



Non-parametric ODE-Based Disease Progression Model of Brain Biomarkers in Alzheimer's Disease

Matías Bossa¹ , Abel Díaz Berenguer¹ , and Hichem Sahli^{1,2} 

¹ Department of Electronics and Informatics (ETRO), Vrije Universiteit Brussel (VUB), 1050 Brussels, Belgium

{mnbossa, aberenguer, hsahli}@etrovub.be

² Interuniversity Microelectronics Centre (IMEC), 3001 Leuven, Belgium

Abstract. Data-driven disease progression models of Alzheimer's disease are important for clinical prediction model development, disease mechanism understanding and clinical trial design. Among them, dynamical models are particularly appealing because they are intrinsically interpretable. Most dynamical models proposed so far are consistent with a linear chain of events, inspired by the amyloid cascade hypothesis. However, it is now widely acknowledged that disease progression is not fully compatible with this conceptual model, at least in sporadic Alzheimer's disease, and more flexibility is needed to model the full spectrum of the disease. We propose a Bayesian model of the joint evolution of brain image-derived biomarkers based on explicitly modelling biomarkers' velocities as a function of their current value and other subject characteristics. The model includes a system of ordinary differential equations to describe the biomarkers' dynamics and sets a Gaussian process prior to the velocity field. We illustrate the model on amyloid PET SUVR and MRI-derived volumetric features from the ADNI study.

Keywords: Disease progression model · Alzheimer's disease (AD) · Magnetic resonance imaging (MRI) · Amyloid PET · Ordinary differential equations (ODE) · Gaussian process (GP)

1 Introduction

Alzheimer's disease (AD) is a growing health-economic worldwide issue, accounting for most cases of dementia [13]. Despite the great amount of effort devoted to AD prevention and drug development during the last three decades, the few pharmacological treatments available show a modest benefit. The study of the AD process is further hindered by the fact that dementia can be caused by multiple pathologies, and that AD often co-occurs with them [11], being age and genetic variations the main risk factors [10].

For more than two decades, the most widely accepted model of the pathophysiological process underlying AD was the so-called amyloid cascade hypothesis. This hypothesis states that the process starts with an abnormal accumulation of the β -amyloid ($A\beta$) peptide, triggering a chain of pathological events

in a predictable way. The corresponding model of biomarker dynamics states that the main AD biomarkers become abnormal in a temporally ordered manner [6, 7]. However, large cohort studies showed that all possible combinations of biomarker abnormalities are frequently present in the cognitively normal population [8], evidencing that the amyloid cascade hypothesis is insufficient to explain the observed heterogeneity in sporadic AD [4, 5].

A new conceptual model of AD was recently proposed [4], which posited a non-deterministic disease path. According to this model, $A\beta$ and tau levels interact between them and with genetic and environmental factors to increase or reduce the risk of disease progression. These interactions would be responsible for the huge heterogeneity observed in biomarker trajectories and the discrepancies between observation and the amyloid cascade hypotheses.

Quantitative tools that estimate the biomarker dynamics are needed to shed light on the AD process and to build better clinical tools for diagnosis, prognosis and therapy efficacy assessment.

1.1 Disease Progression Models of Alzheimer’s Disease

The first AD progression models describing long-term trajectories from short-term biomarker observations were based on Jack’s model [7], *i.e.*, they assumed that all subjects follow the same disease progression pattern but with different onset times and at different speeds. Jedynak defined a disease progression score aimed at quantifying disease progression [9]. Subjects were temporally ordered according to this score and a parametric sigmoid-shaped curve was used to fit the progression of biomarkers. In [3] the authors proposed a semi-parametric model to determine the population mean of biomarker trajectories and the temporal order of subjects. A similar but more flexible model used Gaussian Process (GP) to model also the individual departures from the mean [12]. In general, all these models may suffer from identifiability issues when trained with short-term observations, because of the need to simultaneously estimate the disease onset times and the biomarker trajectories. Sometimes identifiability issues were mitigated using mixed-effect modelling to restrict the variance of the subject-level parameters.

