



# Orientation Relationship in Finite Dimensional Space

Jayant Jha, Atanu Biswas

## ► To cite this version:

Jayant Jha, Atanu Biswas. Orientation Relationship in Finite Dimensional Space. Scandinavian Journal of Statistics, In press, [10.1111/sjos.12479](https://doi.org/10.1111/sjos.12479). [hal-02883709](https://hal.science/hal-02883709)

**HAL Id: hal-02883709**

**<https://hal.science/hal-02883709v1>**

Submitted on 29 Jun 2020

**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.



HAL Authorization

# Orientation Relationship in Finite Dimensional Space

Jayant Jha<sup>1</sup> and Atanu Biswas<sup>2</sup>

<sup>1</sup> INS, Aix-Marseille University

Marseille, France

<sup>2</sup> Applied Statistics Unit, Indian Statistical Institute

Kolkata, India

E-mails: jayantjha@gmail.com and atanu@isical.ac.in

## Abstract

In the present paper, we discuss the regression of a point on the surface of a unit sphere in  $d$  dimensions given a point on the surface of a unit sphere in  $p$  dimensions, where  $p$  may not be equal to  $d$ . Point projection is added to the rotation and linear transformation for regression link function. The identifiability of the model is proved. Then, parameter estimation in this set up is discussed. Simulation studies and data analyses are done to illustrate the model.

**Some keywords and phrases:** Exit distribution; Geodesic distance; Möbius transformation; Projective linear transformation; von Mises-Fisher distribution.

## 1 Introduction

Orientation of a point in  $d$ -dimensional space  $\mathbb{R}^d$  is the unit vector in the direction of that point. Due to the constraint on the norm of such vectors, these points can be taken on the surface of a unit sphere. Such random variables are called spherical random variables. Spherical random variables have wide application in geology (e.g. Chang, 1986), astronomy (e.g. Hall *et al.*, 1987), shape analysis (e.g. Bryner *et al.*, 2012) and crystallography (e.g. Chapman *et al.*, 1995), among others.

Spherical regression is the modelling of orientation,  $y \in \mathbb{S}^{d-1}$ , of a random variable in  $d$ -dimensions, given another spherical random variable  $x \in \mathbb{S}^{p-1}$ , in  $p$ -dimensions. The study of such a regression is mostly available in the statistical literature when both the covariate and response have the same dimension. For the unit sphere in 2-dimensions, circular-circular regression is discussed by Rivest (1997), Downs and Mardia (2002), Minh and Farnum (2003) and Kato *et al.* (2008), among others. For the regression of a spherical random variable on similar dimensional unit sphere, Mackenzie (1957), Downs and Mardia (2002), Downs (2003), Stephens (1979) and Chang (1986) have provided rotational models based on orthogonal transformation. Downs (2003) discussed spherical-spherical regression using stereographic projection on 3-dimensional sphere. Rosenthal *et al.* (2014) used *projective linear transformation* for capturing such a relationship in similar dimensions. These are all parametric models. Di Marzio *et al.* (2014) has used nonparametric regression, where they regressed each coordinate of the response on the coordinates of the covariates. Due to the nature of their regression modelling, the regression can be carried out when the dimensions of covariates and responses are different.

The application of regression modelling on unit sphere in similar dimensions is the vector-cardiogram data discussed in Downs (2003) and Rosenthal *et al.* (2014). In two dimensions, the wind direction prediction is discussed in Kato *et al.* (2008). In Kato *et al.* (2008), the geometry explaining the Möbius Transformation based link function is shown differently for  $|\beta_1| < 1$  and  $|\beta_1| > 1$ . To extend the geometry to the setup of spherical-spherical regression, the geometry in the two cases has to be unified first. By showing that a single geometry can explain both the cases, we extended the model of Kato *et al.* (2008) to unit hypersphere in any dimension when the dimension of the covariate and the response are the same. However, it is also important to study the spherical-spherical regression model with possibly different dimensions of covariates and responses as it can cover all the generalisation for spherical regression in terms of dimensions. Thus, we studied the spherical-spherical regression when the dimensions of the covariate and response are different. As the time of the day and date of the year are circular variables, such a regression can cover the aspects when the prediction of a spherical random variable (such as shapes of clouds) is done based on these covariates.

In this paper, we propose a parametric model which generalises the spherical-spherical regression to all such cases where the dimensions of the covariate and response may differ. We

have considered the geometry of Möbius Transformation based link function and extended the geometry in higher dimension so that it can model the concentration of responses in a part of the sphere. This generalises the Möbius Transformation based circular-circular regression to hyperspheres. We have also shown how to combine the projective linear transformation based spherical-spherical regression model with this geometry. This helps in modelling all such cases where the elongation is axial as well as there is a concentration of responses on a part of the spherical surface. Also, unlike the non-parametric setup of DiMarzio *et al.* (2014), we have considered a parametric regression for such cases which makes the interpretation of parameters easier. Therefore, we have introduced a parametric model which handles rotation, axial elongation as well as concentration of responses on a part of the sphere with different parameters for arbitrary dimensions of spherical variables. We have also provided different summary statistics as well as plots for model diagnostics.

We discuss the model in Section 2. The identifiability of the model is also proved in the same section. We need to study the cases for  $p = d$  and  $p \neq d$  separately. In Section 3, we discuss the properties of the model. Estimation of underlying model parameters is discussed in the same section. Some simulation studies for three cases, namely  $(x \in \mathbb{S}^2, y \in \mathbb{S}^2)$ ,  $(x \in \mathbb{S}^1, y \in \mathbb{S}^2)$  and  $(x \in \mathbb{S}^2, y \in \mathbb{S}^1)$  are reported in Section 4. Section 5 is devoted to data analyses. Section 6 concludes.

## 2 The model

### 2.1 The regression link function

#### 2.1.1 When $p = d$

In the context of circular-circular regression, Kato *et al.* (2008) mentioned that if  $\beta_1$  is a point inside the unit disc, then the conditional mean direction corresponding to covariate  $x$ , excluding the final rotation, is the point on the circle which is on the line joining  $-x$  and  $\beta_1$ . See Kato *et al.* (2008) for details. Extending the same geometry for spherical covariate and spherical response, taking values on the surface of a sphere of similar dimension, the proposed model is explained.

Let  $y$  and  $x$  be the response and covariate on the circle, respectively. Then the regression

function of the model of Kato *et al.* (2008) is

$$y = \beta_0 \frac{x + \beta_1}{1 + \beta_1 x} \epsilon,$$

where  $x, y, \beta_0, \beta_1, \epsilon \in \mathbb{C}$  and  $|x| = |y| = |\beta_0| = |\epsilon| = 1$ . Also, it was assumed that  $\epsilon$  follows a wrapped Cauchy distribution with a real parameter.

In a similar way, we attempt to provide a spherical-spherical regression model on any finite dimensional sphere where the dimensions of both the covariate and response variables are the same. The conditional mean direction  $\mu(I, \beta_1, x)$  corresponding to the covariate  $x$  for a fixed parameter  $\beta_1$ , such that  $\beta_1 \in \mathbb{R}^k$ , is the point on the sphere  $\mathbb{S}^{k-1}$  other than  $-x$  which is at the intersection of the line joining  $-x$  to  $\beta_1$ . The response  $y$  given  $x$  follows a distribution whose parameter is a rotation of  $\mu(I, \beta_1, x)$  which is found by multiplying  $\mu(I, \beta_1, x)$  by a square matrix  $\beta_0 \in SO(k)$  where  $SO(k)$  is the *special orthogonal* group of matrices of dimension  $k \times k$ . The mathematical expression for the proposed model is as follows.

The line joining  $-x$  and  $\beta_1$  is  $a(-x) + (1-a)\beta_1$ , where  $a \in \mathbb{R}$ . Let  $\mu(I, \beta_1, x)$  is the point where this line meets the unit sphere. Then the modulus of this point will be 1. So,

$$\begin{aligned} | -ax + (1-a)\beta_1 |^2 &= 1 \\ \Rightarrow a^2 + (1-a)^2 \beta_1 \cdot \beta_1 - 2a(1-a)\beta_1 \cdot x &= 1 \\ \Rightarrow a^2(1+r^2+2r\cos\gamma) - 2ar(r+\cos\gamma) - (1-r^2) &= 0, \end{aligned}$$

where  $r = (\beta_1 \cdot \beta_1)^{1/2}$  and  $\gamma$  is the angle between  $\beta_1$  and  $x$ . In other words,  $r$  is the Euclidean norm  $\|\cdot\|$  of  $\beta_1$ , and  $\gamma$  is the angle between  $\beta_1$  and the covariate  $x$ .

Clearly, one of the solutions for  $a$  is 1. Also  $-x \in \mathbb{S}^{k-1}$  as  $x \in \mathbb{S}^{k-1}$ . So the other solution of  $a$  will be  $(r^2 - 1)/(1 + r^2 + 2r\cos\gamma)$ . This can be obtained from the property of quadratic equation that the product of the two roots is equal to the ratio of the coefficients of  $r^2$  and  $r^0$  for a quadratic equation in  $r$ . Thus,

$$\begin{aligned} \mu(I, \beta_1, x) &= \frac{1-r^2}{1+r^2+2r\cos\gamma}x + \frac{2+2r\cos\gamma}{1+r^2+2r\cos\gamma}\beta_1 \\ &= \frac{x(1-\beta_1 \cdot \beta_1) + 2\beta_1(1+\beta_1 \cdot x)}{(x+\beta_1) \cdot (x+\beta_1)}. \end{aligned} \quad (2.1)$$

The conditional mean direction of  $y$  will now be a rotation of this point  $\mu(I, \beta_1, x)$  by a fixed matrix parameter  $\beta_0$ , and hence the proposed regression model can be taken as

$$y|x \sim \text{Exit}_k(\rho\mu(\beta_0, \beta_1, x)), \quad (2.2)$$

where  $\text{Exit}_k(\cdot)$  is an Exit distribution in  $k$ -dimensions and  $\rho \in [0, 1]$  is a scalar quantity. The derivation of the conditional mean direction (2.1) is similar to the transformation used in McCullagh (1989, Section 10). However, the motivations for the use of the transformations are different in this paper and McCullagh (1989) because McCullagh (1989) used this transformation to propose a univariate family of distributions on  $(-1, 1)$ . In addition, a different parametrization is used in this paper. In the Exit distribution,  $\rho$  controls the concentration of error. When  $\rho$  is close to 1, the concentration of error is very high while as  $\rho$  gets closer to 0, the error is distributed more uniformly over the unit sphere. Note that Exit distribution with parameter  $\eta$  on  $k$ -dimensional sphere is the distribution of the point from which a particle following Brownian Motion exits a unit ball if it starts from  $\eta$  where  $\|\eta\| \leq 1$  and  $\eta \in \mathbb{R}^k$ . The probability density function (pdf) of  $W$  following Exit distribution  $\text{Exit}_k(\eta)$  is of the form

$$f_W(w) = \frac{1}{A_{k-1}} \frac{1 - \|\eta\|^2}{\|w - \eta\|^k}, \quad w \in \mathbb{S}^{k-1},$$

where  $A_{k-1}$  is the surface area of  $\mathbb{S}^{k-1}$ . See Durrett (1984) and Kato (2009) for details. Here  $\beta_0 \in SO(k)$ . Note that, from (2.1), it follows that  $\mu(\beta_0, \beta_1, x) = \mu(I, \beta_0\beta_1, \beta_0x)$ . An alternative to the Exit distribution can be the von Mises-Fisher distribution; see Fisher *et al.* (1993).

