



Coming Together of Bayesian Inference and Skew Spherical Data

Najmeh Nakhaei Rad^{1,2,3}, Andriette Bekker³, Mohammad Arashi^{3,4} and Christophe Ley^{5*}

¹ Department of Mathematics and Statistics, Mashhad Branch, Islamic Azad University, Mashhad, Iran, ² DSI-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS), Johannesburg, South Africa, ³ Department of Statistics, University of Pretoria, Pretoria, South Africa, ⁴ Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran, ⁵ Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium

This paper presents Bayesian directional data modeling via the skew-rotationally-symmetric Fisher-von Mises-Langevin (FvML) distribution. The prior distributions for the parameters are a pivotal building block in Bayesian analysis, therefore, the impact of the proposed priors will be quantified using the Wasserstein Impact Measure (WIM) to guide the practitioner in the implementation process. For the computation of the posterior, modifications of Gibbs and slice samplings are applied for generating samples. We demonstrate the applicability of our contribution via synthetic and real data analyses. Our investigation paves the way for Bayesian analysis of skew circular and spherical data.

OPEN ACCESS

Edited by:

Jian Qing Shi,
Southern University of Science and
Technology, China

Reviewed by:

Daizong Ding,
Fudan University, China
Chong Zhong,
Hong Kong Polytechnic University,
Hong Kong SAR, China

*Correspondence:

Christophe Ley
christophe.ley@ugent.be

Specialty section:

This article was submitted to
Data Science,
a section of the journal
Frontiers in Big Data

Received: 02 September 2021

Accepted: 27 December 2021

Published: 08 February 2022

Citation:

Nakhaei Rad N, Bekker A, Arashi M
and Ley C (2022) Coming Together of
Bayesian Inference and Skew
Spherical Data.
Front. Big Data 4:769726.
doi: 10.3389/fdata.2021.769726

Keywords: Fisher-von Mises-Langevin distribution, Gibbs sampling, MCMC method, skew-rotationally-symmetric distributions, slice sampler, spherical data, Wasserstein Impact Measure

1. INTRODUCTION

Big and complex data sets are collected from various scientific fields such as atmospheric environment, social science, psychological and biomedical studies, bioinformatics, epidemiology, digital imaging information, and machine learning, to name just a few. Big Data can refer to data with big volume or velocity, high-dimensional data (Ahmed, 2017), unstructured or unusual data (Härdle et al., 2018), complex data, etc. Therefore, there is a need for developing statistical techniques other than the traditional analytical frameworks to model, interpret, and use such data in different fields of science. Data with directions are categorized as unusual data that cannot be analyzed and modeled under the Cartesian coordinate system. With the aid of directional statistics, data science meets another level of analytical methods. In this research work, we focus on the analysis of complex directional data. Bayesian methods have received extensive attention in data science because prior information can be added to enhance modeling. Therefore, here, we consider Bayesian analysis of complex directional data.

A robust roadmap with the symmetric Fisher-von Mises-Langevin (FvML) distribution as the key element from a Bayesian perspective is briefly reviewed. According to Kikuchi's collection of directional data (Kikuchi, 1982), the first attention paid to Bayesian methods for directional data was in a paper by Mardia and El-Atoum (1976). They used the Bayesian approach to estimate the location parameter of the FvML distribution when the concentration parameter was known. The author Bagchi made several contributions in this area: (i) Bagchi (1988) formulated a conjugate prior for the mean direction and a non-informative prior for the concentration parameter of the von Mises distribution; (ii) Bagchi and Guttman (1988) focused on Bayesian inference for the multi-variate FvML distribution; (iii) Bagchi and Kadane (1991) derived the Bayes estimate for the cosine of the direction parameter of the von Mises distribution when

the concentration parameter is known; and (iv) Bagchi (1994) developed empirical Bayesian techniques to estimate the mean direction of the FvML distribution (see also Guttorp and Lockhart, 1988; Dowe et al., 1996). Damien and Walker (1999) presented a full Bayesian inference for the von Mises distribution implementing a Gibbs sampler while (Rodrigues et al., 2000) provided an empirical or approximate Bayesian inference for the von Mises distribution. Nuñez-Antonio and Gutiérrez-Peña (2005) presented Bayesian analysis of the FvML distribution when all of the parameters were unknown, as well as an algorithm to generate samples from the posterior distribution based on a sampling-importance-resampling method. Muralidharan and Parikh (2007) provided Bayes estimates for both location and concentration parameters of the von Mises distribution. Bhattacharya and SenGupta (2009) presented Bayesian analysis of a generalized von Mises distribution introducing a new algorithm based on importance sampling and Markov chain Monte Carlo (MCMC) to draw samples from the posterior distribution. Mardia (2010) moved the attention to the bivariate von Mises distribution (on the torus) from a Bayesian viewpoint. Infinite mixtures of FvML distributions using standard conjugate priors for the parameters and Dirichlet priors for the mixing probabilities received attention from Bangert et al. (2010). Hornik and Grün (2013) defined conjugate and Jeffreys priors for the FvML distribution while Taghia et al. (2014) worked on Bayesian inference for the FvML mixture model. From 2017 and onwards the following contributions can be highlighted: Straub (2017) presented Bayesian analysis for the FvML distribution in 3D; Røge et al. (2017) presented Bayesian inference in the case of infinite FvML mixture model assumption; Mulder et al. (2020) provided Bayesian inference for mixtures of von Mises distributions using a reversible jump MCMC sampler and focused on non-informative priors. Lastly, the interested reader is referred to Pewsey and García-Portugués (2021) for Bayesian inference of other directional distributions.

Numerous directional data sets tend to show non-trivial features such as skewness. Therefore, the underlying distribution is not always symmetric, which emphasizes the focus on skewed directional distributions. This inspired us to investigate Bayesian analysis for the general class of skew-rotationally-symmetric distributions (Ley and Verdebout, 2017b), an asymmetric extension of all rotationally symmetric distributions, when the location, concentration, and skewness parameters are unknown.

In section 2, the skew-rotationally-symmetric distribution and special cases are described. The novel contribution is given in section 3 where the posterior distributions are obtained for the skew-FvML as the likelihood model, for four different scenarios of the prior distributions for the parameters of the model. Moreover, an algorithm is provided for generating samples from these posterior distributions. The impact of the priors is explored in section 4, by implementing the Wasserstein Impact Measure (WIM). In section 5, a synthetic data analysis is conducted to show the accuracy of the Bayes estimates based on the assumptions of the skew-FvML model. We demonstrate the applicability of the Bayesian framework for well known real datasets in section 6 for dimensions $p = 2$ and 3.

2. SKEW-ROTATIONALLY-SYMMETRIC DISTRIBUTIONS

Most of the distributions on the unit hypersphere $\mathbb{S}^{p-1} = \{\mathbf{v} \in \mathbb{R}^p : \mathbf{v}^\top \mathbf{v} = 1\}$, $p \geq 2$, share the common feature of being rotationally-symmetric about their location $\boldsymbol{\mu} \in \mathbb{S}^{p-1}$. The distribution of a random variable $\mathbf{X} \in \mathbb{S}^{p-1}$ is said to be rotationally-symmetric about $\boldsymbol{\mu}$ if for any orthogonal matrix $\mathbf{O}_{p \times p}$ satisfying $\mathbf{O}\boldsymbol{\mu} = \boldsymbol{\mu}$ it is concluded that $\mathbf{O}\mathbf{X}$ is equal in distribution to \mathbf{X} . The FvML distribution, the most common distribution in spherical studies, is obtained by conditioning on the p -variate normal distribution (see Mardia and Jupp, 2000; Ley and Verdebout, 2017a). Suppose \mathbf{X} takes values on the non-linear manifold \mathbb{S}^{p-1} and has the FvML distribution, then its probability density function (pdf) is given by

$$f(\mathbf{x}; \boldsymbol{\mu}, \tau) = C(\tau) \exp(\tau \boldsymbol{\mu}^\top \mathbf{x}), \quad \mathbf{x}, \boldsymbol{\mu} \in \mathbb{S}^{p-1}, \quad (1)$$

where

$$C(\tau) = (2\pi)^{-p/2} \frac{\tau^{p/2-1}}{I_{p/2-1}(\tau)}, \quad (2)$$

$\tau \geq 0$ is the concentration parameter, and I_α is the modified Bessel function of order α . For $\tau = 0$, (1) simplifies to the uniform distribution. If $p = 2$ in (1), it results in the von Mises distribution and when $p = 3$, the Fisher distribution (Fisher, 1953) is obtained.

However, in practice, all real-life phenomena cannot be represented by symmetric models. The interested reader is referred to Downs (2003) for medical research of heart disease diagnosis, Leong and Carlile (1998) for application in neurosciences, Shearman et al. (2000) for biological research on mammalian circadian rhythms, Mardia (2013) and Ameijeiras-Alonso and Ley (2020) for application in bioinformatics especially protein structure prediction, Fisher and Lee (1994) for some studies on wind direction, Ameijeiras-Alonso et al. (2021) for biomechanics studies, and Pewsey (2002) and Ley and Verdebout (2014) for animal movement studies.

Therefore, in this paper, the focus will be on the skew-rotationally-symmetric (SRS) distributions, introduced by Ley and Verdebout (2017b) as

$$f_{\text{SRS}}(\mathbf{x}) = 2f(\mathbf{x}^\top \boldsymbol{\mu}) \Pi(\boldsymbol{\gamma}^\top \boldsymbol{\Upsilon}_\mu^\top \mathbf{x}), \quad \mathbf{x}, \boldsymbol{\mu} \in \mathbb{S}^{p-1}, \quad (3)$$

where $f(\mathbf{x}^\top \boldsymbol{\mu})$ is a rotationally-symmetric pdf about $\boldsymbol{\mu} \in \mathbb{S}^{p-1}$, $\Pi: \mathbb{R} \rightarrow [0, 1]$ is a monotone increasing function satisfying $\Pi(-t) + \Pi(t) = 1$ for all $t \in \mathbb{R}$, and $\boldsymbol{\Upsilon}_\mu^\top$ represents the semi-orthogonal matrix such that $\boldsymbol{\Upsilon}_\mu \boldsymbol{\Upsilon}_\mu^\top = \mathbf{I}_p - \boldsymbol{\mu} \boldsymbol{\mu}^\top$ and $\boldsymbol{\Upsilon}_\mu^\top \boldsymbol{\Upsilon}_\mu = \mathbf{I}_{p-1}$, where \mathbf{I}_p is the $p \times p$ identity matrix. The parameter $\boldsymbol{\gamma} \in \mathbb{R}^{p-1}$ is a skewness parameter vector such that $\boldsymbol{\gamma} = \mathbf{0}$ provides the symmetric pdf $f(\mathbf{x}^\top \boldsymbol{\mu})$ and non-zero values of $\boldsymbol{\gamma}$ provide skewed pdfs. This construction allows using the full potential of existing rotationally-symmetric distributions by turning them into skewed versions.

Substituting (1) in (3) and letting $\boldsymbol{\Upsilon}_\mu^\top \mathbf{x} = ((1 - (\mathbf{x}^\top \boldsymbol{\mu})^2)^{1/2} \mathbf{U}_\mu(\mathbf{x}))$ with $\mathbf{U}_\mu(\mathbf{x})$ the sign vector which

is uniformly distributed on \mathbb{S}^{p-2} , the skew-FvML (SFvML) distribution is obtained as

$$f_{SFvML}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\gamma}, \tau) = 2C(\tau) \exp(\tau \mathbf{x}^\top \boldsymbol{\mu}) \Pi \left((1 - (\mathbf{x}^\top \boldsymbol{\mu})^2)^{1/2} \boldsymbol{\gamma}^\top \mathbf{U}_\mu(\mathbf{x}) \right), \quad \mathbf{x}, \boldsymbol{\mu} \in \mathbb{S}^{p-1}, \quad (4)$$

where $C(\tau)$ is defined in (2).

By using the standard cosine transformation

$$(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)^\top = (\cos \theta_1, \sin \theta_1 \cos \theta_2, \dots, \sin \theta_1 \dots \sin \theta_{p-2} \sin \theta_{p-1})^\top, \quad (5)$$

and choosing $\boldsymbol{\Upsilon}_\mu^\top \mathbf{x} = \mathbf{1}(x \geq \mu) - \mathbf{1}(x \leq \mu)$ in (4), for $p = 2$, the skew-von Mises (SvM) distribution follows as

$$f_{SvM}(\theta; \mu, \tau, \gamma) = \frac{1}{\pi I_0(\tau)} \exp(\tau \cos(\theta - \mu)) \Pi(\gamma \sin(\theta - \mu)) \quad (6)$$

where $\theta, \mu \in (-\pi, \pi], \tau > 0$ and $\gamma \in \mathbb{R}$. Here, the scalar product $\mathbf{x}^\top \boldsymbol{\mu}$ is $\cos(\theta - \mu)$. By choosing $\Pi(x) = \frac{1+x}{2}, x \in [-1, 1]$, in (6), the sine-skewed von Mises distribution introduced by Abe and Pewsey (2011) is obtained where $\gamma \in [-1, 1]$.

The following lemma can be used to generate a sample from the SFvML distribution.

LEMMA 1. (Ley and Verdebout, 2017b) Generate \mathbf{Y} from the rotationally-symmetric FvML distribution in (1). Then for any uniformly distributed sign vector $\mathbf{U}_\mu(\mathbf{Y}), \mathbf{U}_{\mu;\Pi}(\mathbf{Y})$ is defined as

$$\mathbf{U}_{\mu;\Pi}(\mathbf{Y}) = \begin{cases} \mathbf{U}_\mu(\mathbf{Y}) & \text{if } U \leq \Pi \left((1 - (\mathbf{Y}^\top \boldsymbol{\mu})^2)^{1/2} \boldsymbol{\gamma}^\top \mathbf{U}_\mu(\mathbf{Y}) \right), \\ -\mathbf{U}_\mu(\mathbf{Y}) & \text{if } U > \Pi \left((1 - (\mathbf{Y}^\top \boldsymbol{\mu})^2)^{1/2} \boldsymbol{\gamma}^\top \mathbf{U}_\mu(\mathbf{Y}) \right), \end{cases}$$

where U is uniformly distributed on $(0, 1)$ and independent of \mathbf{Y} . Then the vector \mathbf{X} with pdf (4) is obtained as

$$\mathbf{X} = (\mathbf{Y}^\top \boldsymbol{\mu}) \boldsymbol{\mu} + (1 - (\mathbf{Y}^\top \boldsymbol{\mu})^2)^{1/2} \boldsymbol{\Upsilon}_\mu \mathbf{U}_{\mu;\Pi}(\mathbf{Y}).$$

In the next section, Bayesian inference with the SFvML distribution as the key element is presented with all location, concentration, and skewness parameters $\boldsymbol{\mu}, \tau$, and $\boldsymbol{\gamma}$ unknown.