The first dynamical model that relaxes the *unique trajectory* condition, allowing an arbitrary combination of variables as initial conditions, used a Riemannian framework to transport the mean trajectory to fit the subject’s observations [16]. Contrary to the previous works, it is the initial value of the variables, and not the onset time, which was modelled as a random effect.

Finally, differential equation models parameterize biomarker velocities instead of biomarker trajectories and are therefore *implicit* models. Two works [2, 15] tackle the problem of how to estimate long-term biomarker trajectories from short-term observations of a single biomarker. A recent work used a system of ordinary differential equations (ODEs) to simulate the effect of amyloid treatments on the disease course [1], being the first multivariate ODE-based AD progression model.

In this work, we propose a probabilistic AD progression model that uses a system of ODEs to describe biomarker dynamics. In our formulation, and similar to [16] and [1], all combinations of trajectory starting values are allowed. Another important common feature is that onset times are not model parameters, reducing the risk of non-identifiability. But contrary to all ODE-based approaches, we model the biomarker velocities non-parametrically, using GPs, which adds flexibility and imposes less inductive bias.

2 Methods

2.1 Definitions and Model Overview

We propose a Bayesian generative model to describe the trajectories of brain biomarkers throughout AD. Let $\mathbf{x}_s(t) = [x_{1,s}(t), x_{2,s}(t), \dots, x_{L,s}(t)]$ be a set of L brain features and $\mathbf{y}_s(t) = [y_{1,s}(t), y_{2,s}(t), \dots, y_{Q,s}(t)]$ a set of Q covariates for subject s at time t . The features $x_{l,s}(t)$ represent magnitudes associated with the disease status that evolve as the disease progress. For example, they could be brain atrophy, amyloid plaques or neurofibrillary tangles. In general, we can distinguish the aforementioned brain features, associated to the disease process, from the brain biomarkers extracted from MRI or PET images. However, we will consider one observable \hat{x}_l per feature x_l and will refer to features and biomarkers interchangeably. The covariates $y_{q,s}(t)$ are assumed to have no observation error. They could be fixed over time (*e.g.* genetics), change in time according to a predefined or known pattern (*e.g.* age), or be controlled externally (*e.g.* treatments).

The link between a set of observed biomarkers $\hat{\mathbf{x}}$ and the feature vector \mathbf{x} is specified by a likelihood function $\mathcal{L}(\hat{\mathbf{x}}|\mathbf{x}, \Theta)$, where Θ are model parameters. In our case, the likelihood functions will be independent Gaussian distributions. Let $\hat{x}_{l,s;i} \sim \mathcal{N}(x_{l,s}(t_{l,s;i}), \sigma_l^2)$ be the i th observation of biomarker l for subject s at observation time $t_{l,s;i}$. Note that the number observations and observation times may be different for each subject and each biomarker.

The main hypothesis in this work is that the state of features and covariates at a given time determines unequivocally the rate of progression, *i.e.* the expected rate of change of all the brain features. Specifically, trajectories should be a solution of the of the initial value problem defined by the system of ODEs

$$\frac{d\mathbf{x}_s(t)}{dt} = \mathbf{v}(\mathbf{z}_s(t)), \quad (1)$$

where $\mathbf{z}_s(t) = [\mathbf{x}_s(t), \mathbf{y}_s(t)]$, with initial condition

$$\mathbf{z}_s(0) = [\mathbf{x}_{0,s}, \mathbf{y}_s(0)]. \quad (2)$$

We propose to model each component l of the velocity field $\mathbf{v}(\cdot)$ using a GP prior

$$v_l \sim \mathcal{GP}(0, k_l(\cdot, \cdot)), \quad (3)$$

with the exponential kernel, $k_l(\mathbf{z}_m, \mathbf{z}_n) = k_{\alpha_l, \rho_l}(\mathbf{z}_m, \mathbf{z}_n) = \alpha_l^2 \exp(-\frac{(\mathbf{z}_m - \mathbf{z}_n)^2}{2\rho_l^2})$.

