



Progression models for Parkinson's Disease

Raphäel Couronné

► To cite this version:

Raphäel Couronné. Progression models for Parkinson's Disease. Artificial Intelligence [cs.AI]. Sorbonne Université, 2021. English. NNT: . tel-03491211v1

HAL Id: tel-03491211

<https://theses.hal.science/tel-03491211v1>

Submitted on 17 Dec 2021 (v1), last revised 22 Feb 2022 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PH.D. THESIS

submitted in part fulfilment of the requirements for

the degree of Doctor of Philosophy of

SORBONNE UNIVERSITÉ

In applied Mathematics

EDITE de Paris (ED 130)

Informatics, Telecommunications, Electronics

Aramis laboratory, ICM, Inserm, CNRS, Sorbonne Université, INRIA

Algorithms, models and methods for images and signals of the human brain

Progression models for Parkinson's Disease

Raphaël COURONNÉ

September 23th 2021

Advisor	STANLEY DURRLEMAN	Senior Research Scientist at INRIA
Reviewers	BRUNO JEDYNAK	Professor at Portland State University
	ARTHUR TENENHAUS	Professor at CentraleSupélec
Examiners	STÉPHANE LEHÉRICY	Professor at Sorbonne Université
	NEIL OXTOPY	Research Associate at University College London
	MARIA VAKALOPOULOU	Assistant Professor at CentraleSupélec

A mes parents.

PROGRESSION MODELS FOR PARKINSON'S DISEASE**Abstract**

In this work, we developed statistical methods to model disease progression from patient's repeated measurements, with a focus on Parkinson's Disease (PD). A key challenge lies in the inherent heterogeneity of PD across patients, to the extent that PD is now suspected to encompass multiple subtypes or motor phenotypes. To gain insights on disease progression, research studies propose to gather a broad range of marker measurements, at multiple timepoints for each patients. These data allow to investigate the disease's patterns of progression via statistical modeling. In a first part, we modeled the progression of scalar markers of PD. We extended on a disease progression model, namely the longitudinal spatiotemporal model. We then proposed to address data missingness, and to model the joint progression of markers of different nature, such as clinical scores, and scalar measurements extracted on imaging modalities. With this method, we modeled early motor progression in PD, and, in a second work, the heterogeneity of idiopathic PD progression, with a focus on sleep symptoms. In a second, independent, part of the manuscript, we tackled the longitudinal modeling of medical images. For these higher dimensionality data, Deep Learning is often used, but mostly in cross sectional setups, ignoring the possible inner dynamics. We proposed to leverage Deep Learning as a dimensionality reduction tool to build a spatiotemporal coordinate system of disease progression. We first took advantage of this flexibility to handle multimodal data. Then we leveraged the self-supervision induced by assuming monotonicity over time, to offer higher flexibility in modeling temporal variability.

Keywords: longitudinal data, spatio-temporal trajectories, machine learning, parkinson's disease, computer vision, generative models

MODÉLISATION DE LA PROGRESSION DE LA MALADIE DE PARKINSON**Résumé**

Dans ce travail, nous développons des méthodes statistiques pour modéliser la progression de la Maladie de Parkinson (MP) à partir de données répétées. La progression de la MP, très hétérogène, est complexe à modéliser. Pour mieux comprendre la progression des maladies neurodégénératives, des études effectuent un suivi de patients dans le temps, avec une batterie de tests à chaque visite. Ces données permettent d'étudier les différents types de progression par analyse statistique. Dans une première partie, nous modélisons la progression de marqueurs scalaires de la maladie de Parkinson. Nous nous basons sur un modèle longitudinal, le modèle longitudinal spatiotemporel. Nous proposons de gérer les valeurs manquantes, ainsi que de modéliser la progression jointe de marqueurs de différentes natures, comme les scores cliniques, ou les marqueurs extraits de l'imagerie. Avec ce modèle, nous nous concentrons d'abord sur la modélisation des symptômes moteurs précoces dans la MP. Puis nous étudions l'hétérogénéité de la MP, avec un accent sur les troubles du sommeil. Dans une seconde partie indépendante, nous étudions la modélisation de données longitudinales provenant de l'imagerie. Nous proposons d'utiliser un réseau de neurone comme méthode de réduction de dimension afin de construire un système de coordonnées spatiotemporel de progression de la maladie. Nous tirons parti de la flexibilité des réseaux de neurones pour modéliser la progression de données multimodales. Enfin, en supposant la monotonie des marqueurs au cours du temps, nous nous appuyons sur l'ordre des visites plutôt que l'âge des patients pour modéliser plus finement la variabilité temporelle de nos données.

Mots clés : données longitudinales, trajectoires spatiotemporelles, apprentissage automatique, maladie de parkinson, vision par ordinateur, modèles génératifs,

ARAMIS Lab

Institut de la Moelle Epiniere et du Cerveau – 47 Boulevard de l'Hopital – 75013 Paris – France

Remerciements

Je souhaite tout d'abord remercier les membres du jury, qui ont accepté de prendre de leur temps pour étudier et se prononcer sur mon travail de thèse. Arthur Tenenhaus et Bruno Jedynak, qui en tant que rapporteurs, ont pu apprécier mon travail et dont je suis très reconnaissants pour leurs retours. Ensuite, Maria Vakalopoulou et Neil Oxtoby, qui ont accepté d'examiner cette thèse, et dont je me réjouis d'avoir l'opportunité de rencontrer. Stéphane Lehericy, avec qui j'ai eu la chance de pouvoir travailler dans le cadre de cette thèse. Enfin, Stanley Durrleman qui a supervisé ce travail, pour son aide précieuse aux moments clés de cette thèse et dont je suis extrêmement reconnaissant de m'avoir donné l'opportunité de réaliser cette thèse.

En effet, cette thèse n'a pas été juste une aventure scientifique, mais aussi et surtout une opportunité de faire des rencontres inoubliables, dans le cadre multidisciplinaire, riche et accueillant de l'Institut de la Moelle Épinière et du Cerveau. Je souhaite remercier chaleureusement Marie Vidailhet, Jean-Christophe Corvol, qui avec Stéphane, ont donné de leur temps pour partager leurs connaissances d'experts mondiaux de la maladie Parkinson, et d'esquisser les objectifs à viser dans notre collaboration. Merci aussi à Graziella Mangone et Stéphane Epelbaum, pour avoir rajouté une réalité concrète à ce sujet de thèse. Merci à Olivier et à Fabrizio, qui avec Stanley conduisent cette formidable équipe Aramis. A Ninon, dont l'aide incommensurable les veilles de deadline ne peut être suffisamment développée ici. C'est grâce à vous que l'équipe Aramis est ce qu'elle est. Aussi, un petit mot pour Anne, dont j'ai pu apprécier l'immense altruisme pendant les mois où je l'ai connue.

En effet, travailler à Aramis ce n'est pas seulement la chance de pouvoir travailler dans un cadre riche, et d'avoir l'opportunité de continuer à apprendre sur des sujets qui nous passionnent. Mais aussi et surtout de pouvoir apprendre des autres tout au long de cette aventure, échanger, sur des sujets scientifiques ou pas, à l'heure du café autour d'un canapé, à l'Agraf, dans l'open space, ou dans le jardin d'une villa Florentine. Et c'est ma conviction profonde, que cette vie de labo est inscrite dans l'ADN de l'équipe, qui donne envie chaque matin d'y venir avec le sourire, confiant de retrouver une petite famille. Je souhaite donc remercier les collègues, qui pour beaucoup sont maintenant des amis, qui ont illuminés ces années au sein du laboratoire, dont je n'aurai la place ici d'énumérer toutes les qualités.

Tout d'abord, Maxime, Igor et Alexandre, premiers membres du laboratoire que j'ai rencontré. Que je ne vois pas seulement comme des mentors, avec qui j'ai beaucoup appris et aimé travaillé. Mais aussi des amis qui j'ai pu partager des moments mémorables, d'un Pink Floyd sur une route canadienne, aux festivités estivales ou palitos post-travail. Merci à Igor pour les innombrables week-end et sorties organisés d'une main de maître (meilleure

balle au prisonnier ever), Maxime d'avoir partagé avec succès sa passion de l'escalade, Alexandre pour avoir animé avec énergie la vie de l'équipe. Mais aussi à Jorge, Alexandre Routier, Hao, que j'espère revoir à la salle ou à une soirée dansante ou au spot qui sait. Pascal, pour son humour inimitable, je suis toujours un fan ! A Jeremy, et Catalina, pour des randos-stop quand tu veux. A Benoit, avec qui les projets politiques et duos endiablés de Guitare ont été des utopies trop belles pour ne durer plus qu'un été !

Mais aussi les nouveaux, Tristan pour la réparation automobile, Thibault (1v1 ZvT), P-E, Nemo, Ravi, Marius, Rémy, Charley, Benoit, Vito, Omar, partenaires de bloc, et tous les autres que j'aurais aimé plus connaître, le futur d'Aramis pèse sur vos épaules !

Évidemment mes contemporains, qu'il m'est si difficile de quitter. J'ai tout d'abord beaucoup aimé travailler avec Paul dont j'estime tout particulièrement les qualités scientifiques, et à qui je souhaite le meilleur pour la suite. Avec Thomas aussi, et même un peu trop sur ma fin de thèse, mais c'est aussi qu'avec lui comprendre la maladie de Parkinson devient un jeu passionnant. A Elina, pour son érudition scientifique mais aussi et surtout sociétoludique. A Etienne, pour avoir fait de Leaspy quelque chose de solide, Julia, Baptiste, Mauricio. Et comment ne pas remercier Juliana, Simona et Tiziana ? En ce qu'elles ont gravé dans la roche cette notion de famille. Ces années n'auraient pas été les mêmes sans elles, sans leur solidarité et amitié dans cette aventure. Simona, « ma pote » comme elle se décrit dans le rap jeu. Tu as toujours été là pour moi depuis le début. Je me souviens de notre rencontre et amitié instantanée comme si c'était hier ! (c'était ça où se détester vu la disposition des chaises de l'open space). Juliana, qui inonde le labo de ses rayons d'empathie. Tu es mon Sam de cette thèse du destin, je ne sais comment assez te le rendre. Tiziana, dont je respecterai ici l'orthographe, et dont le CV est chargé. Tour à tour coloc, partenaire de rando, compétitrice sportive, importatrice de produits Fiumicino-Orly... On aura tout fait pourrait-elle dire. Mais comment me lasser de ton sourire contagieux ? Bref, pour vous j'irai jusqu'au bout du monde, la, maintenant. « La fête n'est pas finie ».

Merci à Etienne, pour les retraites salvatrices et découvertes culinaires, et à Arnaud le S, coloc parfait en fusion télépathique, pour les soirées animés, les repas aux petits oignons, puisse l'ICM te rendre ce que tu as donné. Merci aussi à Arnaud Gloaguen et Vincent Chagot pour le Bleau, et la famille Caqueux pour ouvrir de temps à autre les portes du paradis marseillais.

Au quintes, les viés, qui m'ont connu et soutenu, avant tous les autres. Marc, Guillaume, Lucas, Luca, pour des trips en RL qui rendent la vie plus intense, une compétition sportive acharnée quoique braccassée, des bœufs endiablés, des délires tarés, et le gout du Risk. Pas seulement une rencontre de grands esprits, mais surtout un socle solide sur lequel je peux compter tout au long de ma vie. Si je suis tel que je suis maintenant, c'est grâce à vous !

A mes parents surtout, à qui je dois tout. Pour leurs conseils, leur amour, leur dévouement, et leur immense investissement mis dans mon éducation ainsi que celle de mes sœurs. J'espère qu'au travers de cette thèse, j'aurai su leur montrer à quel point j'en suis reconnaissant, et les rendre fiers. A Lara pour les dessins inspirants qui trônent sur mon bureau, et Alice, pour les bébés chats aux yeux bleus.

A Audrey enfin, qui m'a toujours soutenu depuis notre rencontre. A toutes ses atten-

tions, aux moments de tous les jours qui changent la perception de la vie. Sans toi, cette thèse n'aurait pas pu voir le jour, et le courage dont tu fais preuve tous les jours m'inspire.

« C'est pas Aramis qui me quitte, c'est moi j'quitte Aramis »

Table of Contents

Abstract	v
Remerciements	vii
Table of Contents	xi
List of Tables	xv
List of Figures	xvii
Introduction en langue française	1
Préambule	2
Motivation	2
Contexte	4
Objectifs de la thèse	7
Organisation du manuscrit	9
Introduction	11
Preamble	12
Motivation	12
Background	14
Thesis Objective	17
Manuscript Overview	18
I Background	21
1 Parkinson’s Disease	23
1.1 Preamble	24
1.2 Clinical Overview of Parkinson’s Disease	25
1.3 Materials	27
1.4 Goals in Parkinson’s Disease Longitudinal Analysis	31
2 Spatiotemporal Model of Disease Progression	41
2.1 Mathematical Background	42
2.2 Building a Riemannian Mixed Effect Model	53

2.3 Estimation	64
II PD progression models with low dimensional un-structured data	69
3 Disease progression model with missing values	71
3.1 Introduction	72
3.2 Methods	73
3.3 Results	75
3.4 conclusion	78
4 Clinical applications on Parkinson's Disease	79
4.1 Tools for clinical applications	80
4.2 Progression Model of Motor Abnormalities in Prodromal PD	86
4.3 Variability of clinical and imaging biomarkers in Parkinson's disease . . .	89
III Longitudinal Auto-encoders	109
5 Longitudinal Autoencoder	111
5.1 Deep Learning in disease progression modeling	112
5.2 Background	113
5.3 A longitudinal autoencoder	115
5.4 Conclusion	118
6 Multimodal longitudinal autoencoder	121
6.1 Introduction	122
6.2 Methods	123
6.3 Decoding : Non linear mixed effect model	123
6.4 Encoding	124
6.5 Regularization, cost function and optimization	124
6.6 Experimental results	125
6.7 Conclusion and perspectives	128
7 Longitudinal self-supervision	131
7.1 Introduction	132
7.2 Methodology	133
7.3 Longitudinal progression model	133
7.4 Modularity	135
7.5 Experimental results	136
7.6 Validation on synthetic data	136
7.7 Application to Alzheimer's disease	138
7.8 Conclusion	139

IV	Software development	141
8	Leaspy	143
8.1	Introduction	144
8.2	Inputs	144
8.3	Classes and methods	145
8.4	Performance Benchmark	146
8.5	Applications	146
	Conclusion and Perspectives	151
	Contributions	152
	Limitations and Perspectives	153
	Conclusion	156
	Publications	157
	Bibliography	159

List of Tables

1.1	Table of the main symptoms in Parkinson's Disease, labeled as primary or secondary depending of the symptoms frequency.	26
1.2	Biomarker inclusion in the considered groups of the 3 cohorts PPMI, ICE-BERG and DIGPD. "X" denotes longitudinal measures for biomarkers, while "(X)" denotes only cross-sectional measures, or very small number of measurements. "-" stands for no data. Note that prodromal group of PPMI include iRBD patients (n=39), as well as patients with hyposmia (n=26).	29
2.1	Description of chosen 1D manifolds. f is the chosen transformation from $\mathbb{R}$ to the 1D-Manifold I , $(f^{-1})'$ refers to the pushforward of $\mathbb{R}$ to I , and $G(p)$ to the metric derived from this pushforward.	57
4.1	On the left, estimated Prodromal Years between arrival of abnormalities and conversion to PD. On the right, estimated duration in years between abnormalities and conversion to PD using independant linear regressions (R. Postuma, Lang, et al. 2012). Our results are in a general agreement with those of a previous study from R. Postuma, Lang, et al. 2012, with still one difference in the ordering between Gait abnormalities and Bradykinesia.	88
4.2	Characteristics of the data sets. For each biomarker, we report the average and standard deviation at baseline. In PPMI, medications are not given from baseline, and therefore the first value seen for MDS-UPDRS III On accounts for a later age, in average 2.0 ± 1.3 years later than baseline. This explains in PPMI the higher value of MDS-UPDRS III On than MDS-UPDRS III Off. Abbreviations: SBR = striatal binding ratio, y = years.	96
4.3	Association between individual parameters in the multi-domain clinical course maps and the RBD status (PDRBD+ and PDRBD-) with correction for cohort effect, sex and baseline age. Statistically significant associations are shown in bold (corrected p-value are shown). Terms into brackets correspond to confidence interval.	101

4.4	Mann-Whitney test between biomarkers values at baseline between PD patients of different cohorts: PPMI and ICEBERG, PPMI and DIGPD, and ICEBERG and DIGPD. Age distribution are matched, but disease duration differs significantly between cohorts. PPMI has the lower disease duration at baseline, followed by ICEBERG and DIGPD. This explains lower severity of biomarkers at baseline in PPMI with respect to both DIGPD and ICEBERG. However, ICEBERG has lower disease duration than DIGPD, but higher biomarker severity at baseline, which is counter-intuitive and shows a bias. Compared to PPMI, it shows a noticeably high difference of MDS-UPDRS III Off at baseline.	107
4.5	Association between individual parameters in the multimodal course maps and the RBD status (PDRBD+ and PDRBD-) with correction for cohort effect, sex and baseline age. Statistically significant associations are shown in bold (corrected p-value are shown). Note that for this experiment, the multimodal course map has been re-trained without RBDSQ as endpoint.	108
6.1	Mean 10-fold reconstruction error for the 2 cognitive scores experiments for each modality respectively	126
7.1	Benchmark of proposed methods on Starmen dataset	137

List of Figures

- 1 Illustration d'un dataset longitudinal à N patients, vus à différentes visites. Les couleurs de cerveau expriment le stade d'avancement de la maladie. Les largeurs et hauteurs de cerveaux spécifiques à chaque sujet représentent la variabilité inter-patient. Un trait plus large indique la partie de la trajectoire patient observée lors de son suivi, alors que une ligne en pointillé indique l'absence de données. Les cerveaux en transparence aux extrémités expriment notre hypothèse de monotonie. C'est à dire que tous les patients sont supposé être complètement sains à un certain âge, et évoluer inexorablement vers le stade maximal de la maladie. 5
- 2 Représentation de l'effet du réalignement temporel des sujets selon leur stade de la maladie, par rapport à une trajectoire moyenne abstraite (en bleu). Le modèle longitudinal spatiotemporel permet de désenchevêtrer la variabilité temporelle, qui modélise les variations en terme de stade d'avancement dans la maladie, et la variabilité spatiale, qui modélise les différences inter sujets. Ici, la variabilité spatiale est représentée avec des hauteurs et largeurs de cerveaux spécifiques à chaque sujet. Nous observons aussi que l'âge n'est pas informatif du stade d'avancement dans la maladie. 7
- 3 Illustration of a longitudinal dataset with N patients, seen at different visits. Color of the brain accounts for disease stage. Different height and width of the brain account for possible inter-patient variability. A wide line indicate the time-span of each patient follow-up, while dotted lines indicate unseen patient data. Transparent brains at extremities relate to our monotonicity assumption, that is to say patient with the disease were healthy at some point, and will eventually reach abnormality in all biomarkers. 14
- 4 Effect of temporal re-alignment of subjects with regard to their disease stage, according to an abstract average trajectory (in blue). The longitudinal spatiotemporal model allows to disentangle temporal variability, accounting for variations in terms of disease stage, and spatial variability, which is subject-dependent. In this case, spatial variability is denoted with a subject-specific height and width of the brain. Note that age is not informative of disease stage. 16

1.1	Schematic overview of the topological propagation of Lewy Bodies in the brain during Parkinson's Disease. The red shading represents the presence of Lewy Bodies. On the left, in a first phase, Lewy Bodies are found in the viagal nerve and olfactory bulb, explaining smell loss and the presence of non-motor symtoms. Then Lewy Bodies propagate until they damage the substantia nigra, leading to sleep and motor disturbances. Ultimately, propagation until the cortex areas induce dementia. Note that this is an average scheme derived via post-mortem observations, and it does not encompass the whole heterogeneity of the disease, as many patient differ from this scenario in terms of symptoms ordering. (Reproduced from Braak et al. 2003).	32
1.2	Timeline of successive phases in prodromal Parkinson's Disease. The prodromal phase may begin up to 10 years before conversion to Parkinson's Disease, with first symptoms in sleep and autonomic dysfunction (especially constipation). (Reproduced from Ronald B Postuma and Berg 2016).	33
1.3	Identified prodromal markers and their score. Markers most indicative of conversion to PD are "(PSG verified) RBD" (130×), "abnormal substantia nigra on imaging" (42×), and "orthostatic hypertension" (19×). (Reproduced from Heinzel et al. 2019).	34
1.4	Combined progression trajectory of motor and non-motor manifestations from prodromal stages to phenoconversion based on actual measurements in a population of RBD patients. For each biomarker, different regressions are performed with respect to age at conversion: linear regression, weighted least squares (WLS) regression, and non-linear third-degree polynomial regression, and more complex ones are used if they provide a significant increase in R^2 ($> 10\%$). (Reproduced from Fereshtehnejad, Yao, et al. 2019).	35
1.5	Schematic overview of Brain First versus Body First Theory. In (A), for the patients with a body-first form of the disease, abnormalities would begin in the guts, on the form of Lewy Bodies, and then propagate through the autonomous nervous system to the locus coreuleus and substantia nigra. iRBD would appear as a prodromal form of this subtype. In (B), subjects with a brain first form of the disease, progression would begin in the amygdala, to then propagate to the substantia nigra. This form is close to the one represented in Figure1.1. (Reproduced from Horsager, Andersen, Knudsen, Skjærbæk, et al. 2020a).	39
2.1	Illustration of a random slope random intercept model. In this model, t_0 is fixed a-priori.	44
2.2	Representation of different manifolds with a tangent space at point p . On top, intrinsic view of a manifold, with a representation of the tangent space $T_p M$ at point p . On the bottom, we show the sphere of $\mathbb{R}^3$ as a 2-dimensional manifold immersed in $\mathbb{R}^3$, which is an example of extrinsic view of a manifold. Note that the two approach are equivalent. In the following, we will use the intrinsic view for vizualization purposes.	48

2.3	Representation of a geodesic γ of M parameterized by its speed $\dot{\gamma}$ at position p . In addition we show a vector $w \in T_p M$	50
2.4	Illustration of the construction of a parallel variation $\eta^w(\gamma)(t)$ of γ with regard to w . First, w is transported along γ , such that its parallel transport at time point t writes $P_{\gamma, t_0, t}(w)$, at position $\gamma(t)$. Then from $\gamma(t)$, we follow the Exponential map $\text{Exp}_{\gamma(t)}^M(P_{\gamma, t_0, t}(w))$ in the direction $P_{\gamma, t_0, t}(w)$ to build $\eta^w(\gamma)(t)$. The curve $t \mapsto \eta^w(\gamma, \cdot)$ is built from the values $\eta^w(\gamma)(t)$. Note that this curve is not necessarily a geodesic of M	51
2.5	A Spatiotemporal model in a Smooth Riemannian Manifold M	53
2.6	Examples of 1D geodesics on the 1D-Manifolds used. On the left $\mathbb{R}$, on the middle $]0, 1[$, and on the right $\mathbb{R}^+$. This leads respectively to linear, sigmoid and exponential decay curves.	57
2.7	Description of the model on geodesics of $]0, 1[^2$ equipped with the induced metric. On the top row, biomarker model of progression with regard to time. On the bottom, intrinsic view of M with geodesics as parametric curves of $\mathbb{R}^2$. On the first column, we show the average trajectory as a geodesic of M , parameterized by p, v, t_0 . On the second column, visualization of a spatial variability via the exp-parallelization of γ in the direction $w_i = (-0.2, 0.05)$. On the two last columns, effects of temporal variability. On the third column, effect of the onset individual parameter $\tau_i = 3$, which shifts the curves in time. This spatial variability shifts biomarker curves in time, while ensuring that this "spatial" variation is orthogonal to the speed v at p ("temporal" variability) in order to preserve identifiability of the model. On the last column, effect of α_i which dilates the curves from the reference time $t_0 + \tau_i$. In diamond, patient i measurements that we aim to fit with these successive transformations from the average trajectory.	61
2.8	Comparison between a random slope random intercept model $y_{i,j} = (t_{i,j} - t_0)(\beta_1 + b_{1i}) + (\beta_2 + b_{2i}) + \epsilon_{i,j}$ and a linear univariate instantiation of the spatiotemporal model. Note that the spatiotemporal model is not linear in the sense of parameters as there is a non-linear effect between speeds of progression α_i and onset τ_i : $y_{i,j} = p_i + v(\alpha_i(t - t_0 - \tau_i)) + \epsilon_{i,j}$	63
2.9	Illustration of the difference of parameterization between a multivariate model of $]0, 1[^5$ and 5 univariate model of $]0, 1[$. The different univariate models have different timelines, with each additional " t_0 " reference time parameters. In addition, although not shown here, each univariate model will have τ_i and ξ_i parameters, leading to 10 individual parameters. On the other hand, the multivariate model will have τ_i and ξ_i , as well as sources of dimension $q < K$, therefore having at maximum $\text{dim} = 2 + (5 - 1) = 6$	63
3.1	Representation of the model on a schematic manifold. Each point $\eta^{w_i}(\gamma_0, t)$ is obtained via the continuous transportation of the tangent vector w_i along the reference geodesic γ_0 (in red). $\eta^{w_i}(\gamma_0, \cdot)$ (in purple) is a "parallel" of γ_0 . The model for patient i consists then in the trajectory $\eta^{w_i}(\gamma_0, \cdot)$ travelled at the subject-specific time $\psi_i(t)$	73

- 3.2 Bootstrap distribution ($b=100$) of errors of estimation for population parameters $(\mathbf{p}_0, \mathbf{v}_0, t_0)$ and individual parameters τ_i and ξ_i according to subsampling frequency on an artificial dataset. Parameter values are taken close to estimation on real-world data. The second modality is assigned to NAs at various frequency to compare performance worsening between the naive method (boxplots on the left), and the generative modeling method (boxplots on the right). 75
- 3.3 Bootstrap distribution ($b=100$) of errors of estimation for population parameters $(\mathbf{p}_0, \mathbf{v}_0, t_0)$ and individual parameters τ_i and ξ_i according to subsampling frequency on a real (ADNI) dataset. The 4 modalities used are subscores of the ADAS-COG accounting respectively for memory, praxis, language and concentration. 2nd, 3rd and 4th modalities are assigned to NAs at various frequency to compare performance worsening between the naive method and the generative modeling method. 77
- 3.4 In wide plain lines the mean geodesic estimated for PPMI PD patients on 2 modalities : MDS-UPDRS Score and Right Caudate SBR obtained from DatScan, and described by population parameters $\mathbf{p}_0 = [p_0^0, p_0^1]$, $\mathbf{v}_0 = [v_0^0, v_0^1]$ and t_0 . The observations (in dotted lines) and individual models (in narrow lines) of 3 patients are also plotted. 78
- 4.1 Data points are assumed to lie on manifold M , and patients follow trajectories $\eta_{\gamma}^{w_i}(t)$ (3) parallel to the average trajectory $\gamma(t)$ (1). For a group G_2 of patients, we propose to average their spatial representations $\{w_i, i \in G_2\}$ to build the group average trajectory (2). In plain line we show the intervals of trajectories where data was seen, in terms of disease stage, and in dotted lines, parts of trajectories where no data was seen. With color red and grey we highlight groups of patients: G_2 encompasses red trajectories, while G_1 englobes grey ones. Modeling the data with the model in the plain line zone accounts for interpolation task, that we can perform at the 3 levels: individual, group or average, for e.g. missing data imputation. Modeling the data with the model outside the plain line zone accounts for extrapolation, may it be in the future, for prognosis, or in the past, to try to estimate how the disease start. 82
- 4.2 Illustration of the use of pathological thresholds to build a timeline of abnormalities (shown in D). First a longitudinal spatiotemporal model is trained on the patient longitudinal data (C). Then a logistic regression method determines the thresholds that best separate control from patients data at baseline for each endpoint (A,B). The thresholds are reported on the y-axis of the average trajectory (C). The Disease Age at which the endpoint become pathological is determined (C). We report the series of PA for each endpoint together with the bootstrap confidence interval on the reference timeline (D). 85

- 4.3 Example of 4 iRBD patient data. Each graph shows the values of a given patient. In color we show the normalized values of the motor score MDS-UPDRS part III (yellow), and subscores of the MDS-UPDRS part III: Rigidity (kaki), Bradykinesia (red), Gait (green), and Voice-Face (blue). A Parkinsonism level (PSLVL) score is plotted in plain grey. We see that these scores are inherently noisy, and that patient are not necessarily at the same stages of the diseases. 86
- 4.4 Normative scenario of biomarker's progression in prodromal PD patients with regard to "Prodromal Age". We compute each biomarker abnormality cutoff value as a the 95% quantile on the control group. We then look at intersections between the normative scenario and these cutoffs values of abnormalities to get timings of abnormalities (diamonds icons). For scale, we add on the normative scenario the reparametrized events of conversion to the last 3 stages of Parkinsonism Level (PSLVL): 3 (may be PD), 2 (Likely PD), and 1 (PD). 87
- 4.5 On the left, bootstrap distribution of "Prodromal Ages" at abnormality for each motor marker. Bootstrap distribution is shown with a kernel density estimation. Both population variability and estimation variability are captured in the bootstrap procedure. On the right, bootstrap distribution of "Prodromal Ages" at abnormality for each motor marker, relatively to the reparameterized event of conversion to Parkinson's Disease. Boxplots representing Prodromal Ages at conversion to PSLVL values of 3, 2, 1 are also added. In this case distributions are shown with boxplots. Relatively to PD conversion, we find that Voice-Face modality is affected first, with median time to conversion of 9.2 years. Then Gait, Bradykinesia, MDS-UPDRS and Rigidity at respectively 7.3, 6.1, 4.6 and 2.9 years before conversion. Note that we show in these graphs the variability of the population parameters, and not the inter-patient variability, which would be higher, and with more overlapping between the different biomarkers. 88
- 4.6 Positioning of a subject in the PD Course Map. A) The typical progression of two biomarkers (blue and yellow) is shown by plain logistic curves together with the corresponding biomarker values of one subject at 5 successive visits. Biomarker values are normalized so that 0 corresponds to the most normal value and 1 the maximal pathological changes (see Methods). B-to-D) A series of operations allow us to transform the model to fit the subject data and therefore determine the subject's specific trajectory. The inter-marker spacings change the time interval between the two curves (B), the time-shift translates the curves (C) and the acceleration factor stretches the curves (D). The last two operations map the actual age of the patient to his Parkinsonian Age. We estimate the shape and position of the logistic curves, so that the average time-shift, inter-marker spacing, and log-acceleration factor for all subjects in the training longitudinal data set is zero. 93

4.7	Multimodal disease course map trained on PPMI showing the progressive onset of clinical endpoints and imaging biomarkers from pre-symptomatic to post-diagnosis stages. Left panel: the progression of the endpoints from the most normal (0) to the maximum pathological changes (1) is shown as Parkinsonian Age progresses. Shaded areas correspond to 95% bootstrap confidence interval. Right panel: the timing at which endpoints become abnormal is shown with 95% confidence intervals. Vertical lines indicate the estimated Parkinsonian Age of first symptoms (light red) and of diagnosis to PD (dark red).	97
4.8	Correlation matrix of inter-marker spacings showing a division of endpoints into three groups: body-related clinical endpoints, brain-related clinical endpoints and imaging biomarkers. Green color means positive correlations. Pink color means negative correlations.	98
4.9	Multi-domain clinical course map trained on PPMI PD patients shows the different timing of symptom onset for PDRBD+ and PDRBD- patients. Parkinsonian Ages at which clinical endpoints become pathological for PDRBD+ (plain colors) and PDRBD- (transparent colors) are shown. Vertical lines indicate the estimated Parkinsonian Age of first symptoms (light red) and of diagnosis to PD (dark red).	100
4.10	Multi-domain clinical course map trained on PPMI PD patients shows the different timing of symptom onset for iRBD patients. The abnormalities begin in sleep and cognition, followed by dysautonomia and activities of daily living, and finish with motor symptoms.	102
4.11	Individual parameter distributions per subgroup, shown via four 2D kernel density estimations. Distribution of progression parameters position PDRBD+ patients as an intermediate phenotype compared to PDRBD- and iRBD. A) Shows temporal individual parameters, while B-C-D) show spatial individual parameters, aka intermarker spacings, accounting for different patient phenotypes after patients temporal realignment. In A) we show Onset in the x-axis against Speed of progression in the y-axis. We observe no significant difference in terms of speed of progression between the groups, while Onset is earlier in PDRBD+ than PDRBD-. iRBD patients have late onset. Indeed, they are recruited at a late age, and with fewer symptoms than PD patients. Consequently the model re-align them at the beginning of the PD Course Map. In B) we show non-motor activities of daily living (y-axis) against dysautonomia (x-axis), which are correlated, as already mentioned in Figure 4.8. iRBD appear as having an early non-motor phenotype, compared to PDRBD-, with PDRBD+ in between. In D) this gradation is naturally seen in the y-axis accounting for sleep, while iRBD show very late motor symptoms. Finally in C), cognition and motor aspects of daily living show few difference between PDRBD+ and PDRBD- patients, while iRBD patients show late subjective motor symptoms, and earlier cognitive symptoms.	103

- 4.12 Clinical disease course map trained on the pooling of PPMI, ICEBERG and DIGPD. It describes the progressive onset of clinical endpoints and imaging biomarkers from pre-symptomatic to post-diagnosis stages. Left panel: the progression of the endpoints from the most normal (0) to the maximum pathological changes (1) is shown as Parkinsonian Age progresses. Shaded areas correspond to 95% bootstrap confidence interval. Right panel: the timing at which endpoints become abnormal is shown with 95% confidence intervals. Vertical lines indicate the estimated Parkinsonian Age of first symptoms (light red) and of diagnosis to PD (dark red). 106
- 5.1 Autoencoder schematic view. Encoder and decoder are composed of many layers: $\Psi = \Psi_1 \circ \dots \circ \Psi_M$, $\Phi = \Phi_1 \circ \dots \circ \Phi_L$, with $\Phi_m : x \mapsto \sigma_m(W_m x + b_m)$ and $\Psi_l : x \mapsto \sigma_l(W_l x + b_l)$ 114
- 5.2 Schematic view of the longitudinal autoencoder. The encoder is a sequence-wise autoencoder, which encodes all patient i 's visits $x_{i,j}$ and times $t_{i,j}$ into a patient representation w_i and affine time reparameterization parameters α_i and τ_i . The point-wise decoder maps points $(\psi_i(t), w_i^1, \dots, w_i^{p-1})$ of the latent space back into the data space. Brains in transparency indicate how the latent space maps to the data space via the decoder. ψ encodes the disease stage, represented by brain colors, while w encodes patient specificity, represented by patient specific brain height and width in the graph. 116
- 5.3 Each row represents a synthetic subject. Columns show visits at increasing times for these subjects. A subject has specific arm's angles. Disease progression is modeled as a decrease of all arm's lengths. (Reproduced from (Louis et al. 2019)) 117
- 5.4 We learn a model on the synthetic cross data, with $\dim(\mathcal{Z}) = 3$, and visualize in these 2 graphs the induced latent space. On the left, we show multiple trajectories over time (x-axis) corresponding to 5 different points in the spatial part of the latent space. Each row is a trajectory with fixed spatial component w , and ψ varies from $\psi = -1$ to $\psi = 1$. The first row depicts the average trajectory $\psi \mapsto \Psi(\psi, 0, 0)$, while the following rows depict parallel trajectories, shifted from $\pm e_2$ or $\pm e_3$ to visualize their effects. These parallel directions of progression show the same arm length reduction scenario, with different arm positions. On the right we show spatial individual variables w_i^1 and w_i^2 colored by left (top) and right (bottom) arm angle of the train data. We observe that w^1 captured the left arm angle, while w^2 captured the right arm angle. (Reproduced from (Louis et al. 2019)) 118
- 5.5 On the left learned average trajectory of the cognitive scores, with a 10-fold resampling procedure ($\dim(\mathcal{Z}) = 2$). On the center, spatial variation in the direction e_1 . On the right, spatial variation of the longitudinal spatiotemporal model, given by Leaspy. (Reproduced from (Louis et al. 2019)) 119

5.6	Average trajectory $\psi \mapsto \Psi(\psi, 0, \dots, 0)$ learned from MRIs of ADNI MCI converters patients. Rows show respectively sagittal, coronal and horizontal planes of the average trajectory. The growth of the ventricles, characteristic of aging and Alzheimer's disease is clearly visible. (Reproduced from (Louis et al. 2019))	119
6.1	Description of the proposed longitudinal autoencoder.	124
6.2	Left: average trajectories for the 10 folds, with increasing partitioning of the input features. Right: average trajectories for the 10 folds, with increasing pruning of the praxis+concentration modality.	126
6.3	Left: average trajectory and reconstruction examples for the scalar data. Right: average trajectory and some reconstructions for the image data.	127
6.4	Left: description of the prediction setup. Right: the MRI prediction errors.	128
6.5	Left: average trajectory. Right: prediction error, in the same setup as in section 3.2	129
7.1	Input data y_{ij} is encoded simultaneously in a space encoder (Deepset) and a point-wise time encoder to get latent representations (z_i^s, z_j^ψ) . z_i^s can be computed from <i>any</i> subset of visits, and in practice randomized fixed-size subsets of visits are drawn in the spirit of stochastic optimization. Decoder can be either agnostic, or specific (e.g., velocity fields for deformations).	134
7.2	Each row represents a synthetic subject across time.	136
7.3	Gradient directions in the latent space (extracted from a forward pass in the decoder). Row 1: gradient wrt to the latent space associated with disease progression. Rows 2 and 3: first two principal directions of the PCA in the orthogonal of the latent time ($\mathcal{Z}^s$ for us). 4th row: the direction in the orthogonal of the latent time that correlates most with it (PLS), as a way to challenge the model disentanglement.	137
7.4	Estimated average trajectory of scores. The effects of latent dimensions z^s (resp. z with β -VAE) are shown with degraded colors.	138
7.5	PLS analysis with respect to ventricle volume ratio $\mathcal{V}$: z^ψ (left), z^s (right).	139
8.1	Computation cost benchmark relative to respectively: number of visits in the dataset, number of sources and number of biomarkers. Leaspy is linear in the 3 variables. On the left, we set $K = 10$, $Ns = 4$ and vary the number of visits. On the middle, we set $N = 50$ patients, each with $m = 10$ visits, $K = 100$ and change the number of sources Ns from 1 to 99. On the right, we set $N = 50$ patients, each with $m = 10$ visits, we set $Ns = 4$ and change the number of biomarkers from $K = 5$ to $K = 100$	146
8.2		148

-
- 8.3 Example of real data on top, and synthetic data on the bottom. A model is calibrated and personalized on the real data, which in this case include 5 clinical scores. Then from the learned distribution of individual parameters, we sample synthetic individual parameters accounting for synthetic patients. Coupled with a deterministic or sampling scheme for patient ages, we can sample synthetic patient longitudinal data. 149
- 8.4 Poitwise correlations of real-data biomarkers on the top left, and synthetic biomarkers on the bottom right. We observe that correlation structure of the biomarker is preserved in synthetic data. 150

Introduction en langue française

Outline of the current chapter

Préambule	2
Motivation	2
Maladie de Parkinson	2
Connaissances et défis dans la maladie de Parkinson	3
Leviers d’actions	3
Contexte	4
Cohortes longitudinales	4
Défis dans l’analyse statistique de la maladie de Parkinson	4
La modélisation de progression de maladies, et le réalignement temporel	5
Objectifs de la thèse	7
Objectifs	7
Contributions	8
Organisation du manuscrit	9
Partie I: Contexte	9
Partie II Modèles de progression de la maladie de Parkinson avec des données en faible dimension et non structurées	10
Partie III Autoencoder longitudinal	10
Partie IV Développement logiciel	10

Préambule

Mettre en évidence, et quantifier l'hétérogénéité des schémas de progression de la Maladie de Parkinson (MP) pourrait permettre de mettre en lumière les mécanismes pathologiques sous-jacents, et ainsi améliorer grandement notre compréhension de la maladie. En effet, depuis sa découverte en 1817, la maladie de Parkinson est loin d'être pleinement comprise, et les traitements n'ont pas significativement évolués depuis l'utilisation de la Levodopa, un précurseur de la dopamine, dans les années 50.

Récemment, la médecine a grandement bénéficié d'outils nouveaux provenant d'avancées en physique, comme l'imagerie par résonance magnétique (IRM), l'émission de positron par tomographie (PET), ainsi que le séquençage génétique. Ces outils ont permis d'atteindre de nouvelles frontières dans la caractérisation des processus anatomiques et biologiques des maladies, au point que certaines peuvent être diagnostiquées et suivies précisément. Cependant, la complexité des maladies neurodégénératives est telle, que des équipes de recherche entières étudient des sous-parties d'un mécanisme pathologique plus global, et constituent actuellement notre principal espoir pour développer des thérapies efficaces.

L'apprentissage automatique offre de nouveaux moyens d'accroître nos connaissances sur ces maladies, en analysant de façon statistique et algorithmique des données patients. Ces marqueurs peuvent être de différentes natures : batterie de tests cliniques, modalités d'imagerie, prélèvements biologiques ou encore génétique. Les marqueurs indicatifs de l'état d'avancement de la maladie sont typiquement suivis dans le temps (Jiang et al. 2017). Ces algorithmes montrent déjà de bonnes performances en diagnostic et pronostic si on leur spécifie des tâches claires, comme la segmentation de tissus (Havaei et al. 2017; Menze et al. 2014; Milletari, Navab, and Ahmadi 2016; Pereira et al. 2016), ou encore la classification de patients. Ces algorithmes ont un fort potentiel pour accroître nos connaissances sur des maladies complexes.

Motivation

Maladie de Parkinson

La maladie de Parkinson est un exemple de maladie neurodégénérative complexe, avec des symptômes moteurs croissants qui peuvent altérer considérablement la qualité de vie des patients. La maladie est habituellement diagnostiquée autour de 60 ans, mais peut aussi toucher des personnes plus jeunes : 20% des patients sont diagnostiqués avant 50 ans (*Forme précoce de la maladie de Parkinson* n.d.). La MP touche 6.3 millions de gens dans le monde (Rocca 2018), et l'impact de la maladie en terme de qualité et espérance de vie, ainsi que coûts sociaux et monétaire sont amenés à augmenter avec le vieillissement de la population (Findley 2007).

Connaissances et défis dans la maladie de Parkinson

Dans les dernières décénies, des avancées importantes ont permis de mieux comprendre la pathophysiologie de la maladie. Il est maintenant communément accepté que la maladie commence longtemps avant le diagnostic reposant sur des symptômes moteurs anormaux, avec une phase "prodromale", principalement asymptomatique (Mahlknecht, Seppi, and Poewe 2015; Ronald B Postuma and Berg 2016). Cette phase prodromale peut commencer jusqu'à 10, voir même 20 ans avant le diagnostic (Kalia and Anthony E Lang 2016; Ronald B Postuma and Berg 2016; Savica, Rocca, and J Eric Ahlskog 2010), avec d'abord des dommages au système nerveux autonome, lié à des troubles du sommeil et de constipation, suivi par la perte de neurones dopaminergiques dans la substance noire. Ces dommages dans la substance noire sont considérés comme la cause directe des symptômes moteurs parkinsoniens, et peuvent être mesurés par imagerie du cerveau. Cependant, les mesures révèlent qu'au diagnostic, 60% des neurones dopaminergiques ont déjà disparus (P. N. Lee et al. 2007). Cela souligne l'importance de trouver des traitements pouvant agir dans les phases précoces de la maladie.

De plus, la maladie de Parkinson est hétérogène par nature, au point qu'elle est maintenant supposée englober plusieurs sous-maladies, ou au moins plusieurs sous-types de la maladie. Un schéma de progression a été proposé pour expliquer la topologie des dommages dans le cerveau (Braak et al. 2003). Ce schéma propose d'expliquer la progression de la maladie par la diffusion progressive de Corps de Lewy dans le cerveau, mais beaucoup de patients différent de ce schéma dans leur manifestation clinique. Afin de grouper les patients dans des sous-types cohérents, il a été proposé de grouper par phénotype moteur (Joseph Jankovic et al. 1990; Zetuskys, Joseph Jankovic, and Pirozzolo 1985), présence de troubles du sommeil (Berg, Borghammer, et al. 2021; Horsager, Andersen, Knudsen, Skjærbæk, et al. 2020a; Yoritaka et al. 2009), génétique, mais aucune convention n'a été adopté pour l'instant.

En fin de compte, la recherche sur la maladie de Parkinson vise à :

- Développer un traitement efficace, à administrer le plus vite possible.
- Dans l'absence d'un tel traitement, administrer au mieux les traitements actuels, en prenant en compte leurs effets secondaires.
- Prédire l'arrivée de symptômes spécifiques qui altèrent la qualité de vie des patients, comme la démence, les chutes ou la dyskinésie, pour adapter au mieux les traitements.

Leviers d'actions

A la différence des approches "ascendantes" où la connaissance de la maladie est construite par la compréhension de ses multiples sous-mécanismes, des approches "descendantes" proposent de raffiner incrémentalement notre connaissance de la maladie de Parkinson en testant des hypothèses sur les données. Avec suffisamment de précision, les approches descendantes peuvent alimenter la réflexion au niveau des mécanismes pathologiques, afin d'accélérer la recherche sur la maladie de Parkinson. Actuellement, ces approches consistent à :

- Identifier des marqueurs précoces de la maladie afin de mieux comprendre et suivre la phase pré-diagnostic de la maladie. Cela pourrait aussi permettre une intervention précoce (Murman 2012), et expérimenter de nouveaux traitements, dont l'efficacité pourrait être liée à la rapidité de leur utilisation.
- Extraire des sous-types de patients. Un sous-groupe de la maladie pourrait signifier un mécanisme de progression différent. Par exemple, les patients avec le gène GBA, aussi présent dans la maladie de Gaucher, pourraient potentiellement bénéficier des avancées thérapeutiques sur cette maladie (Riboldi and Di Fonzo 2019).

Contexte

Cohortes longitudinales

Pour réaliser ces tâches, nous avons besoin de collecter des données, le plus possible et de la meilleure qualité possible. Ainsi, ces dernières décennies, des données patients ont été collectées à grande échelle, et agrégées dans des cohortes de recherche pour être ensuite traitées avec des méthodes statistiques. Un effort particulier a été investi pour créer et partager publiquement les cohortes Alzheimer's Disease Neuroimaging Initiative (ADNI) pour la maladie d'Alzheimer (Petersen et al. 2010), ainsi que Parkinson's Progression Markers Initiative (PPMI) pour la maladie de Parkinson (Marek, Jennings, et al. 2011). Ces deux cohortes sont "longitudinales", au sens où elles suivent des patients dans le temps. Ainsi, des mesures exhaustives sont réalisées à plusieurs visites espacées dans le temps, afin de pouvoir suivre les progressions individuelles (cf Figure 3). Cela permet de considérer ces progression comme des "trajectoires" dans l'espace des mesures, et développer des méthodes statistiques en conséquence. Des modélisations et analyses pertinentes de ces cohortes longitudinales pourraient permettre d'identifier de nouveaux marqueurs de progression, ou d'identifier des critères d'inclusions plus fins pour les patients de futures cohortes (Berg, Ronald B Postuma, et al. 2015; Heinzl et al. 2019), ou d'essais cliniques.

Défis dans l'analyse statistique de la maladie de Parkinson

Avec ces données longitudinales, l'identification de marqueurs précoces ainsi que l'extraction de sous-types de patients est loin d'être facile. Pour l'extraction de sous-groupes, la plupart des études cliniques réalisent un clustering à la première visite, et dans un second temps seulement comparent les progressions de chaque groupe (Eisinger et al. 2017; Fereshtehnejad, Romenets, J. B. Anang, et al. 2015; Fereshtehnejad, Zeighami, et al. 2017). Par construction, ces études ne peuvent pas extraire des groupes en étudiant directement les différences de dynamiques. Concernant l'identification de marqueurs précoces, une méthode intuitive consiste à régresser les marqueurs en arrière dans le temps, et regarder quand les courbes intersectent les valeurs des sujets sains. Cependant, il n'existe pas de temps de référence dans les maladies neurodégénératives, comme la date de naissance le serait pour des courbes de croissance. Les études cliniques utilisent souvent comme

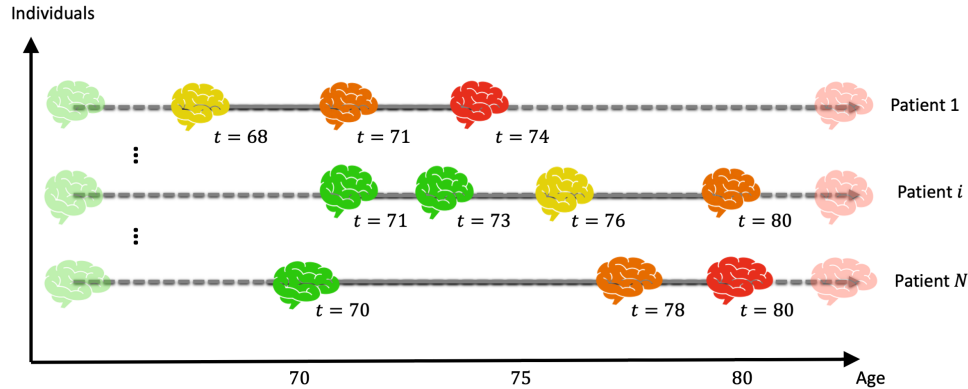


Figure 1 – Illustration d'un dataset longitudinal à N patients, vus à différentes visites. Les couleurs de cerveau expriment le stade d'avancement de la maladie. Les largeurs et hauteurs de cerveaux spécifiques à chaque sujet représentent la variabilité inter-patient. Un trait plus large indique la partie de la trajectoire patient observée lors de son suivi, alors que une ligne en pointillé indique l'absence de données. Les cerveaux en transparence aux extrémités expriment notre hypothèse de monotonie. C'est à dire que tous les patients sont supposé être complètement sains à un certain âge, et évoluer inexorablement vers le stade maximal de la maladie.

référence l'âge au diagnostic (Fereshtehnejad, Yao, et al. 2019; R. Postuma, Lang, et al. 2012), ce qui peut cacher des différences significatives entre les patients en terme de stade dans la maladie et vitesse de progression.

En prenant du recul, ces défis sont plus généraux que l'analyse longitudinale de la maladie de Parkinson, et se posent dans le contexte de l'analyse longitudinale de données médicales. La difficulté de l'acquisition de données médicales, et la grande hétérogénéité inter-patients expliquent en grande partie ces défis. Ainsi, comme ces cohortes longitudinales suivent des patients pour une durée limitée, elles subissent à la fois une censure à gauche, typiquement avec une absence de donnée pre-diagnostic, ainsi qu'une censure à droite. De plus, cette durée de suivie est limitée par rapport à la durée totale de la maladie, ce qui complique l'analyse. Ainsi ces suivis de patients peuvent être liés à des phases de la maladie qui ne se recouvrent pas (cf Figure 3), et ce non-alignement temporel entre les patients complique fortement leur comparaison, par exemple pour l'extraction de sous-types.

La modélisation de progression de maladies, et le réalignement temporel

Pour faire face à ces particularités, un pan des statistiques pour la medecine a émergé, et se concentre sur les dynamiques temporelles de maladies, appelé "disease progression modeling". En 2010, (Clifford R. Jack, Knopman, et al. 2010) a proposé un premier schéma conceptuel de la progression de biomarqueurs dans la maladie d'Alzheimer, prenant la forme de courbes sigmoïdes, depuis un état normal vers un état anormal. Ces courbes sigmoïdes arrivent en cascade, sous-tendant un ordonnancement entre les marqueurs, et reflétant les mecanismes pathologiques sous-jacents (par exemple perte de neurones dans

l'hippocampe, suivi par des troubles cognitifs).

Depuis lors, des modèles plus complexes ont été développés pour estimer statistiquement la progression de maladies. Les "Event Based Model" (Fonteijs et al. 2011) sont des modèles discrets de progression qui utilisent des bases de données cross-sectionnelles. Ils estiment l'ordonnancement d'événements définis à l'avance, et ainsi construisent un scénario normatif de progression.

Aussi, des modèles de progression continus ont été développés en utilisant des cohortes longitudinales. Ces modèles doivent gérer le non-alignement temporel entre les patients, dont l'âge à une visite donnée est peu informatif sur le stade de la maladie dans le cas de maladies neurodégénératives, comme décrit sur la Figure 3. Introduit dans (Jedynak et al. 2012), et utilisé dans nombre de modèles de progression de maladies (Michael C. Donohue et al. 2014; Guerrero et al. 2016), les modèles continus gèrent cette difficulté en reparamétrisant l'âge de façon affine : $\psi : t \mapsto \alpha_i t + \tau_i$, qui réaligne les patients au mieux en fonction de leur stade de progression ψ , en terme de valeurs de biomarqueurs. Cette reparamétrisation temporelle est pratique car elle utilise peu de paramètres, et ψ devient alors un stade de maladie "abstrait". De plus, comme α_i et τ_i sont typiquement considérés comme des effets aléatoires dans un modèle à effets mixte, la trajectoire définie seulement par les effets fixes peut être vue comme une trajectoire long-terme de la population générale, abstraite, informée par chaque suivi individuel de patients.

Schiratti, Allasonniere, et al. 2015b proposent un modèle spatio-temporel, générique dans le sens où il fait l'hypothèse que les observations appartiennent à une variété Riemannienne, ce qui inclut les formes (maillage), l'imagerie du cerveau (Bone, Colliot, and Durrleman 2018), ou simplement des marqueurs scalaires. Le modèle suppose l'existence d'une trajectoire moyenne de progression, modélisée par une géodésique sur cette variété Riemannienne. Une géodésique peut être interprétée comme la trajectoire du plus court chemin dans un espace courbe (cf Figure 4). Les trajectoires individuelles sont modélisées comme des "parallèles" de cette géodésique, sur la variété, et progressant en terme de "stade de la maladie". Ce modèle est donc un modèle à effet mixtes, et utilise la reparamétrisation affine des âges.

Surtout, le modèle offre un cadre théorique pour construire un système de coordonnées spatiotemporelles, que l'on peut visualiser intuitivement comme un système de coordonnées tubulaire. La variabilité temporelle indique à quelle position la visite se place sur l'axe de la trajectoire moyenne, i.e. en terme de stade de progression. D'un autre côté les variations "spatiales" encodent pour la spécificité du patient, et sont orthogonales à la trajectoire moyenne. En d'autres termes, la variation spatiale est la variation par rapport à la trajectoire moyenne, après avoir effectué le réalignement temporel. A noter que les paramètres des variations temporelles et spatiales sont tous considérés comme des effets aléatoires du modèle à effet mixte. Pour des données scalaires, les variations spatiales peuvent être calculées analytiquement comme des transformations géométriques de la trajectoire moyenne.

Ainsi, bien que tous les modèles soient biaisés de façon inhérente par leurs hypothèses sous-jacentes, leur design ou encore leur méthode d'estimation, leur multiplicité permet d'obtenir des résultats avec confiance, quand ces résultats sont reproduits indépendamment. L'initiative Europond propose de regrouper et partager ces modèles à la communauté

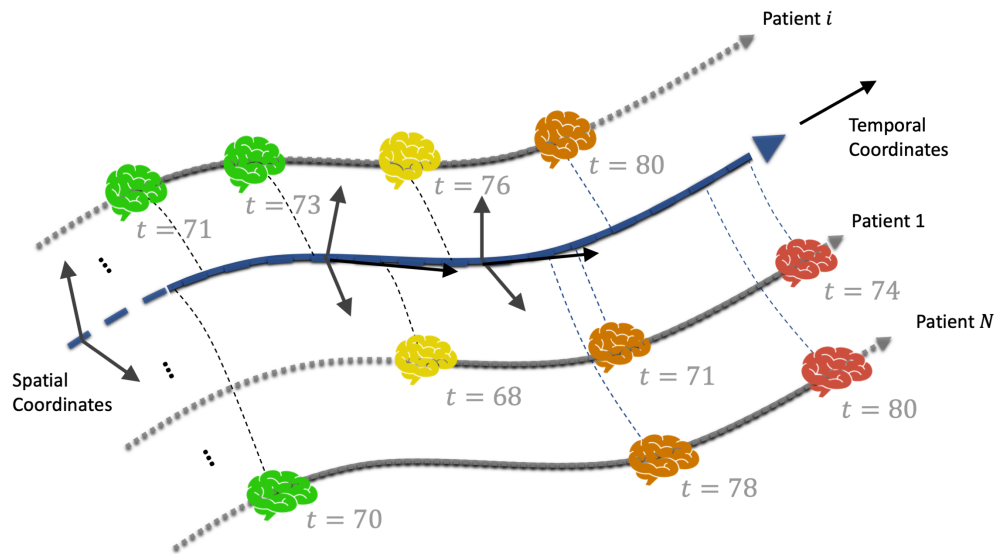


Figure 2 – Représentation de l’effet du réaligement temporel des sujets selon leur stade de la maladie, par rapport à une trajectoire moyenne abstraite (en bleu). Le modèle longitudinal spatiotemporel permet de désenchevêtrer la variabilité temporelle, qui modélise les variations en terme de stade d’avancement dans la maladie, et la variabilité spatiale, qui modélise les différences inter sujets. Ici, la variabilité spatiale est représentée avec des hauteurs et largeurs de cerveaux spécifiques à chaque sujet. Nous observons aussi que l’âge n’est pas informatif du stade d’avancement dans la maladie.

(www.europond.eu). De plus, l’organisation de data challenges promeut une comparaison des méthodes, comme sur le challenge TadpoleRazvan V Marinescu et al. 2018, ou avec le Data Challenge de PPMI.

Objectifs de la thèse

Objectifs

Dans cette thèse, nous visons à construire des modèles de progression de maladies pour mieux capturer l’hétérogénéité de la maladie de Parkinson, en terme de schémas individuels de progression. Bien que cela ne réponde pas directement à l’extraction de sous-types et à l’identification de biomarqueurs précoces, la construction de modèles prenant en compte les spécificités des maladies neurodégénératives est une première étape. Nous faisons l’hypothèse d’un découplage entre variabilité temporelle et variabilité spatiale en construisant des modèles avec coordonnées spatio-temporelles, comme dans (Schiratti, Allassonniere, et al. 2015b). Ces représentations temporelles et spatiales apprises pourront ensuite être utilisées dans des applications cliniques comme la prédiction ou le clustering. Dans cette thèse, nous divisons cet objectif en 3 étapes :

- Étendre le modèle longitudinal spatiotemporel pour pouvoir l’appliquer sur la maladie

de Parkinson.

- Appliquer le modèle à la maladie de Parkinson pour répondre à des hypothèses cliniques. En pratique, nous visons à mieux reconstruire les données, et à analyser les corrélations entre nos paramètres individuels de progression, interprétables, et des cofacteurs d'intérêt qui représentent la physiologie des patients.
- Étendre la flexibilité et les potentielles applications du modèle longitudinal spatio-temporel à des données de hautes dimensions, multimodales, ou non labellisées en utilisant le cadre de l'autoencoder, et trouver un compromis acceptable entre flexibilité et interprétabilité.

