



**HAL**  
open science

## Relative Regression on Riemannian Manifolds

C. Samir

► **To cite this version:**

C. Samir. Relative Regression on Riemannian Manifolds. Université Clermont Auvergne (UCA). 2020.  
hal-04291438

**HAL Id: hal-04291438**

**<https://hal.science/hal-04291438>**

Submitted on 17 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Relative Regression on Riemannian Manifolds

Salah Khardani

*Intelligent Networks and Nanotechnology Laboratory*

Chafik Samir\*

*University of Clermont Auvergne*

Submitted to JMA 2020

## Abstract

The considerations of this paper are restricted to random variables with values on Riemannian manifolds  $\mathcal{M}$ , and hence we propose a geometric framework to estimate their relative regression function. Suppose we are given observations  $(X_i, Y_i)_{i=1 \dots n}$ , where  $X_i \in \mathcal{M}$  and  $Y_i \in \mathbb{R}_*^+$ . In this work we define and study a new estimator of the regression function on Riemannian Manifold  $\mathcal{M}$ . Precisely, we use the mean squared relative error (MSRE) as a loss function to construct a nonparametric estimator of the regression operator on Riemannian Manifolds. Under some standard assumptions in Riemannian Manifolds data analysis, we establish the almost sure consistency, with rates, of the constructed estimator as well as its asymptotic normality. Then, a simulation study, on finite-sized samples, was carried out in order to show the efficiency of our estimation procedure.

**Keywords**— Riemannian manifolds, Consistency, Asymptotic normality, Relative regression, Mean squared relative error.

---

\*Corresponding author: [chafik.samir@uca.fr](mailto:chafik.samir@uca.fr)

# 1 Introduction

Regression analysis is a fundamental statistical tool for determining how a measured variable is related to one or more potential explanatory variables. The most widely used regression model is linear regression, due to its simplicity, ease of interpretation, and ability to model many phenomena. However, if the response variable takes values on a nonlinear manifold (non Euclidean space), a linear model is not applicable. Theoretically, when studying the statistical problem of estimating a regression or density function defined on the non-Euclidean, all techniques used in the Euclidean space cannot be properly used. For solving this problem, the exponential map will be introduced, which was actually one of the main topics in [Park, 2004].

Non-parametric models have been used extensively in last years in fitting data lying on Riemannian manifolds, including kernel-based methods and approaches.

This research is motivated by the necessity for automated methods to analyze such objects. Indeed, for many applications in several branches of science, involving medical imaging and computer vision, (directional data, transformations, tensors, and shape) it is desirable to be able to characterize objects and predict their behavior at unobserved times or in the future. Several works have studied the regression problem on manifolds. [Kim and Park, 2013] studied a geometric structures arising from kernel density on Riemannian manifolds. [Park, 2012] studied a density kernel estimator defined by an exponential map on  $S^d$ . [Hendriks, 2003] developed a generalization of estimation with Fourier series to the case of a compact Riemannian manifold without boundary. [Healy and Kim, 1996] studied a nonparametric convolution density estimation on the unit sphere. [Pelletier, 2005] proposed a family of nonparametric estimators for the density function based on kernel weight when the variables are random object valued in a closed Riemannian manifold. The Pelletier's estimators is consistent with the kernel density estimators in the Euclidean, i.e., it coincides to the kernel estimator considered by [Rosenblatt, 1956] and [Parzen, 1962] when the manifold is  $\mathbb{R}^d$ .

[Prentice, 1989] introduces an extension of spherical regression using a Stiefel manifold as the sample space. [Lee and Ruymgaart, 1996] investigate density and curve estimation on compact smooth submanifolds of a Euclidean space using caps. In a more general setting, [Bhattacharya and Patrangenaru, 2002] studied a non-parametric estimation of location and dispersion on Riemannian manifolds. Under classical assumptions of the kernel and the bandwidth, [Pelletier, 2006] obtained the asymptotic bias and variance and proved the same  $L^2$  rate convergence as kernel regression estimators on Euclidean spaces using a geodesic distance. [Yanovich, 2016] studied the asymptotic properties of local sampling on Riemannian manifolds.

Usually, to estimate the nonparametric regression model, many previous works used the mean squares error (MSE) as a criterion for constructing the predictors, see some details in [Calsavara et al., 2019]. This method is very sensitive to outliers, and therefore, the presence of large outliers can lead to inappropriate results.

The mean square relative error (MSRE) criterion is more convenient as a measure of performance since the notion of relative regression is more recent than the others, although the results are still limited. [Park and Stefanski, 1998] studied prediction for situations in which relative prediction error is more important than the usual prediction error and derived the form of the best mean squared relative error predictor. [Jones et al., 2008] studied nonparametric prediction via relative error regression. [Khardani, 2019] studied nonparametric relative regression under random censorship model. They established the uniform consistency with rate over a compact set and asymptotic normality of the estimator. [Mechab and Laksaci, 2016] studied this regression model when the observations are weakly dependent. Recently, [Khardani and Slaoui, 2019] studied the asymptotic properties of a consistent estimator of this model by using the kernel method for twice censored data.

In this paper, we show that it makes sense to define a relative regression on  $\mathcal{M}$  as the result of a least squares minimization. The proposed method of calculating the regression estimator will be shown to enjoy a number of nice properties: it is a natural analogue of the Euclidean solution which could be considered as a generalization and the solution exists and is unique in many common situations. In particular, we consider the nonparametric estimation of a relative regression function on a Riemannian manifold by using a generalization of the kernel method, introduced by [Pelletier, 2005, Pelletier, 2006] for classical regression estimation.

The paper is organized as follows: in Section 2 we present some geometric background and we define the model and the estimator. In Section 3 we give some assumptions and results for the proposed estimator. Applications (Simulations and real data) are drawn in Section 4. The technical details and the proofs are deferred to Section 5.

