Solid Earth

Open Access

Modelling complex geological circular data with the projected normal distribution and mixtures of von Mises distributions

R. M. Lark[1], **D. Clifford**[2], **and C. N. Waters**[1]

[1]British Geological Survey, Keyworth, Nottingham, UK
[2]CSIRO Computational Informatics, Ecoscience Precinct, G.P.O. Box 2583, Brisbane, Qld 4001, Australia

Correspondence to: R. M. Lark (mlark@bgs.ac.uk)

Abstract. Circular data are commonly encountered in the earth sciences and statistical descriptions and inferences about such data are necessary in structural geology. In this paper we compare two statistical distributions appropriate for complex circular data sets: the mixture of von Mises and the projected normal distribution. We show how the number of components in a mixture of von Mises distribution may be chosen, and how one may choose between the projected normal distribution and the mixture of von Mises for a particular data set. We illustrate these methods with a few structural geological data, showing how the fitted models can complement geological interpretation and permit statistical inference. One of our data sets suggests a special case of the projected normal distribution which we discuss briefly.

1 Introduction

Circular data are commonly encountered in geology. A circular variable may be a direction, such as the direction of dip of a fault or bedding plane, or a palaeomagnetic vector. Alternatively a circular variable may be an orientation (which could be expressed by either one of two opposite directions) such as the orientation of vertical faults or bedding planes or the orientation of primary palaeocurrent lineations where the direction of flow is unknown. Directional variables are distributed on the unit circle, and so have the particular property that the upper bound, 2π radians, and the lower bound, zero, are equivalent. They therefore cannot be treated as though they were distributed over some subset of the real numbers, and require special treatment for statistical analysis (Mardia and Jupp, 2000).

The von Mises distribution is commonly used for the statistical analysis of circular data. The distribution has two parameters, a mean direction and a concentration parameter, κ. The latter parameter measures the dispersion of the variable about the mean. At its minimum, $\kappa = 0$, the von Mises distribution is the uniform distribution over the unit circle. As κ increases so the von Mises distribution increasingly resembles the Gaussian with variance $1/\kappa$.

The von Mises distribution has been widely used in earth science, and software for analysis of circular data including methods based on the von Mises distribution have been made available (e.g. Jones, 2006). For example, Coblentz and Richardson (1995) examined global data on maximum horizontal compressive stress, and examined local evidence for coherence of stress direction by the Rayleigh test, which is equivalent to a comparison of the von Mises distribution with $\kappa > 0$ against a uniform alternative. Witts et al. (2012) used a similar procedure to identify trends in palaeocurrent data from the dip and dip azimuth of sand bars in a Cenozoic sedimentary succession in Indonesia. Sen and Mamtani (2006) used the κ parameter of the von Mises distribution to characterise the preferential orientation of biotite in thin sections of granite which was related to variations in regional strain. O'Brien et al. (2012) used von Mises distributions to characterise the orientations of fault plane solutions – with confidence intervals – before, during and after seismic swarms.

The von Mises distribution is symmetrical and unimodal. Circular data in earth sciences may often have a more complex distribution than this. There may be continuous variation in the orientation of particular features; for example, the preferential directions of structures in sedimentary deposits in palaeochannels are likely to respond to channel orientation

and flow direction, which will vary on metre to kilometre scales. Distributions of circular data may therefore be asymmetric and multimodal.

One way to model such complex variation is to treat it as a mixture of von Mises distributions (MVM). The MVM model that we considered, with g components, has $3g - 1$ parameters: g mean directions and values of κ and the $g - 1$ independent proportions of the components. By including sufficient components it is possible to model a distribution of angles with multiple modes and asymmetry.

An alternative model for more complex distributions of circular data is the projected normal (PN) distribution. This distribution, and its flexibility, can be understood intuitively. Consider a distribution of points on a scatter plot formed by observations drawn from a pair of correlated normal random variables. One may draw a line from each point to the origin, forming a vector with some angle θ. If the correlation between the two variables was zero, and they each had mean zero, then the angles will have a uniform distribution around the circle. If the means were both zero but the correlation were quite large, then an 'antipodal' circular variable with two peaks in its distribution, separated from each other by π radians (180 degrees), would be generated. By allowing the means of the variables to vary the distribution of angles can be made unimodal, or bimodal but asymmetric and non-antipodal. The model has considerable flexibility despite its simple conception. In mathematical terms, if $\boldsymbol{y}$ is a realisation of a bivariate random variable, $\boldsymbol{Y}$, on the plane $\mathbb{R}^2$, and $\Pr\{\boldsymbol{Y} = 0\} = 0$, then its radial projection $||\boldsymbol{y}||^{-1}\boldsymbol{y}$ is a random vector on the unit circle which can be converted to a vector of random angles relative to some direction treated as zero. In the PN distribution $\boldsymbol{y}$ is a realization of a bivariate normal variable $\mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ on the plane. This distribution is discussed by Mardia and Jupp (2000) and a clear and succinct summary is provided by Wang and Gelfand (2013). If $\boldsymbol{\mu} = 0$ and $\boldsymbol{\Sigma} \propto \mathbf{I}$ then the PN distribution is equivalent to a uniform distribution on the circle. Allowing $\boldsymbol{\mu} \neq 0$ gives rise to a non-uniform but unimodal and symmetrical distribution on the circle, and further generalisation so that $\boldsymbol{\Sigma}$ is any valid covariance matrix gives rise to a flexible distribution on the circle which can be asymmetrical and bimodal. Wang and Gelfand (2013) explore the PN distribution in a Bayesian setting, including regression models in which the parameter $\boldsymbol{\mu}$ is modelled as a linear function of covariates.