The function in (2.1) can also be used together with orthogonal projection and projective linear transformation by first transforming the points and then multiplying by a square matrix of determinant 1 and then projecting the points back on the unit sphere. Note that the model defined in this paper is equivalent to the orthogonal projection or rotational model and projective linear transformation model mentioned in Rosenthal *et al.* (2014) when  $\beta_1$  is at the origin. Thus, those models can be considered as a subset of this model. The restriction on  $\beta_0 \in SO(k)$  should be relaxed if the model is used in conjunction with projective linear transformation model.

### 2.1.2 When $p \neq d$

Let  $x \in \mathbb{S}^{p-1}$  and  $y \in \mathbb{S}^{d-1}$ . For  $p = d$ , Rosenthal *et al.* (2014) used the *Projective Linear Transformation* as the link function in the regression of  $y$  on  $x$ . Specifically, they modelled the conditional mean direction as  $y = \frac{Ax}{\|Ax\|}$ , where  $A$  is a non-singular square matrix such that determinant of  $A$  is 1. This regression link function first projected the points on the

surface of a sphere to an ellipsoid based on the eigenvalues and eigenvectors of  $A$ , then the points on the sphere were chosen as the unit vector in the direction of the points on the ellipsoid. In the present paper, we have extended this concept when  $p \neq d$  by imposing some restrictions on  $A$ , mentioned in Section 2.3.2. We have also added a projective point  $\beta$  in  $p$ -dimensions similar to the case of  $p = d$  as in Section 2.1.1 to capture the concentration of responses. This will be particularly helpful when the responses are concentrated on some part of the sphere. Our regression link function is thus an extension of the idea proposed in Rosenthal *et al.* (2014) when the dimensions of covariate ( $x$ ) and response ( $y$ ) differ, i.e. when  $p \neq d$ . We further extend the idea by drawing a line from the antipodal point of  $x$  through a fixed projection point in the  $p$ -dimensional space.

The regression link function is  $y = \frac{Ax^*}{\|Ax^*\|}$ , where  $x^* = \frac{x(1-\beta \cdot \beta) + 2\beta(1+\beta \cdot x)}{(x+\beta) \cdot (x+\beta)}$ , as in Section 2.1.1. Here  $A$  is a rectangular matrix with dimension  $d \times p$ ,  $\beta \in \mathbb{R}^p$ . As  $x \in \mathbb{S}^{p-1}$ , so  $x^* \in \mathbb{S}^{p-1}$ .

## 2.2 Geometric properties of the model

### 2.2.1 When $p = d$

This model has the same geometry as of Kato *et al.* (2008) model when  $\|\beta_1\| \leq 1$ . When  $\beta_1$  is a unit vector one gets a standard location model where the mean direction of  $y$  does not depend on  $x$ . However, if  $\beta_1 = 0$ , then  $\mu(I, \beta_1, x) = x$ ; therefore, only rotation takes place during regression.

The geometry for  $\|\beta_1\| > 1$  is explained differently from the case of Kato *et al.* (2008). Here again the predicted mean direction is on the line joining  $-x$  and  $\beta_1$ . Here, also as  $\|\beta_1\|$  goes to 1, the points get highly concentrated around  $\beta_1/\|\beta_1\|$ . Looking at the geometry, it can be directly argued that the regression function is a continuous function of  $x$  and  $\beta_1, \beta_0$ .

Further, for  $\beta_0 = I$ , the angle between the predicted direction  $y$  and  $x$  is  $\cos^{-1}(y \cdot x)$ . Thus, for all the points for which the angle between  $\beta_1$  and  $x$  is the same, the angle of the predicted direction  $y$  and  $\beta_1$  will also be the same. This can easily be seen from (2.1). For three dimensional sphere, the angle between  $\beta_1$  and all the points on a circle with center on the line joining 0 and  $\beta_1$  will be the same, and hence, the angle of the projection with  $\beta_1$  will also be the same. Hence, the conditional mean directions are also on a circle with the center on the line joining 0 and  $\beta_1$ .

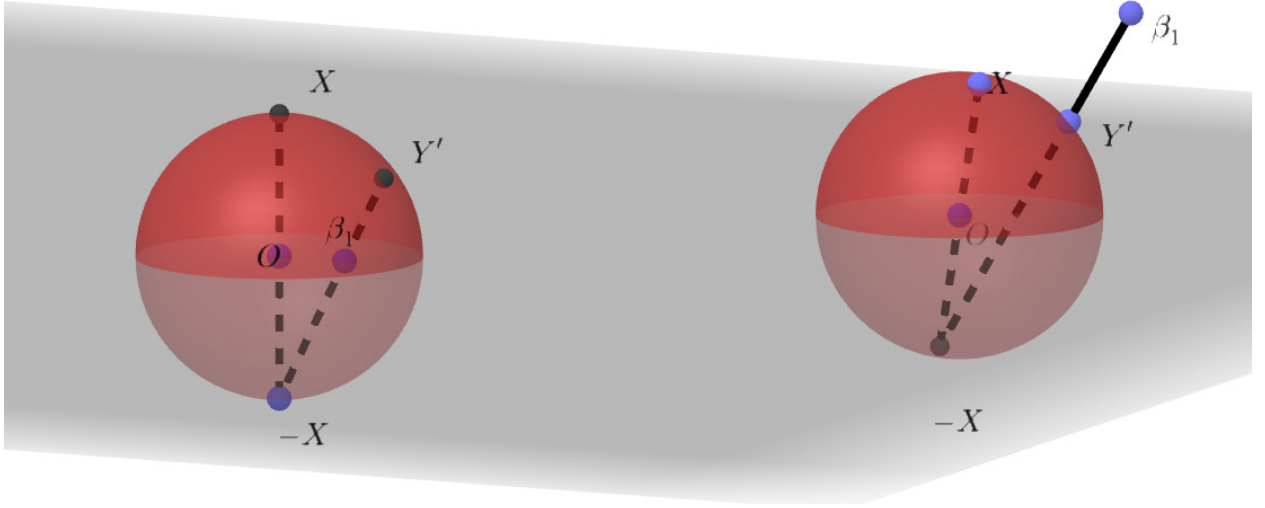


Figure 1: The projection of  $X$  to  $Y'$  via first reflection to  $-X$  and then projection via  $\beta_1$ . (Left:) When  $\beta_1$  is inside the sphere; (Right:) When  $\beta_1$  is outside the sphere.

The geometry of the proposed regression for 3-dimensional case is shown in Figure 1. The left hand figure shows the case when  $\beta_1$  is inside the sphere and the right hand figure illustrates the geometry when  $\beta_1$  is outside the sphere. In the figures,  $\mu(I, \beta_1, X)$  is denoted as  $Y'$ . In both these cases,  $\beta_0$  is taken to be an identity matrix  $I$  of dimension  $3 \times 3$ .

The geometry of regression, when  $\beta_0$  is an axis-angle rotation, is shown in Figure 2. The effect of rotation by  $\beta_0 \in SO(3)$  can be shown first by choosing an axis and then rotating all the points on the unit sphere in 3-dimension around this axis by a fixed angle. In Figure 2, this is shown by denoting the axis by  $\xi$  and the angle by  $\alpha$ . Now, any axis on a 3-dimensional sphere can be determined by taking a point on the sphere and joining the center of the sphere to this point. Thus, the axis of rotation can be identified through two angles  $\theta_1, \theta_2 \in [0, \pi)$ . In the figure, the predicted mean direction when the function in (2.1) is applied to  $X$  is denoted by  $\mu(\beta_0, \beta_1, X) = Y$ . So,  $y$  given  $x$  follows Exit distribution with parameter  $\rho\mu(\beta_0, \beta_1, X)$ .

The change in concentration of the predicted mean directions with varying  $\beta_1$  is shown in Figure 3 for varying  $\|\beta_1\| < 1$ .

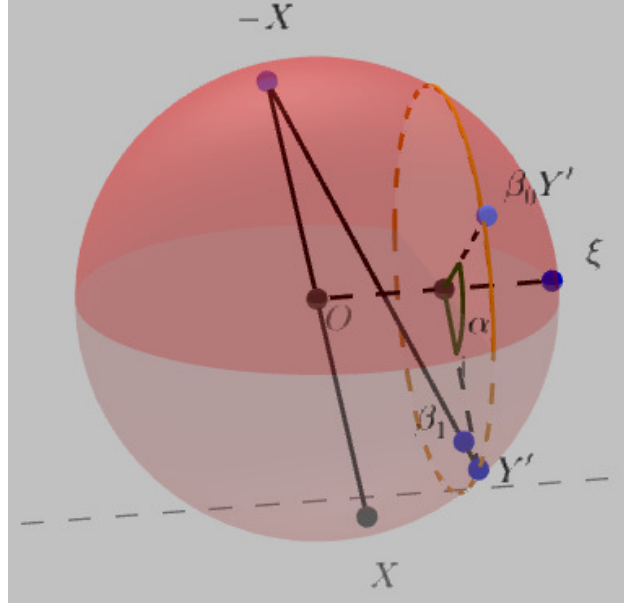


Figure 2: The projection of  $X$  to  $Y'$  via first reflection to  $-X$  and then projection via  $\beta_1$  when  $\beta_1$  is inside the sphere and then axis-angle rotation with unit vector  $\xi$  as axis and  $\alpha$  angle.

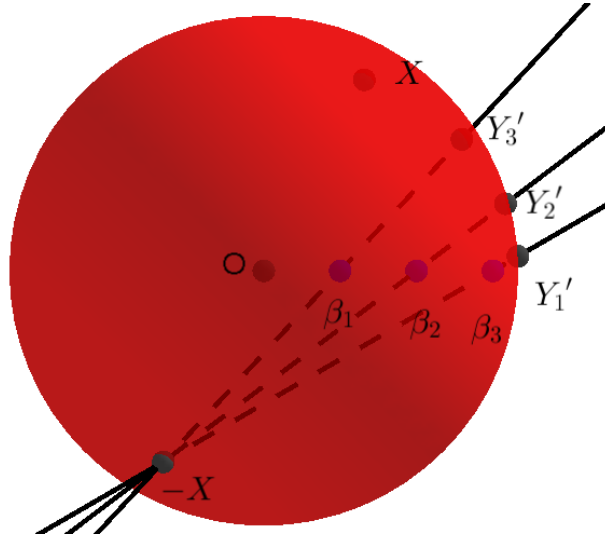


Figure 3: The predicted mean directions  $Y'_1, Y'_2, Y'_3$  for  $X$  corresponding to  $\beta_j = (0.3j, 0, 0)^T$  for  $j = 1, 2, 3$ .

### 2.2.2 When $p \neq d$

There is a four step transformation in the model when  $p \neq d$ . The last three of them are due to the Singular Value Decomposition of  $A$  and the first one is due to the projective point. Let the Singular Value Decomposition of  $A$  is  $A = U\Sigma V^T$ , where  $U$  is an orthogonal matrix of dimension  $d \times d$ ,  $V$  is an orthogonal matrix of dimension  $p \times p$  and  $\Sigma$  is the rectangular diagonal matrix of dimension  $d \times p$  having the singular values of  $A$  at its diagonals. When  $\beta$  is at the origin, regression link function induces the following geometrical transformations to  $x$ . First  $x$  is rotated in a  $p$ -dimensional space by  $V^T$ . Then, the matrix  $\Sigma$  elongates the point in  $d$  dimensions. Then, the points are again rotated in the  $d$ -dimensional space. The division by the norm takes the regressed point on the surface of the sphere.

