Riemannian metric learning for progression modeling of longitudinal datasets

Benoît Sauty, Stanley Durrleman

RIEMANNIAN METRIC LEARNING FOR PROGRESSION MODELING OF LONGITUDINAL DATASETS.

*Benoît Sauty** *Stanley Durrleman**

*Inria, Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, AP-HP, Paris, France

ABSTRACT

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

***Index Terms*—** Disease Modeling - Riemannian manifolds - Mixed-effects models - Alzheimer's disease

1. INTRODUCTION

1.1. Motivation

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

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

1.2. Related work

One approach to construct disease progression models for biomarkers is to formulate the changes using differential equations [1–3], with the caveat that covariates and all sources of random variations should be integrated in the differential equation system which makes interpretation difficult. Another approach called event-based models is proposed in [4, 5], in which cutoff points of abnormality are inferred from observed biomarkers, and disease stage is mapped to a discrete set of biomarker-abnormality events. While providing good robustness, the dichotomization of variables can be ill suited for continuous biomarkers. Lately, recurrent neural networks have also been used to model the progression of scalar [6, 7], imaging [8] and multi-modal data [9, 10] offering a flexible framework for regression.

A more interpretable approach is proposed with mixed-effects models [11], which account for both the average trajectory of the population, called the fixed-effects, and individual variations to that trajectory that account for inter-subjects variability, called the random-effects. The inter-subject variability can be thought of as the combined effect of the pace at which individuals evolve, and of their instrinsic biological caracteristics. That variability can be modeled by a time reparametrization to "align" patients on a common pathological timeline and spaceshifts that define how each patient compares to the average trajectory. Early models used linear modeling [12, 13] while, later, non linearities were added with polynomial [14], logistic [15], exponential [16] and semi-parametric regressions [17].

To model the progression of a disease, one less restrictive assumption is to consider that the observed biomarkers follow continuous trajectories in the space of observations that is assumed to be a Riemannian manifold [18]. Particular cases have been derived for the analysis of longitudinal scalar measurements in [19,20], where this model is studied with a fixed metric that is known *a priori* to yield logistic trajectories. Besides, for multivariate data, the metric is usually set to be a product metric, so that the average trajectory is estimated as a product of independant 1-dimensionnal trajectories.

1.3. Contributions

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

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

2. TRAJECTORY MODEL

2.1. A general metric

The set of possible observation points $\mathbb{M}$ is endowed with a Riemannian manifold structure within the total normalized observation space $[0,1]^N$. As suggested in [7] we consider the family of Riemannian metrics on this manifold, which are defined via their cometrics as a Gaussian interpolation on a set of control points $(c_i)_{1\leqslant i\leqslant N_\delta} \in \mathbb{M}^{N_\delta}$,

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

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

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

2.2. Reference geodesic and parallel curves

For any metric, given a reference geodesic $\gamma : t \in I \mapsto \gamma(t)$ with $I \subset \mathbb{R}$, we can define a family of trajectories called the Exp-parallel curves. A vector field X is said to be *parallel* along a curve $\gamma : I \to \mathbb{M}$ if $\nabla_{\gamma(t)} X = 0$ where ∇ is the connection on $\mathbb{M}$, and for $t_0 \in I$ and $w_0 \in T_{\gamma(t_0)}\mathbb{M}$, one can show there exists a unique vector field w parallel along γ such that $w(t) = w_0$. Thus, given a vector w_0, called a spaceshift, we can define the parallel curve $\eta^w : t \mapsto \mathrm{Exp}(w(t))$ where Exp is the exponential map on $\mathbb{M}$, that maps the vectors of the tangent bundle to the manifold.

The following figure illustrates the benefits of learning the metric (right) when compared to a fixed logistic metric (left) where $\mathbb{M} \subset [0,1]^2$ and $t \in I$ represents the pathological time.