The MVM and PN distributions are flexible models for complex distributions of circular data. We are not aware of any examples of their use in structural geology, although they would clearly be suitable in circumstances where simpler symmetrical and unimodal distributions would not be appropriate. Some practical questions remain for their application. First, how many components of the MVM model are justified for a particular data set? Second, given that MVM for $g > 2$ has more parameters than the PN model, how can one decide when the more complex model is justified?

In this paper we demonstrate the use of the mixture of the von Mises and PN distributions as a model for three sets of circular data from structural geology. We address the question of how to select the number of components, g, of the MVM for a particular data set, and the decision whether the MVM or PN model is preferable in a particular case. We use maximum likelihood to estimate parameters of the PN distribution, and show how to test a hypothesis of the uniformity of the PN model over two data sets.

2   Methods

2.1   Selection of a mixture of von Mises distributions

The mixture of von Mises distributions, as used here, comprises g different von Mises distributions, each with an associated mixture weight $\alpha_i, i = 1, 2, \ldots, g$ which is the probability that an event is drawn from the ith distribution. The parameters of the MVM distribution, for specified g, can be estimated by maximum likelihood estimation (MLE). The MVM distribution is identifiable (Holzmann et al., 2004), but a numerical approximation to the likelihood function is needed because the MLE of the κ parameter includes a ratio of Bessel functions. In this study we used the movMF procedure from the package of that name developed by Hornik and Grün (2013) for the R platform (R development core team, 2013). This finds the MLE by an expectation maximisation algorithm (e.g. Banerjee et al., 2005).

We are interested here in how many components to specify in the MVM model for a data set. The general problem is whether the improvement in the likelihood that we obtain by fitting $g + 1$ rather than g components with an additional three parameters is justified. In general, two nested models, where the simpler (null) model is a particular case of the more general one with parameters fixed to some value, may be compared in their log-likelihood ratio, $\ell_A - \ell_N$, where ℓ_A and ℓ_N are the maximised log likelihoods obtained for the more complex and null models respectively. It is necessary, for inference, to know the distribution of the ratio when the null model holds. In regular cases, asymptotically, $L = 2(\ell_A - \ell_N)$ is distributed as χ^2 with degrees of freedom equal to the number of additional parameters in the more complex model. However, mixture models are not regular because the null model is at the boundary of the parameter space for the alternative. The distribution of L can be counterintuitive at the boundary of the parameter space; see for example Clifford (2006). A solution to this problem has been proposed for the comparison of the two-component MVM distribution with a single von Mises distribution, in the case where the κ parameter is common to both components of the MVM (Fu et al., 2008). This does not address the general problem of comparing MVM models for increasing values of g. For this reason we computed the distribution of $l_{g,g+1} = \ell_{g+1} - \ell_g$ for any comparison between an MVM

with g components (log likelihood ℓ_g) and one with $g+1$ components (log likelihood ℓ_{g+1}) by Monte Carlo simulation. This approach has been used previously to identify the number of components in a mixture model, e.g. Aitkin et al. (1981). The distribution of the log-likelihood ratio $l_{g,g+1}$ under some model with g components may depend on the parameters of the components and how well separated they are on the circle. For this reason we accounted for uncertainty in the estimated parameters of the null model by a bootstrapping step. A bootstrap sample from the data was drawn and an MVM model with g components was fitted. A single parametric bootstrap sample from the fitted model was then drawn and the log-likelihood ratio for MVM models with g and $g+1$ components was calculated. This step was repeated to generate the full Monte Carlo sample of the log-likelihood ratio. We are not aware of a comparable combination of a bootstrap sample from the data with a parametric bootstrap, but note that it is comparable, although not identical, to the "double bootstrapping" procedure of Mclachlan and Peel (1997). The full procedure is described below.

We fitted MVM distributions with g and $g+1$ components to our n data, using the `movMF` procedure. We computed the log of the ratio of likelihoods for the two distributions, which we denote by $\hat{l}_{g,g+1}$. To obtain a distribution for this ratio under the null model we undertook the following steps.

1. A bootstrap sample was drawn from the data and the parameters of the MVM model with g components were estimated with the `movMF` procedure.

2. The `rmovMF` procedure from the `movMF` library was used to generate a random sample from the MVM distribution (g components) with the parameters estimated from the bootstrap sample.

3. We then fitted the MVM distribution with g and $g+1$ components to the simulated values, and computed $\hat{l}_{g,g+1}$.

4. Steps 1–3 above were repeated 1000 times in total.

The proportion of values of the log-likelihood ratio in the resulting sample larger than $\hat{l}_{g,g+1}$ was computed, $\hat{P}_g$. If $\hat{P}_g \geq 0.05$ then the null model (g components) was accepted. This procedure was followed starting with $g=1$ and testing an alternative distribution with two components. If the null model was rejected then a distribution with three components was compared to the distribution with two and so on until the null model for some g was accepted.