The addition of the projective point  $\beta$  to this model adds another aspect to the link function. The transformation via projection through  $\beta$  takes the antipodal point of  $x$  and projects it through  $\beta$  by drawing a line. The point of intersection of this line with the unit sphere, denoted by  $x^*$ , is the transformed point. Closer the  $\|\beta\|$  to 1, more is the concentration of predicted points near  $\frac{\beta}{\|\beta\|}$ . This is the geometry of circular-circular regression described in Kato *et al.* (2008) when the projecting point is inside the unit disc. The last three steps follow subsequently as earlier.

The geometry of the regression when  $p = 2$  and  $d = 3$  is shown in Figure 4 and the geometry of the regression when  $p = 3$  and  $d = 2$  is shown in Figure 5. The geometry of the regression using Rosenthal *et al.* (2014) model along with projection point when  $p = d = 3$  can be shown similarly. The four steps in each of the Figures 4 and 5 are drawn counter-clockwise, starting from bottom left.

## 2.3 Identifiability of the Model

### 2.3.1 When $p = d$

**Theorem 2.1.** *The model (2.2) is identifiable when  $p = d$ .*

*Proof:* When  $\beta_0 \in SO(k)$ , the identifiability of the model can be proved in the following way. Suppose there exists two different models for which the predicted mean directions of  $y$  given  $x$  are the same for all  $x$ . Let the model parameters with respect to the first model

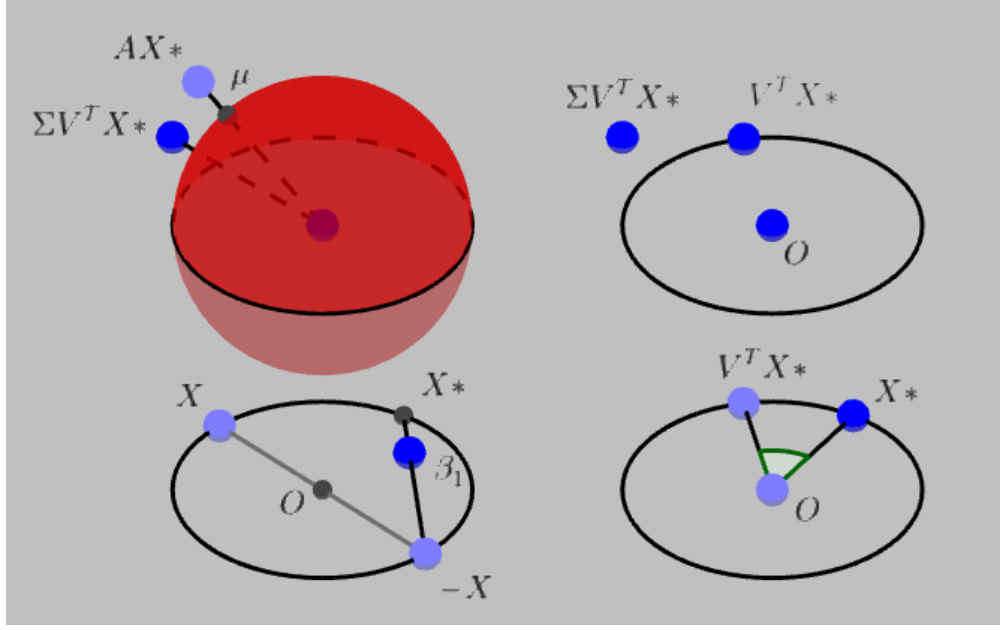


Figure 4: The projection of  $X \in \mathbb{S}^1$  to  $\mu \in \mathbb{S}^2$  counter-clockwise from bottom left. (Bottom Left:) first projection via  $\beta$  when  $\beta$  is inside the circle, (Bottom Right:) Multiplication by  $V^T$ , (Top Right:) Multiplication by  $\Sigma$  and (Top Left:) Final point after multiplication by  $U$  and projecting on the sphere in 3-dimensions.

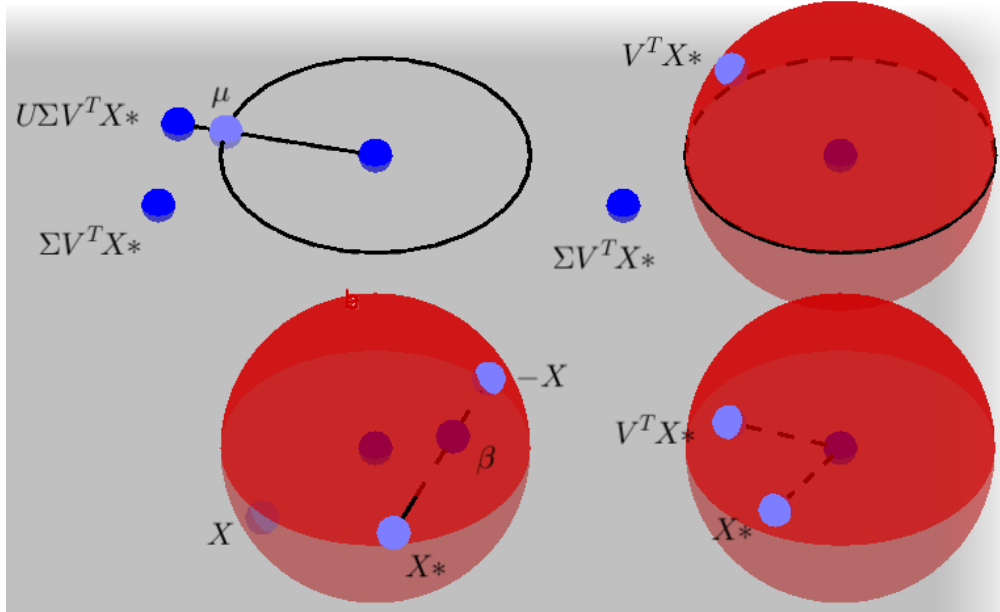


Figure 5: The projection of  $X \in \mathbb{S}^2$  to  $\mu \in \mathbb{S}^1$  counter-clockwise from bottom left. (Bottom Left:) first projection via  $\beta$  when  $\beta$  is inside the sphere, (Bottom Right:) Multiplication by  $V^T$ , (Top Right:) Multiplication by  $\Sigma$  and (Top Left:) Final point after multiplication by  $U$  and projecting on the circle in 2-dimensions.

are  $\beta_0$  and  $\beta_1$  and those corresponding to the second model are  $\beta'_0$  and  $\beta'_1$ . First, suppose  $\beta_0 = \beta'_0 = I$ . Let the predicted mean directions be the same for any  $x$  for both the models, i.e.  $\mu(I, \beta_1, x) = \mu(I, \beta'_1, x)$ . Then,  $-x, \beta_1, \beta'_1$  are collinear. Now, if we take another point  $x'$  on the sphere such that  $-x' \neq \mu(I, \beta_1, x)$ , then  $-x', \beta_1, \beta'_1$  will not be collinear and hence the predicted mean direction point with respect to the models will be different. Hence, the model is identifiable when  $\beta_0 = I$ .

When at least one of  $\beta_0$  and  $\beta'_0$  is not equal to  $I$ , then  $\beta_0\beta'^{-1}_0$  again belongs to  $SO(k)$ . Transforming the coordinate system by choosing the eigenvectors as the principal axes such that  $\beta_0\beta'^{-1}_0$  is an identity matrix, the result follows directly from the case when  $\beta_0 = \beta'_0 = I$ .

Alternatively, suppose  $\mu(\beta_0(1), \beta_1, x) = \mu(\beta'_0, \beta'_1, x)$  for all  $x$ , where  $\beta_0(1) \in SO(k)$ . Then, without loss of generality, this is equivalent to  $\mu(I, \beta_1, x) = \mu(I, \beta_0\beta'_1, \beta_0x)$  for all  $x$  with  $\beta_0 = \beta_0(1)^{-1}\beta'_0$ . Now,  $\mu(I, \beta_1, \beta_1/\|\beta_1\|) = \beta_1/\|\beta_1\|$ . Hence,  $\beta_0\beta'_1$  should lie on the line joining  $\beta_1/\|\beta_1\|$  and  $-\beta_0\beta_1/\|\beta_1\|$ . Also,  $\mu(I, \beta_1, -\beta_1/\|\beta_1\|) = -\beta_1/\|\beta_1\|$ . Thus,  $\beta_0\beta'_1$  should lie on the line joining  $-\beta_1/\|\beta_1\|$  and  $\beta_0\beta_1/\|\beta_1\|$ . But, these two lines are parallel. Hence,  $\beta'_1 = \beta_1$  and  $\beta_0 = I$ .

When our model is used along with the Projective Linear Transformation Model, then  $y$  given  $x$  follows Exit Distribution with parameter  $\frac{\rho A \mu(I, \beta_1, x)}{\|A \mu(I, \beta_1, x)\|}$ . The identifiability of the model can be proved in the following way. Let

$$\frac{Ay'_{(2)}}{\|Ay'_{(2)}\|} = \frac{By'_{(1)}}{\|By'_{(1)}\|}$$

for all  $x \in \mathbb{S}^{k-1}$ . Here  $y'_{(2)}$  is the transformed point using  $\beta_{(2)1}$  as the parameter in the link function and  $y'_{(1)}$  corresponds to  $\beta_{(1)1}$  as the parameter, and  $A, B \in PGL(k)$ , the projective general linear group as defined in Rosenthal *et al.* (2014). Then, for all  $x$ ,

$$\frac{\|By'_{(1)}\|}{\|Ay'_{(2)}\|} B^{-1} Ay'_{(2)} = y'_{(1)}. \quad (2.3)$$

Taking norm on both sides of (2.3), we get

$$\frac{\|By'_{(1)}\|}{\|Ay'_{(2)}\|} \|B^{-1} Ay'_{(2)}\| = 1. \quad (2.4)$$

Using (2.4) in (2.3), we get

$$\frac{B^{-1} Ay'_{(2)}}{\|B^{-1} Ay'_{(2)}\|} = y'_{(1)}.$$

Now,  $A, B \in PGL(k)$  implies  $B^{-1}A \in PGL(k)$ . Writing  $C = B^{-1}A$ , it reduces to

$$\frac{Cy'_{(2)}}{\|Cy'_{(2)}\|} = y'_{(1)}.$$

Let  $C = PD_\lambda P'$ , where  $P$  is an orthogonal matrix of eigenvectors of  $C$  and  $D_\lambda$  is a diagonal matrix of eigenvalues of  $C$ . Then, transforming the coordinates, without loss of generality, we can take  $C = D_\lambda$ . Thus,

$$\frac{D_\lambda y'_{(2)}}{\|D_\lambda y'_{(2)}\|} = y'_{(1)}.$$

When  $x = \frac{\beta_{(2)1}}{\|\beta_{(2)1}\|}$ ,

$$\frac{D_\lambda \beta_{(2)1}}{\|D_\lambda \beta_{(2)1}\|} = \frac{\frac{\beta_{(2)1}}{\|\beta_{(2)1}\|}(1 - \beta_{(1)1} \cdot \beta_{(1)1}) + 2\beta_{(1)1}(1 + \frac{\beta_{(2)1} \cdot \beta_{(1)1}}{\|\beta_{(2)1}\|})}{(\frac{\beta_{(2)1}}{\|\beta_{(2)1}\|} + \beta_{(1)1}) \cdot (\frac{\beta_{(2)1}}{\|\beta_{(2)1}\|} + \beta_{(1)1})}. \quad (2.5)$$

When  $x = \frac{-\beta_{(2)1}}{\|\beta_{(2)1}\|}$ ,

$$\frac{-D_\lambda \beta_{(2)1}}{\|D_\lambda \beta_{(2)1}\|} = \frac{\frac{-\beta_{(2)1}}{\|\beta_{(2)1}\|}(1 - \beta_{(1)1} \cdot \beta_{(1)1}) + 2\beta_{(1)1}(1 - \frac{\beta_{(2)1} \cdot \beta_{(1)1}}{\|\beta_{(2)1}\|})}{(\frac{-\beta_{(2)1}}{\|\beta_{(2)1}\|} + \beta_{(1)1}) \cdot (\frac{-\beta_{(2)1}}{\|\beta_{(2)1}\|} + \beta_{(1)1})}. \quad (2.6)$$

Taking negative of (2.6) and equating its right hand side to the right hand side of (2.5), we get  $\beta_{(2)1} = m\beta_{(1)1}$  for some scalar  $m$ .

