

Article

Möbius Transformation-Induced Distributions Provide Better Modelling for Protein Architecture

Mohammad Arashi ^{1,2,*} , Najmeh Nakhaei Rad ^{2,3} , Andriette Bekker ²  and Wolf-Dieter Schubert ⁴

¹ Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad 4897, Iran

² Department of Statistics, University of Pretoria, Pretoria 0002, South Africa; najme.rad@gmail.com (N.N.R.); andriette.bekker@up.ac.za (A.B.)

³ DSI-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS), Johannesburg 2000, South Africa

⁴ Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria 0002, South Africa; wolf-dieter.schubert@up.ac.za

* Correspondence: arashi@um.ac.ir

Abstract: Proteins are found in all living organisms and constitute a large group of macromolecules with many functions. Proteins achieve their operations by adopting distinct three-dimensional structures encoded within the sequence of the constituent amino acids in one or more polypeptides. New, more flexible distributions are proposed for the MCMC sampling method for predicting protein 3D structures by applying a Möbius transformation to the bivariate von Mises distribution. In addition to this, sine-skewed versions of the proposed models are introduced to meet the increasing demand for modelling asymmetric toroidal data. Interestingly, the marginals of the new models lead to new multimodal circular distributions. We analysed three big datasets consisting of bivariate information about protein domains to illustrate the efficiency and behaviour of the proposed models. These newly proposed models outperformed mixtures of well-known models for modelling toroidal data. A simulation study was carried out to find the best method for generating samples from the proposed models. Our results shed new light on proposal distributions in the MCMC sampling method for predicting the protein structure environment.

Keywords: bioinformatics; cosine model; mixture distributions; Möbius transformation; sine model; toroidal data



Citation: Arashi, M.; Nakhaei Rad, N.; Bekker, A.; Schubert, W.-D. Möbius Transformation-Induced Distributions Provide Better Modelling for Protein Architecture. *Mathematics* **2021**, *9*, 2749. <https://doi.org/10.3390/math9212749>

Academic Editor: Manuel Alberto M. Ferreira

Received: 16 September 2021

Accepted: 24 October 2021

Published: 29 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Proteins constitute a diverse set of biological macromolecules that are often referred to as the workhorses of cells because of their central role in most biological processes. Chemically, proteins are biopolymers consisting of linear sequences of amino acid covalently linked by peptide bonds, such that each polypeptide is a single large molecule. Nineteen of the natural amino acids (all but proline) have an amino group ($-NH_2$), a carboxylic acid group ($-COOH$), an amino acid-specific side-chain, and a hydrogen atom attached to a central carbon atom (C_α). Each peptide bond links the carboxylate group of one amino acid to the amino group of the next. Protein structure is often described in terms of four levels of organisation. The primary structure is the sequence of amino acids. The secondary structure refers to the local folding of the polypeptide backbone into helices, strands, or loops. The tertiary structure describes the complex three-dimensional folding of a polypeptide. Finally, the quaternary structure describes the involvement of one or more polypeptides in creating a functional protein. The amino nitrogen, C_α , and the carbonyl carbon of all residues constitute the protein backbone.

The 3D coordinates of proteins, as provided by electron microscopy, NMR, or X-ray crystallography, directly reveal the conformation of the backbone atoms, with knowledge of standard chemical bond angles and lengths incorporated during the refinement process.

Generally, the backbone conformation is analysed using the backbone torsion or the dihedral angles, denoted by ϕ , ψ , and ω , as introduced by Ramachandran [1] (Figure 1A), where ω is usually close to 180° or occasionally 0° . Alternatively, virtual bond and torsion angles θ and τ may be used to describe a protein backbone representation based on only C_α positions (Figure 1B).

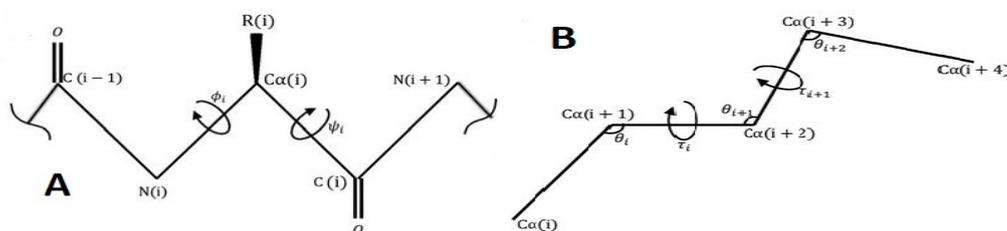


Figure 1. Two representations of protein backbone structures based on torsion or pseudo-torsion angles.

A major challenge in molecular biology and computational biochemistry involves predicting protein 3D structure. The encoding gene provides the primary structure of a protein, and the secondary structure may be predicted computationally with high reliability using artificial neural networks [2], based on the propensity of amino acids to form different secondary structures.

However, predicting the 3D structure of a protein, especially if it is larger than 100 amino acids or if a homologue with a known structure and significant sequence identity is not available, remains challenging. This challenge is addressed by de novo structure prediction, which requires parametrized physical force fields. The probability of observing a particular conformation x of the molecule, $p(x|\beta)$ is considered and expressed as the Boltzmann distribution:

$$p(x|\beta) = \frac{\exp(\beta U(x))}{Z_\beta},$$

where Z_β is the normalization constant, $U(x)$ is the potential energy of the molecule, $\beta = (k_b T)^{-1}$ is the thermodynamic beta, k_b is the Boltzmann constant, and constant T is the temperature. The 3D structure of a molecule can be derived from $p(x|\beta)$ by determining the mode of the distribution. Molecular dynamics (MD) is a simulation-based method used to probe for the mode of distribution. However, many millions to trillions of steps are required to simulate a single folding event. By contrast with MD, Monte Carlo (MC)-based methods are more time-efficient. In the Markov Chain Monte Carlo (MCMC) method, a Markov chain is constructed using the Metropolis–Hastings (MH) algorithm ([3,4]), with $p(x|\beta)$ as the stationary distribution. A symmetric proposal distribution is utilized in the MH algorithm.

Choosing a good proposal distribution is one of the challenges in MCMC-based simulation. Gaussian perturbations are the most straightforward proposal distributions that can be used [5]. The results are more accurate when the proposal distribution is closer to the stationary distribution; therefore, protein structural information is incorporated into most proposal distributions. Using the information on angles and bond lengths observed in real proteins is a simple way to define a suitable proposal distribution. Fragment libraries for backbone angles and rotamer libraries for side-chain angles can be selected as default choices for proposal distributions [6–8].

Various tractable statistical distributions for modelling protein dihedral angles are briefly reviewed. These models can be used as proposal distributions for MCMC protein sampling. They can also be utilized as a prior for determining a protein structure from data. However, these models do not generate folded proteins because they work under some simplifying assumptions, both in terms of their functional form and dependency structure (see [9]). The ultimate goal of our contribution is to propose more flexible models for the proposal distribution.

1.1. Brief Overview

An overview of the models available for toroidal data that forms the departure point for the investigation in this paper follows (see [10]).

The first probability distribution on the torus was proposed by Mardia in [11]. It is the bivariate von Mises distribution:

$$f(\theta_1, \theta_2) = C \exp(\kappa_1 \cos(\theta_1 - \iota_1) + \kappa_2 \cos(\theta_2 - \iota_2) + (\cos(\theta_1 - \iota_1), \sin(\theta_1 - \iota_1))A(\cos(\theta_2 - \iota_2), \sin(\theta_2 - \iota_2))^T),$$

where C is the normalizing constant, $\iota_1, \iota_2 \in [-\pi, \pi)$ are location parameters, $\kappa_1, \kappa_2 \geq 0$ are concentration parameters, and matrix $A_{2 \times 2}$ is the circular–circular dependence parameter. To move beyond the complexity created by the large number of parameters in this founding distribution, a few special cases in the literature have been considered. Rivest in [12] introduced the subclass:

$$f(\theta_1, \theta_2) \propto \exp(\kappa_1 \cos(\theta_1 - \iota_1) + \kappa_2 \cos(\theta_2 - \iota_2) + \alpha \cos(\theta_1 - \iota_1) \cos(\theta_2 - \iota_2) + \beta \sin(\theta_1 - \iota_1) \sin(\theta_2 - \iota_2)), \tag{1}$$

where $\alpha, \beta \in \mathbb{R}$. Singh et al. in [13] proposed the *sine model* as a special case of (1) with one less parameter, letting $\alpha = 0$ and $\beta = \kappa_3$:

$$f(\theta_1, \theta_2) = C \exp(\kappa_1 \cos(\theta_1 - \iota_1) + \kappa_2 \cos(\theta_2 - \iota_2) + \kappa_3 \sin(\theta_1 - \iota_1) \sin(\theta_2 - \iota_2)), \tag{2}$$

where

$$C^{-1} = 4\pi^2 \sum_{i=0}^{\infty} \binom{2i}{i} \left(\frac{\kappa_3^2}{4\kappa_1\kappa_2} \right)^i I_i(\kappa_1)I_i(\kappa_2), \tag{3}$$

where $I_\alpha(z)$ is the modified Bessel function of the first kind of order α . Another submodel of (1), the *cosine model*, was introduced by Mardia et al. in [14] by setting $\alpha = \beta = -\kappa_3$:

$$f(\theta_1, \theta_2) = C \exp(\kappa_1 \cos(\theta_1 - \iota_1) + \kappa_2 \cos(\theta_2 - \iota_2) - \kappa_3 \cos(\theta_1 - \iota_1 - \theta_2 + \iota_2)), \tag{4}$$

where

$$C^{-1} = 4\pi^2 \left\{ I_0(\kappa_1)I_0(\kappa_2)I_0(\kappa_3) + 2 \sum_{i=1}^{\infty} I_i(\kappa_1)I_i(\kappa_2)I_i(\kappa_3) \right\}. \tag{5}$$

It is worth noting that Kent et al. in [15] introduced another version of the cosine model, with a negative interaction given by:

$$f(\theta_1, \theta_2) = C \exp(\kappa_1 \cos(\theta_1 - \iota_1) + \kappa_2 \cos(\theta_2 - \iota_2) - \kappa_3 \cos(\theta_1 - \iota_1 + \theta_2 - \iota_2)),$$

with the same normalizing constant as for the model with a positive interaction in (4). Kent et al. in [15] also introduced a submodel of (1), which is a hybrid between the sine and cosine models, given by:

$$f(\theta_1, \theta_2) \propto \exp(\kappa_1 \cos(\theta_1 - \iota_1) + \kappa_2 \cos(\theta_2 - \iota_2) + \beta\{(\cosh \lambda - 1) \cos(\theta_1 - \iota_1) \cos(\theta_2 - \iota_2) + \sinh \lambda \sin(\theta_1 - \iota_1) \sin(\theta_2 - \iota_2)\}), \tag{6}$$

where $\kappa_1, \kappa_2 \geq 0, \lambda \in \mathbb{R}$, and for simplicity, $\beta = 1$. Mardia and Frelsen in [16] compared the properties of these three submodels in (2), (4), and (6). The multivariate extensions of the sine model can be found in [17]. In another attempt to expand the platform of toroidal distributions, Wehrly and Johnson in [18] used a marginal specification approach to construct bivariate models with more flexible specified circular marginals. Later, Jones et al. in [19] obtained various toroidal models using the general form in [18]. In this way, Fernández-Durán in [20] proposed another general toroidal model by using a *copula* pdf that García-Portugués imposed periodic restrictions on in [21], and Jones et al. [19] defined

it as a *circula* pdf, arguing that it is characterised by a circular uniform distribution. For more details, see [22].

The main incentive for defining toroidal models in recent years has been the demand from other sciences, especially bioinformatics, to model dihedral angles in order to analyse protein structures ([13,14,23,24]). However, toroidal data can also be observed in other fields, for example, in meteorology (wind directions at two different times of day) and medicine (peak systolic blood pressure during two separate time periods). For the interested reader, some applications of toroidal models can be found in [25–29].

Most of the proposed toroidal models are pointwise symmetric, whereas the data that they model usually represent asymmetric patterns. This inspired Ameijeiras-Alonso and Ley [24] to introduce bivariate sine-skewed distributions (BSS):

$$f_{BSS}(\theta_1, \theta_2) = f(\theta_1 - \mu_1, \theta_2 - \mu_2)(1 + \lambda_1 \sin(\theta_1 - \mu_1) + \lambda_2 \sin(\theta_2 - \mu_2)), \tag{7}$$

where $f(.,.)$ is a toroidal density symmetric (pointwise) about $-\pi \leq \mu_1, \mu_2 < \pi$, and the skewness parameters $-1 \leq \lambda_i \leq 1, i = 1, 2$, satisfy $|\lambda_1| + |\lambda_2| \leq 1$.

