Multimodal modelling of Alzheimer's Disease progression

Benoît Sauty

Sorbonne Université

Doctoral Thesis

Multimodal modelling of Alzheimer's Disease progression

Benoît Sauty

Supervisor:	Stanley Durrleman	Senior researcher at *INRIA Paris*
Examiners:	Isabelle Bloch	Professor at *Telecom ParisTech* (President)
	Emma Robinson	Senior Lecturer at *King's College London*
Reviewers:	Marco Lorenzi	Research Associate at *INRIA Sofia*
	Bruno Jedynak	Professor at *Portland State University*

A thesis submitted in fulfillment of the requirements
for the degree of PhD

in the

ARAMIS Lab
Institut du Cerveau - Paris Brain Institute (ICM), Inserm U 1127, CNRS UMR 7225, AP-HP
Hôpital de la Pitié Salpêtrière and Inria

September 13, 2023

Abstract

Alzheimer's disease (AD) is a multi-facet pathology, that can be monitored through a variety of data types. This thesis aims to leverage multimodal longitudinal data, especially imaging scans and cognitive tests, to provide a statistical description of the progression of AD and to enable individual forecasting of future decline. Mixed-effect disease progression models (DPMs) are commonly used for these tasks. In this context, our first contribution questions the frequent assumption that biomarkers follow linear or logistic functions over time, and we propose a geometric framework that assumes the data lie on a manifold and follow geodesics over time. We learn the Riemannian metric of the observation space and are able to model a wider variety of biomarkers, without priors on the shape of the trajectory over time. Using variational auto-encoders, we then extend this framework to neuroimaging data (MRI or PET scans), in order to provide high-dimensional progression models that describe the patterns of structural and functional alterations of the brain over the course of AD. We then apply this family of DPMs to clinical studies data in order to investigate the heterogeneity of AD progression, due to APOE-ε4 genotype and sex on patterns of brain alterations. Lastly, we use said DPMs with a set of imaging and fluid biomarkers to identify the specific combinations of input features that best forecast cognitive declines in patients at different stages of the disease. The thesis demonstrates that DPMs can effectively model the progression of AD using a great variety of multimodal longitudinal data and provide valuable insights into the disease's clinical manifestations and progression. These findings can inform clinical trial design and facilitate more accurate prognosis and individualized treatment strategies for patients with AD.

Abstract en Français

La maladie d'Alzheimer (MA) est une pathologie multi-facette qui peut être surveillée grâce à une grande variété de modalités de données. Cette thèse vise à exploiter des données longitudinales multimodales, principalement des données d'imagerie et des tests cognitifs, pour fournir une description statistique de la progression de la MA et permettre une prévision individuelle de la dégradation future. Les modèles de progression à effet-mixtes de la maladie (DPMs) sont couramment utilisés pour ces tâches. Dans ce contexte, notre première contribution remet en question l'hypothèse fréquente selon laquelle les biomarqueurs suivent des fonctions linéaires ou logistiques au fil du temps, et nous proposons un cadre géométrique qui suppose que les données se trouvent sur une variété et suivent des géodésiques au fil du temps. Nous apprenons la métrique riemannienne de l'espace d'observation et sommes capables de modéliser une plus grande variété de biomarqueurs, sans hypothèses préalables sur la forme de la trajectoire au fil du temps. En utilisant des auto-encodeurs variationnels, nous étendons ensuite ce cadre aux données de neuroimagerie (IRM ou TEP), afin de fournir des modèles de progression en grande dimension qui décrivent les motifs d'altérations structurelles et fonctionnelles du cerveau au cours de la MA. Nous appliquons ensuite cette famille de DPMs à des données réelles afin d'étudier l'hétérogénéité de la progression de la MA, en décrivant l'influence du génotype APOE-ε4 et du sexe sur les motifs d'altérations cérébrales. Enfin, nous utilisons ces DPMs avec un ensemble de biomarqueurs d'imagerie et extrait du fluide cérébrospinal pour identifier les combinaisons spécifiques de *features* qui permettent de prévoir les déclins cognitifs chez les patients à différents stades de la maladie. La thèse démontre que les DPMs peuvent modéliser efficacement la progression de la MA en utilisant une grande variété de données longitudinales multimodales et fournir des informations précieuses sur les manifestations cliniques et la progression de la maladie. Ces résultats peuvent informer la conception d'essais cliniques et faciliter des stratégies de traitement individualisées et plus précises pour les patients atteints de la MA.

Acknowledgements

I would like to express my sincere gratitude to my thesis supervisor, Stanley, for his invaluable guidance, support, and encouragement throughout my research. His expertise and dedication have been instrumental in shaping this work, and I am deeply grateful for his mentorship.

I would like to thank the reviewers, Bruno Jedynak and Marco Lorenzi, who provided me with valuable feedback and comments on the manuscript. Their insights showcase the thorough attention they put into this work, and I am very grateful that they helped improve the quality of this thesis. I would also like to thank Emma Robinson and Isabelle Bloch, whom I was glad to count as jury members.

I would like to extend my thanks to my colleagues from the ARAMIS lab, for all the friendly discussions and support. More specifically, I would like to acknowledge the immense help that Ninon provided me, introducing me to the MICCAI community and being very generous with her time. I would like to also extend my thanks to the old-timers of the lab who also briefed me on the ins and outs of neuroimaging and on how to handle Sorbonne's tedious processes: Alexandre, Igor, Thomas, Tiziana, Simona, Raphaël, Juliana, Elina and Etienne, and all the members of the "Longitudinal team", whom I spent conferences and countless meetings with! Naming everyone in the lab would be a risky exercise, but I would just like to say that all members contributed to making the ICM a great place to work at, and I am very grateful to have had the opportunity to share these years with you.

Last but not least, I would like to commend my family and close friends for being extremely supportive of my various endeavours, however disjointed they may seem. Your cheerful support means a lot to me.

Contents

Résumé en Français

Motivation

La maladie d'Alzheimer (MA) est une maladie complexe et invalidante qui touche des millions de personnes dans le monde. Malgré des efforts de recherche importants, aucun traitement curatif pour la MA n'a encore été découvert, rendant la détection précoce et le suivi de la maladie cruciaux pour améliorer le suivi des patients et informer les essais cliniques.

Les avancées récentes des technologies médicales, y compris la disponibilité de multiples modalités d'imagerie et d'une grande variété de données cliniques, offrent des opportunités sans précédent pour étudier la progression de la maladie et identifier des marqueurs biologiques pour la détection et l'intervention précoces. La maladie progresse selon une cascade d'événements, chacun étant mieux observé à travers certaines modalités de données, sans qu'aucune modalité ne soit informative de l'ensemble de la progression. Tirer parti de l'abondance de données disponibles, de manière multimodale, est crucial pour décrire toutes les facettes de la maladie. En outre, le suivi des patients au fil du temps est particulièrement utile dans le contexte des maladies dégénératives, car la dynamique de la progression est potentiellement plus importante que l'état d'un patient à un moment donné. Dans ce contexte, fournir des modèles de progression multimodaux basés sur les données de la maladie peut aider à décrire les processus pathologiques, même avant l'apparition des symptômes, et à prévoir la progression des patients en se basant sur leurs visites précédentes. Une description individuelle de la progression ouvre la voie pour la médecine de précision et peut informer la conception des essais cliniques, à la fois en identifiant les cibles thérapeutiques prometteuses, et en enrichissant les essais avec des sujets correspondant au stade de la maladie ciblé et au profil de progression.

Ce chapitre présente, en français, une brève description des principales notions sur lesquelles cette thèse se construit : notre connaissance actuelle de la maladie d'Alzheimer, les types de données médicales disponibles qui peuvent être utiles ainsi que la principale base de données publiques, et les modèles statistiques utilisés pour la modélisation de la progression de la maladie, que nous utilisons et améliorons.

Maladie d'Alzheimer

La présentation clinique de la maladie d'Alzheimer se caractérise par un déclin progressif des fonctions cognitives, notamment la mémoire, le langage, les fonctions exécutives et les capacités visuospatiales. Les premiers stades de la maladie se manifestent souvent par des oublis, tels que l'incapacité de se souvenir d'événements récents ou de conversations, la perte d'objets et des difficultés à suivre les instructions. À mesure que la maladie progresse, les patients peuvent éprouver des difficultés à résoudre des problèmes, à prendre des décisions et à planifier. Ils peuvent également avoir du mal à accomplir des tâches qui impliquent plusieurs étapes, comme préparer un repas ou s'habiller. Aux stades avancés, ils peuvent être incapables d'effectuer des activités de base de la vie quotidienne, comme se laver ou se nourrir, et peuvent ne plus reconnaître leurs proches.

En plus du déclin cognitif, les personnes atteintes de la maladie d'Alzheimer peuvent également présenter des symptômes comportementaux et psychologiques, tels que la dépression, l'anxiété, l'agitation et les hallucinations. Ces symptômes sont souvent éprouvants

tant pour les patients que pour leurs aidants et peuvent avoir un impact significatif sur leur qualité de vie. À mesure que la maladie progresse, les personnes atteintes peuvent rencontrer des difficultés d'interaction et de communication et se désengager des activités sociales.

La pathogenèse de la maladie d'Alzheimer implique une interaction complexe entre divers processus pathologiques, y compris l'accumulation de plaques de protéines appelées "bêta-amyloïde" (Aβ), de filaments de protéines appelées "tau", d'inflammation chronique et de mort neuronale qui conduisent à l'atrophie cérébrale. L'ordre chronologique de ces anomalies n'est pas entièrement clair, car elles interagissent probablement et s'influencent mutuellement.

Il est communément accepté que l'accumulation de plaques de bêta-amyloïde est un événement précoce dans la pathogenèse de la maladie d'Alzheimer. La bêta-amyloïde est une protéine qui forme des agrégats insolubles toxiques pour les neurones et perturbe la fonction neuronale normale. L'accumulation de bêta-amyloïde est susceptible de déclencher une cascade d'événements qui conduisent à la formation de filaments neurofibrillaires, composés de fibres torsadées constituées d'une protéine appelée tau. Tau perturbe également la fonction neuronale normale, entraînant la dégénérescence et la mort des neurones. L'inflammation chronique et le stress oxydatif sont également susceptibles de jouer un rôle dans la pathogenèse de la maladie d'Alzheimer, car ils endommagent les neurones et favorisent la formation de plaques de bêta-amyloïde et de filaments neurofibrillaires, exacerbant les processus dégénératifs. Ces processus sont également responsables d'une privation cholinergique, qui désigne la perte de neurones cholinergiques et la diminution résultante du neurotransmetteur acétylcholine (ACh) dans le cerveau. L'ACh joue un rôle crucial dans la fonction cognitive, notamment l'apprentissage, la mémoire, l'attention et l'éveil, et sa diminution est susceptible de contribuer aux symptômes cognitifs et comportementaux de la maladie d'Alzheimer.

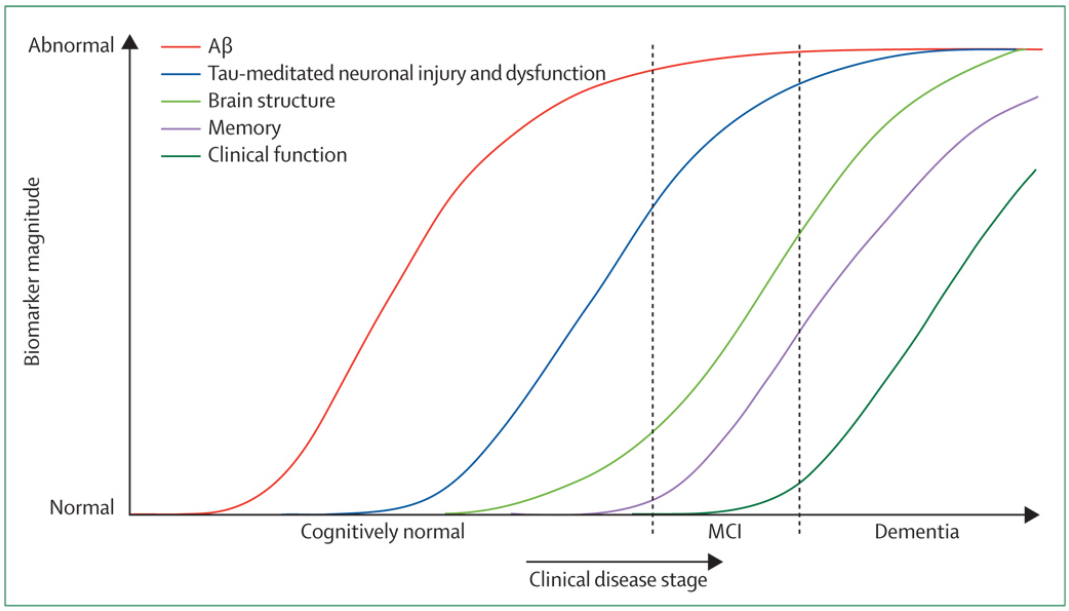


Figure 1: Représentation schématique de la cascade de déclin des marqueurs biologiques au cours de la maladie d'Alzheimer, selon [Jack et al., 2010]. La longue période entre l'accumulation des protéines Aβ et tau dans le cerveau, l'atrophie du cerveau, et le déclin cognitif correspondant, suggèrent l'importance d'utiliser des données multimodales afin de mieux prédire la progression des patients et proposer un suivi adapté.

Avant que le patient ne soit atteint de déficience cognitive, il existe une longue phase appelée "stade prodromal", caractérisée par l'accumulation silencieuse d'altérations structurelles et fonctionnelles du cerveau qui n'auront de conséquences que beaucoup plus tard dans la vie du patient. Par exemple, la concentration d'Aβ franchit souvent un seuil d'anomalie jusqu'à 20 ans avant l'apparition des symptômes cognitifs et les concentrations de protéines tau jusqu'à 15 ans avant. Au stade prodromal, l'état cognitif des patients est généralement considéré comme soit normal cognitivement (CN) soit légèrement altéré cognitivement (MCI), avant que le diagnostic de la maladie d'Alzheimer ne puisse être validé. La figure 3 résume une description empirique de la cascade d'événements à travers le spectre de la maladie d'Alzheimer.

Plusieurs facteurs de risque sont impliqués dans la pathogenèse. L'âge est le facteur de risque le plus établi : l'incidence double tous les cinq ans après l'âge de 65 ans, avec une prévalence estimée de 1 sur 10 individus de plus de 65 ans et près de 1 sur 3 individus de plus de 85 ans. Des facteurs génétiques modifient également le risque de développer la maladie. Les mutations dans les gènes qui sont impliqués dans la production et le traitement de la bêta-amyloïde déclenchent les formes familiales précoces. En outre, des variants dans le gène de l'apolipoprotéine E (APOE), qui joue un rôle dans le métabolisme lipidique et transporte des protéines dans le cerveau, sont associés à un risque accru de développer la maladie. La figure 2 répertorie les facteurs de risque modifiables (alimentation, mode de vie, facteurs environnementaux, etc) identifiés.

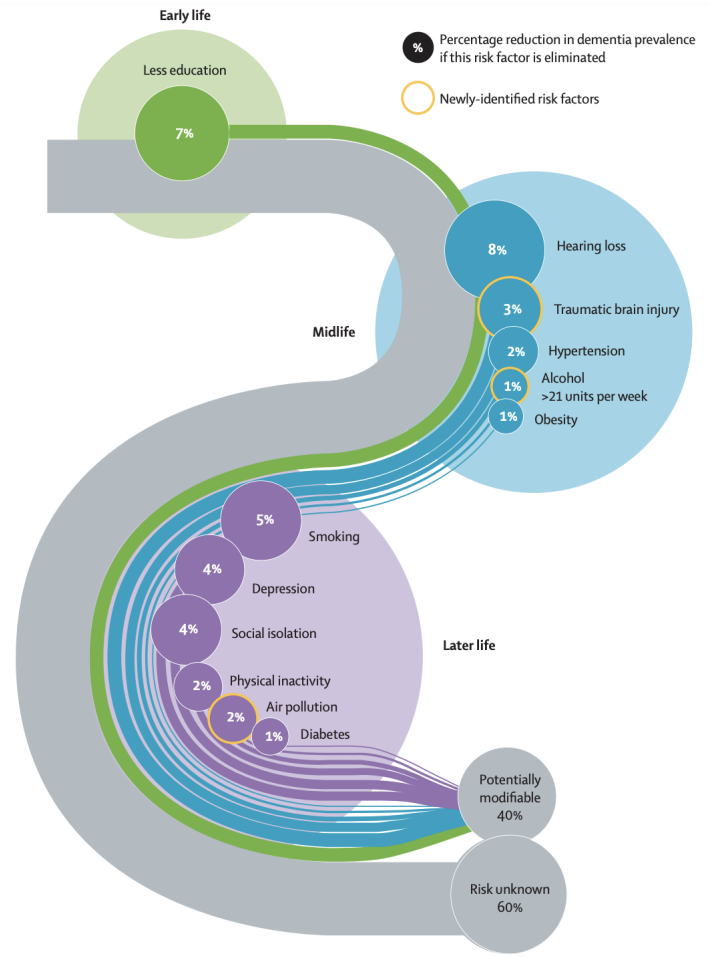


Figure 2: Liste des facteurs de risques potentiellement modifiables identifiés par [Livingston et al., 2020]. Les importances relatives de chaque facteur de risque varient entre ethnies et sont présentées dans l'article original.

Bien qu'il n'existe actuellement aucun remède pour la maladie d'Alzheimer, un certain nombre de stratégies de traitement et de pratiques de soins ont été développées pour gérer les symptômes et améliorer la qualité de vie des personnes affectées. Ces stratégies de traitement comprennent des interventions pharmacologiques, des thérapies cognitives et comportementales, des programmes d'exercice et des interventions nutritionnelles. Les interventions pharmacologiques peuvent être symptomatiques ou être un traitement modificateur de la maladie (DMT). Les premières ne traitent que les symptômes tels que la cognition ou la dépression, mais n'ont pas d'effet durable sur la progression de la maladie, ce qui signifie que les effets disparaissent lorsque le traitement s'arrête et que les processus sous-jacents tels que la neurodégénérescence et l'accumulation de protéines ne sont pas affectés. Les secondes, en revanche, agissent sur l'un des mécanismes de la maladie et ont un impact durable sur l'un des biomarqueurs, même après la fin du traitement. Les traitements symptomatiques ont montré une efficacité modérée et ont même été rendus inéligibles au remboursement par la sécurité sociale française depuis 2018. De plus, la recherche de médicaments s'est surtout concentrée sur les DMT au cours de la dernière décennie. En 2022 et 2023, deux premiers médicaments qui "nettoient" l'accumulation d'Aβ dans le cerveau ont été approuvés par la FDA, l'agence américaine du médicament, mais pas en Europe en raison de l'efficacité non démontrée sur le ralentissement du déclin cognitif.

Données médicales

Les "données médicales" regroupent des examens cliniques, des mesures de biomarqueurs, des tests génomiques et des scans d'imagerie. Elles sont devenues une ressource de plus en plus importante pour améliorer le suivi des patients, faire avancer la recherche médicale et informer les politiques de santé publique. Dans cette section, nous examinerons les modalités de données les plus utiles pour la surveillance et la modélisation de la progression de la MA.

Les **examens cliniques** incluent des tests cognitifs, des examens neurologiques et psychologiques, ainsi que les évaluations génétiques. Les tests cognitifs peuvent être administrés par des professionnels de santé formés, et permettent d'identifier les signes précoces de déclin cognitif, de suivre la progression de la maladie et évaluer l'efficacité des interventions telles que les médicaments ou les programmes de rééducation cognitive. Ils peuvent cibler des domaines de compétence spécifiques et sont utiles pour distinguer la MA des autres formes de démence. Les examens neurologiques et psychologiques permettent de détecter les changements d'humeur, les dépressions ou toute forme de trouble dans les activités du quotidien. Enfin, les évaluations génétiques sont principalement utiles dans le domaine de la recherche afin d'étudier l'effet des facteurs de risques génétiques.

Les **biomarqueurs** sont des mesures quantifiables de l'état d'un organe ou d'une fonction du corps humain. Les ponctions lombaires, qui prélèvent un échantillon de liquide céphalorachidien (CSF), mesurent les quantités de protéines bêta-amyloïdes et tau solubles qui circulent dans l'environnement du cerveau, et permettent d'en déduire la quantité de protéines pathologiques dans le cerveau. Bien que les biomarqueurs de CSF permettent le diagnostic de la MA, de tels tests sont généralement réservés à des fins de recherche ou à des environnements cliniques spécialisés, car ils sont désagréables pour les patients, nécessitent un équipement coûteux et une expertise spécifique. Des équivalents de ces biomarqueurs peuvent également être mesurés dans le sang ou le plasma, et sont également étudiés pour leur potentiel à aider au diagnostic tout en étant moins invasifs et plus faciles à obtenir que les échantillons de CSF. Les biomarqueurs sont notamment utilisés dans le cadre d'essais cliniques pour évaluer la progression et la réponse au traitement.

Les **données d'imagerie** sont également très importantes pour suivre la progression de la MA, car elles permettent de visualiser les changements qui se produisent dans le cerveau au fil du temps. L'imagerie par résonance magnétique (IRM) est une technique d'imagerie médicale qui permet de visualiser les structures anatomiques internes du cerveau avec une grande précision. Différents types d'IRM permettent de visualiser différentes structures, et de mettre en évidence l'atrophie cérébrale, l'amincissement cortical, la présence de micro-hémorragies ou encore de dépôts de fer dans le cerveau au cours de la maladie. Une autre méthode d'imagerie fréquemment utilisée est la tomographie par émission de positons (TEP). Cette technique utilise des traceurs radioactifs pour visualiser les changements biochimiques dans le cerveau, tels que la distribution des protéines bêta-amyloïde et tau, ou le métabolisme du cerveau. Ces modalités sont très utiles pour visualiser les premières phases de la maladie, mais nécessitent beaucoup de logistique pour la production, le transport et le stockage des traceurs radioactifs, qui limitent l'application à grande échelle de la TEP.

Les modalités d'imagerie mentionnées ici sont les plus fréquemment utilisées en pratique hospitalière et dans un contexte de recherche, et permettent aux praticiens et chercheurs de disposer d'images en 3D des processus et structures internes au cerveau. Il est cependant courant d'utiliser des méthodes "d'extraction de paramètres", qui transforment ces images de grande dimension en quelques biomarqueurs d'intérêts : par exemple le volume de l'hippocampe, la quantité totale de matière grise, le métabolisme moyen du cerveau ou l'épaisseur du cortex frontal. Des paramètres régionaux ou globaux peuvent révéler des dynamiques d'altérations spécifiques à la maladie et très informatives sur la progression individuelle du patient.

Une *pipeline* de traitement d'image est essentielle pour réaliser des analyses quantitatives. Les images acquises sont généralement hétérogènes, car elles dépendent du fabricant et de la génération du matériel, du protocole d'acquisition et de la procédure de reconstruction d'image. Mettre en place des étapes reproductibles pour chaque expérience est fondamental pour la recherche en neuroimagerie. Ces étapes comprennent généralement les corrections de biais de champ, la normalisation d'intensité, le détourage du crâne, le recalage et la segmentation d'image. Chacune de ces étapes peut recourir à une gamme d'algorithmes disponibles. Pour assurer la reproductibilité, le logiciel open-source "Clinica" est utilisé dans cette thèse pour combiner tous les outils de traitement usuels dans des *pipelines*. La combinaison exacte des étapes dépend de la tâche à accomplir et sera toujours présentée dans la section correspondante.

Des bases de données de recherche ont été assemblées afin de permettre l'étude quantitative de la MA. Parmi celles-ci, certaines permettent de suivre des patients au cours du temps en obtenant des données multimodales. Dans cette catégorie, L'*Alzheimer's Disease Neuroimaging Initiative* (ADNI) est une collaboration internationale publique-privée qui vise à développer des biomarqueurs pour la maladie d'Alzheimer (MA) en utilisant des méthodes d'imagerie médicale. Les participants à ADNI sont des individus atteints de MA, des personnes atteintes d'un trouble cognitif léger et des personnes âgées en bonne santé. Les participants subissent des examens d'imagerie cérébrale (IRM et TEP principalement), des mesures de biomarqueurs de CSF, un test génétique, ainsi que des tests neuropsychologiques réguliers pour suivre la progression de la maladie. Les données recueillies sont partagées publiquement et permettent aux chercheurs de développer des modèles de la progression de la maladie en utilisant une approche multimodale, combinant plusieurs types d'imagerie et de mesures cliniques pour une meilleure compréhension de la MA.

Modèles de progression de maladie

Le "*Disease progression modeling*", ou modélisation de progression de maladies, est un domaine apparu au début des années 2010 pour répondre au besoin de descriptions quantitatives des maladies dégénératives. De telles descriptions permettent de mieux comprendre la maladie, d'exhiber les biomarqueurs qui évoluent à chaque stade de la maladie, d'identifier des cibles thérapeutiques potentielles et de fournir un prognostique plus précis au patient et sa famille. L'inspiration initiale pour ce domaine est venue des modèles hypothétiques de la progression de la maladie d'Alzheimer, tels que proposés par [Jack et al., 2010] (voir illustration Figure 1). Ces modèles empiriques mettent en évidence la longue phase prodromale, ainsi que l'aspect multimodal de la progression de la maladie. La quantité de données disponibles est à la fois trop élevée pour une évaluation humaine qualitative et trop faible pour des approches d'apprentissage automatique entièrement basées sur les données. De plus, l'hétérogénéité des types de données complique l'intégration par des méthodes statistiques traditionnelles. Les *Disease Progression Models* (DPM), ou modèles de progression de maladies, cherchent à trouver un équilibre entre une connaissance *a priori* des caractéristiques de la maladie et des modèles appris à partir des données.

Les modèles à effets mixtes sont largement utilisés pour la modélisation longitudinale de données, lorsque des observations répétées sont effectuées sur des individus. Ils permettent de prendre en compte à la fois la variation intra-individuelle (pour un individu au fil du temps) et la variation inter-individuelle (entre différents individus). Les modèles à effets mixtes utilisent des effets aléatoires pour capturer la variabilité entre les individus, tandis que les effets fixes représentent les relations systématiques entre les variables explicatives (e.g. l'age, le type de scanner, le sexe) et la variable réponse (e.g. le volume du cerveau, la quantité d'Aβ, le score MMSE). Cette approche permet de modéliser les trajectoires individuelles au cours du temps, tout en tenant compte des différences entre les individus. Dans le contexte des données médicales, les effets fixes permettent de modéliser la progression moyenne des biomarqueurs au cours du temps, tandis que les effets aléatoires permettent de décrire la progression de chaque individu par rapport à cette moyenne. Par exemple, un individu peut décliner plus rapidement ou lentement, ou commencer à décliner plus tôt ou plus tard que la moyenne, ou tout simplement avoir un niveau de base différent de la moyenne pour des raisons physiologiques.

En pratique, nous utilisons et améliorons un modèle de progression appelé "*Disease course mapping*", ou "cartographie du cours de maladie" en français, un modèle à effets mixtes non linéaire, qui modélise la progression de biomarqueurs au cours du temps comme suivant des trajectoires logistiques au cours du temps (comme dans la figure 1). Comme mentionné précédemment, la variabilité inter-individuelle consiste en une combinaison de l'âge de début de déclin, de la vitesse de déclin, et de la variabilité physiologique et anatomique entre les patients, qui sont indépendants du temps. Cette introduction ne rentre pas dans les détails de modélisation, mais cette approche est rigoureusement décrite en Section 1.2.

Présentation des parties

L'objectif de cette thèse est d'utiliser des données longitudinales multimodales pour décrire la progression de la maladie d'Alzheimer et permettre une prévision individuelle du déclin futur. Plus particulièrement, l'objectif est de modéliser la progression des scores cognitifs et des données d'imagerie, afin de mettre en avant le lien entre les altérations du cerveau et le déclin cognitif. Les contributions peuvent être séparées en deux catégories : les contributions méthodologiques qui améliorent les cadres de modélisation de DPM (modèles

de progression de maladie) existants (parties I, II et IV), et les contributions cliniques qui utilisent les DPM pour illustrer un aspect cliniquement pertinent de la maladie (parties III et IV).

Divers cadres de modélisation de DPM ont été suggérés pour modéliser les biomarqueurs dans la maladie d'Alzheimer, et impliquent généralement certaines hypothèses sur le comportement attendu des biomarqueurs au cours du temps. Plus précisément, *Disease course mapping*, une approche couramment utilisée pour la modélisation de la maladie d'Alzheimer, suppose que les biomarqueurs suivent des courbes logistiques au fil du temps. Bien que cette hypothèse ait été largement acceptée en raison de l'héritage de modèles de progression hypothétiques, nous questionnons sa validité. *Disease course mapping* appartient à une famille de modèles géométriques dans lesquels l'espace d'observation est supposé être une variété riemannienne (un espace courbé), et la métrique de cette variété définit la forme des trajectoires de biomarqueurs. Jusqu'à présent, la métrique riemannienne a été choisie pour générer des courbes de progression logistiques. Notre contribution est de lever cette hypothèse en apprenant la métrique riemannienne de l'espace d'observation à partir des données réelles, ce qui permet de mieux apprendre les courbes de progression des biomarqueurs. Pour fournir une analogie, le déclin du patient au cours du temps peut être vu comme un skieur descendant une montagne. La courbure des vallées et des pans de descente de la montagne définissent la trajectoire et la vitesse du skieur. En modelant la montagne comme on le souhaite, on peut obtenir quasiment n'importe quelle trajectoire de skieur. Dans notre cas, nous faisons l'hypothèse qu'il existe une montagne qui a donné lieu aux trajectoires observées et nous décrivons la montagne, uniquement à partir de l'observation des trajectoires de différents skieurs suivis au cours du temps. La partie I présente les outils géométriques et décrit un cadre pour apprendre la métrique riemannienne de l'espace d'observation dans le contexte des DPM à effets mixtes. La comparaison de cette approche avec le *Disease course mapping* "standard" ne révèle pas de différences significatives de performance, ce qui suggère que l'hypothèse logistique est plausible.

Une autre contrainte des DPMs, est qu'ils sont restreints à la modélisation de la progression des biomarqueurs. Peu de modèles permettent d'exploiter les informations structurelles fournies par des images en grandes dimensions. Pourtant, les images révèlent des altérations fonctionnelles et structurelles du cerveau qui se produisent avant l'apparition des premiers symptômes, et peuvent révéler une interaction complexe entre les régions qui n'est pas prise en compte lors de l'extraction de paramètres régionaux. Pour pallier cette limitation, nous proposons une méthode qui combine un réseau de neurones convolutif, pour compresser les images en une représentation de petite dimension, avec un modèle à effets mixtes latents, qui décrit la progression des représentations en petites dimensions des images réelles. La partie II décrit notre modèle "*Longitudinal Variational Autoencoder*" ainsi que le protocole d'inférence proposé, et démontre qu'il retrouve des schémas d'altérations bien connus pour la maladie d'Alzheimer et le vieillissement sain en ce qui concerne l'atrophie cérébrale et l'hypométabolisme.

Travailler sur les modèles de progression pour les données d'imagerie a révélé de grandes différences dans les motifs d'altération entre sous-groupes de patients. Plus précisément, un déclin plus important chez les porteurs d'allèles APOE-ε4 par rapport aux non-porteurs et chez les femmes par rapport aux hommes. Cela a conduit à la troisième partie de ce travail, qui vise à décrire l'influence à la fois du génotype APOE-ε4 et du sexe féminin sur les modèles d'atrophie cérébrale, d'amincissement cortical et de perte de métabolisme au cours de la maladie d'Alzheimer (MA). Au cours du vieillissement normal, les hommes présentent une détérioration plus prononcée que les femmes, tant en ce qui concerne la structure cérébrale que les fonctions cognitives. Par opposition, au cours de la MA, les

femmes présentent un déclin cognitif plus prononcé que les hommes et une atrophie de l'hippocampe plus marquée. L'hétérogénéité de la MA, du point de vue de l'atrophie du cerveau et de la baisse de métabolisme, n'a pour l'instant jamais été analysé pour l'ensemble du cerveau en tenant compte à la fois du sexe et du génotype APOE afin de démêler l'influence de chaque facteur. La partie III propose donc une utilisation descriptive des DPMs pour modéliser ces caractéristiques de la MA et comparer les distributions des âges de début et des taux de déclin entre les sous-groupes, mettant en lumière l'importance de l'APOE-$\varepsilon 4$ et du sexe en tant que facteurs importants influençant la manifestation et la progression de la MA.

L'un des principaux objectifs derrière l'analyse des altérations cérébrales, via les données de neuroimagerie, est d'obtenir des informations sur les symptômes cognitifs qui peuvent être attendus. Pour atteindre l'objectif initial de cette thèse, nous avons utilisé des DPM pour prédire le déclin cognitif chez les patients à différents stades de la maladie en utilisant à la fois des données d'imagerie multimodales et des biomarqueurs fluides. Cependant, la disponibilité de nombreuses modalités de données ne signifie pas nécessairement qu'elles sont toutes aussi informatives sur la progression future d'un patient. Notre contribution est d'identifier, pour chaque stade de la maladie, les modalités et données d'*input* qui améliorent ou diminuent le plus la qualité des prédictions cognitives pour différents horizons de prédiction. Nous sélectionnons des cohortes pour cette étude, comme pour les essais cliniques, en fonction de la charge amyloïde et le niveau cognitif à la première visite. Notre travail peut aider les essais cliniques à déterminer quelles modalités de données sont les plus précieuses à acquérir et quels sujets sont les plus susceptibles de connaître une progression tout au long de l'essai, afin d'obtenir un essai statistiquement plus performant. La partie IV décrit l'approche et les résultats de ce processus de sélection de modalités.

Introduction

Alzheimer's disease (AD) is a complex and debilitating illness that affects millions of people worldwide. Despite significant research efforts, a cure or effective treatment for AD has yet to be discovered, making early detection and monitoring of the disease crucial for improving patient outcomes and informing clinical trials.

Recent advancements in medical technology, including the availability of multimodal imaging and a variety of clinical data, provide unprecedented opportunities to study the disease's progression and identify potential biomarkers for early detection and intervention. The disease is thought to unravel according to a cascade of events, each best monitored by certain modalities of data, with no one modality being informative of the entire progression. Leveraging the abundance of available data, in a multimodal manner, is crucial to the description of all facets of the disease. Besides, following patients over time is particularly useful in the context of degenerative diseases, as the dynamics of the progression is as important as the state of a patient at a given time. In this context, providing data-driven multimodal progression models of the disease can help both in describing the pathological processes, even before the onset of the actual symptoms, and in forecasting the progression of individual patients based on previous visits. Doing so may pave the way for precision medicine and may inform the design of clinical trials, both by identifying pathways that cause the disease to progress, and by enriching the trials in subjects that correspond to the targeted disease stage and progression profile.

This chapter provides a brief presentation of the main notions that this thesis builds on: our current knowledge of Alzheimer's disease, the types of available medical data that may be useful as well as the main public databases, and the general statistical frameworks of disease progression modelling that we will improve upon.

Our methodological contributions consist in improving existing modelling frameworks to describe the complex progression of biomarkers or neuroimaging scans over time. In a second phase, we apply these disease progression models to real life data to explore the heterogeneity of AD progression with regard to risk factors. Last, we devise forecasting tools for cognitive decline that leverage information from patients' history of multimodal data. The last section of this introduction provides a more detailed overview of this thesis.

Alzheimer's Disease

Alzheimer's disease (AD) is a complex, progressive, and devastating neurodegenerative disorder. It is characterized by a gradual decline in cognitive function, memory loss, and the inability to perform activities of daily living. It is the most common form of dementia, accounting for approximately 60-70% of all cases (Alzheimer's Association, 2021). It is estimated to affect 55 million people worldwide, with this number expected to triple by 2050. Despite significant advances in the understanding of the disease pathology and therapeutic interventions, AD remains a significant public health challenge worldwide, both regarding cost to society and impact on the daily lives of patients and caregivers.

Clinical presentation The clinical presentation of AD is characterized by a progressive decline in cognitive function, including memory, language, executive function, and visuospatial abilities. The early stages of AD are often characterized by forgetfulness, including the inability to remember recent events or conversations, misplacing objects, and difficulty in following instructions. As the disease progresses, patients may experience difficulties in problem-solving, decision-making, and planning. They may also have difficulty with tasks that involve multiple steps, such as preparing a meal or getting dressed. In advanced stages, they may become unable to perform basic activities of daily living, such as bathing, toileting, and feeding, and may not recognize close relatives and friends.

In addition to cognitive decline, individuals with AD may also exhibit behavioural and psychological symptoms, including depression, anxiety, agitation, and hallucinations. These symptoms are often distressing for both the patients and their caregivers and can significantly impact their quality of life. As the disease progresses, individuals may experience difficulties with social interaction and communication as they become withdrawn and disengaged from social activities.

It bears noting that the clinical presentation of AD is highly variable regarding the ordering and intensity of the changes, depending on a variety of genetic, environmental or biological factors.

Pathogenesis and diagnosis The pathogenesis of Alzheimer's disease (AD) involves a complex interplay between various pathological processes, including the accumulation of beta-amyloid (Aβ) plaques, neurofibrillary tangles, chronic inflammation, and neuronal loss that lead to brain atrophy. The chronological ordering of these abnormalities is not entirely clear, as they likely interact and influence one another in complex ways.

It is generally believed that the accumulation of Aβ plaques is an early event in the pathogenesis of AD. Beta-amyloid is a protein that forms insoluble aggregates that are toxic to neurons and disrupt normal neuronal function. The accumulation of beta-amyloid is thought to trigger a cascade of events that lead to the formation of neurofibrillary tangles, consisting of twisted fibres made up of a protein called tau. Tau also disrupts normal neuronal function, leading to the degeneration and death of neurons. Chronic inflammation and oxidative stress are also thought to play a role in the pathogenesis of AD, as they damage neurons and promote the formation of Aβ plaques and neurofibrillary tangles, exacerbating the degenerative processes. These processes have also been shown to lead to cholinergic deprivation, which refers to the loss of cholinergic neurons and the resulting decrease in the neurotransmitter acetylcholine (ACh) in the brain. ACh plays a crucial role in cognitive function, including learning, memory, attention, and arousal, and its depletion is believed to contribute to the cognitive and behavioural symptoms of AD.

Before the patient becomes cognitively impaired, there is a long phase referred to as prodromal stage, that is characterized by the silent accumulation of brain structural and functional alterations that will only bear consequences much later in the patient's life. For instance, Aβ concentration often crosses an abnormality threshold up to 20 years before the onset of cognitive symptoms and tau protein concentrations up to 15 years ahead. In the prodromal phase, the cognitive state of patients is typically referred to either as Cognitively Normal (CN) or Mildly Cognitively Impaired (MCI), before the AD diagnosis can be validated. Fig. 3 summarizes an empirical description of the cascade of events across AD spectrum.

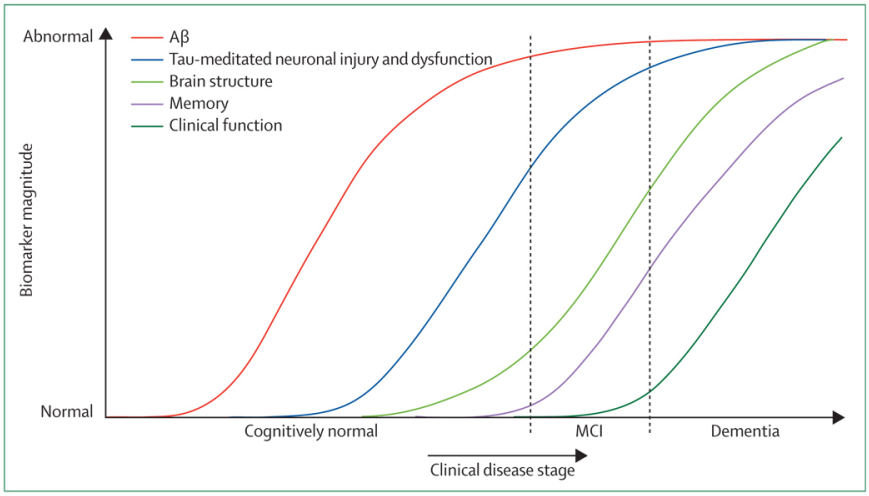


Figure 3: Schematic representation of the cascade of biomarkers' decline over the course of AD, according to [Jack et al., 2010]. The large amount of time between the first signs of accumulation of Aβ and tau proteins in the brain, and the corresponding cognitive decline, make early monitoring of the disease paramount for adequate care and forecast.

For a long time, AD diagnosis was only possible post-mortem when autopsy revealed the Aβ plaques in the brain. The diagnosis of AD is now mainly based on clinical evaluations. The criteria for AD diagnosis proposed by the National Institute on Aging and Alzheimer's Association include the presence of cognitive impairment, which is confirmed by cognitive testing, and the exclusion of other potential causes of dementia. To that effect, neuroimaging and laboratory tests may be used to support the diagnosis and exclude other possible aetiologies. These tests include magnetic resonance imaging (MRI), positron emission tomography (PET), and cerebrospinal fluid (CSF) analysis, that confirm the presence of pathological levels of tau and Aβ as well as specific patterns of brain atrophy. It was noted, in the National Alzheimer's Coordinating Center database, that about 15% of clinical diagnosis are inconsistent with neuropathological diagnosis [Gaugler et al., 2013], highlighting the need for multiple modalities of data in order to provide an accurate diagnosis.

Risk factors AD is a complex and multifactorial disorder with a range of risk factors that have been implicated in its pathogenesis. Age is the most well-established risk factor for AD, as the incidence doubles every five years after the age of 65, with an estimated prevalence of 1 in 10 individuals over the age of 65 and nearly 1 in 3 individuals over the age of 85.

Genetic factors also mitigate the risk of developing AD. Mutations in genes that are involved in the production and processing of beta-amyloid, including the amyloid precursor protein (APP) and presenilin 1 and 2 genes, have been linked to early-onset familial forms

of AD. In addition, variants in the apolipoprotein E (APOE) gene, which plays a role in lipid metabolism and transports proteins in the brain, have been linked to an increased risk of developing AD in both familial and sporadic forms of the disease. Individuals who inherit one copy of the APOE-ε4 allele have a four-fold increased risk of developing AD, while those who inherit two copies have a ten-fold increased risk [Tanzi, 2012]. On the other hand, individuals with copies of the APOE-ε2 allele show lower incidence of AD.

Female sex is also a significant risk factor for Alzheimer's disease. Women are more likely than men to develop AD, as they represent 2/3 of patients. This increased risk may only be partly attributed to the longer life expectancy of women, as sex-specific factors are likely to contribute to the increased risk of AD in women, such as differences in hormone levels and the effects of menopause. The extent of the sexual dimorphism of AD regarding the different hallmarks of the disease is, to this day, not fully quantified and understood.

Environmental, social and lifestyle factors have also been linked to the risk of developing AD. Head injury has been shown to increase the risk of developing AD later in life, particularly if the injury is severe or results in loss of consciousness. Cardiovascular disease and its risk factors, such as hypertension, hypercholesterolemia, and diabetes, have also been associated with an increased risk of AD, possibly due to their effects on cerebral blood flow and oxidative stress. Lifestyle factors, such as a healthy diet and regular exercise, have been linked to a reduced risk of developing AD, while low educational attainment has been associated with an increased risk. Fig. 4 lists the identified modifiable risk factors for AD.

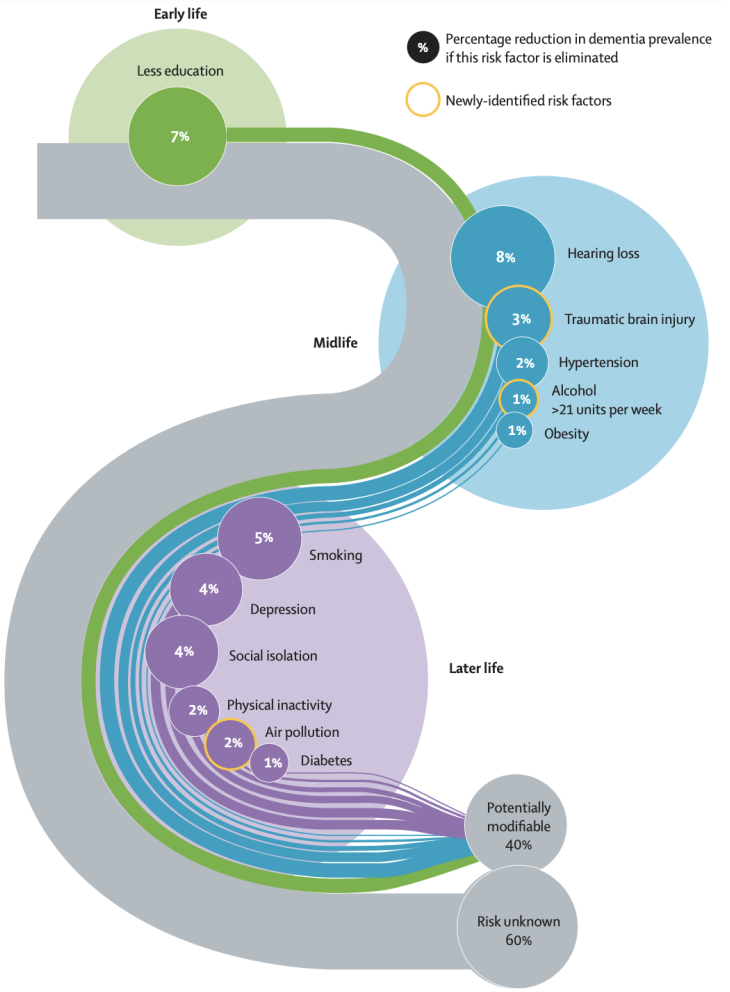


Figure 4: List of all potentially modifiable risk factors according to [Livingston et al., 2020]. These figures differ between ethnicities.

Treatment strategies and care practices Although there is currently no validated cure for Alzheimer's disease, there are a number of treatment strategies and care guidelines that have been developed to manage the symptoms and improve the quality of life of affected individuals. These treatment strategies include pharmacological interventions, cognitive and behavioural therapies, exercise programs, and nutritional interventions. The pharmacological interventions can be either symptomatic or disease-modifying (DMT). The former only act on symptoms such a cognition or depression, but do not have a lasting effect on the progression of the disease, meaning that the effects are depleted when treatment stops, and that the underlying processes such as neurodegeneration and accumulation of proteins are not affected. The latter, on the other hand, act on one of the cogs of the disease and have a lasting impact on one of the biomarkers, even after the end of the treatment. Symptomatic treatments have shown moderate efficacy, and have even been deemed ineligible for reimbursement by the French social security system as of 2018. Besides, research for candidate drugs has been mostly focusing on disease modifying treatments for the past decade, due to the greater promises. However, at the Alzheimer's Disease/Parkinson Disease (AD/PD) conference in 2023, pharma groups and influential group leaders announced their will to invest more to discover efficient symptomatic treatments, in part due to the disappointing results of DMTs.

Figure 5: List of all clinical trials for Alzheimer's Disease as of 2022 according to [Cummings et al., 2022]. Agents underlined are new to the pipeline since 2020.

The candidate DMTs can be classified as biologics (e.g. monoclonal antibodies, vaccines, and gene therapy) and small molecules (drugs typically taken orally and < 500 Daltons in molecular weight) [Cummings et al., 2022]. 83% of current trials targets are DMT, 6.0% target behavioural and neuropsychiatric symptoms, while 9.8% target cognitive enhancement. Among the candidate DMTs, 33.6% are biologics and 66.4% are small

molecules, which can both target any chosen mechanism of the pathological pathways. Fig. 5 summarizes all the ongoing clinical trials according to the phase and targeted mechanism.

In 2021, a monoclonal antibody targeting Amyloid called *Adecanumab* was provisionally approved by the FDA, despite conflicting evidence regarding its efficacy for the mitigation of cognitive decline. It was not approved by the European Medical Administration and Japanese Medical Administration. The decision of the FDA sparked controversy because of the not-well-understood influence of the drug on pre-clinical and mild AD patients. It proved to almost entirely clear the brain of amyloid plaques, without a clear mitigation of cognitive decline. Besides, it constitutes an expensive and intrusive treatment for patients, with potentially heavy adverse effects. The wide scale availability of the drug is hoped to provide a more convincing conclusion for cognitive decline for patients at the earliest stages of the disease. Since then, another monoclonal antibody called *Lecanemab* was demonstrated to both clear amyloid and reduce cognitive decline in a phase III trial, and received accelerated approval by the FDA in early 2023.