Note that the problem of the number of components in an MVM distribution can be addressed by comparing alternative models on information criteria such as Akaike's information criterion which we use for other purposes below (e.g. Mclachlan and Peel, 2000). However, information criteria are an informal basis for model comparison and, where possible, we prefer to use a statistic such as the log-likelihood

ratio, which allows a formal hypothesis test. Note that by using this bootstrapping procedure we also avoid testing the log-likelihood ratio against an asymptotic distribution by using an empirical distribution for the same sample size as our data.

2.2 Maximum likelihood estimation of the projected normal distribution and comparison with the selected mixture of von Mises distributions

A PN distribution of angles can be specified by the parameters of the bivariate normal distribution whose radial projection corresponds to points on the unit circle. To give a unique parameterisation for any distribution on the circle we followed Wang and Gelfand (2013) by fixing the variance of one of the normal variables to 1.0. There are four parameters to be estimated: the two elements of the mean vector, $\boldsymbol{\mu}$, the correlation coefficient, ρ, which takes values in the interval $(-1, 1)$ and the standard deviation of the second normal variable, τ, which takes values $\tau > 0$. These were estimated by maximum likelihood, using the PN density given by Wang and Gelfand (2013):

$$f(\theta|\boldsymbol{\mu}, \rho, \tau) = C(\boldsymbol{\eta})^{-1}\Big[\phi_2\{\mu_1, \mu_2|\mathbf{0}, \boldsymbol{\Sigma}\} +$$
$$aD(\boldsymbol{\eta})\Phi_1\{D(\boldsymbol{\eta})\}\phi_1\Big\{a(\mu_1\sin\theta - \mu_2\cos\theta)/\sqrt{C(\boldsymbol{\eta})}\Big\}\Big], \quad (1)$$

where θ is an angle, obtained from the radial projection of the bivariate normal distribution by the arctan∗ function defined by Wang and Gelfand (2013), $\boldsymbol{\eta}$ is the vector of values $[\theta, \mu_1, \mu_2, \rho, \tau]^{\mathrm{T}}$, $\phi_2(x_1, x_2|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the bivariate normal probability density function (pdf) with specified parameters, $\phi_1(x)$ is the standard normal pdf and $\Phi_1(x)$ is the standard normal distribution function,

$$\boldsymbol{\Sigma} = \left[\begin{array}{cc} \tau^2 & \tau\rho \\ \tau\rho & 1 \end{array}\right],$$

$$a = \left(\tau\sqrt{1-\rho^2}\right)^{-1},$$

$$C(\boldsymbol{\eta}) = a^2\left(\cos^2\theta - \rho\tau\sin 2\theta + \tau^2\sin^2\theta\right)$$

$$D(\boldsymbol{\eta}) = \frac{a^2\{\mu_1(\cos\theta - \rho\tau\sin\theta) + \mu_2\tau(\tau\sin\theta - \rho\cos\theta)\}}{C(\theta)^{-\frac{1}{2}}}.$$

The maximum likelihood estimates $\hat{\boldsymbol{\mu}}$, $\hat{\rho}$ and $\hat{\tau}$ maximise the likelihood function $\mathcal{L}(\boldsymbol{\mu}, \rho, \tau|\theta) = f(\theta|\boldsymbol{\mu}, \rho, \tau)$, given the data in θ. The likelihood was maximised with the `optim` procedure in R.

For $g \geq 2$ the MVM distribution has more parameters than the PN. Having fitted both to a data set, the question remains whether the greater complexity of the MVM is justified by its goodness of fit. One common approach to selecting between models of differing complexity, where the models are not nested and so cannot be formally compared in the likelihood ratio, is to use information criteria which combine the

maximised log likelihood for each fitted model with a term that penalises models for the number of parameters that must be estimated. One such criterion is Akaike's information criterion (AIC) due to Akaike (1973):

$$A = -2\ell + 2N_p,\tag{2}$$

where ℓ is the maximised (full) log likelihood for a model with N_p parameters. The model with the smallest A is selected, so effectively the selection is based on the likelihood with a penalty for the model complexity as measured by N_p. Another information criterion is the Bayes information criterion (BIC) (Kass and Raftery, 2006):

$$B = -2\ell + N_p\{\log(2\pi n)\},\tag{3}$$

where n is the number of observations. In both cases one selects the model for which the information criterion is smallest. The BIC penalises extra parameters more heavily than does AIC unless n is small. More fundamentally, the AIC selects a model which appears to be closest to the underlying but unknown model which generates the data, whereas the BIC selects a model with a maximised posterior probability (Wit et al., 2012). The AIC is a basis for a pragmatic choice of model which seems to explain the data and offer a sound basis for prediction, whereas the BIC aims to identify the "true" model (Spiegelhalter et al., 2014). The two criteria are therefore not directly commensurate, and the question of which is "best" depends on the principles and purposes of model selection (Wit et al., 2012). A detailed discussion of the two criteria is out of the scope of the present paper, so for our present purposes we present results for both.