The problem is completely specified once we define priors for the model hyperparameters, *i.e.*, the observation variance σ_l^2 , the kernel parameters α_l and ρ_l , and the subject-level parameters $\mathbf{x}_{0,s}$.

However, the velocity field $\mathbf{v}(\cdot)$ is a function-valued parameter which could be difficult to estimate and very hard or impossible to marginalize out given that it is involved in the ODE system (1). We propose the following approximation to transform Eq. (3) into a likelihood function and to model $\mathbf{v}(\cdot)$ implicitly, as it is usual in GP regression. Let $\hat{x}_{l,s;i}$ and $\hat{x}_{l,s;i+1}$ be two observations of the same biomarker l at two consecutive time points, $t_{l,s;i}$ and $t_{l,s;i+1}$, respectively. Assuming that the time difference $\Delta t_{l,s;i} = t_{l,s;i+1} - t_{l,s;i}$ is small with respect to the biomarker dynamics, we can approximate the velocity field using the observation differences. Dropping the indexes s and l for clarity, we have

$$\begin{aligned} x(t_{i+1}) &\simeq x(t_i) + v(\mathbf{z}(t_i))\Delta t_i \\ \hat{x}_{i+1} - \epsilon_{i+1} &\simeq \hat{x}_i - \epsilon_i + v(\mathbf{z}(t_i))\Delta t_i \\ \hat{v}_i &= \frac{\hat{x}_{i+1} - \hat{x}_i}{\Delta t_i} \simeq v(\mathbf{z}(t_i)) + \frac{\epsilon_{i+1} - \epsilon_i}{\Delta t_i}, \end{aligned} \quad (4)$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ is the observation error. Then $\hat{v}_i \sim \mathcal{N}(v(\mathbf{z}(t_i)), 2\sigma^2/\Delta t_i^2)$ and we can replace Eq. (3) with

$$\hat{v} \sim \mathcal{GP}(0, \hat{k}(\cdot, \cdot)), \quad (5)$$

where the new kernel $\hat{k}(\cdot, \cdot)$ is the same as $k(\cdot, \cdot)$ plus a noise term, and Eq. (5) represents a likelihood function because its l.h.s. is an observation. Note that the Gaussianity and independence of biomarker observations was critical to define the approximate velocities in Eq. (4).

2.2 Proposed Model

The complete set of parameter priors is given by

$$\begin{aligned} \sigma_l &\sim \mathcal{N}^+(0, \tau_{\sigma,l}) \\ \alpha_l &\sim \mathcal{N}^+(0, \tau_{\alpha,l}) \\ \rho_l &\sim \Gamma(5, 5) \\ [x_{0,s}]_l &\sim \mathcal{U}(0, 1), \end{aligned}$$

where $\mathcal{N}^+(\cdot, \cdot)$ is the half-Gaussian distribution and $\Gamma(5, 5)$ is used as a weakly informative prior that penalizes extremely large and extremely small values of the length scale parameter ρ . Weakly informative priors for σ_l and α_l are determined by setting $\tau_{\sigma,l}$ and $\tau_{\alpha,l}$ equal to the mean subject-level standard deviation of observations and the variance of the estimated velocities $\hat{v}_{l,s;i}$, respectively.

For simplicity, each component of the subject-level parameters $\mathbf{x}_{0,s}$ is restricted to be in the unit segment after normalizing the biomarker values to fit in the unit hypercube. Note that they can be modelled as random effects if desired, which would provide better estimations in case that a biomarker is completely missing for a given subject.

