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 at 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 cortical atrophy propagation during the course of Alzheimer's Disease. Model parameters show the variability of this average scheme of atrophy in terms of age at atrophy onset, pace of propagation and trajectories of propagation across brain regions. It provides a description of the patterns of cortical atrophy at the individual level, paving the way to individual predictions.

1 Introduction

Intensive investigations have been conducted to understand the progression of Alzheimer's Disease (AD) especially before the clinical symptoms. 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 remains unclear at the group level and even more at the individual level. Personalized models of lesion propagation would enable to relate structural or metabolic alterations to the disease 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 dementia, *i.e.* a description of the spatial modifications of the brain over time.

Statistical models are well suited to estimate distributions of spatiotemporal patterns of propagation out of series of short-term longitudinal observations. However, the absence of time correspondence between the patients is a clear obstacle for these types of approaches. Using data series of several individuals requires to reposition 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, [4, 11], without taking into account the dynamics of changes, or develop average scenarios [5, 6]. Recently, a generic approach to align patients has been proposed in [10] for unstructured data: 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. It includes the cortical thickness distributed on the mesh of the pial surface or SUVR distributed on a regular voxel grid associated to 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 pattern of signal changes across neighbouring nodes, highlighting a propagation pattern across the network as in [9]. Extending the model in [10] for data distributed on a network is not straightforward, as the number of model parameters, defined at each node of the network, may explode with the resolution of the mesh. At infinite resolution, the parameters take the form of a smooth continuous map defined on the image mesh. In this paper, we propose to constrain these maps to belong to a finite-dimensional RKHS to ensure smooth spatial variation of model parameters. In practice, these maps are generated by the convolution of parameter values at a sparse set of control nodes on the network. 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 [10]. We introduce new parameters to account for smooth changes in the profiles of changes at neighbouring spatial locations.

To this end, we introduce a mixed-effect generative model that specifies a spatial distribution of the propagation, allowing to learn a group-average spatiotemporal trajectory from series of repeated observations. The model evaluates individual parameters (time reparametrization and spatial adjustment of the propagation) that enables the reconstruction of individual disease propagation through time. This highly non-linear problem is tackled by a stochastic version of the EM algorithm, specifically the MCMC-SAEM, introduced in [1, 7], with 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 a population of MCI converters from the ADNI database.

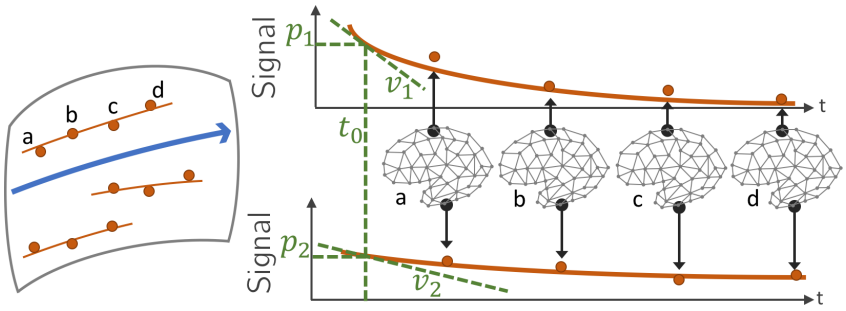


Fig. 1: Manifold representation of the mesh observations (left). Orange dots are patient real observations. The blue line is the reconstruction of the mean propagation. The signal value at each node (right), parametrized by (p, t, v), allows the reconstruction of the propagation over the network (orange 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 $\left((\mathbf{y}_{i,j})_1, ..., (\mathbf{y}_{i,j})_{N_v}\right)$ referred to as a *measurement map*.

