Directional Log-spline Distributions

José T.A.S. Ferreira

Endeavour Capital Management, London, U.K.

Miguel A. Juárez and Mark F.J. Steel*

Department of Statistics

University of Warwick, UK

Abstract

We introduce a new class of distributions to model directional data, based on hyperspherical log-splines. The class is very flexible and can be used to model data that exhibit features that cannot be accommodated by typical parametric distributions, such as asymmetries and multimodality. The distributions are defined on hyperspheres of any dimension and thus, include the most common circular and spherical cases. Due to the flexibility of hyperspherical log-splines, the distributions can closely approximate observed behaviour and are as smooth as desired. We propose a Bayesian setup for conducting inference with directional log-spline distributions where we pay particular attention to the prior specification and the matching of the priors of the log-splines model and an alternative model constructed through a mixture of von Mises distributions. We compare both models in the context of three data sets: simulated data on the circle, circular data on the movement of turtles and a spherical application on the arrival direction of cosmic rays.

Keywords: Directional distributions, hyperspherical splines, mixture of distributions, prior matching, von Mises distributions.

1 Introduction

In this article we introduce a new class of flexible distributions for modelling directional phenomena, *i.e.* phenomena that are characterised by directions in multidimensional real spaces. The most common examples are in the plane and in three-dimensional space, and are usually called circular and spherical, respectively. These types of applications are common in fields such as geology, meteorology, biology and astronomy. Examples of directional applications in higher dimensional spaces have been documented in areas such as Quantum Physics.

Without loss of generality, directional phenomena can be seen as taking place on unit hyperspheres $\mathcal{S}^m = \{d \in \Re^{m+1} : \|d\| = 1\}$, $m = 1, 2, \ldots$, where $\|\cdot\|$ denotes the usual Euclidean or l_2-norm. For example, circular events take place on the unit circle ($m = 1$), and spherical ones on the unit sphere ($m = 2$). Thus, modelling directional data is usually done through a probability distribution on $\mathcal{S}^m$.

A number of distributions on hyperspheres, particularly for $m = 1, 2$, have been proposed in the literature and we refer the reader to monographs such as Mardia and Jupp (2000) and Jammala-madaka and SenGupta (2001). As expected, most of the proposals have focused on simple parametric

---

Address for correspondence: Mark Steel, Department of Statistics, University of Warwick, Coventry, CV4 7AL, U.K. Tel.: +44-24-7652 3369; Fax: +44-24-7652 4532; Email M.F.Steel@stats.warwick.ac.uk.

forms, which are most adequate for applications where complications such as asymmetries and/or multimodality are absent. Here, we focus on more flexible classes of distributions, where the available toolbox is much less developed.

Early generalisations of existing distributions in order to accommodate multimodality were often based on "multiplying the angle" as in Mardia and Spurr (1973) and Wood (1982). Potentially, the most straightforward way to define flexible distributions on hyperspheres is to use mixtures of simpler, parametric distributions (Mooney et al. 2003). Mixture modelling is conceptually quite feasible and has a proven track record. However, mixture modelling has the drawback that, due to the rigid form of the parametric distributions, the number of components required can be quite considerable, leading to both inferential and predictive problems. A different approach is to try to model directional data using kernel density estimation (Hall et al. 1987). Another alternative has been to wrap flexible distributions defined on the real line to the unit circle (Kent and Tyler 1988, Ravindran and Ghosh 2004). This has the slightly unattractive feature that modelling is done on a different topological space than the one of interest. Further, as far as we are aware, there has been no attempt to generalise this methodology to higher dimensional spaces. Recently, a more direct approach to circular distributions has been introduced by Fernández-Durán (2004). The latter author makes use of recent results to define circular density functions as non-negative trigonometric sums. Despite being conceptually interesting, these distributions have two characteristics that make them less appealing: they do not generalise to higher dimensional spaces and they inherit the behaviour of trigonometric sums. The method is limited to the circular case as results on non-negative trigonometric sums are not available when $m > 1$. Also, it is well known that approximating a function, in this case a density, with trigonometric sums usually results in a *wiggly* approximation, unlikely to be supported by most real applications.

We propose the class of directional log-spline distributions. The class applies to any dimension m and its members are characterised by having density functions of hyperspherical log-spline form, *i.e.* the logarithm of the densities are splines on hyperspheres (Wahba 1990, Taijeron et al. 1994). The class borrows the flexibility of hyperspherical splines that can accurately approximate any continuous function. By controlling the smoothness of the spline basis function, it is possible to control the smoothness of the directional density. Finally, hyperspherical splines were shown to be numerically stable (Taijeron et al. 1994). The main computational difficulty involved is that of an m-dimensional integral, which is unlikely to be prohibitive for the practically relevant cases.

Inference with hyperspherical log-spline distributions is implemented within the Bayesian paradigm. We introduce a prior structure on the directional distribution through the parameters of the hyperspherical spline that generates it, thus facilitating the introduction of prior knowledge. The prior distribution is derived using invariance under orthogonal transformations. In addition, we elicit the prior hyperparameters through the induced prior on the mean resultant length (a key measure of concentration), which also allows us to match the prior with that of a competing model. Inference for these models must be conducted using numerical methods, and we define a reversible jump Markov chain Monte Carlo (MCMC) algorithm (Green 1995) that allows inference when the number of basis functions used by the spline is unknown.

We also define a Bayesian model for directional data using mixtures of von Mises distributions, also with an unknown number of components. This model is rather flexible and serves as benchmark for the one employing log-spline distributions. Prior matching allows for a fair comparison of both models through Bayes factors. The models are compared using generated data and in the context of

two real applications. The first one is circular and considers the movement of turtles. The other is spherical and studies the arrival directions of low mu showers of cosmic rays.

Section 2 introduces the class of hyperspherical log-spline distributions, a Bayesian model for their application and details about the numerical methods constructed for conducting inference. Section 3 is the counterpart of Section 2 for the mixtures of von Mises distributions. Section 4 discusses prior matching and elicitation. In Section 5 we analyse a simulated data set as well as two applications with real data and, finally, Section 6 provides some concluding remarks.

