Applied Mathematical Modelling 68 (2019) 629–642

Contents lists available at ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

Bayesian cylindrical data modeling using Abe–Ley mixtures

N. Sadeghianpourhamami[a,*], D.F. Benoit[b], D. Deschrijver[a], C. Develder[a]

[a] Department of Information Technology, Ghent University – imec, IDLab, Technologiepark Zwijnaarde 15, Ghent 9052, Belgium
[b] Faculty of Economics and Business Administration, Ghent University, Tweekerkenstraat 2, Ghent 9000, Belgium

ARTICLE INFO

Article history:
Received 1 December 2017
Revised 13 November 2018
Accepted 26 November 2018
Available online 1 December 2018

Keywords:
Cylindrical data
Cylindrical mixture probabilistic model
Abe–Ley distribution
WeiSSVM distribution
Metropolis–Hastings
Bayesian inference

ABSTRACT

This paper proposes a Metropolis–Hastings algorithm based on Markov chain Monte Carlo sampling, to estimate the parameters of the Abe–Ley distribution, which is a recently proposed Weibull-Sine-Skewed-von Mises mixture model, for bivariate circular-linear data. Current literature estimates the parameters of these mixture models using the expectation-maximization method, but we will show that this exhibits a few shortcomings for the considered mixture model. First, standard expectation-maximization does not guarantee convergence to a global optimum, because the likelihood is multi-modal, which results from the high dimensionality of the mixture's likelihood. Second, given that expectation-maximization provides point estimates of the parameters only, the uncertainties of the estimates (e.g., confidence intervals) are not directly available in these methods. Hence, extra calculations are needed to quantify such uncertainty. We propose a Metropolis–Hastings based algorithm that avoids both shortcomings of expectation-maximization. Indeed, Metropolis–Hastings provides an approximation to the complete (posterior) distribution, given that it samples from the joint posterior of the mixture parameters. This facilitates direct inference (e.g., about uncertainty, multi-modality) from the estimation. In developing the algorithm, we tackle various challenges including convergence speed, label switching and selecting the optimum number of mixture components. We then (i) verify the effectiveness of the proposed algorithm on sample datasets with known true parameters, and further (ii) validate our methodology on an environmental dataset (a traditional application domain of Abe–Ley mixtures where measurements are function of direction). Finally, we (iii) demonstrate the usefulness of our approach in an application domain where the circular measurement is periodic in time.

1. Introduction

Various scientific fields consider bivariate measurements that have a linear and a circular component. This amounts to data that can naturally be represented on a cylinder. A main challenge in modeling such cylindrical data is accounting for cross-correlation between the circular and linear variables. Additionally, to capture skewness and heterogeneity in the data, mixture models are required, which aggravates the modeling difficulty.

Various parametric distributions have been proposed to jointly model bivariate cylindrical data. Some early examples are models by Mardia and Sutton [1], based on the conditional distribution of a trivariate normal distribution, as well as

* Corresponding author.
E-mail addresses: nasrin.sadeghianpourhamami@ugent.be (N. Sadeghianpourhamami), dries.benoit@ugent.be (D.F. Benoit), dirk.deschrijver@ugent.be (D. Deschrijver), chris.develder@ugent.be (C. Develder).

https://doi.org/10.1016/j.apm.2018.11.039
0307-904X/

Johnson and Wehrly [2], based on the principle of maximum entropy, subject to constraints on certain moments. Further extensions of Mardia and Sutton's models have been proposed by Kato and Shimizu's team, based on the conditional of a trivariate normal [3] and the conditional of a trivariate t-distribution [4]. An extension to Johnson and Wehrly's model is defined by Wang [5], with a distribution generated from a combination of the von Mises and transformed Kumaraswamy distributions. A semi-parametric extension to Johnson and Wehrly's model is introduced by Fernández-Durán [6] using non-negative trigonometric sums. Finally, non-parametric models for cylindrical data, based on kernel density estimation are explored by García-Portugués et al. [7] and Carnicero et al. [8].

Recently, Abe and Ley [9] have defined a new cylindrical distribution (now commonly referred to as the Abe–Ley distribution) that is based on the combination of the sine-skewed von Mises [10] and the Weibull distributions. Compared to the other aforementioned cylindrical models, the merits of Abe–Ley are highlighted as having (1) flexible shapes, (2) cross-correlation among linear and circular variables, (3) well-known marginal and conditional distributions and (4) support for data skewness.

Mixtures of Abe–Ley distributions have been used successfully to model environmental data (e.g., for analyzing dynamics of waves [11] and marine currents in the Adriatic Sea [12]) where measurements are a function of the direction (represented by an angle). One of the challenges in estimating parameters for such mixture models is that

(L1) it is difficult to give closed-form expressions for the maximum likelihood estimates (MLEs), which is typically addressed by resorting to numerical methods [9].

Effectively, in current literature the parameters of Abe–Ley mixtures are estimated using expectation-maximization (EM) based on MLE. However, these EM-based methods for parameter estimation of Abe–Ley mixture models (e.g., [11–13]) have the following limitations:

(L2) the EM methods are based on optimizing the log-likelihood and hence are susceptible to converging to local maxima, and
(L3) being a point estimate, the uncertainties of the estimates (e.g., confidence intervals or standard errors) are not directly available in EM methods.

To address limitation (L2), typically a short run strategy [14] is used to avoid converging to a local maximum while estimating the parameters of the Abe–Ley mixture models [11–13]: the EM algorithm then runs multiple times using different random initialization and stops without waiting for full convergence. However, converging to the global optimum is still not guaranteed in case of a mixture of Abe–Ley distributions, due to the high dimensionality and the complexity of the likelihood function.

To alleviate limitation (L3), further calculations are needed to approximate the uncertainty of the estimation (e.g., confidence intervals). One approach is to approximate the sampling distribution of the estimated parameters via bootstrap methods, and use that approximated distribution to compute confidence intervals. Approximating the sampling distribution involves randomly sampling from the data with replacement, to create a so-called bootstrap sample (typically of similar size as the original data). For each bootstrap sample, the parameters of interest (in our case, parameters of the Abe–Ley mixture) are estimated using the EM algorithm. The retained EM estimate's instances from each bootstrap sample are then used to approximate the parameter distributions. When using bootstrapping to approximate the sampling distribution for a mixture distribution, we note that one also needs to tackle the label switching issue, caused by the invariance of the likelihood of a K-component mixture model to any permutation of its component indices (see Section 3.2.2 for further explanation of the label switching issue).

To circumvent aforementioned limitations of EM-based methods, we propose a Bayesian approach based on Markov Chain Monte Carlo (MCMC) to estimate the parameters of a Abe–Ley mixture distribution. MCMC methods perform the integration of the posterior distribution of the parameters by sampling from it, rather than optimizing the likelihood, thus circumventing aforementioned limitations (L1) and (L2) of EM [15,16]. Additionally, MCMC-based approaches give joint posterior *distributions* of the parameters. Such distributional info captures the multi-modality (i.e., provides information on both local and global maxima) of the posterior distributions, and also offers insight into the uncertainty of the parameter values (which can be estimated directly by inference from the posterior distribution, without the need for additional calculations). Such Bayesian approaches have been previously successfully applied for modeling circular data (e.g., [17] and [18]), as well as estimating the parameters of finite mixture models for linear variables (e.g., [19] and [20]). Yet, to the best of our knowledge, we are the first to effectively apply a Bayesian approach to estimate the parameters of a (quite complex) bivariate circular-linear distribution and its mixture models.

In the next Section 2, we describe the Abe–Ley distribution and the mixture model. Subsequently, we discuss the following contributions:

1. We propose a Metropolis–Hastings (MH) algorithm to estimate the parameters of the Abe–Ley mixture model. Given that we are dealing with a mixture, we note that the MH algorithm is complicated by the need to sample the component weights, in addition to the model parameters for each of the components themselves (Section 3.1).
2. We successfully tackle the challenges of the proposed Bayesian MH approach, including (i) convergence speed, (ii) the label switching issue (due to the invariance of the likelihood to permutations of mixture component parameters) and (iii) determining the optimal number of mixture components (Section 3.2).

3. We first validate the effectiveness of our approach by showing we can successfully estimate the model parameters for datasets sampled from an a priori known Abe–Ley mixture distribution, i.e., with a known number of mixture components and known parameter values (Section 4).
4. We then apply the proposed approach to two real-world datasets (one traditional application domain with measurements as function of angles, and one new application domain where the circular measurement is periodic time). We show that the Abe–Ley mixture models, estimated using the proposed MH algorithm, effectively capture the heterogeneity in the data under consideration (by referring to the previous analysis of those datasets in literature, see Section 5).
5. We illustrate the existence of multi-modality and skewness in the posterior density of the parameters of the Abe–Ley distribution (Section 5), which makes EM methods susceptible to converging to local optima. From this, we conclude that the proposed Bayesian approach is more reliable than EM methods in estimating the parameters of the Abe–Ley mixtures (Section 6).

2. Probabilistic model description

In this section, we first introduce the Abe–Ley density function and the Abe–Ley mixture model. We then explain the sampling process from the Abe–Ley distribution as proposed in [9, Section 3.3], with a minor correction. (We later use the sampling in Section 4 to test the effectiveness of our estimation.)