2.3 Modelling variations in the mean vector of the projected normal distribution

Wang and Gelfand (2013) showed that the PN distribution for circular observations can be fitted in the form of a linear model in which the elements of the mean vector $\boldsymbol{\mu}$ are expressed as linear functions of predictors. The parameters τ and ρ remain fixed, but considerable flexibility in the form of the circular distribution is still possible. The predictors could be continuous variables (e.g. coordinates to model a spatial trend) or categorical variables (e.g. stratigraphic units), or a combination of both. In this paper we consider a case where the predictor is a categorical variable with $P = 2$ levels; the observations are circular data that describe the orientation either of anticlinal axial planes or of Landsat-derived lineaments. If there are P levels of the categorical predictor variable then the modelled values of the mean vector for n observations are

$$\boldsymbol{\mu} = \mathbf{MX},\tag{4}$$

where $\mathbf{X}$ is a $P \times n$ design matrix with element $\mathbf{X}\{i, j, \}$ set to 1 if the ith observation is in the jth level of the categorical predictor and set to 0 otherwise and

$$\mathbf{M} = \begin{bmatrix} \mu_{1,1} & \mu_{1,2} & \dots & \mu_{1,P} \\ \mu_{2,1} & \mu_{2,2} & \dots & \mu_{2,P} \end{bmatrix}$$

where $\mu_{1,j}$ is the modelled value of μ_1 for an observation corresponding to the jth level of the categorical predictor. The modelled values of μ_2 are defined similarly to the elements of the second row of $\mathbf{M}$. The elements of $\mathbf{M}$ may be estimated by maximum likelihood by substituting Eq. (4) for $\boldsymbol{\mu}$ in the likelihood function.

3 Case studies

3.1 West Cumbria dip directions

3.1.1 The data

The first two data sets consist of the observations of dip direction of the bedding planes of two sedimentary units in the British Geological Survey's map sheets at 1 : 50 000 in West Cumbria, northwest England. The units are the Sherwood Sandstone Group (Triassic sandstone) and the Windermere Supergroup (Ordovician–Silurian mudstone, sandstone and limestone). A total of 90 observations were available for the Sherwood Sandstone Group and 572 for the Windermere Supergroup. The data are shown in Fig. 1a and b by rose diagrams which show the relative frequency of observations in bins of width $\pi/10$ radians. Note that this binning was done only for the presentation of the data as rose diagrams; all statistical analysis was done on the raw circular observations.

3.1.2 Mixture of von Mises distribution and comparison with projected normal distribution

Table 1 shows the results from comparison of MVM distributions with increasing numbers of components. In the case of the Sherwood Sandstone Group there was evidence to select the MVM distribution with two components over a single von Mises distribution, but not to reject the model with two components in favour of one with three. In the case of the Windermere Supergroup a mixture model with five components was selected. Table 2 shows the fitted parameters for the MVM models, the mean direction for each component is in degrees clockwise from north, and κ is a dispersion parameter: the component has a narrow distribution if this is large.

Table 3 shows the results from the comparison of the selected MVM model with a PN model for both the Sherwood Sandstone Group and the Windermere Supergroup dip directions, and Fig. 2a and b shows the probability densities for these two distributions wrapped around the circle. In both cases the AIC and the BIC were both smaller for the more complex MVM distribution.
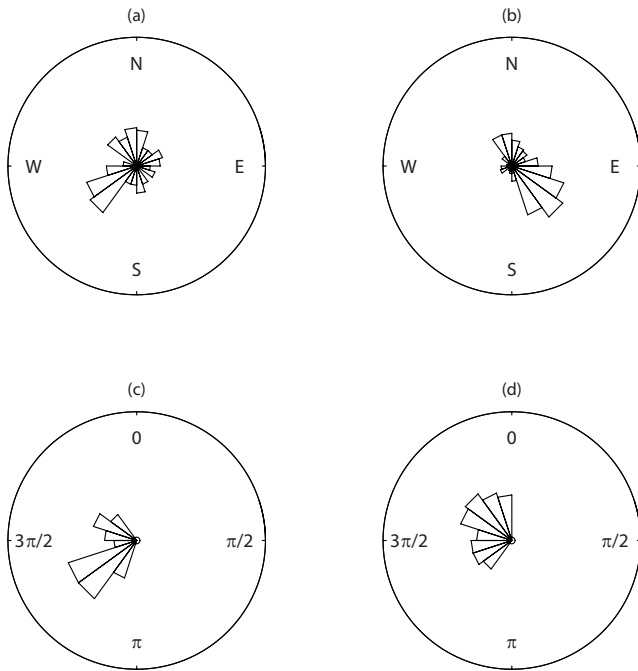
Figure 1. Data: direction of dip for **(a)** the Sherwood Sandstone group (90 data) and **(b)** the Windermere Supergroup (572 data); orientations (doubled from the original range of $[0, \pi)$) for **(c)** Bangladesh anticline axial planes (32 data) and **(d)** Bangladesh Landsat-derived lineaments (40 data). Note that segments of the rose diagrams are proportional to relative frequency within each data set separately, so are not comparable between data sets with respect to numbers of observations. If all data appeared within a single bin of the rose diagram then the corresponding segment would be equal in length to the radius of the circle.