If  $m > 0$ , when  $x_b = \beta_{(2)1}/\|\beta_{(2)1}\|$ , we have

$$\frac{D_\lambda x_b}{\|D_\lambda x_b\|} = x_b. \quad (2.7)$$

Equating each of the coordinates of both sides in (2.7) and using the condition that the determinant of  $D_\lambda$  is 1, we get  $D_\lambda = I$ . Similarly,  $D_\lambda = I$  can be proved for  $m < 0$ . The identifiability of the model follows by the identifiability in the case of  $\beta_0 \in SO(k)$ .  $\square$

### 2.3.2 When $p \neq d$

We show the identifiability of the model by imposing some conditions on the matrix  $A$  and then, proving a theorem.

**Theorem 2.2.** *Under the conditions C1, C2 and C3 on  $A$  mentioned below, the model (2.2) is identifiable when  $p \neq d$ .*

The conditions on  $A$  are the following:

C1.  $\text{rank}(A) = k = \min(p, d)$ .

C2. The singular values of  $A$  are such that  $\lambda_1 > \lambda_2 > \dots > \lambda_k > 0$ .

C3.  $\prod_{i=1}^k \lambda_i = 1$ .

We shall denote such a set of  $d \times p$  matrices by  $\mathcal{G}$ . Now we have the following Theorem.

*Proof:* Using these conditions when  $\beta$ , the projective point, is at the origin, the identifiability can be proved very easily. As the singular values of the matrix  $A$  are non-degenerate and non-zero, the Singular Value Decomposition of  $A$  will be unique. By condition C3, we impose a restriction that among all the matrices  $B$  with  $B = cA$ ,  $c$  being a scalar, the one representative is chosen for which the multiplication of singular values is 1. Hence, when  $\beta = (0, 0, \dots, 0)$ , the model is identifiable.

The identifiability of the model when  $\beta$  is not at the origin and  $\|\beta\| \neq 1$  can be proved in the following way. Let  $A, B \in \mathcal{G}$  be the multiplicative matrices and  $\beta_1, \beta_2 \in \mathbb{R}^p$  are the projecting points respectively for two models such that

$$\frac{Ax_{(1)}}{\|Ax_{(1)}\|} = \frac{Bx_{(2)}}{\|Bx_{(2)}\|}, \quad (2.8)$$

where  $x_{(1)}, x_{(2)} \in \mathbb{S}^{p-1}$  be the projected points corresponding to  $x$  with respect to  $\beta_1$  and  $\beta_2$ . Let  $x = \frac{\beta_1}{\|\beta_1\|}$ . Then,

$$\frac{A\beta_1}{\|A\beta_1\|} = B \frac{\frac{\beta_1}{\|\beta_1\|}(1 - \beta_2 \cdot \beta_2) + 2\beta_2(1 + \frac{\beta_1 \cdot \beta_1}{\|\beta_1\|})}{\|B(\frac{\beta_1}{\|\beta_1\|}(1 - \beta_2 \cdot \beta_2) + 2\beta_2(1 + \frac{\beta_1 \cdot \beta_1}{\|\beta_1\|}))\|}. \quad (2.9)$$

Also, for  $x = \frac{-\beta_1}{\|\beta_1\|}$ ,

$$\frac{-A\beta_1}{\|A\beta_1\|} = B \frac{\frac{-\beta_1}{\|\beta_1\|}(1 - \beta_2 \cdot \beta_2) + 2\beta_2(1 + \frac{-\beta_1 \cdot \beta_1}{\|\beta_1\|})}{\|B(\frac{-\beta_1}{\|\beta_1\|}(1 - \beta_2 \cdot \beta_2) + 2\beta_2(1 + \frac{-\beta_1 \cdot \beta_1}{\|\beta_1\|}))\|}. \quad (2.10)$$

Solving (2.9) and (2.10), we get  $\beta_2 = m\beta_1$ , where  $m$  is scalar.

Without loss of generality, take  $\beta_1 = (b, 0, 0, \dots, 0)$  and  $\beta_2 = (mb, 0, \dots, 0)$ , and suppose that  $x = (x_1, x_2, \dots, x_p)$ . Then,

$$Ax_{(1)} = A \frac{x(1 - b^2) + 2\beta_1(1 + bx_1)}{(x + \beta_1) \cdot (x + \beta_1)}, \quad (2.11)$$

and

$$Bx_{(2)} = B \frac{x(1 - m^2b^2) + 2m\beta_1(1 + mbx_1)}{(x + m\beta_1) \cdot (x + m\beta_1)}. \quad (2.12)$$

Let  $A = \langle (a_{ij}) \rangle$  and  $B = \langle (b_{ij}) \rangle$ . The equation (2.8) reduces to

$$[(1 - b^2)A - (1 - m^2b^2)tB]x = ((2mt + 2m^2tbx_1)B - (2bx_1 + 2)A)\beta_1 \quad (2.13)$$

for all  $x \in \mathbb{S}^{p-1}$ . Here  $t \in \mathbb{R}$ . When  $x = (1, 0, \dots, 0)$ , we get

$$\frac{b_{11}}{a_{11}} = \frac{b_{21}}{a_{21}} = \dots = \frac{b_{d1}}{a_{d1}} = c(\text{say}).$$

Taking  $x = (0, -1, 0, 0, \dots, 0)$  and  $x = (0, 1, 0, 0, \dots, 0)$  in (2.13) and summing the two equations we get,

$$\frac{b_{22}}{b_{21}} = \frac{b_{32}}{b_{31}} = \dots = \frac{b_{d2}}{b_{d1}}.$$

In a similar fashion, if we go on taking  $x$  as  $(0, 0, \dots, 1, 0, \dots, 0)$  and  $(0, 0, \dots, -1, 0, \dots, 0)$  and summing, we get the rank of  $B=1$  which violates C1 or  $m=1$  which again implies  $A = qB$  for some scalar  $q$ . By the condition C3 on  $\mathcal{G}$ , we get  $q = 1$ , and thus the model is identifiable.  $\square$

Note that from the theorems, the model is identifiable whenever the response  $y$ , given the covariate  $x$ , follows any distribution  $f(y; \mu_{y.x}, \rho)$  for which there exists an  $x : f(y; \mu_{y.x}, \rho_1) \neq f(y; \mu'_{y.x}, \rho_2)$  for any  $x$  if  $\mu_{y.x} \neq \mu'_{y.x}$  and  $\rho_1 \neq \rho_2$ .

## 3 Distributional properties

### 3.1 Estimation of parameters

#### 3.1.1 When $p = d$

The estimation of parameters of the model can be done through the maximum likelihood estimation procedure. Method of moments can also be used, where the trigonometric moments can be compared with the sample values and these equations can be used to find the estimates of the parameters. In the simulation study and data analysis, we have used a gradient-descent method called BFGS method under `optim` function in R, which is used to find the maximum or minimum of a function, to get the maximum likelihood estimators (MLEs) under Exit distribution setup. In order to ensure that the global maximum is obtained, it is preferable to start from several different initial values.

When  $p = d$ ,  $x$  are known and deterministic, and  $y/x$ , follows von Mises-Fisher (VMF) distribution with concentration parameter  $\kappa$  the MLE of the parameters under the rigid rotation model and asymptotic properties of the estimators are given in Chang (1986, 1989) and Rivest (1989). Thus, for a given value of  $x$ , when  $y \sim VMF(Ax, \kappa)$ , where  $A$  is a rotation matrix, the MLE of  $A$  is  $U_1 U_2^\top$ , where  $U_1 \Phi U_2^\top$  is the singular value decomposition of  $n^{-1} \sum_{i=1}^n y_i x_i^\top$ .

If our model is used together with the rotational model, then given  $\beta_1$ , the MLE of  $\beta_0$ , which is a rotational matrix, can be found in the following way. Given  $\beta_1$ , let the Singular Value Decomposition of  $n^{-1} \sum_{i=1}^n y_i \mu(I, \beta_1, x_i)^\top$  is  $V_1 \phi V_2^\top(\beta_1)$ . Then, the MLE of  $\beta_0$  is  $V_1 V_2^\top$ . Thus, the function has to be maximized over the space of  $\beta_1$  and given the estimator,  $\hat{\beta}_1$ , the MLE of  $\beta_0$  can be obtained.

### 3.1.2 When $p \neq d$

The maximum likelihood estimation of the parameters when  $p = d$ , and  $\beta$  is at the origin and  $A \in PLG(d)$ , and the error following von Mises-Fisher distribution, is discussed in Rosenthal *et al.* (2014), where  $PLG(d)$  is the Projective General Linear Group in  $d$ -dimensions. Extending the same approach, the maximum likelihood estimation of model parameters when  $p \neq d$  and  $\beta$  is at the origin and the error follows von Mises-Fisher distribution can be obtained in the following way.

Step 1: Initialize the algorithm using  $A = \sum_{i=1}^n y_i x_i^\top$ .

Step 2: Compute  $\nabla_A f_n$ .

Step 3: Update  $A$  by using  $A \rightarrow A + \epsilon \nabla_A f_n$ , where  $\epsilon$  is very small.

Step 4: Stop, if convergence reached. Else, go to Step 2.

Note that  $\nabla_A f_n$  can also be computed in a similar manner as in Michael *et al.* (2014). Using the same computations,

$$\nabla_A f_n = \frac{1}{n} \sum_{i=1}^n \left( \frac{y_i x_i^\top}{\|Ax_i\|} - \frac{Ax_i x_i^\top A^\top y_i x_i^\top}{\|Ax_i\|^3} \right).$$

Now, given  $\hat{A}$ , the MLE of  $\kappa$ , denoted by  $\hat{\kappa}$ , can be obtained immediately by solving the following equation:

$$\frac{I_{d/2}(\kappa)}{I_{d/2-1}(\kappa)} = -\frac{1}{n} \sum_{i=1}^n \left( \frac{y_i^\top \hat{A} x_i}{\|\hat{A} x_i^\top y_i\|} \right),$$

where  $I_\nu(\cdot)$  is the incomplete Bessel function of order  $\nu$ .

When  $\beta$  is not at the origin, the MLE in simulation and data analysis can be found by using `optim` function in R. Also, in case of Exit distribution for error, the MLE can again be obtained by using `optim` function in R as has been done in simulation and data analyses in Sections 4 and 5.