Contributions

Nous choisissons comme point de départ le modèle longitudinal spatiotemporel de Schiratti, Allasonniere, et al. 2015b, et tout au long de cette thèse nous nous appuyons sur la notion de systèmes de coordonnées spatio-temporelles pour les données longitudinales (cf Figure 4).

Motivés par une application finale sur les marqueurs de la maladie de parkinson, nous étendons dans un premier travail le modèle spatiotemporel de Schiratti, Allasonniere, et al. 2015b. L'instanciation de ce modèle pour données scalaires, aussi appelé "modèle de propagation", a été construit dans l'optique de modéliser la progression de scores cognitifs dans la maladie d'Alzheimer et repose sur des hypothèses fortes. Chaque évolution de biomarqueur est modélisée par une courbe sigmoïde, schéma classique de progression de biomarqueur de maladie neurodégénérative depuis l'hypothèse de Jack Jr et al. 2010, avec des valeurs progressant de 0 (normal) à 1 (anormal). Deuxièmement, le modèle de propagation fait l'hypothèse que les biomarqueurs évoluent à la même vitesse : de cette façon chaque progression de biomarqueur peut être paramétrisée seulement par un délai par rapport à un biomarqueur de référence. C'est un moyen simple d'ordonner les biomarqueurs, selon leur "onset". Cependant, dans le cas de biomarqueurs de différentes nature, comme le DATScan, qui mesure le signal dopaminergique dans le striatum, et les scores cliniques moteurs, cette hypothèse n'est pas valide. Pour relâcher cette rigidité, nous proposons de considérer plusieurs types de progression par marqueurs dans une modélisation jointe de la progression. De plus, nous gérons les valeurs manquantes en utilisant l'aspect génératif du modèle, en faisant l'hypothèse qu'elles sont manquantes de manière complètement aléatoire (MMCA), i.e. que leur absence n'est pas informative de leur valeur.

Cela permet une première application clinique sur la maladie de Parkinson. Dans une première étude, nous modélisons la progression des anormalités motrices de patients prodromaux, et estimons une timeline de progression en définissant des seuils d'anormalités pour chaque biomarqueur. Nous comparons nos résultats avec ceux de R. Postuma, Lang, et al. 2012, et confirmons que les altérations du visage et de la parole sont les plus précoces du MDS-UPDRS III, alors que la rigidité est identifiée comme un marqueur tardif. Dans une deuxième étude clinique, nous réalisons une analyse plus large pour quantifier l'hétérogénéité de la maladie de Parkinson, et incluons un ensemble exhaustif de marqueurs pour représenter les différentes facettes de la progression de la maladie : moteur,

non-moteur, cognition et imagerie. Nous étudions l'hétérogénéité en terme de variabilité spatiale et temporelle par rapport à la trajectoire moyenne, et identifions les troubles du sommeil et de la cognition comme les symptômes avec le plus de variance inter-patients. Les troubles du sommeil surtout sont intéressants car ils peuvent arriver très tôt dans la progression de la maladie. Par conséquent, nous étudions les associations entre trouble du sommeil et schémas de progression individuels, et mettons en évidence des différences significatives entre patients avec et sans troubles du sommeil, principalement en terme de symptômes non-moteurs.

Dans une deuxième partie de notre travail, nous basculons de la géométrie Riemannienne vers l'apprentissage profond, plus flexible pour modéliser la progression de maladie. Nous apprenons toujours des modèles de progression avec coordonnées spatiotemporelles, ici dans l'espace latent d'un autoencoder à la manière de (Maxime Louis, Couronné, et al. 2019). En effet, construire un cadre théorique pour l'analyse longitudinale de données en grande dimension est une tâche ardue. L'utilisation de l'autoencoder permet d'alléger cette difficulté, au prix des preuves de convergences, mais aussi et surtout de l'orthogonalité entre variations temporelles et spatiales. Nous proposons tout d'abord de gérer les données multimodales en agrégeant les représentations apprises de chaque modalité. Cependant, la modélisation de la progression jointe en grandes dimensions est complexe, et ainsi l'approche de (G. Lee et al. 2019) se concentre sur des marqueurs scalaires, même si elle repose sur de l'apprentissage profond. Enfin, nous étendons le modèle au cas auto-supervisé, en utilisant seulement l'ordonnancement entre les visites donnée par le cadre longitudinal, et s'abstrayant de devoir exhiber une fonction liant âge et stade de la maladie, comme la reparamétrisation affine mentionnée plus haut. De plus, nous encourageons le désenchevêtrement dans l'espace latent entre variabilité temporelle et spatiale, en apprenant une unique représentation spatiale par patient, sur un sous-ensemble aléatoire de ses visites. L'idée d'ordonner les visites est proche de LLSL (Zhao, Z. Liu, et al. 2020), qui contraignent l'espace latent en forçant les trajectoires des patients à être parallèles via une fonction de coût sur les angles. Cependant, ils se concentrent sur la variabilité temporelle, alors que nous cherchons aussi à donner du sens à la variabilité spatiale apprise.

Organisation du manuscrit

Partie I: Contexte

Le **chapitre 1** présente de façon plus poussée la maladie de parkinson, et les défis actuels, comme l'extraction de sous-types ainsi que l'étude de la phase prodromale de la maladie, qui sont considérés comme des étapes cruciales vers une meilleure compréhension de la maladie.

Le **chapitre 2** introduit le modèle longitudinal spatiotemporel, avec des notions sous-jacentes de géométrie riemannienne, son instanciation avec des marqueurs scalaires, ainsi que la procédure d'estimation

Partie II Modèles de progression de la maladie de Parkinson avec des données en faible dimension et non structurées

Le **chapitre 3** présente notre contribution sur le modèle longitudinal spatiotemporel de (Schiratti, Allassonniere, et al. 2015b). In fine, nous visons à appliquer le modèle aux données de PPMI, à la fois sur les marqueurs cliniques et d'imagerie. Nous proposons de relacher les hypothèse du modèle de propagation de Schiratti, Allassonniere, et al. 2015b, qui suppose que les formes et vitesses des courbes sont partagées pour tous les biomarqueurs. Ensuite nous proposons de gérer les valeurs manquantes en nous reposant sur l'aspect génératif du modèle.

Le **chapitre 4** présente les applications cliniques du modèle. Nous passons en revue les outils mis à disposition par le modèle longitudinal spatiotemporel, en utilisant les trajectoires apprises au niveau individuel, d'un sous groupe de patient, ou de l'ensemble de la population. Ensuite nous présentons une première application de ces méthodes dans le but d'étudier la progression des anormalités moteurs des patients prodromaux, i.e. les patients avec un risque élevé de développer la maladie de Parkinson. Dans ce travail nous construisons une timeline d'apparition des symptômes. Pour finir, nous étudions en détail l'hétérogénéité en terme de schémas de progression dans la maladie de Parkinson, avec un accent sur les troubles du sommeil, que nous trouvons être corrélés avec les symptômes non-moteurs.

Partie III Autoencoder longitudinal

Le **chapitre 5** introduit l'autoencoder longitudinal. Dans ce cadre, nous tirons parti de l'idée de l'autoencoder pour apprendre un lien non-linéaire entre l'espace des données et un espace de faible dimension, supposé euclidien, contraint de jouer le jeu d'un système de coordonnées spatiotemporel. L'autoencoder permet de s'abstraire de méthode de réduction de dimensions ad-hoc, pour se concentrer sur le modèle graphique dans l'espace latent, que nous construisons à la manière de Schiratti, Allassonniere, et al. 2015b. Cette flexibilité permet de gérer de façon agnostique des images ou des scores cliniques, mais nous perdons aussi la notion d'exp-parallélisation qui permettait de s'assurer de l'orthogonalité entre variabilité spatiale et temporelle.

Le **chapitre 6** explore le cas de données multimodales dans le cadre de l'autoencoder longitudinal, en agrégeant les représentation apprises sur chaque modalités, comme des scores cliniques ou encode des modalités d'imagerie.

Le **chapitre 7** propose de modifier l'architecture proposé au chapitre 5, afin de relacher des contraintes sur le modèle graphique de l'espace latent. Nous proposons de s'appuyer sur le cadre longitudinal pour apprendre la variabilité temporelle seulement par l'intermédiaire de l'ordonnancement entre les visites, et plus des âges des patients à leurs visites.

Partie IV Développement logiciel

Le **chapitre 8** présente notre contribution logicielle sous la forme d'une librairie Python nommée Leaspy, qui implémente le modèle longitudinal spatiotemporel utilisé au chapitre 4.

Introduction

Outline of the current chapter

Preamble	12
Motivation	12
Parkinson's Disease	12
Knowledge and challenges in Parkinson's Disease	12
Top-down action-levers	13
Background	14
Longitudinal Cohorts	14
Challenges in Parkinson's Disease statistical analysis	14
Disease progression modeling and time realignment	15
Thesis Objective	17
Goals	17
Contributions	17
Manuscript Overview	18
Part I: Background	18
Part II: Models of PD progression with low dimensional un-structured data	19
Part III: Longitudinal Autoencoders	19
Part IV: Software Development: Leaspy	19

Preamble

Uncovering and accurately quantifying the heterogeneity in Parkinson's Disease progression patterns would greatly benefit our current understanding of the disease, by providing insights on the underlying disease pathways. Indeed, since its characterization in 1817, Parkinson's Disease is still far from fully understood, and as a consequence treatments have not significantly improved since the first use of Levodopa, a dopamine precursor, in the 1950s.

Recently, the field of medicine has witnessed the enormous potential of new inventions derived from breakthroughs in physics, such as Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET) and genome sequencing techniques. These tools reach new frontiers in characterization of anatomical and biological processes, to the extent that some diseases may be diagnosed and monitored precisely. But the complexity of neurodegenerative diseases is so high that entire research teams are targeting sub-processes inside a bigger picture, which is at the moment our only hope for successful therapeutics.

Machine Learning (ML) offers a way to statistically and algorithmically bring new insights to these diseases from patient's wide spectrum of measurements (Jiang et al. 2017). These algorithms are already performing well in diagnosis and prognosis of well defined tasks, such as tissue segmentation (Havaei et al. 2017; Menze et al. 2014; Milletari, Navab, and Ahmadi 2016; Pereira et al. 2016), or patient classification. They have great potential to build up knowledge in complex diseases.

Motivation

Parkinson's Disease

Parkinson's Disease (PD) is a striking example of such complex disease, with slowly progressing motor impairments that considerably alter the patient's quality of life. The disease is usually diagnosed around 60 years old but also affects younger people with 20 % of patient being diagnosed before 50 years old (*Forme précoce de la maladie de Parkinson* n.d.). Around 6.3 million people worldwide live with Parkinson's Disease (Rocca 2018), and the impact of the disease in terms of quality of life, life expectancy and social and monetarial costs (Findley 2007) is likely to rise with the ageing of the population.

Knowledge and challenges in Parkinson's Disease

Significant improvements in our understanding of the disease have been made these last decades. It is now acknowledged that the disease begins long before the diagnosis reporting motor symptoms onsets, in a mostly asymptomatic "prodromal" phase (Mahlknecht, Seppi, and Poewe 2015; Ronald B Postuma and Berg 2016). This prodromal phase begins up to 10-20 years before diagnosis (Kalia and Anthony E Lang 2016; Ronald B Postuma and Berg 2016; Savica, Rocca, and J Eric Ahlskog 2010), with first increasing damage autonomic nervous system, possibly linked with sleep and constipation symptoms, followed by neurodegeneration in the substantia nigra. Damage in the substantia nigra is

considered the direct cause for PD motor symptoms, and can be monitored via a brain imaging. However, measurements reveal that at diagnosis, patient have already lost 60% of their dopaminergic neurons (P. N. Lee et al. 2007). This highlights the need for early intervention.

Then, Parkinson's Disease is very heterogenous by nature, to the point that it is now considered as an umbrella term under which different and overlapping clinical and pathological subtypes fit. A scheme of progression has been proposed (Braak et al. 2003) to explain the topology of damage in the brain from a growing accumulation of proteins, but many patients progression differ from this model. In order to classify patients in more consistent disease groups, many methods have been proposed, such as classifying by motor phenotypes (Joseph Jankovic et al. 1990; Zetuský, Joseph Jankovic, and Pirozzolo 1985), prevalence of sleep symptoms (Berg, Borghammer, et al. 2021; Horsager, Andersen, Knudsen, Skjærbæk, et al. 2020a; Yoritaka et al. 2009), or genetics, but no convention has been adopted to date.

Ultimately, Parkinson's Disease research aims to:

- Develop a cure, to be administered as early as possible.
- In the absence of a cure, best administer the existing treatments, taking into account their secondary effects.
- Predict the arrival of specific symptoms altering patient's quality of life, such as dementia, dyskinesia or falls, in order to best adapt care.

Top-down action-levers

In order to accelerate Parkinson's Disease research, top down approaches propose to refine incrementally our knowledge on Parkinson's Disease from hypothesis testing on the data. These new insights add up, and with enough precision might feed the reflection at the disease sub-processes levels. Presently, levers in statistical analysis of Parkinson's Disease include:

- **Identify early biomarkers** in order to provide insights on the pre-diagnostic phase of the disease, and follow patient from an earlier stage. This would also allow earlier intervention, (Murman 2012) and experiment new disease modifying drugs, which effectiveness might increase with the precocity of their administration.
- **Subtype patients** into consistent subgroups. A consistent subgroup could account for a specific pathway of disease progression, and an associated specific underlying processes. For example, patients with the GBA gene, also present in Gaucher Disease for which a treatment exists, could potentially share some of the disease mechanisms. This association approach could offer new insights to develop treatments (Riboldi and Di Fonzo 2019).

Improvements on these subjects could lay new bricks in the foundations of our current knowledge of the disease.

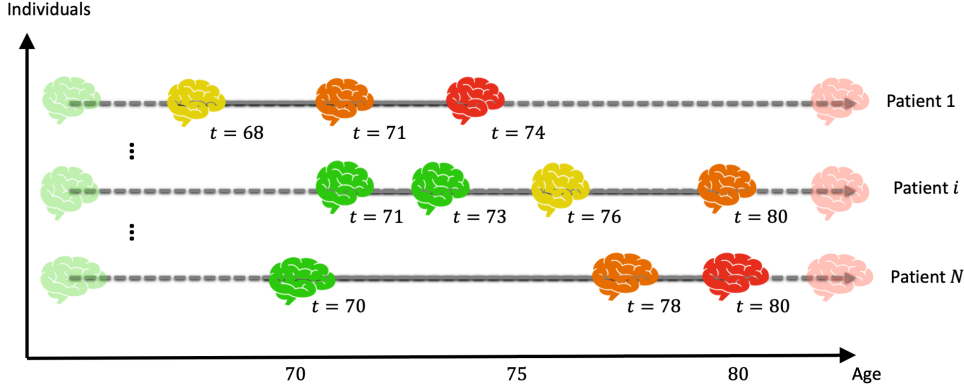


Figure 3 – Illustration of a longitudinal dataset with N patients, seen at different visits. Color of the brain accounts for disease stage. Different height and width of the brain account for possible inter-patient variability. A wide line indicates the time-span of each patient follow-up, while dotted lines indicate unseen patient data. Transparent brains at extremities relate to our monotonicity assumption, that is to say patient with the disease were healthy at some point, and will eventually reach abnormality in all biomarkers.

Background

Longitudinal Cohorts

To perform these tasks, we need to acquire data, as many as possible, and of the highest quality possible. To this end, in the last decades, patient data were gathered, and aggregated in cohorts of patients, to allow for statistical analysis. Well known and publicly shared cohorts include Alzheimer's Disease Neuroimaging Initiative (ADNI) (Petersen et al. 2010) for Alzheimer's Disease, and Parkinson's Progression Markers Initiative (PPMI) (Marek, Jennings, et al. 2011) for Parkinson's Disease.

These two cohorts are "longitudinal", in the sense that they follow patients in time, performing measurements at multiple visits for a same patient (see Figure 3). This longitudinal follow-up allows to see patient's dynamics. We can then appreciate the repeated measurements of patients as "trajectories" in a space of measurements, and develop statistical methods accordingly.

Hopefully, successful building and analysis of longitudinal cohorts may lead to identifying novel biomarker of progression or detect finer inclusion criteria for patients (Berg, Ronald B Postuma, et al. 2015; Heinzl et al. 2019) for future cohorts, and possibly targeted clinical trials.

Challenges in Parkinson's Disease statistical analysis

With these longitudinal data, identification of early biomarkers and subtyping of patients is still not straightforward. Indeed, for subtyping, most clinical studies perform a clustering task at baseline (Eisinger et al. 2017; Fereshtehnejad, Romenets, J. B. Anang, et al. 2015; Fereshtehnejad, Zeighami, et al. 2017) and in a second step assess the progression of each group. By design, these studies cannot capture the differences in dynamics. Regarding the

assessment of biomarkers precocity, an intuitive method consist in regressing the markers back in time. However, no such reference time exist in neurodegenerative diseases, and studies usually choose age at diagnosis (Fereshtehnejad, Yao, et al. 2019; R. Postuma, Lang, et al. 2012), which still might hide significant differences between patients in terms of disease stage, and speed of progression.

Taking a step back, these are more general challenges that arise in the context of longitudinal analysis of medical data, mainly because acquiring data is a hard task in the first place, and because of the disease inherent heterogeneity. Indeed, as longitudinal datasets follow patients for a limited time, they include both a left-censorship, with no data before inclusion in the cohort, typically at diagnosis, and a right-censorship. Then, the limited time span of patients follow-up compared to the overall disease duration complicates the analysis. In addition, these patient snapshots may relate to different non-overlapping phases of the disease (see Figure 3). This temporal un-alignment between patients complicates the inter-patient comparison, for e.g. subtyping.

Disease progression modeling and time realignment

To cope with these specific issues, a new sub-field of statistics for medicine has emerged, focusing on the temporal dynamics of diseases, namely disease progression modeling. In 2010, Clifford R. Jack, Knopman, et al. 2010 proposed a first conceptual plot of prodromal biomarker progression in Alzheimer’s Disease, on the form of sigmoids from a normal state to abnormality. Sigmoid-like progression occur in cascade, with an ordering between the biomarkers, reflecting the underlying disease process (e.g. loss of neurons followed by cognitive symptoms).

Since then, more complex methods have been developed to statistically estimate disease progression. A first discrete model is the Event Based Models (Fonteyjn et al. 2011), which leverages cross-sectional datasets to estimate the sequence of predefined events, and build a normative timeline of evolution.

Then, continuous models of progression have been developed by leveraging longitudinal datasets. They have to cope with the temporal disalignment between patients, whose age at a given visit tells few information on the disease stage in the case of neurodegenerative diseases, as shown in Figure 3. Introduced in Jedynak et al. 2012 and used in many disease progression models (Michael C. Donohue et al. 2014; Guerrero et al. 2016), continuous models typically cope with this issue by learning for each patient an affine reparameterization of age: $\psi : t \mapsto \alpha_i t + \tau_i$ that best realign subjects according to their disease severity ψ , in terms of biomarker measurements. This time reparameterization is practical as it includes few parameters, and ψ becomes then an abstract disease stage. In addition, as α_i and τ_i are typically considered as random effects in a mixed effect model, the trajectory defined solely by the fixed effect can be viewed as an abstract, long-term average trajectory of the population informed by each patient snapshots.

Schiratti, Allasonniere, et al. 2015b propose a longitudinal spatio-temporal model, generic in the sense that it assumes that data on a smooth Riemannian Manifold, encompassing meshes, brain images (Bone, Colliot, and Durrleman 2018) or simply scalar values. The model posits the existence of a population average trajectory of progression

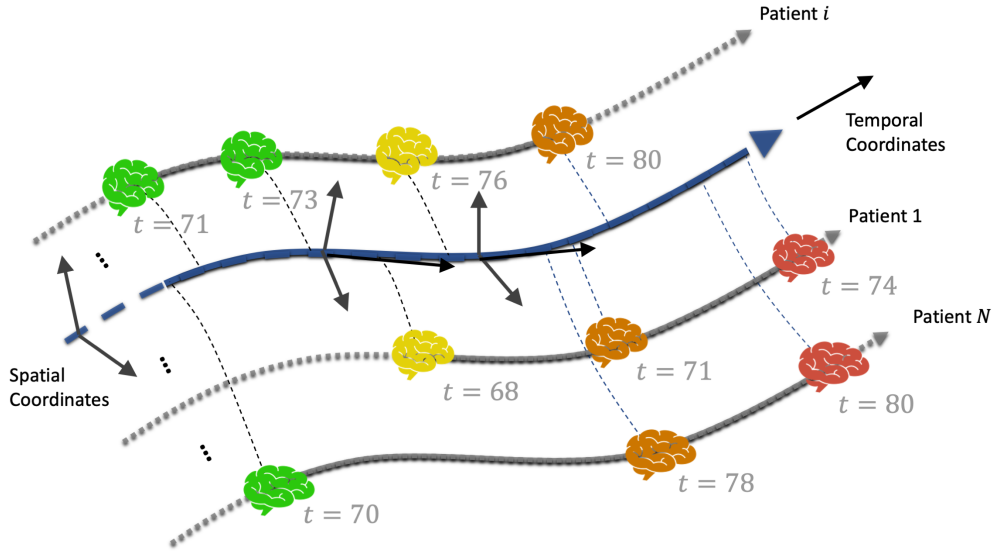


Figure 4 – Effect of temporal re-alignment of subjects with regard to their disease stage, according to an abstract average trajectory (in blue). The longitudinal spatiotemporal model allows to disentangle temporal variability, accounting for variations in terms of disease stage, and spatial variability, which is subject-dependent. In this case, spatial variability is denoted with a subject-specific height and width of the brain. Note that age is not informative of disease stage.

as a geodesic of the manifold, which may be interpreted as a shortest path trajectory in a curved space (see Figure 4). Individual trajectories are modeled as "parallels" on the manifold around this population geodesic, progressing over an abstract disease stage. The model is a mixed-effect model, and posits disease stage as an affine reparameterization of age. Most importantly, it offers a theoretical framework to build a system of tubular spatio-temporal coordinates. Temporal variation account for position on the average trajectory axis, i.e. in terms of disease stage. On the other hand, "spatial" variations account for patient-specific variations which are orthogonal to the average trajectory. That is to say variations from the average trajectory after having performed the affine temporal re-alignment. Note that both patient temporal parameters and spatial parameters are considered as random effects in the mixed effect model. In practice, spatial variations are available in closed form as geometrical transformation of the average trajectory.

Although all models are inherently biased by their underlying assumptions and design, the multiplicity of approaches yields confidence when results are obtained in independent models. Europond initiative propose to regroup and promote the sharing of disease progression models to the community (www.europond.eu). In addition, data challenges promote sane benchmarking of methods, such as in the Tadpole challenge Razvan V Marinescu et al. 2018, or PPMI Data challenges.

Thesis Objective

Goals

In this thesis, we aim at building disease progression models to better capture the heterogeneity of Parkinson’s Disease in terms of patient’s patterns of progression. While this does not address directly the subtyping and early biomarker identification levers, we believe that it is a first step toward these tasks. We rely on a conceptual decoupling between temporal variability and spatial variability by building models with spatio-temporal coordinates as in (Schiratti, Allasonniere, et al. 2015b). Ultimately, we aim at learning useful temporal and spatial representations that may then be used for downstream tasks in PD such as prediction or clustering. On this thesis, we divided this task in three goals:

- Extend the longitudinal spatiotemporal model for relevant real data analysis
- Provide clinical insights using the model. In practice, we aimed to provide better fit to the data, as well as inspect correlation between interpretable individual parameters and cofactors of interest representing the physiology of patients.
- Extend the flexibility and application of the longitudinal spatiotemporal model to high dimensional, multimodal or unlabeled data, leveraging autoencoders, while trying to find an acceptable trade-off between flexibility and loss of interpretability.

Contributions

We choose as starting point the longitudinal spatiotemporal model of Schiratti, Allasonniere, et al. 2015c, and will keep thorough this thesis the notion of spatiotemporal coordinate (see Figure 4).

Driven by an application to Parkinson’s Disease markers, we extend in a first work the spatiotemporal model of Schiratti, Allasonniere, et al. 2015c. Its proposed propagation model for scalar data designed to model Alzheimer cognitive score progression makes strong assumptions. First, each biomarker progression follows a sigmoid curve, which is the standard shape of biomarker progression in neurodegenerative disease since the hypothesis of Jack Jr et al. 2010, with values ranging from 0 (normal) to 1 (anormal). Second, the propagation model assumes biomarkers evolve at the same pace: in this way, each biomarker progression can be parameterized only via a shift of a reference biomarker. This is a convenient way to order biomarkers according to their “onset”. However, in the case of markers of different nature, such as DATScan SBR measures the dopaminergic signal in the striatum, and motor clinical scores, this assumption seems overly simplistic. To relax this rigidity, we propose to consider different types of progression per marker in their joint longitudinal modeling. In addition, we handle missing values by leveraging the generative nature of the model, assuming that they are Missing Completely At Random (MCAR), i.e. that their missingness is not informative of the marker values.

This paves the way for a clinical application to Parkinson’s Disease. In a first study, we model the progression of motor abnormalities in prodromal PD patients, and extract a timeline of progression by defining thresholds of abnormalities for each biomarker. We

compare our results with those of R. Postuma, Lang, et al. 2012, and confirm face and voice alterations as early markers, while rigidity was identified as a late maker. In a second clinical study, we perform a broader analysis to assess the heterogeneity of Parkinson’s Disease, and include an exhaustive set of biomarkers chosen to represent multiple facets of disease progression: motor, non-motor, cognition, and imaging. We study the heterogeneity in terms of spatial and temporal variations from the population trajectory, and identify sleep and cognition as the symptoms with most variability between patients. Sleep especially is interesting as it may appear early in the disease. Consequently we assess the associations between sleep symptoms and individual’s patterns of progression, and found that sleep symptoms discriminate patient profile of progression, with a significant difference in non-motor symptoms.

In a second part of our work, we shift to the flexible autoencoder framework for disease progression modeling, and learn models with spatiotemporal set of a coordinate in the latent space of an autoencoder as in (Maxime Louis, Couronné, et al. 2019). Indeed, defining the right theoretical framework to model longitudinal high-dimensional data is a hard task. The autoencoder alleviates these issues at the cost of proofs of convergence and theoretical orthogonality between spatial and temporal variations. We first propose to handle multimodality by aggregating the representations of each modality. However the joint modeling of multiple modalities with possibly high dimension is especially hard and for example the Deep Learning approach of (G. Lee et al. 2019) only use scalar markers. Finally, We extend the model to the self-supervised case, leveraging only the visit ordering given by the longitudinal setup, breaking free from an explicit mapping between age and disease stage, such as the above-mentioned affine reparameterization. Furthermore, we encourage disentanglement in the latent space between temporal and spatial variability, by learning a single spatial representation per patient, from a random subset of its visits. The idea of ranking visits is close to LSSL (Zhao, Z. Liu, et al. 2020), which constrains the latent space by forcing patient trajectories to be parallel via a cosine loss. However, they only analyze the temporal variability, while we also seek to make sense of the spatial variability.

Manuscript Overview

Part I: Background

Chapter 1 offers a deeper description of Parkinson’s Disease, and presents current challenges, such as disease subtyping and prodromal phase study, which are considered crucial steps toward a better understanding of the disease.

Chapter 2 introduces the longitudinal spatiotemporal model, with underlying notions of Riemannian Geometry, its instantiations for scalar data, and the estimation procedure.

Part II: Models of PD progression with low dimensional un-structured data

Chapter 3 presents our contribution on the longitudinal spatiotemporal model of Schiratti, Allasonniere, et al. 2015c. We aim at applying the model to PPMI Data on both clinical and imaging measurements. We propose to relax modeling assumptions of the propagation model of Schiratti, Allasonniere, et al. 2015c of shared speeds and curve shapes for all biomarker. Then we propose to handle missing values by leveraging the generative nature of the model.

Chapter 4 presents the clinical applications of the model. We first review the tools offered by the longitudinal spatiotemporal model, using individual, subgroups and population trajectories. Then we present a first application of these methods to study the progression of motor abnormalities in prodromal patients, i.e. patients at high risk of developing Parkinson's Disease, where we build a timeline of motor symptoms apparition. lastly, we investigate in depth the heterogeneity in terms of patterns of progression in PD, with a focus on sleep symptoms, that we find strongly associated with non-motor symptoms.

Part III: Longitudinal Autoencoders

Chapter 5 introduces the longitudinal autoencoder. In this framework, we propose to leverage autoencoders to learn a non-linear mapping between the measurement space and a low dimensional latent space, assumed Euclidean, constrained to act as spatiotemporal coordinates of disease progression. Autoencoders alleviate the need for ad-hoc dimensionality reduction methods, and allow to focus on the graphical model in the latent space, that we set similarly to the one of Schiratti, Allasonniere, et al. 2015b. This flexibility allows to cope similarly with images or clinical scores but comes at the cost of theoretical properties such as the loss of exp-parallelization that ensured orthogonality between temporal and spatial variabilities.

Chapter 6 explores the case of multimodality in the longitudinal autoencoder framework, by aggregating representations learned on different modalities, such as clinical scores and imaging modalities.

Chapter 7 proposes to modify the architecture proposed in chapter 5 to relax constraints set on the graphical model in the latent space. We propose to leverage the longitudinal setup to rely solely on the visit ordering to learn the disease stage.

Part IV: Software Development: Leaspy

Chapter 8 Presents our software contribution in the form of the Python library Leaspy, which implements the longitudinal spatiotemporal model.

Part I

Background

Parkinson’s Disease

Outline of the current chapter

1.1 Preamble	24
1.2 Clinical Overview of Parkinson’s Disease	25
1.2.1 History	25
1.2.2 Diagnostic	25
1.2.3 Symptoms and Complications	25
1.2.4 Pathophysiology	26
1.2.5 Treatments	27
1.3 Materials	27
1.3.1 Longitudinal Cohorts	27
1.3.2 Markers	29
1.3.3 Biases in Parkinson’s Disease longitudinal studies	30
1.4 Goals in Parkinson’s Disease Longitudinal Analysis	31
1.4.1 Early Biomarkers of progression	31
1.4.2 Subtyping Parkinson’s Disease	35
1.4.3 Other research focus	38

In this chapter, we describe in depth the current challenges and results in Parkinson’s Disease longitudinal analysis. These insights aim at highlighting the interdisciplinarity of this thesis, and the close collaboration with clinicians at the Brain and Spine Institute.

1.1 Preamble

As of 2017, the prevalence of Parkinson's Disease is of 6.3 million patients worldwide, and 1.2 million patients in Europe (Rocca 2018). In France, 150,000 people with PD are treated in with approximately 25,000 new cases occurring each year (France n.d.). The risk of death is twice as high as in non-diseased people of comparable age and sex (E. D. Louis et al. 1997), with a significant impact on young people and women (Schrag et al. 2003). The average age of diagnosis in France is 58 years (France n.d.). Given the ageing of the population, the incidence of the disease is rising, leading to an increase in costs (Rocca 2018).

Presently, improvement in patient care would greatly benefit from (Poewe et al. 2017):

- Diagnosing the disease early enough (60% of neurons are already lost at diagnosis)
- Developing preventive treatment before the onset of symptoms
- Developing better treatments of deficits
- Managing patients at the different stages of their disease

which however requires better understanding of the disease mechanisms. Various approaches aim at tackling these challenges:

- Identifying early prognostic and predictive markers, possibly combining different modalities
- Identifying genetic and environmental risk factors
- Using digital applications to monitor patients on a daily basis
- extract consistent subtypes of patients

As the physiopathology of Parkinson's Disease remains for the most part not understood, especially in its early stages, these approaches are more and more data-based. They typically rely on building cohorts of patients followed in time from their disease onset, and if possible even before, to extract statistical signal.

More specifically, we describe in the following two tasks of interest. First, as PD is very heterogeneous (Berg, Borghammer, et al. 2021; Zetuský, Joseph Jankovic, and Pirozzolo 1985), subtyping patients in more consistent categories could allow for more personalized prognosis and treatment (Fereshtehnejad, Zeighami, et al. 2017; Lewis et al. 2005; P. Liu et al. 2011). Second, identifying biomarkers that announce and describe PD in its earliest stage (Berg, Ronald B Postuma, et al. 2015; Heinzl et al. 2019), such as predicting on which subtype they belong (Berg, Borghammer, et al. 2021), would offer more opportunities to develop and apply treatments.

1.2 Clinical Overview of Parkinson's Disease

1.2.1 History

Parkinson's Disease has been first described in 1817 by James Parkinson, which at the time was known as the "Shaking Palsy" (Parkinson 2002). In the end of the 19th Century, Charcot refined the description of Parkinson's Disease, categorizing the motor symptoms of PD into rigidity, weakness and bradykinesia. He prescribed anticholinergics, which inhibit the parasympathetic nervous system, and relieved involuntary movement of muscles.

Pathophysiology of Parkinson's Disease was better understood with the discovering of Lewy Bodies, abnormal microscopic aggregates of protein, by Friedrich Lewy in 1912 (Holdorff, Rodrigues e Silva, and Dodel 2013). This helped identifying Substantia Nigra as the main structure affected in the brain in Parkinson's Disease in 1919 (Goedert et al. 2013).

In the 1950s, research work of Arvid Carlsson allowed better understanding of the underlying biochemical modifications in the brain, by describing the neurotransmitter effect of dopamine in Parkinson's Disease. Parkinsonian symptoms would then be a consequence of loss of dopaminergic neurons in the Substantia Nigra. He showed that Levodopa (L-dopa), a dopamine precursor capable of crossing the blood-brain barrier, could reverse the effects of parkinsonian symptoms, which led to a Nobel prize in 2000. Since 1967, L-Dopa is used in clinical practice, and led to a revolution in PD patient care.

1.2.2 Diagnostic

No specific test exists to diagnose Parkinson's disease (Clinic n.d.). A neurologist will diagnose Parkinson's disease based on medical history, a review of signs and symptoms, and a neurological and physical examination.

In the Parkinson's Progression Marker Initiative (PPMI), a reference longitudinal cohort in Parkinson's Disease, to be considered as PD patients, subjects must have at least two of the following: resting tremor, bradykinesia and rigidity, which matched the main symptoms of PD described in (J. Jankovic 2008). This diagnostic must be confirmed with imaging (dopamine transporter SPECT scan) detecting dopamine transporter deficit. Also, patients with dementia are excluded, as it may suggest another Lewy Body disease. Note that unilateral symptoms, and positive reaction to levodopa are strong indicators of Parkinson's Disease.

1.2.3 Symptoms and Complications

Parkinson's disease is highly heterogeneous, and as such, symptoms can vary a lot between individuals. Early symptoms are mild and may remain undetected for a long time. Also, 60% of patients show a clear asymmetry of symptoms at diagnostic (P. N. Lee et al. 2007). We already mentioned resting tremor, bradykinesia and rigidity as the main symptoms, but a lot more are reported, either motor or non-motor (K. Ray Chaudhuri et al. 2006)., which we report on Table 1.1. Note that non-motor symptoms may appear even before motor symptoms and are assumed to account for a prodromal phase of the disease.

	Motor	Non-Motor	Cognition	Drug-Induced
Primary	Bradykinesia Rigidity Resting Tremor	Constipation Sleep Disorders REM Sleep Behaviour Disorders Anxiety Bladder (Urinary incontinence)		Dyskinesia
Secondary	Gait Dysfunction Freezing of Gait Speech Change Writing Change Loss of automatic movements	Smell dysfunction Sexual dysfunction Swallowing problems Chewing and eating problems Blood Pressure change Apathy	Thinking Difficulties Dementia	Hallucination Depression ICD Daytime sleepiness

Table 1.1 – Table of the main symptoms in Parkinson's Disease, labeled as primary or secondary depending of the symptoms frequency.

1.2.4 Pathophysiology

The brain dopamine system can be divided schematically into 3 main functional systems:

- Dopaminergic neurons from the substantia nigra pars compacta that project to the striatum (caudate nucleus and putamen): **nigrostriatal circuit**, involved in motor coordination
- Dopamine neurons in the ventral tegmental area (VTA) that project to the limbic system (nucleus accumbens, amygdala): **mesolimbic circuit**, involved in memory and in our motivation behaviour (reward pathway)
- Dopaminergic neurons of the VTA projecting to the cortex, in particular the frontal cortex: **mesocortical circuit**, involved in the management of our behavioral reactions.

In normal subject, this nigrostriatal dopamine projection inhibits neurons in the striatum carrying D2 dopamine receptors (indirect pathway) and activates those carrying D1 dopamine receptors (direct pathway). In addition, this nigrostriatal projection inhibits cholinergic interneurons in the striatum. Thus, the nigrostriatal dopamine pathway ensures the proper functioning of the subcortical-thalamocortical motor loop, the result of which is the cortical activation that causes movement. In Parkinson's disease, the progressive degeneration of the nigrostriatal pathway unbalances this subcortical-thalamocortical loop. The activation of D1 receptors (direct pathway) is reduced as well as the inhibition of D2 receptors (indirect pathway). The medial globus pallidus (GPi) thus receives increased activation via the indirect pathway and reduced inhibition via the direct pathway. This results in abnormally high activity of GPi neurons whose gabaergic projection excessively inhibits thalamic projections. The common efferece of the subcortical-thalamocortical motor loop becomes less functional, resulting in the parkinsonian motor deficit.

Note that the neurodegenerative process goes beyond the nigro-striatal pathway, explaining the occurrence of other motor signs (e.g. axial signs) and non-motor signs (e.g. cognitive impairment) that are resistant to dopaminergic treatment.

Oxidative stress is considered to play a major role in this neuronal destruction, a damaging effect due to the production of free radicals or their inadequate elimination. In

most cases, the etiology is unknown. The cause is complex and likely to be multifactorial, either genetic or environmental.

1.2.5 Treatments

Parkinson's disease is one of the few neurodegenerative diseases that can be treated symptomatically. The principle of treating Parkinson's disease is based on compensating for the dopamine deficit or correcting its consequences. There are various pharmacological routes: exogenous supply of dopamine precursors (L-dopa), dopamine agonists, enzymatic inhibitors of dopamine catabolism of monoamine oxidase type B (MAO-B), inhibitors of Catechol-O-methyltransferase (COMT) which potentiate the effects of L-dopa. Administration is either per os or subcutaneous depending on the drug and the stage of the disease. All antiparkinsonian drugs have a symptomatic effect by relieving motor symptoms (akinesia, tremor, rigidity). However, after a phase of improvement known as the "honeymoon period", patients experience motor complications due to the dopaminergic therapy, such as motor fluctuations and dyskinesias, which alter their functional outcome (Group 2004). But few pharmacological interventions exist for people with advanced Parkinson's disease. The development of new therapeutic strategies (e.g. cell therapy) is a major challenge for the coming years. Patient-controlled deep brain stimulation is also being considered for advanced forms of the disease. The latter has proven its beneficial effect (Benabid 2003; Deep-Brain Stimulation for Parkinson's Disease Study Group et al. 2001; Lau et al. 2015; Welter et al. 2015). Its invasive nature is however a limit to its expansion. We are moving towards a personalised follow-up of parkinsonian treatment based on a better understanding of pathophysiological mechanisms, also on the integration of biomarkers such as genetics, risk factors and imaging data.

1.3 Materials

1.3.1 Longitudinal Cohorts

In our work, we used 3 longitudinal datasets, that we detail below, and in Table 1.2.

- **Parkinson's Progression Markers Initiative (PPMI)** (Marek, Jennings, et al. 2011): Multicentric study in North America.
- **Cohort Study to Identify Predictor Factors of Onset and Progression of Parkinson's Disease (ICEBERG)** : Mono-centric study conducted at the Pitié-Salpêtrière hospital in France.
- **Drug interaction with genes in Parkinson Disease (DIGPD)** (Corvol, Artaud, Cormier-Dequaire, Rascol, Durif, Derkinderen, Marques, Bourdain, Brandel, Pico, et al. 2018b): Multicentric study in France.

PPMI

The Parkinson's progression markers initiative (PPMI), a landmark multicentric study in North America, publicly available through the PPMI web site www.ppmi-info.org (Marek, Chowdhury, et al. 2018; Marek, Jennings, et al. 2011). Its main objective is to identify new PD markers of progression, a critical step to accelerate research and measure advances in disease modifying therapeutics.

As such, it collects clinical and imaging data and biologic samples from various groups of patients. The study recruited first 423 PD patients, 196 Healthy Controls and 64 Subjects with Scans without Evidence of a Dopaminergic Deficit (SWEDD). It later recruited 65 prodromal subjects, 39 of which having idiopathic RBD, and 26 patients with hyposmia. It also aims at recruiting 600 subjects with specific genetic mutations such as LRRK2, GBA and SNCA, both with and without Parkinson's Disease.

Beyond allowing data analysis, PPMI establish standardized protocols for acquisition, transfer and analysis of clinical and imaging data and biological samples, which come from multiple sites and across multiple cohorts. Its notoriety and availability fosters reproducible research by sharing these processes with the community. Ideally, other cohorts with similar processes as PPMI could be used for replication, or pooled together to form bigger cohorts.

ICEBERG

Cohort Study to Identify Predictor Factors of Onset and Progression of Parkinson's Disease (ICEBERG) is a mono-centric study conducted at the Pitié-Salpêtrière hospital in Paris, France. It includes a strong focus on REM Sleep Behaviour Disorders. Its primary objective consist in the identification of factors that predicts the progression of the disease, ie rates of change of clinical, imaging and biomic outcomes. It also aims to investigate the occurrences of complications such as falls, freezing, dyskinesias, motor fluctuations, cognitive impairment and dysautonomia. It includes 360 subjects over a duation of 4 years, with measurements on an exhaustive set of tests (clinical, imaging, biological samples). It includes 200 PD patients, 50 subjects with idiopathic Rem-sleep behavior disorder (iRBD), 30 subjects related to a patient with genetically confirmed Parkinson Disease, and 50 Healthy controls. Additionally, 30 Patients with Parkinson Disease with a genetic mutation in parkin, LRRK2, SNCA or GBA are included. A description is available at <https://clinicaltrials.gov/ct2/show/NCT02305147>.

DIGPD

Drug interaction with genes in Parkinson Disease (DIGPD) is a multicentric study in France, aiming at identifying genes associated to disease and treatment complication and response. It includes 330 subjects followed over 6 years. On these 330 subjects, 200 patients have idiopathic Parkinson's disease, 50 idiopathic RBD, 50 are healthy subjects, and 30 have first degree parent of the affected patient with gene mutations (LRRK2 or GBA). A description is available at <https://clinicaltrials.gov/ct2/show/NCT01564992>.

Category	Modality	PPMI			ICEBERG			DIGPD
		Control	Prodromal	iPD	Control	iRBD	iPD	iPD
Demographics	Number of patients	196	65	423	70	56	165	415
	Male/Female	70/126	14/51	146/277	31/39	6/50	63/102	167/248
	Baseline Age	60.7+-11.2	68.8+-5.8	61.6+-9.7	62.4+-9.3	68.0+-5.5	62.2+-9.3	62.3+-9.8
	Disease Duration	-	-	0.6+-0.5	-	-	1.5+-1.0	2.6+-1.5
	Number of visits	8.5+-2.2	12.2+-2.9	13.9+-3.6	2.6+-1.3	3.1+-1.4	3.1+-1.5	5.1+-1.9
	Duration	6.3+-2.4	4.6+-1.3	6.2+-2.1	1.7+-1.3	2.3+-1.5	2.3+-1.6	4.5+-2.1
Motor Scores	MDS-UPDRS III Off (/132)	X	X	X	X	X	X	(X)
	MD-SUPDRS III On (/132)	X	X	X	X	X	X	X
Non-Motor Scores	SCOPA-AUT (/69)	X	X	X	X	X	X	X
	MDS-UPDRS I (/52)	X	X	X	X	X	X	X
	MDS-UPDRS II (/52)	X	X	X	X	X	X	X
	RBDSQ (/13)	X	X	X	-	-	-	-
	RBDHK (/100)	-	-	-	X	X	X	-
Cognition	MoCA (/30)	X	X	X	X	X	X	-
	MMSE (/30)	-	-	-	(X)	(X)	(X)	X
Imaging	MRI	X	-	X	-	-	-	-
	DATScan	-	-	X	-	(X)	(X)	-
	DATScan SBR	-	-	X	-	(X)	(X)	-

Table 1.2 – Biomarker inclusion in the considered groups of the 3 cohorts PPMI, ICEBERG and DIGPD. "X" denotes longitudinal measures for biomarkers, while "(X)" denotes only cross-sectional measures, or very small number of measurements. "-" stands for no data. Note that prodromal group of PPMI include iRBD patients (n=39), as well as patients with hyposmia (n=26).

1.3.2 Markers

Markers of disease progression

Neurodegenerative cohorts include a battery of clinical tests, which monitor and measure clinical symptoms on specific scales. In particular, cohorts built to investigate Parkinson's Disease typically monitor motor symptoms in details, but also non-motor symptoms such as dysautonomia, activities of daily living, depression, and sleep, or cognitive symptoms. They include imaging modalities, such as DATScan to monitor the loss of dopamine neurons in the Substantia Nigra. Note that a "Striatal Binding Ratio" is extracted from DATScan imaging, scalar measure of the density of neuros in pre-defined areas of the Substantia Nigra, typically left and right sides of both Putamen and Caudate. Cohorts also may include IRMs and fMRIs. Also, biological samples are collected, to measure for example the concentration of α -synuclein. We present the biomarker of the studied cohorts in Table 1.2, per category.

New promising markers

For the curiosity of the reader, we mention in this section a few new biomarkers in Parkinson's Disease, although they have not been used in this thesis.

In (Horsager, Andersen, Knudsen, Skjærbæk, et al. 2020b), the authors perform specific measurements to validate their theory of "Body First versus Brain First". To that end, they use very specific imaging biomarkers, unfortunately not present in publicly available datasets. They use 11C-donepezil PET/CT to assess cholinergic (parasympathetic) innervation, 123I-metaiodobenzylguanidine (MIBG) scintigraphy to measure

cardiac sympathetic innervation, neuromelanin-sensitive MRI to measure the integrity of locus coeruleus pigmented neurons, and 18F-dihydroxyphenylalanine (FDOPA) PET to assess putaminal dopamine storage capacity. They also measure Colon volume and transit times with CT scans and radiopaque markers. However the study is restricted to 37 patients.

On the other hand, AI based biomarkers offer cheap ways to extract new biomarkers. First, with the new developments in pose estimation (Y. Du, W. Wang, and Liang Wang 2015; Shahroudy et al. 2016; Shi et al. 2019; Yan, Xiong, and Lin n.d.), it is tempting to track PD subjects posture via a camera and estimate their motor symptoms using AI (Mehta et al. 2021). Also, this can be done via smartphone's gyroscopes (Ellis et al. 2015). This method implies more noise in the data, and the smartphone model factor has to be taken into account, but the simplicity of use makes it scalable to large cohorts.

More specific to Parkinson's Disease, and probably occurring earlier, are speech alterations. These have been recently studied in (Jeancolas 2019; H. Zhang et al. 2019), which show that patterns extracted from speech are predictive of Parkinson's Disease.

1.3.3 Biases in Parkinson's Disease longitudinal studies

These longitudinal cohorts provide an invaluable contribution toward a better understanding of the disease. However, we stress that they face issues that should not be forgotten when performing statistical analysis. We argue that they often have intrinsic biases, mainly due to the limitation of our current understanding of the disease. We detail in this section some of them.

First, they include a "temporal" bias, by design, as PD patients are recruited based on their motor symptoms. In Parkinson's Disease, diagnosis is suspected when motor symptoms are over a given threshold, and is confirmed via DATScan Imaging. This means that the longitudinal biomarker study of Parkinson's Disease typically begin at the onset of motor symptoms. However, there is increasing evidence that PD also present non-motor symptoms, even before conversion. If so, with non-motor symptoms that are part of the disease, then we miss signal in the pre-diagnosis phase, or also called prodromal phase.

Second, if we build cohorts focusing on pre-diagnostic PD we have biases in population. Indeed detecting prodromal patients of Parkinson's Disease is hard, and criteria with high sensitivity (Berg, Ronald B Postuma, et al. 2015), such as patients with idiopathic RBD, have rather low specificity, as many patients will rather either develop RBD after diagnostic or not at all (around half of PD patients). These prodromal cohorts will then not be representative of the overall heterogeneity of the disease, but rather of a subset.

Third, there is a bias in cognition. Dementia is an exclusion criteria for PD patients in PPMI and DIGPD, while in ICEBERG it is a MMSE<26. However, it might be that these cognition symptoms are a side effect of Parkinson's Disease, and excluding them removes heterogeneity in the datasets. Patients that ultimately develop dementia could form a PD subtype.

Lastly, there is a recruitment bias for genetic subgroups. Depending on the population, GBA+ PD patients account to 8-14% of proven PD, while LRRK2+ account to 2% to 40% (C. Klein and Westenberger 2012). Their scarcity makes them harder to recruit in

cohorts targeting precisely specific genetic subgroup. As a consequence, inclusion criteria are relaxed, such as maximum disease duration at baseline. This might generate biases with subjects having idiopathic Parkinson’s Disease, easier to recruit.

A way to bypass these biases would be to rely on prospective studies (Ronald B Postuma and Berg 2016), such as UK Biobank (*UK Biobank - UK Biobank* n.d.). However, these studies do no focus on a particular disease, and therefore do not provide advanced biomarkers with high frequency.

1.4 Goals in Parkinson’s Disease Longitudinal Analysis

In this section, in order to situate our work, we present the main research focus in longitudinal analysis of Parkinson’s Disease:

- **Identify early biomarkers of progression.** Extracting early and informative biomarkers of disease progression, would allow earlier recruitment of patients, further reducing left censorship in the longitudinal data.
- **Model the heterogeneity of PD.** Accurate modeling of the longitudinal heterogeneity of PD would allow for better understanding of the pathophysiological processes. Currently, this task is mainly approached via subtyping PD subjects into more consistent subgroups, but disease progression modeling offers an interesting alternative.
- **Prediction.** Machine Learning is practical to predict future values, or the arrival of complications of interest such as falls, freezing of gait, dyskinesias, motor fluctuations, or cognitive impairment, at e.g. 4 years from now.

1.4.1 Early Biomarkers of progression

First hypotheses on prodromal Parkinson’s Disease

Evidence of a pre-symptomatic, also called prodromal, phase of PD have emerged since the 2000s. From the analysis of brain dopamine and nigrostriatal damage in Parkinson’s Disease (Bédard et al. 1969; Bernheimer et al. 1973; Damier et al. 1999; Fearnley and Lees 1991), we already knew that when motor deficits appear, 50 to 60% of the dopaminergic neurons in the substantia nigra are already lost (Mahlknecht, Seppi, and Poewe 2015). Initial estimates based on these findings suggested a 5 to 6 year duration of dopaminergic loss before conversion to PD (Savica, Rocca, and J Eric Ahlskog 2010). This means that there would be a period where dopaminergic neurons are damaged, but motor symptoms are mild enough such that PD has not been diagnosed yet. This is of particular relevance for disease-modifying and neuroprotective therapies, which should be administered at the earliest stage of the disease.

In 2003, Braak (Braak et al. 2003) posits a topographical propagation model of Lewy Bodies in the brain from post-mortem samples. According to his eponym hypothesis, lesions would begin in the dorsal motor nucleus of the glossopharyngeal and vagal nerves and

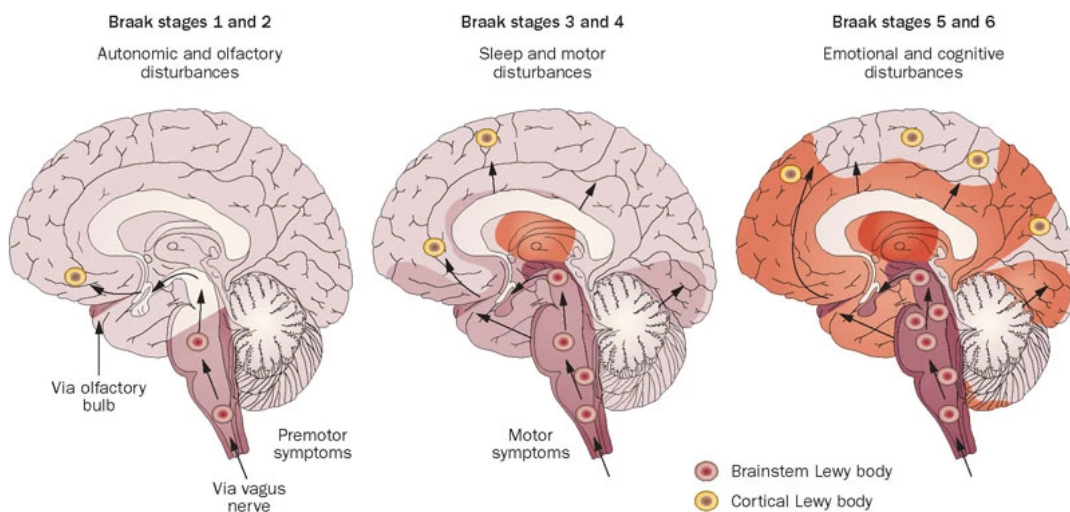


Figure 1.1 – Schematic overview of the topological propagation of Lewy Bodies in the brain during Parkinson's Disease. The red shading represents the presence of Lewy Bodies. On the left, in a first phase, Lewy Bodies are found in the viagal nerve and olfactory bulb, explaining smell loss and the presence of non-motor symptoms. Then Lewy Bodies propagate until they damage the substantia nigra, leading to sleep and motor disturbances. Ultimately, propagation until the cortex areas induce dementia. Note that this is an average scheme derived via post-mortem observations, and it does not encompass the whole heterogeneity of the disease, as many patient differ from this scenario in terms of symptoms ordering. (Reproduced from Braak et al. 2003).

anterior olfactory nucleus, and propagate progressively to the cortical area. (cf Figure1.1). This suggests an early involvement of the peripheral autonomic nervous system, and that PD starts even before damage to the nigrostriatal cells has initiated. As a consequence, the disease would first appear with non-motor symptoms (WC Koller 1992; Tolosa et al. 2009), which is backed by the clinical observations of non-motor symptoms such as sleep disturbance or constipation.

Identification of new prodromal biomarker in PD

From the 2010s, the study of this pre-symptomatic Parkinson's Disease, called prodromal PD, has gain interest in the field. According to (Kalia and Anthony E Lang 2016; Ronald B Postuma and Berg 2016; Savica, Rocca, and J Eric Ahlskog 2010) pre-clinical symptoms of PD appear up to 10-20 years prior to conversion. Classic diagnosis of PD cannot be performed yet, but there is increasing evidence that low-intensity signals of the underlying disease process are present.

It has been known for some time that specific symptoms can increase the risk of developing PD, such as Rapid Eye Movement (REM) Sleep Behaviour Disorder, refered as RBD, olfactory loss, constipation, depression and anxiety, erectile dysfunction (Ronald B Postuma and Berg 2019). In the last decade, extensive research explored in more details

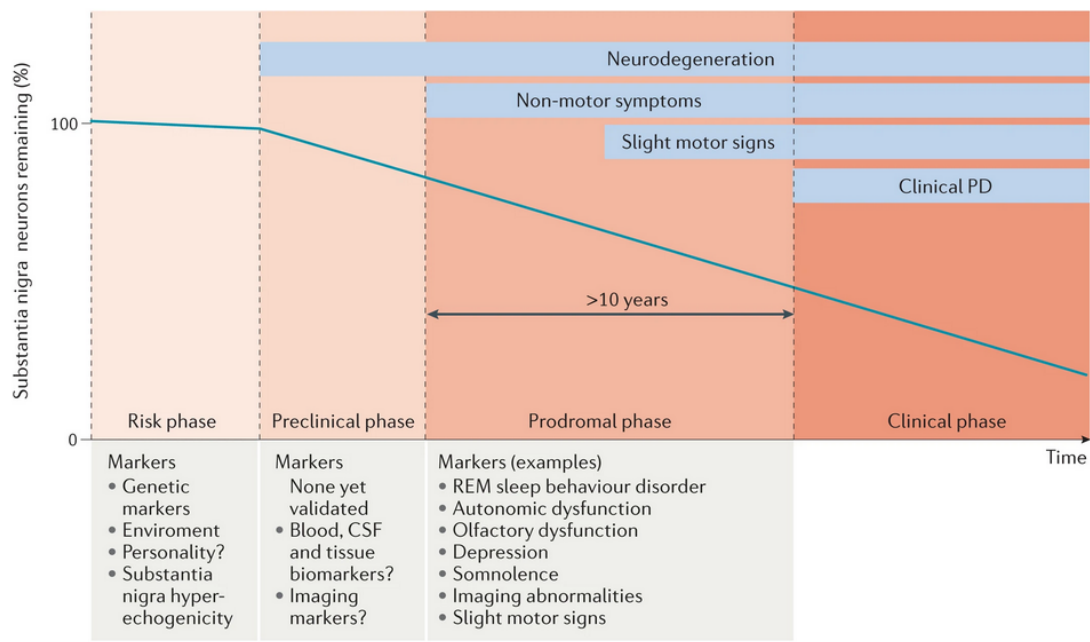


Figure 1.2 – Timeline of successive phases in prodromal Parkinson’s Disease. The prodromal phase may begin up to 10 years before conversion to Parkinson’s Disease, with first symptoms in sleep and autonomic dysfunction (especially constipation). (Reproduced from Ronald B Postuma and Berg 2016).

the markers of prodromal PD. It refined the analysis of their significance, and also exhibited the effects of new markers : orthostatic hypotension, urinary dysfunction, low-level motor symptoms, genetics, and abnormal dopaminergic imaging. (Berg, Ronald B Postuma, et al. 2015; Heinzl et al. 2019; Ronald B Postuma and Berg 2019; Shin et al. 2020). We report in Figure1.2 the estimated timeline of symptoms arrival in prodromal PD.

The MDS scoring for prodromal Parkinson’s Disease

To accelerate the research on prodromal PD, a generic score, the MDS Research Criteria for PD, has been proposed (Berg, Ronald B Postuma, et al. 2015; Heinzl et al. 2019). It is based on a naive bayes classifier, trained to distinguish those who convert to PD and those who don’t from various identified markers. Naive bayes allow the intuitive product of likelihood ratios, ranging from gender (1.2) to dopaminergic loss (43.3) and RBD (130) (see Figure 1.3).

However, the authors underline that naive Bayes assume independence of cofactors. If it is not the case, probabilities can be under or overestimated. For the majority of markers the independence assumption seems to hold, but for some, such as sleep or depression, it is very probable that there is at least correlation, if not causation.