3.1.3 Key findings

In the case of the Sherwood Sandstone Group the clear difference between the PN distribution and the MVM is that in the latter there is a stronger contrast between the tightly distributed subset of dip directions towards the southwest (mean direction is 4.07 radians or 233 degrees clockwise from north, $\kappa = 38$) and a second subset with a mean direction close to north (6.03 radians or 346 degrees clockwise from north) and a wider dispersion ($\kappa = 0.6$). Although the PN distribution is bimodal, the contrast between the two modes is less pronounced, and the MVM distribution, which is selected on statistical criteria, better captures this heterogeneity in the data. Geologically the southwest dips are found in the west of the study area in a relatively small area where there is a relatively simple consistent structure and good exposure of the geology. The approximately northern dips are found mainly in northern Cumbria, and their dispersion may reflect the fact both that they are spread out along the crop of the Sherwood Sandstone Group, encompassing greater structural variability, and that they are subject to greater observation error because the dips are smaller.
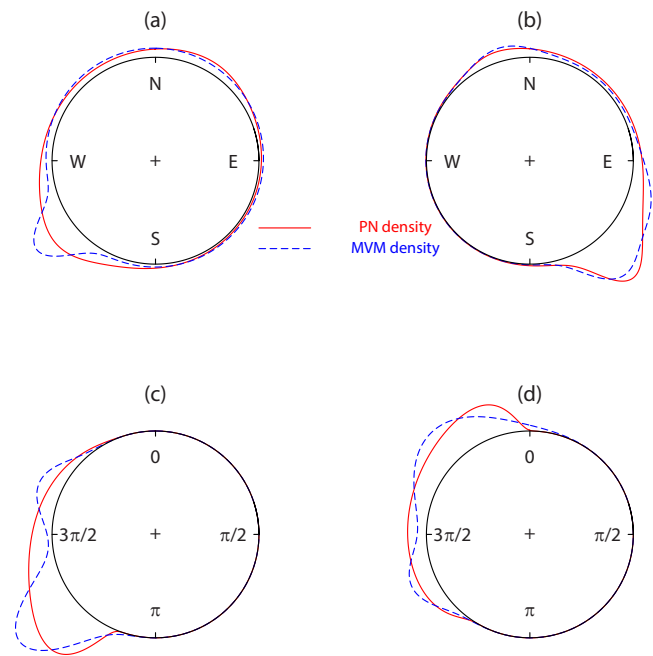
Figure 2. Density of fitted projected normal distributions (PN) and a mixture of von Mises distributions (MVM) wrapped around the circle: direction of dip for **(a)** the Sherwood Sandstone group and **(b)** the Windermere Supergroup; orientations (doubled from the original range of $[0, \pi)$) for **(c)** Bangladesh anticline axial planes and **(d)** Bangladesh Landsat-derived lineaments.

The greater complexity of the selected MVM model for the Windermere Supergroup may partly reflect the fact that there are more data available to support a complex model, but the fitted model also makes geological sense. In the study area the Windermere Supergroup is subject to cylindrical folding which would be expected to give rise to dips of approximately equal frequency in a northwest and southeast direction. Both fitted distributions as illustrated in Fig. 2b show modes in these directions, but the dominant mode is in a southeast direction (see the fourth component for the Windermere Supergroup in Table 2 with a mean direction of 141 degrees and a probability of occurrence of 0.35). The structure of the Windermere Supergroup in the study areas is complicated by a major fault trending northeast–southwest. This may introduce different dominant dip directions locally, which may explain the rather more complex form of the selected MVM distribution. In particular, some of the folding in the vicinity of the fault is overturned, so that both limbs of the fold dip to the southeast. This accounts for the asymmetry between the two dominant modes of both fitted distributions. The MVM model, giving three distinct modes as seen in Fig. 2b, in addition to the broader distribution of dip directions over the interval from due south to northwest, captures this complexity better than the PN, which shows two nearly antipodal modes.

Table 1. Selection of a mixture of von Mises (MVM) distributions.

Number of components			
g (null) model	$g+1$ (proposed) model	$\hat{l}_{g,g+1}$	$\hat{P}_g$
Sherwood Sandstone Group			
1	2	19.8	< 0.001
2	3	2.0	0.273
Windermere Supergroup			
1	2	196.3	< 0.001
2	3	13.2	0.001
3	4	18.2	0.024
4	5	17.1	0.031
5	6	11.6	0.064
Bangladesh anticline axial planes			
1	2	10.3	0.003
2	3	0.1	0.823
Bangladesh Landsat-derived lineaments			
1	2	5.8	0.022
2	3	2.0	0.263

Table 2. Parameters of a fitted mixture of von Mises (MVM) distributions. The ith component occurs with probability α_i and has mean μ_i and dispersion parameter κ_i. Values for direction of dip are degrees clockwise from north, and values for the Bangladesh anticline axial planes and Landsat-derived lineaments are doubled orientation angles.

i (component number)	μ_i	κ_i	α_i
Sherwood Sandstone Group direction of dip			
1	346	0.6	0.68
2	233	38.0	0.32
Windermere Supergroup direction of dip			
1	112	11.2	0.12
2	47	1.04	0.18
3	345	16.6	0.14
4	141	25.2	0.35
5	116	11.9	0.21
Bangladesh anticline axial plane			
1	231	21.0	0.74
2	294	19.8	0.26
Bangladesh Landsat-derived lineament			
1	316	6.7	0.73
2	246	20.5	0.27

3.2 Bangladesh anticlinal axial planes and Landsat-derived lineaments

3.2.1 The data

These two data sets, from eastern Bangladesh, are presented by Davis and Sampson (2002). Unlike the data in the previous section these data are orientations without a preferred direction. As described by Davis and Sampson (2002), all orientation data can be expressed by pairs of values: m and $m + \pi$ radians representing the bearing for each end of a linear feature such as a fault or plane. Following Krumbein (1939), these values can be doubled to give a single value $2m$ (since $2m + 2\pi = 2m$ for values on the circle). The doubled orientations are therefore distributed over the whole circle, and can be analysed with methods appropriate for circular data. All analyses reported here are for these doubled angles.

