Navigating the Abyss: Spatio-temporal Dependencies of Turtle Migration with Sun, Moon, & Earth's Magnetic Field(Supplimentary Material)

Debashis Chatterjee[1] and Prithwish Ghosh[2*]

[1]Department of Statistics, Visva Bharati, Santiniketan, Bolpur, 731235, West Benal, India.

[2]Department of Statistics, North Carolina State University 5109, SAS Hall, 2311 Stinson Dr, Raleigh, NC 27607, United States.

*Corresponding author(s). E-mail(s): pghosh4@ncsu.edu;
Contributing authors: debashis.chatterjee@visva-bharati.ac.in;

Abstract

We employ a directional statistical approach to investigate the characteristics of migratory Turtle paths across the American continents. Beginning with a dataset featuring migratory paths, latitude, longitude, and observation dates of Turtle obtained from a Turtle-path tracking web resource, we compile a derived dataset. This new dataset encompasses directional changes in Turtle paths alongside spatiotemporal positions of the Sun, Moon, , and Earth's magnetic field. The dataset is partitioned into 6 optimal sub-regions, with each subdivision's latitude and longitude distributions fitting based on vonMises fishers distribution. Where, we propose a directional mixture model incorporating Von Mises and uniform distributions for general Turtle path data. We investigate whether directional changes in environmental factors such as the positions of the Sun, Moon, and Earth's magnetic field directly influence a Turtle's decision to alter its path at a particular location on Earth. We quantify changes in the positions of the Sun and Moon, represented by azimuth and zenith angles, and perform analogous calculations for the Earth's magnetic field.

Keywords: Circular Statistics, vonMises fishers, Mixture Modelling, Turtle Migration

1 Materials and Methods

1.1 Novel Hypothesis on Decision of Turtle to Change its Migratory Path based on the Chane of Earth Magnetic field

We seek a rigorous statistical answer to whether a suitable change of the Earth's Magnetic field at a spot on the Earth directly influences a Turtle's decision to fly on that spot to change its paths, and our statistical result affirms that. We consider a novel circular-circular regression model for that and test whether the regression model is validated.

1.2 Principal Component Analysis (PCA)

PCA is a technique used for dimensionality reduction and feature extraction. It transforms the original variables into a new set of uncorrelated variables called principal components. The PCA algorithm finds the directions (principal components) along which the data varies the most. [1]

Given a dataset with n observations and p variables represented by the matrix X, where each row corresponds to an observation and each column corresponds to a variable, PCA computes the covariance matrix C of X.

The covariance matrix C is then decomposed into its eigenvectors and eigenvalues. The eigenvectors represent the directions of maximum variance in the data, while the eigenvalues represent the magnitude of the variance along each eigenvector.[2]

The principal components are obtained by projecting the original data onto the eigenvectors. These components are ordered by the corresponding eigenvalues, with the first principal component capturing the most variance in the data, the second capturing the second most, and so on.

One approach to impute missing values using PCA is to reconstruct the missing values using the information captured by the principal components. The missing values for each observation can be estimated by projecting the observed values onto the principal components and then back-transforming them to the original variable space.

The imputed values are obtained by multiplying the projected values by the loading matrix, representing the relationship between the original variables and the principal components. The loading matrix can be obtained from the eigenvectors of the covariance matrix.

This process allows for imputing missing values while preserving the underlying data structure captured by the principal components. [3]

Consider a dataset $x_1, x_2, x_3, \ldots, x_9$, where the parameters are "$del_{solar_{azimuth}}$","$del_{solar_{zenith}}$","$del_{lunar_{azimuth}}$","del_i","del_d" in the dataset $\mathbf{X}$ with n samples and 5 features arranged as an $n \times p$ matrix.

First, center the data by subtracting the mean of each feature:

$$\mathbf{X}_{\text{centered}} = \mathbf{X} - \bar{\mathbf{X}}$$

where $\bar{\mathbf{X}}$ is a vector of length p containing the means of each feature.

Calculate the covariance matrix $\mathbf{S}$ of the centered data:

$$\mathbf{S} = \frac{1}{n-1}(\mathbf{X}^T_{\text{centered}} \cdot \mathbf{X}_{\text{centered}})$$

Perform an eigenvalue decomposition on $\mathbf{S}$:

$$\mathbf{S} = \mathbf{V} \cdot \mathbf{\Lambda} \cdot \mathbf{V}^T$$

where $\mathbf{V}$ is a matrix containing the eigenvectors and $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues.