In this paper, Möbius transformation will form the foundation for the construction of competitive models. A map $T : \mathbb{C} \rightarrow \mathbb{C}$ is a Möbius transformation if it has the following form:

$$T(z) = \frac{az + b}{cz + d},$$

where \mathbb{C} is the set of complex numbers, $a, b, c, d \in \mathbb{C}$ are complex numbers, and $ad - bc \neq 0$. Let $S \subset \mathbb{C}$ be a unit circle, then Möbius transformation maps a point on the unit circle θ onto another $\tilde{\theta}$. Jones in [30] subsequently applied the Möbius transformation to introduce a new family of distributions on the disc. Kato and Jones in [31] used the Möbius transformation to introduce a new distribution on the circle by transforming the von Mises distribution. Wang and Shimizu in [32] applied the Möbius transformation to cardioid random variables. Kato and Pewsey in [33] employed this transformation to define the unimodal bivariate wrapped Cauchy distribution by transforming the bivariate circular distribution in [34]:

$$f(\theta_1, \theta_2) = c(c_0 - c_1 \cos(\theta_1 - \mu_1) - c_2 \cos(\theta_2 - \mu_2) - c_3 \cos(\theta_1 - \mu_1) \cos(\theta_2 - \mu_2) - c_4 \sin(\theta_1 - \mu_1) \sin(\theta_2 - \mu_2))^{-1}, \tag{8}$$

where $c = (1 - \rho^2)(1 - r_1^2)(1 - r_2^2)/4\pi^2$, $c_0 = (1 + \rho^2)(1 + r_1^2)(1 + r_2^2) - 8|\rho|r_1r_2$, $c_1 = 2(1 + \rho^2)r_1(1 + r_2^2) - 4|\rho|(1 + r_1^2)r_2$, $c_2 = 2(1 + \rho^2)r_2(1 + r_1^2) - 4|\rho|(1 + r_2^2)r_1$, $c_3 = -4(1 + \rho^2)r_1r_2 + 2|\rho|(1 + r_1^2)(1 + r_2^2)$, $c_4 = 2\rho(1 - r_1^2)(1 - r_2^2)$, $\mu_1, \mu_2 \in [-\pi, \pi)$, $r_1, r_2 \in [0, 1)$, and $-1 < \rho < 1$. Kato and McCullagh in [35] introduced the Cauchy distribution on the sphere by using a Möbius transformation.

1.2. Our Contribution

In this paper, two new distributions are introduced on the torus by applying a restricted version of the Möbius transformation developed by Kato and Pewsey in [33], namely the circular Möbius transformation that transforms θ into $\tilde{\theta}$ through the following mapping:

$$\theta = \mathcal{U}(\tilde{\theta}, \mu, \nu, r) = \mu + \nu + 2 \arctan \left\{ \frac{1-r}{1+r} \tan \left(\frac{\tilde{\theta} - \nu}{2} \right) \right\}, \tag{9}$$

where $-\pi \leq \mu, \nu \leq \pi$, $r \in [0, 1)$, and μ is the rotation parameter. When $\mu = 0$, ν and r attract the point θ towards ν . By increasing r , the concentration of the points around ν increases. If $r = 0$, the transformation is identity mapping, and when $r \rightarrow 1$, $\mathcal{U}(\tilde{\theta}, \mu, \nu, r)$

tends to ν . More details about the circular Möbius transformation can be found in [29,36]. The inverse of (9) can be obtained as follows:

$$\tilde{\theta} = \nu + 2 \arctan \left\{ \frac{1+r}{1-r} \tan \left(\frac{\theta - \mu - \nu}{2} \right) \right\}.$$

More specifically, our novel contribution includes the following highlights:

- New Möbius transformation-induced toroidal distributions are developed, acting as alternatives for existing models and efficiently outperforming them in the data application in this paper;
- The proposed distributions reflect the protein structure more accurately than the existing models and can serve as proposal distributions for MCMC sampling of proteins since we should incorporate protein structure information into proposal distributions to obtain more accurate results;
- Sine-skewed versions of these proposed models are introduced to meet the increasing demand for the modelling of asymmetric toroidal data;
- The marginals of the new models lead to new multimodal circular distributions.

The remainder of this paper is organised as follows. Section 2 introduces two new distributions emanating from the sine and cosine models in (2) and (4), respectively. Section 3 introduces the sine-skewed versions of the newly proposed transformed sine and cosine models. Section 4 outlines the maximum likelihood method for obtaining the parameter estimates for the proposed models. Three real datasets, including information on angles in protein structures, are analysed in Section 5 to determine the performance of the proposed models relative to known competitors, and demonstrate their well-deserved designation as possible models for toroidal data. In Section 6, a simulation study is conducted for two reasons: (1) to explore the best method of generating samples from the newly transformed sine and cosine models, and (2) to evaluate the numerical method, followed by the acquisition of the maximum likelihood estimates (MLEs) of the parameters.

2. Two New Models on the Torus

This section highlights two new flexible models for toroidal data, obtained by transforming the sine and cosine models in (2) and (4) via a Möbius transformation.

2.1. Transformed Cosine Model

Let $(\tilde{\Theta}_1, \tilde{\Theta}_2)$ have pdf (4) with $\nu_1 = \nu_2 = 0$. Suppose that

$$(\Theta_1, \Theta_2) = (\mathcal{U}(\tilde{\Theta}_1, \mu_1, \nu_1, r_1), \mathcal{U}(\tilde{\Theta}_2, \mu_2, \nu_2, r_2)),$$

where $\mathcal{U}(\cdot)$ is defined in (9), $\mu_1, \mu_2, \nu_1, \nu_2 \in (-\pi, \pi]$, $r_1, r_2 \in [0, 1)$ and without loss of generality $\nu_1 = \nu_2 = 0$. Then, (Θ_1, Θ_2) has a pdf of

$$f(\theta_1, \theta_2) = \frac{C(1-r_1^2)(1-r_2^2)}{(1+r_1^2-2r_1 \cos(\theta_1-\mu_1))(1+r_2^2-2r_2 \cos(\theta_2-\mu_2))} \times \exp \left\{ \frac{1}{(1+r_1^2-2r_1 \cos(\theta_1-\mu_1))(1+r_2^2-2r_2 \cos(\theta_2-\mu_2))} (C_0 + C_1 \cos(\theta_1-\mu_1) + C_2 \cos(\theta_2-\mu_2) + C_3 \cos(\theta_1-\mu_1) \cos(\theta_2-\mu_2) + C_4 \sin(\theta_1-\mu_1) \sin(\theta_2-\mu_2)) \right\}, \tag{10}$$

where $\kappa_1, \kappa_2 \geq 0, \kappa_3 \in \mathbb{R}, C$ is defined in (5), and

$$\begin{aligned}
 C_0 &= -2\kappa_1 r_1(1+r_1^2) - 2\kappa_2 r_2(1+r_2^2) - 4\kappa_3 r_1 r_2, \\
 C_1 &= \kappa_1(1+r_1^2)(1+r_2^2) + 2\kappa_3 r_2(1+r_1^2) + 4\kappa_2 r_1 r_2, \\
 C_2 &= \kappa_2(1+r_1^2)(1+r_2^2) + 2\kappa_3 r_1(1+r_2^2) + 4\kappa_1 r_1 r_2, \\
 C_3 &= -2\kappa_1 r_2(1+r_1^2) - 2\kappa_2 r_1(1+r_2^2) - \kappa_3(1+r_1^2)(1+r_2^2), \\
 C_4 &= -\kappa_3(1-r_1^2)(1-r_2^2),
 \end{aligned}
 \tag{11}$$

where $\mu_1, \mu_2 \in [-\pi, \pi)$ are location parameters, $\kappa_1, \kappa_2 \geq 0$ are concentration parameters, κ_3 is the circular–circular dependence parameter, and r_1 and r_2 regulate the concentrations of the marginal distributions. In (10), when $r_1 = r_2 = 0$, the cosine model (4) is obtained. If $\kappa_1, \kappa_2, \kappa_3 = 0$ yields the bivariate wrapped Cauchy distribution, then $\theta_1 \perp \theta_2$ follows. The pdf and contour plots of (10) are shown in Figure 2 for $\mu_1 = \mu_2 = 0$ and different values of $\kappa_1, \kappa_2, \kappa_3, r_1$ and r_2 , and reveal unimodal and bimodal behaviour.

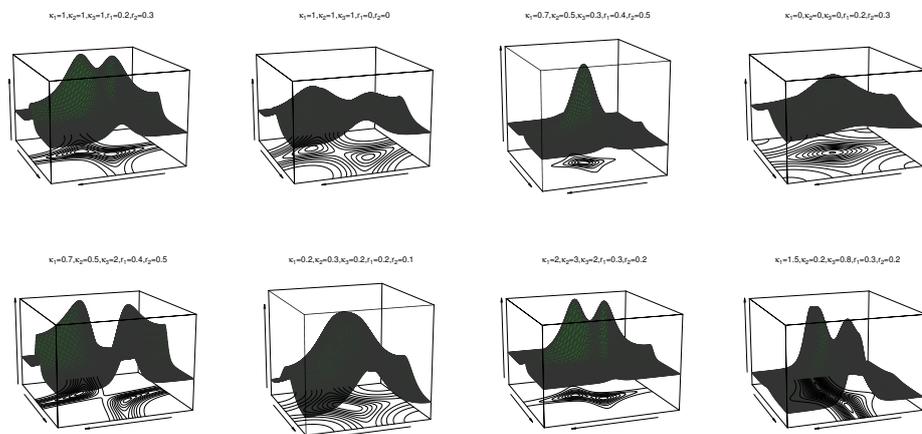


Figure 2. Pdf and contour plots of the transformed cosine model (10) for $\mu_1 = \mu_2 = 0$ and different values of $\kappa_1, \kappa_2, \kappa_3, r_1$, and r_2 .

Proposition 1. Assuming the transformed cosine model (10), when $r_1, r_2 \rightarrow 0$, then (Θ_1, Θ_2) has approximately a bivariate normal distribution if and only if $\kappa_3 \leq \frac{\kappa_1 \kappa_2}{\kappa_1 + \kappa_2}$.

Proof. See Appendix A. \square

In the following, the marginal pdf and conditional pdf of the transformed cosine model (10) and their properties are discussed. The marginal pdf of θ_1 for the transformed cosine model in (10) is as follows:

$$f_{\Theta_1}(\theta_1) = \frac{2\pi C(1-r_1^2)}{1+r_1^2-2r_1 \cos(\theta_1-\mu_1)} I_0(h(\theta_1)) \exp\left\{ \frac{\kappa_1(1+r_1^2) \cos(\theta_1-\mu_1) - 2\kappa_1 r_1}{1+r_1^2-2r_1 \cos(\theta_1-\mu_1)} \right\},
 \tag{12}$$

where

$$h(\theta_1) = \left\{ \kappa_2^2 + \kappa_3^2 - \frac{2\kappa_2 \kappa_3 ((1+r_1^2) \cos(\theta_1-\mu_1) - 2r_1)}{1+r_1^2-2r_1 \cos(\theta_1-\mu_1)} \right\}^{1/2},
 \tag{13}$$

and C is as defined in (5). The marginal pdf of Θ_1 in (12) is symmetric to μ_1 , small values of κ_3 approximate the transformed von Mises distribution [31], and $r_1 = 0$, which simplifies to the marginal pdf of the cosine model [14]. It is clear that for $r_1 = 0$ and small values of κ_3 , the von Mises distribution is approximated. If $\kappa_1 = \kappa_2 = \kappa_3 = 0$ in (12), then the Möbius-transformed uniform distribution is obtained. For $\kappa_1 = \kappa_2 = \kappa_3 = r_1 = 0$, the distribution is uniform. When $\kappa_1 = \kappa_2 = 0$ in (12), the distribution is the transformed von Mises distribution [31], and when $\kappa_1 = \kappa_2 = r_1 = 0$, the von Mises distribution is

obtained. The plots of this generalized marginal pdf of Θ_1 are shown in Figure 3 (left) for $\mu_1 = 0$ and different values of $\kappa_1, \kappa_2, \kappa_3$ and r_1 , reflecting unimodal and bimodal graphs. In the following theorem, the modality of the marginal density function Θ_1 is addressed.

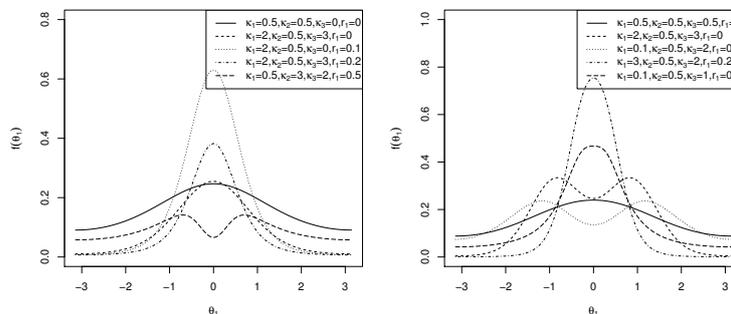


Figure 3. Plots of the marginal pdf of Θ_1 in (12) (left) and in (18) (right) for $\mu_1 = 0$ and different parameter values.

