 (cit. on p. 13).

[Caruana, 1993] Richard Caruana. “Multitask Learning: A Knowledge-Based Source of Inductive Bias”. In: *Proceedings of the Tenth International Conference on Machine Learning*. Morgan Kaufmann, 1993, pp. 41–48 (cit. on p. 119).

[Cash, 2015] D. M. Cash, C. Frost, L. O. Iheme, D. ?nay, M. Kandemir, J. Fripp, O. Salvado, P. Bourgeat, M. Reuter, B. Fischl, M. Lorenzi, G. B. Frisoni, X. Pennec, R. K. Pierson, J. L. Gunter, M. L. Senjem, C. R. Jack, N. Guizard, V. S. Fonov, D. L. Collins, M. Modat, M. J. Cardoso, K. K. Leung, H. Wang, S. R. Das, P. A. Yushkevich, I. B. Malone, N. C. Fox, J. M. Schott, and S. Ourselin. “Assessing atrophy measurement techniques in dementia: Results from the MIRIAD atrophy challenge”. In: *Neuroimage* 123 (Dec. 2015), pp. 149–164 (cit. on p. 65).

[Castro, 2020] Daniel C. Castro, Ian Walker, and Ben Glocker. “Causality matters in medical imaging”. In: *Nature Communications* 11.1 (July 2020), p. 3673 (cit. on p. 92).

[Chen, 2018] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. “Neural Ordinary Differential Equations”. In: *Advances in Neural Information Processing Systems 31*. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. Curran Associates, Inc., 2018, pp. 6571–6583 (cit. on p. 83).

[Clark, 2011] Christopher M Clark, Julie A Schneider, Barry J Bedell, Thomas G Beach, Warren B Bilker, Mark A Mintun, Michael J Pontecorvo, Franz Hefti, Alan P Carpenter, Matthew L Flitter, Michael J Krautkramer, Hank F Kung, R Edward Coleman, P Murali Doraiswamy, Adam S Fleisher, Marwan N Sabbagh, Carl H Sadowsky, Eric P Reiman, Simone P Zehntner, Daniel M Skovronsky, and AV45-A07 Study Group. “Use of florbetapir-PET for imaging beta-amyloid pathology”. In: *JAMA* 305.3 (Jan. 2011), pp. 275–283 (cit. on p. 3).

[Comon, 1994] Pierre Comon. “Independent Component Analysis, a new concept?” In: *Signal Processing* 36 (Apr. 1994), pp. 287–314 (cit. on p. 13).

[Cummings, 2019a] J. Cummings, K. Blennow, K. Johnson, M. Keeley, R. J. Bateman, J. L. Molinuevo, J. Touchon, P. Aisen, and B. Vellas. “Anti-Tau Trials for Alzheimer’s Disease: A Report from the EU/US/CTAD Task Force”. In: *J Prev Alzheimers Dis* 6.3 (2019), pp. 157–163 (cit. on p. 79).

[Cummings, 2019b] J. Cummings, G. Lee, A. Ritter, M. Sabbagh, and K. Zhong. “Alzheimer’s disease drug development pipeline: 2019”. In: *Alzheimers Dement (N Y)* 5 (2019), pp. 272–293 (cit. on p. 63).

[Cutajar, 2017] Kurt Cutajar, Edwin V. Bonilla, Pietro Michiardi, and Maurizio Filippone. “Random Feature Expansions for Deep Gaussian Processes”. In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by Doina Precup and Yee Whye Teh. Vol. 70. Proceedings of Machine Learning Research. International Convention Centre, Sydney, Australia: PMLR, June 2017, pp. 884–893 (cit. on pp. 18, 21).

[Daerr, 2017] S. Daerr, M. Brendel, C. Zach, E. Mille, D. Schilling, M. J. Zacherl, K. B?rger, A. Danek, O. Pogarell, A. Schildan, M. Patt, H. Barthel, O. Sabri, P. Bartenstein, and A. Rominger. “Evaluation of early-phase [18F]-florbetaben PET acquisition in clinical routine cases”. In: *NeuroImage Clinical* 14 (2017), pp. 77–86 (cit. on p. 98).

[Davatzikos, 2009] Christos Davatzikos, Feng Xu, Yang An, Yong Fan, and Susan M. Resnick. “Longitudinal progression of Alzheimer’s-like patterns of atrophy in normal older adults: the SPARE-AD index”. In: *Brain* 132.8 (May 2009), pp. 2026–2035. eprint: `https://academic.oup.com/brain/article-pdf/132/8/2026/742316/awp091.pdf` (cit. on p. 91).

[Delacourte, 1999] A. Delacourte, J. P. David, N. Sergeant, L. Buée, A. Wattez, P. Vermersch, F. Ghozali, C. Fallet-Bianco, F. Pasquier, F. Lebert, H. Petit, and C. Di Menza. “The biochemical pathway of neurofibrillary degeneration in aging and Alzheimer’s disease”. In: *Neurology* 52.6 (Apr. 1999), pp. 1158–1165 (cit. on p. 63).

[Desikan, 2006] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany. “An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest”. In: *NeuroImage* 31.3 (July 2006), pp. 968–980 (cit. on pp. 71, 95).

[Donohue, 2014] 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, and Paul S. Aisen. “Estimating long-term multivariate progression from short-term data”. In: *Alzheimer’s & Dementia* 10.5, Supplement (2014), S400–S410 (cit. on pp. 12, 15, 63, 80).

[Echavarri, 2011] C. Echavarri, P. Aalten, H. B. Uylings, H. I. Jacobs, P. J. Visser, E. H. Gronenschild, F. R. Verhey, and S. Burgmans. “Atrophy in the parahippocampal gyrus as an early biomarker of Alzheimer’s disease”. In: *Brain Struct Funct* 215.3-4 (Jan. 2011), pp. 265–271 (cit. on p. 30).

[Egan, 2019] M. F. Egan, J. Kost, T. Voss, Y. Mukai, P. S. Aisen, J. L. Cummings, P. N. Tariot, B. Vellas, C. H. van Dyck, M. Boada, Y. Zhang, W. Li, C. Furtek, E. Mahoney, L. Harper Mozley, Y. Mo, C. Sur, and D. Michelson. “Randomized Trial of Verubecestat for Prodromal Alzheimer’s Disease”. In: *N. Engl. J. Med.* 380.15 (Apr. 2019), pp. 1408–1420 (cit. on pp. 63, 75).

