

# IMPLEMENTATION D'UN MODÈLE DE PROGRESSION MULTIVARIÉ (Leaspy) POUR L'ÉTUDE DE L'ÉVOLUTION ET L'IDENTIFICATION DE SOUS-GROUPES SUR LA MALADIE DE CADASIL

Sofia Kaisaridi<sup>1</sup>, Hugues Chabriat<sup>2</sup> & Sophie Tezenas du Montcel<sup>3</sup>

<sup>1</sup> Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France, [sofia.kaisaridi@icm-institute.org](mailto:sofia.kaisaridi@icm-institute.org)

<sup>2</sup> Centre Neurovasculaire Translationnel-Centre de Référence CERVCO, FHU NeuroVasc, Hôpital Lariboisière, AP-HP, Université de Paris, INSERM, Unité Mixte de Recherche 1161, Paris, France, [hugues.chabriat@aphp.fr](mailto:hugues.chabriat@aphp.fr)

<sup>3</sup> Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France, [sophie.tezenas@aphp.fr](mailto:sophie.tezenas@aphp.fr)

**Résumé.** Les modèles de progression des maladies sont un outil prometteur pour analyser des données longitudinales présentant de multiples modalités. De tels modèles peuvent être utilisés pour estimer une progression de la maladie à long terme et reconstruire des trajectoires individuelles qui peuvent rendre compte de la variabilité entre les patients, mais aussi entre les modalités. Nous avons implémenté la version multivariée, combinant 14 tests cliniques, d'un modèle bayésien à effets mixtes (Leaspy), qui implique une reparamétrisation temporelle pour analyser l'évolution d'une artériopathie cérébrale autosomique dominante à l'origine d'infarctus sous-corticaux et d'une leucoencéphalopathie (CADASIL). Nous avons analysé les données de 395 patients avec 2007 visites, recrutés au Centre National de Référence français CERVCO, en utilisant Leaspy pour évaluer la variabilité temporelle, résultant de différents rythmes de progression et de décalages temporels, ainsi que la variabilité spatiale, prenant la forme d'une modification dans la séquence des événements. Notre analyse multivariée nous a permis de modéliser la progression de la maladie comme un vecteur à 14 dimensions, et ainsi de rendre compte des relations entre les variables dépendantes, qui ont montré des variations dans le point temporel de début de détérioration. Nous avons ensuite implémenté un modèle de mélange gaussien (GMM) pour les paramètres individuels afin d'identifier des sous-types sous-jacents de la maladie. Nous avons identifié un sous-groupe de patients avec un début précoce et une progression rapide, présentant des symptômes moteurs et des déficits neurologiques plus précoces, et un groupe avec un début tardif et une progression lente, montrant des symptômes cognitifs plus précoces. Pour conclure le modèle multivarié parvient à mettre en évidence les différences dans l'évolution des scores cliniques, suggérant un déclin progressif mais également hétérogène dans CADASIL et le profilage a révélé deux sous-types impliquant à la fois la variabilité temporelle et spatiale.

**Mots-clés.** Modèle de progression des maladies, Multivariée, Bayésienne, Effets mixtes, MCMC-SAEM, Classification

**Abstract.** Disease progression models are a promising tool to analyze longitudinal data presenting multiple modalities. Such models can be used to estimate a long-term disease

progression and to reconstruct individual trajectories that can account for the variability between patients but also between modalities. We implemented the multivariate version, combining 14 clinical tests, of a bayesian mixed-effect model (**Leaspy**) that involves a time reparametrisation to analyze the evolution of cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL). We analyzed data from 395 patients with 2007 visits, recruited at the French National Referral Centre CERVCO, using **Leaspy** to assess the temporal variability, resulting from different pace of progression and temporal offset, along with the spatial variability, that takes the form of a modification in the sequence of events. Our multivariate analysis allowed us to model the disease progression as a 14-dimensional vector, and thus account for the relationships between the dependent variables, that showed variations in the starting deterioration timepoint. We then implemented a GMM clustering algorithm for the individual parameters to identify underlying subtypes for the disease. We identified a subgroup of patients with early onset and fast progression, showing motor symptoms and neurological deficits earlier, and a group with late onset and slow progression, showing cognitive symptoms earlier. To conclude the multivariate model manages to highlight the differences in the evolution of clinical scores suggesting a gradual but also heterogenous decline in CADASIL and the profiling revealed two subtypes that imply both the temporal and the spatial variability.