2.1. Probability density functions

The Abe–Ley distribution is a combination of the Weibull distribution and the sine-skewed von Mises distribution. Its probability density is defined as:

$$f(\theta, x|\boldsymbol{\zeta}) \mapsto \frac{\alpha\beta^{\alpha}}{2\pi\cosh(\kappa)}(1+\lambda\sin(\theta-\mu))x^{\alpha-1}\exp[-(\beta x)^{\alpha}(1-\tanh(\kappa)\cos(\theta-\mu))], \tag{1}$$

with random variables $(\theta,\ x)\in[0,\ 2\pi)\times[0,\ \infty)$, and distribution parameters $\boldsymbol{\zeta}=(\alpha,\beta,\mu,\kappa,\lambda)$ [9]. The parameters of the Abe–Ley distribution comprise $\alpha,\ \beta>0$, which are linear shape and scale parameters respectively, a circular location parameter $0\le\mu<2\pi$, the parameter $\kappa\ge 0$ that controls the circular concentration and regulates the dependence structure, and finally $-1\le\lambda\le 1$ that controls the circular skewness.

The mixture of a K-component Abe–Ley distribution has the following density function:

$$f(\theta, x|\boldsymbol{\vartheta}) = \sum_{k=1}^{K}\tau_k f_k(\theta, x|\boldsymbol{\zeta_k}) \tag{2}$$

where $f_k(\theta,\ x|\boldsymbol{\zeta_k})$ denotes the probability density of the kth component characterized by parameter set $\boldsymbol{\zeta_k}$, and τ_k is the weight of the kth component. Thus, $\boldsymbol{\tau}=(\tau_1,\tau_2,\ldots,\tau_K)$ is the weight distribution that takes a value in the unit simplex ε_K which is a subspace of $(\Re^+)^K$ defined by the following constraints:

$$\tau_k \ge\quad 0,\quad \tau_1+\tau_2+\ldots+\tau_K=1. \tag{3}$$

Hence, $\boldsymbol{\vartheta}=(\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_K,\boldsymbol{\tau})$ is the parameter vector of the mixture model.

2.2. Random number generation

One of the strong assets of the Abe–Ley distribution is that it has well-known conditional and marginal distributions, which simplifies the random number generation process. The marginal density of the circular component θ is a sine-skewed wrapped Cauchy distribution and the conditional density $f(x|\theta)$ is defined as

$$\begin{aligned} f(x|\theta) = \alpha\cdot\Big[\beta\{1-\tanh(\kappa)\cos(\theta-\mu)\}^{1/\alpha}\Big]^{\alpha} \\ \cdot x^{\alpha-1}\cdot\exp\Big[-\Big\{\beta\,(1-\tanh(\kappa)\cos(\theta-\mu))^{1/\alpha}x\Big\}^{\alpha}\Big] \end{aligned} \tag{4}$$

Abe and Ley [9] state (4) to be a Weibull distribution with shape parameter $\beta\,(1-\tanh(\kappa)\,\cos(\theta-\mu))^{1/\alpha}$, whereas according to the standard definition of the Weibull distribution, (4) has shape parameter α and scale parameter $\beta\,(1-\tanh(\kappa)\,\cos(\theta-\mu))^{-1/\alpha}$ (see [9] for mathematical details). Accounting for this corrected terminology, randomly generating numbers following the Abe–Ley distribution [9] can be achieved as follows:

Step 1: Generate a random variable[1] Θ_1 from a wrapped-Cauchy distribution with location parameter μ and concentration $\tanh(\kappa/2)$.

[1] Capital letters indicate sampled data instances (as opposed to lowercase variable notations).

Step 2: Generate U from a uniform distribution on [0,1] and define

$$\Theta = \begin{cases} \Theta_1 & \text{if } U < (1+\lambda \sin(\Theta_1 - \mu))/2 \\ -\Theta_1 & \text{if } U \geq (1+\lambda \sin(\Theta_1 - \mu))/2 \end{cases}$$

to ensure Θ follows the sine-skewed wrapped Cauchy distribution.

Step 3: Generate X from a Weibull with shape parameter α and scale parameter $\beta(1-\tanh(\kappa)\cos(\Theta-\mu))^{-1/\alpha}$.

To draw N samples from a K-component Abe–Ley mixture, we repeat the aforementioned 3 steps for each mixture component k characterized by $\boldsymbol{\zeta}_k$, where the expected number of samples from the kth component is $\tau_k N$.

3. Parameter estimation using Bayesian inference

In this section, the proposed Metropolis–Hastings algorithm for estimating the parameters of a mixture of Abe–Ley distributions is explained and the associated challenges are addressed.

3.1. Metropolis–Hastings algorithm for estimating Abe–Ley mixture parameters

Let us assume $\boldsymbol{S} = \{(\theta_1, x_1), (\theta_2, x_2), \ldots, (\theta_N, x_N)\}$ is a set of N observations from a K-component Abe–Ley mixture distribution defined by (2) where $\boldsymbol{\vartheta} = (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_K, \boldsymbol{\tau})$ is the unknown parameter vector.

Calculating the posterior density by solving analytical equations is impossible, since it involves calculating intractable integrals. To overcome this challenge, typically Markov-Chain Monte-Carlo (MCMC) methods are used to generate samples from the posterior distribution. A well-known MCMC-based algorithm is Metropolis–Hastings (MH).

To estimate a parameter $y \in \boldsymbol{\zeta}$ of the target distribution (Abe–Ley in our case) from a set of observations $\boldsymbol{S}$, MH iteratively refines its estimate y^i in iteration i, starting from an initial value y^0 for $i = 0$. Given y^i in each iteration i, a new draw y^* is obtained from a predefined proposal distribution $q(y^*|y^i)$. Then, y^* is accepted with acceptance probability

$$A = \min\left\{1, \frac{p(y^*|\boldsymbol{S})\, q(y^{(i)}|y^*)}{p(y^{(i)}|\boldsymbol{S})\, q(y^*|y^{(i)})}\right\} \tag{5}$$

where p denotes the target density.

Finally, the accepted draws are returned as the output of MH algorithm. Note that, when estimating the components of the mixture models, the MH algorithm should also estimate the component allocations and weight distributions of each component (τ_k) in addition to its parameters $(\boldsymbol{\zeta}_k)$.

Algorithm 1 summarizes our approach, which basically adapts the MH algorithm to estimate the parameters of the K-component Abe–Ley mixture and obtain the component membership of each observation. Similar to the original MH, the input to our algorithm comprises (1) the target distributions (i.e., the Abe–Ley mixture model defined by (2)), (2) the observations $\boldsymbol{S}$, with $|\boldsymbol{S}| = N$, and (3) the prior and proposal distributions for each parameter.

Table 1 summarizes the choice of the priors and the proposal distributions for each parameter of the Abe–Ley distribution. The priors are non-informative (i.e., the hyper-parameter selection is such that the resulting priors are almost uniform across the parameter domains). The choices of proposal distributions (used for sampling the new value for parameters) are (truncated) normal distributions defined on the permitted domain of the parameters. The Abe–Ley density function is used for determining the component allocation in each iteration of the algorithm. The output of the algorithm comprises draws for the parameter vector $\boldsymbol{\vartheta}$ and the allocation vector $\boldsymbol{l} = (l_1, \ldots, l_N)$ that indicates to which component each observation belongs.

Algorithm 1 starts by randomly initializing the component parameters on their permitted domains to obtain $\boldsymbol{\vartheta}^0 = (\boldsymbol{\zeta}_1^0, \ldots, \boldsymbol{\zeta}_K^0, \boldsymbol{\tau}^0)$ and assigns each observation $(\theta_n, x_n) \in \boldsymbol{S}$ to a component with probability $\boldsymbol{\tau}^0$ to obtain the initial allocation vector $\boldsymbol{l}^0$ (Line 1). The algorithm then runs for $(M_0 + M)$ iterations, where M_0 is the number of initial samples to disregard (burn-in samples). Each iteration consists of two parts. In the first part (Lines 3–8), the parameters $\boldsymbol{\zeta}_1^i, \ldots, \boldsymbol{\zeta}_K^i$ of each component of the mixture are drawn using the MH algorithm: for each component, new parameter values are sampled from the proposal distributions defined in Table 1 (Line 5) and are accepted with probability A defined by (5) (Line 8).

In the second part of the iteration, the allocation vector $\boldsymbol{l}$ and component weight vector $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_K)$ are sampled (Lines 9-10). To identify $\boldsymbol{l}^i$ (the allocation vector in iteration i), first the probability of each observation (θ_n, x_n) belonging to a component k of the mixture is calculated independently using $p((\theta_n, x_n)|\boldsymbol{\zeta}_k^i)$, where $\boldsymbol{\zeta}_k^i$ is the parameter vector of component k drawn at iteration i. Note that $p((\theta_n, x_n)|\boldsymbol{\zeta}_k^i)$ is an Abe–Ley distribution defined by (1) and not an Abe–Ley mixture model. In other words, $p((\theta_n, x_n)|\boldsymbol{\zeta}_k^i)$ denotes the probability of observation n coming from an Abe–Ley distribution with parameters $\boldsymbol{\zeta}_k^i$. The observation is then assigned to a component k with probability $p((\theta_n, x_n)|\boldsymbol{\zeta}_k^i)$. (Line 9). Once $\boldsymbol{l}^i$ is identified, the number of observations allocated to each component of the mixture is counted to calculate the parameter vector of a Dirichlet distribution. The component weight vector $\boldsymbol{\tau}^i = (\tau_1^i, \ldots, \tau_K^i)$ is then sampled from that Dirichlet distribution (Line 10). Finally, the first M_0 draws are discarded and M remaining draws are returned for $(\boldsymbol{\vartheta}, \boldsymbol{l})$ (lines 11-12).