The first set of data comprises 32 observations of the orientation of the axial plane of a series of anticlines. The second set comprises orientations of 40 major lineaments identified by interpretation of Landsat imagery. Davis and Sampson (2002) compare the orientations in the two data sets by an analysis assuming that each has an underlying von Mises distribution. In this paper we compare the MVM and PN distributions for the Landsat data set, and then evaluate evidence that the two sets of orientations are different by fitting PN distributions. As described above the orientation data were doubled so that they are distributed on the circle. Figure 1c and d show the doubled orientation data as rose diagrams.

3.2.2 Mixture of von Mises distribution and comparison with projected normal distribution

As shown in Table 1, both these variables are better fitted by an MVM with two components than by a single von Mises distribution, but adding a third component is not justified. As shown in Table 3, the AIC was smallest for the MVM model in the case of the anticline planes, but for the PN model in the case of the Landsat-derived lineaments; in this latter case the likelihood for the PN distribution was larger than for the more complex MVM distribution. Note that the same inference is supported by the BIC. Figure 2d shows how the asymmetric and bimodal distribution seen in the rose diagram (Fig. 1d) is fitted by the two models. The MVM represents the complex variation of the data with two distinct components, the PN distribution is also asymmetric and bimodal, but note in particular that the MVM has a tail giving non-zero density between zero and about $\pi/4$ radians where there are no observations. The Landsat lineaments data illustrate the flexibility of the PN distribution, which in this case fits the data better than the MVM distribution but with fewer parameters.

3.2.3 Projected normal distributions for the combined Bangladesh data sets

We considered the two data sets on orientations as a combined data set, allowing us to examine the evidence that the

Table 3. Comparison of a mixture of von Mises (MVM) with g components and projected normal (PN) distributions.

	MVM					PN			
	g	N_p	ℓ	AIC	BIC	N_p	ℓ	AIC	BIC
Sherwood Sandstone Group	2	5	−137.39	284.78	288.54	4	−145.33	298.66	301.67
Windermere Supergroup	5	14	−614.88	1257.76	1279.54	4	−635.38	1278.76	1284.98
Bangladesh anticline axial planes	2	5	−14.86	39.72	41.24	4	−18.3	44.6	45.81
Bangladesh Landsat-derived lineaments	2	5	−34.84	79.68	81.68	4	−33.67	75.34	76.94

two variables correspond to the same structural features. Alternative models were fitted to the combined data set, and these are detailed in Table 4 along with the maximised log likelihood. Figure 3 shows the density functions wrapped around the circle for all three models.

Model 1 is a PN distribution with single values for all parameters, i.e. all parameters pooled for both the anticlines and the Landsat lineaments. Model 2 has separate PN distributions for the two orientations, but with common values of the variance parameters τ and ρ. This corresponds to Eq. (4), with "observation type" the categorical predictor variable with two levels: anticlinal axial plane and Landsat-derived lineament. The model therefore provides different values of the mean vector $\boldsymbol{\mu}$ for the orientation of the anticline planes and the Landsat-derived lineaments.

A PN distribution was fitted to each data set separately, as described in the previous section, for comparison with an alternative MVM distribution. These distributions considered together may be treated as a model for the combined data set with all parameters of the PN distribution differing between the anticlines and Landsat lineaments. This is denoted as Model 3 in Table 4 and Fig. 3. The log likelihood for this model is the sum of the two log likelihoods for the separate fittings. Note that the parameter ρ approaches the boundary value 1 for both the anticline planes and the Landsat lineaments. We checked that the maximum likelihood estimation was reliable by computing the profile likelihood for this parameter, and found that the likelihood increased smoothly as ρ approached 1.

This sequence of models can be regarded as nested: Models 1 and 2 are particular cases of Model 3 with certain parameters set to common values. A simpler model may be compared with a more complex one by computing the L statistic, twice the log-likelihood ratio. Under the null model, the simpler one nested in the alternative, the asymptotic distribution of this statistic is χ^2 with degrees of freedom equal to the difference in the number of independent parameters. However, in this case the distribution of L is not a simple χ^2 for any comparison with Model 3, because ρ is at the boundary (in principle one might obtain a sample distribution for L under these circumstances by bootstrapping, but this is a topic for further research). In Table 4 we present the result for a comparison of Model 1 with Model 2. Note that
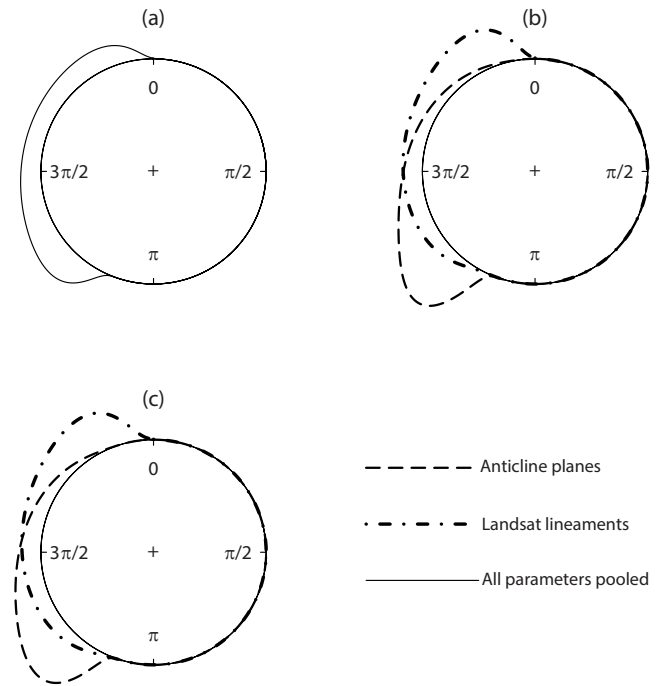
Figure 3. Density of projected normal distributions fitted to combined Bangladesh data and wrapped around the circle. **(a)** Model 1, with all parameters pooled for the combined data set; **(b)** Model 2, with $\boldsymbol{\mu}$ modelled separately for anticline axial planes and Landsat lineaments; **(c)** Model 3, with all parameters separate for anticline axial planes and Landsat lineaments.