[Ellis, 2009] K. A. Ellis, A. I. Bush, D. Darby, et al. “The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer’s disease”. In: *Int Psychogeriatr* 21.4 (Aug. 2009), pp. 672–687 (cit. on p. 119).

[Etters, 2008] L. Etters, D. Goodall, and B. E. Harrison. “Caregiver burden among dementia patient caregivers: a review of the literature”. In: *J Am Acad Nurse Pract* 20.8 (Aug. 2008), pp. 423–428 (cit. on p. 1).

[Falahati, 2014] Farshad Falahati, Eric Westman, and Andrew Simmons. “Multivariate Data Analysis and Machine Learning in Alzheimer’s Disease with a Focus on Structural Magnetic Resonance Imaging”. In: *Journal of Alzheimer’s Disease* 41 (2014). 3, pp. 685–708 (cit. on p. 91).

[Fischl, 2000] B. Fischl and A. M. Dale. “Measuring the thickness of the human cerebral cortex from magnetic resonance images”. In: *Proc Natl Acad Sci U S A* 97.20 (Sept. 2000), pp. 11050–11055 (cit. on p. 6).

[Fonteijn, 2012] H. M. Fonteijn, M. Modat, M. J. Clarkson, J. Barnes, M. Lehmann, N. Z. Hobbs, R. I. Scahill, S. J. Tabrizi, S. Ourselin, N. C. Fox, and D. C. Alexander. “An event-based model for disease progression and its application in familial Alzheimer’s disease and Huntington’s disease”. In: *NeuroImage* 60.3 (Apr. 2012), pp. 1880–1889 (cit. on pp. 5, 12, 63, 92).

[Fox, 2004] N. C. Fox and J. M. Schott. “Imaging cerebral atrophy: normal ageing to Alzheimer’s disease”. In: *Lancet* 363.9406 (Jan. 2004), pp. 392–394 (cit. on p. 3).

[Frisoni, 2010] G. B. Frisoni, N. C. Fox, C. R. Jack, P. Scheltens, and P. M. Thompson. “The clinical use of structural MRI in Alzheimer disease”. In: *Nat Rev Neurol* 6.2 (Feb. 2010), pp. 67–77 (cit. on pp. 3, 30).

[Gamberger, 2017] D. Gamberger, N. Lavrač, S. Srivatsa, R. E. Tanzi, and P. M. Doraiswamy. “Identification of clusters of rapid and slow decliners among subjects at risk for Alzheimer’s disease”. In: *Sci Rep* 7.1 (July 2017), p. 6763 (cit. on p. 75).

[Garbarino, 2019] Sara Garbarino and Marco Lorenzi. “Modeling and Inference of Spatio-Temporal Protein Dynamics Across Brain Networks”. In: *IPMI 2019 - 26th International Conference on Information Processing in Medical Imaging*. Vol. 11492. LNCS. Hong-Kong, China: Springer, June 2019, pp. 57–69 (cit. on p. 64).

[Gauthier, 2019] S. Gauthier, J. Alam, H. Fillit, T. Iwatsubo, H. Liu-Seifert, M. Sabbagh, S. Salloway, C. Sampaio, J. R. Sims, B. Sperling, R. Sperling, K. A. Welsh-Bohmer, J. Touchon, B. Vellas, and P. Aisen. “Combination Therapy for Alzheimer’s Disease: Perspectives of the EU/US CTAD Task Force”. In: *J Prev Alzheimers Dis* 6.3 (2019), pp. 164–168 (cit. on p. 79).

[Ghahramani, 2001] Z. Ghahramani and M.J Beal. “Graphical models and variational methods”. In: *Advanced mean field methods: theory and practice*. Ed. by Manfred Opper and David Saad. Neural Information Processing. MIT, 2001 (cit. on p. 82).

[Ghassemi, 2019] Marzyeh Ghassemi, Tristan Naumann, Peter Schulam, Andrew L. Beam, Irene Y. Chen, and Rajesh Ranganath. “Practical guidance on artificial intelligence for healthcare data”. In: *The Lancet Digital Health* 1.4 (Aug. 2019), e157–e159 (cit. on p. 106).

[Greve, 2014] D. N. Greve, C. Svarer, P. M. Fisher, L. Feng, A. E. Hansen, W. Baare, B. Rosen, B. Fischl, and G. M. Knudsen. “Cortical surface-based analysis reduces bias and variance in kinetic modeling of brain PET data”. In: *Neuroimage* 92 (May 2014), pp. 225–236 (cit. on p. 95).

[Hackmack, 2012] K. Hackmack, F. Paul, M. Weygandt, C. Allefeld, and J. D. Haynes. “Multi-scale classification of disease using structural MRI and wavelet transform”. In: *NeuroImage* 62.1 (Aug. 2012), pp. 48–58 (cit. on p. 14).

[Hao, 2016] W. Hao and A. Friedman. “Mathematical model on Alzheimer’s disease”. In: *BMC Syst Biol* 10.1 (Nov. 2016), p. 108 (cit. on p. 63).

[Hardy, 2002] J. Hardy and D. J. Selkoe. “The amyloid hypothesis of Alzheimer’s disease: progress and problems on the road to therapeutics”. In: *Science* 297.5580 (July 2002), pp. 353–356 (cit. on p. 1).

[Health, 2019] The Lancet Digital Health. “Walking the tightrope of artificial intelligence guidelines in clinical practice”. In: *The Lancet Digital Health* 1.3 (July 2019), e100 (cit. on p. 106).

[Henley, 2019] D. Henley, N. Raghavan, R. Sperling, P. Aisen, R. Raman, and G. Romano. “Preliminary Results of a Trial of Atabecestat in Preclinical Alzheimer’s Disease”. In: *N. Engl. J. Med.* 380.15 (Apr. 2019), pp. 1483–1485 (cit. on p. 63).

[Herholz, 2012] K. Herholz. “Use of FDG PET as an imaging biomarker in clinical trials of Alzheimer’s disease”. In: *Biomark Med* 6.4 (Aug. 2012), pp. 431–439 (cit. on p. 3).