3. METHODOLOGY

Let $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$ be a random sample of size n with pdf (4) where the standard normal cumulative density function (cdf) Φ replaces Π . The likelihood function is

$$L(\boldsymbol{\mu}, \boldsymbol{\gamma}, \tau | \mathbf{X}) = 2^n C^n(\tau) \exp(\tau \boldsymbol{\mu}^\top \sum_{i=1}^n \mathbf{x}_i) \prod_{i=1}^n \Phi \left((1 - (\mathbf{x}_i^\top \boldsymbol{\mu})^2)^{1/2} \boldsymbol{\gamma}^\top \mathbf{U}_\mu(\mathbf{x}_i) \right). \quad (7)$$

Subsequently, four scenarios are presented to define the prior distributions for the location parameter $\boldsymbol{\mu}$, the concentration parameter τ , and the skewness parameter $\boldsymbol{\gamma}$.

3.1. Prior Distributions

As above, let \mathbf{X} denote a set of observations, and a generative model of the data be defined through a set of unknown parameters $\boldsymbol{\Omega} = (\boldsymbol{\mu}, \boldsymbol{\gamma}, \tau)$ (see (7)). In this section the prior distributions for $\boldsymbol{\Omega} = (\boldsymbol{\mu}, \boldsymbol{\gamma}, \tau)$ are outlined.

For the skewness vector $\boldsymbol{\gamma}$ the following prior distributions are proposed: (i) the multi-variate normal distribution with location parameter $\boldsymbol{\xi}$ and covariance matrix $\text{diag}(\boldsymbol{\sigma})$, (ii) the multi-variate skew-normal distribution (Azzalini, 1985) with location parameter $\boldsymbol{\xi}$, covariance matrix $\text{diag}(\boldsymbol{\sigma})$ and skewness parameter $\boldsymbol{\lambda}$, i.e.,

$$\pi_1(\boldsymbol{\gamma} | \boldsymbol{\xi}, \text{diag}(\boldsymbol{\sigma})) \propto \prod_{i=1}^{p-1} \frac{1}{\sigma_i} \phi \left(\frac{\gamma_i - \xi_i}{\sigma_i} \right) \propto \phi_{p-1}(\boldsymbol{\gamma} - \boldsymbol{\xi}; \text{diag}(\boldsymbol{\sigma})), \quad (8)$$

and

$$\pi_2(\boldsymbol{\gamma} | \boldsymbol{\xi}, \text{diag}(\boldsymbol{\sigma}), \boldsymbol{\lambda}) \propto \prod_{i=1}^{p-1} \frac{1}{\sigma_i} \phi \left(\frac{\gamma_i - \xi_i}{\sigma_i} \right) \Phi \left(\lambda_i \frac{\gamma_i - \xi_i}{\sigma_i} \right) \propto \phi_{p-1}(\boldsymbol{\gamma} - \boldsymbol{\xi}; \text{diag}(\boldsymbol{\sigma})) \Phi_{p-1}(\boldsymbol{\gamma} - \boldsymbol{\xi}; \mathbf{D}), \quad (9)$$

where $\mathbf{D} = \text{diag}(\frac{\sigma_1}{\lambda_1}, \dots, \frac{\sigma_{p-1}}{\lambda_{p-1}})$, $\xi_i \in \mathbb{R}, \sigma_i > 0, \lambda_i \in \mathbb{R}$, ϕ is the standard normal pdf, ϕ_n and Φ_n are the pdf and cdf of the n -variate standard normal distribution, respectively. Next, the following priors for $\boldsymbol{\mu}$ and τ are considered.

Case 1: Nuñez-Antonio and Gutiérrez-Peña's Prior

In this case, we adopt the joint prior distribution of Nuñez-Antonio and Gutiérrez-Peña (2005) with direction parameter $\boldsymbol{\mu}_0$, concentration parameters ζ and η for $(\boldsymbol{\mu}, \tau)$, i.e.,

$$\pi(\boldsymbol{\mu}, \tau | \boldsymbol{\mu}_0, \zeta, \eta) \propto \left(\frac{\tau^{p/2-1}}{I_{p/2-1}(\tau)} \right)^\zeta \exp(\eta \tau \boldsymbol{\mu}^\top \boldsymbol{\mu}_0), \quad (10)$$

where $\boldsymbol{\mu}_0 \in \mathbb{S}^{p-1}$ and $0 < \eta < \zeta$. The normalization constant can only be obtained for some special cases. Straub (2017) computed the normalization constant of (10) for $\zeta = 1$ and $p = 3$.

Case 2: FvML and Gamma Prior

In this case, the FvML and gamma distributions with parameters $\boldsymbol{\mu}_0, \tau_0, \alpha$, and β are proposed as priors for $(\boldsymbol{\mu}, \tau)$ (Muralidharan and Parikh, 2007), i.e.,

$$\pi(\boldsymbol{\mu}, \tau | \boldsymbol{\mu}_0, \tau_0, \alpha, \beta) \propto \exp(\tau_0 \boldsymbol{\mu}^\top \boldsymbol{\mu}_0) \tau^{\alpha-1} \exp(-\beta \tau), \quad (11)$$

where $\tau_0, \alpha, \beta > 0$ and $\boldsymbol{\mu}_0 \in [-\pi, \pi)$.

3.2. Posterior Distributions

Subsequently, the posterior distribution $\pi(\boldsymbol{\Omega} | \mathbf{X}) \propto \pi(\boldsymbol{\Omega})L(\boldsymbol{\Omega} | \mathbf{X})$ is determined for the different prior assumptions on $\boldsymbol{\Omega} = (\boldsymbol{\mu}, \boldsymbol{\gamma}, \tau)$. Firstly, assume the prior distribution set up as described under case 1 and different prior distributions for the skewness parameter.

Scenario 1

Assume the prior distribution of the skewness parameter, $\boldsymbol{\gamma}$ is given by (8), then for given \mathbf{X} the posterior distribution of $(\boldsymbol{\mu}, \boldsymbol{\gamma}, \tau)$ can be obtained by using (7), (8), and (10) as

$$\begin{aligned} \pi(\boldsymbol{\mu}, \boldsymbol{\gamma}, \tau \mid \mathbf{X}, \boldsymbol{\mu}_0, \zeta, \eta, \boldsymbol{\xi}, \boldsymbol{\sigma}) &\propto \left(\frac{\tau^{p/2-1}}{I_{p/2-1}(\tau)}\right)^{\zeta+n} \\ &\exp\left(\tau \boldsymbol{\mu}^\top \left(\sum_{i=1}^n \mathbf{x}_i + \eta \boldsymbol{\mu}_0\right)\right) \phi_{p-1}(\boldsymbol{\gamma} - \boldsymbol{\xi}; \text{diag}(\boldsymbol{\sigma})) \\ &\times \prod_{i=1}^n \Phi\left((1 - (\mathbf{x}_i^\top \boldsymbol{\mu})^2)^{1/2} \boldsymbol{\gamma}^\top \mathbf{U}_\mu(\mathbf{x}_i)\right). \end{aligned} \tag{12}$$

The full conditionals for $\boldsymbol{\mu}$, $\boldsymbol{\gamma}$, and τ are, respectively:

$$\begin{aligned} \pi(\boldsymbol{\mu} \mid \boldsymbol{\gamma}, \tau, \mathbf{X}) &\propto \exp\left(\tau \boldsymbol{\mu}^\top \left(\sum_{i=1}^n \mathbf{x}_i + \eta \boldsymbol{\mu}_0\right)\right) \\ &\prod_{i=1}^n \Phi\left((1 - (\mathbf{x}_i^\top \boldsymbol{\mu})^2)^{1/2} \boldsymbol{\gamma}^\top \mathbf{U}_\mu(\mathbf{x}_i)\right), \\ \pi(\boldsymbol{\gamma} \mid \boldsymbol{\mu}, \tau, \mathbf{X}) &\propto \phi_{p-1}(\boldsymbol{\gamma} - \boldsymbol{\xi}; \text{diag}(\boldsymbol{\sigma})) \\ &\prod_{i=1}^n \Phi\left((1 - (\mathbf{x}_i^\top \boldsymbol{\mu})^2)^{1/2} \boldsymbol{\gamma}^\top \mathbf{U}_\mu(\mathbf{x}_i)\right), \\ \pi(\tau \mid \boldsymbol{\mu}, \boldsymbol{\gamma}, \mathbf{X}) &\propto \left(\frac{\tau^{p/2-1}}{I_{p/2-1}(\tau)}\right)^{\zeta+n} \exp\left(\tau \boldsymbol{\mu}^\top \left(\sum_{i=1}^n \mathbf{x}_i + \eta \boldsymbol{\mu}_0\right)\right). \end{aligned}$$

Scenario 2

If we assume the prior distribution (9) for $\boldsymbol{\gamma}$, the posterior distribution of $(\boldsymbol{\mu}, \boldsymbol{\gamma}, \tau)$ can be obtained by using (7), (9), and (10) as follows

$$\begin{aligned} \pi(\boldsymbol{\mu}, \boldsymbol{\gamma}, \tau \mid \mathbf{X}, \boldsymbol{\mu}_0, \zeta, \eta, \boldsymbol{\xi}, \boldsymbol{\sigma}) &\propto \left(\frac{\tau^{p/2-1}}{I_{p/2-1}(\tau)}\right)^{\zeta+n} \\ &\exp\left(\tau \boldsymbol{\mu}^\top \left(\sum_{i=1}^n \mathbf{x}_i + \eta \boldsymbol{\mu}_0\right)\right) \phi_{p-1}(\boldsymbol{\gamma} - \boldsymbol{\xi}; \text{diag}(\boldsymbol{\sigma})) \\ &\times \prod_{i=1}^n \Phi\left((1 - (\mathbf{x}_i^\top \boldsymbol{\mu})^2)^{1/2} \boldsymbol{\gamma}^\top \mathbf{U}_\mu(\mathbf{x}_i)\right) \Phi_{p-1}(\boldsymbol{\gamma} - \boldsymbol{\xi}; \mathbf{D}). \end{aligned} \tag{13}$$

The full conditionals for $\boldsymbol{\mu}$, $\boldsymbol{\gamma}$, and τ are, respectively:

$$\begin{aligned} \pi(\boldsymbol{\mu} \mid \boldsymbol{\gamma}, \tau, \mathbf{X}) &\propto \exp\left(\tau \boldsymbol{\mu}^\top \left(\sum_{i=1}^n \mathbf{x}_i + \eta \boldsymbol{\mu}_0\right)\right) \\ &\prod_{i=1}^n \Phi\left((1 - (\mathbf{x}_i^\top \boldsymbol{\mu})^2)^{1/2} \boldsymbol{\gamma}^\top \mathbf{U}_\mu(\mathbf{x}_i)\right), \\ \pi(\boldsymbol{\gamma} \mid \boldsymbol{\mu}, \tau, \mathbf{X}) &\propto \phi_{p-1}(\boldsymbol{\gamma} - \boldsymbol{\xi}; \text{diag}(\boldsymbol{\sigma})) \Phi_{p-1}(\boldsymbol{\gamma} - \boldsymbol{\xi}; \mathbf{D}) \\ &\times \prod_{i=1}^n \Phi\left((1 - (\mathbf{x}_i^\top \boldsymbol{\mu})^2)^{1/2} \boldsymbol{\gamma}^\top \mathbf{U}_\mu(\mathbf{x}_i)\right), \\ \pi(\tau \mid \boldsymbol{\mu}, \boldsymbol{\gamma}, \mathbf{X}) &\propto \left(\frac{\tau^{p/2-1}}{I_{p/2-1}(\tau)}\right)^{\zeta+n} \exp\left(\tau \boldsymbol{\mu}^\top \left(\sum_{i=1}^n \mathbf{x}_i + \eta \boldsymbol{\mu}_0\right)\right). \end{aligned}$$

Scenario 3

If the prior distribution of the skewness parameter $\boldsymbol{\gamma}$ is given by (8), for given \mathbf{X} , the posterior distribution of $(\boldsymbol{\mu}, \boldsymbol{\gamma}, \tau)$, by using (7), (8), and (11), is

$$\begin{aligned} \pi(\boldsymbol{\mu}, \boldsymbol{\gamma}, \tau \mid \mathbf{X}, \boldsymbol{\mu}_0, \tau_0, \alpha, \beta, \boldsymbol{\xi}, \boldsymbol{\sigma}) &\propto \frac{\tau^{np/2+\alpha-n-1}}{I_{p/2-1}^n(\tau)} \exp\left(\boldsymbol{\mu}^\top \left(\tau \sum_{i=1}^n \mathbf{x}_i + \tau_0 \boldsymbol{\mu}_0\right) - \beta \tau\right) \\ &\phi_{p-1}(\boldsymbol{\gamma} - \boldsymbol{\xi}; \text{diag}(\boldsymbol{\sigma})) \\ &\times \prod_{i=1}^n \Phi\left((1 - (\mathbf{x}_i^\top \boldsymbol{\mu})^2)^{1/2} \boldsymbol{\gamma}^\top \mathbf{U}_\mu(\mathbf{x}_i)\right). \end{aligned} \tag{14}$$

The full conditionals for $\boldsymbol{\mu}$, $\boldsymbol{\gamma}$, and τ are, respectively:

$$\begin{aligned} \pi(\boldsymbol{\mu} \mid \boldsymbol{\gamma}, \tau, \mathbf{X}) &\propto \exp\left(\boldsymbol{\mu}^\top \left(\tau \sum_{i=1}^n \mathbf{x}_i + \tau_0 \boldsymbol{\mu}_0\right)\right) \\ &\prod_{i=1}^n \Phi\left((1 - (\mathbf{x}_i^\top \boldsymbol{\mu})^2)^{1/2} \boldsymbol{\gamma}^\top \mathbf{U}_\mu(\mathbf{x}_i)\right), \\ \pi(\boldsymbol{\gamma} \mid \boldsymbol{\mu}, \tau, \mathbf{X}) &\propto \phi_{p-1}(\boldsymbol{\gamma} - \boldsymbol{\xi}; \text{diag}(\boldsymbol{\sigma})) \\ &\prod_{i=1}^n \Phi\left((1 - (\mathbf{x}_i^\top \boldsymbol{\mu})^2)^{1/2} \boldsymbol{\gamma}^\top \mathbf{U}_\mu(\mathbf{x}_i)\right), \\ \pi(\tau \mid \boldsymbol{\mu}, \boldsymbol{\gamma}, \mathbf{X}) &\propto \frac{\tau^{np/2+\alpha-n-1}}{I_{p/2-1}^n(\tau)} \exp\left(\tau \left(\boldsymbol{\mu}^\top \sum_{i=1}^n \mathbf{x}_i - \beta\right)\right). \end{aligned}$$