1.3 World Magnetic Model

The World Magnetic Model (WMM) provides a comprehensive representation of the Earth's magnetic field on a global scale [4]. This model utilizes a spherical harmonic expansion up to the 12^{12} degree and order, capturing the magnetic scalar potential generated by the Earth's core, which contributes to the geomagnetic main field. Alongside the 168 spherical-harmonic "Gauss" coefficients, the model incorporates an equal number of spherical-harmonic Secular-Variation (S.V.) coefficients. All quantities in this section adhere to the following measurement conventions: angles are in radians, lengths are in meters, magnetic intensities are in nanoteslas (nT, where one tesla is one weber per square meter or one kg.s^{-2}.A^{-1}), and times are in years [5].

The primary magnetic field B_m is a potential field and can, therefore, be expressed in geocentric spherical coordinates (longitude λ, latitude ϕ', radius r) as the negative spatial gradient of a scalar potential.

$$B_m(\lambda, \phi', r, t) = - \bigtriangledown V(\lambda, phi', r, t) \tag{1}$$

where t represents the time. The potential can be expressed as a series expansion in spherical harmonics: [6].

$$V(\lambda, \phi', r, t) = a \left(\sum_{n=1}^{N} \sum_{m=1}^{M} (\frac{a}{r})^{n+1} (g_n^m(t) \cos(m\lambda) + h_n^m(t) \sin(m\lambda) \breve{p}_n^m(t) (\sin \phi') \right) \tag{2}$$

We selected N = 36 as the truncation level for the internal expansion of the World Magnetic Model. Here, 'a' (6371200 m) represents the geomagnetic reference radius, closely approximating the mean Earth radius. The variables (λ, ϕ', r) denote the longitude, latitude, and radius in a geocentric spherical reference frame, respectively. Additionally, $g_n^m(t)$ and $h_n^m(t)$ represent the time-dependent Gauss coefficients of degree n and order m, describing the Earth's main magnetic field. The parameters are defined as follows:

$$g_n^m(t) = g_n^m + \dot{g}_n^m (t - t_0) + \ddot{g}_n^m (t - t_0)^2 \tag{3}$$

$$h_n^m(t) = h_n^m + \dot{h}_n^m (t - t_0) + \ddot{h}_n^m (t - t_0)^2 \tag{4}$$

In this equation, $g_n^m, h_n^m, \dot{g}_n^m, \dot{h}_n^m, \ddot{g}_n^m, \ddot{h}_n^m$ from Equation 3 are considered constants. For any real number μ, $\breve{p}_n^m(\mu)$ represents the Schmidt semi-normalized associated Legendre functions and is defined as:

$$\breve{p}_n^m(\mu) = \sqrt{2 \frac{(n-m)!}{(n+m)!}} p_{n,m}(\mu) \quad , if \quad m > 0 \tag{5}$$

$$\breve{p}_n^m(\mu) = p_{n,m}(\mu) \quad , if \quad m = 0 \tag{6}$$

Finally, in the case of a data set containing hourly mean observatory data, offsets must be incorporated at each observatory to consider the local magnetic field, primarily induced in the Earth's crust, which is beyond the model's scope. Consequently, at a given observatory, the magnetic field (B) is described as follows:

$$B_m(\lambda, phi', r, t) = - \bigtriangledown V(\lambda, phi', r, t) + O(\lambda, phi', r, t). \tag{7}$$

The offset vector $O(\lambda, \phi', r, t)$, commonly known as the crustal bias, remains constant over time. The parameterization is employed to model datasets selected from satellite measurements and hourly mean values observed at the observatory. The equations governing the internal component of the field are as follows:

$$X'(\lambda, \phi', r, t) = -\frac{\partial V}{r \partial \phi'} = -\sum_{n=1}^{N} (\frac{a}{r})^{n+2} \sum_{m=0}^{M} (g_n^m(t) \cos(m\lambda) + h_n^m(t) \sin(m\lambda)) \frac{d\breve{p}_n^m(\sin \phi')}{d\phi'} \tag{8}$$

$$Y'(\lambda, \phi', r, t) = -\frac{\partial V}{r \cos \phi' \partial \lambda} = \frac{1}{r \cos \phi'} \sum_{n=1}^{N} (\frac{a}{r})^{n+2} \sum_{m=0}^{M} m (g_n^m(t) \cos(m\lambda) + h_n^m(t) \sin(m\lambda)) \breve{p}_n^m(\sin \phi') \tag{9}$$

$$Z'(\lambda, \phi', r, t) = \frac{\partial V}{\partial r} = -\sum_{n=1}^{N} (n+1) (\frac{a}{r})^{n+2} \sum_{m=0}^{M} (g_n^m(t) \cos(m\lambda) + h_n^m(t) \sin(m\lambda)) \breve{p}_n^m(\sin \phi') \tag{10}$$