[Honig, 2018] L. S. Honig, B. Vellas, M. Woodward, M. Boada, R. Bullock, M. Borrie, K. Hager, N. Andreasen, E. Scarpini, H. Liu-Seifert, M. Case, R. A. Dean, A. Hake, K. Sundell, V. Poole Hoffmann, C. Carlson, R. Khanna, M. Mintun, R. DeMattos, K. J. Selzler, and E. Siemers. “Trial of Solanezumab for Mild Dementia Due to Alzheimer’s Disease”. In: *N. Engl. J. Med.* 378.4 (Jan. 2018), pp. 321–330 (cit. on pp. 63, 75).

[Howard, 2020] R. Howard and K. Y. Liu. “Questions EMERGE as Biogen claims aducanumab turnaround”. In: *Nat Rev Neurol* 16.2 (Feb. 2020), pp. 63–64 (cit. on p. 63).

[Hyman, 2012] B. T. Hyman, C. H. Phelps, T. G. Beach, E. H. Bigio, N. J. Cairns, M. C. Carrillo, D. W. Dickson, C. Duyckaerts, M. P. Frosch, E. Masliah, S. S. Mirra, P. T. Nelson, J. A. Schneider, D. R. Thal, B. Thies, J. Q. Trojanowski, H. V. Vinters, and T. J. Montine. “National Institute on Aging-Alzheimer’s Association guidelines for the neuropathologic assessment of Alzheimer’s disease”. In: *Alzheimers Dement* 8.1 (Jan. 2012), pp. 1–13 (cit. on p. 2).

[Insel, 2020] Philip S Insel, Elizabeth C Mormino, Paul S Aisen, Wesley K Thompson, and Michael C Donohue. “Neuroanatomical spread of amyloid β and tau in Alzheimer’s disease: implications for primary prevention”. In: *Brain Communications* 2.1 (Feb. 2020). fcaa007. eprint: `https://academic.oup.com/braincomms/article-pdf/2/1/fcaa007/33639942/fcaa007.pdf` (cit. on p. 63).

[IturriaMedina, 2016] Y. Iturria-Medina, R. C. Sotero, P. J. Toussaint, J. M. Mateos-P?rez, A. C. Evans, and Azlheimer’s Disease Neuroimaging Initiative. “Early role of vascular dysregulation on late-onset Alzheimer’s disease based on multifactorial data-driven analysis”. In: *Nat Commun* 7 (June 2016), p. 11934 (cit. on p. 63).

[IturriaMedina, 2017] Yasser Iturria-Medina, Félix M. Carbonell, Roberto C. Sotero, Francois Chouinard-Decorte, and Alan C. Evans. "Multifactorial causal model of brain (dis)organization and therapeutic intervention: Application to Alzheimer's disease". In: *NeuroImage* 152 (2017), pp. 60–77 (cit. on p. 64).

[Jack, 2010] C. R. Jack, D. S. Knopman, W. J. Jagust, L. M. Shaw, P. S. Aisen, M. W. Weiner, R. C. Petersen, and J. Q. Trojanowski. "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade". In: *Lancet Neurol* 9.1 (Jan. 2010), pp. 119–128 (cit. on pp. 3, 12).

[Jack, 2013a] C. R. Jack and D. M. Holtzman. "Biomarker modeling of Alzheimer's disease". In: *Neuron* 80.6 (Dec. 2013), pp. 1347–1358 (cit. on pp. 62, 72).

[Jack, 2013b] C. R. Jack, D. S. Knopman, W. J. Jagust, R. C. Petersen, M. W. Weiner, P. S. Aisen, L. M. Shaw, P. Vemuri, H. J. Wiste, S. D. Weigand, T. G. Lesnick, V. S. Pankratz, M. C. Donohue, and J. Q. Trojanowski. "Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers". In: *Lancet Neurol* 12.2 (Feb. 2013), pp. 207–216 (cit. on pp. 3, 6, 62, 72, 90).

[Jack, 2016] C. R. Jack, D. A. Bennett, K. Blennow, M. C. Carrillo, H. H. Feldman, G. B. Frisoni, H. Hampel, W. J. Jagust, K. A. Johnson, D. S. Knopman, R. C. Petersen, P. Scheltens, R. A. Sperling, and B. Dubois. "A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers". In: *Neurology* 87.5 (Aug. 2016), pp. 539–547 (cit. on p. 2).

[Jack, 2017] C. R. Jack, H. J. Wiste, S. D. Weigand, T. M. Therneau, V. J. Lowe, D. S. Knopman, J. L. Gunter, M. L. Senjem, D. T. Jones, K. Kantarci, M. M. Machulda, M. M. Mielke, R. O. Roberts, P. Vemuri, D. A. Reyes, and R. C. Petersen. "Defining imaging biomarker cut points for brain aging and Alzheimer's disease". In: *Alzheimers Dement* 13.3 (Mar. 2017), pp. 205–216 (cit. on p. 97).

[Jack, 2018] C. R. Jack, D. A. Bennett, K. Blennow, M. C. Carrillo, B. Dunn, S. B. Haeberlein, D. M. Holtzman, W. Jagust, F. Jessen, J. Karlawish, E. Liu, J. L. Molinuevo, T. Montine, C. Phelps, K. P. Rankin, C. C. Rowe, P. Scheltens, E. Siemers, H. M. Snyder, R. Sperling, C. Elliott, E. Masliah, L. Ryan, and N. Silverberg. "NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease". In: *Alzheimers Dement* 14.4 (Apr. 2018), pp. 535–562 (cit. on pp. 2, 65, 80, 91, 105).

[Jagust, 2009] W. J. Jagust, S. M. Landau, L. M. Shaw, J. Q. Trojanowski, R. A. Koeppe, E. M. Reiman, N. L. Foster, R. C. Petersen, M. W. Weiner, J. C. Price, and C. A. Mathis. "Relationships between biomarkers in aging and dementia". In: *Neurology* 73.15 (Oct. 2009), pp. 1193–1199 (cit. on p. 97).