These markers were designed to predict conversion to PD. The question of the specificity of these markers to Parkinson’s Disease against other synucleopathies arises too (Moscovich et al. 2020). For example, some might be shared with Dementia with Lewy Bodies, or

		LR+	LR-
Risk markers	Male sex	1.2 (male)	0.8 (female)
	Regular pesticide exposure	1.5	NA
	Occupational solvent exposure	1.5	NA
	Nonuse of caffeine	1.35	0.88
	Nonsmoking		
	Current smoker	NA	0.51
	Never smoker	1.2	NA
	Former smoker	NA	0.91
	First-degree relative with PD	2.5	NA
	or	LR+ dependent on age-related penetrance, see Table 2	NA
Prodromal markers	Known gene mutation (with intermediate-strength penetrance)		
	or	1.57 (highest quartile of PRS scores)	0.45 (lowest quartile)
	Polygenic risk score (PRS)		
	SN hyperchogenicity	3.4	0.38
	Diabetes mellitus (type II)	1.5	0.97
	Physical inactivity	1.3	0.91
	Low plasma urate levels	1.8 (in men)	0.88 (in men)
	PSG-proven RBD	130	0.65
	Possible RBD (questionnaire)	2.8	0.89
	Dopaminergic PET/SPECT clearly abnormal (eg, <65% normal, 2 SDs below mean)	43.3	0.66
	Subthreshold parkinsonism (UPDRS-III >3 excluding action tremor or MDS-UPDRS-III >6 excluding postural and action tremor)	9.6	0.55
	or	3.5	0.60
	Abnormal quantitative motor testing		
	Olfactory loss	6.4	0.40
	Constipation	2.5	0.82
	Excessive daytime somnolence	2.7	0.86
	Orthostatic hypotension (OH) – neurogenic OH	18.5	0.88
	Symptomatic OH	3.2	0.80
	Erectile dysfunction	3.4 (in men)	0.87 (in men)
	Urinary dysfunction	2.0	0.90
	Depression ($\pm$ anxiety)	1.6	0.88
	Global cognitive deficit	1.8	0.88

NA, not applicable.

Figure 1.3 – Identified prodromal markers and their score. Markers most indicative of conversion to PD are "(PSG verified) RBD" (130 $\times$), "abnormal substantia nigra on imaging" (42 $\times$), and "orthostatic hypertension" (19 $\times$). (Reproduced from Heinzel et al. 2019).

multi-system atrophy, but at different intensities and predictive power.

iRBD as a proxy of prodromal Parkinson's Disease

Having identified RBD as a strong risk factor of developing PD, several studies recruit iRBD patient to perform a longitudinal analysis (see subsection 1.3.1), as a proxy of general PD prodromal progression.

In (R. Postuma, Lang, et al. 2012) and (Fereshtehnejad, Yao, et al. 2019), the author re-align the subjects with their age of conversion to PD, and then look back in time to estimate the progression of biomarkers, as well as the duration before diagnosis where these biomarkers become abnormal.

In (R. Postuma, Lang, et al. 2012) the author focus on the progression of motor markers (subscores of MDS-UPDRS part III) in iRBD patients, and their specificity to predict conversion to Parkinson's Disease (see Figure 1.3). They find that Voice/Face is affected first, at 9.8 (6.7, 29.8) years before conversion, followed by Purdue Pegboard at

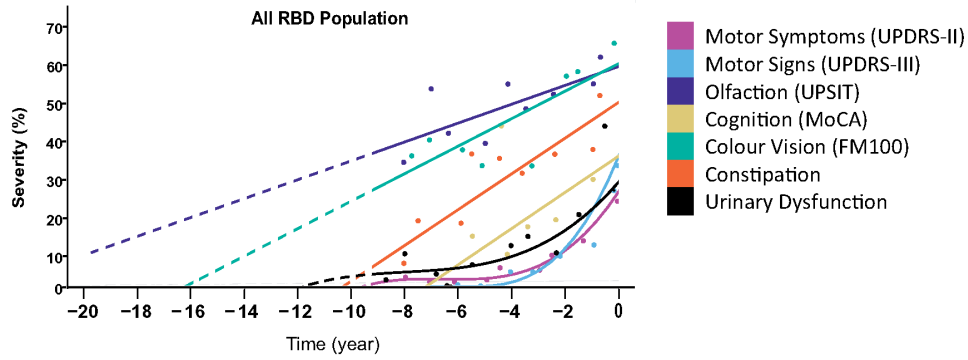


Figure 1.4 – Combined progression trajectory of motor and non-motor manifestations from prodromal stages to phenoconversion based on actual measurements in a population of RBD patients. For each biomarker, different regressions are performed with respect to age at conversion: linear regression, weighted least squares (WLS) regression, and non-linear third-degree polynomial regression, and more complex ones are used if they provide a significant increase in R^2 ($> 10\%$). (Reproduced from Fereshtehnejad, Yao, et al. 2019).

8.6 (4.5, 87.0) and Alternate Tap Test 8.2 (5.9, 20.2), Times up and go 6.3 (4.8, 12.4) then Axial-Gait 7.1 (5.6, 10.8), Bradykinesia 5.4 (4.8, 6.5), Rigidity 4.4 (3.9, 5.5) and Gait 4.4 (3.9, 5.4), and lastly tremor.

In (Fereshtehnejad, Yao, et al. 2019), the author perform a more exhaustive modeling of prodromal biomarker progression on iRBD patients (see fig 1.4). Expressivity of model is also increased with the use and choice of linear, but also mixed effect and polynomial models of progression. They also re-align subjects in regard with their date of conversion. They find that symptoms are first seen in olfaction (-22 years), then Erectile dysfunction (-16 years), Alternative Tap (-12.9 years), Color vision (-12.8 years) and Constipation (-10.8 years), and then MDS-UPDRS II (-9.3 years) and MDS-UPDRS III (-6.5 years).

1.4.2 Subtyping Parkinson's Disease

Why and how to subtype Parkinson's Disease

There is increasing evidence that clinical and prodromal Parkinson's Disease are very heterogeneous, and as such can be classified into subtypes, with possibly different disease mechanisms, symptoms and patterns of progression (Berg, Borghammer, et al. 2021).

Successful PD subtyping could have important practical implications for clinicians and researchers. Subtyping at baseline could allow for more personalized medicine, adapting treatments and prognosis (Qian and Huang 2019). Subtyping from the prodromal stage of the disease would offer the possibility to perform these tasks even before motor symptom start. In all cases, this would be an opportunity for a better understanding of disease aetiology, pathophysiology, possibly leading to developing a curative treatment.

To be successful however, subtyping should be relatively easy to implement, with a small number of clinical scales to be performed at screening. It should include the smallest number of subtypes to preserve statistical power and promote reproducibility. At the same

time, it has to preserve clinical fidelity, and if possible reflects best the pathophysiological processes that determine the observed heterogeneity of PD (Marras and Anthony Lang 2013).

Subtyping with Motor phenotypes

The idea that subtypes of PD exist, differing in motor symptoms, dates back from as early as the 1980s (Joseph Jankovic et al. 1990; Zetuský, Joseph Jankovic, and Pirozzolo 1985). Two Parkinson's motor phenotypes were already clearly distinguished: one with bradykinesia, postural instability and gait difficulty (PIGD) and another with tremor as the dominant feature, as a milder form of PD. Also, an intermediate group also referred to patients in between. The PiGD group report significantly greater subjective intellectual, motor, and occupational impairment than the tremor group.

To standardize the two main motor subtypes, PiGD and Tremor-Dominant, Stebbins et al. 2013 posits cutoffs ratios to separate patients between PiGD and Tremor subtypes.

In 2005 (Lewis et al. 2005) proposed a finer grain data-driven clustering of PD subtypes, leading to 4 subtypes. Since then, other motor subtyping of the disease have been proposed (Eisinger et al. 2017; Selikhova et al. 2009) with similar subgroups.

Subtyping with Motor and Non-Motor

If the existence of PD motor subtypes seems to make consensus, subtyping with inclusion of other biomarkers is still under debate. Non-motor symptoms, especially, are very heterogeneous in the early stages of the disease, and have also been considered to perform PD subtyping (Marras and K Ray Chaudhuri 2016; Mu et al. 2017; Sauerbier et al. 2016).

A reason for their higher heterogeneity compared to motor symptoms might be that at disease onset, patients have been realigned with regard to their motor symptom onset, and not their non-motor symptom onset, as discussed in subsection 1.3.3.

Subtyping with all markers

Clustering studies now incorporate exhaustive sets of markers (Fereshtehnejad, Romenets, J. B. Anang, et al. 2015; Fereshtehnejad, Zeighami, et al. 2017; P. Liu et al. 2011; Linbo Wang et al. 2020), showing significant effects of RBD, orthostatic hypotension and cognition in separating patients. However adding more biomarkers adds difficulty to the task, especially in the a-posteriori interpretation of clusters.

Indeed, in a review article, Thenganatt and Joseph Jankovic 2014 explain that despite the increasing literature on PD subtyping, it remains unclear if identified motor subtypes, as PiGD and Tremor-dominant, can be consistently extracted with a data-driven method.

Issues of Parkinson's Disease data-driven clustering

Parkinson's Disease data-driven clustering faces many issues at the moment.

First and most important would be consistency. In (P. Liu et al. 2011), authors assessed the relationship between the data driven subtypes they derived and empirically assigned motor phenotypes. Their 4 cluster solution was in accordance with the work of

Lewis et al. 2005 and Reijnders et al. 2009 with similar subtypes. However they did not match the 3 traditional motor phenotypes: tremor dominant, PIGD and indeterminate, as defined by Joseph Jankovic et al. 1990, so that motor phenotypes did not tend to group together within their four clusters but were spread across them. This highlights the non-accordance between the traditional 'motor phenotype' method and the most common subtypes identified by cluster analysis methods (Marras and Anthony Lang 2013).

Closely related is the replicability issue. Replicability is not straightforward between different cohorts, and separate analysis present all sorts of biases: inclusion criteria, variable selection, methodology (Qian and Huang 2019), while a-priori subtyping, (e.g. with motor phenotypes), in a broader line, seem less prone to overfitting the dataset.

Also, on the methodological aspect precisely, taking into account disease duration or not may be of capital importance. Indeed, markers, especially non-motor ones may present paths that are not mutually exclusive (Erro et al. 2020). Not including the disease duration might result in clustering the disease stages, missing the different patterns of progression. Fereshtehnejad, Zeighami, et al. 2017 for example managed to cluster subjects per speed of progression. In this spirit, we could argue that cross-sectional clustering, as done in most studies (Eisinger et al. 2017; Fereshtehnejad, Romenets, J. B. Anang, et al. 2015; Fereshtehnejad, Zeighami, et al. 2017) cannot capture different patterns of progression (such as speed) by nature, and only clusters a photography at age t_i^* for each patient i of the cohort. These cross sectional studies can then only hope that the future progressions will be different between clusters, but cannot cluster directly according to this criteria. As a consequence, cluster inclusion may change over time (Eisinger et al. 2017). What is more, these patients may already be at different stages ψ_i^* of the disease at baseline. Linear alignment according to disease duration might be possible, but care should be taken when extrapolating to unseen patient ages, and with the linear assumption. For all these reasons, we believe that a longitudinal clustering method could prove beneficial to the field.

To a broader extent, the community has not yet agreed on well-defined PD subtypes. This raises the question of which subtype classification is most relevant to the underlying cause of PD patterns of progression (Marras and Anthony Lang 2013). In a review, Mestre et al. 2021 even propose to foster individual modeling of PD progression, instead of focusing on extracting subtypes which do not yet show a clear relevance.

PD Subtypes in the literature

PIGD vs (Intermediate) vs Tremor-dominant As presented in section 1.4.2, a known motor phenotype in the community separate Tremor dominant subjects, from PIGD, with eventually an intermediate cluster.

Genetics Although not considered as subtypes of the disease, the main genetic variants, LRRK2, GBA, and SNCA show supposedly specific patterns of progression. GBA for instance are supposed to present earlier motor decline than idiopathic PD patients. SNCA show a very early form of PD. LRRK2 would present a purer motor form of PD, with gait abnormalities.

REM Sleep Behaviour Disorder RBD, already described in 1.3.2, has been shown to be associated with neurodegenerative disease, particularly the synucleinopathies (Boeve et al. 2007). RBD is predominantly related to brain pontomedullary dysfunction and is therefore likely a marker of BraakStage II in PD (cf Figure 1.1). RBD is frequent in PD according to (Moscovich et al. 2020; Sixel-Döring et al. 2011), which report 46% of RBD in PD patients. RBD can occur early, and the majority of patients that have RBD are supposed to develop a synucleinopathy (R. Postuma, Gagnon, et al. 2009; Sixel-Döring et al. 2011). As such, RBD is not only a predictor of the disease, but can be considered a subtype of PD as well (Berg, Borghammer, et al. 2021; Horsager, Andersen, Knudsen, Skjærbæk, et al. 2020a; Yoritaka et al. 2009).

Brain First versus Body First Recently, Horsager, Andersen, Knudsen, Skjærbæk, et al. 2020a proposed that PD is in fact two diseases, with a brain first trajectory, that corresponds to Braak's theory, and a body first trajectory, with disease beginning in the guts, then damaging the autonomic system, while the dopamine is still relatively untouched (see fig 1.5). They combine precise imaging markers to measure parasympathetic, sympathetic, locus coeruleus, and the nigro-striatal innervation on 37 de Novo PD patients to support this theory. Incidentally, PD-RBD+ would correspond to the body first group, while PD-RBD- would correspond to the brain first type.

1.4.3 Other research focus

Dementia prognosis in Parkinson's Disease

In 2003, it has been shown in an 8 year study that up to 78% of patient developed Dementia. The 4 year prevalence of Dementia in PD was already nearly 3 times higher than in the non-PD group (Aarsland et al. 2003).

Since then, identifying the risk factors of Dementia in PD has been explored (J. B. Anang et al. 2014). Risk factors such as mild cognitive impairment at baseline, RBD and orthostatic blood pressure drop were the most significant. Other factors were found significant : hallucinations before baseline, abnormal color vision, gait abnormalities, falls, freezing, Purdue Pegboard Test and alternate tap test. Dementia in the iRBD population has also been assessed with similar risk factors (Ronald B Postuma, Iranzo, et al. 2019).

ICD prognosis in Parkinson's Disease

Impulse control disorders (ICDs) are a class of psychiatric disorders involving problems in the self-control of emotions and behaviors. They include addiction, gambling and compulsive behaviours. They are believed to be mainly due to side effects of PD treatment, being correlated with dose of Dopamine Agonists (Corvol, Artaud, Cormier-Dequaire, Rascol, Durif, Derkinderen, Marques, Bourdain, Brandel, Pico, et al. 2018a). Faouzi et al. 2021 propose to predict ICDs with machine learning.

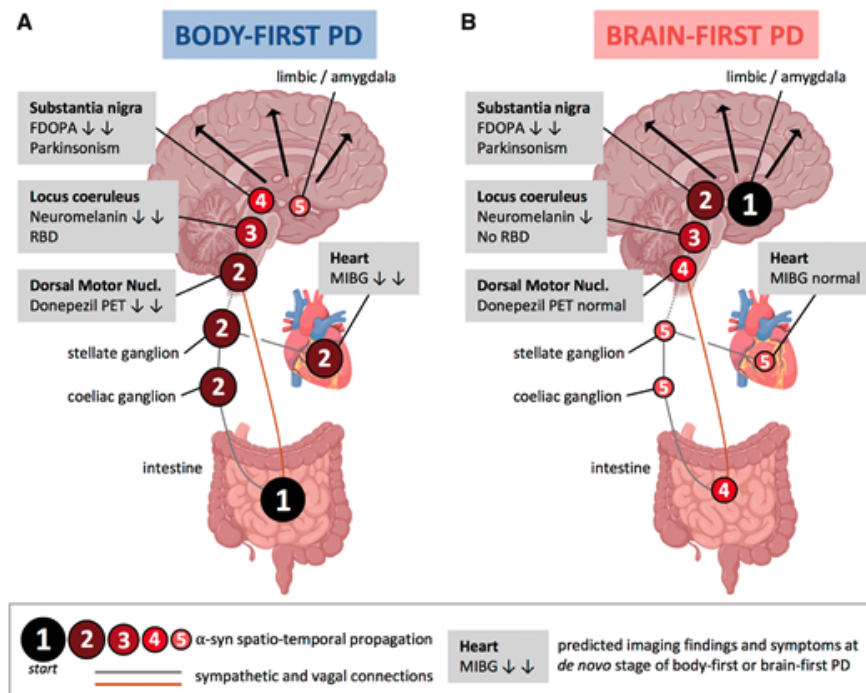
Figure 1

Figure 1.5 – Schematic overview of Brain First versus Body First Theory. In (A), for the patients with a body-first form of the disease, abnormalities would begin in the guts, on the form of Lewy Bodies, and then propagate through the autonomous nervous system to the locus coreuleus and substantia nigra. iRBD would appear as a prodromal form of this subtype. In (B), subjects with a brain first form of the disease, progression would begin in the amygdala, to then propagate to the substantia nigra. This form is close to the one represented in Figure1.1. (Reproduced from Horsager, Andersen, Knudsen, Skjærbæk, et al. 2020a).

Chapter

2

Spatiotemporal Model of Disease Progression

Outline of the current chapter

2.1 Mathematical Background	42
2.1.1 Longitudinal Dataset	42
2.1.2 Mixed Effect Models	42
2.1.3 Disease Progression Modeling	44
2.1.4 Notions of Riemannian Geometry	47
2.2 Building a Riemannian Mixed Effect Model	53
2.2.1 Model Geometry	53
2.2.2 Statistical Description	54
2.2.3 A mixed effect model	55
2.2.4 Instantiations	56
2.2.5 Example on $]0, 1[^2$	60
2.2.6 Remarks	60
2.3 Estimation	64
2.3.1 Statistical Tasks	64
2.3.2 Computation of the likelihood	64
2.3.3 Calibration	64
2.3.4 Personalization	67
2.3.5 Simulation	67

In this chapter, we aim at introducing the spatiotemporal model of progression of (Schiratti, Allasonniere, et al. 2015b). For more details we invite the reader to refer to the thesis associated to this work (Schiratti 2017), as well as to more recent theses leveraging this work (Bône 2020; Koval 2020)

2.1 Mathematical Background

2.1.1 Longitudinal Dataset

Let us first define longitudinal datasets notations, and state the hypotheses on the data on which we will rely to build our longitudinal spatiotemporal model.

Let us have N patients, each with m_i visits. At each visit j of patient i , K modalities are measured, and reported in the observation vector $y_{i,j} = (y_{i,j,k}, k \in [1, K])$. We assume that at each visit (i, j) , we know the age of the patient, denoted $t_{i,j}$. We write $\mathcal{Y} = \{y_{i,j,k}, i \in [1, N], j \in [1, m_i], k \in [1, K]\}$ and $\mathcal{T} = \{t_{i,j}, i \in [1, N], j \in [1, m_i]\}$, so that the couple $(\mathcal{Y}, \mathcal{T})$ refers to a longitudinal dataset $\mathcal{D}$.

In our application to neuro-degenerative diseases, we use ad-hoc biomarkers that measure different facets of disease severity. As such, and in the absence of treatment, we may assume **monotonicity** in the data with regard to time (see 1.3.2 for details on the biomarkers), i.e. $j_2 \leq j_1 \Leftrightarrow y_{i,j_2,k} \leq y_{i,j_1,k}, \forall k \in [1, K]$ and $i \in [1, N]$. Note that this monotonicity will be enforced by design in the longitudinal spatio-temporal model (see 2.2.4)

Although in the general case $y_{i,j}$ might refer to complex data, such as meshes or images, in our instances of the model, we will restrict to an application to k **scalar biomarkers**, such that $y_{i,j,k} \in \mathbb{R}$ and $y_{i,j} \in \mathbb{R}^K$.

2.1.2 Mixed Effect Models

In order to build a progression model, one of the simplest method consists in fitting an Ordinary Least Square regression to map time (and possibly other covariates) to biomarker values, such that

$$y_{i,j} = At_{i,j} + \epsilon_{i,j}, \quad \epsilon_{i,j} \sim \mathcal{N}(0, \sigma^2), \quad A \in \mathbb{R}$$

assuming all visits are independent and identically distributed (iid). However, in our case it is clear that the iid assumption does not hold because data are structured. Multiple samples from the same patients are in general not independent, and in practice are likely to be correlated.

Mixed effect modeling (Fisher 1919/ed) is the reference method when analyzing data with a group structure. In mixed effect models, we consider that the response follows a known functional form f that depends on unknown effects θ and z .

$$y_i = f(\theta, z_i) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

These effects are of two different kinds. Some of the effects are "fixed" (same for all individuals). The other effects are "random" (z), and depend of the individuals (or subgroups). This modeling is also referred as "hierarchical modeling", where higher levels (fixed effects) explain more variance than lower levels (random effects).

With longitudinal data, grouped data occur by design, as measurements for a same individual i are repeatedly observed at different times $t_{i,j}$. Using time as the only cofactor, a baseline model is the random slope random intercept model, which consists in a population

slope and intercept, as well as individual's slope and intercept which describe individual variations around the population average (see section 2.1.2).

Throughout this work, we will see the individuals random effects z_i as representations of interest that describe individuals patterns of progression. This raises the issue of building an adequate modeling of the heterogeneity in patterns of progression, to keep the right balance between modeling power, and interpretability of these parameters. In the following sections, we will present the standard mixed effects models. We will then introduce more complex models to handle more adequately the progression heterogeneity.

Linear Mixed Effect model

In Linear Mixed Effects models (LME) (Laird and Ware 1982a), the functional is assumed to be linear in both the fixed and random effects. Random effects are typically assumed to follow a diagonal multivariate normal distribution. The linear mixed effect model then writes

$$y_i = X_i\beta + Z_ib_i + \epsilon_i \quad (2.1)$$

with p predictor variables and $q \leq p$ random effects:

$$\left\{ \begin{array}{l} y_i \text{ patient } i \text{ observations of size } (m_i \times 1) \\ b_i \sim \mathcal{N}(0, \sigma_b \mathbb{I}), \text{ patient } i \text{ unknown random effects of size } (q \times 1) \\ \epsilon_i \sim \mathcal{N}(0, \sigma \mathbb{I}) \text{ gaussian noise, typically diagonal} \\ \beta \text{ unknown vector of fixed effects of size } (p \times 1) \\ X_i \text{ matrix of covariates of size } (n_i \times p) \\ Z_i \text{ matrix of covariates of size } (n_i \times q) \end{array} \right.$$

For longitudinal data, a special case of LME is often used, the random-slope random-intercept model. It separates time-varying and time-in dependant effects with $X = (1, t_{i,j})$ and $Z = (1, t_{i,j})$:

$$y_{i,j} = (t_{i,j} - t_0)(\beta_1 + b_{1i}) + (\beta_2 + b_{2i}) + \epsilon_{i,j} \quad (2.2)$$

For parameter estimation, two cases are distinguished. In the case where the covariance of the random effects is known the Henderson's mixed model equations (MME) can be solved via matrix inversion to yield the best linear unbiased estimates (BLUE) and predictors (BLUP) for β and u . In the general case covariance is unknown, and Laird and Ware (Laird and Ware 1982a) propose to use the Expectation-Maximization algorithm (Lindstrom and Douglas M. Bates 1988).

Non Linear Mixed Effect model

In pharmacokinetics, growth curves, or disease progression modeling, the linear approximation does not always applies and it can be helpful to posit a nonlinear functional form to better fit the data. Extending the LME framework with non-linear functionals, namely to Non-linear Mixed Effects models (NMLE) has been proposed (Lindstrom and Douglas M. Bates 1990a; Pinheiro and Douglas M Bates 1995). However, it requires a more complex approach for parameter estimation. In the general setting, the model writes

:

$$y_i = f(X_i, b_i, \beta) + \epsilon_i \quad (2.3)$$

$$\begin{cases} X_i \text{ matrix of covariates of size } (n_i \times p) \\ b_i \sim \mathcal{N}(\beta, \sigma_b \mathbb{I}), \text{ patient } i \text{ unknown random effects of size } (q \times 1) \\ \beta \text{ unknown vector of fixed effects of size } (p \times 1) \\ \epsilon_i \sim \mathcal{N}(0, \sigma \mathbb{I}) \text{ gaussian noise, typically diagonal} \end{cases}$$

For longitudinal data, a simple non-linear model consist in applying the functional f on the random slope-random intercept model:

$$y_{i,j} = f((t_{i,j} - t_0)(\beta_1 + b_{1i}) + (\beta_2 + b_{2i})) + \epsilon_{i,j}$$

Note that estimation is more complex than in the linear case.

2.1.3 Disease Progression Modeling

Challenges in modeling neurodegenerative longitudinal data

LME and NMLE cannot be applied straightforward on neurodegenerative diseases longitudinal data, which include multiple issues. Indeed a classical random slope random intercept model is hardly enough. First, such model would completely miss the high non-linearity of biomarker progression. Second, even with an ad-hoc functional, one would still need to define a reference time-point t_0 . This t_0 would be arbitrary, as no reference timepoint or disease stage has been defined in Alzheimer's or Parkinson's. Then, patients are temporally unaligned regarding the disease advancement. They are included in longitudinal studies at possibly different disease stages, and progress at different speeds. Finally, we not only

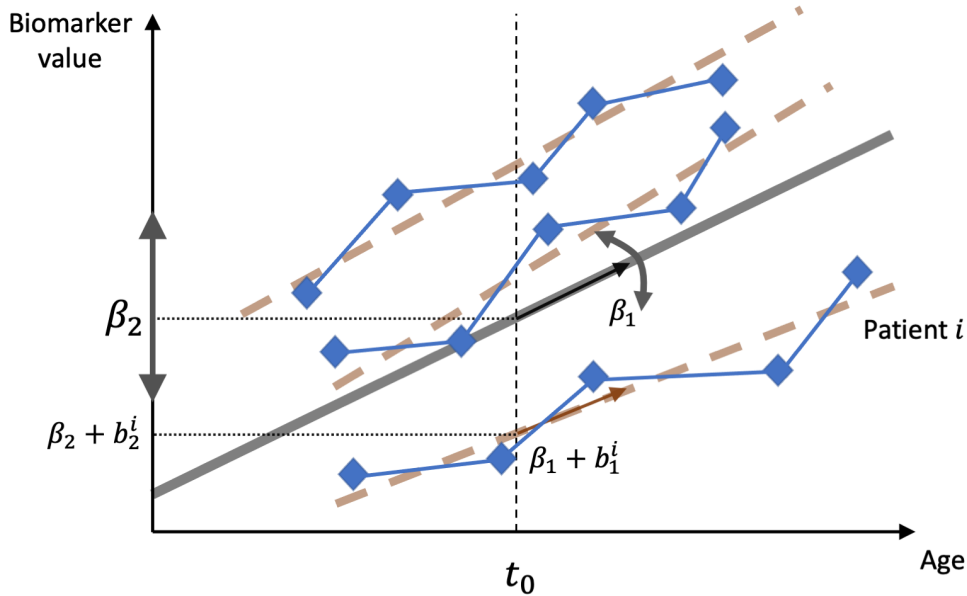


Figure 2.1 – Illustration of a random slope random intercept model. In this model, t_0 is fixed a-priori.

aim at modeling clinical marker progression, but also more complex measurements such as imaging modalities. In the NMLE setup, choosing a functional f with good properties becomes tricky.

A first conceptual model of neurodegenerative disease progression

In 2010, (Clifford R. Jack, Knopman, et al. 2010) proposed a first conceptual plot of prodromal biomarker progression in Alzheimer’s Disease, on the form of sigmoids. This model states clearly that biomarker evolve continuously over the span of the disease. Importantly, they are dynamic in a sequence, each for a relatively short segment of the disease timeline. This cascade of events in the biomarkers seen reflects the underlying disease processes. This view is however hypothetical and comes from the practitioners experience. Since then statisticians have quantitatively tackled the problem by developing methods that learn from cohorts of patients, by building on these first insights.

Discrete model: Event Based Models

Close to the "cascade of events" perspective, Event Based Models (EBM) (Fonteyn et al. 2011) offer an intuitive discrete modelization of disease progression, by estimating from the data the ordering of an a-priori defined set of events, with a first application on Alzheimer’s Disease. This allows to derive a data-driven time line of disease progression with detail, and the uncertainty of the estimated order informs about the confidence in the sequence of events. Since then EBM models have been extended ((*A Data-Driven Model of Biomarker Changes in Sporadic Alzheimer’s Disease* - PubMed n.d.; Alexandra L. Young et al. 2015) and SuStaIn (*Uncovering the Heterogeneity and Temporal Complexity of Neurodegenerative Diseases with Subtype and Stage Inference* / *Nature Communications* n.d.) proposed to perform patient subtyping by estimating a mixture of timeline. It has been applied to Huntington disease (Wijeratne et al. 2018), and also more recently to Parkinson’s Disease (Neil P. Oxtoby et al. 2021). While EBM offer a generative model of disease progression, they can rely solely cross-sectional cohorts, i.e. cohorts with 1 visit per patient. This extends their applicability, as cross-sectional cohorts are more frequent, but miss the time delay between events, which might inform on finer temporal dynamics seen in individuals.

Continuous models

Based on longitudinal cohorts, continuous models of biomarkers progression, such as Non Linear Mixed Effect Models have naturally been proposed. However, they have to cope with the temporal disalignment between patients, as there is no obvious biological “time zero” in Alzheimer’s disease. Patients may be included in studies at possibly different disease stages, and evolve at different speeds. This raises the question of the mathematical definition of a disease stage. Jedynak et al. 2012 propose the creation of an abstract disease progression score (DPS), by temporally re-aligning patients in time according to their biomarker values. Time reparameterization is performed via an affine reparameterization of age t with $s_i(t) = \alpha_i t + \beta_i$ with subject dependant α_i and β_i . Biomarker evolutions are then modeled as sigmoid functions of $s(t)$, and estimation is carried jointly on all

parameters modulo identifiability constraints. The DPS, normalized to have 0 mean and unit variance may then be used as a biomarker of disease severity. In (Guerrero et al. 2016) and (Michael C. Donohue et al. 2014), authors build non-linear mixed-effect models using the same form of time reparameterization, by treating patient parameters as random effects. With this perspective, a long-term average trajectory of disease progression is derived from the fixed effects of the model, built from possibly short-term patient snapshots. While the above-mentioned models were focusing on scalar values, such as clinical tests, or measurements extracted from imaging modalities, the authors of (Răzvan V. Marinescu et al. 2019) propose to directly model vertexwise or voxelwise trajectories from imaging modalities. DIVE estimates a common progression pattern for all patients by clustering vertices of voxels in consistent groups, each group being parameterized by a sigmoid. Then patients follow this average trajectory modulo an affine reparameterization of time.

The longitudinal spatiotemporal: a generic continuous model of progression

Schiratti, Allasonniere, et al. 2015b proposes a Non Linear Mixed Effect Model, namely the longitudinal spatio-temporal model, generic in the sense that it handles data lying on a smooth Riemannian Manifold, may it be meshes, brain images (Bone, Colliot, and Durrleman 2018) or simply intervals of the real line. The model estimates a disease stage via an affine time reparameterization of patient ages. Most importantly, it offers a theoretical framework to build a system of spatio-temporal coordinates. Temporal variation account for position on the average trajectory axis, i.e. in terms of disease stage. On the other hand, "spatial" variations account for variations which are orthogonal to the average trajectory. That is to say variations from the average trajectory after having performed the affine temporal re-alignment. Note that both patient affine reparameterization parameters and spatial parameters are considered as random effects in the mixed effect model. In practice, spatial variations are available in close form as geometrical transformation of the average trajectory.

In the following, we will describe and use the model of Schiratti, Allasonniere, et al. 2015b. It relies on Riemannian Geometry as an insightful theory to tackle disease progression modeling, leveraging the notions of geodesics and exp-parallelization in a Riemannian Manifold.

2.1.4 Notions of Riemannian Geometry

In machine learning, real data often show underlying structure. We highlight that we do not speak there about non-iid dataset, but rather of a possibly lower dimensionality of the data we work with. For example, the space of natural images of size 128x128 is thought to be smaller than $\mathbb{R}^{128 \times 128}$. We can reasonably assume that this is the case too for medical data, such as bounded clinical scores of MRIs of the brain (Manifold Learning of Brain MRIs by Deep Learning). In this spirit, Schiratti, Allasonniere, et al. 2015b assume that the space of measurement is defined by smooth constraints, and do not behave as a euclidian space. From the point of view of differential geometry, this space of measurements, defined by smooth constraints, can be seen as a smooth manifold. Riemannian Geometry offers a theoretically grounded framework to describe such smooth manifolds, introducing the notions of distance and angles via a Riemannian Metric on the manifold. Explanation of the concepts of Riemannian Geometry can be found in *Do Carmo: Riemannian Geometry* (Carmo 1992). Importantly, Riemannian geometry generalizes the notion of straight lines in smooth manifolds with geodesics. Geodesics can be parameterized via a position on the manifold, and a speed at this point. The geodesic then follows the path of minimal length along this direction, length of infinitesimal increments depending on the metric at the given point. This notion of geodesics allow to perform regression with complex objects that lie on the manifolds (images, shapes of the heart or hippocampus). This raises then the question of how to choose and parametrize manifold and Riemannian metric to be consistent with the data at hand (see 2.2.4).

Smooth Manifolds and Tangent spaces

Definition 2.1.1 (Manifolds) *A topological space M is called a d -dimensional manifold if, for every point $p \in M$ there exist a neighbourhood $U(p)$ and a homeomorphism x :*

$$x : \begin{cases} M \rightarrow \mathbb{R}^d \\ U(p) \rightarrow x(U(p)) \end{cases}$$

We also write $\dim(M) = d$. Intuitively, a d -dimensional manifold is a topological space which locally looks like $\mathbb{R}^d$. (U, x) is called a local coordinate chart. An atlas $\mathcal{A}$ of M is an indexed family of charts $\{(U_\alpha, x_\alpha)\}$ that covers M , i.e. such that $\{\bigcup_\alpha U_\alpha\} = M$.

Definition 2.1.2 (Smooth Manifold) *A manifold is called a smooth manifold if its atlas $\mathcal{A}$ is C^∞ , that is, if the transition map $y \circ x^{-1}$ between any two charts $(U, x), (V, y) \in \mathcal{A}^2$ is C^∞ .*

Thus, a smooth manifold is a topological manifold with a globally defined differential structure.

Definition 2.1.3 (Tangent Space) *Given a smooth manifold M , we can differentiate a coordinate chart at each point p . The tangent space $T_p M$ at p is defined as the space of all possible velocities $\dot{c}(0)$ for any curve $c :]-\epsilon, +\epsilon[\rightarrow M$ such that $c(0) = p$.*

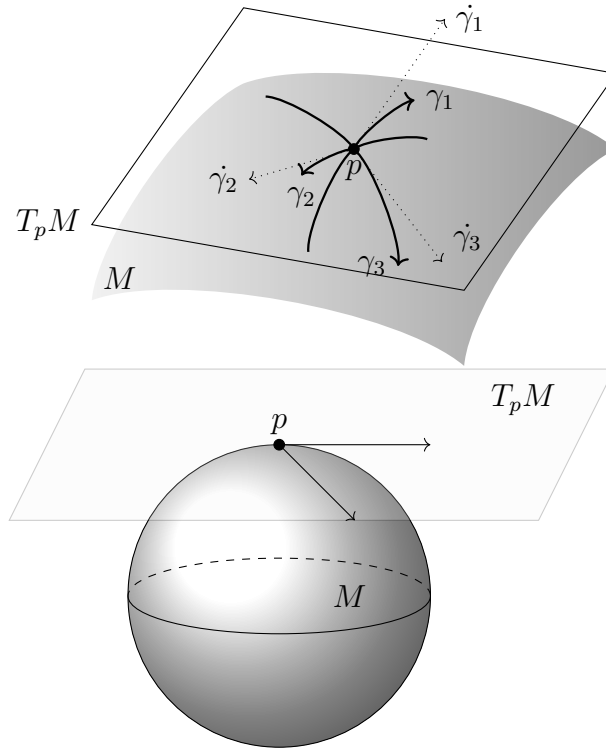


Figure 2.2 – Representation of different manifolds with a tangent space at point p . On top, intrinsic view of a manifold, with a representation of the tangent space $T_p M$ at point p . On the bottom, we show the sphere of $\mathbb{R}^3$ as a 2-dimensional manifold immersed in $\mathbb{R}^3$, which is an example of extrinsic view of a manifold. Note that the two approach are equivalent. In the following, we will use the intrinsic view for vizualization purposes.

This tangent space $T_p M$ provides a linear approximation of the manifold in the neighborhood of p . With $p \in M$ and $x = (x_1, \dots, x_n)$ a coordinate system around p , we can define (Schiratti, Allasonniere, et al. 2015b) a basis $\mathcal{B}$ of the tangent space $T_p M$ with the tangent vectors $\frac{\partial}{\partial x_1}(p), \dots, \frac{\partial}{\partial x_n}(p)$. Working in the tangent space is very practical, as it allows to use the classical tools of linear algebra. However, it should be highlighted that this tangent space $T_p M$ is attached to p , and in general $T_p M \neq T_q M$ for $q \neq p$, unlike in the Euclidian space.

Definition 2.1.4 (Derivation) *A derivation D on a smooth manifold M at p is a linear map $D : C^\infty(M) \rightarrow \mathbb{R}$ which satisfies the Leibniz rule: $\forall (f, g) \in C^\infty(M), D(fg) = f(p)D(g) + D(f)g(p)$.*

Note that if $M \subset \mathbb{R}^n$ is an open subset of $\mathbb{R}^n$, and $p \in M$ then $T_p M = M$, which will be our case in 2.2.4.

Metrics & Push-Forward

Definition 2.1.5 (Riemannian Metric) *A Riemannian metric on a smooth manifold M is a continous function g of M in the set of symmetric positive-definite bilinear forms*

$\langle \cdot, \cdot \rangle_p$ of $T_p M$.

In other words, for each point p in M , g associates a scalar product of $T_p M$. For $(X, Y) \in (T_p M)^2$ we note $\langle X, Y \rangle = g_p(X, Y)$. This allows to generalize the euclidian scalar product to smooth manifolds, and thus define norms and angles on smooth manifolds. With $p \in M$ and $x = (x_1, \dots, x_n)$ a coordinate system around p , we can define (cf (Schiratti, Allasonniere, et al. 2015b)) a basis of the tangent space $T_p M$ and $g(p)$ writes

$$g(p) : (X, Y) \rightarrow \sum_{i,j} g_{i,j}(p)(X_i, Y_j)$$

$$g_{i,j}(p) = \left\langle \frac{\partial}{\partial x_i}(p), \frac{\partial}{\partial x_j}(p) \right\rangle \quad (2.4)$$

The existence of a Riemannian metric for a smooth manifold M is proven in ??.

Definition 2.1.6 (Riemannian Manifold) *We call a Riemannian manifold (M, g) a manifold M equipped with a Riemannian metric g on M .*

Definition 2.1.7 (Push-Forward) *Let (M_1, g) be a Riemannian manifold, M_2 be a smooth manifold and $f : M_1 \rightarrow M_2$ a diffeomorphism. The push-forward of the metric g^{M_1} on M_2 is the Riemannian metric f_*g defined for all p in M_2 by $(f_*g)_p(X, Y) = g_{f^{-1}(p)}^{M_1}(D_p(f^{-1}) \cdot X, D_p(f^{-1}) \cdot Y)$ for $(X, Y) \in T_p M^2$*

In other words, if a diffeomorphism exists between two smooth manifolds M_1 and M_2 , it can carry a Riemannian metric from one manifold to the other. The push-forward is a simple way to get a Riemannian metric when closed form of the diffeomorphisms are available, and we will use this method in 2.2.4.

Parallel Transport

Definition 2.1.8 (Parallel Transport) *Let $\gamma : t \mapsto \gamma(t)$ be a smooth curve on M and X a vector field along $\gamma(t)$. The vector field X is said to be parallel along $c(t)$ if $\nabla_{\gamma(t)}(X) = 0$.*

Proposition 2.1.1 (Unicity of Parallel Transport) *(Do Carmo Valero, 1992, Proposition 2:6, (Carmo 1992)). Let $\gamma :]0, 1[\rightarrow M$ be a smooth curve on M . Let $w_0 \in T_{\gamma(0)} M$. There exists a unique vector field $t \mapsto w(t)$ parallel among $\gamma(t)$ such that $w(0) = w_0$, written $P_{\gamma, t_0, t}(w)$.*

This unique tangent vector w_0 is "transported" along γ . We write the parallel transport of w_0 from time t_0 to time t along γ :

$$P_{\gamma, t_0, t}(w) : \begin{cases} I \rightarrow T_{\gamma(t)} M \\ t \mapsto w(t) \end{cases} \quad (2.5)$$

Note that ∇ refers to the Levi-Civita connection, which generalizes differentials to smooth manifolds. As such, it follows the Leibniz derivation rule (see Do Carmo (Carmo 1992) for details).

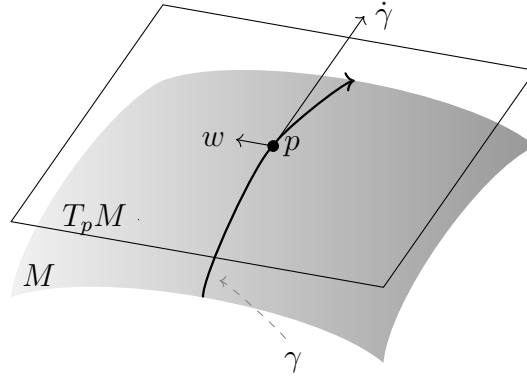


Figure 2.3 – Representation of a geodesic γ of M parameterized by its speed $\dot{\gamma}$ at position p . In addition we show a vector $w \in T_p M$.

Geodesics & Exponential

Definition 2.1.9 (Geodesics) A geodesic of M is a curve $\gamma : I \subset \mathbb{R} \rightarrow M$ with no acceleration : $\nabla_{\dot{\gamma}} \dot{\gamma} = 0$, i.e. whose tangent vectors $\dot{\gamma}$ remains parallel if transported along it.

The notion of geodesics is crucial in Riemannian Geometry. One can see geodesics as straight lines in a curved space, or equivalently being the shortest path between two points in this curved space.

Example 2.1.1 ($\mathbb{R}$) In the real line $\mathbb{R}$, geodesics are of the form $t \rightarrow at + b$

Example 2.1.2 ($M \subset \mathbb{R}$) With $M \subset \mathbb{R}$ and open interval of $\mathbb{R}$, the geodesics of M are of the form $f^{-1}(at + b)$ with $(a, b) \in \mathbb{R}^2$, $f : M \rightarrow f(M) \subset]0, +\infty[$ an increasing diffeomorphism (cf (Schiratti, Allasonniere, et al. 2015b))

Existence and unicity of a geodesic given a set of initial condition is shown in (Carmo 1992). For each position $p \in M$, and time $t_0 \in \mathbb{R}$, there exists an open set U in $T_p M$ such that, for each $v \in U$, there exists a unique geodesic $\gamma : I \rightarrow M$, defined on an open neighborhood $I \in \mathbb{R}$ of t_0 and such that

$$\begin{cases} \gamma(t_0) = p \\ \dot{\gamma}(t_0) = v \end{cases} \quad (2.6)$$

We can then write a geodesic with its associated Riemannian Exponential as follows.

Definition 2.1.10 (Riemannian Exponential) Proposition 2:7 (Carmo 1992) Let $p \in M$, $t_0 \in \mathbb{R}$ and $v \in U \subset T_p M$, the mapping

$$\text{Exp}_{p,t_0}(v)(\cdot) : \begin{cases} I \rightarrow M \\ t \mapsto \text{Exp}_{p,t_0}(v)(t) \end{cases} \quad (2.7)$$

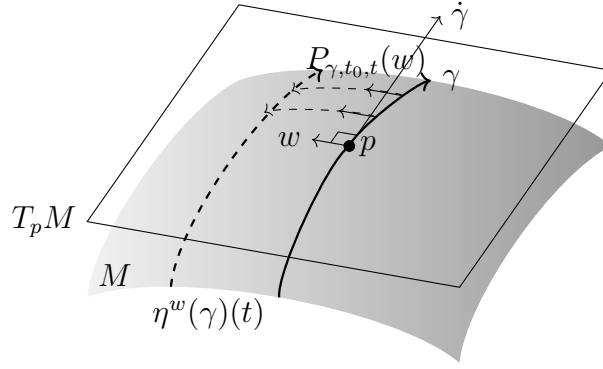


Figure 2.4 – Illustration of the construction of a parallel variation $\eta^w(\gamma)(t)$ of γ with regard to w . First, w is transported along γ , such that its parallel transport at time point t writes $P_{\gamma, t_0, t}(w)$, at position $\gamma(t)$. Then from $\gamma(t)$, we follow the Exponential map $\text{Exp}_{\gamma(t)}^M(P_{\gamma, t_0, t}(w))$ in the direction $P_{\gamma, t_0, t}(w)$ to build $\eta^w(\gamma)(t)$. The curve $t \mapsto \eta^w(\gamma, \cdot)$ is built from the values $\eta^w(\gamma)(t)$. Note that this curve is not necessarily a geodesic of M .

denotes the unique geodesic of M passing at point $p \in M$ at time $t_0 \in \mathbb{R}$ with velocity $v \in T_p M$.

A Riemannian exponential, parameterized by (p, t_0, v) d

Definition 2.1.11 (Riemannian Exponential) *The Exponential map, or Riemannian Exponential at $p \in M$ is the mapping*

$$\text{Exp}_p(v)(\cdot) : \begin{cases} I \rightarrow M \\ t \mapsto \text{Exp}_{p,0}(v)(1) \end{cases} \quad (2.8)$$

It appears as a special case of the previous definition, with $t_0 = 0$ and looking at the geodesic at $t = 1$.

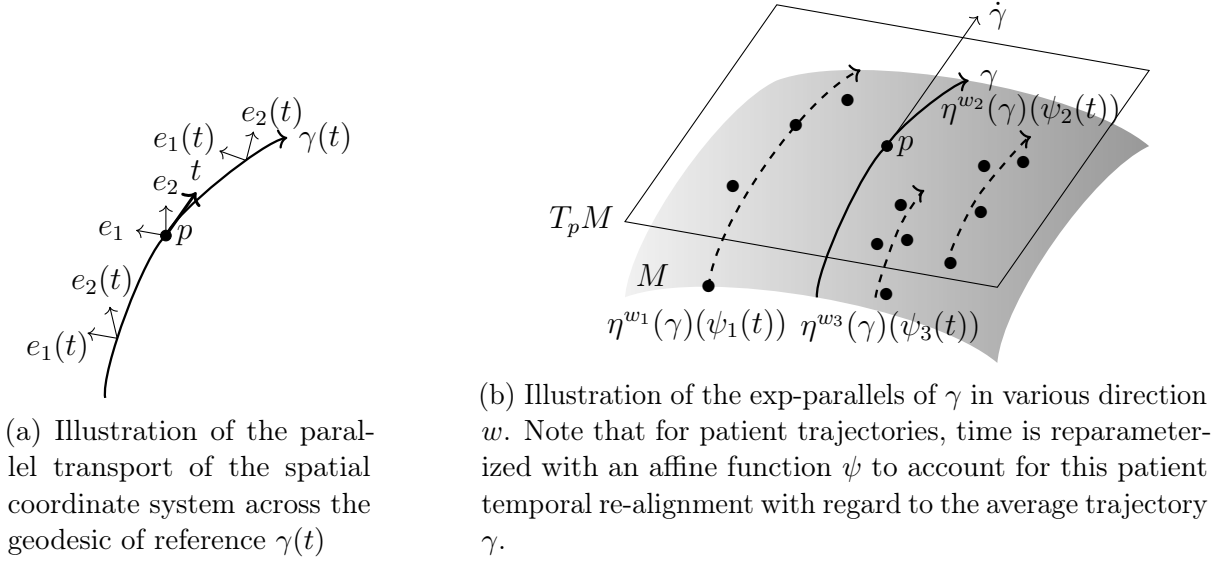
Parallel variation of a geodesic

The notion of variation of a differentiable curve γ on a manifold is defined in (Carmo 1992), Chapter 9. The idea consists in leveraging parallel transport of $w \in T_{\gamma(t_0)}M$ along γ . The parallel variation of γ in the direction of w writes:

$$\eta^w(\gamma, \cdot) \begin{cases} I \rightarrow M \\ t \mapsto \text{Exp}_{\gamma(t)}^M(P_{\gamma, t_0, t}(w)) \end{cases} \quad (2.9)$$

This notion of parallel variation of a geodesic is of particular interest in building a mixed effect model. Indeed, we will consider in the following that each patient i follow a curve $\eta^{w_i}(\gamma, \cdot) : t \mapsto \text{Exp}_{\gamma(t)}^M(P_{\gamma, t_0, t}(w))$ that is in this sense "parallel" to a geodesic of reference γ of M . Parameters (p, v, t_0) defining this geodesic will be used as fixed parameters, while individual parameter w_i will account for the random effect, parameterizing the parallel variation, that we call spatial variation, around γ .

In our use case, we will have a closed form of both the geodesics and their parallel transport. Note that in general this is not the case, and parallel transport can then be approximated using a numerical scheme called Schild's ladder (Lorenzi, Ayache, and Pennec 2011).

Figure 2.5 – A Spatiotemporal model in a Smooth Riemannian Manifold M .

2.2 Building a Riemannian Mixed Effect Model

Schiratti, Allasonniere, et al. 2015b propose to build a disease progression model leveraging the tools of Riemannian Geometry, assuming data $\mathcal{Y}$ lie in a n -dimensional Riemannian Manifold (M, g) . The model proposes to estimate an average trajectory as a geodesic $\gamma_{p,v,t_0}(t) = \text{Exp}_{p,t_0}(v)(t)$ of M , seen as the overall template of disease progression. Individual trajectories are modeled as parallel variations (see 2.1.4) from this average trajectory.

2.2.1 Model Geometry

Assume we know the average trajectory $\gamma_{p,v,t_0}(t) = \gamma(t)$. We aim at positioning visits of patients y with regard to this scenario. Let us define an orthonormal basis $\mathcal{B} = (e_0, \dots, e_n)$ of $T_p M$ with e_0 in the direction of $\dot{\gamma}(t) = v$. We can parallel transport the vectors of $\mathcal{B}$ along γ , that we call $\mathcal{B}(t)$. Parallel transport along γ preserves the orthogonality of $\mathcal{B}(t)$, with $e_0(t)$ parallel to $\dot{\gamma}(t)$ (see Figure 2.5 (a)). These local coordinates, adapted to the geodesic γ , are known as Fermi coordinates, and act similarly to cylindrical coordinates in euclidian spaces. With $t \in \mathbb{R}$, $w \in T_p M$ and $w \perp \dot{\gamma}(t_0) = v$, a patient trajectory write:

$$\eta_i^w(\gamma, t) = \text{Exp}_\gamma(P_{\gamma,t_0,t}(w))$$

where $P_{\gamma,t_0,t}(w)$ denotes the parallel transport of w along γ , and $\text{Exp}_\gamma(P_{\gamma,t_0,t}(w_i))$ the Riemannian exponential in direction $P_{\gamma,t_0,t}(w)$ at $\gamma(t)$. This decomposition can be seen as the orthogonal projection of y to γ , disentangling the temporal component t parallel to v from the spatial component w belonging to the hyperplane of v . The curves η , parallel variations of γ derived from the parallel transport of $w \in \text{Span}(v)^\perp$ are said to be exp-parallel to γ .

Note that for the spatio-temporal reference frame, t_0 does not have a particular meaning,

in the sense that the model could be defined for any other $t^* = t_0 + \delta t$

2.2.2 Statistical Description

Let $(\mathcal{Y}, \mathcal{T})$ be a longitudinal dataset, which is the collection of repeated measurements $(y_{i,j})$ for each visit j of patient $i = 1 \dots N$. Each measurement corresponds to a time $t_{i,j}$ and we assume that $y_{i,j} \in M$ with (M, g) a Riemannian Manifold. Measurements are assumed to be sample points from individual trajectories, which in turns are modeled as exp-parallelizations of a reference geodesic. Consequently, patients have a **constant spatial coordinate** w in the spatiotemporal reference frame. Let us note p, t_0, v the parameters of a reference geodesic γ of (M, g) , the model writes:

$$y_{i,j} = \eta^{w_i}(\gamma)(\psi_i(t)) + \epsilon_{i,j} \quad (2.10)$$

with

$$\begin{cases} \alpha_i = \exp(\xi_i) \\ \psi_i(t) = \alpha_i(t - \tau_i) + t_0 \\ w_i = A_{0,v^\perp} s_i \\ \epsilon_{i,j} \sim \mathcal{N}(0, \sigma) \\ \tau_i \sim \mathcal{N}(0, \sigma_\tau) \\ \xi_i \sim \mathcal{N}(0, \sigma_\xi) \\ s_i \sim \mathcal{N}(0, \mathbb{I}) \end{cases}$$

This model is a two stage hierarchical model, as individual trajectories η are assumed to be independant spatiotemporal transformations of the population geodesic γ . As explained in subsection 2.2.1, this modeling disentangles temporal variability and spatial variability.

- **temporal variability** $t \mapsto \psi(t)$ is a time-warping function that encodes the temporal variability of the individual patient trajectories. It affinely reparameterize a patient time index $t_{i,j}$ with an onset τ_i and a pace of progression α_i (see effects on Figure 2.7, columns 3 and 4), thus mapping $t_{i,j}$ to a time-point $\psi_{i,j}$ on the reference geodesic, that we will interpret as a disease stage. Note that this temporal re-alignment is of particular interest for neurodegenerative diseases, where age is not an accurate marker of disease stage.
- **spatial variability** w_i account for the spatial coordinate of patient i in the reference spatiotemporal frame. It encodes for η the exp-parallelization of γ , along which patient i progress (For an example of spatial variability in $]0, 1[^2$ see Figure 2.7 column 2). To ensure temporal and spatial disentanglement, necessary for the model identifiability, the condition $w_i \perp v$ is imposed. This is ensured by defining $w_i = A_{0,v^\perp} s_i$, with $A_{0,v^\perp}$ of rank $Ns < K$, and $\text{Span}(A_{0,v^\perp}) \in v^\perp$, with $s_i \in \mathbb{R}^{Ns}$. Specifically, the mixing matrix $A_{0,v^\perp} s_i$ is constructed as a linear combination of the basis vectors $\mathcal{B}^{-e_0}$, with $\mathcal{B}^{-e_0}$ referring to the basis $\mathcal{B} = (e_0, e_1, \dots, e_K)$ without its first direction e_0 parallel to v . Furthermore, we can set $Ns \ll K$ to perform an ICA on w_i , as few parameters might suffice to describe the spatial variability,

and to avoid overfitting. In practice, this spatial variability accounts for possibly different patterns of progression across patients, even after temporal-realignment. For example, this can take the form of some biomarkers being affected earlier, and other later than in the reference trajectory.

Note that the distribution of α_i is log-normal, in order to ensure positivity of the speeds of progression.

2.2.3 A mixed effect model

Individual parameters, which describe geometric variations from the average trajectory, are modeled as independant samples from normal distributions. Conveniently, the model falls within the scope of mixed effect model, with:

- Fixed effects $\theta = (p, t, v, \sigma_\tau, \sigma_\xi, \sigma, A_{0,v^\perp})$
- Random effects $z = (z_i)$, with $z_i = (\xi_i, \tau_i, s_i)$

In addition, the model is assumed to be bayesian to ensure existence of a MAP. Priors are set for the model parameters θ (see (Schiratti, Allasonniere, et al. 2015b) for details).

2.2.4 Instanciations

Once we have defined this statistical model, the question arise on which manifold do we assume that our measurements lie ? And which Riemannian metrics should we use on this manifold ?

We first present the simple one-dimensional manifolds that we use to build the model, which metrics are given by the associated push-forward (see 2.1). We then show the closed form of the geodesic of the product manifold of these 1D manifolds, which we will choose as our average trajectory γ . Random effects of the model, accounting for patient specific trajectories are modeled as variations around γ .

Property 2.2.1 (Geodesics of open intervals of $\mathbb{R}$) *Schiratti, Allasonniere, et al. 2015b.* Let $M \in \mathbb{R}$ an open interval of $\mathbb{R}$ and g a Riemannian metric on M . The geodesics of the one-dimensional Riemannian manifold (M, g) are of the form $t \rightarrow f^{-1}(at + b)$ with $a, b \in \mathbb{R}^2$, f an increasing diffeomorphism in C^1 .

In this section we present the one dimensional manifolds we use in the model. First, note that we can easily extract the closed form of the geodesics of 1D-Manifolds, provided the knowledge of a diffeomorphism from $\mathbb{R}$ to M (see 2.2.4). In addition, the knowledge of such diffeomorphism f yields a natural Riemannian metric, given by the push-forward f_*g^{eucl} (see section 2.1.4) of the euclidian metric g^{eucl} of $\mathbb{R}$ to M . Manifolds and their properties are details in Table 2.1.