**Keywords.** Disease Progression Model, Multivariate, Bayesian, Mixed-effects, MCMC-SAEM, Classification

## 1 Introduction

Les modèles à effets mixtes sont prometteurs pour analyser des données répétées représentant plusieurs modalités et sont souvent utilisés dans les études épidémiologiques. De tels modèles peuvent être utilisés pour estimer une progression à long terme d'une maladie et reconstruire des trajectoires individuelles qui tiennent compte de la variabilité entre les patients (effets aléatoires), mais aussi entre les modalités (effets fixes). Comprendre comment la maladie progresse et quelle est la variabilité attendue entre les individus est essentiel. L'hétérogénéité de la maladie est souvent influencée par certaines covariables, mais parfois elle est plus complexe comme elle est structurée autour du sous-type auquel l'individu appartient. Si les modèles classiques font l'hypothèse d'une évolution homogène des patients, ceci n'est pas forcément réaliste car il existe souvent des sous-groupes des patients évoluant de façon différente. La découverte de clusters inconnus d'individus est un défi complexe qui s'ajoute à l'estimation des paramètres du modèle. Dans ce cadre les modèles de progression de maladie (*Disease Progression Models*) sont un outil émergent qui reconstruit les chronologies des maladies chroniques à long terme, fournissant ainsi un aperçu unique des processus pathologiques et de leurs mécanismes sous-jacents (Young et al. (2024)). Ils sont interprétables, facilitant ainsi la compréhension des maladies tout en permettant la classification, la prédiction et la stratification.

Dans cette étude nous avons utilisé un modèle de progression des maladies, non-linéaire

à effets mixtes qui passe par une reparamétrisation du temps (**Leaspy**) comme proposé par Schirradi et al. (2017). Nous avons implémenté une version logistique multivariée du modèle sur les données de la maladie CADASIL en évaluant l'intensité de 14 scores cliniques dans le but de construire une cartographie du cours de la maladie. En utilisant les paramètres individuelles, nous avons identifié deux sous-groupes présentant des trajectoires de maladie différentes.

## 2 Modèle de progression (**Leaspy**)

Nous considérons des données longitudinales, c'est-à-dire des patients avec des mesures répétées dans le temps correspondant à 14 scores cliniques mesurant différents aspects de la maladie. Le modèle de progression utilisé (**Leaspy**), a déjà fait ses preuves pour décrire la progression des maladies de façon multivariée, et a déjà été implémenté pour diverses maladies neurodégénératives (Ortholand et al. (2023), Koval et al. (2021,2022)). Nous supposons un ensemble de données longitudinales  $(y_{ijk})_{1 \leq i \leq n, 1 \leq j \leq N_i, 1 \leq k \leq d}$  qui décrivent les valeurs pour chaque patient  $i$  dans la visite  $j$  de la caractéristique  $k$ . Le nombre total de visites  $N_i$  peut être différent pour chaque patient  $i$ . Les données  $y_{ijk}$  sont supposées être des points sur une variété riemannienne  $\mathcal{M}$ . Les paramètres de population (effets fixes) décrivent la trajectoire moyenne comme une géodésique  $\gamma_0$  sur la  $\mathcal{M}$  avec  $\gamma_0(t_0) = \mathbf{p}$  et  $\dot{\gamma}_0(t_0) = \mathbf{v}$ . Ici nous utilisons la forme logistique du modèle :

$$y_{ijk} = \left(1 + \left(\frac{1}{p_k} - 1\right) \exp\left(-\frac{v_k(e^{\xi_i}(t_{ij} - t_0 - \tau_i) + t_0) + w_{ik}}{p_k(1 - p_k)}\right)\right)^{-1} + \epsilon_{ijk} \quad (1)$$

