Statistical learning of spatiotemporal patterns from longitudinal manifold-valued networks

I. Koval[3,1], J.-B. Schiratti[3,1], A. Routier[1], M. Bacci[1], O. Colliot[1,2], S. Allassonnière[3], S. Durrleman[1], the Alzheimer's Disease Neuroimaging Initiative

[1] Inria Paris-Rocquencourt, Inserm U1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMRS 1127, Institut du Cerveau et de la Moelle épinière, ICM, F-75013, Paris, France
[2] AP-HP, Pitié-Salpêtrière Hospital, Departments of Neurology and Neuroradiology, F-75013, Paris, France
[3] INSERM UMRS 1138, Centre de Recherche des Cordeliers, Université Paris Descartes, Paris, France

Abstract. We introduce a mixed-effects model to learn spatiotemporal patterns on a network by considering longitudinal measures distributed on a fixed graph. The data come from repeated observations of subjects at different time points which take the form of measurement maps distributed on a graph such as an image or a mesh. The model learns a typical group-average trajectory characterizing the propagation of measurement changes across the graph nodes. The subject-specific trajectories are defined via spatial and temporal transformations of the group-average scenario, thus estimating the variability of spatiotemporal patterns within the group. To estimate population and individual model parameters, we adapted a stochastic version of the Expectation-Maximization algorithm, the MCMC-SAEM. The model is used to describe the propagation of cortical atrophy during the course of Alzheimer's Disease. Model parameters show the variability of this average pattern of atrophy in terms of trajectories across brain regions, age at disease onset and pace of propagation. We show that the personalization of this model yields accurate prediction of maps of cortical thickness in patients.

1 Introduction

There is a great need to understand the progression of Alzheimer's Disease (AD) especially before the clinical symptoms to better target therapeutic interventions [8]. During this silent phase, neuroimaging reveals the disease effects on brain structure and function, such as the atrophy of the cortex due to neuronal loss. However, the precise dynamics of the lesions in the brain are not so clear at the group level and even less at the individual level. Personalized models of lesion propagation would enable to relate structural or metabolic alterations to the clinical signs, offering ways to estimate stage of the disease progression in the pre-symptomatic phase. Numerical models have been introduced to describe the temporal and the spatial evolution of these alterations, defining a *spatiotemporal*

trajectory of the disease, *i.e.* a description of the changes in the brain over time, such as lesion progressions, tissue deformation and atrophy propagation.

Statistical models are well suited to estimate distributions of spatiotemporal patterns of propagation out of series of short-term longitudinal observations [3,6]. However, the absence of time correspondence between patients is a clear obstacle for these types of approaches. Using data series of several individuals requires to re-align the series of observations in a common time-line and to adjust to a standardized pace of progression. Current models either consider a sequential propagation [16], without taking into account the continuous dynamics of changes, or develop average scenarios [7,9]. Recently, a generic approach to align patients has been proposed for a set of biomarkers in [15]: the temporal inter-subject variability results from individual variations of a common time-line granting each patient a unique age at onset and pace of progression. On top of the time-alignment of the observations, there exists a spatial variability of the signal propagation that characterizes a distribution of trajectories.

In order to exhibit a spatial representation of the alterations, we study medical images or image-derived features taking the form of a signal discretized at the vertices of a mesh, for instance the cortical thickness distributed on the mesh of the pial surface or Standardized Uptake Value Ratio (SUVR) distributed on the regular voxel grid of a PET scan. The spatial distribution of the signal is encoded in a distance matrix, giving the physical distance between the graph nodes. A sensible prior to include in the model is to enforce smooth variations of the temporal profile of signal changes across neighbouring nodes, highlighting a propagation pattern across the network as in [13]. Extending directly the model in [15] may lead to an explosion of the number of parameters proportional to the mesh resolution. At infinite resolution, the parameters take the form of a smooth continuous map defined on the image domain. In this paper, we propose to constrain these maps to belong to a finite-dimensional Hilbert Space, penalizing high frequency variations. In practice, these maps are generated by the convolution of parameter values at a sparse set of control nodes on the network with a smoothing kernel. The number of control nodes, whose distribution is determined by the bandwidth of the kernel, controls the complexity of the model regardless of the mesh resolution. Furthermore, the propagation of non-normalized signal could not adequately be modeled by the same curve shifted in time as in [15]. We introduce new parameters to account for smooth changes in the profiles of changes at neighbouring spatial locations.