## 2 Preliminaries and estimators

### 2.1 Geometric background

Let  $\mathcal{M}$  be a finite dimensional manifold. An important geometrical construct for the statistical analysis is the definition of the tangent space. Since  $\mathcal{M}$  could be seen a Riemannian submanifold in  $\mathbb{L}^2(I, \mathbb{R}^{N+1})$  [Samir et al., 2012], at any point  $p \in \mathcal{M}$ , we define the tangent space and we denote  $T_p\mathcal{M}$  [Helgason, 1978]. We equip the tangent space of  $\mathcal{M}$  with a smoothly varying Riemannian metric that measures infinitesimal

lengths on  $\mathcal{M}$ . In many cases, the inner product is first defined on  $\mathbb{L}^2$  and then induced on the tangent space of  $\mathcal{M}$ . More precisely, let  $v$  and  $w$  be two tangent vectors in  $T_p\mathcal{M}$ , the metric is defined as,

$$\langle v, w \rangle = \int_0^1 (v(s), w(s))_{\mathbb{R}^{N+1}} ds. \quad (1)$$

We denote by  $g$  the metric defined on  $\mathcal{M}$  with nice physical interpretation and by  $d_g$  the distance induced by  $g$ . Another important step in our analysis is to compute geodesic paths between points with respect to the chosen metric. With respect to the geometric structure,  $\mathcal{M}$  is considered as submanifold in  $\mathbb{L}^2(I, \mathbb{R}^{N+1})$  and consequently we can use some standard results from Riemannian geometry, see for example [Siegel, 1964, Lang, 1995], including geodesics and exponential map. Therefore, we assume that  $(\mathcal{M}, g)$  is geodesically complete, at least locally. So, for any point  $p \in \mathcal{M}$ , there exist a ball  $B_\rho(p)$  where all geodesics exist and are unique. For simplicity, we call  $B_\rho(p)$  the cut locus of  $p$  with maximum  $\rho$  and  $inj_g(\mathcal{M})$  the radius of injectivity of  $\mathcal{M}$ .

For the rest of the paper, we will assume that  $inj_g(\mathcal{M})$  is strictly positive. There are many examples of manifolds that have this property, we cite for example an Euclidean space with infinite  $inj_g(\mathcal{M})$ , the unit finite-dimensional sphere with  $inj_g(\mathcal{M}) = \pi$ , the manifold of symmetric positive definite matrices, the Riemannian  $n$ -hyperboloid with constant negative curvature, etc. The interested reader can refer to [Jost, 2011, Lang, 1995] for more details. Under these conditions, the geodesic path  $\chi_t(p, q)$ , between any two points  $p$  and  $q$  with  $q \in B_\rho(p)$  on  $\mathcal{M}$  can be expressed in terms of a tangent direction  $v \in T_p\mathcal{M}$  with,  $t \in [0, 1]$ ,  $\chi_0(p, q) = p$  and  $\chi_1(p, q) = q$ . This formulation translates the constant-speed parametrization of the geodesic passing through  $p$  with velocity vector  $v$  at  $t = 0$ . As a result, the exponential map  $\exp : T_p\mathcal{M} \rightarrow \mathcal{M}$  is given as  $\exp_p(v) = q$ . So, instead of working on  $\mathcal{M}$  directly, we can use the tangent space  $T_p(\mathcal{M})$  that has a natural differential structure. This is made possible because  $\exp$  is a bijective isometry on  $B_\rho(p)$ . It is worth noticing that, when  $\mathcal{M}$  is Euclidean,  $\exp$  is simply the identity. The velocity vector along a geodesic path  $\chi_t$  can be obtained as  $\frac{\partial \chi_t}{\partial t} \Big|_{t=0}$ . Thus, given two points  $p$  and  $q$ , the inverse exponential map (also known as the logarithmic map) allows the recovery of the tangent vector  $v$  between them.

To summarize, the exponential map takes points from  $\mathcal{M}$  in the tangent space  $\mathcal{M}$ , preserving distance from  $p$ ; it also preserves the tangential direction from  $p$ . Concretely, the exponential map only preserves angles and distances for points in the tangent plane which have the geodesic distance  $d_g(p, \cdot) \leq \rho(p)$  from  $p$ ; however,

we shall implicitly assume this condition holds whenever it is needed. Given the above tools for constructing geodesics and inverse exponential maps on  $\mathcal{M}$ , we will indicate in the next section how these tools may be used to solve the problem of regression from a given set of observations on  $\mathcal{M}$ . Moreover, we will need to introduce the notion of the volume density function  $\theta_p(q)$  on  $\mathcal{M}$  as defined by [Besse, 1978], we have

$$\theta_p : q \longrightarrow \theta_p(q) = \frac{\mu_{\exp_p^* g}}{\mu_{g_p}}(\exp_p^{-1}(q)) \quad (2)$$

as the quotient of the canonical measure of the Riemannian metric  $\exp_p^* g$  on  $T_p(\mathcal{M})$ .

Now, let  $X$  be a random object on  $\mathcal{M}$ , *i.e.* a measurable map from a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  to  $(M, B)$ , where  $B$  denotes the Borel  $\sigma$ -field of  $\mathcal{M}$ . Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be independent, identically distributed (i.i.d.) pairs of random variables valued in  $\mathcal{M} \times \mathbb{R}^+$  with the same distribution as  $(X, Y)$ . For regression analysis, our main objective is estimating  $Y$  given  $X = p$ . As it is largely admitted, the natural solution of this problem in Euclidean spaces is given by the estimators of the regression function  $r$  defined by  $Y = r(X) + \varepsilon$ . Following [Pelletier, 2005, Pelletier, 2006], we present a solution on  $\mathcal{M}$  and we discuss the estimators in the next section.

## 2.2 Model and the estimator

Let us introduce  $n$  pairs of random variables  $(X_i, Y_i)$  for  $i = 1 \dots n$  that we assume drawn from the pair  $(X, Y)$  which is valued in  $\mathcal{M} \times \mathbb{R}$ . Furthermore, we assume that the variables  $X$  and  $Y$  are connected by the following relation

$$Y = r(X) + \varepsilon$$

where  $r$  is an unknown function and  $\varepsilon$  is a random error variable independent of  $X$ .