In terms of care procedures for practitioners and family members, it is suggested to include regular medical check-ups, monitoring of cognitive function, and management of co-morbid conditions such as depression and anxiety. Other care procedures may include providing a safe and supportive living environment, engaging in meaningful activities that promote cognitive function and social interaction, and providing emotional and social support to the person with Alzheimer's disease and their caregivers.

It is noteworthy that while this thesis is not directly focused on the treatment strategies of AD or the organization of clinical trials, the development of multimodal disease progression models can offer significant benefits for these purposes. These models have the potential to assist pharmaceutical companies in identifying relevant therapeutic targets and recruiting a smaller number of patients at specific stages of the disease. Furthermore, they can aid in the identification of relevant biomarkers for tracking disease progression and assessing treatment effectiveness. They can facilitate patient stratification, allowing for personalized medicine by identifying subgroups with distinct disease trajectories or responses to treatment. Additionally, they enable simulation and prediction, supporting the optimization of trial design, dosage selection, and patient selection. In summary, the motivation behind multimodal disease progression models largely stems from the desire to enhance the effectiveness and efficiency of clinical trials for AD.

Medical data

Medical data include clinical data, genomic data, imaging data, and various types of sensor data. As the healthcare industry continues to adopt new technologies and methods for data collection and analysis, medical data has become an increasingly important resource for improving patient outcomes, advancing medical research, and informing public health policy. In this section, we will review the modalities of data that are most useful for the monitoring and modelling of AD progression.

Clinical assessments

Cognitive testing Cognitive testing is a critical component of the evaluation and management of most neurodegenerative diseases. It involves the administration of standardized assessments that measure various aspects of cognitive function, such as memory, attention, language, and executive function. The goal of cognitive testing in AD is to identify early signs of cognitive decline, track disease progression, and monitor response to treatment. Commonly used cognitive tests in AD include the Mini-Mental State Examination (MMSE, from 30 to 0), the Montreal Cognitive Assessment (MoCA, from 30 to 0), the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog, from 0 to 85) and the Clinical Dementia Rating sum-of-boxes (CDR-sb, from 0 to 3). These tests are typically administered by trained healthcare professionals and can be conducted in a variety of settings, including clinics, hospitals, and research studies.

For example, cognitive testing can help identify specific cognitive domains that are affected early in the disease course, such as episodic memory or executive function, and can help distinguish AD from other types of dementia or cognitive impairment. Cognitive testing can also be used to track disease progression and assess the effectiveness of interventions, such as medications or cognitive rehabilitation programs.

Neurological and physical exams Additional neurological assessments (e.g. reflex testing or reaction to light) can complete the insights of standardized cognitive tests. Physical exams can help identify motor deficits, gait abnormalities, and other signs of neurological dysfunction. Other clinical assessments that may be used in AD include psychiatric evaluations, such as assessments of mood and behaviour, and assessments of activities of daily living, which can provide information about the individual's functional abilities and needs. Besides, such exams are often administered to both the patient and the spouse in order to potentially reveal diverging experiences of the condition.

Genetic testing Genetic testing evaluates the presence of genetic risk factors associated with Alzheimer's disease (AD). Genetic testing for the APOE-ε4 allele can be performed using a simple blood or saliva test and can provide valuable information about an individual's risk of developing AD. In addition to APOE-ε4 testing, other genetic tests may also be used that target the presenilin 1 and 2 genes, the amyloid precursor protein gene (APP) and the triggering receptor expressed on myeloid cells 2 (TREM2) gene. Full genome sequence are rarely acquired and are only useful for research purposes.

Biomarkers

Biomarkers are measurable biological indicators that can provide insights into disease processes, progression, and response to treatment. Common biomarkers used in the evaluation of AD include levels of amyloid beta and tau proteins that can be measured either in the cerebrospinal fluid (CSF) using lumbar puncture, in the blood through regular blood testing, or through imaging techniques.

CSF biomarkers For AD, established fluid biomarkers are t-tau (total tau), p-tau (phosphorylated tau) and Aβ42 (a form of Aβ that can form neuritic plaques). For AD patients, CSF Aβ is lowered, while t-tau and p-tau are increased. It is hypothesized that a decrease in soluble Aβ42 in the CSF (the natural form of the protein) is indicative of a higher amount of insoluble plaques (the pathological form of the protein) in the brain, but the precise mechanism that link CSF Aβ42 concentration with actual amyloid load in the brain are not fully understood. The ratio of concentrations between Aβ42 and Aβ40 –another peptide derived from the APP—is often claimed to be a better predictor of AD than the individual concentrations. Besides, t-tau concentration is a very generic indicator of neurodegeneration, while p-tau is more specific to AD. Tau proteins can be phosphorylated at different epitopes, and several pathophysiologically relevant variants have been identified: p-tau181, p-tau199, p-tau217 and p-tau231. Current research focuses on identifying the most sensitive epitopes for diagnosis and the ones that should be targeted for therapeutic intervention. Several other, less validated, biomarkers that reveal neuroinflammation can be evaluated in the CSF, such as Neurofilament light chain, Glial fibrillary acidic protein, Cytokines and chemokines. However, although CSF biomarkers are very discriminative of AD, such testing is typically reserved for research purposes or specialized clinical settings, as it requires expensive equipment and specific expertise.

Blood-based biomarkers Blood-based or plasma-based biomarkers are also being investigated for their potential to aid in the diagnosis while being less invasive and easier to obtain than CSF samples. Almost all the same biomarkers can be discovered in blood sampling, but the soluble p-tau concentration, especially p-tau217 and p-tau231 are believed to be the best markers of AD progression [Milà-Alomà et al., 2022]. Several studies aim to compare the CSF and blood measurements and elucidate the combinations of biomarkers that are most predictive of AD. Similarly to CSF biomarkers, such testing is still limited to research purposes but holds a lot of potential to be implemented in clinical practice or provide quantitative outcomes for clinical trials.

Imaging biomarkers Imaging biomarkers are a class of biomarkers that are measured on medical images, and quantify specific changes in the brain. Several imaging modalities have been studied in the context of AD research, in order to extract biomarkers, including:

- Magnetic resonance imaging (MRI): MRI is a non-invasive imaging technique that uses magnetic fields and radio waves to measure the positions of water molecules in the brain, in turn creating detailed images of the brain's different tissues. In AD, MRI can be used to measure brain atrophy and structural changes.

- Functional MRI (fMRI): fMRI is a subtype of MRI that measures changes in blood oxygenation levels in the brain, which can be used to infer changes in neuronal activity that arises from the specific tasks performed by a subject. fMRI can be used to investigate functional connectivity between different brain regions, which may be disrupted in AD.

- Positron emission tomography (PET): PET is a nuclear medicine imaging technique that uses small amounts of radioactive substances, administered intravenously, to visualize metabolic and biochemical processes in the brain. The most common tracer tracks glucose metabolism in the brain, but several PET tracers have also been developed to target specific pathological changes in AD, including amyloid plaques and tau protein tangles.

Neuroimaging

This section will present with greater details the different modalities that allow the acquisition of images of the brain that are often used to monitor AD progression. In practice, Magnetic Resonance modalities are non-invasive and widely used, while nuclear modalities offer more insights into processes that are specific to AD, although the production and transportation of the radioactive tracers is a technical challenge that limits their widespread use. For each modality, we describe the acquisition protocol and exhibit how AD specific alterations are revealed.

Magnetic Resonance Imaging As previously mentioned, Magnetic resonance imaging (MRI) uses a strong magnetic field and radio waves to generate detailed images of internal structures in the body. In the context of Alzheimer's disease (AD) monitoring, MRI has become an important tool for visualizing brain changes that occur in individuals with AD. Depending on the radio sequence sent as input, different structures can be highlighted. T1-weighted (T1), T2-weighted (T2), T2*-weighted (T2*) Fluid-Attenuated Inversion Recovery (FLAIR), and Diffusion-Weighted Imaging (DWI) are different types of MRI sequences that are commonly used in clinical practice and research to evaluate brain structure and function.

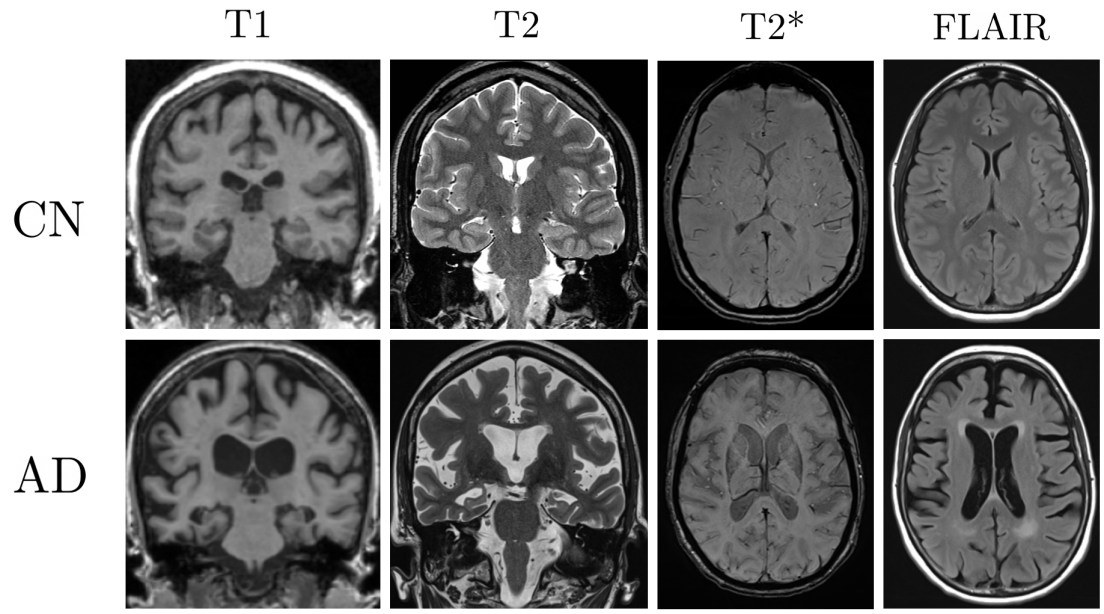


Figure 6: Example of healthy subjects (CN) and AD subjects for the most common MRI sequences. (Coronal views for T1 and T2, axial views for T2* and FLAIR)

- T1-weighted images provide good contrast between different brain tissue, and are used to identify anatomical structures such as the cortex, white matter, and subcortical structures. T1-weighted images are particularly useful for detecting brain atrophy, which is a common feature of AD. In AD, T1-weighted images notably show hippocampal atrophy and cortical thinning, which are characteristic of the disease.

- T2-weighted images, on the other hand, are sensitive to changes in water content and can detect abnormalities such as edema, hemorrhage, and demyelination. In AD, T2-weighted images may reveal white matter hyperintensities, which are associated with small vessel disease and have been linked to cognitive impairment.

- T2*-weighted images are sensitive to changes in local magnetic field inhomogeneities. In AD, it can detect the presence of microhemorrhages and iron deposits in the brain.
- Fluid-attenuated inversion recovery (FLAIR) is a specialized MRI sequence that is particularly sensitive to changes in cerebrospinal fluid (CSF). FLAIR can be used to detect small lesions and other changes in the brain that may be missed on T1 or T2-weighted images. In AD, FLAIR images may show periventricular hyperintensities, which are associated with white matter damage.
- Diffusion-weighted imaging (DWI) is a technique that measures the movement of water molecules in tissue. DWI is particularly sensitive to changes in the microstructure of white matter, and can detect abnormalities such as axonal injury and demyelination. In AD, DWI may reveal changes in the integrity of white matter tracts, which can affect brain connectivity and contribute to cognitive decline.
- Functional MRI (fMRI) is a technique that measures changes in blood flow and oxygenation in the brain in response to neural activity. It is based off of a T2* sequence that measures blood oxygenation level while the patient performs a cognitive task. Statistical testing between activation levels and superimposing the results with a structural MRI image provides a map of the activated regions for a variety of tasks.

Positron Emission Tomography As introduced earlier, PET uses radioactive tracers to visualize metabolic activity and molecular changes in the body. There are three main types of PET imaging agents commonly used in AD research and clinical practice: glucose, amyloid, and tau.

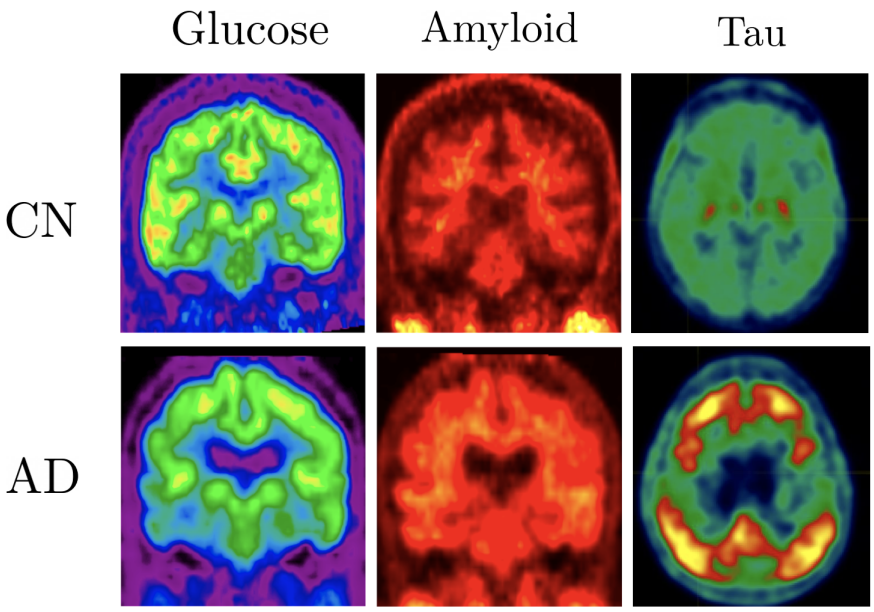


Figure 7: Example of healthy subjects (CN) and AD subjects for the three categories of PET imaging. The glucose marker is fluorodeoxyglucose, the amyloid marker is florbetabir and the tau marker is flortaucipir. (Coronal views for FDG and Amyloid, axial view for Tau).

- Glucose PET imaging agents, such as [18F]-fluorodeoxyglucose (FDG), are glucose analogues that are taken up by metabolically active cells. In AD, FDG-PET may show decreased uptake, specifically in the temporal and parietal lobes, reflecting reduced glucose metabolism and neuronal dysfunction in these areas.

- Amyloid PET imaging agents, such as Pittsburgh Compound-B (PiB), [18F]-florbetapir and [18F]-flutemetamol, bind to beta-amyloid plaques in the brain. In AD, amyloid-PET may show increased uptakes in the cortical regions, reflecting the presence of beta-amyloid deposits.
- Tau PET imaging agents, such as [18F]-flortaucipir (FTP), bind to aggregated tau protein in the brain. In AD, FTP-PET may show increased uptake in regions where tau tangles accumulates, such as the medial temporal lobe.

Image processing A proper image processing pipeline is an essential step in order to perform quantitative analysis. Acquired images are usually heterogeneous as they depend on the manufacturer and generation of the hardware, the acquisition protocol and the image reconstruction procedure. Providing a list of reproducible steps for each experiment is a foundation of the neuroimaging research domain. Such steps usually include the following:

- Bias field correction: MR images can be corrupted by magnetic field inhomogeneities, which need to be corrected using intensity normalization algorithms.
- Intensity rescaling and standardization: images usually have different intensity ranges and distributions of the same tissues, which can be corrected through a rescaling between 0 and 1 and standardization techniques such as histogram matching.
- Skull stripping: the brain is extracted from the image after removal of the skull.
- Image registration: images from different sources are aligned, either globally (through rigid and affine registration) or locally (through non-rigid registration), so that voxels of corresponding positions correspond to similar information.
- Image segmentation and parcellation: after defining regions of interest (ROIs) within the brain, tools have been developed to measure parameters such as volume, cortical thickness, grey matter densities or regional tracer uptakes. FreeSurfer is a commonly used tool that can perform both automated and manual segmentation of various brain structures through the analysis of surfaces that separate the brain structures. SPM (Statistical Parametric Mapping) also offers automated segmentation and normalization of brain structures. It aligns brains on a reference coordinate space and learns the distribution of grey matter in each volumetric region. This step allows transforming high-dimensional imaging data into scalar values of biomarkers.

Each of these steps can resort to a range of available algorithms. In order to ensure reproducibility, we use the Clinica software [Routier et al., 2021] that wraps all the common processing tools into pipelines. The exact combination of steps depends on the task at hand, and will always be presented in the corresponding section.

Longitudinal databases

Several public databases have been assembled to foster data-driven research. This thesis focuses on progression models that require longitudinal data as input, meaning that each subject provides repeated measurements over time, in order to showcase a progression.

- The Alzheimer's Disease Neuroimaging Initiative (ADNI), which collects clinical, genetic, biomarkers and imaging data in North America. It was acquired in 4 waves : ADNI1, ADNIGO, ADNI2 and ADNI3, with slightly different protocols and inclusion criteria for new patients and follow-up patients at each stage. Using all waves, it features 747 AD patients, 688 MCI subjects and 643 healthy controls with an average of 7.2 visits per patient. More information can be found at `adni.loni.usc.edu`.

- The Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing (AIBL). It features 211 AD patients, 133 MCI and 768 healthy comparable controls with an average of 2.8 visits per patients. Is it often chosen as a replication cohort for studies on ADNI since it follows the protocol of ADNI1. More information can be found at `aibl.csiro.au`.
- The MEMENTO cohort, which includes MRI, clinical, genetic and biomarkers data from patients in France. It features 310 AD patients for a total of 2323 participants and an average of 2.1 visits per patients with imaging data.
- The Open Access Series of Imaging Studies (OASIS) is a retrospective cohort that includes healthy controls and individuals with various neurological and psychiatric conditions. It features 1098 patients with an average of 2.0 visits per patient.

The work conducted in this thesis was exclusively performed using data from the ADNI, because it features more data modalities and more patients, with a longer follow-up, which is paramount for accurate longitudinal modelling. The exact demographics and modalities used for each contribution will be listed in the corresponding sections.

Disease Progression Models for AD

Disease Progression Modelling (DPM) is a field that emerged in the early 2010s as a response to the need for quantitative depictions of the pathological unravelling of diseases. The initial inspiration for this field came from the hypothetical models of Alzheimer's disease progression, as notably proposed by [Jack et al., 2010]. These clinically-inspired models illustrate the length of the prodromal phase, as well as the multimodal aspect of disease progression.

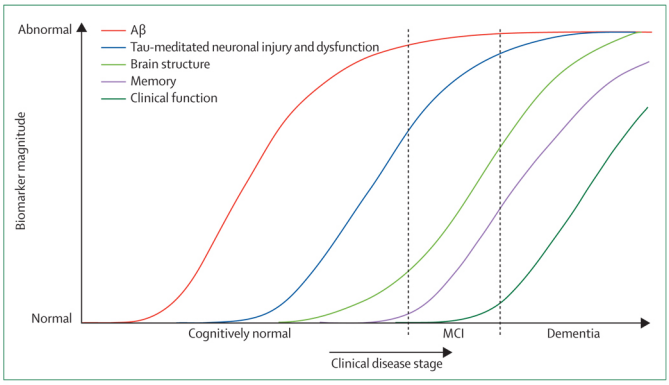


Figure 8: Hypothetical model of [Jack et al., 2010].

The amount of data available to clinicians is both too high for qualitative human evaluation and too low for fully data-driven machine-learning approaches. Besides, the heterogeneity of data types complicates integration through traditional statistical methods. Data-driven DPM aim to find a balance between *prior* knowledge of disease features and patterns learned from data. Several approaches have been proposed for DPMs, and we will hereby list the most frequently used ones. The following section will introduce the general framework of mixed-effects models and the specific regression framework that we will improve upon in this thesis.

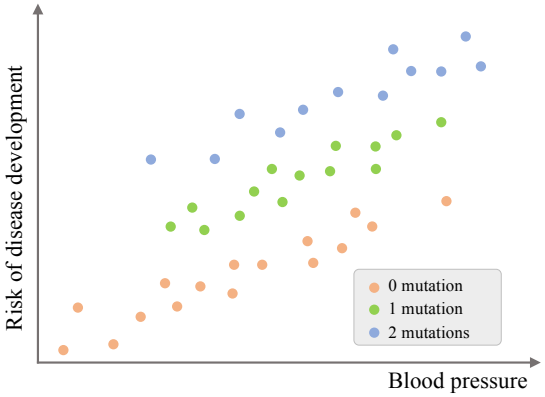
Event-based models characterize the sequence of events during the progression of the disease [Fonteijn et al., 2012, Young et al., 2014]. Cut-off points of abnormality are inferred from observed biomarkers, and disease stage is mapped to a discrete set of biomarker-abnormality events, effectively ordering the cascade of events. The main limitations are that the variability between individuals is only accounted for as the uncertainty in the cascade of events, and that the time delay between successive events is not measured.

Dynamic models formulate the changes in biomarkers using differential equations [Ito et al., 2011, Samtani et al., 2012, Oxtoby et al., 2018, Abi Nader et al., 2021], with the caveat that covariates and all sources of random variations should be integrated in the differential equation system, which makes interpretation difficult.

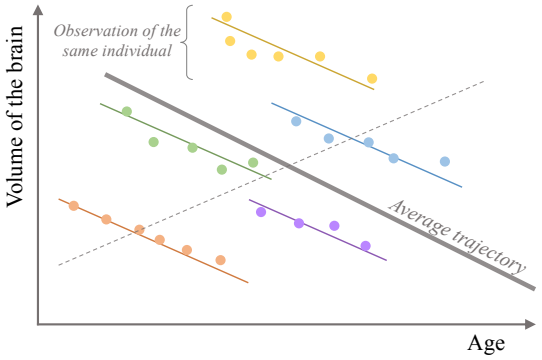
Gaussian-process models are based on the probabilistic estimation of biomarkers' trajectories, and on the quantification of the uncertainty of the predicted individual pathological stage [Lorenzi et al., 2019]. They define a non-parametric, Gaussian process regression model for individual trajectories, introduce a monotonicity constraint, and model individual time transformations encoding the information on the latent pathological stage. That time reparametrization function makes the model less reliant on the age of the patients, which is a poor proxy for disease stage. The long-term model can be used as a statistical reference representing the transition from normal to pathological stages, thus allowing probabilistic diagnosis in the clinical scenario. Such methods have been extended to the modelling of multimodal changes in the brain using brain networks [Garbarino et al., 2021] and full-size images [Abi Nader et al., 2020].

Mixed-effects models

Mixed-effects models are a broad class of hierarchical statistical models that account for both fixed and random effects in the analysis of data with nested or clustered structures. Fixed effects refer to the factors of interest that are assumed to be constant across all levels, while random effects account for the variability due to individual differences or differences between groups. Mixed-effects models are commonly used in social, behavioural, and biomedical research to analyse data with repeated measures, covariates that mitigate the observed variable or multilevel data structures. Fig. 9 illustrates how mixed-effects can provide a more accurate depiction of underlying biological processes by accounting for the hierarchical structure of data.



(a) The risk of disease development might depend on the blood pressure (fixed effect) and the number of mutations of a given allele (random effect).



(b) For longitudinal data, a linear regression (dashed line) might return results that are opposed to the real biological progression over time (fixed effects) due to high inter-individual variability (random effects).

Figure 9: Examples of mixed-effects usefulness for (a) independent data and (b) longitudinal data with repeated observations per individual. Courtesy of [Koval et al., 2017b].

Mixed-effects models are particularly useful for analysing longitudinal data, where repeated measurements are taken over time on the same individuals. In these cases, mixed-effects models can account for the correlation between observations from the same individual and can model the trajectory of change in the outcome variable over time.

Linear mixed-effects models Linear mixed-effects models (LME) are the first mixed-effects models to be widely used [Laird and Ware, 1982, Verbeke, 1997]. They can be understood as an extension of a traditional linear regression – the fixed effects–, with an added degree of freedom – the random effects that model the group-specific derivation from the linear trend. The general form of a linear mixed-effects model can be expressed as:

$$Y = X\beta + Z\gamma + \varepsilon$$

where Y is the vector of observations, X and Z are the input variables for the fixed and random effects, β and γ are the corresponding coefficients, and ε is a vector of independent and identically distributed errors. Assumptions can be made about the distribution of the random effects γ

Non-linear mixed-effects models Non-linear mixed-effects models are an extension of linear mixed-effects models that allow for the modelling of non-linear relationships between the response variable and the covariates. The general form of a non-linear mixed-effects model can be expressed as:

$$Y = f(X, Z) + \varepsilon$$

where Y is the vector of observations, X and Z are the input variables that correspond respectively to fixed effects and random effects, $f(\cdot)$ is a non-linear function of X and Z and ε is a vector of independent and identically distributed errors. In some formulations, the non-linear function f can be parametrized by θ as part of the fixed effects.

Temporal formulation for biological phenomenons In the context of modelling longitudinal evaluations of biological phenomenons, it seems straightforward to describe individual progressions through the combination of fixed effects that describe the average progression over time and random effects that represent the individual variations to the mean scenario, as illustrated in Fig. 9b. Overall brain size decreases in a linear fashion over time (fixed effects) although each individual has a different basal volume (random effects). In practice, different patients show different basal volumes but also different paces of atrophy so a refinement of the LME for biological data is proposed in the *random-slope, random-intercept* model, where random effects consist of individual slopes (e.g. pace of atrophy) and intercepts (e.g. basal volume), in which case the fixed effects are defined as the average slope and intercept. For repeated measurements $y_{i,j}$ of patients $i = 1, \cdots, N$ at times $t_{i,j}, j = 1, \cdots, N_i^{visits}$, such a model writes

$$y_{i,j} = (\alpha^1 + \beta_i^1)t_{i,j} + (\alpha^2 + \beta_i^2) + \varepsilon_{i,j}$$

The fixed effects are $\boldsymbol{\alpha}$ and the average trajectory writes $t \mapsto \alpha^1 t + \alpha^2$. The random effects $\boldsymbol{\beta}_i, i = 1, \cdots, N$ represent variations to the model intercept and slope to account for the variability in the population. A normal prior is added for the distribution of $\boldsymbol{\beta}$.

Such modelling is particularly suited to dynamics with a temporal alignment, i.e. in which a given time corresponds to a given event or threshold in the progression. It can, for instance, be accurate for studies on the effect of a drug from a reference time-point at which the drug is administered, or for dynamics that have a known starting point. In the context of disease progression, on the other hand, patient present unaligned dynamics since they declare the disease at different onset ages at which the decline begins. Coming back to the example of Fig. 9b, it can be argued that such a difference of onset can be accounted for by the intercept β_i^2: there is no way to decide if a patient started atrophy 1 year earlier or if it had a basal volume lower by one unit of annual atrophy. As will be made evident in Section 1.2, this identifiability problem can be lifted in the context of multivariate models, and the age at onset becomes particularly relevant for non-linear dynamics. If we add an individual temporal shift τ_i to the random effects, we can re-align all patients on a common pathological timeline through $t \mapsto t + \tau_i$. Adding a second temporal parameter, [Jedynak et al., 2012] introduced an affine time-reparametrization of the real age $t \mapsto \alpha_i t + \tau_i$ onto the pathological age, where α_i can be understood as the pace of decline. For instance, $\alpha_i > 1$ means that the pathological age of the patient will increase faster than his actual clinical age. Similarly, $\tau_i > 0$ means that the pathological age is higher than the true age of the patient.

The progression of most biomarkers has long been hypothesized to follow sigmoid shapes, as illustrated by the well-known progression in Fig. 3. [Jedynak et al., 2012] provided one of the first examples of logistic regression, coupled with the mentioned affine reparametrization for the modelling of AD biomarkers, which was later adapted for voxel-wise neuroimaging biomarkers by [Bilgel et al., 2014]. [Raket, 2020] proposed to use exponential regression and [Donohue et al., 2014] to use semi-parametric regressions. In all

these models, putting all patients on a common pathological timeline allows learning the long-term progression using a collection of short-term measurements from real patients.

Disease course mapping

The progression model that this thesis builds upon is called "Disease course mapping", and is implemented in the open-source software Leaspy (https://pypi.org/project/leaspy/). It is a generative Bayesian spatio-temporal mixed-effect model, in which the fixed effects correspond to the group-average trajectory, and the random effects describe how each individual trajectory relates to the average scenario. The random effects are:

- the *timeshift* (or *onset age*) τ_i, that accounts for patients that start earlier or later than average,
- the *pace of decline* (or *acceleration factor*) α_i, that accounts for patients that decline faster or slower than average,
- the *spaceshift* w_i, that accounts for the variability between patients in disease presentation and intrinsic differences that are independent of time. The *spaceshift* can alter the ordering of the events and the starting point, which can differ due to anatomical differences, without impacting the disease stage.

Disease course mapping hypothesizes that the observation space can be embedded with a Riemannian manifold structure, on which the average trajectory follows a geodesic, and all individual trajectories are parallel curves defined by the random effects. This formalism is described in details in Chapter 2, but for now, the model can be simply written

$$y_{i,j} = f\Big(\psi_i(t_{i,j}), w_i, \theta\Big) + \varepsilon_{i,j}$$

where $\psi_i : t \mapsto \alpha_i(t - \tau_i)$ is the patient-wise affine temporal warp, w_i is the spaceshift that accounts for variability between patients, ε is the noise, f is the regression function and θ are the fixed-effects of the model that describe the average progression. To simplify, f can be chosen to be linear, logistic or exponential, depending on the expected behaviour of the modelled biomarker. As previously mentioned, logistic modelling is often preferred for AD biomarkers, so we will illustrate Disease course mapping using logistic shapes.

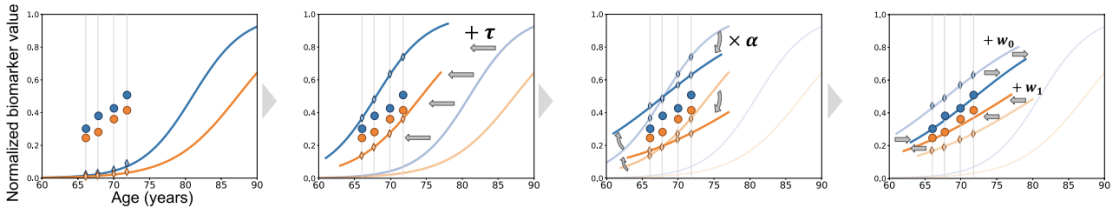


Figure 10: Illustration of the random effects "transforming" the average trajectory to account for the variability between patients, for a model with two features (blue and orange). Dots are the actual measurements and plain lines are the modelled trajectory. More details on this model are provided in [Schiratti et al., 2015a]. Figure courtesy of [Koval et al., 2021b].

Details about the statistical parametrization, estimation procedure and validation are provided in Chapter 1. Without getting into details in this introduction, we can summarize the benefits of Disease course mapping:

- it describes both population-average and individual progression patterns,

- it provides clinically interpretable parameters (onset and pace of decline),
- it disentangles inter-subjects variability from temporal progression,
- it allows imputing missing values or forecast future progression,
- it is resilient to missing and irregularly spaced data,
- it learns long-term patterns from short-term individual measurements.

Machine Learning tasks

Disease progression models (DPMs) offer a comprehensive understanding of the underlying processes of Alzheimer's disease (AD) and enable the prediction of individual progression at future timepoints. However, multiple machine learning (ML) and deep learning (DL) approaches have been developed for similar tasks, such as diagnosis prediction, biomarker forecasting, disease subtypes classification, patients classification according to progression patterns, and region of interest identification. These approaches take biomarkers or full images as inputs, and optimize a ML algorithm for a specific task. While such task-based approaches are highly adaptable, the limited scope of their focus precludes broader insights into the general progression of the disease. Moreover, the quantity of available data is frequently insufficient for fully data-driven ML models to learn the entire disease progression without prior knowledge. Therefore, while relevant ML and DL approaches will be presented in further chapters when they compete with specific aspects of DPMs, it is important to distinguish between DPMs and task-specific ML models.

Contributions and manuscript overview

The goal of this thesis is to leverage multimodal longitudinal data in order to describe AD progression and allow individual forecasting of future decline. More specifically, it aims to model the progression of both cognitive scores and neuroimaging data in order to exhibit a quantitative link between brain alterations and cognitive decline. The contributions can be separated into two categories: methodological contributions that improve upon existing DPM frameworks (Parts I, II and IV), and clinical contributions that use DPMs to illustrate a clinically relevant aspect of the disease (Parts III and IV).

As shown in this introduction, various frameworks of DPMs have been suggested for modelling biomarkers in AD, and typically entail certain presumptions regarding the anticipated behaviour of biomarkers. More specifically, the Disease course mapping, a flexible approach that has shown promise in modelling AD, makes a strong assumption that biomarkers follow sigmoid shapes over time. Although this hypothesis has been widely accepted due to the legacy of hypothetical progression models, we call its validity into question. The Disease course mapping belongs to a family of geometric models in which the observation space is assumed to be a Riemannian manifold, and the metric of which defines the shape of biomarker trajectories. So far, the Riemannian metric has been pre-set to yield sigmoid curves. Our primary contribution is to mitigate this prior assumption by learning the Riemannian metric of the observation space in a data-driven manner, thereby learning the progression curves of biomarkers more effectively. Part I introduces the relevant geometric tools and describes a framework to learn the Riemannian metric of the observation space in the context of mixed-effects DPMs. Comparing our model with Disease course mapping did not reveal significant differences in performance, which suggests that the sigmoid assumption is plausible, and especially relevant for the modelling of cognitive scores.

Another constraint of the DPMs outlined in this introduction, is their restriction to modelling the progression of biomarkers or mesh-valued features derived from images. There are few models that have been created to leverage structural information provided by full-scale images. Yet, images reveal functional and structural alterations of the brain that occur before the onset of the first symptoms, and may reveal a complex interplay between regions that is not accounted for when extracting regional features with processing software. To address this limitation, we proposed a method that combines a convolutional neural network, for dimension reduction purposes, with a latent mixed-effect model, which characterizes the progression of the embeddings of consecutive images. Part II describes the proposed "Longitudinal Variational Auto Encoder" model and inference strategy, and demonstrates that it recovers well known alterations patterns for both AD and healthy ageing regarding brain atrophy and hypometabolism.

Working on progression models for imaging data revealed great differences in alteration patterns between subgroups of patients (namely APOE-ε4 carriers vs. non carriers and men vs. women). This led to the third part of this work, which aimed to describe the influence of both APOE-ε4 genotype and female sex on patterns of brain atrophy, cortical thinning and loss of metabolism over the course of AD. For comparison, men exhibit more pronounced impairment than women at older age, both regarding brain structure and cognitive function, over the course of healthy ageing. The sexual dimorphism of AD regarding cognitive decline, hippocampal atrophy and temporal lobe thinning, has been studied, but no study has provided this analysis for the entire brain and accounted for both sex and APOE in order to disentangle the influence of each factor. Part III thus proposes a descriptive use of DPMs, to model these characteristic features of AD and compare the

distributions of onset ages and rates of decline across subgroups, shedding light on the significance of APOE-ε4 and sex as important factors influencing the manifestation and progression of AD.

One main objective in analysing brain changes through neuroimaging data is to gain insights into the cognitive symptoms that may be expected. To achieve the original goal of this thesis, we utilized DPMs to predict cognitive decline in patients at various stages of the disease using both multimodal imaging and fluid biomarkers. However, the availability of numerous data modalities in a research context does not necessarily mean that all of them are equally informative about a patient's future progression. Our contribution was to identify, for each disease stage, the specific input features that could improve or diminish the quality of cognitive predictions for different time intervals. The selection of cohorts for this research was similar to that used in clinical trials, based on factors such as baseline amyloid and cognitive status. Our work can assist clinical trial organizers in determining which data modalities are most valuable to acquire and which subjects are more likely to experience disease progression throughout the trial, effectively yielding a more powered trial. Part IV describes the approach and findings of this model selection process.

Part I

Riemannian Progression Models for Scalar Data

Chapter 1

Geometric disease progression models

This chapter of the thesis presents geometric tools that underlie the "Disease course mapping" mixed-effect Disease Progression Model (DPM) framework, which serves as the foundation for the research presented. We first briefly introduce key elements of Riemannian geometry, and then describe the "Disease course mapping" model. We provide geometric and statistical descriptions of the model, as well as information about the estimation procedures used. The geometric notions are paramount to the understanding of Chapter 2 while a more superficial understanding of the progression model presented in Section 1.2 will be sufficient for the following chapters. Each chapter will also concisely re-state the necessary notions in order for each section to be readable as a standalone.

Contents

1.1 Riemannian geometry

Riemannian geometry is a branch of mathematics that deals with the study of smooth curved spaces called Riemannian manifolds. It is named after the German mathematician Bernhard Riemann, who laid the foundation for this field in the mid-19th century. This section provides a simplified overview of some key concepts in Riemannian geometry that are useful for the introduction of the DPM framework in Section 1.2, upon which this thesis builds. A common assumption for modelling complex data is that the observed data belong to a particular subspace of the feature space, and that the geometry of said subspace informs the distribution of the data. In the context of longitudinal data, that translates to the hypothesis that repeated measurements follow curves on the observation space. Curved spaces can thus lead to all shapes of progression trajectories. Riemannian

manifolds offer a suitable formalism as it provides a rigorous mathematical description of curves, angles and distances, while allowing non-linear dynamics.

1.1.1 Manifolds

A *manifold* is a topological space in which each point has a neighbourhood that is homeomorphic to Euclidean space, and this can be accomplished using a collection of mappings called an atlas. The atlas maps regions of the space to linear spaces, allowing for calculus to be performed on each linear space and then derived to the corresponding region, resulting in a locally differentiable structure. If the local differentiable structures are continuous from one mapping to the other, then the differentiable structure is said to be globally differentiable, and $\mathbf{M}$ is called a *smooth manifold*.

For a smooth manifold $\mathbf{M}$ of $\mathbb{R}^n$, each point $\mathbf{p} \in \mathbf{M}$ is associated with its tangent space $\mathbf{T_pM}$, which is a linear approximation of $\mathbf{M}$ in the neighbourhood of $\mathbf{p}$. The tangent space contains all possible derivations of $\mathbf{M}$ at $\mathbf{p}$, which can be thought of as the vectors at $\mathbf{p}$ in all directions of the tangent space.

1.1.2 Riemannian metrics

Given a smooth manifold $\mathbf{M}$, a smoothly varying map that associates each point $\mathbf{p} \in \mathbf{M}$ to an inner product $g_\mathbf{p}$ on the tangent space $\mathbf{T_pM}$, is called a *metric* on the manifold. This collection $g_\mathbf{M} = (g_\mathbf{p})_{\mathbf{p}\in\mathbf{M}}$ generalizes the Euclidean scalar product to manifolds. Equipped with this metric, $(\mathbf{M}, g_\mathbf{M})$ is called a *Riemannian manifold*. This key concept allows defining the notions of distances and angles on differentiable structure.

1.1.3 Geodesics

A *geodesic* is a curve on a manifold that locally minimizes the length between two points. In other words, it is a curve that follows the shortest path between two points on the manifold. The notion of a geodesic is a generalization of the concept of a straight line in Euclidean geometry. The equations of geodesics can be obtained by solving a second-order differential equation known as the geodesic equation, which depends on the metric tensor of the manifold.

1.1.4 Exponential Mapping

Considering $\mathbf{p} \in \mathbf{M}$, a tangent vector (referred to as *velocity*) $\mathbf{v} \in \mathbf{T_pM}$ and a geodesic γ such that $\gamma(t) = \mathbf{p}$ and $\dot{\gamma}(t) = \mathbf{v}$. It can be shown that such geodesic is unique so that we rewrite it $\gamma := \mathrm{Exp}_{\mathbf{p},t}(\mathbf{v}) : t \mapsto \mathrm{Exp}_{\mathbf{p},t}(\mathbf{v})(t)$. The *exponential mapping* associates the vector $\mathbf{v}$ to the point reached by this geodesic at time $t+1$. It writes $\mathbf{v} \in \mathbf{T_p} \mapsto \mathrm{Exp}_\mathbf{p}(\mathbf{v}) = \mathrm{Exp}_{\mathbf{p},t}(\mathbf{v})(t+1)$. It is essentially a step in the direction of $\mathbf{v}$ on the manifold.

1.1.5 Parallel Transport and connections

Parallel transport is a method of moving a tangent vector along a curve on a manifold while keeping it parallel to itself. A *connection* is a mathematical tool that provides a way to differentiate vector fields along curves or in the direction of other vector fields. The connection takes into account the curvature of the space and provides rules for differentiating vectors in a manner consistent with the underlying geometry. A connection can be intuitively thought of as a generalization of the gradient for vector fields in curved spaces. The Levi-Civita connection ∇ is a specific connection, uniquely defined by the Riemannian metric of the space, that enables us to define how vectors are transported along curved spaces. Given a manifold $\mathbf{M}$ and a smooth curve $\gamma : I \subset \mathbb{R} \to \mathbf{M}$, a vector field X is said to be *parallel* along γ if $\nabla_{\dot{\gamma}} X = 0$. Given $\mathbf{w}_0 \in \mathbf{T}_{\gamma(t_0)}\mathbf{M}$, one can show there exists a

unique vector field $\mathbf{w}(t)$ parallel along γ such that $\mathbf{w}(t_0) = \mathbf{w}_0$. This corresponds to the *transport* of $\mathbf{w}_0$ along γ such that the vector field $\mathbf{w}(t)$ remains *parallel* to $\mathbf{w}_0$. Besides, given a smooth curve $\gamma : I \subset \mathbb{R} \to \mathbf{M}$, we say that γ is a geodesic of $\mathbf{M}$ if $\nabla_{\dot\gamma}\dot\gamma = 0$, i.e. a smooth curve with zero acceleration.

1.2 Disease course mapping

Here we provide more details on the DPM briefly introduced in the introduction as "Disease course mapping". This mixed-effect framework was introduced in [Schiratti et al., 2015b, Schiratti et al., 2017, Koval et al., 2017a] and this section essentially summarizes these sources. To that effect, we consider the repeated observations of p individuals, such that the i-th individual has been observed n_i times at times $t_{i,1} < \ldots < t_{i,n_i}$. The observation at time t_{ij} is denoted $\mathbf{y}_{ij} \in \mathbb{R}^N$, where $N \in \mathbb{N}^*$. Finally, let us denote $\mathbf{y} = (t_{ij}, \mathbf{y}_{ij})_{\substack{1\leq i\leq p \\ 1\leq j\leq k_i}}$ the set of longitudinal observations.

1.2.1 Geometric description

Using the Riemannian settings introduced in Section 1.1, we consider that observations $\mathbf{y}_{ij}$ belong to a Riemannian manifold $\mathbf{M} \subset \mathbb{R}^N$. We also consider that there exists a geodesic $\gamma_0 : I \subset \mathbb{R} \to \mathbb{R}^N$, reaching p_0 at t_0 ($\gamma_0(t_0) = p_0$) with velocity v_0 ($\dot\gamma_0(t_0) = v_0$), which represents the group-average trajectory, i.e., the global temporal dynamic of disease progression. The individual trajectories are then described as spatio-temporal variations of this mean trajectory in the sense that they derive from it thanks to:

- the *spatial* variability accounted for by the vector $w_i \in T_{p_0}M$, which represents the direction in which the group-average trajectory is shifted to approximate the data $(\mathbf{y}_{ij})_{1\leq j\leq k_i}$ of the i-th individual. It is possible to parallel transport the vector w_i defined at t_0 along the curve γ_0 at any time t, resulting in the vector $P_{\gamma_0,t_0,t}(w_i)$, as shown in Figure 1.1. The exponential mapping of this collection of vectors, $\mathrm{Exp}_{\gamma_0(t)}(P_{\gamma_0,t_0,t}(w_i))$, defines the individual trajectory $\eta_{w_i}(t) := \mathrm{Exp}\gamma_0(t)(P_{\gamma_0,t_0,t}(w_i))$, which corresponds to the Exp-parallel of the group-average geodesic γ_0 in the direction w_i.

- the *temporal* variability is defined by the *pace of decline* $\alpha_i \in \mathbb{R}$ and the *time shift* $\tau_i \in \mathbb{R}$. The individual trajectory $\eta_{w_i}(t)$ may progress at a different speed than $\gamma_0(t)$, so we introduce an affine temporal reparametrization $\psi_i : t \mapsto \alpha_i(t - \tau_i)$. The observation of the i-th individual at time t_{ij} corresponds, on the disease timeline, to the age $\psi_i(t_{ij})$. α_i controls the speed of the dynamic, where $\alpha_i > 1$ (resp. $\alpha_i < 1$) corresponds to faster (resp. slower) progression, and the time shift τ_i enables shifting the temporal progression of a given number of years, where $\tau_i > t_0$ (resp. $\tau_i < t_0$) corresponds to profiles that present late (resp. early) progression.

In case the observation space $\mathbf{M}$ is of high dimension, w_i can be decomposed in an Independent Component Analysis (ICA) manner, such that $\mathbf{w}_i = As_i$ where $A \in \mathbb{R}^{N\times N_s}$ is called the mixing matrix and $s_i = (s_{ij})_{1\leq j\leq N_s}$ are the sources. The idea is that instead of living in a high dimensional space, $\mathbf{w}_i$ can be represented is a subspace spanned by the vectors given by the columns of the mixing matrix A.

Finally, the individual measurements of the i-th individual at time t_{ij} writes

$$\mathbf{y}_{ij} = \eta^{\mathbf{w}_i}(\psi_i(t_{ij})) + \epsilon_{ij}$$

where ϵ_{ij} is the residual noise not captured by the model.

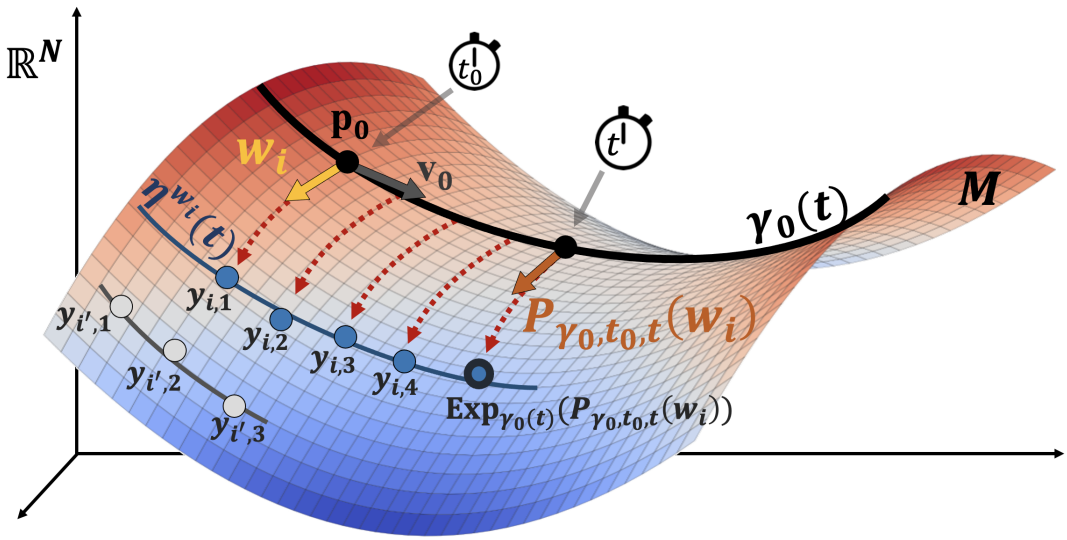


Figure 1.1: Visual depiction of the geometric model. The longitudinal observations $(\mathbf{y}_{ij})$ belong to a Riemannian manifold $\mathbf{M} \subset \mathbb{R}^N$. The black curve is the group-average geodesic γ_0, characterized by p_0, v_0 and t_0. Blue dots and grey dots represent the actual measurements from two patients, and the blue and grey curves are the parallel curves that best match these observations. The vectors $P_{\gamma_0,t_0,t}(\mathbf{w}_i)$ are the parallel transports of the spaceshift $\mathbf{w}_i \in \mathbb{T}_{p_0}\mathbf{M}$, and the exponential mapping of these vectors yields the theoretical trajectories $\eta^{\mathbf{w}_i}$. The temporal parameters (τ_i, α_i) define the pace at which these curves are travelled.