The equations above, with the magnetic field vector observations on the left-hand side, constitute the system of condition equations. Consequently, if there are d data points, the system comprises d linear equations involving the p parameters of the parent model:

$$y_{d\times 1} = A_{d\times p} \cdot m_{p\times 1}, \tag{11}$$

Where **y** is the column vector $(d \times 1)$ of observations, **A** is the matrix $(d \times p)$ of coefficients corresponding to the unknowns, which are functions of position, and **m** is the column vector $(p \times 1)$ of unknowns, representing the Gauss coefficients of the model. Since there are more observations than unknowns $(d > p)$, the system is over-determined and, thus, does not have an exact solution. The final main-field coefficients for 2005.0 were obtained by polynomial extrapolation of the main-field Gauss coefficients from the parent model to this date, using Equation 3. Equations 8, with the time-varying Gauss coefficients g_n^m, h_n^m replaced by their time derivatives $\dot{g}_n^m, \dot{h}_n^m$, were then utilized to determine the final secular-variation coefficients.[7]

1.4 Haversines method for Turtle path Change

To answer the Hypothesis stated in section 1.1, We wish to extract the directional change in the migration path of the Turtles from dataset 1.

A Turtle's decision to change its migration path is invariant intra-species but not invariant inter-species. (In other words, given a particular Turtle species, we may assume the path directions are independent and identically distributed (iid) directional random variables, but if species change, the assumption of iid will be violated).

1.5 Spatio-temporal Calculation of the Position of Sun and Moon

A new model used to understand the migration of Turtle paths depends on the sun's position, as Turtles use the sun's position in the sky as a compass to navigate their way during their migratory journeys. The results indicate that the proposed model can estimate the Position of the Sun concerning the change of longitude, latitude, and day of the year.
[8]

Calculation for the zenith and azimuth angles for Sun and Moon

Suppose the observer's coordinates, or latitude and longitude, are (ϕ_0, λ_0), and the subsolar point's coordinates are (ϕ_s, λ_s), then the x-, y- and z-components of the unit vector, S, pointing from the observer to the center of the Sun.

- $\phi_s = \delta$
- λ_s = -15(T_{GMT} - 12 + E_{min}/60)
- $S_x = \cos\phi_s \sin(\lambda_s - \lambda_0)$
- $S_y = \cos\phi_0 \sin\phi_s$ - $\sin\phi_0 \cos\phi_s \cos(\lambda_s - \lambda_0)$
- $S_z = \cos\phi_0 \sin\phi_s$ - $\cos\phi_0 \cos\phi_s \cos(\lambda_s - \lambda_0)$

In this context, the influence of parallax is disregarded, effectively assuming an infinite Earth-Sun distance. It can be demonstrated that this assumption holds, and as such, there exists. $S_x^2 + S_y^2 + S_z^2 = 1$
Here T_{GMT} is the Greenwich Mean Time or UTC

Remark 1. *The derivation of S_x, S_y, and S_z becomes straightforward when working within the Earth-Centered Earth-Fixed (ECEF) [9] coordinate system. The ECEF system is a geocentric right-handed Cartesian system, and the process can be outlined as follows:*

Begin at the subsolar point in the ECEF system and construct a unit vector pointing upward. At the observer's coordinates in the ECEF system, create three unit vectors pointing east, north, and upward, respectively. Compute the dot product of each unit vector with the vector from Step 1 to obtain the right-hand sides of S_x, S_y, and S_z. Please note that this procedure disregards the influence of parallax, assuming an infinite Earth-Sun distance for simplicity and ease of calculation.