Usually, we use the least square error  $\mathbb{E}[(Y - r(X))^2 | X = p]$  as a loss function to estimate the regression function. However, this kind of loss function is very sensitive to outliers. Indeed, this loss function treats all variables as having an equal weight. Thus, the presence of large outliers can lead to irrelevant results. For this reason, in this contribution, we overcome this drawback by using an alternative loss function based on the relative squared error which is defined, for  $Y > 0$ , by

$$\mathbb{E} \left[ \left( \frac{Y - r(X)}{Y} \right)^2 | X \right], \quad \text{for } Y > 0. \quad (3)$$

Clearly, this criterion is more interesting than the squared error, when the range of predicted values is large. In particular, this kind of obtained regression is less sensitive to the presence of outliers, see [Khardani and Slaoui, 2019]. Moreover, the solution of the minimization of Equation (3) can be explicitly expressed by

$$r(p) := \frac{\mathbb{E}[Y^{-1}|X = p]}{\mathbb{E}[Y^{-2}|X = p]}.$$

Then, our proposal in this paper is to estimate  $r(p)$  by:

$$r_n(p) := \frac{\psi_{1,n}(p)}{\psi_{2,n}(p)} \quad (4)$$

where

$$\mathbb{E}[Y^{-\ell}|X = p] := \frac{\psi_\ell(p)}{f(p)}$$

$$\psi_{l,n}(p) = \frac{1}{n} \sum_{i=1}^n Y_i^{-\ell} \frac{1}{h_n^d} \frac{1}{\theta_{X_n}(p)} K\left(\frac{d_g(p, X_i)}{h_n}\right) \quad \text{for } l \in \{1, 2\}.$$

and

$$f_n(p) = \frac{1}{n} \sum_{k=1}^n \frac{1}{h_n^d} \frac{1}{\theta_{X_k}(p)} K\left(\frac{d_g(p, X_k)}{h_n}\right)$$

### 3 Assumptions and main results

The closed ball in  $(\mathcal{M}, g)$  of center  $p$  and of radius  $h$  will be denoted by  $B_{\mathcal{M}}(p, h)$ , and the closed ball in  $\mathbb{R}^d$  centered at the origin and of radius  $h$  will be denoted by  $B(h)$ .

Throughout the paper, we denote by  $r_\ell(\cdot) = \mathbb{E}[Y^{-\ell}|X = \cdot]$  the conditional  $\ell$ -inverse moments of  $T$  given  $X$  and by  $\psi_\ell(\cdot) = \mathbb{E}[Y^{-\ell}|X = \cdot]f(\cdot)$ , with  $f(\neq 0)$  the density of  $X$ ;  $\ell \in \{1, 2, 3, 4\}$ . When no confusion is possible, we denote by  $M$  and/or  $M'$  any generic positive constant. Define  $J_0 := \{p \in \mathcal{M} \text{ such that } \psi_2(\cdot) > 0\}$  and let  $S$  be a compact subset of  $J_0$ . Then we suppose that

$$\inf_{p \in S} \psi_2(p) > 0$$

and

$$A = \left\{ x \in S, \frac{\{\psi_2(x) - 2r(x)\psi_3(x) + r^2(x)\psi_4(x)\}}{\psi_2^2(x)} \neq 0 \right\}.$$

In this section, we want to establish the uniform almost complete convergence (*a.s.*) of  $r_n(x)$  to  $r(x)$  over a compact  $S$ . Firstly, we consider the following assumptions : **Assumption H1.** :  $f$  is bounded and two-times continuously differentiable at  $p$  such that:

- (i)  $\inf_{p \in S} f(p) > M > 0$ .
- (ii)  $\inf_{p, q \in S} \theta_p(q) > M' > 0$ .

**Assumption H2.**  $h_n < \frac{h_*}{2}$  where  $h_* = \min\{inj(\mathcal{M}), \frac{\pi}{2\sqrt{\kappa}}\}$ . where  $\kappa$  is the least upper bound of sectional curvatures of  $\mathcal{M}$  if this upper bound is positive, and  $\kappa = 0$  otherwise.

**Assumption H3.** The smoothing parameter  $(h_n)$  satisfies:

$$(i) \lim_{n \rightarrow +\infty} \frac{\log n}{nh_n^d} = 0, \quad (ii) \lim_{n \rightarrow +\infty} nh_n^{d+4} = 0.$$

**Assumption H4.**

- (i) The kernel  $K$  is Continuously differentiable compactly supported density function and is Holder of order  $\gamma$ .
- (ii)  $\int_{\mathbb{R}} \|v\|^2 K(\|v\|) dv < \infty$ .
- (iii)  $\left\langle \int_{B(0,1)} K(\|v\|) v dv, \text{grad} \psi_\ell(v) \right\rangle = 0$  for  $\ell = 1, 2$ .

**Assumption H5.** The inverse conditional moments of the response variable  $r_\ell(p) = \mathbb{E}[Y^{-\ell} | X = p]$  is two-times continuously differentiable at  $p$  and  $\sup_{p \in \mathcal{M}} |(\nabla^2 \psi_\ell(p))_{i,j}| < +\infty$ ,  $r_2(p) > M'$ ,  $\forall p \in \mathcal{M}$ , for  $\ell \in \{1, 2\}$ .

**Assumption H6.** The inverse conditional moments verify:  
 $\forall p \in \mathcal{M}$ , and  $\forall \ell \geq 1$ ,  $r_\ell(p) = \mathbb{E}[Y^{-\ell} | X = p]$  exists.

## Discussion of the assumptions.

We impose the condition (H2) on the bandwidth :  $h_n < inj_g(\mathcal{M})$  where  $inj_g(\mathcal{M})$  denotes the injectivity radius of  $\mathcal{M}$  (Chavel, 1993, p108). Since  $M$  is compact,  $inj_g(\mathcal{M})$  is strictly positive by the theorem of Whitehead. The condition  $h_n < inj_g(\mathcal{M})$  guarantees for each  $p \in \mathcal{M}$  the existence of a normal coordinate neighborhood at  $p$  containing  $B_{\mathcal{M}}(p, h_n)$ , the ball in  $\mathcal{M}$  centered at  $p$  and of a radius  $h$ . Assumption



(H4) is a standard assumption when dealing kernel estimators. Furthermore, Assumptions (H5 - H6) are regularity conditions which permit us to evaluate the bias and variance terms of the estimator (4). The fact that  $\theta_p(p) = 1$  for all  $p \in \mathcal{M}$  guarantees that H1(ii) for a small compact neighborhood of  $p$ .

**Theorem 3.1** *Under Assumptions (H0)-(H5), we have*

$$\sup_{p \in S} |r_n(p) - r(p)| = O \left( \max \left\{ \sqrt{\frac{\log n}{nh_n^d}}, h_n^2 \right\} \right) \quad a.s. \quad (5)$$

The following corollary is a particular case of the previous theorem.