1.2.2 Statistical description

This Riemannian setting describes both the average and individual spatio-temporal trajectories that model the evolution of the biomarkers. On the one hand, γ_0 is parametrized by $t_0, \mathbf{p}_0$ and v_0 and describes the long-term disease progression. On the other hand, $t \mapsto \eta^{\mathbf{w}_i}(\psi_i(t))$ is parametrized by α_i, τ_i and $\mathbf{w}_i$ and gives the geometrical description of the individual trajectory on the manifold.

Since α_i is always positive, we parametrize it as $\alpha_i = e^{\xi_i}$. We choose the Gaussian priors for the noise $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and random effects $\tau_i \sim \mathcal{N}(t_0, \sigma_\tau^2)$ and $\xi_i \sim \mathcal{N}(0, \sigma_\xi^2)$, $s_i \sim \mathcal{N}(0, \mathbf{I})$. The parameters $t_0, \mathbf{p}_0, v_0$, together with the variances $\sigma_\varepsilon, \sigma_\tau, \sigma_\xi$, form the fixed-effects of the model.

1.2.3 Identifiability conditions

To ensure the identifiability of the model in the presence of spatio-temporal variability, the space-shift $\mathbf{w}_i$ has to be orthogonal to the velocity $\mathbf{v}_0$. Otherwise, the projection of $\mathbf{w}_i$ on $\gamma_0(t)$ might interfere with the temporal progression which is controlled by α_i and τ_i. To this end, given $g^{\mathbf{M}}$ the Riemannian metric of the manifold $\mathbf{M}$, we must ensure that $g^{\mathbf{M}}(\mathbf{w}_i, \mathbf{v}_0) = 0$. In practice, the Householder method is used to build an orthonormal basis from which the columns of the mixing matrix A are built. This ensures the orthogonality condition between $\mathbf{w}_i$ and $\mathbf{v}_0$.

1.2.4 Product of 1D models

Here, the model has been introduced for any manifold. This general writing allows keeping the same framework for multiple manifolds, and therefore multiple data types. In Disease

course mapping, the manifold is considered to be a product manifold of 1D manifolds such that $\gamma_0(t) = (\gamma_1(t), \dots, \gamma_n(t))$. In this case, the associated metric is a product of 1D metrics.

Given that w_{ik} the k-th coordinate of $\mathbf{w}_i$, the authors of [Schiratti et al., 2017] show that in the case of a product manifold, the k-th coordinate of $\eta^{\mathbf{w}_i}(\psi_i(t_{ij}))$ writes

$$\eta_k^{\mathbf{w}_i}(t) = \gamma_k(\frac{\mathbf{w}_{ik}}{\dot{\gamma}_k(t_0)} + \psi_i(t_{ij}))$$

This assumption essentially means that the spaceshift only affects the ordering of the cascade of events. In Chapter 2, we revert to the more general setting and learn a metric that is not a product metric, which means that the spaceshift influences not only the ordering of the events, but also the shape of the trajectories.

1.2.5 Choosing a model

The Riemannian metric is essentially a continuous scalar product on the tangent bundle. Setting it *a priori* will yield specific geodesics. To illustrate this, let's consider a manifold of observation $\mathbf{M} = [0,1]^N$, scaled between 0 and 1, with 0 being the most normal and 1 the most abnormal possible value. For the k-th dimension, for $p \in [0,1]$, setting $g^k(p) : (u,v) \mapsto uv$ yields linear geodesics, $g_p^k : (u,v) \mapsto \frac{uv}{p^2(1-p)^2}$ yields logistic geodesics and $g_p^k : (u,v) \mapsto \frac{uv}{p^2}$ yields decreasing exponential geodesics. [Koval, 2020] summarizes all the closed-form expressions of the geodesics and parallel curves for these models. In order for the values predicted by the model $\eta^{\mathbf{w}_i}(\psi_i(t_{ij}))$ to match real observations $\mathbf{y}_{ij}$, the geodesic γ_0 must be chosen such that the bundle of parallel curves matches the individual trajectories.

For one dimensional data ($N = 1$), univariate models can be defined by setting all the spaceshifts to 0. Such models will be used in Part III.

1.3 Estimation

This section introduces the algorithms used to estimate the parameters of the statistical mixed-effects models. These procedures are explained in great details in Chapter 2 of [Koval, 2020]. For the purpose of this thesis, a general understanding of the different steps will suffice, but we refer the eager reader to this source for a more in-depth description.

1.3.1 Statistical Learning

Using standard statistical notations, which differ from the notations of the previous section, we consider the model

$$\mathbf{y}_i = f(\theta, \mathbf{z}_i) + \varepsilon_i$$

where $\mathbf{y}_i$ is the output variable (here the observations y_{ij}), $\mathbf{z}_i$ are the random variables associated with subject i (here the *random-effects*), θ are the model parameters (here the *fixed-effects*), and f is the model. The goal is to find the parameters that best describe the observations, in the sense that they maximize the likelihood

$$p(\mathbf{y}, \theta) = \int p(\mathbf{y}, \mathbf{z}, \theta) \mathrm{d}\mathbf{z} = \int p(\mathbf{y}|\mathbf{z}, \theta) p(\mathbf{z}, \theta) \mathrm{d}\mathbf{z}$$

Generally, no closed-form is available for the likelihood, and we resort to an iterative algorithm to estimate the *maximum likelihood*, or *maximum a posteriori*, for the model parameters, called **Expectation-Maximization** (EM). The E step defines a function that compute the expectation over the random variables, given a current value of parameters θ^k at the $k-$th step of the algorithm, and the M step finds the new parameters θ^{k+1} that

maximize the former function. A pseudo-code description of the alternance between E and M steps is given in Algorithm 1

Algorithm 1: Expectation-Maximization

$\theta \leftarrow \theta_0$
$k \leftarrow 0$
while *Convergence of* θ **do**
  $k \leftarrow k + 1$
  Expectation step : Compute $Q(\theta|\theta^{(k)}) = \mathbb{E}_{p(\mathbf{z}|\mathbf{y};\theta^{(k)})}[\log p(\mathbf{y}, \mathbf{z}; \theta)]$
  Maximization step : Update $\theta^{(k+1)} = \text{argmax}_\theta \, Q(\theta|\theta^{(k)})$
end
Result: return $\theta^{(k)}$

For the non-linear models that we consider, calculating the integral over the random variables during the E step is not possible. Instead, realizations can be drawn from the distribution $p(\mathbf{z} \mid \mathbf{y}; \theta^{(k)})$ to estimate the expected value of the log likelihood function of θ with respect to the current $\theta^{(k)}$ parameters: $Q(\theta \mid \theta^{(k)})$. This approximation allows for the E step to be replaced by a Stochastic Approximation step. The resulting algorithm, called the **Stochastic Approximation Expectation Maximization** (SAEM) algorithm, was first introduced in [Delyon et al., 1999]. We refer the reader to this seminal work for more details on the proof of convergence. It is shown that the Stochastic and Approximation steps asymptotically converge to the same set of solutions as the traditional Expectation step. Algorithm 2 outlines the steps of the SAEM algorithm.

Algorithm 2: Stochastic Approximation Expectation-Maximization

$\theta \leftarrow \theta_0$
$Q_0 \leftarrow 0$
$(\epsilon_k)_{k\geq 0}$ such that $\sum_{k\geq 0} \epsilon_k = +\infty$ and $\sum_{k\geq 0} \epsilon_k^2 < +\infty$
$k \leftarrow 0$
while *Convergence of* θ **do**
  $k \leftarrow k + 1$
  Stochastic step $\mathbf{z}^{(k)} \sim p(\mathbf{z} \mid \mathbf{y}; \theta^{(k)})$
  Approximation step : Compute
   $Q_k(\theta) = Q_{k-1}(\theta) + \epsilon_k(\log p(\mathbf{y}, \mathbf{z}^{(k)}; \theta) - Q_{(k-1)})$
  Maximization step : Update $\theta^{(k+1)} = \text{argmax}_\theta \, Q_k(\theta)$
end
Result: return $\theta^{(k)}$

The SAEM algorithm offers computational cost savings by requiring only one sample, $\mathbf{z}^{(k)}$, per iteration, rather than using Monte Carlo estimate for $\mathbb{E}_{p(\mathbf{z}|\mathbf{y};\theta^{(k)})}[\log p(\mathbf{y}, \mathbf{z}; \theta)]$. During the early phase of the Approximation step, the algorithm focuses on maximizing $Q_k(\theta) = \log p(\mathbf{y}, \mathbf{z}^{(k)}; \theta)$ without recording previous values of $Q_k(\theta)$. This memoryless period is referred to as the *burn-in* phase and is critical to the algorithm's convergence. After this phase, the parameters $\theta^{(k+1)}$ are updated based on the current and previous values of $\theta^{(k)}$ and $\theta^{(k-1)}$, respectively.

A key challenge in the SAEM algorithm is that $p(\mathbf{z} \mid \mathbf{y}; \theta^{(k)})$ may be unknown. Using Bayes' rule, we can derive an expression for this probability distribution as

$$p(\mathbf{z} \mid \mathbf{y}; \theta^{(k)}) = \frac{p(\mathbf{y} \mid \mathbf{z}; \theta^{(k)}) p(\mathbf{z}; \theta^{(k)})}{p(\mathbf{y}; \theta^{(k)})} = \frac{p(\mathbf{y} \mid \mathbf{z}; \theta^{(k)}) p(\mathbf{z}; \theta^{(k)})}{\int p(y \mid \mathbf{z}; \theta^{(k)}) p(\mathbf{z}; \theta^{(k)}) \mathrm{d}z}$$

Here, $p(\mathbf{y} \mid \mathbf{z}; \theta^{(k)})$ is the generative model, $p(\mathbf{z}, \theta^{(k)})$ is the prior of the hidden variables, and the denominator is a constant. As a result, $p(\mathbf{z} \mid \mathbf{y}; \theta^{(k)})$ is only known up to the normalizing constant. To approximate the sample $\mathbf{z}^{(k+1)} \sim p(\mathbf{z}^{(k)} \mid \mathbf{y}; \theta^{(k)})$ when the normalizing constant is unknown, we use the Hasting Metropolis algorithm presented in Algorithm 3. This algorithm relies on a Markov Chain method to draw random variables $\mathbf{z}^{(k)}$ from the probability distribution $p(\mathbf{z}^{(k)} \mid \mathbf{y}; \theta^{(k)})$. The complete procedure is then referred to as the **Monte Carlo Markov Chain SAEM** (MCMC-SAEM) algorithm. It is proven to converge to a critical point of the observed likelihood $p(\mathbf{y}|\theta)$ [Kuhn and Lavielle, 2004, Allassonnière et al., 2010] that is likely to be a local maximum, as the randomness of the algorithm makes it diverge from saddle points.

Algorithm 3: Hasting Metropolis algorithm

Given $\mathbf{z}^{(k)}$

begin

Choose a proposition law $q_k(\cdot \mid \mathbf{z}^{(k)})$

Draw $\mathbf{z}^c \sim q_k(\cdot \mid \mathbf{z}^{(k)})$

Update $\mathbf{z}^{(k+1)} = \mathbf{z}^c$ with probability $\tau = \min\left(\frac{p(\mathbf{z}^c|\mathbf{y};\theta^{(k)}) q_k(\mathbf{z}^{(k)}|\mathbf{z}^c)}{p(\mathbf{z}^{(k)}|\mathbf{y};\theta^{(k)}) q_k(\mathbf{z}^c|\mathbf{z}^{(k)})}, 1\right)$

($\mathbf{z}^{(k+1)} = \mathbf{z}^{(k)}$ otherwise)

end

Result: return $\mathbf{z}^{(k+1)}$

In practice, sufficient statistics are introduced to update the model parameters at each step (see [Kuhn and Lavielle, 2005, Allassonnière et al., 2010]), and a Gibbs sampler allows to sample random variables successively along each dimension, but these aspects of the estimation procedures are not relevant to our work.

In Chapter 2 and 3, we introduce a variant of the MCMC-SAEM procedure, while the other chapters use the estimation procedure exactly as developed for Disease course mapping [Schiratti et al., 2017, Koval, 2020].

1.3.2 Calibration

Coming back to the context of DPMs, θ can be thought of as the collection of the fixed-effects (both geometric fixed-effects that describe the average geodesic, and the variances of the random-effects), and $\mathbf{z}$ the individual random-effects (spatial and temporal). Given $\mathbf{y}$, the *calibration* step aims to estimate the parameters $\hat{\theta}$ that best describe the group-average spatio-temporal trajectories and its variability. Since the hidden variables $\mathbf{z}$ are unknown, the observed likelihood is intractable and the MCMC-SAEM algorithm is used to estimate $\hat{\theta} = \arg\max p(\mathbf{y}, \theta)$. We refer the reader to [Kuhn and Lavielle, 2004, Allassonnière et al., 2010, Koval, 2020] for considerations on the proof of convergence and identifiability properties of the estimated models.

1.3.3 Personalization

Given $\hat{\theta}$, the *personalization* step aims to estimate the individual realisations of the random variables $\mathbf{z}_i^*$ that derive the parallel curves' bundle to reconstruct individual measurements. Considering k_i measurements of a patient $\mathbf{y}_i = (\mathbf{y}_{ij}, t_{i,j})_{1 \leq j \leq k_i}$, the optimal random-effect

value $\mathbf{z}_i^*$ maximizes the likelihood $p(\mathbf{y}_i, \mathbf{z}_i, \hat{\theta}) = p(\mathbf{y}_i|\mathbf{z}_i, \hat{\theta})p(\mathbf{z}_i, \hat{\theta})$. This optimization procedure can be realized by quasi Newton's methods such as L-BFGS or Powell's algorithm. Both numerical schemes find $\mathbf{z}_i^*$ without an analytical expression of the gradient of the likelihood with respect to $\mathbf{z}_i$. The personalization step allows finding the individual parameters for any subject, whether they were included in the data used for calibration or not.

1.3.4 Reconstruction, missing data imputation and future prediction

Given, $\hat{\theta}, \mathbf{y}_i$ and $\mathbf{z}_i$, the model predicts a trajectory $t \mapsto f(\hat{\theta}, \mathbf{z}_i^*, t)$, which corresponds to $\eta^{\mathbf{w}_i}(\psi_i(t))$.

When that predictive model is used for $t \in \{t_{i,1}, \cdots, t_{i,k_i}\}$, the resulting $\tilde{\mathbf{y}}_{ij} = f(\hat{\theta}, \mathbf{z}_i^*, t_{ij})$ is called the *reconstruction* of the data $\mathbf{y}_{ij}$, and the difference $||\tilde{\mathbf{y}}_{ij} - y_{ij}||$ the *reconstruction error*.

If t is strictly between two time points t_{ij}, $\tilde{\mathbf{y}} = f(\hat{\theta}, \mathbf{z}_i^*, t)$ is an interpolated value and corresponds to *missing values imputation*. This is particularly useful in the context of medical data, when visits are often irregularly spaced.

Lastly, if $t > t_{ik_i}$, the estimated $\tilde{\mathbf{y}}$ is the *prediction* of the measurement at this future time point. This prediction can only be accurate if the average scenario actually describes long-term progression. For this reason, if the training dataset does not feature patients at an advanced disease stage, predicting long-term progression is unrealistic.

1.3.5 Validation

This framework was introduced by [Schiratti et al., 2015b, Schiratti et al., 2017] and was used to described cortical thinning in AD [Koval et al., 2018], hippocampal atrophy [Bône et al., 2018], brain hypometabolism and cognitive decline [Koval et al., 2021b]. It has been demonstrated to reach test-retest levels of noise for these tasks, hinting that the models could not be improved without overfitting. Besides, several applications have been proposed, such as for Parkinson's Disease [Couronné, 2021] and Huntington's Disease [Koval et al., 2020]. Using these models as prognostic tools and disease staging allows reducing sample sizes in clinical trials for AD [Maheux et al., 2023] and Huntington's Disease [Koval et al., 2022], by selecting only the subjects that are likely to progress over the course of the trial.

Chapter 2

Riemannian metric learning

Explicit descriptions of the progression of biomarkers across time usually involve priors on the shapes of the trajectories. To circumvent this limitation, we propose a geometric framework to learn a manifold representation of longitudinal data. Namely, we introduce a family of Riemannian metrics that span a set of curves defined as parallel variations around a main geodesic, and apply that framework to disease progression modelling with a mixed-effects model, where the main geodesic represents the average progression of biomarkers and parallel curves describe the individual trajectories. Learning the metric from the data allows fitting the model to longitudinal datasets and provides few interpretable parameters that characterize both the group-average trajectory and individual progression profiles. Our method outperforms the 56 methods benchmarked in the TADPOLE challenge for cognitive scores prediction. This work was published and presented at the International Symposium on Biomedical Imaging (ISBI) 2022.

Contents

2.1 Introduction

2.1.1 Motivation

Modelling progressive diseases plays a crucial role in the development of new treatments. For a given individual, the progression of biological phenomenons can be measured by several biomarkers across time, and the collection of these observations allows deriving a progression model that helps us to both understand the average behaviour of the features, and how each individual compares to this reference scenario, allowing to make assumptions about their future progression.

Longitudinal databases are well suited for the study of neurodegenerative diseases, and are usually multimodal as they include imaging scans and a collection of clinical observations. However, in order to remain in low dimension, we focus on the scalar parameters that can be easily extracted from the other types of data, such as brain regions volumes, white matter density map and cortical thicknesses from MRI scans.

2.1.2 Related work

One approach to construct disease progression models for biomarkers is to formulate the changes using differential equations [Ito et al., 2011, Samtani et al., 2012, Abi Nader et al., 2021], with the caveat that covariates and all sources of random variations should be integrated in the differential equation system, which makes interpretation difficult. Another approach called event-based models is proposed in [Fonteijn et al., 2012, Young et al., 2014], in which cut-off points of abnormality are inferred from observed biomarkers, and disease stage is mapped to a discrete set of biomarker-abnormality events. While providing good robustness, the dichotomization of variables can be ill-suited for continuous biomarkers. Lately, recurrent neural networks have also been used to model the progression of scalar [Nguyen et al., 2020, Louis, 2019], imaging [Cui et al., 2019] and multi-modal data [Couronné et al., 2019, Louis et al., 2019] offering a flexible framework for regression.

A more interpretable approach is proposed with mixed-effects models[Laird and Ware, 1982], which account for both the average trajectory of the population, called the fixed-effects, and individual variations to that trajectory that account for inter-subjects variability, called the random-effects. The inter-subject variability can be thought of as the combined effect of the pace at which individuals evolve, and of their intrinsic biological characteristics. That variability can be modelled by a time reparametrization to "align" patients on a common pathological timeline and spaceshifts that define how each patient compares to the average trajectory. Early models used linear modelling [Verbeke, 1997, Cnaan et al., 1997] while, later, non-linearities were added with polynomial [Wu and Zhang, 2002], logistic [Jedynak et al., 2012], exponential [Raket, 2020] and semi-parametric regressions [Donohue et al., 2014].

To model the progression of a disease, one less restrictive assumption is to consider that the observed biomarkers follow continuous trajectories in the space of observations that is assumed to be a Riemannian manifold [Schiratti et al., 2015b]. Particular cases have been derived for the analysis of longitudinal scalar measurements in [Koval et al., 2017b, Schiratti et al., 2017], where this model is studied with a fixed metric that is known *a priori* to yield logistic trajectories. Besides, for multivariate data, the metric is usually set to be a product metric, so that the average trajectory is estimated as a product of independent 1-dimensionnal trajectories.

2.1.3 Contributions

In this work we extend the geometric approach, in which the data is embedded in a Riemannian manifold, by losing the *a priori* on the metric. We introduce a family of parametrized Riemannian metrics and propose a trajectory model that defines, for each metric, a set of curves that are composed of variations around a reference geodesic of the manifold. We then illustrate how this set of curves can be used for disease progression models through a statistical mixed-effects model where the reference geodesic represents the average progression of the observed biomarkers through time, and the parallel curves describe the trajectory of each patient. The metric can then be estimated to fit the model to longitudinal datasets and yield few interpretable parameters that describe both the average trajectory and how the individual progressions relate to this reference evolution. Namely, the main benefits of learning the metric are that :

- the trajectories can take a wide variety of shapes and provide better understanding of the qualitative progression and asymptotic evolution of biomarkers,
- the multivariate metric is not simplified to a product metric, so the parallel curves can differ in shapes from the average trajectory.

The first benefit circumvents most priors on the shape of trajectories, either logistic, linear or even exponential, as it allows plateaux and multiple inflection points and does not require monotonicity. The second benefit can, for instance, allow the model to use the spatial variability to separate subjects with different subtypes of the disease, and each region of the manifold could display adapted trajectories. Provided enough data, *a posteriori* analysis of the spatio-temporal parameters can provide insights into the variability in disease progression.

2.2 Trajectory model

2.2.1 A general metric

The set of possible observation points $\mathbf{M}$ is endowed with a Riemannian manifold structure within the total normalized observation space $[0,1]^N$. As suggested in [Louis, 2019] we consider the family of Riemannian metrics on this manifold, which are defined via their cometrics (inverse of the metric) as a Gaussian interpolation on a set of control points $(_i)_{1\leq i\leq N_\delta} \in \mathbf{M}^{N_\delta}$,

$$\forall \mathbf{p} \in \mathbf{M}, \mathrm{G}^{-1}(\mathbf{p}) = \sum_{i=1}^{N_\delta} L_i^T L_i \exp\left(\frac{||\mathbf{p} -_i ||^2}{\delta^2}\right)$$

where δ, the kernel width, decides how coarse our metric is and is the only hyperparameter to tune. $(L_i)_{1\leq i\leq N_\delta}$ is a family of upper-triangular matrices, such that $(L_i^T L_i)_{1\leq i\leq N_\delta}$ constitutes a family of definite positive matrices (by Cholesky's decomposition theorem) that approximate the inverse metric at each control point. We use a grid of regularly spaced control-points that are δ apart in every direction so that we have $N_\delta = (\lfloor \delta^{-1} \rfloor + 1)^N$. We choose to parametrize the inverse metric rather than the metric itself as it appears in the Hamiltonian equations. For a position $\mathbf{p} \in \mathbf{M}$ and a velocity $\mathbf{v} \in T_p\mathbf{M}$ tangent to $\mathbf{M}$ at $\mathbf{p}$, we introduce the Hamiltonian $\mathrm{H}(\mathbf{p}, \mathbf{v}) := \frac{1}{2}\mathbf{v}^T \mathrm{G}^{-1}(\mathbf{p})\mathbf{v}$ and obtain the following equations for geodesics

$$\dot{\mathbf{p}} = \frac{\partial \mathrm{H}}{\partial \mathbf{v}} = \mathrm{G}^{-1}(\mathbf{p})\mathbf{v} \quad \text{and} \quad \dot{\mathbf{v}} = -\frac{\partial \mathrm{H}}{\partial \mathbf{p}} = -\frac{1}{2}\mathbf{v}^T \frac{\partial \mathrm{G}^{-1}(\mathbf{p})}{\partial \mathbf{p}} v$$

2.2.2 Reference geodesic and parallel curves

For any metric, given a reference geodesic $\gamma : t \in I \mapsto \gamma(t)$ with $I \subset \mathbb{R}$, we can define a family of trajectories called the Exp-parallel curves. A vector field X is said to be *parallel* along a curve $\gamma : I \to \mathbb{M}$ if $\nabla_{\gamma(t)} X = 0$ where ∇ is the connection on $\mathbf{M}$, and for $t_0 \in I$ and $\mathbf{w}_0 \in T_{\gamma(t_0)}\mathbf{M}$, one can show there exists a unique vector field w parallel along γ such that $w(t) = \mathbf{w}_0$. Thus, given a vector $\mathbf{w}_0$, called a spaceshift, we can define the parallel curve $\eta^w : t \mapsto \mathrm{Exp}(w(t))$ where Exp is the exponential map on $\mathbb{M}$, that maps the vectors of the tangent bundle to the manifold.

Learning the metric allows obtaining a wide variety of trajectories for the reference geodesic, and allows the spaceshifts to impact the shape of the parallel curves, as opposed to the fixed product metric where the spaceshift only changes the timing and ordering of feature progression.

2.3 Statistical model

In practice, we consider the repeated scalar observations of p individuals, such that the i-th individual has been observed $k_i \in \mathbb{N}^*$ times at times $t_{i,1} < \cdots < t_{i,k_i}$. The observation at time $t_{i,j}$ is denoted $y_{i,j} \in \mathbf{M} \subset [0,1]^N$.

2.3.1 Generative statistical model

The average trajectory of the population can be entirely characterised by two parameters, $\mathbf{p}_0 \in \mathbf{M}$ and $\mathbf{v}_0 \in T_{\mathbf{p}_0}(\mathbf{M})$ that are respectively a position in the observation space and a velocity in the tangent plane at this point. If we add a reference time t_0, this defines a unique geodesic $\gamma_0 : t \in \mathbb{R} \mapsto \gamma_0(t) \in \mathbf{M}$ such that $\gamma_0(t_0) = \mathbf{p}_0$ and $\dot{\gamma}(t_0) = \mathbf{v}_0$. We recall Fig. 2.1 which summarizes this geometric description.

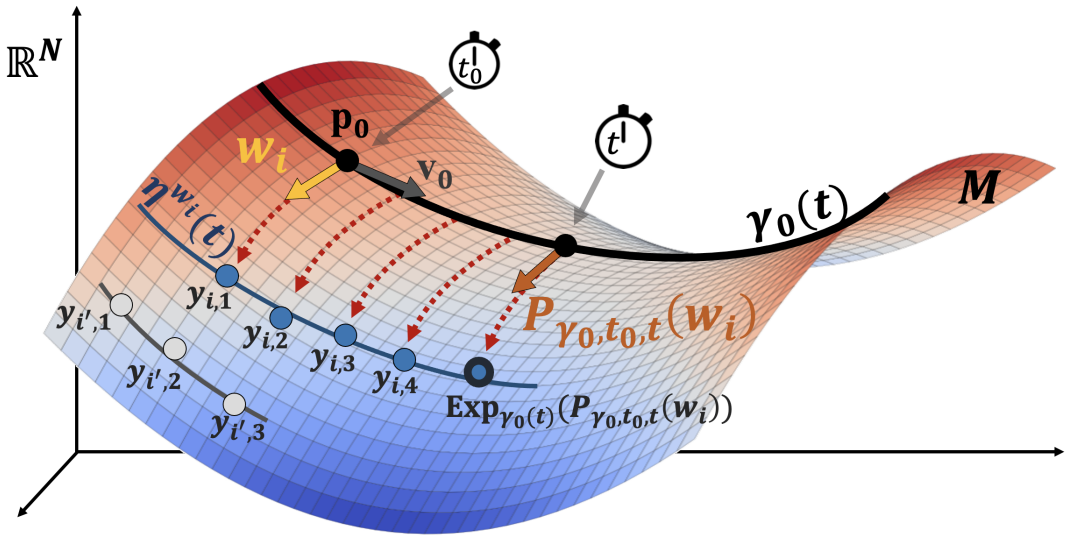


Figure 2.1: Visual depiction of the geometric model. The black curve is the group-average geodesic γ_0, characterized by $\mathbf{p}_0$, $\mathbf{v}_0$ and t_0, and the blue and grey curves are the parallel curves that best match the observations. More detailed are provided in Section 1.2.

As in [Schiratti et al., 2015b], each patient is described by a couple of individual parameters $\alpha_i \in \mathbb{R}_+^*$ and $\tau_i \in \mathbb{R}$ that align the individual timeline to the reference timeline via the affine time-warp function $\psi_i : t \mapsto \alpha_i(t - \tau_i) + t_0$, as well as a spaceshift $\mathbf{w}_i \in T_{\mathbf{p}_0}(\mathbf{M}) \subset \mathbb{R}^N$. τ_i is the onset age, α_i represents the pace at which the patient evolves and the spaceshift defines which Exp-parallel curve represents the progression of the patient. For convenience purposes, we write the acceleration factor $\alpha_i = \exp(\xi_i)$ with $\xi_i \in \mathbb{R}$. For every subject i, given both ψ_i and w_i, we can now model the scalar measurements $(y_{i,j})_{1 \leq j \leq k_i}$ as the sample points at times $(\psi_i(t_{i,j}))_{1 \leq j \leq k_i}$ with additional noise ε

$$y_{i,j} = \eta_{\gamma_0,\mathrm{G}}^{w_i}(\psi_i(t_{i,j})) + \varepsilon_{i,j}$$

The only difference with the general formulation in Section 1.2 is that the metric G and the corresponding reference geodesic explicitly parametrize the curves when the traditional "Disease course mapping" model is usually only used for a set metric and the curves take closed form expressions.

2.3.2 Mixed-effects formulation

We choose a Gaussian noise $\varepsilon_{i,j} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. Then, the previously introduced latent variables are modelled as random-effects $z = (\xi, \tau, s)$, with $\xi_i \sim \mathcal{N}(0, \sigma_\xi^2)$, $\tau_i \sim \mathcal{N}(t_0, \sigma_\tau^2)$ and $s_i \sim \mathcal{N}(0, 1)$. The source $s_i \in \mathbb{R}^{N_s}$ reconstructs w_i through $w_i = As_i$, where the "mixing matrix" $A \in \mathbb{R}^{N \times N_s}$ is also estimated and serves a purpose of dimensionality reduction. Then, the fixed-effects are written $\theta = (\mathbf{p}_0, \mathbf{v}_0, t_0, A, (L_i)_{1 \leq i \leq N_\delta}, \sigma_\varepsilon, \sigma_\xi, \sigma_\tau)$.

2.4 Estimation

We proceed to Maximum A Posteriori (MAP) estimation of the parameters. The Expectation (E) step from a regular Expectation Maximisation (EM) algorithm would require the computation of intractable integrals for the likelihood, so we resort to the Stochastic Approximation EM (SAEM) in which the E step is replaced by a Simulation (S) and Approximation (A) step. For curved Exp-parallel models, the S step can be replaced by a single transition of an ergodic Monte-Carlo Markov Chain (MCMC) whose stationary distribution is estimated iteratively using a Metropolis-Hastings sampler. This global algorithm is called the Monte-Carlo Markov Chain Stochastic Approximation Expectation-Maximization (MCMC-SAEM) [Kuhn and Lavielle, 2004, Kuhn and Lavielle, 2005], and is proven to converge towards a global maximum of the averaged likelihood

$$\theta \mapsto p(y|\theta) = \int q(y, z|\theta) \mathrm{d}z$$

We refer the reader to [Schiratti et al., 2015b] for an extensive presentation of the estimation algorithm for non-linear mixed-effects models. The specificity of our approach resides in the additional (L_i) parameters for the metric, which are estimated via a line-search gradient descent between each iteration of the MCMC-SAEM procedure. Besides, there is no closed-form expression for γ_0, so we integrate the Hamiltonian differential system using a Runge-Kutta numerical scheme, and the parallel transport of the spaceshifts is computed using the Fanning scheme [Louis et al., 2018].

2.5 Experiments and results

2.5.1 Synthetic data

To validate this approach, we simulated 2-dimensional data according to trajectories that are known to occur naturally, namely a logistic shape with asymptotes other than 0 and 1, and a sum of logistics that represents features with multiple inflection points. We engineered a metric that yields these trajectories as geodesics by fitting our model to a family of shifted and accelerated versions of these shapes. With this metric, we generated a realistic dataset of 200 patients with an average follow-up of 10 visits spanned over 5 years. We choose $\sigma_\varepsilon = 0.02, \sigma_\tau = 15, \sigma_\xi = 0.5, t_0 = 70$ and sampled the individual parameters, then used the engineered geodesic $(\mathbf{v}_0, \mathbf{p}_0)$ and parameters $((L_i), A)$ for the generative model $y_{i,j} = \eta^{w_i}_{\gamma_0,\mathrm{G}}(\psi_i(t_{i,j})) + \varepsilon_{i,j}$ with random age at baseline.

Metric	$\hat{\sigma_\varepsilon}$ (RMSE)	$\hat{t_0}$	$\lvert\hat{\xi} - \xi\rvert_{avg}$	$\lvert\hat{\tau} - \tau\rvert_{avg}$
Linear	.022 ±.002	71 ±1.1	.40 ±.08	5.1 ±1.1
Logistic	.023 ±.001	70 ±1.0	.38 ±.05	8.1 ±0.9
Learned	**.020 ± .001**	70 ±1.1	**.22 ± .08**	**2.0 ± 0.8**

Table 2.1: Results over 5 estimations. Errors on τ and ξ are averaged across patients.

Our model recovered both the fixed effects and the individual parameters more accurately than their "fixed metric" counterparts (Table 1.). Reconstruction error is almost perfect for the fixed-metric models, but the individual parameters are not correctly estimated because the average trajectory does not reflect the dynamics of the data. Our approach is expected to better recover the model parameters for any data that can be modelled with a mixed-effect model, without prior knowledge, at the expense of additional computational cost.

Fig. 2.2 illustrates the trajectories modelled when learning the metric (right), compared to a logistic metric (left) where $\mathbf{M} \subset [0,1]^2$ and $t \in I$ represents the pathological time.

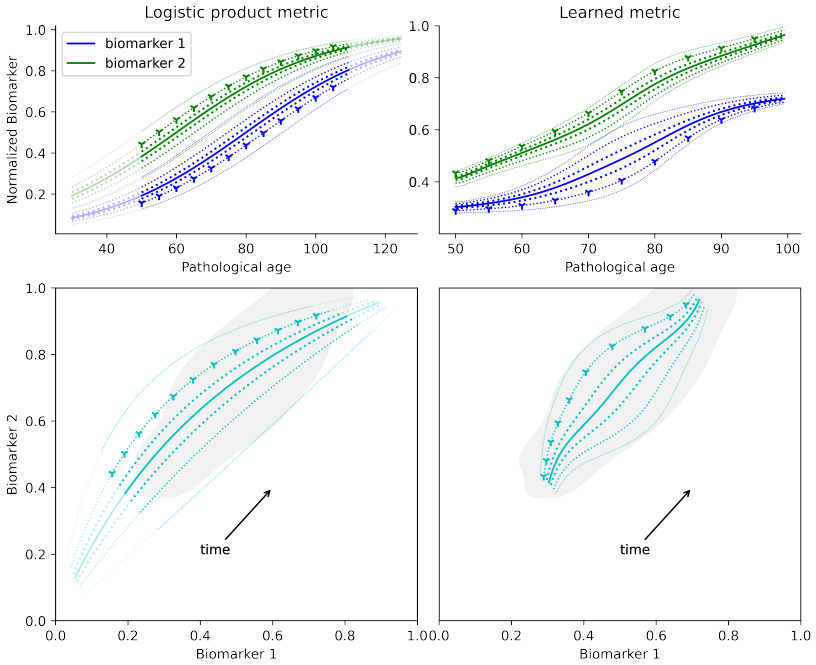


Figure 2.2: Top panels show the progression of the features across time for the main geodesic (plain) and its Exp-parallel curves (dotted). Bottom panels show these trajectories in $\mathbf{M}$, with implicit time, and a kernel density estimation (80% threshold) of the observed data in shown in the grey shaded area. This illustrates the fact that even for normalized biomarkers between 0 and 1, the true distribution of the data do not necessarily populate the entire space, and the asymptotes at 0 and 1 may not be relevant at all since very few data actually reach that part of the space. One parallel curve is highlighted with crosses to display the effect of a spaceshift. Logistic trajectories are extrapolated (shaded) beyond the data points to exhibit the forced 0 and 1 asymptotes.

2.5.2 Real life data

We applied that approach to data from the Alzheimer Disease Neuroimaging Initiative (ADNI) to model three features that are known to be good markers of the decline through Alzheimer's disease : Hippocampal volume, Ventricles volume and the Alzheimer's Disease Assessment Scale (ADAS) cognitive score. We used all patients with multiple visits for a total of 1,452 patients and 9,465 visits. All features are normalized.

Shape of the learned trajectories Fig. 2.3 displays the learned progression model for the chosen features. As expected, we recover a "sigmoid-like" trajectory for the cognitive score and "linear-like" trajectories for the imaging markers. Both our method and the logistic fixed metric outperform the 56 methods benchmarked in the TADPOLE challenge [Marinescu et al., 2020] with improved prediction of the ADAS score (respectively 3.73 and 3.66 MAE compared to 4.20 for the best documented method in TADPOLE). Since the reconstruction errors and prediction errors are not significantly different, it is interesting to question whether the estimated individual parameters offer greater quantitative insights with the learned metric.

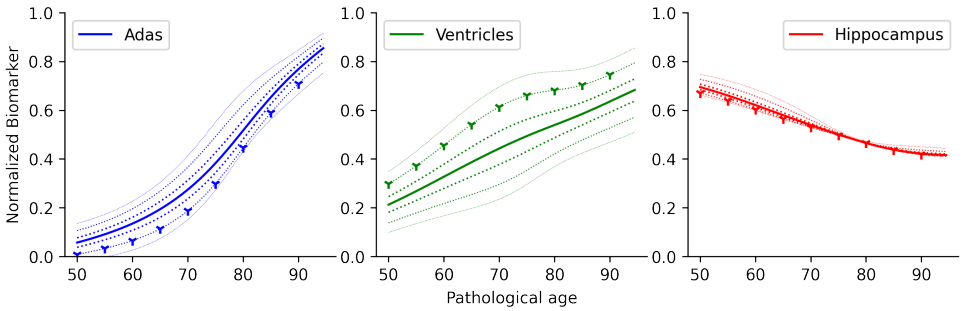


Figure 2.3: Normative scenario predicted by our model for the 3 features : main logistic (plain) and parallels (dotted). One parallel curve is highlighted with crosses.

Interpretability of the random-effects Fig.2.4 illustrates the distribution of individual parameters for patients diagnosed Cognitively Normal (CN), Mild Cognitive Impairment (MCI) and Alzheimer Disease (AD) on their last visit. Fitting a Gaussian Mixture Model (GMM) shows that the distribution for each diagnosis cluster are better separated with our approach. 20 repetitions of 5-fold validation with GMM for diagnosis prediction gave us an average **62.5%±0.8%** test accuracy (see the confusion matrix reported in Fig. 2.5) compared to 53.7%±1.2% for the logistic metric model.

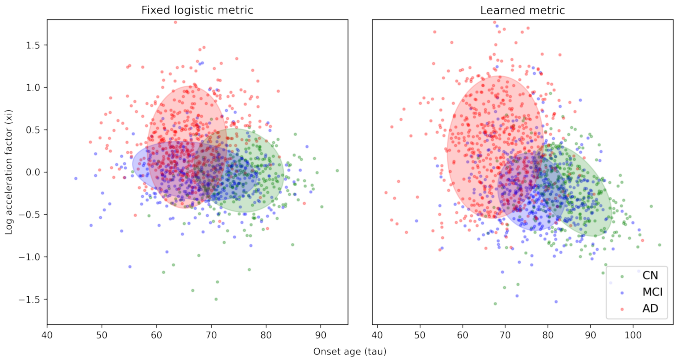


Figure 2.4: Estimated individual parameters for the logistic (left) and learned (right) metric with unsupervised Gaussian Mixture Model to predict diagnosis. As expected, the AD patients show earlier and steeper decline than the CN subjects, with the MCI cohort somewhat in the middle.

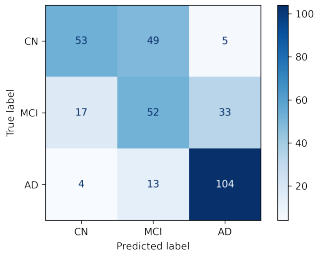


Figure 2.5: Confusion matrix (test set) for diagnosis prediction.

Considering only the last diagnosis available makes the labels overlap which means perfect accuracy is not reachable. However, the improvement of clustering abilities shows that the individual parameters from our approach offer more interpretability.

On the modelling of cognitive scores' progression To further the qualitative analysis of the shape of cognitive trajectories, we also modelled the progression of 4 cognitive scores extracted from the ADAS-cog scale that reflect concentration, memory, language and praxis abilities of subjects. The comparison between the logistic modelling and our approach is striking: the average progression is extremely similar. The temporalities of decline are different, since each model learns the timeline that best suits its spatio-temporal framework, but the ordering and qualitative shape of decline are almost identical. This solidifies the relevance of logistic modelling for cognitive scores.

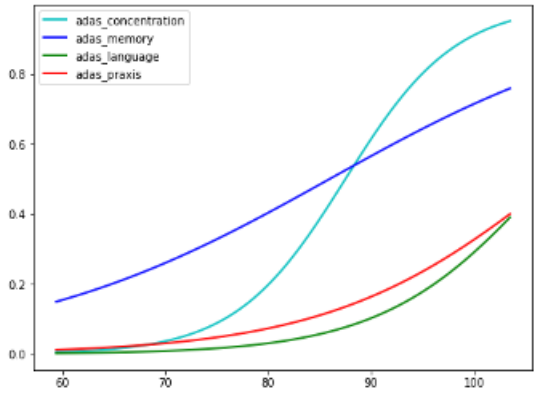


Figure 2.6: Sigmoid geodesics over time for a series of subscores of the ADAS-cog scale, that reflect different aspects of cognition.

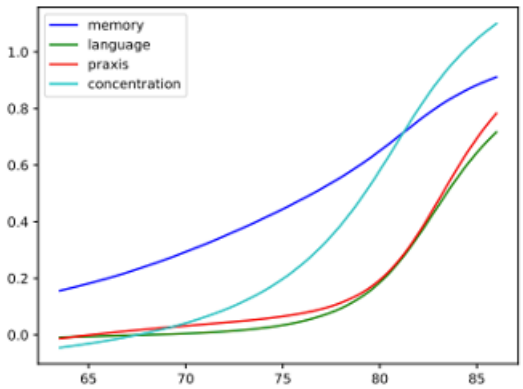


Figure 2.7: Learned geodesics using our Riemannian metric framework, for the same subscores. The curves are very similar.

2.6 Conclusion

In this work, we introduced an efficient geometric representation of longitudinal data to formulate explicit disease progression models with mixed effect modelling[1]. Learning the Riemannian metric allows describing with accuracy both the average trajectory of a biomarker and the individual progression profiles. Our approach reaches state-of-the-art reconstruction and prediction errors, like the fixed metric approaches, but offers greater interpretability.

[1]Implementation publicly available at `https://github.com/bsauty/riemannian_metric_learning`.

Part II

Progression Models for Imaging Data

Chapter 3

Longitudinal Variational Autoencoders

Mixed-effects models have been consistently used to model clinical assessments or biomarkers extracted from medical images, allowing missing data imputation and prediction at any timepoint. However, such progression models have seldom been used for entire medical images. In this work, a Variational Auto Encoder is coupled with a temporal linear mixed-effect model to learn a latent representation of the data such that individual trajectories follow straight lines over time and are characterised by a few interpretable parameters. A Monte Carlo estimator is devised to iteratively optimize the networks and the statistical model. We apply this method on a synthetic data set to illustrate the disentanglement between time dependant changes and inter-subjects variability, as well as the predictive capabilities of the method. We then apply it to 3D MRI and FDG-PET data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) to recover well documented patterns of structural and metabolic alterations of the brain over the course of AD. This work was published and presented at the international conference on Medical Image Computation and Computer-Assisted Intervention (MICCAI) 2022, and got an oral presentation at the 2022 International Symposium of Clinical Biostatistics (ISCB).

Contents

3.1 Introduction

Estimating progression models from the analysis of time dependent data is a challenging task that helps to uncover latent dynamics. For the study of neurodegenerative diseases, longitudinal databases have been assembled where a set of biomarkers (medical images, cognitive scores and covariates) are gathered for individuals across time. Understanding their temporal evolution is of crucial importance for early diagnosis and drug trials design, especially the imaging biomarkers that can reveal a silent prodromal phase.

In this context, several approaches have been proposed for the progression of scalar measurements such as clinical scores or volumes of brain structures [Fonteijn et al., 2012,

Schiratti et al., 2015a, Jedynak et al., 2012] or series of measurements across brain regions forming a network [Koval et al., 2021a, Bauer et al., 2011]. These approaches require the prior segmentation and extraction of the measurements from the images. Providing progression models for high dimensional structured data without prior processing is still a challenging task. The difficulty is to provide a low dimensional representation of the data, where each patient's trajectory admits a continuous parametrization over time. It should allow sampling at any time point, be resilient to irregularly spaced instances, and disentangle temporal alterations from the inter-patients variability.

3.1.1 Related work

When dealing with high dimensional data, it is often assumed that the data can be encoded into a low dimensional manifold where the distribution of the data is simple. Deep Generative Neural Networks such as Variational Auto Encoders (VAE) [Kingma and Welling, 2013] allow finding such embeddings. Several approaches have explored longitudinal modelling for images within the context of dimensionality reduction.

Recurrent Neural Networks (RNN) provide a straightforward way to extract information from sequential data. Convolutional networks with a recurrent structure have been used for diagnosis prediction using MRI [Cui et al., 2019] or PET [Liu et al., 2018] scans in Alzheimer's Disease (AD). The main caveat of these approaches is that the recurrent structure is highly sensible to the temporal spacing between instances, which is troublesome in the context of disease modelling, where visits are often missing or irregularly spaced.

Mixed-effects models provide an explicit description of the progression of each patient, allowing to sample at any timepoint. Through a time reparametrization, all patients are aligned on a common pathological timeline, and individual trajectories are parametrised as small variations (random effects) around a reference trajectory (fixed effects) that can be seen as the average scenario. Now considered a standard tool in longitudinal modelling [Jedynak et al., 2012, Bernal-Rusiel et al., 2013b, Schiratti et al., 2015a], mixed-effects models have yet been scarcely used for images within the context of dimension reduction. In [Louis et al., 2019], a RNN outputs the parameters of a mixed-effect model that describes patients' trajectories as straight lines in the latent space of a VAE across time.

Self supervised methods have proposed to alleviate the need for labels, in our case the age of the patients at each visit. In [Couronné et al., 2021], the encoder of a VAE learns a latent time variable and a latent spatial variable to disentangle the temporal progression from the patient's intrinsic characteristics. Similarly, in [Zhao et al., 2021], the encoder is penalized with a cosine loss that imposes one direction in the latent space that corresponds to an equivalent of time. Both these methods allow the model to learn a temporal progression that does not rely on the clinical age of the patients, offering potential for unlabelled data, at the cost of interpretability of the abstract timeline and the ability to sample at any given timepoint.

Gaussian processes have been proposed in order to provide a temporal structure. [Abi Nader et al., 2020] combined a dimensionality reduction approach using matrix factorization and monotonic Gaussian Process to model the progression of volumetric images. Deep learning architectures have also been proposed in order to endow the latent space with a temporal structure. Namely, Gaussian Process VAEs (GPVAE) [Casale et al., 2018] introduced a more general prior for the posterior distribution in the latent space, in the form of a Gaussian Process (GP) that depends on the age of the patients [Fortuin et al., 2020] as well as a series of covariates [Ashman et al., 2020, Ramchandran et al., 2021].

These approaches pose challenges as to the choice of parametrization for the Gaussian Process, and do not provide an expected trajectory for each patient.

Diffeomorphic methods provide progression models for images. The main approaches are based off of the geodesic regression framework [Niethammer et al., 2011, Banerjee et al., 2016] and allow learning a deformation map that models the effect of time on the images for a given subject. While providing high resolution predictions, these methods show limited predictive abilities further in time when compared to mixed-effects models, that aggregate information from all the subjects at different stages of the disease [Bône et al., 2017].

3.1.2 Contributions

In this context, we propose to endow the latent space of a VAE with a linear mixed-effect longitudinal model. While in [Louis et al., 2019], the networks predict the random effects from visits grouped by patients, we propose to enrich a regular VAE that maps each individual visits to a latent representation, with an additional longitudinal latent model that describes the progression of said representations over time. A novel Monte Carlo Markov Chain (MCMC) procedure to jointly estimate the VAE and the structure of its representation manifold is proposed. To sum up the contributions, we :

1. use the **entire 3D scan** without segmentation or parcellation to study relations across brain regions in an unsupervised manner,

2. proceed to **dimension reduction** using a convolutional VAE with the added constraint that latent representations must comply with the structure of a generative statistical model of the trajectories,

3. provide a **progression model** that disentangles temporal changes from changes due to inter-patients variability, and allows sampling patients' trajectories at any timepoint, to infer missing data or predict future progression,

4. demonstrate this method on a synthetic data set and on both **MRI** and **PET scans** from the Alzheimer's Disease Neuroimaging Initiative (ADNI), recovering known patterns in normal or pathological brain ageing.