Scenario 4

When the prior distribution of $\boldsymbol{\gamma}$ is the skew-normal distribution in (9), the posterior distribution of $(\boldsymbol{\mu}, \boldsymbol{\gamma}, \tau)$ by using (7), (9), and (11) is

$$\begin{aligned} \pi(\boldsymbol{\mu}, \boldsymbol{\gamma}, \tau \mid \mathbf{X}, \boldsymbol{\mu}_0, \tau_0, \alpha, \beta, \boldsymbol{\xi}, \boldsymbol{\sigma}) &\propto \frac{\tau^{np/2+\alpha-n-1}}{I_{p/2-1}^n(\tau)} \exp\left(\boldsymbol{\mu}^\top \left(\tau \sum_{i=1}^n \mathbf{x}_i + \tau_0 \boldsymbol{\mu}_0\right) - \beta \tau\right) \\ &\phi_{p-1}(\boldsymbol{\gamma} - \boldsymbol{\xi}; \text{diag}(\boldsymbol{\sigma})) \\ &\times \prod_{i=1}^n \Phi\left((1 - (\mathbf{x}_i^\top \boldsymbol{\mu})^2)^{1/2} \boldsymbol{\gamma}^\top \mathbf{U}_\mu(\mathbf{x}_i)\right) \Phi_{p-1}(\boldsymbol{\gamma} - \boldsymbol{\xi}; \mathbf{D}). \end{aligned} \tag{15}$$

The full conditionals for $\boldsymbol{\mu}$, $\boldsymbol{\gamma}$, and τ are, respectively:

$$\begin{aligned} \pi(\boldsymbol{\mu} \mid \boldsymbol{\gamma}, \tau, \mathbf{X}) &\propto \exp\left(\boldsymbol{\mu}^\top \left(\tau \sum_{i=1}^n \mathbf{x}_i + \tau_0 \boldsymbol{\mu}_0\right)\right) \\ &\prod_{i=1}^n \Phi\left((1 - (\mathbf{x}_i^\top \boldsymbol{\mu})^2)^{1/2} \boldsymbol{\gamma}^\top \mathbf{U}_\mu(\mathbf{x}_i)\right), \\ \pi(\boldsymbol{\gamma} \mid \boldsymbol{\mu}, \tau, \mathbf{X}) &\propto \phi_{p-1}(\boldsymbol{\gamma} - \boldsymbol{\xi}; \text{diag}(\boldsymbol{\sigma})) \Phi_{p-1}(\boldsymbol{\gamma} - \boldsymbol{\xi}; \mathbf{D}) \\ &\times \prod_{i=1}^n \Phi\left((1 - (\mathbf{x}_i^\top \boldsymbol{\mu})^2)^{1/2} \boldsymbol{\gamma}^\top \mathbf{U}_\mu(\mathbf{x}_i)\right), \\ \pi(\tau \mid \boldsymbol{\mu}, \boldsymbol{\gamma}, \mathbf{X}) &\propto \frac{\tau^{np/2+\alpha-n-1}}{I_{p/2-1}^n(\tau)} \exp\left(\tau \left(\boldsymbol{\mu}^\top \sum_{i=1}^n \mathbf{x}_i - \beta\right)\right). \end{aligned}$$

3.3. Sampling From the Posterior Distributions

A general algorithm is presented to obtain the Bayes estimates of the parameters $\Omega = (\mu, \gamma, \tau)$ based on the modified sampling-resampling method (Smith and Gelfand, 1992) and modified Gibbs sampling.

After a sufficient burn-in period, the generated sample $((\mu_1, \gamma_1, \tau_1), (\mu_2, \gamma_2, \tau_2), \dots, (\mu_N, \gamma_N, \tau_N))$ is approximately distributed according to the posterior distribution of $\Omega = (\mu, \gamma, \tau)$. As can be seen in **Algorithm 1**, it is sufficient to generate samples of size k from prior distributions of (μ, γ, τ) which is one of the advantages of this algorithm. By increasing N and k in **Algorithm 1** the approximation increases. When the joint prior distributions are not independent, **Algorithm 1** still has a good performance (Muralidharan and Parikh, 2007). For the joint prior of (μ, τ) in (10), the slice sampler can be used (see McElreath, 2020).

4. THE WASSERSTEIN IMPACT MEASURE

The prior distributions are a crucial part in Bayesian analysis. If the sample size is small, or available data provide only indirect information about the parameters of interest, the prior distribution becomes more important (Carlin and Louis, 2008). Different criteria can be used for prior selection, we refer the reader to Vehtari et al. (2017). Ghaderinezhad et al. (2022) implemented the Wasserstein Impact Measure (WIM) as a measure of the impact of the choice of the prior in a Bayesian approach. In fact it is a convenient way for quantifying prior impact which will help us to choose between two or more priors in a given situation. Suppose Ω is the vector of parameters and $F_1(\cdot)$ and $F_2(\cdot)$ are two cumulative distribution functions (cdfs) of

two posterior distributions $\pi_1(\Omega|\cdot)$ and $\pi_2(\Omega|\cdot)$. The Wasserstein distance between these two posteriors related to two different prior sets is obtained as follows:

$$d_W(\pi_1, \pi_2) = \int_{D_\Omega} |F_1(\Omega; X) - F_2(\Omega; X)| d\Omega, \quad (16)$$

with D_Ω the domain of all possible values of Ω . The Wasserstein distance between two posteriors indicates, at any finite sample size n , how close the posterior distributions are and how similar the related inference will be. This is particularly interesting when considering a simple vs. a complicated, computationally intense prior; if the WIM between them is small, then one can safely use the simpler version. When $n \rightarrow \infty$ the distance tends to 0.

In this section, a simulation study is conducted to compare the different sets of proposed priors in section 3 for $p = 2$ using this measure. Since the cdfs of the posteriors in (12)–(15) are not computable, **Algorithm 1** and Monte Carlo integration are used to obtain the Wasserstein distance. Also, the transport package (Schuhmacher et al., 2020) in the R software offers functions for computing the Wasserstein distance between two sets of samples from different distributions. Most of the functions in this package have been designed for data with two or higher dimensions. For various combinations of the parameters we draw 200 random observations from the *SvM* in (6).

To compare the impact of the normal distribution and skew-normal distribution in (8) and (9) (for $p = 2$) as priors for the skewness parameter γ , the following steps were performed:

- (1) μ and τ were considered as known parameters.
- (2) For the unknown skewness parameter γ , the $N(0, 5)$ and $SN(0, 5, \lambda)$ with $\lambda = -3, -2, -1, 1, 2, 3$ were considered as priors.
- (3) For a generated sample from (6) (the skewing function is the standard normal cdf), with $\mu = 3, \tau = 1$ and $\gamma = 5$, the posteriors $\pi_1(\gamma|\cdot)$ and $\pi_2(\gamma|\cdot)$ emanating from $N(0, 5)$ and $SN(0, 5, \lambda)$, respectively, were considered.
- (4) The posteriors $\pi_1(\gamma|\cdot)$ and $\pi_2(\gamma|\cdot)$ were sampled using **Algorithm 1**, for $n = 10, 15, 20, 25, 30, 35, 40, 50, 100$.
- (5) The Wasserstein distance between the posteriors $\pi_1(\gamma|\cdot)$ and $\pi_2(\gamma|\cdot)$ was estimated, using the transport package and Monte Carlo method with 1,000 repetitions.

Figure 1 (top) shows the calculated Wasserstein distance for different values of λ and n . As expected,

- when λ is close to 0, there are no nearly differences between the posteriors $\pi_1(\Omega|\cdot)$ and $\pi_2(\Omega|\cdot)$ for different values of n .
- by increasing n , the Wasserstein distance decreases. Hence, for large values of n the difference between the posteriors is minimal.

To illustrate the impact of the prior selection for μ and τ the following approach was followed. Assume the normal distribution as the prior for the skewness parameter γ . The posteriors $\pi_2(\Omega|\cdot)$ and $\pi_3(\Omega|\cdot)$, emanating from the informative priors (10) (case 1) and (11) (case 2) were compared with the posterior resulting from the non-informative prior $\mu \sim Uniform(0, 2\pi)$ and $\pi(\tau) \propto 1$,

Algorithm 1: Steps to generate samples from the posteriors by using priors and full conditionals.

- 1: Set the initial values $\Omega^{(0)} = (\mu^{(0)}, \gamma^{(0)}, \tau^{(0)})$ and fix the sample size as $B + N$, where the first B samples are burn-in.
 - 2: Simulate $\mu_1, \mu_2, \dots, \mu_k$, a sample of size k , from $\pi(\mu)$, the prior distribution of μ .
 - 3: Let $\rho_{1i} = \frac{\pi(\mu_i|X, \gamma^{(0)}, \tau^{(0)})}{\pi(\mu_i)}$ for $i = 1, 2, \dots, k$, where $\pi(\mu | X, \gamma, \tau)$ is the full conditional of μ .
 - 4: Select $\mu^{(1)}$ corresponding to $\max \rho_{1i}, i = 1, 2, \dots, k$.
 - 5: Simulate $\tau_1, \tau_2, \dots, \tau_k$ from $\pi(\tau)$, the prior distribution of τ .
 - 6: Let $\rho_{2i} = \frac{\pi(\tau_i|X, \mu^{(1)}, \gamma^{(0)})}{\pi(\tau_i)}$ for $i = 1, 2, \dots, k$, where $\pi(\tau | X, \mu, \gamma)$ is the full conditional of τ .
 - 7: Select $\tau^{(1)}$ corresponding to $\max \rho_{2i}, i = 1, 2, \dots, k$.
 - 8: Simulate $\gamma_1, \gamma_2, \dots, \gamma_k$ from $\pi(\gamma)$, the prior distribution of γ .
 - 9: Let $\rho_{3i} = \frac{\pi(\gamma_i|X, \mu^{(1)}, \tau^{(1)})}{\pi(\gamma_i)}$ for $i = 1, 2, \dots, k$, where $\pi(\gamma | X, \mu, \tau)$ is the full conditional of γ .
 - 10: Select $\gamma^{(1)}$ corresponding to $\max \rho_{3i}, i = 1, 2, \dots, k$.
 - 11: Repeat Steps 1-10 with $\Omega^{(1)} = (\mu^{(1)}, \gamma^{(1)}, \tau^{(1)})$ to obtain $\Omega^{(2)} = (\mu^{(2)}, \gamma^{(2)}, \tau^{(2)})$ and continue until $\Omega^{(B+N)}$ is obtained.
-

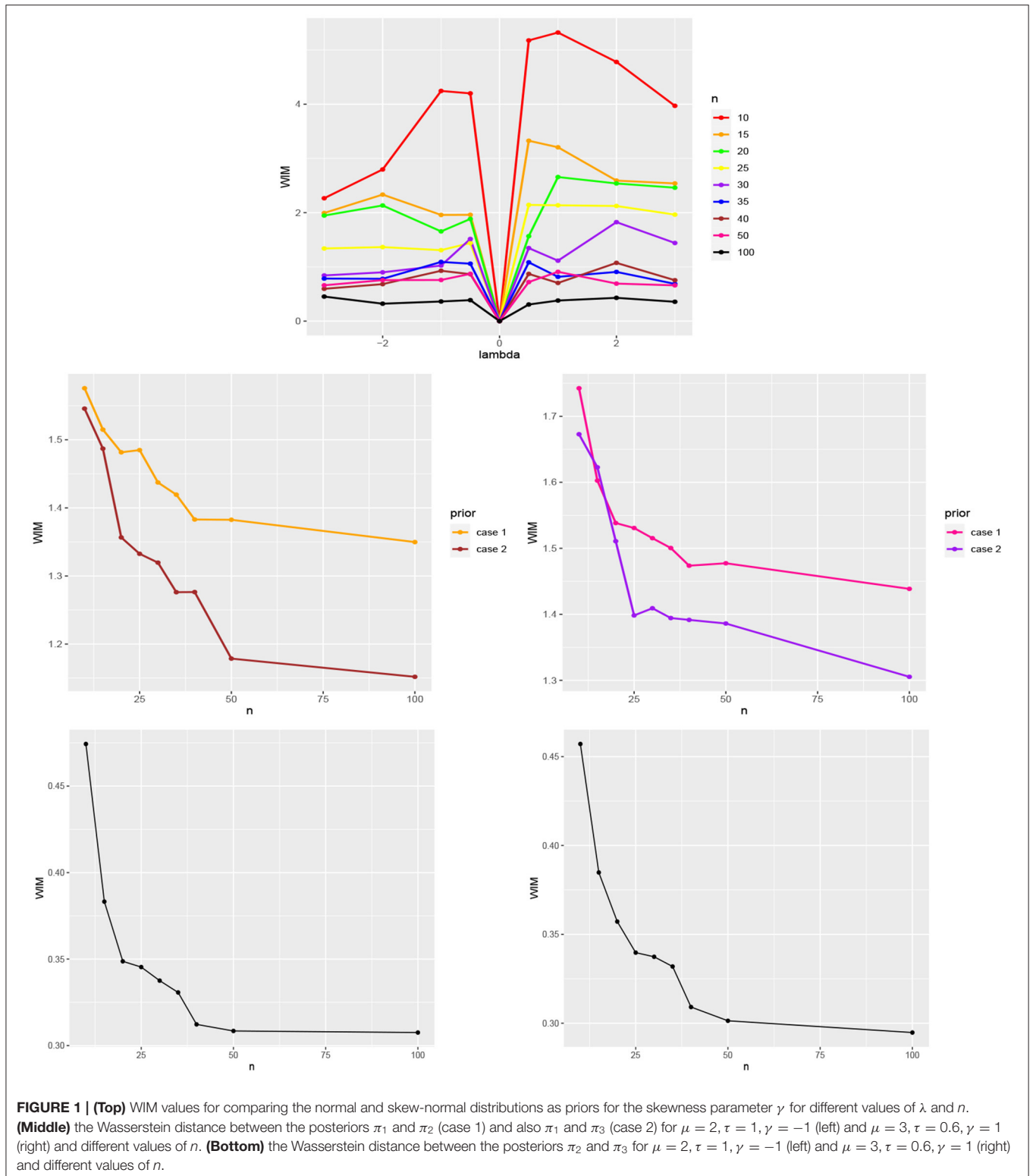


FIGURE 1 | (Top) WIM values for comparing the normal and skew-normal distributions as priors for the skewness parameter γ for different values of λ and n . **(Middle)** the Wasserstein distance between the posteriors π_1 and π_2 (case 1) and also π_1 and π_3 (case 2) for $\mu = 2, \tau = 1, \gamma = -1$ (left) and $\mu = 3, \tau = 0.6, \gamma = 1$ (right) and different values of n . **(Bottom)** the Wasserstein distance between the posteriors π_2 and π_3 for $\mu = 2, \tau = 1, \gamma = -1$ (left) and $\mu = 3, \tau = 0.6, \gamma = 1$ (right) and different values of n .

denoted by $\pi_1(\Omega|\cdot)$. The posteriors were sampled using **Algorithm 1** for $n = 10, 15, 20, 25, 30, 35, 40, 50, 100$. The Wasserstein distances were calculated between them with the transport package and Monte Carlo method with 1,000 repetitions.