**Corollary 3.2** *Under Assumptions (H0)-(H5) and for  $(h_n) = \left( \left( \frac{\log n}{n} \right)^{1/(d+4)} \right)$ , we have*

$$\sup_{p \in S} |r_n(p) - r(p)| = O \left( \left( \frac{\log n}{n} \right)^{2/(d+4)} \right) \quad a.s.$$

**Theorem 3.3** *Under assumptions (H0)-(H6), then for any  $p \in A$ , we have*

$$\left( \frac{nh_n^d}{\sigma^2(p)} \right)^{1/2} (r_n(p) - r(p)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{as } n \rightarrow \infty,$$

where

$$\sigma^2(p) = \frac{\{\psi_2(p) - 2r(p)\psi_3(p) + r^2(p)\psi_4(p)\}}{\psi_2^2(p)} \int_{B(1)} K^2(\|z\|) dz.$$

and

$$A = \left\{ p \in \mathcal{M}, \frac{\{\psi_2(p) - 2r(p)\psi_3(p) + r^2(p)\psi_4(p)\}}{\psi_2^2(p)} \neq 0 \right\},$$

$\xrightarrow{\mathcal{D}}$  denotes convergence in distribution,  $\mathcal{N}$  the Gaussian-distribution.

## 4 Applications

As part of potential applications, we would like to point out that the method developed in this work can be used for more general problems under the assumptions detailed in section 3. As a first example, we want to simulate a regression problem on Riemannian manifolds with positive constant curvature. This setting is very common in different applications where data are longitudinal. We cite for example, wind

direction, temperature values and changes, hurricane-force winds from the center, tropical-storm-force winds, bird migration, etc.

We now examine in details the geometric structure of finite-dimensional Riemannian manifolds with positive constant curvature  $k > 0$ , and provide explicit formulas for computation. In order to avoid trivial cases we assume  $N \geq 2$ . The most used space is the hypersphere for two reasons: i) manifolds with constant  $k > 0$  can be bijectively mapped to the unit sphere of the same dimension, and ii) lot is known about the geometry of the sphere. Here we provide some of the useful definitions from spherical geometry. For more details, we refer the reader to [John M. Lee, 1997]. Since there is no restrictions for the dimension, we denote by  $\mathbb{S}^N$  the unit sphere in  $\mathbb{R}^{N+1}$ .

$$\mathbb{S}^N = \{p \in \mathbb{R}^{N+1} \mid \langle p, p \rangle_2 = 1\}, \quad (6)$$

where  $\langle \cdot, \cdot \rangle_2$  is the standard inner product. Thus, for any  $p$  and  $q$  we have

$$\langle p, q \rangle = \sum_{i=1}^{N+1} p^i q^i. \text{ Every Riemannian manifold isometrically diffeomorphic to } \mathbb{S}^N$$

will be identified with  $\mathbb{S}^N$ . For  $p \in \mathbb{S}^N$ , the tangent space to  $\mathbb{S}^N$  in  $p$  is given by  $T_p(\mathbb{S}^N) = \{v \in \mathbb{R}^{N+1} \mid \langle p, v \rangle_2 = 0\}$ . The restriction of the inner product to closed ball with radius small than  $\pi$  in  $T_p(\mathbb{S}^N)$  is positive definite. This metric is differentiable and therefore  $\mathbb{S}^N$  has a locally complete Riemannian structure. Starting from a point  $p \in \mathbb{S}^N$  and a unit vector  $v \in T_p(\mathbb{S}^N)$  with  $\|v\| \lesssim \pi$ , we can construct a unique geodesic  $\gamma : \mathbb{R} \rightarrow \mathbb{S}^N$  and the arc-length (geodesic distance) between distinct and non antipodal  $p$  and  $q$  on  $\mathcal{M}$  is the angle  $\beta = \arccos(\langle p, q \rangle_2)$ .

The exponential map is a bijective isometry from the tangent space  $T_p(\mathcal{M})$  to  $\mathcal{M}$ . For any  $w \in T_p(\mathcal{M})$ , we write

$$\text{Exp}_p(w) = \cos(\|w\|_2)p + \sin(\|w\|_2) \frac{w}{\|w\|_2}. \quad (7)$$

Its inverse, the log map is defined from  $\mathcal{M}$  to  $T_p(\mathcal{M})$  as

$$\text{Log}_p(q) = \frac{\beta}{\sin(\beta)}(q - \cos(\beta)p). \quad (8)$$

For any two elements  $p$  and  $q$  on  $\mathcal{M}$  the map  $\Gamma : T_p(\mathcal{M}) \rightarrow T_q(\mathcal{M})$  parallel transports a vector  $w$  from  $p$  to  $q$  and is given by:

$$\Gamma_{p \rightarrow q}(w) = w - 2 \frac{(p+q)}{\|p+q\|_2^2} \langle w, q \rangle_2 \quad (9)$$

Now that we have explicit formulas for geometric tools that can insure conditions in section 3, we can apply our method for regression on  $\mathbb{S}^N$ . An important issue in any kernel-based estimation is the choice of the optimal bandwidth. In order to apply our method for the simulation study and the real application considered in next sections, we use a standard method for the choice of the bandwidth. In particular, an optimal choice of  $h_n$  which minimizes the asymptotic mean squared error (MSE) of the estimator  $r_n$  at a point  $x$  is  $(h_n) = \left( \left( \frac{\log n}{n} \right)^{1/(d+8)} \left( \frac{d}{8} \right)^{2/(d+8)} \right)$ . In our experiments, we use the most common method based on least square cross-validation method to search for an asymptotically optimal data-driven bandwidth and thus, adaptive data-driven estimators. Consequently, what we call training data are points used to find optimal bandwidths where the rest is kept for evaluation. For both simulations and real application, the performance of an estimate  $\hat{r}$  of  $r$  is measured using the Root Mean Square Error (RMSE).

## 4.1 Simulations

In this section, we consider two applications, in each investigating the spherical regression models towards different statistical goals. The first involves a simulated data set with a scalar covariate,  $y \in \mathbb{R}$ . We use simple data-generating models to illustrate the flexibility within this regression framework for the mean direction and dispersion to depend on the covariate; to investigate the performance of regression. For simplicity, we consider an equivalent situation where  $X$  is a random position on  $\mathbb{S}^N$  ( $N = 2$ ) representing the sphere and  $Y \in [0, 1]$ . Consequently, the regression problem consists in using available observations  $(X_i, Y_i)$ ,  $i = 1, \dots, n$  in order to predict  $Y$  for unknown positions. To better understand this situation, we display two examples of noisy data on  $\mathbb{S}^2$  (earth here) Figure 2(a), use some observations to better approximate the model Figure 2(b) and then use the proposed method for regression Figure 2(c). We illustrate this idea with Y-value colors for a better visualization.