avec une erreur résiduelle  $\epsilon_{ij} \sim \mathcal{N}(\mathbf{0}_d, \sigma^2 \mathbf{I}_d)$ . Pour simplifier, notre modèle construit la trajectoire moyenne de chaque score sous la forme d'une courbe logistique décrite par les paramètres  $p_0$ ,  $v_0$  et  $t_0$ , où  $p_0$  et  $v_0$  sont la position et la vitesse (dérivée de la courbe) au temps  $t_0$ , le point médian de la logistique. La reparamétrisation du temps individuel prend la forme suivante :  $\psi_i(t) = \alpha_i(t - t_0 - \tau_i) + t_0$  où  $\tau_i$  est le décalage temporel et  $\alpha_i = e^{\xi_i}$  est le facteur d'accélération. Le profil temporel des patients est décrit par deux paramètres : le décalage temporel  $\tau_i$  qui correspond à l'âge estimé du début de la maladie, mesuré en années, et le taux de progression  $\xi_i$  qui indique si la progression globale est accélérée (valeurs positives) ou ralentie (valeurs négatives). Le profil spatial est défini par les paramètres d'espacement intermarqueurs  $(\omega_i)_i$  qui tiennent compte de l'ordre différent des événements au sein de notre population. Pour chaque patient, nous avons un  $\omega$  pour chaque score clinique inclus dans le modèle avec des valeurs négatives indiquant qu'un score spécifique commence à se détériorer plus tôt pour un patient spécifique que pour la population moyenne. Néanmoins pour une interprétabilité accrue, le modèle utilise une analyse en composantes indépendantes avec  $N_s$  sources indépendantes, ce qui conduit à des paramètres de décalage spatial  $\mathbf{w}_i = A s_i$  tels que les colonnes  $A_l = \sum_{k=1}^{d-1} \beta_{lk} B_k$  sont une combinaison linéaire d'une base orthonormale  $(B_k)_{1 \leq k \leq d-1}$  du sous-espace orthogonal à  $Span(\mathbf{v})$ .

Le modèle statistique hiérarchique suppose que les paramètres de population  $z_{pop} =$

$(\mathbf{g}, \mathbf{v}, t_0, A)$  et les paramètres individuels  $z_i = (\xi_i, \tau_i, \mathbf{w}_i)$  sont latents et suivent des distributions gaussiennes :

$$\left\{ \begin{array}{l} \xi_i \sim \mathcal{N}(0, \sigma_\xi^2) \\ \tau_i \sim \mathcal{N}(0, \sigma_\tau^2) \\ s_i \sim \mathcal{N}(\mathbf{0}_{N_s}, \mathbf{I}_{N_s}) \end{array} \right. \quad (2) \quad \left\{ \begin{array}{l} g_k = \frac{1}{p_k} - 1 \quad \text{et} \quad \mathbf{g} \sim \mathcal{N}(\bar{\mathbf{g}}, \sigma_g^2 \mathbf{I}_d) \\ v_k = e^{\tilde{v}_k} \quad \text{et} \quad \tilde{\mathbf{v}} \sim \mathcal{N}(\bar{\mathbf{v}}, \sigma_v^2 \mathbf{I}_d) \\ t_0 \sim \mathcal{N}(\bar{t}_0, \sigma_t^2) \\ \beta_{lk} \sim \mathcal{N}(\tilde{\beta}_{lk}, \sigma_\beta^2) \end{array} \right. \quad (3)$$

Enfin les paramètres du modèle statistique sont  $\theta = (\sigma_\xi, \sigma_\tau, \bar{\mathbf{g}}, \bar{\mathbf{v}}, \bar{t}_0, \tilde{\beta}_{lk})$  tandis que  $\sigma_g, \sigma_t, \sigma_v, \sigma_\beta$  sont fixes.

## 3 Application sur des données réelles

### 3.1 CADASIL

L'artériopathie cérébrale autosomique dominante à l'origine d'infarctus sous-corticaux et d'une leucoencéphalopathie (CADASIL) est la maladie héréditaire la plus courante touchant les petits vaisseaux cérébraux (SVD) et elle est causée par des mutations du gène NOTCH3 (Rutten et al. (2020)). Les patients atteints de CADASIL présentent un large éventail de symptômes telles que des crises de migraines avec aura, des accidents vasculaires cérébraux, des troubles du comportement, une incapacité motrice et des troubles cognitifs de la dysfonction exécutive jusqu'à une démence sévère (Chabriat et al. (2009)). Les premiers signes de déclin cognitif sont observés principalement au niveau des fonctions exécutives et de la mémoire à court terme et des altérations significatives apparaissent progressivement dans tous les domaines cognitifs (Buffon et al. (2006), Amberla et al. (2004), Peters et al (2005), Brookes et al. (2016)). CADASIL se manifeste cliniquement entre 40 et 70 ans et conduit à une démence vasculaire sous-corticale pour la grande majorité des patients (Viitanen et Kalimo (2000)). Les manifestations cliniques présentées tout au long de la maladie varient en intensité au fil du temps et chez différents groupes de patients (Brice et al. (2020,2022)). Malheureusement, malgré ces informations sur l'histoire naturelle du CADASIL la cinétique exacte du déclin tout au long de la maladie reste encore très mal connues. Dans notre étude, nous avons implémenté un modèle Leaspy multivarié en prenant en compte 14 scores cliniques et d'évaluations afin de capturer l'ensemble du spectre du déclin causé par la maladie.