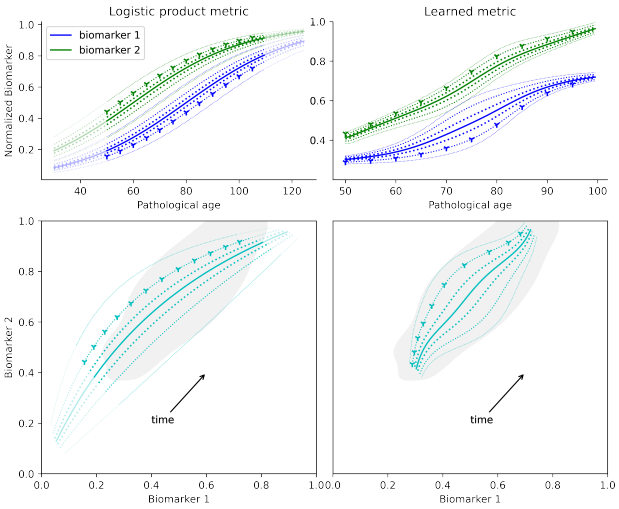

Fig. 1. Top row shows the progression of the features across time for main geodesic (plain) and its Exp-parallel curves (dotted). Bottom row shows these trajectories in $\mathbb{M}$, with implicit time, and a kernel density estimation (80% threshold) of the observed data in grey. One parallel curve is highlighted with crosses to display the effect of a spaceshift. Logistic trajectories are extrapolated (shaded) beyond the datapoints to exhibit the forced 0 and 1 asymptotes.

The protocol to get these curves is detailed in the Results section. Learning the metric allows to obtain a wide variety of trajectories for the reference geodesic, and allows the spaceshifts to impact the shape of the parallel curves, as opposed to the fixed product metric where the spaceshift only changes the timing and ordering of feature progression.

3. STATISTICAL MODEL

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

3.1. Generative statistical model

The average trajectory of the population can be entirely characterised by two parameters $p_0 \in \mathbb{M}$ and $v_0 \in T_{p_0}(\mathbb{M})$ that are respectively a position in the observation space and a velocity in the tangent plane at this point. If we add a reference time t_0, this defines a unique geodesic $\gamma_0 : t \in \mathbb{R} \mapsto \gamma_0(t) \in \mathbb{M}$ such that $\gamma_0(t_0) = p_0$ and $\dot{\gamma}(t_0) = v_0$. Fig. 2. summarizes this geometric description.

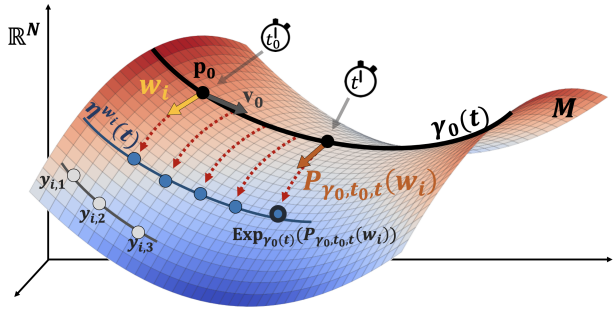

Fig. 2. The geodesic $\gamma_0(t)$ represents the group-average trajectory and the individual trajectories $\eta^{w_i}(t)$ correspond to the exponential mapping of the vectors $P_{\gamma_0,t_0,t}(w_i)$, that are the parallel transport of the vector $w_i \in T_{p_0}\mathbb{M}$ along γ_0

As in [18], each patient is described by a couple of individual parameters $\alpha_i \in \mathbb{R}_+^*$ and $\tau_i \in \mathbb{R}$ that align the individual timeline to the reference timeline via the affine time-warp function $\psi_i : t \mapsto \alpha_i(t - \tau_i) + t_0$, as well as a spaceshift $w_i \in T_{p_0}(\mathbb{M}) \subset \mathbb{R}^N$. τ_i is the onset age, α_i represents the pace at which the patient evolves and the spaceshift defines which Exp-parallel curve represents the progression of the patient. For convenience purposes, we write the acceleration factor $\alpha_i = \exp(\xi_i)$ with $\xi_i \in \mathbb{R}$. For every subject i, given both ψ_i and w_i, we can now model the scalar measurements $(y_{i,j})_{1 \leqslant j \leqslant k_i}$ as the sample points at times $(\psi_i(t_{i,j}))_{1 \leqslant j \leqslant k_i}$ with additional noise ε