[Jedynak, 2012] B. M. Jedynak, A. Lang, B. Liu, E. Katz, Y. Zhang, B. T. Wyman, D. Raunig, C. P. Jedynak, B. Caffo, and J. L. Prince. "A computational neurodegenerative disease progression score: method and results with the Alzheimer's disease Neuroimaging Initiative cohort". In: *NeuroImage* 63.3 (Nov. 2012), pp. 1478–1486 (cit. on pp. 6, 12, 13, 63, 91).

[Kametani, 2018] F. Kametani and M. Hasegawa. "Reconsideration of Amyloid Hypothesis and Tau Hypothesis in Alzheimer's Disease". In: *Front Neurosci* 12 (2018), p. 25 (cit. on p. 38).

[Kaufman, 2018] Sarah K. Kaufman, Kelly Del Tredici, Talitha L. Thomas, Heiko Braak, and Marc I. Diamond. "Tau seeding activity begins in the transentorhinal/entorhinal regions and anticipates phospho-tau pathology in Alzheimer's disease and PART". In: *Acta Neuropathologica* 136.1 (July 2018), pp. 57–67 (cit. on p. 79).

[Khanal, 2016] Bishesh Khanal, Marco Lorenzi, Nicholas Ayache, and Xavier Pennec. "A biophysical model of brain deformation to simulate and analyze longitudinal MRIs of patients with Alzheimer's disease". In: *NeuroImage* 134 (July 2016), pp. 35–52 (cit. on p. 91).

[Khanal, 2017] Bishesh Khanal, Nicholas Ayache, and Xavier Pennec. "Simulating Longitudinal Brain MRIs with known Volume Changes and Realistic Variations in Image Intensity". In: *Frontiers in Neuroscience* 11.Article 132 (Feb. 2017), p. 18 (cit. on p. 91).

[Kim, 2009] J. Kim, J. M. Basak, and D. M. Holtzman. "The role of apolipoprotein E in Alzheimer's disease". In: *Neuron* 63.3 (Aug. 2009), pp. 287–303 (cit. on pp. 79, 106, 120).

[Kingma, 2013] Diederik P. Kingma and Max Welling. "Auto-Encoding Variational Bayes". In: *CoRR* abs/1312.6114 (2013). arXiv: 1312.6114 (cit. on pp. 20, 22, 41, 42, 83).

[Kingma, 2015a] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. 2015 (cit. on pp. 41, 83).

[Kingma, 2015b] Diederik P. Kingma, Tim Salimans, and Max Welling. "Variational Dropout and the Local Reparameterization Trick". In: *CoRR* abs/1506.02557 (2015) (cit. on p. 20).

[Klein, 2019] Gregory Klein, Paul Delmar, Nicola Voyle, Sunita Rehal, Carsten Hofmann, Danielle Abi-Saab, Mirjana Andjelkovic, Smiljana Ristic, Guoqiao Wang, Randall Bateman, Geoffrey A. Kerchner, Monika Baudler, Paulo Fontoura, and Rachelle Doody. "Gantenerumab reduces amyloid-β plaques in patients with prodromal to moderate Alzheimer's disease: a PET substudy interim analysis". In: *Alzheimer's Research & Therapy* 11.1 (Dec. 2019), p. 101 (cit. on pp. 63, 78).

[Klöppel, 2008] Stefan Klöppel, Cynthia M. Stonnington, Carlton Chu, Bogdan Draganski, Rachael I. Scahill, Jonathan D. Rohrer, Nick C. Fox, Jr Jack Clifford R., John Ashburner, and Richard S. J. Frackowiak. "Automatic classification of MR scans in Alzheimer's disease". In: *Brain* 131.3 (Jan. 2008), pp. 681–689. eprint: `https://academic.oup.com/brain/article-pdf/131/3/681/898663/awm319.pdf` (cit. on p. 91).

[Klunk, 2004] W. E. Klunk, H. Engler, A. Nordberg, Y. Wang, G. Blomqvist, D. P. Holt, M. Bergstr?m, I. Savitcheva, G. F. Huang, S. Estrada, B. Aus?n, M. L. Debnath, J. Barletta, J. C. Price, J. Sandell, B. J. Lopresti, A. Wall, P. Koivisto, G. Antoni, C. A. Mathis, and B. L?ngstr?m. "Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B". In: *Ann Neurol* 55.3 (Mar. 2004), pp. 306–319 (cit. on p. 3).

[Klunk, 2015] W. E. Klunk, R. A. Koeppe, J. C. Price, T. L. Benzinger, M. D. Devous, W. J. Jagust, K. A. Johnson, C. A. Mathis, D. Minhas, M. J. Pontecorvo, C. C. Rowe, D. M. Skovronsky, and M. A. Mintun. "The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET". In: *Alzheimers Dement* 11.1 (Jan. 2015), pp. 1–15 (cit. on p. 98).

[Knopman, 2003] D. S. Knopman, J. E. Parisi, A. Salviati, M. Floriach-Robert, B. F. Boeve, R. J. Ivnik, G. E. Smith, D. W. Dickson, K. A. Johnson, L. E. Petersen, W. C. McDonald, H. Braak, and R. C. Petersen. "Neuropathology of cognitively normal elderly". In: *J Neuropathol Exp Neurol* 62.11 (Nov. 2003), pp. 1087–1095 (cit. on p. 2).

[Kochhann, 2010] R. Kochhann, J. S. Varela, C. S. M. Lisboa, and M. L. F. Chaves. "The Mini Mental State Examination: Review of cutoff points adjusted for schooling in a large Southern Brazilian sample". In: *Dement Neuropsychol* 4.1 (2010), pp. 35–41 (cit. on p. 75).

[Koval, 2017] Igor Koval, Jean-Baptiste Schiratti, Alexandre Routier, Michael Bacci, Olivier Colliot, Stéphanie Allassonnière, and Stanley Durrleman. "Statistical learning of spatiotemporal patterns from longitudinal manifold-valued networks". In: *Medical Image Computing and Computer Assisted Intervention*. Medical Image Computing and Computer Assisted Intervention. Quebec City, Canada, Sept. 2017 (cit. on pp. 7, 15, 36, 37).