### 3.2 Test of independence

First consider  $p = d$ . If  $x$  follows  $\text{Exit}_k(\eta)$ , and  $y|x$  follows  $\text{Exit}_k(\rho\mu_{y.x})$  where  $\mu_{y.x} = \mu(\beta_0, \beta_1, x)$ , then the joint density  $f(x, y)$  can be denoted as

$$f(x, y) = f(x)f(y|x) = \frac{1}{A_{k-1}^2} \frac{1 - \|\eta\|^2}{\|x - \eta\|^k} \frac{1 - \|\xi\|^2}{\|y - \xi\|^k},$$

where  $\xi = \rho\mu_{y.x}$ . From the joint distribution, the test for independence of  $y$  and  $x$  can be performed as Kato *et al.* (2008) has done for two circular random variables. If  $\|\beta_1\| = 1$ , then  $\mu_{y.x} = \beta_0\beta_1$ , and hence  $f(y|x) \equiv \text{Exit}_k(\rho\beta_0\beta_1)$ , and thus  $y$  is independent of  $x$ . Hence, the test of independence for  $y$  and  $x$  can be performed in the following way. Defining the null hypothesis  $H_0 : \|\beta_1\| = 1$  and the alternate hypothesis  $H_1 : \|\beta_1\| \neq 1$ , the test rejects the null hypothesis of independence if  $\|\hat{\beta}_1\|$  is far from 1.

Formally, the test statistic

$$\Lambda = -2 \ln \left( \frac{L_0}{L_1} \right)$$

asymptotically follows the chi-squared distribution with  $(k^2 - k + 2)/2$  degrees of freedom, where  $L_0$  is the maximum likelihood under the null hypothesis (i.e when  $\beta_1 = (1, 0, \dots, 0)^\top$ ) and  $L_1$  is the maximum likelihood under  $H_0 \cup H_1$ . Under the null hypothesis, the number of unknown parameters are  $k$  (the angular coordinates of  $\beta_1$  and  $\rho$ ), while under  $H_0 \cup H_1$ , the number of unknown parameters are  $(k^2 + k + 2)/2$  ( $k$  corresponding to  $\beta_1$ ,  $(k^2 - k)/2$  corresponding to  $\beta_0$  and 1 corresponding to  $\rho$ ). Thus, the degrees of freedom of the asymptotic test statistic is  $(k^2 - k + 2)/2$ .

When  $p \neq d$ , the same test of independence can be carried out. Under the null hypothesis, the number of parameters to be estimated are  $d - 1$  corresponding to the mean point and 1 corresponding to  $\rho$ , while under  $H_0 \cup H_1$ , the number of unknown parameters are  $\frac{p^2 - p}{2}$  (corresponding to  $V$ ),  $\frac{d^2 - d}{2}$  (corresponding to  $U$ ),  $\min(p, d) - 1$  (corresponding to  $\Sigma$ ),  $p$

(corresponding to  $\beta_1$ ) and 1 corresponding to  $\rho$ . Hence, the degrees of freedom is  $f = \frac{p^2+p}{2} + \frac{d^2-d}{2} + \min(p, d) - d$ .

### 3.3 Goodness of fit

#### 3.3.1 Geodesic distance as a summary measure

The geodesic distances between the predicted points and the observed points can also be used as a measure to gauge the goodness of fit of the model. Let corresponding to the observed point  $y_{O_i}$  (denoted by  $O_i$ ), the predicted point is  $y_{P_i}$  (denoted by  $P_i$ ), where  $y_{P_i}, y_{O_i} \in \mathbb{S}^{k-1}$ . Suppose the Euclidean distance between these points be  $d_i$ , and the angle subtended by the arc  $O_iP_i$  at the centre of the unit sphere be  $\zeta_i$ . Then, the length of the arc  $O_iP_i$ , which is the geodesic distance between the points, will be the angle between the points, in radian. Using sine rule for the triangle  $O_iOP_i$ , we immediately get

$$d_i \equiv \begin{cases} \frac{\sin \zeta_i}{\cos \frac{\zeta_i}{2}} = 2 \sin \frac{\zeta_i}{2} & \text{if } \zeta_i \neq \pi \\ 2, & \text{if } \zeta_i = \pi, \end{cases}$$

Then, the goodness of fit can be measured by

$$D = \frac{1}{n} \sum_{i=1}^n \zeta_i = \frac{2}{n} \sum_{i=1}^n \sin^{-1} \frac{d_i}{2}. \quad (3.1)$$

Thus,  $D \in [0, \pi]$  and smaller values of  $D$  imply better fit and if  $D = 0$ , it is a perfect fit.

#### 3.3.2 Checking assumptions about link function and error distribution

If a random variable  $X \in S^{d-1}$  follows exit distribution with parameter  $\eta = (\eta_1, \dots, \eta_d)$ , then for any harmonic function  $g(X)$ ,  $E\{g(X)\} = g(\eta)$  (see Durrett (1984), pp. 36). This property can be used to check the assumption of the link function in the model. The property implies that  $E(X_j) = \eta_j$  where  $X = (X_1, X_2, \dots, X_d)$ .

Also, for a spherical random variable  $E(\sum_{j=1}^d X_j^2) = 1$ . If we consider  $g(X) = X_i^2 - X_k^2$ , which is a harmonic function, then  $E(X_i^2 - X_k^2) = \eta_i^2 - \eta_k^2$ . Therefore if we add the equations:

$$E\left(\sum_{j=1}^d X_j^2\right) = 1$$

and  $E(X_i^2 - X_k^2) = \eta_i^2 - \eta_k^2$  for all  $k \neq i$ , we get  $E(X_i^2) = \frac{1+d\eta_i^2-\|\eta\|^2}{d}$ . Therefore, we can find the variance of  $X_i$  and denote it by  $\sigma_i^2$ .

Denoting the predicted and observed  $i$ th response by  $(y_{p,i1}, \dots, y_{p,id})$  and  $(y_{o,i1}, \dots, y_{o,id})$  respectively, we may check the goodness of fit. For checking the goodness of fit marginally in  $k$ th dimension, we plot  $\Delta_{ik} = \frac{y_{p,ik} - \rho y_{o,ik}}{\sigma_{ik}}$  for  $i = 1, \dots, n$ . We use the estimates of the parameters when the parameters are unknown. We also calculate  $c_k = \frac{\sum_{i=1}^n \Delta_{ik}}{\sqrt{n}}$  as summary statistic to check if marginally, the points are not far from normal distribution. This can be done for each of the  $d$  dimensions. If there is some pattern in error, then in the plots, we may find many points far away from 0 and the summary statistics  $c_k$ 's having large values.

After checking the assumption for the link function, we draw the histogram for  $1 - y_p^\top y_o$ . If the error distribution follows exit distribution, then the frequencies in histogram should be monotonically decreasing.

### 3.4 Comparison with existing models

The problem with the rigid rotation model is that every point is transformed similarly. The resultant fit may be good for some data sets; but, for many data sets, e.g. for the data used in shape and image analyses, the transformation of different covariates are different. The Projective Linear Transformation model tries to capture this by first multiplying the covariates with a square matrix whose determinant is 1 and then projecting the points back on the sphere in the direction of the line joining the center of the unit sphere to this point. On multiplication by the matrix, it can be easily observed that the points on the sphere get elongated in a direction which depends on the eigenvalues and eigenvectors of the multiplicative matrix and then the back-projection on the sphere enables the model to capture different transformations of different covariates. The rigid-rotation model is a special case of the Projective Linear Transformation Model when all the eigenvalues of the multiplicative square matrix  $\beta_0$  are 1. Thus, the points on the unit sphere are rotated uniformly and there is no elongation in any particular direction. In the Projective Linear Transformation Model, the absolute values of all the eigenvalues may be different and hence, the elongation happens in the direction of the eigenvector corresponding to the largest eigenvalue, and hence more points get concentrated along the axis of this eigenvector. But the drawback of this model is that this concentration is the same in the direction of the eigenvector as well as in the negative direction of this eigenvector. Hence, this model is unable to capture the regression when there is a concentration of responses around a particular point on the sphere or only

in one part of the unit sphere. As explained in Figure 3, our model can capture this type of responses as well by differing the parameter  $\beta_1$ . Our model can also be used together with rigid rotation as well as Projective Linear Transformation as explained in Section 2.

Downs (2003) used Möbius transformation, but the idea of projecting points on a plane from the surface of a sphere induces a discontinuity at the north pole. Due to continuity of the geometry, our model does not exhibit this problem. Moreover, the model of Downs (2003) has not been extended for dimensions greater than 3.

Note that when  $p \neq d$ , then there is no existing parametric model in the literature. Hence, the comparison is only restricted to the non-parametric models of Di Marzio *et al.* (2014).

## 4 Simulation studies

### 4.1 When $p = d$

Data from our proposed regression model for different given values of  $\beta_1 \in \mathbb{R}^3$ ,  $n$  (sample size) and  $\rho$  for a fixed  $\beta_0 \in SO(3)$  were simulated 1000 times. The MLEs of the parameters were obtained for our proposed model and the rotational model for all the cases.

Exit distribution was used to model the error in response. The data from Exit distribution were simulated using the method proposed by Müller (1956). Covariates were taken uniformly over the unit sphere in 3-dimensions. The simulation algorithm is as follows:

Step 1:  $x_1, x_2, x_3$  are independently generated from  $N(0,1)$ .

Step 2: The covariate taken uniformly on the unit sphere will be  $\frac{1}{\sqrt{x_1^2 + x_2^2 + x_3^2}}(x_1, x_2, x_3)$ .

Step 3:  $u_1, u_2, u_3$  are independently generated from  $N(0,1)$ .

Step 4:  $\rho\mu_{y.x} \rightarrow \rho\mu_{y.x} + (1 - \rho)\frac{1}{\sqrt{u_1^2 + u_2^2 + u_3^2}}(u_1, u_2, u_3)$ .

Step 5:  $\rho \rightarrow \|\rho\mu_{y.x} + (1 - \rho)\frac{1}{\sqrt{u_1^2 + u_2^2 + u_3^2}}(u_1, u_2, u_3)\|$ .

Step 6: if  $\rho > 0.99$ , stop else go to Step 3.

Since our model is nothing but a rotational model when  $\beta_1$  is at the origin, the comparison between both the models can be done through  $\chi^2$ -statistic for testing  $H_0 : \beta_1 = (0, 0, 0)^\top$  against  $H_1 : \beta_1 \neq (0, 0, 0)^\top$ . Under the null hypothesis, the statistic  $\Lambda = -2 \ln \left( \frac{L_0}{L_1} \right)$ , where  $L_0$  is the maximum likelihood under null hypothesis and  $L_1$  is the maximum likelihood under

$H_0 \cup H_1$ , asymptotically follows a  $\chi^2$  distribution with  $k$  degrees of freedom when  $x, y \in \mathbb{S}^{k-1}$ . In the Tables 1 and 2,  $(b_1, b_2, b_3)$  correspond to the coordinates of  $\beta_1$ , and  $\rho$  is the parameter of Exit distribution. Also  $(\theta_1, \theta_2)$  denote the axes of rotation, and  $\alpha$  denotes the angle of rotation. The rotation of  $x$  in terms of axis and angle can be written by the Rodrigues' formula (cf. Rodrigues, 1840). Let  $e = (\sin \theta_2 \cos \theta_1, \sin \theta_2 \sin \theta_1, \cos \theta_1)$ . Then, for any point  $x \in \mathbb{S}^2$ , the rotated point is given by  $x_{rot} = (\cos \alpha)x + (\sin \alpha)(e.x) + (1 - \cos \alpha)(e.x)e$ . As  $\theta_1$  and  $\theta_2$  are axial parameters,  $2\theta_1$  and  $2\theta_2$  are circular parameters. Hence, in the simulation tables, the circular means of the MLEs of  $2\theta_1$  and  $2\theta_2$  over 1000 simulations are reported. The estimates corresponding to only rotation and no projection are denoted by NPE (No Projection Estimates) while the estimates corresponding to our model are denoted by PE (Projection Estimates). The standard deviations (sd's) of the simulated estimates for the linear random variables are obtained, and these are used to obtain the standard errors (se's) of the estimates which are given in parentheses along with the means of the estimates. Circular variances of the simulated estimates are obtained for the circular random variables. The circular variance (or dispersion) is defined as  $(1 - \bar{R})$ , where  $\bar{R}$  is the mean resultant length of the estimated values of the circular parameter from all the simulations. In the Tables, the standard errors of the linear parameters and circular variances of the circular parameters are reported in the parentheses.