Manifold-valued measurements distributed on a fixed graph
Let $d \in \{2, 3\}$ and $\mathcal{V} = (\mathbf{x}_1, ..., \mathbf{x}_{N_v})$ be a set of N_v distinct points in $\mathbb{R}^d$. The elements of $\mathcal{V}$ are called *vertices*. Let $\mathcal{E}$ be a subset of pairs of vertices which is assumed to be symmetric : $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{E}$ if and only if $(\mathbf{x}_j, \mathbf{x}_j) \in \mathcal{E}$. The pair $(\mathcal{V}, \mathcal{E})$ forms a *non-oriented graph* in $\mathbb{R}^d$. We assume that there exists a common fixed-graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = (\mathbf{x}_1, ..., \mathbf{x}_{N_v})$, such that for each $k \in \{1, .., N_v\}$, the coordinate $(\mathbf{y}_{i,j})_k$ is a measurement at node k. As the graph corresponds to spatial related measurements distributed on a mesh, the edges embed a spatial configuration. Therefore any edge $(\mathbf{x}_i, \mathbf{x}_j)$ is valued with respect to the geodesic distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ defined on the underlying mesh. Given a graph distance d, we define 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. The collection $\{(\mathcal{G}, \mathcal{D}, \mathbf{y}_{i,j})_{1\leq i\leq p,\, 1\leq j\leq k_i}\}$ is considered as a family of networks, distributed on the same underlying graph $\mathcal{G}$, associated to a unique spatial distribution defined by the matrix distance $\mathcal{D}$.

We assume 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). Therefore, the space of measurements is best described as a *Riemannian manifold* [3,8]. We assume that there exists a one-dimensional geodesically complete Riemannian manifold (M, g^M) such that, for all i, j, the observation $\mathbf{y}_{i,j}$ is a point in the product manifold M^{N_v}, equipped with a Rie-

mannnian product metric. It follows that for each i, j, $(\mathcal{G}, \mathcal{D}, \mathbf{y}_{i,j})$ is a manifold-valued network. It leads to consider any geodesic γ on M as the product of one-dimensional geodesics $\gamma(t) = (\gamma_1(t), ..., \gamma_{N_v}(t))$ where $\gamma_k(t)$ encodes for the propagation of the signal at the kth node of the network.

Smoothness of the propagation
For network models, we need to ensure that the progression of the signal complies with the temporal and spatial smoothness of the signal propagation. We expect the signal to be continuous at each node. The temporal constraint is handled by the continuous form of the one-dimensional geodesic $t \mapsto \gamma_k(t)$. On the other hand, we expect the signal to be similar for neighbour nodes. Moreover, we consider that each node k is described by N_p parameters $(p_1^k, ..., p_{N_p}^k)$ 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 sub-set of uniformly distributed nodes $\mathcal{V}_C = (\mathbf{x}_{d_1}, ..., \mathbf{x}_{d_{N_c}})$, called control nodes. They define a parameter evaluation $p_j(\mathbf{x})$ encoding for all the nodes $(p_j^k)_{1 \leq k \leq N_v}$: for all $\mathbf{x} \in \mathcal{V}$, $p_j(\mathbf{x}) = \sum_{i=1}^{N_C} K(\mathbf{x}, \mathbf{x}_{d_i}) \beta_j^i$ and for all $i \in \{1, \ldots, N_c\}$, $p_j(\mathbf{x}_{d_i}) = p_j^{d_i}$ where K is a Gaussian Kernel and the $(\beta_j^i)_{1 \leq i \leq N_c, 1 \leq j \leq N_p}$ are the new model parameters. 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 collection of manifold-valued networks $(\mathcal{G}, \mathcal{D}, \mathbf{y})$, we wish to model the propagation of a signal through the vertices of the common fixed-graph $\mathcal{G}$. In [10], the authors introduced a hierarchical model, to learn trajectories of changes from manifold-valued longitudinal observations. The model describes a group-average trajectory in the space of measurements, defined by a geodesic $\boldsymbol{\gamma}$ on a geodesically complete Riemannian manifold $(\mathbb{M}, g^{\mathbb{M}})$, that allows to estimate a typical scenario of progression. Individual trajectories derive from the group-average scenario through spatiotemporal transformations: the *parallel shifting* and the *time reparametrization*.

First, to describe disease pace and onset specific of 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 average disease onset related to the mean trajectory. The parameter τ_i corresponds to the time-shift between the mean and the individual disease 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 align the patients on the same disease time-line in order to reconstruct a group-average propagation.