[Koval, 2018] Igor Koval, Jean-Baptiste Schiratti, Alexandre Routier, Michael Bacci, Olivier Colliot, Stéphanie Allassonnière, and Stanley Durrleman. "Spatiotemporal Propagation of the Cortical Atrophy: Population and Individual Patterns". In: *Frontiers in Neurology* 9 (2018), p. 235 (cit. on pp. 7, 63, 92).

[Landau, 2010] S. M. Landau, D. Harvey, C. M. Madison, E. M. Reiman, N. L. Foster, P. S. Aisen, R. C. Petersen, L. M. Shaw, J. Q. Trojanowski, C. R. Jack, M. W. Weiner, and W. J. Jagust. "Comparing predictors of conversion and decline in mild cognitive impairment". In: *Neurology* 75.3 (July 2010), pp. 230–238 (cit. on p. 97).

[Landau, 2012] S. M. Landau, M. A. Mintun, A. D. Joshi, R. A. Koeppe, R. C. Petersen, P. S. Aisen, M. W. Weiner, and W. J. Jagust. "Amyloid deposition, hypometabolism, and longitudinal cognitive decline". In: *Ann Neurol* 72.4 (Oct. 2012), pp. 578–586 (cit. on p. 97).

[Li, 2019] D. Li, S. Iddi, W. K. Thompson, and M. C. Donohue. "Bayesian latent time joint mixed effect models for multicohort longitudinal data". In: *Stat Methods Med Res* 28.3 (Mar. 2019), pp. 835–845 (cit. on p. 63).

[Lo, 2011] R. Y. Lo, A. E. Hubbard, L. M. Shaw, et al. "Longitudinal change of biomarkers in cognitive decline". In: *Arch Neurol* 68.10 (Oct. 2011), pp. 1257–1266 (cit. on p. 5).

[Long, 2019] J. M. Long and D. M. Holtzman. "Alzheimer Disease: An Update on Pathobiology and Treatment Strategies". In: *Cell* 179.2 (Oct. 2019), pp. 312–339 (cit. on p. 1).

[Lopes Alves, 2020] Isadora Lopes Alves, Lyduine E. Collij, Daniele Altomare, Giovanni B. Frisoni, Laure Saint-Aubert, Pierre Payoux, Miia Kivipelto, Frank Jessen, Alexander Drzezga, Annebet Leeuwis, Alle Meije Wink, Pieter Jelle Visser, Bart N.M. van Berckel, Philip Scheltens, Katherine R. Gray, Robin Wolz, Andrew Stephens, Rossella Gismondi, Christopher Buckely, Juan Domingo Gispert, Mark Schmidt, Lisa Ford, Craig Ritchie, Gill Farrar, Frederik Barkhof, José Luis Molinuevo, and the AMYPAD Consortium. "Quantitative amyloid PET in Alzheimer's disease: the AMYPAD prognostic and natural history study". In: *Alzheimer's & Dementia* 16.5 (2020), pp. 750–758. eprint:

https://alz-journals.onlinelibrary.wiley.com/doi/pdf/10.1002/alz.12069 (cit. on p. 119).

[Lorenzi, 2015a] M. Lorenzi, X. Pennec, G. B. Frisoni, and N. Ayache. "Disentangling normal aging from Alzheimer's disease in structural magnetic resonance images". In: *Neurobiol. Aging* 36 Suppl 1 (Jan. 2015), pp. 42–52 (cit. on pp. 38, 80).

[Lorenzi, 2015b] M. Lorenzi, G. Ziegler, D. C. Alexander, and S. Ourselin. "Efficient Gaussian Process-Based Modelling and Prediction of Image Time Series". In: *Inf Process Med Imaging* 24 (2015), pp. 626–637 (cit. on p. 20).

[Lorenzi, 2017] Marco Lorenzi, Maurizio Filippone, Giovanni B. Frisoni, Daniel C. Alexander, and Sebastien Ourselin. "Probabilistic disease progression modeling to characterize diagnostic uncertainty: Application to staging and prediction in Alzheimer's disease". In: *NeuroImage* (2017) (cit. on pp. 6, 12, 15, 63, 73, 80, 86, 91, 117).

[Lorenzi, 2018] Marco Lorenzi and Maurizio Filippone. "Constraining the Dynamics of Deep Probabilistic Models". In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. Stockholmsmässan, Stockholm Sweden: PMLR, Oct. 2018, pp. 3233–3242 (cit. on pp. 6, 19, 21, 117).

[Mallat, 1989] S. G. Mallat. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 11.7 (July 1989), pp. 674–693 (cit. on p. 14).

[Marcus, 2010] D. S. Marcus, A. F. Fotenos, J. G. Csernansky, J. C. Morris, and R. L. Buckner. "Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults". In: *J Cogn Neurosci* 22.12 (Dec. 2010), pp. 2677–2684 (cit. on p. 119).

[Marinescu, 2019a] R. V. Marinescu, A. Eshaghi, M. Lorenzi, A. L. Young, N. P. Oxtoby, S. Garbarino, S. J. Crutch, and D. C. Alexander. "DIVE: A spatiotemporal progression model of brain pathology in neurodegenerative disorders". In: *NeuroImage* 192 (May 2019), pp. 166–177 (cit. on pp. 6, 7, 15, 36, 63, 91, 92).

[Marinescu, 2019b] Razvan Marinescu, Arman Eshaghi, Daniel Alexander, and Polina Golland. "BrainPainter: A software for the visualisation of brain structures, biomarkers and associated pathological processes". In: *arXiv preprint arXiv:1905.08627* (2019) (cit. on pp. 74, 104).

[Marquand, 2014] A. F. Marquand, M. Brammer, S. C. Williams, and O. M. Doyle. “Bayesian multi-task learning for decoding multi-subject neuroimaging data”. In: *NeuroImage* 92 (May 2014), pp. 298–311 (cit. on p. 20).

[McKhann, 1984] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. 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”. In: *Neurology* 34.7 (July 1984), pp. 939–944 (cit. on p. 2).

[Mendelson, 2017] Alex F. Mendelson, Maria A. Zuluaga, Marco Lorenzi, Brian F. Hutton, and Sébastien Ourselin. “Selection bias in the reported performances of AD classification pipelines”. In: *NeuroImage: Clinical* 14 (2017), pp. 400–416 (cit. on p. 91).

