

Time-warped Geodesic Regression

Yi Hong¹, Nikhil Singh¹, Roland Kwitt³, and Marc Niethammer^{1,2}

¹University of North Carolina (UNC) at Chapel Hill, NC, US

²Biomedical Research Imaging Center, UNC-Chapel Hill, NC, US

³Department of Computer Science, University of Salzburg, Austria

Abstract. We consider geodesic regression with parametric time-warps. This allows, for example, to capture saturation effects as typically observed during brain development or degeneration. While highly-flexible models to analyze time-varying image and shape data based on generalizations of splines and polynomials have been proposed recently, they come at the cost of substantially more complex inference. Our focus in this paper is therefore to keep the model and its inference as simple as possible while allowing to capture expected biological variation. We demonstrate that by augmenting geodesic regression with parametric time-warp functions, we can achieve comparable flexibility to more complex models while retaining model simplicity. In addition, the time-warp parameters provide useful information of underlying anatomical changes as demonstrated for the analysis of corpora callosa and rat calvariae. We exemplify our strategy for shape regression on the Grassmann manifold, but note that the method is generally applicable for time-warped geodesic regression.

1 Introduction

In recent years, an increased amount of research has been devoted to the development of geodesic regression approaches for shapes, images, and general Riemannian manifolds [15,10,16]. These models are attractive because they allow for a compact representation of, e.g., time-series data, while – just as for linear regression – remaining relatively simple. However, these models are only sensible if changes are well described by a geodesic. This is generally not the case when dealing with periodic motions or processes which exhibit saturations, such as brain maturation or degeneration. To model such behaviors, extensions to spline- and polynomial-models for shapes have been developed [17,11]. These approaches are conceptually appealing, due to greater modeling flexibility; however, inference becomes significantly more difficult. Our motivation for time-warped geodesic regression is therefore to obtain a simple model, which allows for easy inference, while remaining flexible enough to capture plausible biological changes.

Contribution. We propose a time-warped geodesic regression approach which increases flexibility of geodesic regression at a moderate increase in model complexity. This is realized by using parametric models for time-warping. This strategy is suitable to model saturations. We note that general time-warps have been

used previously, e.g., in the context of spatio-temporal atlas building [7]. However, our focus here is on simple, parametric time-warping models in the context of geodesic regression. While our approach is general, we demonstrate its application for regression on the Grassmann manifold. This allows regression of dynamical systems, shapes, or any entity with a subspace representation.

The paper is structured as follows: §2 reviews the general theory of geodesic regression and introduces time-warping in the context of geodesic regression. §3 describes time-warped geodesic regression on the Grassmann manifold. Experimental results for human corpora callosa and rat skulls are shown in §4 and compared to results achieved on the same datasets using geodesic regression [10] and its polynomial extension [11] for shapes represented in Kendall shape space.

2 Time-warped Geodesic Regression (TW-GR)

Geodesic regression. Geodesic regression generalizes Euclidean least-squares regression to manifolds. Similar to least-squares regression, the parameter estimates of geodesic regression include the initial conditions of the geodesic that best fits the measurements in a least-squares sense (with respect to the underlying metric). Given N measurements X_i at timepoints $\bar{t}_i, i = 1, \dots, N$, represented as points on the manifold \mathcal{M} , the geodesic $X(\bar{t})$ that passes closest to the data in the least-squares sense minimizes the energy functional

$$E(X_{\bar{t}_0}, \dot{X}_{\bar{t}_0}) = \sum_{i=1}^N d^2(X(\bar{t}_i), X_i) . \quad (1)$$

Here, $d^2(\cdot, \cdot)$ denotes the squared geodesic distance and \bar{t}_0 denotes the reference starting point for the geodesic. $X(\bar{t}_i)$ denotes the point on the geodesic at time \bar{t}_i which is obtained by shooting the geodesic to $\bar{t} = \bar{t}_i$, as per the Riemannian exponential map on the manifold, starting from the initial conditions at $\bar{t} = \bar{t}_0$. As in the Euclidean case, the parameter estimates, $X_{\bar{t}_0}$ and $\dot{X}_{\bar{t}_0}$ can be interpreted as the initial *intercept* and *slope* that parameterize the geodesic.

Time-warped geodesic regression. The idea of time-warped geodesic regression is to use a simple model to warp the time-points when comparison to data is performed. These time-warps should be diffeomorphic so that the time-ordering of data-points does not change. Note that this is conceptually similar to an error-in-variables model where uncertainties in the independent variables are modeled. However, here we are not directly concerned with modeling such uncertainties, but instead in obtaining a somewhat *richer* model for geodesic regression which is easy to compute given algorithms to solve the geodesic regression problem.