Algorithm 1: Metropolis–Hastings.

Input : The target distribution $f(\theta, x|\boldsymbol{\vartheta})$ with unknown parameter vector $\boldsymbol{\vartheta}$ (defined by~(2)); data samples $\boldsymbol{S}$ with $|\boldsymbol{S}| = N$; priors and proposal distributions for each parameter in $\boldsymbol{\vartheta}$ (see Table 1)

Output: the draws for $\boldsymbol{\vartheta}$ (i.e., $\boldsymbol{\vartheta}^i$); the allocation vector $\boldsymbol{l} = (l_1, \ldots, l_N)$ /* $\boldsymbol{l}$ denotes to which component each of the observations are assigned (if (θ_n, x_n) is part of component k, then $l_n = k$) */

Initialize $\boldsymbol{\vartheta}^0 = (\boldsymbol{\zeta}_1^0, \ldots, \boldsymbol{\zeta}_K^0, \boldsymbol{\tau}^0)$ and $\boldsymbol{l}^0$; /* allocate samples in $\boldsymbol{S}$ to each component with probability $\boldsymbol{\tau}^0$ to obtain $\boldsymbol{l}^0$ */

foreach $i = 0, \ldots, M_0 + M - 1$ **do**

  /* Given $\boldsymbol{l}^i$, estimate the mixture parameters */

  foreach $k = 1, \ldots, K$ **do**

    foreach $y \in \boldsymbol{\zeta}_k$ **do**

      /* update parameters of each component in this loop */

      Sample the component parameter y^* from a proposal distribution $y^* \sim q(y^*|y^{(i)})$;

      Sample u from a Uniform distribution on [0,1];

      Calculate the acceptance ratio as $A = \min\{1, \frac{p(y^*)\, q(y^{(i)}|y^*)}{p(y^{(i)})\, q(y^*|y^{(i)})}\}$;

      $y^{(i+1)} = \begin{cases} y^* & u < A \\ y^{(i+1)} & u \geq A \end{cases}$

  Classify each observation $(\theta_n, x_n) \in \boldsymbol{S}$ conditional on knowing $\boldsymbol{\zeta}_1^i, \ldots, \boldsymbol{\zeta}_K^i$, by sampling l_n independently for each $n = 1, \ldots, N$ from: $p\big(l_n = k|\boldsymbol{\zeta}_1^i, \ldots, \boldsymbol{\zeta}_K^i, (\theta_n, x_n)\big) \propto p\big((\theta_n, x_n)|\boldsymbol{\zeta}_k^i\big)$ to obtain $\boldsymbol{l}^i$;

  Sample $\boldsymbol{\tau}^i = (\tau_1^i, \ldots, \tau_K^i)$ from the Dirichlet distribution $D(e_1(\boldsymbol{l}^i), \ldots, e_K(\boldsymbol{l}^i))$, where $e_k(\boldsymbol{l}^i) = e_0 + N_k(\boldsymbol{l}^i), k = 1, \ldots, K$, and $N_k(\boldsymbol{l}^i)$ is the number of data points allocated to component k of the mixture at iteration i and e_0 is the prior of the Dirichlet distribution;

Disregard the first M_0 draws;

return M draws for $\boldsymbol{\vartheta}$ and the allocation vector $\boldsymbol{l}$;

Table 1
Choice of priors and proposal distributions at the input of Algorithm 1.

Parameter	Prior (hyper-parameters)	Proposal*
α	Gamma(shape = 0.001, scale = 1000)	Truncated normal on $[0, \infty)$
β	Gamma(shape = 0.001, scale = 1000)	Truncated normal on $[0, \infty)$
κ	Gamma(shape = 0.001, scale = 1000)	Truncated normal on $[0, \infty)$
μ	von Mises(mean = 0, precision = 0.001)	normal *mod* 2π
λ	Beta(shape1 = 1, shape2 = 1) (rescaled)	Truncated normal on $[-1, 1]$
$\boldsymbol{\tau}$	Dirichlet(vector of ones with length K)	Not-applicable

* The means of the proposal distributions at each iteration are the draws in the previous iteration (or the initial component values at the first iteration); the variances of the proposal distributions are adjusted every 50 iterations (see Section 3.2.1 for details.)

3.2. Addressing the challenges of a Bayesian approach

In this section, we explain how we tackle three computational aspects in our proposed approach: (1) improving the convergence rate of the MH algorithm via an *adaptive* Metropolis–Hastings algorithm, (2) addressing the label switching issue (caused by invariance of the mixture likelihood to a permutation of component parameters) and (3) Bayesian model selection (for determining the optimal number of mixture components). Note that the second and third challenges are inherent to parameter estimation for mixture models in general, both in Bayesian approaches as well as EM-based methods.

3.2.1. Adaptive Metropolis–Hastings

A crucial factor in developing an efficient MH algorithm is the definition of the proposal distribution, $q(y^*|y^i)$. In most applications, a symmetric, unimodal distribution such as the Gaussian distribution is chosen. The MH algorithm requires that the variance of the proposal distribution is preset and does not change during the execution of the MCMC procedure. However, the choice of this variance parameter has an important impact on the efficiency of the MCMC algorithm. When the variance of the proposal distribution is too large, the acceptance probability (defined by (5)) will tend to 0. As a result, the Markov Chain will retain its current value and only jump to new values with vanishing small probability. On the other hand, when the variance of the proposal distribution is too small, the acceptance probability will become close to 1. In this case, the Markov Chain is constantly sampling new values, but these values are very close to the current value, and it will take an excessive amount of time before the entire posterior distribution is sampled.

The variance of the proposal distribution has to be set so that both inefficient behaviors outlined above are avoided. Standard practice [21] is to sample from the posterior distribution for some time and then evaluate the acceptance probabilities. The variances of the proposal distributions should then be adjusted in order to achieve an acceptance probability of $0.3 - 0.5$. These values ensure that the algorithm does not output the same value, while still making reasonably large jumps. However, in the current mixture model, this approach is difficult to implement. As shown in Algorithm 1, the parameters of each component of the mixture are sampled in blocks. Each component has five parameters, $\boldsymbol{\zeta} = (\alpha, \beta, \mu, \kappa, \lambda)$. In the MH algorithm we have to set a variance for each parameter of the K components, so $5 \cdot K$ in total. Tuning these variances is tedious and time consuming.

To overcome the aforementioned challenge, we use a modified version of the MH algorithm proposed by Haario et al. [22] that automatically adjusts the variance of the proposal distribution to maximize efficiency. The basic idea of the adaptive Metropolis–Hastings algorithm is that every R iterations (e.g., $R = 50$), the acceptance probabilities are calculated and evaluated. Whenever the acceptance probabilities are above (below) some threshold (e.g., 0.44), the variance of the proposal distribution is increased (decreased) with an amount $s = \min(0.01, \sqrt{R/i})$ for the next R iterations. At that point, the acceptance probabilities are re-evaluated, and the variances are adjusted again, if necessary. Note that the adjustment amount s is a function of the current iteration i of the MCMC chain, such that at the beginning of the chain larger adjustments are possible, while subsequent adjustments are forced to become continuously smaller.

A detailed description of the adaptive Metropolis–Hastings algorithm and the implications for the mathematical foundations of the algorithm can be found in [22].

3.2.2. Label switching issue

Note that the likelihood of the K-component mixture model in (2) is invariant to any permutation of its component indices (which amounts to a total of $K!$ permutations). Under symmetric priors (i.e., when exchangeable priors are chosen for the component parameters), the resulting posterior will also be invariant to $K!$ permutations in the labeling of the component parameters. In other words, the posterior will have $K!$ symmetric modes. As a result, labels of the components can permute multiple times in subsequent iterations of the MCMC sampling, resulting in a label switching problem. Since in our Bayesian approach, that posterior is used (as distribution for component parameters) for inference of the model parameters, the label switching issue makes such inference very challenging.

Early attempts to solve the label switching issue focus on proposing identifiability constraints via prior distributions to force a unique labeling (e.g., [19,23]). However, as shown by Stephens [24], identifying such constraints is not always feasible, especially when systematically separating the posterior modes is not possible. Hence, two categories of relabeling methods are proposed to post-process the MCMC samples: (1) deterministic relabeling methods that find the optimal permutation in each iteration of the MCMC sampler by minimizing a loss function (e.g., Stephens' algorithm [24], the pivotal reordering algorithm [25,26], default [27] and iterative versions [28] of algorithms for equivalence class representatives, data-based algorithms [28]), and (2) probabilistic relabeling methods that treat permutation of the parameters as missing data with associated uncertainty and estimate its density using an EM type approach (e.g., [29]).

