A coherent framework for learning spatiotemporal piecewise- geodesic trajectories from longitudinal manifold-valued data

Juliette Chevallier, Vianney Debavelaere, Stéphanie Allassonnière

Noname manuscript No.
(will be inserted by the editor)

A Coherent Framework for Learning Spatiotemporal Piecewise-Geodesic Trajectories from Longitudinal Manifold-Valued Data

Juliette Chevallier · Vianney Debavelaere · Stépahnie Allassonnière

Received: date / Accepted: date

Abstract This paper provides a coherent framework for studying longitudinal manifold-valued data. We introduce a Bayesian mixed effects model which allows to estimate both a group-representative piecewise-geodesic trajectory in the Riemannian space of shape and inter-individual variability. We prove the existence of the maximum a posteriori estimate and its asymptotic consistency under reasonable assumptions. Due to the non-linearity of the proposed model, we use a stochastic version of the expectation maximization algorithm to estimate the model parameters. Our simulations show that our model is not noise-sensitive and succeed in explaining various paths of progression.

Keywords Bayesian estimation · EM like algorithm · Longitudinal data · MCMC methods · Nonlinear mixed-effects model · Spatiotemporal analysis

1 Introduction

Longitudinal studies are powerful tools to achieve a better understanding of temporal progressions of biological or natural phenomenons. For instance, longitudinal psychometric data are often used to explore differences in the progression of Alzheimer's and more generally neurodegenerative diseases. Other important applications such as pattern recognition, disease and treatment monitoring, study of face expression dynamics, *etc.* come also from longitudinal studies.

Moreover, efforts in medicine and medical follow-up rely more and more on the understanding of the global disease progression and not only on punctual states of health, often with the help of medical images. In order to provide everyone the best possible treatment, there is a need for prediction methods that allow to grasp quickly the efficiency of a possible treatment without doing invasive measures, such as biopsies. Hence, designing models that deal with medical images and more generally extracted features and shapes from this images is very important for application-related uses.

Reaction-diffusion based tumor growth models have demonstrated their efficiency for cancer monitoring [17, 25]. However, such methods cover images but not shapes or whatever else type of data, e.g. scores. Moreover, even for images, as these models rely on reaction-diffusion equations, they can only apply for situations in which the observed phenomenon is linked to diffusion dynamics. As a consequence, these models can apply to untreated cancer but not to treated-one nor to multiple sclerosis monitoring, neurodegenerative disease follow-up, like Alzheimer's or Parkinson's diseases, or more complicated framework. Likewise, coupling nonlinear partial differential equation (PDE) models and optimization is efficient to make prediction when patients are only under monitoring, *i.e.* without treatment [9, 30]. The idea of these methods is to model the tumor growth with a "simple" PDE model, involving few parameters which are estimated from series of CT-scan or MRI. For therapy planing purpose, such methods can reinforce the decision to wait without specific treatment, in case of slow progression of the tumor, thus preventing heavy treatment. However, they cannot predict the response to a given molecule and so help in the choice of the appropriate one, nor in the choice of an appropriate sequence of molecules.

Anatomical data – and most of structured data – have been successfully modeled as points on a Riemannian manifold, *i.e.* as points on a smooth manifold equipped with a Riemannian metric tensor. These spaces are often called shape spaces. The choice of the shape space and its metric

Juliette Chevallier
CMAP, École polytechnique, Palaiseau, France
E-mail: juliette.chevallier@polytechnique.edu

Vianney Debavelaere
CMAP, École polytechnique, Palaiseau, France

Stépahnie Allassonnière
CRC, Université Paris Descartes, Paris, France

tensor is driven by the type of data in the study [6, 7, 16, 36, 37]. Geometrical properties of shape manifolds have been properly defined over the last decades. Moreover, according to the Whitney embedding theorem [13], as the shape spaces are second-countable, they are always embedded in a real d-dimensional Euclidean space, the space of measurements, which leads us to consider the shape manifold as a submanifold of this Euclidean space. Therefore, the temporal evolution of empirical data may be modeled as a parametric curve in the space of measurements and more precisely as a noisy version of an underlying parametric curve living on the Riemannian shape submanifold. Given a cohort of individuals followed over a time period, we thus observe discreet samples of such a curve for each subject. We call this set of observations a longitudinal data set. Figure 1 illustrates this understanding of the data.

Mixed effects models have proved their efficiency in the study of longitudinal data sets [20], especially for medical purposes [24, 26]. Indeed, mixed effects models provide a general and flexible framework to study complex data which depends on unobserved variables, such as longitudinal data sets. They consist of two parts: fixed effects which describe the data at the population level and random effects which are associated with individual experimental units. In the framework of longitudinal analysis, through mixed effects models, one can explain in two steps, both a representative path of the evolution of the whole population and individual-specific progressions. Given a longitudinal data set, a representative trajectory and its variability are first estimated. Then, we can define subject-specific trajectories according to global progression.

The temporal alignment in longitudinal data analysis is an efficient way to compare trajectories [33, 34]. Here, the authors propose to use the temporal registration to align the different trajectories. Despite good results for comparing trajectories, the interpretation of the temporal parameterization is lost. However, within medical applications, the time parametrization reveals information on the data patient's state of health and has to be considered.

The recent generic approach of Schiratti et al. [31, 32] to align patients takes all this into account. This model was built with the aim of granting temporal and spatial inter-subject variability through individual variations of a common time-line and parallel shifting of a representative trajectory. Each individual trajectory has its own intrinsic geometric pattern through spacial variability and its own time parametrization through time variability. In term of modeling, the time variability allows some individuals to follow the same progression path but at a different age and with possibly a different pace. Schiratti et al. [31, 32] have made a strong hypothesis to build their model as they assume the characteristic evolution to be geodesic, *i.e.* that the characteristic trajectory is the shortest path between the initial representative state and the final one. This hypothesis helps for the parametrization of the model which becomes both generic, *i.e.* allowing for many different types of data, and numerically estimable. However, such an assumption significantly reduces the effective framework of their model. Like the described-above PDEs based models, such a model can be applied to situation with a unique dynamic like neurodegenerative disease, but not to situations in which the dynamic can fluctuate. For instance, this model cannot be used for multiple sclerosis monitoring in which the disease progression is accompanied by recession nor for monitoring tumor regression or recurrence in response to treatments.

In this paper, we will relax this assumption to make the model applicable to a wider variety of situations and data sets: we address each situation in which the evolution can fluctuate several times.

We propose in this paper a coherent and generic statistical framework which includes the model of Schiratti et al. [31, 32]. Following their approach, we define a nonlinear mixed effects model for the definition and estimation of spatiotemporal piecewise geodesic trajectories from longitudinal manifold-valued data. We estimate a representative piecewise geodesic trajectory of the global progression and together with spacial and temporal inter-individual variability. Particular attention is paid to estimation of the correlation between the different phases of the evolution.

Estimation is formulated as a well-defined *Maximum A Posteriori* (MAP) problem which we prove to be consistent under mild assumptions. Numerically, the MAP estimation of the parameters is performed through the expectation-maximization (EM) algorithm [11]. However, as our model is strongly nonlinear and we have to use a stochastic version of the the EM algorithm, namely the Markov chain Monte Carlo stochastic approximation expectation maximization (MCMC-SAEM) algorithm [21]. Theoretical results regarding its convergence have been proved by Delyon et al. [10] and Allassonnière et al. [2] and its numerical efficiency has been demonstrated for these types of models ([31, 32], MONOLIX – `http://lixoft.com/`).

Due to the versatility of the Riemannian geometry, the proposed model provides a comprehensive support for a wide range of practical situations, from unidimensional data to shape analysis. Moreover, the same algorithm can be used in all these situations.

The paper is organized as follows: in Section 2 we define our generic nonlinear mixed effects model for piecewise geodesically distributed data. Riemannian geometry allows us to derive a method that makes light assumptions about the data and applications we are able to deal with. In Section 3, we explain how to use the MCMC-SAEM algorithm to produce MAP estimates of the parameters. We also prove a consistency theorem, whose proof is postponed in appendix A. We then make the generic formulation ex-

plicit for one-dimension manifolds and piecewise logistically distributed data in Section 4.1 and for shape analysis in Section 4.2. These two particular cases are built in the target of chemotherapy monitoring. In Section 5, some experiments are performed for the piecewise logistic model and for the piecewise geodesic shape model: both on synthetic and on real data from the Hôpital Européen Georges Pompidou (HEGP, Georges Pompidou European Hospital) for the piecewise logistic model. These experiments highlight the robustness of our model to noise and its performance in understanding individual paths of progression.

2 Generic mixed effects Model for Piecewise-Geodesically Distributed Data on a Riemannian Manifold

In the following, we describe a *generic* method to build mixed effects models for piecewise-geodesically distributed data. This leads us to a large variety of possible situations that we will be able to deal with within the same framework. This model has been first introduced in Chevallier et al. [8].

We consider a longitudinal data set $\boldsymbol{y}$ obtained by repeating multivariate measurements of $n \in \mathbb{N}^*$ individuals. Each individual $i \in [\![1,n]\!]$ is observed $k_i \in \mathbb{N}^*$ times, at the time points $\boldsymbol{t}_i = (t_{i,j})_{j\in[\![1,k_i]\!]}$, and we denote $\boldsymbol{y}_i = (y_{i,j})_{j\in[\![1,k_i]\!]}$ the sequence of observations for this individual. We also denote $k = \sum_{i=1}^{n} k_i$ the total numbers of observations and assume that each observation $y_{i,j}$ is a point of $\mathbb{R}^d$, where $d \in \mathbb{N}$ can be considered as the dimension of the problem. Thus, our observed data consists in a sequence in $\mathbb{R}^{kd}$, $\boldsymbol{y} = (y_{i,j})_{(i,j)\in[\![1,n]\!]\times[\![1,k_i]\!]}$, where $[\![1,n]\!] \times [\![1,k_i]\!]$ denotes for compactness the set $\{(i,j) | i \in [\![1,n]\!] \wedge j \in [\![1,k_i]\!]\}$.

We generalize the idea of Schiratti et al. [31, 32] and build our model in a hierarchical way. Our data points are seen as noisy samples along trajectories and we suppose that each individual trajectory derives from a group-representative scenario through spatiotemporal transformations. Key to our model is that the group-representative trajectory in no longer assumed to be geodesic but piecewise-geodesic. Thus, the characteristic trajectory is no more the shortest path between the initial and the final representative states but a concatenation of shortest paths between several intermediate states. In particular, this allows us to consider situations in which the evolution can fluctuate. We present at Figure 1 an example of situation our generic model can address.

To ensure that the optimization of those trajectories can be computationally performed in a reasonable amount of time, we build a parametric model. That is to say that the trajectories depend on a finite number of variables. In the following (see Subsection 2.3), we will denote $\boldsymbol{z}_{\text{pop}}$ the variables driving the group-representative scenario and $\boldsymbol{z}_i$ those associated to the individual $i \in [\![1,n]\!]$. For the sake of clarity,

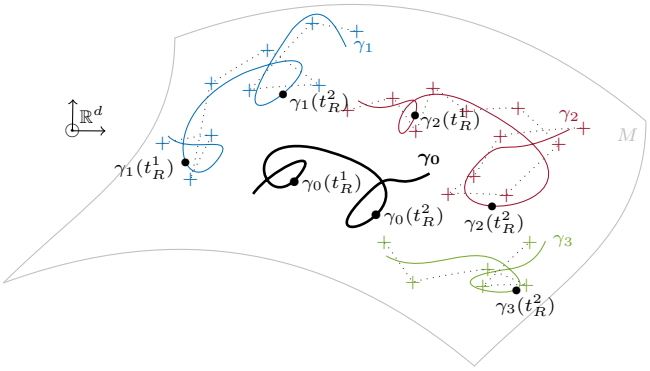


Fig. 1 *The generic piecewise geodesic curve model*. The observed data (crosses) consist in noisy samples along manifold-valued trajectories. Each individual path γ_i (solid colored lines) is a spatiotemporal variation of a piecewise geodesic representative trajectory γ_0 (bold black line). In particular, the individual trajectories are not necessarily piecewise geodesic.

we first detail the construction of the trajectories from a geometrical point of view. Then, we state our generative model in a statistical perspective.

2.1 The Group-Representative Trajectory

Let $m \in \mathbb{N}^*$ and $\boldsymbol{t_R} = \left(-\infty < t_R^1 < \ldots < t_R^{m-1} < +\infty\right)$ a subdivision of $\mathbb{R}$, called the *breaking-up times* sequence. In order the representative trajectory γ_0 to be geodesic on each of the m sub-intervals of $\boldsymbol{t_R}$, we build γ_0 component by component.

2.1.1 A Piecewise-Geodesic Curve

In this context, let M_0 be a geodesically complete submanifold of $\mathbb{R}^d$, $\left(\bar{\gamma}_0^\ell\right)_{\ell\in[\![1,m]\!]}$ a family of geodesics on M_0 and $\left(\phi_0^\ell\right)_{\ell\in[\![1,m]\!]}$ a family of isometries defined on M_0. For all $\ell \in [\![1,m]\!]$, we set $M_0^\ell = \phi_0^\ell(M_0)$ and $\gamma_0^\ell = \phi_0^\ell \circ \bar{\gamma}_0{}^\ell$. The isometric nature of the mapping ϕ_0^ℓ ensures that the manifolds $(M_0^\ell)_{\ell\in[\![1,m]\!]}$ remain Riemannian and that the curves $\gamma_0^\ell \colon \mathbb{R} \to M_0^\ell$ remain geodesic. In particular, each γ_0^ℓ remains parametrizable [13]. We define the representative trajectory γ_0 by

$$\begin{aligned}\forall t \in \mathbb{R}, \quad \gamma_0(t) \;=\; & \gamma_0^1(t)\,\mathbb{1}_{]-\infty,t_R^1]}(t) \\ & + \sum_{\ell=2}^{m-1} \gamma_0^\ell(t)\,\mathbb{1}_{]t_R^{\ell-1},t_R^\ell]}(t) + \gamma_0^m(t)\,\mathbb{1}_{]t_R^{m-1},+\infty[}(t)\,.\end{aligned}$$

In other words, given a *manifold-template* of the geodesic components M_0, we build γ_0 so that the restriction of γ_0 to each sub-interval of $\boldsymbol{t_R}$ is the deformation of a geodesic curve $\bar{\gamma}_0^\ell$ living on M_0 by the corresponding isometry ϕ_0^ℓ. In practice, M_0 is chosen in order to catch the geometric nature of the observed data : if we are studying a score as in

Section 4.1, M_0 will be the standard finite segment $]0,1[$ for instance. The choice of the isometries ϕ_0^ℓ and the geodesics $\bar{\gamma}_0^\ell$ have to be done with the aim of having an "as regular as possible" (at least continuous) curve γ_0 at the breaking-up time points t_R^ℓ. In the following section, we propose a way to meet this criterion in one dimension and in the shape framework. However, the freedoms in the choice of ϕ_0^ℓ and $\bar{\gamma}_0^\ell$ induce a wide panel of models.

2.1.2 Boundary Conditions

Because of the piecewise nature of our representative trajectory γ_0, constraints have to be formulated on each interval of the subdivision $\boldsymbol{t_R}$. Following the formulation of the *local existence and uniqueness theorem* [13], constraints on geodesics are generally formulated by forcing a value and a tangent vector at a given time-point. However, as soon as there is more that one geodesic component, *i.e.* $m > 1$, such an approach cannot ensure the curve γ_0 to be at least continuous. That is why we re-formulate these constraints in our model as boundary conditions. Let $\bar{\boldsymbol{A}} = (\bar{A}^0, \dots, \bar{A}^m) \in (M_0)^{m+1}$. Let $t_0 \in \mathbb{R}$ be a real value representing an initial time and $t_1 \in \mathbb{R}$ representing a final one. We impose that $\bar{\gamma}_0^1(t_0) = \bar{A}^0$, $\bar{\gamma}_0^m(t_1) = \bar{A}^m$ and that

$$\forall \ell \in [\![1, m-1]\!], \quad \bar{\gamma}_0^\ell(t_R^\ell) = \bar{A}^\ell \quad \text{and} \quad \bar{\gamma}_0^{\ell+1}(t_R^\ell) = \bar{A}^\ell \,.$$

Note that we can apply the constraints on γ_0^ℓ instead of $\bar{\gamma}_0^\ell$ by defining $A^\ell = \phi_0^\ell(\bar{A}^\ell)$ for each ℓ. Notably, the $2m$ constraints are defined step by step. In the case where the geodesics could be written explicitly, such constraints do not complicate the model. In more complicated case, as the one shown for shapes in Section 4.2, we use shooting or matching methods to enforce this constraints.

From this representative curve, we derive a modeling of the individual trajectories that mimics the individual evolution of subjects and best fits the individual observations.

2.2 Individual Trajectories: Space and Time Warping

We want the individual trajectories to represent a wide variety of behaviors and to derive from the group characteristic path by spatiotemporal transformations. To do that, we define for each component of the piecewise-geodesic curve γ_0 a couple of transformations: the *diffeomorphic component deformations* and the *time component reparametrizations* which characterize respectively the spatial and the temporal variability of propagation among the population. Moreover, we decree as few constraints as possible in the construction: at least continuity and control of the slopes at the (individual) breaking-up points.

2.2.1 Time Component Reparametrizations

For compactness, we denote t_0 by t_R^0 from now on.

To allow different paces in the progression and different rupture times for each individual, we introduce some temporal transformations $\psi_i^\ell \colon \mathbb{R} \to \mathbb{R}$, called *time-warp*, that are defined for the subject $i \in [\![1,n]\!]$ and for the geodesic component $\ell \in [\![1,m]\!]$ by

$$\psi_i^\ell(t) = \psi_{i\,(\alpha_i^\ell, \tau_i^\ell)}^\ell(t) = \alpha_i^\ell(t - t_R^{\ell-1} - \tau_i^\ell) + t_R^{\ell-1} \,,$$

where $(\alpha_i^\ell, \tau_i^\ell) \in \mathbb{R}^+ \times \mathbb{R}$. The parameters τ_i^ℓ correspond to the time-shifts between the representative and the individual progression onset ; the α_i^ℓ are the acceleration factors that describe the pace of individuals, being faster or slower than the population characteristic. For all individuals $i \in [\![1,n]\!]$, let $\boldsymbol{t_{R,i}} = (t_{R,i}^\ell)_{\ell \in [\![1,m-1]\!]}$ denote the individual sequence of rupture times which is the subdivision of $\mathbb{R}$ such that for all $\ell \in [\![1,m-1]\!]$, $\psi_i^\ell(t_{R,i}^\ell) = t_R^\ell$ *i.e.* such that

$$t_{R,i}^\ell = t_{R,i\,(\alpha_i^\ell, \tau_i^\ell)}^\ell = t_R^{\ell-1} + \tau_i^\ell + \frac{t_R^\ell - t_R^{\ell-1}}{\alpha_i^\ell} \,.$$

To ensure good adjunction at the rupture times, we demand that for all $\ell \in [\![1,m]\!]$, $\psi_i^\ell(t_{R,i}^{\ell-1}) = t_R^{\ell-1}$. Hence the time reparametrizations are constrained and only the acceleration factors α_i^ℓ and the first time shift τ_i^1 are free: all other time shift, $\ell \in [\![2,m]\!]$, are defined by $\tau_i^\ell = t_{R,i}^{\ell-1} - t_R^{\ell-1}$.

In the following, we will sometimes refer to the individual initial and final times which are defined, for all $i \in [\![1,n]\!]$, by $t_0^i = t_0 + \tau_1$ and $t_1^i = t_R^{m-1} + \tau_i^m + \frac{t_1 - t_R^{m-1}}{\alpha_i^m}$.

2.2.2 Diffeomorphic Component Deformations

Concerning the space variability, we introduce m diffeomorphisms $\phi_i^\ell \colon M_0^\ell \to \phi_i^\ell(M_0^\ell)$ to enable the different components of the individual trajectories to vary irrespectively of each other. We enforce the adjunctions to be at least continuous and therefore the mappings ϕ_i^ℓ to satisfy

$$\forall \ell \in [\![1, m-1]\!], \quad \phi_i^\ell \circ \gamma_0^\ell(t_R^\ell) = \phi_i^{\ell+1} \circ \gamma_0^{\ell+1}(t_R^\ell) \,.$$

Note that, as the individual paths are no longer required to be geodesic, the mappings ϕ_i^ℓ do not need to be isometric.

For all individuals $i \in [\![1,n]\!]$ and all component $\ell \in [\![1,m]\!]$, we set $\gamma_i^\ell = \phi_i^\ell \circ \gamma_0^\ell \circ \psi_i^\ell$ and define the corresponding individual curve γ_i by

$$\begin{aligned} \forall t \in \mathbb{R}, \quad \gamma_i(t) \;=\; & \gamma_i^1(t)\, \mathbb{1}_{]-\infty, t_{R,i}^1]}(t) \\ & + \sum_{\ell=2}^{m-1} \gamma_i^\ell(t)\, \mathbb{1}_{]t_{R,i}^{\ell-1}, t_{R,i}^\ell]}(t) + \gamma_i^m(t)\, \mathbb{1}_{]t_{R,i}^{m-1}, +\infty[}(t) \,. \end{aligned}$$

Finally, the observations $\boldsymbol{y}_i = (y_{i,j})_{j\in[\![1,k_i]\!]}$ are assumed to be distributed along the curve γ_i and perturbed by an additive Gaussian noise $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(0,\sigma^2 I_{k_i d})$ where $\sigma \in \mathbb{R}^+$:

$$\forall (i,j) \in [\![1,n]\!]\times[\![1,k_i]\!], \quad y_{i,j} \;=\; \gamma_i(t_{i,j}) \;+\; \varepsilon_{i,j}\,,$$

where $\varepsilon_{i,j} \sim \mathcal{N}(0,\sigma^2 I_d)$. By construction, for each $(i,j) \in [\![1,n]\!]\times[\![1,k_i]\!]$, there exist $\ell \in [\![1,m]\!]$ such that $\gamma_i(t_{i,j})$ lies on the submanifold $\phi_i^\ell(M_0^\ell)$ of $\mathbb{R}^d$. Thus, the previous sum is well-defined. In particular, we do not assume that the noisy-observation remain on the manifold.