We introduce a mixed-effect generative model that learns a distribution of spatiotemporal trajectory from series of repeated observations. The model evaluates individual parameters (time reparametrization and spatial shifts) that enables the reconstruction of individual disease propagation through time. This non-linear problem is tackled by a stochastic version of the EM algorithm, the MCMC-SAEM [1,11] in a high-dimensional setting. It considers fixed-effects describing a group-average trajectory and random effects characterizing individual trajectories as adjustment of the mean scenario. It is used to detect the cortical thickness variations in MRI data of MCI converters from the ADNI database.
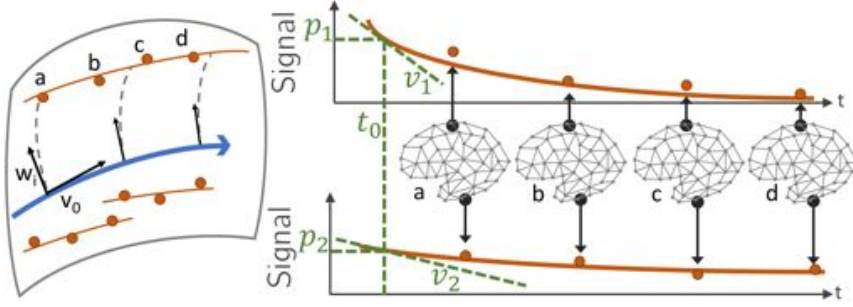
Fig. 1: Manifold representation of the mesh observations (left). Orange dots are patient real observations. The blue line is the reconstruction of the mean propagation. The signal value at each node (right), parametrized by (p, t, v), allows the reconstruction of the propagation over the network (orange lines)

2   Manifold-valued networks

In the following, we consider a longitudinal dataset $\mathbf{y} = (\mathbf{y}_{i,j})_{1 \leq i \leq p,\, 1 \leq j \leq k_i}$ of p individuals, such that the ith individual is observed at k_i repeated time points $t_{i,1} < \ldots < t_{i,k_i}$. We assume that each observation $\mathbf{y}_{ij}$ takes the form of $N_v \in \mathbb{N}^*$ scalar measures $((\mathbf{y}_{i,j})_1, ..., (\mathbf{y}_{i,j})_{N_v})$ referred to as a *measurement map*.

Manifold-valued measurements distributed on a fixed graph
Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a non-oriented graph where $\mathcal{V} = (\mathbf{x}_1, ..., \mathbf{x}_{N_v})$ is a set of vertices of a mesh in $\mathbb{R}^3$ and $\mathcal{E}$ is a subset of pairs of vertices defining the graph edges. We assume that $\mathcal{G}$ is a common fixed graph such that the kth coordinate of each measurement map y_{ij} corresponds to the vertex $\mathbf{x}_k \in \mathcal{V}$. As the graph corresponds to measurements spatially distributed on a mesh, the edges embed a spatial configuration. Therefore, any edge $(\mathbf{x}_i, \mathbf{x}_j)$ is valued with d, a geodesic distance on the graph, defining a distance matrix $\mathcal{D}$ such that for all $i, j \in \{1, .., N_v\}$, $\mathcal{D}_{i,j} = d(\mathbf{x}_i, \mathbf{x}_j)$. Each measurement map $\mathbf{y}_{i,j}$ produces a network $(\mathcal{G}, \mathcal{D}, \mathbf{y}_{i,j})$, *i.e.* a fixed graph with one-dimensional values associated to each vertex and with distances associated to each edge.

We consider that the measurements of the patients at each node k corresponds to observations of a signal function $t \mapsto \boldsymbol{\gamma}_k(t)$ at particular time points. Thus, the function $t \mapsto \boldsymbol{\gamma}(t) = (\boldsymbol{\gamma}_1(t), ..., \boldsymbol{\gamma}_{N_v}(t))$ describes the evolution of the signal over the whole network. We assume that each signal function is continuous and that measurement map $\mathbf{y}_{i,j} \in \mathbf{y}$ lies in a space defined by smooth constraints, as expected for bounded or normalized observations (eg. volume ratios, thickness measures, SUVR). Therefore, the space of measurements is best described as a *Riemannian manifold* [5, 12], leading to consider each function $t \mapsto \boldsymbol{\gamma}_k(t)$ as a one-dimensional geodesically complete Riemannian manifold (M, g^M) such that all spatial observation $\mathbf{y}_{i,j}$ is a point in the product manifold M^{N_v}. It follows that for each i, j, $(\mathcal{G}, \mathcal{D}, \mathbf{y}_{i,j})$ is a manifold-valued network.