2 Directional Log-Spline Distributions

Let $d \in \mathcal{S}^m$ and f denote a density function on $\mathcal{S}^m$. In this article, we propose modelling f through a hyperspherical log-spline as

$$f(d|c, \mathcal{B}, K) = \exp\left\{ c_0 + \sum_{k=1}^{K} c_k R_m^l(d|d_k) \right\}, \tag{1}$$

where K is a positive integer, $c_0 \in \Re$, $c = (c_1, \ldots, c_K)' \in \Re^K$, $\mathcal{B} = \{d_1, \ldots, d_K\}$ is a set of vectors in $\mathcal{S}^m$ and $R_m^l(d|d_k) \in \mathcal{C}(l)$, the set of functions with l continuous derivatives, are real valued spline basis functions defined on $\mathcal{S}^m$ for $k = 1, \ldots, K$. Vectors d_k, $k = 1, \ldots, K$, are usually called knots. Ferreira and Steel (2005) use hyperspherical log-splines in a different context, namely to model the dispersion of a new class of flexible multivariate distributions on $\Re^m$.

The simplicity of hyperspherical splines is very appealing. Despite this feature, hyperspherical splines are very flexible and can closely approximate any smooth function. As will be seen below, hyperspherical splines are constructed in a manner very similarly to splines in real spaces. In most applications, some level of control on the smoothness of the density is likely to be required. As with splines on real spaces, it is possible to control the smoothness of a hyperspherical spline. By choosing $R_m^l(d|d_k) \in \mathcal{C}(l)$, $f(d|c, \mathcal{B}, K)$ will also be $\mathcal{C}(l)$ as the exponential function is $\mathcal{C}(\infty)$. Therefore, controlling the smoothness of f can be trivially done by choosing an appropriate l. Once l is chosen, the form of the basis functions $R_m^l(d|d_k)$ remains to be specified.

Basis functions for spline interpolation and smoothing on the circle ($m = 1$) and the sphere ($m = 2$) were determined in Wahba (1975) and Wahba (1981), respectively. Taijeron et al. (1994) generalised the above results and provided the solution for the problems in any dimension $m > 1$. In the form of (1), the solution is provided when $R_m^l(d|d_k)$, $k = 1, \ldots, K$, represent a family of reproducing kernels related to Green's function for the Laplacian on the hyperspheres. For $m \neq 1$, closed forms for $R_m^l(d|d_k)$ are not known, leading to the use of approximating reproducing kernels for which a closed form is available. For our modelling of f the use of the approximating reproducing kernels does not lead to any modelling restrictions. Thus, we will also use the notation $R_m^l(d|d_k)$ for the approximating kernel.

For the case in S^1, the reproducing kernel is given by

$$R_1^l(d|d_k) = \frac{(-1)^{l/2}(2\pi)^{l-1}}{(2\pi)!} B_{l+2}\left(\frac{\theta}{2\pi}\right),$$

with θ the angle between d and d_k and B_{l+2} the Bernoulli polynomial of degree $l + 2$ (Abramowitz and Stegun 1972).

For $m > 1$, Taijeron et al. (1994) indicate that the approximating reproducing kernels can be defined by

$$R_m^l(d|d_k) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{m-1}\left[\frac{q(l,m-1,d'd_k)}{l!} - \frac{1}{(l+1)!}\right],$$

where

$$q(r,s,z) = \int_0^1 (1-h)^r(1-2hz+h^2)^{-s/2}dh.$$

For $r = 0, 1, \ldots$ and $m > 1$, closed forms for R_m^l are available and can be calculated easily with a symbolic mathematics package. Examples can be found in Wahba (1981) and Taijeron et al. (1994).

To choose the coefficient vector c we use the facts that $f(d|c, \mathcal{B}, K)$ is proportional to

$$\exp\left\{c^* + \sum_{k=1}^K c_k R_m^l(d|d_k)\right\}, \tag{2}$$

with $c^* \in \Re$, and that the logarithm of (2) can be thought of as a spline interpolating a vector of values $v = (v_1, \ldots, v_K)' \in \Re^K$ associated with the knots. Vector c can then be obtained as the solution of the linear system of equations

$$\begin{pmatrix} 1_K & R \\ 0 & 1'_K \end{pmatrix}\begin{pmatrix} c^* \\ c \end{pmatrix} = \begin{pmatrix} v \\ 0 \end{pmatrix}, \tag{3}$$

where 1_K is the K-dimensional vector of ones and R is a $K \times K$ matrix with entry (i, j) equal to $R_m^l(d_i|d_j)$. Vector c is such that its elements add up to zero. The incorporation of c^* ensures that the matrix on the left-hand side of (3) is non-singular as long as d_k, $k = 1, \ldots, K$ are distinct. So choosing c is equivalent to the easier task of choosing v. The value of c_0 is determined by the fact that f in (1) is a density, and therefore integrates to one, which leads to

$$c_0 = -\log\int_{\mathcal{S}^m}\exp\left\{\sum_{k=1}^K c_k R_m^l(d|d_k)\right\}\mathrm{d}d,$$

and can be computed numerically.

The flexibility of spline models is well documented in the literature (see *e.g.* Wahba, 1990). By allowing an appropriate number of knots, positioned at suitable locations, it is possible to approximate any function in $\mathcal{C}(l)$. This characteristic guarantees the flexibility of the class of directional log-spline distributions.

2.1 Bayesian Model

In the remaining part of this section, we assume that the process generating the directional data has an underlying log-spline distribution with density f.

The prior distribution is elicited by considering the construction of the density function as the interpolating log-spline explained above. Assuming a fixed spline basis function, the density function is completely determined by the number of knots K, the knot locations $\mathcal{B}$ and the knot values v. Thus, the prior on f can be completely determined by a prior on these parameters.

If the number of knots is not fixed, then the parameter space where inference is to take place has variable dimension. Eliciting a prior distribution is facilitated by conditioning on the dimension of the parameter space, *i.e.* by focusing on

$$P_{K,\mathcal{B},v} = P_{\mathcal{B},v|K}P_K.$$

4

We assume that $\mathcal{B}$ and v are independent, given K.

In a Bayesian curve fitting spline setting, two prior distributions on K have often been suggested: a Uniform prior on the set of integers between one and some $K_{\max}$, and a Poisson prior, restricted to $K > 0$ (see DiMatteo *et al.* 2001 and references therein). Here we use a Poisson prior with parameter λ_K, but restricted to $K > 1$, since the model reduces to the Uniform model for $K = 1$. Choosing a relatively small λ_K naturally penalises models with a very large number of knots (thus avoiding overfitting), without, however, imposing an exact upper bound.