3.2 Methodology

3.2.1 Representation learning with VAEs

Auto Encoders are a standard tool for non-linear dimensionality reduction, comprised of an *encoder* network q_ϕ that maps high dimensional data $x \in \mathcal{X}$ to $z \in \mathcal{Z}$, in a smaller space referred to as the *latent space*, and a *decoder* network $p_\theta : z \in \mathcal{Z} \mapsto \hat{x} \in \mathcal{X}$. VAEs[Kingma and Welling, 2013] offer a more regularized way to populate the latent space. Both encoder and decoder networks output variational distributions $q_\phi(z|x)$ and $p_\theta(x|z)$, chosen to be multivariate Gaussian distributions. Adding a prior $q(z)$, usually the unit Gaussian distribution $\mathcal{N}(0, \mathrm{I})$, on $\mathcal{Z}$ allows to derive a tractable Evidence Lower BOund for the log-likelihood $ELBO = \mathcal{L}_{recon} + \beta\mathcal{L}_{KL}$ where $\mathcal{L}_{recon}$ is the ℓ_2 reconstruction error, $\mathcal{L}_{KL}$ is the Kullback-Leibler (KL) divergence between the approximate posterior and the prior on the latent space and β balances reconstruction error and latent space regularity[Higgins et al., 2016].

3.2.2 Longitudinal statistical model

In this section, we propose a temporal latent variables model that encodes disease progression in the low-dimensional space $\mathcal{Z}$. Given a family of observations from N patients $\{x_{i,j}\}_{1\leq i\leq N}$, each observed at ages $t_{i,j}$ for $1 \leq j \leq n_i$ visits, and their latent representations $\{z_{i,j}\}$, we define a statistical generative model with

$$z_{i,j} = p_0 + \left[e^{\xi_i}(t_{i,j} - \tau_i)\right] v_0 + w_i + \varepsilon_{i,j}$$

where e^{ξ_i} and τ_i, respectively the *acceleration factor* and the *onset age* of patient i, allow an affine time warp aligning all patients on a common pathological timeline, and $w_i \in \mathcal{Z}$ is the *space shift* that encodes inter-subjects variability, such as morphological variations across regions that are independent of the progression. These parameters position the individual trajectory with respect to the typical progression that is estimated at the population level. These three parameters form the **random effects** of the model ψ_r. Vectors w_i and v_0 need to be orthogonal in order to uniquely identify temporal and spatial variability.

We choose the Gaussian priors for the noise $\varepsilon_{i,j} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and random effects $\tau_i \sim \mathcal{N}(t_0, \sigma_\tau^2)$, $\xi_i \sim \mathcal{N}(0, \sigma_\xi^2)$ and $w_i \sim \mathcal{N}(0, \mathrm{I})$. The parameters $p_0 \in \mathcal{Z}$, $v_0 \in \mathcal{Z}$, $t_0 \in \mathbb{R}$ are respectively a reference *position*, *velocity* and *time* and describe the average trajectory. Together with the variances $\sigma_\varepsilon, \sigma_\tau, \sigma_\xi$, they form the **fixed-effects** of the model ψ_f. We note $\psi = (\psi_r, \psi_f)$.

3.2.3 Longitudinal VAE

We combine dimension reduction using a regular β-VAE and the aforementioned statistical model to add a temporal structure to the latent space. To do so, we consider a composite loss that accounts for both the VAE loss and the goodness-of-fit of the mixed-effect model:

$$\mathcal{L} = \mathcal{L}_{recon} + \beta\mathcal{L}_{KL} + \gamma\mathcal{L}_{align} \text{ where } \begin{cases} \mathcal{L}_{recon} &= \sum_{i,j} ||x_{i,j} - \hat{x}_{i,j}||^2 \\ \mathcal{L}_{KL} &= \sum_{i,j} KL(q_\phi(z|x_{i,j})||\mathcal{N}(0, \mathrm{I})) \\ \mathcal{L}_{align} &= \sum_{i,j} ||z_{i,j} - \eta^i_\psi(t_{i,j})||^2 \end{cases}$$

where $z_{i,j}$ and $\hat{x}_{i,j}$ are the modes of $q_\phi(x_{i,j})$ and $p_\theta(z_{i,j})$, and $\eta^i_\psi(t_{i,j}) = p_0 + \left[e^{\xi_i}(t_{i,j} - \tau_i)\right] v_0 + w_i$ is the expected position of the latent representation according to the longitudinal model and γ balances the penalty for not aligning latent representations with the linear model. Since the loss is invariant to rotation in $\mathcal{Z}$, we set $p_0 = 0$ and $v_0 = (1, 0, \cdots, 0)$.

Since $\mathcal{L}_{align}$ is a ℓ_2 loss in the latent space, it can be seen as the log-likelihood of a Gaussian prior $z_{i,j} \sim \mathcal{N}(\eta^i_\psi(t_{i,j}), \mathrm{I})$ in the latent space, which defines an elementary Gaussian Process, and which supports the addition of GP priors in the latent space of VAEs to model longitudinal data [Fortuin et al., 2020, Ramchandran et al., 2021]. Besides, $\mathcal{X}$ can be seen as a Riemannian manifold, the metric of which is given by the pushforward of the Euclidean metric of $\mathcal{Z}$ through the decoder, such that trajectories in $\mathcal{X}$ are geodesics, in accordance with the Riemannian modeling of longitudinal data (see Chapter 4). The metric on $\mathcal{X}$ thus allows to recover non-linear dynamics, as is often the case for biomarkers. Our approach thus bridges the gap between the deep learning approach to longitudinal data and a natural generalization of well studied disease progression models to images.

Network implementation and estimation. Both the encoder and decoder are chosen to be vanilla convolutional Neural Networks (4 layers of Convolution with stride, BatchNorm and ReLU for the encoder and the transposed network architecture for the decoder) with a dense layer towards the latent space, as described in Fig. 3.1. The implementation

is in PyTorch and available at `https://github.com/bsauty/longitudinal-VAEs`.

Algorithm 4: Monte Carlo estimation of the Longitudinal VAE

Input : Longitudinal visits $\{(x_{i,j}, t_{i,j})\}$ and hyperparameters β and γ.
Output: Estimation of (ϕ, θ) for the VAE and ψ for the temporal model.
Init : Initialize (ϕ, θ) as a regular β-VAE.
Set $k = 0$ and $z^0 = q_\phi(x)$
while *not converged* **do**
  Simulation
    Draw candidates $\psi_r^c \sim p(.|\psi_f^k)$ `// Sampling with prior` $p(.|\psi_f^k)$
    $\forall i, j$ compute $\eta^i_{\psi_k}(t_{i,j})$ `// Expected latent trajectories`
    Compute likelihood ratio $\omega = \min\left(1, \frac{q(\psi_r^c|z^k, \psi_f^k)}{q(\psi_r^k|z^k, \psi_f^k)}\right)$
    if $u \sim \mathcal{U}(0,1) > \omega$ **then** $\psi_r^{k+1}\psi_r^c$ **else** $\psi_r^{k+1}\psi_r^k$
  Approximation Compute sufficient statistics S^k for ψ_r^k
  Maximisation $\psi_f^{k+1} \leftarrow \psi_f^*(S^k)$
  VAE optimization Run one epoch using $\mathcal{L}$ with the target latent representation $\eta^i_{\psi^k}$ for $\mathcal{L}_{align}$ and update $z^{k+1} \leftarrow q_\phi(x)$
end

The difficulty lies in the joint estimation of (θ, ϕ) and ψ, which are co-dependant since $\mathcal{L}_{align}$ depends on η^i_ψ and ψ depends on the encoded representation $z = q_\phi(x)$. Given $\{z_{i,j}\}$, we can proceed to a Maximum a Posteriori estimation of ψ with the MCMC-SAEM procedure in which the estimation step of an EM algorithm is replaced by a stochastic approximation. See [Kuhn and Lavielle, 2004, Allassonnière et al., 2010] for details. Given the target trajectories $\{\eta^i_\psi\}$, the weights from both networks of the VAE are optimized through backpropagation using $\mathcal{L}$ with an optimizer with randomized batches. Both estimation schemes are iterative, so we designed a Monte Carlo estimator for (ϕ, θ, ψ), presented in Algorithm 1, alternating between both schemes.

Once calibrated with a training set, we freeze the VAE parameters (ϕ, θ) and fixed effects ψ_f, and learn the individual parameters ψ_r, via gradient descent of the likelihood, to personalize the model for new subjects.

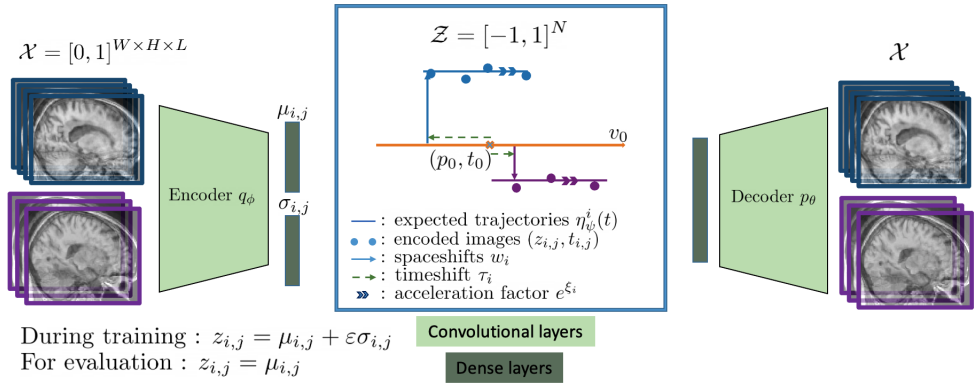


Figure 3.1: Images $\{x_{i,j}\}$ are encoded into $\mathcal{Z}$ such that the $\{z_{i,j}\}$ are close to the estimated latent trajectories. Individual trajectories (straight lines) are parametrized with w_i,τ_i and e^{ξ_i} as variations around the reference trajectory (orange arrow).

Hyperparameters are $\dim(\mathcal{Z})$, which should be small enough to allow the mixed-effect model to be interpretable but big enough to reach good reconstruction; β, which should minimize overfitting while not impairing reconstruction quality; and γ, which should also not be too big to avoid loosing contextual information in $\mathcal{Z}$. These parameters were set using grid search. Lastly, the MCMC-SAEM is computationally inexpensive compared to backpropagation, so memory footprint and runtime are similar to training a regular VAE. All training was performed with a Quadro RTX4000 8Go GPU.

3.3 Experiments and results

3.3.1 Results on synthetic experiments

We first validated our approach on a synthetic data set of images of silhouettes of dimension 64x64 [Couronné and Vernhet, 2021]. Over time, the silhouette raises its left arm. Different silhouettes are generated by varying the relative position of the three other limbs, all of them raising their left arm in time. The motion is modulated by varying the time stamp at which the motion starts and the pace of motion. This is done using an affine reparametrization of the time stamp $t_{i,j}$ of the silhouette with Gaussian log-acceleration factor ξ_i and onset age τ_i. This data set contains $N = 1,000$ subjects with $n = 10$ visits each, sampled at random time-points.

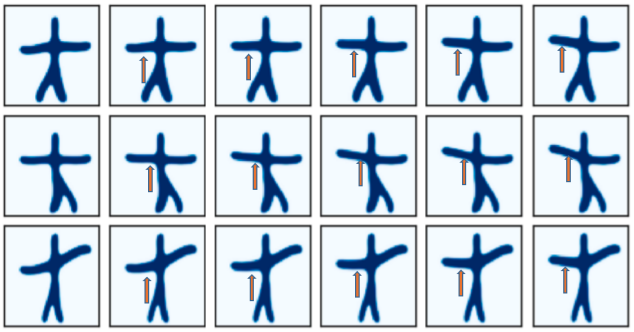


Figure 3.2: Each row represents an individual that evolves over time from left to right.

We choose $\dim(\mathcal{Z}) = 4$ to evaluate the ability of our model to isolate temporal changes (motion of the left arm) from the independent spatial generative factors (the position of the other 3 limbs).

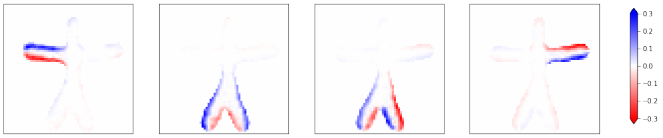


Figure 3.3: Gradients of the image at $p_0 = 0$ in the 4 directions of the latent space v_0, w_1, w_2, w_3. Each direction is thus associated with a generative factor, with almost perfect disentanglement between time dependant changes and inter-subject variability.

The 5-fold reconstruction mean squared errors (MSE) (times 10^{-3}) for train/test images are $7.88 \pm .22/7.93 \pm .29$, showing little over-fitting. Prediction error for missing data, when trained on half-pruned data set, is $8.1 \pm .78$ which shows great extrapolation capabilities.

A thorough benchmark of six former approaches on this data set was provided in [Couronné et al., 2021] displaying similar MSE to ours. Although a couple of approaches [Couronné et al., 2021, Zhao et al., 2021] also disentangle time from space, ours is the only one to yield the true generative factors, with the direction of progression encoding the

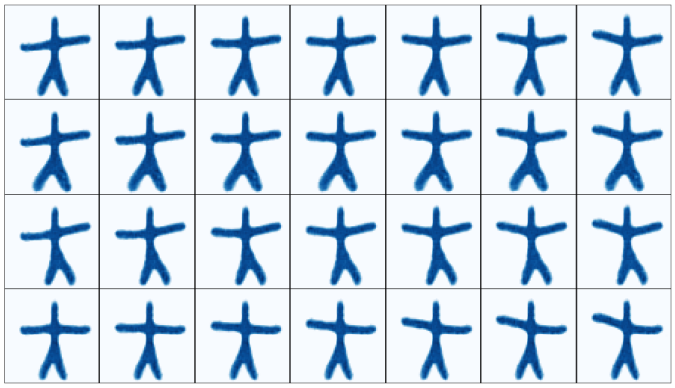


Figure 3.4: The average trajectory over time (left to right) on the first row, followed by its translation in the directions $w_1 = (0, 1, 0, 0)$, $w_2 = (0, 0, 1, 0)$ and $w_3 = (0, 0, 0, 1)$ in the latent space.

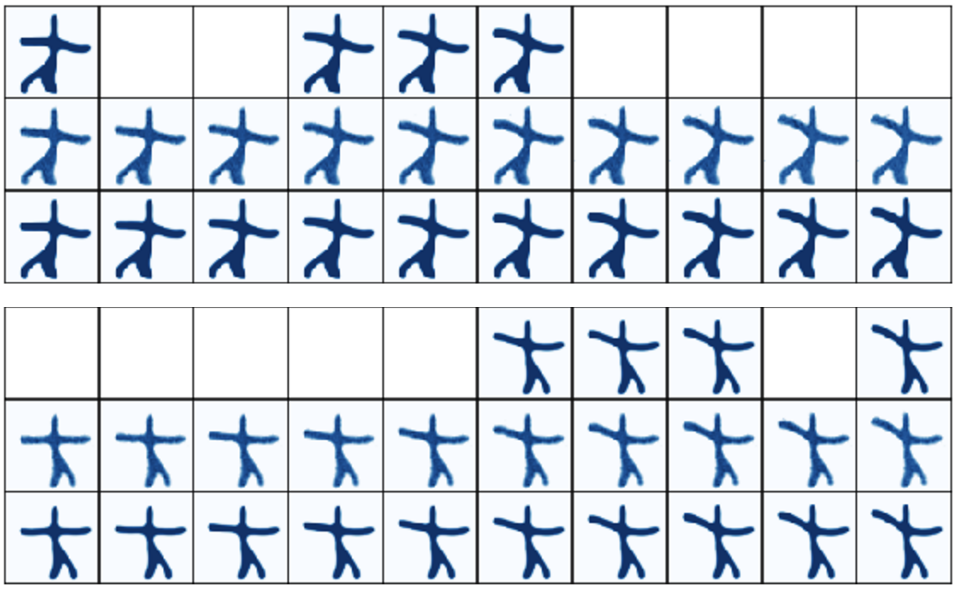


Figure 3.5: Data of two test subject (first rows), its reconstructed trajectory (second rows) and the ground truth (third rows).

motion of the left arm and the 3 spatial directions orthogonal to it encoding legs spreading, legs translation, and right arm position respectively (see Fig. 3.3 and 3.4).

3.3.2 Results on 3D MRI and PET scans

We then applied the method to 3D T1w MRI and FDG-PET scans from the public ADNI database (http://adni.loni.usc.edu). For MRI, we selected two cohorts: patients with a confirmed AD diagnosis at one visit at least ($N = 783$ patients for a total of N_{tot} =3,685 images) and Cognitively Normal (CN) patients at all visits for modelling normal ageing ($N = 886$ and $N_{tot} = 3,205$). We considered PET data for AD patients only ($N = 570$ and $N_{tot} = 1,463$). Images are registered using the T1-linear and PET-linear pipelines of the Clinica software[Routier et al., 2021] and resampled to 80x96x80 resolution.

We set $\dim(\mathcal{Z}) = 16$ for both modalities, as it is the smallest dimension that captured the reported dynamics with satisfying resolution. For MRI, error (10^{-3}) for train/test reconstruction and imputation on half-pruned data set for the AD model are $14.15 \pm .12/15.33 \pm .23$ and $18.65 \pm .76$, again showing little over-fit and good prediction abilities. In Fig. 3, the reference trajectory for AD patients reveals the structural alterations that

	CN cohort	AD cohort
Patients	625	784
Visits	2946	3686
Sex (M/F)	292/333	450/334
APOE (+/-)	195/428	495/269
Follow-up (y)	3.8 ± 2.4	2.9 ± 2.2
Age at baseline	72.8 ± 6.4	74.4 ± 7.3
Education (y)	16.5 ± 2.5	15.6 ± 2.8
MMSE	29.1 ± 1.2	24.0 ±− 4.5
ADAS13	9.6 ± 4.7	26.6 ± 11.4

Table 3.1: Demographics for the t1-MRI cohorts.

CN cohort	AD cohort
503	313
1567	1206
293/210	184/129
158/345	205/108
2.9 ± 1.9	2.8 ± 2.0
73.6 ± 7.2	74.7 ± 7.1
16.2 ± 2.8	15.6 ± 2.7
28.5 ± 1.7	24.5 ± 4.1
11.9 ± 6.4	25.7 ± 10.8

Table 3.2: And FDG-PET cohorts.

are typical of AD progression. The control trajectory displays alterations more in line with normal ageing.

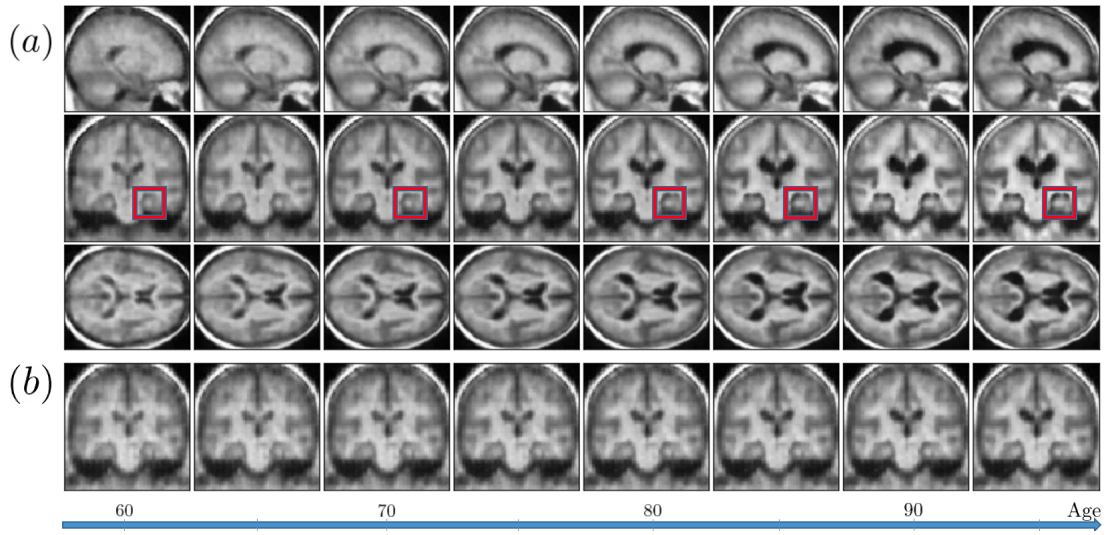


Figure 3.6: (a) Sagittal, coronal and axial views of the population trajectory over pathological time (left to right) for the AD cohort. Enlargement of the ventricles and atrophy of the cortex and the hippocampus are visible. Red squares around the hippocampus are positioned at the 5 stages of the Schelten's scale used in the clinics to evaluate AD progression. (b) Coronal view of the estimated normal ageing scenario, with matched reparameterized age distribution. Atrophy is also visible but to a smaller extent.

We tested differences in the mean of individual parameters between subgroups using Mann-Whitney U test within a 5-fold cross-validation. AD average onset age occurred earlier for women than for men: 72.2±.4 vs. 73.7±.6 years, $p < 3.10^{-7}{}_{\pm 5.10^{-8}}$). APOE-$\varepsilon$4 mutation carriers experience also earlier onset than non-carriers (71.8±.2 vs 73.1±.4, $p < 3.3.10^{-2}{}_{\pm 6.10^{-3}}$) and greater pace of progression (.1±3.10^{-2} vs -.08±2.10^{-2}, $p < 1.4.10^{-4}{}_{\pm 6.10^{-3}}$). The normal ageing model shows an earlier onset for men (71.2±.4 vs 73.7±.6, $p < 3.10^{-10}{}_{\pm 6.10^{-11}}$). These results are in line with the current knowledge in AD progression [Jack et al., 2015, Gurvich et al., 2018] and normal ageing [Coffey et al., 1998]. For PET scans, the 5-fold train/test reconstruction MSE (10^{-2}) are $4.71 \pm .32/5.10 \pm .23$.

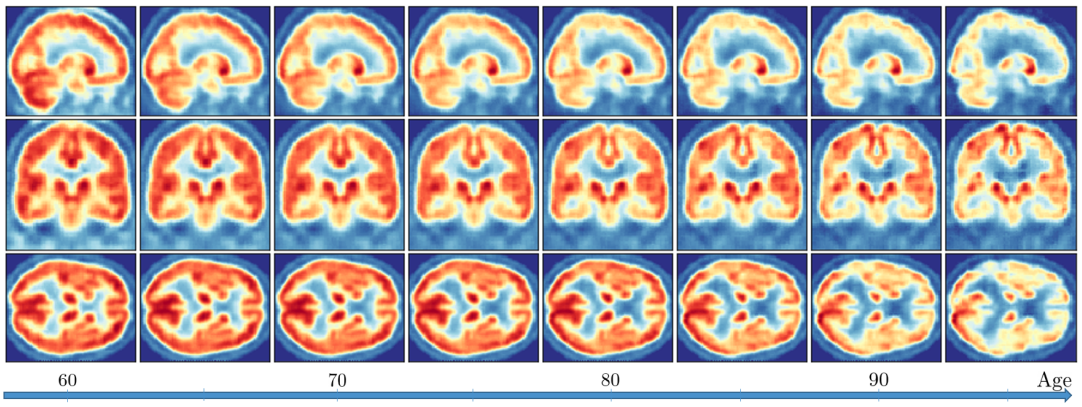


Figure 3.7: Sagittal, coronal and axial views of the average trajectory for FDG-PET scans, showing decreased level of metabolism across brain regions.

As is common in atlasing methods, these images average anatomical details from different subjects to provide a population trajectory, and are thus not as sharp as true images. Besides, the bottleneck layer of the VAE is very low-dimensional, which necessarily looses a certain amount of cortical and structural details.

3.4 Discussion

3.4.1 Effect of λ on the latent space

One crucial hyperparameter introduced in this model is the factor that weights the penalty for the longitudinal model's goodness of fit. If set too low, the longitudinal structure is ignored and the model is a vanilla VAE. On the other hand, if set too big, the reconstruction error becomes negligible and the latent space collapses, in a similar fashion to posterior collapse for β-VAE when β is set too big. It thus yields perfectly linear trajectories for patients, but at the expense of a very bad reconstruction error which removes the contextual information from the latent representations, which defeats the purpose of the longitudinal model. Fig. 3.8 illustrates the effect that increasing λ has on the latent space representations.

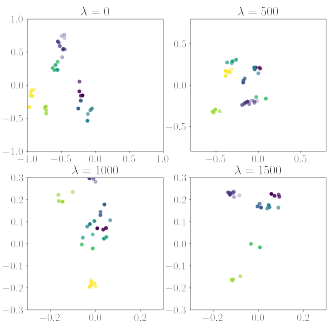


Figure 3.8: Illustration of the two first dimensions of the latent encodings for 6 patients from the AD cohort and the t1-MRI model. Each patient is represented by a colour, and successive visits are displayed with increasing transparency.

Besides, with increasing pressure to perfectly align successive visits, the ability of the network to perform well on test data is impaired. Not only does the training reconstruction go up as lambda increases, but the network also overfits more and more.

3.4.2 Limitations of the ℓ_2 norm for the reconstruction

The ℓ_2 norm, which measures pixel-wise dissimilarity, fails to capture the complex spatial relationships present in 3D volumetric data, resulting in blurry and imprecise reconstructed images. To address these limitations, alternative loss functions could be employed.

One approach is the use of perceptual loss, which incorporates high-level features extracted from a pre-trained neural network. By considering the perceptual similarity between the input and reconstructed images, the model can capture more meaningful information and produce sharper, more accurate reconstructions. Mutual information (MI) can also be used as a loss function. Maximizing MI allows the model to capture the underlying statistical dependencies between the input and output, leading to potentially more accurate reconstructions and improved fidelity. Adversarial losses offer another technique to overcome the limitations of the ℓ_2 norm. Adversarial training involves introducing a discriminator network that aims to distinguish between the reconstructed images and real images. The generator network, which produces the reconstructions, then attempts to generate more realistic images to fool the discriminator. This process encourages the generator to produce visually plausible and sharper reconstructions. Combining multiple loss functions can be beneficial as well.

Moreover, incorporating domain-specific knowledge and prior information can help mitigate the limitations of using the ℓ_2 norm. By incorporating anatomical atlases or anatomical priors into the training process, the reconstructions can be constrained to align with known anatomical structures, resulting in more accurate and clinically meaningful results.

An approach that alleviates the limitations of the pixel-wise ℓ_2 norm would be very likely to yield increased signal-to-noise ratios for medical images, and thus likely provide more accurate longitudinal models.

3.5 Conclusion

We proposed a generative Variational Autoencoder architecture that maps longitudinal data to a low dimensional Euclidean space, in which a linear spatio-temporal structure is learned to accurately disentangle the effects of time and inter-patient variability, while providing interpretable individual parameters (onset age and acceleration factor). This is the first approach to provide a mixed-effect progression model for 3D MRI or PET scans, without modelling voxels or mesh vertices individually, and it relies on vanilla deep learning architectures that only require the tuning of the loss balance. We showed that it bridges the gap between former approaches to handle longitudinal images, namely GP-VAEs, and Riemannian disease progression models. The method applied to MRI and PET data retrieves known patterns of normal and pathological brain ageing, but without the need to extract specific biomarkers. It does not only save time, but also makes the approach independent of prior choice of biomarkers.

Chapter 4

Metric learning with LVAE

Interpretable progression models for longitudinal neuroimaging data are crucial to understanding neurodegenerative diseases. Well validated geometric progression models for biomarkers do not scale for such high dimensional data, so we proposed to combine a Variational Autoencoder with a latent linear mixed-effects model. However, deep-learning methods are often criticized for not being interpretable enough, and therefore ill-suited for clinical translation. Here, we illustrate that the Longitudinal Variational Autoencoder framework can be understood as an extension of the geometric models presented in Part I, and the deep net is only a fancy way to parametrize the metric of the observation manifold. To do so, we demonstrate that imposing a Euclidean prior on the latent space allows the network to learn the geometry of the observation manifold, and model non-linear dynamics. This work was presented at the GeoMedIA workshop on geometric Deep-Learning for neuroimaging (2022).

Contents

4.1 Introduction

Many flavours of geometric disease progression models for biomarkers assume that the data lie on a manifold and evolve according to non-linear geodesics across time. Generalization of such models to high dimensional neuroimaging data has called for the use of deep learning dimension reduction techniques. On the other hand, much attention has been put on finding geometric interpretations of deep learning generative models, in order to refine the learned distribution. Through the geometric interpretation of the model introduced in Chapter 3 called "Longitudinal Variational Autoencoder", we will bridge the gap between the well validated disease progression models for biomarkers and the novel deep-learning-based methods.

4.1.1 Geometric disease progression models

Mixed-effects models provide one of the most popular disease progression framework for longitudinal data. Individual trajectories are modelled as small variations around the population-average trajectory. Early models used linear modelling, while non-linearities were later added with polynomial, logistic and exponential regressions. One less restrictive assumption is to consider that the observed biomarkers follow continuous trajectories in the space of observations that is assumed to be a Riemannian manifold [Schiratti et al., 2015b]. This approach provides spatio-temporal models that describe the average trajectory as a geodesic and the inter-patients variability as the parallel transport of this trajectory on the manifold. Particular cases have been derived with a metric that is set *a priori* in order to yield closed-form trajectories [Koval et al., 2017b, Schiratti et al., 2017], but less restrictive assumptions have been made in order to learn the appropriate metric

from the data [Gruffaz et al., 2021, Sauty and Durrleman, 2022c], at the expense of an added computational burden for the computation of the exponentials as Hamiltonian flows.

Those models however do not scale with high dimensional data and are mostly limited to the study of biomarkers. For neurodegenerative diseases, studying the alterations of the brain is a crucial task in order to understand the early stages of the disease. However, progression models for neuroimaging data are still understudied, and the use of deep learning as a "black-box" constitutes a barrier to widespread application by clinicians.

4.1.2 Geometry-aware deep generative models

Deep generative models aim to provide a neural network generator $p_\theta : z \in \mathcal{Z} \mapsto x \in \mathcal{X}$, parametrized by θ, from a latent space $\mathcal{Z} \subset \mathbb{R}^n$ to a target space $\mathcal{X} \subset \mathbb{R}^N$ (usually $n << N$) such that p_θ maps a simple distribution in $\mathcal{Z}$ to the complex data distribution in $\mathcal{X}$. The Jacobian of the generator $\mathrm{J} = \frac{\partial p_\theta}{\partial z}$ provides a linear mapping from tangent vectors of $\mathcal{Z}$ to tangent vectors of $\mathcal{X}$ and $\mathrm{M} = \mathrm{J}^T\mathrm{J}$ thus defines a smoothly varying inner product on the tangent bundle $T\mathcal{Z}$, and can be seen as a Riemannian metric on $\mathcal{Z}$. For VAEs, the generator network – the decoder – is coupled with an encoder network q_ϕ, parametrized by ϕ, that provides a variational estimate of the posterior distribution $q_\phi(z|x)$ for $x \in \mathcal{X}$. When adding a prior on the latent space, we obtain a lower bound for the log likelihood that can be optimized through backpropagation [Kingma and Welling, 2013].

In [Arvanitidis et al., 2017, Chen et al., 2018], the Riemannian metric M is used to describe the geometry of the latent space of a VAE, significantly improving interpolants and distance computation. [Chadebec et al., 2020] illustrates how such geometric description allows to better understand the latent space of VAE for low sample sizes, which are common in the context of neuroimaging data, and how it allows for better data generation and trajectory interpolation. [Shao et al., 2018] experimentally note that the curvature of the learned latent space is almost flat. [Falorsi et al., 2018], on the other hand, proposed to introduce Riemannian geometry tools in the input space of the VAE itself, in order to allow the generation of manifold valued data.

4.1.3 Longitudinal Variational Autoencoders

In Chapter 3, we introduced a progression model for imaging data that embeds a linear mixed-effect model in the latent space of a VAE. This model can then be calibrated on longitudinal databases of neuroimaging data such as T1-MRI or FDG-PET scans in order to provide a simple parametrization of the individual trajectories in $\mathcal{Z}$, that translate into non-linear trajectories in the image domain. An iterative MCMC optimization scheme allows ensuring that the VAE reconstructs images correctly while allowing to match the simple linear mixed-effect description of latent encodings to the complex trajectories in the observation space. Such an approach allows sampling individual trajectories at any timepoint – future or past – for prediction and missing data imputation.

Contributions The geometric nature of the VAE has been recognised and extensively studied. The most widespread paradigm is to embed the latent space with a Riemannian manifold structure, equipped with the pull-back of the Euclidean metric of the observation space through the decoder. On the other hand, it is standard to consider that medical data lie on a non-Euclidean manifold, on which trajectories are parametrized using mixed-effects models, to provide interpretable non-linear progression models. For high dimensional neuroimaging data, the use of dimension reduction techniques such as VAEs is required.

In this work, we provide a geometric interpretation of the LVAE progression model. We demonstrate that imposing a Euclidean prior on the latent space allows characterising the Riemannian metric of the observation manifold as the push-forward of the Euclidean metric through the decoder, providing a simple way to compute exponentials and geodesics in the image domain. This "reversed" conception of the geometry of the VAEs, first introduced in [Louis et al., 2019], bridges the gap between well-validated geometric progression models for biomarkers and recent approaches that integrate deep learning tools in order to model longitudinal neuroimaging data.

4.2 Geometry of the observation manifold

With the former notations, $p_\theta(\mathcal{Z})$ is a n-dimensional immersed submanifold of $\mathbb{R}^N$ if the activations functions are smooth and monotonic and if the weight matrix of each layer has maximal rank [Shao et al., 2018]. The first condition is a design choice, and the second condition can be checked after training. If we notate g the Euclidean metric on $\mathcal{Z}$, the push-forward of g on $p_\theta(\mathcal{Z})$ is defined, for any smooth vector fields U, V on $p_\theta(\mathcal{Z})$, as

$$p_\theta^*(g)(U,V) = g((p_\theta)_*(U), (p_\theta)_*(V))$$

where $(p_\theta)_*(U)$ and $(p_\theta)_*(V)$ are the pull-back of U and V on $\mathcal{Z}$, which are defined by $(p_\theta)_*(U) : f \mapsto U(f \circ p_\theta^{-1})$ for smooth functions $f : \mathcal{Z} \mapsto \mathbb{R}$. Any geodesic $\gamma : [0,1] \to \mathcal{Z}$ on $(\mathcal{Z}, g)$ thus translates to a geodesic $p_\theta : [0,1] \to \mathcal{X}$ on $(p_\theta(\mathcal{Z}), p_\theta^*(g))$. The decoder network is an isometry between $(\mathcal{Z}, g)$ and $(p_\theta(\mathcal{Z}), p_\theta^*(g))$ so computation of Riemannian exponential or parallel transport can be done in $\mathcal{Z}$ inexpensively, before being push-forward to the observation space. In the context of LVAE, the linear trajectories of the latent space and inter-patients translations thus translate to geodesics and parallel transport in the corresponding submanifold of the observation space.

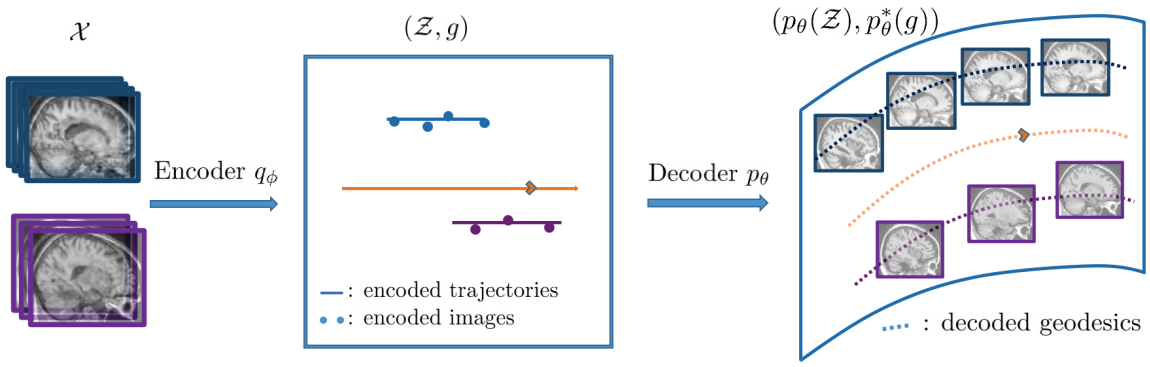


Figure 4.1: Schematic representation of the LVAE model. A linear trajectory is fitted on the latent representations (dots in $\mathcal{Z}$) of successive images, which then decodes into a geodesic of the observation space. Orange line is the population-average trajectory, travelled across time. Blue patient has 4 visits and violet patient has 3, displaying a progression of brain alterations.

4.3 Conclusion

We provided a geometric interpretation of the LVAE model that illustrates how the neural network can be understood as a way to parametrize the Riemannian metric of the observation manifold. The LVAE thus provides an extension of existing well-validated geometric models and provides a cheap way to compute nonlinear dynamics for high dimensional neuroimaging data.

Part III

Heterogeneity in Alzheimer's Disease.

Chapter 5

Influence of sex and APOE genotype on atrophic patterns

Alzheimer's Disease (AD) is a heterogeneous disease that disproportionately affects women and people with the APOE-ε4 susceptibility gene. We aim to describe the not-well-understood influence of both risk factors on the dynamics of brain atrophy in AD and healthy ageing. Regional cortical thinning and brain atrophy were modelled over time using non-linear mixed-effect models and the FreeSurfer software with t1-MRI scans from the Alzheimer's Disease Neuroimaging Initiative. Covariance analysis was used to disentangle the effect of sex and APOE genotype on the regional onset age and pace of atrophy, while correcting for educational level. A map of the regions mostly affected by neurodegeneration is provided. Results were confirmed on gray matter density data from the SPM software. Women experience faster atrophic rates in the temporal, frontal, parietal lobes and limbic system and earlier onset in the amygdalae, but slightly later onset in the postcentral and cingulate gyri as well as all regions of the basal ganglia and thalamus. APOE-ε4 genotypes leads to earlier and faster atrophy in the temporal, frontal, parietal lobes and limbic system in AD patients, but not in healthy patients. Higher education was found to slightly delay atrophy in healthy patients, but not for AD patients. A cohort of amyloid positive patients with MCI showed a similar impact of sex as in the healthy cohort, while APOE-ε4 showed similar associations as in the AD cohort. Female sex is as strong a risk factor for AD as APOE−ε4 genotype regarding neurodegeneration. Women experience a sharper atrophy in the later stages of the disease, although not a significantly earlier onset. These findings may have important implications for the development of targeted intervention. This work was published in Frontiers in Neurology, and has been presented at the AD/PD 2023 conference.

Contents

5.1 Introduction

Alzheimer's Disease (AD) is a neurodegenerative pathology that accounts for about 70% of the more than 55 millions dementia cases worldwide [Alzheimer's Association et al., 2021]. A silent phase, referred to as prodromal phase, shows accumulation of pathological proteins in the brain that lead to structural alterations in the brain without specific symptoms, eventually leading to a progressive cognitive decline that translates into a loss of the patient's autonomy for daily tasks, with impaired mnesic, visuospatial and communicative functions as well as unusual behaviours such as paranoia and aggressiveness. For people under the age of 75, the incidence of AD has been estimated below 1.0%, while it reaches 8.4% [Hebert et al., 1995, Hebert et al., 2010] for patients over 85, making older age the most important risk factor of the disease. Other risk factors have been identified with

certainty and are briefly recalled hereafter.

The second most important risk factor is often said to be the presence of hereditary gene mutations [Tanzi, 2012]. More precisely, the inheritance of AD can be of two types. On one hand, rare autosomal dominant mutations in APP, PSEN1, and PSEN2, that encode amyloid precursors and presenilin proteins, leading to early-onset (<60 years old) familial AD, which accounts for less than 1% of cases [Alzheimer's Association et al., 2021], and are often dismissed from studies on risk factors because of their low prevalence and understood causality. On the other hand, a frequent gene polymorphism: the $\varepsilon 4$ variant of the APOE gene, can influence susceptibility for roughly 50% of the common late-onset AD [Tanzi, 2012]. Genome wide association studies also identified a handful of other loci as potential risk factors, but thus far, the only gene variant considered to be an established risk factor for late onset AD is the APOE$-\varepsilon 4$ allele [Saunders et al., 1993, Farrer et al., 1997, Alzheimer's Association et al., 2021]. The APOE gene provides the blueprint for the alipoprotein E that transports cholesterol in the bloodstream and helps bind Aβ to cerebrospinal fluid, effectively clearing the brain of excess Aβ. The presence of $\varepsilon 4$ allele hinders Aβ clearance and leads to reduced neuronal injury response [Mahley and Huang, 2012], and in turns correlate with higher AD risks.

The other genetic factor that is known to impact AD risk rates is the sex of patients. Two thirds of AD patients are women [Alzheimer's Association et al., 2021] and their cognitive decline is faster than for men [Laws et al., 2016]. By contrast, healthy ageing tends to display delayed cognitive decline and structural alterations for women. [Beam et al., 2018] review conflicting studies on the topic of incidence rates and reach the conclusion that before the age of 75, no significant difference in incidence rates is detected, while incidence rates get significantly higher for women than for men after that inflection point. The longer longevity of women has long been thought to be the sole cause for this discrepancy. However, the loss of the protective effects of estrogens during menopause, demonstrated in vitro and in vivo [Członkowska et al., 2006, Genazzani et al., 2007], has also been suggested as a cause. Evidence also highlight a faster loss of autonomy but a longer lifespan after diagnosis for women [Sinforiani et al., 2010]. These facts raise the question of a sexual dimorphism of AD. A precise understanding of the distinct patterns of structural and functional brain alterations over time is still lacking. Accurate descriptions of those patterns could help to implement better practices to care for both men and women. In addition, women are still underrepresented in clinical trials [Martinkova et al., 2021] and only 15% of trials report results stratified by sex [Schwartz and Weintraub, 2021], although this proportion increases in more recent studies. For instance, the recent trials for Lecanemab reported a better response to treatment for men, and featured just over 50% women in both arms. [Van Dyck et al., 2023]. On the other hand, trials for Donanemab [Mintun et al., 2021] and Aducanumab [Budd Haeberlein et al., 2022] display similar enrolment demographics, but do not report results by sex. As suggested in [Mielke, 2018], several gendered factors other than sex can also mediate the observed associations, although data is rarely available in large scales datasets.

Another often-mentioned risk factor is the education level. According to the cognitive reserve hypothesis, having more years of education increases the connections between neurons and enables the brain to compensate for the early changes of Alzheimer's by using alternate routes of neuron-to-neuron communication to complete a cognitive task [Roe et al., 2007, Stern, 2012]. Women are believed to benefit even more from higher education than men, albeit receiving less education [Subramaniapillai et al., 2021]. [Mungas et al., 2018] confirm that education is an indicator of cognitive reserve, but that the protective effects on cognition are depleted as neurodegeneration progresses.

Building on the observation that the influence of sex on the progression of AD is not well understood, and that APOE-ε4 status is the strongest known genetic variant associated with AD, this work aims to describe the disentangled contributions of both those genetic factors on patterns of structural brain alterations, while accounting for biases due to varying education level. Such alterations, that are visible through MRI scans, reveal neurodegeneration occurring before the onset of the cognitive symptoms, and are thus crucial to understanding the early stages of the disease and for the design of drug trials. Leverageing the information from repeated measurements of cortical thicknesses and volumes of subcortical regions allows describing each patient's atrophic dynamics and compare across populations. Our goal is thus to provide a map of the brain regions that show significant correlation between sex or APOE-ε4 status and onset age and pace of regional atrophy. We provide this analysis for AD diagnosed and cognitively normal patients in order to emphasize the specificities of Alzheimer's progression.

5.1.1 Related work

Cross-sectional studies Over the course of AD progression, the link between cognitive decline and various covariates such as body mass index [Subramaniapillai et al., 2021], cardiac pathologies, APOE-ε4 genotype [Altmann et al., 2014, Sampedro et al., 2015, Neu et al., 2017, Buckley et al., 2018] and sex have been explored, highlighting a higher impact of comorbidities and APOE-ε4 genotype for women on the severity of the disease symptoms. [Ferretti et al., 2018] and [Laws et al., 2018] review the state of knowledge about the differences across sexes in AD clinical manifestations, biomarkers patterns and risk factors.

On the other hand, healthy ageing displays comparatively more spared alteration profiles for women compared to men, for both brain atrophy and cognitive decline. Seminal works dating back to the beginning of in-vivo imaging modalities report higher size decreases in men for average brain [Coffey et al., 1998] and frontal and temporal lobes [Cowell et al., 1994]. Many studies have since provided more insights into the differences between sexes for cognitive decline and link with comorbidities [McCarrey et al., 2016, Armstrong et al., 2019], exhibiting opposite correlation than in Alzheimer's studies. [Jack et al., 2015] exhibit associations between male sex, worse memory and higher hippocampal atrophy.

Longitudinal studies The collection of repeated measurements of neuroimaging data and clinical assessment allows a finer understanding of pathological pathways. Well-established modelling approaches include event-based models [Fonteijn et al., 2012, Young et al., 2014], Gaussian-Process models [Lorenzi et al., 2019, Abi Nader et al., 2020], Deep Learning methods [Couronné et al., 2019, Sauty and Durrleman, 2022b] and mixed-effects models. Application of such longitudinal models are frequent for the assessment of cognitive decline [Ito et al., 2011, Samtani et al., 2012, Bilgel et al., 2014, Jedynak et al., 2012, Fonteijn et al., 2012, Schiratti et al., 2015b, Koval et al., 2017b, Raket, 2020] but less used for imaging features. [Bernal-Rusiel et al., 2013a] advocate the use of linear mixed-effects models (LME) for progression models of neuroimaging data, and apply LME to a mesh of control points of cortical thickness in a mass-univariate setting in [Bernal-Rusiel et al., 2013b]. [Sabuncu et al., 2014] correlate these progression models with diagnosis in order to exhibit regions of the cortex that are most representative of the cognitive state of patients. Other studies also apply progression models to a few ROIs of the brain [Risacher et al., 2010, Skup et al., 2011, Tustison et al., 2019, Li et al., 2020] in order to display the regions that correlate most with diagnosis. Applying this kind of modelling to a study of sex differences, [Hua et al., 2010, Crivello et al., 2014] find that atrophy rates are higher

for female than male patients, while [Sangha et al., 2021, Cieri et al., 2022] explain that brain volumes and cortical thicknesses are consistently smaller for cognitively normal males after accounting for age, and that the differences get smaller as cohorts are chosen at a more advanced state of cognitive decline, until they get indistinguishable for confirmed AD patients, with a sharper decline for women. Several studies also find that sex modulates the interactions between amyloid SUVR and brain volume change [Armstrong et al., 2019], between CSF biomarkers and hippocampal atrophy [Koran et al., 2017] and between APOE genotype and hippocampal atrophy [Shen et al., 2019].

Those studies do not take into account important covariates such as APOE status and educational level, and do not disentangle the age of onset and the pace of atrophy. Many studies report the impact of APOE polymorphisms on brain atrophy in AD progression and link $\varepsilon 4$ with faster cortical thinning [Gutiérrez-Galve et al., 2009, Abushakra et al., 2020, Spampinato et al., 2016] and hippocampal atrophy [Liu et al., 2015, Li et al., 2016, Spampinato et al., 2016, Manning et al., 2014], with increasing effects as patients get older [Kim et al., 2018]. In the context of non-demented ageing, only $\varepsilon 4$ homozygotes are found to increase hippocampal and total gray matter atrophy [Crivello et al., 2010], with a stronger correlation between hippocampal volume and memory loss for $\varepsilon 4$ carriers [Gorbach et al., 2020]. GWAS also highlight APOE as the genetic variant mostly associated with brain atrophy and cortical thinning across lifespan, with a stronger effect for subjects with brain disorders [Brouwer et al., 2022].