[Molchanov, 2017] Dmitry Molchanov, Arsenii Ashukha, and Dmitry Vetrov. “Variational Dropout Sparsifies Deep Neural Networks”. In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by Doina Precup and Yee Whye Teh. Vol. 70. Proceedings of Machine Learning Research. International Convention Centre, Sydney, Australia: PMLR, June 2017, pp. 2498–2507 (cit. on pp. 20, 22, 40).

[Montine, 2012] T. J. Montine, C. H. Phelps, T. G. Beach, E. H. Bigio, N. J. Cairns, D. W. Dickson, C. Duyckaerts, M. P. Frosch, E. Masliah, S. S. Mirra, P. T. Nelson, J. A. Schneider, D. R. Thal, J. Q. Trojanowski, H. V. Vinters, and B. T. Hyman. “National Institute on Aging-Alzheimer’s Association guidelines for the neuropathologic assessment of Alzheimer’s disease: a practical approach”. In: *Acta Neuropathol* 123.1 (Jan. 2012), pp. 1–11 (cit. on p. 2).

[Murphy, 2010] M. P. Murphy and H. LeVine. “Alzheimer’s disease and the amyloid-beta peptide”. In: *J. Alzheimers Dis.* 19.1 (2010), pp. 311–323 (cit. on p. 63).

[Murray, 2011] M. E. Murray, N. R. Graff-Radford, O. A. Ross, R. C. Petersen, R. Duara, and D. W. Dickson. “Neuropathologically defined subtypes of Alzheimer’s disease with distinct clinical characteristics: a retrospective study”. In: *Lancet Neurol* 10.9 (Sept. 2011), pp. 785–796 (cit. on p. 120).

[Navitsky, 2018] Michael Navitsky, Abhinay D. Joshi, Ian Kennedy, William E. Klunk, Christopher C. Rowe, Dean F. Wong, Michael J. Pontecorvo, Mark A. Mintun, and Michael D. Devous. “Standardization of amyloid quantitation with florbetapir standardized uptake value ratios to the Centiloid scale”. In: *Alzheimer’s & Dementia* 14.12 (2018), pp. 1565–1571 (cit. on p. 98).

[Nelson, 2011] P. T. Nelson, E. Head, F. A. Schmitt, P. R. Davis, J. H. Neltner, G. A. Jicha, E. L. Abner, C. D. Smith, L. J. Van Eldik, R. J. Kryscio, and S. W. Scheff. “Alzheimer’s disease is not "brain aging": neuropathological, genetic, and epidemiological human studies”. In: *Acta Neuropathol* 121.5 (May 2011), pp. 571–587 (cit. on p. 2).

[Nichols, 2019] Emma Nichols, Cassandra E. I. Szoeke, Stein Emil Vollset, et al. “Global, regional, and national burden of Alzheimer’s disease and other dementias, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016”. In: *The Lancet Neurology* 18.1 (Jan. 2019), pp. 88–106 (cit. on p. 1).

[Okamura, 2014] N. Okamura, R. Harada, S. Furumoto, H. Arai, K. Yanai, and Y. Kudo. “Tau PET imaging in Alzheimer’s disease”. In: *Curr Neurol Neurosci Rep* 14.11 (Nov. 2014), p. 500 (cit. on p. 3).

[Oxtoby, 2017] N. P. Oxtoby, S. Garbarino, N. C. Firth, J. D. Warren, J. M. Schott, and D. C. Alexander. “Data-Driven Sequence of Changes to Anatomical Brain Connectivity in Sporadic Alzheimer’s Disease”. In: *Front Neurol* 8 (2017), p. 580 (cit. on p. 63).

[Oxtoby, 2018] N. P. Oxtoby, A. L. Young, D. M. Cash, T. L. S. Benzinger, A. M. Fagan, J. C. Morris, R. J. Bateman, N. C. Fox, J. M. Schott, and D. C. Alexander. “Data-driven models of dominantly-inherited Alzheimer’s disease progression”. In: *Brain* 141.5 (May 2018), pp. 1529–1544 (cit. on p. 63).

[Oxtoby, 2021] N. P. Oxtoby, L. A. Leyland, L. M. Aksman, G. E. C. Thomas, E. L. Bunting, P. A. Wijeratne, A. L. Young, A. Zarkali, M. M. X. Tan, F. D. Bremner, P. A. Keane, H. R. Morris, A. E. Schrag, D. C. Alexander, and R. S. Weil. “Sequence of clinical and neurodegeneration events in Parkinson’s disease progression”. In: *Brain* (Feb. 2021) (cit. on p. 118).

[Paszke, 2019] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett. Curran Associates, Inc., 2019, pp. 8024–8035 (cit. on pp. 23, 70, 83).

[Pedregosa, 2011] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cit. on p. 23).

[Petrella, 2019] J. R. Petrella, W. Hao, A. Rao, and P. M. Doraiswamy. "Computational Causal Modeling of the Dynamic Biomarker Cascade in Alzheimer's Disease". In: *Comput Math Methods Med* 2019 (2019), p. 6216530 (cit. on p. 63).

[Poewe, 2017] Werner Poewe, Klaus Seppi, Caroline M. Tanner, Glenda M. Halliday, Patrik Brundin, Jens Volkmann, Anette-Eleonore Schrag, and Anthony E. Lang. "Parkinson disease". In: *Nature Reviews Disease Primers* 3.1 (Mar. 2017), p. 17013 (cit. on p. 118).

[Pontecorvo, 2019] M. J. Pontecorvo, M. D. Devous, I. Kennedy, M. Navitsky, M. Lu, N. Galante, S. Salloway, P. M. Doraiswamy, S. Southekal, A. K. Arora, A. McGeehan, N. C. Lim, H. Xiong, S. P. Truocchio, A. D. Joshi, S. Shcherbinin, B. Teske, A. S. Fleisher, and M. A. Mintun. "A multicentre longitudinal study of flortaucipir (18F) in normal ageing, mild cognitive impairment and Alzheimer's disease dementia". In: *Brain* 142.6 (June 2019), pp. 1723–1735 (cit. on pp. 79, 106).