Corollary 1. *The marginal distribution of Θ_1 in (12) is symmetric around $\theta_1 = \mu_1$ and unimodal (with mode at μ_1) if and only if $\frac{A(|\kappa_2 - \kappa_3|)}{|\kappa_2 - \kappa_3|} \leq \left(\frac{2r_1(1-r_1)^2}{(1-r_1^2)^2} + \kappa_1 \right) / \kappa_2 \kappa_3$, where $A(\kappa) = I_1(\kappa) / I_0(\kappa)$. Moreover, the marginal distribution of Θ_1 in (12) is bimodal (with the modes at $\mu_1 - \theta_1^*$ and $\mu_1 + \theta_1^*$) if and only if $\frac{A(|\kappa_2 - \kappa_3|)}{|\kappa_2 - \kappa_3|} > \left(\frac{2r_1(1-r_1)^2}{(1-r_1^2)^2} + \kappa_1 \right) / \kappa_2 \kappa_3$, and θ_1^* is the root of $\kappa_2 \kappa_3 (1 - r_1^2)^2 A(h(\theta_1^*)) / h(\theta_1^*) - 2r_1(1 + r_1^2 - 2r_1 \cos(\theta_1^* - \mu_1)) - \kappa_1(1 - r_1^2)^2 = 0$, where $h(\theta)$ is as defined in (13).*

Proof. See Appendix A. \square

The conditional pdf $f(\theta_2 | \Theta_1 = \theta_1)$ results in the transformed von Mises distribution [31] given by the following:

$$f(\theta_2 | \Theta_1 = \theta_1) = \frac{1 - r_2^2}{2\pi I_0(h(\theta_1))} \frac{1}{1 + r_2^2 - 2r_2 \cos(\theta_2 - \mu_2)} \times \exp \left\{ \frac{h(\theta_1) \cos \tau ((1 + r_2^2) \cos(\theta_2 - \mu_2) - 2r_2) + h(\theta_1) \sin \tau (1 - r_2^2) \sin(\theta_2 - \mu_2)}{1 + r_2^2 - 2r_2 \cos(\theta_2 - \mu_2)} \right\}, \tag{14}$$

where $h(\theta_1)$ is as defined in (13), and

$$\tan \tau = \frac{-\kappa_3(1 - r_1^2) \sin(\theta_1 - \mu_1)}{\kappa_2(1 + r_1^2 - 2r_1 \cos(\theta_1 - \mu_1)) - \kappa_3((1 + r_1^2) \cos(\theta_1 - \mu_1) - 2r_1)} \tag{15}$$

Note that for $r_1 = r_2 = 0$, (14) simplifies to the von Mises distribution with the parameters τ and $h(\theta_1)$ in (13).

2.2. Transformed Sine Model

Let $(\tilde{\Theta}_1, \tilde{\Theta}_2)$ have a bivariate pdf (2), with $\nu_1 = \nu_2 = 0$. Suppose that

$$(\Theta_1, \Theta_2) = (\mathcal{U}(\tilde{\Theta}_1, \mu_1, \nu_1, r_1), \mathcal{U}(\tilde{\Theta}_2, \mu_2, \nu_2, r_2)),$$

where $\mathcal{U}(\cdot)$ is as defined in (9), $\mu_1, \mu_2, \nu_1, \nu_2 \in (-\pi, \pi)$, $r_1, r_2 \in [0, 1)$, and without loss of generality $\nu_1 = \nu_2 = 0$. Then, (Θ_1, Θ_2) has a pdf as follows:

$$f(\theta_1, \theta_2) = \frac{C(1 - r_1^2)(1 - r_2^2)}{(1 + r_1^2 - 2r_1 \cos(\theta_1 - \mu_1))(1 + r_2^2 - 2r_2 \cos(\theta_2 - \mu_2))} \times \exp \left\{ \frac{1}{(1 + r_1^2 - 2r_1 \cos(\theta_1 - \mu_1))(1 + r_2^2 - 2r_2 \cos(\theta_2 - \mu_2))} (C_0 + C_1 \cos(\theta_1 - \mu_1) + C_2 \cos(\theta_2 - \mu_2) + C_3 \cos(\theta_1 - \mu_1) \cos(\theta_2 - \mu_2) + C_4 \sin(\theta_1 - \mu_1) \sin(\theta_2 - \mu_2)) \right\}, \tag{16}$$

where $\kappa_1, \kappa_2 \geq 0, \kappa_3 \in \mathbb{R}, C$ is as defined in (3), and

$$\begin{aligned} C_0 &= -2\kappa_1 r_1(1 + r_2^2) - 2\kappa_2 r_2(1 + r_1^2), \\ C_1 &= \kappa_1(1 + r_1^2)(1 + r_2^2) + 4\kappa_2 r_1 r_2, \\ C_2 &= \kappa_2(1 + r_1^2)(1 + r_2^2) + 4\kappa_1 r_1 r_2, \\ C_3 &= -2\kappa_1 r_2(1 + r_1^2) - 2\kappa_2 r_1(1 + r_2^2), \\ C_4 &= \kappa_3(1 - r_1^2)(1 - r_2^2), \end{aligned} \tag{17}$$

where $\mu_1, \mu_2 \in [-\pi, \pi)$ are location parameters, $\kappa_1, \kappa_2 \geq 0$ are concentration parameters, κ_3 is the circular–circular dependence parameter, and r_1 and r_2 regulate the concentrations of the marginal distributions. If $r_1 = r_2 = 0$ in (16), then the sine model in (2) follows. The pdf and contour plots of (16) are shown in Figure 4 for $\mu_1 = \mu_2 = 0$ and for different values of $\kappa_1, \kappa_2, \kappa_3, r_1$ and r_2 . As can be seen, this transformed sine pdf (16) can have both unimodal and bimodal forms.

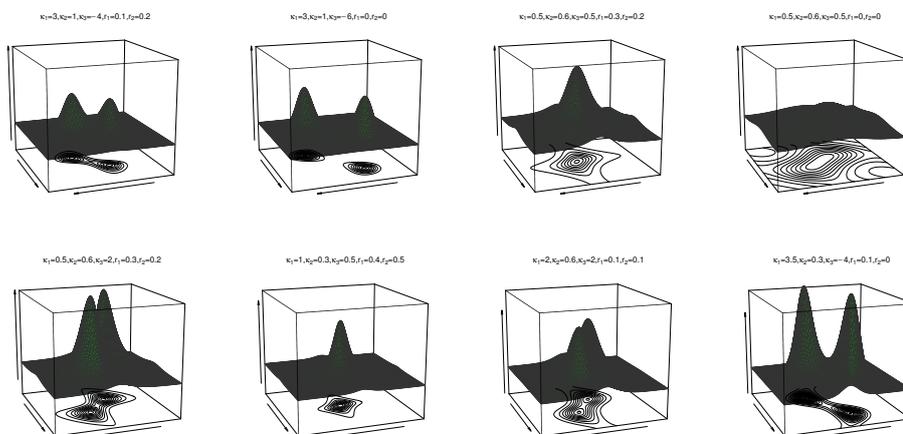


Figure 4. Pdf and contour plots of the transformed sine model (16) for $\mu_1 = \mu_2 = 0$ and different values of $\kappa_1, \kappa_2, \kappa_3, r_1$, and r_2 .

Proposition 2. Assuming the transformed sine model in (16), when $r_1, r_2 \rightarrow 0$, then (Θ_1, Θ_2) has an approximately bivariate normal distribution if and only if $\kappa_3^2 < \kappa_1 \kappa_2$.

Proof. Similarly, Theorem 1 is proved using the results in [13]. □

In this case, the marginal pdf of Θ_1 for the transformed sine model in (16) is as follows:

$$f_{\Theta_1}(\theta_1) = \frac{2\pi C(1 - r_1^2)}{1 + r_1^2 - 2r_1 \cos(\theta_1 - \mu_1)} I_0(h(\theta_1)) \exp \left\{ \frac{\kappa_1(1 + r_1^2) \cos(\theta_1 - \mu_1) - 2\kappa_1 r_1}{1 + r_1^2 - 2r_1 \cos(\theta_1 - \mu_1)} \right\}, \tag{18}$$

where

$$h(\theta_1) = \left\{ \kappa_2^2 + \left(\frac{\kappa_3(1 - r_1^2)}{1 + r_1^2 - 2r_1 \cos(\theta_1 - \mu_1)} \right)^2 \sin^2(\theta_1 - \mu_1) \right\}^{1/2}, \tag{19}$$

and C , as shown in (3). The marginal pdf of Θ_1 is symmetric around μ_1 . If $\kappa_3 = 0$, the distribution is the transformed von Mises distribution [31]. If $r_1 = 0$ in (18), the marginal distribution of the sine model [13] is obtained. The plots of the marginal pdf of Θ_1 in (18) are shown in Figure 3 (right) for $\mu_1 = 0$ and different values of $\kappa_1, \kappa_2, \kappa_3$, and r_1 . As can be seen, the distribution can be both unimodal and bimodal. In the following theorem, the modality of the marginal pdf of Θ_1 in (18) is explored.

Corollary 2. *The marginal distribution of Θ_1 in (18) is symmetric around $\theta_1 = \mu_1$ and unimodal (with mode at μ_1) if and only if $\frac{A(\kappa_2)}{\kappa_2} \leq \left(\frac{2r_1(1-r_1)^2}{(1-r_1^2)^2} + \kappa_1\right) / \kappa_3^2$, where $A(\kappa) = I_1(\kappa) / I_0(\kappa)$. Moreover, the marginal distribution of Θ_1 in (18) is bimodal (with the modes at $\mu_1 - \theta_1^*$ and $\mu_1 + \theta_1^*$) if and only if $\frac{A(\kappa_2)}{\kappa_2} > \left(\frac{2r_1(1-r_1)^2}{(1-r_1^2)^2} + \kappa_1\right) / \kappa_3^2$, and θ_1^* is the root of $\kappa_3^2(1-r_1^2)^2 \cos \theta_1^* A(h(\theta_1^*)) / h(\theta_1^*) - 2r_1(1+r_1^2 - 2r_1 \cos(\theta_1^* - \mu_1)) - \kappa_1(1-r_1^2)^2 = 0$, where $h(\theta)$ is as defined in (19).*

Proof. See Appendix A. \square

The conditional pdf $f(\theta_2 | \Theta_1 = \theta_1)$ is given by:

$$f(\theta_2 | \Theta_1 = \theta_1) = \frac{1-r_2^2}{2\pi I_0(h(\theta_1))} \frac{1}{1+r_2^2-2r_2 \cos(\theta_2-\mu_2)} \times \exp\left\{\frac{h(\theta_1) \cos \tau((1+r_2^2) \cos(\theta_2-\mu_2) - 2r_2) + h(\theta_1) \sin \tau(1-r_2^2) \sin(\theta_2-\mu_2)}{1+r_2^2-2r_2 \cos(\theta_2-\mu_2)}\right\}, \tag{20}$$

where $h(\theta_1)$ is as defined in (19), and

$$\tan \tau = \frac{\kappa_3}{\kappa_2} \frac{(1-r_1^2) \sin(\theta_1-\mu_1)}{1+r_1^2-2r_1 \cos(\theta_1-\mu_1)} \tag{21}$$

Interestingly, the conditional distribution is the transformed von Mises distribution [31]. When $r_1 = r_2 = 0$ in (20), the von Mises distribution with parameters τ and $h(\theta_1)$ is obtained.

3. Sine-Skewed Transformed Sine and Cosine Distributions

In practice, it is possible to have skewed toroidal datasets, despite the well-known toroidal distributions being pointwise symmetric. Therefore, it would be interesting to extend this methodology to the recent model of Ameijeiras-Alonso and Ley in [24]. In this section, the skewed versions of the proposed transformed sine and cosine models in (16) and (10) are introduced. In addition, Abe and Pewsey’s skew model in [37] is applied to extend models on the circle manifold using marginal density functions.

By substituting (10) in (7), the *sine-skewed transformed cosine (BSSTC)* distribution can be defined as follows:

$$f_{BSSTC}(\theta_1, \theta_2) = \frac{C(1-r_1^2)(1-r_2^2)}{(1+r_1^2-2r_1 \cos(\theta_1-\mu_1))(1+r_2^2-2r_2 \cos(\theta_2-\mu_2))} \times \exp\left\{\frac{1}{(1+r_1^2-2r_1 \cos(\theta_1-\mu_1))(1+r_2^2-2r_2 \cos(\theta_2-\mu_2))} (C_0 + C_1 \cos(\theta_1-\mu_1) + C_2 \cos(\theta_2-\mu_2) + C_3 \cos(\theta_1-\mu_1) \cos(\theta_2-\mu_2) + C_4 \sin(\theta_1-\mu_1) \sin(\theta_2-\mu_2))\right\} \times (1 + \lambda_1 \sin(\theta_1 - \mu_1) + \lambda_2 \sin(\theta_2 - \mu_2)), \tag{22}$$

where $\kappa_1, \kappa_2 \geq 0, \kappa_3 \in \mathbb{R}, C$ is as defined in (5), and C_0-C_4 are as defined in (11). The pdf and contour plots of the sine-skewed transformed cosine model for $\kappa_1 = 0.2, \kappa_2 = 0.3, \kappa_3 = 0.2, r_1 = 0.2, r_2 = 0.1, \mu_1 = \mu_2 = 0$, and different values of λ_1 and λ_2 are shown in Figure 5 (top).

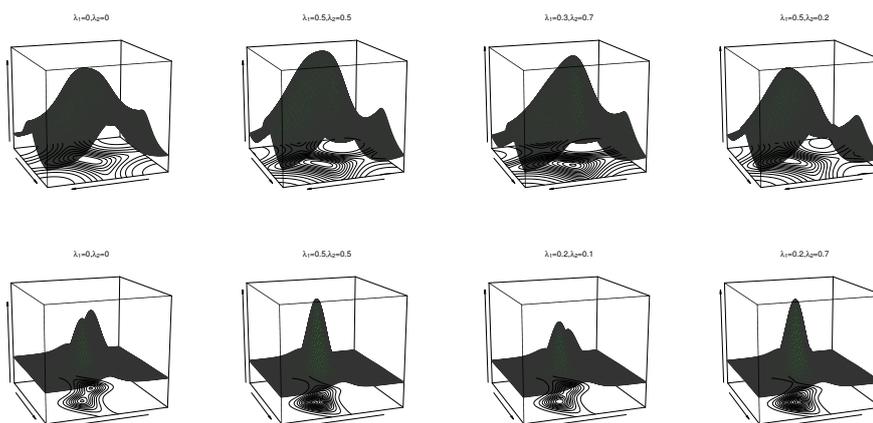


Figure 5. Pdf and contour plots of the sine-skewed transformed cosine model in (22) (top) and the sine-skewed transformed sine model in (24) (bottom) for different values of λ_1 and λ_2 .