The solar zenith angle is now simply

$$Z = \cos^{-1} S_z \tag{12}$$

and the solar azimuth angle following the South-Clockwise convention is

$$\gamma_s = a \tan 2(-s_x, -s_y) \tag{13}$$

Remark 2. *The azimuth angle of the sun and moon provides valuable directional information relative to an observer, measured in degrees clockwise from the north. It reveals their lateral positions in the sky. The sun's azimuth angle continuously changes throughout the day as it moves across the sky. It starts at 90 degrees (east) during sunrise, increases until solar noon (varies by location and time of year), and sets at 270 degrees (west) during sunset. At solar noon, the sun is due south in the northern hemisphere and north in the southern hemisphere. Similarly, the azimuth angle of the moon changes as it orbits the Earth. During a full moon, it rises around 90 degrees, reaches its highest point at approximately 180 degrees, and sets at 270 degrees. During a new moon, the moon's azimuth at rising and setting aligns more closely with the sun's position. In conjunction with the altitude angle (vertical angle above the horizon), the azimuth angle allows observers to precisely determine the sun and moon's positions in the sky at any time and location. This knowledge is crucial for various applications, including astronomy, navigation, and aligning solar panels for maximum sunlight exposure.*

Remark 3. *The Sun's zenith angle is the angle between the Sun and the vertical point directly above an observer on Earth. At solar noon, it is 0 degrees when the Sun is now overhead. During sunrise and sunset, it is 90 degrees when the Sun is at the horizon. The zenith angle changes throughout the day due to Earth's rotation and varies with the observer's location, time of year, and time of day. It is crucial for solar energy applications, affecting the amount of solar and lunar position change.*

To compute the column Z and γ_s of dataset 2, we use the position of the Sun for each species. For given Turtles, we sort the dataset concerning latitude. Then, we took the first difference of magnitude of the Sun Position(computed using the formulae discussed before).

1.6 Directional Statistical Preliminaries

Von Mises Distribution

A circular random variable θ follows the Von Mises distribution, also known as the Circular Normal Distribution if it is characterized by the probability density function (pdf): [10]

$$f(\theta; \mu, k) = \frac{1}{2\pi I_0(k)} e^{k \cos\theta(\theta-\mu)}, \tag{14}$$

In this equation, θ lies in the range $[0, 2\pi)$, μ is constrained to $[0, 2\pi)$, and $(k > 0)$. The normalizing constant $I_0(k)$ is the modified Bessel function of the first kind and order zero, given by:

$$\frac{1}{2\pi} \int_1^{2\pi} \exp(k \cos\theta) d\theta = \sum_{r=0}^{\infty} \left(\frac{k}{2}\right)^{2r} \left(\frac{1}{r!}\right)^2 \tag{15}$$

To determine the cumulative distribution of the circular normal or the Von Mises Distribution, we integrate the pdf, resulting in the following cumulative distribution function (cdf):

$$F(\theta) = \frac{1}{2\pi I_0(k)} \left(\theta I_0(k) + 2 \sum_{p=1}^{\infty} \frac{I_p(k) \sin p(\theta - \mu)}{p} \right), \tag{16}$$

where θ is confined to the interval $[0, 2\pi)$.

1.7 Test for Directional Correlation

In dealing with with the circular variable θ, M_1, and M_2 and trying to retain many of these properties,

$$\rho_{incc}(\theta_1\ ,\ M_{11}) = \frac{E\{\sin(\theta_i - \bar{\theta})\sin(M_{1i} - \bar{M_1})\}}{\sqrt{Var(\sin(\theta_i - \bar{\theta})Var(\sin(M_{1i} - \bar{M_1}))}} \tag{17}$$

$$\rho_{decc}(\theta_1\ ,\ M_{21}) = \frac{E\{\sin(\theta_i - \bar{\theta})\sin(M_{2i} - \bar{M_2})\}}{\sqrt{Var(\sin(\theta_i - \bar{\theta})Var(\sin(M_{2i} - \bar{M_2}))}} \tag{18}$$

Properties of ρ_{incc} and ρ_{decc}

1. $\rho_{incc}(\theta_1\ ,\ M_{11})$ and $\rho_{decc}(\theta_1\ ,\ M_{21})$ does not depend on the zero direction used for both variable
2. $|\rho_{incc}(\theta_1\ ,\ M_{11})| \leq 1$ and $|\rho_{decc}(\theta_1\ ,\ M_{21})| \leq 1$
3. $\rho_{incc}(\theta_1\ ,\ M_{11})$ and $\rho_{decc}(\theta_1\ ,\ M_{21}) = 0$ if θ, M_1 and θ, M_2 are independent although the converse need not to be true.