Figure 1 (middle) illustrates the obtained Wasserstein distance between the posteriors $\pi_1(\Omega|\cdot)$ and $\pi_2(\Omega|\cdot)$ (case 1) and between $\pi_1(\Omega|\cdot)$ and $\pi_3(\Omega|\cdot)$ (case 2) for $\mu = 2, \tau = 1, \gamma = -1$ (left) and $\mu = 3, \tau = 0.6, \gamma = 1$ (right). **Figure 1** (bottom) shows the Wasserstein distance between the posteriors

TABLE 1 | Bayes estimates of parameters for $\rho = 2$ based on scenario 1 with prior parameters $\mu_0 = 1, \tau_0 = 9, \alpha = 0.5, \beta = 5, \xi = -4, \sigma = 1$, scenario 2 with prior parameters $\mu_0 = 1, \xi = 10, \eta = 0.5, \xi = 0.5, \sigma = 0.5, \lambda = -1$, scenario 3 with prior parameters $\mu_0 = 1, \tau_0 = 9, \alpha = 0.5, \beta = 5, \xi = -4, \sigma = 1$, and scenario 4 with prior parameters $\mu_0 = 0.5, \tau_0 = 0.01, \alpha = 0.5, \beta = 9, \xi = 0.5, \sigma = 0.5, \lambda = -2$ and for $\rho = 3$ based on scenario 3 with prior parameters $\mu_{0_1} = 1, \tau_{0_1} = 5, \mu_{0_2} = 2, \tau_{0_2} = 9, \alpha = 12, \beta = 2, \xi_1 = 1, \sigma_1 = 2, \xi_1 = -2$, and $\sigma_1 = 2$.

	Parameter	Actual	Mean	sd	Q ₁	Median	Q ₃	Mode
Scenario 1 ($\rho = 2$)								
$n = 500$	μ	2.0	2.0931	0.1498	1.8640	2.0719	2.3978	1.9445
	τ	1.0	1.0089	0.2867	0.3646	1.0031	1.5859	1.0884
	γ	-1.0	-0.9848	0.0019	-0.9879	-0.9850	-0.9808	-0.9832
$n = 100$	μ	2.0	2.0005	0.2908	1.3715	1.9963	2.5974	1.9910
	τ	1.0	1.0247	0.2975	0.3902	1.0376	1.5037	1.1037
	γ	-1.0	-0.9749	0.0044	-0.9825	-0.9749	-0.9660	-0.9745
$n = 50$	μ	2.0	1.9953	0.3032	1.4256	1.9830	2.6032	1.9127
	τ	1.0	1.0617	0.3153	0.3991	1.1056	1.5552	0.9032
	γ	-1.0	-0.9755	0.0041	-0.9820	-0.9760	-0.9655	-0.9787
$n = 20$	μ	2.0	1.9386	0.2645	1.5802	1.8836	2.4149	2.1606
	τ	1.0	1.0367	0.2317	0.5845	1.0376	1.4549	1.1174
	γ	-1.0	-0.9753	0.0036	-0.9806	-0.9753	-0.9694	-0.9734
Scenario 2 ($\rho = 2$)								
$n = 500$	μ	3.0	2.9830	0.0598	2.8158	2.9821	3.0978	2.9747
	τ	0.6	0.5715	0.1532	0.2072	0.5661	0.8942	0.5404
	γ	1.0	1.0269	0.0969	0.8801	1.0058	1.2686	0.9833
$n = 100$	μ	3.0	3.1035	0.1040	2.9108	3.0964	3.3410	3.1665
	τ	0.6	0.5967	0.0986	0.4041	0.6027	0.7983	0.6432
	γ	1.0	1.0639	0.0878	0.8412	1.0734	1.1998	1.0157
$n = 50$	μ	3.0	3.1011	0.1428	2.8433	3.1214	3.3861	2.9560
	τ	0.6	0.6389	0.1510	0.3696	0.6456	0.8706	0.5258
	γ	1.0	1.1032	0.1707	0.8495	1.0739	1.4655	0.9629
$n = 20$	μ	3.0	2.9777	0.2014	2.6674	2.9620	3.2630	2.9674
	τ	0.6	0.6571	0.2084	0.3139	0.6539	0.9103	0.5216
	γ	1.0	1.0203	0.1147	0.8313	1.0037	1.2143	1.1558
Scenario 3 ($\rho = 2$)								
$n = 500$	μ	2.0	2.0865	0.1512	1.8514	2.0701	2.4726	2.0578
	τ	1.0	1.0185	0.3038	0.3688	0.9740	1.7048	0.8468
	γ	-1.0	-0.9515	0.3780	-1.5384	-1.0019	-0.0617	-1.1953
$n = 100$	μ	2.0	2.0976	0.1591	1.8746	2.0766	2.4898	1.9338
	τ	1.0	1.0518	0.1968	0.7561	1.0252	1.4816	0.9890
	γ	-1.0	-0.9717	0.3878	-1.5099	-1.0616	0.0234	-0.9795
$n = 50$	μ	2.0	2.0671	0.1407	1.8745	2.0492	2.3671	1.9274
	τ	1.0	1.0503	0.1853	0.7767	1.0282	1.4259	0.9899
	γ	-1.0	-1.0243	0.3483	-1.5132	-1.0963	-0.3674	-0.9612
$n = 20$	μ	2.0	2.0406	0.1480	1.8427	2.0338	2.3473	1.9457
	τ	1.0	1.0650	0.1608	0.8113	1.0457	1.3459	0.9899
	γ	-1.0	-1.0472	0.4333	-1.5279	-1.1187	-0.0711	-0.9801
Scenario 4 ($\rho = 2$)								
$n = 500$	μ	3.0	3.1034	0.1231	2.7620	3.1036	3.3263	3.1364
	τ	0.6	0.5714	0.1592	0.2248	0.5508	0.9029	0.5621
	γ	1.0	0.9923	0.0837	0.8458	0.9904	1.1678	1.0903
$n = 100$	μ	3.0	3.2601	0.0829	3.1212	3.2432	3.4170	3.2175
	τ	0.6	0.5967	0.1442	0.4130	0.5747	0.8620	0.5985
	γ	1.0	0.9095	0.0592	0.8041	0.9173	0.9903	0.9261
	μ	3.0	3.1193	0.0888	2.9994	3.1142	3.2818	3.0141

(Continued)

TABLE 1 | Continued

	Parameter	Actual	Mean	sd	Q ₁	Median	Q ₃	Mode
n = 50	τ	0.6	0.6047	0.1597	0.4307	0.5747	0.8464	0.4968
	γ	1.0	1.0161	0.0861	0.9059	0.9945	1.1983	0.9820
	μ	3.0	2.9734	0.1287	2.8209	2.9589	3.2492	2.8303
n = 20	τ	0.6	0.6150	0.2097	0.4322	0.5775	1.1058	0.4968
	γ	1.0	1.0066	0.0779	0.8978	0.9888	1.1682	0.9820
Scenario 3 ($\rho = 3$)								
n = 500	μ_1	1.0	0.9512	0.0547	0.8766	0.9495	1.0569	0.9386
	μ_2	2.0	2.0385	0.0635	1.9296	2.0296	2.1261	1.9755
	τ	2.0	1.9858	0.9318	0.5731	1.8519	4.3790	1.8534
	γ_1	1.0	0.9081	0.0112	0.8836	0.9097	0.9257	0.9123
n = 100	γ_2	-1.0	-0.9724	0.5095	-1.0492	-0.9856	-0.8903	-1.0477
	μ_1	1.0	0.9996	0.0347	0.9473	0.9970	1.0729	0.9877
	μ_2	2.0	2.0020	0.0329	1.9378	2.0016	2.0690	2.0731
	τ	2.0	2.1327	0.6247	1.5256	2.0795	2.6173	1.8732
n = 50	γ_1	1.0	0.9092	0.0119	0.8845	0.9116	0.9284	0.9130
	γ_2	-1.0	-0.9082	0.0113	-0.9257	-0.9091	-0.8835	-0.9173
	μ_1	1.0	1.1198	0.0171	1.0921	1.1166	1.1533	1.1262
	μ_2	2.0	2.1168	0.0157	2.0908	2.1146	2.1449	2.1096
n = 20	τ	2.0	2.0631	0.3130	1.5308	2.0340	2.6085	2.2340
	γ_1	1.0	0.9089	0.0122	0.8809	0.9095	0.9265	0.9126
	γ_2	-1.0	-0.9073	0.0106	-0.9242	-0.9095	-0.8842	-0.9173
	μ_1	1.0	0.8940	0.0244	0.8390	0.9081	0.9248	0.8927
n = 50	μ_2	2.0	1.9098	0.0760	1.8415	1.8957	2.1224	1.8639
	τ	2.0	1.9192	0.3611	1.3785	1.8825	2.5855	1.8419
	γ_1	1.0	0.9275	0.0109	0.9068	0.9259	0.9482	0.9248
	γ_2	-1.0	-0.9282	0.0086	-0.9426	-0.9282	-0.9124	-0.9360

$\pi_2(\Omega|.)$ and $\pi_3(\Omega|.)$ for $\mu = 2, \tau = 1, \gamma = -1$ (left) and $\mu = 3, \tau = 0.6, \gamma = 1$ (right), respectively. From **Figure 1** (middle and bottom) the following observations can be obtained:

- The impact of the informative priors (10) (case 1) and (11) (case 2) for μ and τ is clearly visible in comparison with the assumed non-informative priors.
- Comparatively, the posterior resulting from prior (11) (case 2) is closer to the non-informative priors.
- By increasing n , the posteriors resulting from the informative priors (10) (case 1) and (11) (case 2) tend to the case of non-informative priors.
- There is less difference between the informative priors (10) (case 1) and (11) (case 2), than with respect to the non-informative priors.
- By increasing n , the posteriors resulting from the informative priors (10) (case 1) and (11) (case 2) approach each other.

We can thus conclude that from moderate sample sizes on, both priors for all three parameters are rather similar (hence one could use the less computationally intense of both priors), but differ clearly from a non-informative one. In order to judge how large the obtained WIM values actually are, bootstrap re-sampling could be done with the original data; we leave this for

future research. Our analysis here is also limited to the chosen parameter values; more simulations need to be done to get a complete picture.

A similar analysis can be performed for $p = 3$.

5. SYNTHETIC DATA ANALYSIS

In this section, to assess the performance of the Bayesian approach for obtaining the estimates of $\Omega = (\mu, \tau, \gamma)$, a synthetic data analysis was conducted to obtain the Bayes estimates of the parameters of the SvM distribution (6). We generated samples of size $N = 20, 50, 100, 500$ from the posterior distributions (12)-(15) (scenarios 1-4) with a burn-in period of 5,000 and $k = 500$, using **Algorithm 1** (the values of these parameters are written down in the respective tables). It is noteworthy that steps 2-7 in **Algorithm 1** are combined for scenarios 1 and 2. Bayes estimates of the parameters μ, τ and γ were obtained under the squared error, absolute error and zero-one loss functions by calculating mean, median, and mode of the generated samples, respectively.