### 3.2 Données

Nous avons analysé les données de 395 patients effectuant au total 2007 visites avec une médiane de 4 visites par patient (IQR : 3-7), recrutés au Centre National de Référence CERVCO en France. La durée de suivi a varié de 6 mois à 19 ans, avec une moyenne de 7.5 ans (écart-type : 4.7 ans), et l'âge d'inclusion de 25 à 80 ans, (moyenne de 52,2 ans et écart-type : 11.7). Ces patients avaient complété au moins deux visites, le temps de visite n'était pas manquant et ils présentaient au moins l'un des 14 scores déterminés par consensus entre

les neurologues (N.A., D.H. et H.C.) et les neuropsychologues (A.J., S.R., C.M. et C.R.) du centre de référence, obtenus lors du suivi.

Les scores sélectionnés peuvent nous fournir une description du déclin global qui est observé pendant toute la durée de la maladie. L'efficacité cognitive a été évaluée à l'aide de l'échelle de Mattis (MDRS : Mattis Dementia Rating Scale) et de la sous-échelle VADAS-Cog (VADAS-Cog : Vascular Dementia Assessment Scale cognitive subscale). Le score d'Initiation/Persévérance (*MDRSinitiation*) du MDRS a été pris en compte. Trois des composants de la VADAS-Cog ont été inclus dans notre analyse : le Digit Cancellation Test (*DigitCancel*), le Symbol Digit Test (*SymbolDigit*) et le Backward Digit Span (*BackwardDigit*). Nous avons également pris en compte les 3 scores obtenus du Trail Making Test (TMT) évaluant la vitesse cognitive et la flexibilité mentale, le temps pour la partie A du TMT (*TMTAT*), le temps et les erreurs pour la partie B du TMT (*TMTBT*, *TMTBE*). Les performances de la mémoire ont été analysées à l'aide de 3 scores de l'épreuve de rappel libre/ rappel indicé du Grober et Buschke (GB) : le rappel libre total (*GBfree*), l'indice de sensibilité au repérage (*GBcueing*) et le rappel total indicé (*GBdelayed*). La sévérité du handicap a été évaluée à l'aide de l'échelle de Rankin modifiée (*Rankin*) et l'indépendance dans les activités quotidiennes par l'indice de Barthel (*Barthel*). Les déficits neurologiques focaux ont été aussi inclus dans notre analyse, représentés par le score NIH pour les AVC (*NIHSS*). Enfin, la qualité de vie a été évaluée à l'aide de l'échelle visuelle analogique de l'EuroQol à 3 niveaux (*EQVAS*). Ces 14 scores ont été transformés et normalisés pour obtenir une évolution ascendante de 0 à 1, répondant aux exigences de notre modèle.

### 3.3 Résultats

#### 3.3.1 Évolution du CADASIL

Basé sur le modèle *Leaspy* le plus performant pour l'évolution du CADASIL, nous observons qu'il existe trois groupes de scores qui se modifient à des stades différents de la maladie (Figure 1). Le premier groupe de scores, dont l'évolution commence à un stade précoce de la maladie, comprend des mesures de dysfonctionnement exécutif comme les sous-scores VADAS-Cog, Symbol Digit Test (*SymbolDigit*), Backward Digit Span (*BackwardDigit*) et Cancellation Task (*DigitCancel*), TMT B temps (*TMTBT*) et aussi certains domaines de performance de la mémoire comme le GB Total Free Recall (*GBfree*). Aux stades intermédiaires de la maladie, nous pouvons voir que la qualité de vie EQ VAS (*EQVAS*), la sévérité du handicap mesuré par l'échelle de Rankin modifiée (*Rankin*), et le temps du TMT A (*TMTAT*) commencent également à montrer une évolution. Enfin, à un stade plus avancé de la maladie, les caractéristiques qui commencent à se détériorer correspondent aux divers domaines comme l'efficacité cognitive représentée par l'Initiation du Mattis DRS (*MDRSinitiation*) et les erreurs du TMT B (*TMTBE*), la performance de mémoire mesuré par le GB Index of Sensitivity to Cueing (*GBcueing*) et le GB Delayed Total Recall (*GBdelayed*) et aussi les activités quotidiennes (*Barthel*) et les AVC (*NIHSS*).