Our Hypothesis can be statistically formulated as follows:

$$H_0: \quad \rho_c = 0 \tag{19}$$

$$H_1: \quad \rho_c \neq 0. \tag{20}$$

Here we took $(\theta_1\ ,\ M_{11}), \cdots, (\theta_n\ ,\ M_{1n})$ and $(\theta_1\ ,\ M_{21}), \cdots, (\theta_n\ ,\ M_{2n})$ be some random sample of observations which are of two attributes both measured as angles concerning the same zero direction and the same sense of rotation. Let's take (θ, M_1) and (θ, M_2) be the joint probability density function on the torus $0 \leq \theta < 2\pi$, $0 \leq M_1 \leq 2\pi$ and $0 \leq M_2 \leq 2\pi$. Let $\bar{M_2}$, $\bar{M_1}$ and $\bar{\theta}$ denote the mean direction of three variables.

In the equation 21 $\bar{M_2}$, $\bar{M_1}$ and $\bar{\theta}$ are sample mean directions. Below correlation coefficient if $(\theta_1\ ,\ M_{11}), \cdots, (\theta_n\ ,\ M_{1n})$ and $(\theta_1\ ,\ M_{21}), \cdots, (\theta_n\ ,\ M_{2n})$ are random sample, given by [11]

$$r_{incc,n} = \frac{\sum_{i=1}^{n} \sin(\theta_i - \bar{\theta}) \sin(M_{1i} - \bar{M_1})}{\sqrt{\sum_{i=1}^{n} \sin^2(\theta_i - \bar{\theta}) \sin^2(M_{1i} - \bar{M_1})}} \tag{21}$$

$$r_{decc,n} = \frac{\sum_{i=1}^{n} \sin(\theta_i - \bar{\theta}) \sin(M_{2i} - \bar{M_2})}{\sqrt{\sum_{i=1}^{n} \sin^2(\theta_i - \bar{\theta}) \sin^2(M_{2i} - \bar{M_2})}} \tag{22}$$

The sample correlation is an estimation of ρ_c. When the joint distributions of (θ, M_1) and (θ, M_2) are not fully specified, we can use the sample measure $r_{incc,n}$ and $r_{decc,n}$ for testing the hypothesis about ρ_c mentioned in (19) when n is sufficiently large. For details on the test statistics and its distributional properties under H_0, we refer to [12].

References

[1] S. Wold, K. Esbensen, P. Geladi. Principal component analysis (1987)

[2] C. Labrín, F. Urdinez. Principal component analysis (2020)

[3] M. Greenacre, P.J. Groenen, T. Hastie, A.I. dEnza, A. Markos, E. Tuzhilina. Principal component analysis (2022)

[4] D. Bartlett, *Essentials of positioning and location technology* (Cambridge University Press, 2013)

[5] A. Chulliat, S. Macmillan, P. Alken, C. Beggan, M. Nair, B. Hamilton, A. Woods, V. Ridley, S. Maus, A. Thomson, The us/uk world magnetic model for 2015-2020. US/UK world magnetic model (2015)

[6] J. Rebollo, S. Khater, W.J. Coupe, *A recursive multi-step machine learning approach for airport configuration prediction*, in *AIAA Aviation 2021 Forum* (2021), p. 2406

[7] World magnetic model. https://www.ngdc.noaa.gov/geomag/WMM/data/WMMReports/TRWMM_2005.pdf. Accessed: 2023-11-30

[8] Y. El Mghouchi, A. El Bouardi, Z. Choulli, T. Ajzoul, New model to estimate and evaluate the solar radiation. International Journal of Sustainable Built Environment **3**(2), 225–234 (2014)

[9] S. Soulayman, Comments on solar azimuth angle. Renewable Energy **123**, 294–300 (2018)

[10] M.K. V., *Statistics of Directional Data* (Academic Press, New York, 1972)

[11] N.I. Fisher, A.J. Lee, A correlation coefficient for circular data. Biometrika **70**(2), 327–332 (1983)

[12] S.R.J. A SenGupta, *Topics in Circular Statistics*. Volume 5 (World Scientific Publishing Co. Pte. Ltd, 2001)