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

3.2. Mixed-effects formulation

We choose a Gaussian noise $\varepsilon_{i,j} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. Then, the previously introduced latent variables are modeled as random-effects $z = (\xi, \tau, s)$, with $\xi_i \sim \mathcal{N}(0, \sigma_\xi^2)$, $\tau_i \sim \mathcal{N}(t_0, \sigma_\tau^2)$ and $s_i \sim \mathcal{N}(0, 1)$. The source $s_i \in \mathbb{R}^{N_s}$ reconstructs w_i through $w_i = As_i$, where the "mixing matrix" $A \in \mathbb{R}^{N \times N_s}$ is also estimated and serves a purpose of dimensionality reduction. Then, the fixed-effects are written $\theta = (p_0, v_0, t_0, A, (L_i)_{1 \leqslant i \leqslant N_\delta}, \sigma_\varepsilon, \sigma_\xi, \sigma_\tau)$.

4. ESTIMATION

We proceed to a Maximum A Posteriori (MAP) estimation of the parameters. The Expectation (E) step from a regular Expectation Maximisation (EM) algorithm would require the computation of intractable integrals for the likelihood, so we resort to the Stochastic Approximation EM (SAEM) in which the E step is replaced by a Simulation (S) and Approximation (A) step. For curved Exp-parallel models, the S step can be replaced by a single transition of an ergodic Monte-Carlo Markov Chain (MCMC) whose stationary distribution is estimated iteratively using a Metropolis-Hastings sampler. This global algorithm is called the Monte-Carlo Markov Chain Stochastic Approximation Expectation-Maximization (MCMC-SAEM) [21, 22], and is proven to converge towards a global maximum of the averaged likelihood [23]

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

We refer the reader to [18] for an extensive presentation of the estimation algorithm for non linear mixed-effects models for manifold valued longitudinal observations. The specificity of our approach resides in the additional (L_i) parameters for the metric, which are estimated via a line-search gradient descent between each iteration of the MCMC-SAEM procedure. Besides, there is no closed-form expression for γ_0, so we integrate the Hamiltonian differential system using a Runge-Kutta numerical scheme, and the parallel transport of the spaceshifts is computed using the Fanning scheme [24].

5. EXPERIMENTS AND RESULTS

To validate this approach, we simulated 2 dimensional data according to trajectories that are known to occur naturally, namely a logistic shape with asymptotes other than 0 and 1, and a sum of logistics that represents features with multiple inflection points. We engineered a metric that yields these trajectories as geodesics by fitting our model to a family of shifted and accelerated versions of these shapes. With this metric we generated a realistic dataset of 200 patients with an average follow-up of 10 visits spanned over 5 years. We choose $\sigma_\varepsilon = 0.02, \sigma_\tau = 15, \sigma_\xi = 0.5, t_0 = 70$ and sampled the individual parameters, then used the engineered geodesic (v_0, p_0) and parameters $((L_i), A)$ for the generative model $y_{i,j} = \eta_{\gamma_0,\mathrm{G}}^{w_i}(\psi_i(t_{i,j})) + \varepsilon_{i,j}$ with random age at baseline.

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

Table 1. Results over 5 successive estimations. The errors on τ and ξ are averaged across patients.

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

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

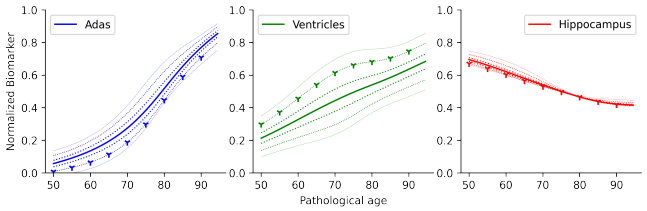

Fig. 3. Normative scenario predicted by our model for the 3 features : main logistic (plain) and parallels (dotted). One parrallel curve is highlighted with crosses.

As expected, we recover a "sigmoid-like" trajectory for the cognitive score and "linear-like" trajectories for the imaging markers. Both our method and the logistic fixed metric outperform the 56 methods benchmarked in the TADPOLE challenge [25] with improved prediction of the ADAS score (respectively 4.33 and 4.27 MAE compared to 4.70 for the best method in TADPOLE). Since the reconstruction errors and prediction errors are not significatively different, it is interesting to question whether the estimated individual parameters offer greater quantitative insights with the learned metric.