The likelihood functions are

$$\begin{aligned}\hat{x}_{l,s;i} &\sim \mathcal{N}([\mathbf{x}_s(t_{l,s;i})]_l, \sigma_l^2) \\ \hat{\mathbf{v}}_l &\sim \mathcal{N}(\mathbf{0}, K_{\alpha_l, \rho_l}(\hat{Z}_l, \hat{Z}_l) + \text{diag}(\mathbf{s}_{\sigma_l}^2)),\end{aligned}\quad (6)$$

where $\hat{\mathbf{v}}_l$ includes all consecutive observation differences from all subjects, *i.e.*,

$$\hat{\mathbf{v}}_l = \left[\frac{\hat{x}_{l,1;2} - \hat{x}_{l,1;1}}{\Delta t_{l,1;1}}, \frac{\hat{x}_{l,1;4} - \hat{x}_{l,1;3}}{\Delta t_{l,1;3}}, \dots, \frac{\hat{x}_{l,2;2} - \hat{x}_{l,2;1}}{\Delta t_{l,2;1}}, \dots \right]^T, \quad (7)$$

\hat{Z}_l is a matrix with the corresponding features and covariates,

$$\hat{Z}_l = \begin{pmatrix} \hat{x}_{1,1;1} & \hat{x}_{2,1;1} & \cdots & y_{1,1;1} & y_{2,1;1} & \cdots \\ \hat{x}_{1,1;3} & \hat{x}_{2,1;3} & \cdots & y_{1,1;3} & y_{2,1;3} & \cdots \\ \vdots & \vdots & & \vdots & \vdots & \\ \hat{x}_{1,2;1} & \hat{x}_{2,2;1} & \cdots & y_{1,2;1} & y_{2,2;1} & \cdots \\ \vdots & \vdots & & \vdots & \vdots & \end{pmatrix} = \begin{pmatrix} (\hat{\mathbf{z}}_{1,1})^T \\ (\hat{\mathbf{z}}_{1,3})^T \\ \vdots \\ (\hat{\mathbf{z}}_{2,1})^T \\ \vdots \end{pmatrix},$$

the noise term is given by $\mathbf{s}_{\sigma_l} = \sqrt{2}\sigma_l[\Delta t_{l,1;1}^{-1}, \Delta t_{l,1;3}^{-1}, \dots, \Delta t_{l,2;1}^{-1}, \dots]$, and the matrix $K_{\alpha_l, \rho_l}(\hat{Z}_l, \hat{Z}_l)$ is obtained by applying the kernel to all combinations of rows in \hat{Z}_l .

To compute $\mathbf{x}_s(t_{l,s;i})$ in the likelihood term (Eq. (6)) we used forward Euler integration

$$\mathbf{x}_s(t_{l,s;i+1}) = \mathbf{x}_s(t_{l,s;i}) + \mathbf{v}(\mathbf{z}_s(t_{l,s;i}))\Delta t_{l,s;i}$$

with initial condition given by Eq. (2), where

$$[\mathbf{v}(\mathbf{z})]_l = \mathbf{k}_{\alpha_l, \rho_l}(\mathbf{z}, \hat{Z}_l)^T \left(K_{\alpha_l, \rho_l}(\hat{Z}_l, \hat{Z}_l) + \text{diag}(\mathbf{s}_{\sigma_l}^2) \right)^{-1} \hat{\mathbf{v}}_l$$

Note that the trajectory of all biomarkers should be computed simultaneously, even when only a single component l is needed in Eq. (6). This doesn't represent any problem as far as the covariates $\mathbf{y}_s(t_{l,s;i})$ are available at all time points because the observations $\hat{x}_{l',s;i}$ are not used in this computations for $l' \neq l$. This implies that the biomarkers don't need to be acquired at the same time points. This is an important consideration for long longitudinal studies, such as ADNI, for which each imaging modality is scheduled at a different rate and the time gap between the first acquisition of two modalities differ between subjects.

However, the matrix \hat{Z}_l should be complete for each component l . This implies that, at least at the first of the two time points used to compute the differences in Eq. (4), the complete set of observations is needed, and that the vectors $\hat{\mathbf{v}}_l$ in Eq. (7) may have different lengths for each biomarker l .