The marginal pdf of θ_1 for BSSTC in (22) is as follows:

$$f_{\Theta_1;BSS}(\theta_1) = \frac{C(1 - r_1^2)}{1 + r_1^2 - 2r_1 \cos(\theta_1 - \mu_1)} \exp\left\{ \frac{\kappa_1(1 + r_1^2) \cos(\theta_1 - \mu_1) - 2\kappa_1 r_1}{1 + r_1^2 - 2r_1 \cos(\theta_1 - \mu_1)} \right\} \quad (23)$$

$$\times \left\{ 2\pi I_0(h(\theta_1)) \left(1 + \lambda_1(1 - r_1^2) \sin(\theta_1 - \mu_1) / \left(1 + r_1^2 - 2r_1 \cos(\theta_1 - \mu_1) \right) \right) \right.$$

$$\left. + \lambda_2 A(h(\theta_1)) \cos(\tau + \mu_2) \right\}$$

where $h(\theta_1)$ and τ are obtained from (13) and (15), respectively. When $\lambda_2 = 0$, $f_{\Theta_1;BSS}(\theta_1)$ is the Möbius-transformed sine-skewed version [37] of the marginal pdf of the cosine model. The plots of the skewed pdf in (23) are shown in Figure 6 (left) for $\mu_1 = \mu_2 = 0$ and different values of $\kappa_1, \kappa_2, \kappa_3, r_1, \lambda_1$, and λ_2 . As can be observed, the distribution can be both unimodal and bimodal.

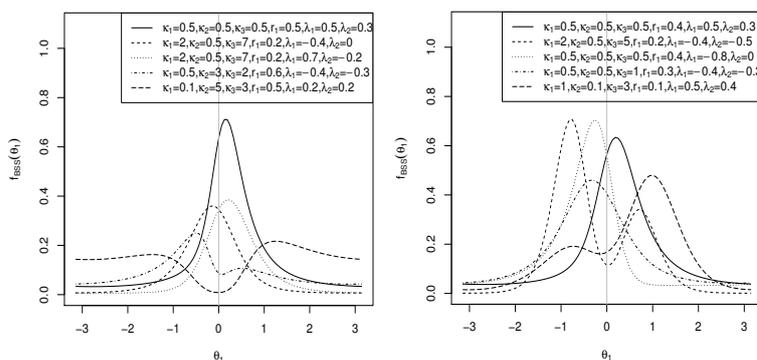


Figure 6. Plots of the marginal pdf of Θ_1 for BSSTC (left) and BSSTS (right) for $\mu_1 = \mu_2 = 0$ and different parameter values.

Similarly, from (16) and (7), the sine-skewed transformed sine (BSSTS) distribution can be obtained as follows:

$$f_{BSSTS}(\theta_1, \theta_2) = \frac{C(1 - r_1^2)(1 - r_2^2)}{(1 + r_1^2 - 2r_1 \cos(\theta_1 - \mu_1))(1 + r_2^2 - 2r_2 \cos(\theta_2 - \mu_2))} \quad (24)$$

$$\times \exp\left\{ \frac{1}{(1 + r_1^2 - 2r_1 \cos(\theta_1 - \mu_1))(1 + r_2^2 - 2r_2 \cos(\theta_2 - \mu_2))} (C_0 + C_1 \cos(\theta_1 - \mu_1) \right.$$

$$+ C_2 \cos(\theta_2 - \mu_2) + C_3 \cos(\theta_1 - \mu_1) \cos(\theta_2 - \mu_2) + C_4 \sin(\theta_1 - \mu_1) \sin(\theta_2 - \mu_2)) \left. \right\}$$

$$\times (1 + \lambda_1 \sin(\theta_1 - \mu_1) + \lambda_2 \sin(\theta_2 - \mu_2)),$$

where $\kappa_1, \kappa_2 \geq 0, \kappa_3 \in \mathbb{R}, C$ is as defined in (3), and C_0-C_4 are defined in (17). The pdf and contour plots of the sine-skewed transformed sine model for $\kappa_1 = 2, \kappa_2 = 0.6, \kappa_3 = 2, r_1 = 0.1, r_2 = 0.1, \mu_1 = \mu_2 = 0$, and different values of λ_1 and λ_2 are shown in Figure 5 (bottom).

The marginal pdf of θ_1 for BSSTS is of the same density as in (23), where $h(\theta_1)$ and τ are obtained from (19) and (21). When $\lambda_2 = 0, f_{\Theta_1;BSS}(\theta_1)$ is the Möbius-transformed sine-skewed version [37] of the marginal pdf of the sine model. The plots of the skewed pdf in (23) are shown in Figure 6 (right) for $\mu_1 = \mu_2 = 0$ and different values of $\kappa_1, \kappa_2, \kappa_3, r_1, \lambda_1$, and λ_2 . Figure 6 illustrates that the distribution can have both unimodal and bimodal forms.

To expand the skewed circular models, the following models are introduced based on the k sine-skewed model of [37]. The skewed version of the marginal distribution of Θ_1 in (12) is the following:

$$f_{SS}(\theta_1) = \frac{2\pi C(1 - r_1^2)}{1 + r_1^2 - 2r_1 \cos(\theta_1 - \mu_1)} I_0(h(\theta_1)) \exp\left\{ \frac{\kappa_1(1 + r_1^2) \cos(\theta_1 - \mu_1) - 2\kappa_1 r_1}{1 + r_1^2 - 2r_1 \cos(\theta_1 - \mu_1)} \right\} \times (1 + \lambda \sin(k(\theta_1 - \mu_1))), \tag{25}$$

where C is as defined in (5), $h(\theta_1)$ is as defined in (13), and $-1 \leq \lambda \leq 1$. $\lambda > 0$ leads to left-skewed distributions, and $\lambda < 0$ provides right-skewed distributions. The plots of the skewed pdf in (25) are shown in Figure 7 (left) for $k = 1, \mu_1 = 0$, and different values of $\kappa_1, \kappa_2, \kappa_3, r_1$, and λ .

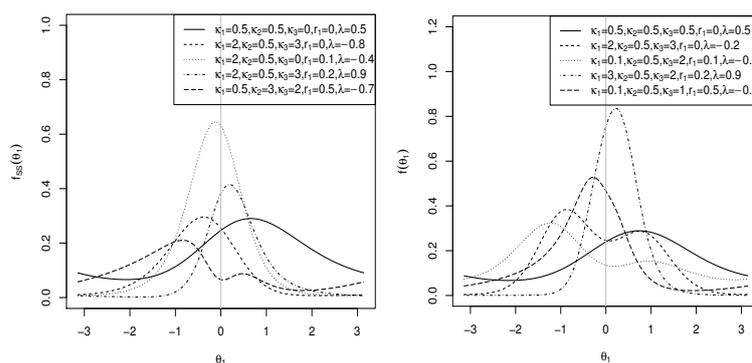


Figure 7. Plots of the sine-skewed versions of marginal pdfs of Θ_1 in (12) (left) and (18) (right) for $k = 1, \mu_1 = 0$ and different parameter values.

Similarly, the sine-skewed version [37] of the marginal pdf of Θ_1 in (18) is of the same density as in (25), where C is as defined in (3), $h(\theta_1)$ is as defined in (19), and $-1 \leq \lambda \leq 1$. The plots of the sine-skewed version of the marginal pdf in (18) are shown in Figure 7 (right) for $k = 1, \mu_1 = 0$, and different values of $\kappa_1, \kappa_2, \kappa_3, r_1$, and λ . As can be seen, the distribution is both unimodal and bimodal. Multimodal results for $k > 1$.

4. Maximum Likelihood Estimation

In this section, the maximum likelihood method is outlined to obtain the estimates of parameters for both the transformed cosine and sine models. Suppose that $\zeta = (\mu_1, \mu_2, \kappa_1, \kappa_2, \kappa_3, r_1, r_2)^T$ are the parameters associated with the transformed cosine model (10). The log-likelihood function of the transformed cosine model is represented as follows:

$$\begin{aligned}
 l(\zeta) = & n \log C + n \log(1 - r_1^2) + n \log(1 - r_2^2) - \sum_{i=1}^n \log(1 + r_1^2 - 2r_1 \cos(\theta_{1i} - \mu_1)) \\
 & - \sum_{i=1}^n \log(1 + r_2^2 - 2r_2 \cos(\theta_{2i} - \mu_2)) + \sum_{i=1}^n \frac{1}{(1 + r_1^2 - 2r_1 \cos(\theta_{1i} - \mu_1))(1 + r_2^2 - 2r_2 \cos(\theta_{2i} - \mu_2))} \\
 & \times (C_0 + C_1 \cos(\theta_{1i} - \mu_1) + C_2 \cos(\theta_{2i} - \mu_2) + C_3 \cos(\theta_{1i} - \mu_1) \cos(\theta_{2i} - \mu_2) \\
 & + C_4 \sin(\theta_{1i} - \mu_1) \sin(\theta_{2i} - \mu_2)),
 \end{aligned} \tag{26}$$

where C is as defined in (5), and C_0 – C_4 are as defined in (11). The MLE of the parameters, $\hat{\zeta} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\kappa}_1, \hat{\kappa}_2, \hat{\kappa}_3, \hat{r}_1, \hat{r}_2)^T$, can be determined by maximizing (26) with respect to $\zeta = (\mu_1, \mu_2, \kappa_1, \kappa_2, \kappa_3, r_1, r_2)^T$.

Supposing that $\zeta = (\mu_1, \mu_2, \kappa_1, \kappa_2, \kappa_3, r_1, r_2)^T$ are the parameters associated with the transformed sine model (16), the log-likelihood function of the transformed sine model can be represented as follows:

$$\begin{aligned}
 l(\zeta) = & n \log C + n \log(1 - r_1^2) + n \log(1 - r_2^2) - \sum_{i=1}^n \log(1 + r_1^2 - 2r_1 \cos(\theta_{1i} - \mu_1)) \\
 & - \sum_{i=1}^n \log(1 + r_2^2 - 2r_2 \cos(\theta_{2i} - \mu_2)) + \sum_{i=1}^n \frac{1}{(1 + r_1^2 - 2r_1 \cos(\theta_{1i} - \mu_1))(1 + r_2^2 - 2r_2 \cos(\theta_{2i} - \mu_2))} \\
 & \times (C_0 + C_1 \cos(\theta_{1i} - \mu_1) + C_2 \cos(\theta_{2i} - \mu_2) + C_3 \cos(\theta_{1i} - \mu_1) \cos(\theta_{2i} - \mu_2) \\
 & + C_4 \sin(\theta_{1i} - \mu_1) \sin(\theta_{2i} - \mu_2)),
 \end{aligned} \tag{27}$$

where C is as defined in (3), and C_0 – C_4 are as defined in (17). The maximization of (27) with respect to $\zeta = (\mu_1, \mu_2, \kappa_1, \kappa_2, \kappa_3, r_1, r_2)^T$ results in the MLE of the parameters, $\hat{\zeta} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\kappa}_1, \hat{\kappa}_2, \hat{\kappa}_3, \hat{r}_1, \hat{r}_2)^T$.

By setting the partial derivatives of the log-likelihood functions in (26) and (27) with respect to ζ to zero, the MLEs of $\zeta = (\mu_1, \mu_2, \kappa_1, \kappa_2, \kappa_3, r_1, r_2)^T$ can be derived for the transformed cosine and sine models. Given the fact that no closed-form expressions exist, it is necessary to use numerical methods to obtain the MLEs. Operationally, the maximization of (26) and (27) with respect to ζ is obtained by the DEoptim package in the R software [38] based on the differential evolution (DE) algorithm [39]. Extensive studies have validated its significant performance as a global optimization algorithm for continuous numerical minimization problems [40]. It is worth noting that this package was also used to obtain the MLEs of the parameters for sine-skewed versions and mixtures of transformed cosine and sine models.

5. Protein Structure Application

To demonstrate the performance of the proposed models in modelling the dihedral angles and the planar and torsion angles in a protein structure, three datasets are considered, which are available at <http://scop.mrc-lmb.cam.ac.uk/scop/>. SCOP.1 contains 10,188 planar and torsion angles (θ, τ) (see Figure 1A) for about 63 protein domains that were randomly selected from three remote protein classes in the structural classification of proteins (SCOP). SCOP.3 includes 4607 planar and torsion angles (θ, τ) from approximately 40 protein chains, and the TCBIG.VAL.right set consists of 2673 dihedral angles (ϕ, ψ) (see Figure 7B) [41]. The Ramachandran plots [1] for each dataset are presented in Figure 8. As can be seen, the datasets are at least bimodal, so bimodal or mixture distributions will be good choices for fitting.