The case $M = \mathbb{R}$

In this case f is the identity, and the scalar product is still euclidian, i.e. $g = g^{eucl}$. The geodesics write : $\gamma_{p,v,t_0} : \begin{cases} \mathbb{R} \rightarrow \mathbb{R} \\ t \rightarrow p + v(t - t_0) \end{cases}$

The case $M =]0, 1[$

With bounded observations, such as clinical scores on a specific scale, measurements can be normalized between 0 and 1. We can choose f as the logit function and $f^{-1} :$

$$\begin{cases} \mathbb{R} \rightarrow]0, 1[\\ p \rightarrow \ln\left(\frac{p}{1-p}\right) \end{cases} \quad \text{with differential } D_p(f^{-1}) : p \rightarrow \ln\left(\frac{1}{p(1-p)}\right).$$

The Riemannian metric is given by the definition of the pushforward

$$\begin{aligned} (f_*g^{eucl})_p(u, v) &= g_{f^{-1}(p)}(D_p(f^{-1})(u), D_p(f^{-1})(v)) \\ &= \langle D_p(f^{-1})(u), D_p(f^{-1})(v) \rangle_{eucl} \\ &= \left\langle \frac{u}{p(1-p)}, \frac{v}{p(1-p)} \right\rangle_{eucl} \\ &= \frac{1}{p^2(1-p)^2} \langle u, v \rangle_{eucl} \end{aligned}$$

Thus the metric induced by the push-forward writes

$$g_p^{]0,1[} : \begin{cases} T_p\mathbb{M} \times T_p\mathbb{M} \rightarrow \mathbb{R}^+ \\ (u, v) \mapsto g_p(u, v) = uG(p)v = u \frac{1}{p^2(1-p^2)} v \end{cases} \quad (2.11)$$

I	f	f^{-1}	$(f^{-1})'$	$G(p)$	γ_{p,v,t_0}
$\mathbb{R}$	p	p	1	1	$p + v(t - t_0)$
$\mathbb{R}^+$	$\exp(-p)$	$\ln(-p)$	$-\frac{1}{p}$	$\frac{1}{p^2}$	$p \exp(-\frac{v}{p}(t - t_0))$
$]0, 1[$	$\frac{1}{1+\exp(-p)}$	$\ln(\frac{p}{1-p})$	$\frac{1}{p(1-p)}$	$\frac{1}{p^2(1-p)^2}$	$(1 + (\frac{1}{p} - 1) \exp(\frac{-v(t-t_0)}{p(1-p)}))^{-1}$

Table 2.1 – Description of chosen 1D manifolds. f is the chosen transformation from $\mathbb{R}$ to the 1D-Manifold I , $(f^{-1})'$ refers to the pushforward of $\mathbb{R}$ to I , and $G(p)$ to the metric derived from this pushforward.

and the geodesics write

$$\gamma_{p,v,t_0} : \begin{cases} \mathbb{R} \rightarrow M =]0, 1[\\ t \mapsto (1 + (\frac{1}{p} - 1) \exp(\frac{-v(t-t_0)}{p(1-p)}))^{-1} \end{cases} \quad (2.12)$$

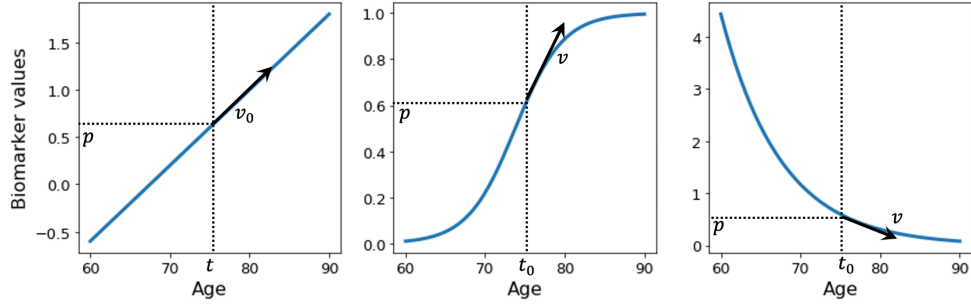


Figure 2.6 – Examples of 1D geodesics on the 1D-Manifolds used. On the left $\mathbb{R}$, on the middle $]0, 1[$, and on the right $\mathbb{R}^+$. This leads respectively to linear, sigmoid and exponential decay curves.

The case $M = \mathbb{R}^+$

Similarly to the previous case, we derive the metric and form for the geodesics using the diffeomorphism $p \rightarrow \exp(-p)$ (see Table 2.1).

Manifold Product of 1D Manifolds

In the previous section, we exhibited the geodesics of the chosen 1D manifolds, with parameter p, v, t_0 , which can be used to model the progression of one biomarker. To model the joint progression of K biomarkers at a time, Schiratti, Allasonniere, et al. 2015b propose to consider the product manifold $\mathcal{M} = M_1 \times M_k \times \dots \times M_K$ of 1D manifolds. This is practical, as the tangent space of this manifold is the cartesian product of the tangent spaces, and the metric is simply the product metric of these manifolds. This

Riemannian metric is diagonal, and can be easily computed as the sum of the 1D manifolds metrics : $g(u, v) = \sum g_i(u_i, v_i)$ with $(u_i, v_i) \in M_i^2$. The geodesics of the product manifold write $t \rightarrow \gamma(t) = (\gamma_1(t), \gamma_2(t) \dots \gamma_K(t))$, with parameter $p = (p_1, \dots, p_K)$, $v = (v_1, \dots, v_K)$ and t_0 . Note that t_0 is shared between all the geodesics in the 1D manifolds to ensure identifiability.

Spatial Variability in the Manifold Product

Schiratti, Allasonniere, et al. 2015b assume that all patient progress in the same direction in the manifold. To this end, the concept of parallel variations in the manifold is used (see section 2.1.4): Each patient i is assumed to follow a curve, parallel of the average geodesic γ_{p,v,t_0} . Schiratti, Allasonniere, et al. 2015b (IV.2.2.2 *Products of one-dimensional manifolds*) show that in the simple case of a product of 1D manifolds, this parallel curve, parameterized by $w_i \in T_p M$ writes

$$\eta_k^{w_i}(t) = \gamma_k\left(\frac{w_{i,k}}{\dot{\gamma}_k(t_0)} + t\right)$$

An example of spatial variability in the case of scalar data can be visualized in Figure 2.7, row 1 column 2.

Basis of $T_p M$ and Model Identifiability

Importantly, to ensure identifiability of the model, temporal and spatial variabilities have to be disentangled. Space shifts are thus required to meet the orthogonality condition : $g_p^M(w_i, v) = 0 \forall i \in [1, N]$, i.e. $w_i \in \text{Span}(\dot{\gamma}(t_0))^\perp$, the $N-1$ dimensional subspace of $T_p M$ orthogonal to v (in the sense) of the metric g_p .

In practice, we use the Householder method to build a basis $\mathbb{B} = (v, B_1, \dots, B_K)$ of $T_p M$. Note that this basis will be updated at each optimization step, as it depends on p and v . In addition, Schiratti, Allasonniere, et al. 2015b propose to perform an Independent Component Analysis (ICA) on the w , $w_i = A s_i$. Columns of the ICA matrix A will then be chosen as linear combinations of $(B_1, \dots, B_K)$ to respect the orthogonality condition. At the same time, the ICA offer a way to perform dimensionality reduction to avoid overfitting.

Individual patient trajectories

The trajectory of patient i writes

$$\eta_k^{w_i}(t) = \gamma_k\left(\frac{w_{i,k}}{\dot{\gamma}_k(t_0)} + t\right)$$

We reparametrize patient i time with the affine function $\psi_i : t \rightarrow \alpha_i(t - \tau_i)$, so that $\eta_k^{w_i}(t)$ is now traveled at time $\psi_i(t) = \alpha_i(t - \tau_i)$

$$\eta_k^{w_i}(\psi_i(t)) = \gamma_k\left(\frac{w_{i,k}}{\dot{\gamma}_k(t_0)} + \psi_i(t)\right)$$

For the aboved mentioned manifold product of 1D manifolds, the model writes for patient i seen at visit j on biomarker k writes

$$\left\{ \begin{array}{l} \mathbb{R} : (y_{i,j})_k = w_{i,k} + p_k + \alpha_i(t - \tau_i - t_0) + (\epsilon_{i,j})_k \\]0, 1[: (y_{i,j})_k = (1 + (\frac{1}{p_k} - 1) \exp^{(-\frac{v_k \alpha_i(t_{i,j} - t_0 - \tau_i) + (w_i)_k}{p_k(1-p_k)})})^{-1} + (\epsilon_{i,j})_k \\ \mathbb{R}^+ : (y_{i,j})_k = p_k \exp(-\frac{w_{i,k}}{p_k} - \frac{v_k \alpha_i}{p_k} t_{i,j} - t_0 - \tau_i) + (\epsilon_{i,j})_k \end{array} \right.$$

Note that in the case $K = 1$, the manifold is a 1D manifold, and the geodesic γ is a 1D geodesic. We have $w_i = 0 \forall i \in [1, N]$, as $Span(\dot{\gamma}(t_0))^\perp = \{0\}$.

2.2.5 Example on $]0, 1[^2$

To illustrate the method, we consider a small example in $M =]0, 1[^2$, with the metric induced by the push-forward of the logit transformation: $G(x, y) = \frac{1}{x^2(1-x)^2} + \frac{1}{y^2(1-y)^2}$.

Let $(\mathcal{Y}, \mathcal{T})$ be a longitudinal dataset, with at time $t_{i,j}$, observation $y_{i,j} \in]0, 1[^2$. As presented in section 2.2.4, we consider geodesics of $M =]0, 1[^2$ of the form a family of two sigmoid functions (γ_1, γ_2) . In this simple case, $T_p M = \mathbb{R}^2$, $\forall p \in M$.

We show on the left column of Figure 2.7 a geodesic of $]0, 1[^2$ with parameters $t_0 = 70$, $p = (0.8, 0.25)$, $v = (0.1, 0.1)$, both in the data space (extrinsic manifold view, on the top), and in the manifold $]0, 1[^2$ on the bottom.

Then, we show the geometric transformation encoded by individual parameters of a patient i , which longitudinal measurements are shown on the top right plot. In the second column, we show the effect of $w_i = (-0.2, 0.05)$ which encode the spatial variation of the average geodesic, to build the exp-parallelization $\eta^{w_i}(\gamma)$. In the third column, the onset $\tau_i = 3$ shifts the time at which $\eta^{w_i}(\gamma)$ is seen. It amounts to a delay of all the biomarker curves at once, as shown on the first row, third column. Then, a speed $\alpha_i = 0.5$ modify the speed at which $\eta^{w_i}(\gamma)$ is travelled. In this case, with $\alpha_i < 1$, progress is slower than the average, and a smaller portion of $\eta^{w_i}(\gamma)$ is seen.

2.2.6 Remarks

Input Data

Typically, the model, implemented in Leaspy (Couronne R n.d.) is run on longitudinal cohorts with number of patients ranging from a few dozens to a few thousands, with each having between 3 and 12 visits, on a dozen of different modalities. Another important factor is the duration on which patients are seen, in regard to the assumed total duration of their disease. The more each patient snapshots overlap with one another in terms of disease stage, the better the estimation. Biomarker values of patients across time are assumed to be monotonous. We handle increasing biomarker, and decreasing ones are inverted. To use the sigmoid model, we typically perform a min-max normalization, using theoretical minimum and maximum of biomarkers if available.

Choosing the 1D manifolds

With the product manifold, the model assumes one shape of curve per modality. This requires a prior knowledge on the data to support this choice. Alternatively, one could perform an exhaustive cross-validation of all possible choices of curves, according to some performance metric such as the reconstruction error. However, this problem is combinatorial, and in practice we choose the shape a-priori.

Setting the number of source

The number of sources is a parameter controlling the flexibility of the model, and thus its potential over fitting. Setting no source leads to a very rigid model with no spatial variability, and only a temporal variability. On the contrary, setting $K-1$ sources saturates

$\text{Span}(\dot{\gamma}(t_0))^\perp$, and might lead to overfitting if there is too few visit per subject. This parameter thus needs to be chosen accordingly for the problem at hand, or could be cross validated to get the best trade-off between reconstruction and number of parameters.

Basis of TpM and mixing matrix A

The basis of TpM is dependant of p and v , and is estimated from scratch at each update of either p or v (e.g. via the Householder method). In the MCMC-SAEM estimation, Schiratti, Allasonniere, et al. 2015b assume that the basis varies continuously with p and v , so that $A_{0,v^\perp}$, which is built as a linear combination of elements (e_k) of basis $\mathcal{B}^{-e_0}$, namely the (β_k) , also varies continuously with p and v .

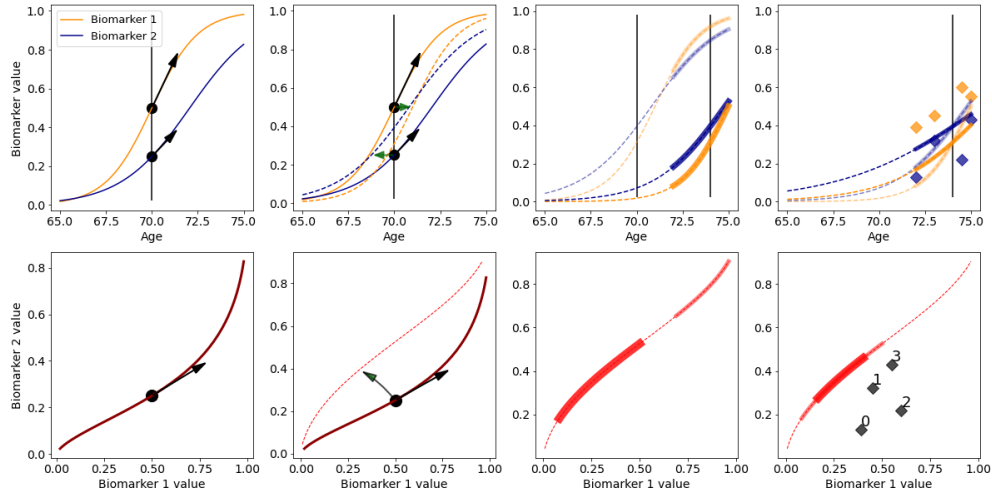


Figure 2.7 – Description of the model on geodesics of $]0, 1[^2$ equipped with the induced metric. On the top row, biomarker model of progression with regard to time. On the bottom, intrinsic view of M with geodesics as parametric curves of $\mathbb{R}^2$. On the first column, we show the average trajectory as a geodesic of M , parameterized by p, v, t_0 . On the second column, visualization of a spatial variability via the exp-parallelization of γ in the direction $w_i = (-0.2, 0.05)$. On the two last columns, effects of temporal variability. On the third column, effect of the onset individual parameter $\tau_i = 3$, which shifts the curves in time. This spatial variability shifts biomarker curves in time, while ensuring that this "spatial" variation is orthogonal to the speed v at p ("temporal" variability) in order to preserve identifiability of the model. On the last column, effect of α_i which dilates the curves from the reference time $t_0 + \tau_i$. In diamond, patient i measurements that we aim to fit with these successive transformations from the average trajectory.

The Propagation model

In (Schiratti, Allasonniere, et al. 2015b) they chose to consider the propagation model, a particular case where all speeds v_k are assumed to be the same. At the cost of model expressivity, this allow to parametrize the average trajectory via only a (shared) speed of progression, the position of a biomarker of reference, and for each other biomarkers, a time delay with regard to the biomarker of reference.

extract time deltas between modalities.

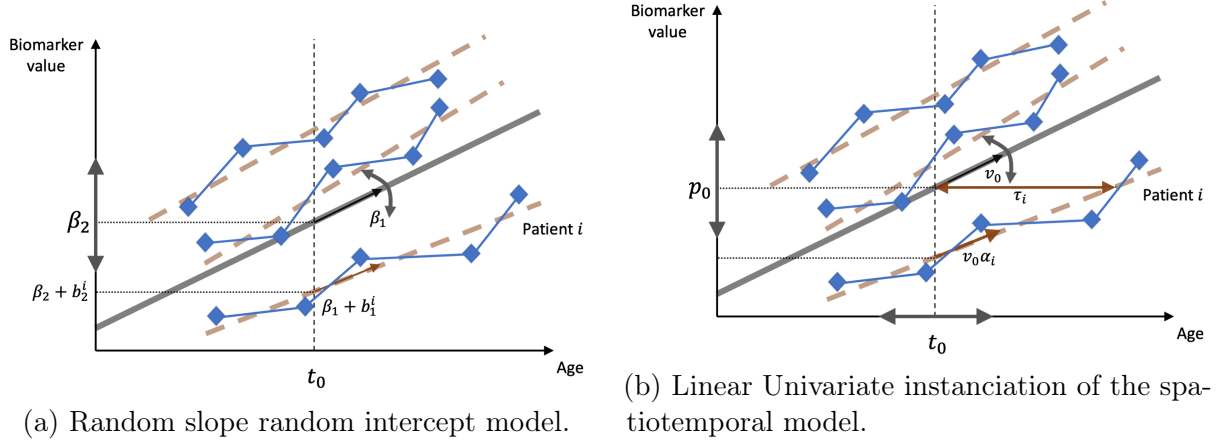


Figure 2.8 – Comparison between a random slope random intercept model $y_{i,j} = (t_{i,j} - t_0)(\beta_1 + b_{1i}) + (\beta_2 + b_{2i}) + \epsilon_{i,j}$ and a linear univariate instantiation of the spatiotemporal model. Note that the spatiotemporal model is not linear in the sense of parameters as there is a non-linear effect between speeds of progression α_i and onset τ_i : $y_{i,j} = p_i + v(\alpha_i(t - t_0 - \tau_i)) + \epsilon_{i,j}$.

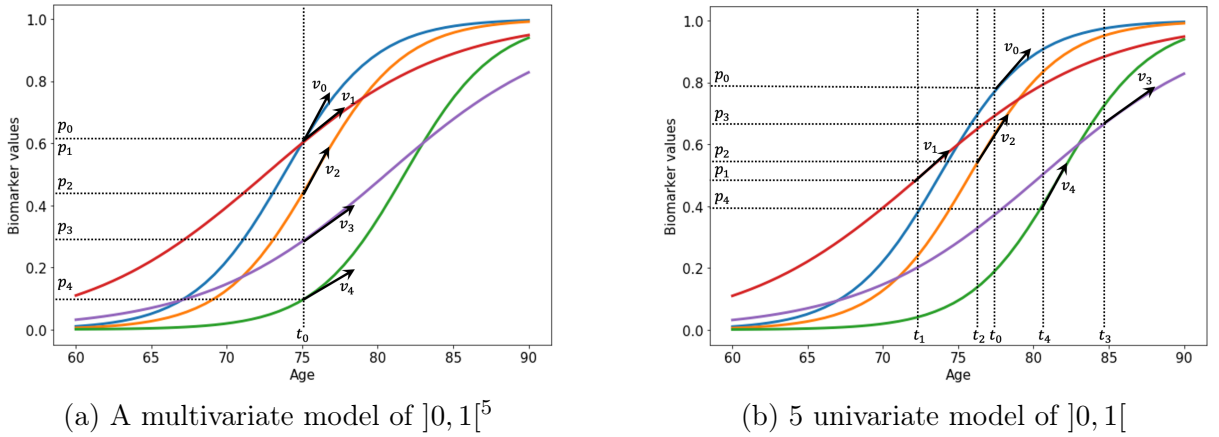


Figure 2.9 – Illustration of the difference of parameterization between a multivariate model of $]0, 1[^5$ and 5 univariate model of $]0, 1[$. The different univariate models have different timelines, with each additional " t_0 " reference time parameters. In addition, although not shown here, each univariate model will have τ_i and ξ_i parameters, leading to 10 individual parameters. On the other hand, the multivariate model will have τ_i and ξ_i , as well as sources of dimension $q < K$, therefore having at maximum $\dim = 2 + (5 - 1) = 6$.

2.3 Estimation

2.3.1 Statistical Tasks

Given a longitudinal dataset $\mathcal{Y} = \{y_{i,j}\}$, $\mathcal{T} = \{t_{i,j}\}$, we distinguish the following estimation tasks:

- **Calibration:** Estimate the MAP of θ , given a calibration dataset $\mathcal{Y}$ and $\mathcal{T}$. We use the MCMC-SAEM (Kuhn and Lavielle 2004) algorithm.

$$\theta^{cal} = \operatorname{argmax}_{\theta} \int_z p(\mathcal{D}, z, \theta;) dz$$

- **Personalization:** Estimate the MAP of z , given a personalization dataset $\mathcal{Y}$, $\mathcal{T}$ and θ . Note that the dataset may change between calibration and personalization, for instance for replication purposes.

$$z = \operatorname{argmax}_z p(z|\mathcal{D}; \theta^{cal})$$

- **Simulation:** Simulate $\mathcal{Y}^{syn}$, $\mathcal{T}^{syn}$ given a real dataset $\mathcal{Y}$ and $\mathcal{T}$ in input. The synthetic dataset should mimic the geometrical variability of the real dataset, encoded by the synthetic random effects z^{syn} :

$$z^{syn} \sim \mathcal{N}(\bar{z}, \sigma_z)$$

2.3.2 Computation of the likelihood

Knowing $\mathcal{D}, z, \theta$, the likelihood writes $p(\mathcal{D}, z, \theta) = p(\mathcal{D}|z, \theta)p(z, \theta)$ so that

$$L = \log p(\mathcal{D}, z, \theta) = \log p(\mathcal{D}|z, \theta) + \log p(z, \theta) = L_{fit} + L_{reg}$$

with

$$L_{fit} = \sum_{i=0}^p \sum_{j=0}^{m_i} \sum_{k=1}^K \|y_{i,j,k} - (\eta^{w_i}(\gamma)(\psi_i(t_{i,j})))_k\|^2 \quad (2.13)$$

and L_{reg} encompassing both random effect regularizations $\log p(z|\theta)$, as well as bayesian priors $\log p(\theta)$ (not taken into account in the sampling step).

2.3.3 Calibration

Rational for MCMC-SAEM

We aim at maximizing the likelihood $p(\mathcal{D}, \theta) = \int_z p(\mathcal{D}, \theta, z) dz$. As the intermediate variable z is not observed, we may want to use EM base methods, that iteratively compute the form of the expectation $\int_z \log(p(\mathcal{D}, z, \theta|\theta^{(k)}))$ and then update θ to maximize this expectation, until convergence. However, in our case there is no closed form of the

expectation $\int_z \log(p(\mathcal{D}, z, \theta | \theta^{(k)}))$ due to the non-linear relationship between x and z . Therefore, SAEM (Delyon, Lavielle, and Moulines 1999) propose to replace the expectation by a stochastic approximation $z \sim p(z | \mathcal{D}, \theta^{(t)})$. If this approximation is also not available in closed form, which is the case in our model, (Kuhn and Lavielle 2004) propose the MCMC-SAEM. Using the Bayes rule yields

$$p(z | \mathcal{D}; \theta^{(k)}) = \frac{p(\mathcal{D} | z; \theta^{(k)})p(z; \theta^{(k)})}{p(\mathcal{D}; \theta^{(k)})}$$

and $p(\mathcal{D}, z | \theta^{(k)})$ is known up to the normalizing constant $p(\mathcal{D}; \theta^{(k)})$. Thus, MCMC-SAEM propose to build a Markov Chain using the Metropolis Hastings procedure to sample $z^{(k+1)} \sim p(z^{(k)} | \mathcal{D}, \theta^{(k)})$.

Appartenance to the curved exponential family

MCMC-SAEM theoretical convergence is proven in (Allasonnière, Kuhn, and Trouvé 2010; Kuhn and Lavielle 2004) for models which belong to the curved exponential family, i.e. models such that:

$$\log(\mathcal{D}, z, \theta) = \langle \psi(\theta), S(\mathcal{D}, z) \rangle - \log C(\theta)$$

where ψ and C are smooth functions of the parameters and S is a function of the observations and latent variables called sufficient statistics. Under its current form, the model does not belong to this family, and we perform an "exponentialization" trick (see (Schiratti, Allasonniere, et al. 2015b)) so that it becomes the case, injecting θ parameters in z . Thus $z_{pop} = (p, t_0, v, \beta)$ (β referring to the coefficients of $\mathcal{B}$ to get the matrix $A_{0,v^\perp}$) will be treated like random variables, gaussian with a small, fixed variance around their associated fixed effect: $t_0 \sim \mathcal{N}(\bar{p}, \sigma_p)$, $p \sim \mathcal{N}(\bar{p}, \sigma_p)$, $v \sim \mathcal{N}(\bar{v}, \sigma_v)$, $\beta \sim \mathcal{N}(\bar{\beta}, \sigma_\beta)$.

In the MCMC-SAEM, these random variables will be sampled in the expectation step via Gibbs sampling in the same manner that individual random variables $z_{ind} = (\tau, \xi, s)$.

The other fixed effects $(\sigma_\tau, \sigma_\xi, \sigma)$ are updated in closed form from the sufficient statistics in the maximization step.

Bayesian priors

In theory, to ensure the existence of a MAP, the model is assumed to be Bayesian, posing a-priori distributions for the model parameter θ , adding a regularization term in the likelihood. In practice, the a-priori variance can be large, and in our Leaspy implementation we do not include them.

MCMC-SAEM algorithm

Algorithm 1 : MCMCSAEM

input : $\theta^{(0)}, z^{(k-1)} = (z_1^{(k-1)}, \dots, z_L^{(k-1)})$ with L variable blocks, initial parameters
 $\theta^{(0)}, z^{(0)}$

output : $\theta^{(k)}$

Expectation Step;

pour chaque Variable block b faire

Let $z = (z_1^{(k)}, \dots, z_{b-1}^{(k)}, z_b^{(k-1)}, z_{b+1}^{(k-1)}, \dots)$;

Draw candidate $z^b \sim \mathcal{N}(z_b^{(k-1)}, \sigma_z^b)$ for block b ; // Gibbs Sampling

Let $z^* = (z_1^{(k)}, \dots, z_{b-1}^{(k)}, z_b, z_{b+1}^{(k-1)}, \dots)$;

Compute the likelihood ratio $\alpha = \frac{p(\mathcal{D}|z^*, \theta^{(k-1)}, t)p(z^*|\theta^{(k-1)})}{p(\mathcal{D}|z, \theta^{(k-1)}, t)p(z|\theta^{(k-1)})}$;

Draw $u \sim \mathcal{U}(0, 1)$; // Metropolis Hastings

si $u > \alpha$ **alors**

$z_b^{(k)} \leftarrow z_b$;

sinon

$z_b^{(k)} \leftarrow z_b^{(k-1)}$;

fin

fin

Adapt proposal variances ; // Aim for 30% (Atchadé 2006)

Stochastic Approximation Step

Compute sufficient statistics $S(\mathcal{D}, z^{(k)})$;

$S^{(k)} \leftarrow S^{(k)} + \epsilon^{(k)} S(\mathcal{D}, z^{(k)} - S^{(k-1)})$;

Maximization Step

$\theta^{(k+1)} = \theta(S^{(k)})$

Heuristics for initialization

Smart initialization may greatly speed up convergence of the MCMC-SAEM algorithm, especially for the fixed effects of the model. For scalar data, we use intuitive initialization heuristics. We choose t_0 as the average time $t_{i,j}$ of all visits in the dataset. Then we perform linear regression on each biomarker of each individual with respect to intercept and time $t - t_0$ to obtain slopes $a_{i,k}$ and intercepts $b_{i,k}$. We choose biomarker k 's position p_k as the average of intercepts $b_{i,k}$ and biomarker k 's speed v_k as the average of slopes $a_{i,k}$. The mixing matrix A is initialized at random in the Span of $v = (v_0, \dots, v_K)$ in $T_p M$. Individual parameters are initialized at random from the estimated variance of onsets and speeds only. We set σ_{tau} as the standard deviation of times $t_{i,j}$, and $\sigma_{xi}=1$. We generate from gaussian distributions $\tau \sim \mathcal{N}(0, \sigma_{tau})$, $\xi \sim \mathcal{N}(0, \sigma_{xi})$, and $s \sim \mathcal{N}(0, 1)$.

Heuristics during optimization: Tempering Scheme

We use tempering in order to speed up the optimization process. We use a deacreasing linear tempering profile, from $T = 3$ during the n_1 first iterations, and then $T = 1$ during the $n_{iter} - n_1$ last iterations.

2.3.4 Personalization

In the personalization step, we estimate random effects with the knowledge of a calibrated model $\hat{\theta}$. Let $\mathcal{D} = (y_{i,j}, t_{i,j})$ be a longitudinal dataset of patients, seen or unseen in the calibration procedure. For each patient i , we aim at estimating $z_i = \operatorname{argmax}_{z_i} p(\hat{\theta}, z_i; \{y_{i,j}\}, \{t_{i,j}\}) = p(\{y_{i,j}\}, \{t_{i,j}\} | z_i, \hat{\theta}) p(z_i | \hat{\theta})$, $j \in [1, m_i]$. In practice, we could use the MCMC-SAEM by sampling only z_i 's, but rather rely on classical optimization procedure such as L-BFGS or Powell's method for simplicity. In addition, we also use optimization procedure informed with the gradient of $p(\hat{\theta}, z_i; \{y_{i,j}\}, \{t_{i,j}\})$, obtained via automatic differentiation, to speed up the computations. Note that as in standard linear mixed effect models, the more visits we have for a patient, the more we weight its attachment term $p(\{y_{i,j}\}, \{t_{i,j}\} | z_i, \hat{\theta})$ against its regularization term $p(z_i | \hat{\theta})$.

2.3.5 Simulation

In the calibration step, we estimate a spatiotemporal reference frame, which encodes an average trajectory $\gamma(t)$, as well as the forms of the geometrical transformations $\eta^{w_i}(\gamma)(\psi_i(t))$ of $\gamma(t)$, encoded in individual parameters z_i . We also estimate the distributions of these random effects z_i , namely σ_τ , σ_ξ , σ_s . We can then draw sample from this learned distribution to generate synthetic patient's individual parameters $z_i^{synthetic}$. Coupled with a deterministic scheme, or a distribution for time points $t_{i,j}$, we can generate visits $y_{i,j}^{synthetic}$ for these synthetic patients.

In low dimension, the method produces quite realistic samples (cf 8.5.3). However, note that biases naturally occur in longitudinal studies (see 1.3.3), which harm the assumption that all individual variables are independant, and that they are independant from the distrubution of times $t_{i,j}$. Also, note that we could repeat the process by learning possibly more complex distributions of z_i from personalized z_i directly, such as with a kernel density estimation, or gaussian mixtures.

Part II

PD progression models with low
dimensional un-structured data

Learning disease progression models with longitudinal data and missing values

Outline of the current chapter

3.1 Introduction	72
3.2 Methods	73
3.2.1 The general model	73
3.2.2 Manifold Product for multivariate Data	74
3.2.3 Missing Data	74
3.2.4 Estimation	75
3.3 Results	75
3.3.1 Experiment methodology	75
3.3.2 Synthetic Data	76
3.3.3 Real Data	76
3.3.4 Application to PPMI	77
3.4 conclusion	78

In this chapter, we relax some of the assumptions of the spatiotemporal propagation model of (Schiratti, Allasonniere, et al. 2015b) for scalar data. Motivated by an application to Parkinson’s Disease clinical markers, we allow for different curves and speeds of progression per biomarkers to be considered. Additionally, we propose to handle missing data occurring by design in the PPMI cohort by leveraging the generative aspect of our model, and experimentally assess the model’s resilience to missing values. This work has been accepted for publication in the proceedings of the 2019 International Symposium on Biomedical Imaging (ISBI).

Abstract

Statistical methods have been developed for the analysis of longitudinal data in neurodegenerative diseases. To cope with the lack of temporal markers - i.e. to account for subject-specific disease progression in regard to age - a common strategy consists in realigning the individual sequence data in time. Patient's specific trajectories can indeed be seen as spatiotemporal perturbations of the same normative disease trajectory. However, these models do not easily allow one to account for multimodal data, which more than often include missing values. Indeed, it is rare that imaging and clinical examinations for instance are performed at the same frequency in clinical protocols. Multimodal models also need to allow a different profile of progression for data with different structure and representation.

We propose to use a generative mixed effect model that considers the progression trajectories as curves on a Riemannian Manifold. We use the concept of product manifold to handle multimodal data, and leverage the generative aspect of our model to handle missing values. We assess the robustness of our methods toward missing values frequency on both synthetic and real data. Finally we apply our model on a real-world dataset to model Parkinson's disease progression from data derived from clinical examination and imaging.

3.1 Introduction

Linear mixed effect model estimated via EM have been introduced for the analysis of longitudinal data (Laird and Ware 1982b), and later were extended for more flexibility to the non-linear (Lindstrom and Douglas M. Bates 1990b) case. Well adapted with an objective time (e.g. relative to an event), they are less adapted to data that do not include such consistent time event, such as neurodegenerative disease progression. In (Durrleman et al. 2012; Hong et al. 2014) the concept of Time Warps is introduced to account for age variability at onset, and in (Lorenzi, Pennec, et al. 2015) a morphological age-shift. However these Time Shifts are not estimated in the context of a statistical model. Generalization of LME to Riemannian manifolds were proposed (Singh et al. 2013, 2014), that allows to consider features defined by smooth constraints, such as images or mesh (Bone, Colliot, and Durrleman 2018). In (Schiratti, Allasonniere, et al. 2015b), a generic spatio-temporal model is introduced in the Bayesian framework, modeling the course of biomarker's progression as a geodesic, as well as individual variations via parallel transport, travelled at subject-specific onset and speed with an affine time reparametrization. Although this approach allows multivariate data, it assumes the same profile of progression (e.g. linear, logistic, exponential, etc..) for all coordinates, and does not account for missing values, leading to the removal of all visits with at least one missing values from the analysis, or to the use of ad-hoc data imputation procedure. This can be problematic for multimodal data where missing values (denoted NAs) occur by design of the experiment. We propose to build on this model to extend its application range, assuming missingness is unrelated to the data (Missing Completely at Random) (Ibrahim and Molenberghs 2009). We allow the modeling of the joint progression of features that are assumed to offer different

evolution profile, and handle the missing values in the context of a generative model.

3.2 Methods

3.2.1 The general model

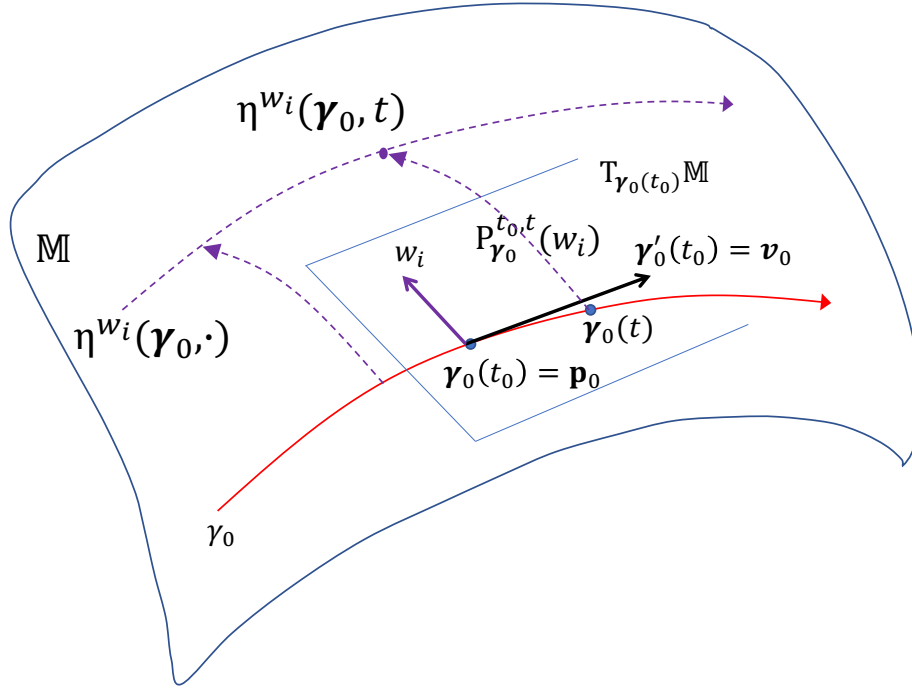


Figure 3.1 – Representation of the model on a schematic manifold. Each point $\eta^{w_i}(\gamma_0, t)$ is obtained via the continuous transportation of the tangent vector w_i along the reference geodesic γ_0 (in red). $\eta^{w_i}(\gamma_0, \cdot)$ (in purple) is a "parallel" of γ_0 . The model for patient i consists then in the trajectory $\eta^{w_i}(\gamma_0, \cdot)$ travelled at the subject-specific time $\psi_i(t)$.

In (Schiratti, Allasonniere, et al. 2015c) each data point is seen as a point in a Riemannian manifold, denoted $y_{i,j}$, observation of the i -th subject at its j -th visit. These points are then considered as noisy samples along an individual trajectory, namely a curve on the manifold, which in turn is seen as a random spatiotemporal transformation of a reference geodesic on the manifold. The model $y_{i,j}$ projected on the k th modality can be written as:

$$(y_{i,j})_k = (\eta^{w_i}(\gamma_0)(\psi_i(t_{i,j})))_k + (\varepsilon_{i,j})_k \quad (3.1)$$

where

- $\gamma_0 : t \rightarrow \text{Exp}_{p_0}^{t_0, t}(v_0)$ is the population average trajectory in the form a the geodesic

passing at point p_0 with velocity v_0 at time t_0 (Exp denotes the Riemannian exponential as a concise way to write geodesics),

- $\eta^{w_i}(\gamma_0) : t \rightarrow \text{ExpP}_{\gamma_0}^{t_0,t}(w_i) = \text{ExpP}_{\gamma_0}(P_{\gamma_0}^{t_0,t}(w_i))$ is the exp-parallelisation of the geodesic γ_0 in the subject-specific direction w_i , called space-shift, as depicted in Fig 3.1. ($P_{\gamma_0}^{t_0,t}(w_i)$ denotes the parallel transport of the vector w_i along the curve γ_0 from $\gamma_0(t_0)$ to $\gamma_0(t)$),
- $\psi_i : t \rightarrow \alpha_i(t - t_0 - \tau_i) + t_0$ is a time-reparameterizing function, where α_i is a subject-specific acceleration factor and τ_i a subject-specific time-shift.

η^{w_i} and ψ_i define a spatiotemporal transformation of the average trajectory. To assure a unique decomposition through both the spatial and temporal transformation, the w_i are chosen orthogonal to $\mathbf{v}_0$ in the tangent space at $\mathbf{p}_0$. A spatiotemporal transformation of the reference trajectory to the i th patient trajectory is then parametrized by the individual parameters τ_i , α_i and w_i . A time-shift τ_i represents the delay at onset relative to t_0 for the individual i , to distinguish between individuals with early or late onset. The α_i models the speed at which the trajectory of individual i is travelled. Then the space-shifts w_i accounts for variations in position of the individual trajectory, and model difference in patterns of disease progression between individuals. Normal distributions are chosen as priors for τ_i , w_i and ξ_i with $\alpha_i = \exp(\xi_i)$. These parameters are the random effects of the model, whereas γ_0 is the fixed effect, parametrized by $\mathbf{p}_0$, $\mathbf{v}_0$ and t_0 .

3.2.2 Manifold Product for multivariate Data

Dealing with a longitudinal and multimodal dataset, we wish to analyze at once the temporal progression of a family of N features, with possibly different evolution profile. Thus at the difference of (Schiratti, Allasonniere, et al. 2015c) we consider manifold product that are not necessarily the product of the same univariate manifold M . Each feature k is described by repeated univariate observations $y_{i,j,k}$ on M_k that are considered as random perturbations along each trajectory. For each feature we prescribe a user-defined profile of progression (e.g. straight line, exponential decay, logistic). Ignoring missing values at the moment, each individual observations can be represented as a N -dimensional vector $(y_{i,j})_{1 \leq i \leq p}$, that is considered as random perturbation of quantities lying on the product manifold $\mathbb{M} = M_1 \times M_2 \times \dots \times M_N$ equipped with the product metric. The product manifold gives geodesics of the form $\{\gamma : t \in \mathbb{R} \rightarrow (\gamma_1(t), \gamma_2(t), \dots, \gamma_N(t))\}$ on $\mathbb{M} = M_1 \times M_2 \times \dots \times M_N$ equipped with the product metric. γ_k is the (univariate) geodesic which goes through point $p_k \in M_k$ at time t_0 and velocity v_k .

3.2.3 Missing Data

When missing values occur at time $t_{i,j}$, only a subset of $y_{i,j}$ if visible, we note $m_{i,j}$ these modalities. We decide to handle missing data in the context of our generative model, and compute the likelihood with visible data. The goodness of fit at a given visit $v_{i,j}$, at time $t_{i,j}$ for the k th modality writes $\|y_{i,j,k} - (\eta^{w_i}(\gamma_0)(\psi_i(t_{i,j})))_k\|^2$, and for the entire goodness of fit :

$$\mathcal{L}_{fit} = \sum_{i=0}^p \sum_{j=0}^{l_i} \sum_{k \in m_{i,j}} \|y_{i,j,k} - (\eta^{w_i}(\gamma_0)(\psi_i(t_{i,j})))_k\|^2 \quad (3.2)$$

with l_i the number of visits of the i th patient. We see in Eq 3.2 that the likelihood is informed only by available data while taking into account all the information available and without imputing missing values with ad-hoc procedures.

3.2.4 Estimation

Estimation of model parameters is done via the use of a stochastic version of the Expectation-Maximization Algorithm, namely the MCMC-SAEM algorithm (Kuhn and Lavielle 2004), that seeks to maximize the likelihood $\mathcal{L} = \mathcal{L}_{fit} + \mathcal{L}_{prior}$. MCMC-SAEM iterates in 3 steps : simulation, approximation and maximization. It simulates first candidate individual variables, that are then accepted or rejected according to a probability function of the likelihood ratio. Then sufficient statistics are extracted from the current variables. Finally the current estimates of the parameters are maximized.

3.3 Results

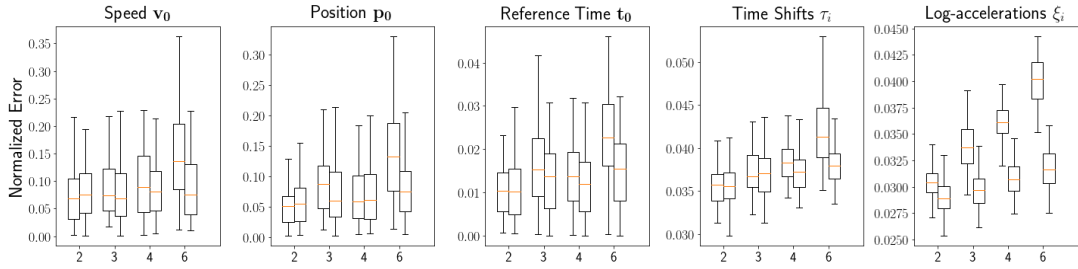


Figure 3.2 – Bootstrap distribution ($b=100$) of errors of estimation for population parameters $(\mathbf{p}_0, \mathbf{v}_0, t_0)$ and individual parameters τ_i and ξ_i according to subsampling frequency on an artificial dataset. Parameter values are taken close to estimation on real-world data. The second modality is assigned to NAs at various frequency to compare performance worsening between the naive method (boxplots on the left), and the generative modeling method (boxplots on the right).

3.3.1 Experiment methodology

We propose to evaluate the method by pruning existing datasets and comparing the performance in the estimation between removing all visits with at least one missing value (naive method) or taking into account these missing values via our generative model (generative modeling method). From our experience that NAs occur mainly by design in

neurodegenerative diseases datasets, we decide to prune the datasets by assigning chosen modalities to missing values at various (visit) frequencies.

In the 2 following experiments we use normalized scores on $M =]0, 1[$ with a metric ensuring that geodesics take the form of a logistic curve for each coordinate, so with $p_k = (\mathbf{p}_0)_k$ and $v_k = (\mathbf{v}_0)_k$ the multivariate model writes :

$$(y_{i,j})_k = \left(1 + \left(\frac{1}{p_k} - 1\right) \exp\left(-\frac{v_k \alpha_i (t_{i,j} - t_0 - \tau_i) + (w_i)_k}{p_k (1 - p_k)}\right)\right)^{-1} + (\epsilon_{i,j})_k \quad (3.3)$$

3.3.2 Synthetic Data

We produce synthetic data by simulating random-effects from their prior distribution and generating sample with the model. We generate a synthetic cohort of $p=300$ patients with 12 visits of 2 modalities each, occurring regularly on 4 years. Patient's age at beginning of the study are chosen arbitrarily as samples from a $\mathcal{N}(78, 5)$. We choose as initial parameters $\mathbf{p}_0^* = [0.4, 0.3]$, $t_0^* = 78$, $\mathbf{v}_0^* = [0.03, 0.04]$, $\sigma = 0.1$, $\sigma_\xi = 1$, $\sigma_\tau = 5$.

For each period in $[2, 3, 4, 6]$ we prune the dataset by assigning the second modality to a missing value every period of time, yielding datasets with patients that have respectively 6, 4, 3 and 2 visits with NAs. For each one of the obtained dataset, we bootstrap at the patient level the estimation procedure (6000 iterations) to obtain bootstrap distribution of relative estimation errors for both methods. The relative error is computed for each step of the bootstrap as followed : $\frac{\|\mathbf{v}_0 - \mathbf{v}_0^*\|}{\|\mathbf{v}_0^*\|}$, $\frac{\|\mathbf{p}_0 - \mathbf{p}_0^*\|}{\|\mathbf{p}_0^*\|}$, $\frac{\|t_0 - t_0^*\|}{\|t_0^*\|}$ for the main population parameters, and $\frac{\|\xi - \xi^*\|}{\|\xi\|}$ and $\frac{\|\tau - \tau^*\|}{\|\tau\|}$ for individual parameters.

Results are reported in Fig 3.2. We observe that population parameter's estimation is quite robust to pruning, with a significant difference in performance only visible from period = 6 (2 visits per subjects). On the individual parameters the difference is more striking, the generative modeling approach showing more robustness toward pruning already with only 1 over 2 visits removed.

3.3.3 Real Data

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. For up-to-date information, see www.adni-info.org.

We perform a similar experiment to a real dataset from ADNI1 cohort, consisting in 4 normalized neuropsychological test scores extracted from the ADAS-Cog, respectively associated with memory, language, praxis and concentration. Criteria for patient selection in ADNI1 was mild cognitive impairment at baseline and conversion to AD during the course of the study (MCI-converter), which led to 248 individuals. Patients are followed for an average of 6 visits and the dataset does not include any missing values. True parameters are not known, so we use as a proxy parameters estimated from a run of the estimation procedure on the entire dataset. Similarly to the previous experiment (6000 iterations, same estimation parameters), we subsample patients with NAs at various frequencies, discarding patients that are left with less than 2 visits without NAs and observe the

bootstrap distribution of the resulting estimation error in Fig 3.3. Results show the same trend as with synthetic data, although estimation error is higher.

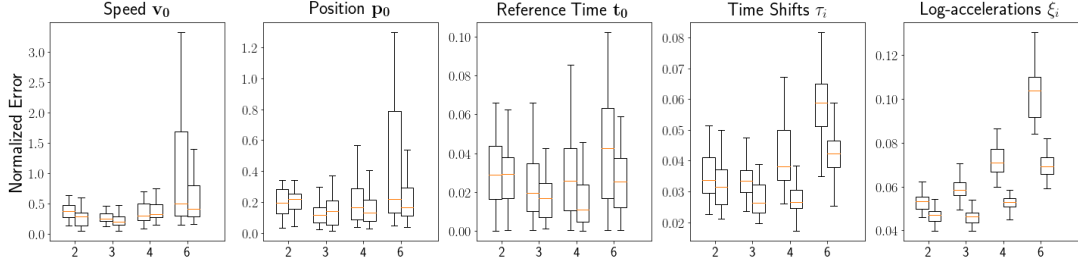


Figure 3.3 – Bootstrap distribution ($b=100$) of errors of estimation for population parameters ($\mathbf{p}_0, \mathbf{v}_0, t_0$) and individual parameters τ_i and ξ_i according to subsampling frequency on a real (ADNI) dataset. The 4 modalities used are subscores of the ADAS-COG accounting respectively for memory, praxis, language and concentration. 2nd, 3rd and 4th modalities are assigned to NAs at various frequency to compare performance worsening between the naive method and the generative modeling method.

3.3.4 Application to PPMI

Data used in the preparation of this article were obtained from the Parkinson’s Progression Markers Initiative (PPMI) database. (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org. From the PPMI cohort we extract 362 parkinsonian patients followed in average for 12 visits spread out on 4.6 years, yielding a total of 4441 visits. We model the joint progression of Parkinson’s disease for 2 biomarkers, a motor score, namely the MDS-UPDRS part 3 (MDS) and an imaging score, the Right Caudate Striatal Binding Ratio (SBR). We normalize the data between 0 and the theoretical max for the motor score MDS-UPDRS part 3, and the empirical max for the Striatal Binding Ratio. We choose to prescribe a logistic profile for the motor score, as such assessments are designed to be sensitive to the transition from normal to disease state. By contrast, there is no such assumption for imaging data, that we assume to decay in a linear fashion. MDS contains 151 NAs, while SBR includes 3202 NAs. We run our multivariate sigmoid model to model the progression of these modalities at once (6000 iterations), which we represent in Fig 3.4, and obtain a resulting noise variance $\sigma_{MDS, SBR}$ is 0.00341, the same magnitude than the noise variance on univariate features only, with $\sigma_{MDS} = 0.00318$ and $\sigma_{SBR} = 0.00502$. We observe a positive correlation between the acceleration factor ξ_i and age at diagnostic ($p = 3.010^{-2}$), meaning subjects with later onset will progress faster. Furthermore, studying correlations with biological covariables we find the alpha-synuclein mean level to correlate with subject’s onset ($p = 4.710^{-4}$).

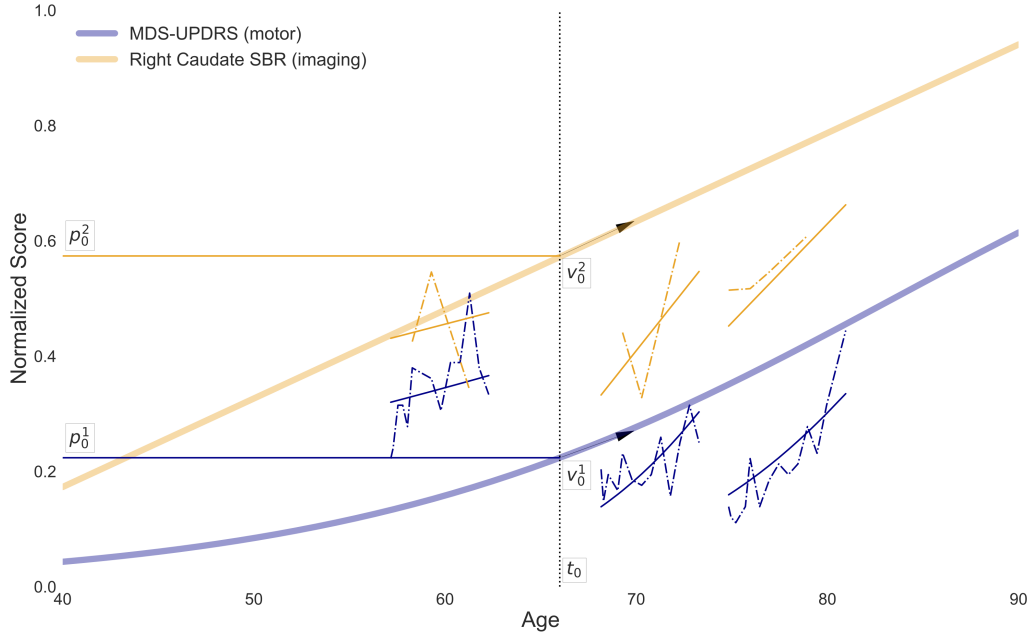


Figure 3.4 – In wide plain lines the mean geodesic estimated for PPMI PD patients on 2 modalities : MDS-UPDRS Score and Right Caudate SBR obtained from DatScan, and described by population parameters $\mathbf{p}_0 = [p_0^0, p_0^1]$, $\mathbf{v}_0 = [v_0^0, v_0^1]$ and t_0 . The observations (in dotted lines) and individual models (in narrow lines) of 3 patients are also plotted.

3.4 conclusion

We extended on a Bayesian non-linear mixed-effect model to allow the joint estimation of disease progression model on data with heterogeneous evolution profile. In practice such multimodal data include missing values by design. Instead of using ad-hoc method for data imputation, the generative statistical modeling allows to estimate model parameters by comparing generated data with observations only when they are available. Robustness analysis of the method is performed via the increasing pruning of existing dataset, while the variance of the performance is represented as a bootstrap distribution. The proposed method shows lower performance error in both synthetic and real-world data. This advocates for an extended use of the model, applicable to multimodal data with sparse design. We use thus our model to analyse the main PPMI modalities (motor, non-motor, imaging), and find that individual parameters correlates with age of diagnosis and alpha-synuclein levels.

Clinical applications on Parkinson's Disease

Outline of the current chapter

4.1 Tools for clinical applications	80
4.1.1 Data inclusion	80
4.1.2 Overview of clinical applications	81
4.2 Progression Model of Motor Abnormalities in Prodromal PD	86
4.2.1 Introduction	86
4.2.2 Method	87
4.2.3 Results and Discussion	87
4.3 Variability of clinical and imaging biomarkers in Parkinson's disease	89
4.3.1 Introduction	89
4.3.2 Materials and Methods	91
4.3.3 Results	95
4.3.4 Discussion	104
4.3.5 Supplementary Materials	106

*In this chapter, we present the clinical applications performed in this thesis. The first section recalls the main tools offered by the model for longitudinal data analysis. The second section presents a first work on motor abnormalities in early Parkinson's Disease, which led to an oral presentation at **Alzheimer's Disease/Parkinson's Disease (ADPD)** in 2019. Finally, the last section introduces our last work on quantifying PD heterogeneity over a broad range of markers, in PPMI, ICEBERG and DIGPD cohorts.*

4.1 Tools for clinical applications

In the two previous chapter, we introduced and extended a bayesian mixed-effect longitudinal model for scalar valued biomarkers. In this section, we briefly present how we used the longitudinal model to extract quantitative insights on longitudinal cohorts. We will rely on those methods for Parkinson's Disease data analysis in sections 4.2 and 4.3.

4.1.1 Data inclusion

In order to model Parkinson's Disease progression, we first need to select the clinically relevant data, used to train the model. In practice, to estimate a disease progression model, we select all subjects with the disease under study, excluding the healthy controls, on a carefully chosen set of biomarker linked with disease severity. We explain our rational in more details below.

Biomarkers

Recall that in neurodegenerative diseases longitudinal cohorts, exhaustive sets of biomarkers are measured. Some are cross-sectional, measured only at baseline (e.g. genotype, right-handedness...). Other are measured longitudinally, such as subject vitals, cognition, or motor scores. Among the biomarker measured longitudinally, most important are those assumed to directly measure disease severity (motor scales for Parkinson's Disease), or indirectly (protein Tau in Alzheimer's). Intuitively, a good longitudinal biomarker indicative of disease severity would evolve from 0 (healthy), to a value, say 1, indicating maximum severity for this particular biomarker.

As we want to model the disease progression, we focus on biomarkers indicative of disease severity, thus assumed to increase over time in patients, fitting our monotonicity assumption. In practice we choose biomarkers offering an overview for the specific task at hand. Including more biomarker would allow for finer granularity which may come handy, such as with clustering tasks, but a reasonable duration for model estimation limits the number of biomarker to a maximum of a few dozens.

Patient inclusion

An age matched, healthy control group is often available in longitudinal cohorts for comparison. One could consider to estimate a model including both healthy controls and patients with the disease under study. Doing so, an average scenario of progression of this bi-modal group would be estimated. Healthy controls have biomarker values close to 0 provided biomarkers measure disease severity. Consequently, they would be temporally re-aligned at the very beginning of the average scenario, and assumed to follow the same trajectory as patients, i.e. to become ill at some point. This assumption generally does not hold, and a joint modeling of the 2 groups should take into account this bimodal distribution.

We prefer the second option of estimating an average scenario solely on the patient group. We then see this average scenario, that depends on the arbitrary set of biomarkers we chose, as well as the patients included, as a proxy of disease progression.

Note that even if we do not include healthy controls, biomarkers are imperfect, and some are strongly linked with normal ageing (e.g. cognition), which then becomes hard to disentangle from disease progression. What we model is then rather the overall temporal variability of patients, encompassing both disease progression and normal ageing.

4.1.2 Overview of clinical applications

Recall that given a dataset $\mathcal{D} = (\mathcal{Y}, \mathcal{T})$, and assuming our assumptions of monotonicity and normalization hold with the chosen biomarkers (see section 1.3.2) we can estimate fixed effects θ of the model in a calibration procedure to obtain the average trajectory $\gamma(t)$. Then we can estimate the individual parameters z_i of a set of patients in a personalization procedure, yielding their individual trajectories $\eta^{w_i}(\gamma)(\psi_i(t))$. We write $z_G, w_G, \alpha_G, \tau_G$ for the respective averages of $z_i, w_i, \alpha_i, \tau_i$ over the patients belonging to group G , and assuming we know cutoffs $\{c_k, k = 1, \dots, K\}$ that indicate abnormality thresholds for each biomarker k , we can perform the following clinically relevant tasks:

- **Group regression** : $\gamma(t)$
Example : How do biomarker progress in average in Parkinson's Disease?
- **Progression heterogeneity** : θ
Example : What is the distribution of speed of progression across patients ? Of age at disease onset ? How do they differ in terms of patterns of progression ?
- **Population sequence of abnormalities** : $\gamma(t), \{c_k, k = 1, \dots, K\}$
Example : What is the typical ordering of symptoms in Parkinson's Disease and its variability ?
- **Individual regression** : $\eta^{w_i}(\gamma)(\psi_i(t))$
Example : What motor score will this particular patient have 4 years from now ?
- **Subgroup regression** : $\eta^{w_G}(\gamma)(\psi_G(t))$
Example : How do biomarker progress in this particular set of patients ?
- **Subgroup sequence of abnormality** : $\eta^{w_G}(\gamma)(\psi_G(t)), \{c_k, k = 1, \dots, K\}$
Example : What is the typical ordering of symptoms for this particular subgroup ?
- **Covariate analysis** : z_i
Example : Is a genetic subtype linked with specific patterns of progression of the disease ? Are older patient at baseline following different patterns of progression ?

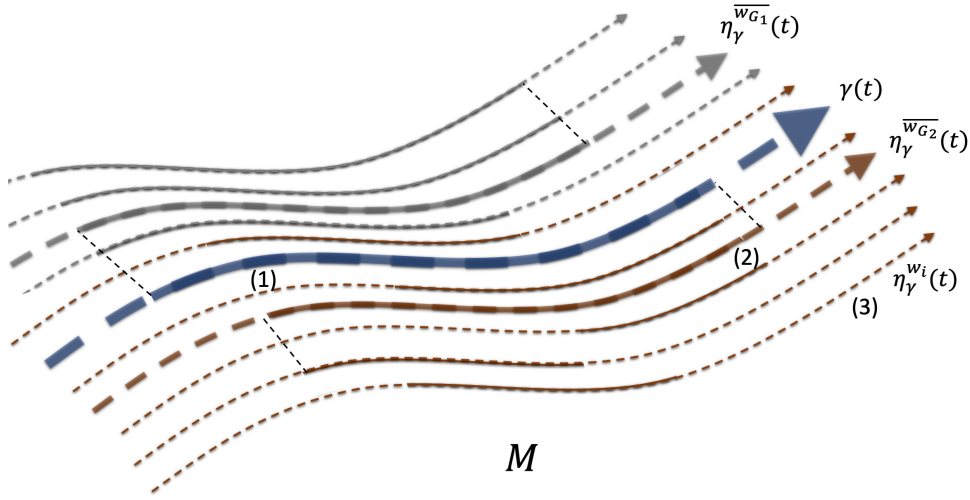


Figure 4.1 – Data points are assumed to lie on manifold M , and patients follow trajectories $\eta_{\gamma}^{w_i}(t)$ (3) parallel to the average trajectory $\gamma(t)$ (1). For a group G_2 of patients, we propose to average their spatial representations $\{w_i, i \in G_2\}$ to build the group average trajectory (2). In plain line we show the intervals of trajectories where data was seen, in terms of disease stage, and in dotted lines, parts of trajectories where no data was seen. With color red and grey we highlight groups of patients: G_2 encompasses red trajectories, while G_1 englobes grey ones. Modeling the data with the model in the plain line zone accounts for interpolation task, that we can perform at the 3 levels: individual, group or average, for e.g. missing data imputation. Modeling the data with the model outside the plain line zone accounts for extrapolation, may it be in the future, for prognosis, or in the past, to try to estimate how the disease start.