We refer the interested reader to [30] for an explanation and performance comparison of the relabeling methods in terms of CPU times as well as to what extent they agree on the component labels. We note that the data-based relabeling algorithm [28] performs better in terms of both performance criteria (as shown in [28,30]), and hence use it for the relabeling of the samples from Algorithm 1. The data-based relabeling is based on the key idea that in a converged MCMC, while the labels of each cluster might change from one iteration to the other, the clusters remain almost the same. Leveraging such minute difference between the clusters of each iteration, one may keep track of the k clusters throughout each MCMC iteration to identify cluster movements. Further details of the relabeling algorithm are outlined in [28, Algorithm 5].

3.2.3. Bayesian model selection

In many modeling problems, the number of mixture components is not known and needs to be identified as a part of the model selection process. Earlier attempts tried to estimate the true number of components either by calculating the marginal likelihoods (e.g., [31,32]) or by trans-dimensional MCMC samplers (e.g., reversible jump MCMC [33]). In recent approaches however, the view of model selection is changed from *identifying the true model* to *finding a useful model* [34]. In the latter case, model usefulness is seen as its predictive ability for future or unseen data (i.e., out-of-sample prediction accuracy [35]). Vehtari et al. [36] quantify the out-of-sample prediction accuracy as expected log pointwise predictive density (elpd). However, since future data is not available, to calculate elpd, first a log point-wise prediction density (lpd) is calculated using the observed samples. Here, lpd is an over-estimation of elpd for future data, which can be corrected with a bias term [35]. The lpd measure is calculated from the posterior samples using leave-one-out cross-validation (LOO-CV) as

$$lpd_{LOO\text{-}CV} = \sum_{i=1}^{n} \log\left(\frac{1}{S}\sum_{s=1}^{S} p\left(y_i|\theta^{is}\right)\right) \tag{6}$$

where n is the number of observations, S is the number of samples from the posterior, θ^{is} is a sample s from the posterior samples drawn based on all but observation y_i, and $p(y_i|\theta^{is})$ is the probability of observation y_i given posterior parameter θ^{is}.

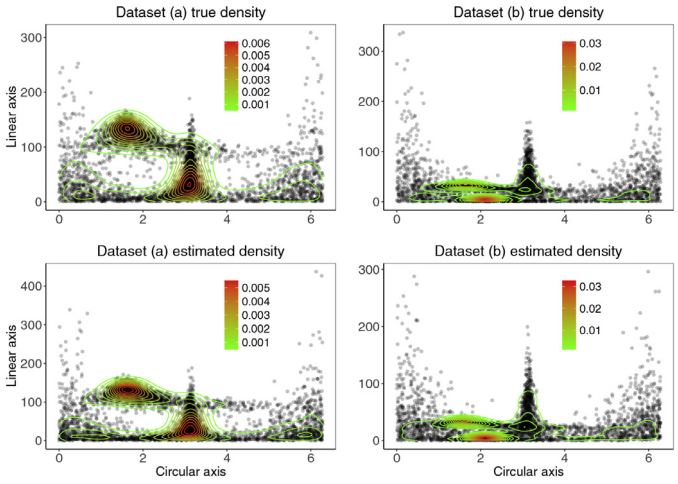

Fig. 1. Sampled datasets from mixture of true (top row) and estimated (bottom row) Abe–Ley distributions. Contour plots indicate two-dimensional kernel densities.

The above calculations are based on $n-1$ observations. If n is large, the overestimation is negligible, otherwise it is corrected using a bias b that denotes the improvement of an estimation when n observations are considered.

Note that the calculations of $lpd_{LOO\text{-}CV}$ are computationally expensive for a large number of observations. Hence, Vehtari et al. [36] also propose an efficient approximation of $lpd_{LOO\text{-}CV}$ using Pareto-smoothed importance sampling (PSIS). Still, the approximations by PSIS-LOO are not reliable when the estimated shape parameter of the generalized Pareto distribution exceeds 0.7 [36]. In that case, 10-fold cross-validation is used to estimate the elpd values as outlined in [36, Section 2.3].

For every modeling endeavor there is a trade-off between the interpretability of a model and the predictive performance. Here we focus mainly on the latter, hence we use elpd values to find the optimum number of mixture components. To avoid overfitting, we use graphical inspection of the elpd measure's evolution for increasing the number of mixture components. We identify a knee point in that graph as a point beyond which the increase in the number of components K results in a smaller, or at least not better, elpd value compared to that for smaller K.

4. Validation on a sample dataset with known true parameters

Before moving to applying our approach to real-world data, we first validate its capability of correctly estimating the parameters from synthetic samples generated using a known Abe–Ley mixture model. We generate such data using the random number generation process explained in Section 2.2. We then run Algorithm 1 for a total of $M_0 + M = 100,000$ iterations, from which we disregard the $M_0 = 20,000$ initial draws that are considered as burn-in. Additionally, to reduce auto-correlation between the samples, we use thinning by a factor 5, (i.e., only keeping every 5th draw of the MCMC chain) in Algorithm 1. Hence, we finally retain 20,000 draws with a burn-in of 5000 initial samples. The choice of 5 is based on the auto-correlation plots of the posterior draws.[2] We use trace plots to examine the convergence and mixing performance of an MCMC chain. As a spread measure of the posterior distribution, we use a 95% Bayesian credible interval.

The top row of Fig. 1 shows the data sampled from a mixture of three (dataset (a)) and four (dataset (b)) Abe–Ley distributions along with the true Abe–Ley mixture densities in the form of contour plots. Both datasets contain 4500 samples. The number of samples from each component of the mixture is the same in dataset (a) but different per component in dataset (b) (see component weights from Table 2).

As mentioned earlier, one of the defining advantages of the MH algorithm is that — unlike likelihood-based estimations, which are point estimates — the MH algorithm outputs samples from the posterior distribution of the model parameters, making the uncertainty of the estimates directly inferable from the posterior densities, without the need for extra calculations (such as in bootstrapping). The posterior densities of the parameters for a mixture of 3 Abe–Ley distributions (dataset (a)) are shown in Fig. 2. The accompanying trace plot for each parameter is used to analyze the convergence and mixing performance of the MH algorithm. As seen from the trace plots, the burn-in of 5000 initial samples is sufficient to disregard the unstable initial draws of the algorithm. The retained draws are from the higher probability region of the posterior and are close to the true values of the parameters, indicating that the chain has converged. The density plots in Fig. 2 are based on the 15,000 retained draws. Finally, the trace plots also confirm that the draws among various iterations are not identical: the chain is mixing well and effectively exploring the posterior. For some parameters however, the mixing of the chain is not apparent: this is due to the scaling issue, caused by a large difference between the starting and the true value

[2] Note that the auto-correlation plots are excluded from this paper to maintain a reasonable paper length.

Table 2
Estimated and true parameters for sampled data from mixture of Abe–Ley distribution (The true values are shown in parenthesis).

Dataset	k	α	β	κ	μ	λ	τ
(a)	1	0.98 (1)	0.07 (0.07)	0.93 (1)	0.03 (0)	0.00 (0)	0.33 (0.33)
	2	1.96 (2)	0.21 (0.2)	2.97 (3)	3.14 (3.14)	−0.98 (−1)	0.34 (0.33)
	3	9.91 (10)	0.01 (0.01)	1.97 (2)	1.58 (1.57)	0.82 (0.8)	0.33 (0.33)
(b)	1	1.00 (1)	0.07 (0.07)	0.95 (1)	0.06 (0)	−0.06 (0)	0.3 (0.31)
	2	1.99 (2)	0.19 (0.2)	3.00 (3)	3.14 (3.14)	−0.79 (−1)	0.18 (0.18)
	3	10.10 (10)	0.04 (0.04)	2.01 (2)	1.58 (1.57)	0.85 (0.8)	0.24 (0.24)
	4	2.97 (3)	1.02 (1)	3.00 (3)	2.09 (2.09)	0.53 (0.5)	0.27 (0.27)