To account for saturations of scalar-valued outputs, logistic regression is frequently used. In our case the dependent variables are complex: images, shapes, or as in our example problem elements of the Grassmann manifold. Hence, it is easier to warp the time-axis. For saturations, points on the geodesic which are close should then be mapped to points far apart in time that they are compared to, i.e., small changes are mapped to large intervals in time (cf. Fig. 1). In

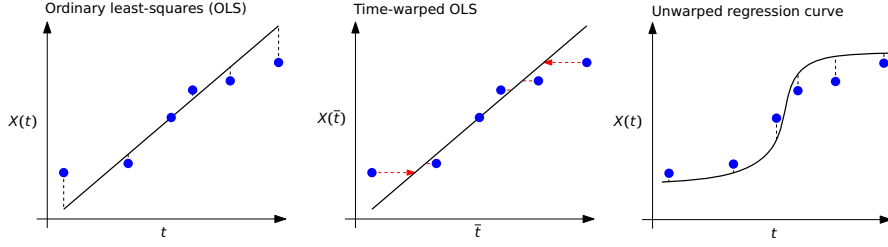


Fig. 1: Illustration of regression lines before and after time-warp using least-squares. *Left:* Ordinary least-squares (OLS) in the original time coordinates t . *Middle:* OLS in the warped time coordinates \bar{t} . The red lines indicate at what time points of the regression line the measurements are evaluated after the time-warp. *Right:* The unwarped regression curve in the original time coordinates t obtained by using time-warped OLS.

principle the mapping could be described by a general diffeomorphism. In fact, such an approach is followed in [7] for general spatio-temporal atlas-building for shapes. Our motivation for proposing an approach to geodesic regression with *parametric* time-warps is to keep the model simple while, at the same time, gaining more flexibility. Extensions to general non-parametric approaches can easily be obtained. Given that $\bar{t} = f(t; \boldsymbol{\theta})$ denotes the time-warp, parametrized by $\boldsymbol{\theta}$, and $X(\bar{t})$ denotes the regression geodesic in the warped time coordinates \bar{t} , the general time-warped geodesic regression problem seeks to minimize

$$E(X_{\bar{t}_0}, \dot{X}_{\bar{t}_0}, \boldsymbol{\theta}) = \frac{1}{\sigma^2} \sum_{i=1}^N d^2(X(f(t_i; \boldsymbol{\theta})), X_i) , \quad (2)$$

such that $X(\bar{t})$ describes a geodesic with $X(\bar{t}_0) = X_{\bar{t}_0}$ and $\dot{X}(\bar{t}_0) = \dot{X}_{\bar{t}_0}$. Eq. (2) may include regularization terms with respect to $\dot{X}_{\bar{t}_0}$ as appropriate.

The easiest way to minimize this energy is by alternating optimization. This requires the derivative of the energy with respect to $\boldsymbol{\theta}$ for fixed $X(\bar{t})$, given by

$$\nabla_{\boldsymbol{\theta}} E = -\frac{2}{\sigma^2} \sum_{i=1}^N \langle \text{Log}_{X(f(t_i; \boldsymbol{\theta}))} X_i, \dot{X}(f(t_i; \boldsymbol{\theta})) \rangle \nabla_{\boldsymbol{\theta}} f(t_i; \boldsymbol{\theta}) , \quad (3)$$

where $\text{Log}_{X(f(t_i; \boldsymbol{\theta}))} X_i$ denotes the log-map, i.e., the initial velocity of the geodesic connecting $X(f(t_i; \boldsymbol{\theta}))$ and X_i in unit time and $\dot{X}(f(t_i; \boldsymbol{\theta}))$ is the velocity of the regression geodesic at the warped-time point. The inner-product depends on the metric being used. This leaves to choose a good parametric model for $f(t; \boldsymbol{\theta})$. As we require the transform to be diffeomorphic we choose a parametric model which is diffeomorphic by construction. A possible choice is the generalized logistic function [9], e.g., with asymptotes 0 for $t \rightarrow -\infty$ and 1 for $t \rightarrow \infty$, i.e.,

$$\bar{t} = f(t, \boldsymbol{\theta}) = \frac{1}{(1 + \beta e^{-k(t-M)})^{1/m}} , \quad (4)$$

where $\boldsymbol{\theta} = (k, M, \beta, m)$, k controls the growth rate, M is the time of maximum growth if $\beta = m$, β and m define the value of f at $t = M$, and $m > 0$ affects the

asymptote of maximum growth. Using this function $f(t; \boldsymbol{\theta})$, we map the original infinite time interval to a warped time-range from 0 to 1. Initial parameters of this function need to be chosen such that all data evenly covers the warped region from 0 to 1. The algorithm using alternating optimization is as follows:

- 0) Initialize $\boldsymbol{\theta}$ such that the warped time is evenly distributed within $(0, 1)$.
- 1) Compute new time-points $\{f(t_i; \boldsymbol{\theta})\}$ to be used in $\bar{t} \in (0, 1)$ for fitting.
- 2) Compute standard geodesic regression solution with time-points from 1).
- 3) Update $\boldsymbol{\theta}$ by numerical optimization using the gradient given in Eq. (3).
- 4) Check convergence. If not converged goto 1).

3 TW-GR on the Grassmann manifold

While time-warped geodesic regression is general for Riemannian manifolds, we describe the specialization to the Grassmann manifold $\mathcal{G}(p, n)$, i.e., the manifold of p -dimensional linear subspaces of \mathbb{R}^n , as an example. Details about the Riemannian structure of the Grassmannian can be found in [5,8] for instance.

In our case, a point $\mathcal{Y} \in \mathcal{G}(p, n)$ is represented by an orthonormal matrix $\mathbf{Y} \in \mathbb{R}^{n \times p}$ where the span of the column vectors spans \mathcal{Y} . Hence, $\mathbf{Y}^\top \mathbf{Y} = \mathbf{I}_p$, where \mathbf{I}_p is the $p \times p$ identity matrix. Under the canonical metric on the Grassmannian [8], the arc-length of the geodesic connecting two subspaces $\mathcal{Y}, \mathcal{Z} \in \mathcal{G}(p, n)$ is related to the *canonical angles* $\phi_1, \dots, \phi_p \in [0, \pi/2]$ between \mathcal{Y} and \mathcal{Z} as

$$d^2(\mathcal{Y}, \mathcal{Z}) = \|\boldsymbol{\phi}\|_2^2 . \quad (5)$$

In what follows, we slightly abuse notation and use $d^2(\mathbf{Y}, \mathbf{Z})$ as a surrogate for Eq. (5) with $\mathcal{Z} = \text{span}(\mathbf{Z}), \mathcal{Y} = \text{span}(\mathbf{Y})$.

Regression. Our objective is to fit a geodesic, represented by an initial point \mathbf{Y}_{r_0} and an initial velocity $\dot{\mathbf{Y}}_{r_0}$, to a collection of points $\{\mathbf{Y}_i\}_{i=1}^N$ at N measurement instances $\{r_i\}_{i=1}^N$. The geodesic regression problem minimizes

$$E(\mathbf{Y}_{r_0}, \dot{\mathbf{Y}}_{r_0}) = \frac{1}{\sigma^2} \sum_{i=1}^N d^2(\mathbf{Y}(r_i), \mathbf{Y}_i), \quad \sigma > 0 \quad (6)$$

subject to the initial condition constraints $\mathbf{Y}(r_0)^\top \mathbf{Y}(r_0) = \mathbf{I}_p$, $\mathbf{Y}(r_0)^\top \dot{\mathbf{Y}}(r_0) = 0$, and the geodesic equation for a curve on $\mathcal{G}(p, n)$: $\dot{\mathbf{Y}}(r) + \mathbf{Y}(r)[\dot{\mathbf{Y}}(r)^\top \dot{\mathbf{Y}}(r)] = 0$. This geodesic equation also defines the Riemannian exponential map (exp-map) on the Grassmannian as an ODE for convenient numerical computations. In other words, integrating the geodesic equation, starting with initial conditions, evolves (or shoots) the geodesic forward in time. Computational steps for Eq. (5) or the Riemannian log-map (used in Eq. (3)), can be found in [3] for instance. Also note that for time-warped regression on the Grassmann manifold, the time instances $\{r_i\}_{i=1}^N$ are the warped time-points using the logistic function.

In our implementation, we use a shooting strategy to solve this regression problem (see the supplementary material [1]). A theoretical characterization of geodesic curve fitting on the Grassmannian is discussed in [2], but no numerical strategy is provided. Alternatively, a numerical solution based on Jacobi fields [10] can also be used for geodesic fitting.