## 4.2 Real Application

The second dataset concern the air quality as the covariate and the position (longitude and latitude) as the manifold-valued measurement. In this application, we use a dataset consisting of 400 measured at different locations in Europe from January 2015 to July 2019. The dataset is a part of the data archive provided by the European Environment Agency and is available at <https://www.eea.europa.eu>. The dataset contain observations about the air quality in individual countries, regions

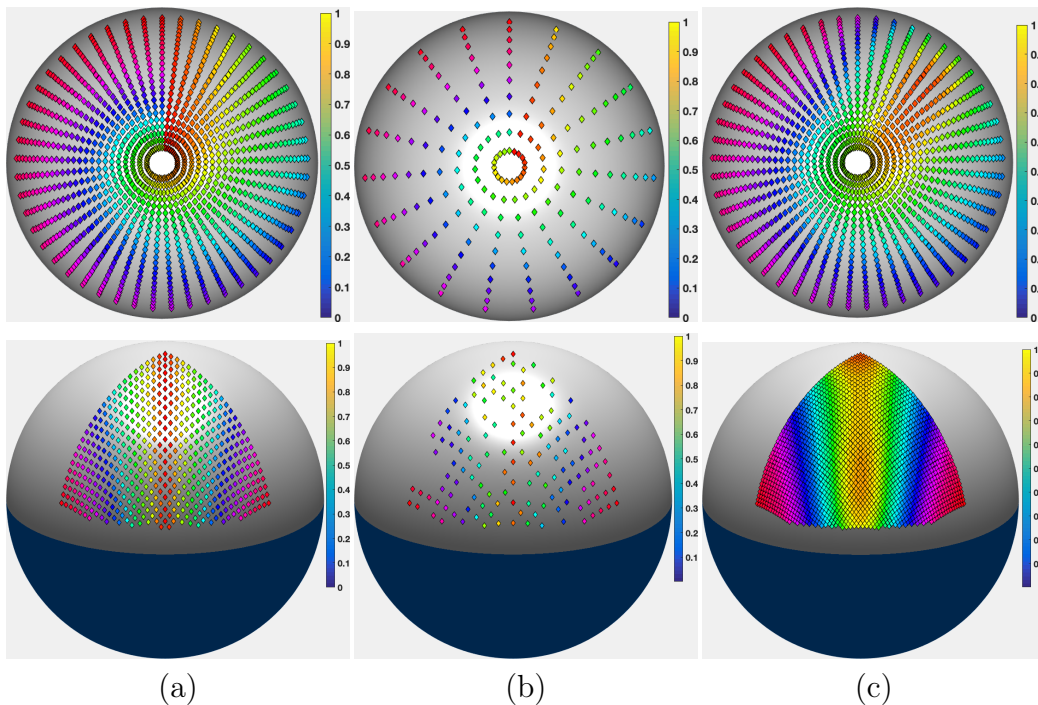


Figure 1: Simulation 1 (top) & 2 (bottom) for regression on unit sphere  $\mathbb{S}^2$ . For each example: (a) represents original data with noise, (b) examples from training, and (c) the predicted values.

and cities. The quality is based on concentration values of different key pollutants to give an idea about the impact on health. Since data do not cover all locations (monitoring stations) and some values are either missing or wrong when data are not formally verified, the relative spherical regression is more convenient. 75% was used for training and the remaining 25% was reserved for testing the performance of the proposed model. For all applications, the proposed model showed a good performance for predicting covariate. In particular, simulation results in simulation study show robustness against noise and outliers. Note that the optimal bandwidth is data-dependent but because the geodesic distance on distance on the sphere is defined on wide ball, the impact of  $h_n$  is moderate. We expect this property to hardly impact the performance if the radius of injectivity is relatively small. Nevertheless, there are many Riemannian manifolds where the geodesics are available with explicit formulas or can be efficiently approximated. Thus, the proposed model can be generalized for other manifolds with different geometric structures but the conclusions

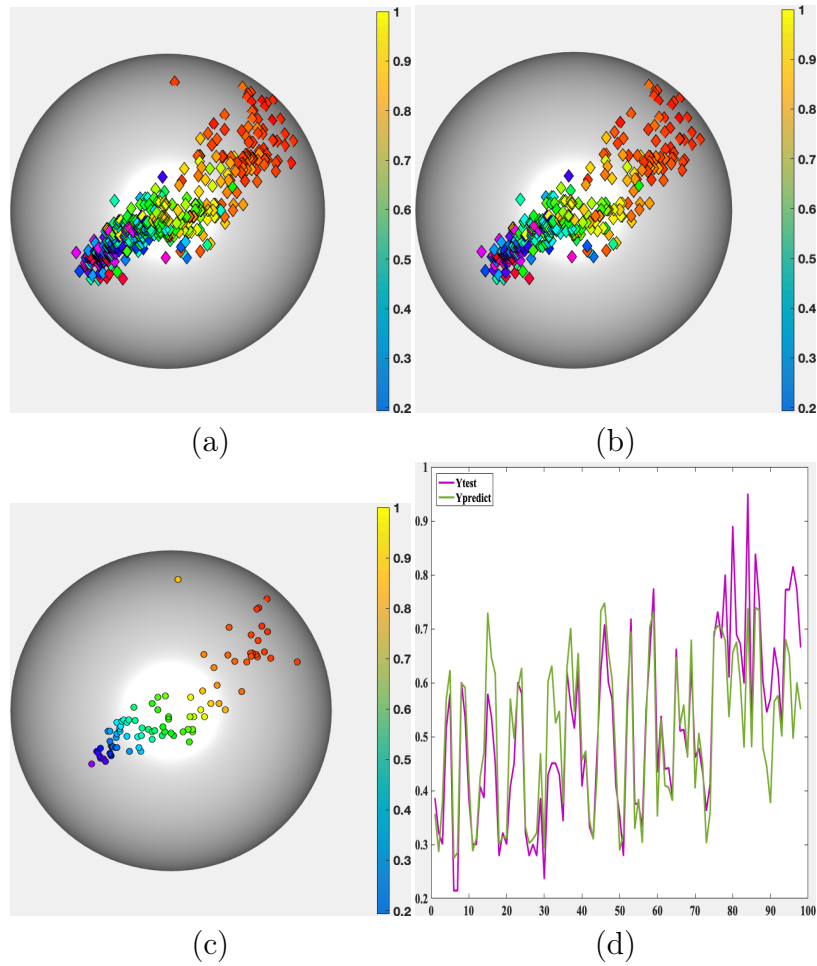


Figure 2: Regression on air quality: (a) represents examples from original data, (b) training, (c) test supposed unknown, and (d) original covariate in red and the prediction in green. In this application, the RMSE between test and predicted by the proposed model is 0.0072.

remain the same.