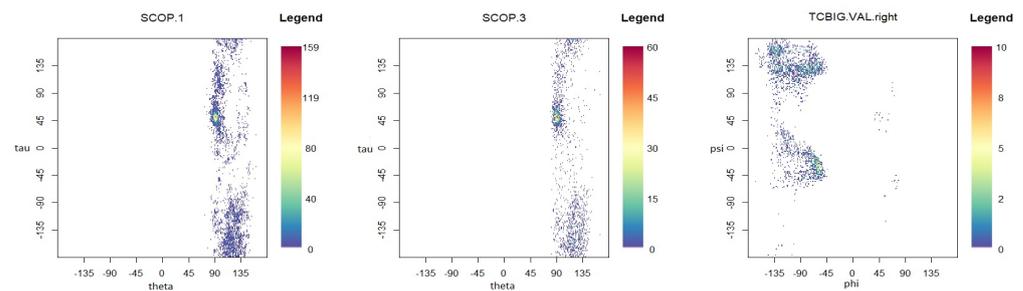


Figure 8. Ramachandran plots for each dataset.

The transformed sine and cosine models in (16) and (10), along with their competitors—the sine model, and a mixture of sine models (see (2); [13]), the cosine model, and a mixture of cosine models (see (4); [14]), and a mixture of bivariate wrapped Cauchy models (see (8); [33])—were fitted to the SCOP.1 and SCOP.3 datasets. A mixture distribution with two components was investigated as follows:

$$g_M(\theta_1, \theta_2) = pf_1(\theta_1, \theta_2) + (1 - p)f_2(\theta_1, \theta_2)$$

where $p \in [0, 1]$ and $f_1(\cdot, \cdot)$ and $f_2(\cdot, \cdot)$ are two toroidal distributions. The estimation of parameters, identifiability, and choosing the number of mixing components and parameters are among the well-known challenges in the application of mixture distributions. Furthermore, when the empirical density of the data is highly asymmetric, it can result in a misleading statistical inference of the parameters [42]. Multimodal distributions, which represent the random behaviour of data with multi-mode presence, can provide better model fitting. This is observed here using the bimodal transformed sine model.

The sine-skewed versions of the aforementioned distributions [24] form part of these evaluations. The results, including the MLEs of parameters, log-likelihood, Akaike information criterion (AIC), and the Bayesian information criterion (BIC), are shown in Tables 1 and 2. Based on these results, the bimodal transformed sine model in (16) provides the best fit for the data, and its performance is better than that of the mixture models for these datasets. Based on the symmetry test of Ameijeiras-Alonso and Ley in [24] and the values of log-likelihood in Tables 1 and 2, there is no evidence that rejects the fact that underlying distributions for SCOP.1 and SCOP.3 are pointwise symmetric. The results of the mixture of transformed sine and the mixture of transformed cosine models are not reported in Tables 1 and 2 because $\hat{p} \approx 1$. Scatter plots of the data, together with contour plots of the fitted distributions are provided in Figures 9 and 10.

With the last dataset TCBIG.VAL.right, good results are not observed upon application of the single component distributions. Therefore, a mixture model might offer a solution. Subsequently, only mixtures of the aforementioned distributions were considered. For comparison, goodness-of-fit was evaluated for mixtures of distributions from transformed sine and cosine models, and for mixtures of distributions from existing models. The results are listed in Table 3. As can be seen, the mixture of transformed sine models provides the best fitting of the data. Scatter plots of the data and contour plots of the fitted distributions are shown in Figure 11.

Table 1. Maximum likelihood estimates and corresponding log-likelihood, AIC, and BIC for SCOP.1 ($n = 10,188$).

Model	$\hat{\rho}$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\kappa}_3$	$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\rho}$	Log-Likelihood	AIC	BIC
Sine [13]	–	25.2085	0.3679	7.3700	–	–	1.8976	2.4624	–	–	–	–15,890.80	31,790.16	31,827.74
Sine-skewed sine [24]	–	18.8058	0.0852	4.8449	–	–	1.8701	–3.1415	0.4051	–0.3718	–	–18,089.71	36,193.42	36,244.02
Mixture of sine	–	4.9938	0.4603	2.3512	–	–	2.0560	2.5011	–	–	0.3476	–15,719.26	31,460.22	31,540.04
	–	0.0217	0.0413	–4.4594	–	–	1.0912	–1.8997	–	–	0.6524			
Cosine [14]	–	11.6274	6.7×10^{-17}	0.6507	–	–	1.8807	–0.8652	–	–	–	–19,919.04	39,848.07	39,884.22
Sine-skewed cosine [24]	–	11.6274	1.7×10^{-8}	0.6507	–	–	1.8807	–0.8651	–0.7557	0.0789	–	–19,919.04	39,852.07	39,902.68
Mixture of cosine [14]	–	9.6015	2.6459	0.0087	–	–	1.7967	0.8676	–	–	0.5266	–18,120.09	36,262.18	36,341.70
	–	8.4761	0.0820	2.3228	–	–	2.1309	0.9647	–	–	0.4734			
Mixture of bivariate wrapped Cauchy [33]	–0.2892	–	–	–	0.9551	0.5649	1.6129	1.5337	–	–	0.4463	–17,099.36	34,220.72	34,300.24
	–0.1289	–	–	–	0.8513	0.5433	2.1128	–2.6980	–	–	0.5537			
Transformed sine	–	2.1585	0.3489	3.1712	0.6036	0.0131	1.8573	2.4321	–	–	–	–15,558.98	31,131.97	31,182.56
Sine-skewed transformed sine	–	2.1582	0.3487	3.1712	0.6037	0.0131	1.8573	2.4321	–0.1894	0.0556	–	–15,558.98	31,135.97	31,201.02
Transformed cosine	–	4.5122	1.9×10^{-16}	2.7905	0.2632	0.4164	1.8806	–0.6888	–	–	–	–16,920.43	33,854.86	33,905.46
Sine-skewed transformed cosine	–	4.4704	4.2×10^{-5}	2.8185	0.2656	0.4228	1.8805	–0.6871	0.6225	–0.1849	–	–16,920.43	33,858.86	33,923.92

Table 2. Maximum likelihood estimates and corresponding log-likelihood, AIC, and BIC for SCOP3 ($n = 4607$).

Model	$\hat{\rho}$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\kappa}_3$	\hat{r}_1	\hat{r}_2	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}	Log-Likelihood	AIC	BIC
Sine [13]	–	27.0312	0.3243	8.0789	–	–	1.8810	2.4618	–	–	–	–6970.09	13,950.18	13,982.36
Sine-skewed sine [24]	–	26.8304	0.3224	8.0732	–	–	1.8960	2.4724	–0.4124	–0.0159	–	–6941.31	13,896.62	13,941.67
Mixture of sine	–	7.3842	2.0013	–6.3567	–	–	2.0918	–1.4321	–	–	0.6632	–6893.41	13,901.15	13,879.61
	–	2.8774	0.0347	–1.7125	–	–	1.9306	–1.1124	–	–	0.3368			
Cosine [14]	–	11.5883	3.8×10^{-16}	0.6404	–	–	1.8537	–0.9851	–	–	–	–9028.72	18,067.45	18,099.62
Sine-skewed cosine [24]	–	11.5883	5.0×10^{-9}	0.6404	–	–	1.8537	–0.9850	–0.0168	–0.6387	–	–9028.72	18,071.45	18,116.49
Mixture of cosine [14]	–	29.9375	1.9210	0.0213	–	–	1.6840	0.8043	–	–	0.5648	–6959.76	13,941.52	14,012.31
	–	17.3302	0.0211	1.9456	–	–	2.0575	0.8866	–	–	0.4352			
Mixture of bivariate wrapped Cauchy [33]	–0.2347	–	–	–	0.9169	0.5546	1.5969	1.1037	–	–	0.4712	–7137.52	14,297.04	14,367.83
	–0.1279	–	–	–	0.8388	0.5100	1.9792	–2.0869	–	–	0.5288			
Transformed sine	–	3.8755	0.3414	3.6786	0.4950	1.3×10^{-9}	1.8589	2.4490	–	–	–	–6905.08	13,824.17	13,869.22
Sine-skewed transformed sine	–	3.8764	0.3415	3.7066	0.4883	2.6×10^{-8}	1.8591	2.4491	–0.1544	0.0796	–	–6905.08	13,828.17	13,886.08
Transformed cosine	–	4.1351	2.4×10^{-16}	2.8283	0.2884	0.4183	1.8604	–0.6560	–	–	–	–7567.28	15,148.56	15,193.61
Sine-skewed transformed cosine	–	4.1350	6.4×10^{-10}	2.8283	0.2884	0.4183	1.8604	–0.6560	0.6868	–0.1567	–	–7567.27	15,152.56	15,210.46

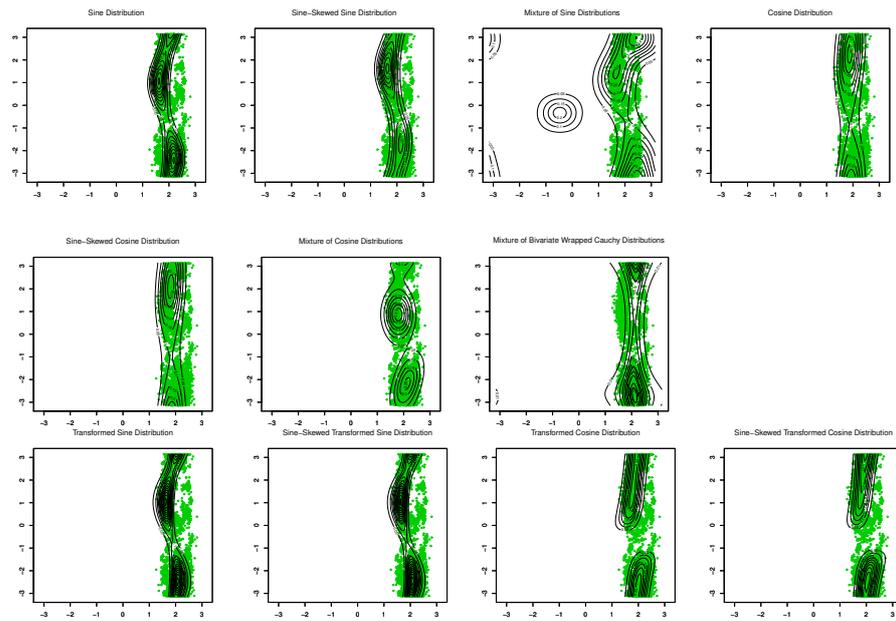


Figure 9. Contour plots of fitted pdfs together with scatter plot for SCOP.1 ($n = 10,188$). The last row includes the proposed models.

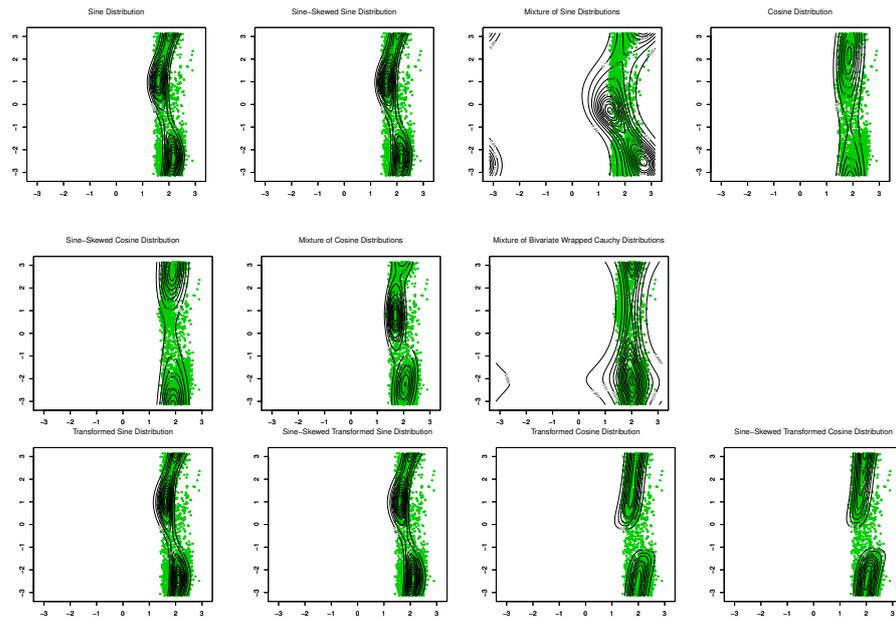


Figure 10. Contour plots of fitted pdfs together with scatter plot for SCOP.3 ($n = 4607$). The last row includes the proposed models.

Table 3. Maximum likelihood estimates and corresponding log-likelihood, AIC, and BIC for TCBIG.VAL.right ($n = 2673$).