3 Experiments

3.1 Data

The model was fitted to the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database¹. All subjects from the ADNI dataset having at least 4 valid AV45 (Florbetapir) PET scans and 4 MRI T1 scans were used to fit the model, resulting in a total of 198 participants (88 Cognitively Normal and 110 with Mild Cognitive Impairment), and 874 PET and 1225 MRI measurements.

Three features were selected: mean AV45-PET SURV (average PET signal in cortical grey matter normalized by whole cerebellum), and the ratios of hippocampal and ventricular volume to intracranial volume (ICV)². Covariates included age and the presence of a copy of the E4 allele of the apolipoprotein-E (APOE) gene. The covariate vector \mathbf{y} had only 1 dimension (age) because velocities fields for APOE E4 carriers and non-carriers were kept apart and estimated separately, sharing only the hyperparameters.

3.2 Results

The posterior distributions of the model parameters were obtained with Markov chain Monte Carlo (MCMC) sampling using Stan software [17]. To explore the model predictive performance we did a leave-one-site-out experiment consisting in removing all subjects from a given hospital, except for a few observations used for predicting the rest of the biomarker trajectory, but not for velocity estimation. All the observations from left-out sites within a 5-year interval centred at the PET-MRI time overlap were used to estimate the rest of the trajectory. This interval was defined in order that all subjects to have at least one PET observation. As AV45 started to be acquired after MRI, there are few observations after this period. Therefore, prediction time in the past is larger than in the future.

Figure 1 shows predicted trajectories along with observations for a subset of subjects. Specifically, we selected the subject with most data not shown to the model from each site. Then, the 8 subjects with the longest unobserved trajectories from the APOE E4 non-carrier group were selected.

Apart from prediction, the model allows to test an endless amount of hypotheses, such as the mean difference in a given biomarker rate of change between two given sub-populations. For the sake of illustration, we have focused on the hippocampal rate of change, shown in Fig. 2. The top row panels show a representation of the velocity field in the MRI plane (hippocampal and ventricular volumes) and the bottom row panels show the rate of change of hippocampal volume for different conditions. The most prominent pattern is that APOE E4 carriers present higher rates of hippocampal atrophy than E4 non-carriers. These dynamics are only mildly modulated by brain amyloid levels, as can be observed in the bottom row panels. The strong influence of genetic factors in AD dynamics and the importance of considering their effect in AD progression modelling was recently highlighted in [4].

¹ adni.loni.usc.edu.

² The MRI volumes were computed using FreeSurfer (5.1).

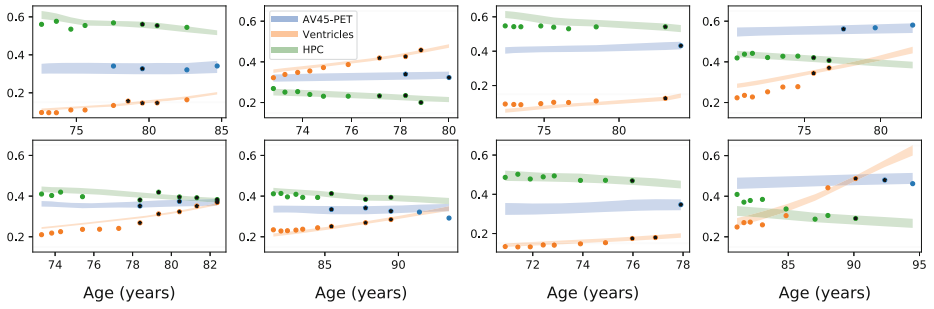


Fig. 1. Normalized biomarker trajectories for selected individuals. Blue: AV45-PET mean SUVR/3; Orange: Ventricular volume/ICV $\times 10$; Green: Hippocampal (HPC) volume/ICV $\times 100$. Dots correspond to observations, black stars denote observations used for prediction while the rest of the observations were hidden during the model inference. Shaded areas represent the 90% highest density interval of the prediction posterior probability. (Color figure online)