## 5 Proofs and auxiliary results

**Lemma 1** *Under assumptions of Theorem 3.1 and for  $l \in \{1, 2\}$ , we have*

$$\sup_{p \in S} |\mathbb{E}(\psi_{l,n}(p)) - \psi_l(p)| = O(h_n^2).$$

**Lemma 2** *Under assumptions of Theorem 3.1 and for  $l \in \{1, 2\}$  we obtain*

$$\sup_{p \in S} |\psi_{l,n}(p) - \mathbb{E}(\psi_{l,n}(p))| = O\left(\sqrt{\frac{\log n}{nh_n^d}}\right).$$

**Proof of Theorem 3.1.** We consider the following decomposition:

$$r_n(p) - r(p) = \frac{1}{\psi_{2,n}(p)} [\psi_{1,n}(p) - \psi_1(p)] + [\psi_2(p) - \psi_{2,n}(p)] \frac{r(p)}{\psi_{2,n}(p)} \quad (10)$$

The proof of this theorem is based on the decomposition given by (10) and is a direct consequence of Lemmas (1) and (2), for which the proofs are given in the appendix.

**Proof of Theorem 3.3:** We write

$$r_n(p) - r(p) = \frac{1}{\psi_{2,n}(p)} [D_n + W_n(\psi_{2,n}(p) - \mathbb{E}[\psi_{2,n}(p)])] + W_n, \quad (11)$$

where

$$\begin{aligned} W_n &= \frac{1}{\mathbb{E}[\psi_{2,n}(p)] \psi_2(p)} \{ \mathbb{E}[\psi_{1,n}(p)] \psi_2(p) - [\mathbb{E}[\psi_{2,n}(p)]] \psi_1(p) \} \\ D_n &= \frac{1}{\psi_2(p)} [ [\psi_{1,n}(p) - \mathbb{E}[\psi_{1,n}(p)]] \psi_2(p) + [\mathbb{E}[\psi_{2,n}(p)] - \psi_{2,n}(p)] \psi_1(p) ]. \end{aligned}$$

Then, it follows from (11), that

$$\begin{aligned} r_n(p) - r(p) - W_n &= \frac{1}{\psi_{2,n}(p)} [D_n + W_n(\psi_{2,n}(p) - \mathbb{E}[\psi_{2,n}(p)])] \\ &= : \frac{D_n + W_n \mathcal{J}_{2n}(p)}{\psi_{2,n}(p)}, \end{aligned}$$

where

$$\mathcal{J}_{jn}(p) = \psi_{j,n}(p) - \mathbb{E}[\psi_{j,n}(p)], \quad \text{for } j = 1, 2.$$

Therefore, Theorem 3.3 is a consequence of the convergence rate of Lemmas (1) and (2) and the following intermediate results ( Lemmas 3–4), for which the proofs are also postponed to the appendix.

**Lemma 3** Under assumptions of Theorem 3.3, we obtain

$$\left(\frac{nh_n^d}{\sigma^2(p)\psi_2^2(p)}\right)^{1/2} (D_n - \mathbb{E}[D_n]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

**Lemma 4** Under assumptions of Theorem 3.3, we obtain

$$\psi_{2,n}(p) \xrightarrow{\mathbb{P}} \psi_2(p),$$

$$\left(\frac{nh_n^d}{\sigma^2(p)\psi_2^2(p)}\right)^{1/2} W_n \rightarrow 0,$$

and

$$\left(\frac{nh_n^d}{\sigma^2(p)\psi_2^2(p)}\right)^{1/2} W_n(\psi_{2,n}(p) - \mathbb{E}[\psi_{2,n}(p)]) \xrightarrow{\mathbb{P}} 0,$$

where  $\xrightarrow{\mathbb{P}}$  denotes convergence in probability.

## 5.1 Appendix

**Proof of Lemma 1** Using a conditional expectation, we have

$$\begin{aligned} \left| \mathbb{E}[\psi_{\ell,n}(p)] - \psi_{\ell}(p) \right| &= \left| \mathbb{E}\left[ Y_i^{-\ell} \frac{1}{h_i^d} \frac{1}{\theta_{X_i}(p)} K\left(\frac{d_g(p, X_i)}{h_i}\right) - \psi_{\ell}(p) \right] \right| \\ &= \left| \int_M r_{\ell}(q) \frac{1}{h_n^d} \frac{1}{\theta_q(p)} K\left(\frac{d_g(p, q)}{h_n}\right) f(q) d\nu_g(q) - \psi_{\ell}(p) \right| \\ &= \left| \int_{B_M(p, h_n)} \frac{1}{h_n^d} \frac{1}{\theta_q(p)} K\left(\frac{d_g(p, q)}{h_n}\right) (\psi_{\ell}(q) - \psi_{\ell}(p)) d\nu_g(q) \right| \\ &= \left| \frac{1}{2} \left( \nabla^2 \psi_{\ell}(p)_{i,j} \int_{B(1)} K(\|u\|) u^i u^j du \right) h_n^2 + o(h_n^2) \right| \end{aligned}$$

Assumptions **(H4 – H5)** conclude the proof.