Spatial smoothness of the propagation

Besides the temporal smoothness of the propagation, we expect the signal to be similar for neighbour nodes. We consider that each node is described by N_p parameters that parametrize the signal trajectory. In order to ensure smooth variations of the parameters values at neighbouring nodes, we assume that they result from the interpolation of the parameter values at a sparse subset of uniformly distributed nodes $\mathcal{V}_C = (\mathbf{x}_{d_1}, ..., \mathbf{x}_{d_{N_c}})$, called control nodes. For each parameter p, potentially estimated at each node, the control nodes define a parameter evaluation function $p(\mathbf{x})$ encoding for all the nodes: $\forall \mathbf{x} \in \mathcal{V}, \forall \mathbf{y} \in \mathcal{V}_C, \forall i \in \{1, \ldots, N_c\}$ $p(\mathbf{x}) = \sum_{i=1}^{N_C} K(\mathbf{x}, \mathbf{x}_{d_i})\beta_i$ and $p(\mathbf{y}) = p_{\mathbf{y}}$ where the $(\beta_i)_{1 \le i \le N_c}$ are the new model parameters and K is a Gaussian Kernel such that $\forall \mathbf{x}, \mathbf{y} \in \mathcal{V} K(\mathbf{x}, \mathbf{y}) = f(\frac{d(\mathbf{x},\mathbf{y})}{\sigma})$, d being the geodesic distance on the graph and σ the kernel bandwidth.

This convolution guarantees the spatial regularity of the signal propagation. Moreover this smooth spatial constraint enables a reduction of the number of parameters, reducing the dimensional complexity from N_p independent parameters at each node, to N_p parameters only at the control nodes.

3   The statistical model

A propagation model

Given a set of manifold-valued networks $(\mathcal{G}, \mathcal{D}, \mathbf{y})$, the model describes a group-average trajectory in the space of measurements, defined by a geodesic $\boldsymbol{\gamma}$ that allows to estimate a typical scenario of progression. Individual trajectories derive from the group-average scenario through spatiotemporal transformations: the *exp-parallelization* and the *time reparametrization.*

First, to describe disease pace and onset specific to each subject, we introduced a temporal transformation, called the time-warp, that is defined, for the subject i, by $\psi_i(t) = \alpha_i(t_{i,j} - \tau_i - t_0) + t_0$ where t_0 is the reference time-point in the space of measurements. The parameter τ_i corresponds to the time-shift between the mean and the individual age at onset and α_i is the acceleration factor that describes the pace of an individual, being faster or slower than the average. This *time reparametrization* allows to reposition the dynamics of the average scenario in the real time-line of the ith individual.

The *exp-parallelization* allows to translation the observations in the space of measurements, from the mean scenario to individual trajectories, encoding a variation in the trajectory of changes across the nodes of the graph. This exp-parallelization is handled by a family of individual vectors $(\mathbf{w}_i)_{1 \le i \le p}$, called space-shifts. As shown on Figure 1 (left), the orange dots refer to individual observations in the space of measurements. The group-average trajectory estimated from the longitudinal measurements corresponds to the blue line. The space shifts characterize a spatial shift perpendicular to v_0 that describes the velocity of the mean scenario.

Finally, the parameters $(\alpha_i, \tau_i, \mathbf{w}_i)$ allow the reconstruction of the individual trajectories from the mean scenario of propagation.

Given a noise $\boldsymbol{\varepsilon}_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \sigma^2 \mathrm{Id}_{N_v})$, the mixed-effect model writes, for a arbitrary vertex function $\boldsymbol{\gamma}_k(t)$:

$$(\mathbf{y}_{i,j})_k = \gamma_k \left(\frac{(\mathbf{w}_i)_k}{\dot{\gamma}_k(t_0)} + \alpha_i(t_{i,j} - t_0 - \tau_i) + t_0 \right) + (\boldsymbol{\varepsilon}_{i,j})_k \qquad (1)$$