The choice of the isometries ϕ_0^ℓ and the diffeomorphisms ϕ_i^ℓ induces a large range of piecewise-geodesic models. For example, if $m=1$, $\phi_0^1 = Id$ and if ϕ_i^1 denotes the application that maps a curve onto its parallel curve for a given non-zero tangent vector $\boldsymbol{w}_i$, we feature the model proposed by Schiratti et al. [31, 32]. In Section 4, we propose two other specific models which can be used for chemotherapy monitoring, as a first example of fluctuate dynamic.

2.3 Toward a Coherent and Tractable Statistical Generative Model

Now that we have presented the geometrical objects we will deal with, we set up a comprehensive statistical framework in order to estimate the different parameters that control the trajectories.

We first introduce some notations in order to clearly state our statistical generative model. Let $\boldsymbol{z}_i^\psi = (\alpha_i^\ell, \tau_i^\ell)_{\ell\in[\![1,m]\!]}$ denote the individual temporal variables and similarly $\boldsymbol{z}_i^\phi$ denote the individual spatial variables, *i.e.* the variables associated to the variation of the m diffeomorphic deformations ϕ_i^ℓ. Likewise, let $\boldsymbol{z}_{\text{pop}}$ denote the population variable, *i.e.* the variable associated to the variation of the m isometric mappings ϕ_0^ℓ.

Let $p_{\text{ind}} \in \mathbb{N}$ be the dimension of each vector $\boldsymbol{z}_i = (\boldsymbol{z}_i^\psi, \boldsymbol{z}_i^\phi)$ such that $\forall i \in [\![1,n]\!]$, $\mathcal{Z}_i \subset \mathbb{R}^{p_{\text{ind}}}$ denotes the space of random effects. Similarly, let $p_{\text{pop}} \in \mathbb{N}$ be the dimension of $\boldsymbol{z}_{\text{pop}}$ and $\mathcal{Z}_{\text{pop}} \subset \mathbb{R}^{p_{\text{pop}}}$ denotes the space of fixed effects.

To cover many situations, we do not explicit here the individual spatial variables $\boldsymbol{z}_i^\phi$. However, for examples, we propose an instantiation of this generic model for one dimension manifolds and piecewise logistically distributed data at Section 4.1 and for shape analysis at Section 4.2. Moreover, our generic approach encompass a large variety of models as such proposed by Schiratti et al. [32], Bône et al. [5] and Koval et al. [18].

2.3.1 Modeling Constraints...

In a modeling perspective, we are interested in understanding the individual behaviors with respect to the characteristic one. Thus, we focus on the variance of the random effects $\boldsymbol{z}_i = (\boldsymbol{z}_i^\psi, \boldsymbol{z}_i^\phi)$ rather than their distributions. Moreover, as we want the representative path to characterize the pattern of behavior of the individual trajectories, we have to slightly modify the individual parameters $\boldsymbol{z}_i$ in such a way that for all i, $\mathbb{E}(\boldsymbol{z}_i) = 0$. In particular, if our model were linear, this would have ensure the representative trajectory to be the mean (in the statistical meaning) of the individual ones. Concerning the individual temporal variables for instance, the acceleration parameters $(\alpha_i^\ell)_{\ell\in[\![1,m]\!]}$ have to be positive and equal to one on average while the time shifts $(\tau_i^\ell)_{\ell\in[\![1,m]\!]}$ are of any signs and must be zero on average. For these reasons, we set $\alpha_i^\ell = \mathrm{e}^{\xi_i^\ell}$ and consider the "new" temporal variable, still denoted $\boldsymbol{z}_i^\psi$ for compactness, $\boldsymbol{z}_i^\psi = (\xi_i^\ell, \tau_i^\ell)_{\ell\in[\![1,m]\!]}$. We proceed in the same way for the individual spatial variables z_i^ϕ, when required (for centered or positive variables).

To sum up, we assume that there exists a symmetric positive definite matrix $\Sigma \in \mathscr{S}^+_{p_{\text{ind}}}(\mathbb{R})$ such that $\boldsymbol{z}_i \sim \mathcal{N}(0,\Sigma)$, and now want to estimate Σ. Hence, the parameters we are interested in are $\theta = (\boldsymbol{z}_{\text{pop}}, \Sigma, \sigma) \in \mathscr{Z}_{\text{pop}} \times \mathscr{S}^+_{p_{\text{ind}}}(\mathbb{R}) \times \mathbb{R}^+$.

2.3.2 ...and Computational Feasibility

Given a n-sample, we target $\hat{\theta}_n$ an estimation of our parameters. Following the classical approach for maximum likelihood estimation in nonlinear mixed effects models, we use the MCMC-SAEM algorithm. However, the theoretical convergence of this algorithm is proved only if the model belongs to the curved exponential family [2, 10]. This framework is also important for numerical performances. Without further hypothesis, our model does not satisfy this constraint. Therefore, we proceed as in Kuhn and Lavielle [19]: we assume that $\boldsymbol{z}_{\text{pop}}$ is the realization of independent Gaussian random variables with fixed small variances and estimate the means of those variables. So, the parameters we want to estimate are $\theta = \left(\overline{z_{\text{pop}}}, \Sigma, \sigma\right)$ and we define the set of admissible parameters by $\Theta = \mathbb{R}^{p_{\text{pop}}} \times \mathscr{S}^+_{p_{\text{ind}}}(\mathbb{R}) \times \mathbb{R}^+$.

The fixed and random effects $\boldsymbol{z} = (\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)_{i\in[\![1,n]\!]}$ are considered as latent variables, *i.e.* as hidden variables that are not directly observed but can be inferred by the observations. Our model writes in a hierarchical way as

$$\begin{cases} y \,|\, z, \theta \;\sim\; \displaystyle\bigotimes_{i=1}^{n}\bigotimes_{j=1}^{k_i} \mathcal{N}\left(\gamma_i(t_{i,j})\,,\, \sigma^2\right), \\[2ex] z \,|\, \theta \;\sim\; \mathcal{N}\left(\overline{z_{\text{pop}}}\,,\, D_{\text{pop}}^{-1}\right) \displaystyle\bigotimes_{i=1}^{n} \mathcal{N}(0,\Sigma)\,, \end{cases}$$

where $\boldsymbol{\sigma}_{\text{pop}} \in \mathbb{R}_+^{p_{\text{pop}}}$ is an hyperparameter of the model and $D_{\text{pop}} = \boldsymbol{\sigma}_{\text{pop}}^2 I_{p_{\text{pop}}} \in \mathscr{M}_{p_{\text{pop}}}(\mathbb{R})$ is the diagonal matrix of size p_{pop} whose diagonal entries are given by the vector $\boldsymbol{\sigma}_{\text{pop}}^2$. The products $\otimes$ mean that the corresponding entries are considered to be independent. In other words, we assume that each of the measurement noises is independent of all the others. Of course, it may not be the case in practice. But, as all the observations for a given subject come from a single

curve, this assumption is reasonable in our context. Moreover, this assumption leads us to a more computationally tractable algorithm.

3 Parameters Estimation

As said just above, we want to estimate $\theta = (\overline{z_{\text{pop}}}, \Sigma, \sigma) \in \mathbb{R}^{p_{\text{pop}}} \times \mathcal{S}^+_{p_{\text{ind}}}(\mathbb{R}) \times \mathbb{R}^+$. As we want our model to be consistent with low sample size high-dimensional data analysis, we consider a Bayesian framework, *i.e.* we assume the following priors

$$(\Sigma, \sigma) \sim \mathscr{W}^{-1}(V, m_\Sigma) \otimes \mathscr{W}^{-1}(v, m_\sigma)\,,$$

where $V \in \mathcal{S}^+_{p_{\text{ind}}}(\mathbb{R})$, $v, m_\Sigma, m_\sigma \in \mathbb{R}$ and where $\mathscr{W}^{-1}(V, m_\Sigma)$ denotes the inverse Wishart distribution with scale matrix V and degrees of freedom m_Σ. Regularization has indeed proven its fruitfulness in this context [14]. In order for the inverse Wishart to be non-degenerate, the degrees m_Σ and m_σ must satisfy $m_\Sigma > 2p_{\text{ind}}$ and $m_\sigma > 2$. In practice, we yet use degenerate priors but with well-defined posteriors. In the spirit of the one-dimension inverse Wishart distribution, we define the density function distribution of higher dimension as

$$f_{\mathscr{W}^{-1}(V,m_\Sigma)}(\Sigma) = \frac{1}{\Gamma_{p_{\text{ind}}}\left(\frac{m_\Sigma}{2}\right)} \left(\frac{\sqrt{|V|}}{2^{\frac{p_{\text{ind}}}{2}} \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2} \operatorname{tr}\left(V\Sigma^{-1}\right) \right) \right)^{m_\Sigma},$$

where $\Gamma_{p_{\text{ind}}}$ is the multivariate gamma function and, for all matrices A, $|A|$ denotes the determinant of the matrix A.

The estimates are obtained by maximizing the posterior density on θ conditionally on the observations $\boldsymbol{y} = (y_{i,j})_{(i,j)\in[\![1,n]\!]\times[\![1,k_i]\!]}$.

In the following paragraphs, we first show that the model is well-posed *i.e.* that for any finite sample the *maximum* we are looking for exists. We then prove a consistency theorem which ensures that the set of parameters which well-explain the observations is non-empty and that the MAP estimator converges to this set. Last, we explain how to use the MCMC-SAEM algorithm to produce MAP estimates.

3.1 Existence of the Maximum a Posteriori Estimator

The inverse Wishart priors on the variances not only regularize the log-likelihood of the model, they also ensure the existence of the MAP estimator.

Theorem 1 (Existence of the MAP estimator) *Given a piecewise geodesic model and the choice of probability distributions for the parameters and latent variables of the model, for any data set $(t_{i,j}, y_{i,j})_{(i,j)\in[\![1,n]\!]\times[\![1,k_i]\!]}$, there exists*

$$\widehat{\theta}_{MAP} \in \underset{\theta\in\Theta}{\operatorname{argmax}}\, q\left(\theta|\boldsymbol{y}\right).$$

The demonstration of the theorem uses the following lemma.

Lemma 1 *Given a piecewise geodesic model and the choice of a probability distribution for the parameters and latent variables of the model, the posterior $\theta \mapsto q\left(\theta|\boldsymbol{y}\right)$ is continuous on the parameter space Θ.*

Proof Let $\mathscr{Z} = \mathscr{Z}_{\text{pop}} \times \prod_{i=1}^n \mathscr{Z}_i$ denote the space of latent variables. Using Bayes rule, for all $\theta \in \Theta$,

$$q(\theta\,|\,\boldsymbol{y}) = \frac{1}{q(\boldsymbol{y})} \left(\int_{\mathscr{Z}} q\left(\boldsymbol{y}\,|\,\boldsymbol{z}, \theta\right) q\left(\boldsymbol{z}\,|\,\theta\right) \mathrm{d}\boldsymbol{z} \right) q_{prior}(\theta)\,.$$

The density functions $\theta \mapsto q_{prior}(\theta)$ and $\theta \mapsto q\left(\boldsymbol{y}|\boldsymbol{z},\theta\right) q\left(\boldsymbol{z}|\theta\right)$ are continuous on Θ for all $\boldsymbol{z} \in \mathscr{Z}$. Moreover, for all $\theta \in \Theta$ and all $\boldsymbol{z} \in \mathscr{Z}$,

$$q\left(\boldsymbol{y}\,|\,z,\theta\right) = \frac{1}{(\sigma\sqrt{2\pi})^k} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \sum_{j=1}^{k_i} \left(y_{i,j} - \gamma_i(t_{i,j})\right)^2 \right)$$

and so, for all $\theta \in \Theta$ and $\boldsymbol{z} \in \mathscr{Z}$,

$$q\left(\boldsymbol{y}|\boldsymbol{z},\theta\right) q\left(\boldsymbol{z}|\theta\right) \leqslant \frac{1}{(\sigma\sqrt{2\pi})^k}\, q\left(\boldsymbol{z}|\theta\right)$$

which is positive and integrable as a probability distribution. As a consequence, $\boldsymbol{z} \mapsto q\left(\boldsymbol{y}|\boldsymbol{z},\theta\right) q\left(\boldsymbol{z}|\theta\right)$ is integrable – and positive – on $\mathscr{Z}$ for all $\theta \in \Theta$ and $\theta \mapsto q\left(\boldsymbol{y}|\theta\right)$ is continuous. □

Proof (Theorem 1 – Existence of the MAP) We use the Alexandrov one-point compactification $\overline{\Theta} = \Theta \cup \{\infty\}$ of the parameters space Θ, where a sequence $(\theta_n)_{n\in\mathbb{N}}$ converges toward the point ∞ if and only if it eventually steps out of every compact subset of Θ. Thus, given the result of Lemma 1, it suffices to prove that $\lim_{\theta\to\infty} \log q\left(\theta|\boldsymbol{y}\right) = -\infty$. We keep the notation of the previous proof and proceed similarly. In particular, for all $\theta \in \Theta$,

$$\log q\left(\theta|\boldsymbol{y}\right) \leqslant -\log q(\boldsymbol{y}) - k\log(\sqrt{2\pi}) - k\log(\sigma) + \log q_{prior}(\theta)\,.$$

By computing the prior distribution q_{prior}, we remark that there exists λ which does not depend on the parameter θ such as

$$\log q\left(\theta|\boldsymbol{y}\right) \leqslant \lambda(\boldsymbol{y}) - (k+m_\sigma)\log(\sigma) - \frac{m_\sigma}{2}\left(\frac{v}{\sigma}\right)^2 - \frac{m_\Sigma}{2}\left[\log(|\Sigma|) + \frac{m_\Sigma}{2}\operatorname{tr}\left(V\Sigma^{-1}\right)\right].$$

Let $\mu(V)$ denote the smallest eigenvalue of V, $\rho(\Sigma^{-1})$ the largest eigenvalue of Σ^{-1}, which is also its operator norm, and $\left\langle\, \Sigma \,\middle|\, V \,\right\rangle_F$ the Frobenius inner product of Σ with V. As

$$\log(|\Sigma^{-1}|) - \operatorname{tr}\left(V\Sigma^{-1}\right) \leqslant p_{\text{ind}} \log\left(\|\Sigma^{-1}\|\right) - \mu(V)\,\|\Sigma^{-1}\|$$

it comes that

$$\lim_{\|\Sigma\|+\|\Sigma^{-1}\|\to+\infty} \left\{ \frac{m_\Sigma}{2} \left[\log(|\Sigma^{-1}|) - \operatorname{tr}\left(V\Sigma^{-1}\right) \right] \right\} = -\infty.$$

Likewise,

$$\lim_{\sigma+\sigma^{-1}\to+\infty} \left\{ -(k+m_\sigma)\log(\sigma) - \frac{m_\sigma}{2}\left(\frac{v}{\sigma}\right)^2 \right\} = -\infty$$

hence the result. □

We have detailed the previous proof in order to emphasize the necessity of prior distribution on the variances Σ and σ to ensure the existence of the *maximum a posteriori*.

3.2 Consistency of the Maximum a Posteriori Estimator

We are now interested in the consistency of the MAP estimator without making strong assumptions on the distribution of the observations $\boldsymbol{y}$. In particular, we do not assume that the observations are generated by the model.

We denote $P(\mathrm{d}\boldsymbol{y})$ the distribution governing the observations and Θ_* the set of admissible parameters inducing a model distribution close to $P(\mathrm{d}\boldsymbol{y})$:

$$\Theta_* = \Big\{ \theta_* \in \Theta \;\Big|\; \mathbb{E}_{P(\mathrm{d}\boldsymbol{y})}\left[\log q(\boldsymbol{y}|\theta_*)\right] = \sup_{\theta\in\Theta^\omega} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y})}\left[\log q(\boldsymbol{y}|\theta)\right] \Big\}.$$

The MAP estimator is said consistent if it converges to the set Θ_* (on every compact of Θ possibly). Classical results of consistency assume that the space Θ_* is non-empty (see the Wald's consistency theorem [38]). However, such an hypothesis is not entirely satisfactory: we have no guarantee that Θ_* is actually non-empty. We propose below a reasonable framework in which the convergence of the MAP estimator toward the corresponding non-empty set Θ_* is guaranteed.

3.2.1 Two Kinds of Latent Variables

To this end and for any $\omega \in \mathbb{R}^+$, we define the space Θ^ω of admissible parameters such that on average, the fixed effects are bounded by ω:

$$\Theta^\omega = \left\{ \theta = \left(\overline{z_{\mathrm{pop}}}, \Sigma, \sigma\right) \in \Theta \mid \|\overline{z_{\mathrm{pop}}}\|_2 \leqslant \omega \right\},$$

where $\Theta = \mathbb{R}^{p_{\mathrm{pop}}} \times \mathscr{S}^+_{p_{\mathrm{ind}}}(\mathbb{R}) \times \mathbb{R}^+$. As the assumption only concern the average behavior of the population variable $\boldsymbol{z}_{\mathrm{pop}}$, it is not restrictive. Moreover, fixed effects are most of the time bounded (but potentially with high bounds) in applications. In this new framework, for all $\omega \in \mathbb{R}^+$, we consider

$$\Theta^\omega_* = \left\{ \theta \in \Theta^\omega \;\middle|\; \mathbb{E}_{P(\mathrm{d}\boldsymbol{y})}\left[\log q(\boldsymbol{y}|\theta)\right] = \mathbb{E}^*(\omega) \right\},$$

where

$$\mathbb{E}^*(\omega) = \sup_{\theta\in\Theta^\omega} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y})}\left[\log q(\boldsymbol{y}|\theta)\right].$$

To state the consistency of the MAP estimator, we first have to give some notations. For all $i \in [\![1,n]\!]$, we assume the existence of two subsets of $\mathscr{Z}_i$, $\mathscr{Z}_i^{\mathrm{reg}}$ and $\mathscr{Z}_i^{\mathrm{crit}}$, such that $\mathscr{Z}_i = \mathscr{Z}_i^{\mathrm{reg}} \times \mathscr{Z}_i^{\mathrm{crit}}$. In other words, we assume that each component of each individual latent variable $\boldsymbol{z}_i$ may be of two sorts: *regular* or *critical*. We will respectively denote $\boldsymbol{z}_i^{\mathrm{reg}}$ and $\boldsymbol{z}_i^{\mathrm{crit}}$ this sub-variables leading to write, up to permutations, $\boldsymbol{z}_i = (\boldsymbol{z}_i^{\mathrm{reg}}, \boldsymbol{z}_i^{\mathrm{crit}})$. Likewise, we assume that the components of the population latent variables can be regular or critical, *i.e.* that there exists $\mathscr{Z}_{\mathrm{pop}}^{\mathrm{reg}}, \mathscr{Z}_{\mathrm{pop}}^{\mathrm{crit}} \subset \mathscr{Z}_{\mathrm{pop}}$ such that $\boldsymbol{z}_{\mathrm{pop}} = (\boldsymbol{z}_{\mathrm{pop}}^{\mathrm{reg}}, \boldsymbol{z}_{\mathrm{pop}}^{\mathrm{crit}}) \in \mathscr{Z}_{\mathrm{pop}}^{\mathrm{reg}} \times \mathscr{Z}_{\mathrm{pop}}^{\mathrm{crit}}$. To stay consistent with the previous notations, we denote $p_{\mathrm{ind}}^{\mathrm{reg}}$, $p_{\mathrm{ind}}^{\mathrm{crit}}$, $p_{\mathrm{pop}}^{\mathrm{reg}}$ and $p_{\mathrm{pop}}^{\mathrm{crit}}$ the dimension of the ambient space of the matching sets: $\mathscr{Z}_i^{\mathrm{reg}} \subset \mathbb{R}^{p_{\mathrm{ind}}^{\mathrm{reg}}}$ and so on.

3.2.2 Consistency of the Maximum a Posteriori Estimator

In the following, we want to study the effect of the variables $(\boldsymbol{z}_{\mathrm{pop}}, \boldsymbol{z}_i)$ on the trajectories. To this end, we introduce for all i the notation $\gamma_i(\boldsymbol{z}_{\mathrm{pop}}, \boldsymbol{z}_i) = \left(\gamma_i(t_{i,j})\right)_{j\in[\![1,k_i]\!]} \in \mathbb{R}^{k_i}$ and more generally the functions $\gamma_i \colon \mathscr{Z}_{\mathrm{pop}} \times \mathscr{Z}_i \to \mathbb{R}^{k_i}$. Let $\ell \in [\![1,n]\!]$, consider a ℓ-tuple of individuals and denote by $k^\ell = \sum_{i=1}^\ell k_i$ the total number of measures for this ℓ-tuple. Let $\boldsymbol{y}^\ell = (y_i)_{i\in[\![1,\ell]\!]} \in \mathbb{R}^{k^\ell}$ and $\boldsymbol{z}^\ell = (\boldsymbol{z}_{\mathrm{pop}}, \boldsymbol{z}_i)_{i\in[\![1,\ell]\!]} \in \mathbb{R}^{p_{\mathrm{pop}}+\ell\, p_{\mathrm{ind}}}$ be the vectors made up of the ℓ corresponding vectors. As in the one-by-one case, we define by $\boldsymbol{\gamma}^\ell \colon \mathscr{Z}_{\mathrm{pop}} \times \mathscr{Z}_i^\ell \to \mathbb{R}^{k^\ell}$ the function which maps the vector $\boldsymbol{z}^\ell$ to the one $\left(\gamma_i(\boldsymbol{z}_{\mathrm{pop}}, \boldsymbol{z}_i)\right)_{i\in[\![1,\ell]\!]}$.