5.1.2 Contributions

Overall, many studies explored the differences between men and women regarding AD progression and healthy brain ageing, as well as the influence of APOE polymorphisms. However, many of them rely on cross-sectional data that miss important information about the progression of the disease. Longitudinal tools have been proposed to leverage such information, and have been successfully applied to biomarkers for disease progression models. However, these studies have either focused on the patterns of cognitive decline, or on a chosen set of ROIs to assess structural alterations, thus not providing a detailed map of the differences across the brain. Some longitudinal studies have proposed analysis of cortical thinning or brain atrophy with a high spatial resolution, however the focus has always been put on finding regions that correlate best with diagnosis in order to obtain valuable information about the brain alterations most specific to AD. To the best of our knowledge, no study has yet provided a quantitative comparison between male and female patterns of structural alterations over time, while accounting for APOE-$\varepsilon 4$ genotype and education level, in order to provide a map of the brain regions that showcase distinct age-related alterations profiles. In addition, former works usually focus on the impact of one single factor, while only a joint analysis of both the genetic factors allows isolating the contributions from each one.

In this context, it seems relevant to further explore the differences between men and women regarding the onset and pace of alterations of the brain in both normal and AD progression, with a high degree of spatial resolution, while accounting for APOE-$\varepsilon 4$ genotype and education level. Longitudinal studies allow to leverage information about the progression over time, while mixed-effect models allow for an interpretable modelling of each feature's trajectory across time. To sum up our contributions :

- We model regional cortical thinning and brain atrophy using non-linear mixed-effect models, and propose a formulation that disentangles the onset age and the pace of atrophy,

- for each region, we proceed to a multivariate regression for the individual parameters with regard to female sex, APOE-ε4 allele count and education level, in order to isolate the effect of each factor,
- we obtain a brain map of p-values that assess how significantly each factor influences the patterns of atrophy over time, when corrected for the other covariates,
- we display the regions that differ the most regarding onset age and pace of structural atrophy for the chosen covariate, for both healthy and AD progression.

This statistical pipeline addresses the lack of understanding of the influence of sex and APOE on structural brain alterations, and disentangles the joint effects of the multiple risk factors.

5.2 Material and Methods

Data sets We performed the analysis on publicly available data from all waves (GO,1,2,3) of the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu), as it provides repeated MRI scans for both cognitively normal elderlies and AD patients. The ADNI was launched in 2003 as a public-private partnership and its primary goal has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). All scans are quality-checked by trained anatomists. We selected all patients with at least 2 visits with a MRI scan and selected three cohorts : an AD cohort with patients with at least one confirmed AD diagnosis, a healthy cohort with patients who are diagnosed cognitively normal at every visit and an intermediate cohort of amyloid positive patients (see [Hansson et al., 2018] for the chosen cutoffs) with at least one MCI diagnosis but no AD conversion. Patients with inconsistent AD diagnosis that revert to CN or MCI (107 patients) and amyloid negative patients (50 patients) are removed from the pathological cohort. Table 7.4 reports the demographics for the selected ADNI patients. The AD cohort is composed mostly of late-onset AD (566 patients) but also features a few early-onset cases (24 patients).

	AD cohort		Aβ+ MCI cohort		Healthy cohort	
	Male	Female	Male	Female	Male	Female
Patients (N)	329	261	262	176	210	264
Visits (N_{scan})	1,447	1,092	1,164	794	1,063	1,169
Total follow-up (y)	3.7 ± 2.9	3.4 ± 3.0	4.5 ± 3.4	4.6 ± 3.3	4.6 ± 3.3	4.1 ± 3.0
Age at baseline	75.0 ± 7.1	73.6 ± 7.7	75.4 ± 7.5	71.5 ± 7.8	74.0 ± 6.1	72.7 ± 6.0
Education (y)	16.1 ± 2.8	14.7 ± 2.6	16.4 ± 2.8	15.5 ± 2.7	17.2 ± 2.4	16.1 ± 2.8
APOE-ε4 (2/1/0)	69/163/97	44/133/84	32/99/121	24/71/81	5/51/154	7/75/182
MMSE	24.0 ± 4.0	23.2 ± 4.6	27.7 ± 2.0	28.2 ± 2.1	29.0 ± 1.2	29.2 ± 1.1
ADAS-Cog13	26.5 ± 10.6	29.2 ± 11.8	15.3 ± 6.9	12.6 ± 6.7	10.0 ± 4.6	8.3 ± 4.3

Table 5.1: Demographics for cohorts selected from ADNI. Numeric fields are in the form mean ± standard deviation. (y) is years. (2/1/0) refers to the amount of APOE-ε4 alleles. Cognitive scores are for all visits (not just baseline).

Data processing In order to validate our findings with an independent measure of atrophy, we extract gray matter density maps from the same t1-MRI data from ADNI. These features are obtained using the t1-volume pipeline of Clinica [Routier et al., 2021]. This pipeline is a wrapper of the Segmentation, Run Dartel (using a ADNI-based template) and Normalize to MNI Space routines implemented in the SPM12 software and yields a map of the average gray matter densities in each anatomical region defined in the AAL2 atlas [Tzourio-Mazoyer et al., 2002], which splits the brain into 80 gyri and sulci for the cortex and 14 subcortical regions. It displays a similar granularity to the Destrieux atlas and allows comparison. Gray matter density is fundamentally different from cortical thicknesses and regional volumes, and could potentially exhibit different progression patterns, especially since it is well documented that women have on average more gray matter and less white matter than men after correcting for brain size, in most of the brain [Luders et al., 2006, Chen et al., 2007, Lotze et al., 2019]. Each scan is processed independently, using the cross-sectional routines, since the amount and temporal spacing between successive scans is very variable. Indeed, the longitudinal routines build subject-specific templates to reduce within-subject variability, and are particularly useful when the successive visits are acquired at the same timepoint for all patients. In our case, the longitudinal templates would not be built the same way for each patient, and using the cross-sectional processing tools ensures that the thicknesses and volumes are estimated the same way for everyone, even when visits are missing.

MRI images from ADNI are acquired with either 1.5T (3440 samples) or 3T (3289 samples) scanners. As noted in [Han et al., 2006], Freesurfer outputs on 3T images are significantly thicker than for 1.5T images, which can hurt longitudinal modelling and bias population-level comparisons. As is common in longitudinal studies, we proceed to additive bias correction within each diagnosis group for each extracted feature. The corresponding matched histograms are presented in supplementary materials (Fig. 7 and 8).

Longitudinal modeling Mixed-effects models describe each patient's progression over time as a small variation – the random effects of the model – around the average population trajectory – the fixed effects of the model. This study focuses on univariate models, that are calibrated on the repeated measurements for one region of the cortex or subcortical structure at a time. Linear mixed-effects models are widely used for longitudinal modelling and assume that features evolve according to straight lines over time. Each patient is thus characterized by a slope and an intercept. The slope can be interpreted as the pace of decline of a patient for the given feature, but the intercept is less meaningful. In that setting, we choose to parametrize the inter-subjects variability as the combined effects of an individual onset age, and a pace of decline as suggested in [Schiratti et al., 2015b]. The onset age can be understood as a horizontal intercept and describes the age at which a patient crosses the population-average-value threshold, and the pace of decline as the slope. Both these temporal parameters allow to align all patients on a common progression timeline. Namely, given a family of feature observations $\{y_{i,j}\}$ at times $\{t_{i,j}\}$ for $1 \leq i \leq N$ indexes the N patients and $1 \leq j \leq N_i$ indexes the N_i visits of patient i, we resort to the mixed-effect generative model

$$y_{i,j} = f\left(e^{\xi_i}(t_{i,j} - \tau_i) + t_0\right) + \varepsilon_{i,j}$$

where e^{ξ_i} and τ_i, respectively, the *progression pace* and *onset age* of patient i are called the individual parameters that provide an affine time warp to account for the variability in pace and onset of decline between patients. These individual parameters form the random effects of the model. We choose the Gaussian priors for the noise $\varepsilon_{i,j} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and random effects $\tau_i \sim \mathcal{N}(t_0, \sigma_\tau^2)$ and $\xi_i \sim \mathcal{N}(0, \sigma_\xi^2)$. In a linear mixed-effects framework, f

is an affine function $t \mapsto p_0 + v_0 * (t - t_0)$, while for non-linear mixed-effects it is usually chosen to be the logistic function $t \mapsto \left(1 + \left(\frac{1}{p_0}\right) \exp\left(\frac{-v_0 t}{p_0(1-p_0)}\right)\right)^{-1}$. The parameters p_0, v_0 and t_0 are respectively a reference *position*, *velocity* and *time* and describe the average trajectory. Together with the variances $\sigma_\varepsilon, \sigma_\tau, \sigma_\xi$, they form the fixed-effects of the model.

In practice, logistic regression is one of the most popular shape of trajectory, so all features should be normalized to the range $[0, 1]$. We thus discard outliers using the so-called three sigma rule and add a min-max normalization within each cohort, providing realistic asymptotes. Since cortical thicknesses and brain volumes decrease over time, we flip data around .5, using the rotation $x \mapsto 1 - x$, in order to ensure increasingness, which is required for logistic modelling.

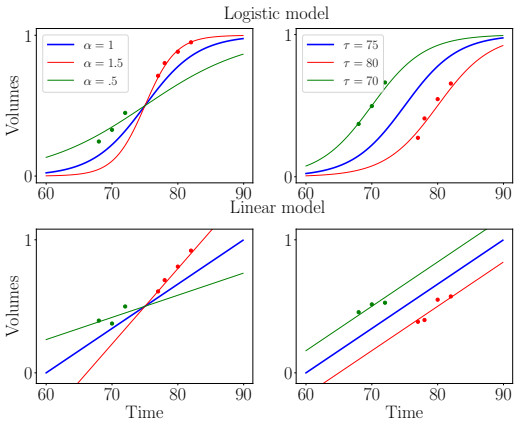


Figure 5.1: Illustration of logistic (first row) and linear (second row) mixed-effects models. Blue curve is the average trajectory (parametrized by t_0, p_0 and v_0) and red and green curves illustrate the effect of progression pace ($\alpha = e^\xi$) and onset age (τ), which allow matching individual measurements (isolated dots) to a continuous trajectory.

Calibration of the model These longitudinal statistical models are part of a family of geometric models that have been studied in [Schiratti et al., 2015a] and [Koval et al., 2017a]. We can proceed to a Maximum a Posteriori estimation of both random and fixed effects with the MCMC-SAEM procedure, in which the estimation step of an Expectation-Maximization algorithm is replaced by a stochastic approximation. See [Kuhn and Lavielle, 2004] and [Allassonnière et al., 2010] for details on this procedure, description of the complete likelihood and proof of convergence and stability. One model is calibrated on each of the three cohorts. All analysis is done using the Leaspy software.

Logistic versus linear regression As mentioned above, the choice of f in the mixed-effect formulation should reflect the average progression of the feature. A few methods exist to learn the exact shape of the progression profile [Gruffaz et al., 2021, Sauty and Durrleman, 2022c] but add a strong computational burden. We compared the two most widespread available shapes, namely linear and logistic curves, in order to assess the stability of convergence as well as the estimated noise of the longitudinal fit. The logistic curve marginally surpassed the linear curve for both these criteria.

Goodness of fit In order to ensure that the longitudinal models accurately describe the atrophy of each feature, we compare the distributions of reconstruction errors (the differences between the predicted values of the models at the time of each visit and the actual measurements) for the three cohorts to the distribution of measurements noise. imaging data are indeed subject to variations in the experimental conditions and in the processing

pipeline. In order to evaluate this noise, some visits are acquired twice within a few hours. For ADNI, we dispose of 1604 duplicate images from 285 different patients. We provide boxplots that display the empirical distributions of test/re-test errors and of the reconstruction errors for the three cohorts in supplementary materials. Distributions largely overlap, and reconstruction errors fall close to the measurements' uncertainty, which hints that our models could not be improved without overfitting. Additionally, it is important to note that the reconstruction errors are not biased regarding sex, APOE status, education level, ADNI phases nor field strength.

Statistical analysis Once the univariate models are calibrated, we have a family of patient-wise onset ages and progression paces for each feature. We can thus proceed to statistical testing of the differences between subgroups. To do so, we perform covariance analysis using ordinary least squares to regress the individual parameters with regard to female sex, number of APOE-$\varepsilon 4$ alleles and education level of each patient. This evaluates the impact of each covariate after correcting for the other ones, and also the significance of the corresponding association. Since we are in a multiple testing setting, we use False Discovery Rate (FDR) [Benjamini and Yekutieli, 2005] to obtain a map of corrected p-values that assess how certain it is that the chosen covariate influences the onset or pace of atrophy for each region. Given the high number of statistical tests (one for each of the 148 regions of the Destrieux atlas), a more stringent correction method such as Family-Wise Error Rates would be unsuitable as it would discard significant correlations. This *a posteriori* analysis is necessary in our non-linear Bayesian setting, since the covariates can not be integrated in the longitudinal model as they would for a general linear model. Mixture models have been proposed in [Poulet and Durrleman, 2021] in order to account for covariates in the distributions of the random-effects, but it requires a lot more data in order to estimate such distributions for each combination of covariates.

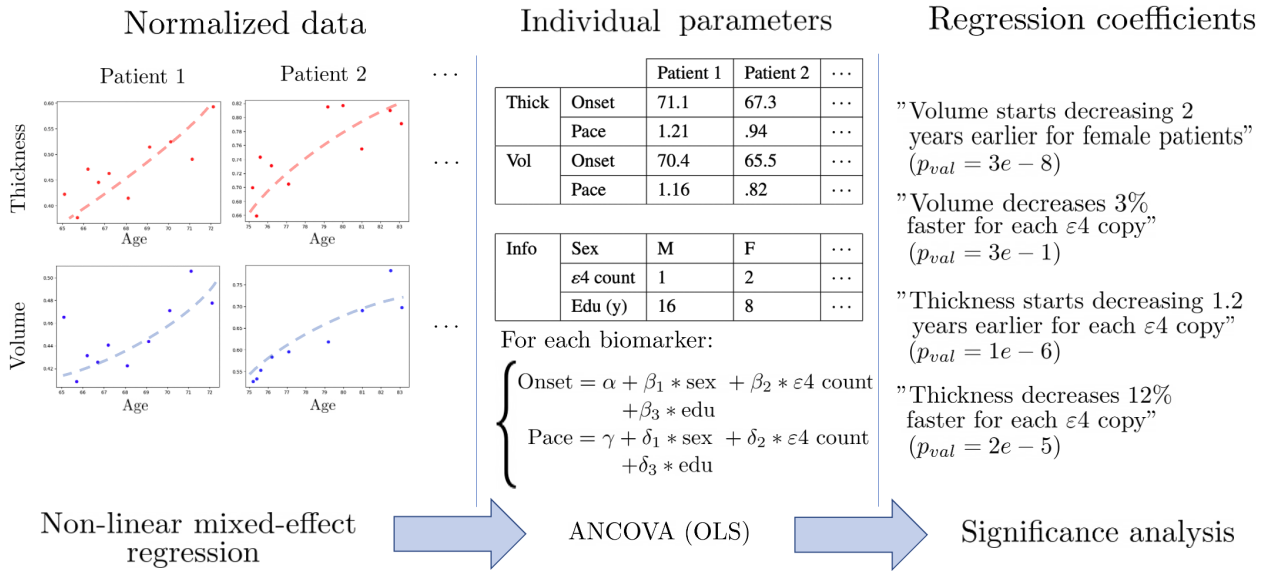


Figure 5.2: Schematic description of the statistical pipeline for two dummy features called "Volume" and "Thickness". First, the normalized data is modeled by univariate non-linear mixed-effects models, in order to obtain an individual age at onset and a pace of decline for each feature. The dots are the actual measurements over time, and the dashed shapes are the fitted trajectories. Then, Covariance Analysis is performed using ordinary least square regression (OLS) in order to learn the influence of each covariate on the onset and pace of decline. Finally, statistical analysis is performed in order to assess the significance of each association.

Interpretability of the p-values For each feature, patients are aligned on a common timeline with the individual onset age and pace of atrophy. One can, for instance, observe that for the inferior temporal gyrus, cortical thinning occurs on average 1.5 years earlier for each copy of the $\varepsilon 4$ allele, and on average 23% faster for women, after correcting for the other covariates. However, these multiple timelines are learned independently and cannot be compared directly in order to exhibit the regions that are mostly affected by a covariate, as it may not be statistically significant. Resorting to statistical testing, using t-tests under the null hypothesis that the regression coefficient is 0, allows circumventing this issue as it normalizes all the effect sizes by the natural temporal scale associated with the region, and accounts for the uncertainties in estimation of the regression coefficients. In addition, for a fixed sample size, p-values are monotonically related to effect sizes, so lower corrected p-values can be interpreted as indicative of a bigger impact of the covariate, and allow comparison across brain regions and covariates.

We provide a map of cortical regions and a list of sub-cortical volumes that display significant differences when stratified for sex or APOE-$\varepsilon 4$ status, regarding onset age and pace of atrophy, for AD, amyloid positive MCI and healthy ageing. Corrected p-values higher than .05 are not displayed, and the remaining p-values are presented in logarithmic scale (base 10) for cortical thinning maps.

5.3 Results

Scales of log p-values vary between each plot and should be taken into account to evaluate the strength of the considered risk factor. For visualization purposes, cortical thinning maps are presented by cohort (AD in Fig. 5.3, control cohort in Fig. 5.4 and MCI in Fig. 5.5) while all results for subcortical volumes are compiled in Table 7.5. Scatterplots of the raw data are provided in supplementary materials to illustrate the non-linear progression of repeated measurements over the course of disease progression, and the relevance of the chosen modelling framework. It should be noted that absolute pairwise-correlations between covariates are .02 for APOE/sex, .05 for APOE/education and .18 for sex/education, which means that regressing with regard to the education level might mitigate the strength of the observed sexual dimorphism, but interactions between covariates are less likely for the other pairs.

Impact of sex on cortical thinning for AD progression Almost all regions of the cortex display a significantly higher progression pace for female patients, with the exception of the motor cortex, sensory areas and inferior frontal lobe. Regions that are most accelerated in females are the entire temporal lobe, the middle and superior frontal lobe, the entire occipital lobe and the anterior and medial parietal lobe. Onset ages are, on the other hand, more homogeneous across sexes. The motor cortex, the cingulate gyrus and the medial parietal lobe display an earlier onset for men, while the inferior temporal lobe and inferior frontal lobe display an earlier onset age for female patients. All other regions do not allow rejecting the null hypothesis.

Impact of APOE-$\varepsilon 4$ genotype on cortical thinning for AD progression Almost all regions of the cortex also display a significantly higher progression pace for APOE-$\varepsilon 4$ carriers. The most affected regions are located roughly in the same areas as for the impact of sex, but with different regions of highest intensity, and an overall lower effect. The temporal lobe, the parietal lobe and the frontal lobe also display an earlier onset for APOE-$\varepsilon 4$ carriers. The region that presents the most advanced onset for $\varepsilon 4$ carriers is the hippocampal gyrus.

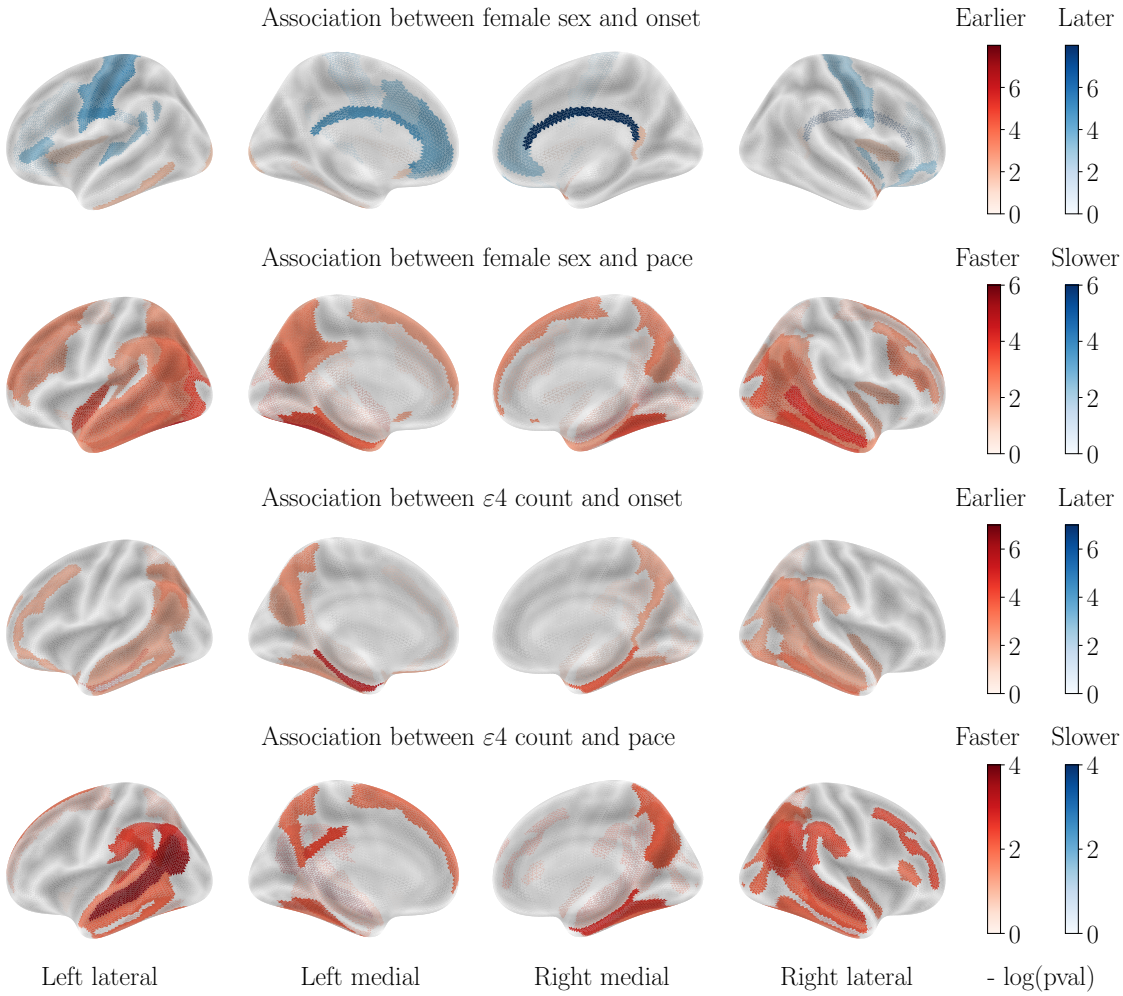


Figure 5.3: Cortical thinning over the course of AD progression. Legend bars show negative log p-values. It should be noted that blue values indicate that the considered covariate (female sex or APOE-ε4 genotype) is a protective factor, and red suggests a risk factor.

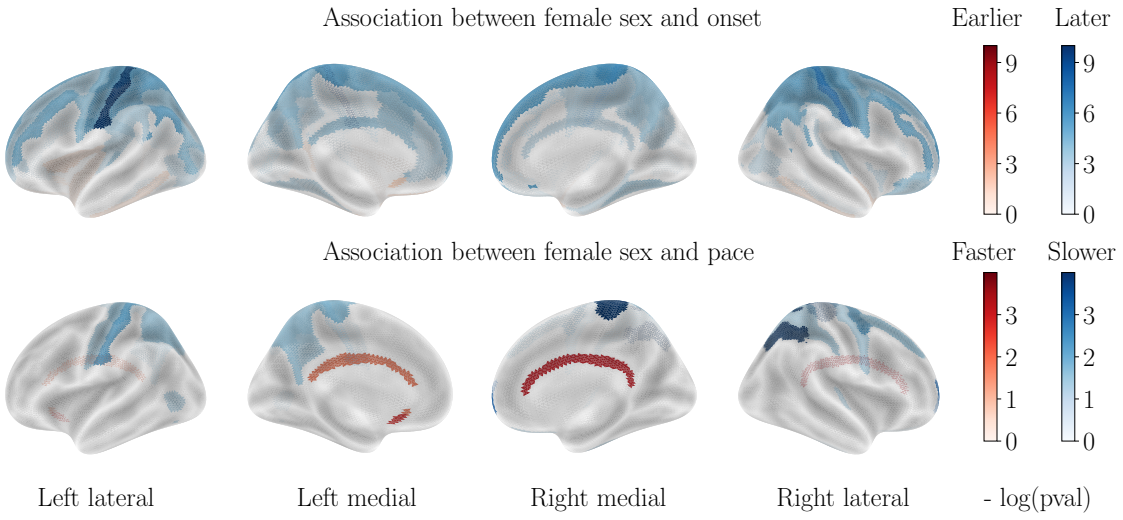


Figure 5.4: Cortical thinning over the course of cognitively normal ageing. Associations with APOE-ε4 status are not displayed, as they are not significant.

Impact of sex on cortical thinning for healthy ageing Contrary to what was seen in AD progression, the sexual dimorphism for healthy ageing manifests mainly through a significantly earlier onset age for male subjects, especially in the parietal and frontal lobes. Only the inferior temporal lobe and subcallosal gyrus display a slightly earlier onset for women. On the other hand, progression paces are similar across sexes except for the cingulate gyrus that is much accelerated in female patients, while a few regions of the parietal lobe display a higher pace of atrophy for male patients.

Impact of APOE-ε4 genotype on cortical thinning for healthy ageing In order to assess the impact of the APOE-ε4 genotype on cognitively normal ageing, we also stratified regarding this factor. Interestingly, the ε4 allele carriers do not display significantly different patterns of cortical thinning across healthy ageing, which leads to believe that APOE-ε4 by itself does not cause the accelerated atrophy of the brain but only serves as one cog in the unravelling of AD. It should be noted however that the small amount of ε4 carriers in the healthy cohort can bias this result.

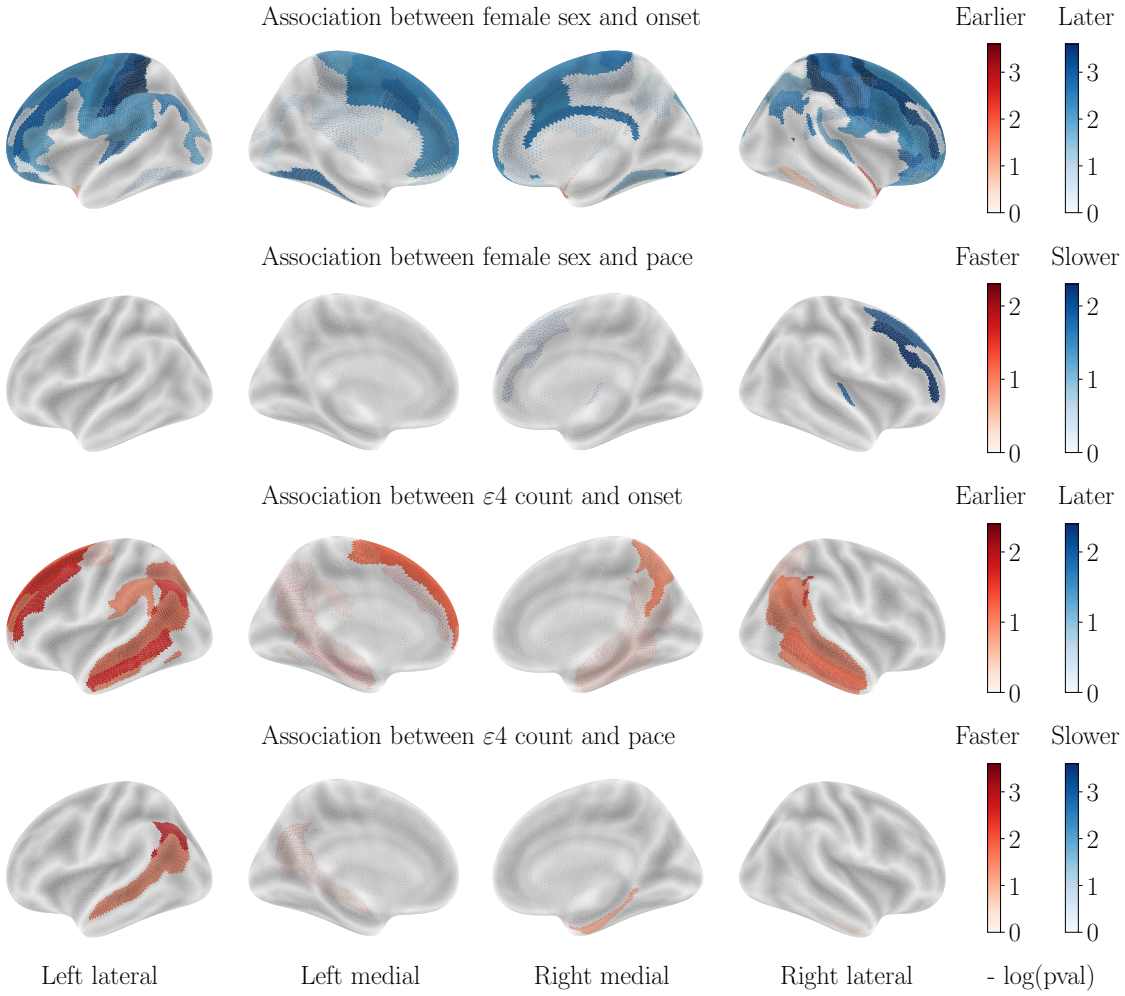


Figure 5.5: Cortical thinning over the course of MCI patients' ageing.

Influence of both covariates for patients with MCI The progression paces show little significant correlation with both covariates, however, for onset age the correlations with sex are similar to those displayed for healthy ageing while the correlations with APOE-ε4 status are similar to those displayed for AD progression, although of a weaker effect.

Correlations with the patterns of atrophy for subcortical structures For healthy ageing, male sex correlates with an earlier onset age for most regions with no significant differences in pace of atrophy, while no significant correlation is found for APOE-ε4 genotype. For the AD cohort on the other hand, female sex correlates with an earlier onset for most regions, and higher paces of the hippocampi and bilateral amygdalae while APOE-ε4 genotype correlates with earlier onset and higher pace for the hippocampi and bilateral amygdalae. We reach the same conclusion regarding the MCI cohort as for cortical thinning : sex correlates with atrophy similarly to the healthy cohort, while APOE-ε4 genotype correlates similarly to the AD cohort. Regions without significant correlations are not displayed.

		Correlations with sex			Correlations with APOE-ε4		
		p_{val} CN	p_{val} MCI	p_{val} AD	p_{val} CN	p_{val} MCI	p_{val} AD
Amygdala (L)	Onset	-	-	8.1e-03	-	4.9e-03	1.9e-07
	Pace	-	-	1.2e-06	-	2.2e-03	7.0e-06
Amygdala (R)	Onset	-	-	7.4e-05	-	3.2e-02	1.4e-06
	Pace	-	-	2.3e-05	-	2.5e-05	1.9e-02
Caudate (L)	Onset	5.5e-07	6.7e-04	1.7e-04	-	-	-
	Pace	-	-	-	-	-	-
Caudate (R)	Onset	1.2e-04	1.8e-04	8.3e-04	-	-	-
	Pace	-	-	-	-	-	-
Hippocampus (L)	Onset	4.7e-06	1.2e-02	-	-	3.0e-03	1.5e-07
	Pace	-	-	3.9e-05	-	5.7e-04	1.2e-02
Hippocampus (R)	Onset	1.1e-04	1.2e-02	-	-	2.1e-03	1.1e-09
	Pace	-	-	1.2e-04	-	3.9e-04	2.6e-02
Pallidum (L)	Onset	4.6e-04	6.9e-04	4.9e-02	-	-	-
	Pace	-	-	-	-	-	-
Pallidum (R)	Onset	6.5e-03	1.8e-04	3.1e-05	-	-	-
	Pace	-	-	-	-	-	-
Thalamus (L)	Onset	1.3e-06	7.8e-12	1.5e-07	-	-	-
	Pace	-	1.1e-02	-	-	-	-
Thalamus (R)	Onset	4.1e-06	4.6e-09	5.5e-08	-	-	-
	Pace	-	-	-	-	-	-

Table 5.2: Significant correlations with subcortical atrophy. For all the significant features, APOE-ε4 genotype correlate with lower onset age and higher pace, and female sex correlates with higher onset age and higher pace, as illustrated by the colors (red for earlier or faster and blue for later or slower).

Differences between left and right hemisphere For AD progression, the left hemisphere displays more regions that differ significantly for both sex and APOE-ε4 stratification, but the associated t-values do not significantly differ between hemispheres. This confirms that atrophic patterns are asymmetric but not completely lateralized [Derflinger et al., 2011].

Impact of education level Fig. 5.6 illustrates the impact of education level on cortical thinning patterns, after accounting for sex and APOE$-\varepsilon$4 status. Correlations are only found in the healthy cohort, and higher education level correlates with higher onset age for the postcentral gyrus, the medial parietal and medial occipital lobes, as well as the left

superior temporal lobe, and lower pace of atrophy for parts of the frontal and occipital lobes. No significant correlation is found for the AD cohort, in line with former studies [Koval et al., 2021b]. For subcortical volumes, only the hippocampal atrophy is delayed for more educated healthy patients, and no significant correlation is found for AD patients. One hypothesis is that education level is not, in itself, helping delay the atrophy for healthy controls, but acts as a proxy of lifestyle healthiness, which influences metabolic pathways and brain atrophy. In AD cohort, that small preserving effect is likely to be cancelled by the influence of other covariates that cause massive atrophy.

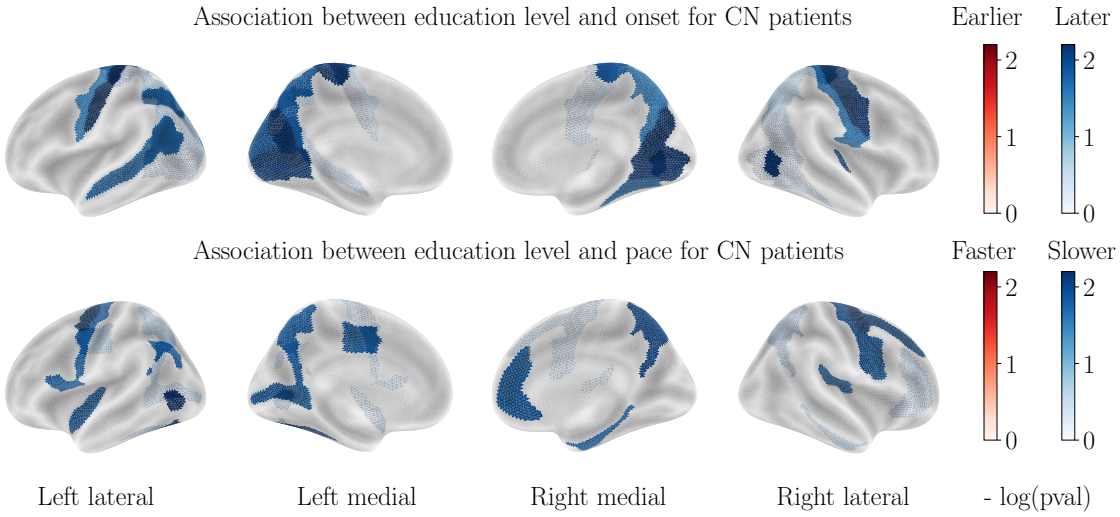


Figure 5.6: Correlation between education level and cortical thinning dynamics for CN patients. Correlations observed for the AD patients are not significant and thus not displayed.

Validation on gray matter density maps from ADNI Despite the strong sexual dimorphism in regional gray/white matter ratios, we recover most results from the main analysis. For AD patients, the same regions display a strong dimorphism : almost all the temporal lobe, frontal lobe and occipital lobe display a much stronger atrophic pace for women, while the postcentral gyrus displays an earlier onset for men. APOE$-\varepsilon 4$ allele also correlates with a faster pace and earlier onset in the medial temporal and parietal lobes, as well as in the temporal lobe and parts of the frontal lobe. For CN patients, we recover the absence of association with APOE genotype, while men display a much earlier onset for almost all regions (with emphasis on the parietal lobe and postcentral gyrus), while a few isolated regions of the frontal and temporal lobe display a faster pace of decline for men. The conclusions regarding the MCI cohort and the impact of education level are also confirmed. The differences with the main analysis are thus only a matter of a few isolated regions and of the order of magnitude of the statistical associations. We provide the association figures for cortical thinning and the tables for subcortical structures in supplementary materials.

5.4 Discussion

Our study allows isolating the contributions from sex, APOE status and education level on both onset age and pace of atrophy for each region of the cortex and brain, while accounting for the other covariates. We find that :

- Sex is as important a risk factor as APOE-$\varepsilon 4$ genotype for cortical thinning and brain atrophy for AD patients, as both stratification yield effect sizes that are not

significantly different. This finding is important, given that most studies refer to APOE polymorphism as the biggest risk factor for AD, with little to no attention paid to the sexual dimorphism. Besides, focus is often put on specific regions such as the hippocampus and enthorinal cortex, as well as cognitive abilities, while the dimorphism in structural alteration is manifest for a wide variety of brain regions, that may also play an important role in the progression of the disease.

- Female AD patients decline much faster than male patients, however, the onset of the atrophy is still a little earlier for male patients, in line with the healthy ageing patterns. This point is revealed by the use of longitudinal studies and has important implications for practitioners for the follow-up of patients. It suggests that the disease strikes female patients with more intensity, but not earlier. Besides, former studies found that female sex leads to both earlier and faster cognitive decline [Koval et al., 2021b], which suggests different compensation mechanisms in males and females, that translate cortical thinning and brain atrophy into cognitive alterations.

- Higher education seems to be a slightly protective factor for healthy patients, but is not significant in AD progression. Given the strong protective effect of higher education on cognition [Stern, 2012], it is interesting to note that such protective effect does not happen at the structural level, but rather a functional level.

- For the amyloid positive MCI cohort, the influence of sex is similar to that of the healthy cohort, while the effect of APOE-ε4 genotype is similar to that of the AD cohort, although of an overall weaker effect. This is in line with the idea that the sexual dimorphism of AD manifests itself at the later stages of AD, rather than at the prodromal stage, with a harsher decline in structure and function, but not an earlier onset. It is also coherent with the absence of significant difference in prevalence before the age of 75.

It is important to note that our study focuses specifically on structural alterations in the brain, and that this is only one piece of the larger puzzle of AD pathology. Nevertheless, structural changes in the brain are a key feature of the disease, and understanding the dynamics of brain atrophy is an important step towards developing effective interventions. Besides, exhibiting a different influence of risk factors on neuronal death than what is expected from previous knowledge on studies of cognitive decline hints that said risk factors differentially influence each cog of AD pathogenesis.

Several limitations and design choices need to be acknowledged in order to put results in perspective.

Clinical characterization of diagnosis in ADNI For the control cohort, we used the CN patients in ADNI, because it allows a fair comparison between AD progression and cognitively normal ageing, without acquisition biases. Since ADNI is not an epidemiological study on healthy ageing, the cognitively normal patients may not represent accurately the general population because of ADNI's inclusion criteria. For the MCI cohort, it is important to note that MCI does not always lead to AD as other factors can cause patients to have it. The prevalence of MCI symptoms are found to be between 15 and 20% for patients over 60 years old [Gauthier et al., 2006], and roughly 8 to 15% of these convert to AD each year [Petersen, 2016]. In order to make the cohort less heterogeneous, we decided to only keep amyloid positive patients in order to interpret this cohort as describing AD pathogenesis. [Petersen et al., 2010] assert that the recruited cohorts for CN, MCI and AD patients successfully describe the associated clinical status.

Validation cohort Reproducing this study is complicated because of the lack of a publicly available longitudinal database of the same scale as ADNI, and the difficulty to fit longitudinal models to pooled databases with varying protocols. Our statistical pipeline requires splitting the patients between CN and AD patients, while only keeping patients with more than 2 visits with a t1-MRI scan. For instance with those criteria, in the Australian imaging Biomarkers and Lifestyle Study of ageing (AIBL), often used to validate findings on ADNI, the AD cohort only contains 38 patients with a total 105 visits, and the healthy cohort 128 patients with 431 visits, before the additional stratification by sex or APOE-ε4 genotype. The sample sizes are thus too small to provide enough statistical power to reproduce the regional map of significant correlations.

Impact of the chosen atlas It must be taken into account that the choice of a different atlas might yield different results. One anatomical region that is positively associated with a covariate can span many ROIs or on the other hand be split between ROIs and prevent the detection of the true underlying association. On the other hand, our mixed-effect formulation cannot be fitted directly to vertex-wise thickness measurements because the onset age (or horizontal intercept) will not converge for features that display little progression compared to the feature noise.

Regions normalization Since subcortical regions' volumes are correlated with the volumes of the skull and brain, it is standard practice to normalize by the intracranial volume in order to allow a fair comparison between subjects. The impact of such normalization is discussed in former studies : in [Whitwell et al., 2001, Westman et al., 2013, Voevodskaya et al., 2014] authors suggest that subcortical volumes should be normalized with total intracranial volume, while cortical thicknesses should not, for diagnosis prediction and progression models. On the other hand, [Zhou et al., 2014] find that cortical thicknesses should be normalized with either intracranial volume or average thickness to help predict the cognitive status of a patient. [Luders et al., 2006] discuss the effect of normalizing cortical thicknesses to allow a fair comparison between healthy male and female subjects. Allometric scaling is often used to circumvent this issue, but [Williams et al., 2021] find that allometric and linear scaling yield similar effect sizes and coefficients when evaluating the effect of sex and age on brain measurements.

Interpretability of the onset age As opposed to the straightforward pace of atrophy, the onset age needs to be interpreted with caution. It represents the age at which a patient crosses a chosen threshold (chosen to be the average value for the whole cohort). Brain atrophy does not have a well-defined starting point, and the notion of onset age is thus to be understood with regard to a reference point in the average progression. For instance, the earlier onset ages for men in the CN cohort can be understood as describing the fact that cortical thicknesses are significantly thinner than for women. Since the difference in onset gets way smaller for the AD cohort, it agrees with former findings that the gap between men and women closes as neurodegeneration progresses, with a greater pace of decline for women in the late stages of the disease [Sangha et al., 2021, Cieri et al., 2022].

Clinical implications The discrepancies in alteration patterns of the brain between sexes and APOE genotypes reinforce the idea that the disease manifests differently between subpopulation and care should be provided accordingly. Besides, clinical trials that monitor the effect of a drug on neurodegeneration should put an emphasis on both having more representative demographics after enrolment, and evaluating the impact of drugs in a stratified manner as patients are likely to react differently to the drug depending on both the explored genetic risk factors. It is also interesting to note that most regions significantly affected by the considered risk factors are known to be associated with AD (e.g. medial

temporal lobe, limbic system, frontal lobe, etc). This work could be interpreted jointly with studies that describe the sexual differences in associations between cortical thickness and cognition such as [Cieri et al., 2022] in order to provide a complete description of the sexual dimorphism of neurodegeneration and its link with AD cognitive symptoms.

Perspectives This statistical pipeline can be applied to other modalities of neuroimaging data. For instance, [Sauty and Durrleman, 2023] applies the same statistical pipeline to FDG-PET scans instead of structural MRI in order to assess the sexual dimorphism of AD and influence of APOE genotype regarding brain hypometabolism, which complements the study of brain atrophy. Other imaging modalities specific to AD, such as amyloid-PET or tau-PET, are also under investigation within this pipeline. A strong sexual dimorphism for the spreading patterns of pathological protein tangles in the brain would prompt the latest clinical trials, that choose amyloid load as a primary disease outcome, to put great care in evaluating the differentiated results between male and female patients.

5.5 Conclusion

Structural alterations of the brain happen before the onset of cognitive decline, and the collection of MRI-derived features reveals varying patterns of atrophy between subgroups of patients. Female sex, APOE-ε4 genotype and low education level have been identified as the biggest risk factors for cognitive decline, but their impact on early-stage structural alterations is not yet well understood. In this work, we described the disentangled effects of these 3 covariates on the onset and pace of regional atrophy and cortical thinning. This reveals that female sex is actually as strong a risk factor for neurodegeneration as APOE-ε4 genotype (Fig. 5.3), and women experience much faster atrophic rates after correcting for the APOE genotype. APOE-ε4 genotype leads to earlier and faster atrophy for most regions of the brain for AD patients, but not for healthy patients. Healthy patients with higher education experience slightly delayed atrophy, but not patients with neurodegeneration. The results were validated with congruent findings from the same study applied to the regional loss of gray matter density. This work calls for further exploration of the sexual dimorphism of AD regarding structural alterations.

Chapter 6

Influence of sex and APOE genotype on brain metabolism

Age, sex and APOE-ε4 genotype have been identified as the strongest predictors of the risk of developing Alzheimer's Disease (AD). This work models the pathological progression of regional brain hypometabolism, using mixed-effect models with latent time variable and longitudinal FDG-PET data. Statistical comparisons then disentangle the effects of sex and APOE-ε4 genotype on the onset age and pace of progression of hypometabolism in each brain region, while correcting for education level. They provide a brain map of the regions with earlier and/or faster alterations of the metabolism. We show that females are associated with faster hypometabolism in the caudate nuclei, the thalamus and right temporal and medial-occipital lobes, while APOE-ε4 is associated with earlier hypometabolism in the limbic system (hippocampus, parahippocampus and amygdala) and temporal lobe. This work was published and presented at the 2023 International Symposium on Biomedical Imaging (ISBI).

Contents

6.1 Introduction

6.1.1 Motivation

Alzheimer's Disease (AD) is a neurodegenerative pathology that accounts for more than 60% of the 55 millions dementia cases worldwide. A silent prodromal phase shows accumulation of alterations in the brain without specific symptoms, eventually leading to a progressive cognitive decline that translates into a loss of autonomy for daily tasks. Age, female sex and APOE genotype are the strongest three risk factors for AD [Riedel et al., 2016]. Firstly, AD incidence approximately doubles every five years after the age of 60, reaching a prevalence of almost 50% over 85 years old [Hebert et al., 2010]. Secondly, nearly two thirds of patients are female and it is documented that women's cognitive decline and neurodegeneration happen faster compared to men in AD, while being comparatively better preserved in healthy populations [Alzheimer's Association et al., 2016]. Lastly, The ε4-allele of APOE increases AD risk by approximately 4-fold when inherited in one copy and by greater than 10-fold for two doses of the allele [Tanzi, 2012]. Exhibiting different disease progression patterns across subgroups of patients, such as men vs. women or APOE mutations carriers vs non-carriers, can pave the way for precision medicine and

differentiated diagnosis and care.

Neuroimaging data allow the monitoring of the early stages of neurodegeneration before the onset of the symptoms and clinical diagnosis. Structural modalities such as MRI measure the regional loss of grey/white matter while functional modalities such as ^{18}FDG-Positron Emission Tomography (PET) or functional MRI assess changes in brain activity, unveiling distinct facets of the disease occurring before the clinical phase. Specifically, patients with AD have abnormally low PET measurements of the cerebral metabolic rate for glucose in the brain, and the patterns of hypometabolism are specific to AD [Mosconi, 2005, Herholz, 2010].

Since AD is a progressive disease, the collection of repeated measurements for each patient allows to leverage information about the dynamics of the alterations, otherwise impossible to notice in cross-sectional studies. Longitudinal models have been applied to the study of cognitive decline [Boyle et al., 2006, Koval et al., 2021b] and atrophy in regions of interest [Koval et al., 2021b, Bernal-Rusiel et al., 2013a]. Given a disease progression model, one can study the impact of risk factors on the dynamics of the observed alterations. However, longitudinal models have scarcely been applied to the progression of functional imaging data, and the influence of risk factors on alterations patterns has not been precisely described.

In this context, studying the influence of covariates such as sex, APOE genotype and education level on the longitudinal patterns of progression of regional brain hypometabolism could elucidate another aspect of the heterogeneity of AD. Mixed-effects models offer a suitable framework for the study of progression patterns since they describe each patient's trajectory with few interpretable parameters that can be compared for different groups of patients.

6.1.2 Related work

The impact of covariates such as sex and APOE status on cognitive decline and atrophy have been well studied [Ferretti et al., 2018, Laws et al., 2018, Subramaniapillai et al., 2021], highlighting steeper cognitive decline and neurodegeneration for women and different associations with previous conditions and modifiable risk factors such as BMI. Multiple cross-sectional studies have also established that men and women display different age-related metabolic decrease in healthy ageing and pathological progression [Zhao et al., 2016, Malpetti et al., 2017]. The menopausal transition seems to induce hypometabolism in the same regions as AD [Scheyer et al., 2018], potentially acting as a substrate for a neurological condition. It is also known that APOE$-\varepsilon 4$ carriers show steeper decrease of energy metabolism in the regions affected by AD [Mosconi et al., 2004]. In [Sampedro et al., 2015], authors show that the impact of APOE mutations is stronger for women than for men regarding hypometabolism and atrophy.