Model	$\hat{\rho}$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{\kappa}_3$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\rho}$	Log-Likelihood	AIC	BIC
Mixture of sine	–	4.4364	7.6222	–1.2187	–	–	–1.7736	2.3336	0.6239	–4901.12	9824.25	9889.04
	–	5.4606	7.7290	–3.1154	–	–	–1.4197	–0.4111	0.3761			
Mixture of cosine [14]	–	4.2849	6.1824	4.6×10^{-6}	–	–	–1.4003	–0.4191	0.3787	–5005.01	10,032.03	10,096.82
	–	4.6268	7.8256	9.1×10^{-6}	–	–	–1.7785	2.3358	0.6213			
Mixture of bivariate wrapped Cauchy [33]	–0.3805	–	–	–	0.8545	0.8118	–1.1194	–0.4710	0.3108	–5283.42	10,588.85	10,653.64
	–0.0294	–	–	–	0.7038	0.7778	–1.8772	2.3037	0.6892			
Mixture of transformed sine	–	2.8274	7.3718	–3.0328	0.2930	0.0872	–1.3860	–0.4040	0.3515	–4826.80	9683.61	9771.96
	–	4.0949	1.6545	–0.7133	0.02450	0.4387	–1.7802	2.3495	0.6485			
Mixture of transformed cosine	–	2.3349	9.3385	0.0063	0.4909	0.1287	–1.1835	–0.5358	0.2645	–4882.14	9794.28	9882.64
–	3.9496	0.0117	0.8886	0.0553	0.8668	–1.8501	2.3841	0.7355				

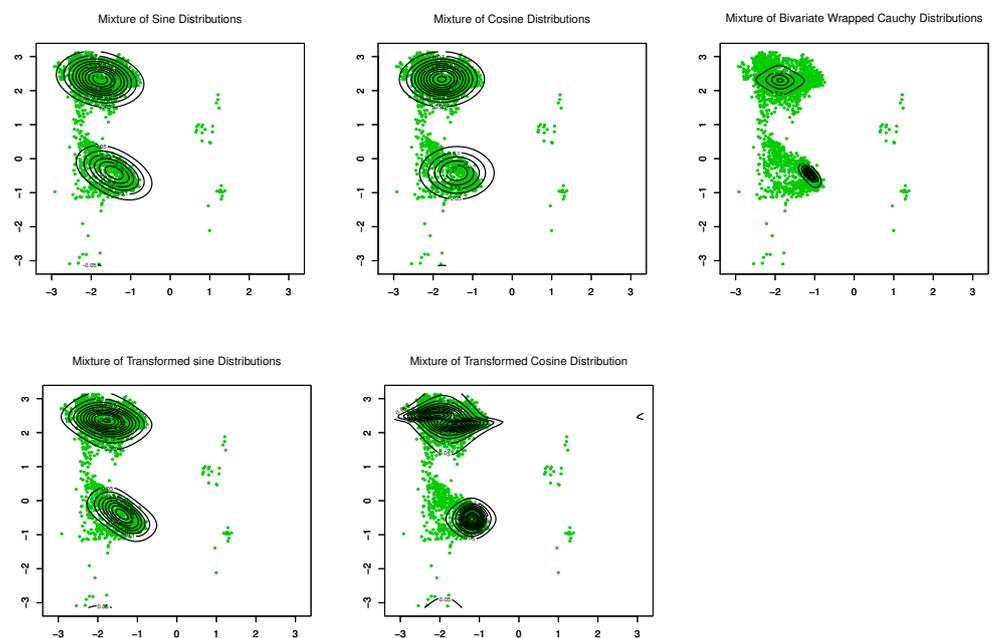


Figure 11. Contour plots of fitted pdfs together with scatter plots for TCBIG.VAL.right ($n = 2673$). The last row includes the proposed models.

The kernel density plots of the three datasets and the best-fit models obtained for each dataset are shown in Figure 12. According to the levels of contours in the kernel densities of the data and fitted curves, our proposed models provide an accurate fit.

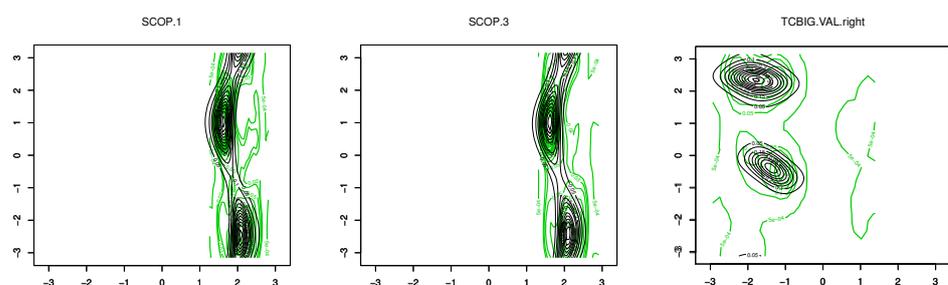


Figure 12. Kernel density plots of the data, and the best-fit models.

6. Simulation Study

The authors of Ref. [16] explored suitable methods for generating samples from cosine (with positive interaction) and sine models. They found that both Gibbs and rejection sampling approaches performed well, but the latter was more efficient. To simulate a sample from the newly proposed transformed sine and transformed cosine distributions in (16) and (10), four packages in R, which are generally based on rejection sampling, including MCMCpack [43], gibbs.met [44], LearnBayes [45], and MHadaptive [46], were used and the results were compared. These packages are based on Metropolis sampling, random walk Metropolis sampling, Metropolis-Hastings MCMC sampling, and Gibbs sampling with Metropolis steps. First, a sample of size $n = 1000$ was generated with each package from the transformed sine model in (16), with the parameters $\kappa_1 = 2.1585$, $\kappa_2 = 0.3489$, $\kappa_3 = 3.1712$, $r_1 = 0.6036$, $r_2 = 0.0131$, $\mu_1 = 1.8573$, and $\mu_2 = 2.4321$ (the best-fit model for the SCOP.1 dataset in the previous section). The results, including scatter plots of simulated samples with contour plots of the distribution, trace plots, and compare-partial plots [47], which use the last 10 percent of the chain, are shown in Figure 13. The runtime of each method is shown in Figure 14 (left) for a sample size of $n = 100$ [48] (system: Intel(R) Core(TM) i7-8550U CPU @ 1.80 GHz RAM 8.00 GB). Second, the MLE of the parameters and bias and the mean squared error (MSE) of the estimates were calculated for each method using the Monte Carlo method, with 500 replications and $n = 1,001,000$. The results are listed in Table 4.

Similarly, for the transformed cosine model in (10) with parameters $\kappa_1 = 3.9891$, $\kappa_2 = 0.6532$, $\kappa_3 = 1.7911$, $r_1 = 0.2305$, $r_2 = 0.5046$, $\mu_1 = -1.5651$, and $\mu_2 = 0.9878$, the aforementioned R packages were applied, first to generate a sample size of $n = 1000$. The results, including scatter plots of simulated samples with contour plots of the distribution, trace plots, and compare-partial plots [47], are shown in Figure 15. The runtime of each method is presented in Figure 14 (right) for a sample size of $n = 100$ [48]. Then, the MLE of the parameters and bias and the MSE of the estimates were calculated for each method using the Monte Carlo method, with 500 replications and $n = 100,1000$. The results are listed in Table 4, which support the performance of the selected approach for obtaining the MLEs of parameters. As shown in Figure 14, the MCMCmetrop1R is the highest-speed method, and gibbs_met is the lowest-speed method. According to the results in Table 4, rejection sampling provides accurate results. Gibbs sampling with Metropolis steps (gibbs_met) is also precise despite the low speed. With increasing n , bias and MSE decrease.

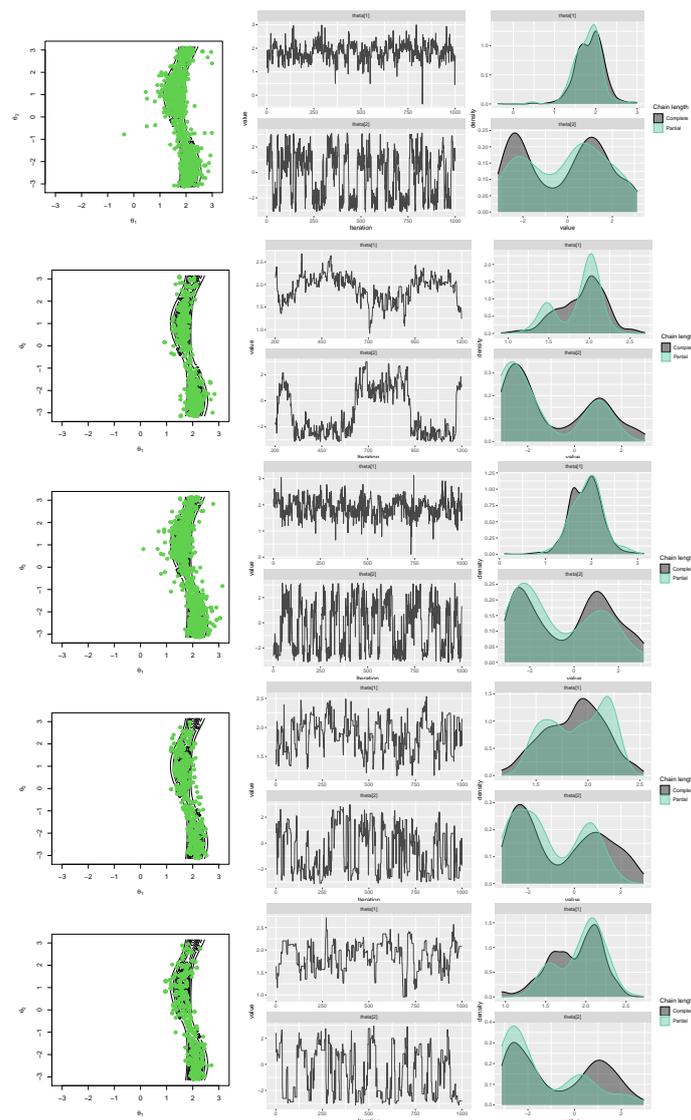


Figure 13. Scatter, trace, and compare-partial plots of the simulated data from the transformed sine model using “gibbs_met” in the “gibbs.met” package (first row), “MCMCmetrop1R” in the “MCMCpack” package (second row), “met_gaussian” in the “gibbs.met” package (third row), “Metro_Hastings” in the “MHadaptive” package (fourth row), and “rwmctrop” in the “LearnBayes” package (fifth row).

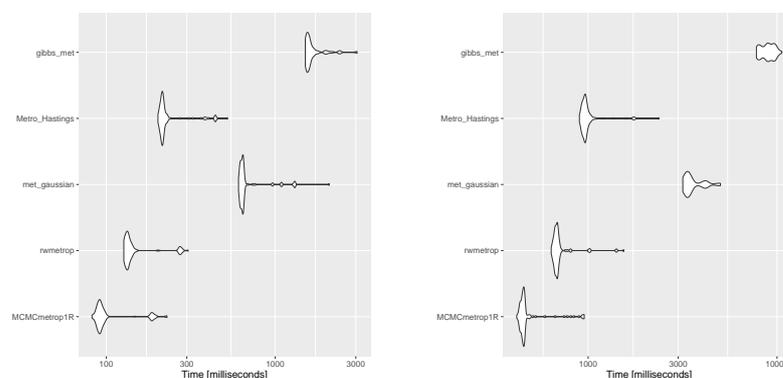


Figure 14. Execution times for generating a sample size of $n = 100$ from a transformed sine model (left) and a transformed cosine model (right) for each method.

Table 4. Maximum likelihood estimates of parameters and bias, and the MSE of the estimates for the simulated data obtained from each method.