For all vectors of the form $(\boldsymbol{a},\boldsymbol{b}) \in \mathbb{R}^{p_a} \times \mathbb{R}^{p_b}$, where p_a and p_b are any integer number and for all indices $v \in [\![1, p_a+p_b]\!]$, $(\boldsymbol{a},\boldsymbol{b})_v$ and $(\boldsymbol{a},\boldsymbol{b})_{-v}$ refer respectively to

$$(\boldsymbol{a},\boldsymbol{b})_v = \left((a_1,\dots,a_{p_a}),(b_1,\dots,b_{p_b})\right)_v = \begin{cases} a_v & \text{if } v \leqslant p_a \\ b_{v-p_a} & \text{else} \end{cases}$$

and

$$(\boldsymbol{a},\boldsymbol{b})_{-v} = \begin{cases} \left((a_1,\dots,a_{v-1},a_{v+1},\dots,a_{p_a}),\boldsymbol{b}\right) & \text{if } v \leqslant p_a \\ \left(\boldsymbol{a},(b_1,\dots,b_{v-p_a-1},b_{v-p_a+1},\dots,b_{p_b})\right) & \text{else} \end{cases}.$$

Last, for all $k \in \mathbb{N}$, $\mathscr{L}_k$ refers to the Lebesgue measure on $\mathbb{R}^k$.

Theorem 2 (Consistency of the MAP estimator) *Assume that there exists an integer $\ell \in [\![1,n]\!]$ such that:*

- (H 1) *The number of observations is bigger than the one of latent variables: $p^\ell < k^\ell$, where $k^\ell = \sum_{i=1}^\ell k_i$ and $p^\ell = p_{pop} + \ell\, p_{ind}$;*
- (H 2) *The times of acquisition $\boldsymbol{t}_i = (t_{i,j})_{j\in[\![1,k_i]\!]}$ are independent and identically distributed;*

(H 3) *The density $P(\mathrm{d}\boldsymbol{y}^{\ell})$ is continuous with polynomial tail decay of degree bigger than the dimension of the truncated space of latent variables, i.e. bigger than $p^{\ell}+1$, apart from a compact subset K of $\mathbb{R}^{k^{\ell}}$;*

(H 4) *The individual trajectories grow super-linearly with respect to the regular variables: for all individuals $i \in [\![1,n]\!]$ and for all $v \in [\![1, p^{reg}_{pop} + p^{reg}_{ind}]\!]$, there exists two functions $a_{i,v}, b_{i,v} \colon \mathbb{R}^{p^{reg}_{pop}+p^{reg}_{ind}-1} \to \mathbb{R}$ which depend only on $(\boldsymbol{z}^{reg}_{pop}, \boldsymbol{z}^{reg}_{i})_{-v}$ and such that*

$$\forall (\boldsymbol{z}_{pop}, \boldsymbol{z}_i) \in \mathscr{Z}_{pop} \times \mathscr{Z}_i, \quad a_{i,v}\left((\boldsymbol{z}^{reg}_{pop}, \boldsymbol{z}^{reg}_{i})_{-v}\right) \geqslant 0\,,$$

where

$$a_{i,v}\left((\boldsymbol{z}^{reg}_{pop}, \boldsymbol{z}^{reg}_{i})_{-v}\right) = 0 \ \textit{iff} \ (\boldsymbol{z}^{reg}_{pop}, \boldsymbol{z}^{reg}_{i})_{-v} = 0\,,$$

and

$$\left\|\boldsymbol{\gamma_i}(\boldsymbol{z}_{pop}, \boldsymbol{z}_i)\right\|_{\infty} \geqslant a_{i,v}\left((\boldsymbol{z}^{reg}_{pop}, \boldsymbol{z}^{reg}_{i})_{-v}\right)\left|(\boldsymbol{z}^{reg}_{pop}, \boldsymbol{z}^{reg}_{i})_{v}\right| + b_{i,v}\left((\boldsymbol{z}^{reg}_{pop}, \boldsymbol{z}^{reg}_{i})_{-v}\right);$$

(H 5) *Critical variables induce critical trajectories: for all individuals $i \in [\![1,n]\!]$ and for all $v \in [\![1, p^{crit}_{pop} + p^{crit}_{ind}]\!]$, there exists a critical trajectory $\gamma^{crit}_{i,v}$ such that*

$$\lim_{\left|(\boldsymbol{z}^{crit}_{pop}, \boldsymbol{z}^{crit}_{i})_{v}\right| \to +\infty} \gamma_i(\boldsymbol{z}_{pop}, \boldsymbol{z}_i) = \gamma^{crit}_{i,v}$$

and $\mathscr{L}_{k_i}(\{y_i = \gamma^{crit}_{i,v}\}) = 0$.

Let $(\hat{\theta}_n)_{n\in\mathbb{N}}$ denote any MAP estimator. Then $\Theta^{\omega}_{*} \neq \varnothing$ and for any $\varepsilon \in \mathbb{R}^{*}_{+}$,

$$\lim_{n\to\infty} \mathbb{P}\left[\delta(\hat{\theta}_n, \Theta^{\omega}_{*}) \geqslant \varepsilon\right] = 0\,,$$

where δ in any metric compatible with the topology on Θ^{ω}.

The proof is postponed in Appendix A.

If the times of observations $\boldsymbol{t}_i$ are identically distributed, the individual numbers of measurements k_i are in particular all equal. Thus, under (H 2), Assumption (H 1) writes in a more concise manner as $p^{\ell} < \ell k_1$. However, as (H 2) is not required for all intermediate results (see the proof in Appendix A), we keep the more general statement for (H 1). The condition (H 2) is for instance met if we assume that the times $\boldsymbol{t}_i$ are regularly spaced, that is to say that for all individuals $i \in [\![1,n]\!]$ and all measurements $j \in [\![1,k_1]\!]$, $t_{i,j}$ follows the uniform distribution $\mathscr{U}\left([T_{j-1}, T_j]\right)$, where T is a maximum of the set $\left\{t_{i,j} | i \in [\![1,n]\!],\, j \in [\![1,k_1]\!]\right\}$ and $(T_0 = 0 < T_1 < \ldots < T_{k_1} = T)$ is a subdivision of $[0,T]$.

The condition $p^{\ell} < k^{\ell}$ means that without enough observations for at least some individuals, we cannot build a consistent model. Such an assumption is quite reasonable as we have no chance to catch the trajectories behavior with certitude with less observations than the constraints over them. The assumption on the distribution $P(\mathrm{d}\boldsymbol{y})$ is really weak and always fulfilled in practice. Moreover, as the theorem holds for all $\omega \in \mathbb{R}^{+}$, the boundary over the average of the population latent variable $\overline{z_{\mathrm{pop}}}$ is not really restrictive.

For compactness, we have stated the theorem by considering that a latent variable may be of only one kind: regular or critical. Actually, a single latent variable can be of two kinds: critical in the neighbourhood of $+\infty$ and regular around $-\infty$, and *vice-versa* (see the proof for details). This remark is all the more important in view of some applications and Section 4 but is treated by our proof.

3.3 Estimation with the MCMC-SAEM Algorithm

As explain at the paragraph 2.3.2, a stochastic version of the EM algorithm is adopted, namely the SAEM algorithm. As the conditional distribution $q(\boldsymbol{z}|\boldsymbol{y}, \theta)$ involves he renormalization constant which is our target function, the simulation step is replaced using a sampling algorithm, leading to consider the MCMC-SAEM algorithm [2, 19]. It alternates between simulation, stochastic approximation and maximization steps until convergence. The simulation step is achieved using a symmetric random walk Hasting-Metropolis within Gibbs sampler [27].

The complete log-likelihood of our model writes

$$\begin{aligned}
\log q(\boldsymbol{y}, \boldsymbol{z}, \theta) \ = \ & -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{k_i}\left(y_{i,j} - \gamma_i(t_{i,j})\right)^2 - k\log(\sigma) \\
& - \frac{1}{2}\sum_{i=1}^{n}\left({}^{t}\boldsymbol{z}_i\,\Sigma^{-1}\,\boldsymbol{z}_i\right) - \frac{n}{2}\log(|\Sigma|) - \frac{1}{2}\operatorname{tr}\left(V\Sigma^{-1}\right) \\
& - \frac{1}{2}\,{}^{t}\left(\boldsymbol{z}_{\mathrm{pop}} - \overline{z_{\mathrm{pop}}}\right)D^{-1}\left(\boldsymbol{z}_{\mathrm{pop}} - \overline{z_{\mathrm{pop}}}\right) - \frac{1}{2}\log(|D|) \\
& + \frac{m_{\Sigma}}{2}\left(\log(|V|) - \log(|\Sigma|)\right) + m_{\sigma}\log\left(\frac{v}{\sigma}\right) \\
& - \frac{m_{\sigma}}{2}\left(\frac{v}{\sigma}\right)^2 + \mathit{csts}\,.
\end{aligned}$$

It is clear to see that this model belongs to the curved exponential family: up to a multiple constant, the sufficient statistics are defined as

$$\boldsymbol{S}_1(\boldsymbol{y}, \boldsymbol{z}) = \boldsymbol{z}_{\mathrm{pop}} \in \mathbb{R}^{p_{\mathrm{pop}}}\,, \quad S_2(\boldsymbol{y}, \boldsymbol{z}) = \frac{1}{n}\sum_{i=1}^{n}{}^{t}\boldsymbol{z}_i\boldsymbol{z}_i \in \mathscr{M}_{p_{\mathrm{ind}}}(\mathbb{R})$$

$$\text{and} \quad S_3(\boldsymbol{y}, \boldsymbol{z}) = \frac{1}{k}\sum_{i=1}^{n}\sum_{j=1}^{k_i}\left(y_{i,j} - \gamma_i(t_{i,j})\right)^2 \in \mathbb{R}\,.$$

By denoting `iter` the increment, $\boldsymbol{z}^{(\mathtt{iter})}$ the current sample and $S_u^{(\mathtt{iter})}$ the current approximation of the u^{th} sufficient statistics, $u \in \{1,2,3\}$, the stochastic approximation step is defined as:

$$S_u^{(\mathtt{iter})} = S_u^{(\mathtt{iter})} + \varepsilon_{\mathtt{iter}}\left(S_u(\boldsymbol{y}, \boldsymbol{z}^{(\mathtt{iter})}) - S_u^{(\mathtt{iter})}\right),$$

Algorithm 1: Overview of the SAEM for the generic piecewise geodesic model.

Input: $\theta^* = \left(\overline{z_{\text{pop}}}^*, \Sigma^*, \sigma^*\right)$, (V, m_Σ), (v, m_σ), `maxIter`, `Nburnin`.

Output: $\theta = \left(\overline{z_{\text{pop}}}, \Sigma, \sigma\right)$.

 # Initialization: $\theta \leftarrow \theta^*$; $S \leftarrow 0$; $(\varepsilon_{\texttt{iter}})_{\texttt{iter}>0}$; $z_{pop} \leftarrow \overline{z_{\text{pop}}}$; $(\boldsymbol{z}_i)_i \leftarrow 0$;

for $iter = 1$ **to** $maxIter$ **do**

  # Simulation: $(\boldsymbol{z}_{\text{pop}}, (\boldsymbol{z}_i)_i) \leftarrow \texttt{sampler}(\boldsymbol{z}_{\text{pop}}, (\boldsymbol{z}_i)_i)$;

  # Stochastic Approximation:

$$\boldsymbol{S}_1 \leftarrow \boldsymbol{S}_1 + \varepsilon_{\texttt{iter}} \left(\boldsymbol{z}_{\text{pop}} - \boldsymbol{S}_1\right) ;$$

$$S_2 \leftarrow S_2 + \varepsilon_{\texttt{iter}} \left(\frac{1}{n} \sum_i {}^t\boldsymbol{z}_i \boldsymbol{z}_i - S_2\right) ;$$

$$S_3 \leftarrow S_3 + \varepsilon_{\texttt{iter}} \left(\frac{1}{k} \sum_{i=1}^{n} \sum_{j=1}^{k_i} (y_{i,j} - \gamma_i(t_{i,j}))^2 - S_3\right) ;$$

  # Maximization: $\overline{z_{\text{pop}}} \leftarrow \boldsymbol{S}_1$; $\Sigma \leftarrow \frac{nS_2 + m_\Sigma V}{n + m_\Sigma}$; $\sigma \leftarrow \sqrt{\frac{kS_3 + m_\sigma v^2}{k + m_\sigma}}$;

end

where $(\varepsilon_{\texttt{iter}})$ is a sequence positive step size (see below).

The maximization step is straightforward given the sufficient statistics of our exponential model: we update the parameters by taking a barycentre between the corresponding stochastic approximation and the prior (when they exist). In other words:

$$\overline{z_{\text{pop}}}^{(\texttt{iter}+1)} = \boldsymbol{S}_1(\boldsymbol{y}, \boldsymbol{z}^{(\texttt{iter})}),$$

$$\Sigma^{(\texttt{iter}+1)} = \frac{nS_2(\boldsymbol{y}, \boldsymbol{z}^{(\texttt{iter})}) + m_\Sigma V}{n + m_\Sigma}$$

$$\text{and} \quad \sigma^{2\,(\texttt{iter}+1)} = \frac{kS_3(\boldsymbol{y}, \boldsymbol{z}^{(\texttt{iter})}) + m_\sigma v^2}{k + m_\sigma}.$$

Finally, given an adapted sampler and the sequence $(\varepsilon_{\texttt{iter}})_{\texttt{iter}}$ defined by:

$$\forall \texttt{iter} \geqslant 1, \quad \varepsilon_{\texttt{iter}} = \mathbb{1}_{\texttt{iter} \leqslant \texttt{Nburnin}} + (\texttt{iter} - \texttt{Nburnin})^{-0.65} \mathbb{1}_{\texttt{iter} > \texttt{Nburnin}}$$

our algorithm writes as Algorithm 1 . Some experimental results are presented in Section 5.

4 Application to Chemotherapy Monitoring

Understanding the global disease progression is the key of chemotherapy monitoring. Indeed, physicians have to choose the best possible treatment and sequence of molecules for each of their patients, in the shortest possible time. Here, we propose two instantiations of the generic piecewise geodesic model, both in view of chemotherapy monitoring: the piecewise logistic curve model and the piecewise geodesic shape model.

We recall that patients are treated and so the evolution of the tumoral growth will fluctuate. Therefore, reaction-diffusion based tumor growth models do not apply in this context. Moreover, the two proposed models allow to bring a representative of the whole population comportment out for any kind of input data: scores, images, shapes, etc.

4.1 The Piecewise Logistic Curve Model: Chemotherapy Monitoring through RECIST Score

In this section, we explicit the generic model with logistic geodesics and $M =]0,1[$. This is motivated by the study of the RECIST score monitoring, which leads to consider one-dimension manifold, with one rupture point. As this explicit model is designed in view of our target application, we first give a short description of RECIST score.

4.1.1 The RECIST Score

Patients suffering from the metastatic kidney cancer, take a drug each day and regularly have to check their tumor evolution. Indeed, during the past few years, the way renal metastatic cancer are monitoring was profoundly changed: a new class of anti-angiogenic therapies targeting the tumor vessels instead of the tumor cells has emerged and drastically improved survival by a factor of three [12]. These new drugs, however, do not cure the cancer, and only succeed in delaying the tumor growth, requiring the use of successive therapies which have to be continued or interrupted at the appropriate moment according to the patient's response. So, the new medicine process has also created a new scientific challenge: how to choose the most efficient drug therapy given he first monitoring times of the response profile of a given patient.

The RECIST (Response Evaluation Criteria In Solid Tumors) score [35] is a set of published rules that measures the tumoral growth. Physicians select at most five lesions, with a sufficient diameter, and sum the longest diameter for all target lesions. This leads them to determine if the tumors in cancer patients respond (completely or partially), stabilize or progress during treatment.

The response to a given treatment has generally two distinct phases: first, tumor's size reduces; then, the tumor grows again. So, we have to build a model which allow to us to catch this behaviors. Moreover, a practical question is to quantify the correlation between both phases and to determine as accurately as possible the individual rupture times t_R^i which are related to an escape of the patient's response to treatment.

4.1.2 The Piecewise Logistic Curve Model

Our observations consist of patient's RECIST score over time, *i.e.* of sequences of bounded one-dimension measures. As explained above, we could make out two phases in the evolution of RECIST scores: a decreasing and a growing one. So, we set $m = 2$ and $d = 1$, which leads us to propose a way to build models for chemotherapy monitoring. This model has been designed after discussions with oncologists of the HEGP.

The Group-Representative Trajectory. Let M_0 be the open interval $]0,1[$, equipped with the logistic metric: $\forall x \in M_0$, $\forall \xi, \zeta \in T_x M_0 \simeq \mathbb{R}$,

$$g_x(\xi,\zeta) = \xi\, \mathscr{G}(x)\, \zeta \quad \text{with} \quad \mathscr{G}(x) = \frac{1}{x^2(1-x^2)}\,.$$

Given three real numbers γ_0^{init}, γ_0^{escap} and γ_0^{fin} we define two affine functions by setting down $\phi_0^1 : x \mapsto \left(\gamma_0^{\text{init}} - \gamma_0^{\text{escap}}\right) x + \gamma_0^{\text{escap}}$ and $\phi_0^2 : x \mapsto \left(\gamma_0^{\text{fin}} - \gamma_0^{\text{escap}}\right) x + \gamma_0^{\text{escap}}$. This allows us to map M_0 onto the intervals $]\gamma_0^{\text{escap}}, \gamma_0^{\text{init}}[$ and $]\gamma_0^{\text{escap}}, \gamma_0^{\text{fin}}[$ respectively: if $\bar{\gamma}_0$ refers to the sigmoid function, $\phi_0^1 \circ \bar{\gamma}_0$ will be a logistic curve, growing from γ_0^{escap} to γ_0^{init}. For compactness, we note t_R the single breaking-up time at the population level and t_R^i at the individual one. Moreover, due to our target application, we force the first logistic to be decreasing and the second one increasing (this condition may be easily relaxed for other framework).

Logistics are defined on open intervals, with asymptotic constraints. We want to formulate our constraints on some non-infinite time-points, as explained in paragraph 2.1.2. So, we set a positive threshold ν, close to zero, and demand the logistics γ_0^1 and γ_0^2 to be ν-near from their corresponding asymptotes. More precisely, we impose the trajectory γ_0 to be of the form $\gamma_0 = \gamma_0^1 \mathbb{1}_{]-\infty,t_R]} + \gamma_0^2 \mathbb{1}_{]t_R,+\infty[}$, where, for all times $t \in \mathbb{R}$,

$$\gamma_0^1(t) = \frac{\gamma_0^{\text{init}} + \gamma_0^{\text{escap}} \mathrm{e}^{(at+b)}}{1+\mathrm{e}^{(at+b)}} \in]\gamma_0^{\text{escap}}, \gamma_0^{\text{init}}[,$$

$$\gamma_0^2(t) = \frac{\gamma_0^{\text{fin}} + \gamma_0^{\text{escap}} \mathrm{e}^{-(ct+d)}}{1+\mathrm{e}^{-(ct+d)}} \in]\gamma_0^{\text{escap}}, \gamma_0^{\text{fin}}[$$

and a, b, c and d are some positive numbers given by the following constraints

$$\gamma_0^1(t_0) = \gamma_0^{\text{init}} - \nu\,, \quad \gamma_0^1(t_R) = \gamma_0^2(t_R) = \gamma_0^{\text{escap}} + \nu$$

and $\quad \gamma_0^2(t_1) = \gamma_0^{\text{fin}} - \nu\,.$

In order the previous logistics to be well-defined, we also have to enforce $\gamma_0^{\text{escap}} + 2\nu \leqslant \gamma_0^{\text{init}}$ and $\gamma_0^{\text{escap}} + 2\nu \leqslant \gamma_0^{\text{fin}}$. Thus, $p_{\text{pop}} = 5$ and

$$\mathscr{Z}_{\text{pop}} = \left\{ \left(\gamma_0^{\text{init}}, \gamma_0^{\text{escap}}, \gamma_0^{\text{fin}}, t_R, t_1\right) \in \mathbb{R}^5 \;\middle|\; \gamma_0^{\text{escap}} + 2\nu \leqslant \gamma_0^{\text{init}} \;\wedge\; \gamma_0^{\text{escap}} + 2\nu \leqslant \gamma_0^{\text{fin}} \right\}.$$

In our context, the initial time of the process is known: it is the beginning of the treatment. So, we assume that the representative initial time t_0 is equal to zero.

Individual Trajectories. For each $i \in [\![1,n]\!]$, given $(\alpha_i^1, \alpha_i^2, \tau_i) \in \mathbb{R}_+^2 \times \mathbb{R}$, the time-warps (*cf.* 2.2.1) write

$$\psi_i^1(t) = \alpha_i^1 (t - t_0 - \tau_i) + t_0 \quad \text{and} \quad \psi_i^2(t) = \alpha_i^2 (t - t_R - \tau_i^2) + t_R\,,$$

where $\tau_i^2 = \tau_i + \left(\frac{1-\alpha_i^1}{\alpha_i^1}\right)(t_R - t_0)$.

In the same way as the time-warp, the diffeomorphisms ϕ_i^1 and ϕ_i^2 (*cf.* 2.2.2) are chosen to allow different amplitudes

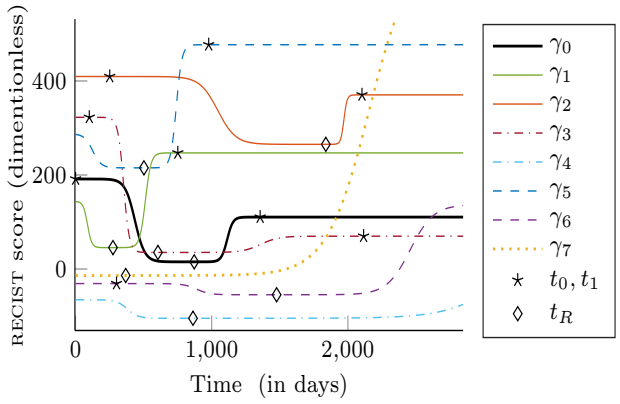


Fig. 2 *The piecewise logistic curve model: Diversity of individual trajectories.* A typical representative trajectory in bold and several individual ones, for different vectors $\boldsymbol{z}_i$. The rupture times are represented by diamonds and the initial/final times by stars.