The prior distribution on the location of the knots is set using invariance arguments. In the absence of further information, the prior should not favour any particular set of directions. Thus, assuming independent and Uniform priors on $\mathcal{S}^m$ for the elements of $\mathcal{B}$ is immediate.

The prior distribution on f is completed by setting the prior on the knot values. For each of the elements of v, we assume an independent Normal prior with mean $\log\{\Gamma(\frac{m+1}{2})\} - (\frac{m+1}{2})\log(2\pi)$ and variance σ_v^2. This prior is centred on the Uniform distribution on the sphere, with the prior flexibility of the log-spline distribution controlled by σ_v^2. As σ_v^2 tends to zero, the model reduces to the Uniform model on $\mathcal{S}^m$.

By assigning a Uniform prior on the location and a jointly independent prior on the values at the knots, we define a prior that is invariant to linear orthogonal transformations (rotations or roto-inversions) and is, therefore, uninformative about the direction. Prior knowledge about particular directions and of any existing asymmetries could be incorporated by making the priors on $\mathcal{B}$ and v dependent and varying with location. Section 4 will discuss how λ_K and σ_v are chosen.

2.2 Inference

Inference for the regression model introduced in the previous section requires numerical methods. Further, as the number of knots K is left unknown, inference must be conducted on a space with variable dimension. Here we briefly describe the details of a reversible jump MCMC sampler that we use in this context.

For conducting inference, it is helpful to use a polar transformation of the direction. Let $d = [d_1(\omega), \ldots, d_{m+1}(\omega)]' \in \mathcal{S}^m \subset \Re^{m+1}$ and let $\omega = (\omega_1, \ldots, \omega_m) \in \Omega = [0, \pi)^{m-1} \times [0, 2\pi)$ be such that

$$d_j(\omega) = \left(\prod_{i=1}^{j-1} \sin \omega_i\right) \cos \omega_j, \quad 1 \leq j \leq m \quad \text{and} \quad d_{m+1}(\omega) = \prod_{i=1}^{m} \sin \omega_i.$$

In order to parameterise the knot location d_k we use polar coordinates parameterised by $\omega^k = (\omega_1^k, \ldots, \omega_m^k)$ as above. The Uniform prior of d_k on $\mathcal{S}^m$ is translated to the prior on ω^k given by

$$p(\omega^k) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{2\pi^{(m+1)/2}} \prod_{j=1}^{m-1} \sin^{m-j} \omega_j^k, \quad k = 1, \ldots, K.$$

We suggest a reversible jump MCMC characterised by four move types: (a) change values, (b) change locations, (c) introduce a new knot and (d) remove an existing knot. Moves (a) and (b) are fixed dimension moves and are defined as random walk Metropolis-Hastings moves, whilst moves (c) and (d) change the dimension of the parameter space.

For move (a), we select one existing knot $k \in \{1, \ldots, K\}$ and then propose a new knot value v_k^* sampled from a Normal distribution with mean v_k, the current value of the knot, and variance tuned so as to obtain an acceptance rate close to 30%.

5

Moves of type (b) also start with the choice of one existing knot $k \in \{1, \ldots, K\}$. A component $j \in 1, \ldots, m-1$ of ω^k is then selected and a new value for ω_j^k* is proposed from a Normal distribution, centred at ω_j^k, variance tuned to obtain a similar acceptance rate as for move (a), and truncated to $[0, \pi)$ if $j \leq m-1$ or $[0, 2\pi)$ if $j = m$. When $m = 1$, the location of each knot is parameterised using only one angle. In this case, ordering the location of the knots is possible and a sampler similar to the change-point locations sampler suggested in Section 4 of Green (1995) could be used. We did not find any significant differences between the samplers and therefore we always use the former sampling scheme which is valid for any dimension.

For the introduction of a new knot, we set $K^* = K + 1$, sample a location d_{K^*} in $\mathcal{S}^m$ from the Uniform distribution and then sample a knot value v_{K^*} from a Normal distribution centred on the current value of the logarithm of (2) at d_{K^*} and variance tuned as previously. The proposed new set of locations and values are then $\mathcal{B}^* = \mathcal{B} \cup d_{K^*}$ and $v^* = (v', v_{K^*})'$.

The removal of an existing knot can only be done when the number of existing knots is larger than two and is performed by selecting $k \in \{1, \ldots, K\}$ from a discrete Uniform distribution and setting $K^* = K - 1$, $\mathcal{B}^* = \mathcal{B} \setminus d_k$ and $v^* = v(1, \ldots, k-1, k+1, \ldots, K)'$.

3   Mixtures of von Mises Distributions

One of the simplest ways to create flexible distributions is to consider mixtures of other, simpler, parametric ones. In order to evaluate the quality of directional log-spline distributions, it seems appropriate to compare them with mixtures of von Mises distributions, as defined here. Mixtures of von Mises distributions are not new and have been used before (see, *i.a.* Mooney et al., 2003). Nevertheless, in this section we introduce a Bayesian setup for inference with mixtures of von Mises distributions that, to the extent of our knowledge, is novel.

Let d and μ be vectors in $\mathcal{S}^m$, $m = 1, 2, \ldots$, and let $\tau \in \Re_+$. Then, d has a von Mises distribution with mean direction μ and concentration parameter τ if its probability density function is

$$f(d|\mu, \tau) = \left(\frac{\tau}{2}\right)^{\frac{m-1}{2}} \frac{1}{\Gamma(\frac{m+1}{2})I_{\frac{m-1}{2}}(\tau)} \exp\{\tau\mu'd\},$$

where I_ν denotes the modified Bessel function of the first kind and order ν defined by

$$I_\nu(\tau) = \frac{1}{2\pi} \int_0^{2\pi} \cos\nu\theta \exp\{\tau\cos\theta\}d\theta.$$

We define $g(d|w, \boldsymbol{\mu}, \boldsymbol{\tau})$ to be a mixture of von Mises distributions is its density has the form

$$g(d|w, \boldsymbol{\mu}, \boldsymbol{\tau}) = \sum_{k=1}^K w_k f(d|\mu_k, \tau_k), \tag{4}$$

with $w = (w_1, \ldots, w_K)'$ a vector of positive constants such that $\sum_{k=1}^K w_k = 1$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)'$ a vector of K directions in S^m and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_K) \in \Re_+^K$. Obviously, g corresponds to a (single-component) von Mises distribution when $K = 1$.