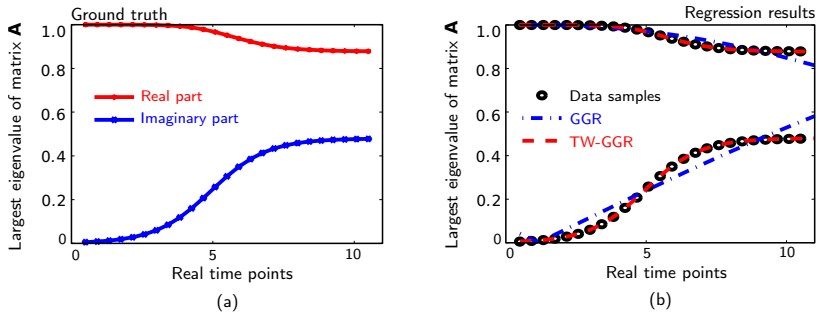


Fig. 2: Comparing GGR and TW-GGR on synthetic data (best-viewed in color).

4 Experimental Results

To demonstrate our method, we use both synthetic and real data to compare time-warped Grassmannian geodesic regression (TW-GGR) with the non-time-warped variant (GGR). Three measurements are used to quantitatively compare the regression results: 1) the regression energy, i.e., the data matching error for all observations; 2) the R^2 statistic on the Grassmannian, which is between 0 and 1, with 1 indicating a perfect fit and 0 indicating a fit no better than the Fréchet mean (see [10,11] for more details); and 3) the mean squared error (MSE) of testing data, reported in a (leave-one-out) crossvalidation experiment. In all experiments, σ in the cost function is set to 1. For the parameters in θ , we fix $\beta, m = 1$ so that M is the time of the maximal growth, and initialize k, M to make sure the warped time points are evenly distributed over the range of $(0, 1)$.

Synthetic data. We generate synthetic data using 25 time points that are uniformly distributed within $(0, 11)$. Each data point corresponds to a sine/cosine signal where the frequency is 10 times the warped time point using a logistic function with $\theta = (1, 1, 1, 5)$. These sine/cosine signals are sampled at 630 points in $[0, 10\pi]$ and we embed them in \mathbb{R}^{24} . For each signal $\mathbf{y}_t \in \mathbb{R}^{630 \times 24}$, we then estimate a (two-state) linear dynamical system (LDS) $\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{v}_t$, $\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{w}_t$, where \mathbf{A} is the state-transition matrix, \mathbf{C} controls the generation of the output \mathbf{y}_t and \mathbf{x}_t is the state at time t . $\mathbf{w}_t, \mathbf{v}_t$ are the normally distributed state- and observation noise, respectively. Using the estimation approach of [6] this system has, by design, a full-rank observability matrix $\mathbf{O} = [\mathbf{C} \ \mathbf{C}\mathbf{A}]^\top$. By using SVD, $\mathbf{O} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$, $\mathbf{U} \in \mathbb{R}^{1260 \times 2}$ acts as a representer for the observability subspace of the system. Further, for this synthetic data, the largest eigenvalue of the state-transition matrix \mathbf{A} reflects the frequency of the sine/cosine signal and serves as our ground truth, cf. Fig. 2(a).

Fig. 2(b) shows the regression results for both GGR and TW-GGR. Note that for each point we have extracted \mathbf{A} from the regression result (which is a representer for the subspace and not the state-transition matrix itself; see supplementary material). The regression energy is reduced from $2.9\text{e-}02$ for GGR to $5.2\text{e-}05$ for TW-GGR, and R^2 is improved from 0.94 for GGR to 1.0 for TW-GGR, indicating a perfect fit for the synthetic data. The mean-squared error

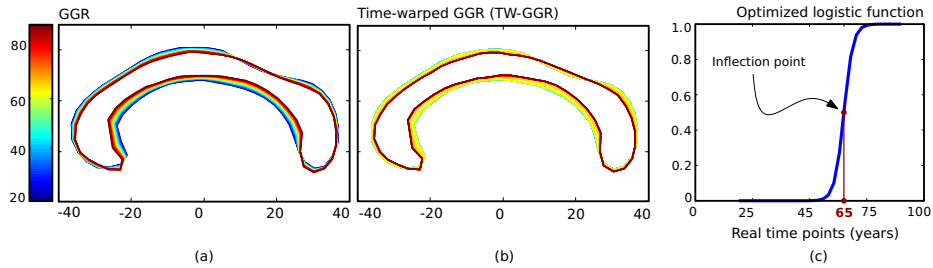


Fig. 3: Comparison between GGR and TW-GGR on corpus callosum data. Shapes are colored by age. (a) and (b) show the shape generated along the geodesic; (c) shows the corresponding optimized logistic function (best-viewed in color).