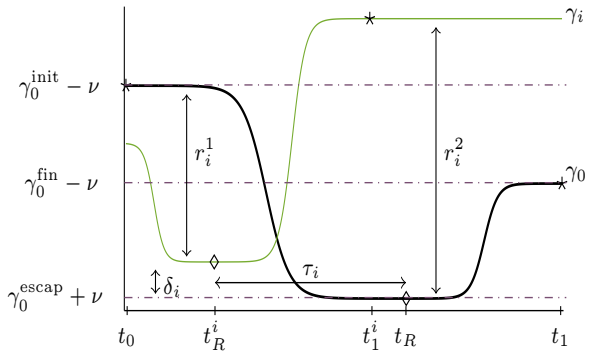


Fig. 3 *The piecewise logistic curve model: From representative to individual trajectory.* Illustration of the non-standard constraints for γ_0 and the transition from the representative trajectory to an individual one: the trajectory γ_i is subject to a temporal and a spacial warp. In other "words", $\gamma_i = \phi_i^1 \circ \gamma_0^1 \circ \psi_i^1 \mathbb{1}_{]-\infty,t_R^i]} + \phi_i^2 \circ \gamma_0^2 \circ \psi_i^2 \mathbb{1}_{]t_R^i,+\infty[}$.

and rupture values: for each $i \in [\![1,n]\!]$, given the two scaling factors r_i^1 and r_i^2 and the space-shift δ_i, we define

$$\forall \ell \in \{1,2\}, \quad \phi_i^\ell(x) = r_i^\ell\,(x - \gamma_0(t_R)) + \gamma_0(t_R) + \delta_i\,.$$

Other choices are conceivable but in the context of our target applications, this one is the most appropriate: as we want to study the correlation between growth and decrease phase, none of the portions of the curves have to be favoured and affine functions allow us to put the same weight on the whole curves. Mathematically, any regular and injective function defined on $]\gamma_0^{\text{escap}}, \gamma_0^{\text{init}}[$ (respectively $]\gamma_0^{\text{escap}}, \gamma_0^{\text{fin}}[$) works.

To sum up, each individual trajectory γ_i depends on the representative curve γ_0 through $\boldsymbol{z}_{\text{pop}} = \left(\gamma_0^{\text{init}}, \gamma_0^{\text{escap}}, \gamma_0^{\text{fin}}, t_R, t_1\right)$ fixed and $\boldsymbol{z}_i = \left(\alpha_i^1, \alpha_i^2, \tau_i, r_i^1, r_i^2, \delta_i\right)$ random effects. This leads to a non-linear mixed effects model. More precisely, we set for all individuals $i \in [\![1,n]\!]$

$$\forall \ell \in \{1,2\},\ \gamma_i^\ell = \phi_i^\ell \circ \gamma_0^\ell \circ \psi_i^\ell \quad \text{and} \quad t_R^i = t_0 + \tau_i + \frac{t_R - t_0}{\alpha_i^1}\,,$$

which leads us to write for all measurements $j \in [\![1,k_i]\!]$,

$$y_{i,j} = \gamma_i^1(t_{i,j})\,\mathbb{1}_{]-\infty,t_R^i]}(t_{i,j}) + \gamma_i^2(t_{i,j})\,\mathbb{1}_{]t_R^i,+\infty[}(t_{i,j}) + \varepsilon_{i,j}\,.$$

Figures 2 and 3 provide illustrations of the model. On each figure, the bold black curve represents the characteristic trajectory γ_0 and the color curves several individual trajectories.

We proceed as in the paragraph 2.3.1 and set $\alpha_i^\ell = \mathrm{e}^{\xi_i^\ell}$ for $\ell \in \{1,2\}$. Likewise, the scaling parameters r_i^ℓ have to be positive and equal to one on average while the space shifts δ_i can be of any signs and must be zero on average. So, we set $r_i^\ell = \mathrm{e}^{\rho_i^\ell}$ for $\ell \in \{1,2\}$ leading to $\boldsymbol{z}_i = \left(\xi_i^1, \xi_i^2, \tau_i, \rho_i^1, \rho_i^2, \delta_i\right)$. In particular, $p_{\text{ind}} = 6$ and we assume that there exists $\Sigma \in \mathscr{S}^+_{p_{\text{ind}}}(\mathbb{R})$ such that $\boldsymbol{z}_i \sim \mathscr{N}(0,\Sigma)$ for all $i \in [\![1,n]\!]$. In view of our target application, this assumption is really important: usually, the random effects are studied independently. Here, we are interested in correlations between the two phases of patient's response to treatment in order to answer question like: does a fast response induce a fast reprogression after the rupture time, which would mean that a fast response would decrease the susceptibility to this drug?

4.1.3 Theoretical Analysis of the Piecewise Logistic Curve Model

Theorem 1 applies as is leading to a well-defined MAP estimator for the piecewise logistic model. Moreover, at the risk of assuming some restriction concerning the distribution of our observations, the piecewise logistic model is consistent.

More precisely, let Θ^{PL} be the space of the admissible parameters for the piecewise logistic model, *i.e.*

$$\Theta^{\text{PL}} = \big\{(\overline{\gamma_0^{\text{init}}}, \overline{\gamma_0^{\text{escap}}}, \overline{\gamma_0^{\text{fin}}}, \overline{t_R}, \overline{t_1}, \Sigma, \sigma) \in \mathbb{R}^{p_{\text{pop}}} \times \mathscr{S}^+_{p_{\text{ind}}}(\mathbb{R}) \times \mathbb{R}^+ \big\}\,.$$

We define

$$\Theta^{\omega,\text{PL}} = \{\,\theta \in \Theta^{\text{PL}} \mid \|(\overline{\gamma_0^{\text{init}}}, \overline{\gamma_0^{\text{escap}}}, \overline{\gamma_0^{\text{fin}}}, \overline{t_R}, \overline{t_1})\| \leqslant \omega\,\}$$

the space of the parameters associated to bounded on average fixed effects, for the piecewise logistic model and, as in the generic framework, the space

$$\Theta_*^{\omega,\text{PL}} = \{\theta \in \Theta^{\omega,\text{PL}} \mid \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log q(\boldsymbol{y}^\ell|\theta)\right] = \mathbb{E}^*(\omega)\}\,,$$

where $\mathbb{E}^*(\omega) = \sup_{\theta \in \Theta^{\omega,\text{PL}}} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log q(\boldsymbol{y}^\ell|\theta)\right]$.

Theorem 3 (Consistency of the MAP, piecewise logistic) *Assume that*

- (H 1) *The number of observations is bigger than the one of latent variables: There exists $\ell \in [\![1,n]\!]$ such that $p^\ell < k^\ell$, where $k^\ell = \sum_{i=1}^\ell k_i$ and $p^\ell = p_{pop} + \ell\, p_{ind}$;*
- (H 2) *The times of acquisition $t_i = (t_{i,j})_{j \in [\![1,k_i]\!]}$ are independent and identically distributed;*
- (H 3) *The density $P(\mathrm{d}\boldsymbol{y}^\ell)$ is continuous with polynomial tail decay of degree bigger $p^\ell + 1$ apart from a compact subset K of $\mathbb{R}^{k^\ell}$;*

Then, the piecewise logistic model satisfies the hypothesis of Theorem 2. In particular, if $(\hat{\theta}_n)_{n\in\mathbb{N}}$ denote any MAP estimator, $\Theta_*^{\omega,\text{PL}} \neq \varnothing$ and for any $\varepsilon \in \mathbb{R}_+^*$,

$$\lim_{n\to\infty} \mathbb{P}\left[\delta(\hat{\theta}_n, \Theta_*^{\omega,\text{PL}}) \geqslant \varepsilon\right] = 0\,,$$

where δ in any metric compatible with the topology on $\Theta^{\omega,\text{PL}}$.

Proof We demonstrate that, for all $i \in [\![1,n]\!]$, the variables $\left(\gamma_0^{\text{init}}, \gamma_0^{\text{escap}}, \gamma_0^{\text{fin}}, \rho_i^1, \rho_i^2, \delta_i\right)$ are regular, that the variables $\left(t_R, t_1, \xi_i^1, \xi_i^2, \tau_i\right)$ are critical, and that $\left(\rho_i^1, \rho_i^2\right)$ are regular in the neighbourhood of $+\infty$ and critical near $-\infty$. See the remark after Theorem 2.

- (H 4) Let $i \in [\![1,n]\!]$. By definition of γ_i,

$$\begin{aligned}\|\gamma_i(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)\|_\infty \;=\; \max\Big\{ &|\,\gamma_0^{\text{escap}} + \nu + \delta_i\,|\,; \\ &|\,\gamma_0^{\text{escap}} + \nu + \delta_i + \mathrm{e}^{\rho_i^1}(\gamma_0^{\text{init}} - \gamma_0^{\text{escap}} - 2\nu)\,|\,; \\ &|\,\gamma_0^{\text{escap}} + \nu + \delta_i + \mathrm{e}^{\rho_i^2}(\gamma_0^{\text{fin}} - \gamma_0^{\text{escap}} - 2\nu)\,|\,\Big\}\,.\end{aligned}$$

  And we can check that for γ_0^{init}, γ_0^{escap}, γ_0^{fin}, ρ_i^1, ρ_i^2 and δ_i and that for ρ_i^1 and ρ_i^2 as soon as $|\rho_i^1|, |\rho_i^2| \geqslant 1$ there exists two functions a_i and b_i as in [Theorem 2 (H 4)].
- (H 5) Let $i \in [\![1,n]\!]$ and $j \in [\![1,k_i]\!]$. By definition of γ_i,

$$\begin{aligned}\lim_{t_R\to+\infty} \gamma_i(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)_j = \Big[&\mathrm{e}^{\rho_i^1}\left(\gamma_0^{\text{init}} - \gamma_0^{\text{escap}} - 2\nu\right) \\ &+ \gamma_0^{\text{escap}} + \nu + \delta_i\Big]\,\mathbb{1}_{[t_0,+\infty[}(t_{i,j})\,,\end{aligned}$$

  where $\gamma_i(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)_j$ denotes the j^{th} coordinate of the vector $\gamma_i(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i) \in \mathbb{R}^{k_i}$. However, by construction,

$\gamma_0^{\text{init}} - \gamma_0^{\text{escap}}$ and γ_0^{escap} follow a normal distribution so

$$\mathscr{L}_{k_i}\Big(\Big\{ y_{i,j} = e^{\rho_i^1}\left(\gamma_0^{\text{init}} - \gamma_0^{\text{escap}} - 2\nu\right) + \gamma_0^{\text{escap}} + \nu + \delta_i \Big\}\Big) = 0\,.$$

Likewise for $t_R \to -\infty$. The same argument holds when t_1, ξ_i^1, ξ_i^2 or τ_i become infinite and when ρ_i^1 or ρ_i^2 go to $-\infty$.

□

4.2 The Piecewise Geodesic Shape Model: Chemotherapy Monitoring through 3D Anatomical Shape

A more precise way to follow-up cancer is to focus on the evolution of the tumors as anatomical shape. For this purpose, the tumors is segmented and transformed into a surface mesh or a curvilinear paths depending on the type of data we consider, typically depending on the number of layers in scanner we have access to. Let $d \in \{2,3\}$ be the dimension of the ambient space.

Shape trajectories within the geodesic-framework developed by Schiratti et al. has already been addressed [5]. We explain here quickly how to adapt this model to the piecewise geodesic framework. In particular, we use consistent notations and admit the notion of exp-parallelism.

4.2.1 The Piecewise Geodesic Shape Model

In the same way as the piecewise logistic curve model, we are targeting chemotherapy monitoring with two distinct phases in the evolution of the tumoral growth and thus we set $m = 2$. We keep the notation t_R for the single breaking-up time at the population level and t_R^i at the individual one. We consider only two geodesic components, so that an easy way to enforce the representative path to be continuous is to define the first component in the past and the second one in the future, from the rupture time t_R. Thus, provided that we follow forward the first component, we can use exactly the same construction that the one introduced by Bône et al.. As a consequence, the following is applicable either for currents [37] or varifolds [7], allowing to consider shapes without any point correspondence.

The Group-Representative Trajectory. Let $y_R \in M \subset \mathbb{R}^d$ be the rupture shape, *i.e.* the representative shape of the population at the rupture time t_R, and likewise $c_R \in \mathbb{R}^{n_{cp}d}$ be a set of n_{cp} rupture control points. Let $m_R^1 \in \mathbb{R}^{n_{cp}d}$ and $m_R^2 \in \mathbb{R}^{n_{cp}d}$ be respectively the backward and the forward momenta at the rupture time. We define the representative path by:

$$\gamma_0 : t \mapsto \mathscr{E}xp_{c_R,t_R,-t}(m_R^1) \circ y_R \; \mathbb{1}_{]-\infty,t_R]}(t) + \mathscr{E}xp_{c_R,t_R,t}(m_R^2) \circ y_R \; \mathbb{1}_{[t_R,+\infty[}(t)\,,$$

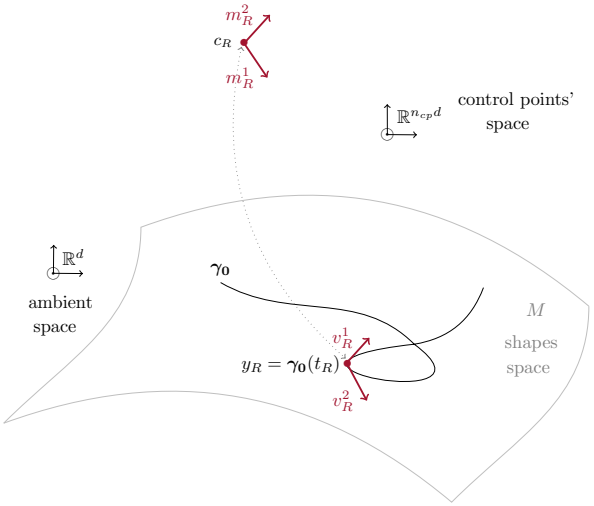


Fig. 4 *The piecewise geodesic shape model: Construction of the group-representative trajectory*. Let t_R be the rupture time and $y_R \in M$ the rupture shape, *i.e.* the shape of the representative path at the rupture time. We define the path γ_0 as the concatenation of the two geodesics starting at the rupture time t_R and the point y_R, in the directions associated to m_R^1 the backward and m_R^2 the forward momenta respectively and where the first one is followed backward.

where $t \mapsto \mathscr{E}xp_{c_R,t_R,t}(m_R)$ denotes the exponential operator associated to the manifold of diffeomorphisms underlying the shape space (See [4, 5, 23] for details about the construction of the shape space). In particular, the resulting backward and forward velocity vectors at the rupture time are respectively define by $v_R^1 = c_R \cdot m_R^1 = \sum_{q=1}^{n_{cp}} c_{R,q}\, m_{R,q}^1$ and $v_R^2 = c_R \cdot m_R^2$. Figure 4 sums up this construction.

Individual Trajectories. At the individual level, as the initial representative time is not explicitly defined in this framework, we slightly modify the first time component reparametrization leading to $\psi_i^1(t) = e^{\xi_i^1}(t - t_R - \tau_i) + t_R$ and $\psi_i^2(t) = e^{\xi_i^2}(t - t_R - \tau_i) + t_R$. The individual rupture time are given by $t_R^i = t_R + \tau_i$ and we check that $\psi_i^1(t_R^i) = \psi_i^2(t_R^i) = t_R$.

The diffeomorphic component deformations consist of exp-parallelism of the representative path [31]. Given a vector w, to define the exp-parallel of a curve γ in the direction of w, we first transport the vector w along the curve γ and then compute the flow given by the transported vector. We note $P_t : \mathbb{R}^{n_{cp}d} \to \mathbb{R}^{n_{cp}d}$ the parallel transport operator, which transport any vector $w \in \mathbb{R}^{n_{cp}d}$ along the curve γ_0 from $\gamma_0(t_R)$ to $\gamma_0(t)$ and we set: $\eta_w : t \mapsto \mathscr{E}xp_{c(t),0,1}(P_t(w))$, where $c(t)$ is the set of control points for γ_0, at the time t:

$$c(t) = \mathscr{E}xp_{c_R,t_R,-t}(m_R^1) \circ c_R \; \mathbb{1}_{]-\infty,t_R]}(t) + \mathscr{E}xp_{c_R,t_R,t}(m_R^2) \circ c_R \; \mathbb{1}_{[t_R,+\infty[}(t)\,.$$

Thus, given a space shift momenta w_i for all individuals, the space deformation of the curve γ_0 is given by $t \mapsto \eta_{w_i}(t) \circ y_R$.

Last, for all individuals $i \in [\![1,n]\!]$, we define a subject-specific trajectory by setting:

$$\gamma_i : t \mapsto \eta_{w_i}(\psi_i^1(t)) \circ y_R \, \mathbb{1}_{]-\infty, t_R^i]}(t) + \eta_{w_i}(\psi_i^2(t)) \circ y_R \, \mathbb{1}_{[t_R^i, +\infty[}(t).$$

Space-shift momenta and identifiability Following [5] and in the spirit of Independent Component Analysis [15], we assume that each space-shift momenta w_i is a linear combination of n_s sources $s_i \in \mathbb{R}^{n_s}$, *i.e.* that $w_i = A_{m_R^\perp} s_i$, where $A_{m_R^\perp} \in \mathscr{M}_{n_{cp}d, n_s}(\mathbb{R})$ calls modulation matrix. As argued in [5, 31], we have to ensure the orthogonality between m_R and w_i in order to ensure the identifiability of the model. This orthogonality condition prevents a confusion between the space shifts and the acceleration factors and can be achieved through projection techniques that we do not detail here.

To sum up, the population random effects are given by $z_{\text{pop}} = (y_R, c_R, m_R^1, m_R^2, t_R, A_{m_R^\perp})$ and the individual ones by $z_i = (\xi_i^1, \xi_i^2, \tau_i, s_i)$. To place ourselves in the hierachical framework detailed at the paragraph 2.3.2, we assume that there exists $\Sigma \in \mathscr{S}_3^+(\mathbb{R})$ such that $z_i \sim \mathscr{N}(0,\Sigma) \otimes \mathscr{N}(0,1)$ and that there exists small fixed variances such that the population latent variable follow a tight Gaussian distribution.

4.2.2 Theoretical Analysis of the Piecewise Geodesic Shape Model

As for the piecewise logistic curve model, Theorem 1 applies and the MAP estimator for the piecewise geodesic shape model is well-defined. We therefore focus on the consistency of this model.

Like previously, we define the space of admissible parameters associated to bounded on average fixed effects:

$$\Theta^{\omega,\text{PS}} = \{\, \theta \in \Theta^{\text{PS}} \mid \|(\overline{\gamma_0^{\text{init}}}, \overline{\gamma_0^{\text{escap}}}, \overline{\gamma_0^{\text{fin}}}, \overline{t_R}, \overline{t_1})\| \leqslant \omega \,\}$$

for all $\omega \in \mathbb{R}$, where

$$\Theta^{\text{PS}} = \Big\{ (\overline{y_R}, \overline{c_R}, \overline{m_R^1}, \overline{m_R^2}, \overline{t_R}, \overline{A_{m_R^\perp}}\Sigma, \sigma) \in \mathbb{R}^{p_{\text{pop}}} \times \mathscr{S}^+_{p_{\text{ind}}}(\mathbb{R}) \times \mathbb{R}^+ \Big\}.$$

As in the generic and piecewise-logistic framework, we also define the space

$$\Theta^{\omega,\text{PS}}_* = \{\theta \in \Theta^{\omega,\text{PS}} \mid \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log q(\boldsymbol{y}^\ell|\theta)\right] = \mathbb{E}^*(\omega)\},$$

where $\mathbb{E}^*(\omega) = \sup_{\theta \in \Theta^{\omega,\text{PL}}} \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log q(\boldsymbol{y}^\ell|\theta)\right]$.

Theorem 4 (Consistency of the MAP, piecewise shapes) *Assume that*

- (H 1) *The number of observations is bigger than the one of latent variables: There exists $\ell \in [\![1,n]\!]$ such that $p^\ell < k^\ell$, where $k^\ell = \sum_{i=1}^\ell k_i$ and $p^\ell = p_{pop} + \ell\, p_{ind}$;*
- (H 2) *The times of acquisition $t_i = (t_{i,j})_{j \in [\![1,k_i]\!]}$ are independent and identically distributed;*
- (H 3) *The density $P(\mathrm{d}\boldsymbol{y}^\ell)$ is continuous with polynomial tail decay of degree bigger $p^\ell + 1$ apart from a compact subset K of $\mathbb{R}^{k^\ell}$;*
- (H 4) *For all individuals $i \in [\![1,n]\!]$, $t \mapsto \|\gamma_i(t)\|$ grows superlinearly or $t \mapsto \gamma_i(t)$ converges uniformly w.r.t the variable y_R toward a function $t \mapsto \gamma_i^*(t)$;*
- (H 5) *For all individuals $i \in [\![1,n]\!]$, the variables c_R, m_R^1, m_R^2, $A_{m_R^\perp}$ and s_i are either regular or critical, in the sens of Theorem 2.*

Then, the piecewise geodesic shapes model satisfies the hypothesis of Theorem 2. In particular, if $(\hat{\theta}_n)_{n\in\mathbb{N}}$ denote any MAP estimator, $\Theta_*^{\omega,\text{PS}} \neq \varnothing$ and for any $\varepsilon \in \mathbb{R}_+^*$,

$$\lim_{n\to\infty} \mathbb{P}\left[\delta(\hat{\theta}_n, \Theta_*^{\omega,\text{PS}}) \geqslant \varepsilon\right] = 0,$$

where δ in any metric compatible with the topology on $\Theta^{\omega,\text{PS}}$.

Proof Let's demonstrate that the variables $y_R, t_R, \xi_i^1, \xi_i^2$ and τ_i are either regular of critical for all individuals $i \in [\![1,n]\!]$.