The results for $p = 2$ and $p = 3$ including the sample mean, standard deviation, quartiles, and mode of the posterior distribution are summarized in **Table 1** for each of the scenarios. As can be seen in **Table 1** the obtained Bayes estimates are

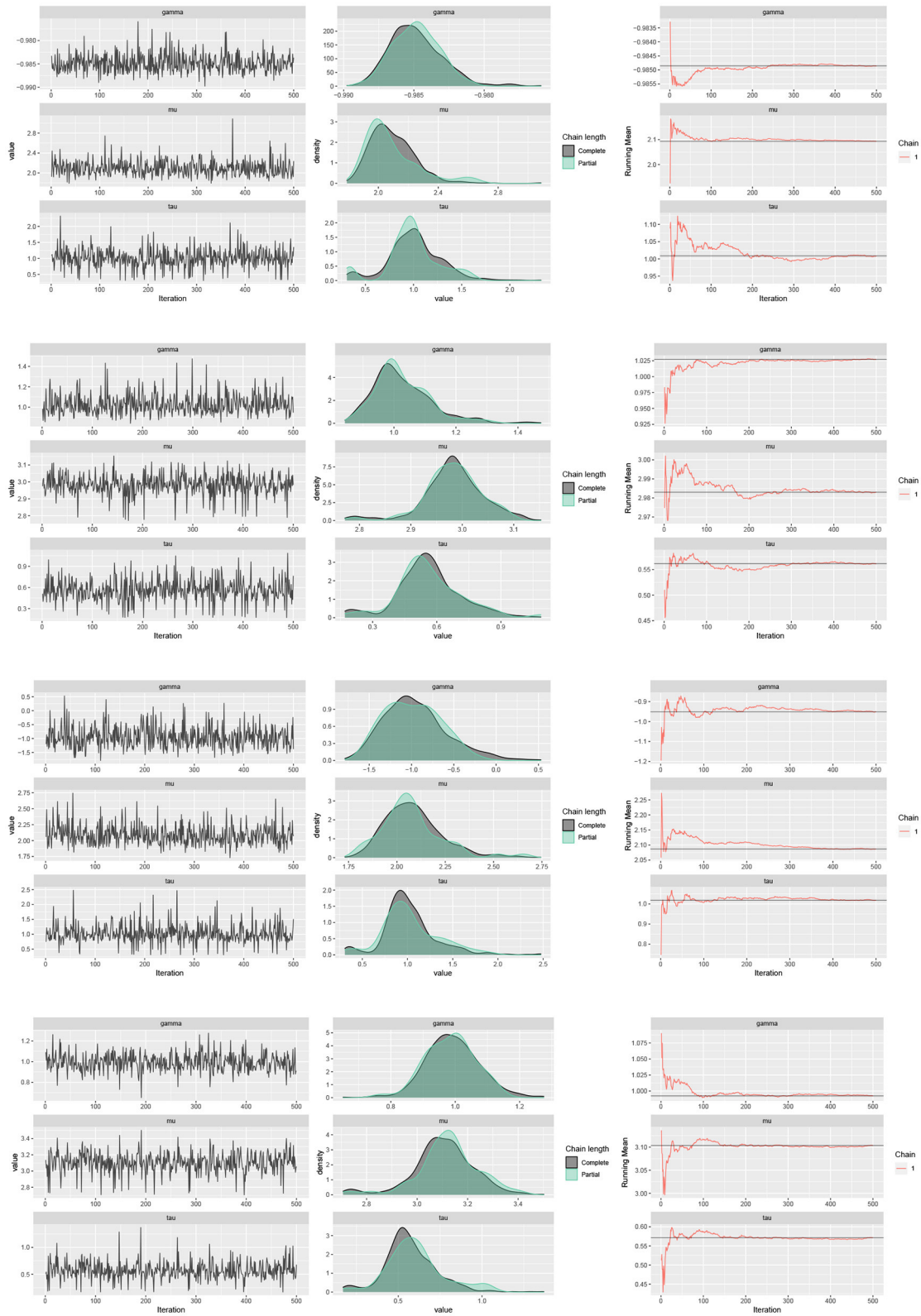


FIGURE 2 | Traceplots, mean running and estimated posterior pdf plots of generated samples for (μ, τ, γ) in **Table 1** for $p = 2, n = 500$ and scenario 1 (first row), scenario 2 (second row), scenario 3 (third row), and scenario 4 (fourth row).

close to the actual values of the parameters. In addition, for small sample sizes our proposed Bayesian approach still provides accurate estimates.

The traceplots of the generated samples from the posteriors, the compare-partial plots and the running mean plots are shown in **Figure 2** ($p = 2$) and **Figure 3** ($p = 3$) for each of the

scenarios and $p = 2$ and 3 using the ggmcmc package in R (Fernández-i-Marín, 2016). A traceplot is an essential plot for evaluating convergence and diagnosing chain problems. It shows the time series of the sampling process and the expected outcome is to get a traceplot that looks completely random. A compare-partial plot provides overlapped kernel density plots

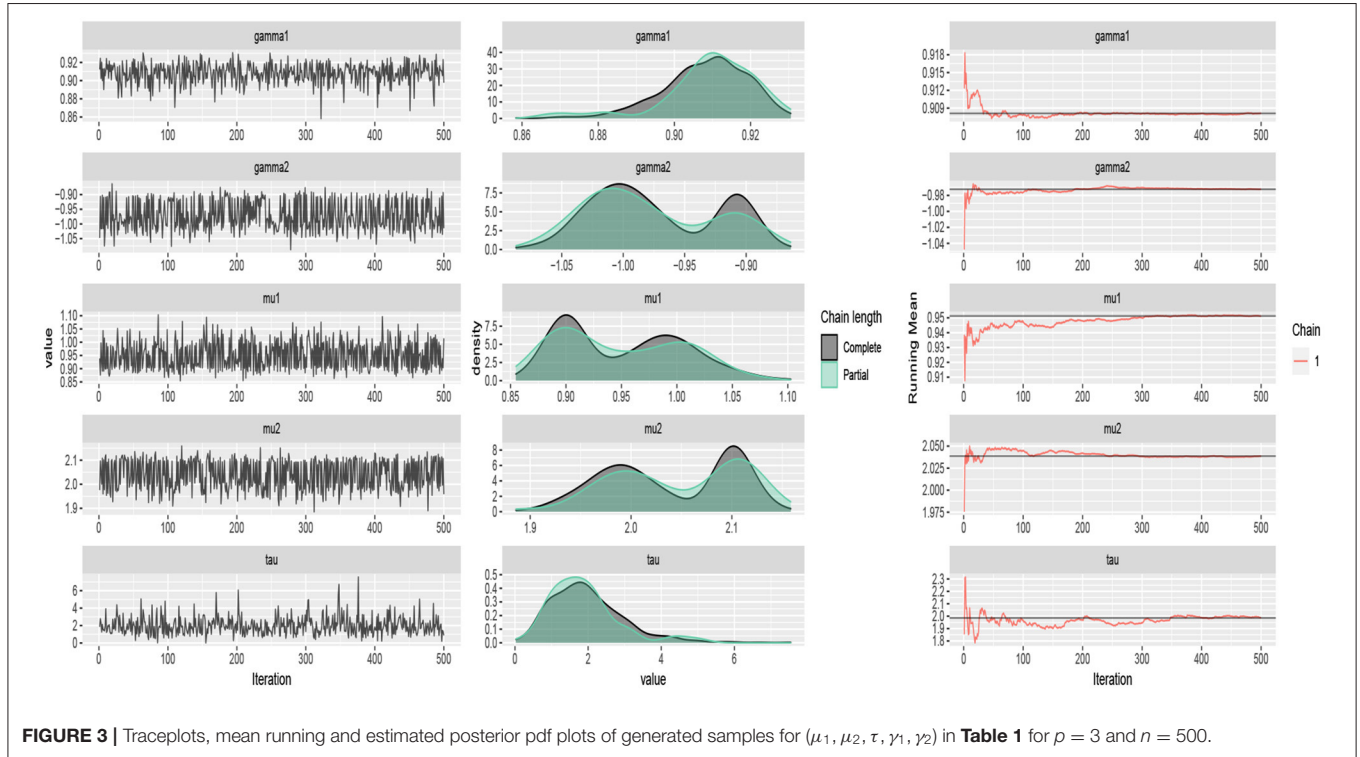


FIGURE 3 | Traceplots, mean running and estimated posterior pdf plots of generated samples for $(\mu_1, \mu_2, \tau, \gamma_1, \gamma_2)$ in **Table 1** for $\rho = 3$ and $n = 500$.

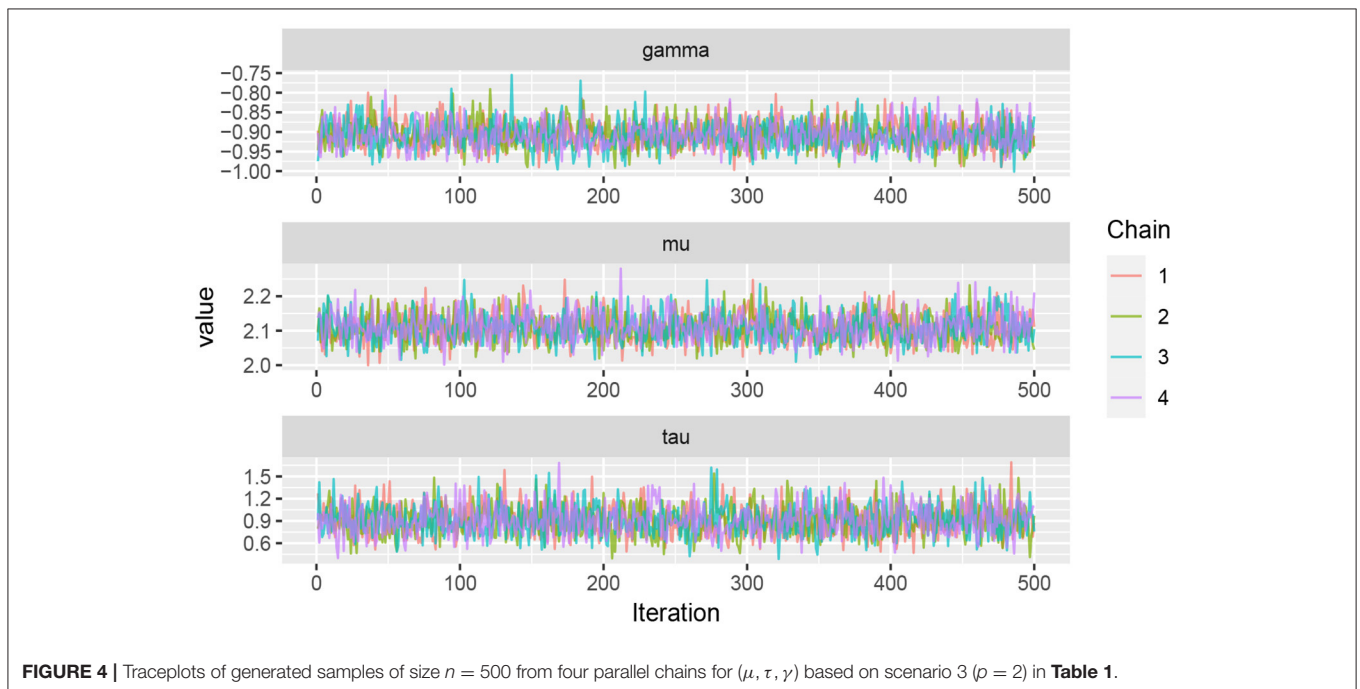
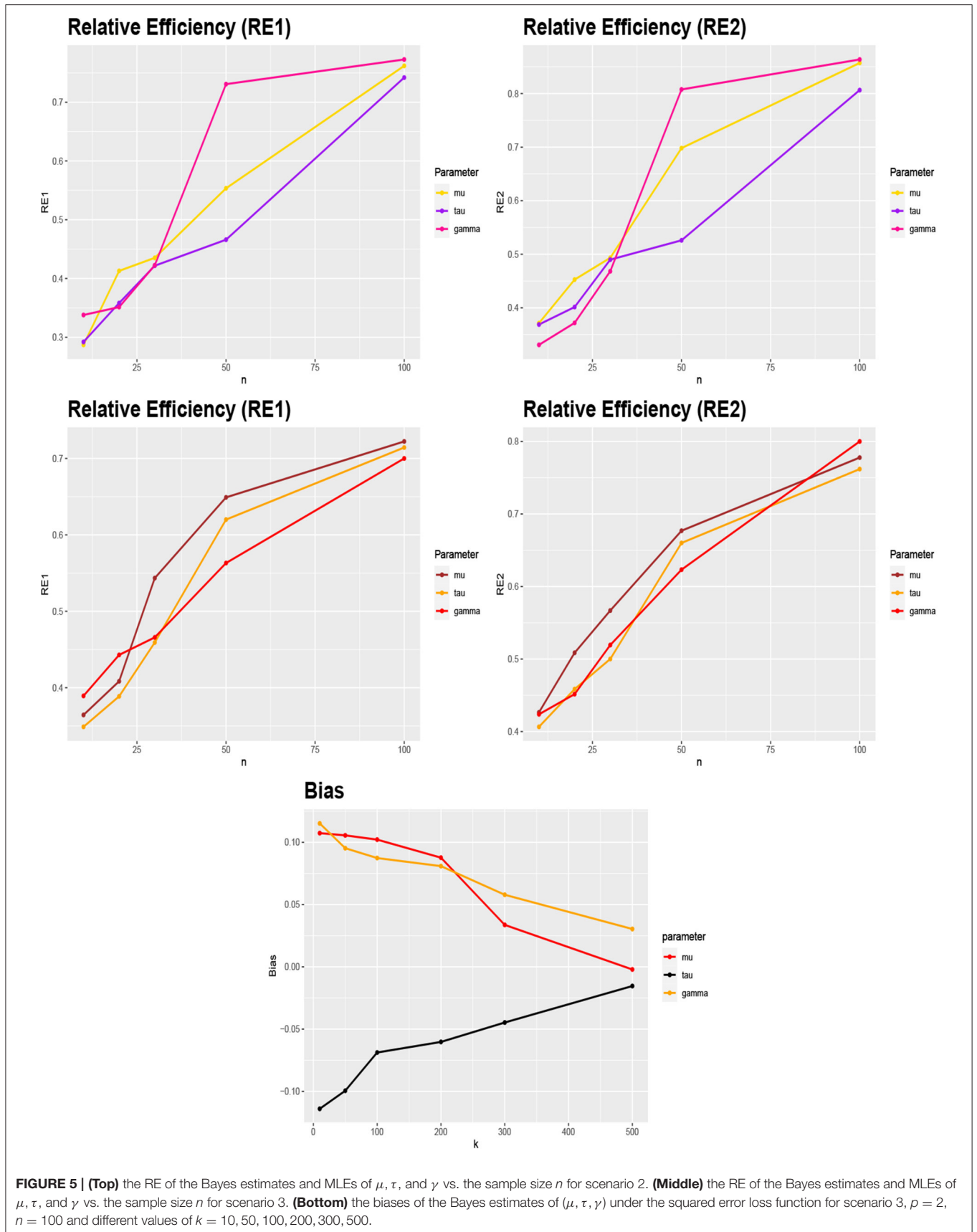
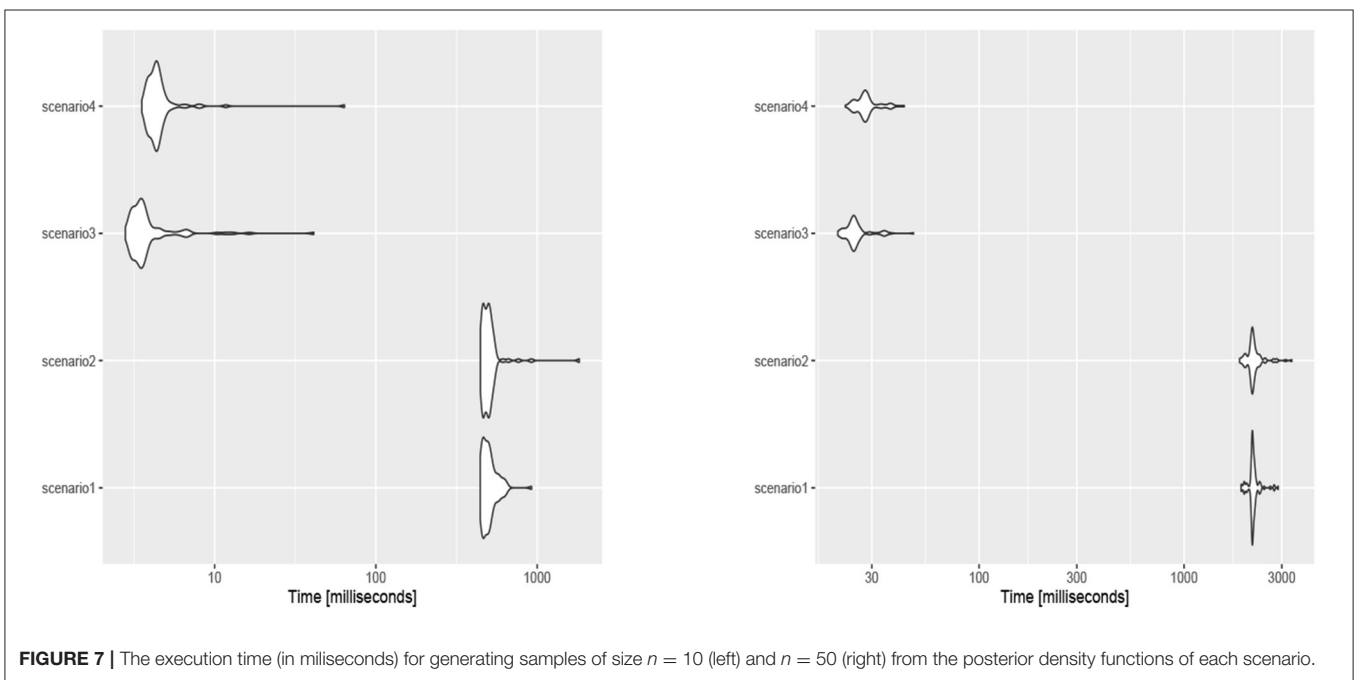
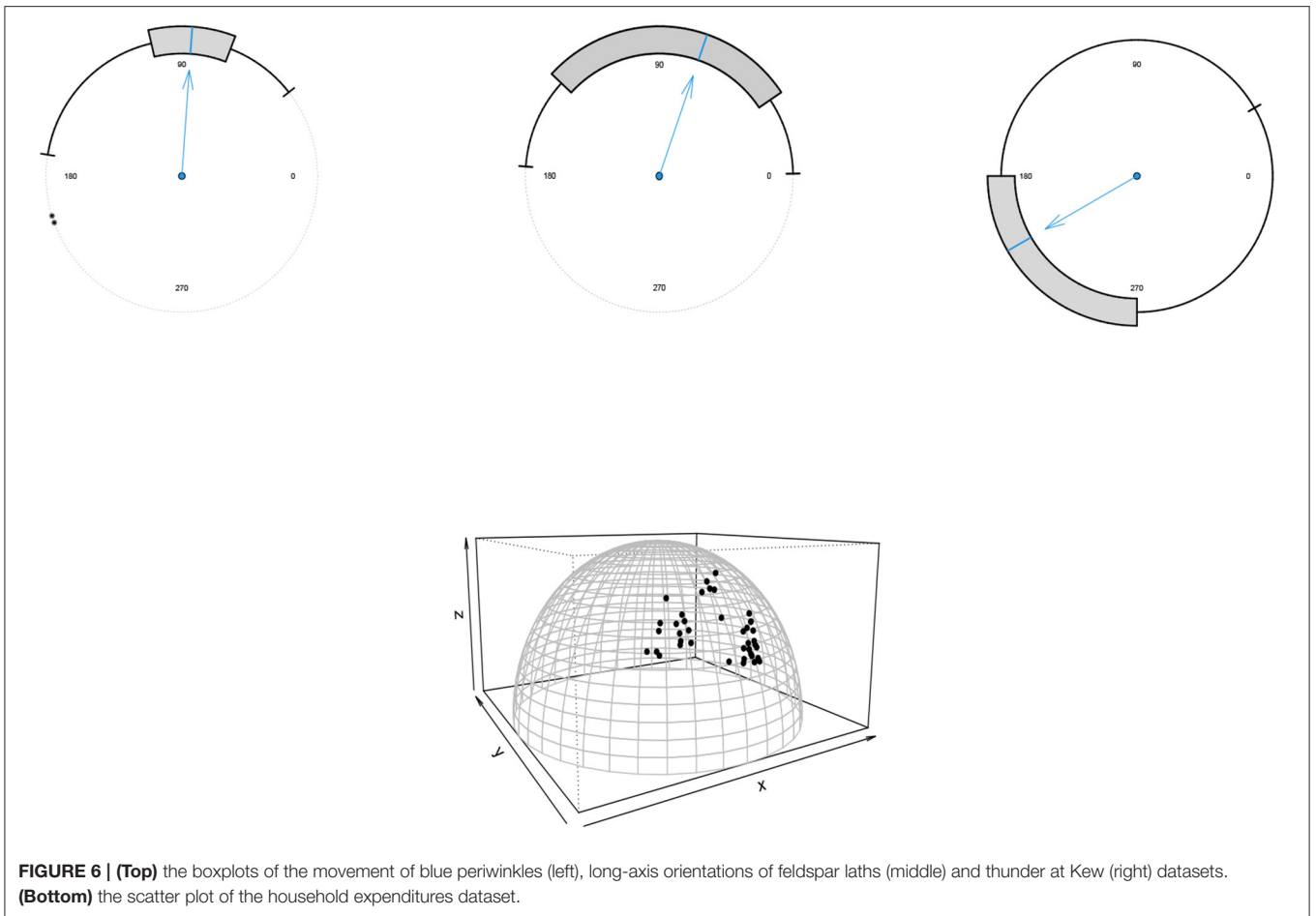