**Proof of Lemma 3.** It is clear that:  $\mathbb{E}[D_n] = 0$ , so  $Var[D_n] = \mathbb{E}[D_n^2]$ . Then

$$\begin{aligned} \mathbb{E}[D_n^2] &= \mathbb{E}\{[\psi_{1,n}(x) - \mathbb{E}(\psi_{1,n}(x))]^2\} + \left\{ \frac{\psi_1(x)}{\psi_2(x)} \right\}^2 \mathbb{E}\{[\psi_{2,n}(x) - \mathbb{E}(\psi_{2,n}(x))]^2\} - 2 \frac{\psi_1(x)}{\psi_2(x)} cov(\psi_{1,n}(x), \psi_{2,n}(x)) \\ &= Var(\psi_{1,n}(x)) + r^2(x) Var(\psi_{2,n}(x)) - 2r(x) Cov(\psi_{1,n}(x), \psi_{2,n}(x)). \end{aligned}$$

We have

$$\begin{aligned}
Var[\psi_{1,n}(x)] &= n^{-1}h_n^{-2d}var\left[\frac{1}{\theta_{X_n}(p)}Y_n^{-1}K\left(\frac{d_g(p, X_n)}{h_n}\right)\right] \\
&= \frac{1}{nh_n^d}\mathbb{E}\left[\frac{1}{h_n^d}\frac{1}{\theta_{X_n}^2(p)}Y_n^{-2}K^2\left(\frac{d_g(p, X_n)}{h_n}\right)\right] - \frac{1}{n}\mathbb{E}^2\left[\frac{1}{h_n^d}\frac{1}{\theta_{X_n}(p)}Y_n^{-1}K\left(\frac{d_g(p, X_n)}{h_n}\right)\right] \\
&= \frac{1}{nh_n^d}\mathbb{E}\left[\frac{1}{h_n^d}\frac{1}{\theta_{X_n}^2(p)}Y_1^{-2}K^2\left(\frac{d_g(p, X_n)}{h_n}\right)\right] - \frac{1}{n}(\psi_1(p) + o(1))^2 \\
&=: \mathcal{K}_{1n}(p) + \mathcal{K}_{2n}(p).
\end{aligned}$$

For  $\mathcal{K}_{2n}(x)$  it is clear,

$$\mathcal{K}_{2n}(x) = \frac{1}{n}(\psi_1(p) + o(1))^2 = o(1). \quad (12)$$

For  $\mathcal{K}_{1n}(x)$ , under **(H4–H6)**, using again the conditional expectation properties and a Taylor expansion, we get,

$$\begin{aligned}
\mathcal{K}_{1n}(p) &= \frac{1}{nh_n^d}\mathbb{E}\left[\frac{1}{h_n^d}\frac{1}{\theta_{X_n}^2(p)}Y_1^{-2}K^2\left(\frac{d_g(p, X_n)}{h_n}\right)\right] \\
&= \frac{1}{nh_n^d}\int_M \frac{1}{h_n^d}\frac{1}{\theta_q^2(p)}K^2\left(\frac{d_g(p, q)}{h_n}\right)\psi_2(q)d\nu_g(q) \\
&= \frac{1}{nh_n^d}\psi_2(p)\int_{B(1)}K^2(\|z\|)dz \quad (13)
\end{aligned}$$

Then, (12),(13) and a Taylor expansion ensures that

$$Var(\psi_{1,n}(p)) = \frac{1}{nh_n^d}\mathbb{E}[Y_i^{-2}|X_i = p]f(p)\int_{B(1)}K^2(\|z\|)dz. \quad (14)$$

and

$$Var(\psi_{2,n}(x)) = \frac{1}{nh_n^d}\mathbb{E}[Y_i^{-4}|X_i = p]f(p)\int_{B(1)}K^2(\|z\|)dz \quad (15)$$

Moreover, we have

$$\begin{aligned}
&Cov(\psi_{1,n}(p), \psi_{2,n}(p)) \\
&= cov\left(\frac{1}{nh_n^d}\sum_{i=1}^n Y_i^{-1}\frac{1}{\theta_{X_i}(p)}K\left(\frac{d_g(p, X_i)}{h_n}\right), \frac{1}{nh_n^d}\sum_{j=1}^n Y_j^{-2}\frac{1}{\theta_{X_j}(p)}K\left(\frac{d_g(p, X_j)}{h_n}\right)\right) \\
&= \frac{1}{nh_n^{2d}}\left\{\mathbb{E}\left[\frac{1}{Y_i^{-3}\theta_{X_i}^2(p)}K^2\left(\frac{d_g(p, X_i)}{h_n}\right)\right] - \mathbb{E}\left[\frac{Y_i^{-1}}{\theta_{X_i}(p)}K\left(\frac{d_g(p, X_i)}{h_n}\right)\right]\mathbb{E}\left[\frac{Y_i^{-2}}{\theta_{X_i}(p)}K\left(\frac{d_g(p, X_i)}{h_n}\right)\right]\right\}
\end{aligned}$$



A simple calculation ensures that

$$\frac{1}{nh_n^{2d}} \mathbb{E} \left[ Y_i^{-1} \frac{1}{\theta_{X_i}(p)} K \left( \frac{d_g(p, X_i)}{h_n} \right) \right] = o(1) \quad \text{and} \quad \frac{1}{nh_n^{2d}} \mathbb{E} \left[ Y_i^{-2} \frac{1}{\theta_{X_i}(p)} K \left( \frac{d_g(p, X_i)}{h_n} \right) \right] = o(1).$$

A Taylor expansion ensures that

$$\text{Cov}(\psi_{1,n}(p), \psi_{2,n}(p)) = \frac{1}{nh_n^d} \mathbb{E} [Y_i^{-3} | X_i = p] f(p) \int_{B(1)} K^2(\|z\|) dz + o \left( \frac{1}{nh_n^d} \right). \quad (16)$$

Then the combination of (14), (15) and (16) ensures that

$$\text{Var}[D_n] = \frac{f(p)}{nh_n^d} \left\{ \mathbb{E} [Y^{-2} | X = p] + r^2(p) \mathbb{E} [Y^{-4} | X = p] - 2r(p) \mathbb{E} [Y^{-3} | X = p] \right\} \int_{B(1)} K^2(\|z\|) dz.$$

Then, it follows that

$$\left( \frac{nh_n^d}{\sigma^2(p) \psi_2^2(p)} \right)^{1/2} (D_n - \mathbb{E}[D_n]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

**Proof of Lemma 4.** It follows from Lemma 1 that

$$\mathbb{E}[\psi_{2,n}(p) - \psi_2(p)] \rightarrow 0$$

Under assumptions **H1**, **H4** and **H5-H6**, we have :

$$\begin{aligned} \text{Var}[\psi_{2,n}(p)] &= n^{-1} h_k^{-2d} \text{var} \left[ \frac{1}{\theta_{X_k}(p)} Y_k^{-2} K \left( \frac{d_g(p, X_k)}{h_k} \right) \right] \\ &= \frac{1}{nh_n^d} \mathbb{E} \left[ \frac{1}{h_k^d \theta_{X_k}^2(p)} Y_k^{-4} K^2 \left( \frac{d_g(p, X_k)}{h_k} \right) \right] - \frac{1}{n} \mathbb{E}^2 \left[ \frac{1}{h_n^d \theta_{X_k}(p)} Y_k^{-2} K \left( \frac{d_g(p, X_k)}{h_k} \right) \right] \\ &= \frac{1}{nh_n^d} \mathbb{E} \left[ \frac{1}{h_k^d \theta_{X_k}^2(p)} Y_1^{-4} K^2 \left( \frac{d_g(p, X_k)}{h_k} \right) \right] - \frac{1}{n} (\psi_2(p) + o(1))^2 \\ &= O(n^{-1} h_n^{-d}) = o(1). \end{aligned}$$

Hence  $\psi_{2,n}(p) \xrightarrow{\mathbb{P}} \psi_2(p)$ .