The comparison of our model with the rotational model is illustrated in Table 1 for three sample sizes  $n = 25, 50, 100$ . In Table 2, the MLEs are reported for different values of  $\beta_1$  for  $n = 25$ . The mean values of the statistic  $\Lambda$  along with the  $p$ -values are also reported in each of the Tables. For the cross-validation predictive performance, for each sample size  $n$  of learning data set, a test data set of size 40% of  $n$  is considered. The mean values of  $D$ , a measure of goodness of fit defined in (3.1) in Section 3.3, is also reported in the Tables.

From Table 1, we observe that as  $n$  increases, the standard deviation and circular variance decreases and the estimates get closer to the true values. In Table 2,  $\beta_1 = (0, 0, 0)^\top$  correspond to the no rotation model. It can be seen from the values of  $\Lambda$  and its  $p$ -values under asymptotic distribution that as  $\beta_1$  goes far from origin, the value of  $\Lambda$  increases.

Our empirical study based on the QQ plots (not reported here for the sake of brevity) shows that the statistic  $\Lambda$  can be approximated by a  $\chi_3^2$  distribution for a sample size as low as 25 while the approximation gets closer to the true distribution as the sample size

Table 1: Estimates of parameters (with standard errors/circular variances in parentheses) for  $\rho = 0.85$ ,  $2\theta_1 = 6$ ,  $2\theta_2 = 4$ ,  $\alpha = 4$  and  $\beta_1 = (0.3, 0, 0)$  for different  $n$ .

	$n = 25$		$n = 50$		$n = 100$	
Para- meters	NPE	PE	NPE	PE	NPE	PE
$2\theta_1$	5.976 (0.027)	6.013 (0.008)	5.964 (0.011)	5.998 (0.003)	5.987 (0.006)	5.999 (0.001)
$2\theta_2$	4.191 (0.573)	4.016 (0.201)	4.019 (0.404)	4.000 (0.047)	3.915 (0.251)	3.989 (0.023)
$\alpha$	3.888 (0.122)	3.941 (0.059)	3.964 (0.040)	3.998 (0.003)	3.988 (0.023)	4.001 (0.001)
$\rho$	0.685 (0.002)	0.862 (0.002)	0.672 (0.002)	0.861 (0.001)	0.653 (0.001)	0.859 (0.000)
$b_1$		0.298 (0.001)		0.300 (0.001)		0.300 (0.000)
$b_2$		0.002 (0.001)		0.001 (0.001)		0.001 (0.000)
$b_3$		0.000 (0.001)		0.001 (0.001)		0.000 (0.000)
$D$	0.273	0.167	0.263	0.165	0.260	0.164
$\Lambda$ ( $p$ -value)	56.3257 (0)		111.064 (0)		223.235 (0)	

Table 2: Estimates of parameters (with standard errors/circular variances in parentheses) for  $\rho = 0.85$ ,  $2\theta_1 = 6$ ,  $2\theta_2 = 4$ ,  $\alpha = 4$  and  $n = 25$  for different  $\beta_1$ .

	$\beta_1 = (0, 0, 0)$		$\beta_1 = (0.3, 0, 0)$		$\beta_1 = (0.6, 0, 0)$	
Para- meters	NPE	PE	NPE	PE	NPE	PE
$2\theta_1$	5.998 (0.004)	5.997 (0.008)	5.976 (0.027)	6.013 (0.008)	5.955 (0.154)	6.032 (0.031)
$2\theta_2$	4.028 (0.113)	4.026 (0.141)	4.191 (0.573)	4.016 (0.201)	5.359 (0.607)	4.082 (0.374)
$\alpha$	3.979 (0.017)	3.981 (0.023)	3.888 (0.122)	3.941 (0.059)	3.505 (0.347)	3.857 (0.131)
$\rho$	0.858 (0.001)	0.869 (0.002)	0.685 (0.002)	0.862 (0.001)	0.461 (0.002)	0.862 (0.001)
$b_1$		0.004 (0.002)		0.298 (0.001)		0.596 (0.002)
$b_2$		0.000 (0.001)		0.002 (0.001)		0.005 (0.003)
$b_3$		0.000 (0.002)		0.000 (0.001)		0.004 (0.003)
$D$	0.164	0.169	0.273	0.167	0.445	0.176
$\Lambda$ ( $p$ -value)	3.2163 (0.3595)		56.3257 (0)		93.47231 (0)	

increases.

## 4.2 When $p \neq d$

There is no parametric model proposed in the statistical literature for handling the case of  $p \neq d$ . Thus, our comparison is only restricted to the non-parametric regression model proposed by Di Marzio *et al.* (2014).

For the comparison, the data set is again divided into two parts. One is the training data set and another is the test data set. The number of observations under the test data set is restricted to 40% of the training data set ( $n$ ) and the comparison is made at different values of  $n, \rho, \lambda, ||\beta||$  where  $\lambda$  denotes the bigger of the two eigenvalues of the matrix  $\beta_0$ .

The comparison is shown for  $p = 2, d = 3$  and also when  $p = 3, d = 2$ . Both the training and test data are taken from our parametric models. On the basis of the training data, MLEs of the parameters are obtained for our model and also for the local constant model and local linear model of Di Marzio *et al.* (2014). For the models of Di Marzio *et al.* (2014), the kernel function  $K_m(h)$  is chosen to be proportional to  $\exp(m \cos \theta)$ , where  $\theta$  is the angular distance between the concerned observation and any other observation. Cross-validation is used to choose the appropriate concentration parameter  $m$  for the kernel. The predicted point corresponding to the test set is then obtained by using this  $m$  and the observations which form the training data.

The comparison between our model, local constant model (NPE1) and the local linear model (NPE2) is again done by  $D$ -values, defined in (3.1). For different values of the model parameters, the corresponding values of  $D$  are shown in Tables 3-6. In Tables 3 and 4, using 1000 simulations and assuming Exit distribution for the error, the  $D$ -values along with the estimates of the parameters (with se's or circular variables) are given. The  $D$  values along with the parameter estimates for  $p = 3, d = 2$  for different set of parameters are shown in Tables 5 and 6. The values in the Tables are based on the average values of  $D$  in 1000 such simulations. Lower the value of  $D$ , better is the fit of the model. In the Tables, the  $D$ -value corresponding to our parametric model is denoted by  $D_{PE}$ , the same corresponding to the local constant model (Di Marzio *et al.*, 2014) is denoted by  $D_{NPE1}$  and the  $D$ -value corresponding to the local linear model (Di Marzio *et al.*, 2014) is denoted by  $D_{NPE2}$ . Here  $\beta_1 = (b_1, b_2)$  when  $p = 2$  and  $\beta_1 = (b_1, b_2, b_3)$  when  $p = 3$ ,  $A = U\Lambda V$  and  $\rho$  is the parameter of

Table 3: Estimates of parameters (with standard errors/circular variances in parentheses) for  $\rho = 0.85$ ,  $2\theta_1 = \pi/2$ ,  $2\theta_2 = \pi/3$ ,  $\theta_3 = \pi/5$ ,  $\phi = \pi/3$ ,  $\lambda = 5$ , and  $\beta_1 = (0.6, 0)$  for different values of  $n$ .

	$n = 25$	$n = 50$	$n = 100$
Parameters			
$2\theta_1$	1.604 (0.126)	1.553 (0.081)	1.565 (0.039)
$2\theta_2$	1.090 (0.100)	1.090 (0.054)	1.073 (0.023)
$\theta_3$	0.692 (0.020)	0.654 (0.008)	0.641 (0.003)
$\phi$	1.025 (0.017)	1.032 (0.007)	1.043 (0.002)
$\rho$	0.874 (0.001)	0.861 (0.001)	0.855 (0.000)
$b_1$	0.574 (0.004)	0.591 (0.003)	0.596 (0.002)
$b_2$	-0.035 (0.004)	-0.007 (0.002)	-0.006 (0.001)
$\lambda$	5.199 (0.025)	5.141 (0.020)	5.058 (0.012)
$D_{PE}$	0.190	0.175	0.168
$D_{NPE1}$	0.292	0.223	0.199
$D_{NPE2}$	0.479	0.471	0.462

exit distribution,  $U$  is a rigid-rotation matrix and hence it can be explained by three angles  $\theta_1, \theta_2 \in [0, \pi)$  and  $\theta_3 \in [0, 2\pi)$ . The rotation matrix  $V$  can be explained by a parameter  $\phi \in [0, 2\pi)$ . As  $\theta_1, \theta_2$  show the axes,  $2\theta_1, 2\theta_2$  are considered to be circular parameters. Thus, in the simulation tables the circular mean has been computed for  $2\theta_1, 2\theta_2, \theta_3, \phi$  and the circular variances are reported for them.

## 5 Real Data Analyses

### 5.1 Vectorcardiogram data

Vectorcardiogram dataset is mentioned in Downs (2003) where two vectorcardiograms, one using the Frank system and the other using the McFee system, were taken from 28 boys and 25 girls. The readings corresponding to Frank system were denoted by  $F_{max} = (F_x, F_y, F_t)^\top$  and the same corresponding to McFee system were denoted by  $M_{max} = (M_x, M_y, M_t)^\top$ , where  $F_{max}$  and  $M_{max}$  are unit vectors in the direction of maximum QRS loop vector. As in Downs (2003), we have chosen  $F_{max}$  to be the dependent variable and  $M_{max}$  to be the independent variable (see Downs (2003) for details). The regression models were fitted both for boys and girls separately along with the rigid rotation models, to illustrate the comparison of our

Table 4: Estimates of parameters (with standard errors/circular variances in parentheses) for  $n = 50$ ,  $2\theta_1 = \pi/2$ ,  $2\theta_2 = \pi/3$ ,  $\theta_3 = \pi/5$ ,  $\phi = \pi/3$ ,  $\lambda = 5$ , and  $\rho = 0.85$  for different values of  $\beta_1$ .

	$\beta_1 = (0.6, 0)$	$\beta_1 = (0.3, 0)$
Parameters		
$2\theta_1$	1.553 (0.081)	1.590 (0.081)
$2\theta_2$	1.090 (0.054)	1.076 (0.059)
$\theta_3$	0.654 (0.008)	0.668 (0.012)
$\phi$	1.032 (0.007)	1.045 (0.003)
$\rho$	0.861 (0.001)	0.862 (0.001)
$b_1$	0.591 (0.003)	0.291 (0.004)
$b_2$	-0.007 (0.002)	-0.010 (0.003)
$\lambda$	5.141 (0.020)	5.166 (0.022)
$D_{PE}$	0.175	0.175
$D_{NPE1}$	0.223	0.221
$D_{NPE2}$	0.471	0.474

Table 5: Estimates of parameters (with standard errors/circular variances in parentheses) for  $\beta_1 = (0.6, 0, 0)$ ,  $2\theta_1 = \pi/2$ ,  $2\theta_2 = \pi/3$ ,  $\theta_3 = \pi/5$ ,  $\phi = \pi/3$ ,  $\lambda = 5$ , and  $\rho = 0.85$  for different values of  $n$ .