3.1   Bayesian Model

In the definition of the Bayesian model with directional log-spline distributions, we also allowed for inference on the number of components K in (1). In the same spirit, we will also conduct inference

6

on the number of von Mises components in (4). Thus, the unknown parameters are the number of components K, the vector of weights w, the vector of mean directions $\boldsymbol{\mu}$ and the vector of concentration parameters $\boldsymbol{\tau}$. We assume a prior structure of the form

$$P_{K,w,\boldsymbol{\mu},\boldsymbol{\tau}} = P_{w,\boldsymbol{\mu},\boldsymbol{\tau}|K} P_K = P_{w|K} P_{\boldsymbol{\mu}|K} P_{\boldsymbol{\tau}|K} P_K.$$

The prior on K is as in Subsection 2.1, except that we now only truncate to $K > 0$ and allow for $K = 1$ (which leads to the standard von Mises model). Given K, we assume a Dirichlet prior for w with parameter vector $\kappa = (\kappa_1, \ldots, \kappa_K)$. A similar prior was used in the context of mixtures of Normal distributions in Richardson and Green (1997). The prior distribution is now completed by specifying the priors on the parameters of the von Mises distributions. Conjugate priors for mean direction and concentration parameters of a single von Mises distribution were used in *e.g.* Mardia and El-Atoum (1976), Guttorp and Lockhart (1988) and Nuñez-Antonio and Gutiérrez-Peña (2005). Here we do not need conjugacy of the prior and use invariance arguments and an elicitation procedure set out in Section 4. The prior on $\boldsymbol{\mu}$ is selected using invariance arguments. Assuming prior ignorance about the direction of the data leads us to choose independent Uniform priors on $\mathcal{S}^m$ for each of the components of $\boldsymbol{\mu}$. Finally, the components of $\boldsymbol{\tau}$ are assumed independent with each assigned an Exponential prior with mean parameter b. Section 4 will discuss how we choose b, along with λ_K and κ.

3.2   Inference

Inference for this class of models again requires the use of numerical methods and, as in Subsection 2.3, we use a reversible jump MCMC sampler and we parameterise directions through polar coordinates.

Our proposed sampler has five move types: (a) change weights, (b) change mean directions, (c) change concentrations, (d) introduce a new component and (e) remove an existing component. Moves (a)-(c) are fixed dimension moves and are very similar to the ones explained in Subsection 2.2, with the addition of the restriction that the elements of w must add up to one.

For the creation of a new component, we propose a new set of parameters $w^* = (\frac{w_1}{1+u}, \ldots, \frac{w_K}{1+u}, u)'$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K, \mu_{K+1})'$ and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_K, \tau_{K+1})'$, where u is sampled from the Uniform distribution in $(0,1)$ and μ_{K+1} and τ_{K+1} are sampled from the respective prior distributions. Finally, removing an existing component is achieved by proposing a new set of parameters, obtained from the current state of the sampler by removing the kth component, $k = 1, \ldots, K$ and rescaling w so that it adds up to one.

Some classical ways of fitting this model are discussed in Fisher et al. (1987) and Mooney et al. (2003). Issues of identifiability are discussed in Kent (1983).

4   Prior Matching

In order to compare the different models above, we will use the formal tool of Bayes factors. The sensitivity of Bayes factors to the prior assumptions is a well-documented feature (see *e.g.* Kass and Raftery, 1995). Therefore, a fair comparison of the models needs to be based on priors that are comparable. We achieve this through prior matching, which entails that the prior distributions of both models represent very similar prior beliefs on a certain quantity of interest. Here we will focus on the mean resultant length, defined as

$$\rho = \|\mathrm{E}(d)\|, \tag{5}$$

i.e. the Euclidean norm of the mean direction. The measure $\rho \in [0, 1)$ is an important characteristic of directional distributions (for any m), describing the dispersion of the distribution of d (see Mardia and Jupp, 2000). If the distribution is tightly concentrated then ρ is close to one, whereas for a Uniform distribution on the sphere, the mean resultant length will be zero.

Focusing on the implied priors for ρ also facilitates prior elicitation, as ρ is a measure that we can interpret. We want the prior to induce a distribution for ρ with most mass fairly close to zero (low concentration), but without totally excluding values of ρ over, say, 0.5 (high concentration). This is in line with a lack of a strong prior belief that a certain direction will dominate and is similar to a fairly vague prior for observations on the real line.

The full description of the models still requires the specification of the hyperparameters λ_K, σ_v, κ and b. Barring a difference in the truncation of K ($K > 1$ for the log-spline model and $K > 0$ for the von Mises mixture), we choose the same Poisson (λ_K) prior for the number of knots in the log-spline model and the number of components in the von Mises mixture model. We choose $\lambda_K = 5$ reflecting our expectation that a small number of components will be sufficient. We also select κ to be a vector of ones. We have experimented with different prior hyperparameters (for example, setting λ_K to be different for both models), but found that we can best achieve the matching and the desired prior for ρ by using these choices for λ_K and κ while adjusting the values of σ_v and b. Figure 1 presents Monte Carlo simulation estimates of the prior densities of ρ for both models in the cases with $m = 1$ and $m = 2$, covering most applications of practical interest. To achieve a reasonable and matched prior, we recommend the following values: $\sigma_v = 1/4$ and $b = 1$ for $m = 1$ and $\sigma_v = 5/3$ and $b = 4/3$ when $m = 2$. For larger m the same approach can easily be followed.
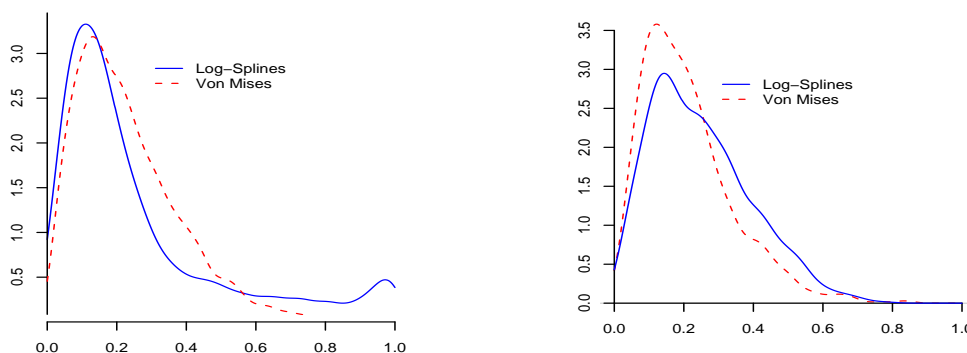

Figure 1: Prior density functions of ρ for log-spline (solid) and mixtures of von Mises (dashed) models. The left panel is for the case $m = 1$ (circular) and the right panel is for $m = 2$ (spherical).