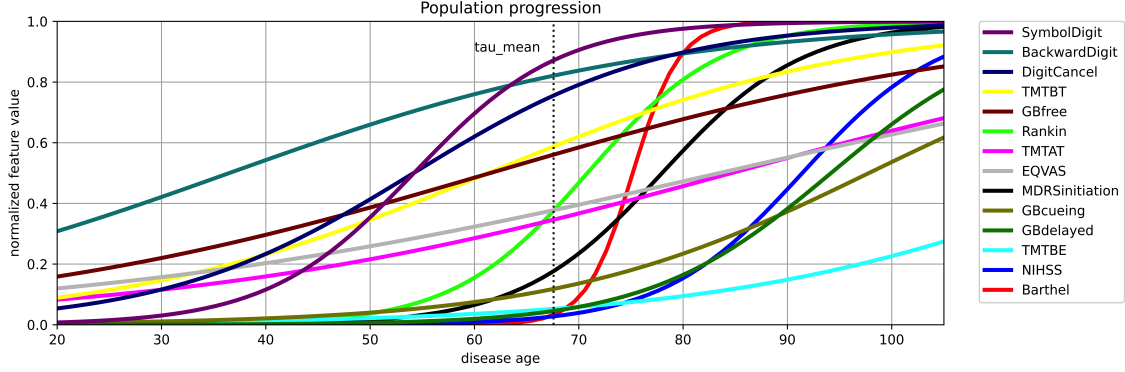


FIGURE 1 – Évolution du CADASIL décrite par la progression longitudinale moyenne de 14 scores cliniques.

### 3.3.2 Sous-groupes de patients avec des trajectoires similaires

Pour identifier les sous-groupes de patients avec des trajectoires similaires, nous avons appliqué un modèle de mélange gaussien (GMM) (Zhou (2021)). À cette fin, nous avons pris en compte les paramètres individuels  $(\xi_i, \tau_i, s_i)$  et nous avons fixé le nombre de clusters à 2 qui optimisent le critère d'information d'Akaike et le critère d'information Bayésien (AIC et BIC).

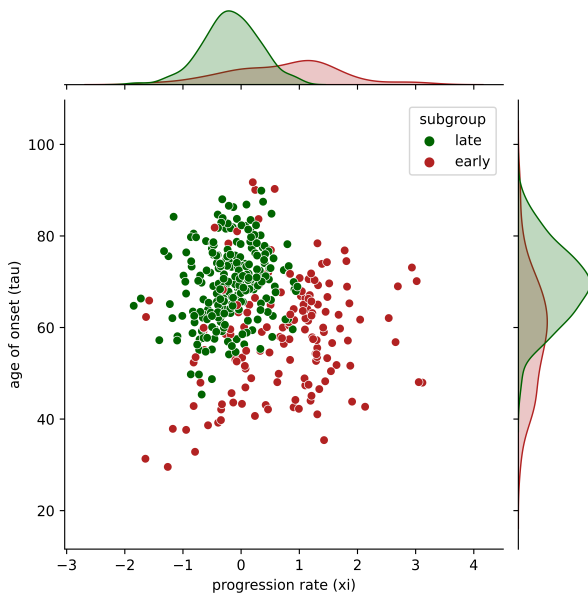


FIGURE 2 – Répartition en sous-groupes en fonction des paramètres temporels individuels ( $\tau$  et  $\xi$ ).