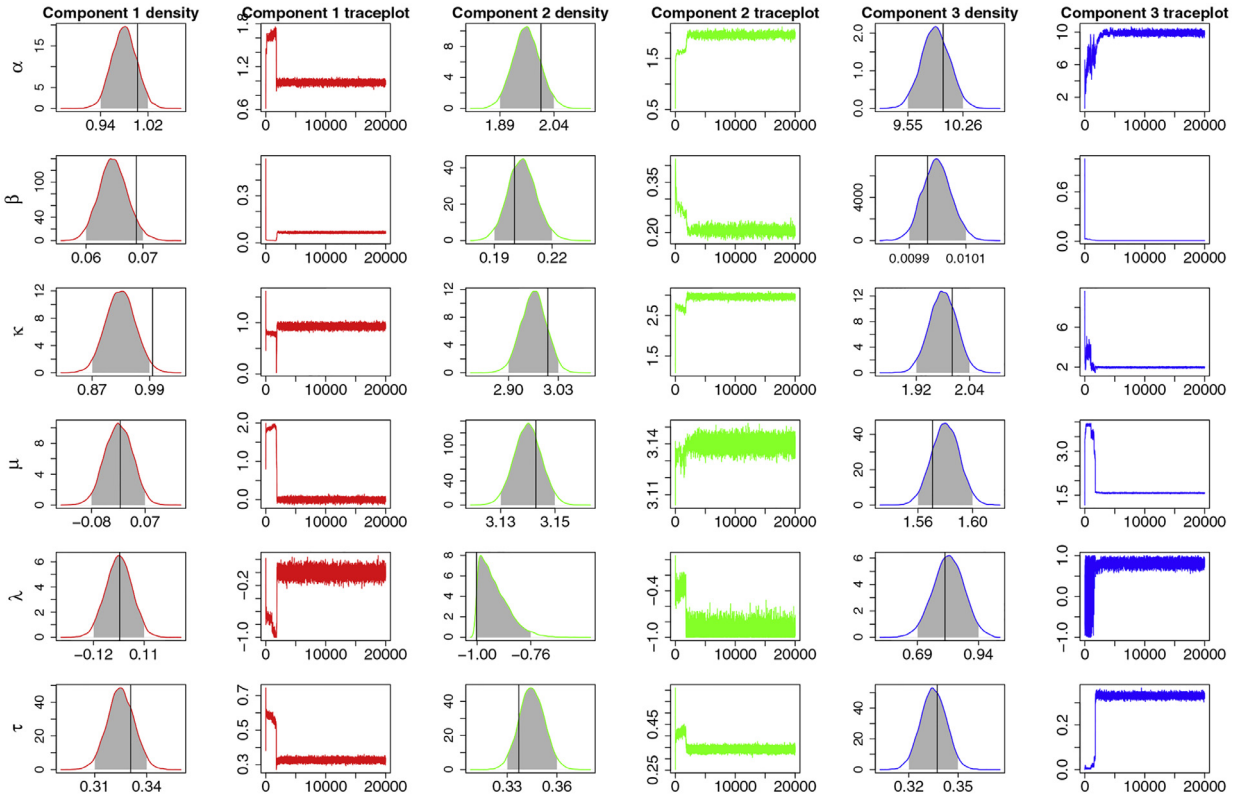

Fig. 2. Posterior densities and trace plots of the parameters for mixture of 3 Abe–Ley (Dataset (a)). Black vertical lines mark the true parameter values and shaded areas are the 95% Bayesian credible intervals.

of the parameter. Also, very low auto-correlation is observed among the draws, which confirms efficient exploration of the posterior.

The shaded regions in the density plots of Fig. 2 indicate the 95% Bayesian credible intervals. The boundaries of the credible intervals are marked with numeric values on the horizontal axis, while the black vertical lines show the true values of the parameters. For some parameters, the true values are in the tails of the posterior. This is only natural, since there is always a 5% chance that the true value will be outside this credible interval, and we have 5 parameters per component plus the component weights (in this case totaling 18).

To demonstrate the estimated predictive density for both datasets, we use the last 4500 MCMC draws of ϑ^i and for each draw, sample a data point from the Abe–Ley mixture distribution parameterized by ϑ^i. The generated data points are composed and presented in Fig. 1. We then use a two-dimensional kernel density estimate to obtain the estimated predictive density (shown in the form of contour plots in Fig. 1). As seen from Fig. 1, the estimated and true predictive density of the Abe–Ley mixtures for both datasets are very similar. This comparison further validates the effectiveness of the proposed approach in estimating the parameters of the Abe–Ley mixture distributions.

While noting that Bayesian estimation is not a point estimate, still, to be able to numerically compare the component-wise densities for the true and the estimated parameters, we summarize the posterior distributions in point forms. To do that, we use maximum a posteriori (MAP) estimation, which corresponds to the mode of the empirical distribution of the posterior. An alternative summarization would be taking the mean of the posterior distribution, but that Bayes estimate

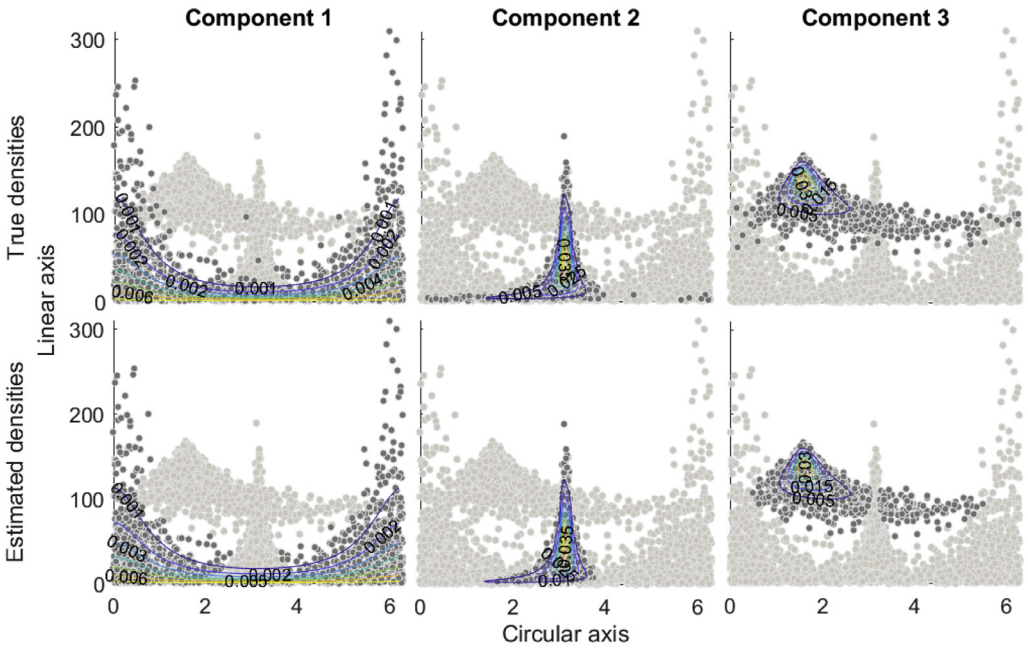

Fig. 3. Component-wise densities for the true and the estimated parameters for Dataset (a).

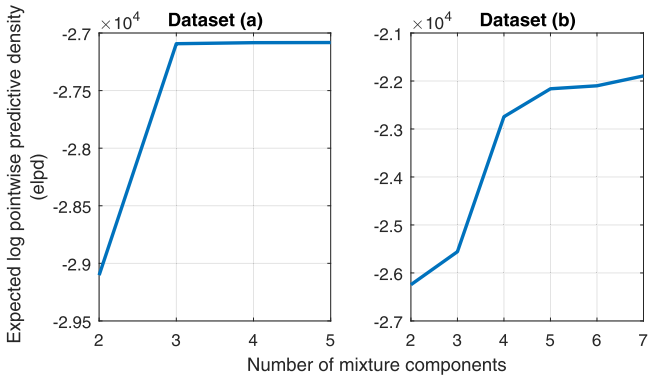

Fig. 4. Determining number of mixture components using elpd measure for sample datasets (the bend in each curve is used for selecting the best number of mixture components).

is not suitable for multi-modal posteriors. The true and estimated parameters we thus obtain are summarized in Table 2. These results indicate that our proposed algorithm can effectively estimate the mixture model parameters for both datasets. For dataset (a), we further demonstrate component-wise densities for the true and the estimated parameters in Fig. 3. This figure shows that the true and estimated densities for the mixture of 3 Abe–Ley distributions are very similar.

To validate the effectiveness in model selection of elpd, approximated by PSIS-LOO, we have calculated the elpd for a varying number of mixture components for both datasets as shown in Fig. 4. The location of the bend (knee) in Fig. 4 indicates the most suitable number of components, which is the same as the number of true mixtures for both datasets.

The examples presented above demonstrate the effectiveness of our proposed algorithm for estimating the parameters of a mixture of Abe–Ley distributions and of elpd as suitable model selection measure. Next, we apply our approach to model the data for two different real-world applications.

5. Modeling real-world datasets with a mixture of Abe–Ley distributions

We now apply our methodology to fit an Abe–Ley mixture to (1) a wave dynamics dataset, and (2) an EV charging dataset. The first application is a traditional application domain of mixtures of cylindrical distributions where the circular measurement is the direction (angle). The dataset in the second example on the other hand is a new one, where the circular measurement is periodic time. For both applications, we run Algorithm 1 for $M_0 + M = 100,000$ iterations and discard a

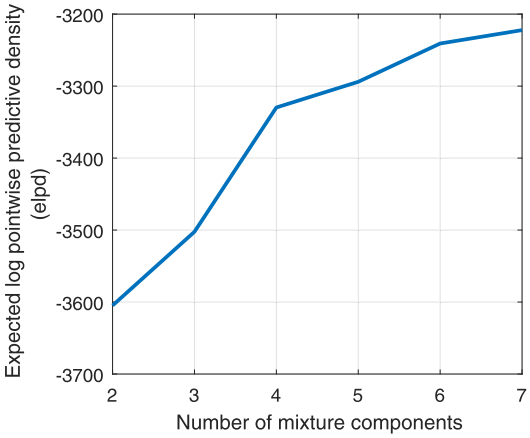

Fig. 5. The elpd values for model selection in modeling wave dynamics. (The bend in the curve is used for selecting the best number of mixture components).

burn-in of $M_0 = 20,000$ draws. Further, we use thinning by a factor 5, i.e., we only keep every 5th draw of the MCMC chain) to reduce the auto-correlation among subsequent draws.