[Prince, 2015] Martin James Prince, Anders Wimo, Maelenn Mari Guerchet, Gemma Claire Ali, Yu-Tzu Wu, and Matthew Prina. *World Alzheimer Report 2015 - The Global Impact of Dementia: An analysis of prevalence, incidence, cost and trends*. English. Alzheimer's Disease International, Aug. 2015 (cit. on p. 62).

[Rahimi, 2008] Ali Rahimi and Benjamin Recht. "Random Features for Large-Scale Kernel Machines". In: *Advances in Neural Information Processing Systems 20*. Ed. by J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis. Curran Associates, Inc., 2008, pp. 1177–1184 (cit. on p. 18).

[Reuter, 2012] M. Reuter, N. J. Schmansky, H. D. Rosas, and B. Fischl. "Within-subject template estimation for unbiased longitudinal image analysis". In: *NeuroImage* 61.4 (July 2012), pp. 1402–1418 (cit. on pp. 71, 95).

[Riihimäki, 2010] Jaakko Riihimäki and Aki Vehtari. "Gaussian processes with monotonicity information". In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Ed. by Yee Whye Teh and Mike Titterington. Vol. 9. Proceedings of Machine Learning Research. Chia Laguna Resort, Sardinia, Italy: PMLR, 13–15 May 2010, pp. 645–652 (cit. on p. 19).

[Risacher, 2017] S. L. Risacher, W. H. Anderson, A. Charil, P. F. Castelluccio, S. Shcherbinin, A. J. Saykin, and A. J. Schwarz. "Alzheimer disease brain atrophy subtypes are associated with cognition and rate of decline". In: *Neurology* 89.21 (Nov. 2017), pp. 2176–2186 (cit. on p. 91).

[Rodrigue, 2009] K. M. Rodrigue, K. M. Kennedy, and D. C. Park. "Beta-amyloid deposition and the aging brain". In: *Neuropsychol Rev* 19.4 (Dec. 2009), pp. 436–450 (cit. on p. 31).

[Ross, 2014] C. A. Ross, E. H. Aylward, E. J. Wild, D. R. Langbehn, J. D. Long, J. H. Warner, R. I. Scahill, B. R. Leavitt, J. C. Stout, J. S. Paulsen, R. Reilmann, P. G. Unschuld, A. Wexler, R. L. Margolis, and S. J. Tabrizi. "Huntington disease: natural history, biomarkers and prospects for therapeutics". In: *Nat Rev Neurol* 10.4 (Apr. 2014), pp. 204–216 (cit. on p. 118).

[Rowe, 2010] C. C. Rowe, K. A. Ellis, M. Rimajova, P. Bourgeat, K. E. Pike, G. Jones, J. Fripp, H. Tochon-Danguy, L. Morandeau, G. O'Keefe, R. Price, P. Raniga, P. Robins, O. Acosta, N. Lenzo, C. Szoeke, O. Salvado, R. Head, R. Martins, C. L. Masters, D. Ames, and V. L. Villemagne. "Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging". In: *Neurobiol. Aging* 31.8 (Aug. 2010), pp. 1275–1283 (cit. on p. 78).

[Saatçi, 2011] Yunus Saatçi. *Scalable Inference for Structured Gaussian Process Models*. 2011 (cit. on p. 20).

[Safieh, 2019] M. Safieh, A. D. Korczyn, and D. M. Michaelson. "ApoE4: an emerging therapeutic target for Alzheimer's disease". In: *BMC Medicine* 17 (2019) (cit. on p. 79).

[Schiratti, 2015] Jean-Baptiste Schiratti, Stéphanie Allassonnière, Olivier Colliot, and Stanley Durrleman. "Learning spatiotemporal trajectories from manifold-valued longitudinal data". In: *NIPS*. 2015, pp. 2404–2412 (cit. on pp. 6, 13, 37, 63).

[Schuff, 2009] N. Schuff, N. Woerner, L. Boreta, T. Kornfield, L. M. Shaw, J. Q. Trojanowski, P. M. Thompson, C. R. Jack, and M. W. Weiner. "MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers". In: *Brain* 132.Pt 4 (Apr. 2009), pp. 1067–1077 (cit. on p. 65).

[Schwarz, 2019] Adam J. Schwarz, Karen L. Sundell, Arnaud Charil, Michael G. Case, Ralf K. Jaeger, David Scott, Luc Bracoud, Joonmi Oh, Joyce Suhy, Michael J. Pontecorvo, Bradford C. Dickerson, and Eric R. Siemers. "Magnetic resonance imaging measures of brain atrophy from the EXPEDITION3 trial in mild Alzheimer's disease". In: *Alzheimer's & Dementia: Translational Research & Clinical Interventions* 5.1 (2019), pp. 328–337. eprint: https://alz-journals.onlinelibrary.wiley.com/doi/pdf/10.1016/j.trci.2019.05.007 (cit. on pp. 63, 91).

[Scott, 2007] K. R. Scott and A. M. Barrett. "Dementia syndromes: evaluation and treatment". In: *Expert Rev Neurother* 7.4 (Apr. 2007), pp. 407–422 (cit. on p. 1).

[Shaw, 2009] L. M. Shaw, H. Vanderstichele, M. Knapik-Czajka, C. M. Clark, P. S. Aisen, R. C. Petersen, K. Blennow, H. Soares, A. Simon, P. Lewczuk, R. Dean, E. Siemers, W. Potter, V. M. Lee, and J. Q. Trojanowski. "Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects". In: *Ann. Neurol.* 65.4 (Apr. 2009), pp. 403–413 (cit. on p. 65).

[Sivera, 2019] Raphaël Sivera, Hervé Delingette, Marco Lorenzi, Xavier Pennec, and Nicholas Ayache. "A model of brain morphological changes related to aging and Alzheimer's disease from cross-sectional assessments". In: *NeuroImage* (2019) (cit. on p. 38).

[Sivera, 2020] Raphaël Sivera, Nicolas Capet, Valeria Manera, Roxane Fabre, Marco Lorenzi, Hervé Delingette, Xavier Pennec, Nicholas Ayache, and Philippe Robert. "Voxel-based assessments of treatment effects on longitudinal brain changes in the Multidomain Alzheimer Preventive Trial cohort". In: *Neurobiology of Aging* 94 (2020), pp. 50–59 (cit. on p. 80).