(MSE) for left-out samples in the crossvalidation experiment reduces from $1.4e-03$ for GGR to $1.5e-06$ for TW-GGR (cf. Table 1).

Corpus callosum aging. To explore the benefits of time-warped regression on real data, we demonstrate our method on a collection of 32 corpus callosum shapes with varying ages from 19 to 90 years [10]. Each shape is represented by 64 2D boundary landmarks, and is projected to a point on the Grassmannian using the left-singular vectors obtained from a SVD decomposition of the 64×2 coordinate matrix. This results in an affine-invariant representation, cf. [3].

Fig. 3 shows the corpus callosum shapes¹ along the geodesic computed with standard GGR and TW-GGR. For GGR, the corpus callosum starts to shrink from 19 years old, which is consistent with the regression results in [10,11]. However, according to biological studies [12,14], the corpus callosum size remains stable during the most active years of the lifespan, which is consistent with the TW-GGR result. As we can see from the shapes and optimized logistic function shown in Figs. 3(b)–(c), TW-GGR estimates that thinning starts at the age of > 50 years, and at the age of 65, the shrinking rate reaches its peak.

Table 1 lists the reduction in energy from 0.3494 for GGR to 0.3379 for TW-GGR. The R^2 statistic improves from 0.12 for GGR to 0.15 for TW-GGR and the crossvalidation MSE reduces from $1.25e-02$ for GGR to $1.22e-02$ for TW-GGR.

Rat calvarium growth. We further demonstrate the applicability of our method on shape analysis of the Vilmann rat data [4]². We use 18 individuals with 8 time points, each in the age range of 7 to 150 days. Each shape is represented by a set of 8 landmarks. Fig. 4(a) shows the landmarks projected onto the Grassmannian by using their SVD-based affine-invariant representation, similar to the previous experiment. From the GGR results in Fig. 4(b), we see that the rat calvarium grows at an approximately constant speed during the first 150 days. However, the TW-GGR results in Figs. 4(c)–(d) indicate that the rat calvarium only grows fast in the first few weeks, reaching its peak at 30 days; then, the rate of growth gradually levels off and becomes steady after 14 weeks. In fact, similar growth

¹ The shown shapes are recovered from the points along the geodesic on the Grassmannian manifold through scaling by the mean singular values of the SVD results.

² available at <http://life.bio.sunysb.edu/morph/data/datasets.html>

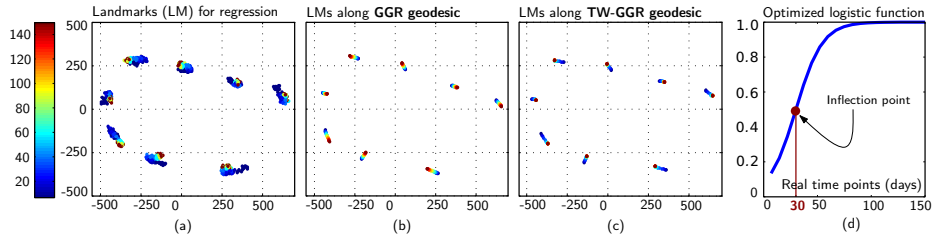


Fig. 4: Comparison of rat calvarium [4] growth, modeled by GGR and TW-GGR. Landmarks are colored by age in days; (d) shows the corresponding optimized logistic function (best-viewed in color).

	Synthetic data		Corpus callosum [10]		Rat calvarium [4]	
	GGR	TW-GGR	GGR	TW-GGR	GGR	TW-GGR
Energy	2.9e-02	5.2e-05	0.3494	0.3379	0.3238	0.1754
R^2	0.94	1.0	0.12	0.15	0.61	0.79
MSE	1.4e-03	1.5e-06	1.25e-02	1.22e-02	2.3e-03	1.3e-03

Table 1: Comparison of GGR and TW-GGR on synthetic and real data.

curves for the rat skull were reported in [13]. Based on their study, the growth velocities of viscerocranium length and neurocranium width rose to the peak in the 26-32 days period.

Table 1 lists the reduction in regression energy from 0.3238 for GGR to 0.1754 for TW-GGR. R^2 increases from 0.61 for GGR to 0.79 for TW-GGR, and MSE reduces from 2.3e-03 for GGR to 1.3e-03 for TW-GGR.