Next, it is clear that the second limit of Lemma 4 is a consequence of the above convergence. Then, it suffices to treat the last result. For this, we use the fact that

$$\text{Var}[(\psi_{2,n}(p) - \mathbb{E}[\psi_{2,n}(p)])] = \text{Var}[\psi_{2,n}(p)] \rightarrow 0.$$

Then, by the application of Lemma 2 we obtain  $W_n = O(h_n^2)$ . Then, we deduce that

$$\left( \frac{nh_n^d}{\sigma^2(p) \psi_2^2(p)} \right)^{1/2} W_n (\psi_{2,n}(p) - \mathbb{E}[\psi_{2,n}(p)]) \xrightarrow{\mathbb{P}} 0.$$

## References

- [Besse, 1978] Besse, A. L. (1978). *Manifolds all of whose geodesics are closed*, volume 93 of *Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas]*. Springer-Verlag, Berlin-New York. With appendices by D. B. A. Epstein, J.-P. Bourguignon, L. Bérard-Bergery, M. Berger and J. L. Kazdan.
- [Bhattacharya and Patrangenaru, 2002] Bhattacharya, R. and Patrangenaru, V. (2002). Nonparametric estimation of location and dispersion on riemannian manifolds. *Journal of Statistical Planning and Inference*, 108(1):23–35. C.R. Rao 80th Birthday Felicitation Volume, Part II.
- [Calsavara et al., 2019] Calsavara, V. F., Rodrigues, A. S., Rocha, R., Tomazella, V., and Louzada, F. (2019). Defective regression models for cure rate modeling with interval-censored data. *Biometrical Journal*, 61(4):841–859.
- [Healy and Kim, 1996] Healy, Jr., D. M. and Kim, P. T. (1996). An empirical Bayes approach to directional data and efficient computation on the sphere. *Ann. Statist.*, 24(1):232–254.
- [Helgason, 1978] Helgason, S. (1978). *Differential Geometry, Lie Groups, and Symmetric spaces*. Academic Press, New York.
- [Hendriks, 2003] Hendriks, H. (2003). Application of fast spherical Fourier transform to density estimation. *J. Multivariate Anal.*, 84(2):209–221.
- [John M. Lee, 1997] John M. Lee (1997). *Riemannian Manifolds: An Introduction to Curvature*. Graduate texts in mathematics. Springer Science.
- [Jones et al., 2008] Jones, M. C., Park, H., Shin, K.-I., Vines, S. K., and Jeong, S.-O. (2008). Relative error prediction via kernel regression smoothers. *J. Statist. Plann. Inference*, 138(10):2887–2898.
- [Jost, 2011] Jost, J. (2011). *Chapter 3 The Laplace Operator and Harmonic Differential Forms*, page 89–131. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [Khardani, 2019] Khardani, S. (2019). Relative error prediction for twice censored data. *Math. Methods Statist.*, 28(4):291–306.

- [Khardani and Slaoui, 2019] Khardani, S. and Slaoui, Y. (2019). Nonparametric relative regression under random censorship model. *Statistics and Probability Letters*, 151:116–122.
- [Kim and Park, 2013] Kim, Y. T. and Park, H. S. (2013). Geometric structures arising from kernel density estimation on riemannian manifolds. *Journal of Multivariate Analysis*, 114:112 – 126.
- [Lang, 1995] Lang, S. (1995). *Differential and Riemannian manifolds*, volume 160 of *Graduate Texts in Mathematics*. Graduate Texts in Mathematics, Springer-Verlag New York.
- [Lee and Ruymgaart, 1996] Lee, J. M. and Ruymgaart, F. H. (1996). Nonparametric curve estimation on Stiefel manifolds. *J. Nonparametr. Statist.*, 6(1):57–68.
- [Mechab and Laksaci, 2016] Mechab, W. and Laksaci, A. (2016). Nonparametric relative regression for associated random variables. *Metron*, 74(1):75–97.
- [Park, 2004] Park, H. (2004). *Mean distance for Brownian motion on Riemannian manifold and density estimation on Riemannian manifold*. Ph.d. dissertation, Hallym University.
- [Park and Stefanski, 1998] Park, H. and Stefanski, L. A. (1998). Relative-error prediction. *Statist. Probab. Lett.*, 40(3):227–236.
- [Park, 2012] Park, H. S. (2012). Asymptotic behavior of the kernel density estimator from a geometric viewpoint. *Comm. Statist. Theory Methods*, 41(19):3479–3496.
- [Parzen, 1962] Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33:1065–1076.
- [Pelletier, 2005] Pelletier, B. (2005). Kernel density estimation on Riemannian manifolds. *Statistics and Probability Letters*, 3:297–304.
- [Pelletier, 2006] Pelletier, B. (2006). Non-parametric regression estimation on closed Riemannian manifolds. *J. Nonparametr. Stat.*, 18(1):57–67.
- [Prentice, 1989] Prentice, M. J. (1989). Spherical regression on matched pairs of orientation statistics. *J. Roy. Statist. Soc. Ser. B*, 51(2):241–248.
- [Rosenblatt, 1956] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, 27(3):832–837.

- [Samir et al., 2012] Samir, C., Absil, P. A., Srivastava, A., and Klassen, E. (2012). A gradient-descent method for curve fitting on Riemannian manifolds. *Foundations of Computational Mathematics*, 12(1):49–73.
- [Siegel, 1964] Siegel, C. L. (1964). *Symplectic Geometry*. Academic Press, New York.
- [Yanovich, 2016] Yanovich, Y. A. (2016). Asymptotic properties of local sampling on manifold. *Journal of Mathematics and Statistics*, 12(3):157–175.