Leur répartition est représentée dans la figure 2 selon les paramètres temporels  $\tau$  et  $\xi$ . Sur les axes supérieur et droit, on peut également voir les densités spécifiques par sous-groupe des paramètres individuels. Le premier sous-groupe, précoce (ou rapide), comprend 37% des patients avec un tau moyen de 58,27 (écart type 12,20) et un xi moyen de 0,73 (0,95). Les 63% restants des patients appartiennent au deuxième sous-groupe, tardif (et lent), et présentent un tau moyen de 69,33 (8,20) et un xi de -0,18 (0,49). Une fois que les clusters ont été formés, afin de décrire le profil des patients attribués à chaque sous-groupe, nous avons réalisé un test U de Mann-Whitney pour identifier les éventuelles différences dans les distributions des paramètres  $\tau_i$ ,  $\xi_i$  et  $\omega_i$  entre les sous-groupes. Après avoir examiné les distributions des  $\omega$ , nous constatons que le sous-groupe précoce (ou rapide) présente des symptômes d'incapacité motrice plus tôt

que l'autre sous-groupe tardif (et lent), car il a des omegas plus petits pour les indices de Barthel et de Rankin. Dans ce sous-groupe, nous observons également un début plus précoce des accidents vasculaires cérébraux, comme le montrent les distributions de l'indice NIHSS. D'autre part, les patients du groupe tardif (et lent) présentent un début plus précoce des symptômes cognitifs, comme le montrent les différences dans les distributions de l'Initiation du MDRS, des scores TMT, du GB Total Free Recall et des sous-scores VADAS-Cog. Les distributions de la qualité de vie EQ VAS et de GB Delayed Total Recall et l'Index of Sensitivity to Cueing sont similaires pour les deux sous-groupes. Toutes les distributions  $\omega$  selon les sous-groupes sont montrées dans la figure 3.