5 Discussion and Conclusions

Table 1 indicates that TW-GGR consistently performs better than GGR on both synthetic and real data. This can be attributed to the increased flexibility due to time-warping. More importantly our approach can model growth and saturation processes which are common phenomena in biological data. On the examples of corpora callosa and rat calvariae shapes, we demonstrated that TW-GGR allows us to obtain models that are compatible with results reported in biological studies. Contrary to that, regression methods such as [10] or [11] have no parameters that could naturally capture these effects. As for the R^2 statistics for corpora callosa and rat calvariae, our results are lower than those reported in [11]. This can be explained by the difference in shape representation (Kendall vs. Grassmannian) which renders the R^2 values not directly comparable. The Grassmann approach is useful when affine-invariance for shape representations is desired. If not, our approach could directly be applied for time-warped geodesic regression on Kendall shape space. We remark that geodesic regression on the Grassmannian has applications beyond regression on shapes, as demonstrated

on the dynamical systems example. Our regression approach focuses on warping the time-axis, *not* the data-axis, i.e., it is independent of the underlying type of manifold. This enables easy adoption of the concept to other Riemannian manifolds and applications which will be explored in future work.

Acknowledgements This work was supported by NSF EECS-1148870, NSF EECS-0925875, and NIH 2P41EB002025.

References

1. Supplementary material, https://bitbucket.org/yi_hong/tw-gr.git 4
2. Batzies, E., Machado, L., Silva Leite, F.: The geometric mean and the geodesic fitting problem on the Grassmann manifold, unpublished manuscript (available at <http://www.mat.uc.pt/preprints/ps/p1322.pdf>) 4
3. Begelfor, E., Werman, W.: Affine invariance revisited. In: CVPR. pp. 2087–2094. IEEE Press (2006) 4, 6
4. Bookstein, F.: Morphometric tools for landmark data: geometry and biology. Cambridge Univ. Press (1991) 6, 7
5. Boothby, W.: An Introduction to Differentiable Manifolds and Riemannian Geometry. Academic Press (1986) 4
6. Doretto, G., Chiuso, A., Wu, Y., Soatto, S.: Dynamic textures. *Int. J. Comput. Vision* 51(2), 91–109 (2003) 5
7. Durrleman, S., Pennec, X., Trouvé, A., Braga, J., Gerig, G., Ayache, N.: Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data. *Int. J. Comput. Vision* 103(1), 22–59 (2013) 2, 3
8. Edelman, A., Arias, T., Smith, S.T.: The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.* 20(2), 303–353 (1998) 4
9. Fekedulegn, D., Mac Siurtain, M., Colbert, J.: Parameter estimation of nonlinear growth models in forestry. *Silva Fennica* 33(4), 327–336 (1999) 3
10. Fletcher, P.: Geodesic regression and the theory of least squares on Riemannian manifolds. *Int. J. Comput. Vision* 105(2), 171–185 (2013) 1, 2, 4, 5, 6, 7
11. Hinkle, J., Fletcher, P.T., Joshi, S.: Intrinsic polynomials for regression on Riemannian manifolds. *J. Math. Imaging Vis.* pp. 1–21 (2014) 1, 2, 5, 6, 7
12. Hopper, K., Patel, S., Cann, T., Wilcox, T., Schaeffer, J.: The relationship of age, gender, handedness and sidedness to the size of the corpus callosum. *Acad. Radiol.* 1, 243–248 (1994) 6
13. Hughes, P., Tanner, J., Williams, J.: A longitudinal radiographic study of the growth of the rat skull. *J. Anat.* 127(1), 83–91 (1978) 7
14. Johnson, S., Farnworth, T., Pinkston, J., Bigler, E., Blatter, D.: Corpus callosum surface area across the human adult life span: Effect of age and gender. *Brain Res. Bull.* 35(4), 373–377 (1994) 6
15. Niethammer, M., Huang, Y., Vialard, F.X.: Geodesic regression for image time-series. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS. vol. 6892, pp. 655–662. Springer, Heidelberg (2011) 1
16. Singh, N., Hinkle, J., Joshi, S., Fletcher, P.T.: A hierarchical geodesic model for diffeomorphic longitudinal shape analysis. In: Gee, J., Joshi, S., Pohl, K., Wells, W., Zöllei, L. (eds.) IPMI 2013, LNCS. vol. 7917, pp. 560–571. Springer (2013) 1
17. Trouvé, A., Vialard, F.X.: Shape splines and stochastic shape evolutions: A second order point of view. arXiv preprint arXiv:1003.3895 (2010) 1