1. Let $i \in [\![1,n]\!]$ and $(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i) \in \mathscr{Z}_{\text{pop}} \times \mathscr{Z}_i$. By continuity of the parallel transport and the trajectory γ_0, there exists $a \in \mathbb{R}^+$ such that
$$\left\|\gamma_{\boldsymbol{i}}(\boldsymbol{z}_{\text{pop}}, \boldsymbol{z}_i)\right\|_\infty \geqslant a \|y_R = \gamma_0(t_R)\|$$
and so, y_R is regular.
2. By continuity of the geodesic flow, for all times $t \in \mathbb{R}$, $\gamma_i(t)$ depends continuously of $\gamma_i(t_R)$. Thus, for all $t \in \mathbb{R}$, by continuity of the parallel transport, $\gamma_i(t)$ depends continuously of y_R. If γ_i^* exists, since the convergence of γ_i toward γ_i^* is uniform w.r.t the variable y_R, γ_i^* is also continuous w.r.t $y_R = \gamma_0(t_R) \sim \mathscr{N}(\overline{y_R}, *)$ and so is a continuous distribution.
For all $i \in [\![1,n]\!]$, all $\ell \in \{1,2\}$ and all $t \in \mathbb{R}$, we have $\lim_{\xi_i^\ell \to -\infty} \psi_i^\ell(t) = t_R$. Therefore, $\lim_{\xi_i^\ell \to -\infty} \gamma_i^\ell(t) = \gamma_i(t_R)$ which is a continuous distribution. So, for all $j \in [\![1,k_i]\!]$, $\mathscr{L}_{k_i}(\{y_{i,j} = \gamma_i(t_R)\}) = 0$.
Last, since $\lim \psi_i^\ell(t) = \pm\infty$ when $|t_R|$, $|\tau_i|$ or ξ_i^ℓ converge toward $+\infty$, we get the result with Assumption (H 4).

□

Proving (H 5) is a very interesting issue but outside of the scope of this paper. Our conjecture is that a positive and restricted curvature for the shape space M will guarantee that, for all individuals $i \in [\![1,n]\!]$, $c_R, m_R^1, m_R^2, A_{m_R^\perp}$ and s_i are regular. Indeed, we guess that in geodesic shooting, sufficient initial momenta will enforce the trajectory to "go away", provided that the underlying manifold is "kind" enough. Note that the computation of the curvature for both currents and varifolds is still an open problem.

5 Experimental Results

Experimentations are performed for both models introduced above: the piecewise logistic curve model and the piecewise geodesic shape model. In order to validate our model and numerical scheme, we first run experiments on synthetic data for the piecewise logistic model. We then test our estimation algorithm on real data from the Hôpital Européen Georges Pompidou (HEGP – Georges Pompidou European Hospital). A medical paper is under progress to provide a more accurate interpretation of this results. Then, we run experiments on synthetic data for the piecewise geodesic shape model to confirm the performance of our model on more complicated data. Real data for this framework are being collected and preprocessed.

Table 1 *Degree of non-linearity*. Relative errors (expressed as a percentage) for the initial population parameters $\overline{z_{\text{pop}}}^{\text{init}}$, according to the type of data set and the sample size n.

	n	$\delta_{\mathscr{L}}(\overline{\gamma_0^{\text{init}}})$	$\delta_{\mathscr{L}}(\overline{\gamma_0^{\text{escap}}})$	$\delta_{\mathscr{L}}(\overline{\gamma_0^{\text{fin}}})$	$\delta_{\mathscr{L}}(\overline{t_R})$	$\delta_{\mathscr{L}}(\overline{t_1})$
A	50	7.08	17.01	5.94	1.97	1.98
	100	2.93	22.33	3.66	2.40	2.42
	250	2.16	24.06	2.12	3.52	3.54
A*	50	5.63	283.14	1.51	1.03	1.01
	100	3.38	259.25	0.07	4.75	4.76
	250	3.67	269.42	0.41	3.94	3.95
B	50	80.47	2.77	39.78	35.04	35.09
	100	88.17	4.39	51.83	36.14	36.19
	250	83.52	12.91	47.90	33.23	33.27
B*	50	59.25	201.98	33.46	28.85	28.89
	100	74.94	213.96	43.50	30.74	30.78
	250	79.14	229.40	47.30	34.39	34.44

5.1 Univariate Synthetic Data

We generate four types of data set, to put our algorithm in different situations. More precisely, we want to quantify its sensitivity to initialisation, sample size and noise.

5.1.1 Influence of the Initialization

The estimation is performed through the SAEM algorithm (Algorithm 1). This iterative algorithm is proven to converge toward a critical point of the observed likelihood. Therefore, as our model does no imply a convex likelihood, one may end up with a local *maximum* depending on the initialization point and the dynamic of our iterations. This choice of initialization appears crucial. In particular the choice of the initial mean population parameters $\overline{z_{\text{pop}}}^{\text{init}}$.

If our model were linear, the representative curve γ_0 would exactly be the one induced by the mean of the individual trajectories γ_i, *i.e.* the one where $z_{\text{pop}} = \text{mean}_{i\in[\![1,n]\!]}\, z_i$. Following this idea, we set in our experiments

$$\overline{\gamma_0^{\text{init}}}^{\text{init}} = \underset{i\in[\![1,n]\!]}{\text{mean}}\, y_{i,1} \;\; ; \;\; \overline{\gamma_0^{\text{escap}}}^{\text{init}} = \underset{i\in[\![1,n]\!]}{\text{mean}} \min_{j\in[\![1,k_i]\!]} y_{i,j} \;\; ;$$
$$\overline{\gamma_0^{\text{fin}}}^{\text{init}} = \underset{i\in[\![1,n]\!]}{\text{mean}}\, y_{i,k_i} \;\; ; \;\; \overline{t_R}^{\text{init}} = \frac{1}{2}\, \underset{i\in[\![1,n]\!]}{\text{mean}}\, t_{k_i} \;\; ;$$
$$\text{and} \quad \overline{t_1}^{\text{init}} = \underset{i\in[\![1,n]\!]}{\text{mean}}\, t_{k_i} \, .$$

Note that the choice of the initial covariance matrix Σ^{init} and the residual noise σ^{init} does not seem to be very influential. We just demand Σ^{init} to be definite positive.

5.1.2 Influence of the Proposal Variances

The SAEM algorithm is very sensitive to the choice of the proposal variances in the sampling step. Thus, we have to carefully tune these variances in order the mean acceptance ratio to stay around the optimal rate, *i.e.* 24% as we are using a symmetric random walk sampler. To decrease the influence of a bad calibration, we adapt the proposal variances over the iterations in the way of Roberts and Rosenthal [28, 29]: every s$^{\text{th}}$ batch of 50 iterations, we increase or decrease the logarithm of the proposal proposal variances by $\delta(s) = \min\left(0.001, \frac{1}{\sqrt{s}}\right)$ depending on whether the mean associated variable acceptance rate is bigger or smaller than the optimal one. Note that we have also tried to adapt the proposal variances as in Atchadé [3] but the results we obtained were not satisfactory. Actually, it appears numerically that if we want the adaptive procedure to increase the efficiency of our algorithm, we must modify the proposal variance neither too often nor with a too big amplitude of change.

5.1.3 Construction of the Data Sets

For each type of data set, given the corresponding ground truth parameters θ^{true}, we generate three data sets of respective size 50, 100 and 250. Last, to put our algorithm on a more realistic situation, the synthetic individual times are non-periodically spaced and individual sizes vary between 12 and 18.

The first type – A – is said quasilinear in the sense that, for these data sets, the representative trajectory γ_0 is "close" to the mean trajectory described above. Hence, we put our algorithm in a favorable situation where the optimal representative trajectory is close to the initial one. The second type –A* – is a noisy version of A. The noise level is approximately 20% (against 2% for the non-noisy data set A).

On the contrary, the thirds type – B – is built in order to be "truly non-linear": the representative trajectory γ_0 is "far" from the curve built by $\overline{z_{\text{pop}}}^{\text{init}}$. Likewise, the fourth type – B* – is a noisy version of B, with a 20 % noise level.

Table 2 *The piecewise logistic curve model: Fixed effects, variability and residual noise*. Mean (standard deviation) relative errors (expressed as a percentage) for the estimated parameters $\overline{z_{\text{pop}}}^{\text{estim}}$; mean (standard deviation) of KullbackLeibler divergences from Σ^{estim} to Σ^{true}, mean (standard deviation) relative errors (expressed as a percentage) for the individual rupture times $t_R^{i\,\text{estim}}$ and mean estimated residual noise σ^{estim} according to the data set and the sample size n. All over 50 runs.

	n	$\overline{\gamma_0^{\text{init}}}$	$\overline{\gamma_0^{\text{escap}}}$	$\overline{\gamma_0^{\text{fin}}}$	$\overline{t_R}$	$\overline{t_1}$	Σ	t_R^i	σ
A	50	6.03 (0.32)	10.25 (0.50)	3.69 (0.25)	**1.95 (0.13)**	2.43 (0.18)	15.54 (5.17)	**0.49 (0.04)**	**2.03**
	100	2.19 (0.17)	3.28 (0.22)	2.07 (0.18)	**1.69 (0.11)**	1.86 (0.17)	8.45 (2.26)	**0.63 (0.06)**	**1.97**
	250	1.30 (0.10)	1.96 (0.13)	1.53 (0.08)	**0.78 (0.06)**	1.67 (0.09)	9.29 (3.13)	**0.57 (0.60)**	**2.06**
A*	50	3.74 (0.26)	25.73 (1.64)	6.84 (0.40)	**3.32 (0.26)**	3.73 (0.26)	16.52 (19.45)	**4.66 (0.45)**	**19.81**
	100	2.35 (0.15)	12.20 (0.64)	1.35 (0.09)	**2.98 (0.22)**	2.29 (0.18)	12.86 (4.26)	**3.85 (0.32)**	**19.03**
	250	1.70 (0.12)	3.94 (0.29)	1.33 (0.09)	**1.36 (0.10)**	1.51 (0.10)	6.72 (2.44)	**3.98 (0.32)**	**20.07**
B	50	71.13 (1.33)	100.24 (8.09)	90.73 (2.54)	**7.78 (0.56)**	46.39 (1.32)	16.53 (7.72)	**5.89 (3.45)**	**3.07**
	100	58.73 (0.98)	58.88 (3.00)	84.99 (1.42)	**8.13 (0.57)**	42.06 (1.04)	13.59 (5.42)	**4.44 (1.93)**	**2.14**
	250	67.49 (0.47)	23.12 (1.54)	57.82 (0.74)	**6.01 (0.33)**	38.09 (0.36)	22.24 (9.77)	**4.96 (1.93)**	**2.49**
B*	50	41.61 (1.26)	29.86 (2.53)	46.38 (1.60)	**9.04 (0.58)**	29.90 (0.58)	27.62 (17.71)	**14.32 (4.06)**	**19.93**
	100	60.39 (0.81)	28.43 (2.06)	58.35 (1.07)	**8.11 (0.54)**	29.75 (0.50)	23.98 (18.07)	**13.97 (3.71)**	**20.56**
	250	55.89 (0.74)	15.56 (0.98)	59.90 (0.58)	**3.26 (0.25)**	39.28 (0.43)	17.70 (5.35)	**11.57 (2.42)**	**21.38**

To measure this degree of non-linearity, we introduce the ratio $\delta_{\mathscr{L}}(\overline{z_{\text{pop}}})$ which is the relative error of $\overline{z_{\text{pop}}}^{\text{init}}$:

$$\delta_{\mathscr{L}}(\overline{z_{\text{pop}}}) = \frac{\|\overline{z_{\text{pop}}}^{\text{init}} - \overline{z_{\text{pop}}}^{\text{true}}\|}{\|\overline{z_{\text{pop}}}^{\text{true}}\|}.$$

Table 1 compiles this ratio for every data set, and for every parameter in $\overline{z_{\text{pop}}}$. In particular, the initialization of $\overline{\gamma_0^{\text{escap}}}$ is in itself a challenge and very sensitive to the noise in the data set: even in the quasilinear case, $\overline{\gamma_0^{\text{escap}}}^{\text{init}}$ is quite far from $\overline{\gamma_0^{\text{escap}}}^{\text{true}}$.

5.1.4 Estimation of the Fixed Effects

Table 2 displays the relative errors for the estimated population parameters. In most case, these errors decrease with the size of the data set. More specific to our model, we observe that these errors are correlated to the subjective linearity of the model. With the exception of $\overline{\gamma_0^{\text{escap}}}$, the errors for estimating population parameters grow linearly with the non-linearity of the model. We suppose that the difference of scale between $\overline{\gamma_0^{\text{escap}}}$ and the others can, at least partly, explain this phenomena: $\overline{\gamma_0^{\text{escap}}}$ is about a few tens of units ; $\overline{\gamma_0^{\text{init}}}$, $\overline{\gamma_0^{\text{fin}}}$ and $\overline{t_R}$ about a few hundreds and $\overline{t_1}$ about one thousand. Thus, a same absolute error will lead to markedly different relative error.

As Table 1 displays the relative error for for the initial population parameters $\overline{z_{\text{pop}}}^{\text{init}}$ and Table 2 the relative errors for the estimated population parameters $\overline{z_{\text{pop}}}^{\text{estim}}$, by comparing this two tables, we are able to quantify the contribution of the estimation-procedure in the knowledge of the population parameters. The first point to note is that this relative error generally decrease. Specifically, the population parameters are well-learned in quasilinear cases (data sets A and A*) and in particular in large data set ($n = 250$). Then, the algorithm we propose is not noise-sensitive: errors for non-noisy and noisy versions of a same type of data set are notably the same. And even better, for the non-linear data sets, the estimation is better performed in the noisy case than in the non-noisy one. It seems that the presence of noise helps the algorithm not to get stuck in potential well.

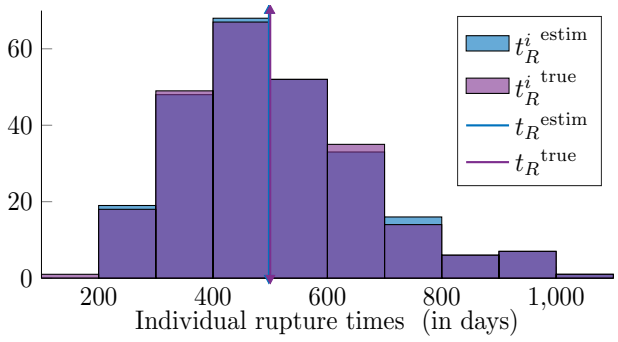
Hence, the degree of non-linearity in the data set seems to play a significant role in the estimation of the population parameters. To be certain that the poor estimation of $\overline{z_{\text{pop}}}$ when the ratio $\delta_{\mathscr{L}}(\overline{z_{\text{pop}}})$ is too big is due to the non-linearity of the data set and not to a bad initialization, we have also performed estimations by assigning $\theta^{\text{init}} = \theta^{\text{true}}$. The results were better but not so significantly.

Last, note that the representative rupture time $\overline{t_R}$ is well-estimated, no matter the subjective linearity of the data set. In the view of chemotherapy monitoring, well-estimating the rupture time, which correspond to an escapement from the treatment, is very important.

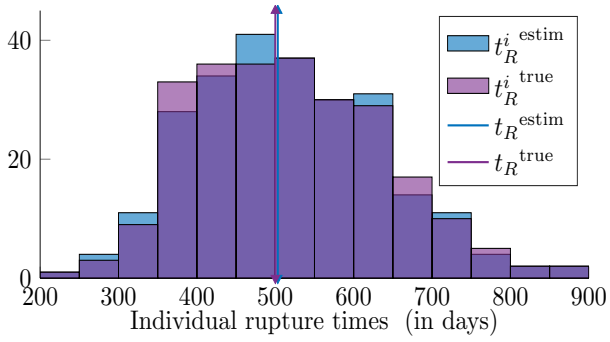
5.1.5 Estimation of the Inter-Individual Variability

In the target of our application, the covariance matrix Σ gives a lot of information on the health status of a patient: pace and amplitude of tumor progression, individual rupture times, *etc*. Therefore, we have to pay special attention to the estimation of Σ.

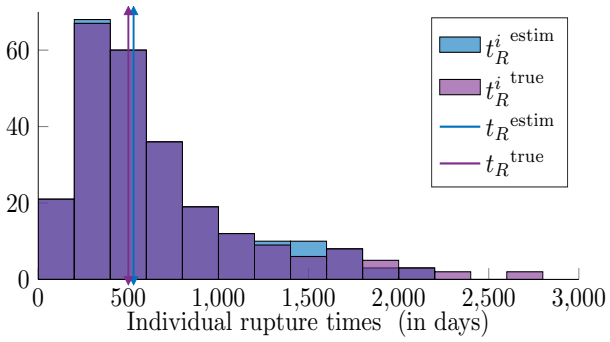
Much as the representative trajectory is not always good-estimated, our algorithm always allows a well-understanding of the inter-individual variability. We present at Table 2 the Kullback-Leibler divergence from Σ^{estim} to Σ^{true}, the relative error of the individual rupture times and the estimated residual noise. As for the estimation of the population parameters, errors decrease with the sample size n and are not significantly different between noisy and non-noisy versions



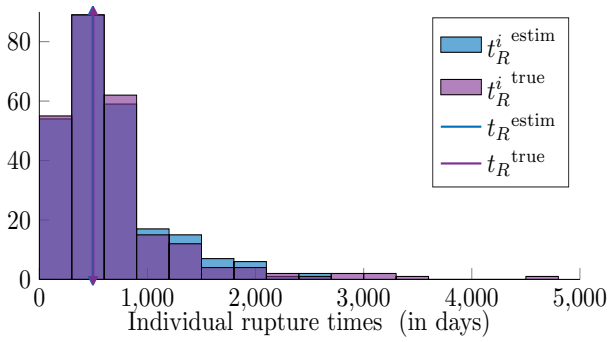
(a) A – *Quasi-linear*



(b) A* – *Noisy and quasi-linear*

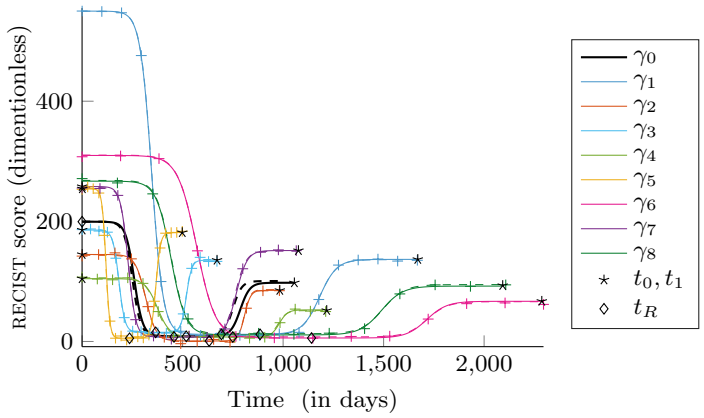


(c) B – *Strongly non-linear*

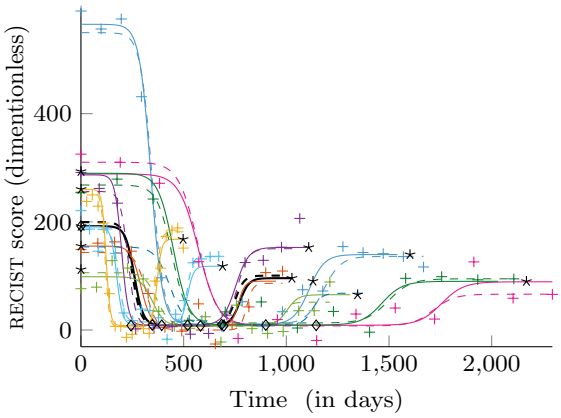


(d) B* – *Noisy and non-linear*

Fig. 5 *The piecewise logistic curve model: Distribution of the individual rupture times*. Each subfigure compares the distribution of the (mean of the) estimated individual rupture times $t_R^{i\ \text{estim}}$ and the distribution of the true individual rupture times $t_R^{i\ \text{true}}$. In bold line, the estimated average rupture time t_R^{estim} and the true average rupture time t_R^{true} are relatively close to each other. $n = 250$.



(a) A – *Quasi-linear*



(b) A* – *Noisy and quasi-linear*

Fig. 6 *The piecewise logistic curve model: Qualitative performance of the estimation and robustness to noise of the MAP estimator*. On both figures, the estimated trajectories are in plain lines and the target curves in dashed lines. The (noisy) observations are represented by crosses. The representative path is in bold black line, the individuals in colour. $n = 250$.

of a same type of data set. Moreover, in that case, the errors seem to not rely on the subjective linearity of the data set.

Moreover, the individual rupture times $\overline{t_R}$, the residual noise σ is always well-estimated.

5.1.6 Reconstruction of the Individual Trajectories

Figure 5 illustrates the well-understanding of the variance within the population, including for the non-linear data set. Determining accurate individual rupture time t_R^i is all the

most important as, in the aim of chemotherapy monitoring, these times are related to an escape of the patient's response to treatment.

An important point was to allow a lot of different individual behaviors. In our synthetic example, Figure 2 illustrates this variability. From a single representative trajectory (γ_0 in bold plain line), we can generate individuals who are cured at the end (dot-dashed lines: γ_3 and γ_4), some whose response to the treatment is bad (dashed lines: γ_5 and γ_6), some who only escape (no positive response to the treatments – dotted lines: γ_7). Likewise, we can generate "patients" with only positive responses or no response at all. The case of individual 4 is interesting in practice: the tumor still grows but so slowly that the growth is negligible, at least in the short-run.

Figure 6 illustrates the qualitative performance of the estimation. We are notably able to understand various behaviors and fit subjects which are far from the characteristic path. Moreover, the noise seems to not reduce the quality of the estimation. We represent only five individuals but 250 subjects have been used to perform the estimation.

5.2 Metastatic Kidney Cancer Monitoring

The algorithm is now run on RECIST score of real patients suffering from kidney cancer. The estimation is performed over a cohort of 176 patients of the HEGP. There is an average of 7 visits per subjects (min: 3, max: 22), with an average duration of 90 days between consecutive visits. We present here a run with a low residual standard variation with respect to the amplitude of the trajectories and complexity of the data set: $\sigma = 9.10$.

Figure 7 illustrates the qualitative performance of the model on ten patients. Although we cannot explain all the paths of progression, the algorithm succeeds in fitting various types of curves: from the curve γ_6 which is flat to the curve γ_3 which is spiky. From Figure 8, it seems that the rupture times occur early in the progression in average.