The *parallel shifting* is handled by a family of individual tangent vector $(\mathbf{w}_i)_{1\leq i\leq p}$, called space-shifts. It encodes the modification of the mean signal propagation on the network. The spatial and temporal parametrization allow to generate an individual scenario of lesion propagation. 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 parameters $(\alpha_i, \tau_i, \mathbf{w}_i)$ allow to reconstruct the individual trajectories (orange lines) 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 model defines a mixed-effect model

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

Parameters estimation with the MCMC-SAEM algorithm

To reconstruct the long-term scenario of the disease propagation, we would like to estimate the parameters of the group-average trajectory $\boldsymbol{\theta} = ((\beta_j^i)_{1\leq i\leq N_c, 1\leq j\leq N_p}, \sigma)$ using a maximum likelihood estimator. The random-effects $\mathbf{z} = (z_i)_{1\leq i\leq p} = (\mathbf{w}_i, \alpha_i, \tau_i)_{1\leq i\leq p}$ are considered as latent variables, whose distributions characterize the variability of the individual trajectories. Due to the non-linearity of the equation (1), we use a Stochastic Approximation Expectation Maximization coupled with a Monte-Carlo Markov Chain sampler (MCMC-SAEM) introduced in [1,7]. It alternates between a simulation step, a stochastic approximation step and a maximization step until convergence.

We denote $\theta^{(k)}$ the current estimation of the parameters and $\mathbf{z}^{(k)}$ the current iterate of the Markov chain of the latent variables. The simulation uses an adaptive version [2] of the Hasting Metropolis within Gibbs sampler to draw $\mathbf{z}^{(k+1)}$ from $(\mathbf{z}^{(k)}, \mathbf{y}, \boldsymbol{\theta}^{(k)})$. As we consider models in the exponential family, for which the convergence of the algorithm has been proven, the second step corresponds to a stochastic approximation of the sufficient statistics of the model. The maximization step is straightforward given this stochastic approximation.

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 $(u, v) \in \mathrm{T}_p M$, $g_p(u, v) = uv/p^2$. With this metric and given $k \in \{1, \ldots, N_v\}$, M is a geodesically complete Riemannian manifold whose geodesics are of the form $t \mapsto p_k \exp(\frac{v_k}{p_k}(t - t_k))$ where $p_k \in M$, $t_k \in \mathbb{R}, v_k \in \mathrm{T}_{p_k} M$. These parameters are represented on Figure 1 (right) at two nodes where the signal decrease through time vary 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 2 and the fact that the parameters (p_j^k) are (p_k, v_k), it leads to define $p(\mathbf{x}) = \sum_{i=1}^{N_C} K(\mathbf{x}, \mathbf{x}_{d_i})\beta_p^i$ and $v(\mathbf{x}) = \sum_{i=1}^{N_C} K(\mathbf{x}, \mathbf{x}_{d_i})\beta_v^i$

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 $\boldsymbol{\theta} = (t_0, (\beta_p^i)_{1\leq i\leq N_c}, (\beta_v^i)_{1\leq i\leq N_c}, \sigma)$ and $\mathbf{z} = (\mathbf{w}_i, \alpha_i, \tau_i)_{1\leq i\leq p}$