Group regression

The average trajectory $\gamma(t)$ describes quantitatively the longitudinal cohort it is trained on. This normative scenario shows a continuous and increasing progression of biomarkers over reparametrized age ψ . We call this reparametrized age ψ a "Disease Age", which can be understood as a disease stage. Individual trajectories are modeled as geometric variations around this average scenario that fit individuals measurements. The model is a mixed-effect model, and individual parameters z_i encoding these geometric variations are assumed to follow a normal $\mathcal{N}(0, \sigma_z \mathbb{I})$, which acts as regularization. Thus, the average trajectory can be understood as a Frechet mean of individuals trajectories, with a-priori assumption of a family of progression curves and geometric transformations. From a clinical point of view, we see $\gamma(t)$ as our average patient.

Individual regressions

Recall that individual parameter $z_i = (w_i, \alpha_i, \tau_i)$ encode temporal variability with (α_i, τ_i) and spatial variability with w_i . The trajectory of patient i is defined by w_i , and traveled at reparameterized time $\psi_i(t) = \alpha_i(t - \tau_i) : \eta^{w_i}(\gamma)(\psi_i(t))$.

Knowing z_i allows to perform interpolation for patient i , that may come handy to

impute missing data via the generative model. It also allows extrapolation tasks, such as predicting future value of a biomarker. The model was used to win a prize in the Tadpole Challenge <https://tadpole.grand-challenge.org/>, which aim at predicting future cognition in Alzheimer's Disease. However, our model assumptions of smooth increasing progression might not hold, especially if we predict at reparameterized age unseen in the training set. This way, predicting future cognition in Alzheimer's disease is especially hard as it can deteriorate quickly from a certain breaking point, that might be hard to predict a-priori.

Progression Heterogeneity

Parameters z_i are assumed to follow a normal $\mathcal{N}(0, \sigma_z \mathbb{I})$. The estimated scaling parameter are thus indicative of the heterogeneity of the population. It informs us of temporal variability: σ_τ indicates how different patient ages are, when patient have been re-aligned for disease stage. σ_α indicates the range of speed in biomarker progression. Also, σ_w is informative of biomarker-wise variations from the average trajectory, which can help detect biomarkers with high heterogeneity in the population. For example in Parkinson's Disease (see 4.3), cognition (MoCA) and sleep (RBDSQ) show the highest spatial variability in the population of Parkinson's Disease patients, with regard to motor (MDS-UDPRS III) or imaging biomarkers.

Note that we assume unimodal distributions of each individual parameters, but it could also be possible to relax the gaussian hypothesis for a mixture of gaussians, which would allow for unsupervised clustering of individuals.

Cofactor Analysis

Knowing the extent of longitudinal heterogeneity in our dataset, it is tempting to correlate this heterogeneity, in the form of z_i with known cofactors, such as patient phenotypes or genetic subtypes. Note that as z_i encode interpretable, geometrical variations from the average scenario, and this is very useful in analyzing the correlation results.

Subgroup Scenario of progression

As mentioned above, we saw $\gamma(t)$ as the "average" patient. In the same manner, we propose to compute a patient subgroup G trajectory using the expectation of z_i over the group G .

$$\begin{cases} \tau_G = \mathbb{E}[\tau_i] \\ \alpha_G = \mathbb{E}[\alpha_i] \\ w_G = \mathbb{E}[w_G] \end{cases}$$

We can then perform the same regression and abnormality arrival analysis as the average scenario of progression.

Timeline of abnormalities in the normative scenario

We have an average model of progression $\gamma(t)$ which describes the continuous progression of biomarkers over time. However this does not answer the question of when biomarkers become abnormal, and in which order ?

In (Schiratti, Allasonniere, et al. 2015b), this issue is handled by enforcing the same speed for all biomarker, which are then modeled by the same sigmoid curve, only shifted in time. However, this comes at the cost of model flexibility, which is useful with biomarkers of different nature, that evolve naturally at different speeds.

If we use sigmoids, we could also consider 0.5, the inflexion point of our biomarker curves, to be used as reference point for each biomarkers. But biomarker scales may have very different meanings, and some might not even reach the value 0.5 in the course of the disease. For example a MoCA value of 15/30 indicate an extreme severity of cognition symptoms, and is rarely reached in Parkinson's Disease.

Provided only $\gamma(t)$, answering the question of the ordering of symptoms would require expert knowledge of which biomarker values indicate abnormalities. We propose to estimate cutoffs of abnormalities for each biomarker directly from the data. We consider 2 different simple methods to estimate these cutoffs that rely on (age-matched) control biomarkers values at baseline. The first option consist in simply computing the 95% percentile of the control biomarkers at baseline. However, it might happen that biomarker distribution at baseline of controls and patients overlap, and this is the case with biomarkers linked with normal ageing, such as cognition. Thus we also consider as second option, consisting in using the cutoff of a balanced logistic regression trained to classify, for each biomarker separately, healthy controls and patients. We argue that both option are valid, keeping in mind how they were obtained to interpret the results.

Once we have these cutoffs $\{c_k, k = 1, \dots, K\}$, we can compute the times at which the average scenario intersects these cutoffs to get "Disease Age" of abnormalities for each biomarker, and order them. To take into account both the variability in data and model estimation, we propose to bootstrap the process, yielding bootstrap distributions of Disease Age at abnormality for each biomarker. In Figure 4.2, we choose the logistic regression method to get the cutoffs, and biomarker 1 becomes abnormal at around 50 in Disease Age, while biomarker 2 becomes abnormal at around 60.

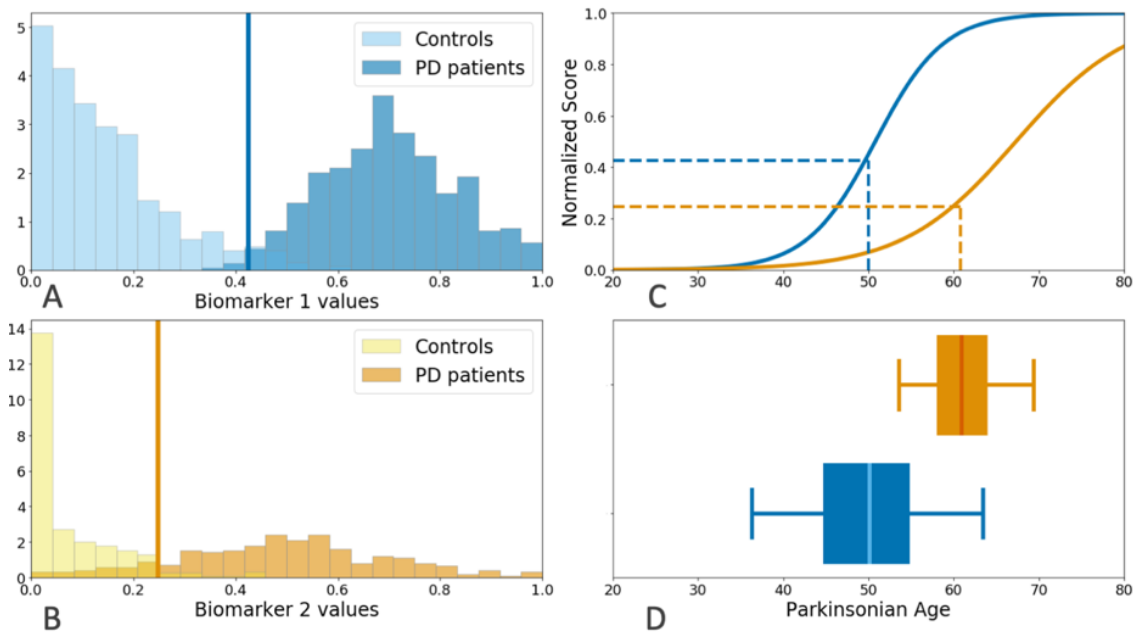


Figure 4.2 – Illustration of the use of pathological thresholds to build a timeline of abnormalities (shown in D). First a longitudinal spatiotemporal model is trained on the patient longitudinal data (C). Then a logistic regression method determines the thresholds that best separate control from patients data at baseline for each endpoint (A,B). The thresholds are reported on the y-axis of the average trajectory (C). The Disease Age at which the endpoint become pathological is determined (C). We report the series of PA for each endpoint together with the bootstrap confidence interval on the reference timeline (D).

4.2 Progression Model of Motor Abnormalities in Prodromal PD

*In this section we illustrate the use of the timeline of abnormalities to better understand the early motor stages of Parkinson's Disease. We build the timeline on prodromal patients of PPMI. Prodromal patients are assumed to develop at some point a Lewy Body disease, mostly Parkinson's Disease. They are chosen in PPMI as either having RBD ($n=38$) or hypsomnia ($n=31$) We focus on the motor biomarkers, and by ordering their timing of abnormality, we estimate which happen at the earliest stage of the disease. This work led to an oral presentation at **Alzheimer's Disease/Parkinson's Disease (ADPD)** in 2019.*

4.2.1 Introduction

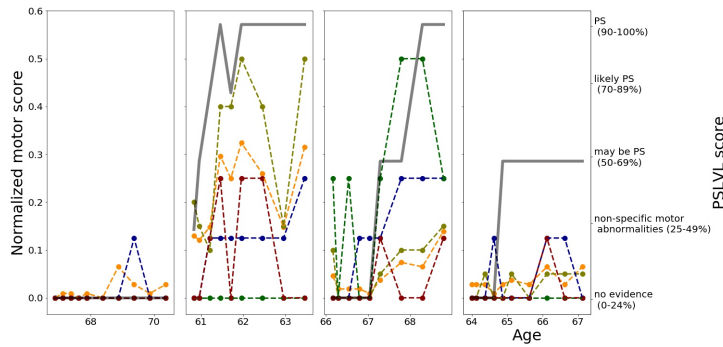


Figure 4.3 – Example of 4 iRBD patient data. Each graph shows the values of a given patient. In color we show the normalized values of the motor score MDS-UPDRS part III (yellow), and subscores of the MDS-UPDRS part III: Rigidity (kaki), Bradykinesia (red), Gait (green), and Voice-Face (blue). A Parkinsonism level (PSLVL) score is plotted in plain grey. We see that these scores are inherently noisy, and that patient are not necessarily at the same stages of the diseases.

We estimate a normative progression of motor abnormalities in Parkinson Disease prodromal phase using sub-scores of the MDS-UPDRS scale. PPMI study includes longitudinal MDS-UPDRS sub-scores for PD ($n=423$), Control ($n=196$) and Prodromal ($n=65$) cohorts. We use a multivariate non-linear bayesian mixed-effect model (Schiratti, Allasonniere, et al. 2015b) to estimate at once the longitudinal progression of MDS and MDS-subscores. It is built on the normalization of multiple short-term data sequences with individual time-reparameterization functions that account for variations in age at symptoms onset and pace of progression.

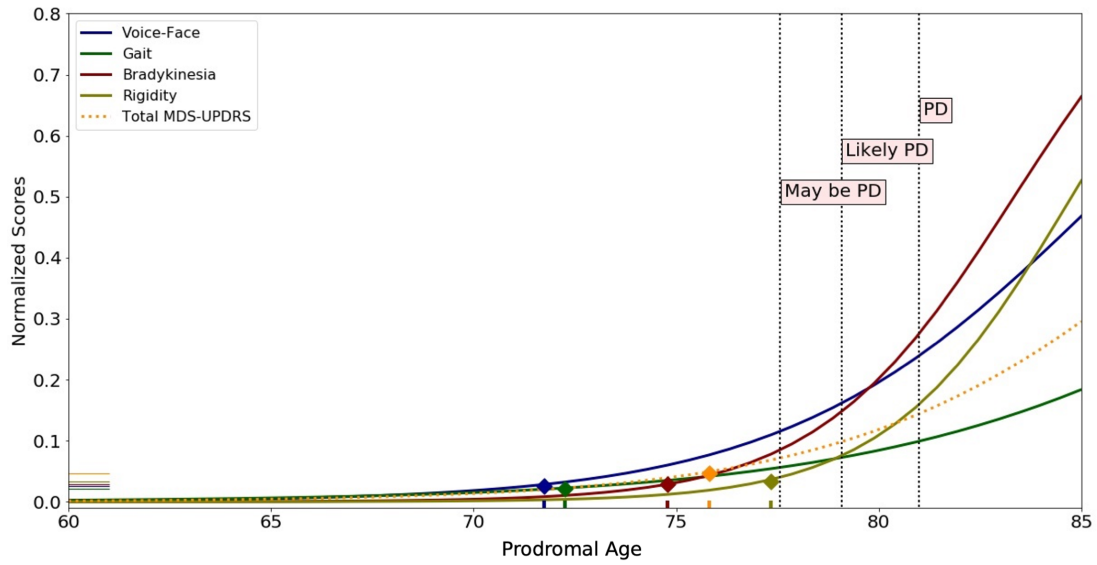


Figure 4.4 – Normative scenario of biomarker's progression in prodromal PD patients with regard to "Prodromal Age". We compute each biomarker abnormality cutoff value as a the 95% quantile on the control group. We then look at intersections between the normative scenario and these cutoffs values of abnormalities to get timings of abnormalities (diamonds icons). For scale, we add on the normative scenario the reparametrized events of conversion to the last 3 stages of Parkinsonism Level (PSLVL): 3 (may be PD), 2 (Likely PD), and 1 (PD).

4.2.2 Method

We estimate a normative scenario of progression from the 65 prodromal subjects, and refer to "Prodromal Age" for the index in the timeline of the normative model of progression. We compute the times of intersection between the normative progression model and abnormal values (above 95 percentile on control group) to assess an abnormality timing for each modality. Finally we bootstrap the process ($p=500$) to account for sampling bias and stochastic variations in the estimation algorithm, and obtain confidence intervals. We compute both the normative progression scenario and abnormality timings (see Figure 4.4). We then assess via bootstrap the distribution of these timings (see Figure 4.5).

4.2.3 Results and Discussion

Relatively to PD conversion, we find that Voice-Face modality is affected first, with median time to conversion of 9.2 years. Then Gait, Bradykinesia, MDS-UPDRS and Rigidity at respectively 7.3, 6.1, 4.6 and 2.9 years before conversion (see Fig 4.5). We were able to assess the progression of motor scores on the PPMI prodromal cohort using a longitudinal model with patient's specific time reparametrization. Results obtained here are specific to prodromal subjects, and may not be representative of the other types of PD progression.

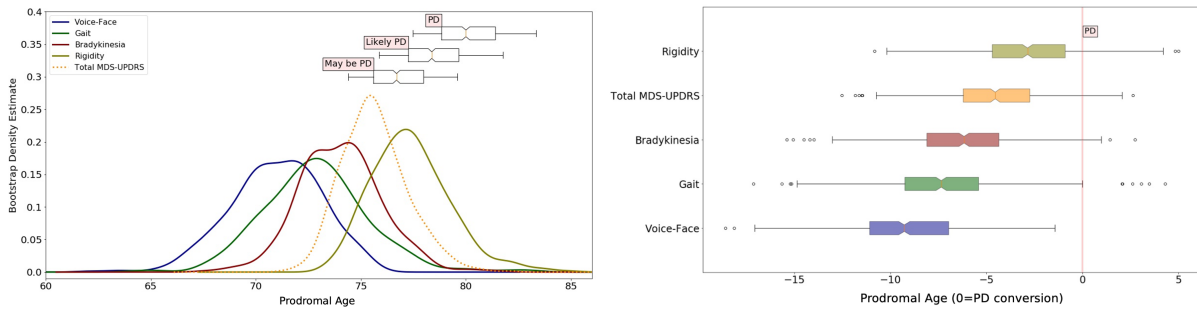


Figure 4.5 – On the left, bootstrap distribution of "Prodromal Ages" at abnormality for each motor marker. Bootstrap distribution is shown with a kernel density estimation. Both population variability and estimation variability are captured in the bootstrap procedure. On the right, bootstrap distribution of "Prodromal Ages" at abnormality for each motor marker, relatively to the reparameterized event of conversion to Parkinson's Disease. Boxplots representing Prodromal Ages at conversion to PSLVL values of 3, 2, 1 are also added. In this case distributions are shown with boxplots. Relatively to PD conversion, we find that Voice-Face modality is affected first, with median time to conversion of 9.2 years. Then Gait, Bradykinesia, MDS-UPDRS and Rigidity at respectively 7.3, 6.1, 4.6 and 2.9 years before conversion. Note that we show in these graphs the variability of the population parameters, and not the inter-patient variability, which would be higher, and with more overlapping between the different biomarkers.

	Our method	R. Postuma, Lang, et al. 2012
Voice/face	9.2 (4.6-14)	9.8 (6.7-29.8)
Gait	7.3 (2.1-12)	4.4 (3.9-5.4)
Bradykinesia	6.1 (1.6-11)	5.4 (4.8-6.5)
MDS-UPDRS part III	4.6 (0.45-8.5)	4.5 (3.7-5.6)
Rigidity	2.9 (2-7.1)	4.4 (3.9-5.5)

Table 4.1 – On the left, estimated Prodromal Years between arrival of abnormalities and conversion to PD. On the right, estimated duration in years between abnormalities and conversion to PD using independant linear regressions (R. Postuma, Lang, et al. 2012). Our results are in a general agreement with those of a previous study from R. Postuma, Lang, et al. 2012, with still one difference in the ordering between Gait abnormalities and Bradykinesia.

4.3 Variability in the progression of clinical and imaging biomarkers in Parkinson's disease

*In this section, we present our clinical application work on leveraging the spatiotemporal model to uncover the heterogeneity in Parkinson's Disease progression, with a focus on sleep symptoms. This work was already supported by accepted abstracts at **Virtual Physiological Human 2020, International Congress of Parkinson's Disease and Movement Disorders 2020 and Compag 2020**. This section presents a broader work encompassing these contributions, for later submission to a clinical journal.*

4.3.1 Introduction

It is widely assumed that Parkinson's disease (PD), like most neurodegenerative diseases, has a long prodromal phase during which neuropathological (Braak et al. 2003) and clinical (Berg, Ronald B Postuma, et al. 2015) changes develop slowly. The presence of this long prodromal phase suggests that disease-modifying interventions aiming to alter the course of the disease would be desirable as soon as possible, perhaps even at the pre-symptomatic stage. Although disease-modifying treatments are not yet available in PD, it is therefore important to generate precise models of the course of the disease in order to facilitate future therapeutic trials at an early stage of the disease.

The most influential model was based on neuropathology (Braak et al. 2003). This model postulates that the pathological process gradually progresses from peripheral autonomic neurons to the medulla oblongata and pons before affecting the substantia nigra (Charles H. Adler and Beach 2016; Braak et al. 2003). In neuropathological studies, the estimated duration of the pre-symptomatic phase in the substantia nigra was estimated at 4.7 (Fearnley and Lees 1991) and 5.0 years (Greffard et al. 2006). Similarly, using neuromelanin-sensitive MRI and calculation of nigral volume loss, neurodegeneration in the nigra was estimated to start 5.3 years before disease diagnosis (Biondetti et al. 2020). PET studies have reported longer pre-symptomatic phases ranging from 5.6 (Hilker et al. 2005) to 10.0 years (Fuente-Fernández et al. 2011) supporting the hypothesis that damage to dopaminergic axons precedes cell death in the SN (Zou et al. 2016).

The common hypothesis is that the sequence of functional and anatomical alterations in the brain translates into the progressive onset of several motor and non-motor symptoms including autonomic dysfunction and cognitive impairments. Unfortunately, the Braak's unitary vision of the pathogenesis does not explain the large heterogeneity that is observed in the disease presentation (Greenland, Williams-Gray, and Barker 2019; G. Halliday et al. 2008; G. M. Halliday and McCann 2010). The sequence and timing of motor and non-motor symptoms shows great variability across patients, and therefore are particularly difficult to predict.

Several approaches have been used to model this variability. Survival analyses were proposed to predict the risk to develop a particular aspect of the disease such as cognitive decline or loss of autonomy within a pre-defined time period. These methods showed accuracies of the order of 85% in predicting such risks from a series of clinical endpoints assessed at one time-point close to diagnosis (Latourelle et al. 2017; G. Liu et al. 2017;

Macleod et al. 2018; Velseboer et al. 2016). They confirmed the variability in disease presentation, and the fact that the patient trajectory was partly predictable using patient's data at time of diagnosis. Models of disease progression aiming to describe the continuous progression of biomarkers over time were used for instance to simulate how treatments might change the progression of the motor impairments assessed by the unified Parkinson's disease rating scale (UPDRS) during a clinical trial (Cilia et al. 2020; Venuto et al. 2016). Such an approach has been followed recently to describe the long-term progression of a series of clinical endpoints from the prodromal to the clinical stage of the disease (Fereshtehnejad, Yao, et al. 2019). The authors proposed a general scheme of PD progression starting with hyposmia and sleep disorders, followed by dysautonomia, motor impairment and eventually cognitive decline. Subtyping is another approach to summarize the heterogeneity in disease progression profiles with a small set of disease subtypes. The general approach is to cluster patients' data at baseline, and then show that the rates of change of the biomarkers differ significantly between clusters (Aleksovski et al. 2018; De Pablo-Fernández et al. 2019; Duarte Folle et al. 2019; Fereshtehnejad, Romenets, J. B. M. Anang, et al. 2015; Lawton et al. 2018). These methods successfully exhibited stereotypic progression profiles, for instance a motor-predominant less severe form of the disease compared to a diffuse severe 'malignant' form with greater cognitive impairment, RBD and dysautonomia (Fereshtehnejad, Zeighami, et al. 2017).

In this paper, we propose a multivariate model showing the progression of four imaging biomarkers and eight clinical endpoints assessing both motor and non-motor symptoms from the prodromal to the clinical stage of the disease. The model describes the range of likely trajectories of these 12 biomarkers during disease progression as observed in 916 patients from three independent data sets. In lieu of the simplifying procedure of subtyping, we considered that each subject followed a different trajectory characterized by a specific age at onset, pace of progression and specific timing and ordering of the different biomarkers. The model describes the range of biomarker trajectories in a continuous fashion.

The proposed model is an application of a novel statistical learning technique called disease course mapping (Schiratti, Allassonniere, et al. 2015a; Schiratti, Allassonniere, et al. 2017). It extends disease-modeling methods to a multivariate analysis. It addresses specific issues of the multivariate analysis of longitudinal data by uniquely identifying variations due to differences in the dynamics of progression (e.g. age at onset and pace of progression) and due to different disease presentation (e.g. relative timing among biomarkers). It has been applied for similar purposes in the field of Alzheimer (Koval, Bône, et al. 2021) and Huntington disease (Koval, Dighiero-Brecht, et al. 2021). It allows the forecasting of future cognitive decline in subjects with mild cognitive impairments with better accuracy than 56 competing methods (Koval, Bône, et al. 2021). It allows more powered clinical trial design by selecting of participants with more homogeneous progression profiles than with current selection methods (Koval, Dighiero-Brecht, et al. 2021). This technique is applied here for the first time in the field of Parkinson disease (PD) to explore the variability of progression profiles within the idiopathic PD population. It is used further to position the trajectories of idiopathic RBD (iRBD) subjects with respect to the distribution of PD trajectories.

4.3.2 Materials and Methods

Participants

We used data from three independent longitudinal studies: Parkinson’s progression markers initiative PPMI, a multicenter study in North America, DIGPD DIGPD (Corvol, Artaud, Cormier-Dequaire, Rascol, Durif, Derkinderen, Marques, Bourdain, Brandel, Pico, et al. 2018b), a multicenter study in France, and ICEBERG a single-center study conducted at the Paris Brain Institute (ICM), France ICEBERG. We included all consecutive patients with idiopathic Parkinson disease (PD) of PPMI (n=423), DIGPD (n=415) and ICEBERG (n=165) studies, idiopathic rapid eye movement Sleep Behavior Disorder patients (iRBD) of PPMI (n=39), as well as the healthy controls of PPMI (n=196) and ICEBERG (n=70). Patients were excluded if they had only one visit available.

Selection of the endpoints

We selected all available endpoints among the following set of 12 endpoints:

- activities of daily living: (1) part I of the MDS-Unified Parkinson’s Disease Rating Scale – MDS-UPDRS (MDS-UPDRS I), (2) part II of the MDS-UPDRS (MDS-UPDRS II)
- motor: part III of the MDS-UPDRS in off state (MDS-UPDRS III Off) and on state (MDS-UPDRS III On),
- autonomic: scale for outcomes in Parkinson disease– autonomic dysfunction sub-scale (SCOPA-AUT)
- cognitive: (1) Montreal cognitive assessment (MoCA), (2) mini-mental state examination (MMSE)
- sleep dysfunction: the RBD screening questionnaire (RBDSQ)
- imaging: striatal binding ratios (SBR) in putamen and caudate nucleus measured by dopamine transporter scan in the ipsi or contra lateral hemisphere (Put. Ipsi, Put. Contra, Caud. Ipsi, Caud. Contra respectively)

For practical purposes, we grouped these endpoints in three categories: ‘body’, ‘brain’ and ‘imaging’. ‘Body’ endpoints referred to MDS-UPDRS I, SCOPA-AUT and RBDSQ, which described the main non-motor symptoms of the disease, and whose early onset was assumed to indicate a “body-first” subtype (Berg, Borghammer, et al. 2021; Horsager, Andersen, Knudsen, Skjærbæk, et al. 2020a). ‘Brain’ endpoints included MDS-UPDRS III in Off and On states, as well as cognition (MoCA and MMSE), which reflected the classic pattern of Parkinson’s Disease progression from the Braak’s model (Braak et al. 2003). Finally, ‘imaging’ endpoints included putamen and caudate SBR.

Definition of the RBD status

We subdivided PD patients in two subgroups depending on whether they developed PD with or without RBD. In PPMI, we labeled PDRBD+ the PD patients which maximal RBDSQ score over the course of the study was higher or equal to 6 in a scale of 13 (Stiasny-Kolster et al. 2007). In ICEBERG and DIGPD, we used the RBD diagnosis reported by a clinician as label.

Data normalization

All endpoints were supposed to decrease or increase monotonically over the disease course. We converted them to a scale between 0 (normal value) and 1 (maximum pathological change).

For clinical endpoints of all cohorts, normal values and maximum pathological changes corresponded to the range of the scale: MDS-UPDRS I: 0-52, MDS-UPDRS II: 0-52, MDS-UPDRS III: 0-108, SCOPA: 0-69, RBDSQ: 0-13, MoCA: 0-31, MMSE: 0-30.

For imaging biomarkers, we set to 0 (normal value) the average SBR of control subjects at baseline, and 1 (maximum pathological change) a null SBR. SBR measurements were inverted and rescaled to this 0 to 1 scale.

Disease course mapping

We applied to these data a disease modeling technique called disease course mapping (Schiratti, Allassonniere, et al. 2015a; Schiratti, Allassonniere, et al. 2017). This technique has been used to map the range of likely trajectories of Alzheimer and Huntington disease (Koval, Bône, et al. 2021; Koval, Dighiero-Brecht, et al. 2021). It was applied here in the area of Parkinson disease.

Disease course mapping takes as input a longitudinal data set with several endpoints. It outputs a typical model of progression in the form of a series of logistic curves starting from the most normal value and ending at the maximum pathological change (see Figure 4.6).

Three series of parameters change the relative shape and position of these curves to account for the phenotypic differences across patients and variations in their progression (see Figure 4.6):

- **one time-shift** showing how early or late is the disease onset assessed by taking all endpoints into account,
- **one acceleration factor**, showing how fast or slow is the progression assessed by taking all endpoints into account,
- **one inter-marker spacing for each endpoint**, showing how early or late is the onset of the endpoint relatively to the other ones, for a normalized age at onset and pace of progression.

A time-shift of 1 year (resp. -1 year) means that endpoints overall, after correcting for inter-marker spacing and pace of progression, reach the same values one year after (resp.

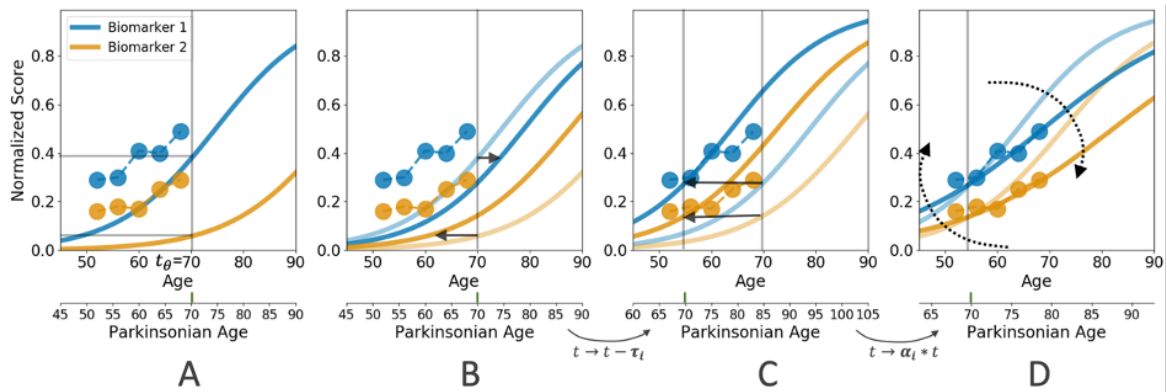


Figure 4.6 – Positioning of a subject in the PD Course Map. A) The typical progression of two biomarkers (blue and yellow) is shown by plain logistic curves together with the corresponding biomarker values of one subject at 5 successive visits. Biomarker values are normalized so that 0 corresponds to the most normal value and 1 the maximal pathological changes (see Methods). B-to-D) A series of operations allow us to transform the model to fit the subject data and therefore determine the subject's specific trajectory. The inter-marker spacings change the time interval between the two curves (B), the time-shift translates the curves (C) and the acceleration factor stretches the curves (D). The last two operations map the actual age of the patient to his Parkinsonian Age. We estimate the shape and position of the logistic curves, so that the average time-shift, inter-marker spacing, and log-acceleration factor for all subjects in the training longitudinal data set is zero.

before) the typical patient. An acceleration factor of 2 (resp. 0.5) means that the same changes in the endpoints take twice less time (resp. twice more time) to occur compared to the typical scenario. An inter-marker spacing of 1 year (resp. -1 year) for a given endpoint means that the changes of this endpoint occur one year after (resp. before) than in the typical scenario, all other thing being equal.

The first two parameters (time-shift and acceleration factor) map the actual age of the patient at a given visit to a Parkinsonian age (PA), which is an index in the timeline of the typical model of progression. It can be understood as a disease stage that is independent of the phenotype of the subject. Inter-marker spacing captures the variations in the disease presentation.

We estimated together the typical model of progression (population parameters) and the distributions of the subject parameters (individual parameters) given longitudinal data (Schiratti, Allassonniere, et al. 2015a). The result was a distribution of likely trajectories derived from the typical trajectory, which was called a PD Course Map. We determined the subject’s trajectory in the PD Course Map by computing estimates of the individual parameters from the repeated observation of this subject. PD Course Map estimation did not require that data were evenly sampled in time, or that the sampling was uniform across subjects. It did not need imputation of missing values as the model likelihood was informed with the available data only (Couronne et al. 2019).

We computed confidence intervals of the population and individual parameters using a bootstrap method with 50 resampling steps.

We shared an open-source python library designed to estimate disease Course Maps from longitudinal data at <https://gitlab.com/icm-institute/aramislab/leaspy/>.

Pathological thresholds

We estimated a pathological threshold for all endpoints. We contrasted values at baseline of the considered endpoint between the control group and the PD group in the PPMI and ICEBERG data set. We used a balanced logistic regression to estimate the cut-off value that best separate the two groups. Values above this threshold were considered pathological (Clifford R. Jack, Wiste, et al. 2017; Martínez-Martín et al. 2015). We used these thresholds to determine the Parkinsonian Age at which a given endpoint become pathological in PD Course Maps.

PD Course Maps

We built two PD Course Maps:

- **A multimodal course map** using both clinical and imaging endpoints in all selected PD patients in the PPMI data set: MDS-UPDRS I, MDS-UPDRS II, MDS-UPDRS III Off, MDS-UPDRS III On, SCOPA-AUT MoCA, RBDSQ, Put. Ipsi, Put. Contra, Caud. Ipsi, and Caud. Contra.
- **A multi-domain clinical course map** using only clinical endpoints in all selected PD patients in the PPMI, ICEBERG and DIGPD data sets: MDS-UPDRS I, MDS-

UPDRS II, MDS-UPDRS III Off, MDS-UPDRS III On, SCOPA-AUT and MoCA and MMSE.

Note that in the multi-domain course map, no MMSE endpoint was available in PPMI, while DIGPD did not include MoCA assessment, and only few MDS-UPDRS III Off observations. We still concatenated all PD patients of the 3 cohorts, and as mentioned above we treated the lack of endpoints as missing values, which are conveniently handled by PD Course Map.

We computed the average Parkinson Age corresponding to the actual age at which patients experience first symptoms and are diagnosed as PD. Eventually, we used the multimodal course map to position the iRBD subjects from PPMI with respect to the PD patients.

Comparison of PDRBD+ and PDRBD- patient progression

To compare PDRBD+ and PDRBD- patients, we analyzed each individual parameter from the multimodal course map using the ANOVA method with RBD status, age at baseline and sex as cofactors. We added the cohort as cofactor for the analysis of the individual parameters from the multi-domain clinical map. In all cases, we used the Bonferroni method to correct for testing each individual parameter. Significance level was set to 5% after correction. We transformed the typical scenario of progression with the average of the individual parameters in each group. We obtained two scenarios that we superimposed on the same timeline to illustrate the differences of progression between both groups. All results are presented as mean $\pm$ standard deviation.

4.3.3 Results

Study Subjects

From the total of 1003 PD patients, we excluded 27 patients in DIGPD (5.5%) and 40 in ICEBERG (24.2%) because they had only one visit. For the same reason, we removed 1 iRBD (2.6%) patient in PPMI. We considered all 70 healthy controls from ICEBERG and 196 from PPMI, as we did not model their longitudinal progression, but only use their baseline biomarker values to estimate cutoffs of abnormality. Thus analyses were performed on 936 PD patients (all subjects in PPMI, 388 PD in DIGPD and 125 PD in ICEBERG), 38 iRBD and all control subjects (196 in PPMI and 70 in ICEBERG).

PD Patient characteristics

Subjects' characteristics are presented in Table 4.2. At baseline, there was no significant difference in age distribution between groups of patients in the three cohorts, at the exception of PPMI iRBD patients, which are older than other groups (in average 7.8 years older than their PD counterpart in PPMI). Gender balance was similar in all groups of patients except PPMI iRBD with 16% of men (proportion of men: 35% in PPMI, 37% in both ICEBERG and DIGPD for PD patients; 36% in PPMI and 44% in ICEBERG for healthy controls). PD patients in ICEBERG and DIGPD had slightly higher disease

		PPMI			ICEBERG		DIGPD
		Control	iRBD	PD	Control	PD	PD
General information	Total Patients	196	38	423	70	125	388
	Male/Female	70/126	6/32	146/277	31/39	46/79	155/233
	Age (y)	60.7±11.2	69.4±5.5	61.6±9.7	62.4±9.3	62.4±9.0	62.1±9.8
	Disease Duration (y)	-	-	0.6±0.5	-	1.5±1.0	2.6±1.5
	Visit Number	8.5±2.2	12.7±1.7	13.9±3.6	2.6±1.3	3.8±1.0	5.4±1.7
	Study Duration (y)	6.3±2.4	4.8±0.8	6.2±2.1	1.7±1.3	3.0±1.2	4.8±1.8
Autonomic & Daily living	MDS-UPDRS I	3.0±3.0	7.1±4.0	5.8±4.2	5.0±3.5	9.6±4.2	7.9±4.8
	SCOPA-AUT	5.8±3.7	14.9±8.2	9.5±6.1	7.2±5.8	12.4±7.2	11.1±6.9
Motor	MDS-UPDRS II	0.5±1.0	2.1±2.4	5.7±4.2	1.2±1.7	8.3±3.9	7.5±4.5
	MDS-UPDRS III Off	1.2±2.2	4.7±3.9	20.4±8.9	5.6±5.1	30.4±8.1	28.6±15.8
	MDS-UPDRS III On	-	25.6±13.9	23.6±11.5	-	26.7±7.8	20.0±10.1
	Hoehn&Yahr Off	0.0±0.1	0.0±0.0	1.5±0.5	-	-	2.0±0.6
Cognition	MoCA	28.2±1.1	25.5±4.2	27.1±2.3	27.8±2.0	27.6±2.0	-
	MMSE	-	-	-	29.4±0.9	29.0±1.0	28.2±1.9
Sleep	RBDSQ	2.8±2.3	9.0±3.0	4.1±2.7	-	-	-
Imaging (SBR)	Putamen contralateral	2.1±0.6	-	0.7±0.3	-	-	-
	Putamen ipsilateral	2.1±0.6	-	1.0±0.4	-	-	-
	Caudate contralateral	3.0±0.6	-	1.8±0.6	-	-	-
	Caudate ipsilateral	3.0±0.6	-	2.2±0.6	-	-	-

Table 4.2 – Characteristics of the data sets. For each biomarker, we report the average and standard deviation at baseline. In PPMI, medications are not given from baseline, and therefore the first value seen for MDS-UPDRS III On accounts for a later age, in average 2.0 ± 1.3 years later than baseline. This explains in PPMI the higher value of MDS-UPDRS III On than MDS-UPDRS III Off. Abbreviations: SBR = striatal binding ratio, y = years.

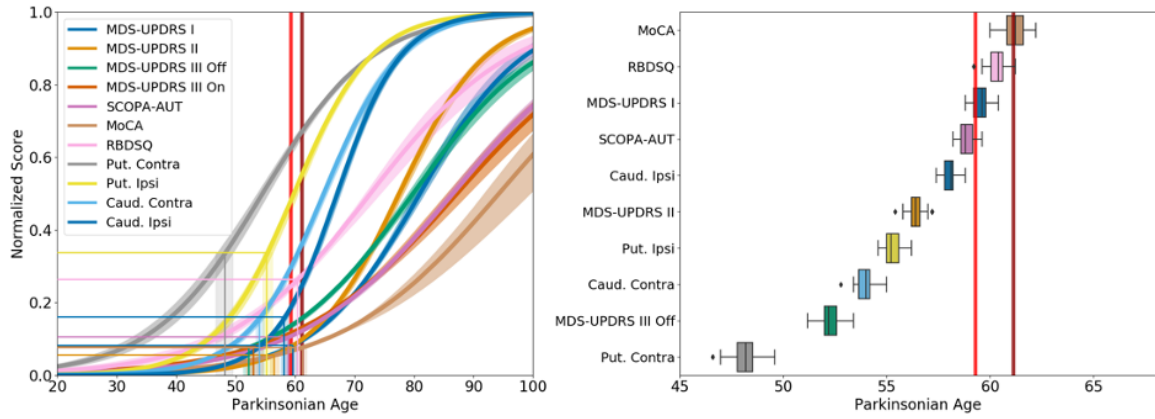


Figure 4.7 – Multimodal disease course map trained on PPMI showing the progressive onset of clinical endpoints and imaging biomarkers from pre-symptomatic to post-diagnosis stages. Left panel: the progression of the endpoints from the most normal (0) to the maximum pathological changes (1) is shown as Parkinsonian Age progresses. Shaded areas correspond to 95% bootstrap confidence interval. Right panel: the timing at which endpoints become abnormal is shown with 95% confidence intervals. Vertical lines indicate the estimated Parkinsonian Age of first symptoms (light red) and of diagnosis to PD (dark red).

durations (1.5 ± 1.0 and 2.6 ± 1.5 years respectively) than in PPMI patients (0.6 ± 0.5 years). In PD patients, follow-up duration was 6.2 ± 2.1 years with 13.9 ± 3.6 visits in PPMI, 4.8 ± 1.8 years with 5.4 ± 1.7 visits in DIGPD, and 3.0 ± 1.2 years with 3.8 ± 1.0 visits in ICEBERG. In iRBD patients of PPMI, follow-up duration was 4.8 ± 0.8 years over 12.7 ± 1.7 visits. In PD patients, there was a total of 529 PDRBD+ patients (278 in PPMI, 61 in ICEBERG and 190 in DIGPD) and 404 PDRBD- patients (145 in PPMI, 64 in ICEBERG, 198 in DIGPD). Most biomarker values at baseline were significantly different between cohorts (Mann Whitney U-Test, Supplementary Table 4.4). Biomarkers severity was lower in PPMI than in DIGPD and ICEBERG because disease duration at baseline was shorter in PPMI. ICEBERG biomarker severity at baseline was higher than in DIGPD, while ICEBERG patients had 1.13 years lower disease duration at baseline than DIGPD patients. This pointed out potential biases between cohorts, which could be conveniently handled by PD Course Map.

A typical PD progression starting with dopaminergic loss and ending with cognitive impairment

The multimodal course map, which described the typical scenario of progression derived from the 423 PD patients of the PPMI cohort, is shown in Figure 4.7. We chose PPMI as our discovery cohort, as it offered the longest follow-up and highest number of visits on an exhaustive set of biomarkers. After realignment of all PD subjects' age at diagnosis on the multimodal course map, the average PA of the age at diagnosis was 61.1 ± 0.3 years. Imaging endpoints became pathological first. Functional alterations started in the putamen 12.9 ± 0.5 years before diagnosis (Put. Contra PA = 48.2 ± 0.7 years)

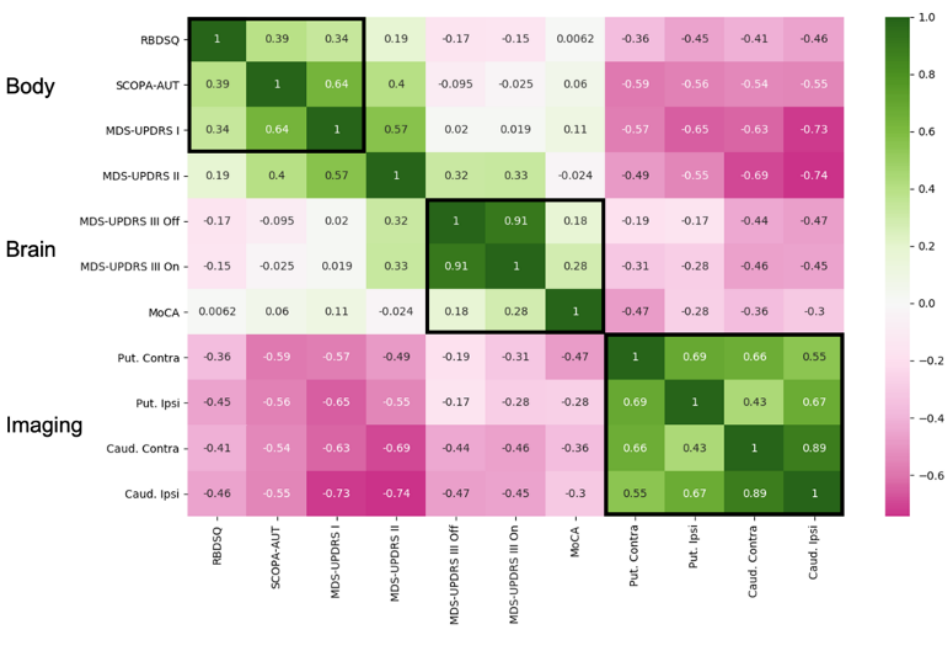


Figure 4.8 – Correlation matrix of inter-marker spacings showing a division of endpoints into three groups: body-related clinical endpoints, brain-related clinical endpoints and imaging biomarkers. Green color means positive correlations. Pink color means negative correlations.

before affecting the caudate 5.8 years later (Caud. Contra PA = 54.0 ± 0.4 years) both in the contralateral hemisphere. The motor impairments measured with the MDS-UPDRS III became pathological 8.8 ± 0.4 years before diagnosis (PA = 52.3 ± 0.5 years). The pathological threshold was estimated between the value 10 and 11 on a scale of 132. By contrast, the patient reported motor impairment in MDS-UPDRS II became pathological 4.6 years before diagnosis (PA = 56.5 ± 0.4 years), so about 4 years after the more objective measurements of motor impairment in MDS-UPDRS III became pathological. The disease course ended with the non-motor impairments: alterations of autonomic functions (SCOPA-AUT PA = 58.9 ± 0.3 years), daily living activities (MDS-UPDRS I PA = 59.6 ± 0.4 years), sleep disorders (RBDSQ, PA = 60.4 ± 0.4 years) and cognitive impairment occurred last right after the age at diagnosis (MoCA, PA = 61.2 ± 0.5 years).

This scenario of progression was confirmed for clinical endpoints using the multidomain course map pooling all data of the three cohorts (936 patients). Multi-domain clinical course map showed similar progression of clinical parameters (Supplementary Figure 4.12).

Body-related signs, brain-related signs and imaging: three components of PD progression

Figure 4.7 shows the sequence of pathological events in a typical PD patient. Individual parameters depicted how this sequence varied across individuals. The age at onset showed a standard deviation of 9.8 years across subjects. The acceleration factor ranged from 0.58 (first decile) to 2.21 (last decile), meaning that PD patient typically progressed between

0.58 times and 2.21 times the average speed of progression. Inter-marker spacing measured how early or late was the onset of one marker relatively to the other ones, once age at onset and pace of progression had been normalized. The inter-marker spacings with the greatest variability were RBDSQ and MoCA with a standard deviation of 10.0 years for both. The third highest variability was found in SCOPA-AUT with a standard deviation of 6.9 years, followed by MDS-UPDRS III On (6.6 years) and MDS-UPDRS I (5.4 years). The standard deviations for the other markers, MDS-UPDRS II, MDS-UPDRS III Off and imaging, ranged between 2.4 and 4.4 years.

The correlation matrix clearly showed three clusters of endpoints that were correlated among themselves (Figure 4.8): the group of body-related clinical endpoints (RBDSQ, SCOPA-AUT, MDS-UPDRS I), the group of brain-related clinical endpoints (MDS-UPDRS III On, MDS-UPDRS III Off, MoCA) and the group of imaging biomarkers (Caud and Put). MDS-UPDRS II inter-marker spacing correlated with markers in both body and brain-related endpoints. MDS-UPDRS II is indeed a mixed endpoint reporting subjective motor impairments in daily-living activities.

Interestingly, the progression of imaging biomarkers did not correlate strongly with the progression of either the body or the brain related clinical endpoints. The negative correlations of imaging biomarkers were stronger with the body-related endpoints, suggesting that earlier dopaminergic loss was associated with a later onset of body-related signs, but was not associated with earlier motor impairments.

PD patients with RBD experience later dopaminergic loss and earlier dysautonomia

Next, we explored disease heterogeneity in PD patients by comparing PDRBD+ and PDRBD- patients. We focused on sleep dysfunction because was found that the heterogeneity in PD progression was predominantly due to different timing in the onset of sleep and cognitive disorders compared to imaging biomarkers and other clinical endpoints and because RBDSQ changes progressed much earlier and faster than MoCA changes (Figure 4.7). In the multi-domain course map, PDRBD+ patients had earlier disease onset (2.8 years), earlier alterations of daily living activity (1.5 years) and autonomic dysfunction (2.5 years) than PDRBD- patients (Figure 4.9, Table 4.3). In contrast, motor symptoms occurred 1.3 years later in Off state and 1.5 years later in On state. The difference in pace of progression was not significant. The same analysis performed in the multimodal course map of the PPMI patients showed the same significant associations with similar effect sizes (Supplementary Table 4.5). In addition, imaging biomarkers tended to show slightly later alterations in PDRBD+ as compared to PDRBD- patients, although this difference was not significant.

The very strong association between age at baseline and onset of individual parameter showed that patients in these studies were recruited at similar disease stages, e.g. Parkinsonian Ages. Pace of progression was also significantly associated with age at baseline: patients who started the disease at a later age progressed 1% faster for each additional year at baseline (CI=[0.63%, 1.4%], $p \leq 0.001$ in multi-domain clinical map). Older patients had SBR alterations significantly later in the contralateral putamen (0.09 years, CI=[0.064, 0.12], $p \leq 0.001$) and at an equivalent stage of disease in the caudate (0.05 years, CI=[0.019,

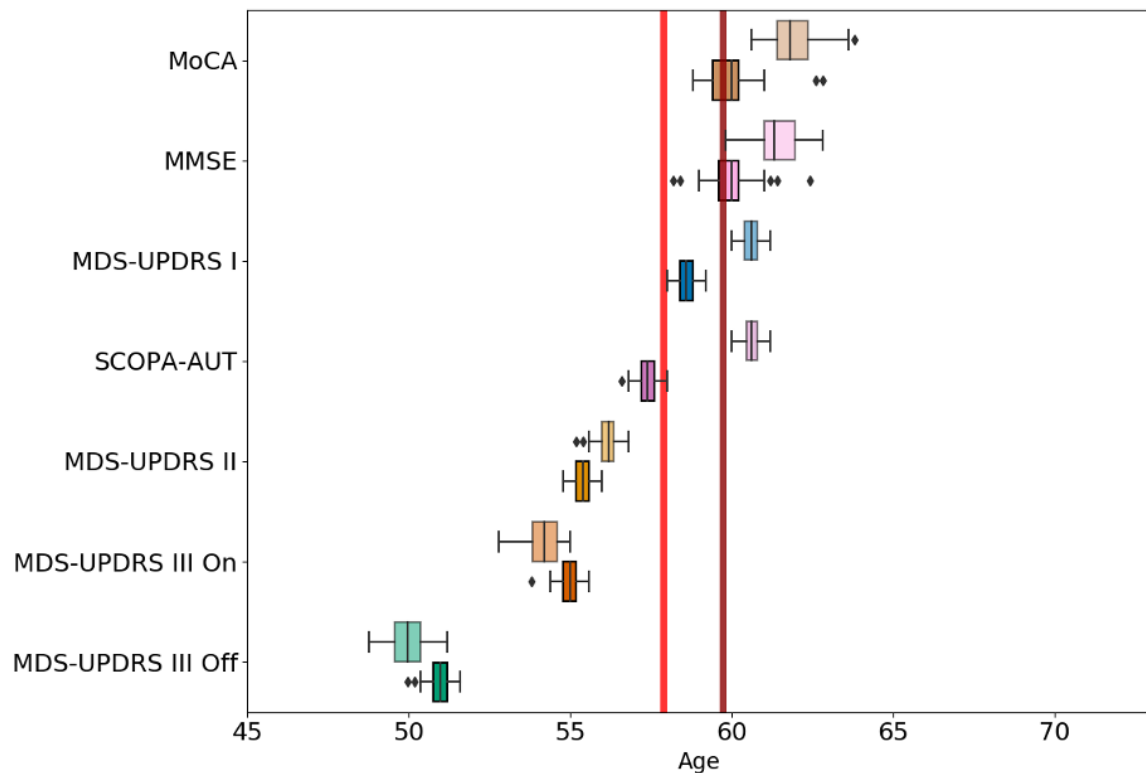


Figure 4.9 – Multi-domain clinical course map trained on PPMI PD patients shows the different timing of symptom onset for PDRBD+ and PDRBD- patients. Parkinsonian Ages at which clinical endpoints become pathological for PDRBD+ (plain colors) and PDRBD- (transparent colors) are shown. Vertical lines indicate the estimated Parkinsonian Age of first symptoms (light red) and of diagnosis to PD (dark red).

	DIGPD	ICEBERG	Gender	Age at Baseline	PDRBD+
Onset	-2.3 (-3.0, -1.6) p=1.5e-08	-3.2 (-4.3, -2.1) p=4.6e-08	1.3 (0.54, 2.0) p=0.0055	0.78 (0.74, 0.81) p=1.2e-219	-2.8 (-3.5, -2.1) p=8.7e-14
Speed	0.025 (-0.056, 0.11) p=4.9	-0.31 (-0.48, -0.25) p=8.2e-09	-0.077 (-0.16, -0.0021) p=0.4	0.01 (0.0063, 0.014) p=3.4e-06	0.089 (0.01, 0.16) p=0.24
MDS-UPDRS I	-0.73 (-1.5, 0.03) p=0.54	-1.9 (-3.0, -0.76) p=0.0085	-2.0 (-2.7, -1.3) p=1e-06	0.096 (0.06, 0.13) p=3e-06	-1.5 (-2.2, -0.8) p=0.00029
MDS-UPDRS II	-0.21 (-0.76, 0.34) p=4.1	-0.19 (-0.99, 0.6) p=5.7	0.62 (0.091, 1.2) p=0.2	0.12 (0.093, 0.15) p=5.6e-17	-0.05 (-0.56, 0.46) p=7.6
MDS-UPDRS III Off	-0.28 (-0.9, 0.33) p=3.3	-1.8 (-2.7, -0.88) p=0.00089	0.39 (-0.2, 0.98) p=1.7	0.047 (0.018, 0.077) p=0.016	1.3 (0.71, 1.9) p=0.00011
MDS-UPDRS III On	-0.25 (-1.0, 0.53) p=4.7	-2.5 (-3.6, -1.3) p=0.00019	0.78 (0.024, 1.5) p=0.39	0.013 (-0.025, 0.051) p=4.5	1.5 (0.79, 2.3) p=0.00045
SCOPA-AUT	0.1 (-0.97, 1.2) p=7.7	-0.91 (-2.5, 0.64) p=2.3	-0.83 (-1.9, 0.2) p=1.0	-0.084 (-0.14, -0.032) p=0.014	-2.5 (-3.5, -1.5) p=9.7e-06
MoCA	1.2 (-0.13, 2.5) p=0.7	4.3 (2.4, 6.2) p=6.5e-05	0.77 (-0.48, 2.0) p=2.0	-0.26 (-0.32, -0.2) p=1e-14	0.75 (-0.46, 2.0) p=2.0
MMSE	0.69 (-0.16, 1.5) p=0.99	2.9 (1.7, 4.2) p=2.7e-05	0.27 (-0.55, 1.1) p=4.7	-0.13 (-0.17, -0.093) p=2.2e-09	0.75 (-0.049, 1.5) p=0.59

Table 4.3 – Association between individual parameters in the multi-domain clinical course maps and the RBD status (PDRBD+ and PDRBD-) with correction for cohort effect, sex and baseline age. Statistically significant associations are shown in bold (corrected p-value are shown). Terms into brackets correspond to confidence interval.

0.076], $p=0.013$) and showed a more symmetrical dopaminergic neuron loss than younger PD patients for caudate in the multimodal course map.

Women experienced earlier alterations in activities of daily living (2.0 years, $CI=[1.3, 2.7]$, $p=1e-06$ in the multi-domain course map, 1.6 years, $CI=[0.43, 2.7]$, $p=0.083$ in the multimodal course map) but a later disease onset by 1.3 years ($CI=[0.54, 2.0]$, $p=0.0055$ in the multidomain course map).

PDRBD+ patients present an intermediate phenotype between PDRBD- and iRBD patients

In iRBD patients, the sequence of pathological events patients (Figure 4.10) presented an inversed order to that of PD patients of multimap (Figure 4.7) and both PDRBD+ and PDRBD- of the multidomain course map (Figure 4.9). Sleep symptoms occurred first at $PA=55.2 \pm 1.1$ years, followed by dysautonomia at $PA=63.6 \pm 0.3$ years, cognitive impairment at $PA=64.6 \pm 0.7$ years, and deterioration of daily living activities at $P1=67.7 \pm 0.2$ years. Motor abnormalities occurred at $PA=71.5 \pm 0.2$ years for MDS-UPDRS II and $PA=73.0 \pm 0.2$ years for the MDS-UPDRS III Off.

Motor symptoms tended to occur later in the disease course and non-motor symptoms earlier in PDRBD+ compared to PDRBD- patients (Figure 4.9). The same trend was observed in iRBD patients although the non-motor symptoms occurred before motor symptoms.

We overlayed in Figure 4.11 the distributions of the individual parameters characterizing progression in iRBD, PDRBD+ and PDRBD-. Speed of progression was similar in the three

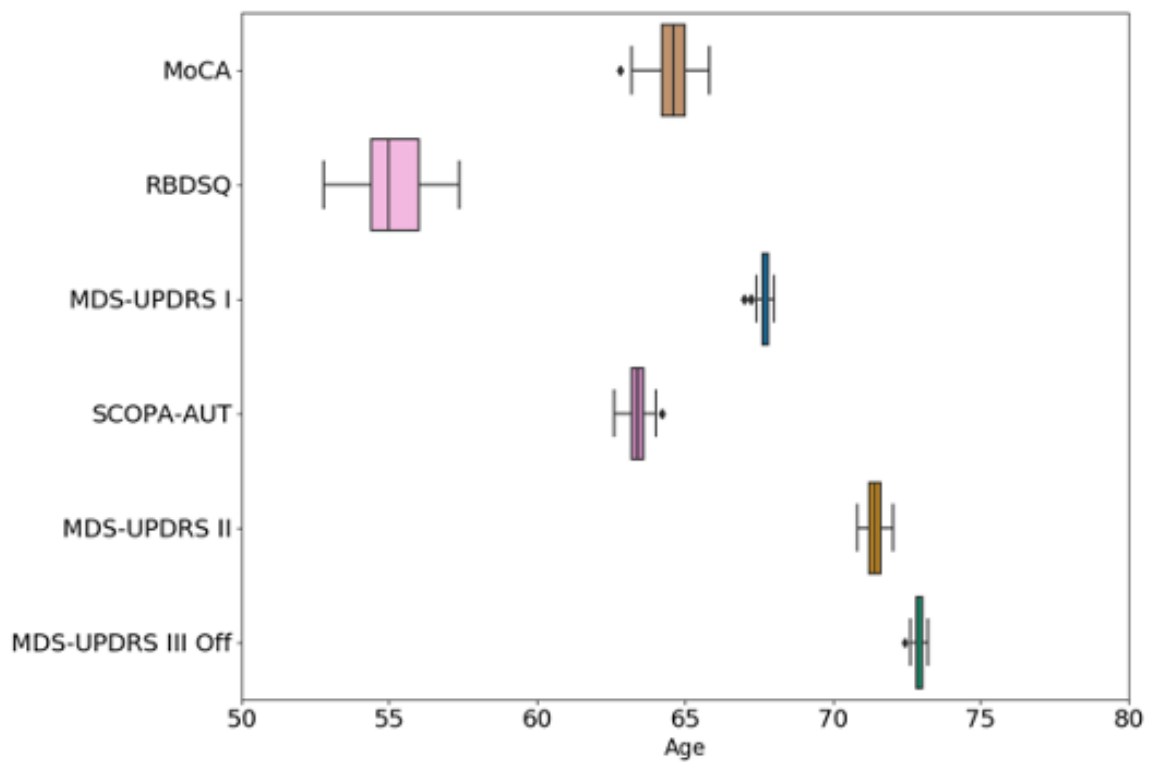


Figure 4.10 – Multi-domain clinical course map trained on PPMI PD patients shows the different timing of symptom onset for iRBD patients. The abnormalities begin in sleep and cognition, followed by dysautonomia and activities of daily living, and finish with motor symptoms.