	$n = 25$	$n = 50$	$n = 100$
Parameters			
$2\theta_1$	1.598 (0.089)	1.606 (0.113)	1.640 (0.089)
$2\theta_2$	1.053 (0.020)	1.068 (0.025)	1.022 (0.017)
$\theta_3$	0.632 (0.046)	0.628 (0.033)	0.631 (0.029)
$\phi$	1.044 (0.002)	1.049 (0.001)	1.056 (0.001)
$\rho$	0.880 (0.001)	0.865 (0.001)	0.852 (0.001)
$b_1$	0.589 (0.004)	0.594 (0.002)	0.624 (0.001)
$b_2$	-0.037 (0.006)	-0.033 (0.004)	-0.011 (0.001)
$b_3$	0.015 (0.006)	0.013 (0.004)	0.011 (0.001)
$\lambda$	5.035 (0.009)	5.070 (0.016)	5.099 (0.014)
$D_{PE}$	0.184	0.165	0.173
$D_{NPE1}$	0.381	0.369	0.247
$D_{NPE2}$	0.334	0.322	0.301

Table 6: Estimates of parameters (with standard errors/circular variances in parentheses) for  $n = 50$ ,  $2\theta_1 = \pi/2$ ,  $2\theta_2 = \pi/3$ ,  $\theta_3 = \pi/5$ ,  $\phi = \pi/3$ ,  $\lambda = 5$ , and  $\rho = 0.85$  for different values of  $\beta_1$ .

	$\beta_1 = (0.6, 0, 0)$	$\beta_1 = (0.3, 0, 0)$
Parameters		
$2\theta_1$	1.606 (0.113)	1.574 (0.051)
$2\theta_2$	1.068 (0.025)	1.057 (0.031)
$\theta_3$	0.628 (0.033)	0.637 (0.009)
$\phi$	1.049 (0.001)	1.050 (0.001)
$\rho$	0.865 (0.001)	0.870 (0.001)
$b_1$	0.594 (0.002)	0.296 (0.002)
$b_2$	-0.033 (0.004)	-0.014 (0.003)
$b_3$	0.013 (0.004)	0.011 (0.004)
$\lambda$	5.070 (0.016)	5.078 (0.013)
$D_{PE}$	0.165	0.159
$D_{NPE1}$	0.369	0.376
$D_{NPE2}$	0.322	0.333

model to the rotational model.

The  $p$ -value of the test assuming Exit distribution based on the  $\Lambda$  statistic for the Boys' data is 0.1160 and that for the Girls' data is 0.1101. The  $p$ -values are not large, although not significant at 5% level. But, we still fit our proposed model to this dataset due to the historical importance of this dataset.

The estimate of regression coefficient  $\beta_1$  for the Boys' data is closer to 1, hence the projection and subsequent rotation makes the predicted values more concentrated about  $\beta_1$  while in the Girls' data,  $\hat{\beta}_1$  is far from 1, and hence the predicted values are less concentrated. The projection points in Girls' data are near 0. However, still  $\Lambda$  for Girls' data is more than that of Boys' data. From our detailed simulations (not reported here), we observe that  $\Lambda$  increases with  $\rho$ . Hence, this can be explained by the values of  $\hat{\rho}$  which is higher in the case of Girls' data.

In fact, the sum of geodesic distances for Boys' data for the rotational model and for our proposed model come to be  $D_{NPE} = 0.3856$  and  $D_{PE} = 0.3576$ , respectively, while the same for the Girls' data for the rotational model and our proposed model are  $D_{NPE} = 0.202$  and  $D_{PE} = 0.1866$ , respectively. Thus, our proposed model provides better fit for the Boys' data

as well as the Girls' data; we obtain about 7-8% reduction in  $D$ -values in both the cases.

The pictorial representation of the extent of goodness of fit by the Lambert azimuthal equal-area projection (see Mulcahy, 2007) is also shown in Figure 6. Here the mean predicted values and the observed values on the sphere are first projected inside a unit disc of radius 2. Then, the mean predicted values are joined to the observed values. The smaller the lengths of these line segments, the better is the fit.

To check the assumptions for link function and error distribution, we calculated marginally  $c = (c_1, \dots, c_3)$  for both the datasets. For boy data,  $c = (0.525, 0.477, 1.313)$  and for girls' data it is  $(0.842, 0.867, 0.515)$ . We also checked the marginal plots of  $\Delta$  (shown in supplementary material) which show that the points are approximately distributed randomly around 0. We also plot the histograms in Figure 7 for circular distances  $(1 - y_p^\top y_0)$  of observed points and predicted points. The histograms are monotonically decreasing. Thus, we may conclude that the assumptions on link function and error distribution are not much violated.

## 5.2 Hottest Day Data

The observations corresponding to the hottest date of the year 2011 were taken for major cities of the world. The covariate is the location of the city on the globe. Assuming that the earth is a perfect sphere, on the basis of the longitude and latitude, the co-ordinates of the points on the surface of the sphere is obtained, and these points are the covariates  $x \in \mathbb{S}^2$ . Dates are converted in angles taking 1st January as the 0 angle in case of responses, and hence the responses are on the circumference of a unit circle. Thus,  $y \in \mathbb{S}^1$ . Then,  $y$  is regressed on  $x$  by the method discussed in this paper.

The *donut-plot*, introduced by Jha and Biswas (2017) is shown for our model as well as the local constant and local linear models. In the *donut plot*,  $p_i = 1 + \cos(\theta_p - \theta_o)(\cos \theta_p \cdot \sin \theta_p)$ , the points in the direction of predicted values are plotted. Points near the circumference of the outer circle show good fit while points near the centre of the circle inside the inner circle show bad fit for the points. If a predicted point is in anti-clockwise direction of the predicted point, then they are shown as circles otherwise they are shown as stars.

The nonparametric regression using the local constant model of Di Marzio *et al.* (2014) resulted in the value of the optimum tuning parameter  $k$  to be 199.99, while the nonparametric regression using the local linear model gave the tuning parameter  $k$  as 114.48. The

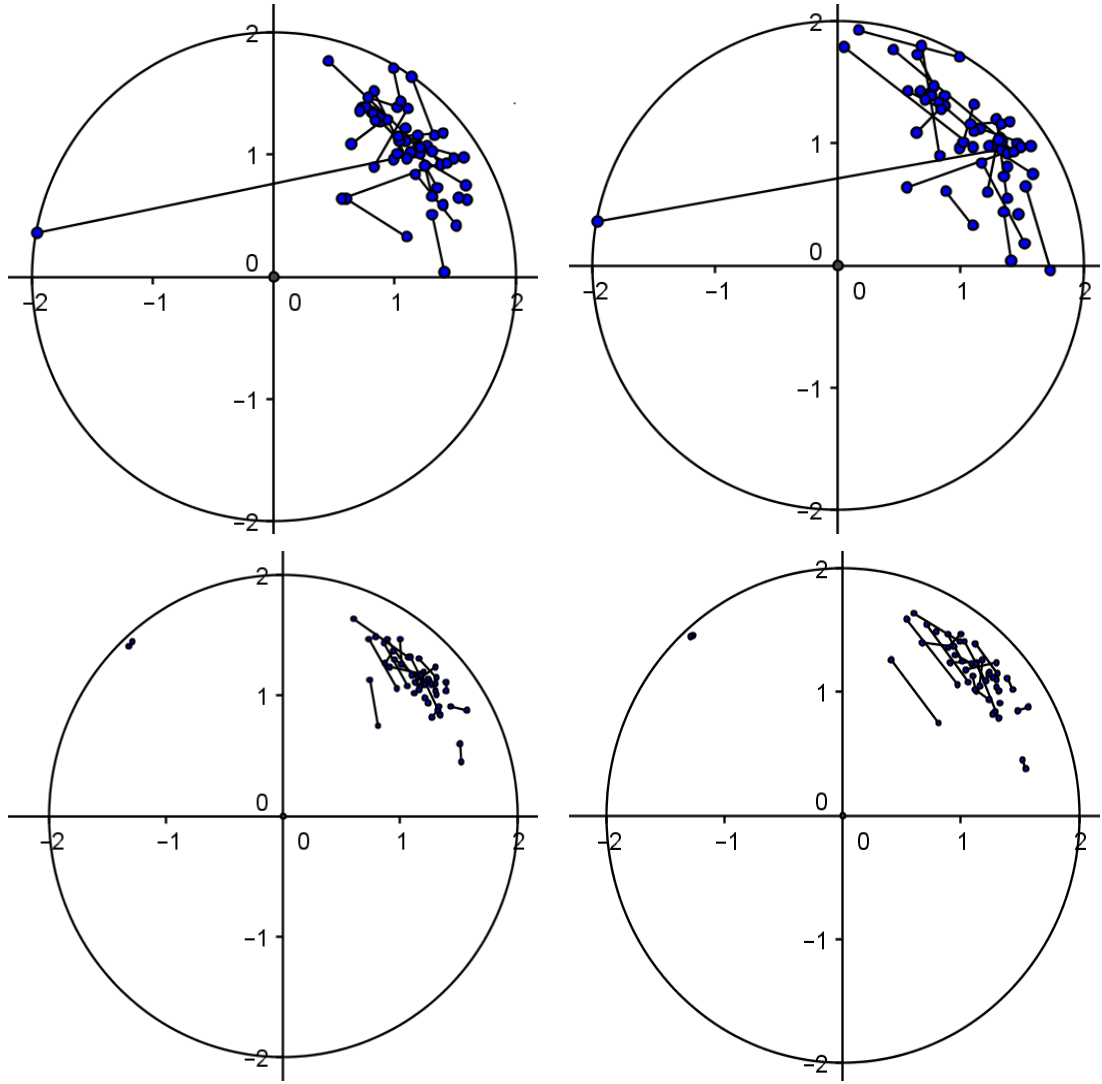


Figure 6: Lambert equal-area projection for vectorcardiogram data. Top row: Boys' data; Bottom row: Girls' data. Left column: proposed model; Right column: rigid-rotation model.

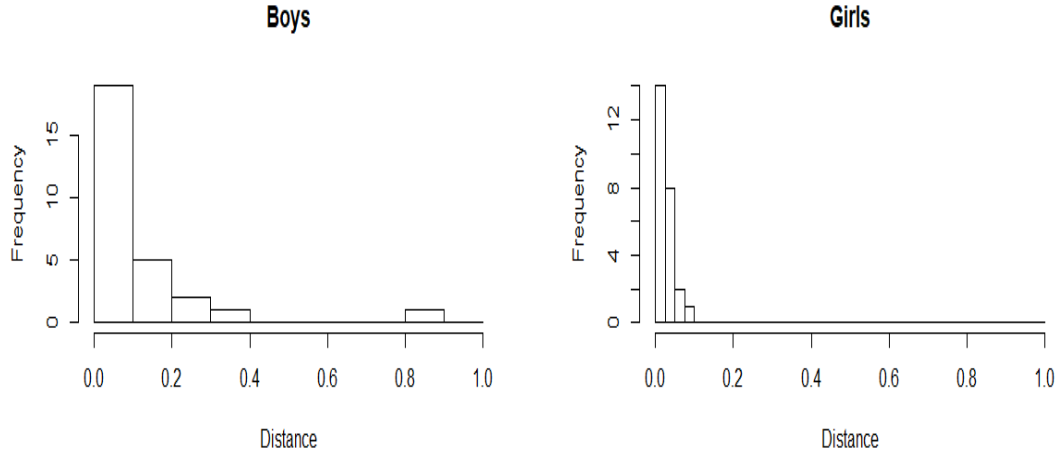


Figure 7: Histogram for circular distances between predicted and observed points for vectorcardiogram data.