5.1. Wave dynamics in the Adriatic Sea

In this section, we consider a dataset of wave dynamics, which is a well-studied application of the Abe–Ley distribution. The dataset comprises semi-hourly wave directions and heights in the Adriatic Sea, recorded in the period 15 February 2010–16 March 2010 as reported by Lagona et al. [11]. Lagona et al. [11] also approximate this data with a mixture of Abe–Ley distributions, whose parameters depend on the states of a latent Markov chain. However, their proposed estimation algorithm is based on the EM method (and thus a point estimate), whereas our proposed approximation algorithm is based on the MCMC method (thus giving a posterior distribution for the mixture parameters). Additionally, the procedure by Lagona et al. [11] includes a temporal dependence for the data, based on a hidden Markov model. In this work, we focus on fitting the distribution only, while addressing temporal dependence could be interesting as future work.

Fig. 5 shows the elpd values for different numbers of Abe–Ley mixtures and suggests $K = 4$ as the optimum number of mixtures because the improvement in elpd when increasing the number of mixtures from 2 to 4 is significantly larger than for the increase from 4 to 7 and beyond. Hence, $K = 4$ is a knee point. Note that in [11] the best number of states (mixtures) is deemed to be $K = 3$ using the BIC measure. To represent and compare the distribution of each component of the mixture models, we use MAP estimation to summarize the estimated parameters from our approach in a point form. The resulting mixtures are depicted in Fig. 7(b) and are compared with the fits from Lagona et al. [11] shown in Fig. 7(a).

Fig. 7 suggests that both approaches identify similar heterogeneity in the data. The first component of Lagona et al. models the high waves coming from the north [11], which in our model are represented by the first and the second components. The second component of Lagona's model (and our third component) is associated with calm sea and finally, their third component (fourth in our model) is associated with Sirocco episodes (caused by wind blowing southeasterly, along the major axis of the Adriatic Sea). However, the posterior densities of the 4 mixture components depicted in Fig. 6 confirm that the Bayesian approach is more reliable than the EM method. The posterior densities for component 1 and 2 in Fig. 6 are multi-modal, making the EM based approaches susceptible to premature convergence to a local maximum.

5.2. Electric vehicle hourly charging requests

Our second study is motivated by the increasing use of EVs and the need to analyze their impact on the power grid. Initial studies only presented empirical distributions of the arrival times of EVs from real-world measurements [37–40]. Here, we take the first step to model the arrival times of EVs using a mixture of Abe–Ley distributions. We use an EV charging session dataset (collected by ElaadNL[3]) that includes the arrival times of electric vehicles at public roadside charging stations across the Netherlands from January to March 2015.

We divide a day into hour-long slots and count the number of EV arrivals in each slot. We also take the mean time-of-arrival of EVs in each slot to characterize the timing aspect of the measurement. Therefore, the resulting data points

[3] ElaadNL is the knowledge and innovation center in the field of charging infrastructure in the Netherlands, providing coordination for the connections of public roadside charging stations to the electricity grid on behalf of 6 participating distribution system operators (DSOs). It also performs technical tests of charging infrastructure, researches and tests smart charging possibilities of EVs, and develops communication protocols for managing EV charging. The EV charging session data is available upon request for non-commercial research purposes, subject to signing an agreement. For more information, please contact Chris Develder (email: chris.develder@ugent.be).

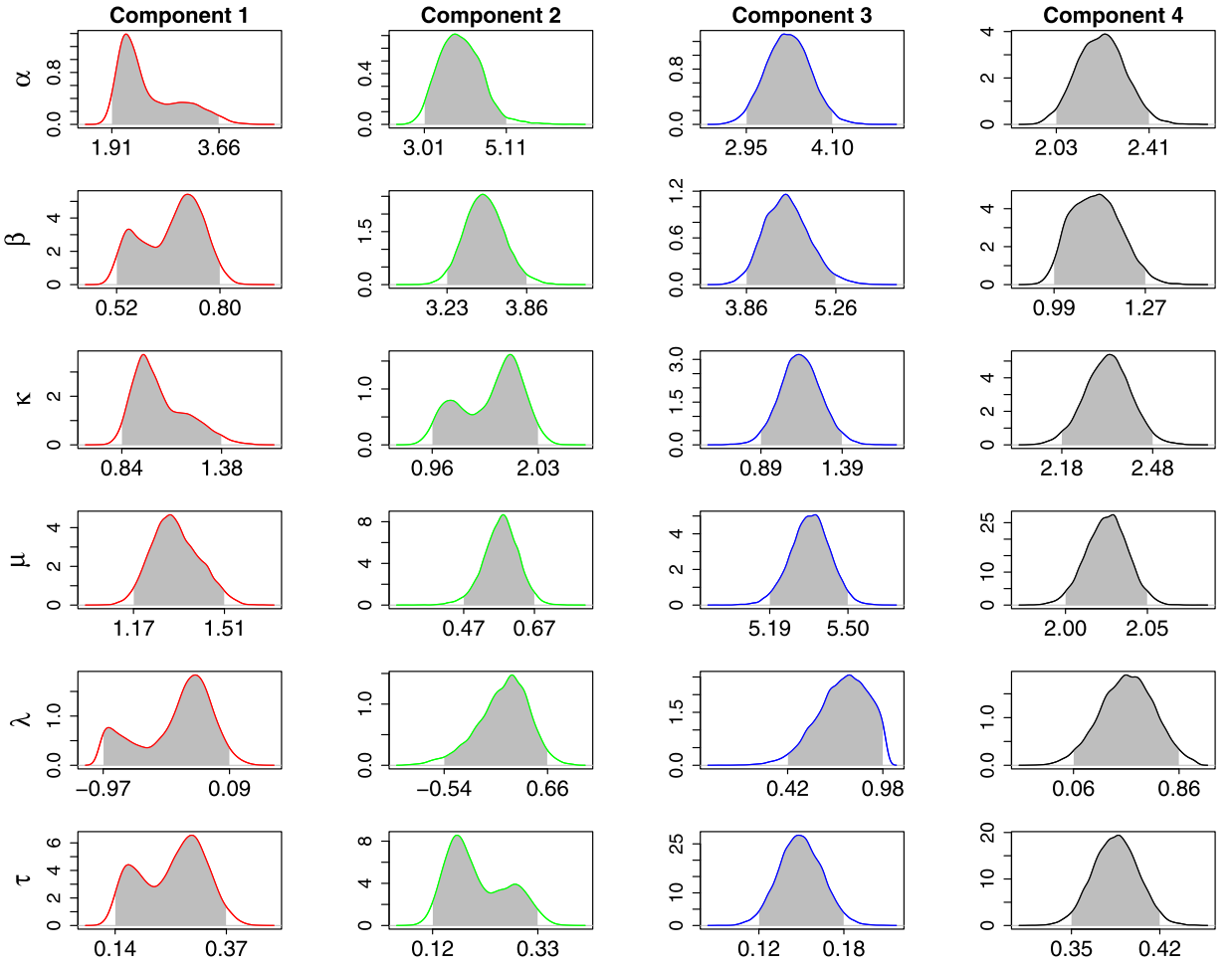

Fig. 6. Posterior densities of the parameters for best Abe–Ley mixture model for wave dynamics.

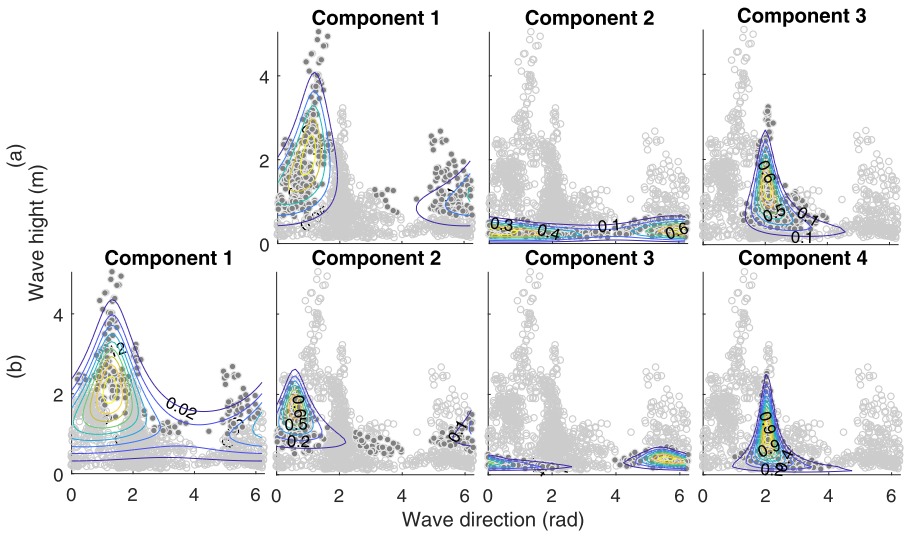

Fig. 7. Component-wise densities for (a) Abe–Ley mixture model estimated by Lagona et al. [11] and (b) Abe–Ley mixture model estimated by our proposed approach.