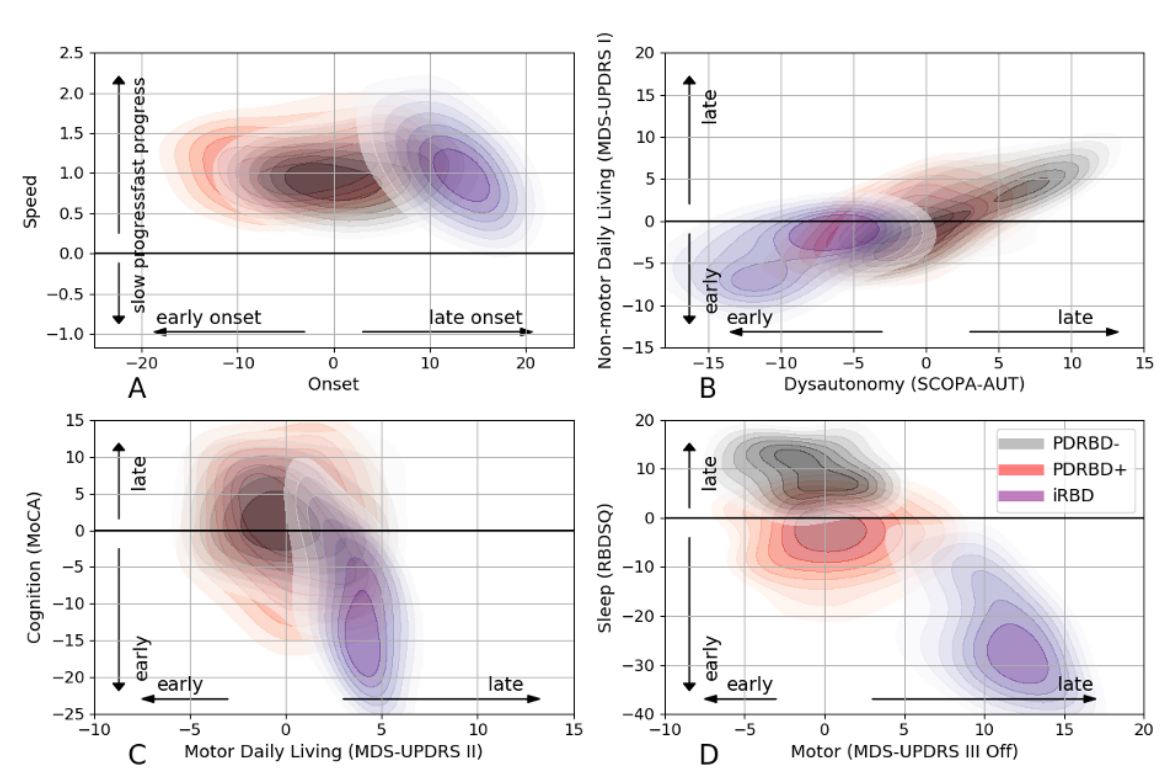


Figure 4.11 – Individual parameter distributions per subgroup, shown via four 2D kernel density estimations. Distribution of progression parameters position PDRBD+ patients as an intermediate phenotype compared to PDRBD- and iRBD.

A) Shows temporal individual parameters, while B-C-D) show spatial individual parameters, aka intermarker spacings, accounting for different patient phenotypes after patients temporal realignment. In A) we show Onset in the x-axis against Speed of progression in the y-axis. We observe no significant difference in terms of speed of progression between the groups, while Onset is earlier in PDRBD+ than PDRBD-. iRBD patients have late onset. Indeed, they are recruited at a late age, and with fewer symptoms than PD patients. Consequently the model re-align them at the beginning of the PD Course Map. In B) we show non-motor activities of daily living (y-axis) against dysautonomia (x-axis), which are correlated, as already mentioned in Figure 4.8. iRBD appear as having an early non-motor phenotype, compared to PDRBD-, with PDRBD+ in between. In D) this gradation is naturally seen in the y-axis accounting for sleep, while iRBD show very late motor symptoms. Finally in C), cognition and motor aspects of daily living show few difference between PDRBD+ and PDRBD- patients, while iRBD patients show late subjective motor symptoms, and earlier cognitive symptoms.

groups but onset varied, with earlier onset for PDRBD+ than PDRBD- (Figure 4.11A). In contrast, iRBD were associated with later onsets, as they were significantly older while having similar or less symptoms overall at baseline. Figures 4.11B-D showed distributions in inter-marker spacings, in years. For the trade-off between sleep disturbances and the onset of motor symptoms, PDRBD+ were positioned in-between iRBD and PDRBD- populations, (Figure 4.11D). PDRBD- patients had earlier motor changes and later sleep disorders, while iRBD had later motor changes and earlier sleep disorders and PDRBD+ patients had intermediate values between the other two groups. This gradation was also seen in daily living activities and dysautonomia (Figure 4.11B). Consistently with Table 4.3, cognition did not separate PDRBD+ and PDRBD- (Figure 4.11C). In general, iRBD patients had very different distributions than the other two groups (Figures 4.11B-D), with earlier modifications of daily living activities and dysautonomia, and later modifications of motor, as an extreme form of PDRBD+. They also showed earlier cognitive decline (Figure 4.11C).

4.3.4 Discussion

We used a longitudinal mixed-effect model to build a normative scenario of progression of Parkinson's Disease, called PD Course Map. This scenario describes quantitatively the progression of Parkinson's Disease, the ordering of abnormalities, as well as the heterogeneity between patients. Heterogeneity is assessed temporally, with high variance in onsets and speeds of progression. But also spatially, with possibly different patterns at equivalent disease stage. Most noticeably, sleep and cognitive symptoms had the highest inter-patient variability in our analysis. Focusing on sleep because of its precocity, we split the PD population of PPMI in 2 subgroups: PDRBD+ and PDRBD- and compared their progressions patterns in both clinical and imaging measurements. PDRBD+ were beginning the disease earlier, and showed a more non-motor form of the disease than PDRBD-. We argue that this came from a clear separation of biomarkers, in 3 blocks: motor, non-motor and imaging. We replicated the results on the pooling of PPMI, ICEBERG and DIGPD, on clinical measurements only. Finally, PD Course Map was applied on PPMI iRBD patients, which appeared as extreme PDRBD+ patients.

By studying PD Course Map in its early stages, we derived an ordering in biomarker abnormality arrival for the overall PD patient population, with a notion of confidence. Damage in the Putamen contralateral occurs first 12.9 years before diagnosis, followed by motor symptoms, non-motor symptoms, and finally cognition. Note that motor appear rather early, even before Caudate abnormalities. It might be that in the preclinical stage, subtle but still significant onset of MDS-UPDRS III alone may not be sufficient for a clinician to diagnose the patient with PD. To the knowledge of the authors, few multimodal models of progression have been derived for Parkinson's Disease. Very recently, (Oxtoby 2021) proposed an event based model to build a normative progression scenario for PD patients at risk of developing dementia. However, most studies only model individual regressions, or focus on specific biomarkers with independent univariate models. We stress that PD Course Map models the joint progression of multiple biomarkers at once. Consistent results with multiple approaches in modeling PD progression will help strengthen

our knowledge in quantifying PD.

In capturing the heterogeneity of PD patients, PD Course Map clearly separated 3 blocks of biomarkers: imaging, motor and non-motor. These blocks are longitudinally correlated in the sense that they explain variations from the average trajectory (e.g. earlier in motor, but later in imaging). These blocks are intuitive, but we could have expected that motor would be more linked to imaging than of non-motor. Results indicate that heterogeneity comes rather independently from these 3 blocks.

Additionally, variance is highest in the non-motor block, thus explaining most of PD's heterogeneity, especially in sleep symptoms. This might support the brain first versus body first theory, with possibly different pathophysiological mechanisms underneath. However, in our case we argue that this highest variability for non-motor symptom could come from patient re-alignment for motor symptom at baseline, as cohort inclusion criteria. If we consider Parkinson's Disease not as a motor disease only, but also encompassing a broader range of symptoms, then our data are biased toward motor onset.

RBD especially, is a rather simple way to cut in PD heterogeneity, as a good proxy of non-motor symptoms. PDRBD+ and PRBD- experience significantly different patterns of progression, mostly in non-motor symptoms. However in the light of the identification of these 3 blocks of variables, we argue that at the scale of our study, this splitting is artificial, as sleep, and non-motor in general rather describe a spectrum of PD profile, rather than well separated entities. iRBD would be positioned at an extremity of this non-motor variability axis.

4.3.5 Supplementary Materials

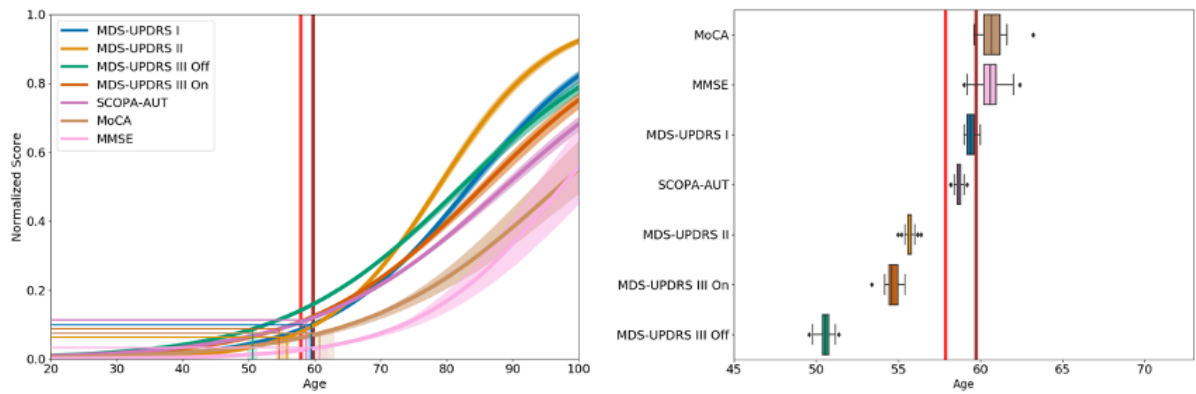


Figure 4.12 – Clinical disease course map trained on the pooling of PPMI, ICEBERG and DIGPD. It describes the progressive onset of clinical endpoints and imaging biomarkers from pre-symptomatic to post-diagnosis stages. Left panel: the progression of the endpoints from the most normal (0) to the maximum pathological changes (1) is shown as Parkinsonian Age progresses. Shaded areas correspond to 95% bootstrap confidence interval. Right panel: the timing at which endpoints become abnormal is shown with 95% confidence intervals. Vertical lines indicate the estimated Parkinsonian Age of first symptoms (light red) and of diagnosis to PD (dark red).

	PPMI vs ICEBERG	PPMI vs DIGPD	ICEBERG vs DIGPD
Age at Baseline	-0.7, p=0.212	-0.39, p=0.266	0.31, p=0.348
Disease Duration (y)	-0.9, p=5.78e-23	-2.02, p=1.16e-94	-1.13, p=1.54e-13
Hoehn and Yahr Off	-	-0.48, p=6.26e-09	-
MDS-UPDRS I	-3.9, p=1.55e-19	-2.19, p=8.3e-13	1.71, p=8.63e-06
MDS-UPDRS II	-2.61, p=9.01e-12	-1.78, p=1e-10	0.82, p=0.00765
MDS-UPDRS III Off	-9.97, p=6.32e-26	-	-
MDS-UPDRS III On	-3.07, p=0.000685 (V04) for PPMI	3.56, p=2.54e-06 (V04) for PPMI	6.63, p=3.14e-13
SCOPA-AUT	-2.89, p=2.12e-05	-1.57, p=0.000471	1.33, p=0.0386
MoCA	-0.51, p=0.02	-	-
MMSE	-	-	0.77, p=4.42e-05

Table 4.4 – Mann-Whitney test between biomarkers values at baseline between PD patients of different cohorts: PPMI and ICEBERG, PPMI and DIGPD, and ICEBERG and DIGPD. Age distribution are matched, but disease duration differs significantly between cohorts. PPMI has the lower disease duration at baseline, followed by ICEBERG and DIGPD. This explains lower severity of biomarkers at baseline in PPMI with respect to both DIGPD and ICEBERG. However, ICEBERG has lower disease duration than DIGPD, but higher biomarker severity at baseline, which is counter-intuitive and shows a bias. Compared to PPMI, it shows a noticeably high difference of MDS-UPDRS III Off at baseline.

	Sex	Age at Baseline	PDRBD+
Onset	0.54 (CI=[-0.28, 1.4]), p=2.4	0.87 (CI=[0.83, 0.91]), p=4.2e-155	-1.9 (CI=[-2.7, -1.1]), p=3.2e-05
Speed	-0.079 (CI=[-0.19, 0.025]), p=1.6	0.0056 (CI=[0.00037, 0.011]), p=0.43	0.093 (CI=[-0.013, 0.19]), p=1.0
MDS-UPDRS I	-1.6 (CI=[-2.7, -0.43]), p=0.083	0.002 (CI=[-0.053, 0.057]), p=11.0	-2.3 (CI=[-3.4, -1.2]), p=0.00039
MDS-UPDRS II	0.81 (CI=[-0.063, 1.7]), p=0.83	0.024 (CI=[-0.018, 0.067]), p=3.2	-1.2 (CI=[-2.1, -0.39]), p=0.049
MDS-UPDRS III Off	0.5 (CI=[-0.39, 1.4]), p=3.2	-0.012 (CI=[-0.056, 0.032]), p=7.1	0.87 (CI=[0.022, 1.7]), p=0.53
MDS-UPDRS III On	1.5 (CI=[0.18, 2.8]), p=0.32	-0.094 (CI=[-0.16, -0.029]), p=0.056	0.95 (CI=[-0.32, 2.2]), p=1.7
SCOPA-AUT	-0.83 (CI=[-2.2, 0.55]), p=2.8	-0.11 (CI=[-0.18, -0.041]), p=0.02	-3.7 (CI=[-5.0, -2.4]), p=5.9e-07
MoCA	2.6 (CI=[0.55, 4.6]), p=0.16	-0.43 (CI=[-0.53, -0.33]), p=3.1e-15	-0.72 (CI=[-2.7, 1.2]), p=5.6
Putamen contralateral	-0.7 (CI=[-1.3, -0.14]), p=0.18	0.091 (CI=[0.064, 0.12]), p=1.6e-09	0.73 (CI=[0.2, 1.3]), p=0.088
Putamen ipsilateral	-0.089 (CI=[-0.54, 0.36]), p=8.4	0.024 (CI=[0.0016, 0.046]), p=0.42	0.63 (CI=[0.2, 1.1]), p=0.048
Caudate contralateral	-0.32 (CI=[-0.9, 0.26]), p=3.4	0.048 (CI=[0.019, 0.076]), p=0.013	0.73 (CI=[0.18, 1.3]), p=0.11
Caudate ipsilateral	0.049 (CI=[-0.46, 0.56]), p=10.0	0.01 (CI=[-0.014, 0.035]), p=5.0	0.69 (CI=[0.21, 1.2]), p=0.059

Table 4.5 – Association between individual parameters in the multimodal course maps and the RBD status (PDRBD+ and PDRBD-) with correction for cohort effect, sex and baseline age. Statistically significant associations are shown in bold (corrected p-value are shown). Note that for this experiment, the multimodal course map has been re-trained without RBDSQ as endpoint.

Part III

Longitudinal Auto-encoders

Longitudinal Autoencoder

Outline of the current chapter

5.1 Deep Learning in disease progression modeling	112
5.2 Background	113
5.2.1 Rigidity and computational cost of the spatiotemporal model	113
5.2.2 Autoencoders	113
5.2.3 A Riemanian perspective on autoencoders	113
5.3 A longitudinal autoencoder	115
5.3.1 Longitudinal Dataset	115
5.3.2 Setting constraints on the latent space	115
5.3.3 Estimation	116
5.3.4 Applications	117
5.4 Conclusion	118

In this chapter, we explain our rational for shifting into the autoencoder framework, in order to model high-dimensional longitudinal data such as medical images. For further details, we invite the reader to refer to the conference article from (Louis et al. 2019).

5.1 Deep Learning in disease progression modeling

Deep Learning has witnessed tremendous success in the field of medical imaging, becoming mainstream since the U-NET (Ronneberger, Fischer, and Brox 2015). Deep Learning algorithms for instance, can now segment with very high accuracy cancer tissues (Havaei et al. 2017; Menze et al. 2014; Milletari, Navab, and Ahmadi 2016; Pereira et al. 2016), and even detect neurodegenerative diseases from MRIs such as Alzheimer's (Ansart et al. 2020; Lu et al. 2021).

It offers also a powerful framework for dimensionality reduction, with autoencoders (Kingma and Welling 2013) and generative adversarial networks (Creswell et al. 2018; Goodfellow et al. 2014), by learning a non-linear mappings between the data space and a latent space of lower dimension.

This property makes Deep Learning appealing for longitudinal data analysis, as it alleviates the difficulty of working with high-dimensional data such as images, by learning efficient dimensionality reduction models, and allows to focus on the latent space. In (Zhao, Adeli, et al. 2019), authors leverage the autoencoder framework to encode 3D MRIs of Alzheimer's patient in a latent space, where patient trajectories in the latent space are regressed with regard to age. In (Zhao, Z. Liu, et al. 2020) they build a similar model, however constraining progression to occur in a learned direction of the latent space, using a cosine loss. In their work they only analyze the temporal representation learned, but such tubular data representation in the latent space also inherently induce a "spatial" representation, as presented in the longitudinal spatiotemporal model.

Most importantly, the autoencoding framework allows the building of complex graphical model in the latent space while working with high-dimensional data (Bouchacourt, Tomioka, and Nowozin 2017; Johnson et al. 2016). For time series, (Krebs, Delingette, Ayache, et al. 2020; Krebs, Delingette, Mailhé, et al. 2019) propose to study a generative model for the cardiac cycle seen as a registration task, with a temporal encoder of sequences and a rotation-compatible latent space, specifically suited to the prior on expected motions. SOM-VAE (Fortuin et al. 2019) aims at finding low dimensional representations of time series based on discrete Markov chain transitions between states in the latent space.

Recently, for longitudinal data, (Maxime Louis, Couronné, et al. 2019) propose to use the longitudinal spatiotemporal model of (Schiratti, Allasonniere, et al. 2015b) and transpose it to the autoencoder framework. A longitudinal spatiotemporal is enforced in the latent space, while the encoder and decoder handle the mapping between the low dimensionality latent space representations into the data space. In this sense, (Louis et al. 2019) see the decoder as defining a pushforward, and thus as a way to learn the metric, provided the decoder is smooth. However, with the use of a decoder, we loose the theoretical properties of orthogonality between temporal and spatial variability, which can then only be empirically assessed. (Louis et al. 2019) apply the model to longitudinal data of Alzheimer's Disease, on both scalar and imaging data.

5.2 Background

5.2.1 Rigidity and computational cost of the spatiotemporal model

The model presented in section 2.2.4 allows to model the progression of longitudinal scalar biomarkers on the manifold product of simple riemanian manifolds. Orthogonality conditions in the tangent space disentangle temporal variability, and spatial variability in the data.

However, the proposed instantiation of the spatiotemporal model for scalar data is very rigid (Schiratti, Allasonniere, et al. 2015b). The inter-patient variability can only be modeled by biomarker specific delays with regard to the average trajectory. It seems reasonable to assume that this variability is not enough to capture the different profiles of progression. For example, causal links may exist between biomarkers, and their dynamics might influence each other. These effects, such as an increase in speed for a subset of the biomarkers only cannot be captured with a diagonal metric.

Additionally, in this setup, we stucked to low-dimensional data (typically $K < 10$) due to the costly estimation with the MCMC-SAEM. (Bône, Colliot, and Durrleman 2018) managed to scale to higher dimension in the context of learning deformations for images and meshes data, resorting to specific heuristics. They perform smart initialization, and couple the MCMC-SAEM with gradient steps, obtained via automatic differentiation. Estimation is costly still, and they later also used autoencoders to perform dimensionality reduction on the deformation space (Bône, Maxime Louis, Colliot, et al. 2019).

5.2.2 Autoencoders

An autoencoder is a neural network that learns to copy its input. It performs successive transformations of an input $x \in \mathbb{R}^n$ to a representation of lower dimensionality $z \in \mathbb{R}^p$ with an encoder Φ , and then back to the data space using a decoder Ψ , yielding an approximation $\hat{x} \in \mathbb{R}^n$ of the input data (see Figure 5.1). This task can be seen as performing a non-linear compressed sensing of x .

In practice, encoder and decoder consist in neural networks with multiple layers. A layer is parameterized via σ, W and b , and perform a matrix multiplication with weight W , matrix addition with bias b , and apply a non-linearity σ (e.g. sigmoid or ReLU non-linearities) : $x \mapsto \sigma(Wx + b)$. Convolutional layers have become the norm (LeCun, Bengio, et al. 1995) for imaging data, allowing for efficient dimensionality reduction of images. The autoencoder is trained to minimize with respect to (Φ, Ψ) the loss function:

$$L = \|x - \Psi(\Phi(x))\|_2^2 + L_{reg}$$

5.2.3 A Riemanian perspective on autoencoders

Dimentionality Reduction

It is commonly assumed that real data, such as natural images, are structured, hence the success of convolutional layers which leverage local smoothness (LeCun, Bengio, et

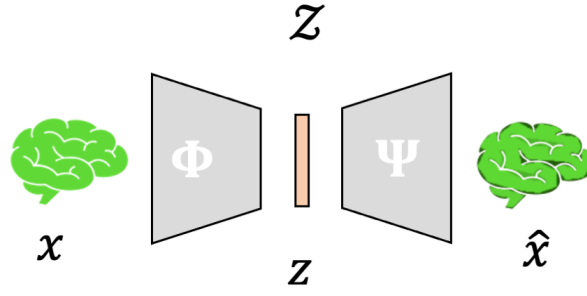


Figure 5.1 – Autoencoder schematic view. Encoder and decoder are composed of many layers: $\Psi = \Psi_1 \circ \dots \circ \Psi_M$, $\Phi = \Phi_1 \circ \dots \circ \Phi_L$, with $\Phi_m : x \mapsto \sigma_m(W_m x + b_m)$ and $\Psi_l : x \mapsto \sigma_l(W_l x + b_l)$.

al. 1995). In this sense, their dimensionality is seen as artificially high, and could be reduced via Machine Learning methods (Bengio, Courville, and Vincent 2013). Principal Component Analysis (Bro and Smilde 2014; Hotelling 1933), Linear Discriminant Analysis (McLachlan 2004) and Non-negative Matrix Factorization (Sra and Dhillon 2006) are standard linear methods for dimensionality reduction, but handling natural images usually requires non-linear methods. Autoencoders allow to perform non-linear dimensionality reduction, and scale easily to high dimensional, large dataset, unlike Isomap (Tenenbaum, Silva, and Langford 2000), t-SNE (Maaten and G. Hinton 2008) or multidimensional scaling (Kruskal 1964). The latent space representation is the subject of extensive research (Burgess et al. 2018; R. T. Chen et al. 2018; Higgins, Amos, et al. 2018; Higgins, Matthey, et al. 2016; Kim and Mnih 2019; Locatello et al. 2019), at it supposedly **captures the heterogeneity of the dataset**, and may be constrained to capture specific parts of this heterogeneity.

Pushforward on the manifold

From a Riemannian geometry perspective (Louis et al. 2019), the decoder Ψ learns a map $z \mapsto x$ from the latent space $\mathcal{Z}$ to the data space. Assuming transformations at each layer are smooth (i.e. each non-linearity σ is smooth), the map Ψ is a smooth map from $\mathcal{Z}$ to $\mathcal{X}$. We can derive from this map a pushforward from $\mathcal{Z}$ to $\mathcal{X}$ 2.1.4 provided we know a metric on $\mathcal{Z}$. With $\mathcal{Z}$ assumed to be euclidian, we can derive a metric from the pushforward of the map Ψ into the data space. Assuming the latent space $\mathcal{Z}$ to be euclidian yields geodesics that are straight lines in $\mathcal{Z}$, which turns very practical for our longitudinal modeling. (Louis et al. 2019) propose to build a mixed effect model in the latent space as in a simple euclidean case.

5.3 A longitudinal autoencoder

5.3.1 Longitudinal Dataset

Let us have N patients, each with m_i visits. At each visit j of patient i , a measurement $y_{i,j} \in \mathbb{R}^n$ is performed, at age $t_{i,j}$. We write $\mathcal{Y} = \{y_{i,j}, i \in [1, N], j \in [1, m_i]\}$ and $\mathcal{T} = \{t_{i,j}, i \in [1, N], j \in [1, m_i]\}$, so that the couple $(\mathcal{Y}, \mathcal{T})$ refers to a longitudinal dataset $\mathcal{D}$.

5.3.2 Setting constraints on the latent space

(Louis et al. 2019) propose to learn latent space representations adapted to the longitudinal setup, by analogy with the longitudinal spatiotemporal model.

Recall that (Schiratti, Allasonniere, et al. 2015b) assume that the average trajectory γ is a geodesic, and that each patient trajectory η^{w_i} is a parallel variation of this average trajectory γ . In our case the latent space of dimension p is assumed Euclidean, with its canonical basis denoted $(e_1, \dots, e_p)$. Consequently γ is simply a straight line of $\mathbb{R}^p$, and $\eta^{w_i}(\gamma)$ are parallels of γ .

We choose e_1 to account for temporal variability, while the other $p - 1$ dimensions $(e_2, \dots, e_p)$ account for inter-patient variability. By analogy with the longitudinal spatiotemporal model, this builds a set of spatiotemporal coordinates (ψ, w_i) , with $\psi \in \mathbb{R}$ denoting a disease stage, while $w_i \in \mathbb{R}^{p-1}$ is fixed for a patient i , and encode for the "spatial" specificity of this patient. Consequently we write the population and individual trajectories:

$$\gamma(\cdot) : \begin{cases} \mathbb{R} \rightarrow \mathbb{R}^p \\ \psi \rightarrow (\psi, 0, \dots, 0) \end{cases} \quad \eta^{w_i}(\gamma)(\cdot) : \begin{cases} \mathbb{R} \rightarrow \mathbb{R}^p \\ \psi \rightarrow (\psi, w_i^1, \dots, w_i^{p-1}) \end{cases} \quad (5.1)$$

and adding patient specific affine reparameterization of age t similarly to (Schiratti, Allasonniere, et al. 2015b), a patient trajectory is traveled at $\psi_i(t) = \alpha_i(t - \tau_i)$ with $\alpha_i = \exp(\xi_i)$:

$$\eta^{w_i}(\gamma)(\psi_i(t)) = (\psi_i(t), w_i^1, \dots, w_i^{p-1})$$

Finally the pointwise decoder maps patient trajectories in the latent space to patient trajectories in the data space:

$$\Psi(\eta^{w_i}(\gamma)(\psi_i(t))) = \Psi((\psi_i(t), w_i^1, \dots, w_i^{p-1}))$$

This way, the network is forced to learn temporal variability in the only dimension we allow him to, the first direction of our latent space e_1 . Learning patient representation w_i is encouraged by enforcing it is shared for all visits of the patient in the other direction of the latent space $(e_2, \dots, e_p)$.

5.3.3 Estimation

Attachment

We use the same attachment term as in the longitudinal spatiotemporal model:

$$L_{fit} = \sum_{i=0}^N \sum_{j=0}^{m_i} \|y_{i,j} - \Psi(\eta^{w_i}(\gamma)(\psi_i(t_{i,j})))\|^2 \quad (5.2)$$

with $(\xi_i, \tau_i, w_i^1, \dots, w_i^p) = \Phi(x)$ and $\alpha_i = \exp(\xi_i)$.

Regularization

In order to preserve the useful mixed effect interpretation of the model, regularization is added on the individual's representations:

$$L_{reg} = \sum_{i=0}^N \frac{\tau_i^2}{\sigma_\tau^2} + \sum_{i=0}^N \frac{\xi_i^2}{\sigma_\xi^2} + \sum_{i=0}^N \frac{\|w_i\|^2}{\sigma_w^2}$$

with hyperparameters $\sigma_\tau, \sigma_\xi, \sigma_w$. In this sense, the "average trajectory" $t \mapsto \gamma(t)$ accounts for $\tau_i = \xi_i = w_i = 0$. Also, to avoid overfitting, a L2 regularization on the weights of the network is applied.

Inference

We seek to minimize $L = L_{fit} + L_{reg}$ with respect to Φ and Ψ . To that end, we do not rely anymore on MCMC methods, as the set of parameters of Φ and Ψ is now orders of magnitude larger. We instead use standard deep learning optimization, based on stochastic

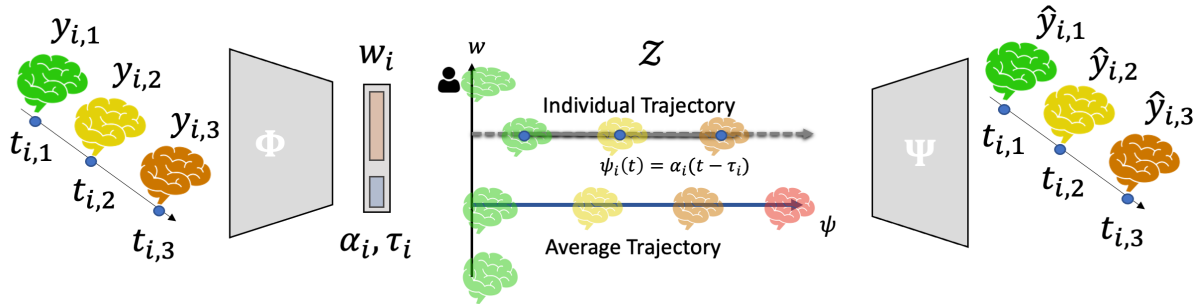


Figure 5.2 – Schematic view of the longitudinal autoencoder. The encoder is a sequence-wise autoencoder, which encodes all patient i 's visits $x_{i,j}$ and times $t_{i,j}$ into a patient representation w_i and affine time reparameterization parameters α_i and τ_i . The point-wise decoder maps points $(\psi_i(t), w_i^1, \dots, w_i^{p-1})$ of the latent space back into the data space. Brains in transparency indicate how the latent space maps to the data space via the decoder. ψ encodes the disease stage, represented by brain colors, while w encodes patient specificity, represented by patient specific brain height and width in the graph.

gradient descent. Gradients are given by the automatic differentiation python library PyTorch. In practice, we train the architecture end to end using Adam optimizer with a learning rate of 0.01.

5.3.4 Applications

Synthetic Dataset

In a first application, (Louis et al. 2019) apply the model to a synthetic longitudinal dataset (see Figure 7.2). Each cross is parameterized by its arms length and angles. A mean scenario of progression for the arm lengths is prescribed, and subjects have specific arm angles, which are sampled from a zero-centered normal distribution. The model is trained on these data, and successfully learns the average progression, as well as the inter-patient variability (see Figure 5.4).

Cognitive scores

In Figure 5.5, we show the model trained on cognitive scores from MCI converters patients in the longitudinal cohort ADNI. The model successfully learns a consistent average trajectory as well as an inter-patient variability close to the one of the longitudinal spatiotemporal model applied on scalar data (Leaspy).

Magnetic Resonance Imaging data

In Figure 5.6, we show the average trajectory of a model trained on 3D MRIs from MCI converters patients in the longitudinal cohort ADNI. We observe that the model captured the main temporal variability of this dataset: the widening of the ventricles.

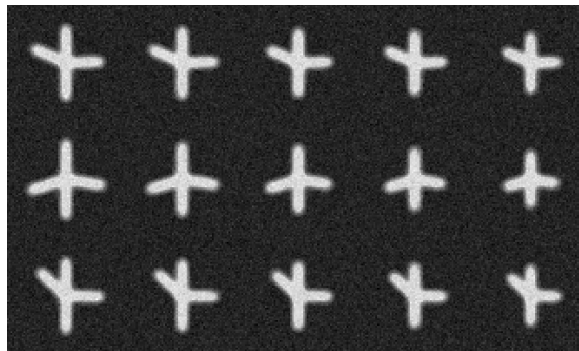


Figure 5.3 – Each row represents a synthetic subject. Columns show visits at increasing times for these subjects. A subject has specific arm’s angles. Disease progression is modeled as a decrease of all arm’s lengths. (Reproduced from (Louis et al. 2019))

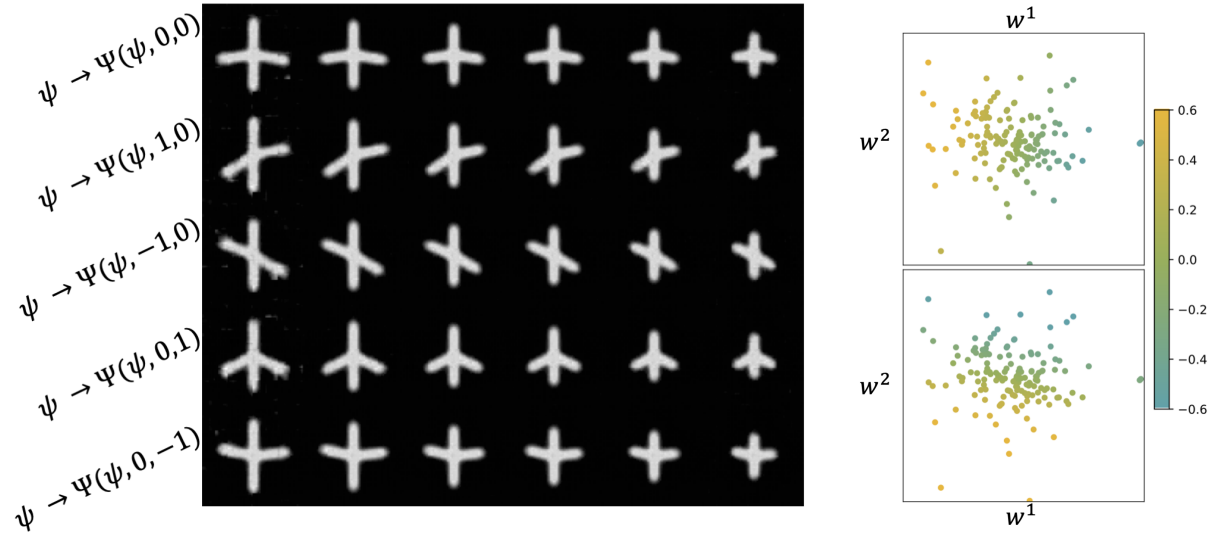


Figure 5.4 – We learn a model on the synthetic cross data, with $\dim(\mathcal{Z}) = 3$, and visualize in these 2 graphs the induced latent space. On the left, we show multiple trajectories over time (x-axis) corresponding to 5 different points in the spatial part of the latent space. Each row is a trajectory with fixed spatial component w , and ψ varies from $\psi = -1$ to $\psi = 1$. The first row depicts the average trajectory $\psi \mapsto \Psi(\psi, 0, 0)$, while the following rows depict parallel trajectories, shifted from $\pm e_2$ or $\pm e_3$ to visualize their effects. These parallel directions of progression show the same arm length reduction scenario, with different arm positions. On the right we show spatial individual variables w_i^1 and w_i^2 colored by left (top) and right (bottom) arm angle of the train data. We observe that w^1 captured the left arm angle, while w^2 captured the right arm angle. (Reproduced from (Louis et al. 2019))

5.4 Conclusion

We saw in this chapter that an autoencoder extension of 2.2.4 is able to conveniently handle high dimensional data, while keeping the attractive representation of patient trajectories as curves on a manifold parallel to an average trajectory.

This framework is very flexible, and we will see in the following sections contributions to extend the model in multiple directions. First, we tackle the multimodality problem by jointly learning and merging representations from multiple modalities, to build trajectories in the latent space in the same manner as (Louis et al. 2019). We then propose to relax the affine time reparameterization, and instead rely solely on the visit ranking to estimate disease stages. Additionally, we shift to the variational autoencoder framework, whereas models of (Louis et al. 2019) and Chapter 6 were not variational. The variational autoencoder is a generative model, unlike classical autoencoders. Additionally, variational autoencoders allow for better disentanglement in the latent space (Higgins, Matthey, et al. 2016).

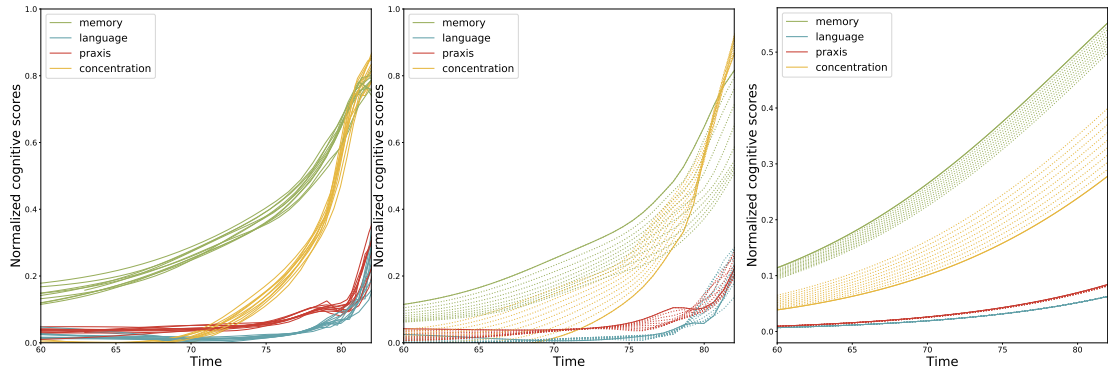


Figure 5.5 – On the left learned average trajectory of the cognitive scores, with a 10-fold resampling procedure ($\dim(\mathcal{Z}) = 2$). On the center, spatial variation in the direction e_1 . On the right, spatial variation of the longitudinal spatiotemporal model, given by Leaspy. (Reproduced from (Louis et al. 2019))

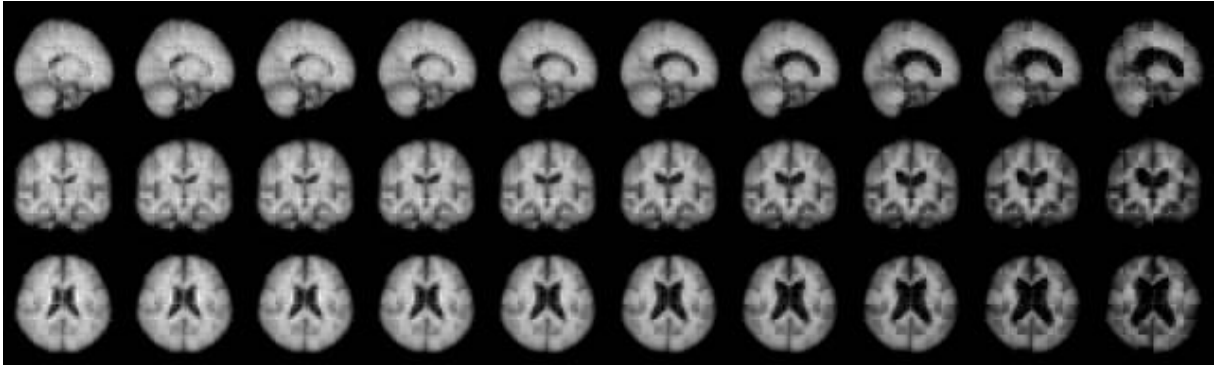


Figure 5.6 – Average trajectory $\psi \mapsto \Psi(\psi, 0, \dots, 0)$ learned from MRIs of ADNI MCI converters patients. Rows show respectively sagittal, coronal and horizontal planes of the average trajectory. The growth of the ventricles, characteristic of aging and Alzheimer's disease is clearly visible. (Reproduced from (Louis et al. 2019))

Chapter

6

Longitudinal autoencoder for multi-modal disease progression modelling

Outline of the current chapter

6.1 Introduction	122
6.2 Methods	123
6.3 Decoding : Non linear mixed effect model	123
6.4 Encoding	124
6.5 Regularization, cost function and optimization	124
6.6 Experimental results	125
6.6.1 Cognitive scores: proof of concept	125
6.6.2 A synthetic dataset	126
6.6.3 Application to Alzheimer's disease future image prediction . .	127
6.7 Conclusion and perspectives	128

In this chapter, we naturally extend the model described in Chapter 5 to handle multimodal data.

Abstract

Imaging modalities and clinical measurement, as well as their time progression can be seen as heterogeneous observations of the same underlying disease process. The analysis of sequences of multi-modal observations, where not all modalities are present at each visit, is a challenging task. In this paper, we propose a multi-modal autoencoder for longitudinal data. The sequences of observations for each modality are encoded using a recurrent network into a latent variable. The variables for the different modalities are then fused into a common variable which describes a linear trajectory in a low-dimensional latent space. This latent space is mapped into the multi-modal observation space using separate decoders for each modality. We first illustrate the stability of the proposed model through simple scalar experiments. Then, we illustrate how information can be conveyed from one modality to refine predictions about the future using the learned autoencoder. Finally, we apply this approach to the prediction of future MRI for Alzheimer’s patients.

6.1 Introduction

The longitudinal pattern of progression of a disease contains more information than a static observation. Leveraging this information is a key problem in machine learning for healthcare, complicated by the nature of clinical datasets. These datasets may contain very heterogeneous observations from various modalities of subjects at multiple time points, such as clinical scores, imaging and biological samples. They include missing values, often by design: not all modalities are observed at each visit. Besides, the number of observations and their time spacing vary between subjects. For these reasons, the analysis of multiple modalities and their time dynamic at once is a challenging task.

Linear mixed effect model estimated via EM and their extension to the non-linear case (Laird and Ware 1982b; Lindstrom and Douglas M. Bates 1990b) were developed for the analysis of unimodal longitudinal data. More recently, recurrent auto-encoder (Rumelhart, G. E. Hinton, Williams, et al. 1988; Srivastava, Mansimov, and Salakhudinov 2015) offer a way to encode trajectories into a low-dimensional embedding, allowing to perform unsupervised clustering of the trajectories (Falissard et al. 2018). Riemannian geometry based approaches such as (Louis et al. 2019; Schiratti, Allasonniere, et al. 2015c) offer ways to learn sub-manifolds of the observation space with a system of coordinate adapted to the progression of the modality observed in the data.

On the other hand, various unsupervised methods exist to fuse information from multiple modalities but from a single time snapshot. In (Chartsias et al. 2018; Ngiam et al. 2011), the authors propose to learn a common embedding for multiple modalities auto-encoding, merging the information from all modalities and allowing the generation of missing modalities. In (Miotto et al. 2016), unsupervised features are learned from heterogeneous health data as a dimensionality reduction method before machine learning tasks.

In (Yang et al. 2017), combining time and multi-modal approaches, the authors propose a setting for multi-modal time-series embedding. But their design does not handle missing

modalities, common in clinical data sets. Besides, the fusion of the information from the different modalities is done at each time step and not on the progression pattern globally, thus decreasing the importance of the dynamics of each modality in the encoding.

To address these limitations, we propose a new setting for longitudinal multi-modal encoding. We extend to the multi-modal case the approach of (Louis et al. 2019). Each modality is first separately encoded using a recurrent neural network. A fusion network is then used to merge the obtained representations into a unique representation, which describes the multi-modal trajectory of the subject as a time-parametrized linear trajectory in a latent space $\mathcal{Z}$. Then, this trajectory is decoded using a different neural network for each modality, which generates continuously varying trajectories of data changes. This setting allows to handle multiple modalities even when not all of them are observed at each visit and it can handle any number of visits and any time spacing between the visits. Finally, extrapolation in the latent space allows for prediction of the future of each modality and we show on a synthetic dataset and on the ADNI database using cognitive scores and MRI jointly that the predictive power is enhanced by the fusion of each modality embeddings.

In section 7.2 we explain the proposed model, in section 6.6 we present experimental results highlighting the stability of the method on synthetic and real data sets and we show how the information from one modality that contributes to the encoding allows to refine prediction of the future of another modality.

6.2 Methods

We set a longitudinal dataset which contains repeated observations of subjects, where the observations at each time point contain a various combination of modalities among $M \in \mathbb{N}$ modalities. For any subject $i \in \{1, \dots, N\}$ where $N \in \mathbb{N}$ and for any modality $m \in \{1, \dots, M\}$, we have a sequence $(y_{ij}^m, t_{ij}^m)_{j=1, \dots, n_i^m}$ of observations y_{ij}^m of observed at times t_{ij}^m .

6.3 Decoding : Non linear mixed effect model

We set $p \in \mathbb{N}$ and consider a p -dimensional latent space $\mathcal{Z} = \mathbb{R}^p$ and its canonical basis $(\vec{e}_i)_{i=1, \dots, p}$. Then, in the spirit of random slopes and intercepts models, we consider trajectories in $\mathcal{Z}$ of the form $l(t) = e^\eta(t - \tau)\vec{e}_1 + \sum_{i=2}^p \lambda^i \vec{e}_i$ where $\eta, \tau, \lambda_2, \dots, \lambda_p \in \mathbb{R}$ are random variables. These trajectories progress in the $\vec{e}_1$ direction and are translated in any direction orthogonal to $\vec{e}_1$, so that the λ s play the role of random intercepts. η controls the pace of progression while τ allows for a time shift between the trajectories. We consider that the i -th subject follows a trajectory of this form with parameters $\varphi_i = (\eta_i, \tau_i, \lambda_i^2, \dots, \lambda_i^p)$.

For each considered modality m , we consider a nonlinear mapping Ψ_{w_m} which maps $\mathcal{Z}$ on a subspace of the m -th modality observation space. This transports the mixed-effect model formulated in $\mathcal{Z}$ into the corresponding observation spaces. Note that the apparent rigidity of the family of trajectories considered in $\mathcal{Z}$ is not restrictive provided the mappings Ψ_{w_m} are flexible enough. In practice, the Ψ_{w_m} are neural networks, de-convolutional for

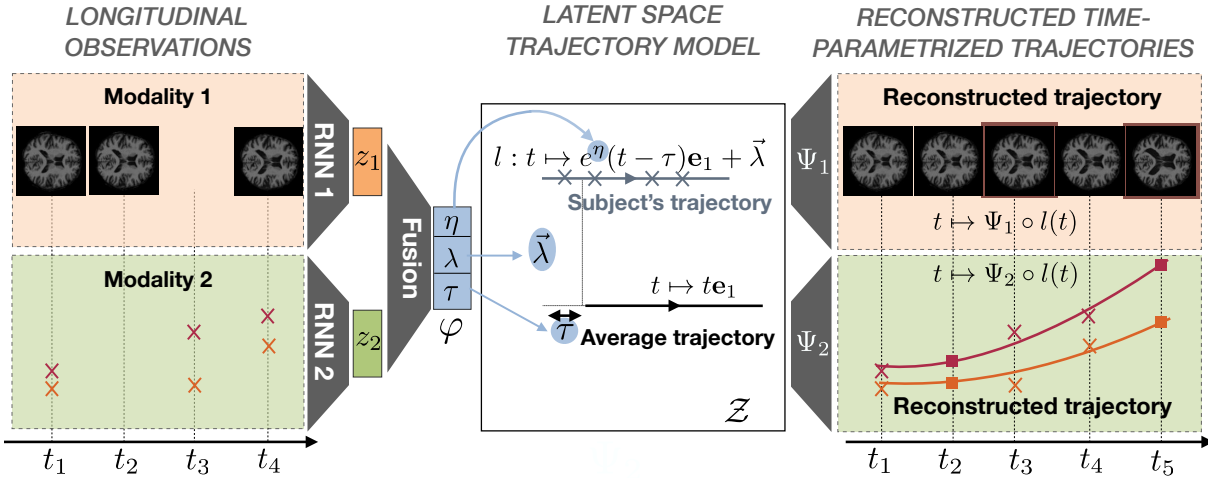


Figure 6.1 – Description of the proposed longitudinal autoencoder.

images and fully connected for scalars. The right half of Figure 6.1 illustrates the procedure. Overall, this setting can be viewed as a non-linear mixed-effect model where the random effects are the φ_i 's and the fixed effects are the parameters of the mappings Ψ_{w_m} .

6.4 Encoding

Individual parameters φ_i are estimated via the use of an encoder network. More precisely, each modality is first processed by a dedicated Recurrent Neural Network (RNN), to get modality-wise representations. To correct for the varying spacings between the observations, we provide to the RNN the visit times, previously normalized to zero-mean and unit variance.

We then concatenate the obtained representations, and use a fully-connected network to merge the representations. The given architecture allows fast inference for new subjects, and is trainable end to end. Besides, the fusion operation is learned so as to produce a single vector which contains the most information about the reconstruction of the whole sequences of all the modalities. The left part of Figure 6.1 illustrates the procedure.

6.5 Regularization, cost function and optimization

To enforce some structure in the latent space and in the family of trajectories obtained, we set the following regularization on the individual variable Φ_i : $r(\eta, \tau, (\lambda_i)_{i=2, \dots, d}) = \eta^2 + \tau^2 + \sum_{j=2}^p (\lambda^j)^2$. This regularization models the η variable to be distributed along a zero-centered normal distribution, which allows the pace of progression to vary typically between 0.2 and 5. times the mean velocity. The τ variable is regularized the same way. This regularization is not arbitrary: during each run, the observation times t_{ij}^m are rescaled to zero-mean unit variance, and thus τ can handle delays between subjects of order the standard deviation of the observation ages.

Overall, the optimized cost function for one subject is the regularization cost added to the ℓ^2 reconstruction cost summed over all modalities:

$$C((w_m)_m, \eta, \tau, (\lambda_i)_i) = r(\eta, \tau, (\lambda_i)_i) + \sum_m \frac{1}{\sigma_m^2} \sum_{j=1}^{n_i^m} \|y_{ij}^m - \Psi_{w_m}(l_i(t_{ij}^m))\|_2^2 \quad (6.1)$$

where the $(\sigma_m)_m$ are trade-off parameters between each modality and the regularization. We set an automatic update rule for these parameters after each batch by setting them to the empirical quadratic errors in reconstruction for the modality over the batch. The estimation is achieved by stochastic gradient descent with the Adam optimizer (Kingma and Ba 2014) and a batch size of 32 subjects. The Decoders are either fully connected or de-convolution networks depending on the kind of modality considered, with standard architectures. The encoders are either Elman networks or Elman networks working on features extracted using a convolution network in the case of images. All networks are trained end to end using back-propagation and the PyTorch library. A complete code to reproduce these experiments will be released upon publication of the paper.

6.6 Experimental results

6.6.1 Cognitive scores: proof of concept

As in (Schiratti, Allasonniere, et al. 2015c), we apply our model on repeated measurement of 4 normalized cognitive score extracted from the ADNI cohort, respectively associated with memory, language, praxis and concentration. We include the 248 MCI-converter subjects, followed for an average of 3 years, over 6 visits. We conduct 2 experiments in order to assess the robustness of the method, and report estimated average trajectories in Figure 6.2, as well as individual reconstruction errors in Table.6.1, computed from a patient-wise 10-fold cross validation.

First, we apply our model on an increasing partitioning of input feature. We consider 3 cases: selecting all scores at once as one modality, selecting separately memory+language and praxis+concentration as two modalities, and selecting each one separately. We note the overall good stability of the average model over multiple multi-modal architectures, with stability decreasing in the 4-modalities scenario, arguing for a concatenation of the consistent features.

In our second experiment, we assess the robustness of the model with the number of visits per subjects. To this end we consider the 2-modalities scenario, and perform a pruning of the dataset, removing an increasing number of visits of the second modality, i.e. praxis+concentration per subjects. Datasets are obtained from pruning frequencies of respectively 10%, 20% and 40%. Here we also observe an overall good stability of the average trajectory over pruning frequency.

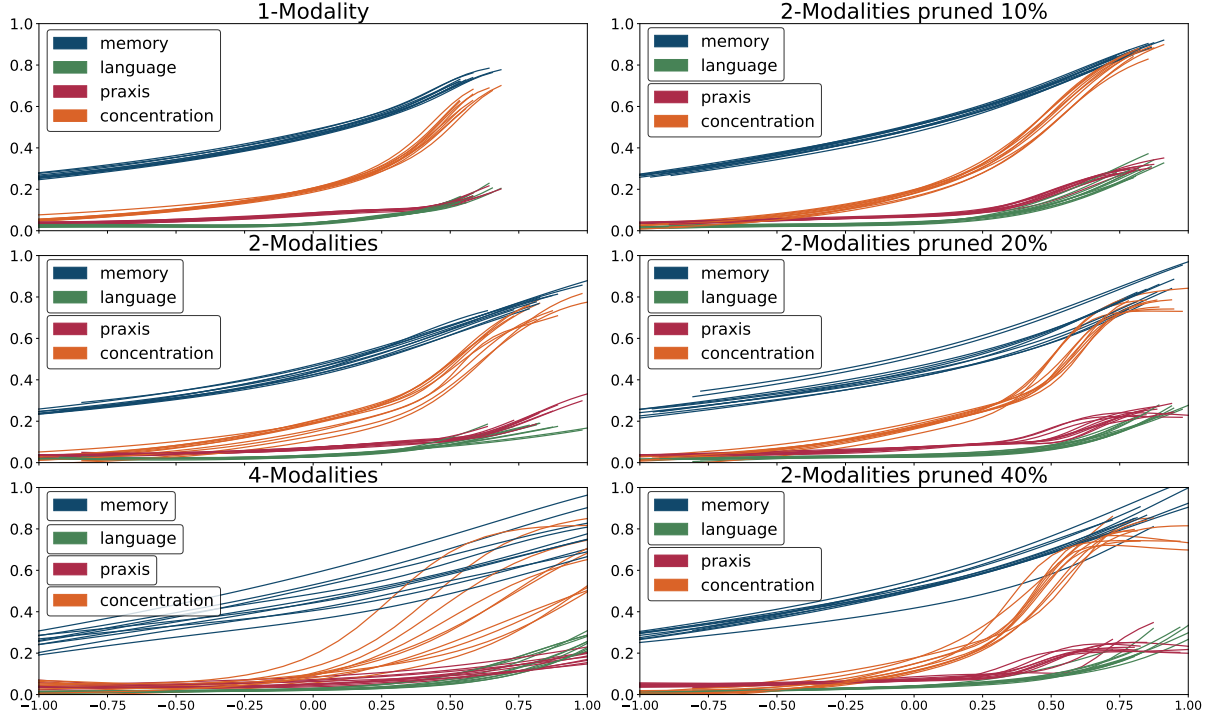


Figure 6.2 – Left: average trajectories for the 10 folds, with increasing partitioning of the input features. Right: average trajectories for the 10 folds, with increasing pruning of the praxis+concentration modality.

6.6.2 A synthetic dataset

To test the proposed setup in realistic conditions, we generate a synthetic multi-modal data set comprising 300 subjects observed 7 times on average. The first modality is a 2D image of a cross, with varying arm lengths and angles while the second modality consists of two scores with a sigmoid-like growth. We set a time reparametrization function s with parameters a_1, a_2 defined by: $s_{a,b}(t) = t + a \text{sign}(t)t^2 + bt^3$. To generate an individual, we sample two sets of parameters $(a_k, b_k)_{k=1,2}$. These serve to reparametrize a scenario of score increase: the k -th score for the subject at time t is given by $\sigma \circ s_{a_k, b_k}$ where σ is the sigmoid function. Then, the arms lengths L_1, L_2 for the images of the subject at time t are given by $L_1 = \sigma \circ s_{(a_2-a_1)+\varepsilon_{a1}, (b_2-b_1)+\varepsilon_{b1}}$, $L_2 = \sigma \circ s_{(a_2+a_1)+\varepsilon_{a2}, (b_2+b_1)+\varepsilon_{b2}}$ where the ε are

	Partitioning				Pruning		
	1-mod	2-mod	4-mod		2-mod 10%	2-mod 20%	2-mod 40%
Train ($\times 10^{-3}$)	6.7	3.8 / 9.7	21.1 / 2.2 / 5.6 / 5.3		4.9 / 11.3	4.1 / 11.5	4.5 / 14.6
Test ($\times 10^{-3}$)	7.8	5.1 / 10.6	24.7 / 3.3 / 7.1 / 5.2		4.9 / 11.7	5.0 / 11.9	5.4 / 15.5

Table 6.1 – Mean 10-fold reconstruction error for the 2 cognitive scores experiments for each modality respectively

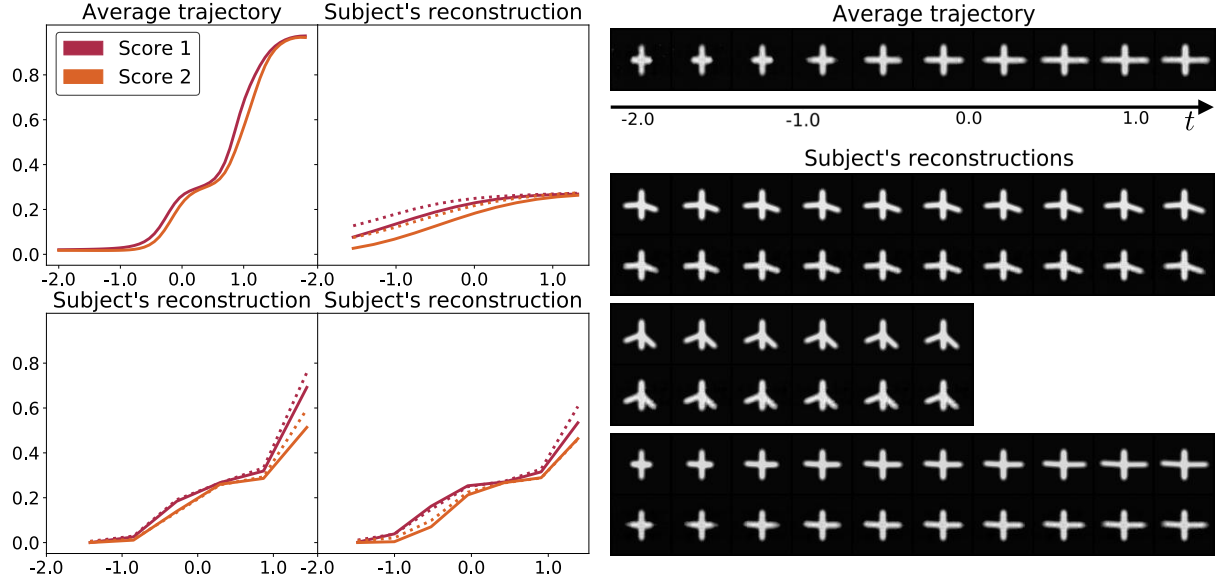


Figure 6.3 – Left: average trajectory and reconstruction examples for the scalar data. Right: average trajectory and some reconstructions for the image data.

samples from a zero-mean normal distribution and constant with time. Finally, the arm angles are sampled along a normal distribution but are not informative of the synthetic disease process. This design is so that the images contain, in an intricate way, information about the progression of the scores materialized through the a_1, a_2, b_1, b_2 variables. The two modalities are different noisy facets of a common underlying process.