*donut-plot* for the nonparametric local constant model is presented in the left panel of the lower half of Figure 10, while the same for the local linear model is drawn on the right panel of the lower half of Figure 10.

The parametric model, introduced in the paper, the local constant and local linear non-parametric models were compared with respect to the  $D$  measure based on all the observations. For the proposed parametric model, the  $D$  value comes out to be 0.721, while the same for the local constant model and local linear model are 0.819 and 0.759, respectively. We have slight gain by the proposed parametric model.

To check the assumptions for link function and error distribution, we find  $c = (-1.294, 4.054)$  for the dataset. Also, the marginal plots of  $\Delta$  are shown in Figure 8. It seems from the plots that there are some points which are far from 0. We also plot the histograms for circular distances of observed points and predicted points. The histogram as seen in Figure 9 is not monotonically decreasing smoothly. Therefore, from the marginal plots and the histogram, it can be said that the assumptions about the distribution link function and error distribution are not completely true and there are some other patterns in the data too.

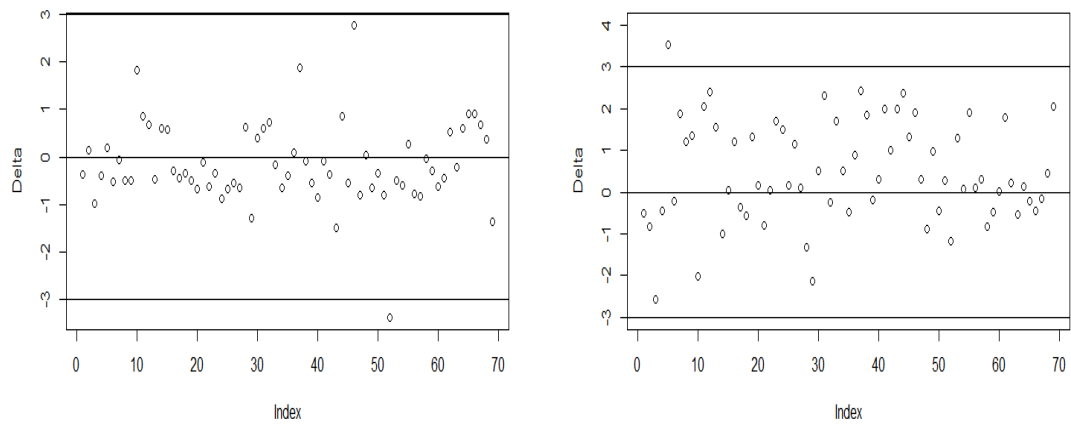


Figure 8: Marginal plots of  $\Delta$  for Hottest day data

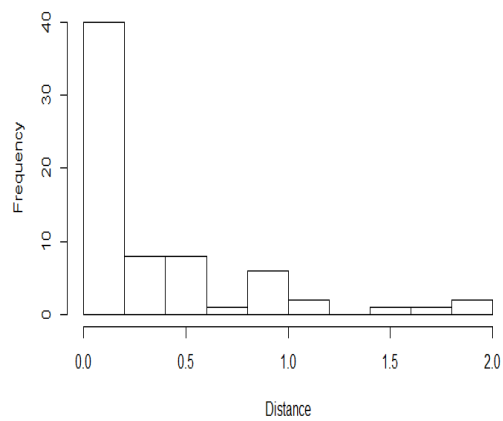


Figure 9: Histogram of circular distance for Hottest day data

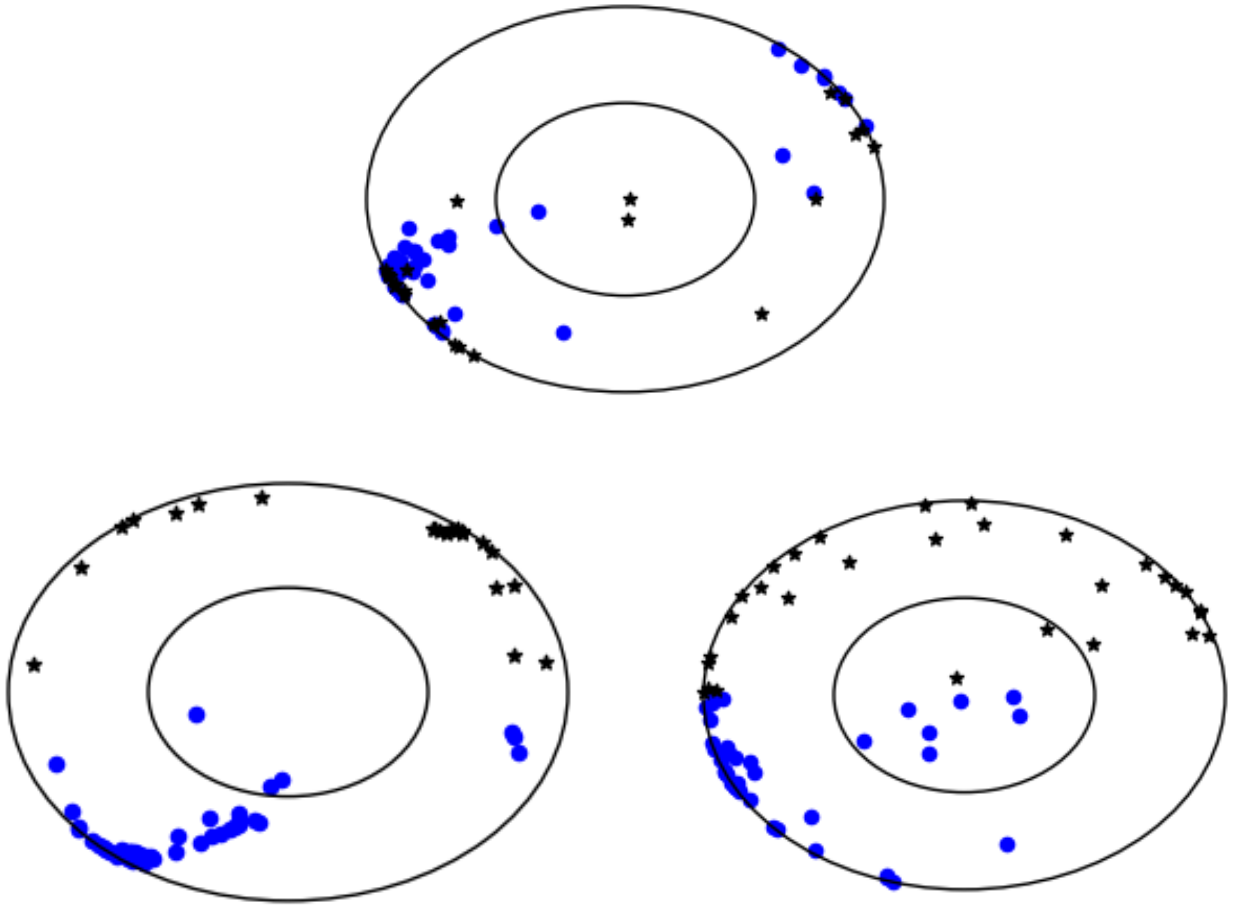


Figure 10: (Upper:) *donut plot* for the proposed model; (Bottom Left:) *donut plot* for the nonparametric local constant model; (Bottom Right:) *donut plot* for the nonparametric local linear model.

## 6 Concluding remarks

In this paper, we have used a model based on the geometry of circular-circular regression with the use of Möbius transformation as mentioned in Kato *et al.* (2008). This approach of spherical-spherical regression is new as far as we know. Also, we covered a fairly general situation –  $p$  can be smaller, equal or larger than  $d$ .

Multiple spherical regression using this model can be carried out by projecting the  $x_\beta$ 's with respect to each covariate and then taking the convex combination of these projected values by assigning suitable weights for different covariates such that the sum of the weights is 1, and then projecting the final point on the surface of the sphere. Then, the square matrix can be multiplied to this final point to get the predicted mean value  $\gamma$ .

Other distributions for error like the von Mises-Fisher distribution or the Watson distribution, etc., can also be used together with the proposed model.

## Acknowledgements

The authors wish to thank the Editor, Associate Editor and the referees for their careful reading and constructive suggestions which led to some improvement over an earlier version of the manuscript.

## References

- Bryner, D., Klassen, E. and Srivastava, A.(2012). Affine invariant elastic shape analysis of planar contours. *IEEE conference on computer vision and pattern recognition* 2012, 390-397.
- Chang, T. (1986). Spherical regression. *Annals of Statistics* 14, 907-924.
- Chang, T. (1989). Spherical regression with errors in variables. *Annals of Statistics* 17, 293-306.
- Chapman, G.R., Chen, G. and Kim, P.T. (1995). Assessing geometrical integrity through spherical regression techniques. *Statistica Sinica* 5, 173-220.
- Di Marzio, M., Panzera, A. and Taylor, C.C. (2014). Nonparametric regression for spherical data. *Journal of the American Statistical Association* 109, 748-763.
- Downs, T.D. and Mardia, K.V. (2002). Circular Regression. *Biometrika* 89, 683-697.
- Downs, T.D. (2003). Spherical regression. *Biometrika* 90, 655-668.
- Durrett, R. (1984). *Brownian Motion and Martingales in Analysis*. Belmont, CA: Wadsworth. MR0750829
- Fisher, N., Lewis, T. and Embleton, J.J.(1993).Statistical analysis of spherical data. *Cambridge University Press*
- Hall, P., Watson, G.S. and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika* 74, 751-762.
- Jha, J. and Biswas, A. (2017). Multiple circular-circular regression. *Statistical Modelling* 17(3), 142-171.
- Kato, S. (2009). A distribution for a pair of unit vectors generated by Brownian motion. *Bernoulli* 15, 898-921.
- Kato, S., Shimizu, K. and Shieh, G.S. (2008). A circular-circular regression model. *Statistica Sinica* 18, 633-645.

- Mackenzie, J.K. (1957). The estimation of an orientation relationship. *Acta Cryst* 10, 61-62.
- McCullagh, P. (1989). Some statistical properties of a family of continuous univariate distributions. *Journal of the American Statistical Association* 84, 125-129.
- Minh, D.L.P. and Farnum, N.R. (2003). Using bilinear transformations to induce probability distributions. *Communications in Statistics – Theory and Methods* 39, 1-9.
- Mulcahy, K. (2007). Lambert azimuthal equal area. *City University of New York*. Retrieved 2007-03-30.
- Müller, M.E. (1956). Some continuous Monte Carlo methods for the Dirichlet problem. *Ann. Math. Statist.* 27, 569-589.
- Rivest, L.-P. (1989). Spherical regression for concentrated Fisher-von Mises distribution. *Annals of Statistics* 17, 307-317.
- Rivest, L.-P. (1997). A decentred predictor for circular-circular regression. *Biometrika* 84, 717-726.
- Rodrigues, O.(1840). Des lois géométriques qui régissent les déplacements d' un système solide dans l' espace, et de la variation des coordonnées provenant de ces déplacements considérées indépendent des causes qui peuvent les produire, *J. Math. Pures Appl.* 5, 380–440.
- Rosenthal, M., Wu, W., Klassen, E. and Srivastava, A. (2014). Spherical regression models using projective linear transformations. *Journal of the American Statistical Association* 109, 1615-1624.
- Stephens, M.A. (1979). Vector correlation. *Biometrika* 66, 41-48.