Solid Earth
Open Access

Discussions

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.

F. Salvini (Referee)

salvini@uniroma3.it

Received and published: 14 April 2014

Dear Editor, the paper: Modelling Complex geological angular data with the Projected Normal distribution and mixtures of Von Mises distributions by R.M.Lark, D. Clifford, C.N. Waters deals with the statistical interpretation of polymodal data distributions on angular data. The interpretation is based on the comparison among multiple Von Mises and Projected Normal distributions. The proposed methodology succeeds in identifying the independent populations, Von Mises like, present in the dataset by comparing the meaning of the bettering of the approximation by increasing the number of independent distributions and with respect to the Projected Normal one. Proposed results

show the efficiency of the proposed method by analyzing sets of data representing angular distributions that were successfully compared to the two proposed approaches. The proposed methodology seems of interest in the general analysis of angular data. Yet the lack of presentation of results (e.g. which are the found mean angle for each distribution?) prevents from a final judgment on what proposed. In its present form, the paper looks like a theoretical statistical debate and I am afraid it would be of little interest to Earth Science geologists. To solve this, I suggest to add the presentation of the results (not just the mere statistical fit/reliability) associated to the statistical analysis. I am sure that most scientists will judge on the reliability of improving their angular data analysis approach with the proposed method if they can really compare distributions results with data. This difficulty is even improved by the poor azimuthal resolution of the data. The used approximation (18°!) is really lower than most structural geology study on, say, fracture azimuths. The Authors should improve this resolution. I personally use 2° resolution for angular data, and many colleagues use a 10° resolution. Of course this point is useless in unimodal distribution, but it is an important influencing parameter in dealing with polymodal data distribution as the authors are presenting. Two close populations with means that differ of less than the adopted resolution will tend to mix and provide as a result a single distribution with wrong deviations. Another important point of the paper were the authors should take a larger care is on the method they propose to establish the number of Von Mises distribution of data. The proposed method might be the correct one (say low probability of fail) for meaningful samples of the analysed population. This assumption must be proved prior to define the number of distributions. As an example, if you have very few data they will scatter on the frequency histograms with a high noise (due to the not significant number of data) and the proposed method will tend to provide an excessive number of distributions. To avoid this might be very simple. You randomly extract a subset of half of the sample data and analyze it with the same proposed methodology. Then you analyze the other half. Is you get acceptably similar results (acceptable differences in population statistical parameters) than you prove (say: you have an acceptable risk) that the number of samples you used

is about twice the required one to be representative of the analyzed population. You can play also the battle of progressively eliminating a percentage of randomly selecte data from the sample data and find when you start getting different results, but this generally will produce a progressive losing in reliability and you just move the question to a subjective threshold. My suggestion to the Authors is to add this test (and providing results!) to their proposed strategy. On the other hand, the paper is well written with few grammatical mistakes (e.g. pag. 2189, line 6?). Some difficulty in reading the paper surely rises from the used correct statistical/mathematical language (I prefer the clear approach in the cited book by Davies). This will produce a lack of interest of the scientists that deals with angular structural data (at least many that I know within the world-wide community). And a consequent underestimation of the paper (that is, I am afraid that many readers will give up from reading it after pag.2183, line 29!). This is a pity, so I suggest to the Authors to add a brief description of the meaning of the used equations to "the rest of us". As a final comment, my suggestion is that the paper is suitable for publication with minor revision, since my suggestions are limited to clarify/improve reading of the paper. Sincerely, Francesco Salvini

----