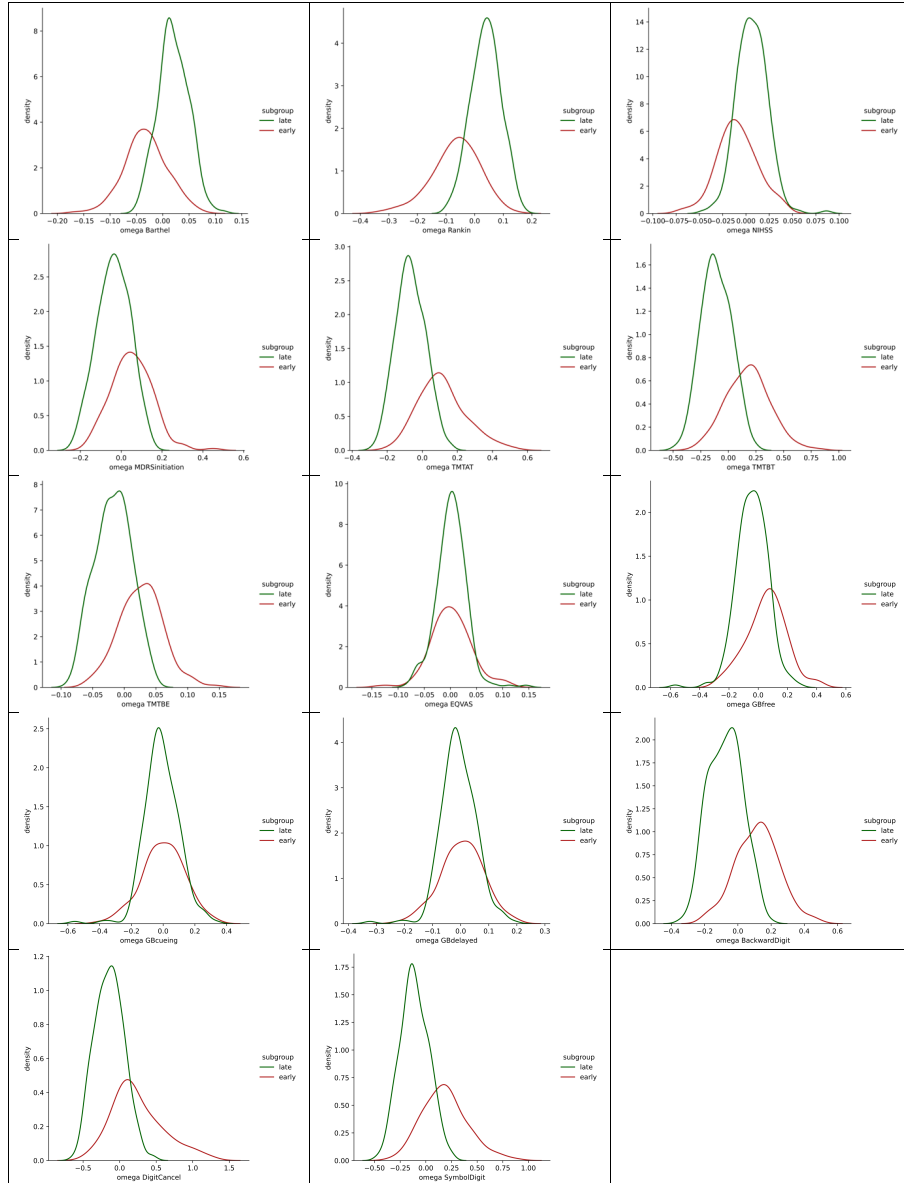


FIGURE 3 – Distributions spécifiques au sous-groupe des paramètres spatiaux individuels ( $\omega$ ).

## 4 Discussion

Cette étude permet de mieux comprendre l’histoire naturelle de la maladie et d’explorer individuellement les patients pour en déterminer des groupes d’évolution différente. Bien que la variété en termes d’âges, de moments de début et de progression puisse poser des difficultés pour construire un modèle fiable, la reparamétrisation de l’âge nous permet de construire une trajectoire de population cohérente avec le cours de la maladie et de comparer les individus se trouvant au même stade de la maladie plutôt que d’avoir le même âge biologique. Un atout particulièrement fort est que nous sommes en mesure d’examiner le début de la maladie en tant que paramètre individuel  $\tau$  estimé par notre modèle, alors que ce début est difficilement défini cliniquement pour les patients atteints du CADASIL. Dans la plupart des études, le début était généralement considéré comme le moment où le premier accident vasculaire cérébral se produit (Amberla et al. (2004), Peters et al. (2005), Chabriat et al. (2009)). La migraine avec aura est également mentionnée comme un symptôme inaugural majeur du CADASIL et peut survenir des années avant le premier accident vasculaire cérébral (Guey et al. (2016)).

Le résultat le plus prometteur de notre étude est qu’elle a révélé deux phénotypes de CADASIL, plus complexes que ce qui était initialement soupçonné, avec des trajectoires qui diffèrent non seulement en termes d’évolution temporelle, mais aussi en ce qui concerne la séquence d’événements. Le profilage effectué en considérant simultanément tous les aspects de l’évolution spatiotemporelle individuelle nous a permis de conclure que les patients présentant une forme plus agressive de la maladie développent également des symptômes précoces d’incapacité motrice et d’accidents vasculaires cérébraux, tandis que chez les patients présentant une forme légère, les symptômes cognitifs apparaissent en premier.

Il y a également plusieurs limites dans cette étude qui laissent place à des investigations supplémentaires. Les clusters individuels ont été formés a posteriori en utilisant les paramètres individuels estimés par notre modèle. Une procédure de classification intégrée à l’étape d’estimation du modèle, en supposant un modèle de mélange peut nous fournir une description plus précise et non-biaisé de la trajectoire de la maladie ainsi que des différents phénotypes sous-jacents.

## 5 Conclusion

Avec notre modèle de progression multivarié, nous avons pu obtenir des résultats fiables et prometteurs sur la cartographie de l’évolution du CADASIL et sur ses variations au sein de différents groupes de patients. Nous considérons cela comme une première étape, car le modèle, dans son état actuel, offre encore de nombreuses possibilités pour approfondir la compréhension de la cinétique de la maladie. Les sous-groupes suggérés ici illustrent que les phénotypes sous-jacents de CADASIL sont potentiellement plus complexes et implique le profil temporel et aussi spatial de patients. Cette perception nous offre des perspectives d’amélioration afin de profiler de manière plus efficace les différents sous-groupes.