To the best of our knowledge no study has yet modeled the joint effects of sex and APOE genotype on the loss of energy metabolism in the brain for AD patients through the use of longitudinal FDG-PET databases. To summarize our contribution, we model the decrease in regional PET measurements using univariate mixed-effects models with an interpretable formulation that describes the onset and pace of decline. We assess the disentangled impact of each covariate on the onset and pace of decline for all anatomical regions of the AAL2 brain atlas. For reference, we compare our results to the previously reported associations between the mentioned risk factors on atrophic patterns across AD progression.

6.2 Materials and methods

6.2.1 Datasets and data processing

We used data from the Alzheimer Disease Neuroimaging Initiative (ADNI). We selected patients with a confirmed AD diagnosis and only kept patients with at least two visits with a FDG-PET scan. Regional measurements are obtained using the "pet-volume" pipeline of Clinica [Routier et al., 2021]. This pipeline first performs intra-subject registration of the PET image into the space of the subject's t1-weighted MR image using SPM12. The PET image is corrected for partial volume effects using the PETPVC toolbox, and then spatially normalized into MNI space using the DARTEL deformation model of SPM, and intensity normalized using the average PET uptake in the pons. Finally, the average PET uptake is computed for a set of regions obtained from the AAL2 atlas in MNI space.

	Men	Women
Patients	152	108
Visits	581	413
Follow-up (y)	2.8 ± 1.9	3.1 ± 2.0
Age at baseline	75.2 ± 6.7	73.8 ± 7.3
Education (y)	16.0 ± 2.9	15.2 ± 2.7
APOE (0/1/2)	53/74/25	32/56/20
MMSE	24.4 ± 3.9	23.5 ± 4.7
ADASCog13	25.7 ± 9.5	28.6 ± 12.0

Table 6.1: Cohort demographics. Numerical fields are in the form *mean* ± *std.* (0/1/2) represents the number of copies of the $\varepsilon 4$ allele. Cognitive measurements are for all visits.

6.2.2 Longitudinal modeling

Mixed-effects models describe each patient's progression over time as a variation – the random effects – around the average population trajectory – the fixed effects. This study focuses on univariate models, calibrated on the repeated measurements for one brain region at a time. Linear mixed-effects models are widely used for longitudinal modeling and assume that features evolve according to straight lines over time. Inter-patients variability is characterised by individual slopes and intercepts. The slope is the pace of decline of a patient for the given feature, but the intercept is less meaningful. In that setting, we parameterise the inter-subjects variability as the combined effects of an individual onset age, and a pace of decline as suggested in [Schiratti et al., 2015b]. The onset age is an horizontal intercept and describes the age at which a patient crosses a chosen threshold, and the pace of decline as the slope. Both these temporal parameters allow to align all patients on a common progression timeline. Namely, given a family of feature observations $\{y_{i,j}\}$ at times $\{t_{i,j}\}$ where $1 \leq i \leq N$ indexes the N patients and $1 \leq j \leq N_i$ indexes the N_i visits of patient i, we resort to the mixed-effect generative model

$$y_{i,j} = p_0 + v_0 \left(\alpha_i * (t_{i,j} - \tau_i) + t_0 \right) + \varepsilon_{i,j}$$

where α_i and τ_i, the *progression pace* and *onset age* of patient i, are called the individual parameters. Gaussian priors are chosen for the noise $\varepsilon_{i,j} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and random effects $\tau_i \sim \mathcal{N}(t_0, \sigma_\tau^2)$ and the log-acceleration $\xi_i = \log(\alpha_i) \sim \mathcal{N}(0, \sigma_\xi^2)$ so that progression paces are distributed around 1. It should be noted that the product of α_i and τ_i makes the time warp non-linear although the trajectories are straight lines.

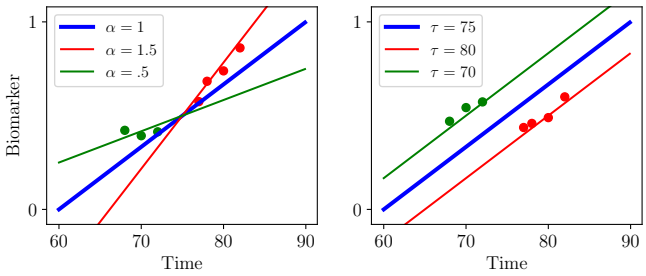


Figure 6.1: Illustration of the mixed-effect model with time-warp. Blue curve is the average trajectory (parameterised by t_0, p_0 and v_0) and red and green curves illustrate the effect of progression pace ($\alpha = e^{\xi}$) and onset age (τ) to match individual measurements (isolated dots) to a continuous trajectory.

6.2.3 Estimation and validation

These longitudinal statistical models are part of a family of geometric models [Schiratti et al., 2015a, Koval et al., 2017a]. Estimation can be performed using Maximum a Posteriori estimation of both random and fixed effects with the MCMC-SAEM procedure in which the estimation step of an EM algorithm is replaced by a stochastic approximation. See [Allassonnière et al., 2010] for details on the likelihood and proof of convergence and stability. The model and its estimation procedure are implemented in the open-source Leaspy software.

This family of models allows the use of logistic or exponential functions instead of straight-lines modeling in order to allow non-linearities [Koval et al., 2017a]. However, the estimated noise was not improved compared to linear functions. Parametric regression methods that learn the exact shape of the progression [Gruffaz et al., 2021, Sauty and Durrleman, 2022c] also exist but add a computational burden without improving the estimated noise of the fit. To further validate the quality of the fit, we use the baseline and follow-up scans of stable cognitively normal and amyloid-negative subjects in ADNI below the age of 75, as a proxy to test/retest data for measurement uncertainty (116 subjects, 232 visits with an average follow-up time of 16 months). Fig. 6.2 displays the distributions of this estimated test/retest and of the reconstruction errors of the longitudinal model. Distributions largely overlap, and reconstruction errors fall close to the measurements uncertainty, which hints that our models could not be improved without overfitting.

6.2.4 Statistical analysis

Once the univariate models are calibrated, we have a family of patient-wise onset ages and progression paces for each feature. We proceed to statistical testing of the differences between subgroups. We perform covariance analysis using ordinary least squares to regress the individual parameters with respect to sex, number of APOE-ε4 alleles and educational level of each patient. This evaluates the association of each covariate after correcting for the other ones, and also the significance of the corresponding association. We use False Discovery Rate to correct the p-values for multiple comparisons.

6.3 Results and discussion

APOE$-\varepsilon$4 mutation carriers display significantly earlier hypometabolism in the limbic system (hippocampus, parahippocampus, amygdala) and caudate nuclei. Female patient

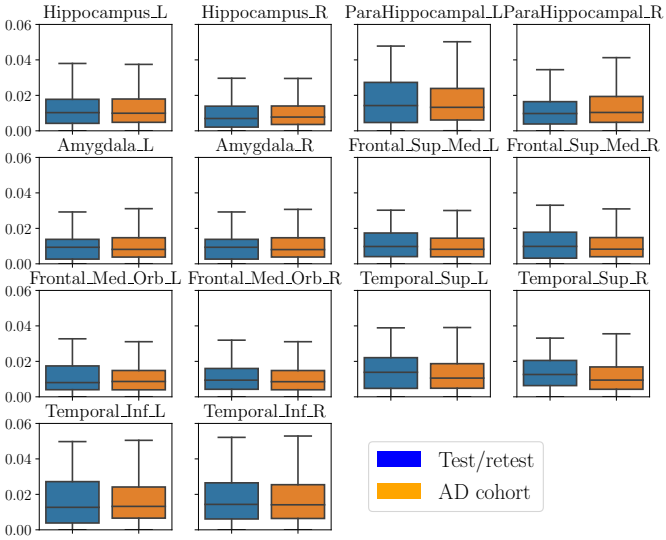


Figure 6.2: Distribution of reconstruction errors for the longitudinal model with measurements uncertainty for reference. For readability, only a subset of the 120 regions is displayed.

display faster progression of hypometabolism in the caudate nuclei, thalamus, right temporal lobe and right medial-occipital lobe. No significant association is recovered in any other parts of the brain. Table 2 reports the regression coefficients and significance for all regions with corrected p-value below the .05 threshold. The impact of sex seems to be lateralized with a stronger association on the cortex of the right hemisphere, while associations with APOE genotype are symmetric.

Higher education significantly correlates with earlier hypometabolism in parts of the frontal lobe, the entire temporal lobe as well as in the hippocampus, parahippocampus, amygdalas, pallidum and thalamus. In these regions, patients with highest education show average onset ages 8 to 12 years smaller than patients with the least education. Although counter-intuitive, these results are documented [Garibotto et al., 2008, Malpetti et al., 2017] and can be explained by the cognitive reserve hypothesis. This hypothesis stipulates that higher education and intellectual attainment allow patients to have better cognitive results for similar brain alterations. Since our cohort only keeps cognitively impaired patients with a confirmed AD diagnosis, it only features educated patients with a higher degree of functional alterations than their less-educated counterparts.

In [Sauty and Durrleman, 2022a], the same analysis is applied to t1-MRI scans in order to evaluate the effects of sex and APOE genotype on the dynamics of regional atrophy. Comparing both works reveals that brain hypometabolism in the pathological brain showcases more homogeneous progression profiles across the sub-groups than the atrophic patterns, especially regarding the impact of sex which proved to be the most important risk factor for brain atrophy. While higher education shows no significant correlation with atrophic patterns, it does display strong correlations with brain hypometabolism. It is known that regional glucose metabolic abnormalities are not solely the result of atrophy in Alzheimer's disease [Ibáñez et al., 1998] and different associations with risk factors are not surprising. The congruent findings between both studies are that the temporal lobe shows faster alterations for women and APOE-ε4 has a strong influence on the alterations in the limbic system.

A few methodological limitations need to be acknowledged. First, changing the brain

Associations between female sex and pace of decline

	coeff	p_{val}
Caudate_L	.18	2.9e-4
Caudate_R	.21	1.0e-4
Thalamus_L	.010	5.0e-2
Thalamus_R	.023	3.9e-2
Temporal_sup_R	.021	2.6e-3
Temporal_mid_R	.012	3.7e-3
Temporal_inf_R	.025	1.6e-2
SupraMarginal_R	.0084	1.9e-2
Occipital_mid_R	.079	2.8e-2
Fusiform_R	.012	7.6e-3
Lingual_R	.029	4.1e-2

Associations between APOE$-\varepsilon 4$ count and onset age

	coeff	p_{val}
Hippocampus_L	-3.2	9.1e-5
Hippocampus_R	-2.8	5.4e-5
ParaHippocampus_L	-3.1	5.1e-4
ParaHippocampus_R	-3.3	4.6e-8
Amygdala_L	-4.1	6.7e-5
Amygdala_R	-4.3	7.3e-5
Caudate_L	-1.6	5.8e-3
Temporal_Pole_Sup_L	-1.3	2.5e-2
Temporal_Pole_Sup_R	-1.7	1.3e-2
Temporal_Pole_Mid_L	-2.1	7.8e-4
Temporal_Pole_Mid_R	-2.5	3.0e-3
Temporal_Inf_L	-2.8	1.1e-2
Temporal_Inf_R	-2.5	3.2e-2

Table 6.2: Both tables report the coefficient of association as well as the corrected $p-$value for each region showing significant impact of the risk factor. First table reads for instance: "women show a steeper decrease in metabolism in the right thalamus by 2.3%" and second table reads for instance: "for each copy of $\varepsilon 4$ allele, the right hippocampus is expected to start to decline 2.8 years earlier". Female sex correlates with faster decline and $\varepsilon 4$ count correlates with earlier decline for all the reported associations. No significant association between sex and onset or APOE status and pace is reported.

atlas and spatial granularity might yield different results. One anatomical region that is positively associated with a covariate can span many ROIs or on the other hand be split between ROIs and prevent the detection of the true underlying association. Secondly, there is no such thing as test-retest for PET measurements and the proxy we used slightly overestimates the measurement error since even healthy patients show a decrease of glucose metabolism over time, which can explain why longitudinal errors are actually lower than the "measurements uncertainty" for most features. Lastly, no open-access cohort provides enough repeated FDG-PET scans to replicate the findings.

6.4 Conclusion

Building on the observation that risk factors such as sex, APOE genotype and education level influence the progression of AD pathology over time, we quantified their impact on brain hypometabolism regarding onset age and pace of decline. $\varepsilon 4$ alleles correlate with earlier hypometabolism in the limbic system and temporal lobes while female sex correlates with faster hypometabolism in the caudate nuclei, thalamus and right temporal and medial-occipital lobes. The sexual dimorphism of hypometabolism is weaker than that of atrophic patterns. These results provide insights into the not well-understood heterogeneity of AD and emphasize the need to provide care and analyse clinical trials results in a way that accounts for sex and APOE genotype.

Part IV

Multimodal Forecast of Cognitive Decline

Chapter 7

Forecasting Cognitive Decline with Multimodal Longitudinal Data

This study discusses the challenge of predicting cognitive decline in Alzheimer's disease (AD) patients. Multimodal medical data (e.g. MR and PET imaging, CSF measurements, clinical assessments) reflect different aspects of Alzheimer's Disease, including early changes in brain structure and function that can occur before the onset of the associated cognitive impairment. We propose to use a feature selection method within a disease progression model to identify the combinations of imaging and non-imaging biomarkers across modalities that allow the best predictions of the cognitive decline. We first demonstrate that the chosen non-linear mixed-effect model outperforms all benchmarked methods in the TADPOLE challenge, with increasing performance as various modalities are added. We then introduce a controlled protocol to compare the added value of each feature for the forecast of cognition, at different stages of the disease, and for varying time-to-predictions. Notable findings include that the volumes of the ventricles are predictive features at the later AD stages but not at early stages, hippocampal volume is mostly important for intermediate stages and cognitively unimpaired subjects, cortical thickness of temporal cortex is most important for short-term predictions in AD patients at any stages, and cortical summaries of glucose and amyloid PET uptakes are only useful for intermediate AD stages. These conclusions may inform the design of efficient prognosis scores that have been shown to decrease sample size in clinical trials and can be adapted to the targeted disease stages and the trial duration.

Contents

7.1 Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder that affects millions of individuals worldwide. It is characterized by an accumulation of pathological proteins and the loss of neurons in the brain that lead to a gradual decline in cognitive function, including memory loss and difficulty with language, reasoning, and perception. Despite its high prevalence and significant personal and societal impact, current diagnostic and treatment strategies are limited. Data-driven forecasting methods and enrichment strategies have the potential to substantially decrease the sample size and inform the design of targeted clinical trials [Maheux et al., 2023, Ballard et al., 2019]. In that spirit, disease progression models offer

prognostic tools, identify the disease stage best suited for clinical intervention and predict individual progression patterns.

Forecasting cognitive decline in AD patients is challenging due to the complex nature of the disease and the variability in its progression. Traditional diagnostic tools, such as cognitive testing and clinical evaluations, may not be sensitive enough to detect early changes in brain function. Multimodal imaging data, including structural and functional imaging, may be valuable in detecting early changes in brain function that occur before the onset of cognitive symptoms. Longitudinal data, which capture changes in the brain over time, are particularly informative in understanding the progression of the disease. However, the combinations of modalities that best forecast the decline at different stages of the disease have not yet been elucidated.

Disease progression models for biomarkers and imaging features can be formulated using differential equations [Ito et al., 2011], time-to-event models for discretized abnormality thresholds [Young et al., 2014], neural networks for scalar [Nguyen et al., 2020], imaging [Cui et al., 2019, Sauty and Durrleman, 2022b] and multi-modal data [Couronné et al., 2019], Gaussian processes [Lorenzi et al., 2019] and various regression frameworks. Well suited to irregularly spaced or missing data, a very flexible approach is proposed with mixed-effects models, which account for both the average trajectory of the population, called the fixed-effects, and individual variations to that trajectory that account for inter-subjects variability called the random-effects. Early models used linear modelling [Verbeke, 1997] while non-linearities were later added with polynomial [Wu and Zhang, 2002], logistic [Jedynak et al., 2012], and exponential [Raket, 2020] regressions. Such models can make prediction about future outcomes. However, little focus has been put on the selection of the most predictive features for cognitive forecast. In this study, we investigate the potential of multimodal longitudinal data in forecasting cognitive scores in AD patients using mixed-effects models, highlighting the most relevant combinations of modalities for each stage of the disease progression.

7.1.1 Related work

The automatic prediction of the patient's current diagnosis from multimodal data at the time of the acquisition has been extensively studied, with an emphasis on T1-MRI [Falahati et al., 2014, Leandrou et al., 2018] and more broadly neuroimaging data [Rathore et al., 2017, Arbabshirani et al., 2017]. The study of future conversion from MCI to AD, which is a more challenging and clinically relevant task, has also been tackled with a variety of approaches. [Ansart et al., 2021] proposes a quantitative review of such studies to exhibit the best-performing methods and set of modalities for this particular task. It concludes that for short-term predictions (<2y), predicting the patient does not change clinical status works as well as prediction models and that using FDG-PET data and cognition improves prediction while T1-MRI does not. This task, however, depends on the current clinical diagnostic practices that may not be reproducible among clinicians and countries and provides little dynamic information about the cognitive changes of patients. Predicting the cognitive outcomes at varying time-points in the future is of greater interest for clinical decision support systems. The most notable step in this direction was the introduction of The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) challenge[Marinescu et al., 2018] that uses data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) to provide a benchmark to compare the forecasting performance of more than 90 multimodal progression models. However, amongst the official submissions, no method significantly outperformed a simple linear mixed model for the prediction of cognitive decline [Marinescu et al., 2020]. The challenge has remained open and new meth-

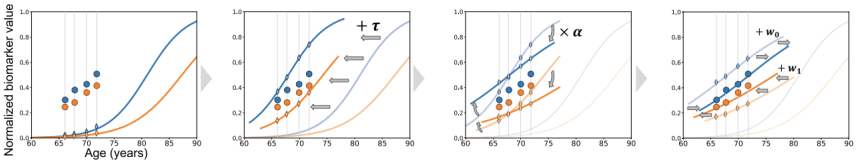


Figure 7.1: Illustration of the random effects "transforming" the average trajectory to account for the variability between patients, for a model with two features (blue and orange). Dots are the actual measurements and plain lines are the modelled trajectory. More details on this model are provided in [Schiratti et al., 2015a]. Figure courtesy of [Koval et al., 2021b].

ods have since provided better estimates, although the added value of each selected feature has not been assessed. In this work, we:

- demonstrate that a mixed-effect logistic model outperforms all benchmarked methods in the TADPOLE challenge, with increasing performance as imaging data from various modalities are added,
- describe a protocol to fairly compare the added value of each imaging feature for the forecast of cognition in different scenarios depending on the targeted disease stage and the time to prediction,
- exhibit, for each prediction task (a given cohort and a given prediction horizon), the features that are most useful or detrimental for the forecast.

7.2 Methodology

7.2.1 Longitudinal geometric model

Mixed-effects models describe each patient's progression over time as a variation – the random-effects – around the average population trajectory – the fixed-effects. As suggested in [Schiratti et al., 2017], the inter-subjects variability is parametrized as the combined effects of an individual onset age τ and pace of decline α – for the temporal variability – and a spaceshift w – for the intrinsic differences that are independent of time. An affine time-warp $t \mapsto \alpha(t-\tau)$ maps the age at visits to a pathological age on the common timeline of the average trajectory. These individual parameters form the random-effects, which we model with Gaussian priors. The reference positions, velocities, and time that define the average trajectory, together with the variances of the random-effects and residual noise, form the fixed-effects of the model. These models have also been demonstrated to generalize well to different databases [Maheux et al., 2023] and reach test-retest noise levels for the modeling of cognitive scores and regional imaging features [Koval et al., 2021b, Sauty and Durrleman, 2022a, Sauty and Durrleman, 2023], hinting that the models could not be improved without over-fitting.

7.2.2 Calibration and goodness of fit

These longitudinal models can be calibrated using a Maximum a Posteriori estimation of both random and fixed effects, using MCMC-SAEM procedure in which the estimation step of an Expectation-Maximization algorithm is replaced by a stochastic approximation. See [Kuhn and Lavielle, 2004] and [Allassonnière et al., 2010] for details on the procedure, description of the complete likelihood and proof of convergence and stability. These models have also been demonstrated to generalize well on different databases [Maheux et al., 2023] and reach test-retest noise levels for the modelling of cognitive scores and regional imaging

features [Koval et al., 2021b, Sauty and Durrleman, 2022a, Sauty and Durrleman, 2023], hinting that the models could not be improved without over-fitting. Once a model is calibrated and the fixed-effects are estimated for a given cohort, random-effects can be learned for unseen patients using a gradient-descent on the likelihood. In our case, we split the dataset in 5-folds and make prediction based on the first two visits for all patients of one fold using the calibration made on all visits from the other folds to prevent data leakage. All analysis is done using the public Leaspy software with a 2.3GHz CPU. Fitting one model and making predictions takes between 20" in a univariate setting and 10' for a multivariate model with 10 features.

7.2.3 Data processing

All analysis in this study is done on publicly available data from the ADNI database (adni.loni.usc.edu), which provides repeated MRI, FDG-PET and AV45-PET scans, as well as various clinical assessments for AD patients. Firstly, we demonstrate the usefulness of using multimodal data on the well-studied TADPOLE benchmark [Marinescu et al., 2018]. This challenge uses a subset of ADNI database to make monthly forecasts of the ADAS-Cog score that can be compared with the actual measurements. We use the processed features from the challenge data for the volumes (hippocampi and ventricles), thicknesses (enthorinal and fusiform cortex), and brain-average PET uptake (FDG and AV45). This challenge allows demonstrating that a model performs well on a real life dataset, but does not allow describing the most predictive features. In practice, the training cohort is very heterogeneous (1667 patients with 8068 visits total, including 615 AD, 594 MCI and 458 healthy patients), and the prediction tasks are not designed with a fixed amount of training visits per patients (7.6 ± 3.9 visits), nor for a fixed time-to prediction ($2.7 \pm .8$ years, with 219 predictions to make).

We thus select four homogeneous cohorts from the ADNI database using the baseline cognitive score and amyloid status based on Aβ42 measurements in CSF or average AV45-PET uptake [Hansson et al., 2018]. Cognitively unimpaired patients are defined as having less than median score ($<$14/85 for ADAS-Cog13) and mildly (resp. strongly) impaired patients as being below (resp. above) the third quantile ($<$ and $>$23/85 for ADAS-Cog13) of the overall distribution. We refer to those cohorts as A-/C-, A+/C-, A+/C+ and A+/C++. Subjects with A- and C+ or C++ are few and likely to be suffering from a variety of conditions, so they are not considered in our study. We process raw T1-MRI, FDG-PET and AV45-PET images using the Clinica software[Routier et al., 2021] that wraps common image processing pipelines to extract volumes (hippocampi, ventricles and striatum), average thickness of the temporal cortex, average FDG uptakes (temporal and occipital lobes) and brain-average AV45 uptake. Volumes are normalized by intracranial volume, and we also use p-tau protein concentration from CSF.

In practice, biomarkers have long been hypothesized to follow sigmoid shapes [Jack Jr et al., 2013] so we elect to use logistic trajectories and features should be normalized to the range $[0, 1]$. We thus discard outliers using the so-called three sigma rule for features that are not naturally bounded and add a min-max normalization within each cohort, providing realistic asymptotes. Since cortical thicknesses, brain volumes and FDG uptakes decrease over time, we flip data around .5, using the rotation $x \mapsto 1 - x$, in order to ensure increasingness, which is required for logistic modeling. For each cohort, we split the dataset in 5-folds and make predictions based on the first two visits for all patients of one fold, using the calibration made on all visits from the other folds.

	A+/C++	A+/C+	A+/C-	A-/C-
$N_{patients}/N_{visits}$	349 / 1621	253 / 1431	321 / 1788	428 / 2596
Age at baseline (y)	74.0 ± 7.6	74.4 ± 6.8	72.5 ± 6.7	71.1 ± 6.8
Average follow-up (y)	2.3 ± 1.8	3.5 ± 2.5	4.2 ± 3.0	4.8 ± 3.1
Sex (M/F)	193/156	173/80	135/186	203/225
ADAS-cog baseline	28.5 ± 7.5	17.1 ± 4.4	9.1 ± 3.8	8.2 ± 3.7
ADAS-cog last visit	39.0 ± 12.6	23.5 ± 11.8	12.4 ± 10.5	9.1 ± 5.1
N_{scans} (MRI/FDG/AV45)	1097/523/362	933/479/318	1018/474/583	1510/629/915

Table 7.1: Demographics for the four selected cohorts.

7.2.4 Statistical testing of the usefulness of each feature

For a given cohort, once a model is calibrated for each possible combination of features and predictions are made using the first two visits, we rank the models by quality of prediction using R2 scores for a given time horizon. We then compare the number of occurrences of each feature in the best and the worst-performing models. We use 8 features, which yield a total of 256 models, and compare the top and bottom 32 models. Under the null hypothesis that a feature does not improve nor worsen prediction, the number of models that use it should follow a binomial distribution under the normal approximation. We thus proceed to Welch t-tests that compare the proportion of models that use the feature in the best and worst performing models, providing a proxy for how much a feature improves or hurts predictive abilities.

7.3 Experiments and results

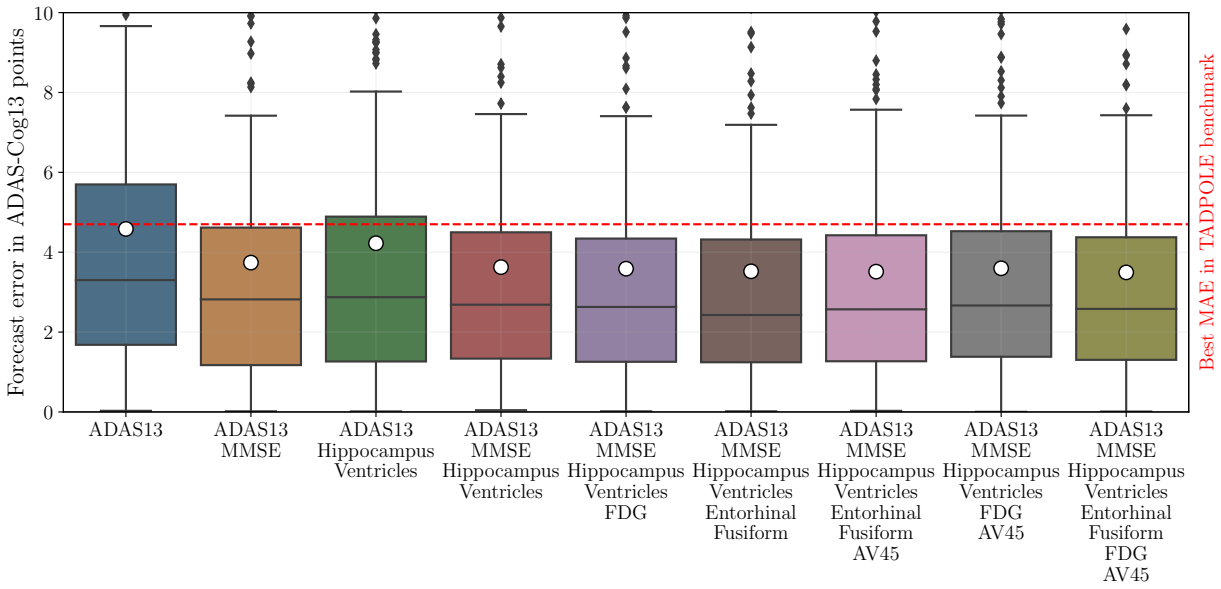
7.3.1 Benchmark on the TADPOLE challenge

As displayed in Fig. 7.2, we evidence that multivariate models perform better, both when adding a second cognitive score (MMSE) or imaging data, with the best-performing model being the one that features all modalities. However, it is likely that all modalities do not convey relevant information about neurodegeneration at the same disease stage, and the added value of some effects are mitigated by the heterogeneity of the cohort. Besides, some patients – especially the ones that are most healthy and decline the least – have a lot more visits to train on, driving the average error down. In order to rigorously assess the added value of each modality for the forecast of cognitive decline at the different stages of AD, a more controlled protocol is necessary.

7.3.2 Experiments on ADNI

Fig. 7.3 represents the average progression of the multivariate model with all the selected features for each cohort. Fig. 7.4 compares the best performing multivariate model to a univariate model that only uses the cognitive score and to a constant prediction model. Predictions are stratified based on the time to the last visit used for training (rounded to the unit). Fig. 7.5 displays, for each cohort and time horizon, the features that increase or decrease prediction accuracy.

For predictions at less than 2 years, the constant prediction outperforms progression models for the C- cohorts because of the very little change happening over such a period of time, which concurs with [Marinescu et al., 2020, Ansart et al., 2021]. For longer time-to-prediction, multimodal models always outperform univariate models, even more as



MAE	4.59	3.74	4.22	3.63	3.59	3.52	3.51	3.60	**3.49**
R2	0.53	0.66	0.50	0.67	0.67	0.68	**0.70**	0.68	**0.70**

Figure 7.2: Distribution of prediction errors for the TADPOLE challenge. Circles are the mean absolute error for each model. Below are the first and second order statistics that measure mean absolute error and explained variance. Models without any cognitive features display MAE of 7.0 and more, so we only display models that include cognition.

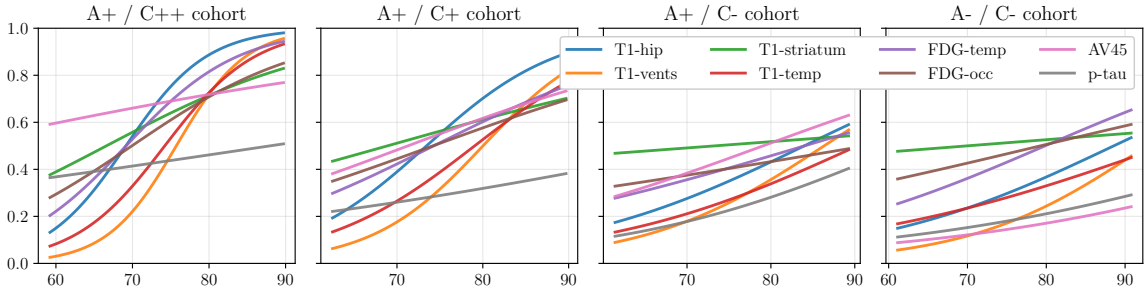


Figure 7.3: Average progression for the 4 learned models and corresponding cohorts. T1-MRI features are the volumes of hippocampi, ventricles, striatum, and thickness of the temporal cortex. FDG-PET features are the average FDG uptake in temporal and occipital cortex, and AV45 is the average AV45 brain uptake. p-tau is the concentration of phosphorylated tau protein in the CSF.

the patients are at an advanced disease stage. The R2 score is more informative of the quality of the model as it penalizes large errors more and represents the ability to explain the variability in the progression of patients, effectively separating the slow and the fast progressors, while the MAE can be very small for patients that do not progress a lot, while providing no specific information about each patient. For instance, the lower MAE for A-/C- and A+/C- cohorts correspond to similar R2 scores as the A+/C+ (see Table 7.1 for the average progression of each cohort).

In the most advanced stages of the disease (A+/C++), ventricles volumes and thickness of temporal cortex are the most useful features. In the A+/C+ cohort, PET and CSF data are significantly more retained in the best models, highlighting the usefulness of multimodality during the "core progression" of AD. In the A+/C- cohort, multimodality

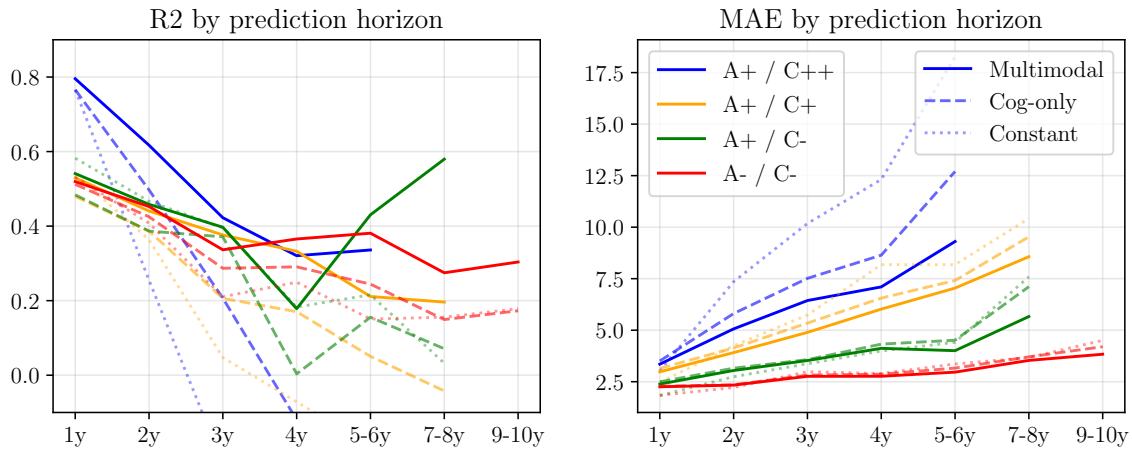


Figure 7.4: Best predictions across all models for each prediction horizons (y. is years). Amount of predicted visits are shown in the right table. Predictions task with sample size <30 visits are discarded.

	A+/C++	A+/C+	A+/C-	A-/C-
1y	382	298	288	385
2y	215	206	263	386
3y	93	122	124	181
4y	54	88	127	206
5-6y	53	95	177	268
7-8y	22	48	109	166
9-10y	3	18	24	56

Table 7.2: Amount of predicted visits for each task.

yields less information about the progression, and only the thickness of temporal cortex and ptau concentration significantly improve the forecast. The healthy A-/C- cohort, on the other hand, only uses the volume of the hippocampus to predict the (small) progression of patients. Interestingly, the ventricles are very useful for advanced AD patients but get progressively detrimental as cohorts are selected at earlier stages of the disease. It is also striking to see that gathering multimodal imaging from patients, at the cost of great expenses and inconveniences for the patient, does not yield much improvement for the early stages of the disease, while it is most useful for A+/C+ patients.

To validate those findings, we run the same pipeline with a standard linear mixed-effects model for prediction and recover the same feature selection patterns, except for the T1-striatum being more retained in the best models for all horizons in the A-/C- cohort. Besides, we decided to rank models by R2 scores for statistical testing, but it could be argued that progression models should minimize the mean error instead of the tail of the distribution. The same figure obtained by maximization of the MAE score yields the exact same trend, with only slight variations in the t-values. Besides, selecting 16, 32, 64 or 128 of the best/worst models do not change the trends of usefulness of features.

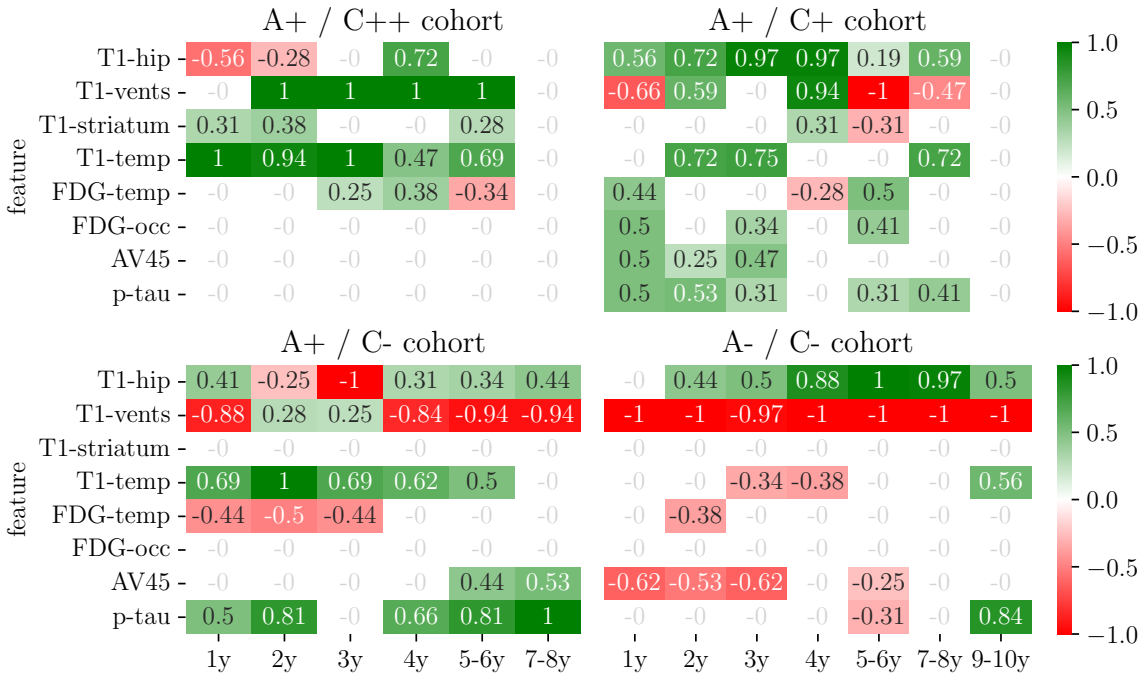


Figure 7.5: Difference in proportion of models that feature each outcome in the best and worst models for a given cohort and time to prediction. 1 means the feature is selected in all the best models and none of the worst, while -1 is the opposite. Non-significant differences are not displayed. Green means that features are significantly more present in the models with best R2 scores, and red means significantly less.

7.4 Discussion and conclusion

Several methodological limitations need to be acknowledged. Firstly, as is typical of neuroimaging databases, the visits we use do not all feature each selected modality, which can in turn mitigate the added-value of choosing a feature in our model (see Table 7.1 for information on the amount of scans for each modality). Secondly, it must be noted that this study used a selection of 8 features that are plausibly associated with AD, but further work that scales this statistical pipeline up for a wide variety of available features without pre-selection is underway. It must also be noted that patients with very long follow-up periods are usually less impaired, which can bias the prediction scores and selected features for the long-term predictions.

As [Maheux et al., 2023] demonstrates, improving the forecast of cognitive decline allows to drastically decrease sample size for clinical trials and could help provide personalized care and prognosis. Understanding the features that should be used and are worth acquiring for each population is a crucial first step in that direction.

Conclusion and perspectives

Conclusion

This thesis aimed to investigate the link between brain alterations and cognitive loss throughout Alzheimer's Disease. We tackled this question from a "disease modelling" perspective, with the goal to develop and apply bespoke models to multimodal longitudinal data. Namely, this work illustrates how the well-validated framework of generative mixed-effects models can be engineered to model biomarkers and full-scale images in a non-linear fashion over time. Besides, estimating the natural course of a disease at both the average and individual level, from individual measurements, reveals insights about the influence of risk factors, about the heterogeneity of disease progression and allows forecasting short and long-term prospects for cognitive declines.

Learning the shape of trajectories of biomarkers

Our first contribution alleviates the need for *prior* assumptions on the shapes of the trajectories of biomarkers over time. In practice, it is often hypothesized that the biomarkers follow either a logistic or linear trend over the course of the disease. We allow a geometric mixed-effect model to learn the shape of the curves from the data itself, which allows modelling very complex non-monotoneous trajectories with plateaux, multiple inflexion points or asymptotes for biomarkers that do not have set minimum and maximum values. This work showed that, although it provides more interpretable trajectories of individual progression, learning the geometry of the observation space does not significantly improve the prediction accuracy, hinting that the logistic hypothesis is somewhat valid. Besides, our model recovers almost perfect sigmoid shapes when modelling cognitive scores, which further solidifies the "logistic hypothesis" for the modelling of bounded cognitive scores.

Adapting the framework to model neuroimaging data

One other limitation of the standard mixed-effects models was their inability to effectively handle high-dimensional data. Indeed, voxel-wise modelling miss important interplay between region and algorithmic complexity is also very high for most approaches. In the context of AD, it is particularly unfortunate, since neuroimaging scans reveal structural and functional alterations in the brain before the onset of cognitive symptoms. We thus embedded a dimensionality-reduction neural network within a mixed-effect model in order to learn the normal and individual progression of brain images over the course of AD. Our approach, called "Longitudinal Variational Autoencoders" recovered known patterns of alterations regarding brain atrophy and loss of brain metabolism. The major limitation of this approach is that successive scans show little change (the intra-patient variability) over time when compared to the noise of the dimensionality reduction tool (the reconstruction error of the auto-encoder), and the temporal information can be lost in the compression of the images. For clinical MRI and PET images, the model is bound to estimate the average progression and can not yet accurately forecast subject-wise future images. This is likely due to the fact that, even for a low-dimensional latent space of 16 dimensions, the amount of available scans is too low to densely populate the latent space. It should be noted, however, that for simple synthetic data, individual trajectories were perfectly estimated, which hints that progression models for images using LVAEs are promising but require more fine-tuning and training data for clinical applicability.

Using disease progression models to describe the heterogeneity of AD

The important perk of mixed-effects models for progression modelling is that they learn the average progression of the disease, and recover individual trajectories as variations of that normative scenario. It allows fair comparisons between individuals regarding disease stage and presentation. Namely, we investigated the influence of the most important risk factors: female sex, APOE-ε4 genotype and low education level, on the patterns of structural and functional alterations. To do so, we describe the impact of each covariate on the onset and pace of regional atrophy and hypometabolism in the brain over the course of AD. This reveals a strong sexual dimorphism in disease presentation, and suggests a much faster atrophy for women than for men, but not an earlier onset.

Combining modalities in order to forecast cognitive declines years ahead

The last part of this work questioned how informative each modality of data is, for the forecast of cognitive decline, at different stages of the disease. Indeed, each modality is best suited to describe certain facets of the disease, and the precise combination of features that best predict future cognitive decline has not been elucidated. Through a process of model selection, we showed that the best combinations vary depending on the stages of the disease. This result is useful for providing guidelines regarding the modalities of data worth acquiring in future studies according to its objectives. It can also significantly decrease sample sizes in clinical trials by adjusting the outcome with a prognosis score with higher specificity and sensibility.

Perspectives

Several research axes naturally arise as possible continuations to the work that was presented in this thesis. The proposed ideas mostly emerge from the hardships of handling small datasets of high-dimensional neuroimaging data, of extracting relevant biomarkers, and of integrating different modalities together for the best possible description of AD progression.

Quantify the limits of using full images or processed features

Full-scale images provide extensive and detailed information regarding structural and functional changes in the brain. However, their utilization presents challenges in terms of data processing, storage, computational requirements, and the need for large datasets. Consequently, the scalability and feasibility of employing full-scale images in extensive studies or clinical settings may be constrained. In contrast, derived biomarkers serve as condensed representations of pertinent information extracted from the images, such as volumes or cortical thicknesses, facilitating handling and analysis. Despite the widespread use and validation of tools like Freesurfer for deriving biomarkers, they still rely on assumptions and algorithms that can potentially introduce biases or inaccuracies in the resulting measures. The accuracy of these derived biomarkers is influenced by various factors, including image quality, segmentation algorithms, and preprocessing techniques.

To summarize, full-scale images can be valuable for investigating regional-specific abnormalities or capturing complex patterns that might not be fully captured by derived biomarkers alone. Conversely, using derived biomarkers offers a more focused and quantifiable assessment, facilitating statistical analysis and interpretation. Both approaches have practical limitations when it comes to data management, generalization, and interpretation. Biomarker-based study claim to alleviate the need for "black-box" deep learning approaches and huge training sets, while image-based studies claim to alleviate the need

for computationally heavy and lossy processing software programs. It would be a very significant contribution to the field of neuroimaging to estimate the "signal-to-noise ratio" of both approaches for disease progression modelling and predictive power for future forecast. Such a study would require defining a benchmark of tasks, and evaluating both approaches in parallel. For instance, one could evaluate how the models react to data subsampling or noise injection to assess the stability of both data pipelines, do interpretability analysis to see what specific patterns associated with disease progression are revealed for each model, and provide guidelines as to which approach is best suited depending on the task at hand and the amount of available data.

High-dimensional data modelling

When modelling the progression of full scale images, a variety of dimension-reduction tools can be used. This amounts to extracting "data-driven biomarkers", that may be more complex than just the volume of specific regions, and modelling their progression. In this thesis, we resorted to autoencoder neural networks for the extraction of latent biomarkers, but several other approaches –linear and non-linear– can be used. For instance, techniques such as Principal Component Analysis (PCA) or Independent Component Analysis (ICA) decompose the variability between images in different components and may allow separating the inter-patients and the intra-patient temporal variability. Other approaches such as Non-negative Matrix Factorization, Sparse Dictionary Learning or Manifold Learning may also extract interesting low-dimensional features from images, without priors. Comparisons between the different types of dimensionality-reduction tools have been proposed in the context of image segmentation or brain age estimation, but to my knowledge no such study has been proposed in the context of progression models for brain images. A proper benchmark of the strengths and weaknesses of each approach would be very insightful. For instance one could, in the spirit of Chapter 3, plug a well-validated disease progression model at the end of a dimension-reduction pipeline and jointly estimate both parts using a longitudinal neuroimaging dataset, and compare the performance of each approach regarding generalizability, reconstruction error, prediction abilities, interpretability and computational efficiency.

Besides, this research axis is linked to the first axis: FreeSurfer primarily outputs meshes that describe the surfaces that separate the distinct brain structures, from which the cortical thicknesses are computed. This mesh is sometimes used in conjunction with progression models in order to describe the alterations of the cortex. Mesh points still constitute high-dimensional representations of the data, and dimension-reduction techniques can be applied similarly to full images. Comparing such "half-processed" models to other dimension reduction tools can also be helpful in understanding the benefits of full-scale images compared to using prior knowledge for biomarkers extraction.

Using transfer learning

As previously mentioned, models that use full-scale images suffer from the need for large training datasets. In the context of deep learning, an autoencoder fed with only a few thousand samples with a million dimensions is extremely under-powered. Unfortunately, longitudinal neuroimaging datasets rarely feature more than a few hundred patients. One way to circumvent this issue could be to use transfer learning to help the network extract relevant latent representations. This can be done either through generic training on unrelated datasets such as ImageNet, or using large medical imaging datasets unrelated to AD, such as the UK BioBank. For instance, pre-training a network to reconstruct a vast amount of MRI scans or predict a simple known variable such as the sex of the patient, even if not from AD patients, could improve the network's convolutional weights after fine-tuning on

the small AD datasets. Literature reviews point out that transfer learning improves diagnosis prediction, segmentation, and classification abilities from neuroimaging data, and list the strengths and weaknesses of each approach [Agarwal et al., 2021, Ardalan and Subbian, 2022]. However, such approaches have not yet been properly implemented in the context of disease progression models. This axis poses significant challenges regarding deep-learning routines and was outside the scope of this thesis, but it would be an important direction to explore to understand the full potential of imaging data for the description of AD, and to circumvent the intrinsic limitations in dataset sizes.

Integrating multimodality

Modelling multiple modalities simultaneously in the context of neuroimaging is a challenging task and, as was done in Parts II and III, different modalities are often modelled independently, focusing on one at a time. This approach overlooks potential synergistic relationships and interactions between different modalities that could provide a more comprehensive understanding of brain alterations. To improve our modelling efforts, a more integrated and holistic approach is needed. One possible strategy is to develop multimodal fusion techniques that can leverage the complementary strengths of each modality and capture the complex interactions and dependencies between them. The use of deep learning architectures, such as multi-stream or graph-based models, could facilitate the integration of multiple modalities by explicitly considering their relationships and shared representations. In approaches such as in Parts I and IV, multimodal biomarkers data can be integrated together, although processing is required to extract scalar values from imaging, CSF or clinical data.

An important challenge in integrated multimodal modelling of neuroimaging data is the lack of simultaneous measurements for all modalities during most visits, as well as the limited availability of longitudinal measurements for certain modalities. This limitation is particularly evident in the case of tau-PET and amyloid-PET, which are notoriously missing from this thesis, due to the scarcity of repeated measurements and overall small sample sizes. Furthermore, the presence of multiple amyloid tracers, such as florbetapir, florbetapen, and PiB, adds to the heterogeneity of the data. While efforts have been made to extract biomarkers from the images and normalize the values for average brain tracer uptakes, such as the "centiloid scale" for amyloid uptake, achieving homogenization becomes more challenging when dealing with full-scale PET images. Addressing this aspect would be of significant importance for further advancements in this research field.