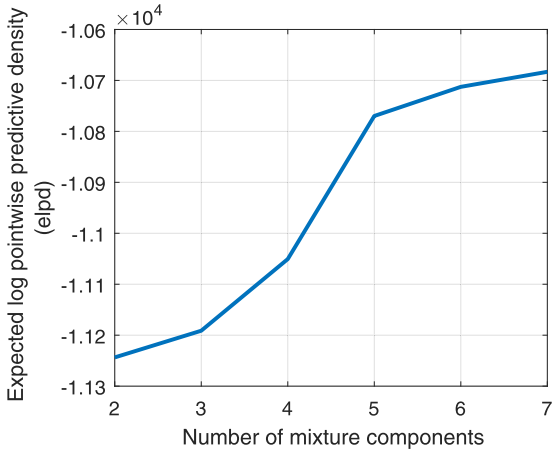

Fig. 8. elpd values for model selection in modeling EV arrivals (the bend in the curve is used for selecting the best number of mixture components).

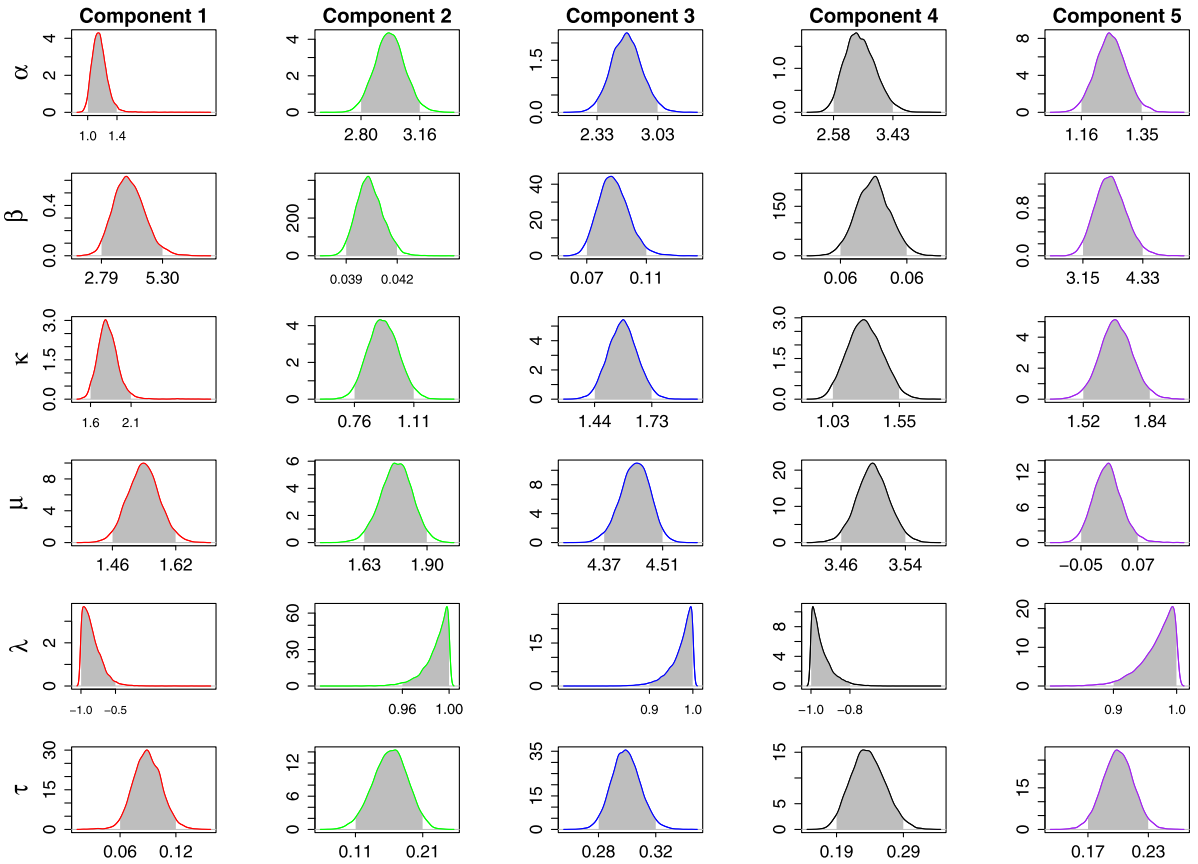

Fig. 9. Posterior densities of the parameters for best Abe–Ley mixture model for EV arrival.

have one linear (number of EV arrivals) and one circular (average time-of-arrival) measurement and are best represented on a cylinder. Note that the linear measurements in this dataset are of discrete nature. However, due to unavailability of the cylindrical distributions that effectively take into account the cross-correlation of the circular and linear measurements for discrete data, we have modeled this dataset with an Abe–Ley distribution as the best existing candidate. To prevent over-fitting, we add a random value generated from a uniform distribution on (0,1) to the EV counts in each slot.

We fit the mixture of Abe–Ley distributions with a varying number of mixture components to the EV dataset and use the elpd measure to select the best number of mixture components, as illustrated in Fig. 8. The bend in the elpd values suggests $K = 5$ mixtures to model the EV arrival distribution. The estimated posterior densities are shown in Fig. 9, where

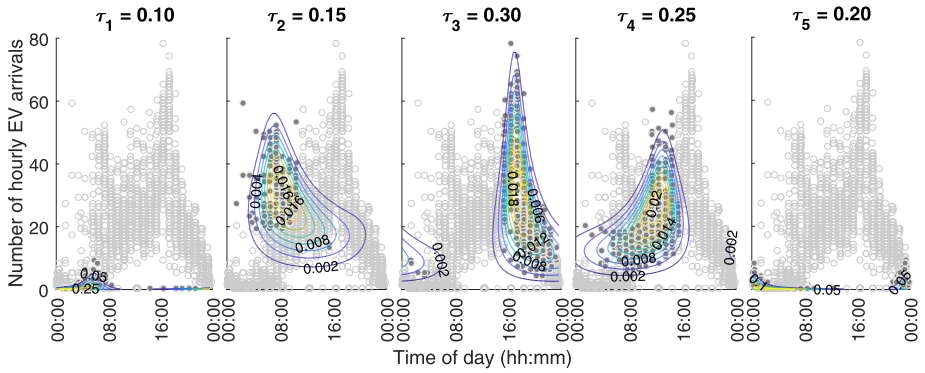

Fig. 10. Component-wise densities for the true and estimated parameters in modeling EV arrivals.

the shaded regions indicate the 95% Bayesian credible intervals. We also use MAP estimation to numerically summarize the posteriors in point forms and show the component densities in Fig. 10.

Next, the resulting mixtures are compared with our previous studies on this dataset. In [39,40], we clustered this data on a 2-dimensional surface (i.e., time-of-arrival vs. time-of-departure) into 3 clusters: charge-near-work (characterized by early morning arrivals, mainly on weekdays), charge-near-home (with late afternoon arrivals), and park-to-charge (with arrivals throughout the day). We also found that the EV arrivals have different empirical distributions on weekdays compared to weekends: weekday arrivals have two peaks (mornings and evenings), whereas the weekend arrivals only peak around noon.

This heterogeneity is very well captured by the mixture of the 5-component Abe–Ley distribution. As seen in Fig. 10, the second component models the morning peaks in EV arrivals, which are typically arrivals on weekdays and from the charge-near-work cluster. The third component of the mixture comprises weekday arrivals, from charge-near-home sessions. The fourth component models the weekend arrivals and day-time charging during the weekdays, which are typically park-to-charge sessions. Finally, the probability of having a very small number of EV arrivals is modeled by the first (for very early morning arrivals) and fifth (for arrivals around midnight) components of the mixture.

6. Conclusion

In this paper, a Metropolis–Hastings algorithm based on MCMC sampling was proposed for estimating the components of Abe–Ley mixture models. A dynamic Metropolis–Hastings algorithm was used to adjust the variance of the proposal distribution to improve both convergence and exploration of the posterior distribution in the proposed algorithm. Two challenges associated with estimating the mixture parameters were also tackled: the label switching issue and the selection of the optimal number of components.

By referring to the posterior distributions of the parameters of the Abe–Ley mixture, we illustrated that a Bayesian based estimation is more reliable than EM methods for estimating these parameters, because:(1) the multi-modality of the posteriors can be inferred directly from the output of the MH algorithm, hence, local maxima are avoided without a need for additional calculations (which also do not guarantee arriving at a global optimum), and (2) in EM-based estimation, the uncertainty of the estimates is not available and further calculations are needed to approximate them. While bootstrapping could address this, in our proposed Bayesian approach, the uncertainty of the estimation is directly inferred from the posteriors.

We validated the effectiveness of our proposed approach as well as the model selection measure by estimating the parameters of actual Abe–Ley mixture models. Further, we applied our proposed approach to model the data from two different real-world application domains: wave dynamics and electric vehicle (EV) arrivals. In both applications, the Abe–Ley mixtures captured the data skewness, the correlation between the circular and linear variables and the data heterogeneity (multi-modality). We found that the resulting mixtures were intuitively appealing and qualitatively in accordance with previous studies of the real-world datasets.