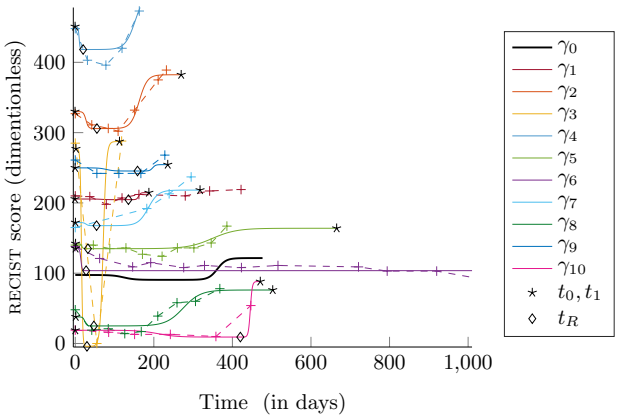


Fig. 7 *RECIST score of patients from the HEGP*. We keep conventions of the previous figures. We represent only 10 patients among the 176.

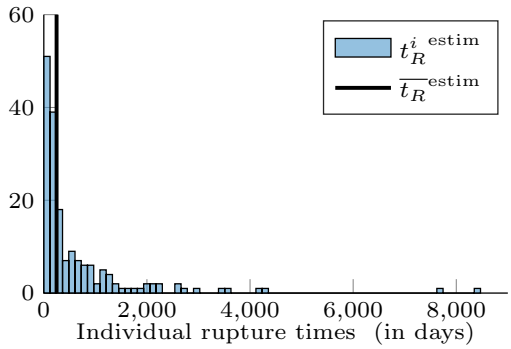


Fig. 8 *Individual rupture times t_R^i (in days)*. Histogram of the rupture times t_R^i for this run. In black bold line, the estimated average rupture time $\overline{t_R}$ is a good estimate of the average of the individual rupture times although there exists a large range of escape.

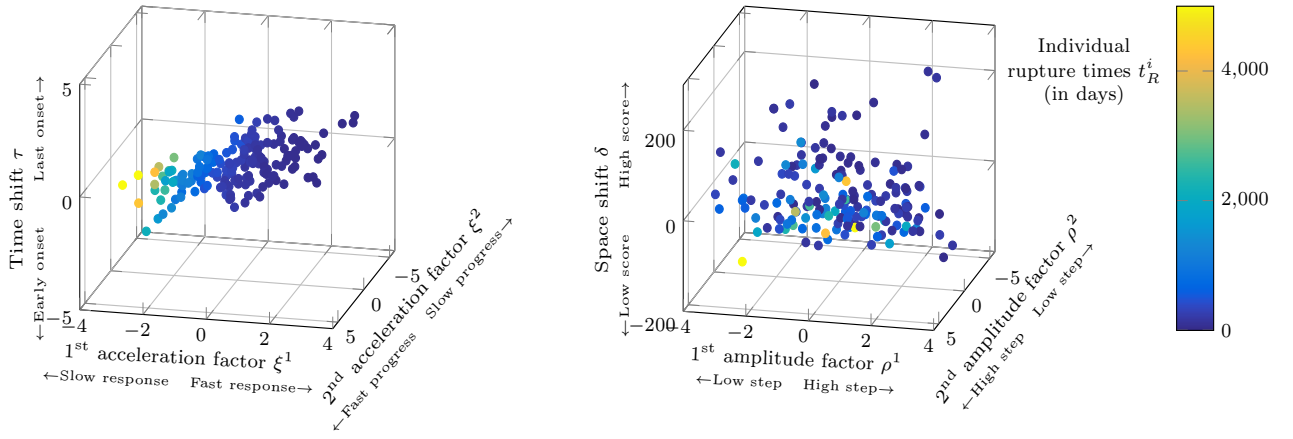
In Figure 9, we plot the individual estimates of the random effects (obtained from the last iteration) in comparison to the individual rupture times. Even though the parameters which lead the space warp, *i.e.* ρ_i^1, ρ_i^2 and δ_i are correlated, the correlation with the rupture time is not clear. In other words, the volume of the tumors seems to not be relevant to evaluate the escapement of a patient. On the contrary, which is logical, the time warp strongly impacts the rupture time.

5.3 Shape Synthetic Data

The dataset consists of 20 synthetic sequences of 3D-shapes built in accordance to the piecewise geodesic shape model described at the paragraph 4.2. The estimation is still performed through the MCMC-SAEM algorithm (Algorithm 1). Real data are not yet available as the segmentation of the tumor has to be done manually, which is complex and time-consuming. This study will motivate new segmentations for future works.

The control points used to construct the data are chosen to be regularly distributed. Thus, the algorithm has no reason to return the same control points: on the contrary, it will return more relevant control points. As momenta and control points share a single dynamic, we rather evaluate the performances on the reconstruction relative error which summaries the goodness of fit of our algorithm.

Table 3 displays the relative errors for the estimated representative rupture shape, representative rupture time and individual rupture times. We emphasis the well-estimation of the rupture times t_R and $(t_R^i)_{i\in[\![1,n]\!]}$, which is critical in the target to our application to chemotherapy monitoring. We also provide the relative errors of reconstruction, *i.e.* the relative residual distances between the estimated trajectories and and their corresponding paths in the data set for both the representative path and the individuals ones. Figure 10



(a) The time warp.

(b) The space warp.

Fig. 9 *Individual random effects*. Fig. 9a: log-acceleration factors ξ_i^1 and ξ_i^2 against times shifts τ_i. Fig. 9b: log-amplitude factors ρ_i^1 and ρ_i^2 against space shifts δ_i. In both figures, the colour corresponds to the individual rupture time t_R^i. These estimates hold for the same run as Fig. 7.

Table 3 *The piecewise geodesic shape model*. Relative errors for the representative rupture shape $\overline{y_R}$ and the representative rupture times $\overline{t_R}$. Mean (standard deviation) relative errors for the individual rupture time $t_R^{i\,\text{estim}}$, within the population. Error of reconstruction for the template and mean (standard deviation) os the error of resonstruction for the individuals. All expressed as a percentage and for a type run.

$\overline{y_R}$	$\overline{t_R}$	Template reconstruction	t_R^i	Individuals reconstruction
1.30	0.01	9.72	0.31 (0.41)	7.94 (5.91)

illustrates the qualitative performance of the reconstruction of the template and Figure 11 the qualitative performance of the reconstruction of subject type. Mainly, the reconstruction turns out to be very efficient.

6 Discussion and Perspective

We have proposed a coherent statistical framework for the spatio-temporal analysis of piecewise geodesic manifold valued measurements. This model allows each individual to have his own intrinsic geometry and his own time parametrization. Unlike previous similar works [31, 32], it allows for piecewise geodesic trajectories. Relaxing the classical geodesic assumption widen the application scope of the model in biology and medicine. The model is built in a hierarchical way as a non-linear mixed effects model whose fixed effects define a representative trajectory of the global evolution in the space of measurements and random effects account for the spatio-temporal variability of the trajectories at the individual level.

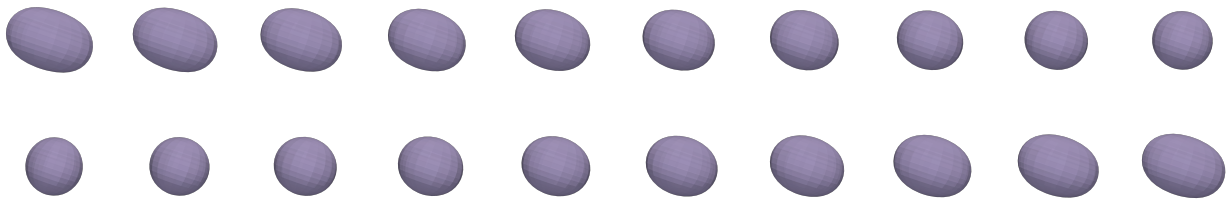
Estimation was formulated as a well-defined MAP problem and numerically performed through the MCMC-SAEM algorithm. Experimentations have highlighted the robustness of our model to noise and its performance in catching individual behaviors. We believe that the complexity of our model ensures its practical identifiability, even if it is not structurally identifiable [22]. Besides, as the posterior likelihood is not convex, the MAP could be difficult to determine numerically. Future work focuses on exploring some possible improvement of the numerical scheme.

Our model can be applied to a wide variety of situations and data sets. In particular, we can address medical follow-up such as neurodegenerative diseases or chemotherapy monitoring. The example of chemotherapy monitoring is especially interesting in a modeling perspective as the patients are treated and tumors may respond, stabilize or progress during the treatment, with different conducts for each phase. At the age of personalized medicine, to give physicians decision support systems is really important. Therefore learning correlations between phases is crucial. This has been taken into account in our experimentations. More generally, the inter-individual variability allows us to personalize the model to new patient and thus perform predictive medicine.

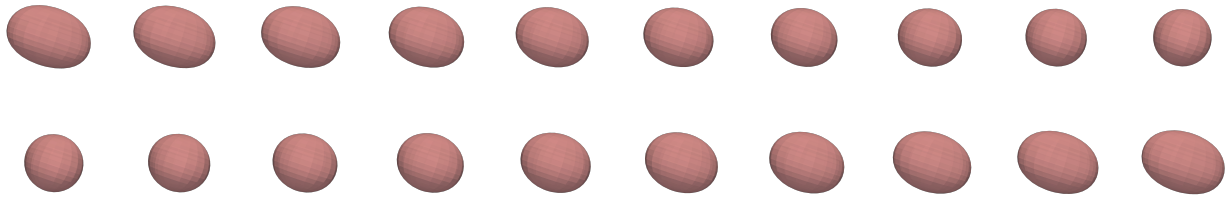
A Proof of the Consistency Theorem for Bounded Population Variable

The proof of the theorem relies on several lemmas. Lemma 4 is the heart of the proof: we control here the behavior of the log-likelihood at the boundary points of the parameters space Θ_*^ω and prove that this set is non-empty. It is based on Lemma 3 which states the integrability of the supremum over the parameter space of the positive part of the log-likelihood. Lemma 2 is derived from Allassonnière et al. [1]. We transpose the proof of the cited article here (with few more details) as this lemma is critical in the proof of Lemma 3 and not such classical.

In the following, we freely (and without reminder) use the notations introduced in Section 3.2. Moreover, (H 1), (H 2), (H 3), (H 4)



(a) Template used to generate the dataset



(b) Template estimated after 600 iterations

Fig. 10 *The piecewise shape model: Reconstruction of the template*. Evolution of the template over time. In purple (Fig. 10a), the template used for the generation of the dataset ; In red (Fig. 10b), the one estimated by the algorithm.

(a) Samples of the dataset: an observed individual trajectory

(b) Estimation of the same individual trajectory, 600 iterations

Fig. 11 *The piecewise shape model: Reconstruction of the individual trajectories*. Evolution of a standard subject over time. We keep the same convention than at Figure 10: Fig. 11a (in purple) shows the shooting of an individual evolution path and Fig. 11b (in red)) the corresponding reconstructed one.

and (H 5) refer to the hypothesis of the consistency theorem (Theorem 2, page 7).

A.1 Lemmas

We first recall that the minimal number of balls of radius $r \in \mathbb{R}_+^*$ required to cover a compact set $K \in \mathbb{R}^p$ is bounded from above by $\left(\frac{\mathscr{D}\mathrm{iam}(K)}{r}\right)^p$.

Lemma 2 (Preliminary of measure theory) *Let $p < q$ be two integers. Then, for any differentiable map $f\colon \mathbb{R}^p \to \mathbb{R}^q$ and any compact subset K of $\mathbb{R}^p$, there exists a constant λ which depends only on p and q such that*

$$\int_{\mathbb{R}^q \setminus f(K)} \log^+ \frac{1}{d\big(\boldsymbol{y}, f(K)\big)}\, \mathrm{d}\boldsymbol{y} \;<\; \lambda \left(\sup_K \|\mathscr{D}f\| + 2\right)^q \mathscr{D}iam(K)^p\,,$$

where d is the euclidean distance on $\mathbb{R}^q$, $\mathscr{D}f$ the differential of f and $\mathscr{D}iam(K)$ the diameter of the compact K. Especially,

$$\int_{\mathbb{R}^q \setminus f(K)} \log^+ \frac{1}{d\big(\boldsymbol{y}, f(K)\big)}\, \mathrm{d}\boldsymbol{y} < +\infty\,.$$

Proof For all $\rho, \rho_1, \rho_2 \in \mathbb{R}_+^*$, $\rho_1 < \rho_2$, let

$$M_{\rho_1,\rho_2} = \big\{\boldsymbol{y} \in \mathbb{R}^q \mid \rho_1 \leqslant d\big(\boldsymbol{y}, f(K)\big) \leqslant \rho_2\big\} \quad \text{and} \quad M_\rho = M_{0,\rho}\,.$$

For all $\rho \in \mathbb{R}_+^*$, due to the compactness of K, there exists a finite set $\Lambda_\rho \subset K$ such that $K \subset \bigcup_{\boldsymbol{x}\in\Lambda_\rho} \mathscr{B}(\boldsymbol{x}, \rho)$ and $|\Lambda_\rho| \leqslant \left(\frac{\mathscr{D}\mathrm{iam}(K)}{\rho}\right)^p$. Let $\tau = \sup_K \|\mathscr{D}f\|$. According to the mean value theorem, $M_{0,\rho} \subset \mathscr{B}\big(f(\boldsymbol{x}), (\tau+2)\rho\big)$ and

$$\begin{aligned}\mathscr{L}_q(M_\rho) &\leqslant \sum_{\boldsymbol{x}\in\Lambda_\rho} \mathscr{L}_q\Big(\mathscr{B}\big(f(\boldsymbol{x}), (\tau+2)\rho\big)\Big) \\ &\leqslant \frac{\sqrt{\pi}^p\,(\tau+2)^p}{\Gamma\left(\frac{p}{2}+1\right)} \times \big(\mathscr{D}\mathrm{iam}(K)\big)^p \times \rho^{q-p}\,.\end{aligned}$$

Let $s \in]0,1[$. Then, from the Abel transformation,

$$\begin{aligned}\int_{\mathbb{R}^q \setminus f(K)} \log^+ \frac{1}{d\big(\boldsymbol{y}, f(K)\big)}\,\mathrm{d}\boldsymbol{y} &= \sum_{n=0}^{+\infty} \int_{M_{s^{n+1}}\setminus M_{s^n}} \log^+ \frac{1}{d\big(\boldsymbol{y}, f(K)\big)}\,\mathrm{d}\boldsymbol{y} \\ &\leqslant \sum_{n=0}^{+\infty} \log \frac{1}{s^{n+1}} \left[\mathscr{L}_q(M_{s^n}) - \mathscr{L}_q(M_{s^{n+1}})\right] \\ &\leqslant -\log(s) \sum_{n=0}^{+\infty} \mathscr{L}_q(M_{s^n})\,.\end{aligned}$$

Hence the result as $s \in]0,1[$. □

Lemma 3 *Assume* (H 1)*,* (H 3)*,* (H 4) *and*

(H′ 5) *Bounded regular variables implies bounded trajectories: For all individuals $i \in [\![1,n]\!]$, if there exists $b \in \mathbb{R}$ such that $\|(\boldsymbol{z}_{pop}^{reg}, \boldsymbol{z}_i^{reg})\|_\infty < b$ then there exists $R \in \mathbb{R}_+^*$ such that $\|\gamma_i(\boldsymbol{z}_{pop}, \boldsymbol{z}_i)\|_\infty < R$.*

Then, for any such ℓ,

$$\mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\sup_{\theta\in\Theta}\left(\sum_{i=1}^{\ell} \log q(\boldsymbol{y}_i|\theta)\right)^+\right] \;<\; +\infty.$$

Proof Let $i \in [\![1,\ell]\!]$, $\Gamma_i = \mathscr{I}m(\gamma_i)$ and $\Gamma^\ell = \prod_{i=1}^{\ell} \Gamma_i$. For all $\theta \in \Theta^\omega$,

$$\begin{aligned}q(\boldsymbol{y}_i|\theta) =\; & \frac{1}{(\sigma\sqrt{2\pi})^{k_i}} \int_{\bar{\mathscr{Z}}_i} \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y}_i - \gamma_i(\boldsymbol{z}_{\mathrm{pop}}, \boldsymbol{z}_i)\|_2^2\right) \\ & \qquad q(\boldsymbol{z}_{\mathrm{pop}}, \boldsymbol{z}_i|\theta)\,\mathrm{d}(\boldsymbol{z}_{\mathrm{pop}}, \boldsymbol{z}_i) \\ \leqslant\; & \frac{1}{(\sigma\sqrt{2\pi})^{k_i}} \exp\left(-\frac{1}{2\sigma^2} d(\boldsymbol{y}_i, \Gamma_i)^2\right),\end{aligned}$$

where d denotes the Euclidean distance on $\mathbb{R}^{k_i}$. Thus for all $\theta \in \Theta^\omega$,

$$\sum_{i=1}^{\ell} \log q(\boldsymbol{y}_i|\theta) \leqslant -\frac{k^\ell}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2} d(\boldsymbol{y}^\ell, \Gamma^\ell)^2,$$

where d denotes now the Euclidean distance on $\mathbb{R}^{k^\ell}$, $k^\ell = \sum_{i=1}^{\ell} k_i$. As the right hand side is maximized for $\sigma^2 = \frac{1}{k^\ell} d(\boldsymbol{y}^\ell, \Gamma^\ell)^2$, there exists a constant $\lambda \in \mathbb{R}_+^*$ such that

$$\sup_{\theta\in\Theta}\left(\sum_{i=1}^{\ell} \log q(\boldsymbol{y}_i|\theta)\right)^+ \leqslant \lambda + k^\ell \log^+ \frac{1}{d(\boldsymbol{y}^\ell, \Gamma^\ell)}\,.$$

1. Assume there exists $i_0 \in [\![1,n]\!]$ such that that $\|(\boldsymbol{z}_{\mathrm{pop}}^{\mathrm{reg}}, \boldsymbol{z}_{i_0}^{\mathrm{reg}})\|_\infty \geqslant b$ for all $b \in \mathbb{R}$.
   For all $r_1, r_2 \in \mathbb{R}$ we define a compact subset Γ_{r_1,r_2}^ℓ of Γ^ℓ by setting

$$\begin{aligned}\bar{\mathscr{I}}(r_1,r_2) = \Big\{\boldsymbol{z}^\ell \in \mathbb{R}^{p^\ell} \mid\; & r_1 \leqslant \|(\boldsymbol{z}_{\mathrm{pop}}^{\mathrm{reg}}, \boldsymbol{z}_i^{\mathrm{reg}})_{i\in[\![1,\ell]\!]}\|_\infty, \\ & \|(\boldsymbol{z}_{\mathrm{pop}}^{\mathrm{crit}}, \boldsymbol{z}_i^{\mathrm{crit}})_{i\in[\![1,\ell]\!]}\|_\infty \leqslant r_2\Big\}\end{aligned}$$

   and $\Gamma_{r_1,r_2}^\ell = \big\{\gamma^\ell(\boldsymbol{z}^\ell) \mid \boldsymbol{z}^\ell \in \bar{\mathscr{I}}(r_1,r_2)\big\}$.