We perform a patient-wise 10-fold estimation of the model this data set. Figure 6.3 shows the obtained average trajectory for the first fold, as well as the reconstructions of some subjects images and scores observations. We evaluate and average for all folds the test and train reconstruction errors. For the cross, the test error is $2.0 \cdot 10^{-8} \pm 8. \cdot 10^{-9}$ while the train error is $1.7 \cdot 10^{-8} \pm 3.9 \cdot 10^{-9}$. For the scores, the test error is $7. \cdot 10^{-3} \pm 3. \cdot 10^{-3}$ while the train error is $7. \cdot 10^{-3} \pm 3. \cdot 10^{-3}$. This shows that the model generalizes well to unseen data.

We use the trained model to predict the future scores on the test data. We do so by decoding the extrapolation of the latent trajectory encoded by the model. We repeat this experiment by gradually removing the last observations of the image modality, to look at the impact of this modality on the predictive power of the model. Figure 6.4 shows the experimental setup and the results. As the time span of the observed images shrinks, the prediction deteriorates: when more image data is available, the score prediction is more accurate. This shows the ability of the model to find a relevant common representation for the progressions of the different modalities.

6.6.3 Application to Alzheimer's disease future image prediction

On the 248 patients of section 6.6.1, we apply the same model on the 217 that have at least 1 MRI observation, leading to a total of 1199 cognitive scores measurements and

1441 MRIs. We work on both the MRI images and the cognitive scores. The MRI images are rigidly aligned and sub-sampled to 64^3 resolution. Note that the subjects do not have both the MRI and the cognitive scores measurements at each visit.

Figure 6.5 shows one of the estimated average trajectory for the MRI modality. We evaluate and average for all folds the test and train reconstruction errors on both modalities. For the MRI, the test error is $2.5 \cdot 10^{-3} \pm 6. \cdot 10^{-5}$ while the train error is $2.4 \cdot 10^{-3} \pm 2. \cdot 10^{-5}$. For the scores, the test error is $2.2 \cdot 10^{-2} \pm 3. \cdot 10^{-3}$ while the train error is $1.7 \cdot 10^{-3} \pm 6. \cdot 10^{-4}$. This shows that the model generalizes well to unseen data.

We then perform the same prediction task as in the previous section: we attempt to predict the future MRI from past data, using a variable amount of score data in the past. Figure 6.5 shows the prediction errors for different time horizon. Once again, the errors increase as we feed the model with less cognitive scores measurements. This shows that the model captures information contained in the cognitive scores progression to refine the MRI prediction.

6.7 Conclusion and perspectives

We extended on a deep autoencoder architecture with a mixed effect latent space to propose a practical framework for modeling multi-modal longitudinal data, trainable end-to-end. This allows for analysis of heterogeneous longitudinal datasets, deriving a model-wise average trajectory, as well as condensed patient representations. We study its robustness toward modalities partitioning and dataset pruning and illustrate its utility in both synthetic and real scenarios. In the future we plan to model the progression of more modalities at once. This work has been partially funded by the European Research Council (ERC) under grant agreement No 678304, European Union’s Horizon2020 research and innovation programme under grant agreement No 666992, and the program ”Investissements

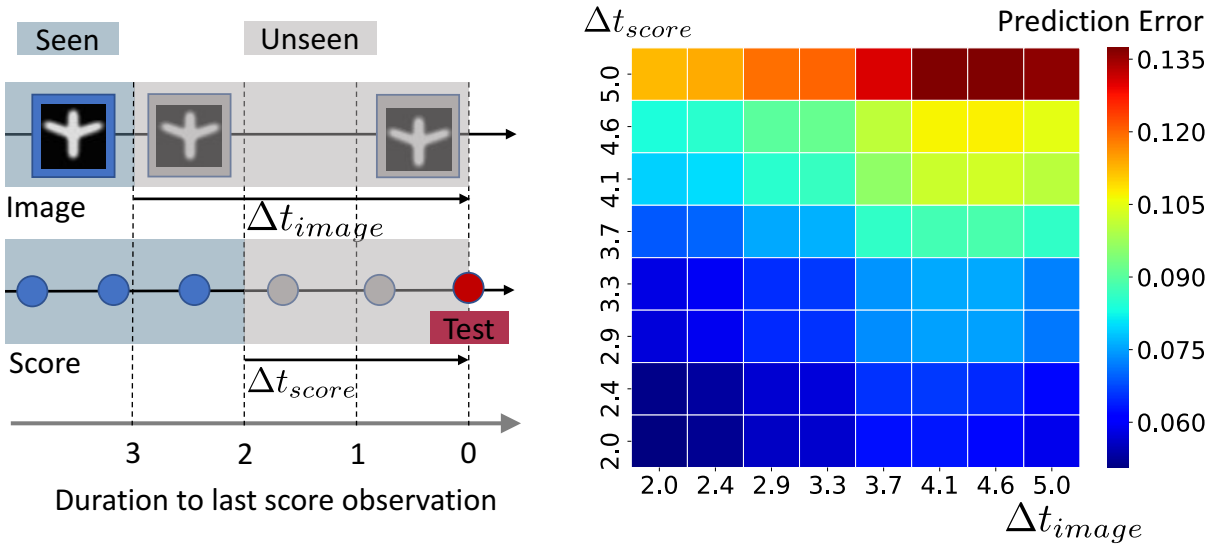


Figure 6.4 – Left: description of the prediction setup. Right: the MRI prediction errors.

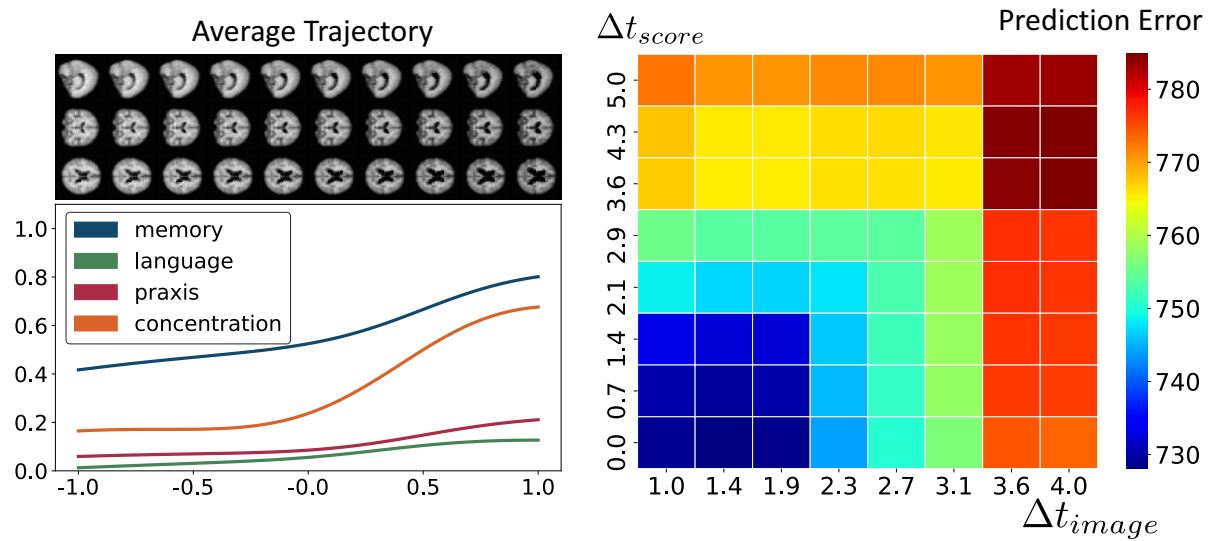


Figure 6.5 – Left: average trajectory. Right: prediction error, in the same setup as in section 3.2

d’avenir” ANR-10-IAIHU-06.

Longitudinal self-supervision to disentangle inter-patient variability from disease progression

Outline of the current chapter

7.1 Introduction	132
7.2 Methodology	133
7.3 Longitudinal progression model	133
7.4 Modularity	135
7.5 Experimental results	136
7.6 Validation on synthetic data	136
7.7 Application to Alzheimer’s disease	138
7.8 Conclusion	139

*In this chapter, we relax some of the assumptions of the longitudinal autoencoder model described in Chapter 5. Temporal variability is assessed by ordering the visits in a self-supervised fashion, instead of using an affine reparameterization of time. This work has been accepted for publication in the proceedings of the **2021 International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)**.*

Abstract

The problem of building disease progression models with longitudinal data has long been addressed with parametric mixed-effect models. They provide interpretable models at the cost of modeling assumptions on the progression profiles and their variability across subjects. Their deep learning counterparts, on the other hand, strive on flexible data-driven modeling, and additional interpretability - or, as far as generative models are involved, *disentanglement* of latent variables with respect to generative factors - comes from additional constraints. In this work, we propose a deep longitudinal model designed to disentangle inter-patient variability from an estimated disease progression timeline. We do not seek for an explicit mapping between age and disease stage, but to learn the latter solely from the ordering between visits using a differentiable ranking loss. Furthermore, we encourage inter-patient variability to be encoded in a separate latent space, where for each patient a single representation is learned from its set of visits, with a constraint of invariance under permutation of the visits. The modularity of the network architecture allows us to apply our model on various data types: a synthetic image dataset with known generative factors, cognitive assessments and neuroimaging data. We show that, combined with our patient encoder, the ranking loss for visits helps to exceed models with supervision, in particular in terms of disease staging.

7.1 Introduction

Understanding the progression of diseases is essential for accurate early diagnosis, prognosis, and patient monitoring. Often, there is a strong interplay between the pathological progression and the inter-subject variability, which makes it all the more necessary to characterize the contribution of each factor. Typically, in the context of neurodegenerative diseases, we may ask whether the atrophy of a particular brain region is predictive of a specific patient advancement in the disease, or rather can be dismissed as a specific characteristic of the individual.

Longitudinal data analysis has been usually addressed in the framework of parametric mixed-effect models. For instance, geometric approaches have been proposed either for the progression of biomarkers (Schiratti, Allasonniere, et al. 2015c) or shape changes (Bône, Maxime Louis, Martin, et al. 2018). This family of models assumes that each subject follows a curve on a Riemannian manifold which translates from a common geodesic. They also assume that the direction of translation is orthogonal to the direction of the progression curve, which ensures that the changes due to the progression of the disease are disentangled from the effects of different physiological or anatomical characteristics of the patient. The family of progression profiles is constrained, e.g. sigmoid curves for biomarkers changes, and an affine function maps the age of the subject to a disease stage.

Generative models such as variational auto-encoders (VAE) (Kingma and Welling 2013) have been consistently used in deep learning as they offer a flexible learning framework, in which disentanglement may be enforced through soft constraints and optimization

schemes, as in β -VAE and their extensions (Higgins, Matthey, et al. 2016; Kim and Mnih 2019; Mathieu et al. 2019). With time series, however, separating static and dynamic representations without inductive bias still remains a challenge. In most research works, authors disentangle time-varying from time-invariant information by leveraging time labels explicitly: in literature focused on style and content of videos (Grathwohl and Wilson 2016; Y. Li and Mandt 2018), in face ageing progression (Hsu, Y. Zhang, and Glass 2017) or medical data, where age is used for supervision (Ravi, Daniel C. Alexander, and Neil P. Oxtoby 2019; Xia, Chartsias, and Tsiftaris 2019). These previous methods are not directly transferable to longitudinal data where duration between visits differs. In (Berchuck, Mukherjee, and Medeiros 2019), authors seek an age direction in the latent *a posteriori*, while in (Zhifei Zhang, Song, and Qi 2017) they estimate the latent age regression jointly with the reconstruction task in a supervised fashion. A Riemannian manifold learning point of view, in the spirit of parametric models is proposed in (Maxime Louis, Couvonné, et al. 2019) as it estimates both a static representation and an affine time reparametrization per patient. All these methods assume that age at observation is a direct marker for the progression timeline, which is not the case for most neurodegenerative disorders. In the recent work closest to ours (Zhao, Z. Liu, et al. 2020), the authors propose to learn the disease stage without relying on the patient age, in a self-supervised fashion. They use a cosine loss to enforce progression in a specific direction of the latent space, learned during optimization. They do not study the disentanglement of their model but rather focus on the correlation with a disease progression timeline.

In this paper, we propose a generic deep longitudinal model, designed to disentangle inter-patient variability from an estimated disease progression timeline. We learn a disease stage as a flexible function that does not rely on age, but solely on the individual order between visits using a differentiable ranking loss, leveraging a much weaker prior. The remaining latent space is further favored to produce representations independent of the progression thanks to a DeepSet network which acts as a permutation invariance function on visits. The main contributions of this paper are therefore (i) an architecture that is tailored to disease progression modeling and disentangles the changes due to progression from the changes due to phenotypic differences across subjects; (ii) a modular method with decoders adapted to data types; (iii) an application on synthetic and real datasets - including imaging and clinical data - showing that one direction of the latent space alone describes temporal progression.

7.2 Methodology

The proposed generic deep longitudinal model is summarized in Fig 7.1.

7.3 Longitudinal progression model

In this section, we propose a temporal latent variable model that encodes the disease progression in a low-dimensional probabilistic space. It assumes that a sequence of observations is generated as the combination of an intrinsic code z^s (as in *space shift*) and

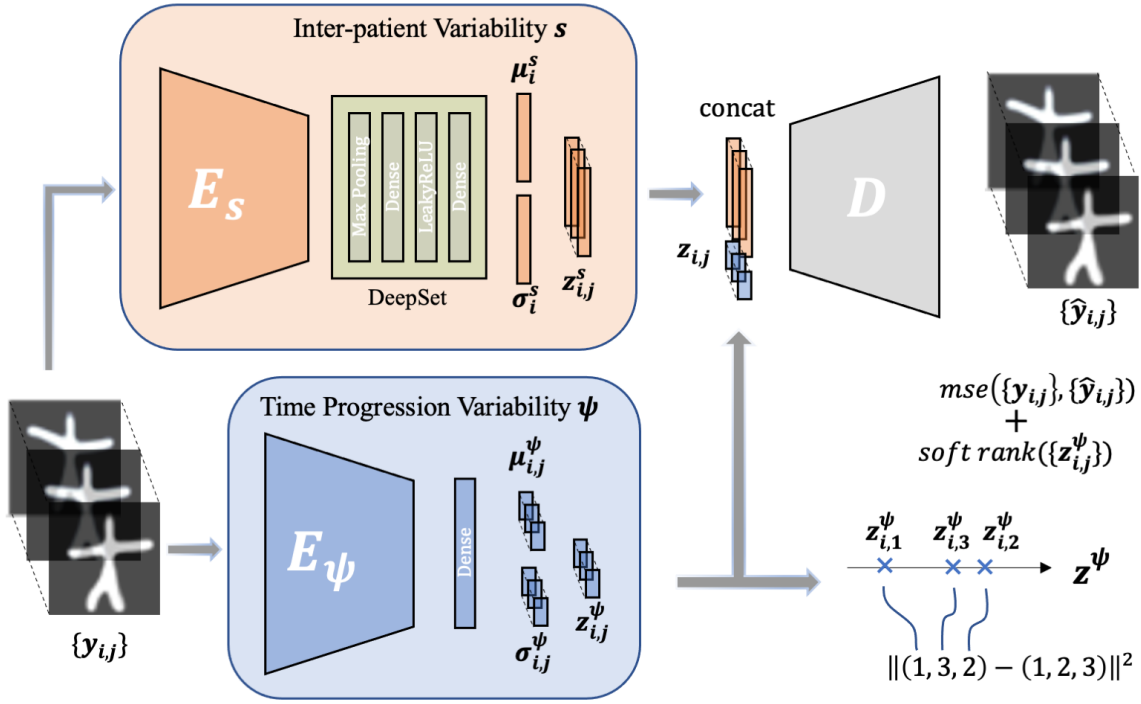


Figure 7.1 – Input data y_{ij} is encoded simultaneously in a space encoder (Deepset) and a point-wise time encoder to get latent representations (z_i^s, z_j^ψ) . z_i^s can be computed from *any* subset of visits, and in practice randomized fixed-size subsets of visits are drawn in the spirit of stochastic optimization. Decoder can be either agnostic, or specific (e.g., velocity fields for deformations).

a disease progression factor z^ψ (where we use ψ instead of t to clearly distinguish the stage from the temporality of visits).

Generative disease progression model Let $\{(t_{i,j}, y_{i,j})\}_{1 \leq i \leq N}$ be a set of N subjects, each observed at the age of $t_{i,j}$ for $1 \leq j \leq n_i$ visits. We assume that the observations $\{y_{i,j}\}$ are generated from a Bayesian generative model as follows:

$$y_{i,j} \stackrel{\text{iid}}{\sim} \mathcal{N}\left\{\Phi(z_i^s, z_j^\psi); \epsilon^2 \mathbb{I}\right\} \text{ with } z_i^s \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \lambda_s^2 \mathbb{I}) \text{ and } z_j^\psi \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \lambda_\psi^2 \mathbb{I}) \quad (7.1)$$

Here Φ denotes an unknown non-linear transform from the strongly decoupled generative factor space, also called latent space, $\mathcal{Z} = \mathcal{Z}^s \times \mathcal{Z}^\psi$, towards our observation space of scores or images $\mathcal{X}$. The generative factor z^ψ is assumed to be independent from the individual variability z^s factor. Notice that we do not assume any relationship between the ages t_{ij} and the associated observations y_{ij} . Parameters ϵ^2 , λ_ψ^2 and λ_s^2 are the diagonal Gaussian variance priors.

Variational Inference with VAE The inference is conducted within the VAE paradigm. The function Φ is approximated by a parametric class of neural network Φ_θ , the decoder.

Two neural network encoders are used to approximate the intractable posterior distribution $p(z_i^s, z_j^\psi | y_{ij}; \theta)$. They respectively model the latent distributions of space shifts and disease progression, such that the approximated parametric distribution can be factorized as a product of independent Gaussians $\mathcal{N}(\mu^s, \Sigma^s) \otimes \mathcal{N}(\mu^\psi, \Sigma^\psi)$.

Set-invariant representation for $\mathcal{Z}^s$ The strong condition on $\mathcal{Z}^s$ is that it should extract from any time-series a time-invariant representation. To do so, we choose to learn the posterior $q_{\eta^s} \equiv \mathcal{N}(\mu^s, \Sigma^s)$ as a DeepSet encoder network (Zaheer et al. 2017) acting on any unordered subset of visits. This rewriting of our disentanglement hypothesis generalizes the use of simple operators such as averaging or maxing out of visits in an intermediate latent representation, or even more elaborate inverse Gaussian product of group-wise non-*iid* distributions (Bouchacourt, Tomioka, and Nowozin 2017); all these can indeed be cast as specific choices of permutation-invariance operators, which were shown in (Zaheer et al. 2017) to be universally approximated by DeepSets.

Ranking visits in $\mathcal{Z}^\psi$ as a regularization constraint The remaining generative factor of our one-dimensional $\mathcal{Z}^\psi$ space must encode the dynamic of the progression preferentially. A disease progression constraint, $\mathcal{C}^{\text{ranking}}$, aims at favoring a natural ordering of visits along the temporal latent dimension in a self-supervised way. Unlike in (Zhao, Z. Liu, et al. 2020), it builds upon the soft-ranking differentiable loss of (Blondel et al. 2020) to enforce isotonic individual progression. The individual ranking errors of visits are penalized according to $\sum_j \|r(z_j^\psi) - j\|_2^2$. From a theoretical perspective, the minimization of this loss is similar to maximizing the Spearman correlation between different visits of a given subject i : it is therefore to be understood as a soft supervision which only relies on the ordering of visits and not on the times of observations, which are never seen by the model.

Final objective The final loss can be written as the sum of two terms: the evidence lower bound, written as the sum of the KL-divergence $\text{KL}[q_\eta(z^\psi, z^s | y_i) || p(z_i)]$ and the data attachment term $-\mathbb{E}[\log p_\theta(y_{ij} | z_{ij})]$ which is proportional to the ℓ^2 reconstruction loss, and the self-supervised ranking of visits $\mathcal{C}^{\text{ranking}}$ discussed previously with a weight γ :

$$\sum_{i=1}^{n_b} \text{KL}[q_\eta(z^\psi, z^s | y_i) || p(z_{ij})] - \mathbb{E}[\log p_\theta(y_{ij} | z_{ij})] + \gamma \mathcal{C}_i^{\text{ranking}}$$

7.4 Modularity

The model we propose in this paper can be seen as a “*meta*” architecture that can be instantiated according to the datatype: clinical data (1D), images (2D and 3D). We may even use decoders that are specifically designed for data, such as diffeomorphometry for brain grey matter. We detail the latter case, directly in the spirit of classical models with stationary velocity fields (Dalca et al. 2019; Krebs, Delingette, Ayache, et al. 2020). An additional parameter, a template $\mathcal{T}$, is learned at the centered reference disease stage

$z^\psi = 0$. From this, any observation y_{ij} can be reconstructed from the latent code (z^ψ, z^s) by a deformation field Φ_v , parametrized with a velocity decoder v , acting on $\mathcal{T}$.

7.5 Experimental results

The network E_ψ is a classical CNN encoder with LeakyReLU non-linearities and a final dense layer toward $\mathcal{Z}^\psi$. The spatial encoder E_s is composed as a DeepSet, whose output does not depend upon the ordering of its inputs: it can be written as $\rho \circ \max_{j \sim \text{visits}}(f)$. We chose f to be a convolutional encoder network which outputs an intermediate representation (per observation), from which the max operator (over visits) retrieves a permutation invariant code. The latter is eventually mapped, *via* a MLP ρ , into the space shifts domain $\mathcal{Z}^s$. Inference was performed using the PyTorch library, with Adam optimizer.

7.6 Validation on synthetic data

To validate the disentangling ability of our model, we first generated a synthetic longitudinal dataset of starmen images, based on the longitudinal diffeomorphic model of (Bône, Colliot, and Durrleman 2018). From a given reference template y_0 , the cross-sectional variability of our population is prescribed by a diffeomorphism localized at four control points: the head, right arm and legs. The common progression timeline, on the other hand, is generated through a displacement of the left arm only.

The dynamics of progression is given by an affine reparametrization of the age t_{ij} at visit j , characterized by individual onset τ_i and acceleration α_i factors, such that the true disease progression is given by $\psi_{ij}^* = t_0 + \alpha_i(t_{ij} - \tau_i - t_0)$. We sample variables in a similar fashion as in (Bône, Colliot, and Durrleman 2018) to obtain a dataset of $N = 1000$ subjects, each with $n = 10$ visits (Fig 7.2).

We benchmark contender approaches mentioned in the introduction on the Starmen dataset (see Table 7.1): β -VAE (Higgins, Matthey, et al. 2016), ML-VAE (Bouchacourt, Tomioka, and Nowozin 2017), LSSL (Zhao, Z. Liu, et al. 2020) and both supervised “Longitudinal Riemannian VAE” (LR-AE) (Maxime Louis, Charlier, and Durrleman 2018) and “Age Regression VAE” (AR-VAE) (Gao et al. 2018). We compare with our model in its generic iconic form on pixels and its diffeomorphic version (wD). The version without ranking loss (woR) is added for ablation purposes. Beyond the reconstruction quality (Mean Square Error), which reveals only LSSL and our model are below the baseline of β -VAE, we are interested in disentanglement capacity. It can be

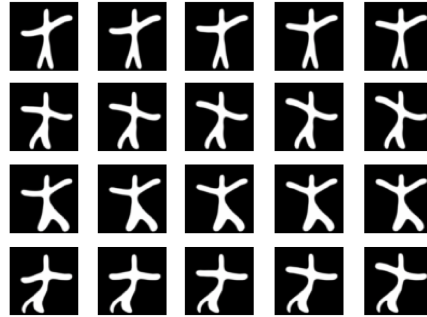


Figure 7.2 – Each row represents a synthetic subject across time.

Metric	β -VAE	ML-VAE	LR-AE	AR-VAE	LSSL	Ours	Ours (wD)	Ours (woR)
MSE (10^{-3})	7.90 ± 0.57	22.7 ± 1.51	10.9 ± 1.53	8.26 ± 0.62	7.32 ± 0.379	8.83 ± 0.88	6.22 ± 1.23	14.2 ± 5.46
PLS z^ψ/z^s	- -	0.660 ± 0.343	0.137 ± 0.209	0.125 ± 0.117	0.098 ± 0.047	0.083 ± 0.026	0.083 ± 0.025	0.149 ± 0.131
Staging ψ^*	0.263 ± 0.348	0.030 ± 0.028	0.971 ± 0.024	0.984 ± 0.008	0.994 ± 0.003	0.997 ± 0.001	0.996 ± 0.002	0.524 ± 0.464

Table 7.1 – Benchmark of proposed methods on Starmen dataset

measured by correlations between the estimated staging z^ψ and the latent space code z^s with a partial least square regression analysis (PLS), so as to ensure the independence of $\mathcal{Z}^s \times \mathcal{Z}^\psi$ (2nd row). Among all methods, ours performs best: even though LSSL uses a similar loss, it does not constrain its orthogonal directions to be independent. Other methods, especially supervised, naturally learn correlated representations.

Finally, the proper staging of ψ^* is evaluated by computing the Spearman ranking correlation between ψ^* and z^ψ : it evaluates the monotonicity of individual trajectories. Only methods with time supervision or ranking strategy (*ie* all but first two columns) manage to grasp a staging close to one (row 3).

In complement to the previous metrics, we visualized in Fig 7.3 the effect of specific directions in the latent space *via* gradient maps. The β -VAE is a low baseline as it does not model the progression and only views the data as static representations: we plot the PCA in the whole space, and observe that no principal direction correlates with the left-arm progression. It is interesting to note that the benefit of the ranking loss (when no supervision is available) is made clear by the study of our model without it (woR), and ML-VAE. They both focus on group-structure only, and fail to grasp progression. LR-AE

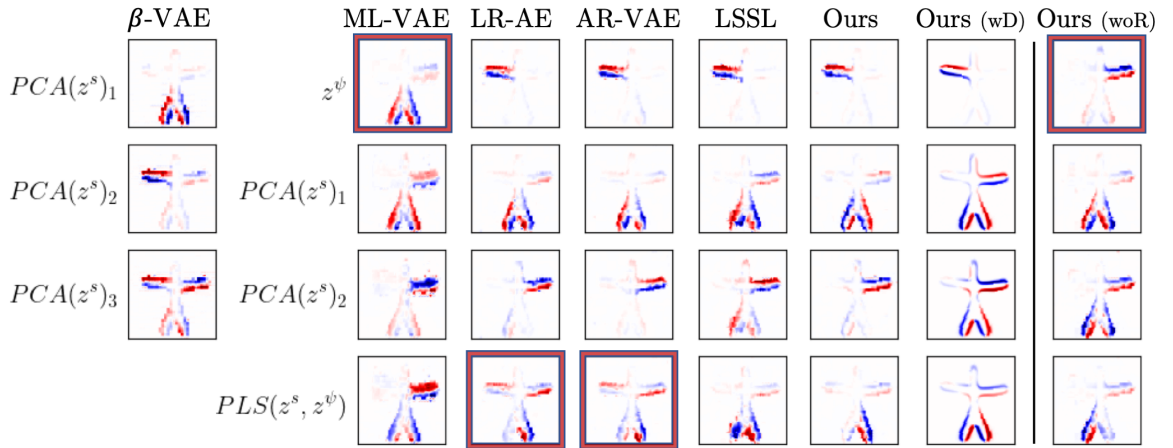


Figure 7.3 – Gradient directions in the latent space (extracted from a forward pass in the decoder). Row 1: gradient wrt to the latent space associated with disease progression. Rows 2 and 3: first two principal directions of the PCA in the orthogonal of the latent time ($\mathcal{Z}^s$ for us). 4th row: the direction in the orthogonal of the latent time that correlates most with it (PLS), as a way to challenge the model disentanglement.

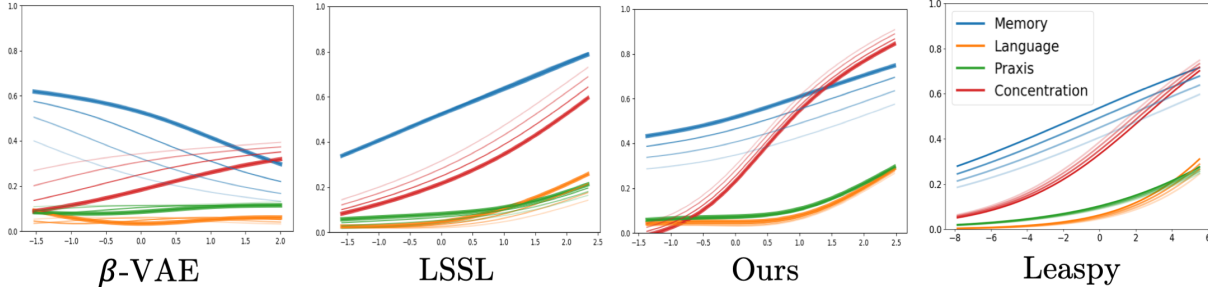


Figure 7.4 – Estimated average trajectory of scores. The effects of latent dimensions z^s (resp. z with β -VAE) are shown with degraded colors.

and AR-AE, because they use supervision of time, are displaying time-related correlations in a space shift direction (last row).

7.7 Application to Alzheimer’s disease

Cognitive scores We apply our model on four subtest scores of the ADAS-Cog scale obtained from the ADNI dataset, namely concentration, praxis, memory and language; normalized between 0 and 1, with higher values indicating lower performance. 248 patients with mild cognitive impairment (MCI) converting to Alzheimer’s disease (AD) during the study are followed for an average of 6 visits over 3.5 years.

MSE (10^{-3}) on 5-fold cross validation yields 7.47 ± 0.778 for our model: slightly less than β -VAE (3.78 ± 0.562) and LSSL (3.64 ± 0.429), but on par with a parametric model of reference, Leaspy (Schiratti, Allasonniere, et al. 2015c) (8.21 ± 0.155). Additionally, we can predict future visits from previous ones: in this scenario, our model reaches a lower MSE (10^{-3}) 29.1 ± 5.53 than LSSL 32.4 ± 5.93 .

Figure 7.4 illustrates the estimated average time progression for each model, as well as the effects of orthogonal directions in the latent space. An agnostic β -VAE fails at extracting a consistent dimension for the time progression while still providing a good reconstruction of data. Our proposed model learns monotonicity in an unsupervised fashion, and is able to estimate a consistent time progression of the ADAS-Cog scores from longitudinal measurement in small number. Interestingly, despite the noisiness of cognitive scores, our model generates a progression very similar to that of Leaspy’ sigmoid geodesics: in particular, the first PCA directions have the same effects on scores (cf degraded color effects).

Neuroimaging data We selected 356 subjects MCI converters from ADNI, with a total of 1898 visits. The 1898 T1-weighted MRI were preprocessed using Clinica (Routier et al. 2019) (non-rigid alignment, skull stripping, intensity rescaling), and converted into mid axial slices of dimension 128×128 .

We do not have access to the true disease progression, however the disease severity at the MCI stage can be monitored through cognitive scores (ADAS-Cog) or markers of the morphological evolution such as atrophy of the hippocampi and increase of ventricle

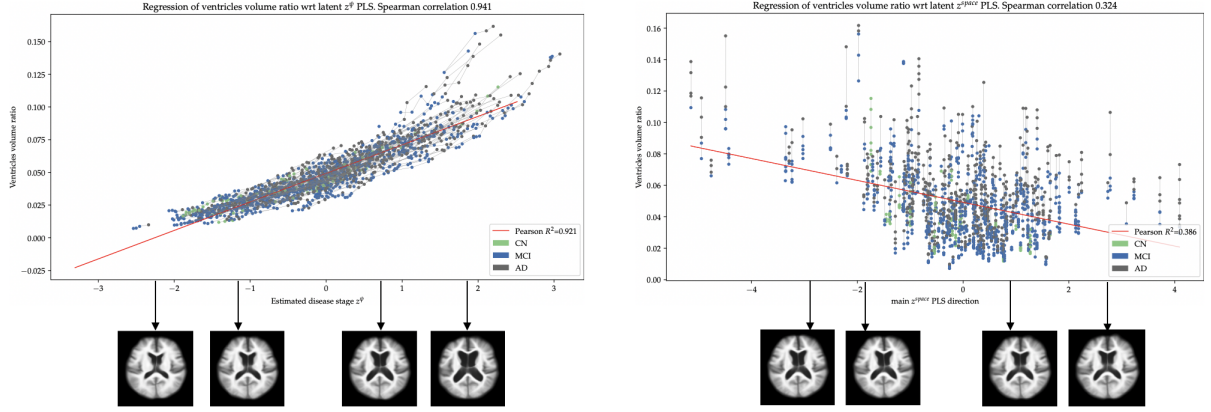


Figure 7.5 – PLS analysis with respect to ventricle volume ratio $\mathcal{V}$: z^ψ (left), z^s (right).

volumes. We computed the ratio of ventricle volumes by brain volumes, as a covariate factor, noted $\mathcal{V}$, which we assimilate with a good proxy of the disease progression.

Figure 7.5 shows interactions between our two latent spaces and the disease stage proxy $\mathcal{V}$ via a correlation (PLS) analysis. First, z^ψ exhibits a quasi-linear regression fit associated with a high Spearman ranking (0.934 ± 0.025), suggesting that z^ψ has indeed captured the main disease progression trend $\mathcal{V}$ and the individual ordering of visits. On the other hand, the best correlated direction (measured as the PLS main direction) between z^s and $\mathcal{V}$ is not localized on the ventricle. This result further implies that $\mathcal{Z}^s$ has indeed captured the variability necessary to perform a good fit of the data without correlating significantly with the disease stage marker.

Furthermore, z_ψ behaves as a clinical score informative of the onset: its distribution is significantly earlier for women ($p < 3.83e^{-2} \pm 7.05e^{-2}$ for Mann-Whitney U test), as observed from clinical data. Its derivative $\frac{\partial z_\psi}{\partial t}$ correlates with the pace of the disease: it is significantly ($p < 4.82e^{-2} \pm 1.06e^{-2}$) faster for APOE4 carriers (1 or 2 alleles), a result in accordance with well-documented disease progression patterns in AD (Bigio et al. 2002).

7.8 Conclusion

In this paper, we proposed a generative variational autoencoder architecture that leverages the repetition of measurements per individual to disentangle between the global disease timeline and inter-patient variability. The disease stage estimation is driven by a differentiable ranking loss, while a permutation invariant function reduces the remaining information in a representative space. As we further demonstrate, inductive biases on the data itself (such as using diffeomorphometry for structural medical imaging) are completely synergetic and improve the quality of the representations learned.

A very interesting avenue would be to further explore the space shift space $\mathcal{Z}^s$, in particular because it has the potential to offer a good representation of neurodegenerative diseases sub-types.

Part IV

Software development

Leaspy

Outline of the current chapter

8.1 Introduction	144
8.2 Inputs	144
8.3 Classes and methods	145
8.3.1 Algorithm Settings	145
8.3.2 Model Settings	145
8.3.3 Data	145
8.3.4 Leaspy (API)	145
8.4 Performance Benchmark	146
8.5 Applications	146
8.5.1 A Basic Example	147
8.5.2 Interactive visualization	147
8.5.3 Privacy Preserving synthetic longitudinal data	148

We need a sound code source to implement the longitudinal spatiotemporal model, and perform clinical applications on Parkinson’s Disease. To that end, we coded the Python library Leaspy. Leaspy is released under the GNU GPLv3 licence. Since then, Leaspy has benefited from inputs of various contributors.

Neurodegeneration cohort studies, such as ADNI or PPMI often come in the form of longitudinal data. However statistical analysis of such structured data is not straightforward. The python library Leaspy (<https://gitlab.com/icm-institute/aramislab/leaspy/>) implements the longitudinal spatiotemporal model (Schiratti, Allasonniere, et al. 2015b), a generative mixed-effect model for disease progression. In practice Leaspy proposes, under some assumptions on the longitudinal dataset, and on the curve shapes, to estimate the progression of scalar biomarkers at both the population and individual levels. A trained model may then be used for interpolation or extrapolation tasks, at both population and individual levels. Additionally, individual parameters inform on interpretable geometrical variation from the average, and may be used as cofactors for subsequent statistical analysis.

8.1 Introduction

Leaspy is a Pytorch based software package for the statistical analysis of monotonic, longitudinal data, designed specifically for biomarker progression modeling in neurodegenerative diseases. Typical use-cases include the joint progression of cognition scores such as MMSE or ADAS in Alzheimer's Disease, or the joint progression of clinical scores such as the MDS-UPDRS part III in Parkinson's Disease.

8.2 Inputs

Let us have a longitudinal dataset $\mathcal{Y} = \{y_{i,j,k}, i \in [1, N], j \in [1, k_i], k \in [1, K]\}$ with $y_{i,j,k}$ scalar measurement of the k th biomarker at the j th visit of patient i , measured at patient age $t_{i,j}$. Leaspy handles data in the form of pandas dataframe with a first "ID" column, a "TIME" column, and K columns for the K studied biomarkers.

8.3 Classes and methods

We describe in this part the minimal set of leaspy classes and methods that are needed to use the library.

8.3.1 Algorithm Settings

- **Instantiate:** `AlgorithmsSettings(n_iter, ...) ↦ algorithmsettings`
Instantiate an `algorithmsettings` object, which includes parameters needed for optimization.

8.3.2 Model Settings

- **Instantiate:** `ModelSettings(name, dim, source_dim, ...) ↦ modelsettings`
Instantiates a `modelsettings` object, which includes parameters that describe the model (e.g. logistic of dimension 4).

8.3.3 Data

Data checks and transform a pandas DataFrame to a `data` object, built to be fed to a `leaspy` for estimation tasks.

- **Instantiate:** `Data.from_dataframe(df) ↦ data`
Processes a pandas DataFrame `df` to build a `data` object.

8.3.4 Leaspy (API)

A `leaspy` object include model parameters θ , and is chosen as the main api of the library to compute estimation tasks.

- **Instantiate:** `Leaspy(modelsettings) ↦ leaspy`
Instantiates a `leaspy` object from a `modelsettings` object.
- **Calibration:** `leaspy.calibrate(algorithmsettings, data)`
Estimates (MCMC-SAEM) model parameters from calibration dataset `data` and update `leaspy` with new estimated parameters.
- **Personalization:** `leaspy.personalize(algorithmsettings, data) ↦ z`
Estimates the individual parameters $\mathbf{z}$ of a personalization dataset `data` knowing model parameters of object `leaspy`.
- **Simulate:** `leaspy.simulate(algorithmsettings, z) ↦ z_syn`
Simulates synthetic individual parameters `z_syn` from real-data individual parameters `z`, knowing model parameters.
- **Save:** `leaspy.save(path)`
Saves the current `leaspy` object in a `.json` file.

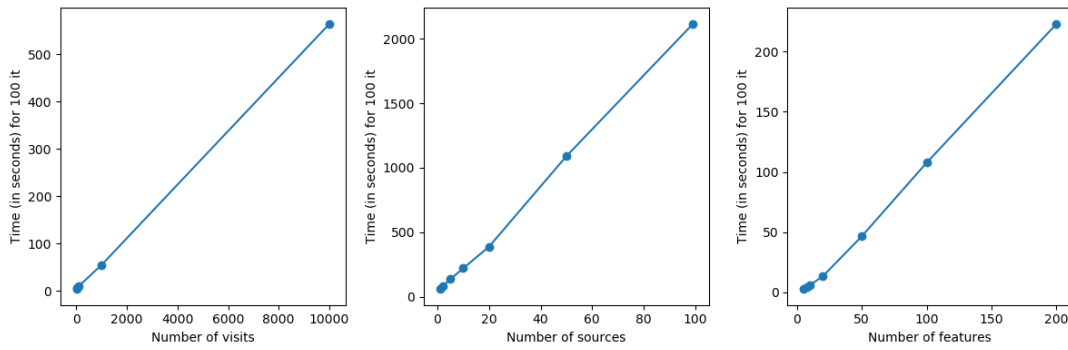


Figure 8.1 – Computation cost benchmark relative to respectively: number of visits in the dataset, number of sources and number of biomarkers. Leaspy is linear in the 3 variables. On the left, we set $K = 10$, $Ns = 4$ and vary the number of visits. On the middle, we set $N = 50$ patients, each with $m = 10$ visits, $K = 100$ and change the number of sources Ns from 1 to 99. On the right, we set $N = 50$ patients, each with $m = 10$ visits, we set $Ns = 4$ and change the number of biomarkers from $K = 5$ to $K = 100$.

- **Load:** `leaspy.load(path)`
Instantiates a `leaspy` object from a `.json` file.

8.4 Performance Benchmark

We show in Figure 8.1 a benchmark of Leaspy in terms of time to compute 100 iterations. Leaspy complexity is linear in the number of visits in the dataset, in the number of biomarkers included, and in the number of sources. Note that as we include more biomarkers, we may want to use more sources as well to preserve flexibility of the model, which may lead to a higher complexity depending on the design choice regarding the number of sources. To give an order of magnitude, 100 iterations of Leaspy on a dataset of 200 patients with each 10 visits, on 10 biomarkers, and choosing 4 sources would last 300 seconds. Typically we choose 10000 iterations, leading to 3.2 hours of computations.

8.5 Applications

8.5.1 A Basic Example

```
1 import pandas as pd
2 from leaspy import Leaspy, Data, AlgorithmSettings
3
4 # Load DataFrame and transform to Data
5 df = pd.read_csv(my_data_path)
6 data = Data.from_dataframe(df)
7
8 # Instantiate a Leaspy object
9 leaspy = Leaspy("logistic") # set model type
10 leaspy.model.load_hyperparameters({'source_dimension': N_S})
11
12 # Define calibration parameters
13 calibration_settings = AlgorithmSettings('mcmc_saem',
14                                         n_iter=1000,
15                                         progress_bar=True)
16
17 # Launch calibration
18 leaspy.calibrate(data, algorithm_settings=calibration_settings)
19
20 # Define personalization parameters
21 personalization_settings = AlgorithmSettings('scipy_minimize')
22
23 # Launch personalization
24 z = leaspy.personalize(data, personalization_settings)
25
26 # Define simulation parameters
27 simulation_settings = AlgorithmSettings('simulation',
28                                         number_of_subjects=100)
29
30 # Launch simulation
31 z_synthetic = leaspy.simulate(z, simulation_settings)
32
33 # Compute model at timepoints for
34 # sets of individual parameters in z
35 values = leaspy.estimate(timepoints, z)
```

8.5.2 Interactive visualization

A leaspy model trained on Parkinson's Disease Data can be visualized at:
<https://dashboard.heroku.com/apps/dashboard-leaspy-park>.

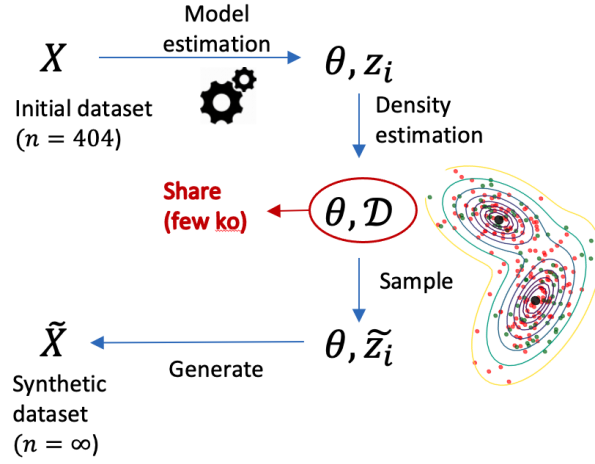


Figure 8.2

8.5.3 Privacy Preserving synthetic longitudinal data

Provided a good fit of our model to the data, we can leverage the generative model to generate privacy preserving data, which can come handy to promote reproducible research. This application was presented at the conference ADPD 2021.

Goal

We aim at providing a longitudinal synthetic data generator for neurodegeneration biomarkers. This generator can be first fitted on any set of progressing biomarker in a cohort, to then be able to generate synthetic data that mimic the real data, while preserving privacy. Although not directly usable for clinical research, this lightweight, easily sharable generator fosters reproducibility by providing plausible data on demand for educational, visualization or benchmarking purposes.

Methods

We included 404 idiopathic Parkinson's disease subjects from the PPMI dataset with six clinical scores, the MDS-UPDRS part I, II and III, MoCA, SCOPA-AUT and RBDSQ, for an average of 13.8 ± 2.7 visits over 6.4 ± 1.7 years. We then used a generative mixed-effect model (Schiratti, Allasonniere, et al. 2015b) to fit the longitudinal progression of these clinical scores. We ensured privacy by using a bayesian gaussian mixture model to fit the empirical individual parameters distribution, so that no individual parameters from real data can be retrieved from the synthesis model.

Results

We fitted a model on PPMI and generated 404 synthetic patients. To assess the quality of synthesis, we trained two classifiers with default parameters to distinguish synthetic visits from real ones, leading to low test accuracy : 60% for Random Forest and 52% for

Logistic Regression. We then showed that biomarker's correlation structure is consistent in synthetic Data.

Conclusions

We proposed a longitudinal synthetic data generator for reproducible research. Synthetic data are difficult to distinguish from real data, and preserve the correlation structure.

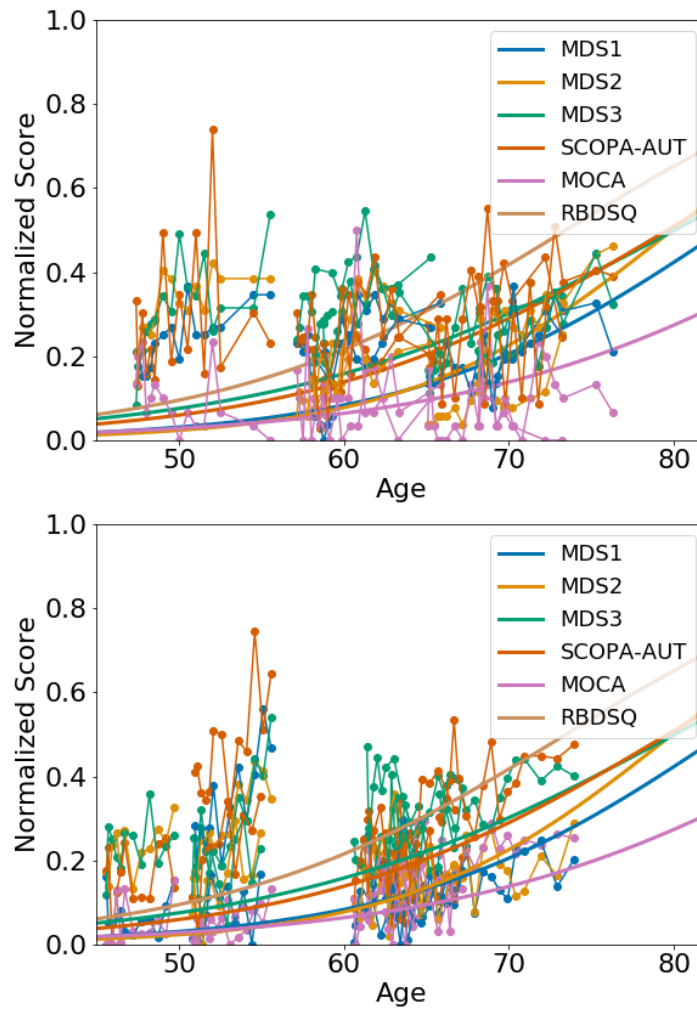


Figure 8.3 – Example of real data on top, and synthetic data on the bottom. A model is calibrated and personalized on the real data, which in this case include 5 clinical scores. Then from the learned distribution of individual parameters, we sample synthetic individual parameters accounting for synthetic patients. Coupled with a deterministic or sampling scheme for patient ages, we can sample synthetic patient longitudinal data.



Figure 8.4 – Poitwise correlations of real-data biomarkers on the top left, and synthetic biomarkers on the bottom right. We observe that correlation structure of the biomarker is preserved in synthetic data.

Conclusion and Perspectives

Preamble to conclusion

This interdisciplinary work requires to understand notions in pathological and clinical Parkinson’s Disease progression, and what longitudinal analysis can bring to the clinic. This also implies theoretical and empirical work to develop new longitudinal models, with inherent contributions to the field of disease progression modeling. Bridging the gap between these two aspects has proven to be both extremely challenging and rewarding, and has been the ultimate goal of this thesis. This work presents our solution to this multi-objective optimization process.

Contributions

A longitudinal spatiotemporal model

We first present the spatiotemporal model, a bayesian longitudinal mixed-effect model, designed to model neurodegenerative disease progression. The model is able to estimate disease trajectories at both population level, and individual levels, disentangling disease progression from inter-patient variability.

We propose to relax assumptions of the propagation model which instantiates the spatiotemporal model in the case of scalar data (Schiratti, Allasonniere, et al. 2015b). We relax the assumption that all markers progress in the same fashion modulo a time delay between them, and rather propose to learn for each marker their own curve, which are still estimated jointly. We also allow missing values in the data and propose to leverage the generative nature of the model to handle them. From a more practical point of view, the aforementioned model can be calibrated on a set of biomarkers of a given cohorts, to estimate the fixed effects. It can then be used for personalization of individual parameters from the same cohorts, or independant cohorts for validation. Once calibrated, we can also simulate synthetic patients from the learned a-posteriori distribution.

Building Parkinson’s Disease Couse Map

In this work, we ultimately aim at building more flexible longitudinal models to better account for Parkinson’s Disease heterogeneity across its progression. Despite considerable efforts to investigate possible data-driven subtyping of the disease, this task still remains far from straightforward in the community even leveraging recent longitudinal models (see 1.4.2). This is also our experience on the task. It might be that PD is so heterogeneous that clear, separate subtypes cannot be extracted. Or that biomarker are not nearly precise enough to distinguish the subtypes. In the latter case, this task could hopefully be tackled with finer biomarkers such as those of (Horsager, Andersen, Knudsen, Skjærbæk, et al. 2020b). In any case, each iteration of longitudinal cohorts and consecutive statistical analysis will improve the state of the art on PD biomarkers (He et al. 2018). In our clinical contributions, we still have access to a broad range of biomarkers, although not the latest state of the art. First attempting to extract subtypes, we shift to estimating and explaining Parkinson’s Disease heterogeneity, that we believe to be a pre-requisite for

subsequent subtyping. We find significant effects of age at baseline, and sleep symptoms on the patterns of progression. In addition, we find results that could corroborate those of (Horsager, Andersen, Knudsen, Skjærbæk, et al. 2020b), with the highest variance between individuals found on an axis: neurodopaminergic loss versus non-motor scores. However, we argue (see 1.3.3) that this could be a consequence of a left censorship bias on the data. Indeed, monitoring usually begins with Parkinson’s Disease diagnosis, and therefore longitudinal data are re-aligned with regard to motor symptoms at baseline. This might suffice to explain that variability from the average trajectory is mainly found in the other modalities.

Deep-Learning Extensions

In the spatiotemporal model, we mathematically understand each parameter’s effect on the model. However, applying the model to high dimensional appears prohibitive, both in terms of computation time and modeling power. On the other hand, autoencoders trade interpretability of the spatial individual parameters for flexibility. We leverage these models to handle effortlessly high dimensional and multimodal data in Part III. Trajectories in the data space are mapped on a latent space, assumed to be Euclidean.

In our last work, we attempt to constrain the network to preserve interpretability to some extent. In Deep Learning, interpretability is often tackled via disentanglement of the latent space. However, despite exhaustive research (R. T. Chen et al. 2018; Higgins, Amos, et al. 2018; Higgins, Matthey, et al. 2016; Mathieu et al. 2019) unsupervised disentanglement remains currently unreachable without leveraging inductive biases in the data (Locatello et al. 2019). Thus, we leverage the longitudinal setup and monotonicity assumptions. The two combined allow for self-supervision of the network via a ranking loss. Disentanglement is additionally fostered by enforcing a constant patient representation along its visits.

Limitations and Perspectives

Estimation

Recent work study the theoretical convergence of the MCMC-SAEM under milder hypothesis. (Allasonnière and Chevallier 2019) proves that convergence still holds when using a tempering scheme in the optimization, which is our case. Additionally, (Debavelaere and Allasonnière 2021) shows that model exponentialization does not bias the MAP. Recall from that we did perform exponentialization so that the model belong to the curved exponential family, which is needed for convergence. Finally, convergence of MCMC-SAEM had already been proved under an assumption of geometric ergodicity of the Markov dynamic. Recently, (Debavelaere, Durrleman, and Allasonnière 2020) relaxes this hypothesis by only assuming a subgeometric ergodicity of the Markov dynamic.

However, we also rely on heuristics at initialization to boost MCMC-SAEM convergence speed. This might bias the MCMC-SAEM toward particular local minimas. What is more, no diagnostic of convergence of the Markov Chain is currently used in Leaspy, and

convergence is generally assessed subjectively. Convergence diagnostic would also allow for less computations.

Then, variable sampling is performed via Gibbs sampling. Leveraging the information of the gradient of the likelihood, computed by autodifferentiation, we could use Hamiltonian Monte Carlo instead of Gibbs Sampling, to aim for faster convergence.

Modeling Hypothesis

In the model building, we assume that we can disentangle inter-patient variability against intra-patient variability : i.e. spatial against temporal variability. This view of patients following parallel curves over their progression is very practical, as it offers representations to compute distances between patients (for e.g. clustering). This assumption seems acceptable in low dimension, with, in the scalar case, spatial variability modeled as shifting the sequence of events around the average. However, this assumption is likely to be challenged in high dimension. In medical images such as brain MRIs for example, patient particularity, such as its brain region shapes, and disease progression, which could take the form of damage in particular regions are very likely to be intertwined to some extent. Full disentanglement between temporal and spatial variability may then appear as a too strong constraint, which could be relaxed. Intertwining would be ideally reduced to its minimal influence to still get good reconstruction.

Then we assume monotonous progression. This assumption might hold in the field of neurodegenerative diseases in the current state of our knowledge. However, with potential disease-modifying drugs, or applications to other diseases, where periods of progression and remission may occur, the monotonicity hypothesis would not hold.

Finally we assume data lie on a smooth manifold. In scalar instantiations of the spatiotemporal model, we assume parametric families of these manifolds, as product manifold of simple open intervals of $\mathbb{R}$. This is practical to get closed form of the average trajectory geodesic, and its exp-parallelizations, while still performing well on noisy clinical data in terms of reconstruction error. However, with this product manifold and its associated diagonal metric, we not only assume a-priori shapes of biomarker progression, but also completely miss the inter-biomarker dynamics. It is reasonable to assume however that some of them actually exist, probably even in the form of causal links, such as imaging abnormalities inducing subsequent cognition symptoms. Learning directly the Riemannian Metric from the data could allow for much more expressivity (Maxime Louis 2019). This is especially interesting if we consider the fact that longitudinal cohorts are growing in size and data quality. However, the number of parameters may quickly rise so that computations are expensive and regularization needed to avoid overfitting. For example (Maxime Louis 2019) propose to learn the following Riemannian Metric from the data:

$$G^{-1}(z) = \sum_{i=1}^s L_{\psi_i} L_{\psi_i}^T \exp\left(-\frac{\|z - c_i\|_2^2}{T^2}\right) + \lambda Id \quad (8.1)$$

with (U_i) a set of s lower triangular matrixes with positive diagonal coefficients, (c_i) a set of s control points in the manifold, T is a temperature parameter to control how close from control points the metric is, and λ is a regularization parameter that set the metric when

far from. This amounts to estimate $s \times K(K + 1)$ parameters for the Riemannian metric.

Flexibility and interpretability of Deep Models

On the other side, fully Deep Learning based model, such as our proposed longitudinal autoencoders are both flexible and simple, and allow to concentrate on the graphical model in the latent space.

However, they still suffer major drawbacks for adoption. First, with more constraints than a classical autoencoder architecture, they provide a slightly higher reconstruction error. In this work this was not our main concern, as we focused on proofs of concepts of the method, if possible in subsampled images to decrease the computation costs. Furthermore, we believe independent work could be performed on the architecture to increase image quality. For example a classical solution is to add a discriminator loss on the reconstructed samples, to ensure more realistic data are generated. Another possibility consists in leveraging Riemannian metric learning to improve sampling quality. The idea is to consider the latent space no more as an Euclidean space, but rather a smooth Riemannian Manifold. Indeed, in an agnostic autoencoder setup, the latent space has few reasons to behave as an Euclidean space. The decoder may be regularized to behave as an isometry in order to force the learning of a flat latent space (N. Chen et al. 2020; Gropp, Atzmon, and Lipman 2020). Then, it has been proposed to learn the metric of the latent space by approximating the Jacobian of the decoder, and pushback the metric in the data space (Arvanitidis, Georgiev, and Schölkopf 2021; Arvanitidis, Hansen, and Hauberg 2017). Finally, (Chadebec et al. 2021) propose to learn the metric directly from the data in the manner of (Maxime Louis 2019), allowing for significantly better sampling quality, and distances in the latent in order to perform clustering.

Second and most importantly, interpretability is hardly achievable with deep learning models, and relevant inductive biases are needed to learn good representations. For example in our longitudinal autoencoder, while modifying ψ allows to move in terms of disease stage on the average trajectory, only visual assessment allows to comprehend the effect of spatial variability with regard to the average trajectory. Ensuring maximal disentanglement between spatial and temporal variability is already a challenging task (see 7). Another lead on improving interpretability of the model could be to lean on visual interpretability, which has become a huge topic in the computer vision community (W. Liu et al. 2020; Selvaraju et al. 2017) and medical imaging (Thibeau-Sutre et al. 2020). Incorporating this visual interpretability, either via attention modules or a-posteriori analysis, could provide valuable insights on the data and the model. This is all the more interesting in the sense that we could leverage the longitudinal setup to regularize visual interpretability method in the temporal dimension.

A-posteriori analysis of the models

In all our clinical analysis, we estimate a population and individual trajectories of disease progression, which may be used for regression tasks, or missing data imputation for example. A-posteriori analysis of the individual parameters encoding for geometrical variations of

the average trajectory is performed via correlations and ANOVA with cofactors of interest. This proves to be informative, and shows for example that medical conditions before diagnosis and demographics are significantly associated with the form of the disease.

However, these are correlations, not causality. Correlation analysis is long and tedious, and significativity can be tricky. It is not always clear how much to correct for p-values and cofactors can be correlated between themselves. Causal inference, by building probabilistic graphical models, may be better suited to understand of the pathological and clinical processes (Lartigue, Durrleman, and Allassonnière 2020). We believe that building interpretable, causal models of progression, leveraging the flexibility of deep learning, could help gain a better understanding of disease underlying processes.

Reproducibility : Data and Models

Reproducibility is key to produce useful research.

First, clinical results should ideally be replicated on independent cohorts. This is increasingly the case, becoming a pre-requisite for credibility. If not, biases such as patient inclusion, cohort design and modalities are numerous in line to alter the results. In the case where sharing real data is not an option, realistic synthetic data, or federated learning could provide good alternatives.

Then, models themselves are also prone to biases, as they are often designed to deal with the data at hand. This can be remedied by code availability on Git and easiness of use. Indeed, Leaspy in the parametric modeling case, and Longitudinal Autoencoders in the more flexible case allow for reproducibility of our main results to the presented cohorts, but hopefully also to unseen cohorts of patients.