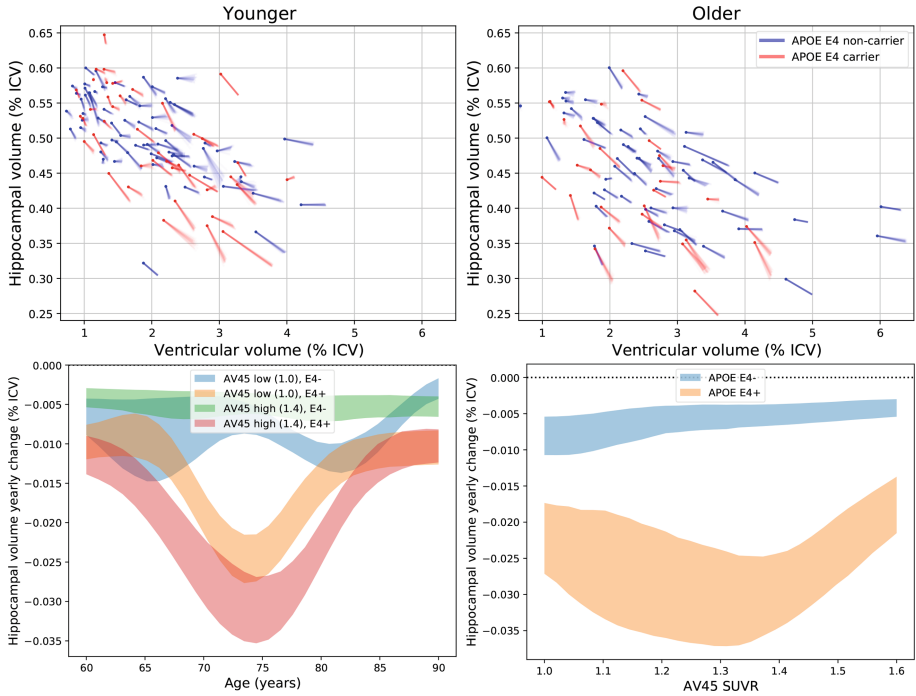


Fig. 2. Top: Representation of velocity fields. Dots correspond to the mean estimated initial values and lines represent velocity (two years of evolution) and their uncertainty (lines are drawn from the posterior distribution). **Bottom:** Hippocampal volume rate of change among subjects with mild neurodegeneration for different conditions of age, APOE and amyloid PET. Shaded areas represent 90% highest density interval of the posterior probability.

4 Discussion

We have presented a statistical model of brain-derived biomarker progression that overcomes important limitations of previous progression models of AD [14]. Remarkably, the proposed model dispenses with the assumption of a common disease trajectory. Biomarker independence is another limiting assumption frequently made. Conversely, the relationship between one biomarker value and another biomarker dynamics is at the core of the proposed model. Finally, this work presents the first non-parametric ODE-based AD progression model.

We have illustrated the model on PET and MRI-derived biomarkers and shown its potential as a tool for AD dynamics understanding and prediction. We showed the feasibility of full Bayesian posterior inference using MCMC in a moderate-sized dataset. The only multivariate ODE-based model of AD progression we are aware of used a variational approximation to estimate posteriors [1].

4.1 Limitations and Future Directions

A limitation of this work is the small number of selected biomarkers. We foresee no mayor computational issues in adding a large number of biomarkers, because GPs scale well with dimensionality. However, more experiments are required to verify the stability of the estimates.

The proposed model can be easily extended with other relevant AD tests, such as cognitive tests. Additionally, a cross-sectional clinical prediction model whose input features are the biomarkers and covariates used in this work could be added on top of the progression model. The progression and the diagnostic model combine together to forecast diagnosis in the future, *i.e.*, producing prognosis predictions.

Acknowledgements. This work was partially funded by INNOVIRIS (Brussels Capital Region, Belgium) under the project: ‘DIMENTIA: Data governance in the development of machine learning algorithms to predict neurodegenerative disease evolution’ (BHG/2020-RDIR-2b).