An alternative and potentially promising approach to improve the forecast of future cognitive decline from multimodal longitudinal data is through ensemble modelling. This technique involves making predictions separately from each modality using a Disease progression model. By leveraging the predictive power of each modality individually, a more comprehensive understanding of cognitive decline can be achieved. To account for the interplay between modalities, a traditional regressor can be employed to combine the predictions from each modality and generate a more accurate estimate of the expected decline for a specific patient. Although this approach partially considers the interactions between modalities, it offers a practical solution for integrating information from different sources and improving the accuracy of cognitive decline predictions.

Valorization

Scientific publications

Journal articles

- [Sauty and Durrleman, 2022a] Benoît Sauty and Stanley Durrleman. **Impact of sex and APOE-ε4 genotype on patterns of regional brain atrophy in Alzheimer's Disease and healthy ageing**. In *Frontiers in Neurology, 2023*

Peer-reviewed conference papers

- [Sauty and Durrleman, 2022c] Benoît Sauty and Stanley Durrleman. **Riemannian metric learning for progression modeling of longitudinal datasets**. In *Proceedings of the International Symposium on Biomedical Imaging (ISBI) 2022.*

- [Sauty and Durrleman, 2022b] Benoît Sauty and Stanley Durrleman. **Progression models for imaging data with Longitudinal Variational Auto Encoders**. In *Proceedings of the International Conference on Medical Image Computation and Computer Assisted Intervention (MICCAI) 2022.*

- [Sauty and Durrleman, 2023] Benoît Sauty and Stanley Durrleman. **Impact of sex and APOE-ε4 genotype on regional brain metabolism in Alzheimer's Disease**. In *Proceedings of the International Symposium on Biomedical Imaging (ISBI) 2023.*

- [Hassanaly et al., 2023] Ravi Hassanaly, Simona Bottani, Benoît Sauty, Olivier Colliot and Ninon Burgos. **Simulation-based evaluation framework for deep learning unsupervised anomaly detection on brain FDG PET**. In *Proceedings of SPIE Medical Imaging 2023.*

Preprints

- Benoît Sauty, Etienne Maheux and Stanley Durrleman. **Forecasting Cognitive Decline in Alzheimer's Disease using Multimodal Longitudinal Imaging**.

Talks

- **Progression models for imaging data with Longitudinal Variational Auto Encoders**. In : *International Symposium of Clinical Biostatisticians, ISCB* August 2022. Newcastle, United Kingdom.

- **Longitudinal Variational Autoencoders learn a Riemannian progression model for imaging data**. In *Workshop on Geometric Deep-Learning for Neuroimaging Data 2022* November 2022. Amsterdam, Netherlands.

- **Impact of sex and APOE-ε4 genotype on patterns of regional brain atrophy in Alzheimer's Disease and healthy ageing** International conference on Alzheimer's Disease and Parkinson's Disease (ADPD) March 2023. Goteborg, Sweden.

Awards

The article **Riemannian metric learning for progression modeling of longitudinal datasets** won runner-up **best paper award** at the ISBI 2022 conference.

Carbon footprint

The concept of carbon footprint has become increasingly relevant in recent years. It refers to the total amount of greenhouse gas emissions caused directly or indirectly by an individual, organization, or product. In this section, we will estimate the carbon footprint of the research work carried out over the course of three years for this doctorate. This will provide insight into the impact of our work on global warming and allow us to assess the potential for reducing our carbon footprint in the future. By estimating our carbon footprint, we hope to raise awareness and encourage more sustainable practices in academic research. Emissions linked to transportation are estimated using the french database ADEME (`https://bilans-ges.ademe.fr/`) and the traveled distance for each transportation mode (long-haul flight, short-haul flight and high speed train). Emissions from the computing resources are evaluated using the "Machine Learning Emissions Calculator" (`https://mlco2.github.io/impact/#compute`) from the duration of training (hours), energy consumption of the used hardware (kW) and the carbon intensity of electricity (CO2-eq/kWh). All the estimates for the office emissions are based on annual data provided by an inside commission from the institute where offices are located (Paris Brain Institute), using the "labo1.5" estimator (`https://labos1point5.org/`). The goal is not to provide the exact amount but to provide reasonable orders of magnitude and highlight the expense that could be systematically reduced.

		Emissions (kg CO2-eq)	% of emissions
Plane	Paris-Singapore	3.500	28.2 %
	Paris-Copenhagen	350	2.8 %
Train	Paris-Amsterdam	16	0.1 %
	Copenhagen-Goteborg	14	0.1 %
	Paris-Milan	22	0.2 %
Computing	GPU (∼1500h)	30	0.2 %
	CPU (∼5000h)	60	0.5 %
Office	Electricity	1.000	8.1 %
	Heating/AC	3.600	29.1 %
	IT hardware	3.800	30.7 %
Total		12.392	100 %

Table 7.3: Rough estimates of the carbon footprint of this doctorate. For reference, ∼2000kgCO2-eq/year/person for all emissions (private and professional) is a symbolic threshold believed to be sustainable at the global scale.

The low carbon intensity of french electricity (∼ 70gCO2-eq/kWh) makes the computing resources specific to our model training negligible compared to the manufacturing and operating cost of all IT equipment (e.g. laptops, displays, network infrastructures, servers). Reducing the infrastructure is not feasible without impacting the research, but cutting on heating and AC is very plausible as the working conditions were very comfortable to say the least. Emissions linked to food are not taken into account as they are arguably not a consequence of this research and are not a systematic burden of this research domain. Besides, being located in central Paris, transportation to and from the workplace are almost carbon-free when using a bicycle.

The elephant in the room is obviously air travel, that amounts for one third of emissions in just a few days of conference. Had I elected to go in-person to ISBI 2022 in Calcutta, India, to ISCB 2022 in Newcastle, UK, to ISBI 2023 in Cartagena, Colombia and MICCAI 2023 in Vancouver, Canada, the carbon footprint of air travel would reach 11.6T of equivalent CO2 and 58% of all emissions. In-person conferences are often incredibly valuable and it is very detrimental that it comes at such a price regarding environmental issues. Frequent intra-continental conferences, of a more reasonable scale, with occasional inter-continental mega-conferences offer a more durable alternative to the current system that insidiously incites young researchers to attend multiple global conferences each year. The fields of ML are prone to such excess while the density of researchers allows continental conferences to be held with great success. Putting emphasis on building "local" communities at the scale of continents could help decrease emissions by almost 50% for researchers that live near high-intensity research clusters.

Financial acknowledgments

This doctorate was financed by the Paris AI Research Institute (PRAIRIE 3IA Institute) under grant number ANR-19-P3IA-0001. The work was also funded in part by the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-JPW2-000 (E-DADS) and ANR10-IAIHU-06 (IHU ICM), as well as by the European Research council reference ERC-678304 and the H2020 programme via grant 826421 (TVB-Cloud).

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.;Cogstate;Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies;Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

A – Supplementary materials for Section 5

Reconstruction error and comparison with acquisition noise

In Fig. 7.6, the distributions of fit error of the three cohorts (CN, MCI and AD) are compared to the acquisition and processing noise of subcortical volumes. In Fig. [7.7,7.8,7.9], the same errors are displayed for cortical thicknesses.

The overlapping of the distributions hints that the models could not improve the reconstruction error without over-fitting. For all features, reconstruction errors gets slightly larger as the selected cohort gets more cognitively altered. That could be attributed to the considered brain regions being smaller due to a higher level of neurodegeneration, and thus yield a higher error in parcellation. That could also explain why for some features, the reconstruction errors of the CN cohort are marginally lower than the re-test errors that are computed with patients at all stages of brain atrophy.

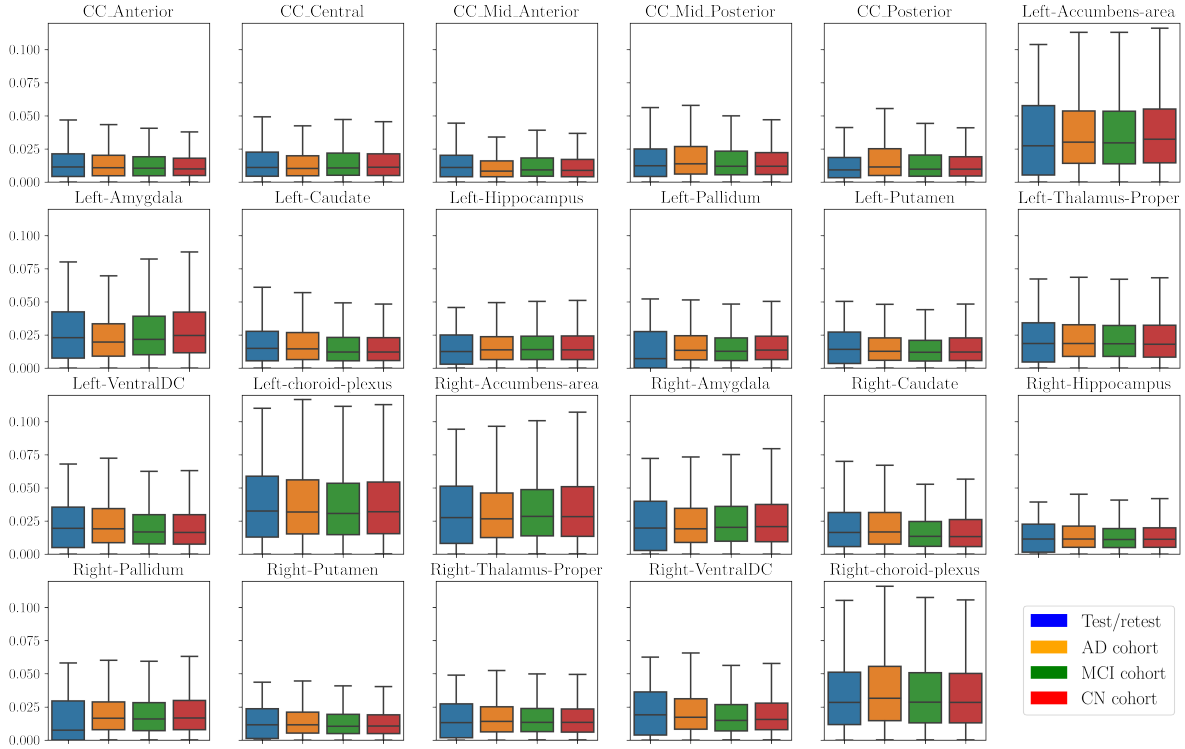


Figure 7.6: Comparison of measurement noise (blue) and fit errors (AD: orange, MCI: green, CN: red) for subcortical volumes.

Longitudinal scatterplots

In order to illustrate the relevance of the chosen non-linear mixed-effect model with a temporal reparametrization, we display the scatterplots of repeated measurements (4.4 visits per patient on average), both with regard to clinical patients' age, and to the learned pathological age. Due to the large amount of cortical regions for each cohort, we chose to only display a fraction of the features for the AD cohort, in order to improve readability.

Spaghetti plots that correspond to the same data are also provided to illustrate the longitudinal aspect of the data. Each colour represents one patient. The noisiness of the data is visible.

Bias correction for MRI field strength

As mentioned in the main article, cortical thickness measurements from FreeSurfer are known to be biased with regard to the field strength of the MRI scanner. As is common in longitudinal designs, in which data from 1.5 and 3T scanners are pooled, we remove the additive bias by matching the means of the distributions. The cortical thickness measurements presented below are normalized and flipped around .5 in order to be increasing over time. This means that 3T scans yield higher thicknesses on average for all the considered features, which is consistent with the literature.

Before homogenization, patients with visits that use different field strengths display significantly higher reconstruction errors from the longitudinal model, for most regions, while the same models after homogenization yield unbiased errors with regard to field strength. This hints that the homogenization removes the intra-patient's discontinuities that arise from varying field strengths.

Results for the gray matter density maps

This secondary analysis is done on the same cohorts from ADNI. Table 7.4 recalls the demographics and amount of available scans for each model.

	AD cohort		Aβ+ MCI cohort		Healthy cohort	
	Male	Female	Male	Female	Male	Female
Patients (N)	329	261	262	176	210	264
Visits (N_{scan})	1,447	1,092	1,164	794	1,063	1,169
Total follow-up (y)	3.7 ± 2.9	3.4 ± 3.0	4.5 ± 3.4	4.6 ± 3.3	4.6 ± 3.3	4.1 ± 3.0
Age at baseline	75.0 ± 7.1	73.6 ± 7.7	75.4 ± 7.5	71.5 ± 7.8	74.0 ± 6.1	72.7 ± 6.0
Education (y)	16.1 ± 2.8	14.7 ± 2.6	16.4 ± 2.8	15.5 ± 2.7	17.2 ± 2.4	16.1 ± 2.8
APOE-ε4 (2/1/0)	69/163/97	44/133/84	32/99/121	24/71/81	5/51/154	7/75/182
MMSE	24.0 ± 4.0	23.2 ± 4.6	27.7 ± 2.0	28.2 ± 2.1	29.0 ± 1.2	29.2 ± 1.1
ADAS-Cog13	26.5 ± 10.6	29.2 ± 11.8	15.3 ± 6.9	12.6 ± 6.7	10.0 ± 4.6	8.3 ± 4.3

Table 7.4: Demographics for cohorts selected from ADNI. Numeric fields are in the form mean ± standard deviation. (y) is years. (2/1/0) refers to the amount of APOE-ε4 alleles. Cognitive scores are for all visits (not just baseline).

Here we report the figures for loss of grey matter in cortical regions and subcortical volumes defined by the AAL2 atlas. Since the AAL2 atlas is a volumetric atlas, data cannot be projected onto a reference cortical surface and make visualization more tedious. We used the fsleyes software to display the lateral views, but could not provide a medial view. Significant associations for the AD cohort are reported in Fig. 7.14, for CN cohort in Fig. 7.15 and for MCI cohort in Fig. 7.16. We use the same conventions for plotting the signed log p-values as in the main manuscript. Table 7.5 reports the associations found for subcortical atrophy.

In order to facilitate comparison with the main analysis, we present results with the same structure and emphasize the differences at the end of each paragraph.

		Correlations with sex			Correlations with APOE-ε4		
		p_{val} CN	p_{val} MCI	p_{val} AD	p_{val} CN	p_{val} MCI	p_{val} AD
Amygdala (L)	Onset	-	-	-	-	4.8e-02	3.7e-07
	Pace	-	-	4.0e-06	-	1.0e-04	2.8e-08
Amygdala (R)	Onset	-	-	-	-	4.9e-02	2.8e-08
	Pace	-	-	2.0e-06	-	1.4e-04	1.6e-06
Caudate (L)	Onset	3.7e-09	9.5e-03	2.0e-03	-	-	-
	Pace	-	-	-	-	-	-
Caudate (R)	Onset	1.2e-09	4.3e-3	1.5e-04	-	-	-
	Pace	-	-	-	-	-	-
Hippocampus (L)	Onset	4.8e-06	-	-	-	1.8e-02	6.4e-08
	Pace	-	-	1.3e-02	-	6.8e-04	2.7e-06
Hippocampus (R)	Onset	1.2e-03	-	-	-	3.9e-02	8.5e-07
	Pace	-	-	1.7e-03	-	2.9e-04	5.0e-07
ParaHippocampus (L)	Onset	-	-	-	-	-	5.6e-04
	Pace	-	-	1.2e-02	-	1.9e-02	7.5e-04
ParaHippocampus (R)	Onset	-	-	-	-	-	1.5e-05
	Pace	-	-	1.9e-02	-	1.3e-02	2.0e-04
Putamen (L)	Onset	1.9e-07	1.8e-02	1.0e-03	-	-	-
	Pace	-	-	1.2e-02	-	-	-
Putamen (R)	Onset	3.3e-07	1.3e-02	1.2e-04	-	-	-
	Pace	-	-	8.3e-3	-	-	-
Thalamus (L)	Onset	6.3e-09	1.3e-05	4.5e-05	-	-	2.2e-02
	Pace	-	-	-	-	-	8.9e-04
Thalamus (R)	Onset	7.0e-08	1.6e-05	6.3e-04	-	-	-
	Pace	-	-	-	-	-	-

Table 7.5: Significant correlations with subcortical structures' atrophic dynamics. For all the significant features, APOE-ε4 genotype are correlated with lower onset age and higher pace, and female sex is correlated with higher onset age and higher pace, although correlations with pace are rare and only present in AD and MCI cohorts. Pallidum is discarded as it shows no significant correlation.

Impact of sex on cortical thinning for AD progression Almost all regions of the cortex display a significantly higher acceleration factor for female patients, with an emphasis on the entire temporal, frontal and occipital lobes and the anterior parietal lobe. Onset ages are, on the other hand, more homogeneous across sexes. The motor cortex, the cingulate gyrus and the medial parietal lobe display an earlier onset for men, while the inferior occipital lobe displays an earlier onset age for female patients. The impact of sex on loss of gray matter is thus almost identical to that of sex on cortical thinning for AD patients.

Impact of APOE-ε4 genotype on cortical thinning for AD progression With the exception of the postcentral gyri, almost all regions of the cortex also display a significantly higher acceleration factor for APOE-ε4 carriers. The effect is overall lower than that of the influence of sex. The temporal lobe, the parietal lobe and the frontal lobe also display an earlier onset for APOE-ε4 carriers. Once again, the results are similar to those of the main analysis.

Impact of sex on cortical thinning for healthy aging Contrary to what was seen in AD progression, the sexual dimorphism for healthy ageing manifests mainly through a significantly earlier onset age for male subjects, especially in the motor cortex, but also in the parietal and frontal lobes. On the other hand, acceleration factors are not statistically distinguishable across sexes, except for a few isolated regions that barely display threshold-level of significance. The findings for the onset age are in line with the main analysis, but the pace of loss of gray matter slightly differs, possibly because of the small effect sizes.

Impact of APOE-ε4 genotype on cortical thinning for healthy ageing APOE-ε4 allele carriers do not display significantly different patterns of loss of gray matter in the cortex across healthy ageing, in line with the findings on cortical thinning.

Influence of both covariates for patients with MCI The onset ages display correlations with sex that are similar to those displayed for healthy ageing and correlations with APOE-ε4 status that are similar to those displayed for AD progression, although of a weaker effect. On the other hand, the acceleration factors are significantly higher for men in almost all the cortex, and significantly higher for APOE$-\varepsilon$4 carriers in the temporal, frontal and parietal lobes. The associations with onset are in line with the main analysis, but the associations with pace of loss of gray matter are novel.

Correlations with the patterns of atrophy for subcortical structures For healthy ageing, male sex correlates with an earlier onset age for most regions with no significant differences in pace of atrophy, while no significant correlation in found for APOE-ε4 genotype. For the AD cohort on the other hand, female sex correlates with an earlier onset for most regions, and higher paces of the hippocampi, parahippocampi, and bilateral amygdalae and putamens while APOE-ε4 genotype correlates with earlier onset and higher pace for the hippocampi, parahippocampi, bilateral amygdalae and left thalamus. We reach the same conclusion regarding the MCI cohort as for the main analysis : sex correlates with loss of gray matter similarly to the healthy cohort, while APOE-ε4 genotype correlates similarly to the AD cohort.

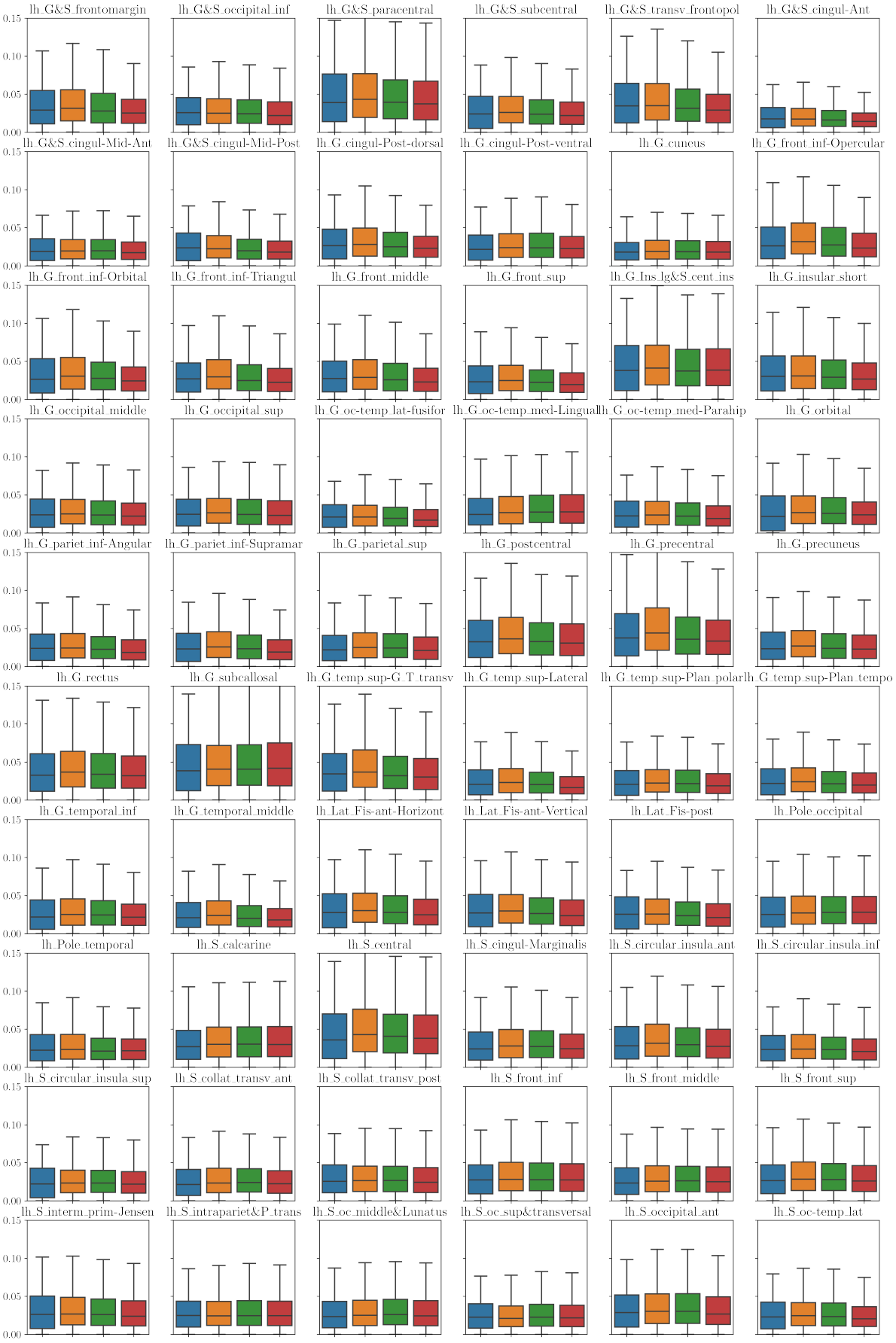


Figure 7.7: Comparison of test/re-test measurement noise (blue) and fit errors (AD: orange, MCI: green, CN: red) for cortical thicknesses. Part 1/3.

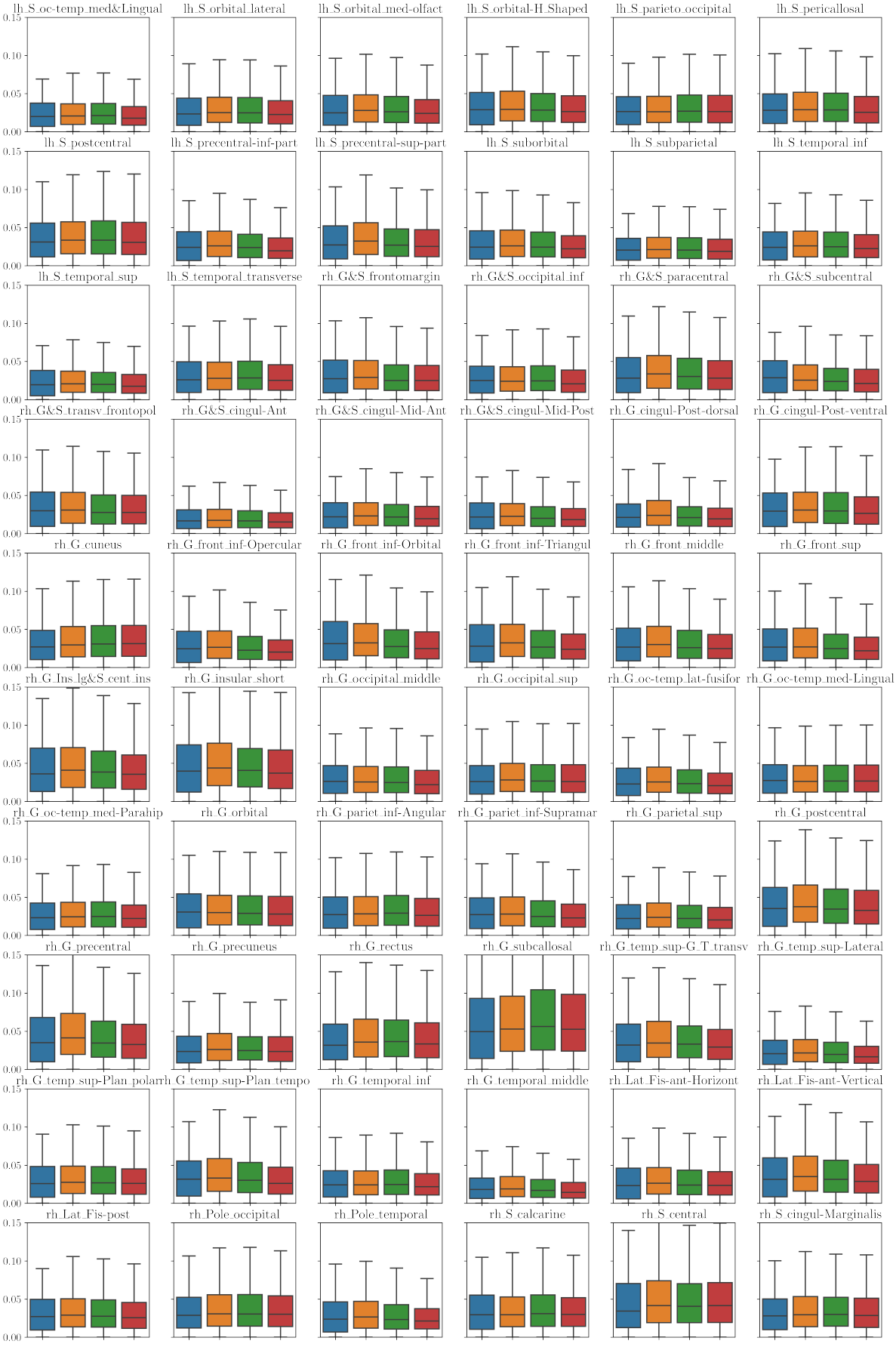


Figure 7.8: Comparison of test/re-test measurement noise (blue) and fit errors (AD: orange, MCI: green, CN: red) for cortical thicknesses. Part 2/3.

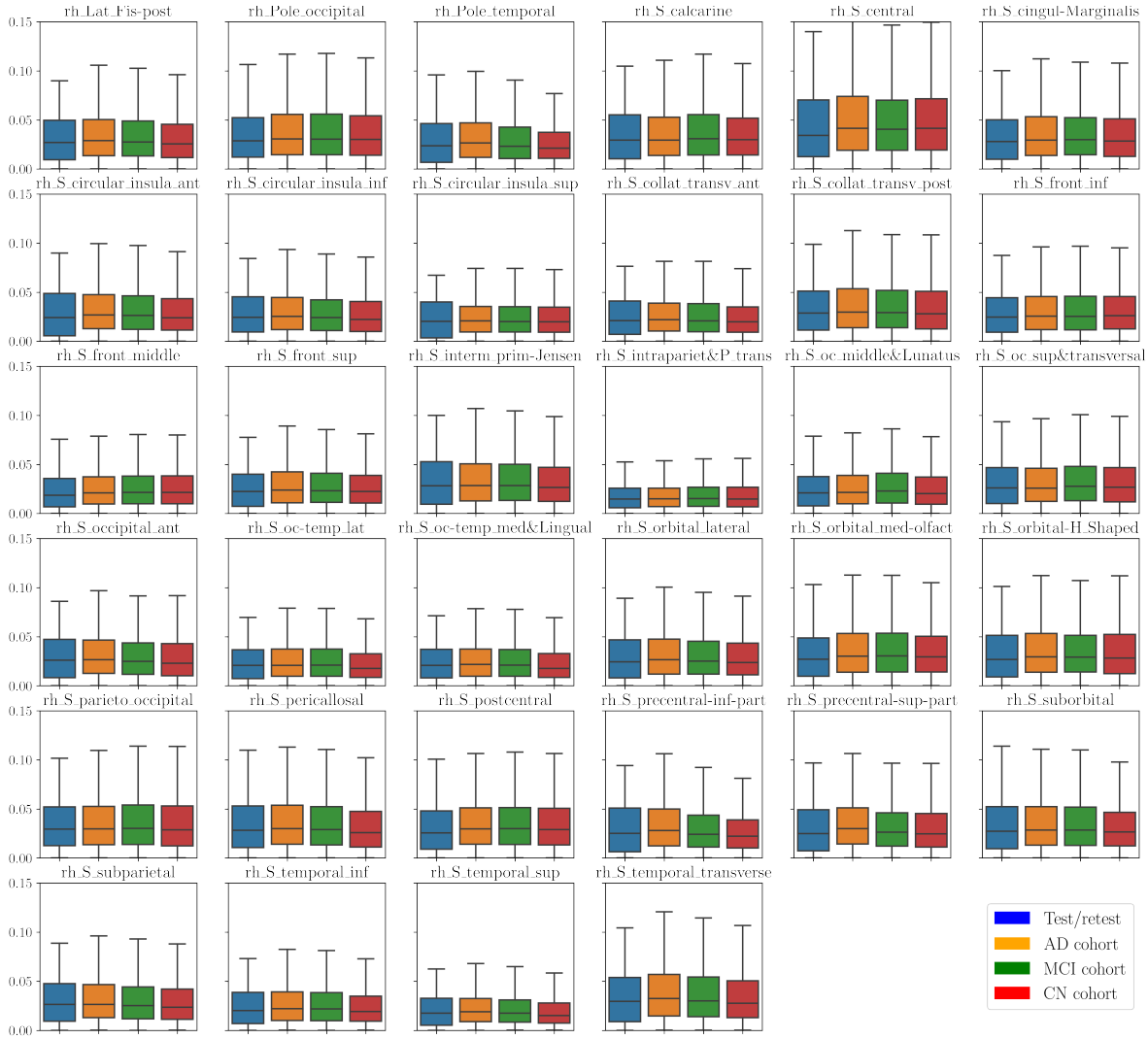


Figure 7.9: Comparison of test/re-test measurement noise (blue) and fit errors (AD: orange, MCI: green, CN: red) for cortical thicknesses. Part 3/3.

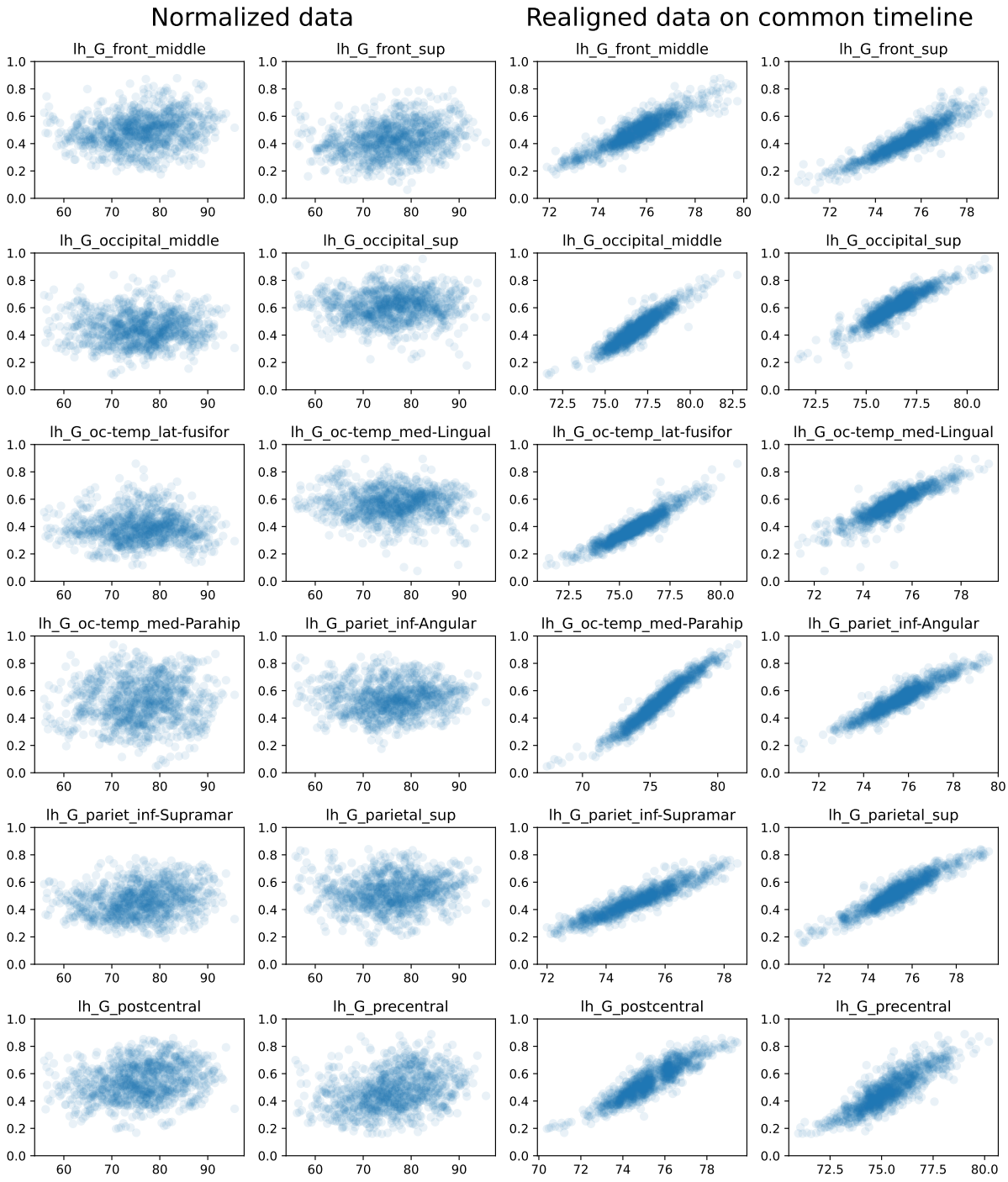


Figure 7.10: Scatterplots of cortical thickness measurements for 12 randomly selected regions of the Destrieux atlas for the AD cohort. The first two columns represent real measurements over time, while the last two columns represent the same measurements over the learned pathological timeline, using the onset age and pace of decline for the affine time reparametrization. It should be noted that, although the raw data share the same time of visits (between 60 and 90), the learned pathological progression is different for each feature.

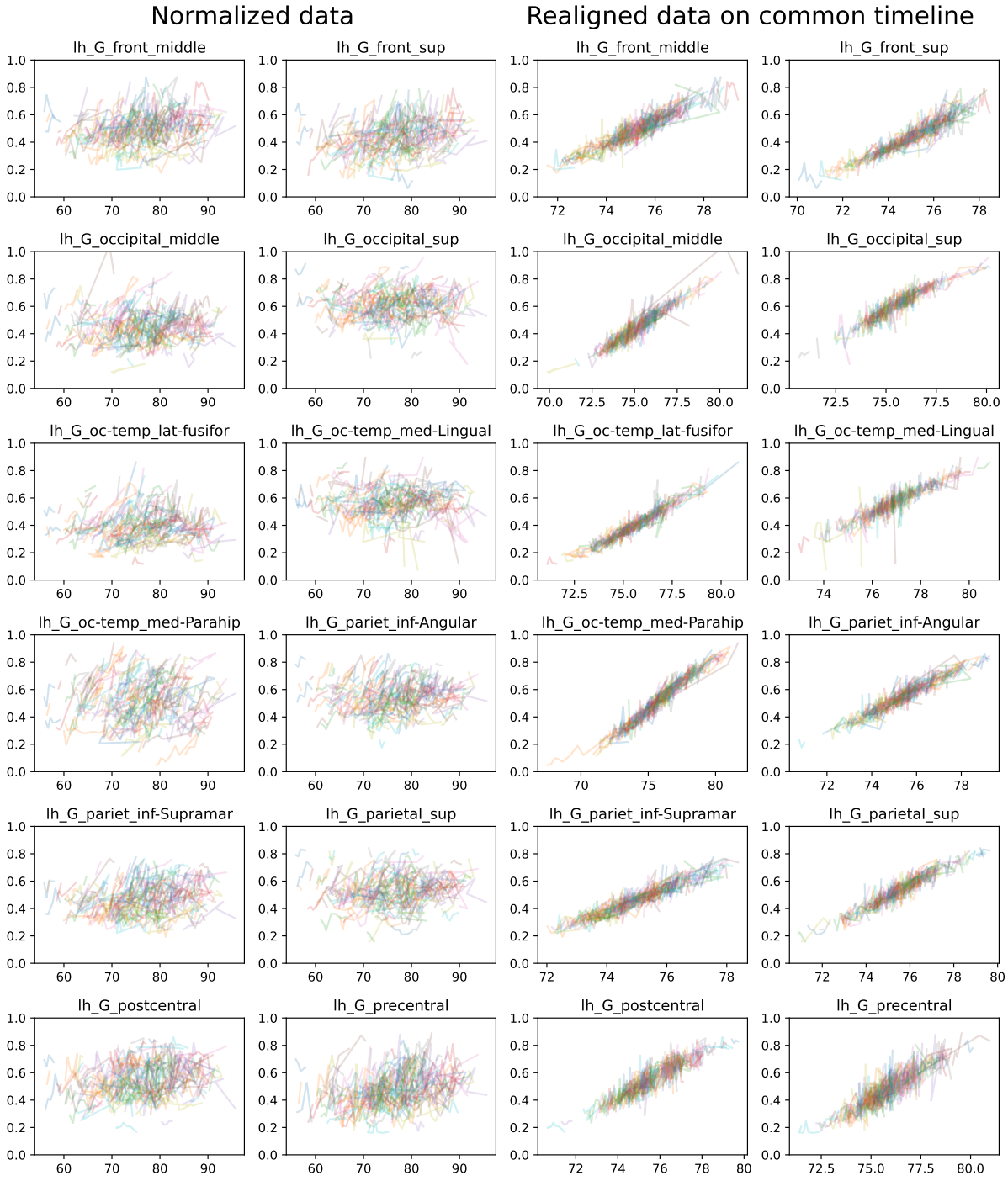


Figure 7.11: Spaghetti plots of longitudinal cortical thickness measurements for the same regions of the Destrieux atlas for the AD cohort. The first two columns represent real measurements over time, while the last two columns represent the same measurements over the learned pathological timeline, using the onset age and pace of decline for the affine time reparametrization.

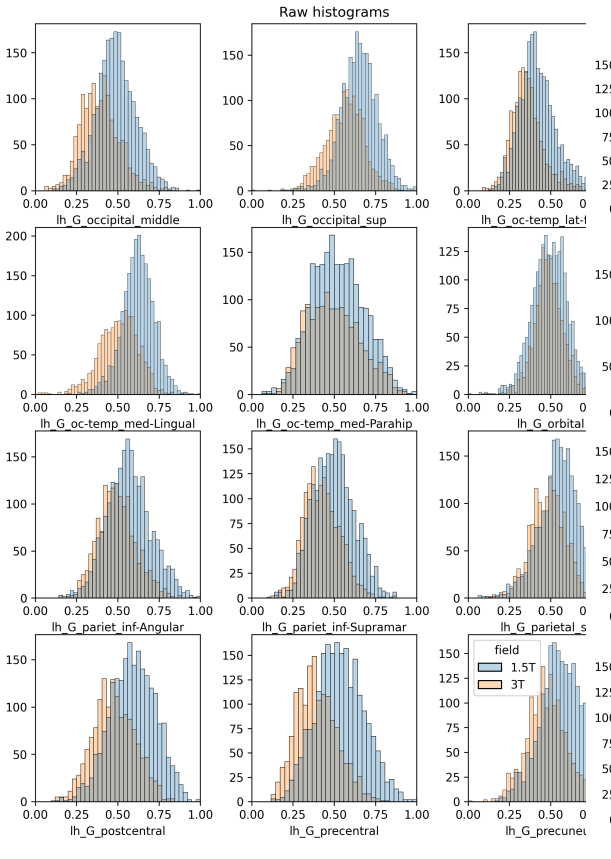


Figure 7.12: Histogram of cortical thickness measurements for 12 randomly selected regions of the Destrieux atlas for the AD cohort. Distributions are stratified by the field strength of the MRI.

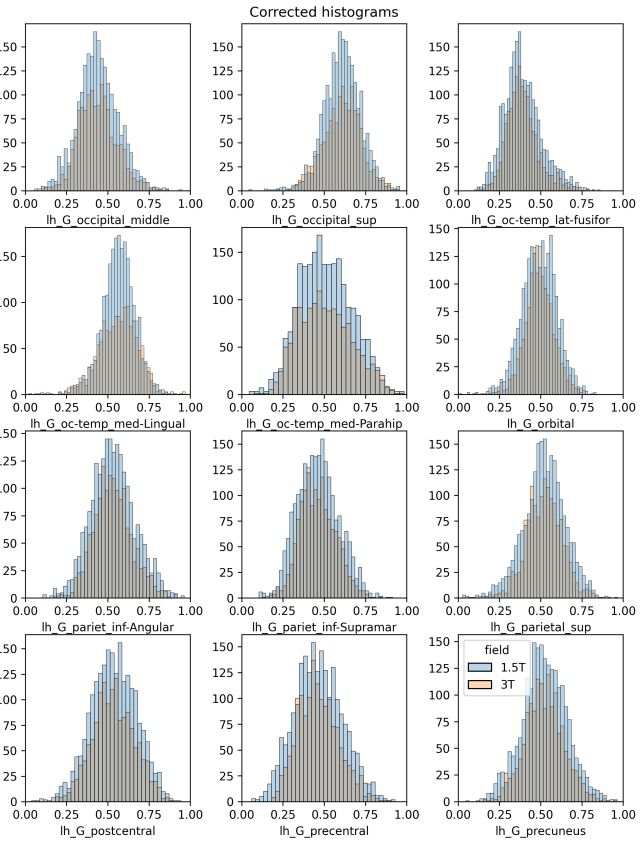


Figure 7.13: Histogram of *corrected* cortical thickness measurements for the same regions. Additive bias is removed and histograms align across field strengths.

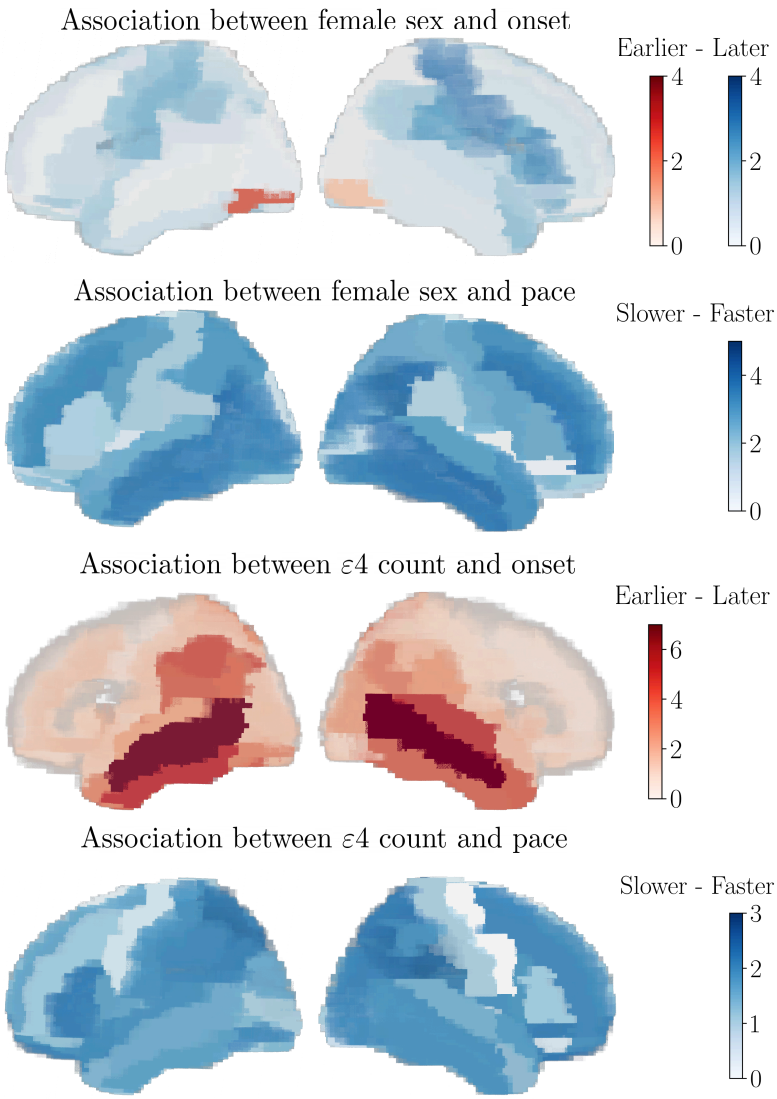


Figure 7.14: Cortical thinning over the course of AD progression. Columns display left lateral and right lateral view. Legend bars show negative log p-values. It should be noted that blue values indicate that the considered stratifying factor is a protective factor for onset age but a risk factor for accelerations regarding disease severity, since one wants the highest possible onset and lowest possible acceleration factor.

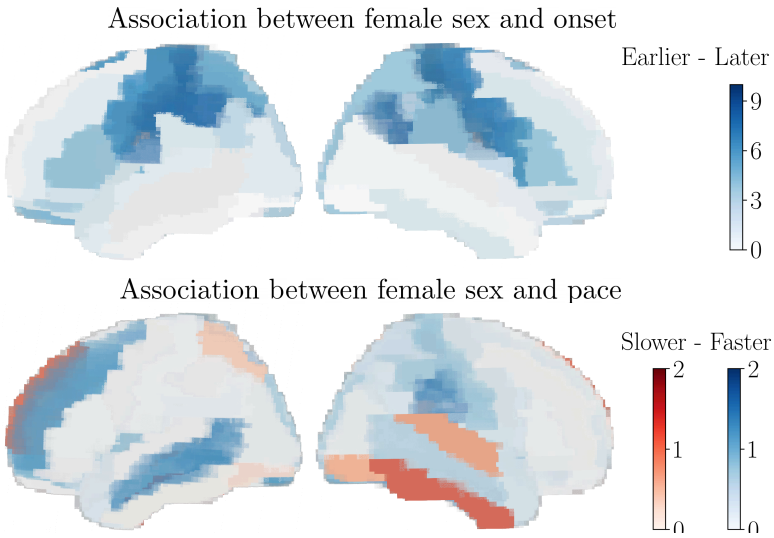


Figure 7.15: Cortical thinning over the course of healthy ageing.