   Especially, $\lim_{r_2\to\infty} \Gamma_{0,r_2}^\ell = \Gamma^\ell$. Moreover, γ^ℓ is differentiable *a.e.*, at least one-side differentiable everywhere and there exists $\tau \in \mathbb{R}$ such that $\sup_{\mathbb{R}^{p^\ell}} \|\mathscr{D}_{\boldsymbol{z}^\ell}\gamma^\ell\| < \tau$. So, according to Lemma 2, for all $r_1, r_2 \in \mathbb{R}$, there exists $\mu \in \mathbb{R}$ which depends only on p^ℓ and k^ℓ such that $\mathbb{E}\left[\log^+ \frac{1}{d(\boldsymbol{y}^\ell, \Gamma_{r_1,r_2}^\ell)}\right] < \mu\,(\tau+2)^{k^\ell}\, {r_2}^{p^\ell}$. As in the proof of Lemma 2, we set $\overline{\Gamma_{r_1,r_2}^\ell} = \{\boldsymbol{y}^\ell \in \mathbb{R}^{k^\ell} \mid d(\boldsymbol{y}^\ell, \Gamma_{r_1,r_2}^\ell) \leqslant 1\}$ and we have for all $r_1, r_2 \in \mathbb{R}$,

$$\begin{aligned}\int_{\mathbb{R}^{k^\ell}} \log^+ \frac{1}{d(\boldsymbol{y}^\ell, \Gamma_{r_1,r_2}^\ell)} P(\mathrm{d}\boldsymbol{y}^\ell) &= \int_{\overline{\Gamma_{r_1,r_2}^\ell}} \log^+ \frac{1}{d(\boldsymbol{y}^\ell, \Gamma_{r_1,r_2}^\ell)} P(\mathrm{d}\boldsymbol{y}^\ell) \\ &\leqslant \bar{\mu}\, {r_2}^{p^\ell} \sup_{\overline{\Gamma_{r_1,r_2}^\ell}} P(\boldsymbol{y}^\ell)\,,\end{aligned}$$

   where $\bar{\mu} = \mu\,(\tau+2)^{k^\ell} \in \mathbb{R}$. Let $R_1, R_2 \in \mathbb{N}$ such that $K \subset \bar{\mathscr{B}}(0,R_1)$ and $R_1 < R_2$. By definition of the distance to a subset, it comes that

$$\begin{aligned}\mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log^+ \frac{1}{d(\boldsymbol{y}^\ell, \Gamma_{0,R_2}^\ell)}\right] \;\leqslant\; & \bar{\mu}\, {R_1}^{p^\ell} \sup_{\overline{\Gamma_{0,R_1}^\ell}} P(\boldsymbol{y}^\ell) \\ & + \bar{\mu} \sum_{r=R_1}^{R_2-1} (r+1)^{p^\ell} \sup_{\overline{\Gamma_{r,r+1}^\ell}} P(\boldsymbol{y}^\ell)\,.\end{aligned}$$

   The first term is finite as $P(\mathrm{d}\boldsymbol{y})$ is continuous. Besides, if $\boldsymbol{y}^\ell \in \overline{\Gamma_{r,r+1}^\ell}$, there exists $\boldsymbol{z}^\ell \in \bar{\mathscr{I}}(r,r+1)$ such that $\|\gamma^\ell(\boldsymbol{z}^\ell) - \boldsymbol{y}^\ell\|_\infty \leqslant 1$. Let $i \in [\![1,n]\!]$ and $v \in [\![1, p_{\mathrm{pop}}^{\mathrm{reg}} + p_{\mathrm{ind}}^{\mathrm{reg}}]\!]$ so that $\|(\boldsymbol{z}_{\mathrm{pop}}^{\mathrm{reg}}, \boldsymbol{z}_i^{\mathrm{reg}})_{i\in[\![1,n]\!]}\|_\infty = |(\boldsymbol{z}_{\mathrm{pop}}^{\mathrm{reg}}, \boldsymbol{z}_i^{\mathrm{reg}})_v|$. Such a couple exists due to the existence of i_0. Moreover, there exists $a_{i,v}\big((\boldsymbol{z}_{\mathrm{pop}}^{\mathrm{reg}}, \boldsymbol{z}_i^{\mathrm{reg}})_{-v}\big)$ and $b_{i,v}\big((\boldsymbol{z}_{\mathrm{pop}}^{\mathrm{reg}}, \boldsymbol{z}_i^{\mathrm{reg}})_{-v}\big)$ as in (H 4) and by definition of $\boldsymbol{z}^\ell$ and the infinite norm,

$$\begin{aligned}\|\boldsymbol{y}^\ell\|_\infty \;\geqslant\; \|\gamma^\ell(\boldsymbol{z}^\ell)\|_\infty - 1 &\geqslant \|\gamma_i(\boldsymbol{z}_{\mathrm{pop}}, \boldsymbol{z}_i)\|_\infty - 1 \\ &\geqslant a_{i,v}\,|(\boldsymbol{z}_{\mathrm{pop}}^{\mathrm{reg}}, \boldsymbol{z}_i^{\mathrm{reg}})_v| + b_{i,v} - 1 \\ &\geqslant a_{i,v} \times r + b_{i,v} - 1\,.\end{aligned}$$

   Consequently,

$$\sup_{\overline{\Gamma_{r,r+1}^\ell}} P(\boldsymbol{y}^\ell) \leqslant \sup\{P(\boldsymbol{y}^\ell) \mid \|\boldsymbol{y}^\ell\|_\infty \geqslant a_{i,v} \times r + b_{i,v} - 1\}$$

   and the series $\sum (r+1)^{p^\ell} \sup_{\overline{\Gamma_{r,r+1}^\ell}} P(\boldsymbol{y}^\ell)$ converge since $P(\mathrm{d}\boldsymbol{y})$ has a polynomial decay tail of degree bigger than $p^\ell + 1$ apart from K by assumption (H 3).
2. Assume that there exists $b \in \mathbb{R}$ such that $\|(\boldsymbol{z}_{\mathrm{pop}}^{\mathrm{reg}}, \boldsymbol{z}_i^{\mathrm{reg}})\|_\infty \leqslant b$ for all $i \in [\![1,n]\!]$. Then, by assumption (H′ 5), there exists $R \in \mathbb{R}_+^*$ such that for all i, $\|\gamma_i(\boldsymbol{z}_{\mathrm{pop}}, \boldsymbol{z}_i)\|_\infty < R$. In particular, $\|\gamma^\ell(\boldsymbol{z}^\ell)\|_\infty < R$ and $\Gamma^\ell \subset \bar{\mathscr{B}}(0,R)$. Thus,

$$\mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log^+ \frac{1}{d(\boldsymbol{y}^\ell, \Gamma^\ell)}\right] \;\leqslant\; \mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log^+ \frac{1}{d\big(\boldsymbol{y}^\ell, \bar{\mathscr{B}}(0,R)\big)}\right].$$

   Yet, by still denoting $\overline{\bar{\mathscr{B}}(0,R)} = \{\boldsymbol{y}^\ell \in \mathbb{R}^{k^\ell} \mid d\big(\boldsymbol{y}^\ell, \bar{\mathscr{B}}(0,R)\big) \leqslant 1\}$ and applying Lemma 2 to the compact $K = \bar{\mathscr{B}}(0,R)$ and $f = Id$, there exists $\mu \in \mathbb{R}$ such that

$$\begin{aligned}&\mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log^+ \frac{1}{d\big(\boldsymbol{y}^\ell, \bar{\mathscr{B}}(0,R)\big)}\right] \\ &\qquad = \int_{\overline{\bar{\mathscr{B}}(0,R)}} \log^+ \frac{1}{d\big(\boldsymbol{y}^\ell, \bar{\mathscr{B}}(0,R)\big)} P(\mathrm{d}\boldsymbol{y}^\ell) \\ &\qquad \leqslant \mu\, 3^{k^\ell}\, R^{p^\ell} \sup_{\overline{\bar{\mathscr{B}}(0,R)}} P(\boldsymbol{y}^\ell) \;<+\infty.\end{aligned}$$

Finally, in both cases, $\mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\sup_{\theta\in\Theta}\left(\sum_{i=1}^{\ell} \log q(\boldsymbol{y}_i|\theta)\right)^+\right] < +\infty$. □

Lemma 4 *Assume* (H 1)*,* (H 3)*,* (H 4) *and* (H 5)*. Let*

$$\overline{\mathscr{S}_{p_{ind}}^+(\mathbb{R})} = \mathscr{S}_{p_{ind}}^+(\mathbb{R}) \cup \{\infty\}$$

be the one point Alexandrov compactification of $\mathscr{S}_{p_{ind}}^+(\mathbb{R})$ and consider the compactification of the parameter space Θ^ω

$$\overline{\Theta^\omega} = \left\{\theta = (\overline{\boldsymbol{z}_{pop}}, \Sigma, \sigma) \in \mathbb{R}^{p_{pop}} \times \overline{\mathscr{S}_{p_{ind}}^+(\mathbb{R})} \times \overline{\mathbb{R}^+} \mid \|\overline{\boldsymbol{z}_{pop}}\| \leqslant \omega\right\},$$

where $\overline{R^+} = [0, +\infty[\cup \{+\infty\}$. Then, we have for all $\omega \in \mathbb{R}$,

(C 1) *$P(\mathrm{d}\boldsymbol{y}^\ell)$ almost surely, for any sequence $\theta_\kappa = \big(\overline{\boldsymbol{z}_{pop}}_\kappa, \Sigma_\kappa, \sigma_\kappa\big)$ of elements from Θ^ω such that $\lim_{\kappa\to\infty}\theta_\kappa \in \overline{\Theta^\omega}\setminus\Theta^\omega$,*

$$\lim_{\kappa\to\infty}\sum_{i=1}^{\ell}\log q\left(\boldsymbol{y}_i|\theta_\kappa\right) = -\infty\,;$$

(C 2) *For any sequence $(\theta_\kappa)\in{\Theta^\omega}^{\mathbb{N}}$ such that $\lim_{\kappa\to\infty}\theta_\kappa\in\overline{\Theta^\omega}\setminus\Theta^\omega$,*

$$\lim_{\kappa\to\infty}\mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log q(y|\theta_\kappa)\right] = -\infty\,;$$

(C 3) *The mapping $\theta\mapsto\mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\log q(y|\theta)\right]$ is continuous on Θ^ω and $\Theta^\omega_*\neq\varnothing$.*

Proof We recall that a sequence $(\Sigma_\kappa)_{\kappa\in\mathbb{N}}$ of $\overline{\mathscr{S}^+_{p_{\mathrm{ind}}}(\mathbb{R})}$ converge toward the point ∞ if it eventually steps out of every compact subset of $\mathscr{S}^+_{p_{\mathrm{ind}}}(\mathbb{R})$. Let prove the three points in order.

1. As

$$\overline{\Theta^\omega}\setminus\Theta^\omega = \Big\{\,(\Sigma,\sigma)\in\overline{\mathscr{S}^+_{p_{\mathrm{ind}}}(\mathbb{R})}\times\overline{\mathbb{R}^+}\;\Big|\;\|\Sigma\| = +\infty\wedge\|\Sigma^{-1}\| = +\infty\wedge\sigma\in\{0,+\infty\}\Big\}$$

we proceed by disjunction. Let

$$\forall\kappa\in\mathbb{N},\quad \theta_\kappa = (\overline{\boldsymbol{z}_{pop}}_\kappa,\Sigma_\kappa,\sigma_\kappa)\in\Theta^\omega\,.$$

(i) Assume that, up to extraction of a subsequence, $\|\Sigma_\kappa\|\to\infty$ or $\|\Sigma_\kappa^{-1}\|\to\infty$.
Let $M = \|\boldsymbol{y}^\ell\|_\infty$. For all individuals $i\in[\![1,n]\!]$ and all $\kappa\in\mathbb{N}$, the marginal density of $\boldsymbol{y}_i$ given θ_κ is given by :

$$q(\boldsymbol{y}_i|\theta_\kappa) = \frac{1}{(\sigma_\kappa\sqrt{2\pi})^{k_i}}\int_{\mathscr{Z}_{\mathrm{pop}}\times\mathscr{Z}_i}\exp\left(-\frac{1}{2\sigma_\kappa^2}\|\boldsymbol{y}_i-\gamma_i(\boldsymbol{z}_{\mathrm{pop}},\boldsymbol{z}_i)\|_2^2\right) q(\boldsymbol{z}_{\mathrm{pop}},\boldsymbol{z}_i|\theta_\kappa)\,\mathrm{d}(\boldsymbol{z}_{\mathrm{pop}},\boldsymbol{z}_i)\,.$$

Let $x\geqslant 1$, $\mathscr{Z}^{\mathrm{reg}}_{i,-1} = \left\{\left(\boldsymbol{z}^{\mathrm{reg}}_{i,2},\ldots,\boldsymbol{z}^{\mathrm{reg}}_{i,p^{\mathrm{reg}}_{\mathrm{ind}}}\right)\mid\boldsymbol{z}_i^{\mathrm{reg}}\in\mathscr{Z}_i^{\mathrm{reg}}\right\}$ and likewise $\mathscr{Z}^{\mathrm{reg}}_{\mathrm{pop},-1}$. Let $\bar{\mathscr{B}}^x_{i,1}$ be the closed ball defined by

$$\bar{\mathscr{B}}^x_{i,1} = \bar{\mathscr{B}}^x_{i,1}\Big((\boldsymbol{z}^{\mathrm{reg}}_{\mathrm{pop}},\boldsymbol{z}_i^{\mathrm{reg}})_{-1}\Big) = \bar{\mathscr{B}}\left(0,\,\frac{xM-b_{i,1}\big((\boldsymbol{z}^{\mathrm{reg}}_{\mathrm{pop}},\boldsymbol{z}_i^{\mathrm{reg}})_{-1}\big)}{a_{i,1}\big((\boldsymbol{z}^{\mathrm{reg}}_{\mathrm{pop}},\boldsymbol{z}_i^{\mathrm{reg}})_{-1}\big)}\right),$$

where $a_{i,1}\big((\boldsymbol{z}^{\mathrm{reg}}_{\mathrm{pop}},\boldsymbol{z}_i^{\mathrm{reg}})_{-1}\big)$ and $b_{i,1}\big((\boldsymbol{z}^{\mathrm{reg}}_{\mathrm{pop}},\boldsymbol{z}_i^{\mathrm{reg}})_{-1}\big)$ are defined as in (H 4). Thus, by slicing the integral in half and bounding the exponential on $\bar{\mathscr{B}}^x_{i,1}$ by 1,

$$\begin{aligned}q(\boldsymbol{y}_i|\theta_\kappa) \leqslant{}& \frac{1}{(\sigma_\kappa\sqrt{2\pi})^{k_i}}\int_{\bar{\mathscr{B}}^x_{i,1}\times\mathscr{Z}_{i,-1}} q(\boldsymbol{z}_{\mathrm{pop}},\boldsymbol{z}_i|\theta_\kappa)\,\mathrm{d}(\boldsymbol{z}_{\mathrm{pop}},\boldsymbol{z}_i)\\ &+\frac{1}{(\sigma_\kappa\sqrt{2\pi})^{k_i}}\int_{\bar{\mathscr{B}}^x_{i,1}{}^{\complement}\times\mathscr{Z}_{i,-1}}\exp\left(-\frac{1}{2\sigma_\kappa^2}\|\boldsymbol{y}_i-\gamma_i(\boldsymbol{z}_{\mathrm{pop}},\boldsymbol{z}_i)\|_2^2\right) q(\boldsymbol{z}_{\mathrm{pop}},\boldsymbol{z}_i|\theta_\kappa)\,\mathrm{d}(\boldsymbol{z}_{\mathrm{pop}},\boldsymbol{z}_i)\,,\end{aligned}$$

where $\mathscr{Z}_{i,-1} = \mathscr{Z}^{\mathrm{reg}}_{\mathrm{pop},-1}\times\mathscr{Z}^{\mathrm{crit}}_{\mathrm{pop}}\times\mathscr{Z}^{\mathrm{reg}}_{i,-1}\times\mathscr{Z}^{\mathrm{crit}}_i$. Moreover, by conditioning,

$$\begin{aligned}&\int_{\bar{\mathscr{B}}^x_{i,1}\times\mathscr{Z}_{i,-1}} q(\boldsymbol{z}_{\mathrm{pop}},\boldsymbol{z}_i|\theta_\kappa)\,\mathrm{d}(\boldsymbol{z}_{\mathrm{pop}},\boldsymbol{z}_i)\\ &\qquad= \int_{\bar{\mathscr{B}}^x_{i,1}} q(\boldsymbol{z}^{\mathrm{reg}}_{\mathrm{pop},1},\boldsymbol{z}^{\mathrm{reg}}_{i,1}|\theta_\kappa)\,\mathrm{d}(\boldsymbol{z}^{\mathrm{reg}}_{\mathrm{pop},1},\boldsymbol{z}^{\mathrm{reg}}_{i,1})\,.\end{aligned}$$

By continuity of $(\boldsymbol{z}^{\mathrm{reg}}_{\mathrm{pop},1},\boldsymbol{z}^{\mathrm{reg}}_{i,1})\mapsto q(\boldsymbol{z}^{\mathrm{reg}}_{\mathrm{pop},1},\boldsymbol{z}^{\mathrm{reg}}_{i,1}|\theta_\kappa)$ and compactness of $\bar{\mathscr{B}}^x_{i,1}$,

$$\begin{aligned}&\int_{\bar{\mathscr{B}}^x_{i,1}\times\mathscr{Z}_{i,-1}} q(\boldsymbol{z}_{\mathrm{pop}},\boldsymbol{z}_i|\theta_\kappa)\,\mathrm{d}(\boldsymbol{z}_{\mathrm{pop}},\boldsymbol{z}_i)\\ &\qquad\leqslant \sup_{\bar{\mathscr{B}}^x_{i,1}} q(\boldsymbol{z}^{\mathrm{reg}}_{\mathrm{pop},1},\boldsymbol{z}^{\mathrm{reg}}_{i,1}|\theta_\kappa)\,\mathscr{L}_1(\bar{\mathscr{B}}^x_{i,1})\,.\end{aligned}$$

Since the marginal of a multivariate distribution is a multivariate distribution whose mean vector and covariance matrix are obtained by dropping the irrelevant variables, $\lim_{\|\Sigma_\kappa\|\to\infty} q(\boldsymbol{z}^{\mathrm{reg}}_{\mathrm{pop},1},\boldsymbol{z}^{\mathrm{reg}}_{i,1}|\theta_\kappa) = 0$ and the first integral goes to zero as $\|\Sigma_\kappa\|\to\infty$.
In the same way of the proof of Theorem 1, the marginal density $q(\boldsymbol{z}^{\mathrm{reg}}_{\mathrm{pop},1},\boldsymbol{z}^{\mathrm{reg}}_{i,1}|\theta_\kappa)$ is controlled by the operator norm of the covariance matrix Σ_κ^{-1} from which we have drop the irrelevant variables. Hence, as $\|\Sigma_\kappa^{-1}\|\to\infty$, the first integral converges toward zero as well.
The second integral is maximized at $\sigma_\kappa^2 = \frac{1}{k_i}\|\boldsymbol{y}_i-\gamma_i(\boldsymbol{z}_{\mathrm{pop}},\boldsymbol{z}_i)\|^2$. Thus, due to the Cauchy-Schwarz inequality, there exists a constant $c\in\mathbb{R}_+^*$ such that for all $(\boldsymbol{z}_{\mathrm{pop}},\boldsymbol{z}_i)\in\bar{\mathscr{B}}^x_{i,1}{}^{\complement}\times\mathscr{Z}_{i,-1}$,

$$\begin{aligned}\|\boldsymbol{y}_i-\gamma_i(\boldsymbol{z}_{\mathrm{pop}},\boldsymbol{z}_i)\|_2^2 &\geqslant c\left(a_{i,1}\times\frac{xM-b_{i,1}}{a_{i,1}}+b_{i,1}-\|\boldsymbol{y}_i\|_\infty\right)^2\\ &\geqslant c\left((x-1)M\right)^2\end{aligned}$$

and by bounding the marginal density $q(\boldsymbol{z}_{\mathrm{pop}},\boldsymbol{z}_i|\theta_\kappa)$ on $\bar{\mathscr{B}}^x_{i,1}{}^{\complement}\times\mathscr{Z}_{i,-1}$ by 1, the second integral is bounded from above by

$$\left(\frac{k_i}{2\pi}\right)^{\frac{k_i}{2}}\mathrm{e}^{-\frac{k_i}{2}}\frac{1}{\left(\sqrt{c}\,(x-1)M\right)^{k_i}}\,.$$

Therefore,

$$\begin{aligned}\limsup_{\kappa\to\infty}\sum_{i=1}^{\ell}\log q(\boldsymbol{y}_i|\theta_\kappa) \leqslant{}& -\frac{k^\ell}{2}\Big[1+\log(2\pi)+\log\left(\sqrt{c}\,(x-1)M\right)\Big]\\ &+\frac{1}{2}\sum_{i=1}^{\ell}k_i\log k_i\,.\end{aligned}$$

Since x can be chosen arbitrarily large, we obtain the result for the case $\|\Sigma_\kappa\|\to+\infty$ as well as $\|\Sigma_\kappa^{-1}\|\to+\infty$.

(ii) Assume that, up to extraction of a subsequence, $\sigma_\kappa\to 0$ or $\sigma_\kappa\to\infty$.
Let $M = \|\boldsymbol{y}^\ell\|_\infty$. With the same notations as in the proof of Lemma 3, for all $\kappa\in\mathbb{N}$,

$$\sum_{i=1}^{\ell}\log q(\boldsymbol{y}_i|\theta_\kappa) \leqslant -\frac{k^\ell}{2}\log(2\pi\sigma_\kappa^2) - \frac{1}{2\sigma^2}d(\boldsymbol{y}^\ell,\Gamma^\ell)^2\,,$$

where $\Gamma^\ell = \mathscr{I}m(\gamma^\ell)$ and d denotes the Euclidean distance on $\mathbb{R}^{k^\ell}$. Let us prove that $d(\boldsymbol{y}^\ell,\Gamma^\ell)>0$ *a.s.* : the result will go along whatever $\sigma_\kappa\to 0$ or $\sigma_\kappa\to+\infty$ with the previous inequality. Let $\mathscr{Z}^\ell = \mathscr{Z}_{\mathrm{pop}}\times\prod_{i=1}^{\ell}\mathscr{Z}_i$.
Due to (H 4), for all $i\in[\![1,n]\!]$,

$$\lim_{\|(\boldsymbol{z}^{\mathrm{reg}}_{\mathrm{pop}},\boldsymbol{z}_i^{\mathrm{reg}})\|_\infty\to\infty}\|\gamma_i(\boldsymbol{z}_{\mathrm{pop}},\boldsymbol{z}_i)\|_\infty = +\infty,$$

and so for all $\varepsilon\in\mathbb{R}_+^*$ non-negative, there exists $R\in\mathbb{R}$ such as for all $\boldsymbol{z}^\ell\in\mathscr{Z}^\ell$ satisfying $\|\boldsymbol{z}^\ell\|>R$, $\|\gamma^\ell(\boldsymbol{z}^\ell)\|>M+\varepsilon$. In particular, by definition of M, $\|\boldsymbol{y}^\ell-\gamma^\ell(\boldsymbol{z}^\ell)\|_\infty>0$ for $\|(\boldsymbol{z}^{\mathrm{reg}}_{\mathrm{pop}},\boldsymbol{z}_i^{\mathrm{reg}})_{i\in[\![1,\ell]\!]}\|_\infty$ sufficiently large.