FIGURE 4 | Traceplots of generated samples of size $n = 500$ from four parallel chains for (μ, τ, γ) based on scenario 3 ($\rho = 2$) in **Table 1**.





that compare the last part of the chain (the last 10% of the values, in green) with the whole chain (in black). Ideally, the initial and final parts of the chain have to be sampling in the same target distribution, so the overlapped densities

should be similar. In addition to the traceplots, the running mean plot of the chains is very useful to find within-chain convergence issues. A time series of the running mean of the chain allows to check whether the chain is slowly or quickly

TABLE 2 | Bayes estimates of parameters based on scenario 1 with prior parameters $\mu_0 = 2, \zeta = 2, \eta = 1, \xi = 5,$ and $\sigma = 2$ for the movement of blue periwinkles dataset.

Parameter	Mean	Mode	sd	Q ₁	Median	Q ₃
μ	0.8786	0.8854	0.04155	0.7952	0.8785	0.9574
τ	1.4033	1.2918	0.1729	1.0746	1.4021	1.7409
γ	5.8574	5.8183	0.3974	5.0180	5.8851	6.5687

TABLE 3 | Bayes estimates of parameters based on scenario 3 with prior parameters $\mu_0 = 0, \tau_0 = 8, \alpha = 5, \beta = 2, \xi = 3, \sigma = 1$ for the long-axis orientations of feldspar laths dataset.

Parameter	Mean	Mode	sd	Q ₁	Median	Q ₃
μ	0.0442	0.0497	0.0134	0.0180	0.0451	0.0697
τ	0.3438	0.2616	0.0970	0.1571	0.3504	0.5301
γ	5.0906	5.0972	0.0109	5.0729	5.0894	5.1158

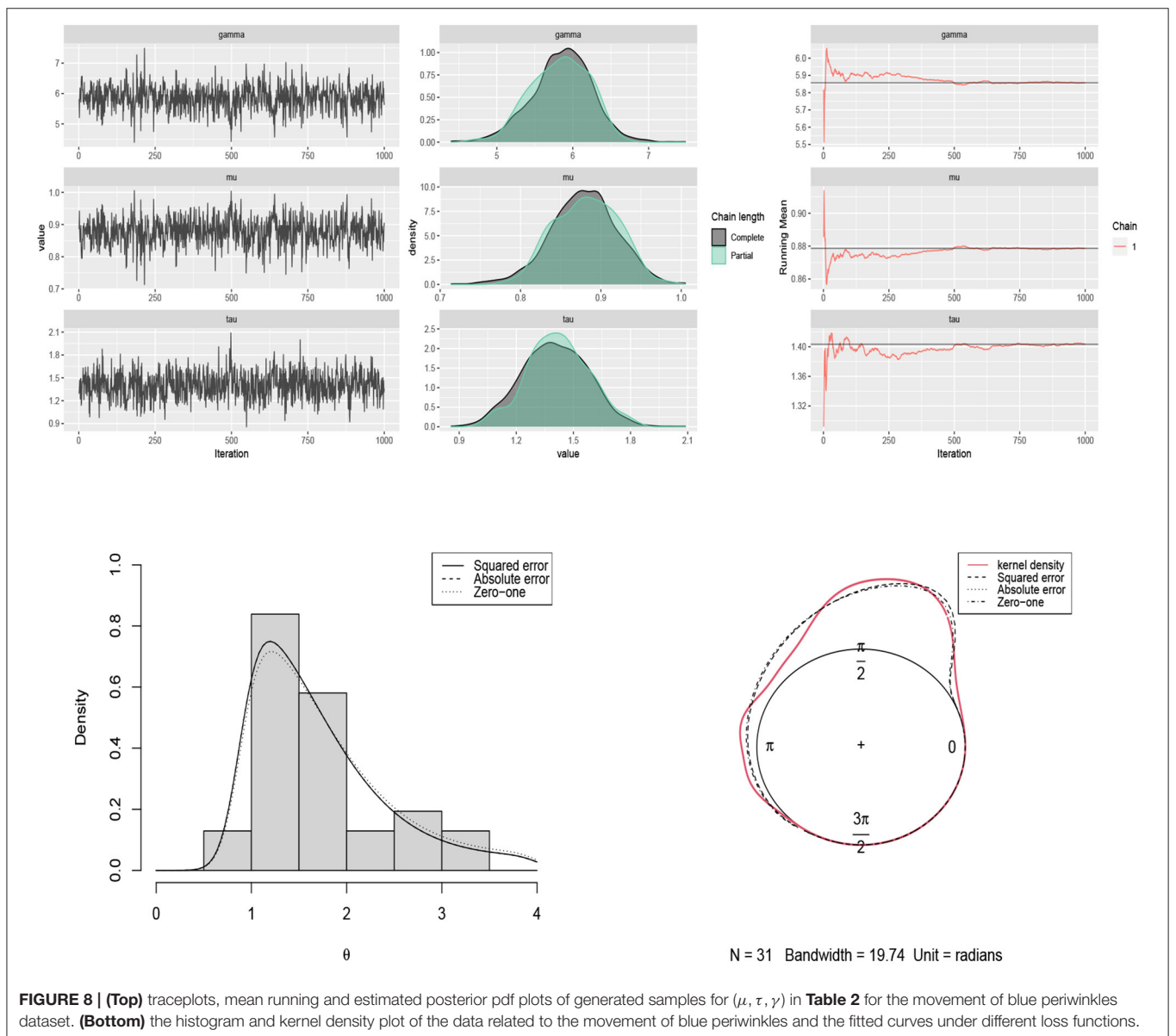


FIGURE 8 | (Top) traceplots, mean running and estimated posterior pdf plots of generated samples for (μ, τ, γ) in **Table 2** for the movement of blue periwinkles dataset. (Bottom) the histogram and kernel density plot of the data related to the movement of blue periwinkles and the fitted curves under different loss functions.

approaching its target distribution. The expected output is a line that quickly approaches the overall mean. **Figures 2, 3** confirm the convergence of the chains and show that the modified Gibbs sampler recovers the values that actually come from the target posterior distributions.

Running multiple independent chains in parallel is necessary to access the representativeness of the chains. If the multiple chains are not well mixed, the convergence of the chains is suspected (Kruschke, 2014; Vehtari et al., 2021). Therefore, four independent chains were run in parallel for scenario 3 ($p = 2$) in **Table 1** to make the inference more robust and reliable. The results are shown in **Figure 4** which confirm the convergence.

To compare the efficiency of Bayes estimates with respect to the maximum likelihood estimations (MLE), the mean squared

errors (MSE) of MLEs and Bayes estimates of parameters under the squared error and absolute error loss functions were obtained for scenario 2 and 3 and $n = 10, 20, 30, 50, 100$ using a

TABLE 4 | Bayes estimates of parameters based on scenario 4 with prior parameters $\mu_0 = 3, \tau_0 = 2, \alpha = 5, \beta = 6, \xi = 0.5, \sigma = 0.5$, and $\lambda = -2$ for the thunder at Kew dataset.