Reproducibility is fostered by enterprise such as Europond (Europond n.d.), which allows experienced researchers and newcomers to benchmark different models fast.

Conclusion

At the crosswalk between medicine and machine learning, the goal of the disease progression modeling field is ambitious : no less than uncovering new insights on (neurodegenerative) diseases, leveraging data to make and possibly validate new hypotheses, based on our current understanding of the disease process.

Indeed, we currently only grasp a tip of the iceberg of the pathological processes underlying neurogenerative diseases. In this work, we provided contributions on disease progression models and Parkinson's Disease understanding. However, taking a step back and looking at the last decade, at the current work of fellow colleagues and researcher, and at the growing size and sharing of medical data, I believe we have good reasons to be hopeful for what's to come.

Publications

Scientific Articles

Peer-reviewed conference articles

- Raphaël Couronné, Marie Vidailhet, Jean-Christophe Corvol, Stéphane Lehericy, Stanley Durrleman. Learning disease progression models with longitudinal data and missing values. (2019). Oral presentation at the International Symposium on Biomedical Imaging, Apr 2019, Venice, Italy. Nominated for best paper awards.
- Maxime Louis, Raphaël Couronné, Igor Koval, Benjamin Charlier, Stanley Durrleman. (2019). Riemannian Geometry Learning for Disease Progression Modelling. Poster presentation at Information Processing in Medical Imaging - (IPMI) Jun 2019, Hong-Kong, China.
- Raphaël Couronné*, Paul Vernhet* (* : equal contribution), Stanley Durrleman. (2019). Riemannian Geometry Learning for Disease Progression Modelling. Poster presentation at Information Processing in Medical Imaging - (IPMI) Jun 2019, Hong-Kong, China.

Journal Articles

- Raphaël Couronné, Philipp Probst, Anne-Laure Boulesteix.(2018). Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinformatics volume 19, Article number: 270
- Manon Ansart, Stéphane Epelbaum, Giulia Bassignana, Alexandre Bône, Simona Bottani, Tiziana Cattai, Raphaël Couronné, Johann Faouzi, Igor Koval, Maxime Louis, Elina Thibaud-Sutre, Junhao Wen, Adam Wild, Ninon Burgos, Didier Dormont, Olivier Colliot, Stanley Durrleman. (2021). Predicting the Progression of Mild Cognitive Impairment Using Machine Learning: A Systematic, Quantitative and Critical Review. Medical Image Analysis, Elsevier, 67, pp.101848.

Abstracts

- Raphaël Couronné, Marie Vidailhet, Graziella Mangone, Stéphane Lehericy, Stanley Durrleman. Motor abnormalities progression in prodromal PD. (2019). Oral presen-

tation at the 14th International Conference on Alzheimer's and Parkinson's Diseases and related neurological disorders (AD/PD), Mar 2019, Lisbon, Portugal.

- Raphaël Couronné, Arnaud Valladier, Marie Vidailhet, Jean Christophe Corvol, Stéphane Lehericy, et al.. Longitudinal comparison of subjects with and without Sleep Disorders in Parkinson's Disease. Oral presentation at the conference Virtual Physiological Human (VPH), Aug 2020, Paris, France.
- Raphaël Couronné, Arnaud Valladier, Marie Vidhaillet, Jean Christophe Corvol, Stéphane Lehericy, et al.. Modeling the progression of Parkinson's Disease: comparison of subjects with and without Sleep Disorders. (2020). Poster presentation at Computational approaches for ageing and age-related diseases (CompAge), Sep 2020, Paris, France
- Raphaël Couronné, Arnaud Valladier, Stanley Durrleman. Privacy Preserving synthetic longitudinal data. (2021). Poster presentation at the 16th International Conference on Alzheimer's and Parkinson's Diseases and related neurological disorders (AD/PD), Mar 2021, Online.

Pre-prints

- Raphaël Couronné, Maxime Louis, Stanley Durrleman. Longitudinal autoencoder for multi-modal disease progression modelling. (2019).

Bibliography

- [1] *A Data-Driven Model of Biomarker Changes in Sporadic Alzheimer’s Disease* - *PubMed*. <https://pubmed-ncbi-nlm-nih-gov.accesdistant.sorbonne-universite.fr/25012224/>.
- [2] Dag Aarsland et al. “Prevalence and characteristics of dementia in Parkinson disease: an 8-year prospective study”. In: *Archives of neurology* 60.3 (2003), pp. 387–392.
- [3] Charles H. Adler and Thomas G. Beach. “Neuropathological Basis of Nonmotor Manifestations of Parkinson’s Disease”. eng. In: *Movement Disorders: Official Journal of the Movement Disorder Society* 31.8 (Aug. 2016), pp. 1114–1119. ISSN: 1531-8257. DOI: 10.1002/mds.26605.
- [4] Darko Aleksovski et al. “Disease Progression in Parkinson Subtypes: The PPMI Dataset”. eng. In: *Neurological Sciences: Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology* 39.11 (Nov. 2018), pp. 1971–1976. ISSN: 1590-3478. DOI: 10.1007/s10072-018-3522-z.
- [5] Stéphanie Allasonnière and Juliette Chevallier. “A New Class of EM Algorithms. Escaping Local Minima and Handling Intractable Sampling”. June 2019.
- [6] Stéphanie Allasonnière, Estelle Kuhn, and Alain Trouvé. “Construction of Bayesian Deformable Models via a Stochastic Approximation Algorithm: A Convergence Study”. In: *Bernoulli* 16.3 (Aug. 2010), pp. 641–678. ISSN: 1350-7265. DOI: 10.3150/09-BEJ229.
- [7] Julius BM Anang et al. “Predictors of dementia in Parkinson disease: a prospective cohort study”. In: *Neurology* 83.14 (2014), pp. 1253–1260.
- [8] Manon Ansart et al. “Predicting the progression of mild cognitive impairment using machine learning: a systematic, quantitative and critical review”. In: *Medical Image Analysis* (2020), p. 101848.
- [9] Georgios Arvanitidis, Bogdan Georgiev, and Bernhard Schölkopf. “A Prior-Based Approximate Latent Riemannian Metric”. In: *arXiv:2103.05290 [cs, stat]* (Mar. 2021). arXiv: 2103.05290 [cs, stat].
- [10] Georgios Arvanitidis, Lars Kai Hansen, and Søren Hauberg. “Latent Space Oddity: on the Curvature of Deep Generative Models”. In: *arXiv preprint arXiv:1710.11379* (2017).

- [11] Yves F. Atchadé. “An Adaptive Version for the Metropolis Adjusted Langevin Algorithm with a Truncated Drift”. en. In: *Methodology and Computing in Applied Probability* 8.2 (June 2006), pp. 235–254. ISSN: 1573-7713. DOI: 10.1007/s11009-006-8550-0.
- [12] P. Bédard et al. “The Nigrostriatal Pathway: A Correlative Study Based on Neuroanatomical and Neurochemical Criteria in the Cat and the Monkey”. eng. In: *Experimental Neurology* 25.3 (Nov. 1969), pp. 365–377. ISSN: 0014-4886. DOI: 10.1016/0014-4886(69)90131-9.
- [13] Alim Louis Benabid. “Deep Brain Stimulation for Parkinson’s Disease”. en. In: *Current Opinion in Neurobiology* 13.6 (Dec. 2003), pp. 696–706. ISSN: 0959-4388. DOI: 10.1016/j.conb.2003.11.001.
- [14] Yoshua Bengio, Aaron Courville, and Pascal Vincent. “Representation learning: A review and new perspectives”. In: *IEEE transactions on pattern analysis and machine intelligence* 35.8 (2013), pp. 1798–1828.
- [15] Samuel I. Berchuck, Sayan Mukherjee, and Felipe A. Medeiros. “Estimating Rates of Progression and Predicting Future Visual Fields in Glaucoma Using a Deep Variational Autoencoder”. In: *Scientific Reports* 9.1 (Dec. 2019). Number: 1 Publisher: Nature Publishing Group, p. 18113. ISSN: 2045-2322. DOI: 10.1038/s41598-019-54653-6. URL: <https://www.nature.com/articles/s41598-019-54653-6> (visited on 12/02/2020).
- [16] Daniela Berg, Per Borghammer, et al. “Prodromal Parkinson disease subtypes—key to understanding heterogeneity”. In: *Nature Reviews Neurology* (2021), pp. 1–13.
- [17] Daniela Berg, Ronald B Postuma, et al. “MDS research criteria for prodromal Parkinson’s disease”. In: *Movement Disorders* 30.12 (2015), pp. 1600–1611.
- [18] H. Bernheimer et al. “Brain Dopamine and the Syndromes of Parkinson and Huntington. Clinical, Morphological and Neurochemical Correlations”. eng. In: *Journal of the Neurological Sciences* 20.4 (Dec. 1973), pp. 415–455. ISSN: 0022-510X. DOI: 10.1016/0022-510x(73)90175-5.
- [19] EH Bigio et al. “Synapse loss is greater in presenile than senile onset Alzheimer disease: implications for the cognitive reserve hypothesis”. In: *Neuropathology and applied neurobiology* 28.3 (2002), pp. 218–227.
- [20] Emma Biondetti et al. “Spatiotemporal Changes in Substantia Nigra Neuromelanin Content in Parkinson’s Disease”. eng. In: *Brain: A Journal of Neurology* 143.9 (Sept. 2020), pp. 2757–2770. ISSN: 1460-2156. DOI: 10.1093/brain/awaa216.
- [21] Mathieu Blondel et al. “Fast Differentiable Sorting and Ranking”. en. In: *International Conference on Machine Learning*. ISSN: 2640-3498. PMLR, Nov. 2020, pp. 950–959. (Visited on 12/02/2020).
- [22] Bradley F Boeve et al. “Pathophysiology of REM sleep behaviour disorder and relevance to neurodegenerative disease”. In: *Brain* 130.11 (2007), pp. 2770–2788.

- [23] Alexandre Bone, Olivier Colliot, and Stanley Durrleman. “Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms”. In: *CVPR 2018 - Computer Vision and Pattern Recognition 2018*. Salt Lake City, United States, June 2018.
- [24] Alexandre Bône. “Learning Adapted Coordinate Systems for the Statistical Analysis of Anatomical Shapes. Applications to Alzheimer’s Disease Progression Modeling.” en. PhD thesis. Sorbonne Université, Jan. 2020.
- [25] Alexandre Bône, Olivier Colliot, and Stanley Durrleman. “Learning Distributions of Shape Trajectories From Longitudinal Datasets: A Hierarchical Model on a Manifold of Diffeomorphisms”. In: 2018, pp. 9271–9280. (Visited on 12/02/2020).
- [26] Alexandre Bône, Maxime Louis, Olivier Colliot, et al. “Learning Low-Dimensional Representations of Shape Data Sets with Diffeomorphic Autoencoders”. en. In: *Information Processing in Medical Imaging*. Ed. by Albert C. S. Chung et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 195–207. ISBN: 978-3-030-20351-1. DOI: 10.1007/978-3-030-20351-1_15.
- [27] Alexandre Bône, Maxime Louis, Benoît Martin, et al. “Deformetrica 4: an open-source software for statistical shape analysis”. In: *ShapeMI @ MICCAI 2018*. Granada, Spain, Sept. 2018. URL: <https://hal.inria.fr/hal-01874752>.
- [28] Diane Bouchacourt, Ryota Tomioka, and Sebastian Nowozin. “Multi-Level Variational Autoencoder: Learning Disentangled Representations from Grouped Observations”. In: *CoRR* abs/1705.08841 (2017). eprint: 1705.08841.
- [29] Heiko Braak et al. “Staging of brain pathology related to sporadic Parkinson’s disease”. In: *Neurobiology of aging* 24.2 (2003), pp. 197–211.
- [30] Rasmus Bro and Age K Smilde. “Principal component analysis”. In: *Analytical methods* 6.9 (2014), pp. 2812–2831.
- [31] Christopher P. Burgess et al. “Understanding disentangling in β -VAE”. In: *arXiv:1804.03599 [cs, stat]* (Apr. 2018). arXiv: 1804.03599. (Visited on 12/02/2020).
- [32] Manfredo do Carmo. *Riemannian Geometry*. en. Mathematics: Theory & Applications. Birkhäuser Basel, 1992. ISBN: 978-0-8176-3490-2.
- [33] Clément Chadebec et al. “Data Augmentation in High Dimensional Low Sample Size Setting Using a Geometry-Based Variational Autoencoder”. In: *arXiv preprint arXiv:2105.00026* (2021).
- [34] Agisilaos Chartsias et al. “Multimodal MR synthesis via modality-invariant latent representation”. In: *IEEE transactions on medical imaging* 37.3 (2018), pp. 803–814.
- [35] K. Ray Chaudhuri et al. “Non-Motor Symptoms of Parkinson’s Disease: Diagnosis and Management”. eng. In: *The Lancet. Neurology* 5.3 (Mar. 2006), pp. 235–245. ISSN: 1474-4422. DOI: 10.1016/S1474-4422(06)70373-8.
- [36] Nutan Chen et al. “Learning Flat Latent Manifolds with VAEs”. In: *arXiv:2002.04881 [cs, stat]* (Aug. 2020). arXiv: 2002.04881 [cs, stat].

- [37] Ricky TQ Chen et al. “Isolating sources of disentanglement in variational autoencoders”. In: *Advances in Neural Information Processing Systems* 31 (2018), pp. 2610–2620.
- [38] Roberto Cilia et al. “Natural History of Motor Symptoms in Parkinson’s Disease and the Long-Duration Response to Levodopa”. eng. In: *Brain: A Journal of Neurology* 143.8 (Aug. 2020), pp. 2490–2501. ISSN: 1460-2156. DOI: 10.1093/brain/awaa181.
- [39] Mayo Clinic. *MayoClinicPark*. <https://www.mayoclinic.org/diseases-conditions/parkinsons-disease/symptoms-causes/syc-20376055>. Accessed: 2021-04-25.
- [40] Jean-Christophe Corvol, Fanny Artaud, Florence Cormier-Dequaire, Olivier Rascol, Franck Durif, Pascal Derkinderen, Ana-Raquel Marques, Frédéric Bourdain, Jean-Philippe Brandel, Fernando Pico, et al. “Longitudinal analysis of impulse control disorders in Parkinson disease”. In: *Neurology* 91.3 (2018), e189–e201.
- [41] Jean-Christophe Corvol, Fanny Artaud, Florence Cormier-Dequaire, Olivier Rascol, Franck Durif, Pascal Derkinderen, Ana-Raquel Marques, Frédéric Bourdain, Jean-Philippe Brandel, Fernando Pico, et al. “Longitudinal Analysis of Impulse Control Disorders in Parkinson Disease”. eng. In: *Neurology* 91.3 (July 2018), e189–e201. ISSN: 1526-632X. DOI: 10.1212/WNL.0000000000005816.
- [42] Raphael Couronne et al. “Learning disease progression models with longitudinal data and missing values”. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE. 2019, pp. 1033–1037.
- [43] Koval I Couronne R. *leaspy*. <https://gitlab.com/icm-institute/aramislab/leaspy/>. Accessed: 2021-04-25.
- [44] Antonia Creswell et al. “Generative adversarial networks: An overview”. In: *IEEE Signal Processing Magazine* 35.1 (2018), pp. 53–65.
- [45] Adrian V. Dalca et al. “Learning Conditional Deformable Templates with Convolutional Networks”. en. In: *arXiv:1908.02738 [cs, eess]* (Oct. 2019). arXiv: 1908.02738. (Visited on 12/02/2020).
- [46] P. Damier et al. “The Substantia Nigra of the Human Brain. II. Patterns of Loss of Dopamine-Containing Neurons in Parkinson’s Disease”. eng. In: *Brain: A Journal of Neurology* 122 (Pt 8) (Aug. 1999), pp. 1437–1448. ISSN: 0006-8950. DOI: 10.1093/brain/122.8.1437.
- [47] Eduardo De Pablo-Fernández et al. “Prognosis and Neuropathologic Correlation of Clinical Subtypes of Parkinson Disease”. eng. In: *JAMA neurology* 76.4 (Apr. 2019), pp. 470–479. ISSN: 2168-6157. DOI: 10.1001/jamaneuro.2018.4377.
- [48] Vianney Debavelaere and Stéphanie Allasonnière. “On the Curved Exponential Family in the Stochastic Approximation Expectation Maximization Algorithm”. Feb. 2021.
- [49] Vianney Debavelaere, Stanley Durrleman, and Stéphanie Allasonnière. “On the Convergence of Stochastic Approximations under a Subgeometric Ergodic Markov Dynamic”. Dec. 2020.

- [50] Deep-Brain Stimulation for Parkinson’s Disease Study Group et al. “Deep-Brain Stimulation of the Subthalamic Nucleus or the Pars Interna of the Globus Pallidus in Parkinson’s Disease”. eng. In: *The New England Journal of Medicine* 345.13 (Sept. 2001), pp. 956–963. ISSN: 0028-4793. DOI: 10.1056/NEJMoa000827.
- [51] Bernard Delyon, Marc Lavielle, and Eric Moulines. “Convergence of a Stochastic Approximation Version of the EM Algorithm”. In: *The Annals of Statistics* 27.1 (Mar. 1999), pp. 94–128. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/aos/1018031103.
- [52] Michael C. Donohue et al. “Estimating Long-Term Multivariate Progression from Short-Term Data”. en. In: *Alzheimer’s & Dementia* 10.5, Supplement (Oct. 2014), S400–S410. ISSN: 1552-5260. DOI: 10.1016/j.jalz.2013.10.003.
- [53] Yong Du, Wei Wang, and Liang Wang. “Hierarchical Recurrent Neural Network for Skeleton Based Action Recognition”. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2015, pp. 1110–1118. DOI: 10.1109/CVPR.2015.7298714.
- [54] Aline Duarte Folle et al. “Clinical Progression in Parkinson’s Disease with Features of REM Sleep Behavior Disorder: A Population-Based Longitudinal Study”. eng. In: *Parkinsonism & Related Disorders* 62 (May 2019), pp. 105–111. ISSN: 1873-5126. DOI: 10.1016/j.parkreldis.2019.01.018.
- [55] Stanley Durrleman et al. “Toward a Comprehensive Framework for the Spatiotemporal Statistical Analysis of Longitudinal Shape Data”. In: *International Journal of Computer Vision* 103.1 (Nov. 2012), pp. 22–59. DOI: 10.1007/s11263-012-0592-x. URL: <https://doi.org/10.1007%2Fs11263-012-0592-x>.
- [56] Robert S Eisinger et al. “Motor subtype changes in early Parkinson’s disease”. In: *Parkinsonism & related disorders* 43 (2017), pp. 67–72.
- [57] Robert J Ellis et al. “A validated smartphone-based assessment of gait and gait variability in Parkinson’s disease”. In: *PLoS one* 10.10 (2015), e0141694.
- [58] Roberto Erro et al. “The Role of Disease Duration and Severity on Novel Clinical Subtypes of Parkinson Disease”. en. In: *Parkinsonism & Related Disorders* 73 (Apr. 2020), pp. 31–34. ISSN: 1353-8020. DOI: 10.1016/j.parkreldis.2020.03.013.
- [59] Europond. *Europond*. <http://europond.eu/software/>. Accessed: 2021-04-25.
- [60] Louis Falissard et al. “Deep clustering of longitudinal data”. In: *arXiv preprint arXiv:1802.03212* (2018).
- [61] Johann Faouzi et al. “Exploratory analysis of the genetics of impulse control disorders in Parkinson’s disease using genetic risk scores”. In: *Parkinsonism & Related Disorders* 86 (2021), pp. 74–77.
- [62] J. M. Fearnley and A. J. Lees. “Ageing and Parkinson’s Disease: Substantia Nigra Regional Selectivity”. eng. In: *Brain: A Journal of Neurology* 114 (Pt 5) (Oct. 1991), pp. 2283–2301. ISSN: 0006-8950. DOI: 10.1093/brain/114.5.2283.

- [63] Seyed-Mohammad Fereshtehnejad, Silvia Rios Romenets, Julius B. M. Anang, et al. “New Clinical Subtypes of Parkinson Disease and Their Longitudinal Progression: A Prospective Cohort Comparison With Other Phenotypes”. eng. In: *JAMA neurology* 72.8 (Aug. 2015), pp. 863–873. ISSN: 2168-6157. DOI: 10.1001/jamaneuro.2015.0703.
- [64] Seyed-Mohammad Fereshtehnejad, Silvia Rios Romenets, Julius BM Anang, et al. “New clinical subtypes of Parkinson disease and their longitudinal progression: a prospective cohort comparison with other phenotypes”. In: *JAMA neurology* 72.8 (2015), pp. 863–873.
- [65] Seyed-Mohammad Fereshtehnejad, Chun Yao, et al. “Evolution of prodromal Parkinson’s disease and dementia with Lewy bodies: a prospective study”. In: *Brain* 142.7 (2019), pp. 2051–2067.
- [66] Seyed-Mohammad Fereshtehnejad, Yashar Zeighami, et al. “Clinical criteria for subtyping Parkinson’s disease: biomarkers and longitudinal progression”. In: *Brain* 140.7 (2017), pp. 1959–1976.
- [67] Leslie J Findley. “The economic impact of Parkinson’s disease”. In: *Parkinsonism & related disorders* 13 (2007), S8–S12.
- [68] R. A. Fisher. “XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance.” en. In: *Earth and Environmental Science Transactions of The Royal Society of Edinburgh* 52.2 (1919/ed), pp. 399–433. ISSN: 2053-5945, 0080-4568. DOI: 10.1017/S0080456800012163.
- [69] Hubert M. Fonteijn et al. “An Event-Based Disease Progression Model and Its Application to Familial Alzheimer’s Disease”. eng. In: *Information Processing in Medical Imaging: Proceedings of the ... Conference* 22 (2011), pp. 748–759. ISSN: 1011-2499. DOI: 10.1007/978-3-642-22092-0_61.
- [70] *Forme précoce de la maladie de Parkinson*. fr-FR. <https://www.parkinson.ca/fr/la-maladie/forme-precoc-de-la-maladie-de-parkinson/>.
- [71] Vincent Fortuin et al. *SOM-VAE: Interpretable Discrete Representation Learning on Time Series*. 2019. arXiv: 1806.02199 [cs.LG].
- [72] Sante Publique France. *SantePubliqueFrance*. <https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-neurodegeneratives/maladie-de-parkinson>. Accessed: 2021-06-15.
- [73] Raúl de la Fuente-Fernández et al. “Age-Specific Progression of Nigrostriatal Dysfunction in Parkinson’s Disease”. en. In: *Annals of Neurology* 69.5 (2011), pp. 803–810. ISSN: 1531-8249. DOI: 10.1002/ana.22284.
- [74] Linlin Gao et al. “Brain Disease Diagnosis Using Deep Learning Features from Longitudinal MR Images”. In: *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data*. Springer. 2018, pp. 327–339.

- [75] Michel Goedert et al. “100 Years of Lewy Pathology”. eng. In: *Nature Reviews. Neurology* 9.1 (Jan. 2013), pp. 13–24. ISSN: 1759-4766. DOI: 10.1038/nrneuro1.2012.242.
- [76] Ian Goodfellow et al. “Generative adversarial nets”. In: *Advances in neural information processing systems*. 2014, pp. 2672–2680.
- [77] Will Grathwohl and Aaron Wilson. “Disentangling space and time in video with hierarchical variational auto-encoders”. In: *arXiv preprint arXiv:1612.04440* (2016).
- [78] Julia C. Greenland, Caroline H. Williams-Gray, and Roger A. Barker. “The Clinical Heterogeneity of Parkinson’s Disease and Its Therapeutic Implications”. eng. In: *The European Journal of Neuroscience* 49.3 (Feb. 2019), pp. 328–338. ISSN: 1460-9568. DOI: 10.1111/ejn.14094.
- [79] Sandrine Greffard et al. “Motor Score of the Unified Parkinson Disease Rating Scale as a Good Predictor of Lewy Body-Associated Neuronal Loss in the Substantia Nigra”. eng. In: *Archives of Neurology* 63.4 (Apr. 2006), pp. 584–588. ISSN: 0003-9942. DOI: 10.1001/archneur.63.4.584.
- [80] Amos Gropp, Matan Atzmon, and Yaron Lipman. “Isometric Autoencoders”. In: *arXiv:2006.09289 [cs, stat]* (Oct. 2020). arXiv: 2006.09289 [cs, stat].
- [81] Parkinson Study Group. “Levodopa and the progression of Parkinson’s disease”. In: *New England Journal of Medicine* 351.24 (2004), pp. 2498–2508.
- [82] R. Guerrero et al. “Instantiated Mixed Effects Modeling of Alzheimer’s Disease Markers”. en. In: *NeuroImage* 142 (Nov. 2016), pp. 113–125. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2016.06.049.
- [83] Glenda Halliday et al. “The Progression of Pathology in Longitudinally Followed Patients with Parkinson’s Disease”. eng. In: *Acta Neuropathologica* 115.4 (Apr. 2008), pp. 409–415. ISSN: 0001-6322. DOI: 10.1007/s00401-008-0344-8.
- [84] Glenda Margaret Halliday and Heather McCann. “The Progression of Pathology in Parkinson’s Disease”. eng. In: *Annals of the New York Academy of Sciences* 1184 (Jan. 2010), pp. 188–195. ISSN: 1749-6632. DOI: 10.1111/j.1749-6632.2009.05118.x.
- [85] Mohammad Havaei et al. “Brain tumor segmentation with deep neural networks”. In: *Medical image analysis* 35 (2017), pp. 18–31.
- [86] Runcheng He et al. “Recent advances in biomarkers for Parkinson’s disease”. In: *Frontiers in aging neuroscience* 10 (2018), p. 305.
- [87] Sebastian Heinzel et al. “Update of the MDS research criteria for prodromal Parkinson’s disease”. In: *Movement Disorders* 34.10 (2019), pp. 1464–1470.
- [88] Irina Higgins, David Amos, et al. *Towards a Definition of Disentangled Representations*. 2018. arXiv: 1812.02230 [cs.LG].
- [89] Irina Higgins, Loic Matthey, et al. “beta-vae: Learning basic visual concepts with a constrained variational framework”. In: (2016).

- [90] Ruediger Hilker et al. “Nonlinear Progression of Parkinson Disease as Determined by Serial Positron Emission Tomographic Imaging of Striatal Fluorodopa F 18 Activity”. eng. In: *Archives of Neurology* 62.3 (Mar. 2005), pp. 378–382. ISSN: 0003-9942. DOI: 10.1001/archneur.62.3.378.
- [91] Bernd Holdorff, Antonio M. Rodrigues e Silva, and Richard Dodel. “Centenary of Lewy Bodies (1912–2012)”. en. In: *Journal of Neural Transmission* 120.4 (Apr. 2013), pp. 509–516. ISSN: 1435-1463. DOI: 10.1007/s00702-013-0984-2.
- [92] Yi Hong et al. “Time-Warped Geodesic Regression”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*. Springer International Publishing, 2014, pp. 105–112. DOI: 10.1007/978-3-319-10470-6_14. URL: https://doi.org/10.1007%2F978-3-319-10470-6_14.
- [93] Jacob Horsager, Katrine B Andersen, Karoline Knudsen, Casper Skjærbæk, et al. “Brain-first versus body-first Parkinson’s disease: a multimodal imaging case-control study”. In: *Brain* 143.10 (2020), pp. 3077–3088.
- [94] Jacob Horsager, Katrine B Andersen, Karoline Knudsen, Casper Skjærbæk, et al. “Brain-first versus body-first Parkinson’s disease: a multimodal imaging case-control study”. In: *Brain* 143.10 (Aug. 2020), pp. 3077–3088. ISSN: 0006-8950. DOI: 10.1093/brain/awaa238. eprint: <https://academic.oup.com/brain/article-pdf/143/10/3077/34032305/awaa238.pdf>. URL: <https://doi.org/10.1093/brain/awaa238>.
- [95] H. Hotelling. “Analysis of a Complex of Statistical Variables into Principal Components”. In: *Journal of Educational Psychology* 24.6 (1933), pp. 417–441. ISSN: 1939-2176(Electronic),0022-0663(Print). DOI: 10.1037/h0071325.
- [96] Wei-Ning Hsu, Yu Zhang, and James Glass. “Unsupervised learning of disentangled and interpretable representations from sequential data”. In: *Advances in neural information processing systems* 30 (2017), pp. 1878–1889.
- [97] Joseph G. Ibrahim and Geert Molenberghs. “Rejoinder on: Missing data methods in longitudinal studies: a review”. In: *TEST* 18.1 (Feb. 2009), pp. 68–75. DOI: 10.1007/s11749-009-0144-z. URL: <https://doi.org/10.1007%2Fs11749-009-0144-z>.
- [98] Clifford R. Jack, David S. Knopman, et al. “Hypothetical Model of Dynamic Biomarkers of the Alzheimer’s Pathological Cascade”. eng. In: *The Lancet. Neurology* 9.1 (Jan. 2010), pp. 119–128. ISSN: 1474-4465. DOI: 10.1016/S1474-4422(09)70299-6.
- [99] Clifford R. Jack, Heather J. Wiste, et al. “Defining Imaging Biomarker Cut Points for Brain Aging and Alzheimer’s Disease”. eng. In: *Alzheimer’s & Dementia: The Journal of the Alzheimer’s Association* 13.3 (Mar. 2017), pp. 205–216. ISSN: 1552-5279. DOI: 10.1016/j.jalz.2016.08.005.
- [100] Clifford R Jack Jr et al. “Hypothetical model of dynamic biomarkers of the Alzheimer’s pathological cascade”. In: *The Lancet Neurology* 9.1 (2010), pp. 119–128.

- [101] J. Jankovic. “Parkinson’s Disease: Clinical Features and Diagnosis”. eng. In: *Journal of Neurology, Neurosurgery, and Psychiatry* 79.4 (Apr. 2008), pp. 368–376. ISSN: 1468-330X. DOI: 10.1136/jnnp.2007.131045.
- [102] Joseph Jankovic et al. “Variable expression of Parkinson’s disease: A base-line analysis of the DAT ATOP cohort”. In: *Neurology* 40.10 (1990), pp. 1529–1529.
- [103] Laetitia Jeancolas. “Détection précoce de la maladie de Parkinson par l’analyse de la voix et corrélations avec la neuroimagerie”. Theses. Université Paris-Saclay, Dec. 2019. URL: <https://tel.archives-ouvertes.fr/tel-02470759>.
- [104] Bruno M. Jedynak et al. “A Computational Neurodegenerative Disease Progression Score: Method and Results with the Alzheimer’s Disease Neuroimaging Initiative Cohort”. en. In: *NeuroImage* 63.3 (Nov. 2012), pp. 1478–1486. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2012.07.059.
- [105] Fei Jiang et al. “Artificial intelligence in healthcare: past, present and future”. In: *Stroke and vascular neurology* 2.4 (2017).
- [106] Matthew J Johnson et al. “Composing graphical models with neural networks for structured representations and fast inference”. In: *arXiv preprint arXiv:1603.06277* (2016).
- [107] Lorraine V Kalia and Anthony E Lang. “Evolving basic, pathological and clinical concepts in PD”. In: *Nature reviews Neurology* 12.2 (2016), pp. 65–66.
- [108] Hyunjik Kim and Andriy Mnih. “Disentangling by Factorising”. en. In: *arXiv:1802.05983 [cs, stat]* (July 2019). arXiv: 1802.05983. (Visited on 12/02/2020).
- [109] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).
- [110] Diederik P Kingma and Max Welling. “Auto-encoding variational bayes”. In: *arXiv preprint arXiv:1312.6114* (2013).
- [111] Christine Klein and Ana Westenberger. “Genetics of Parkinson’s disease”. In: *Cold Spring Harbor perspectives in medicine* 2.1 (2012), a008888.
- [112] WC Koller. “When does Parkinson’s disease begin?” In: *Neurology* 42.4 Suppl 4 (1992), pp. 27–31.
- [113] Igor Koval. “Learning Multimodal Digital Models of Disease Progression from Longitudinal Data : Methods & Algorithms for the Description, Prediction and Simulation of Alzheimer’s Disease Progression”. en. PhD thesis. Institut Polytechnique de Paris, Jan. 2020.
- [114] Igor Koval, Alexandre Bône, et al. “AD Course Map Charts Alzheimer’s Disease Progression”. en. In: *Scientific Reports* 11.1 (Apr. 2021), p. 8020. ISSN: 2045-2322. DOI: 10.1038/s41598-021-87434-1.
- [115] Igor Koval, Thomas Dighiero-Brecht, et al. *Machine Learning Spots the Time to Treat Huntington Disease*. en. Preprint. In Review, Mar. 2021. DOI: 10.21203/rs.3.rs-264531/v1.

- [116] Julian Krebs, Hervé Delingette, Nicholas Ayache, et al. “Learning a Generative Motion Model from Image Sequences based on a Latent Motion Matrix”. en. In: *arXiv:2011.01741 [cs]* (Nov. 2020). arXiv: 2011.01741. (Visited on 12/02/2020).
- [117] Julian Krebs, Hervé Delingette, Boris Mailhé, et al. “Learning a Probabilistic Model for Diffeomorphic Registration”. en. In: *IEEE Transactions on Medical Imaging* 38.9 (Sept. 2019). arXiv: 1812.07460, pp. 2165–2176. ISSN: 0278-0062, 1558-254X. DOI: 10.1109/TMI.2019.2897112. (Visited on 12/02/2020).
- [118] J. B. Kruskal. “Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis”. en. In: *Psychometrika* 29.1 (Mar. 1964), pp. 1–27. ISSN: 1860-0980. DOI: 10.1007/BF02289565.
- [119] Estelle Kuhn and Marc Lavielle. “Coupling a stochastic approximation version of EM with an MCMC procedure”. fr. In: *ESAIM: Probability and Statistics* 8 (2004), pp. 115–131. DOI: 10.1051/ps:2004007.
- [120] Nan M. Laird and James H. Ware. “Random-Effects Models for Longitudinal Data”. In: *Biometrics* 38.4 (1982), pp. 963–974. ISSN: 0006-341X. DOI: 10.2307/2529876.
- [121] Nan M. Laird and James H. Ware. “Random-Effects Models for Longitudinal Data”. In: *Biometrics* 38.4 (Dec. 1982), p. 963. DOI: 10.2307/2529876. URL: <https://doi.org/10.2307%2F2529876>.
- [122] Thomas Lartigue, Stanley Durrleman, and Stéphanie Allasonnière. “Mixture of Conditional Gaussian Graphical Models for unlabelled heterogeneous populations in the presence of co-factors”. In: *arXiv preprint arXiv:2006.11094* (2020).
- [123] Jeanne C. Latourelle et al. “Large-Scale Identification of Clinical and Genetic Predictors of Motor Progression in Patients with Newly Diagnosed Parkinson’s Disease: A Longitudinal Cohort Study and Validation”. eng. In: *The Lancet. Neurology* 16.11 (Nov. 2017), pp. 908–916. ISSN: 1474-4465. DOI: 10.1016/S1474-4422(17)30328-9.
- [124] Brian Lau et al. “The Integrative Role of the Pedunculopontine Nucleus in Human Gait”. eng. In: *Brain: A Journal of Neurology* 138.Pt 5 (May 2015), pp. 1284–1296. ISSN: 1460-2156. DOI: 10.1093/brain/awv047.
- [125] Michael Lawton et al. “Developing and Validating Parkinson’s Disease Subtypes and Their Motor and Cognitive Progression”. eng. In: *Journal of Neurology, Neurosurgery, and Psychiatry* 89.12 (Dec. 2018), pp. 1279–1287. ISSN: 1468-330X. DOI: 10.1136/jnnp-2018-318337.
- [126] Yann LeCun, Yoshua Bengio, et al. “Convolutional networks for images, speech, and time series”. In: *The handbook of brain theory and neural networks* 3361.10 (1995), p. 1995.
- [127] Garam Lee et al. “Predicting Alzheimer’s disease progression using multi-modal deep learning approach”. In: *Scientific reports* 9.1 (2019), pp. 1–12.
- [128] Patricia N. Lee et al. “Asymmetric Developmental Potential along the Animal–Vegetal Axis in the Anthozoan Cnidarian, *Nematostella vectensis*, Is Mediated by Dishevelled”. en. In: *Developmental Biology* 310.1 (Oct. 2007), pp. 169–186. ISSN: 0012-1606. DOI: 10.1016/j.ydbio.2007.05.040.

- [129] SJG Lewis et al. “Heterogeneity of Parkinson’s disease in the early clinical stages using a data driven approach”. In: *Journal of Neurology, Neurosurgery & Psychiatry* 76.3 (2005), pp. 343–348.
- [130] Yingzhen Li and Stephan Mandt. “Disentangled sequential autoencoder”. In: *arXiv preprint arXiv:1803.02991* (2018).
- [131] Mary J. Lindstrom and Douglas M. Bates. “Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data”. In: *Journal of the American Statistical Association* 83.404 (1988), pp. 1014–1022. ISSN: 0162-1459. DOI: 10.2307/2290128.
- [132] Mary J. Lindstrom and Douglas M. Bates. “Nonlinear Mixed Effects Models for Repeated Measures Data”. In: *Biometrics* 46.3 (1990), pp. 673–687. ISSN: 0006-341X. DOI: 10.2307/2532087.
- [133] Mary J. Lindstrom and Douglas M. Bates. “Nonlinear Mixed Effects Models for Repeated Measures Data”. In: *Biometrics* 46.3 (Sept. 1990), p. 673. DOI: 10.2307/2532087. URL: <https://doi.org/10.2307/2532087>.
- [134] Ganqiang Liu et al. “Prediction of Cognition in Parkinson’s Disease with a Clinical-Genetic Score: A Longitudinal Analysis of Nine Cohorts”. eng. In: *The Lancet. Neurology* 16.8 (Aug. 2017), pp. 620–629. ISSN: 1474-4465. DOI: 10.1016/S1474-4422(17)30122-9.
- [135] Ping Liu et al. “Clinical heterogeneity in patients with early-stage Parkinson’s disease: a cluster analysis”. In: *Journal of Zhejiang University Science B* 12.9 (2011), pp. 694–703.
- [136] Wenqian Liu et al. “Towards visually explaining variational autoencoders”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 8642–8651.
- [137] Francesco Locatello et al. “Challenging common assumptions in the unsupervised learning of disentangled representations”. In: *international conference on machine learning*. PMLR. 2019, pp. 4114–4124.
- [138] Marco Lorenzi, Nicholas Ayache, and Xavier Pennec. “Schild’s Ladder for the Parallel Transport of Deformations in Time Series of Images”. en. In: *Information Processing in Medical Imaging*. Ed. by Gábor Székely and Horst K. Hahn. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2011, pp. 463–474. ISBN: 978-3-642-22092-0. DOI: 10.1007/978-3-642-22092-0_38.
- [139] Marco Lorenzi, Xavier Pennec, et al. “Disentangling normal aging from Alzheimer’s disease in structural magnetic resonance images”. In: *Neurobiology of Aging* 36 (Jan. 2015), S42–S52. DOI: 10.1016/j.neurobiolaging.2014.07.046. URL: <https://doi.org/10.1016/j.neurobiolaging.2014.07.046>.
- [140] Elan D Louis et al. “Mortality from Parkinson disease”. In: *Archives of neurology* 54.3 (1997), pp. 260–264.
- [141] M Louis et al. “Riemannian geometry learning for disease progression modelling”. In: *International Conference on Information Processing in Medical Imaging*. 2019.

- [142] Maxime Louis. “Computational and Statistical Methods for Trajectory Analysis in a Riemannian Geometry Setting”. en. PhD thesis. Sorbonne Université, Oct. 2019.
- [143] Maxime Louis, Benjamin Charlier, and Stanley Durrleman. “Geodesic Discriminant Analysis for manifold-valued data”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018, pp. 332–340.
- [144] Maxime Louis, Raphaël Couronné, et al. “Riemannian Geometry Learning for Disease Progression Modelling”. en. In: *Information Processing in Medical Imaging*. Vol. 11492. Series Title: Lecture Notes in Computer Science. Springer International Publishing, 2019, pp. 542–553. (Visited on 12/02/2020).
- [145] Bin Lu et al. “A Practical Alzheimer Disease Classifier via Brain Imaging-Based Deep Learning on 85,721 Samples”. In: *bioRxiv* (2021), pp. 2020–08.
- [146] Laurens van der Maaten and Geoffrey Hinton. “Visualizing Data Using T-SNE”. In: *Journal of Machine Learning Research* 9.86 (2008), pp. 2579–2605. ISSN: 1533-7928.
- [147] Angus D. Macleod et al. “Development and Validation of Prognostic Survival Models in Newly Diagnosed Parkinson’s Disease”. eng. In: *Movement Disorders: Official Journal of the Movement Disorder Society* 33.1 (Jan. 2018), pp. 108–116. ISSN: 1531-8257. DOI: 10.1002/mds.27177.
- [148] Philipp Mhlknecht, Klaus Seppi, and Werner Poewe. “The concept of prodromal Parkinson’s disease”. In: *Journal of Parkinson’s disease* 5.4 (2015), pp. 681–697.
- [149] Kenneth Marek, Sohini Chowdhury, et al. “The Parkinson’s Progression Markers Initiative (PPMI) – Establishing a PD Biomarker Cohort”. en. In: *Annals of Clinical and Translational Neurology* 5.12 (2018), pp. 1460–1477. ISSN: 2328-9503. DOI: 10.1002/acn3.644.
- [150] Kenneth Marek, Danna Jennings, et al. “The Parkinson Progression Marker Initiative (PPMI)”. English (US). In: *Progress in Neurobiology* 95.4 (Dec. 2011), pp. 629–635. ISSN: 0301-0082. DOI: 10.1016/j.pneurobio.2011.09.005.
- [151] Razvan V Marinescu et al. “Tadpole challenge: Prediction of longitudinal evolution in Alzheimer’s disease”. In: *arXiv preprint arXiv:1805.03909* (2018).
- [152] Răzvan V. Marinescu et al. “DIVE: A Spatiotemporal Progression Model of Brain Pathology in Neurodegenerative Disorders”. eng. In: *NeuroImage* 192 (May 2019), pp. 166–177. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2019.02.053.
- [153] Connie Marras and K Ray Chaudhuri. “Nonmotor features of Parkinson’s disease subtypes”. In: *Movement Disorders* 31.8 (2016), pp. 1095–1102.
- [154] Connie Marras and Anthony Lang. “Parkinson’s disease subtypes: lost in translation?” In: *Journal of Neurology, Neurosurgery & Psychiatry* 84.4 (2013), pp. 409–415.
- [155] Pablo Martínez-Martín et al. “Parkinson’s Disease Severity Levels and MDS-Unified Parkinson’s Disease Rating Scale”. eng. In: *Parkinsonism & Related Disorders* 21.1 (Jan. 2015), pp. 50–54. ISSN: 1873-5126. DOI: 10.1016/j.parkreldis.2014.10.026.

- [156] Emile Mathieu et al. “Disentangling disentanglement in variational autoencoders”. In: *International Conference on Machine Learning*. PMLR. 2019, pp. 4402–4412.
- [157] Geoffrey J McLachlan. *Discriminant analysis and statistical pattern recognition*. Vol. 544. John Wiley & Sons, 2004.
- [158] Deval Mehta et al. *Towards Automated and Marker-less Parkinson Disease Assessment: Predicting UPDRS Scores using Sit-stand videos*. 2021. arXiv: 2104.04650 [cs.CV].
- [159] Bjoern H Menze et al. “The multimodal brain tumor image segmentation benchmark (BRATS)”. In: *IEEE transactions on medical imaging* 34.10 (2014), pp. 1993–2024.
- [160] Tiago A Mestre et al. “Parkinson’s Disease Subtypes: Critical Appraisal and Recommendations”. In: *Journal of Parkinson’s disease* (2021).
- [161] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. “V-net: Fully convolutional neural networks for volumetric medical image segmentation”. In: *2016 fourth international conference on 3D vision (3DV)*. IEEE. 2016, pp. 565–571.
- [162] Riccardo Miotto et al. “Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records”. In: *Scientific Reports* 6 (May 2016), 26094 EP -. URL: <https://doi.org/10.1038/srep26094>.
- [163] Mariana Moscovich et al. “How specific are non-motor symptoms in the prodrome of Parkinson’s disease compared to other movement disorders?” In: *Parkinsonism & Related Disorders* 81 (2020), pp. 213–218.
- [164] Jesse Mu et al. “Parkinson’s disease subtypes identified from cluster analysis of motor and non-motor symptoms”. In: *Frontiers in aging neuroscience* 9 (2017), p. 301.
- [165] Daniel L. Murman. “Early Treatment of Parkinson’s Disease: Opportunities for Managed Care”. eng. In: *The American Journal of Managed Care* 18.7 Suppl (Sept. 2012), S183–188. ISSN: 1936-2692.
- [166] Jiquan Ngiam et al. “Multimodal deep learning”. In: *Proceedings of the 28th international conference on machine learning (ICML-11)*. 2011, pp. 689–696.
- [167] Neil P. Oxtoby et al. “Sequence of Clinical and Neurodegeneration Events in Parkinson’s Disease Progression”. eng. In: *Brain: A Journal of Neurology* 144.3 (Apr. 2021), pp. 975–988. ISSN: 1460-2156. DOI: 10.1093/brain/awaa461.
- [168] James Parkinson. “An Essay on the Shaking Palsy. 1817”. eng. In: *The Journal of Neuropsychiatry and Clinical Neurosciences* 14.2 (2002), 223–236, discussion 222. ISSN: 0895-0172. DOI: 10.1176/jnp.14.2.223.
- [169] Sérgio Pereira et al. “Brain tumor segmentation using convolutional neural networks in MRI images”. In: *IEEE transactions on medical imaging* 35.5 (2016), pp. 1240–1251.
- [170] Ronald Carl Petersen et al. “Alzheimer’s disease neuroimaging initiative (ADNI): clinical characterization”. In: *Neurology* 74.3 (2010), pp. 201–209.

- [171] José C Pinheiro and Douglas M Bates. “Approximations to the log-likelihood function in the nonlinear mixed-effects model”. In: *Journal of computational and Graphical Statistics* 4.1 (1995), pp. 12–35.
- [172] Werner Poewe et al. “Parkinson disease”. In: *Nature reviews Disease primers* 3.1 (2017), pp. 1–21.
- [173] RB Postuma, JF Gagnon, et al. “Quantifying the risk of neurodegenerative disease in idiopathic REM sleep behavior disorder”. In: *Neurology* 72.15 (2009), pp. 1296–1300.
- [174] RB Postuma, AE Lang, et al. “How does parkinsonism start? Prodromal parkinsonism motor changes in idiopathic REM sleep behaviour disorder”. In: *Brain* 135.6 (2012), pp. 1860–1870.
- [175] Ronald B Postuma and Daniela Berg. “Advances in markers of prodromal Parkinson disease”. In: *Nature Reviews Neurology* 12.11 (2016), p. 622.
- [176] Ronald B Postuma and Daniela Berg. “Prodromal Parkinson’s disease: the decade past, the decade to come”. In: *Movement disorders* 34.5 (2019), pp. 665–675.
- [177] Ronald B Postuma, Alex Iranzo, et al. “Risk and predictors of dementia and parkinsonism in idiopathic REM sleep behaviour disorder: a multicentre study”. In: *Brain* 142.3 (2019), pp. 744–759.
- [178] Elizabeth Qian and Yue Huang. “Subtyping of Parkinson’s disease-where are we up to?” In: *Aging and disease* 10.5 (2019), p. 1130.
- [179] Daniele Ravi, Daniel C. Alexander, and Neil P. Oxtoby. *Degenerative Adversarial NeuroImage Nets: Generating Images that Mimic Disease Progression*. 2019. arXiv: 1907.02787 [eess.IV].
- [180] JSAM Reijnders et al. “The association between motor subtypes and psychopathology in Parkinson’s disease”. In: *Parkinsonism & related disorders* 15.5 (2009), pp. 379–382.
- [181] Giulietta M. Riboldi and Alessio B. Di Fonzo. “GBA, Gaucher Disease, and Parkinson’s Disease: From Genetic to Clinic to New Therapeutic Approaches”. In: *Cells* 8.4 (Apr. 2019). ISSN: 2073-4409. DOI: 10.3390/cells8040364.
- [182] Walter A Rocca. “The burden of Parkinson’s disease: a worldwide perspective”. In: *The Lancet Neurology* 17.11 (2018), pp. 928–929.
- [183] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.
- [184] Alexandre Routier et al. *Clinica: an open source software platform for reproducible clinical neuroscience studies*. en. Oct. 2019. URL: <https://hal.inria.fr/hal-02308126> (visited on 12/02/2020).
- [185] David E Rumelhart, Geoffrey E Hinton, Ronald J Williams, et al. “Learning representations by back-propagating errors”. In: *Cognitive modeling* 5.3 (1988), p. 1.

- [186] Anna Sauerbier et al. “Non motor subtypes and Parkinson’s disease”. In: *Parkinsonism & related disorders* 22 (2016), S41–S46.
- [187] Rodolfo Savica, Walter A Rocca, and J Eric Ahlskog. “When does Parkinson disease start?” In: *Archives of neurology* 67.7 (2010), pp. 798–801.
- [188] Jean-Baptiste Schiratti. “Methods and Algorithms to Learn Spatio-Temporal Changes from Longitudinal Manifold-Valued Observations”. These de Doctorat. Université Paris-Saclay (ComUE), Jan. 2017.
- [189] Jean-Baptiste Schiratti, Stéphanie Allassonniere, et al. “Learning Spatiotemporal Trajectories from Manifold-Valued Longitudinal Data”. en. In: *Advances in Neural Information Processing Systems* 28 (2015).
- [190] Jean-Baptiste Schiratti, Stéphanie Allassonniere, et al. “Learning spatiotemporal trajectories from manifold-valued longitudinal data”. In: *Neural Information Processing Systems*. 28. 2015.
- [191] Jean-Baptiste Schiratti, Stéphanie Allassonniere, et al. “Learning spatiotemporal trajectories from manifold-valued longitudinal data”. In: *Advances in Neural Information Processing Systems*. 2015, pp. 2404–2412.
- [192] Jean-Baptiste Schiratti, Stéphanie Allassonniere, et al. “A Bayesian Mixed-Effects Model to Learn Trajectories of Changes from Repeated Manifold-Valued Observations”. In: *Journal of Machine Learning Research* 18.133 (2017), pp. 1–33. ISSN: 1533-7928.
- [193] Anette Schrag et al. “Young-versus older-onset Parkinson’s disease: impact of disease and psychosocial consequences”. In: *Movement disorders: official journal of the Movement Disorder Society* 18.11 (2003), pp. 1250–1256.
- [194] Marianna Selikhova et al. “A clinico-pathological study of subtypes in Parkinson’s disease”. In: *Brain* 132.11 (2009), pp. 2947–2957.
- [195] Ramprasaath R Selvaraju et al. “Grad-cam: Visual explanations from deep networks via gradient-based localization”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 618–626.
- [196] Amir Shahroudy et al. “NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2016, pp. 1010–1019. DOI: 10.1109/CVPR.2016.115.
- [197] Lei Shi et al. “Two-Stream Adaptive Graph Convolutional Networks for Skeleton-Based Action Recognition”. In: *arXiv:1805.07694 [cs]* (July 2019). arXiv: 1805.07694 [cs].
- [198] Jung Hwan Shin et al. “Longitudinal change in dopamine transporter availability in idiopathic REM sleep behavior disorder”. In: *Neurology* 95.23 (2020), e3081–e3092.
- [199] Nikhil Singh et al. “A Hierarchical Geodesic Model for Diffeomorphic Longitudinal Shape Analysis”. In: *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2013, pp. 560–571. DOI: 10.1007/978-3-642-38868-2_47. URL: https://doi.org/10.1007/978-3-642-38868-2_47.

- [200] Nikhil Singh et al. “An efficient parallel algorithm for hierarchical geodesic models in diffeomorphisms”. In: *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. IEEE, Apr. 2014. DOI: 10.1109/isbi.2014.6867878. URL: <https://doi.org/10.1109%2Fisbi.2014.6867878>.
- [201] Friederike Sixel-Döring et al. “Associated factors for REM sleep behavior disorder in Parkinson disease”. In: *Neurology* 77.11 (2011), pp. 1048–1054.
- [202] Suvrit Sra and Inderjit Dhillon. “Generalized Nonnegative Matrix Approximations with Bregman Divergences”. In: *Advances in Neural Information Processing Systems*. Ed. by Y. Weiss, B. Schölkopf, and J. Platt. Vol. 18. MIT Press, 2006. URL: <https://proceedings.neurips.cc/paper/2005/file/d58e2f077670f4de9cd7963c857f2534-Paper.pdf>.
- [203] Nitish Srivastava, Elman Mansimov, and Ruslan Salakhudinov. “Unsupervised learning of video representations using lstms”. In: *International conference on machine learning*. 2015, pp. 843–852.
- [204] Glenn T Stebbins et al. “How to identify tremor dominant and postural instability/gait difficulty groups with the movement disorder society unified Parkinson’s disease rating scale: comparison with the unified Parkinson’s disease rating scale”. In: *Movement Disorders* 28.5 (2013), pp. 668–670.
- [205] Karin Stiasny-Kolster et al. “The REM Sleep Behavior Disorder Screening Questionnaire—a New Diagnostic Instrument”. eng. In: *Movement Disorders: Official Journal of the Movement Disorder Society* 22.16 (Dec. 2007), pp. 2386–2393. ISSN: 0885-3185. DOI: 10.1002/mds.21740.
- [206] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. “A Global Geometric Framework for Nonlinear Dimensionality Reduction”. en. In: *Science* 290.5500 (Dec. 2000), pp. 2319–2323. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.290.5500.2319.
- [207] Mary Ann Thenganatt and Joseph Jankovic. “Parkinson disease subtypes”. In: *JAMA neurology* 71.4 (2014), pp. 499–504.
- [208] Elina Thibeau-Sutre et al. “Visualization approach to assess the robustness of neural networks for medical image classification”. In: *Medical Imaging 2020: Image Processing*. Vol. 11313. International Society for Optics and Photonics. 2020, 113131J.
- [209] Eduardo Tolosa et al. “Diagnosis and the premotor phase of Parkinson disease”. In: *Neurology* 72.7 Supplement 2 (2009), S12–S20.
- [210] *UK Biobank - UK Biobank*. <https://www.ukbiobank.ac.uk/>.
- [211] *Uncovering the Heterogeneity and Temporal Complexity of Neurodegenerative Diseases with Subtype and Stage Inference | Nature Communications*. <https://www-nature-com.accesdistant.sorbonne-universite.fr/articles/s41467-018-05892-0>.
- [212] Daan C. Velseboer et al. “Development and External Validation of a Prognostic Model in Newly Diagnosed Parkinson Disease”. eng. In: *Neurology* 86.11 (Mar. 2016), pp. 986–993. ISSN: 1526-632X. DOI: 10.1212/WNL.0000000000002437.

- [213] Charles S. Venuto et al. “A Review of Disease Progression Models of Parkinson’s Disease and Applications in Clinical Trials”. eng. In: *Movement Disorders: Official Journal of the Movement Disorder Society* 31.7 (July 2016), pp. 947–956. ISSN: 1531-8257. DOI: 10.1002/mds.26644.
- [214] Linbo Wang et al. “Association of specific biotypes in patients with Parkinson disease and disease progression”. In: *Neurology* 95.11 (2020), e1445–e1460.
- [215] Marie-Laure Welter et al. “PPNa-DBS for Gait and Balance Disorders in Parkinson’s Disease: A Double-Blind, Randomised Study”. eng. In: *Journal of Neurology* 262.6 (June 2015), pp. 1515–1525. ISSN: 1432-1459. DOI: 10.1007/s00415-015-7744-1.
- [216] Peter A. Wijeratne et al. “An Image-Based Model of Brain Volume Biomarker Changes in Huntington’s Disease”. en. In: *Annals of Clinical and Translational Neurology* 5.5 (2018), pp. 570–582. ISSN: 2328-9503. DOI: 10.1002/acn3.558.
- [217] Tian Xia, Agisilaos Chartsias, and Sotirios A. Tsiftaris. “Consistent Brain Ageing Synthesis”. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Ed. by Dinggang Shen et al. Cham: Springer International Publishing, 2019, pp. 750–758. ISBN: 978-3-030-32251-9.
- [218] Sijie Yan, Yuanjun Xiong, and Dahua Lin. “Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition”. en. In: (), p. 9.
- [219] Xitong Yang et al. “Deep multimodal representation learning from temporal data”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 5447–5455.
- [220] Asako Yoritaka et al. “Parkinson’s disease with and without REM sleep behaviour disorder: are there any clinical differences?” In: *European neurology* 61.3 (2009), pp. 164–170.
- [221] Alexandra L. Young et al. “Multiple Orderings of Events in Disease Progression”. eng. In: *Information Processing in Medical Imaging: Proceedings of the ... Conference* 24 (2015), pp. 711–722. ISSN: 1011-2499. DOI: 10.1007/978-3-319-19992-4_56.
- [222] Manzil Zaheer et al. “Deep Sets”. In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017, pp. 3391–3401.
- [223] Walter J Zetuskys, Joseph Jankovic, and Francis J Pirozzolo. “The heterogeneity of Parkinson’s disease: clinical and prognostic implications”. In: *Neurology* 35.4 (1985), pp. 522–522.
- [224] Hanbin Zhang et al. “PDVocal: Towards Privacy-Preserving Parkinson’s Disease Detection Using Non-Speech Body Sounds”. In: *The 25th Annual International Conference on Mobile Computing and Networking*. MobiCom ’19. Los Cabos, Mexico: Association for Computing Machinery, 2019. ISBN: 9781450361699. DOI: 10.1145/3300061.3300125. URL: <https://doi.org/10.1145/3300061.3300125>.

- [225] Zhifei Zhang, Yang Song, and Hairong Qi. “Age Progression/Regression by Conditional Adversarial Autoencoder”. en. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI: IEEE, July 2017, pp. 4352–4360. ISBN: 978-1-5386-0457-1. DOI: 10.1109/CVPR.2017.463. (Visited on 12/02/2020).
- [226] Qingyu Zhao, Ehsan Adeli, et al. “Variational AutoEncoder For Regression: Application to Brain Aging Analysis”. In: *arXiv:1904.05948 [cs, stat]* (July 2019). arXiv: 1904.05948. URL: <http://arxiv.org/abs/1904.05948> (visited on 12/02/2020).
- [227] Qingyu Zhao, Zixuan Liu, et al. “LSSL: Longitudinal Self-Supervised Learning”. In: *arXiv:2006.06930 [cs, stat]* (June 2020). arXiv: 2006.06930. URL: <http://arxiv.org/abs/2006.06930> (visited on 12/02/2020).
- [228] Jing Zou et al. “Position Emission Tomography/Single-Photon Emission Tomography Neuroimaging for Detection of Premotor Parkinson’s Disease”. eng. In: *CNS neuroscience & therapeutics* 22.3 (Mar. 2016), pp. 167–177. ISSN: 1755-5949. DOI: 10.1111/cns.12493.