Acknowledgments

We thank ElaadNL for providing the data and relevant insights on EV charging in the Netherlands. Special thanks goes out to Nazir Refa (ElaadNL) for preparing the raw data for analysis. We also thank Prof. Francesco Lagona for sharing the wave dynamics data of the Adriatic Sea. Finally, we thank Dr. Thomas Demeester for his expert advice on the bootstrapping approach.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.apm.2018.11.039.

References

[1] K.V. Mardia, T.W. Sutton, A model for cylindrical variables with applications, J. R. Stat. Soc. Ser. B Methodol. 40 (2) (1978) 229–233.
[2] R.A. Johnson, T.E. Wehrly, Some angular-linear distributions and related regression models, J. Am. Stat. Assoc. 73 (363) (1978) 602–606.
[3] S. Kato, K. Shimizu, Dependent models for observations which include angular ones, J. Stat. Plann. Inference 138 (11) (2008) 3538–3549. Special Issue in Honor of Junjiro Ogawa (1915 - 2000): Design of Experiments, Multivariate Analysis and Statistical Inference. doi: 10.1016/j.jspi.2006.12.009.
[4] S. Sugasawa, K. Shimizu, S. Kato, A flexible family of distributions on the cylinder, ArXiv e-prints (2015) https://arxiv.org/abs/1501.06332.
[5] M. Wang, Extensions of Probability Distributions on Torus, Cylinder and Disc, School of Science and Technology, Kieo University, Tokyo, 2013.
[6] J.J. Fernández-Durán, Models for circular-linear and circular-circular data constructed from circular distributions based on nonnegative trigonometric sums, Biometrics 63 (2) (2007) 579–585.
[7] E. García-Portugués, R.M. Crujeiras, W. González-Manteiga, Exploring wind direction and so2 concentration by circular–linear density estimation, Stoch. Environ. Res. Risk Assess. 27 (5) (2013) 1055–1067, doi:10.1007/s00477-012-0642-5.
[8] J. A. Cărnicero, C. Ausín, M. Wiper, Non-parametric copulas for circular-linear and circular-circular data: an application to wind directions, Stoch. Environ. Res. Risk Assess. 27 (2013) 1991–2002.
[9] T. Abe, C. Ley, A tractable, parsimonious and flexible model for cylindrical data, with applications, Economet. Stat. 4 (Supplement C) (2017) 91–104, doi:10.1016/j.ecosta.2016.04.001.
[10] T. Abe, A. Pewsey, Sine-skewed circular distributions, Stat. Pap. 52 (3) (2011) 683–707, doi:10.1007/s00362-009-0277-x.
[11] F. Lagona, M. Picone, A. Maruotti, A hidden Markov model for the analysis of cylindrical time series, Environmetrics 26 (8) (2015) 534–544. Env.2355 doi: 10.1002/env.2355.
[12] F. Lagona, M. Picone, Model-based segmentation of spatial cylindrical data, J. Stat. Comput. Simul. 86 (13) (2016) 2598–2610, doi:10.1080/00949655.2015.1122791.
[13] M. Ranalli, F. Lagona, M. Picone, E. Zambianchi, Segmentation of sea current fields by cylindrical hidden Markov models: a composite likelihood approach, J. R. Stat. Soc. Ser. C Appl. Stat. doi:10.1111/rssc.12240.
[14] J. Bulla, F. Lagona, A. Maruotti, M. Picone, A multivariate hidden Markov model for the identification of sea regimes from incomplete skewed and circular time series, J. Agric. Biol. Environ. Stat. 17 (4) (2012) 544–567, doi:10.1007/s13253-012-0110-1.
[15] Y. Tang, J. Fu, W. Liu, A. Xu, Bayesian analysis of repairable systems with modulated power law process, Appl. Math. Modell. 44 (Supplement C) (2017) 357–373, doi:10.1016/j.apm.2017.01.067.
[16] S. Ali, Mixture of the inverse rayleigh distribution: properties and estimation in a Bayesian framework, Appl. Math. Modell. 39 (2) (2015) 515–530, doi:10.1016/j.apm.2014.05.039.
[17] R. McVinish, K. Mengersen, Semiparametric Bayesian circular statistics, Comput. Stat. Data Anal. 52 (10) (2008) 4722–4730, doi:10.1016/j.csda.2008.03.016.
[18] G. Nuñez-Antonio, E. Gutiérrez-Peña, G. Escarela, A Bayesian regression model for circular data based on the projected normal distribution, Stat. Modell. 11 (3) (2011) 185–201, doi:10.1177/1471082X1001100301.
[19] S. Frühwirth-Schnatter, Markov chain monte carlo estimation of classical and dynamic switching and mixture models, J. Am. Stat. Assoc. 96 (453) (2001) 194–209.
[20] J. Diebolt, C.P. Robert, Estimation of finite mixture distributions through Bayesian sampling, J. R. Stat. Soc. Ser. B Methodol. 56 (2) (1994) 363–375.
[21] G.O. Roberts, A. Gelman, W.R. Gilks, Weak convergence and optimal scaling of random walk metropolis algorithms, Ann. Appl. Probab. 7 (1) (1997) 110–120, doi:10.1214/aoap/1034625254.
[22] H. Haario, E. Saksman, J. Tamminen, An adaptive metropolis algorithm, Bernoulli 7 (2) (2001) 223–242.
[23] S. Richardson, P.J. Green, On Bayesian analysis of mixtures with an unknown number of components (with discussion), J. R. Stat. Soc. Ser. B Stat. Methodol. 59 (4) (1997) 731–792, doi:10.1111/1467-9868.00095.
[24] M. Stephens, Dealing with label switching in mixture models, J. R. Stat. Soc. Ser. B Stat. Methodol. 62 (4) (2000) 795–809.
[25] J.-M. Marin, K. Mengersen, C.P. Robert, Bayesian modelling and inference on mixtures of distributions, Handb. Stat. 25 (2005) 459–507, doi:10.1016/S0169-7161(05)25016-2.
[26] J.-M. Marin, C. Robert, Bayesian Core: A Practical Approach to Computational Bayesian Statistics, Springer Science & Business Media, 2007.
[27] P. Papastamoulis, G. Iliopoulos, An artificial allocations based solution to the label switching problem in Bayesian analysis of mixtures of distributions, J. Comput. Graph. Stat. 19 (2) (2010) 313–331, doi:10.1198/jcgs.2010.09008.
[28] C.E. Rodríguez, S.G. Walker, Label switching in Bayesian mixture models: deterministic relabeling strategies, J. Comput. Graph. Stat. 23 (1) (2014) 25–45, doi:10.1080/10618600.2012.735624.
[29] M. Sperrin, T. Jaki, E. Wit, Probabilistic relabelling strategies for the label switching problem in Bayesian mixture models, Stat. Comput. 20 (3) (2010) 357–366, doi:10.1007/s11222-009-9129-8.
[30] P. Papastamoulis, label.switching: An R package for dealing with the label switching problem in MCMC outputs, J. Stat. Softw. Code Snippets 69 (1) (2016) 1–24, doi:10.18637/jss.v069.c01.
[31] S. Chib, Marginal likelihood from the Gibbs output, J. Am. Stat. Assoc. 90 (432) (1995) 1313–1321.
[32] S. Chib, I. Jeliazkov, Marginal likelihood from the metropolis-hastings output, J. Am. Stat. Assoc. 96 (453) (2001) 270–281.
[33] P.J. Green, Reversible jump Markov chain monte carlo computation and Bayesian model determination, Biometrika 82 (4) (1995) 711–732.
[34] J. Piironen, A. Vehtari, Comparison of Bayesian predictive methods for model selection, Stat. Comput. 27 (3) (2017) 711–735, doi:10.1007/s11222-016-9649-y.
[35] A. Gelman, J. Hwang, A. Vehtari, Understanding predictive information criteria for Bayesian models, Stat. Comput. 24 (6) (2014) 997–1016, doi:10.1007/s11222-013-9416-2.
[36] A. Vehtari, A. Gelman, J. Gabry, Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC, Stat. Comput. 27 (5) (2017) 1413–1432, doi:10.1007/s11222-016-9696-4.
[37] J. Brady, M. O'Mahony, Modelling charging profiles of electric vehicles based on real-world electric vehicle charging data, Sustain. Cities Soc. 26 (2016) 203–216, doi:10.1016/j.scs.2016.06.014.
[38] Y.B. Khoo, C.-H. Wang, P. Paevere, A. Higgins, Statistical modeling of electric vehicle electricity consumption in the Victorian EV trial, Australia, Transp. Res. Part D Transp. Environ. 32 (2014) 263–277, doi:10.1016/j.trd.2014.08.017.
[39] C. Develder, N. Sadeghianpourhamami, M. Strobbe, N. Refa, Quantifying flexibility in EV charging as DR potential: analysis of two real-world data sets, in: Proceedings of the 7th IEEE International Conference on Smart Grid Communications (SmartGridComm), Sydney, Australia, 2016, pp. 600–605. doi: 10.1109/SmartGridComm.2016.7778827.
[40] N. Sadeghianpourhamami, N. Refa, M. Strobbe, C. Develder, Quantitive analysis of electric vehicle flexibility: a data-driven approach, Int. J. Electrical Power Energy Syst. 95 (2018) 451–462, doi:10.1016/j.ijepes.2017.09.007.