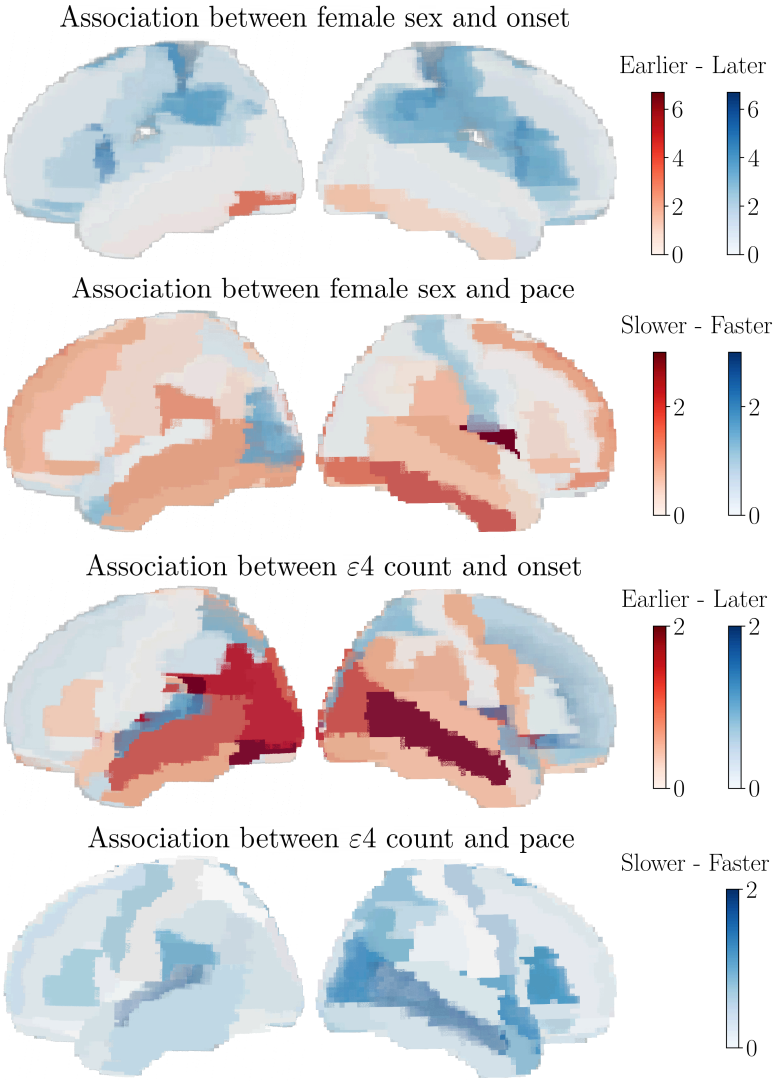


Figure 7.16: Cortical thinning for patients with MCI.

Bibliography

[Abi Nader et al., 2021] Abi Nader, C., Ayache, N., Frisoni, G., Robert, P., and Lorenzi, M. (2021). Simulating the outcome of amyloid treatments in Alzheimer's disease from imaging and clinical data. *Brain communications*, 3(2):fcab091.

[Abi Nader et al., 2020] Abi Nader, C., Ayache, N., Robert, P., Lorenzi, M., Initiative, A. D. N., et al. (2020). Monotonic gaussian process for spatio-temporal disease progression modeling in brain imaging data. *Neuroimage*, 205:116266.

[Abushakra et al., 2020] Abushakra, S., Porsteinsson, A. P., Sabbagh, M., Bracoud, L., Schaerer, J., Power, A., Hey, J. A., Scott, D., Suhy, J., Tolar, M., et al. (2020). Apoe $\varepsilon 4/\varepsilon 4$ homozygotes with early alzheimer's disease show accelerated hippocampal atrophy and cortical thinning that correlates with cognitive decline. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 6(1):e12117.

[Agarwal et al., 2021] Agarwal, D., Marques, G., de la Torre-Díez, I., Franco Martin, M. A., García Zapiraín, B., and Martín Rodríguez, F. (2021). Transfer learning for alzheimer's disease through neuroimaging biomarkers: a systematic review. *Sensors*, 21(21):7259.

[Allassonnière et al., 2010] Allassonnière, S., Kuhn, E., and Trouvé, A. (2010). Construction of bayesian deformable models via a stochastic approximation algorithm: a convergence study. *Bernoulli*, 16(3):641–678.

[Allassonnière et al., 2010] Allassonnière, S., Kuhn, E., and Trouvé, A. (2010). Construction of Bayesian deformable models via a stochastic approximation algorithm: A convergence study. *Bernoulli*, 16(3):641 – 678.

[Altmann et al., 2014] Altmann, A., Tian, L., Henderson, V. W., Greicius, M. D., and Investigators, A. D. N. I. (2014). Sex modifies the apoe-related risk of developing alzheimer disease. *Annals of neurology*, 75(4):563–573.

[Alzheimer's Association et al., 2016] Alzheimer's Association et al. (2016). 2016 alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 12(4):459–509.

[Alzheimer's Association et al., 2021] Alzheimer's Association et al. (2021). 2021 alzheimer's disease facts and figures. *Alzheimer's & Dementia*.

[Ansart et al., 2021] Ansart, M., Epelbaum, S., Bassignana, G., Bône, A., Bottani, S., Cattai, T., Couronné, R., Faouzi, J., Koval, I., Louis, M., et al. (2021). Predicting the progression of mild cognitive impairment using machine learning: A systematic, quantitative and critical review. *Medical Image Analysis*, 67:101848.

[Arbabshirani et al., 2017] Arbabshirani, M. R., Plis, S., Sui, J., and Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage*, 145:137–165.

[Ardalan and Subbian, 2022] Ardalan, Z. and Subbian, V. (2022). Transfer learning approaches for neuroimaging analysis: A scoping review. *Frontiers in Artificial Intelligence*, 5:15.

[Armstrong et al., 2019] Armstrong, N. M., An, Y., Beason-Held, L., Doshi, J., Erus, G., Ferrucci, L., Davatzikos, C., and Resnick, S. M. (2019). Sex differences in brain aging and predictors of neurodegeneration in cognitively healthy older adults. *Neurobiology of aging*, 81:146–156.

[Arvanitidis et al., 2017] Arvanitidis, G., Hansen, L. K., and Hauberg, S. (2017). Latent space oddity: on the curvature of deep generative models. *arXiv preprint arXiv:1710.11379*.

[Ashman et al., 2020] Ashman, M., So, J., Tebbutt, W., Fortuin, V., Pearce, M., and Turner, R. E. (2020). Sparse gaussian process variational autoencoders. *arXiv preprint arXiv:2010.10177*.

[Ballard et al., 2019] Ballard, C., Atri, A., Boneva, N., Cummings, J. L., Frölich, L., Molinuevo, J. L., Tariot, P. N., and Raket, L. L. (2019). Enrichment factors for clinical trials in mild-to-moderate alzheimer's disease. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 5:164–174.

[Banerjee et al., 2016] Banerjee, M., Chakraborty, R., Ofori, E., Okun, M. S., Viallancourt, D. E., and Vemuri, B. C. (2016). A nonlinear regression technique for manifold valued data with applications to medical image analysis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4424–4432.

[Bauer et al., 2011] Bauer, S., May, C., Dionysiou, D., Stamatakos, G., Buchler, P., and Reyes, M. (2011). Multiscale modeling for image analysis of brain tumor studies. *IEEE transactions on biomedical engineering*, 59(1):25–29.

[Beam et al., 2018] Beam, C. R., Kaneshiro, C., Jang, J. Y., Reynolds, C. A., Pedersen, N. L., and Gatz, M. (2018). Differences between women and men in incidence rates of dementia and alzheimer's disease. *Journal of Alzheimer's Disease*, 64(4):1077–1083.

[Benjamini and Yekutieli, 2005] Benjamini, Y. and Yekutieli, D. (2005). False discovery rate–adjusted multiple confidence intervals for selected parameters. *Journal of the American Statistical Association*, 100(469):71–81.

[Bernal-Rusiel et al., 2013a] Bernal-Rusiel, J. L., Greve, D. N., Reuter, M., Fischl, B., Sabuncu, M. R., Initiative, A. D. N., et al. (2013a). Statistical analysis of longitudinal neuroimage data with linear mixed effects models. *Neuroimage*, 66:249–260.

[Bernal-Rusiel et al., 2013b] Bernal-Rusiel, J. L., Reuter, M., Greve, D. N., Fischl, B., Sabuncu, M. R., Initiative, A. D. N., et al. (2013b). Spatiotemporal linear mixed effects modeling for the mass-univariate analysis of longitudinal neuroimage data. *Neuroimage*, 81:358–370.

[Bilgel et al., 2014] Bilgel, M., An, Y., Lang, A., Prince, J., Ferrucci, L., Jedynak, B., and Resnick, S. M. (2014). Trajectories of alzheimer disease-related cognitive measures in a longitudinal sample. *Alzheimer's & Dementia*, 10(6):735–742.

[Bône et al., 2018] Bône, A., Colliot, O., and Durrleman, S. (2018). Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9271–9280.

[Bône et al., 2017] Bône, A., Louis, M., Routier, A., Samper, J., Bacci, M., Charlier, B., Colliot, O., Durrleman, S., Initiative, A. D. N., et al. (2017). Prediction of the progression of subcortical brain structures in alzheimer's disease from baseline. In *Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics*, pages 101–113. Springer.

[Boyle et al., 2006] Boyle, P., Wilson, R., Aggarwal, N., Tang, Y., and Bennett, D. (2006). Mild cognitive impairment: risk of alzheimer disease and rate of cognitive decline. *Neurology*.

[Brouwer et al., 2022] Brouwer, R. M., Klein, M., Grasby, K. L., Schnack, H. G., Jahanshad, N., Teeuw, J., Thomopoulos, S. I., Sprooten, E., Franz, C. E., Gogtay, N., et al. (2022). Genetic variants associated with longitudinal changes in brain structure across the lifespan. *Nature neuroscience*, 25(4):421–432.

[Buckley et al., 2018] Buckley, R. F., Mormino, E. C., Amariglio, R. E., Properzi, M. J., Rabin, J. S., Lim, Y. Y., Papp, K. V., Jacobs, H. I., Burnham, S., Hanseeuw, B. J., et al. (2018). Sex, amyloid, and apoe $\varepsilon 4$ and risk of cognitive decline in preclinical alzheimer's disease: Findings from three well-characterized cohorts. *Alzheimer's & Dementia*, 14(9):1193–1203.

[Budd Haeberlein et al., 2022] Budd Haeberlein, S., Aisen, P., Barkhof, F., Chalkias, S., Chen, T., Cohen, S., Dent, G., Hansson, O., Harrison, K., von Hehn, C., et al. (2022). Two randomized phase 3 studies of aducanumab in early alzheimer's disease. *The Journal of Prevention of Alzheimer's Disease*, 9(2):197–210.

[Casale et al., 2018] Casale, F. P., Dalca, A. V., Saglietti, L., Listgarten, J., and Fusi, N. (2018). Gaussian process prior variational autoencoders. *arXiv preprint arXiv:1810.11738*.

[Chadebec et al., 2020] Chadebec, C., Mantoux, C., and Allassonnière, S. (2020). Geometry-aware hamiltonian variational auto-encoder. *arXiv preprint arXiv:2010.11518*.

[Chen et al., 2018] Chen, N., Klushyn, A., Kurle, R., Jiang, X., Bayer, J., and Smagt, P. (2018). Metrics for deep generative models. In *International Conference on Artificial Intelligence and Statistics*, pages 1540–1550. PMLR.

[Chen et al., 2007] Chen, X., Sachdev, P. S., Wen, W., and Anstey, K. J. (2007). Sex differences in regional gray matter in healthy individuals aged 44–48 years: a voxel-based morphometric study. *Neuroimage*, 36(3):691–699.

[Cieri et al., 2022] Cieri, F., Zhuang, X., Cordes, D., Kaplan, N., Cummings, J., and Caldwell, J. (2022). Relationship of sex differences in cortical thickness and memory among cognitively healthy subjects and individuals with mild cognitive impairment and alzheimer disease. *Alzheimer's research & therapy*, 14(1):1–12.

[Cnaan et al., 1997] Cnaan, A., Laird, N., and Slasor, P. (1997). Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in medicine*, 16(20):2349–2380.

[Coffey et al., 1998] Coffey, C. E., Lucke, J. F., Saxton, J. A., Ratcliff, G., Unitas, L. J., Billig, B., and Bryan, R. N. (1998). Sex differences in brain aging: a quantitative magnetic resonance imaging study. *Archives of neurology*, 55(2):169–179.

[Couronné, 2021] Couronné, R. (2021). *Progression models for Parkinson's Disease*. PhD thesis, Sorbonne Université.

[Couronné et al., 2019] Couronné, R., Louis, M., and Durrleman, S. (2019). Longitudinal autoencoder for multi-modal disease progression modelling. preprint.

[Couronné et al., 2021] Couronné, R., Vernhet, P., and Durrleman, S. (2021). Longitudinal self-supervision to disentangle inter-patient variability from disease progression. In *Int. Conf. on MICCAI*, pages 231–241. Springer.

[Couronné and Vernhet, 2021] Couronné, R. and Vernhet, P. (2021). Starmen longitudinal.

[Cowell et al., 1994] Cowell, P. E., Turetsky, B. I., Gur, R. C., Grossman, R. I., Shtasel, D., and Gur, R. (1994). Sex differences in aging of the human frontal and temporal lobes. *Journal of Neuroscience*, 14(8):4748–4755.

[Crivello et al., 2010] Crivello, F., Lemaître, H., Dufouil, C., Grassiot, B., Delcroix, N., Tzourio-Mazoyer, N., Tzourio, C., and Mazoyer, B. (2010). Effects of apoe-4 allele load and age on the rates of grey matter and hippocampal volumes loss in a longitudinal cohort of 1186 healthy elderly persons. *Neuroimage*, 53(3):1064–1069.

[Crivello et al., 2014] Crivello, F., Tzourio-Mazoyer, N., Tzourio, C., and Mazoyer, B. (2014). Longitudinal assessment of global and regional rate of grey matter atrophy in 1,172 healthy older adults: modulation by sex and age. *PloS one*, 9(12):e114478.

[Cui et al., 2019] Cui, R., Liu, M., et al. (2019). Rnn-based longitudinal analysis for diagnosis of alzheimer's disease. *Computerized Medical Imaging and Graphics*, 73:1–10.

[Cummings et al., 2022] Cummings, J., Lee, G., Nahed, P., Kambar, M. E. Z. N., Zhong, K., Fonseca, J., and Taghva, K. (2022). Alzheimer's disease drug development pipeline: 2022. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 8(1):e12295.

[Członkowska et al., 2006] Członkowska, A., Ciesielska, A., Gromadzka, G., and Kurkowska-Jastrzębska, I. (2006). Gender differences in neurological disease. *Endocrine*, 29(2):243–256.

[Delyon et al., 1999] Delyon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a stochastic approximation version of the em algorithm. *Annals of statistics*, pages 94–128.

[Derflinger et al., 2011] Derflinger, S., Sorg, C., Gaser, C., Myers, N., Arsic, M., Kurz, A., Zimmer, C., Wohlschläger, A., and Mühlau, M. (2011). Grey-matter atrophy in alzheimer's disease is asymmetric but not lateralized. *Journal of Alzheimer's Disease*, 25(2):347–357.

[Donohue et al., 2014] Donohue, M., Jacqmin-Gadda, H., Le Goff, M., Thomas, R., Raman, R., Gamst, A., Beckett, L., Jack Jr, C., Weiner, M., Dartigues, J., et al. (2014). Estimating long-term multivariate progression from short-term data. *Alzheimer's & Dementia*, 10:S400–S410.

[Falahati et al., 2014] Falahati, F., Westman, E., and Simmons, A. (2014). Multivariate data analysis and machine learning in alzheimer's disease with a focus on structural magnetic resonance imaging. *Journal of Alzheimer's disease*, 41(3):685–708.

[Falorsi et al., 2018] Falorsi, L., De Haan, P., Davidson, T. R., De Cao, N., Weiler, M., Forré, P., and Cohen, T. S. (2018). Explorations in homeomorphic variational auto-encoding. *arXiv preprint arXiv:1807.04689*.

[Farrer et al., 1997] Farrer, L. A., Cupples, L. A., Haines, J. L., Hyman, B., Kukull, W. A., Mayeux, R., Myers, R. H., Pericak-Vance, M. A., Risch, N., and Van Duijn, C. M. (1997). Effects of age, sex, and ethnicity on the association between apolipoprotein e genotype and alzheimer disease: a meta-analysis. *Jama*, 278(16):1349–1356.

[Ferretti et al., 2018] Ferretti, M. T., Iulita, M. F., Cavedo, E., Chiesa, P. A., Schumacher Dimech, A., Santuccione Chadha, A., Baracchi, F., Girouard, H., Misoch, S., Giacobini, E., et al. (2018). Sex differences in alzheimer disease—the gateway to precision medicine. *Nature Reviews Neurology*, 14(8):457–469.

[Fonteijn et al., 2012] Fonteijn, H. M., Modat, M., Clarkson, M. J., Barnes, J., Lehmann, M., Hobbs, N. Z., Scahill, R. I., Tabrizi, S. J., Ourselin, S., Fox, N. C., et al. (2012). An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. *NeuroImage*, 60(3):1880–1889.

[Fortuin et al., 2020] Fortuin, V., Baranchuk, D., Rätsch, G., and Mandt, S. (2020). Gp-vae: Deep probabilistic time series imputation. In *International conference on artificial intelligence and statistics*, pages 1651–1661. PMLR.

[Garbarino et al., 2021] Garbarino, S., Lorenzi, M., Initiative, A. D. N., et al. (2021). Investigating hypotheses of neurodegeneration by learning dynamical systems of protein propagation in the brain. *Neuroimage*, 235:117980.

[Garibotto et al., 2008] Garibotto, V., Borroni, B., Kalbe, E., Herholz, K., Salmon, E., Holtoff, V., Sorbi, S., Cappa, S., Padovani, A., Fazio, F., et al. (2008). Education and occupation as proxies for reserve in amci converters and ad: Fdg-pet evidence. *Neurology*.

[Gaugler et al., 2013] Gaugler, J. E., Ascher-Svanum, H., Roth, D. L., Fafowora, T., Siderowf, A., and Beach, T. G. (2013). Characteristics of patients misdiagnosed with alzheimer's disease and their medication use: an analysis of the nacc-uds database. *BMC geriatrics*, 13(1):1–10.

[Gauthier et al., 2006] Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R. C., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D., Chertkow, H., et al. (2006). Mild cognitive impairment. *The lancet*, 367(9518):1262–1270.

[Genazzani et al., 2007] Genazzani, A. R., Pluchino, N., Luisi, S., and Luisi, M. (2007). Estrogen, cognition and female ageing. *Human reproduction update*, 13(2):175–187.

[Gorbach et al., 2020] Gorbach, T., Pudas, S., Bartrés-Faz, D., Brandmaier, A. M., Düzel, S., Henson, R. N., Idland, A.-V., Lindenberger, U., Macià Bros, D., Mowinckel, A. M., et al. (2020). Longitudinal association between hippocampus atrophy and episodic-memory decline in non-demented apoe $\varepsilon 4$ carriers. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 12(1):e12110.

[Gruffaz et al., 2021] Gruffaz, S., Poulet, P.-E., Maheux, E., Jedynak, B., and Durrleman, S. (2021). Learning riemannian metric for disease progression modeling. *Advances in Neural Information Processing Systems*, 34.

[Gurvich et al., 2018] Gurvich, C., Hoy, K., Thomas, N., and Kulkarni, J. (2018). Sex differences and the influence of sex hormones on cognition through adulthood and the aging process. *Brain sciences*, 8(9):163.

[Gutiérrez-Galve et al., 2009] Gutiérrez-Galve, L., Lehmann, M., Hobbs, N. Z., Clarkson, M. J., Ridgway, G. R., Crutch, S., Ourselin, S., Schott, J. M., Fox, N. C., and Barnes, J. (2009). Patterns of cortical thickness according to apoe genotype in alzheimer's disease. *Dementia and geriatric cognitive disorders*, 28(5):461–470.

[Han et al., 2006] Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., Busa, E., Pacheco, J., Albert, M., Killiany, R., et al. (2006). Reliability of mri-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *Neuroimage*, 32(1):180–194.

[Hansson et al., 2018] Hansson, O., Seibyl, J., Stomrud, E., Zetterberg, H., Trojanowski, J. Q., Bittner, T., Lifke, V., Corradini, V., Eichenlaub, U., Batrla, R., et al. (2018). Csf biomarkers of alzheimer's disease concord with amyloid-β pet and predict clinical

progression: a study of fully automated immunoassays in biofinder and adni cohorts. *Alzheimer's & dementia*, 14(11):1470–1481.

[Hassanaly et al., 2023] Hassanaly, R., Bottani, S., Sauty, B., Colliot, O., and Burgos, N. (2023). Simulation-based evaluation framework for deep learning unsupervised anomaly detection on brain fdg pet. In *SPIE MI*.

[Hebert et al., 2010] Hebert, L., Bienias, J., Aggarwal, N., Wilson, R., Bennett, D., Shah, R., and Evans, D. (2010). Change in risk of alzheimer disease over time. *Neurology*, 75(9):786–791.

[Hebert et al., 1995] Hebert, L. E., Scherr, P. A., Beckett, L. A., Albert, M. S., Pilgrim, D. M., Chown, M. J., Funkenstein, H. H., and Evans, D. A. (1995). Age-specific incidence of alzheimer's disease in a community population. *Jama*, 273(17):1354–1359.

[Herholz, 2010] Herholz, K. (2010). Cerebral glucose metabolism in preclinical and prodromal alzheimer's disease. *Expert review of neurotherapeutics*.

[Higgins et al., 2016] Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., and Lerchner, A. (2016). beta-vae: Learning basic visual concepts with a constrained variational framework.

[Hua et al., 2010] Hua, X., Hibar, D. P., Lee, S., Toga, A. W., Jack Jr, C. R., Weiner, M. W., Thompson, P. M., Initiative, A. D. N., et al. (2010). Sex and age differences in atrophic rates: an adni study with n= 1368 mri scans. *Neurobiology of aging*, 31(8):1463–1480.

[Ibáñez et al., 1998] Ibáñez, V., Pietrini, P., Alexander, G., Furey, M., Teichberg, D., Rajapakse, J., Rapoport, S., Schapiro, M., and Horwitz, B. (1998). Regional glucose metabolic abnormalities are not the result of atrophy in alzheimer's disease. *Neurology*.

[Ito et al., 2011] Ito, K., Corrigan, B., Zhao, Q., French, J., Miller, R., Soares, H., Katz, E., Nicholas, T., Billing, B., Anziano, R., et al. (2011). Disease progression model for cognitive deterioration from alzheimer's disease neuroimaging initiative database. *Alzheimer's & Dementia*, 7(2):151–160.

[Jack et al., 2010] Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., Petersen, R. C., and Trojanowski, J. Q. (2010). Hypothetical model of dynamic biomarkers of the alzheimer's pathological cascade. *The Lancet Neurology*, 9(1):119–128.

[Jack et al., 2015] Jack, C. R., Wiste, H. J., Weigand, S. D., Knopman, D. S., Vemuri, P., Mielke, M. M., Lowe, V., Senjem, M. L., Gunter, J. L., Machulda, M. M., et al. (2015). Age, sex, and apoe $\varepsilon 4$ effects on memory, brain structure, and β-amyloid across the adult life span. *JAMA neurology*, 72(5):511–519.

[Jack Jr et al., 2013] Jack Jr, C. R., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., Shaw, L. M., Vemuri, P., Wiste, H. J., Weigand, S. D., et al. (2013). Update on hypothetical model of alzheimer's disease biomarkers. *Lancet neurology*, 12(2):207.

[Jedynak et al., 2012] Jedynak, B. M., Lang, A., Liu, B., Katz, E., Zhang, Y., Wyman, B. T., Raunig, D., Jedynak, C. P., Caffo, B., Prince, J. L., et al. (2012). A computational neurodegenerative disease progression score: method and results with the alzheimer's disease neuroimaging initiative cohort. *Neuroimage*, 63(3):1478–1486.

[Kim et al., 2018] Kim, J., Park, S., Yoo, H., Jang, H., Kim, Y., Kim, K. W., Jang, Y. K., Lee, J. S., Kim, S. T., Kim, S., et al. (2018). The impact of apoe 4 in alzheimer's disease differs according to age. *Journal of Alzheimer's Disease*, 61(4):1377–1385.

[Kingma and Welling, 2013] Kingma, D. P. and Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

[Koran et al., 2017] Koran, M. E. I., Wagener, M., Hohman, T. J., and Initiative, A. N. (2017). Sex differences in the association between ad biomarkers and cognitive decline. *Brain imaging and behavior*, 11:205–213.

[Koval, 2020] Koval, I. (2020). *Learning multimodal digital models of disease progression from longitudinal data: methods & algorithms for the description, prediction and simulation of Alzheimer's disease progression.* PhD thesis, Institut polytechnique de Paris.

[Koval et al., 2021a] Koval, I., Bône, A., Louis, M., Lartigue, T., Bottani, S., Marcoux, A., Samper-Gonzalez, J., Burgos, N., Charlier, B., Bertrand, A., Epelbaum, S., Colliot, O., Allassonnière, S., and Durrleman, S. (2021a). AD Course Map charts Alzheimer's disease progression. *Scientific Reports*, 11(1).

[Koval et al., 2021b] Koval, I., Bône, A., Louis, M., Lartigue, T., Bottani, S., Marcoux, A., Samper-Gonzalez, J., Burgos, N., Charlier, B., Bertrand, A., et al. (2021b). Ad course map charts alzheimer's disease progression. *Scientific Reports*, 11(1):1–16.

[Koval et al., 2020] Koval, I., Dighiero, T., Scahill, R., Durr, A., and Durrleman, S. (2020). Prediction of biomarkers' trajectory in huntington's disease: application to precise clinical trial design. In *CompAge 2020-Computational approaches for ageing and age-related diseases*.

[Koval et al., 2022] Koval, I., Dighiero-Brecht, T., Tobin, A. J., Tabrizi, S. J., Scahill, R. I., Tezenas du Montcel, S., Durrleman, S., and Durr, A. (2022). Forecasting individual progression trajectories in huntington disease enables more powered clinical trials. *Scientific Reports*, 12(1):18928.

[Koval et al., 2017a] Koval, I., Schiratti, J.-B., Routier, A., Bacci, M., Colliot, O., Allassonnière, S., and Durrleman, S. (2017a). Statistical learning of spatiotemporal patterns from longitudinal manifold-valued networks. In Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D. L., and Duchesne, S., editors, *Medical Image Computing and Computer Assisted Intervention MICCAI 2017*, pages 451–459, Cham. Springer International Publishing.

[Koval et al., 2018] Koval, I., Schiratti, J.-B., Routier, A., Bacci, M., Colliot, O., Allassonniere, S., and Durrleman, S. (2018). Spatiotemporal propagation of the cortical atrophy: Population and individual patterns. *Frontiers in neurology*, 9:235.

[Koval et al., 2017b] Koval, I., Schiratti, J.-B., Routier, A., Bacci, M., Colliot, O., Allassonnière, S., Durrleman, S., Initiative, A. D. N., et al. (2017b). Statistical learning of spatiotemporal patterns from longitudinal manifold-valued networks. In *International conference on medical image computing and computer-assisted intervention*, pages 451–459. Springer.

[Kuhn and Lavielle, 2004] Kuhn, E. and Lavielle, M. (2004). Coupling a stochastic approximation version of em with an mcmc procedure. *ESAIM: Probability and Statistics*, 8:115–131.

[Kuhn and Lavielle, 2005] Kuhn, E. and Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Computational statistics & data analysis*, 49(4):1020–1038.

[Laird and Ware, 1982] Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*.

[Laws et al., 2016] Laws, K. R., Irvine, K., and Gale, T. M. (2016). Sex differences in cognitive impairment in alzheimer's disease. *World journal of psychiatry*, 6(1):54.

[Laws et al., 2018] Laws, K. R., Irvine, K., and Gale, T. M. (2018). Sex differences in alzheimer's disease. *Current opinion in psychiatry*, 31(2):133–139.

[Leandrou et al., 2018] Leandrou, S., Petroudi, S., Kyriacou, P. A., Reyes-Aldasoro, C. C., and Pattichis, C. S. (2018). Quantitative mri brain studies in mild cognitive impairment and alzheimer's disease: a methodological review. *IEEE reviews in biomedical engineering*, 11:97–111.

[Li et al., 2016] Li, B., Shi, J., Gutman, B. A., Baxter, L. C., Thompson, P. M., Caselli, R. J., Wang, Y., and Initiative, A. D. N. (2016). Influence of apoe genotype on hippocampal atrophy over time-an n= 1925 surface-based adni study. *PloS one*, 11(4):e0152901.

[Li et al., 2020] Li, Y., Zhang, L., Bozoki, A., Zhu, D. C., Choi, J., and Maiti, T. (2020). Early prediction of alzheimer's disease using longitudinal volumetric mri data from adni. *Health Services and Outcomes Research Methodology*, 20(1):13–39.

[Liu et al., 2018] Liu, M., Cheng, D., Yan, W., Initiative, A. D. N., et al. (2018). Classification of alzheimer's disease by combination of convolutional and recurrent neural networks using fdg-pet images. *Frontiers in neuroinformatics*, 12:35.

[Liu et al., 2015] Liu, Y., Yu, J.-T., Wang, H.-F., Han, P.-R., Tan, C.-C., Wang, C., Meng, X.-F., Risacher, S. L., Saykin, A. J., and Tan, L. (2015). Apoe genotype and neuroimaging markers of alzheimer's disease: systematic review and meta-analysis. *Journal of Neurology, Neurosurgery & Psychiatry*, 86(2):127–134.

[Livingston et al., 2020] Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., Brayne, C., Burns, A., Cohen-Mansfield, J., Cooper, C., et al. (2020). Dementia prevention, intervention, and care: 2020 report of the lancet commission. *The Lancet*, 396(10248):413–446.

[Lorenzi et al., 2019] Lorenzi, M., Filippone, M., Frisoni, G. B., Alexander, D. C., Ourselin, S., Initiative, A. D. N., et al. (2019). Probabilistic disease progression modeling to characterize diagnostic uncertainty: application to staging and prediction in alzheimer's disease. *NeuroImage*, 190:56–68.

[Lotze et al., 2019] Lotze, M., Domin, M., Gerlach, F. H., Gaser, C., Lueders, E., Schmidt, C. O., and Neumann, N. (2019). Novel findings from 2,838 adult brains on sex differences in gray matter brain volume. *Scientific reports*, 9(1):1–7.

[Louis, 2019] Louis, M. (2019). *Computational and statistical methods for trajectory analysis in a Riemannian geometry setting*. Theses, Sorbonnes universités.

[Louis et al., 2018] Louis, M., Charlier, B., Jusselin, P., Pal, S., and Durrleman, S. (2018). A fanning scheme for the parallel transport along geodesics on Riemannian manifolds. *SIAM Journal on Numerical Analysis*, 56(4):2563–2584.

[Louis et al., 2019] Louis, M., Couronné, R., Koval, I., Charlier, B., and Durrleman, S. (2019). Riemannian geometry learning for disease progression modelling. In *Int. Conf. on IPMI*, pages 542–553. Springer.

[Luders et al., 2006] Luders, E., Narr, K. L., Thompson, P. M., Rex, D. E., Woods, R. P., DeLuca, H., Jancke, L., and Toga, A. W. (2006). Gender effects on cortical thickness and the influence of scaling. *Human brain mapping*, 27(4):314–324.

[Maheux et al., 2023] Maheux, E., Koval, I., Ortholand, J., Birkenbihl, C., Archetti, D., Bouteloup, V., Epelbaum, S., Dufouil, C., Hofmann-Apitius, M., and Durrleman, S. (2023). Forecasting individual progression trajectories in alzheimer's disease. *Nature Communications*, 14(1):761.

[Mahley and Huang, 2012] Mahley, R. W. and Huang, Y. (2012). Apolipoprotein e sets the stage: response to injury triggers neuropathology. *Neuron*, 76(5):871–885.

[Malpetti et al., 2017] Malpetti, M., Ballarini, T., Presotto, L., Garibotto, V., Tettamanti, M., Perani, D., database;, A., and database, N.-D. (2017). Gender differences in healthy aging and alzheimer's dementia: A 18f-fdg-pet study of brain and cognitive reserve. *Human brain mapping*.

[Manning et al., 2014] Manning, E. N., Barnes, J., Cash, D. M., Bartlett, J. W., Leung, K. K., Ourselin, S., Fox, N. C., and Initiative, A. D. N. (2014). Apoe $\varepsilon 4$ is associated with disproportionate progressive hippocampal atrophy in ad. *PloS one*, 9(5):e97608.

[Marinescu et al., 2020] Marinescu, R. V., Oxtoby, N. P., Young, A. L., Bron, E. E., Toga, A. W., Weiner, M. W., Barkhof, F., Fox, N. C., Eshaghi, A., Toni, T., et al. (2020). The alzheimer's disease prediction of longitudinal evolution (tadpole) challenge: Results after 1 year follow-up. *arXiv preprint arXiv:2002.03419*.

[Marinescu et al., 2018] Marinescu, R. V., Oxtoby, N. P., Young, A. L., Bron, E. E., Toga, A. W., Weiner, M. W., Barkhof, F., Fox, N. C., Klein, S., Alexander, D. C., et al. (2018). Tadpole challenge: Prediction of longitudinal evolution in alzheimer's disease. *arXiv preprint arXiv:1805.03909*.

[Martinkova et al., 2021] Martinkova, J., Quevenco, F.-C., Karcher, H., Ferrari, A., Sandset, E. C., Szoeke, C., Hort, J., Schmidt, R., Chadha, A. S., and Ferretti, M. T. (2021). Proportion of women and reporting of outcomes by sex in clinical trials for alzheimer disease: a systematic review and meta-analysis. *JAMA Network Open*, 4(9):e2124124–e2124124.

[McCarrey et al., 2016] McCarrey, A. C., An, Y., Kitner-Triolo, M. H., Ferrucci, L., and Resnick, S. M. (2016). Sex differences in cognitive trajectories in clinically normal older adults. *Psychology and aging*, 31(2):166.

[Mielke, 2018] Mielke, M. M. (2018). Sex and gender differences in alzheimer's disease dementia. *The Psychiatric times*, 35(11):14.

[Milà-Alomà et al., 2022] Milà-Alomà, M., Ashton, N. J., Shekari, M., Salvadó, G., Ortiz-Romero, P., Montoliu-Gaya, L., Benedet, A. L., Karikari, T. K., Lantero-Rodriguez, J., Vanmechelen, E., et al. (2022). Plasma p-tau231 and p-tau217 as state markers of amyloid-β pathology in preclinical alzheimer's disease. *Nature Medicine*, 28(9):1797–1801.

[Mintun et al., 2021] Mintun, M. A., Lo, A. C., Duggan Evans, C., Wessels, A. M., Ardayfio, P. A., Andersen, S. W., Shcherbinin, S., Sparks, J., Sims, J. R., Brys, M., et al.

(2021). Donanemab in early alzheimer's disease. *New England Journal of Medicine*, 384(18):1691–1704.

[Mosconi, 2005] Mosconi, L. (2005). Brain glucose metabolism in the early and specific diagnosis of alzheimer's disease. *European journal of nuclear medicine and molecular imaging*.

[Mosconi et al., 2004] Mosconi, L., Sorbi, S., Nacmias, B., De Cristofaro, M., Fayyaz, M., Bracco, L., Herholz, K., and Pupi, A. (2004). Age and apoe genotype interaction in alzheimer's disease: an fdg-pet study. *Psychiatry Research: Neuroimaging*.

[Mungas et al., 2018] Mungas, D., Gavett, B., Fletcher, E., Farias, S. T., DeCarli, C., and Reed, B. (2018). Education amplifies brain atrophy effect on cognitive decline: implications for cognitive reserve. *Neurobiology of Aging*, 68:142–150.

[Neu et al., 2017] Neu, S. C., Pa, J., Kukull, W., Beekly, D., Kuzma, A., Gangadharan, P., Wang, L.-S., Romero, K., Arneric, S. P., Redolfi, A., et al. (2017). Apolipoprotein e genotype and sex risk factors for alzheimer disease: a meta-analysis. *JAMA neurology*, 74(10):1178–1189.

[Nguyen et al., 2020] Nguyen, M., He, T., An, L., Alexander, D., Feng, J., Yeo, B., et al. (2020). Predicting Alzheimer's disease progression using deep recurrent neural networks. *NeuroImage*, 222:117203.

[Niethammer et al., 2011] Niethammer, M., Huang, Y., and Vialard, F.-X. (2011). Geodesic regression for image time-series. In *International conference on medical image computing and computer-assisted intervention*, pages 655–662. Springer.

[Oxtoby et al., 2018] Oxtoby, N., Young, A., Cash, D., Benzinger, T., Fagan, A., Morris, J., Bateman, R., Fox, N., Schott, J., and Alexander, D. (2018). Data-driven models of dominantly-inherited Alzheimer's disease progression. *Brain*, 141(5):1529–1544.

[Petersen, 2016] Petersen, R. C. (2016). Mild cognitive impairment. *CONTINUUM: Lifelong Learning in Neurology*, 22(2 Dementia):404.

[Petersen et al., 2010] Petersen, R. C., Aisen, P., Beckett, L. A., Donohue, M., Gamst, A., Harvey, D. J., Jack, C., Jagust, W., Shaw, L., Toga, A., et al. (2010). Alzheimer's disease neuroimaging initiative (adni): clinical characterization. *Neurology*, 74(3):201–209.

[Poulet and Durrleman, 2021] Poulet, P.-E. and Durrleman, S. (2021). Mixture modeling for identifying subtypes in disease course mapping. In *Information Processing in Medical Imaging: 27th International Conference, IPMI 2021, Virtual Event, June 28–June 30, 2021, Proceedings*, pages 571–582. Springer.

[Raket, 2020] Raket, L. L. (2020). Statistical disease progression modeling in Alzheimer disease. *Frontiers in big Data*, 3.

[Ramchandran et al., 2021] Ramchandran, S., Tikhonov, G., Kujanpää, K., Koskinen, M., and Lähdesmäki, H. (2021). Longitudinal variational autoencoder. In *International Conference on Artificial Intelligence and Statistics*, pages 3898–3906. PMLR.

[Rathore et al., 2017] Rathore, S., Habes, M., Iftikhar, M. A., Shacklett, A., and Davatzikos, C. (2017). A review on neuroimaging-based classification studies and associated feature extraction methods for alzheimer's disease and its prodromal stages. *NeuroImage*, 155:530–548.

[Riedel et al., 2016] Riedel, B., Thompson, P., and Brinton, R. (2016). Age, apoe and sex: triad of risk of alzheimer's disease. *The Journal of steroid biochemistry and molecular biology*.

[Risacher et al., 2010] Risacher, S. L., Shen, L., West, J. D., Kim, S., McDonald, B. C., Beckett, L. A., Harvey, D. J., Jack Jr, C. R., Weiner, M. W., Saykin, A. J., et al. (2010). Longitudinal mri atrophy biomarkers: relationship to conversion in the adni cohort. *Neurobiology of aging*, 31(8):1401–1418.

[Roe et al., 2007] Roe, C. M., Xiong, C., Miller, J. P., and Morris, J. C. (2007). Education and alzheimer disease without dementia: support for the cognitive reserve hypothesis. *Neurology*, 68(3):223–228.

[Routier et al., 2021] Routier, A., Burgos, N., Díaz, M., Bacci, M., Bottani, S., El-Rifai, O., Fontanella, S., Gori, P., Guillon, J., Guyot, A., et al. (2021). Clinica: an open-source software platform for reproducible clinical neuroscience studies. *Frontiers in Neuroinformatics*, 15.

[Sabuncu et al., 2014] Sabuncu, M. R., Bernal-Rusiel, J. L., Reuter, M., Greve, D. N., Fischl, B., Initiative, A. D. N., et al. (2014). Event time analysis of longitudinal neuroimage data. *NeuroImage*, 97:9–18.

[Sampedro et al., 2015] Sampedro, F., Vilaplana, E., De Leon, M. J., Alcolea, D., Pegueroles, J., Montal, V., Carmona-Iragui, M., Sala, I., Sánchez-Saudinos, M.-B., Antón-Aguirre, S., et al. (2015). Apoe-by-sex interactions on brain structure and metabolism in healthy elderly controls. *Oncotarget*, 6(29):26663.

[Samtani et al., 2012] Samtani, M. N., Farnum, M., Lobanov, V., Yang, E., Raghavan, N., DiBernardo, A., Narayan, V., and Initiative, A. D. N. (2012). An improved model for disease progression in patients from the alzheimer's disease neuroimaging initiative. *The Journal of Clinical Pharmacology*, 52(5):629–644.

[Sangha et al., 2021] Sangha, O., Ma, D., Popuri, K., Stocks, J., Wang, L., Beg, M. F., Initiative, A. D. N., et al. (2021). Structural volume and cortical thickness differences between males and females in cognitively normal, cognitively impaired and alzheimer's dementia population. *Neurobiology of Aging*, 106:1–11.

[Saunders et al., 1993] Saunders, A. M., Strittmatter, W. J., Schmechel, D., George-Hyslop, P. S., Pericak-Vance, M. A., Joo, S., Rosi, B., Gusella, J., Crapper-MacLachlan, D., Alberts, M., et al. (1993). Association of apolipoprotein e allele 4 with late-onset familial and sporadic alzheimer's disease. *Neurology*, 43(8):1467–1467.

[Sauty and Durrleman, 2022a] Sauty, B. and Durrleman, S. (2022a). Impact of sex and apoe-ε4 genotype on patterns of regional brain atrophy in alzheimer's disease and healthy ageing.

[Sauty and Durrleman, 2022b] Sauty, B. and Durrleman, S. (2022b). Progression models for imaging data with longitudinal variational auto encoders. In *MICCAI 2022, International Conference on Medical Image Computing and Computer Assisted Intervention*.

[Sauty and Durrleman, 2022c] Sauty, B. and Durrleman, S. (2022c). Riemannian metric learning for progression modeling of longitudinal datasets. In *ISBI 2022-International Symposium on Biomedical Imaging*.

[Sauty and Durrleman, 2023] Sauty, B. and Durrleman, S. (2023). Impact of sex and apoe-ε4 genotype on regional brain metabolism in alzheimer's disease. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*.

[Scheyer et al., 2018] Scheyer, O., Rahman, A., Hristov, H., Berkowitz, C., Isaacson, R., Diaz Brinton, R., and Mosconi, L. (2018). Female sex and alzheimer's risk: the menopause connection. *The journal of prevention of Alzheimer's disease*, 5(4):225–230.

[Schiratti et al., 2015a] Schiratti, J.-B., Allassonniere, S., Colliot, O., and Durrleman, S. (2015a). Learning spatiotemporal trajectories from manifold-valued longitudinal data. In *Neural Information Processing Systems*, number 28 in Advances in Neural Information Processing Systems, Montréal, Canada.

[Schiratti et al., 2015b] Schiratti, J.-B., Allassonniere, S., Colliot, O., and Durrleman, S. (2015b). Learning spatiotemporal trajectories from manifold-valued longitudinal data. In *Neural Information Processing Systems*, number 28.

[Schiratti et al., 2017] Schiratti, J.-B., Allassonnière, S., Colliot, O., and Durrleman, S. (2017). A bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations. *The Journal of Machine Learning Research*, 18(1):4840–4872.

[Schwartz and Weintraub, 2021] Schwartz, J. B. and Weintraub, S. (2021). Treatment for alzheimer disease—sex and gender effects need to be explicitly analyzed and reported in clinical trials. *JAMA network open*, 4(9):e2124386–e2124386.

[Shao et al., 2018] Shao, H., Kumar, A., and Thomas Fletcher, P. (2018). The riemannian geometry of deep generative models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 315–323.

[Shen et al., 2019] Shen, S., Zhou, W., Chen, X., Zhang, J., and Initiative, A. D. N. (2019). Sex differences in the association of apoe $\varepsilon 4$ genotype with longitudinal hippocampal atrophy in cognitively normal older people. *European Journal of Neurology*, 26(11):1362–1369.

[Sinforiani et al., 2010] Sinforiani, E., Citterio, A., Zucchella, C., Bono, G., Corbetta, S., Merlo, P., and Mauri, M. (2010). Impact of gender differences on the outcome of alzheimer's disease. *Dementia and geriatric cognitive disorders*, 30(2):147–154.

[Skup et al., 2011] Skup, M., Zhu, H., Wang, Y., Giovanello, K. S., Lin, J.-a., Shen, D., Shi, F., Gao, W., Lin, W., Fan, Y., et al. (2011). Sex differences in grey matter atrophy patterns among ad and amci patients: results from adni. *Neuroimage*, 56(3):890–906.

[Spampinato et al., 2016] Spampinato, M., Langdon, B., Patrick, K., Parker, R., Collins, H., Pravata', E., and Initiative, A. D. N. (2016). Gender, apolipoprotein e genotype, and mesial temporal atrophy: 2-year follow-up in patients with stable mild cognitive impairment and with progression from mild cognitive impairment to alzheimer's disease. *Neuroradiology*, 58:1143–1151.

[Stern, 2012] Stern, Y. (2012). Cognitive reserve in ageing and alzheimer's disease. *The Lancet Neurology*, 11(11):1006–1012.

[Subramaniapillai et al., 2021] Subramaniapillai, S., Rajagopal, S., Snytte, J., Otto, A. R., Einstein, G., Rajah, M. N., Group, P.-A. R., et al. (2021). Sex differences in brain aging among adults with family history of alzheimer's disease and apoe4 genetic risk. *NeuroImage: Clinical*, 30:102620.

[Tanzi, 2012] Tanzi, R. E. (2012). The genetics of alzheimer disease. *Cold Spring Harbor perspectives in medicine*, 2(10):a006296.

[Tustison et al., 2019] Tustison, N. J., Holbrook, A. J., Avants, B. B., Roberts, J. M., Cook, P. A., Reagh, Z. M., Duda, J. T., Stone, J. R., Gillen, D. L., Yassa, M. A., et al. (2019). Longitudinal mapping of cortical thickness measurements: An alzheimer's disease neuroimaging initiative-based evaluation study. *Journal of Alzheimer's Disease*, 71(1):165–183.

[Tzourio-Mazoyer et al., 2002] Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289.

[Van Dyck et al., 2023] Van Dyck, C. H., Swanson, C. J., Aisen, P., Bateman, R. J., Chen, C., Gee, M., Kanekiyo, M., Li, D., Reyderman, L., Cohen, S., et al. (2023). Lecanemab in early alzheimer's disease. *New England Journal of Medicine*, 388(1):9–21.

[Verbeke, 1997] Verbeke, G. (1997). Linear mixed models for longitudinal data. In *Linear mixed models in practice*. Springer.

[Voevodskaya et al., 2014] Voevodskaya, O., Simmons, A., Nordenskjöld, R., Kullberg, J., Ahlström, H., Lind, L., Wahlund, L.-O., Larsson, E.-M., Westman, E., Initiative, A. D. N., et al. (2014). The effects of intracranial volume adjustment approaches on multiple regional mri volumes in healthy aging and alzheimer's disease. *Frontiers in aging neuroscience*, 6:264.

[Westman et al., 2013] Westman, E., Aguilar, C., Muehlboeck, J.-S., Simmons, A., et al. (2013). Regional magnetic resonance imaging measures for multivariate analysis in alzheimer's disease and mild cognitive impairment. *Brain topography*, 26(1):9–23.

[Whitwell et al., 2001] Whitwell, J. L., Crum, W. R., Watt, H. C., and Fox, N. C. (2001). Normalization of cerebral volumes by use of intracranial volume: implications for longitudinal quantitative mr imaging. *American journal of neuroradiology*, 22(8):1483–1489.

[Williams et al., 2021] Williams, C. M., Peyre, H., Toro, R., and Ramus, F. (2021). Neuroanatomical norms in the uk biobank: The impact of allometric scaling, sex, and age. *Human Brain Mapping*, 42(14):4623–4642.

[Wu and Zhang, 2002] Wu, H. and Zhang, J. (2002). Local polynomial mixed-effects models for longitudinal data. *Journal of the American Statistical Association*, 97(459):883–897.

[Young et al., 2014] Young, A., Oxtoby, N., Daga, P., Cash, D., Fox, N., Ourselin, S., Schott, J., and Alexander, D. (2014). A data-driven model of biomarker changes in sporadic Alzheimer's disease. *Brain*, 137(9):2564–2577.

[Zhao et al., 2016] Zhao, L., Mao, Z., Woody, S., and Brinton, R. (2016). Sex differences in metabolic aging of the brain: insights into female susceptibility to alzheimer's disease. *Neurobiology of aging*.

[Zhao et al., 2021] Zhao, Q., Liu, Z., Adeli, E., and Pohl, K. M. (2021). Longitudinal self-supervised learning. *Medical Image Analysis*, 71:102051.

[Zhou et al., 2014] Zhou, Q., Goryawala, M., Cabrerizo, M., Barker, W., Duara, R., and Adjouadi, M. (2014). Significance of normalization on anatomical mri measures in predicting alzheimer's disease. *The Scientific World Journal*, 2014.