5 Applications

In the sequel, we always make use of cubic hyperspherical spline basis functions, corresponding to $l = 2$. This generates density functions that we feel are sufficiently smooth for most applications. Nevertheless, altering l does not imply any methodological change.

In all cases, the prior used is as explained in Sections 2, 3 and 4, for the appropriate value of the dimension $(m + 1)$. Two of the applications are circular, *i.e.* $m = 1$, and for the last one $m = 2$

(spherical).

Inference is conducted using MCMC chains of 1,200,000 iterations, retaining every 10th sample after a burn-in period of 200,000 draws. Convergence is never a problem: chains of one tenth this length lead to virtually identical results in all applications.

We assess the relative adequacy of different models using Bayes factors. Estimates of marginal likelihoods are obtained using the p_4 measure in Newton and Raftery (1994), with their $\delta = 0.4$. Various values of δ in the range $(0.1, 0.8)$ were tried, without any substantive effect on the results.

5.1   Artificial data

To get a first idea of the behaviour of the log-spline model, we have generated a sample of 250 observations on a circle from a mixture of von Mises distributions as in (4) with $K = 2$, $w_1 = 0.65$, $\mu_1 = \pi/4$, $\tau_1 = 3$, $\mu_2 = (7/4)\pi$ and $\tau_2 = 6$. We used both the log-spline model and the mixture of von Mises model on these data. The posterior distributions on the mean direction and the mean resultant length implied by these models are quite close and are concentrated around their sample counterparts. Interestingly, for the particular data analysed the log-splines model is favoured over the von Mises mixture (even though the latter model was used to generate the data) by a Bayes factor of 3.2.

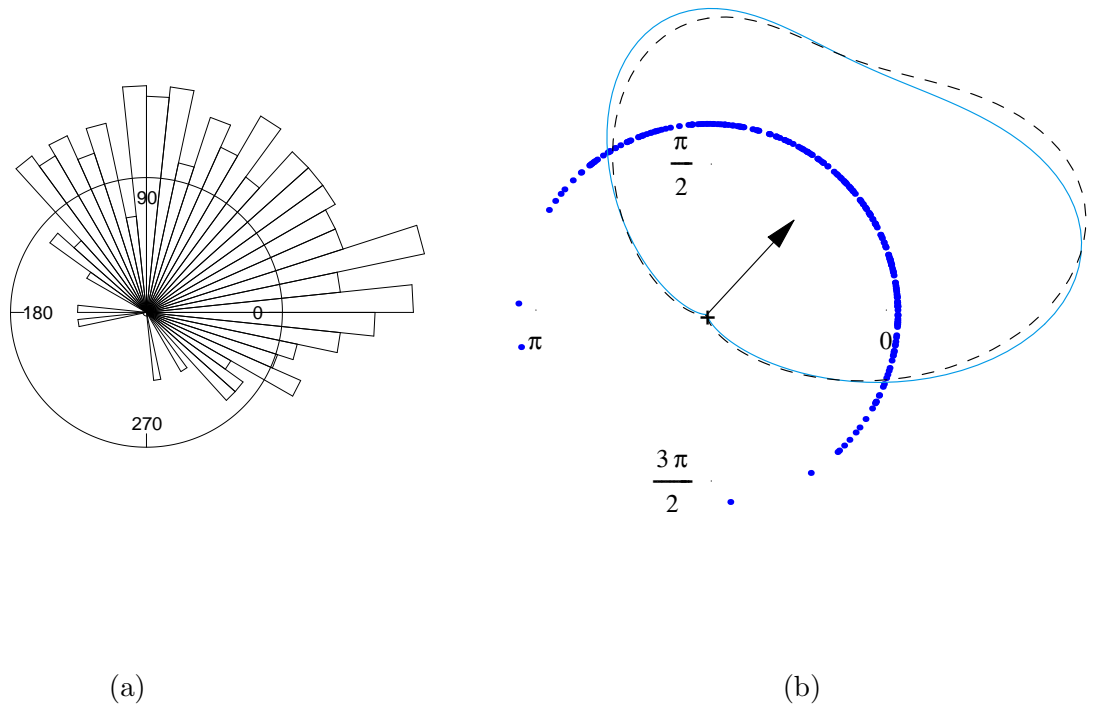(a)   (b)

Figure 2: (a) Rose diagram of the artificial data. (b) Circular plot of the data (dots) and representation of the mean posterior directional log-spline (solid) and mixture of von Mises (dashed) densities. The direction of the vector is the mean direction of the data and its length is the mean resultant length.

In order to get a better feeling for how the models fit the data, Figure 2 displays the data: they are represented through a rose diagram (circular histogram with the area of each sector proportional to the observed frequency) in panel (a) and indicated by dots in (b)). Figure 2(b) also presents the posterior predictive densities corresponding to the two models. Predictive density values are proportional to the distance from the centre of the unit circle (indicated by +) and illustrate that the two models lead to very similar distributions. The mean direction of the data is indicated by the vector in the figure,

9

with length equal to the observed mean resultant length.

In the sampler we find maximum values for K (the number of knots for the log-splines model and the number of components for the mixture of von Mises model) of 7 for both models. The von Mises mixture model tends to have smaller K, with a mean of 2.4 and a median of 2 (in line with the actual model that generated the data), as opposed to 4.3 and 4, respectively, for the log-splines model.