we can reject Model 1 in favour of Model 2 with different parameters $\boldsymbol{\mu}$ for the anticlines and the Landsat lineaments, so we can conclude that there is reason to believe that the orientations of the Landsat lineaments differ from those of the anticline axial planes. This is consistent with the conclusion of Davis and Sampson (2002), but our analysis treats the complex distribution of the data more plausibly.

In Model 3, ρ approaches the boundary at $\rho = 1$. The density function given by Eq. (1) is undefined at this boundary. When $\rho = 1$ and $\tau \mu_2 \neq \mu_1$ the PN distribution, which we denote by PN_1, is continuous with support over half the circle.

Table 4. Projected normal distributions fitted to Bangladesh anticline axial plane and Landsat-derived lineament orientations (doubled directions). Subscript "A" for a parameter denotes that it pertains only to observations of anticline axial plane orientations, subscript "L" denotes a parameter of the distribution of Landsat lineaments, and subscript "P" denotes a pooled parameter.

Model	Parameters								ℓ
Model 1	$\mu_{1,P}$		$\mu_{2,P}$		τ_P		ρ_P		
	0.05		−3.49		3.94		0.83		−67.5
Model 2	$\mu_{1,A}$	$\mu_{1,L}$	$\mu_{2,A}$	$\mu_{2,L}$	τ_P		ρ_P		
	−2.94	2.34	−4.86	−3.39	3.66		0.72		−52.9
Model 3	$\mu_{1,A}$	$\mu_{1,L}$	$\mu_{2,A}$	$\mu_{2,L}$	τ_A $\quad$ τ_L		ρ_A	ρ_L	
	−9.30	2.25	−16.56	−3.5	12.07 $\quad$ 3.76		0.999*	0.999*	−52.0

Comparison between models				
Null	Alternative	L	P	χ^2 degrees of freedom
Model 1	Model 2	29.2	4.6×10^{-7}	2

* These estimates of ρ are at the boundary of the parameter space.

The density of the PN_1 distribution is

$$f_1(\theta|\boldsymbol{\mu},\tau) = \frac{|c| \times \phi_1 \left(c \times (\tau + \cot(\theta - \theta_0)) - \mu_2 \right)}{\sin(\theta - \theta_0)^2} \quad (5)$$

for values of θ between $\theta_0 = \arctan(1/\tau)$ and $\theta_0 + \pi$ on the side of the circle that faces $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and 0 on the side that faces away from $\boldsymbol{\mu}$. The constant c that appears in Eq. (5) is defined as $c = \frac{\tau\mu_2 - \mu_1}{(\tau^2 + 1)}$ and is guaranteed to be non-zero in this case. When $\rho = 1$ and $\tau\mu_2 = \mu_1$ (i.e. when $c = 0$), the PN distribution is discrete, with probabilities of $\Phi_1(-\mu_2)$ and $1 - \Phi_1(-\mu_2)$ in directions θ_0 and $\theta_0 + \pi$ respectively. Similar results hold at the boundary $\rho = -1$.

The PN_1 distribution might be useful because it is more parsimonious than the general PN distribution while retaining flexibility for both unimodal and bimodal distributions. However, as seen above, the support of PN_1 is restricted to an interval of width π and so it does not include the uniform distribution on the circle as a special case, which limits its usefulness for inference.

3.2.4 Key findings

These data were analysed previously, by Davis and Sampson (2002), who assumed a simple von Mises distribution. Our results suggest that this distribution is not appropriate for these data; in both cases there was evidence that a mixture of two von Mises distributions was more suitable than a single von Mises distribution, and in the case of the Landsat-derived lineaments, the projected normal distribution was favoured over the MVM.

Using the PN distribution it was shown that there was a significant difference between the orientations of the anticline axial planes and the Landsat-derived lineaments from eastern Bangladesh. Davis and Sampson (2002) drew the same conclusion, but in an analysis which made the implausible assumption that the distribution of (doubled) orientation angles within the two subsets was a simple von Mises. Our results here support the conclusion that the orientations of the Landsat-derived lineaments have a signficantly different distribution from the those of the anticline axial planes, and that the two variables cannot therefore be regarded as samples from the same population. This suggests that the Landsat interpretation identifies features which are not all aligned with anticlinal axial planes.

4 Conclusions

The two case studies reported above show that both the projected normal (PN) distribution and a mixture of von Mises (MVM) distributions can be used to model variations of circular data in the earth sciences which may be distributed in a complex multimodal and asymmetric way.