On the other hand, if at least a critical variable blows up, then by (H 5) there exists a critical trajectory γ_i^{crit} such that

$$\lim_{\|(z_{\text{pop}}^{\text{crit}}, z_i^{\text{crit}})\|_\infty \to \infty} \|\gamma_i(z_{\text{pop}}, z_i)\|_\infty = \gamma_i^{\text{crit}}$$

and as soon as this variable becomes sufficiently large, $y_i \neq \gamma_i^{\text{crit}}$ *a.s.* Thus $\|y^\ell - \gamma^\ell(z^\ell)\|_\infty > 0$ *a.s.* for $\|(z_{\text{pop}}^{\text{crit}}, z_i^{\text{crit}})_{i\in[\![1,\ell]\!]}\|_\infty$ sufficiently large.
In other words, there exists $R \in \mathbb{R}_+^*$ such that for all $z^\ell \in \mathscr{Z}^\ell$, if $\|z^\ell\|_\infty > R$, then $\|y^\ell - \gamma^\ell(z^\ell)\|_\infty > 0$ *a.s.* So, by contraposition, if there exists $z^\ell \in \mathscr{Z}^\ell$ such that $\|y^\ell - \gamma^\ell(z^\ell)\|_\infty = 0$ (at least *a.s.*) then $\|z^\ell\|_\infty \leqslant R$. Especially, $\{z^\ell \in \mathscr{Z}^\ell \mid y^\ell = \gamma^\ell(z^\ell)\ a.s.\} \subset \bar{\mathscr{B}}(0,R)$. Since (H 3) assumes that $P(\mathrm{d}\mathbf{y})$ has a continuous density and since $\gamma^\ell\left(\bar{\mathscr{B}}(0,R)\right)$ is a sub-manifold of dimension $p^\ell < k^\ell$, it comes that $P\left[z^\ell \in \bar{\mathscr{B}}(0,R)\right] = 0$. Hence,

$$\mathscr{L}_{k^\ell}\left(\left\{y^\ell \mid d(y^\ell, \mathscr{I}m(\gamma^\ell)) = 0\right\}\right) = 0$$

2. Let $f_\kappa(y^\ell) = \sum_{i=1}^\ell \log q(y_i|\theta_\kappa)$. From (C 1), we deduce that, up to extraction, the negative part $\left(f_\kappa(y^\ell)\right)^-$ is almost surely a non-decreasing and non-negative sequence converging to $+\infty$. From the monotone convergence theorem we then have

$$\liminf_{\kappa\to+\infty} \mathbb{E}_{P(\mathrm{d}y^\ell)}\left[\left(f_\kappa(y^\ell)\right)^-\right] = +\infty$$

and so

$$\lim_{\kappa\to+\infty} \mathbb{E}_{P(\mathrm{d}y^\ell)}\left[\left(f_\kappa(y^\ell)\right)^-\right] = +\infty.$$

Concerning the positive part $\left(f_\kappa(y^\ell)\right)^+$, using the dominated convergence theorem, Lemma 3 and the point (C 1), we get

$$\lim_{\kappa\to+\infty} \mathbb{E}_{P(\mathrm{d}y^\ell)}\left[\left(f_\kappa(y^\ell)\right)^+\right] = 0.$$

Actually, for all $i \in [\![1,n]\!]$ the application $(z_{\text{pop}}^{\text{reg}}, z_i^{\text{reg}}) \mapsto \gamma_i^{\text{crit}}$ is continuous by continuity of the function γ_i and so (H′ 5) holds.
Finally, we have proved that

$$\lim_{\kappa\to+\infty} \mathbb{E}_{P(\mathrm{d}y^\ell)}\left[\sum_{i=1}^\ell \log q(y_i|\theta_\kappa)\right] = -\infty$$

and (C 2) follows immediately.

3. The continuity statement is straightforward. If Θ_*^ω is empty, any maximizing sequence θ_κ of $\mathbb{E}_{P(\mathrm{d}y^\ell)}\left[\log q(y^\ell|\theta)\right]$ satisfies (up to extraction of a subsequence) $\theta_\kappa \in \Theta^\omega$, $\|\Sigma_\kappa\| \to +\infty$, $\|\Sigma_\kappa^{-1x}\| \to +\infty$, $\sigma_\kappa \to 0$ or $\sigma_\kappa \to +\infty$, which is on contradiction with conclusion (C 2).

□

A.2 Proof of the Consistency Theorem

We follow in the following proof the classical approach of van der Vaart [38].

Proof As in Lemma 4, let $\overline{\Theta^\omega}$ denote the one point Alexandrov compactification of the parameter space Θ^ω. We have already proved at [Lemma 4 (C 3)] that $\Theta_*^\omega \neq \varnothing$. To achieve the proof, let us first demonstrate that for all $\theta_\infty \in \overline{\Theta^\omega}$ such that $\delta(\theta_\infty, \Theta_*^\omega) \geqslant \varepsilon$ there exists an open set $\mathscr{U} \subset \Theta^\omega$ such that

$$\frac{1}{\ell}\mathbb{E}_{P(\mathrm{d}y^\ell)}\left[\sup_{\theta\in\mathscr{U}\cap\Theta^\omega} \sum_{i=1}^\ell \log q(y_i|\theta)\right] < \mathbb{E}^*(\omega). \tag{0}$$

Let $\varepsilon \geqslant 0$, $(\mathscr{U}_h) \subset \Theta^{\omega^{\mathbb{N}}}$ be a non-increasing sequence of open subsets of Θ^ω for which $\bigcap_{h\geqslant 0}\mathscr{U}_h = \{\theta_\infty\}$ and f_h be the function defined by

$$f_h(y^\ell) = \frac{1}{\ell}\sup_{\theta\in\mathscr{U}_h}\sum_{i=1}^\ell \log q(y_i|\theta).$$

1. If $\theta_\infty \in \Theta^\omega$, through the continuity of the map $\theta \mapsto \sum_{i=1}^\ell \log q(y_i|\theta)$ and the definition of the sequence $(\mathscr{U}_h)$,

$$\lim_{h\to+\infty} f_h(y^\ell) = \frac{1}{\ell}\sum_{i=1}^\ell \log q(y_i|\theta_\infty).$$

So, according to the monotone convergence theorem, Lemma 3 and since $\theta_\infty \notin \Theta_*^\omega$,

$$\lim_{h\to+\infty} \mathbb{E}_{P(\mathrm{d}y^\ell)}\left[f_h(y^\ell)\right] = \frac{1}{\ell}\sum_{i=1}^\ell \mathbb{E}_{P(\mathrm{d}y^\ell)}\left[\log q(y_i|\theta_\infty)\right] < \mathbb{E}^*(\omega).$$

2. If $\theta_\infty \notin \Theta^\omega$, *i.e.* if $\theta_\infty \in \overline{\Theta^\omega}\setminus\Theta^\omega$, we can prove that for all observations $y^\ell \in \mathbb{R}^{k^\ell}$ $\lim_{h\to\infty} f_h(y^\ell) = -\infty$ $P(\mathrm{d}y^\ell)$ *a.s.* We proceed by contradiction : assume that there exists a measurable set $A \in \mathscr{B}(\mathbb{R}^{k^\ell})$ such that $\mathbb{P}(y^\ell \in A) > 0$ and for all $y^\ell \in A$, $\inf_{h\in\mathbb{N}} f_h(y^\ell) > -\infty$. Then, by definition of the *infimum*, for all $y^\ell \in A$ there exists a sequence $(h_n) \in \mathbb{R}^{\mathbb{N}}$ such as $\liminf_{n\to+\infty} f_{h_n}(y) > -\infty$. However for all $y^\ell \in A$, $h \mapsto f_h(y^\ell)$ is non-increasing and reaches its *infimum* limit for $h = +\infty$ and thus $\lim_{n\to+\infty}\mathscr{U}_{h_n} = \mathscr{U}_\infty = \{\theta_\infty\}$. Finally, up to considering a sequence $(\theta_{n,n'}) \in \mathscr{U}_{h_n}^{\mathbb{N}}$ for all subsets $\mathscr{U}_{h_n} \subset \Theta^\omega$ such that for all $n \in \mathbb{N}$,

$$\lim_{n'\to+\infty}\sum_{i=1}^\ell \log q(y_i|\theta_{n,n'}) = \sup_{\theta\in\mathscr{U}_n}\sum_{i=1}^\ell \log q(y_i \mid \theta),$$

concatenating, reindexing those sequences and using the continuity of the map $\theta \mapsto \sum_{i=1}^\ell \log q(y_i|\theta)$ we know that there exists a sequence $(\theta_n) \in \Theta^{\omega^{\mathbb{N}}}$ such that

$$\lim_{n\to\infty}\theta_n = \theta_\infty \qquad \text{and} \qquad \liminf_{n\to+\infty}\sum_{i=1}^\ell \log q(y_i|\theta_n) > -\infty.$$

Moreover, $\theta_\infty = \left(\overline{z_{pop}}_\infty, \Sigma_\infty, \sigma_\infty\right) \in \overline{\Theta^\omega}\setminus\Theta^\omega$ and thus $\|\Sigma_\infty\| = +\infty$, $\|\Sigma_\infty^{-1}\| = +\infty$ or $\sigma_\infty \in \{0, +\infty\}$ in contradiction to [Lemma 4 (C 1)]. So for all observations y, $\lim_{h\to\infty} f_h(y^\ell) = -\infty$ $P(\mathrm{d}y)$ *a.s.* As in the proof of Lemma 4, Hypothesis (H 5) implies (H′ 5) and according to Lemma 3 and the monotone convergence theorem,

$$\lim_{h\to+\infty} \mathbb{E}_{P(\mathrm{d}y^\ell)}\left[f_h(y^\ell)\right] = -\infty < \mathbb{E}^*(\omega).$$

That is, in both cases $\lim_{h\to+\infty}\mathbb{E}_{P(\mathrm{d}y^\ell)}\left[f_h(y^\ell)\right] < \mathbb{E}^*(\omega)$ and there exists an open set $\mathscr{U} \subset \Theta^\omega$ such that

$$\frac{1}{\ell}\mathbb{E}_{P(\mathrm{d}y^\ell)}\left[\sup_{\theta\in\mathscr{U}\cap\Theta^\omega}\sum_{i=1}^\ell \log q(y_i|\theta)\right] < \mathbb{E}^*(\omega)$$

as announced.

Let $K_\varepsilon = \{\theta \in \overline{\Theta^\omega} \mid \delta(\theta, \Theta_*^\omega) \geqslant \varepsilon\}$. Through the compactness of K_ε, there exists an open finite cover $(\mathscr{U}_\alpha)_{\alpha\in[\![1,A]\!]}$ of K_ε satisfying (0). Thus, denoting $q_n = \lfloor \frac{n}{\ell} \rfloor$ and $r_n = n - q_n\ell$ the quotient and the rest of the euclidean division of n by ℓ, we get for all $\theta \in K_\varepsilon$,

$$\sup_{\theta\in K_\varepsilon\cap\Theta^\omega}\sum_{i=1}^n \log q(y_i|\theta) \leqslant \sup_{\alpha\in[\![1,A]\!]}\left(\sum_{q=0}^{q_n}\sup_{\theta\in\mathscr{U}_\alpha\cap\Theta^\omega}\sum_{r=1}^\ell \log q(y_{q\ell+r}|\theta) + \sup_{\theta\in\mathscr{U}_\alpha\cap\Theta^\omega}\sum_{r=\ell+1}^{r_n}\log q(y_{q_n\ell+r}|\theta)\right).$$

However, according to the strong law of large numbers, Assumption (H 2) and (0),

$$\lim_{q_n\to\infty}\frac{1}{q_n}\sum_{q=0}^{q_n}\sup_{\theta\in\mathscr{U}_\alpha\cap\Theta^\omega}\sum_{r=1}^{\ell}\log q(\boldsymbol{y}_{q\ell+r}|\theta)\leqslant\ell\,\mathbb{E}^*(\omega)$$

hence, since $\lim_{n\to+\infty}q_n=+\infty$ and $r_n<\ell$ for all $n\in\mathbb{N}$,

$$\limsup_{n\to\infty}\left[\frac{q_n}{n}\sup_{\alpha\in[\![1,A]\!]}\left(\frac{1}{q_n}\sum_{q=0}^{q_n}\sup_{\theta\in\mathscr{U}_\alpha\cap\Theta^\omega}\sum_{r=1}^{\ell}\log q(\boldsymbol{y}_{q\ell+r}|\theta)\right)\right]$$
$$=\frac{1}{\ell}\times\sup_{\alpha\in[\![1,A]\!]}\left(\mathbb{E}_{P(\mathrm{d}\boldsymbol{y}^\ell)}\left[\sup_{\theta\in\mathscr{U}_\alpha\cap\Theta^\omega}\sum_{r=1}^{\ell}\log q(\boldsymbol{y}_{q_n\ell+r}|\theta)\right]\right)<\mathbb{E}^*(\omega)\,.$$

Otherwise, for all $r\in[\![\ell+1,\ell_n]\!]$, $\log q(\boldsymbol{y}_{q_n\ell+r}|\theta)\leqslant -k^\ell\log q(\sigma\sqrt{2\pi})$ so

$$\frac{1}{n}\sup_{\alpha\in[\![1,A]\!]}\left(\sup_{\theta\in\mathscr{U}_\alpha\cap\Theta^\omega}\sum_{r=\ell+1}^{r_n}\log q(\boldsymbol{y}_{q_n\ell+r}|\theta)\right)\leqslant\frac{k^\ell(r_n-1)}{n}\log(\sigma\sqrt{2\pi})\,.$$

Thereafter

$$\limsup_{n\to\infty}\left[\frac{1}{n}\sup_{\alpha\in[\![1,A]\!]}\left(\sup_{\theta\in\mathscr{U}_\alpha\cap\Theta^\omega}\sum_{r=\ell+1}^{r_n}\log q(\boldsymbol{y}_{q_n\ell+r}|\theta)\right)\right]\leqslant 0$$

and

$$\limsup_{n\to\infty}\frac{1}{n}\sup_{\theta\in K_\varepsilon\cap\Theta^\omega}\sum_{i=1}^{n}\log q(\boldsymbol{y}_i|\theta)<\mathbb{E}^*(\omega)\,.\tag{1}$$

By definition of Θ_*^ω and according to the strong law of large numbers and (H 2), for all $\theta^*\in\Theta_*^\omega$ $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\log q(\boldsymbol{y}_i|\theta^*)=\mathbb{E}^*(\omega)$ *a.s.* Moreover for all $i\in[\![1,n]\!]$,

$$q(\boldsymbol{y}_i|\hat{\theta}_n)=\frac{q(\hat{\theta}_n\mid\boldsymbol{y}_i)\,q(\boldsymbol{y}_i)}{q_{prior}(\hat{\theta}_n)}\geqslant\frac{q(\theta_*\mid\boldsymbol{y}_i)\,q(\boldsymbol{y}_i)}{q_{prior}(\hat{\theta}_n)}=\frac{q(\boldsymbol{y}_i|\theta_*)\,q_{prior}(\theta_*)}{q_{prior}(\hat{\theta}_n)}$$

and so

$$\sum_{i=1}^{n}\log q(\boldsymbol{y}_i|\hat{\theta}_n)\geqslant\sum_{i=1}^{n}\log q(\boldsymbol{y}_i|\theta_*)+\left(\log q_{prior}(\theta_*)-\log q_{prior}(\hat{\theta}_n)\right).$$

Since q_{prior} is upper-bounded on Θ^ω, there exists $M\in\mathbb{R}^+$ such that

$$\frac{1}{n}\left(\log q_{prior}(\theta_*)-\log q_{prior}(\hat{\theta}_n)\right)\geqslant\frac{1}{n}\log\left(\frac{q_{prior}(\theta_*)}{M}\right)$$

i.e. $\liminf_{n\to+\infty}\frac{1}{n}\left(\log q_{prior}(\theta_*)-\log q_{prior}(\hat{\theta}_n)\right)\geqslant 0$ and

$$\liminf_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}\log q(\boldsymbol{y}_i|\hat{\theta}_n)\geqslant\mathbb{E}^*(\omega)\,.\tag{2}$$

The result follows from Equations 1 and 2 by contradiction : Assume that for all $n\in\mathbb{N}$, $\hat{\theta}_n\in K_\varepsilon$ *i.e.* that $\delta(\hat{\theta}_n,\Theta_*^\omega)\geqslant\varepsilon$. Then

$$\sum_{i=1}^{n}\log q(\boldsymbol{y}_i|\hat{\theta}_n)\leqslant\sup_{\theta\in K_\varepsilon\cap\Theta^\omega}\sum_{i=1}^{n}\log q(\boldsymbol{y}_i|\theta)$$

and by taking the limit superior, we get

$$\mathbb{E}^*(\omega)\overset{(2)}{\leqslant}\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\log q(\boldsymbol{y}_i|\hat{\theta}_n)\overset{(1)}{<}\mathbb{E}^*(\omega)$$

i.e. $\mathbb{E}^*(\omega)<\mathbb{E}^*(\omega)$. Hence $\lim_{n\to\infty}\mathbb{P}\left[\delta(\hat{\theta}_n,\Theta_*^\omega)\geqslant\varepsilon\right]=0$. □

Acknowledgements Ce travail bénéficie d'un financement public Investissement d'avenir, réference ANR-11 LABX-0056-LMH. This work was supported by a public grant as part of the Investissement d'avenir, project reference ANR-11 LABX-0056-LMH.

Travail réalisé dans le cadre d'un projet financé par la Fondation de la Recherche Médicale DBI20131228564. Work performed as a part of a project funded by the Fondation of Medical Research, grant number DBI20131228564.

References

1. S. Allassonnière, Y. Amit, and A. Trouvé. Toward a coherent statistical framework for dense deformable template estimation. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 69(1):3–29, 2007.
2. 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.
3. 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.
4. M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, 61(2):139–157, 2005.
5. A. Bône, O. Colliot, and S. Durrleman. Learning distributions of shape trajectories from longitudinal datasets: A hierarchical model on a manifold of diffeomorphisms. In *Computer Vision and Pattern Recognition*, Salt Lake City, United States, 2018.
6. B. Charlier, N. Charon, and A. Trouvé. The fshape framework for the variability analysis of functional shapes. *Foundations of Computational Mathematics*, 17(2):287–357, 2017.
7. N. Charon and A. Trouvé. The varifold representation of nonoriented shapes for diffeomorphic registration. *SIAM Journal on Imaging Sciences*, 6(4):2547–2580, 2013.
8. J. Chevallier, S. Oudard, and S. Allassonnière. Learning spatiotemporal piecewise-geodesic trajectories from longitudinal manifold-valued data. In *Neural Information Processing Systems*, number 30, Long Beach, CA, USA, 2017.
9. T. Colin, A. Iollo, D. Lombardi, and O. Saut. Prediction of the evolution of thyroidal lung nodules using a mathematical model. *ERCIM News*, 82:37–38, 2010.
10. B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics*, 27(1):94–128, 1999.
11. A. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 39(1):1–38, 1977.
12. B. Escudier, C. Porta, M. Schmidinger, N. Rioux-Leclercq, A. Bex, V. S. Khoo, V. Gruenvald, and A. Horwich. Renal cell carcinoma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, 27(suppl 5):v58–v68, 2016.
13. S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian Geometry*. Universitext. Springer-Verlag Berlin Heidelberg, 3 edition, 2004.
14. C. Giraud. *Introduction to High-Dimensional Statistics*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 2014.
15. A. Hyvärinen, J. Karhunen, and E. Oja. *Independent component analysis*, volume 46. John Wiley & Sons, 2004.
16. E. Klassen, A. Srivastava, W. Mio, and S. Joshi. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE transactions on pattern analysis and machine intelligence*, 26(3):372–383, 2004.
17. E. Konukoglu, O. Clatz, B. H. Menze, B. Stieltjes, M.-A. Weber, E. Mandonnet, H. Delingette, and N. Ayache. Image guided personalization of reaction-diffusion type tumor growth models using modified anisotropic eikonal equations. *IEEE transactions on medical imaging*, 29(1):77–95, 2010.
18. I. Koval, J.-B. Schiratti, A. Routier, M. Bacci, O. Colliot, S. Allassonniere, and S. Durrleman. Spatiotemporal propagation of the cortical atrophy: Population and individual patterns. *Frontiers in Neurology*, 9, 2018.

19. E. Kuhn and M. Lavielle. Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, 49(4):1020–1038, 2005.
20. N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974, 1982.
21. M. Lavielle. *Mixed Effects Models for the Population Approach: Models, Tasks, Methods and Tools*. CRC Biostatistics Series. Chapman and Hall, 2014.
22. M. Lavielle and L. Aarons. What do we mean by identifiability in mixed effects models? *Journal of Pharmacokinetics and Pharmacodynamics*, 43(1):111–122, February 2016.
23. M. I. Miller, A. Trouvé, and L. Younes. Geodesic shooting for computational anatomy. *Journal of mathematical imaging and vision*, 24(2):209–228, 2006.
24. J. K. Milliken and S. D. Edland. Mixed effect models of longitudinal alzheimer's disease data: A cautionary note. *Statistics in Medicine*, 19(11-12):1617–1629, 2000.
25. I. Rekik, S. Allassonnière, O. Clatz, E. Geremia, E. Stretton, H. Delingette, and N. Ayache. Tumor growth parameters estimation and source localization from a unique time point: Application to low-grade gliomas. *Computer Vision and Image Understanding*, 117(3):238–249, 2013.
26. B. Ribba, N. Holford, P. Magni, I. Trocniz, I. Gueorguieva, P. Girard, C. Sarr, M. Elishmereni, C. Kloft, and L. Friberg. A review of mixed-effects models of tumor growth and effects of anticancer drug treatment used in population analysis. *CPT: Pharmacometrics & Systems Pharmacology*, 3(5):1–10, 2014.
27. C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer-Verlag New York, 1999.
28. G. O. Roberts and J. S. Rosenthal. Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *Journal of Applied Probability*, 44(2):458–475, 03 2007.
29. G. O. Roberts and J. S. Rosenthal. Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367, 2009.
30. O. Saut, J.-B. Lagaert, T. Colin, and H. M. Fathallah-Shaykh. A multilayer grow-or-go model for gbm: effects of invasive cells and anti-angiogenesis on growth. *Bulletin of mathematical biology*, 76 (9):2306–2333, 2014.
31. J.-B. Schiratti, S. Allassonnière, O. Colliot, and S. Durrleman. Learning spatiotemporal trajectories from manifold-valued longitudinal data. In *Neural Information Processing Systems*, number 28 in Advances in Neural Information Processing Systems, Montréal, Canada, 2015.
32. 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. *Journal of Machine Learning Research*, 18(133):1–33, 2017.
33. J. Su, S. Kurtek, E. Klassen, and A. Srivastava. Statistical analysis of trajectories on riemannian manifolds: bird migration, hurricane tracking and video surveillance. *The Annals of Applied Statistics*, 8(1):530–552, 2014.
34. J. Su, A. Srivastava, F. D. de Souza, and S. Sarkar. Rate-invariant analysis of trajectories on riemannian manifolds with application in visual speech recognition. In *Computer Vision and Pattern Recognition*, pages 620–627, 2014.
35. P. Therasse, S. G. Arbuck, E. A. Eisenhauer, J. Wanders, R. S. Kaplan, L. Rubinstein, J. Verweij, M. Van Glabbeke, A. T. van Oosterom, M. C. Christian, and S. G. Gwyther. New guidelines to evaluate the response to treatment in solid tumors. *Journal of the National Cancer Institute*, 92(3):205–216, 2000.
36. A. Trouvé and L. Younes. *Shape Spaces*, pages 1759–1817. Springer New York, New York, USA, 2015.
37. M. Vaillant and J. Glaunès. Surface matching via currents. In *Information Processing in Medical Imaging*, volume 3565, Glenwood Springs, USA, 2005.
38. A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2000.