Parameters estimation with the MCMC-SAEM algorithm

To reconstruct the long-term scenario of the disease propagation, we estimate the parameters of the group-average trajectory $\boldsymbol{\theta} = ((\beta^i_{param})_{1 \le i \le N_c, 1 \le j \le N_p}, \sigma)$ using a maximum likelihood estimator. The random-effects $\mathbf{z} = (z_i)_{1 \le i \le p} = (\mathbf{w}_i, \alpha_i, \tau_i)_{1 \le i \le p}$ are considered as latent variables, whose distributions characterize the variability of the individual trajectories. Due to the non-linearity in Eq. (1), we use a Stochastic Approximation Expectation Maximization [4] coupled with a Monte-Carlo Markov Chain sampler (MCMC-SAEM) [10]. Let $\theta^{(k)}$ be the current estimation of the parameters and $\mathbf{z}^{(k)}$ the current iterate of the Markov chain of the latent variables. The algorithm alternates between a simulation step, a stochastic approximation step and a maximization step, until convergence [1]. The simulation uses an adaptive version [2] of the Hasting Metropolis within Gibbs sampler to draw $\mathbf{z}^{(k+1)}$ from $(\mathbf{z}^{(k)}, \mathbf{y}, \boldsymbol{\theta}^{(k)})$. This algorithm was chosen as it handles non-linear mixed effects models [11] with proven convergence and consistent estimations in practice.

Model instantiation

As many measurements correspond to positive values (eg. the cortical thickness, volume ratios), we consider in the following the open interval $M =]0, +\infty[$ as a one-dimensional Riemannian manifold equipped with a Riemannian metric g such that for all $p \in M$ and for all $(u, v) \in T_p M$, $g_p(u, v) = uv/p^2$. With this metric and given $k \in \{1, \ldots, N_v\}$, M is a geodesically complete Riemannian manifold whose geodesics are of the form $t \mapsto \boldsymbol{\gamma}_k(t) = p_k \exp(\frac{v_k}{p_k}(t - t_k))$ where $p_k \in M$, $t_k \in \mathbb{R}, v_k \in T_{p_k} M$. These parameters are represented on Figure 1 (right) at two nodes where the decrease of the signal varies spatially. For identifiability reasons, we choose to fix the parameters t_k among the nodes, leading to a shared parameter t'_0 such that for all $k \in \{1, \ldots, N_v\}$ $t_k = t'_0$. As t'_0 can be arbitrarily chosen in $\mathbb{R}$, we fix $t'_0 = t_0$ defined in Section 3. Considering the interpolation functions introduced in Section 2 and the fact that the parameters (p^k_j) are (p_k, v_k), it leads to define $p(\mathbf{x}) = \sum_{i=1}^{N_C} K(\mathbf{x}, \mathbf{x}_{d_i}) \beta^i_p$ and $v(\mathbf{x}) = \sum_{i=1}^{N_C} K(\mathbf{x}, \mathbf{x}_{d_i}) \beta^i_v$

Finally, the model defined in (1) rewrites:

$$(\mathbf{y}_{i,j})_k = p(\mathbf{x}_k) \exp \left(\frac{(\mathbf{w}_i)_k}{p(\mathbf{x}_k)} + \frac{v(\mathbf{x}_k)}{p(\mathbf{x}_k)} \alpha_i(t_{i,j} - t_0 - \tau_i) \right) + (\boldsymbol{\varepsilon}_{i,j})_k \qquad (2)$$

such that $\boldsymbol{\theta} = (t_0, (\beta^i_p)_{1 \le i \le N_c}, (\beta^i_v)_{1 \le i \le N_c}, \sigma)$ and $\mathbf{z} = (\mathbf{w}_i, \alpha_i, \tau_i)_{1 \le i \le p}$
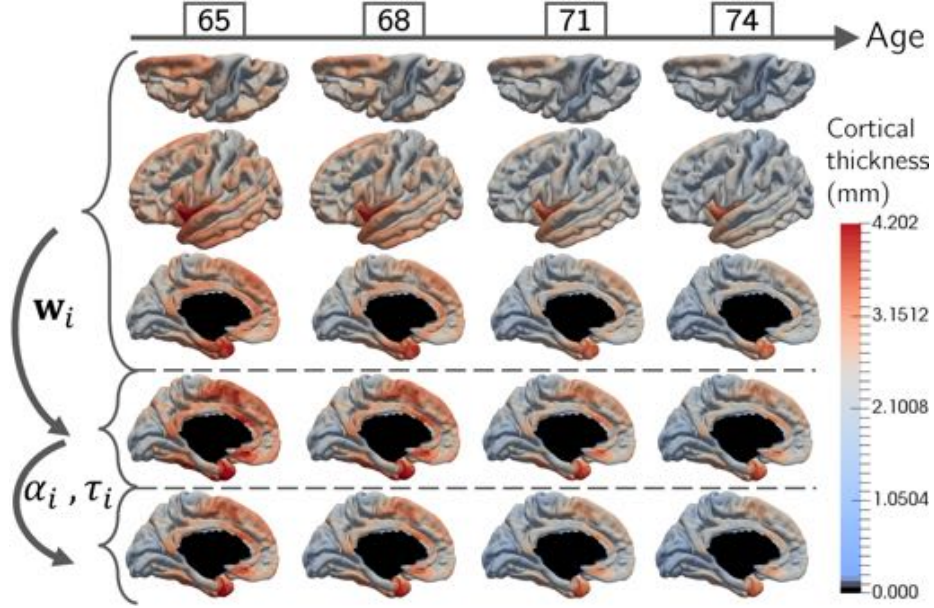
Fig. 2: Cortical thickness at 65, 68, 71 and 74 years old of the mean propagation (first rows). Effect of the space-shift $\mathbf{w}_i$ (fourth row), then with temporal reparametrization α_i, τ_i (fifth row) on the cortical thickness.