It is possible to select the number of components in an MVM distribution by a sequential testing procedure in which the log-likelihood ratio is used as a test statistic to decide whether to add a component to the model, with the distribution of the statistic under the simpler model computed by a Monte Carlo simulation. One may select the number of components for an MVM distribution and compare the result with the PN distribution, which is a more parsimonious model, by computing the Akaike or Bayes information criteria for the two distributions. In our case studies the two information criteria, which are not directly comparable, led to the same choice of model.

In the West Cumbria example, and the case of the anticlinal axial planes from Bangladesh, the MVM model was favoured over the PN, indicating that the complexity of the

circular data in these cases was such that, despite the flexibility of the PN model, and its parsimony, the more complex mixture model was required. In contrast, although the complexity of the orientation of the Landsat-derived lineaments in Bangladesh meant that these data were too complex to be described by a single von Mises distribution, the PN distribution was sufficiently flexible to capture the distribution of the doubled orientation data, and was favoured by information criteria over more complex MVM distributions.

We have shown how the PN distribution can be used in a model with covariates (in this case a categorical variable which distinguishes the anticline axial planes from Landsat-derived lineaments for the structural orientation data from Bangladesh). In this case we were able to show that the two sets of orientations are different, indicating that the two sets of features are different. This approach could be of general interest in earth sciences, for example to identify spatial trends or differences between geological units in the orientation of borehole break-out data. In principle a similar extension of the MVM model in which the mixture weights α_i depend on a set of covariates might be developed. However, such a model would be complex and might not be identifiable. While this could be an interesting topic for further work, the extension of the PN model is more straightforward.

We have identified a more parsimonious form of the PN distribution, denoted PN_1, which is distributed on the half-circle. The properties of this distribution and its practical usefulness are topics for further research.

References

Aitkin, M., Anderson, D., and Hinde, J.: Statistical modelling of data on teaching styles, J. Roy. Stat. Soc. A , 144, 419–461, 1981.

Akaike, H.: Information theory and an extension of the maximum likelihood principle, in: Second International Symposium on Information Theory, edited by: Petov, B. N. and Csaki, F., Akademia Kiado, Budapest, 267–281, 1973.

Banerjee, A., Dhillon, I. S., Ghosh, J., and Sra, S.: Clustering on the unit hypersphere using von Mises-Fisher distributions, J. Mach. Learn. Res., 6, 1345–1382, 2005.

Clifford, D.: Distribution of increases in residual log likelihood for nested spatial models, Commun. Stat. Simulat., 35, 779–788, 2006.

Coblentz, D. D. and Richardson, R. M.: Statistical trends in the intraplate stress field, J. Geophys. Res., 100, 245–255, 1995.

Davis, J. C. and Sampson, R. J.: Statistics and Data Analysis in Geology, John Wiley & Sons, New York, 2002.

Fu, Y., Chen, J., and Li, P.: Modified likelihood ratio test for homogeneity in a mixture of von Mises distributions, J. Stat. Plan. Infer., 138, 667–681, 2008.

Holzmann, H., Munk, A., and Stratmann, B.: Identifiability of finite mixtures – with applications to circular distributions, Sankhya, 66, 440–449, 2004.

Hornik, K., Grün, B., and the R Development Core Team: Package movM F, R package version 0.1–2, 2013.

Jones, T. A.: MATLAB functions to analyze directional (azimuthal) data – I: Single-sample inference, Comput. Geosci., 32, 166–175, 2006.

Kass, R. E. and Raftery, A. E.: Bayes factors, J. Am. Statist. Assoc., 90, 773–795, 1995.

Krumbein, W. C.: Preferred orientation of pebbles in sedimentary deposits, J. Geol., 47, 673–706, 1939.

Mardia, K. V. and Jupp, P. E.: Directional Statistics, John Wiley & Sons, Chichester, 2000.

McLachlan G. J. and Peel, D.: On a resampling approach to choosing the number of components in normal mixture models, in: Computing Science and Statistics, Vol. 28, edited by: Billard, L., Fisher, N. I., Interface Foundation of North America, Fairfax Station, VA, 260–266, 1997.

McLachlan G. J. and Peel, D.: Finite Mixture Models, John Wiley & Sons, New York, 2000.

O'Brien, J. F., Roman, D. C., Dixon, J. P., Power, J. A., and Arnold, R.: Multiple causes for non-eruptive seismic swarms at Mt. Martin, Katmai Volcanic Cluster, Alaska (2004–2008), J. Volcanol. Geoth. Res., 229, 13–22, 2012.

R Development Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3–900051-07–0, available at: http://www. R-project.org/, 2013.

Sen, K. and Mamtani, M. A.: Magnetic fabric, shape preferred orientation and regional strain in granitic rocks, J. Struct. Geol., 28, 1870–1882, 2006.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A.: The deviance information criterion: 12 years on, Journal of the Royal Statistical Society Series B., 76, 485–493, 2014.

Wang, F. and Gelfand, A. E.: Directional data analysis under the general projected normal distribution, Statistical Methodology, 10, 113–127, 2013.

Wit, E., van den Heuvel, E., Romeijn, J-W.: "All models are wrong…": an introduction to model uncertainty, Statistica Neerlandica, 66, 217–236, 2012.

Witts, D., Hall, R., Nichols, G., and Morley, R.: A new depositional and provenance model for the Tanjung formation, Barito Basin, SE Kalimantan, Indonesia, J. Asian Earth Sci., 56, 77–104, 2012.