4 Experimental results

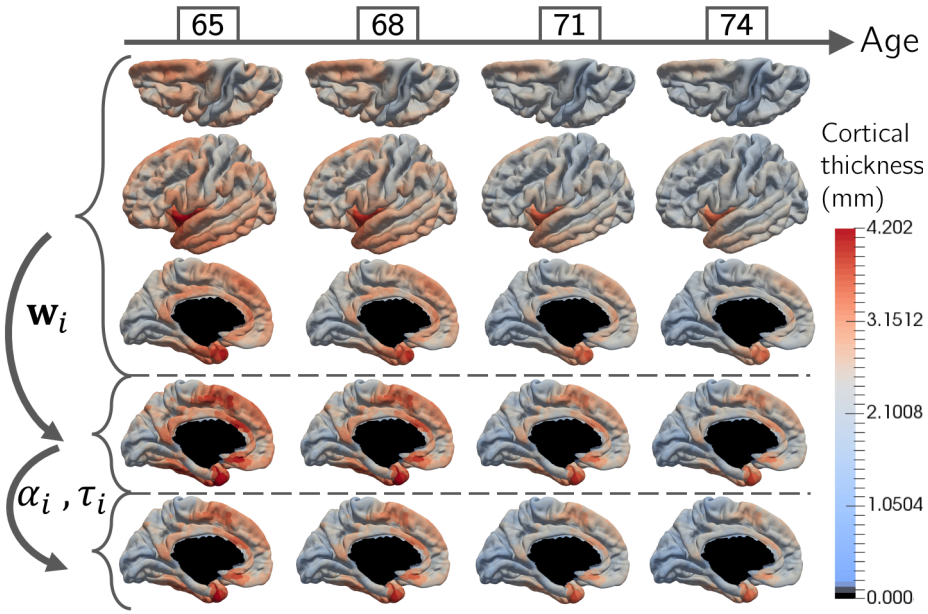


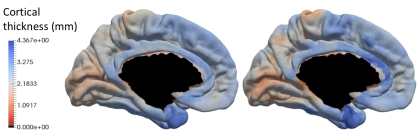
Fig. 2: Cortical thickness at 65, 68, 71 and 74 years old of the mean propagation (first lines). Effect of the space-shift $\mathbf{w}_i$ (fourth line), then with temporal reparametrization α_i, τ_i (fifth line) on the cortical thickness.

Data

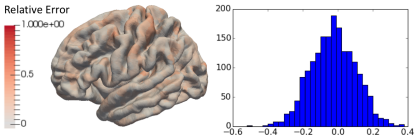
We used this model to highlight typical spatiotemporal patterns of cortical atrophy during the course of Alzheimer's Disease from longitudinal MRI of MCI converters from the ADNI database. This 154 MCI converters correspond to 787 observations, each subject being observed 5 times on average. We aligned the measures on a common atlas with FreeSurfer [4] in order to distribute the measurement maps on the same common fixed-graph $\mathcal{G}$. The later is constituted

[4] Software available here : `http://surfer.nmr.mgh.harvard.edu`

of 1827 nodes that map entirely the surface of the brain left hemisphere, more precisely its pial surface. Out of these 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 $\mathcal{G}$.



(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

Cortical thickness measurements

We used the model instantiation defined in 3 to characterize the cortical thickness decrease. Multiple runs of 30.000 iterations (∼4hours) of this MCMC-SAEM lead to a noise standard deviation $\sigma \simeq 0.27$ with 90% of the data included in $[1.5, 3.6]$ mm. The mean spatiotemporal propagation, described on the first three lines of the Figure 2 as the cortical thickness at respectively 65, 68, 71 and 74 years old shows that the most affected area is the medial-temporal lobe, followed by the temporal neocortex. The parietal association cortex and the frontal lobe are also 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 highly 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 lines 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 a real observation of the cortical thickness and the reconstruction done by the model with the individual parameters $(\mathbf{w}_i, \alpha_i, \tau_i)$. The relative error and its histogram are represented on Figure 3b. It should be mentioned that the input data are noisy due to the acquisition and pre-processes, and, due to the alignment of all the individuals on the same network. The temporal and spatial smoothness introduced by the control nodes and the form of the geodesics smooth the signal propagation over the surface.

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 lon-

gitudinal 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: we were able to highlight areas affected by considerable neuronal loss such as 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. 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.
2. Y. F. Atchadé. 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. M. P. Do Carmo Valero. *Riemannian geometry*. Birkhäuser, 1992.
4. M.C. Donohue, H. Jacqmin-Gadda, M. Le Goff, R.G. Thomas, R. Raman, A.C. Gams, L.A. Beckett, and C.R. Jack et al. Estimating long-term multivariate progression from short-term data. *Alzheimer's & Dementia*, 10(5):400–410, 2014.
5. R Guerrero et al. Instantiated mixed effects modeling of alzheimer's disease markers. *Neuroimage*, (142):113–125, 2016.
6. Y. Iturria-Medina, R.C. Sotero, P.J. Toussaint, J.M. Mateos-Pérez, A.C. Evans, and the ADNI. Early role of vascular dysregulation on late-onset alzheimer's disease based on multifactorial data-driven analysis. *Nature Com.*, (7):11934, 2016.
7. E. Kuhn and M. Lavielle. Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, 49(4):1020–1038, 2005.
8. J. M. Lee. Smooth manifolds. In *Introduction to Smooth Manifolds*, pages 1–29. Springer, 2003.
9. A. Raj, A. Kuceyeski, and M. Weiner. A network diffusion model of disease progression in dementia. *Neuron*, pages 1204—-1215, 2012.
10. J.-B. Schiratti, S. Allassonnière, O. Colliot, and S. Durrleman. Learning spatiotemporal trajectories from manifold-valued longitudinal data. In *Advances in Neural Information Processing Systems*, pages 2404–2412, 2015.
11. A. L. Young et al. Multiple orderings of events in disease progression. In *IPMI*, pages 711–722. Springer, 2015.