4   Experimental results

Data

We used this model to highlight typical spatiotemporal patterns of cortical atrophy during the course of Alzheimer's Disease from longitudinal MRI of 154 MCI converters from the ADNI database, which amounts for 787 observations, each subject being observed 5 times on average. We aligned the measures on a common atlas with FreeSurfer [14] so that the measurement maps are distributed on the same common fixed-graph $\mathcal{G}$ which is constituted of 1827 nodes that map the surface of the left pial surface. Out of the vertices, we selected 258 control nodes uniformly distributed over the surface. They encode the spatial interpolation of the propagation. The distance matrix $\mathcal{D}$ is defined by a geodesic distance on $\mathcal{G}$.

Cortical thickness measurements We used the model instantiation defined in Section 3 to characterize the cortical thickness decrease. Multiple runs of 30.000 iterations ($\sim$4hours) of this MCMC-SAEM lead to a noise standard deviation $\sigma \simeq 0.27$mm with 90% of the data included in $[1.5, 3.6]$ mm. The mean spatiotemporal propagation, described on the first three rows of the Figure 2 as the cortical thickness at respectively 65, 68, 71 and 74 years old shows that the primarily affected area is the medial-temporal lobe, followed by the temporal neocortex. The parietal association cortex and the frontal lobe are also

(a) Real observation of the cortical thickness (right) and its model reconstruction (left)

(b) Relative error between the observation and its reconstruction plotted on the mesh (left) and its histogram (right)
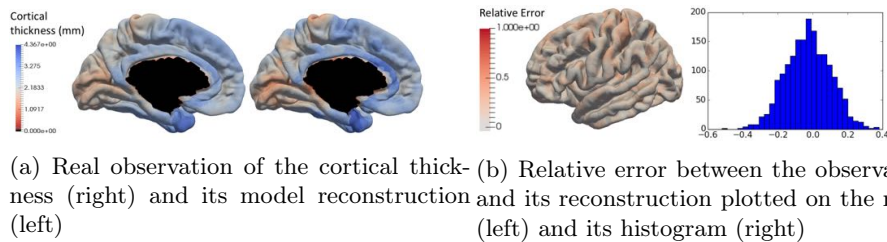
Fig. 3: Comparison of an observation and its reconstruction by the model

subject to important alterations. On the other side, the sensory-motor cortex and the visual cortex are less involved in the lesion propagation. These results are consistent with previous knowledge of the Alzheimer's Disease effects on the brain structure. As the model is able to exhibit individual spatiotemporal patterns with their associated pace of progression, the fourth and fifth rows of the Figure 2 represent consecutively the effect of the parallel shifting and of the time reparametrization on the cortical thickness atrophy. The figure 3a shows the real cortical thickness of a subject and the reconstruction predicted by the model. The relative error and its histogram are represented on Figure 3b. The reconstruction is coherent to the real observation, the remaining error represents essentially an unstructured noise that we precisely try to smooth out.