Method	Distribution	n		κ_1	κ_2	κ_3	r_1	r_2	μ_1	μ_2
MCMCmetrop1R	Transformed sine	$n = 100$	MLE	1.8450	0.2876	2.8360	0.6001	0.0063	1.7744	2.6177
			Bias	-0.0542	-0.1932	-0.1531	-0.0089	-0.0521	-0.0815	0.1185
			MSE	0.0879	0.0921	0.1864	0.0083	0.0044	0.0087	0.0344
		$n = 1000$	MLE	2.1311	0.2649	3.1546	0.6243	0.0395	1.8491	2.4263
			Bias	-0.0298	-0.0900	-0.0135	0.0142	0.0130	-0.0018	-0.0046
			MSE	0.0527	0.0588	0.0002	0.0018	0.0008	0.0005	2.5×10^{-5}
	Transformed cosine	$n = 100$	MLE	3.7753	0.5985	1.5693	0.2564	0.4660	-1.4684	1.0937
			Bias	-0.1137	-0.0572	-0.1022	0.0143	-0.0376	0.0957	0.0421
			MSE	0.0998	0.0932	0.1007	0.0028	0.0116	0.0135	0.0275
		$n = 1000$	MLE	4.0693	0.6305	1.8210	0.2546	0.5557	-1.5613	1.0116
			Bias	0.0758	-0.0319	0.0305	0.0241	0.0329	0.0037	0.0238
			MSE	0.0065	0.0426	0.0085	0.0007	0.0014	0.0007	0.0011
rwmetrop	Transformed sine	$n = 100$	MLE	1.8186	0.2260	3.2619	0.6444	0.0359	1.8815	2.6027
			Bias	-0.2337	-0.0867	0.3837	0.0398	0.0205	0.0370	0.1006
			MSE	0.1094	0.0350	0.0726	0.0016	0.0406	0.0022	0.0291
		$n = 1000$	MLE	1.9735	0.3887	3.1546	0.6204	0.0159	1.8701	2.4270
			Bias	0.0979	0.0360	-0.0165	0.0131	0.0046	0.0129	-0.0046
			MSE	0.0340	0.0482	0.0003	0.0004	0.0001	0.0005	2.1×10^{-5}
	Transformed cosine	$n = 100$	MLE	3.8086	0.5397	1.4956	0.1726	0.5929	-1.5746	1.0417
			Bias	-0.0984	-0.0940	-0.2954	-0.0512	0.0883	-0.0115	0.0524
			MSE	0.0902	0.0893	0.1003	0.0033	0.0088	0.0009	0.0037
		$n = 1000$	MLE	3.9139	0.7320	1.8188	0.2546	0.4667	-1.5766	0.9962
			Bias	-0.0758	0.0703	0.0166	0.0241	-0.0386	-0.0095	0.0084
			MSE	0.0056	0.0674	0.0007	0.0013	0.0014	0.0007	0.0018
met_gaussian	Transformed sine	$n = 100$	MLE	2.5400	0.4480	3.1322	0.6963	0.1087	1.8321	2.3133
			Bias	0.3469	0.0904	-0.0328	0.0837	0.0916	-0.0294	-0.0908
			MSE	0.1773	0.0583	0.0117	0.0092	0.0091	0.0009	0.0141
		$n = 1000$	MLE	1.9022	0.2894	3.3403	0.6649	0.0016	1.8619	2.3822
			Bias	-0.2223	-0.0658	0.1137	0.0607	-0.0123	0.0042	-0.0481
			MSE	0.1268	0.0042	0.0286	0.0043	0.0025	2.1×10^{-5}	0.0024
	Transformed cosine	$n = 100$	MLE	3.4389	0.5409	1.4373	0.2530	0.5214	-1.5411	1.0526
			Bias	-0.3642	-0.1187	-0.2970	0.0143	0.0113	0.0210	0.0623
			MSE	0.1927	0.0993	0.1251	0.0853	0.0082	0.0005	0.0041
		$n = 1000$	MLE	3.6491	0.5341	1.6260	0.2420	0.5134	-1.5720	1.0158
			Bias	-0.2978	-0.1049	-0.0999	0.0195	0.0027	-0.0105	0.0126
			MSE	0.1214	0.0915	0.0875	0.0019	0.0032	0.0005	0.0024
Metro_Hastings	Transformed sine	$n = 100$	MLE	2.1465	0.2728	2.6713	0.6912	0.0010	1.8107	2.2024
			Bias	-0.0197	-0.0784	-0.4598	0.0856	0.0770	-0.0603	-0.1029
			MSE	0.0434	0.0883	0.2498	0.0073	0.0059	0.0046	0.0527
		$n = 1000$	MLE	2.1657	0.2743	3.2124	0.5813	0.0762	1.8487	2.4826
			Bias	0.0362	-0.0456	0.0433	-0.0262	0.0531	-0.0081	0.0404
			MSE	0.0246	0.0556	0.0027	0.0011	0.0039	7.3×10^{-5}	0.0025
	Transformed cosine	$n = 100$	MLE	3.8290	0.5903	2.0419	0.2857	0.5753	-1.5434	0.8373
			Bias	-0.1600	-0.0582	0.2407	0.0577	0.0709	0.0172	-0.1202
			MSE	0.1998	0.0944	0.1208	0.0061	0.0059	0.0006	0.0266
		$n = 1000$	MLE	3.8936	0.6747	1.5822	0.2775	0.4859	-1.5676	0.9935
			Bias	-0.0961	0.0298	-0.2317	0.0470	-0.0196	-0.0025	0.0057
			MSE	0.0091	0.0847	0.0829	0.0030	0.0003	0.0002	0.0051
gibbs_met	Transformed sine	$n = 100$	MLE	2.2340	0.2728	2.8277	0.6198	0.1398	1.8437	2.1688
			Bias	0.0712	-0.0784	-0.3884	0.0143	0.1220	-0.0129	-0.2598
			MSE	0.1171	0.0583	0.1179	0.0092	0.0160	0.0003	0.0692
		$n = 1000$	MLE	2.1901	0.3555	3.1760	0.6016	0.0156	1.8573	2.4342
			Bias	0.0315	0.0066	0.0048	-0.0020	0.0024	6.1×10^{-5}	0.0021
			MSE	0.0963	0.0221	0.0464	0.0008	0.0003	1.7×10^{-5}	0.0018
	Transformed cosine	$n = 100$	MLE	3.6123	0.5732	1.5193	0.2006	0.5852	-1.5886	0.8808
			Bias	-0.2842	-0.0853	-0.2717	-0.0276	0.0869	-0.0238	-0.0758
			MSE	0.1419	0.0893	0.1067	0.0082	0.0087	0.0005	0.0187
		$n = 1000$	MLE	3.6322	0.5978	1.7613	0.2626	0.5793	-1.5685	0.9668
			Bias	-0.2568	-0.0653	-0.0232	0.0221	0.0648	-0.0034	-0.0209
			MSE	0.0915	0.0726	0.0008	0.0031	0.0074	0.0002	0.0046

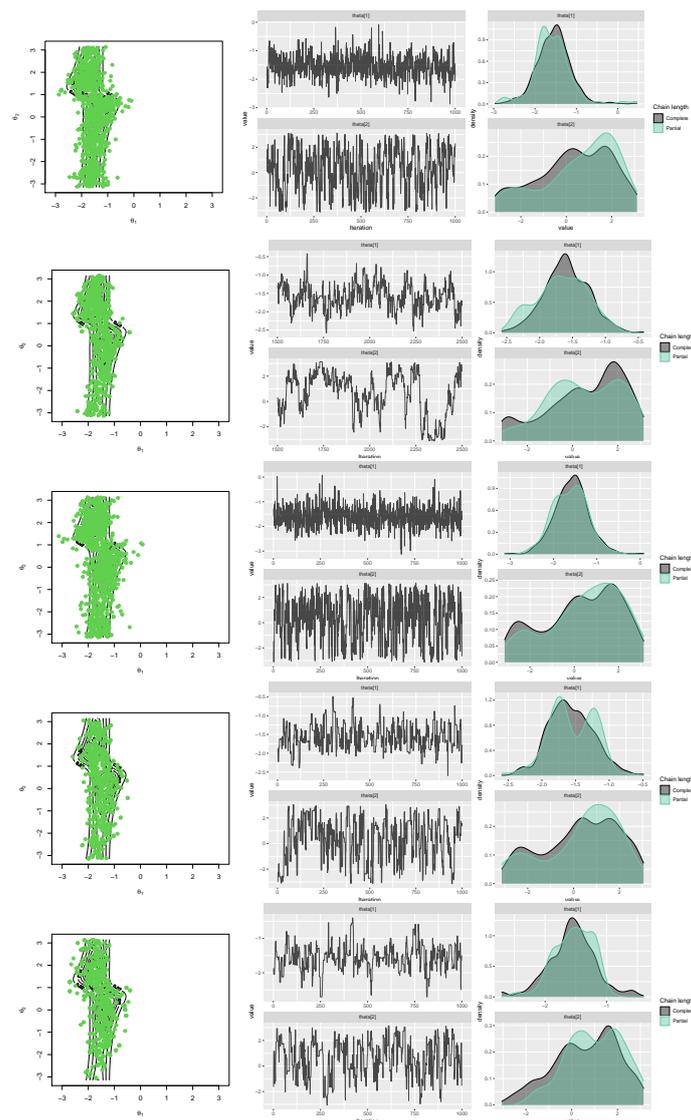


Figure 15. Scatter, trace, and compare-partial plots of the simulated data from the transformed cosine model using “gibbs_met” in the “gibbs.met” package (first row), “MCMCmetrop1R” in the “MCMCpack” package (second row), “met_gaussian” in the “gibbs.met” package (third row), “Metro_Hastings” in the “MHadaptive” package (fourth row), and “rwmtemp” in the “LearnBayes” package (fifth row).

7. Conclusions

In MCMC protein sampling for predicting the 3D structure, when the proposal distribution is closer to the stationary distribution, the results are more accurate. Therefore, a suitable proposal distribution can be defined using the angles and bond lengths observed in natural proteins. Statistical distributions for modelling protein dihedral angles can be used as proposal distributions for MCMC protein sampling. We gave a brief overview of existing symmetric models that formed the basis of the proposed models in this paper ((2) and (4)). In addition, new Möbius transformation-induced toroidal distributions, together with skewed versions, were developed in this study as alternatives to proposal distributions for the MCMC sampling of proteins. We demonstrated their performance with three protein datasets of toroidal nature and graphically illustrated their flexible behaviour. The AIC and BIC confirmed the better performance of our proposed models in comparison with the existing models. These newly proposed models even outperformed mixtures of well-known models for modelling toroidal data. In comparison with the existing toroidal models, these proposed models reflect the protein structural information better and should

be incorporated into proposal distributions. Lastly, to meet the need for sampling of proposal distribution in the MCMC algorithm, suitable methods for generating samples from these new models were explored using different types of the Metropolis sampling. In the future, one can investigate the performance of the Möbius transformation to obtain new cylindrical distributions.

Author Contributions: Conceptualization, M.A. and A.B.; methodology, M.A., N.N.R., A.B. and W.-D.S.; validation, M.A., N.N.R. and A.B.; formal analysis, M.A., N.N.R. and A.B.; investigation, M.A., N.N.R., A.B. and W.-D.S.; writing—original draft preparation, N.N.R.; writing—review and editing, M.A., N.N.R. and A.B.; visualization, M.A., N.N.R., A.B. and W.-D.S.; supervision, M.A. and A.B.; project administration, N.N.R.; funding acquisition, N.N.R. and A.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Research Foundation grant number 71199, 109214, 120839.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are available at <http://scop.mrc-lmb.cam.ac.uk/scop/> (accessed on 20 August 2020).

Acknowledgments: We would like to sincerely thank the three anonymous reviewers for their constructive comments that improved the paper. This work was based on research supported in part by the National Research Foundation (NRF) of South Africa, SARChI Research Chair UID: 71199; Ref.: IFR170227223754 grant No. 109214; Ref.: SRUG190308422768 grant No. 120839, STATOMET at the Department of Statistics at the University of Pretoria, and DSI-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS), South Africa. The opinions expressed and conclusions arrived at are those of the authors and are not necessarily attributed to the CoE-MaSS or the NRF.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Proof of Proposition 1

When $r_1, r_2 \rightarrow 0$, pdf (10) tends to the cosine distribution, which for large values of κ_1 and κ_2 is concentrated near 0. Suppose $\mu_1 = \mu_2 = 0$ (without loss of generality), according to Theorem 1 in [14] and using Taylor expansions, $(\Theta_1, \Theta_2) \sim N_2(0, \Sigma)$, where $\Sigma^{-1} = \begin{pmatrix} \kappa_1 - \kappa_3 & \kappa_3 \\ \kappa_3 & \kappa_2 - \kappa_3 \end{pmatrix}$, with $\kappa_3 \leq \frac{\kappa_1 \kappa_2}{\kappa_1 + \kappa_2}$.

Appendix A.2. Proof of Corollary 1

Without loss of generality, we consider $\mu_1 = 0$. According to (12), we conclude that:

$$\begin{aligned} \frac{f'_{\Theta_1}(\theta_1)}{f_{\Theta_1}(\theta_1)} &= \left\{ \frac{-2r_1}{(1+r_1^2-2r_1\cos\theta_1)} + \frac{\kappa_2\kappa_3(1-r_1^2)^2 A(h(\theta_1))}{h(\theta_1)(1+r_1^2-2r_1\cos\theta_1)^2} \right. \\ &\quad \left. - \frac{\kappa_1(1-r_1^2)^2}{(1+r_1^2-2r_1\cos\theta_1)^2} \right\} \sin\theta_1 \\ &= g(\theta_1) \sin\theta_1. \end{aligned} \tag{A1}$$

In (A1), if $\kappa_3 < 0$, then $g(\theta_1) < 0$. Therefore, for $\theta_1 \in [0, \pi)$, $f'_{\Theta_1}(\theta_1) < 0$, and for $\theta_1 \in [-\pi, 0)$, $f'_{\Theta_1}(\theta_1) \geq 0$. Thus, $f_{\Theta_1}(\theta_1)$ is increasing in $[-\pi, 0)$ and decreases from 0 to π . In addition, $f_{\Theta_1}(\theta_1) = f_{\Theta_1}(-\theta_1)$, which means that $f_{\Theta_1}(\theta_1)$ is symmetric around 0; thus, for $\kappa_3 < 0$, $f_{\Theta_1}(\theta_1)$ is unimodal. If $\kappa_3 > 0$, $h(\theta_1)$ decreases from $-\pi$ to 0 and increases from 0 to π , and $h(0) = |\kappa_2 - \kappa_3|$ and $h(-\pi) = \kappa_2 + \kappa_3$. From Lemma 1 in Singh et al. (2002), $A(t)/t$ is a decreasing function of t ; therefore, $A(h(\theta_1))/h(\theta_1)$ is increasing in $[-\pi, 0)$ and

decreases from 0 to π . It can be concluded that $g(\theta_1)$ is decreasing in $[0, \pi)$ and increasing in $[-\pi, 0)$; hence, if $-2r_1 + \kappa_2\kappa_3 \frac{A(|\kappa_2 - \kappa_3|)}{|\kappa_2 - \kappa_3|} \frac{(1-r_1^2)^2}{(1-r_1)^2} - \kappa_1 \frac{(1-r_1^2)^2}{(1-r_1)^2} < 0$, then $f'_{\Theta_1}(\theta_1) \geq 0$ for $\theta_1 \in [-\pi, 0)$ and $f'_{\Theta_1}(\theta_1) < 0$ for $\theta_1 \in [0, \pi)$; which means that $f_{\Theta_1}(\theta_1)$ is unimodal. If $-2r_1 + \kappa_2\kappa_3 \frac{A(|\kappa_2 - \kappa_3|)}{|\kappa_2 - \kappa_3|} \frac{(1-r_1^2)^2}{(1-r_1)^2} - \kappa_1 \frac{(1-r_1^2)^2}{(1-r_1)^2} > 0$ and $-2r_1 + \kappa_2\kappa_3 \frac{A(\kappa_2 + \kappa_3)}{\kappa_2 + \kappa_3} \frac{(1-r_1^2)^2}{(1+r_1)^2} - \kappa_1 \frac{(1-r_1^2)^2}{(1+r_1)^2} \leq 0$, then $f_{\Theta_1}(\theta_1)$ is first increasing and then decreasing in $[-\pi, 0)$, which means that $f_{\Theta_1}(\theta_1)$ is bimodal. A more detailed proof is provided by the authors upon request.