Parameter	Mean	Mode	sd	Q_1	Median	Q_3
μ	3.0943	3.1601	0.1598	2.8035	3.1032	3.3948
τ	0.9675	1.2178	0.3597	0.2863	0.7868	1.6630
γ	0.9179	1.0629	0.1610	0.7402	1.1234	1.3237

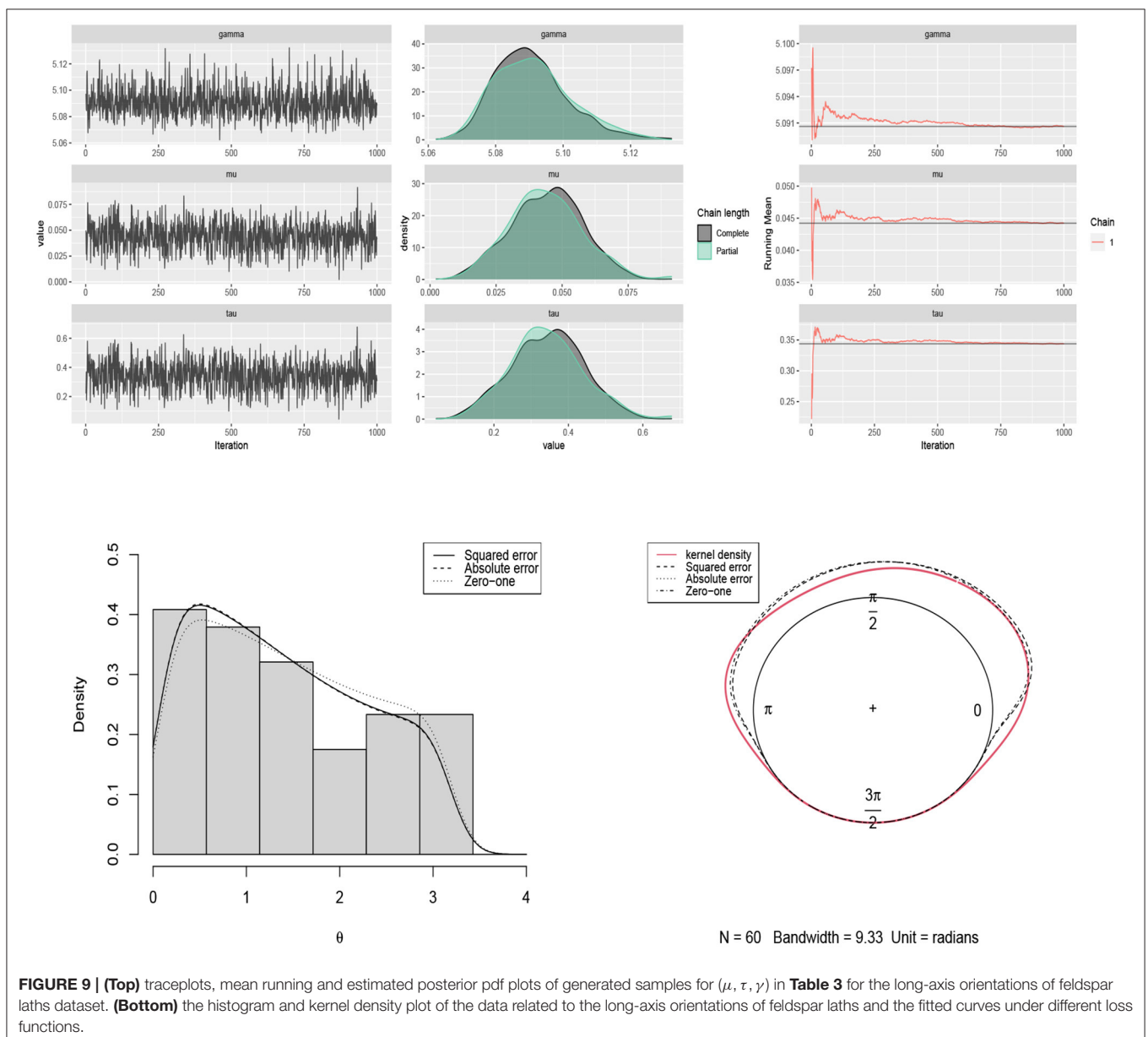


FIGURE 9 | (Top) traceplots, mean running and estimated posterior pdf plots of generated samples for (μ, τ, γ) in **Table 3** for the long-axis orientations of feldspar laths dataset. (Bottom) the histogram and kernel density plot of the data related to the long-axis orientations of feldspar laths and the fitted curves under different loss functions.

Monte Carlo simulation with 500 replications. Then, the relative efficiency (RE) was computed as

$$RE_1 = \frac{MSE(\hat{\Omega})}{MSE(\tilde{\Omega})}, RE_2 = \frac{MSE(\hat{\tilde{\Omega}})}{MSE(\tilde{\Omega})},$$

where $\tilde{\Omega}$ is the MLE of $\Omega = (\mu, \tau, \gamma)$ and $\hat{\Omega}$ and $\hat{\tilde{\Omega}}$ are the Bayes estimates of Ω under the squared error and absolute error loss functions, respectively. The results are shown in Figure 5 for scenario 2 (top) and scenario 3 (middle) and $\mu = 3, \tau = 0.6, \gamma = 1$.

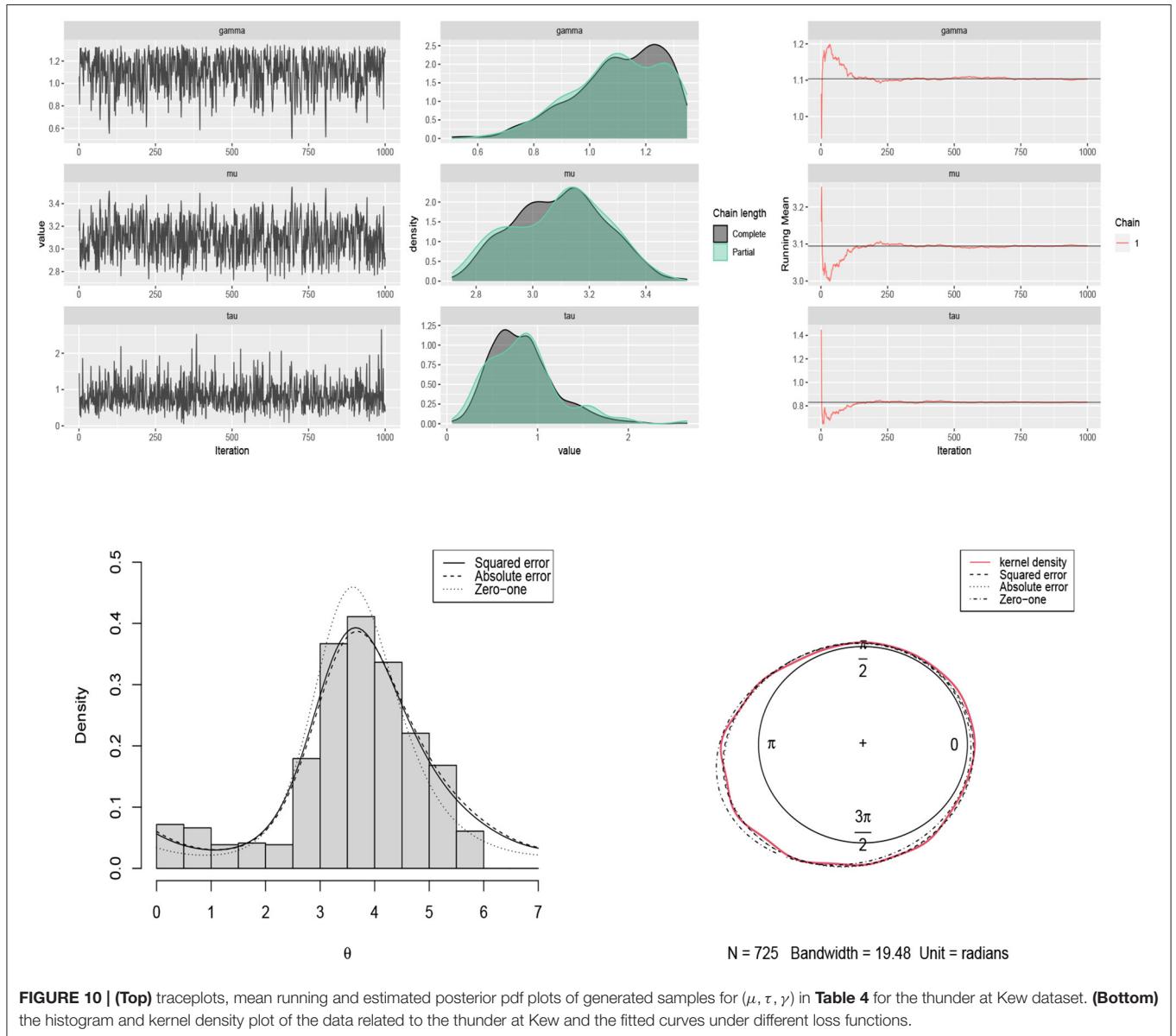
From Figure 5 (top and middle) the following general conclusions can be observed:

- Our proposed Bayesian approach provides more accurate estimates for parameters in comparison with the maximum likelihood method for small values of n .

- The obtained Bayes estimates under the squared error loss function have less MSE than the estimates based on absolute error loss function.

TABLE 5 | Bayes estimates of parameters based on scenario 3 with prior parameters $\mu_{0_1} = 3, \tau_{0_1} = 2, \mu_{0_2} = 3, \tau_{0_2} = 4, \alpha = 20, \beta = 1, \xi_1 = 0, \sigma_1 = 3, \xi_2 = 0$, and $\sigma_2 = 2$ for the household expenditures dataset.

Parameter	Mean	Mode	sd	Q ₁	Median	Q ₃
μ_1	4.9376	4.9550	0.0392	4.8634	4.9362	5.0156
μ_2	3.9085	3.9300	0.0384	3.8327	3.9080	3.9828
τ	9.1075	8.9427	0.9499	6.9295	9.1673	10.7127
γ_1	1.8475	1.8412	0.0189	1.8019	1.8490	1.8773
γ_2	0.1974	0.2449	0.0365	0.1184	0.2008	0.2550



- By increasing n , our proposed Bayesian approach has a similar performance as the maximum likelihood method.

Finally, to investigate the rule of k in **Algorithm 1**, the biases of the Bayes estimates of (μ, τ, γ) under the squared error loss function were obtained for scenario 3 ($p = 2$) in **Table 1**, $n = 100$ and different values of $k = 10, 50, 100, 200, 300, 500$ using a Monte Carlo simulation with 500 replications. The results are shown in **Figure 5** (bottom) which demonstrate that, by increasing k , the bias tends to zero and thus, the accuracy of estimates increases.

6. DATA APPLICATION

In what follows, the proposed Bayesian approach's performance for $p = 2$ is demonstrated through three datasets with different sizes $n = 31, 60, 725$ with the skew-von Mises distribution in (6) as assumed model. The circular boxplots (Buttarazzi et al., 2018)

of the datasets are shown in **Figure 6** (top) and confirm the skew pattern of the datasets.

The obtained results in section 5 show all the assumed scenarios provide accurate estimates for the parameters. However, based on the obtained results in section 5 with the WIM, we propose scenario 3 or 4 for obtaining the Bayes estimates to avoid time intensive computations when the sample size is not small (see **Figure 7**). The justification is that scenarios 1 and 2 need the slice sampler in **Algorithm 1** to generate samples from the joint prior (10). Therefore, the Bayes estimates of the parameters μ, τ , and γ were obtained based on scenario 1 for the movement of blue periwinkles dataset ($n = 31$); scenarios 3 and 4 for the long-axis orientations of feldspar laths ($n = 60$) and the thunder at Kew ($n = 725$) datasets, respectively. See below for the description of said datasets.

For $p = 3$, a dataset of size $n = 40$ including the expenditures of households is considered. **Figure 6** (bottom) shows the scatter plot of the data. For all the datasets to obtain the Bayes estimates

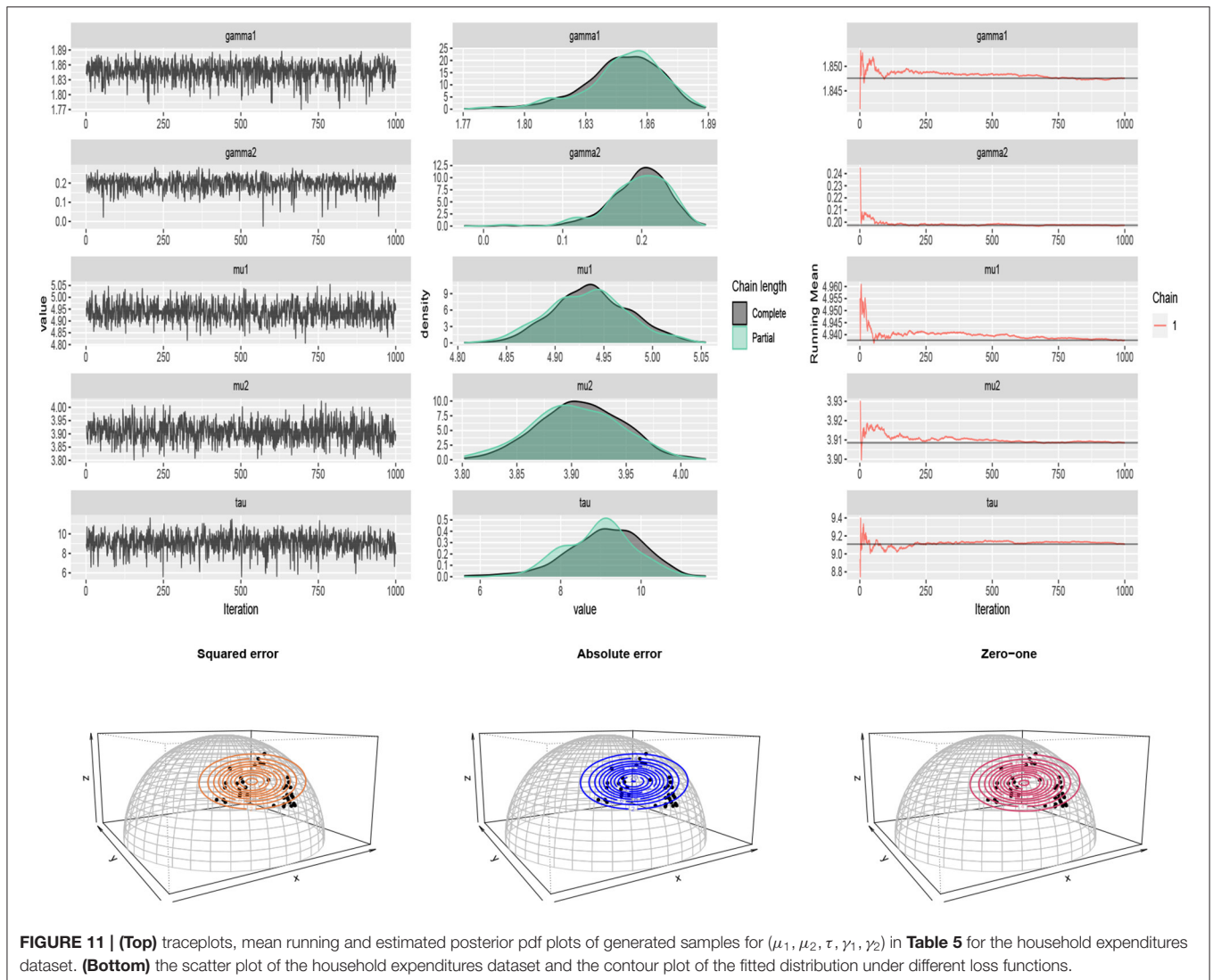


FIGURE 11 | (Top) traceplots, mean running and estimated posterior pdf plots of generated samples for $(\mu_1, \mu_2, \tau, \gamma_1, \gamma_2)$ in **Table 5** for the household expenditures dataset. **(Bottom)** the scatter plot of the household expenditures dataset and the contour plot of the fitted distribution under different loss functions.

we generated samples of size $N = 1,000$ from the posterior distributions using **Algorithm 1** with a burn-in period of 10,000 and $k = 500$.

In what follows, we shall describe the individual datasets in detail. Since the conclusion is more or less the same for all $p = 2$ settings, we already write it down here. It is observed that the proposed Bayesian approach with the skew von Mises distribution as the underlying model provides a good fit to the datasets. Generally, the obtained estimates based on the squared error and absolute error loss functions are more accurate.

