Solid Earth Discuss., 5, C1049–C1051, 2014
www.solid-earth-discuss.net/5/C1049/2014/

Solid Earth
Discussions
Open Access

Interactive comment on "Modelling complex geological angular data with the Projected Normal distribution and mixtures of von Mises distributions" *by* R. M. Lark et al.

Anonymous Referee #1

Received and published: 21 February 2014

Report on "Modelling complex geological angular data with the Projected normal distributions and mixtures of von Mises distributions"

R. M. Lark, D. Clifford, C. N. Waters

In this paper the authors promote the use of two statistical distributions appropriate for complex angular datasets in structural geology.
One is the mixture of von Mises distribution (MVM), the other one is the projected normal (PN) distribution.

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

These two distributions are flexible enough to model complex angular data with multimodal and asymmetrical distributions, contrarily to the widely used simple von Mises distribution.
The authors present the two distributions and give some practical tools for selecting the number of components for the MVM, or determining which of the MVM or PN is the most suitable model for the data.
Their use is demonstrated on four sets of angular data from structural geology.

A parametric bootstrap procedure is implemented to compute the log ratio of likelihoods for determining the right number of components in the MVM, similarly at what is done by Maclachlan *et al*.
Nevertheless in the normal mixture case the BIC criterion is usually used to select the number of components. This is not mentioned, the authors should explain why they did not investigate this criterion, that would be less computational than the bootstrap procedure.

The PN model is estimated by a standard maximum likelihood procedure, it is emphasized that covariates may be taken into account to model the mean and that categorical variables will be considered. This is not a specificity of the PN distribution, it is likely also possible to model the MVM mean with covariates (maybe increasing too much the number of parameters), and I did not see in the applications where categorical variables are considered.

The comparison of MVM and PN is addressed by computing the classical AIC criterion. This criterion is known to favour distribution with numerous parameters, it would be interesting to consider other penalized criteria such as BIC and have a discussion about this choice.

For the two first datasets the MVM model with two components in one case and five in the other one has a smaller AIC than the PN model, and that demonstrates that the MVM enables modelling more complex data than the PN distribution. It would have

SED
5, C1049–C1051, 2014

been interesting to have a table with the parameter estimates, and a discussion on the general lack of ability of the PN distribution to model such processes.

The two following datasets deal with orientation data, it is said that they are doubled so that they are distributed on the circle. I understand that symmetric data have been generated but this does seem to be true as the fitted distributions have their support in half circle. Please precise how you double the data. In order to examine whether the two datasets are similarly distributed, three nested PN models are considered and a likelihood ratio test is performed. Could a similar procedure could be done with the MVM distribution?
For the most general PN model one of the parameter is at the boundary and the likelihood ratio test could not be performed. This is well explained, but no solution is given to bypass this difficulty. Could a procedure like the parametric bootstrap for selecting the number of components of the MVM could apply in this case?

Missprints

page 2189, line 6 $X(i, j)$

Interactive comment on Solid Earth Discuss., 5, 2181, 2013.