5 Discussion and perspectives

We proposed a mixed-effect model which is able to evaluate a group-average spatiotemporal propagation of a signal at the nodes of a mesh thanks to longitudinal neuroimaging data distributed on a common network. The network vertices describe the evolution of the signal whereas its edges encode a distance between the nodes via a distance matrix. The high dimensionality of the problem is tackled by the introduction of control nodes: they allow to evaluate a smaller number of parameters while ensuring the smoothness of the signal propagation through neighbour nodes. Moreover, individual parameters characterize personalized patterns of propagation as variations of the mean scenario. The evaluation of this non-linear high dimensional model is made with the MCMC-SAEM algorithm that leads to convincing results as we were able to highlight areas affected by considerable neuronal loss: the medial-temporal lobe or the temporal neocortex.

The distance matrix, which encodes here the geodesic distance on the cortical mesh, may be changed to account for the structural or functional connectivity information. In this case, signal changes may propagate not only across neighbouring locations, but also at nodes far apart in space but close to each other in the connectome. The model can be used with multimodal data, such as PET scans, introducing numerical models of neurodegenerative diseases that could inform about the disease evolution at a population level while being customizable to fit individual data, predicting stage of the disease or time to symptom onset.

References

1. Allassonnière, S., Kuhn, E., Trouvé, A.: Construction of bayesian deformable models via a stochastic approximation algorithm: a convergence study. Bernoulli 16(3), 641–678 (2010)
2. Atchadé, Y.F.: An adaptive version for the metropolis adjusted langevin algorithm with a truncated drift. Methodology and Computing in applied Probability 8(2), 235–254 (2006)
3. Bilgel, M., Prince, J.L., Wong, D.F., Resnick, S.M., Jedynak, B.M.: A multivariate nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging. NeuroImage 134, 658–670 (2016)
4. Delyon, B., Lavielle, M., Moulines, E.: Convergence of a stochastic approximation version of the em algorithm. Annals of statistics pp. 94–128 (1999)
5. Do Carmo Valero, M.P.: Riemannian geometry. Birkhäuser (1992)
6. Donohue, M., Jacqmin-Gadda, H., Goff, M.L., Thomas, R., Raman, R., Gams, A., Beckett, L., Jack, C., Weiner, M., Dartigues, J.F., Aisen, P., the ADNI: Estimating long-term multivariate progression from short-term data. Alzheimer's & Dementia 10(5), 400–410 (2014)
7. Guerrero, R., Schmidt-Richberg, A., Ledig, C., Tong, T., Wolz, R., Rueckert, D., the ADNI: Instantiated mixed effects modeling of alzheimer's disease markers. Neuroimage (142), 113–125 (2016)
8. Hampel, H., O'Bryant, S.E., Durrleman, S., Younesi, E., Rojkova, K., Escott-Price, V., Corvol, J.C., Broich, K., Dubois, B., Lista, S., for the Alzheimer Precision Medicine Initiative: A precision medicine initiative for alzheimer's disease: the road ahead to biomarker-guided integrative disease modeling. Climacteric 20(2), 107–118 (2017)
9. Iturria-Medina, Y., Sotero, R., Toussaint, P., Mateos-Pérez, J., Evans, A., the ADNI: Early role of vascular dysregulation on late-onset alzheimer's disease based on multifactorial data-driven analysis. Nat. Commun. (7), 11934 (2016)
10. Kuhn, E., Lavielle, M.: Coupling a stochastic approximation version of em with an mcmc procedure. ESAIM: Probability and Statistics 8, 115–131 (2004)
11. Kuhn, E., Lavielle, M.: Maximum likelihood estimation in nonlinear mixed effects models. Computational Statistics & Data Analysis 49(4), 1020–1038 (2005)
12. Lee, J.M.: Smooth manifolds. In: Introduction to Smooth Manifolds, pp. 1–29. Springer (2003)
13. Raj, A., Kuceyeski, A., Weiner, M.: A network diffusion model of disease progression in dementia. Neuron pp. 1204–1215 (2012)
14. Reuter, M., Schmandsky, N., Rosas, H., Fischl, B.: Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 61(4), 1402–1418 (2012)
15. Schiratti, J.B., Allassonnière, S., Colliot, O., Durrleman, S.: Learning spatiotemporal trajectories from manifold-valued longitudinal data. In: Advances in Neural Information Processing Systems. pp. 2404–2412 (2015)
16. Young, A.L., Oxtoby, N.P., Huang, J., Marinescu, R.V., Daga, P., Cash, D., Fox, N., Ourselin, S., Schott, J.M., Alexander, D.C., the ADNI: Multiple orderings of events in disease progression. In: Information Processing in Medical Imaging. pp. 711–722. Springer (2015)