5.2 Turtle data

Our first application to real data is also circular and concerns the orientation of turtles after laying eggs. The data consist of 76 observations which can be found in Mardia and Jupp (2000, p. 9) and have been analysed on a number of occasions, including Stephens (1969) and, more recently, Fernández-Durán (2004). A scatterplot of the data is presented in Figure 3. Summary statistics of the data are 0.36π for the angle of the mean direction (indicated by the direction of the vector in Figure 3) and 0.497 for the mean resultant length (represented by the length of the vector in Figure 3).
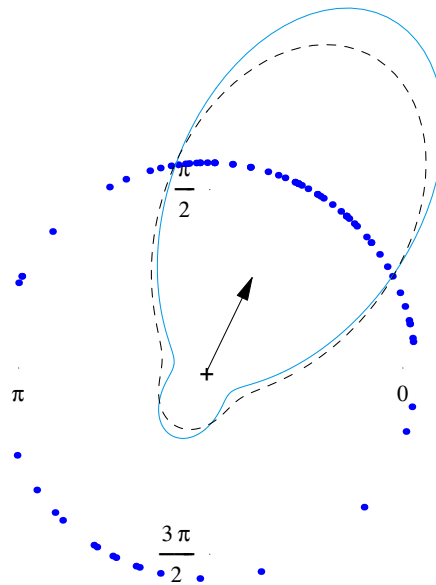

Figure 3: Circular plot of the turtle data (dots) and representation of the mean posterior directional log-spline (solid) and mixture of von Mises (dashed) densities. The direction of the vector is the mean direction of the data and its length is the mean resultant length.

Formal comparison of the two models reveals an advantage for the directional log-spline one, with a Bayes factor of 7. The differences between the directional log-spline and the mixture of von Mises models are illustrated in Figure 3, where we present the posterior predictive densities for both models. The densities are proportional to the distance from the centre of the circle defined by the dots. Clearly, both models allocate most mass to the two areas where the data is concentrated and little mass to areas with almost no data. However, the directional log-spline model produces a mean posterior density that is somewhat more concentrated than the von Mises model. Indeed, the posterior density for the mean resultant length, presented in the right panel of Figure 4 indicates that the concentration tends to be slightly higher for the log-splines model. This is a consequence of the large flexibility of the log-splines model, which allows the density to go virtually to zero in those regions where no observations occur. Figure 4 also presents the posterior density of the mean direction for both models,

10

which are very similar and in line with the mean direction of the data. Incidentally, this figure also illustrates the inadequacy of this summary for multimodal distributions.
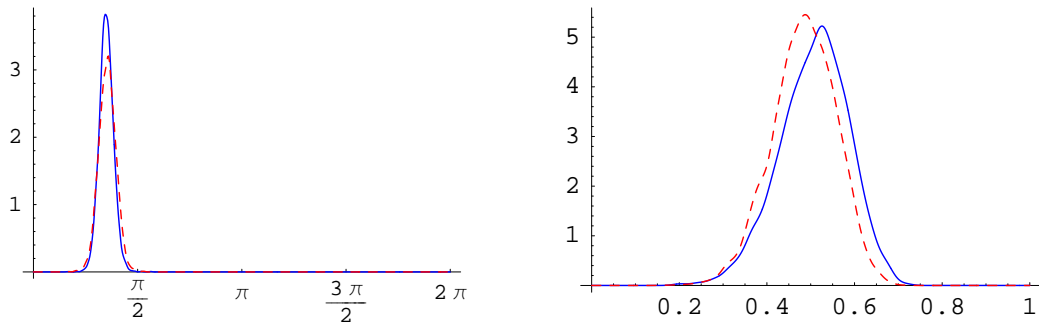

Figure 4: Turtle data: posterior densities of the angle of the mean direction and the mean resultant length (ρ) for log-spline (solid) and mixtures of von Mises (dashed) models.

In terms of K, the number of knots for the log-spline model and the number of components in the mixture of von Mises distributions, the results are quite similar. For both models the posterior median of K is four with a mean of 4.6 for the log-spline model and 4.2 for the von Mises mixture. The main difference is that the sampler for the mixture model visits models with a larger number of components than the log-spline one, with maxima 14 and 10, respectively.

An aspect that is interesting to analyse for a flexible model like the one generated by directional log-spline distributions is the variability of the distribution. Figure 4 already presented the posterior distributions of summary measures of the distribution. To further illustrate this variability, Figure 5 presents one hundred directional log-spline densities, corresponding to models visited by the sampler. Differences between the densities are evident. Nonetheless, we can see that the main characteristics of the data are consistently being detected, leading to the two areas with large posterior mass. The same plot for the von-Mises mixture model looks very similar.
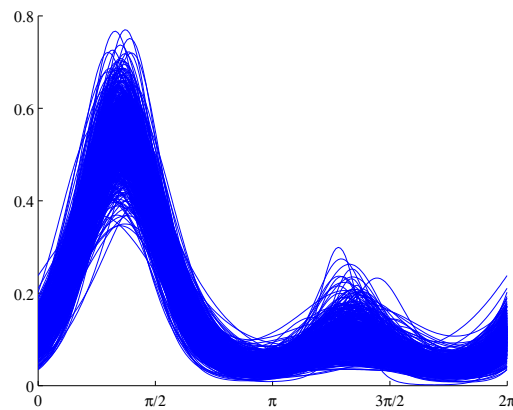

Figure 5: Turtle data: directional log-spline densities corresponding to one hundred samples from the posterior, as a function of the angle of the direction.