## Bibliographie

- Amberla, K., Wäljas, M., Tuominen, S., Almkvist, O., Pöyhönen, M., Tuisku, S., Kalimo, H., et Viitanen, M. (2004), Insidious cognitive decline in CADASIL. *Stroke*, 35(7), pp. 1598-1602.
- Brice, S., Jabouley, A., Reyes, S., Machado, C., Rogan, C., Dias-Gastellier, N., Chabriat, H., et du Montcel, S. T. (2020), Modeling the Cognitive Trajectory in CADASIL. *Journal of Alzheimer's disease : JAD*, 77(1), pp. 291-300.
- Brice, S., Reyes, S., Jabouley, A., Machado, C., Rogan, C., Gastellier, N., Alili, N., Guey, S., Jouvent, E., Hervé, D., Tezenas du Montcel, S., et Chabriat, H. (2022), Trajectory Pattern of Cognitive Decline in Cerebral Autosomal Dominant Arteriopathy With Subcortical Infarcts and Leukoencephalopathy. *Neurology*, 99(10), pp. e1019-e1031.
- Brookes, R. L., Hollocks, M. J., Tan, R. Y., Morris, R. G., et Markus, H. S. (2016), Brief Screening of Vascular Cognitive Impairment in Patients With Cerebral Autosomal-Dominant Arteriopathy With Subcortical Infarcts and Leukoencephalopathy Without Dementia. *Stroke*, 47(10), pp. 2482-2487.
- Buffon, F., Porcher, R., Hernandez, K., Kurtz, A., Pointeau, S., Vahedi, K., Bousser, M. G., et Chabriat, H. (2006), Cognitive profile in CADASIL. *Journal of neurology, neurosurgery, and psychiatry*, 77(2), pp. 175-180.
- Chabriat, H., Joutel, A., Dichgans, M., Tournier-Lasserre, E. et Bousser, M.G. (2009), Cadasil. *The Lancet. Neurology*, 8(7), pp. 643-653.
- Guey, S., Mawet, J., Hervé, D., Duering, M., Godin, O., Jouvent, E., Opherk, C., Alili, N., Dichgans, M., et Chabriat, H. (2016), Prevalence and characteristics of migraine in CADASIL. *Cephalalgia : an international journal of headache*, 36(11), pp. 1038-1047.
- Koval I., Bône A., Louis M., Lartigue T., Bottani S., Marcoux A., Samper-González J., Burgos N., Charlier B., Bertrand A., Epelbaum S., Colliot O., Allasonnière S. et Durrleman S. (2021), AD Course Map charts. *Scientific Reports - Nature*, 11(1), 8020
- Koval, I., Dighiero-Brecht, T., Tobin, A. J., Tabrizi, S. J., Scahill, R. I., Tezenas du Montcel, S., Durrleman, S., et Durr, A. (2022), Forecasting individual progression trajectories in Huntington disease enables more powered clinical trials. *Scientific reports - Nature*, 12(1), 18928.
- Ortholand J., Pradat P.F., Tezenas du Montcel S., et Durrleman S. (2023) Interaction of sex and onset site on the disease trajectory of amyotrophic lateral sclerosis. *Journal of Neurology*, 270(12), pp. 5903-5912.
- Peters, N., Opherk, C., Danek, A., Ballard, C., Herzog, J., et Dichgans, M. (2005), The pattern of cognitive performance in CADASIL : a monogenic condition leading to subcortical ischemic vascular dementia. *The American journal of psychiatry*, 162(11), pp. 2078-2085.
- Rutten, J. W., Hack, R. J., Duering, M., Gravesteyn, G., Dauwerse, J. G., Overzier, M., van den Akker, E. B., Slagboom, E., Holstege, H., Nho, K., Saykin, A., Dichgans, M., Malik, R., et Lesnik Oberstein, S. A. J. (2020), Broad phenotype of cysteine-altering NOTCH3 variants in UK Biobank : CADASIL to nonpenetrance. *Neurology*, 95(13), pp. e1835-e1843.

Schirratti J.B., Allasonnière S., Colliot O., et Durrleman S. (2017), A Bayesian Mixed- Effects Model to Learn Trajectories of Changes from Repeated Manifold-Valued Observations. *Journal of Machine Learning Research*, 18, pp. 1-33.

Viitanen, M., et Kalimo, H. (2000), CADASIL : hereditary arteriopathy leading to multiple brain infarcts and dementia. *Annals of the New York Academy of Sciences*, 903, pp. 273-284.

Young, A. L., Oxtoby, N. P., Garbarino, S., Fox, N. C., Barkhof, F., Schott, J. M., et Alexander, D. C. (2024), Data-driven modelling of neurodegenerative disease progression : thinking outside the black box. *Nature reviews. Neuroscience*, 25(2), pp. 111-130.

Zhou Z.H. (2021), *Machine Learning*. Springer, Singapore. pp. 222-227