Appendix A.3. Proof of Corollary 2

Suppose $\mu_1 = 0$ (without loss of generality). According to (18), the following result can be obtained:

$$\begin{aligned} \frac{f'_{\Theta_1}(\theta_1)}{f_{\Theta_1}(\theta_1)} &= \left\{ \frac{-2r_1}{(1+r_1^2-2r_1\cos\theta_1)} - \frac{\kappa_1(1-r_1^2)^2}{(1+r_1^2-2r_1\cos\theta_1)^2} + \frac{\kappa_3^2(1-r_1^2)^2 A(h(\theta_1))}{h(\theta_1)(1+r_1^2-2r_1\cos\theta_1)^4} \right. \\ &\quad \left. \times \left(((1+r_1^2)^2 + 4r_1^2) \cos\theta_1 - 4r_1(1+r_1^2) \cos^2\theta_1 - 2r_1(1+r_1^2) \sin^2\theta_1 \right) \right\} \sin\theta_1 \\ &= g(\theta_1) \sin\theta_1. \end{aligned} \quad (\text{A2})$$

In (A2), if $\cos\theta_1 \leq 0$, then $g(\theta_1) < 0$ and the sign of (A2) depends on the sign of $\sin\theta_1$. Hence, for $\theta_1 \in (-\pi, -\pi/2]$, $f'_{\Theta_1}(\theta_1) < 0$ and for $\theta_1 \in [\pi/2, \pi]$, $f'_{\Theta_1}(\theta_1) \geq 0$. Thus, $f_{\Theta_1}(\theta_1)$ is increasing in $(-\pi, -\pi/2]$ and decreasing from $\pi/2$ to π . In addition, $f_{\Theta_1}(\theta_1) = f_{\Theta_1}(-\theta_1)$, which means that $f_{\Theta_1}(\theta_1)$ is symmetric around 0; therefore, $f_{\Theta_1}(\theta_1)$ is unimodal. For $\theta \in [0, \pi/2]$, $h(\theta_1)$ is an increasing function of θ_1 , and according to Lemma 1 in [13], $A(h(\theta_1))/h(\theta_1)$ is a decreasing function of θ_1 . We can conclude that if $-2r(1-r_1)^2 - \kappa_1(1-r_1^2)^2 + \kappa_3^2(1-r_1^2)^2 \frac{A(\kappa_2)}{\kappa_2} < 0$, then $f_{\Theta_1}(\theta_1)$ is a decreasing function from 0 to $\pi/2$, and because $f_{\Theta_1}(\theta_1)$ is symmetric around 0, it increases from $-\pi/2$ to 0. If $-2r(1-r_1)^2 - \kappa_1(1-r_1^2)^2 + \kappa_3^2(1-r_1^2)^2 \frac{A(\kappa_2)}{\kappa_2} > 0$, then $f_{\Theta_1}(\theta_1)$ first increases and then decreases in $[0, \pi/2]$ and $[-\pi/2, 0]$ (because it is symmetric around 0), which states that $f_{\Theta_1}(\theta_1)$ is bimodal.

References

- Ramachandran, G.T.; Sasisekharan, V. Conformation of polypeptides and proteins. *Adv. Protein Chem.* **1968**, *23*, 283–437.
- Holley, L.H.; Karplus, M. Protein secondary structure prediction with a neural network. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 152–156. [\[CrossRef\]](#)
- Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092. [\[CrossRef\]](#)
- Hastings, W.K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970**, *57*, 97–109. [\[CrossRef\]](#)
- Irbäck, A.; Mohanty, S. PROFASI: A Monte Carlo simulation package for protein folding and aggregation. *J. Comput. Chem.* **2006**, *27*, 1548–1555. [\[CrossRef\]](#)
- Jones, D.T. Successful ab initio prediction of the tertiary structure of NK-lysin using multiple sequences and recognized supersecondary structural motifs. *Proteins Struct. Funct. Bioinform.* **1997**, *29*, 185–191. [\[CrossRef\]](#)
- Jones, T.A.; Thirup, S. Using known substructures in protein model building and crystallography. *Embo J.* **1986**, *5*, 819–822. [\[CrossRef\]](#)
- Simons, K.T.; Kooperberg, C.; Huang, E.; Baker, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* **1997**, *268*, 209–225. [\[CrossRef\]](#)
- Ley, C.; Verdebout, T. *Applied Directional Statistics: Modern Methods and Case Studies*; CRC Press: Boca Raton, FL, USA, 2018.
- Ley, C.; Verdebout, T. *Modern Directional Statistics*; CRC Press: Boca Raton, FL, USA, 2017.
- Mardia, K.V. Statistics of directional data. *J. R. Stat. Soc. Ser. B (Methodol.)* **1975**, *37*, 349–371. [\[CrossRef\]](#)
- Rivest, L.P. A distribution for dependent unit vectors. *Commun. Stat.-Theory Methods* **1988**, *17*, 461–483. [\[CrossRef\]](#)
- Singh, H.; Hnizdo, V.; Demchuk, E. Probabilistic model for two dependent circular variables. *Biometrika* **2002**, *89*, 719–723. [\[CrossRef\]](#)
- Mardia, K.V.; Taylor, C.C.; Subramaniam, G.K. Protein bioinformatics and mixtures of bivariate von-Mises distributions for angular data. *Biometrics* **2007**, *63*, 505–512. [\[CrossRef\]](#)
- Kent, J.T.; Mardia, K.V.; Taylor, C.C. Modelling strategies for bivariate circular data. In *Proceedings of the Leeds Annual Statistical Research Conference; The Art and Science of Statistical Bioinformatics*, Leeds University Press: Leeds, UK, 2008; pp. 70–73.

16. Mardia, K.V.; Frelsen, J. Statistics of bivariate von Mises distributions. In *Bayesian Methods in Structural Bioinformatics*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 159–178.
17. Mardia, K.V.; Hughes, G.; Taylor, C.C.; Singh, H. A multivariate von Mises distribution with applications to bioinformatics. *Can. J. Stat.* **2008**, *36*, 99–109. [[CrossRef](#)]
18. Wehrly, T.E.; Johnson, R.A. Bivariate models for dependence of angular observations and a related Markov process. *Biometrika* **1980**, *67*, 255–256. [[CrossRef](#)]
19. Jones, M.C.; Pewsey, A.; Kato, S. On a class of circulars: Copulas for circular distributions. *Ann. Inst. Stat. Math.* **2015**, *67*, 843–862. [[CrossRef](#)]
20. Fernández-Durán, J.J. Models for circular–linear and circular–circular data constructed from circular distributions based on nonnegative trigonometric sums. *Biometrics* **2007**, *63*, 579–585. [[CrossRef](#)]
21. García-Portugués, E.; Crujeiras, R.M.; González-Manteiga, W. Exploring wind direction and SO₂ concentration by circular–linear density estimation. *Stoch. Environ. Res. Risk Assess.* **2013**, *27*, 1055–1067. [[CrossRef](#)]
22. Pewsey, A.; García-Portugués, E. Recent advances in directional statistics. *TEST* **2021**, *30*, 1–58. [[CrossRef](#)]
23. Di Marzio, M.; Panzera, A.; Taylor, C.C. Kernel density estimation on the torus. *J. Stat. Plan. Inference* **2011**, *141*, 2156–2173. [[CrossRef](#)]
24. Ameijeiras-Alonso, J.; Ley, C. Sine-skewed toroidal distributions and their application in protein bioinformatics. *Biostatistics* **2020**. Available online: [doi:10.1093/biostatistics/kxaa039](https://doi.org/10.1093/biostatistics/kxaa039) (accessed on 20 January 2021). [[CrossRef](#)]
25. Kato, S.; Shimizu, K.; Shieh, G.S. A circular–circular regression model. *Stat. Sin.* **2008**, *18*, 633–645.
26. Shieh, G.S.; Johnson, R.A. Inferences based on a bivariate distribution with von-Mises marginals. *Ann. Inst. Stat. Math.* **2005**, *57*, 789–802. [[CrossRef](#)]
27. Shieh, G.S.; Zheng, S.; Johnson, R.A.; Chang, Y.F.; Shimizu, K.; Wang, C.C.; Tang, S.L. Modeling and comparing the organization of circular genomes. *Bioinformatics* **2011**, *27*, 912–918. [[CrossRef](#)]
28. Liu, D.; Peddada, S.D.; Li, L.; Weinberg, C.R. Phase analysis of circadian-related genes in two tissues. *BMC Bioinform.* **2006**, *7*, 87. [[CrossRef](#)] [[PubMed](#)]
29. Downs, T.D.; Mardia, K.V. Circular regression. *Biometrika* **2002**, *89*, 683–697. [[CrossRef](#)]
30. Jones, M.C. The Möbius distribution on the disc. *Ann. Inst. Stat. Math.* **2004**, *56*, 733–742. [[CrossRef](#)]
31. Kato, S.; Jones, M.C. A family of distributions on the circle with links to, and applications arising from, Möbius transformation. *J. Am. Stat. Assoc.* **2010**, *105*, 249–262. [[CrossRef](#)]
32. Wang, M.Z.; Shimizu, K. On applying Möbius transformation to cardioid random variables. *Stat. Methodol.* **2012**, *9*, 604–614. [[CrossRef](#)]
33. Kato, S.; Pewsey, A. A Möbius transformation-induced distribution on the torus. *Biometrika* **2015**, *102*, 359–370. [[CrossRef](#)]
34. Kato, S. A distribution for a pair of unit vectors generated by Brownian motion. *Bernoulli* **2009**, *15*, 898–921. [[CrossRef](#)]
35. Kato, S.; McCullagh, P. Some properties of a Cauchy family on the sphere derived from the Möbius transformation. *Bernoulli* **2020**, *26*, 3224–3248. [[CrossRef](#)]
36. McCullagh, P. Möbius transformation and Cauchy parameter estimation. *Ann. Stat.* **1996**, *24*, 787–808. [[CrossRef](#)]
37. Abe, T.; Pewsey, A. Sine-skewed circular distributions. *Stat. Pap.* **2011**, *52*, 683–707. [[CrossRef](#)]
38. Mullen, K.; Ardia, D.; Gil, D.L.; Windover, D.; Cline, J. DEoptim: An R package for global optimization by differential evolution. *J. Stat. Softw.* **2011**, *40*, 1–26. [[CrossRef](#)]
39. Storn, R.; Price, K. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J. Glob. Optim.* **1997**, *11*, 341–359. [[CrossRef](#)]
40. Price, K.; Storn, R.M.; Lampinen, J.A. *Differential Evolution: A Practical Approach to Global Optimization*; Springer: Berlin/Heidelberg, Germany, 2006.
41. Najibi, S.M.; Maadooliat, M.; Zhou, L.; Huang, J.Z.; Gao, X. Protein structure classification and loop modeling using multiple Ramachandran distributions. *Comput. Struct. Biotechnol. J.* **2017**, *15*, 243–254. [[CrossRef](#)] [[PubMed](#)]
42. Moghimbeygi, M.; Golalizadeh, M. Spherical logistic distribution. *Commun. Math. Stat.* **2020**, *8*, 151–166. [[CrossRef](#)]
43. Martin, A.D.; Quinn, K.M.; Park, J.H.; Park, M.J.H. *MCMCpack: Markov Chain Monte Carlo (MCMC) Package*; Version 1.5-0; R Package: Vienna, Austria, 2020. Available online: <https://cran.r-project.org/web/packages/MCMCpack/index.html> (accessed on 25 August 2020).
44. Li, L. *gibbs.met: Naive Gibbs Sampling with Metropolis Steps*; Version 1.1-3; R Package: Vienna, Austria, 2015. Available online: <https://cran.r-project.org/web/packages/gibbs.met/index.html> (accessed on 25 August 2020).
45. Albert, J. *LearnBayes: Functions for Learning Bayesian Inference*; Version 2.15.1; R Package: Vienna, Austria, 2018. Available online: <https://cran.r-project.org/web/packages/LearnBayes/index.html> (accessed on 25 August 2020).
46. Chivers, C.; Chivers, M.C. *MHADaptive: General Markov Chain Monte Carlo for Bayesian Inference Using Adaptive Metropolis-Hastings Sampling*; Version 1.1-8; R Package: Vienna, Austria, 2015. Available online: <https://cran.r-project.org/web/packages/MHADaptive/index.html> (accessed on 25 August 2020).
47. Fernández-i-Marín, X. ggmcmc: Analysis of MCMC samples and Bayesian inference. *J. Stat. Softw.* **2016**, *70*, 1–20. [[CrossRef](#)]
48. Mersmann, O. *Microbenchmark: Accurate Timing Functions*; Version 1.4-7; R Package: Vienna, Austria, 2019. Available online: <https://www.rdocumentation.org/packages/microbenchmark/versions/1.4-7/topics/microbenchmark> (accessed on 25 August 2020).