6.1. Movement of Blue Periwinkles

A real dataset including the movement directions of 31 blue periwinkles, *Nodilittorina unifasciata*, in degrees is considered (Fisher, 1995). The data was collected from a series of experiments which were done on the distances and directions that small blue periwinkles moved after the transplantation to downshore at a specific height where they live normally. The test of Pewsey (2002) confirms that the underlying distribution for this dataset is asymmetric (p-value = 0.0000). This test is defined based on the sample second sine moment $\bar{b}_2 = \frac{1}{n} \sum_{i=1}^n \sin 2(\theta_i - \bar{\theta})$ where $\bar{\theta}$ is the sample mean direction. The large values of $|\frac{\bar{b}_2}{\sqrt{\text{var}(\bar{b}_2)}}|$ compared with the quantiles of the standard normal distribution lead to the rejection of symmetry. For more details see Pewsey (2002).

The Bayes estimates of parameters are obtained by using scenario 1 based on squared error, absolute error and zero-one loss functions. The results are summarized in **Table 2**. The traceplots of generated samples from the posterior, the compare-partial and mean running plots are shown in **Figure 8** (top). The kernel density plot and histogram of the data along with the fitted curves under different loss functions are shown in **Figure 8** (bottom).

6.2. Long-Axis Orientations of Feldspar Laths

Another dataset including the measurements of long-axis orientation of 60 feldspar laths in basalt (Fisher, 1995) is considered. The symmetry test of Pewsey (2002) confirms the skew pattern of the data in **Figure 6** (p-value = 0.0000). The Bayes estimates of parameters are obtained by using scenario 3 based on squared error, absolute error and zero-one loss functions. The results are summarized in **Table 3**. The traceplots of generated samples from the posterior, the compare-partial, and the mean running plots are shown in **Figure 9** (top). The histogram and kernel density plot of the data and the fitted curves under different loss functions are shown in **Figure 9** (bottom).

6.3. Thunder at Kew

A grouped frequency data set consisting of 725 observations about the number of times that thunder was heard at Kew (England) during each two hourly interval of the day in the summers of 1910–1935 is considered (Mardia, 1975). In this case, each 15° on the circle represents 1 h. According

to the test of Pewsey (2002), the underlying distribution for this data set is not symmetric (p-value = 0.0000). The Bayes estimates of parameters are obtained by using scenario 4 based on squared error, absolute error and zero-one loss functions. The results are summarized in **Table 4**. The traceplots of generated samples from the posterior, the compare-partial and mean running plots are shown in **Figure 10** (top). The histogram and kernel density plot of the data and the fitted curves under different loss functions are shown in **Figure 10** (bottom).

6.4. Household Expenditures

For $p = 3$, a sub data from the dataset available in the HSAUR2 package (Everitt and Hothorn, 2017) in R is considered. The entire data was collected from a survey on household expenditures in four commodity groups of housing, food, goods, and service. It includes the expenses of 20 single men and 20 single women. We considered variables housing, food, and service and normalized. After applying cosine transformation (5), the SFvML was fitted on the data and the Bayes estimates of the parameters were obtained. The results are summarized in **Table 5**. The traceplots of generated samples from the posterior and the compare-partial and mean running plots are shown in **Figure 11** (top). The scatter plot of the data and the contour plot of the fitted distribution under different loss functions are shown in **Figure 11** (bottom).

7. CONCLUSION

Since the assumption that data is rotationally-symmetric is often rejected (Pewsey, 2002; Ley and Verdebout, 2014; Ameijeiras-Alonso and Ley, 2020; Ameijeiras-Alonso et al., 2021), in this paper, we have presented a Bayesian analysis for the skew-rotationally-symmetric FvML distribution. For the first time in Bayesian analysis of directional data the impact of the proposed priors in the set up has been compared using the Wasserstein Impact Measure. Using this measure can give guidance to the practitioner to avoid computationally intensive priors if a simpler prior has similar impact. An algorithm has been used based on modified Gibbs sampling and weighted bootstrap resampling for generating samples from posterior distributions. This coming together of Bayesian methods and skew distributions in the directional domain promises new research interest.

DATA AVAILABILITY STATEMENT

All relevant references for data are contained within the article.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was based upon research supported in part by the Visiting Professor programme, University of Pretoria and the National Research Foundation (NRF) of South Africa, SARChI Research Chair UID: 71199; Ref.: IFR170227223754 grant No. 109214; Ref.: SRUG190308422768 grant No. 120839, 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 to be attributed to the CoE-MaSS or the NRF. Christophe Ley's research is supported by the FWO Krediet aan Navorsers grant with reference number 1510391N.

ACKNOWLEDGMENTS

We would like to thank the Associate Editor and the reviewers for their constructive comments, which improved this paper.

REFERENCES

- Abe, T., and Pewsey, A. (2011). Sine-skewed circular distributions. *Stat. Pap.* 52, 683–707. doi: 10.1007/s00362-009-0277-x
- Ahmed, S. E. (2017). *Big and Complex Data Analysis: Methodologies and Applications*. Switzerland: Springer.
- Ameijeiras-Alonso, J., and Ley, C. (2020). Sine-skewed toroidal distributions and their application in protein bioinformatics. *Biostatistics*. kxaa039. doi: 10.1093/biostatistics/kxaa039
- Ameijeiras-Alonso, J., Ley, C., Pewsey, A., and Verdebout, T. (2021). On optimal tests for circular reflective symmetry about an unknown central direction. *Stat. Pap.* 62, 1651–1674. doi: 10.1007/s00362-019-01150-7
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Stat.* 12, 171–178. doi: 10.6092/ISSN.1973-2201/711
- Bagchi, P. (1988). *Bayesian Analysis of Directional Data* (Ph.D. thesis). University of Toronto, Toronto, ON, Canada.
- Bagchi, P. (1994). Empirical Bayes estimation in directional data. *J. Appl. Stat.* 21, 317–326. doi: 10.1080/757583874
- Bagchi, P., and Guttman, I. (1988). Theoretical considerations of the multivariate von Mises-Fisher distribution. *J. Appl. Stat.* 15, 149–169. doi: 10.1080/02664768800000022
- Bagchi, P., and Kadane, J. B. (1991). Laplace approximations to posterior moments and marginal distributions on circles, spheres, and cylinders. *Can. J. Stat.* 19, 67–77. doi: 10.2307/3315537
- Bangert, M., Hennig, P., and Oelfke, U. (2010). "Using an infinite von Mises-Fisher mixture model to cluster treatment beam directions in external radiation therapy," in *2010 Ninth International Conference on Machine Learning and Applications* (Washington, DC: IEEE), 746–751.
- Bhattacharya, S., and SenGupta, A. (2009). Bayesian inference for circular distributions with unknown normalising constants. *J. Stat. Plan. Inference* 139, 4179–4192. doi: 10.1016/j.jspi.2009.06.008
- Buttarazzi, D., Pandolfo, G., and Porzio, G. C. (2018). A boxplot for circular data. *Biometrics* 74, 1492–1501. doi: 10.1111/biom.12889
- Carlin, B. P., and Louis, T. A. (2008). *Bayesian Methods for Data Analysis*. Boca Raton: CRC Press.
- Damien, P., and Walker, S. (1999). A full Bayesian analysis of circular data using the von Mises distribution. *Can. J. Stat.* 27, 291–298. doi: 10.2307/3315639
- Dowe, D. L., Oliver, J. J., Baxter, R. A., and Wallace, C. S. (1996). "Bayesian estimation of the von Mises concentration parameter," in *Maximum Entropy and Bayesian Methods*, (Cambridge: Springer), 51–60.
- Downs, T. (2003). Spherical regression. *Biometrika* 90, 655–668. doi: 10.1093/biomet/90.3.655
- Everitt, B. S., and Hothorn, T. (2017). *Package 'hsaur2'*. Available online at: <https://CRAN.R-project.org/package=HSAUR2>
- Fernández-i-Marín, X. (2016). *ggmcmc: analysis of mcmc samples and Bayesian inference*. *J. Stat. Softw.* 70, 1–20. doi: 10.18637/jss.v070.i09
- Fisher, N., and Lee, A. (1994). Time series analysis of circular data. *J. R. Stat. Soc. Series B (Methodol.)* 56, 327–339. doi: 10.1111/j.2517-6161.1994.tb01981.x
- Fisher, N. I. (1995). *Statistical Analysis of Circular Data*. Cambridge: Cambridge University Press.
- Fisher, R. A. (1953). Dispersion on a sphere. *Proc. R. Soc. London Series A. Math. Phys. Sci.* 217, 295–305. doi: 10.1098/rspa.1953.0064
- Ghaderinezhad, F., Ley, C., and Serrien, B. (2022). The Wasserstein impact measure (WIM): a generally applicable, practical tool for quantifying prior impact in Bayesian statistics. *Comput. Stat. Data Anal.* [Epub ahead of print].
- Guttorp, P., and Lockhart, R. A. (1988). Finding the location of a signal: a Bayesian analysis. *J. Amer. Stat. Assoc.* 83, 322–330. doi: 10.1080/01621459.1988.10478601. [Epub ahead of print].
- Härdle, W., Lu, H. H.-S., and Shen, X. (2018). *Handbook of Big Data Analytics*, Cham: Springer.
- Hornik, K., and Grün, B. (2013). On conjugate families and Jeffreys priors for von Mises-Fisher distributions. *J. Stat. Plan. Inference* 143, 992–999. doi: 10.1016/j.jspi.2012.11.003
- Kikuchi, D. (1982). *Directional Data Abstracts: 1972-1981*. Technical Report 265.
- Kruschke, J. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Oxford: Academic Press.
- Leong, P., and Carlile, S. (1998). Methods for spherical data analysis and visualization. *J. Neurosci. Methods* 80, 191–200. doi: 10.1016/S0165-0270(97)00201-X
- Ley, C., and Verdebout, T. (2014). Simple optimal tests for circular reflective symmetry about a specified median direction. *Stat. Sinica* 24, 1319–1339. doi: 10.5705/ss.2013.083
- Ley, C. and Verdebout, T. (2017a). *Modern Directional Statistics*. Boca Raton: CRC Press.
- Ley, C., and Verdebout, T. (2017b). Skew-rotationally-symmetric distributions and related efficient inferential procedures. *J. Multivariate Anal.* 159, 67–81. doi: 10.1016/j.jmva.2017.02.010
- Mardia, K. V. (1975). Statistics of directional data. *J. R. Stat. Soc. Series B (Methodol.)* 37, 349–393. doi: 10.1111/j.2517-6161.1975.tb01550.x
- Mardia, K. V. (2010). Bayesian analysis for bivariate von Mises distributions. *J. Appl. Stat.* 37, 515–528. doi: 10.1080/02664760903551267
- Mardia, K. V. (2013). "Some aspects of geometry driven statistical models," in *Annual LASR 2013 Proceedings* (Leeds: Leeds University Press), 7–15.
- Mardia, K. V., and El-Atoum, S. (1976). Bayesian inference for the von Mises-Fisher distribution. *Biometrika* 63, 203–206. doi: 10.1093/biomet/63.1.203
- Mardia, K. V., and Jupp, P. E. (2000). *Directional Statistics*. London: John Wiley & Sons, LTD.
- McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course With Examples in R and Stan*. Boca Raton: CRC Press.
- Mulder, K., Jongma, P., and Klugkist, I. (2020). Bayesian inference for mixtures of von Mises distributions using reversible jump MCMC sampler. *J. Stat. Comput. Simulat.* 90, 1539–1556. doi: 10.1080/00949655.2020.1740997
- Muralidharan, K., and Parikh, R. (2007). Some Bayesian inferences for von Mises distribution. *Amer. J. Math. Manag. Sci.* 27, 123–137. doi: 10.1080/01966324.2007.10737692
- Núñez-Antonio, G., and Gutiérrez-Peña, E. (2005). A Bayesian analysis of directional data using the von Mises-Fisher distribution. *Commun. Stat. Simulat. Comput.* 34, 989–999. doi: 10.1080/03610910500308495
- Pewsey, A. (2002). Testing circular symmetry. *Can. J. Stat.* 30, 591–600. doi: 10.2307/3316098
- Pewsey, A. and García-Portugués, E. (2021). Recent advances in directional statistics. *TEST* 30, 1–58. doi: 10.1007/s11749-021-00759-x
- Rodrigues, J., Galvão Leite, J., and Milan, L. A. (2000). Theory & Methods: an empirical Bayes inference for the von Mises distribution.

- Aust. New Zealand J. Stat.* 42, 433–440. doi: 10.1111/1467-842X.00140
- Røge, R. E., Madsen, K. H., Schmidt, M. N., and Mørup, M. (2017). Infinite von Mises-Fisher mixture modeling of whole brain fmri data. *Neural Comput.* 29, 2712–2741. doi: 10.1162/neco_a_01000
- Schuhmacher, D., Bähre, B., Gottschlich, C., Hartmann, V., Heinemann, F., Schmitzer, B., et al. (2020). Transport: Computation of optimal transport plans and Wasserstein distances, R package version 0.11-1. Available online at: <https://cran.r-project.org/package=transport>
- Shearman, L. P., Sriram, S., Weaver, D. R., Maywood, E. S., Chaves, I., Zheng, B., et al. (2000). Interacting molecular loops in the mammalian circadian clock. *Science* 288, 1013–1019. doi: 10.1126/science.288.5468.1013
- Smith, A. F., and Gelfand, A. E. (1992). Bayesian statistics without tears: a sampling–resampling perspective. *Amer. Stat.* 46, 84–88. doi: 10.1080/00031305.1992.10475856
- Straub, J. (2017). *Bayesian Inference With the Von-Mises-Fisher Distribution in 3D*. Available online at: <http://people.csail.mit.edu/jstraub/>
- Taghia, J., Ma, Z., and Leijon, A. (2014). Bayesian estimation of the von Mises-Fisher mixture model with variational inference. *IEEE Trans. Pattern Anal. Mach. Intell.* 36, 1701–1715. doi: 10.1109/TPAMI.2014.2306426
- Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and waic. *Stat. Comput.* 27, 1413–1432. doi: 10.1007/s11222-016-9696-4
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: an improved \hat{R} for assessing convergence of MCMC (with discussion). *Bayesian Anal.* 16, 667–718. doi: 10.1214/20-BA1221
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2022 Nakhaei Rad, Bekker, Arashi and Ley. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.