11

5.3 Cosmic ray data

We now turn our attention to a spherical application dealing with the arrival of low mu showers of cosmic rays. The dataset used here, with 148 observations, originates from Toyoda et al. (1965) and can be found in Fisher et al. (187, pp. 280-281). The data are depicted in Figure 6 from a variety of viewpoints. The mean resultant length of the data is 0.241 and the mean direction is $\{0.071, -0.317, -0.946\}$, where the first component corresponds to the north-south direction.
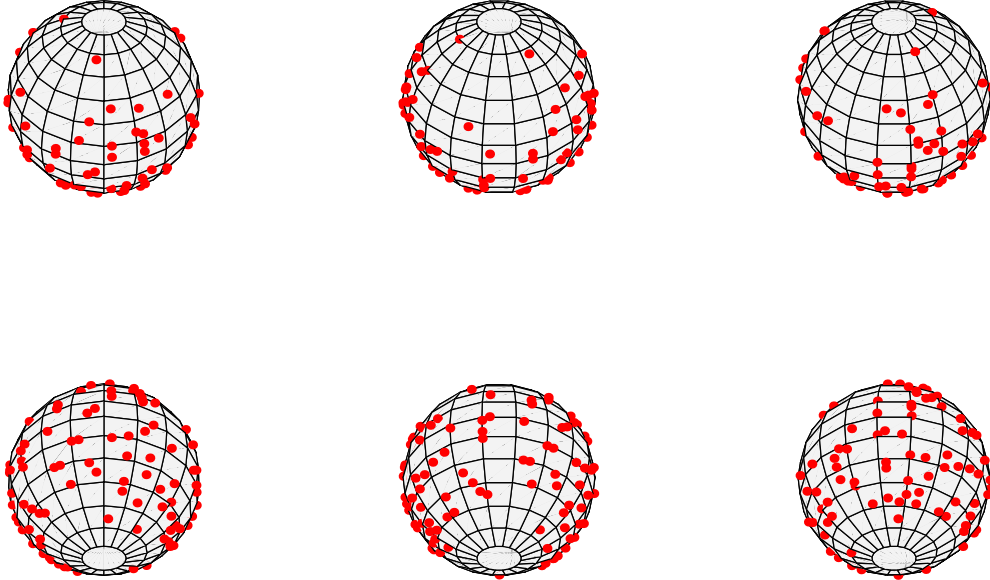

Figure 6: Arrival directions of low mu showers of cosmic rays from different viewpoints. Observations are indicated by dots. The top row shows the North pole and the bottom row the South pole.

The model using directional log-spline distributions is strongly preferred by the data to the one based on mixtures of von Mises distributions, with a Bayes factor estimated at 3×10^{14}. From the posterior distribution of the mean resultant length in Figure 7, we deduce that both models are in line with the sample value, with the posterior of the log-spline model relatively less spread out.

Plots of the posterior predictive density function implied by both models are given in Figure 8, illustrating the differences between the models. These plots are in terms of the polar angles, where the angle ω_1 in $(0, \pi)$ is the colatitude ($\pi/2$-latitude) and the angle running from 0 to 2π (ω_2) denotes the longitude. Density values are lowest on the poles (corresponding to the lines $\omega_1 = 0$ and $\omega_1 = \pi$), where there are very few observations (this is, of course, at least partly induced by the topology of the polar representation). This can be seen from Figure 6, but is even clearer when we consider Figure 8(c), where we plot the observations as a function of the polar angles. Clearly, the two predictive density surfaces share common characteristics, such as the low values at the poles and the main mode around $(\omega_1, \omega_2) = (3\pi/4, 5\pi/4)$. There are also important differences; in particular, the log-spline model seems to induce somewhat more irregular shapes, certainly along the ω_2 dimension. This is, again, indicative of the large flexibility of this model.

The number of components of the von Mises model (maximum 16, mean 4.3 and median 4) tends
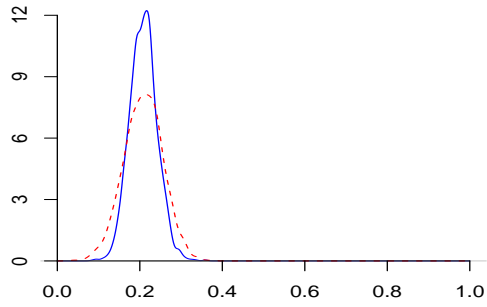
Figure 7: Cosmic ray data: posterior density of the mean resultant length (ρ) for log-spline (solid) and mixtures of von Mises (dashed) models.
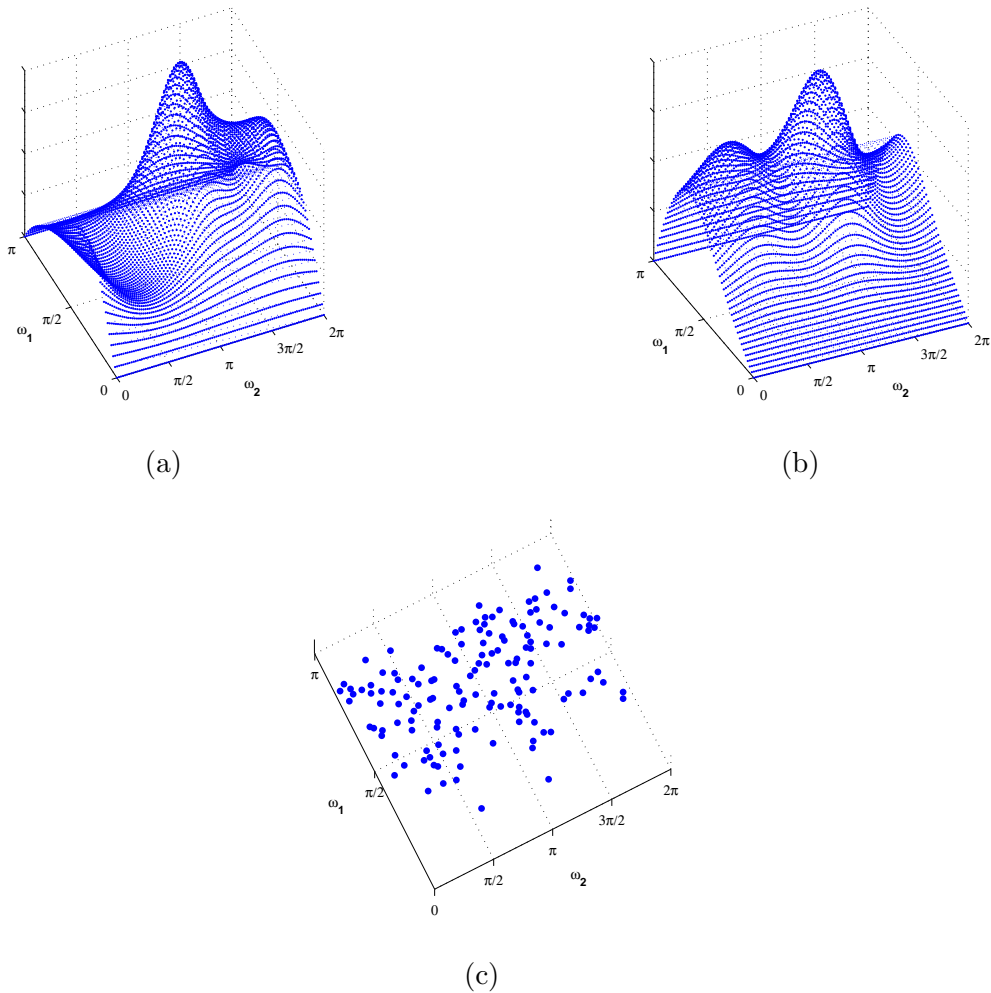

(a)

(b)

(c)

Figure 8: Cosmic ray data: Plots of the predictive densities for log-spline (a) and mixtures of von Mises (b) models, in terms of the polar angles; (c) data plotted as a function of the polar angles.

to be somewhat higher than the number of knots in the log-spline model with a maximum of 11, a mean of 3.4 and a median of 3. In particular the log-spline model manages to produce quite complex

13

distributional shapes with surprisingly small values for K.