[Sperling, 2011] R. A. Sperling, C. R. Jack, and P. S. Aisen. "Testing the right target and right drug at the right stage". In: *Sci Transl Med* 3.111 (Nov. 2011), p. 111cm33 (cit. on pp. 4, 78, 91).

[Terry, 1991] R. D. Terry, E. Masliah, D. P. Salmon, N. Butters, R. DeTeresa, R. Hill, L. A. Hansen, and R. Katzman. "Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment". In: *Ann Neurol* 30.4 (Oct. 1991), pp. 572–580 (cit. on p. 2).

[Venkatraghavan, 2019] Vikram Venkatraghavan, Esther E. Bron, Wiro J. Niessen, and Stefan Klein. “Disease progression timeline estimation for Alzheimer’s disease using discriminative event based modeling”. In: *NeuroImage* 186 (2019), pp. 518–532 (cit. on pp. 6, 92).

[Villemagne, 2013] V. L. Villemagne, S. Burnham, P. Bourgeat, B. Brown, K. A. Ellis, O. Salvado, C. Szoeke, S. L. Macaulay, R. Martins, P. Maruff, D. Ames, C. C. Rowe, and C. L. Masters. “Amyloid Î² deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer’s disease: a prospective cohort study”. In: *Lancet Neurol* 12.4 (Apr. 2013), pp. 357–367 (cit. on pp. 5, 63).

[Wang, 2004] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. “Image Quality Assessment: From Error Visibility to Structural Similarity”. In: *IEEE TRANSACTIONS ON IMAGE PROCESSING* 13.4 (2004), pp. 600–612 (cit. on p. 24).

[Wessels, 2019] A. M. Wessels, P. N. Tariot, J. A. Zimmer, K. J. Selzler, S. M. Bragg, S. W. Andersen, J. Landry, J. H. Krull, A. M. Downing, B. A. Willis, S. Shcherbinin, J. Mullen, P. Barker, J. Schumi, C. Shering, B. R. Matthews, R. A. Stern, B. Vellas, S. Cohen, E. MacSweeney, M. Boada, and J. R. Sims. “Efficacy and Safety of Lanabecestat for Treatment of Early and Mild Alzheimer Disease: The AMARANTH and DAYBREAK-ALZ Randomized Clinical Trials”. In: *JAMA Neurol* (Nov. 2019) (cit. on pp. 63, 75).

[Westwood, 2016] S. Westwood, E. Leoni, A. Hye, S. Lynham, M. R. Khondoker, N. J. Ashton, S. J. Kiddle, A. L. Baird, R. Sainz-Fuertes, R. Leung, J. Graf, C. T. Hehir, D. Baker, C. Cereda, C. Bazenet, M. Ward, M. Thambisetty, and S. Lovestone. “Blood-Based Biomarker Candidates of Cerebral Amyloid Using PiB PET in Non-Demented Elderly”. In: *J. Alzheimers Dis.* 52.2 (Mar. 2016), pp. 561–572 (cit. on p. 78).

[Whitwell, 2010] J. L. Whitwell. “Progression of atrophy in Alzheimer’s disease and related disorders”. In: *Neurotox Res* 18.3-4 (Nov. 2010), pp. 339–346 (cit. on pp. 7, 13).

[Whitwell, 2012] J. L. Whitwell, D. W. Dickson, M. E. Murray, S. D. Weigand, N. Tosakulwong, M. L. Senjem, D. S. Knopman, B. F. Boeve, J. E. Parisi, R. C. Petersen, C. R. Jack, and K. A. Josephs. “Neuroimaging correlates of pathologically defined subtypes of Alzheimer’s disease: a case-control study”. In: *Lancet Neurol* 11.10 (Oct. 2012), pp. 868–877 (cit. on p. 120).

[Wijeratne, 2018] Peter A. Wijeratne, Alexandra L. Young, Neil P. Oxtoby, Razvan V. Marinescu, Nicholas C. Firth, Eileanoir B. Johnson, Amrita Mohan, Cristina Sampaio, Rachael I. Scahill, Sarah J. Tabrizi, and Daniel C. Alexander. "An image-based model of brain volume biomarker changes in Huntington's disease". In: *Annals of Clinical and Translational Neurology* 5.5 (2018), pp. 570–582. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/acn3.558` (cit. on p. 118).

[Young, 2014] A. L. Young, N. P. Oxtoby, P. Daga, D. M. Cash, N. C. Fox, S. Ourselin, J. M. Schott, and D. C. Alexander. "A data-driven model of biomarker changes in sporadic Alzheimer's disease". In: *Brain* 137.Pt 9 (Sept. 2014), pp. 2564–2577 (cit. on pp. 5, 12, 63).

[Young, 2015] A. L. Young, N. P. Oxtoby, J. Huang, R. V. Marinescu, P. Daga, D. M. Cash, N. C. Fox, S. Ourselin, J. M. Schott, and D. C. Alexander. "Multiple Orderings of Events in Disease Progression". In: *Inf Process Med Imaging* 24 (2015), pp. 711–722 (cit. on p. 15).

[Young, 2018] A. L. Young, R. V. Marinescu, N. P. Oxtoby, M. Bocchetta, K. Yong, N. C. Firth, D. M. Cash, D. L. Thomas, K. M. Dick, J. Cardoso, J. van Swieten, B. Borroni, D. Galimberti, M. Masellis, M. C. Tartaglia, J. B. Rowe, C. Graff, F. Tagliavini, G. B. Frisoni, R. Laforce, E. Finger, A. de Mendonca, S. Sorbi, J. D. Warren, S. Crutch, N. C. Fox, S. Ourselin, J. M. Schott, J. D. Rohrer, and D. C. Alexander. "Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference". In: *Nat Commun* 9.1 (Oct. 2018), p. 4273 (cit. on p. 38).

[Zetterberg, 2019] Henrik Zetterberg and Samantha C. Burnham. "Blood-based molecular biomarkers for Alzheimer's disease". eng. In: *Molecular brain* 12.1 (Mar. 2019), pp. 26–26 (cit. on p. 78).