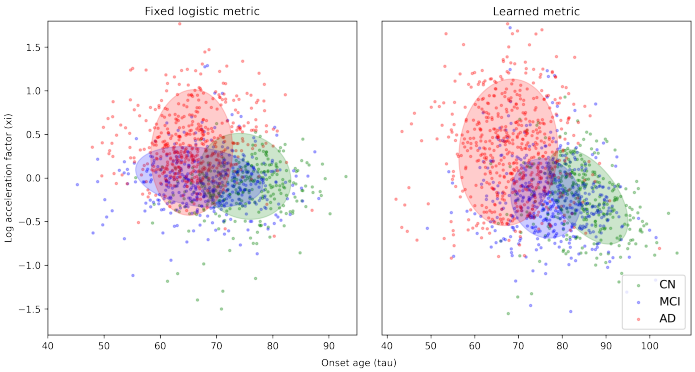

Fig. 4. Estimated individual parameters for the logistic (left) and learned (right) metric with unsupervised Gaussian Mixture Model to predict diagnosis.

Fig.4 illustrates the distribution of individual parameters for patients diagnosed Cognitively Normal (CN), Mild Cognitive Impairment (MCI) and Alzheimer Disease (AD) on their last visit. Fitting a Gaussian Mixture Model (GMM) shows that the distribution for each diagnosis cluster are better separated with our approach. 20 repetitions of 5-fold validation with GMM for diagnosis prediction gave us an average **62.5%±0.8%** test accuracy (confusion matrix reported below) compared to 53.7%±1.2% for the logistic metric model.

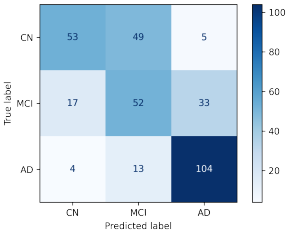

Fig. 5. Confusion matrix (test set) for diagnosis prediction.

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

6. CONCLUSION

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

This work was funded in part by grant number 678304 (ERC) and 826421 (TVB-Cloud) from H2020 programme, and ANR-10-IAIHU-06 (IHU ICM), ANR-19-P3IA-0001 (PRAIRIE) and ANR-19-JPW2-000 (E-DADS) from ANR.

7. COMPLIANCE WITH ETHICAL STANDARDS

There are no conflicts of interest associated with this publication. Ethical approval was not required as confirmed by the license attached with the open access data.

8. REFERENCES

[1] K. Ito, B. Corrigan, Q. Zhao, J. French, R. Miller, H. Soares, E. Katz, T. Nicholas, B. Billing, R. Anziano, et al. Disease progression model for cognitive deterioration from Alzheimer's Disease Neuroimaging Initia-

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

tive database. *Alzheimer's & Dementia*, 7(2):151–160, 2011.

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

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

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

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

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

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

[8] Ruoxuan Cui and Manhua Liu. RNN-based longitudinal analysis for diagnosis of Alzheimer's disease. *Computerized Medical Imaging and Graphics*, 73:1–10, 2019.

[9] Raphäel Couronné, Maxime Louis, and Stanley Durrleman. Longitudinal autoencoder for multi-modal disease progression modelling. preprint, April 2019.

[10] M. Louis, R. Couronné, I. Koval, B. Charlier, and S. Durrleman. Riemannian geometry learning for disease progression modelling. In *Internationnal Conference on IPMI*, pages 542–553. Springer, 2019.

[11] Nan M Laird and James H Ware. Random-effects models for longitudinal data. *Biometrics*, 1982.

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

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

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

[15] B. Jedynak, A. Lang, B. Liu, E. Katz, Y. Zhang, B. Wyman, 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.

[16] Lars Lau Raket. Statistical disease progression modeling in Alzheimer disease. *Frontiers in big Data*, 3, 2020.

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

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

[19] I. Koval, J-B Schiratti, A. Routier, M. Bacci, O. Colliot, S. Allassonnière, and S. Durrleman. Statistical learning of spatiotemporal patterns from longitudinal manifold-valued networks. In *International Conference on MICCAI*, pages 451–459. Springer, 2017.

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

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

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

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

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

[25] R. Marinescu, N. Oxtoby, A. Young, E. Bron, A. Toga, M. Weiner, F. Barkhof, N. Fox, A. Eshaghi, T. Toni, et al. The alzheimer's disease prediction of longitudinal evolution (TADPOLE) challenge: Results after 1 year follow-up. 2020.