The data used in preparation of this article was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012).

References

1. Abi Nader, C., Ayache, N., Frisoni, G.B., Robert, P., Lorenzi, M.: For the Alzheimer’s Disease Neuroimaging Initiative: Simulating the outcome of amyloid treatments in Alzheimer’s disease from imaging and clinical data. *Brain Commun.* **3**(2), 1–17 (2021). <https://doi.org/10.1093/braincomms/fcab091>. <https://academic.oup.com/braincomms/article-pdf/3/2/fcab091/38443216/fcab091.pdf>
2. Budgeon, C., Murray, K., Turlach, B., Baker, S., Villemagne, V., Burnham, S.: For the Alzheimer’s Disease neuroimaging initiative: constructing longitudinal disease progression curves using sparse, short-term individual data with an application to Alzheimer’s disease. *Stat. Med.* **36**(17), 2720–2734 (2017)

3. Donohue, M.C., et al.: Estimating long-term multivariate progression from short-term data. *Alzheimer's Dementia* **10**(5, Supplement), S400–S410 (2014)
4. Frisoni, G.B., et al.: The probabilistic model of Alzheimer disease: the amyloid hypothesis revised. *Nat. Rev. Neurosci.* **23**(1), 53–66 (2022)
5. Herrup, K.: The case for rejecting the amyloid cascade hypothesis. *Nat. Neurosci.* **18**(6), 794–799 (2015)
6. Jack, C.R., et al.: Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol.* **12**(2), 207–216 (2013)
7. Jack, C.R., et al.: Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol.* **9**(1), 119–128 (2010)
8. Jack, C.R., et al.: Age-specific and sex-specific prevalence of cerebral β -amyloidosis, tauopathy, and neurodegeneration in cognitively unimpaired individuals aged 50–95 years: a cross-sectional study. *Lancet Neurol.* **16**(6), 435–444 (2017)
9. Jedynak, B.M., et al.: A computational neurodegenerative disease progression score: Method and results with the Alzheimer's disease neuroimaging initiative cohort. *Neuroimage* **63**(3), 1478–1486 (2012)
10. Knopman, D.S., et al.: Alzheimer Disease. *Nat. Rev. Dis. Primers.* **7**(1), 33 (2021)
11. Kovacs, G.G., et al.: Non-Alzheimer neurodegenerative pathologies and their combinations are more frequent than commonly believed in the elderly brain: a community-based autopsy series. *Acta Neuropathol.* **126**(3), 365–384 (2013)
12. Lorenzi, M., Filippone, M., Frisoni, G.B., Alexander, D.C., Ourselin, S.: Probabilistic disease progression modeling to characterize diagnostic uncertainty: application to staging and prediction in Alzheimer's disease. *Neuroimage* **190**, 56–68 (2019)
13. Nichols, E., et al.: Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the global burden of disease study 2019. *Lancet Public Health* **7**(2), e105–e125 (2022)
14. Oxtoby, N.P., Alexander, D.C.: For the EuroPOND consortium: Imaging plus x: multimodal models of neurodegenerative disease. *Curr. Opinion Neurol.* **30**(4), 371–379 (2017)
15. Oxtoby, N.P., et al.: Learning imaging biomarker trajectories from noisy Alzheimer's Disease data using a Bayesian multilevel model. In: Cardoso, M.J., Simpson, I., Arbel, T., Precup, D., Ribbens, A. (eds.) *BAMBI 2014*. LNCS, vol. 8677, pp. 85–94. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-12289-2_8
16. Schiratti, J.B., Allasonnière, S., Colliot, O., Durrleman, S.: A bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations. *J. Mach. Learn. Res.* **18**(133), 1–33 (2017)
17. Stan Development Team: Stan Modeling Language Users Guide and Reference Manual, Version 2.29 (2019). <https://mc-stan.org>