6   Conclusion

In this paper we have proposed a very flexible class of directional distributions, based on hyperspherical log-splines. Such distributions can easily deal with nonstandard features of directional data, such as multimodality and skewness, and are applicable in hyperspheres of any dimension. We contrast this new class of distributions with mixtures of von Mises distributions, which is, arguably, the main competing model currently available in the literature.

A particular challenge for a Bayesian analysis of these models is to specify reasonable prior distributions, which are comparable across models and thus allow for model choice or model averaging via Bayes factors. The specification of the priors is based on invariance arguments and on the implied prior distribution for the mean resultant length, a measure of concentration of the distribution. The latter also allows us to match the priors for both models.

The use of reversible jump MCMC samplers makes Bayesian inference in these models quite feasible.

We compare the log-spline and the mixture of von Mises models in the context of three data sets. One set of circular observations generated from a mixture of two von Mises distributions, the circular turtle data and spherical data on cosmic rays. In all applications, the data favour the log-spline model (moderately for both circular data sets and overwhelmingly in the cosmic ray application). Our examples illustrate the large amount of flexibility of the log-spline model. Despite this flexibility, the number of knots used in the log-spline model is quite small (with a mean of less than five in these applications), facilitating the computations. In fact, with the priors proposed here, neither model leads to overfitting the data.

References

Abramowitz, M. and Stegun, I. (eds) (1972). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, Dover, New York.

DiMatteo, I., Genovese, C. R. and Kass, R. E. (2001). Bayesian curve-fitting with free-knot splines, *Biometrika* **88**: 1055–1071.

Fernández-Durán, J. J. (2004). Circular distributions based on nonnegative trigonometric sums, *Biometrics* **60**: 499–503.

Ferreira J. T. A. S. and Steel, M. F. J. (2005). Modelling directional dispersion through hyperspherical log-splines, *J. Roy. Statist. Soc. B* **67**: 599–616.

Fisher, N. I., Lewis, T. and Embleton, B. J. J. (1987). *Statistical Analysis of Spherical Data*, Cambridge University Press, Cambridge.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika* **82**: 711–732.

Guttorp, P. and Lockhart, R. A. (1988). Finding the location of a signal: A Bayesian analysis, *J. Amer. Statist. Assoc.* **83**: 322–329.

Hall, P., Watson, G. S. and Cabrera, J. (1987). Kernel density estimation with spherical data, *Biometrika* **74**: 751–762.

Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics*, World Scientific, Singapore.

Kass, R. E. and Raftery A. E. (1995). Bayes factors, *J. Amer. Statist. Assoc.* **90**: 773–795.

Kent, J. T. (1983). Identifiability of finite mixtures for directional data, *Ann. Statist.* **15**: 247–254.

Kent, J. T. and Tyler, D. E. (1988). Maximum likelihood estimation for the wrapped cauchy distribution, *J. Appl. Stat.* **15**: 247–254.

Mardia, K. V. and El-Atoum, S. A. M. (1976). Bayesian inference for the von Mises-Fisher distribution, *Biometrika* **63**: 403–405.

Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*, Wiley, Chichester.

Mardia, K. V. and Spurr, B. D. (1973). Multisample tests for multimodal and axial circular populations, *J. Roy. Statist. Soc. B* **35**: 422–436.

Mooney, J. A., Helms, P. J. and Jolliffe, I. T. (2003). Fitting mixtures of von Mises distributions: a case study involving sudden infant death syndrome, *Comp. Statist. Data Anal.* **41**: 505–513.

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap (with discussion), *J. Roy. Statist. Soc. B* **56**: 3–48.

Nuñez-Antonio, G. and Gutiérrez-Peña, E. (2005). A Bayesian analysis of directional data using the von Mises-Fisher distribution, *Communs Statist. Simul. Comp.* **34**: 1–11.

Ravindran, P. and Ghosh, S. K. (2004). Bayesian analysis of circular data using wrapped distributions, *Institute of Statistics Mimeographs 2564.* North Carolina State University.

Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion), *J. Roy. Statist. Soc. B* **59**: 731–792.

Stephens, M. A. (1969). Techniques for directional data, *Technical Report 150.* Department of Statistics, Stanford University.

Taijeron, H. J., Gibson, A. G. and Chandler, C. (1994). Spline interpolation and smoothing on hyperspheres, *SIAM J. Sci. Comput.* **15**: 1111–1125.

Toyoda, Y., Suga, K., Murakami, K., Hasegawa, H., Shibata, S., Domingo, V., Escobar, I., Kamata, K., Bradt, H., Clark, G. and La Pointe, M. (1965). Studies of primary cosmic rays in the energy region 10^{14} eV to 10^{17} eV (Bolivian Air Shower Joint Experiment), *Proc. Int. Conf. Cosmic Rays*, Vol. 2, The Institute of Physics and The Physical Society, London, pp. 708–711.

Wahba, G. (1975). Smoothing noisy data by spline functions, *Numer. Math.* **24**: 383–393.

Wahba, G. (1981). Spline interpolation and smoothing on the sphere, *SIAM J. Sci. Statist. Comput.* **2**: 5–16.

Wahba, G. (1990). *Spline Models for Observational Data*, CBMS-NSF Regional Conference Series in Applied Mathematics, Philadelphia.

Wood, A. T. A. (1982). A bimodal distribution on the sphere, *Appl. Statist.* **31**: 52–58.