Hydrol. Earth Syst. Sci. Discuss., 9, C106–C109, 2012
www.hydrol-earth-syst-sci-discuss.net/9/C106/2012/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Technical Note: A significance test for data-sparse zones in scatter plots" *by* V. V. Vetrova and W. E. Bardsley

V. V. Vetrova and W. E. Bardsley

web@waikato.ac.nz

Received and published: 19 February 2012

General comment:

This paper is an extension of the method proposed in the paper by Bardsley et al. (1999) to test for sparse zones in scatter plots. Although the method might prove to be a useful tool for hydrologist, I think the paper cannot be published as is.

I fully agree agree with R. Hulf's comment in that the paper needs extra explanations. As is, it is not possible to understand how the test is performed. The author claim that, compared to the test of Bardsley et al. (1999), the new test can now cope with sparse zones containing a single point. However, as the theory of the test is not presented (or at least too briefly), I don't understand what differs between the two tests.

Response: We are a little perplexed as to why it is asserted that our test is with respect to a data-sparse region defined as always containing exactly one data point. It is clear throughout the text (e.g. p.1337, line 19) that the number of data points m in the sparse region is such that m$\geq$1 and not m=1. It may be that in the introduction to our test (Section 2) the example we considered was with respect to a specific case where it happened that m=1 and this may have caused the misinterpretation. We will clarify the introductory example as being simply an example and not defining the test. In the Bardsley et al (1999) paper m=0 always, so the current test is more general.

Specific comments:

1) I don't understand what is $\Delta(m)$: is it an area (as stated p. 1337 l. 20), or is it a proportion of area (as in p. 1338 l. 1)?

Response: Whether the sparse area $\Delta(m)$ is defined in area units or as a proportion does not change the outcome of the test. However, it is desirable to be consistent and in the revised version we will define the area as a proportion, consistent with the usage of the Bardsley et al (1999) paper.

2) I don't understand how the random swapping is performed in the test.

Response: This point was also raised by R. Hut. As noted in our response, we will clarify the definition of random swapping.

3) How is p (the p-value of the test) computed from $\Delta(m)$? This is not clearly explained so far.

Response: The value of p is obtained via the usual randomisation process of finding the proportion of randomisations yielding a better result (in this case a larger area) by random data reordering. We will make this explicit in the revised version.

4) I wonder if Fig. 2 is a very good example since the case m = 0 is tested, which was already possible with the earlier test of Bardsley et al. (1999).

Response: Fig. 2 is useful to include because it illustrates the case where an m=0 situation may or may not lose statistical significance with the subsequent appearance of a data point in a previously empty region.

5) I'm not sure to understand p. 1339 l. 7–8: "the autumn forecasting model has value for forecasting some future low inflows". Do you conclude this because the black line is close to he 1:1 line for low inflow (and because of the test)?

Response: The value derives from the fact that a future data point in the presently empty region would not necessarily cause statistical non-significance, suggesting the upper stochastic boundary is real. The prediction ability is that the downward trend in the upper data boundary implies that a forecast low flow has a good chance of translating into a low flow (below the upper boundary) in reality. A forecast high flow on the other hand might translate to a high or low flow in reality, because the upper data boundary is high on the right hand side of the figure. We will clarify this is the revision.

6) Regarding the spring inflow forecast: according to me (if I understood things correctly), even if the p-value would have been small, the black line is anyway to far from the 1:1 line to say that low inflows can be correctly predicted, so the conclusion that "there is no confirmed predictive ability for spring inflows" seems not to be a conclusion from the test itself.

Response: It is important when dealing with stochastic boundaries to make clear what is meant by "predictive value". In the classical use of prediction, the hope is the future value will be somewhere near the predicted value (plus or minus). When dealing with a stochastic boundary, the boundary may or may not be near the 1:1 line of a classical predictive model. With only a stochastic (upper) boundary the prediction is that the future value will be below the boundary value concerned (perhaps near the boundary or perhaps far below it). Depending on the situation, there is still predictive value in this sense and perhaps water resource decisions could be based on it. It is therefore still of value to check whether the sparse corner has statistical support. As

an aside, we are somewhat fixated in hydrology with a view of nature as a deterministic animal and "calibration" is with respect to passing a line through the central portion of recorded data with the residuals from fitting seen as the effect of some mix of still-to-be-found variables and random error. But it might also happen that nature is sometimes only deterministic to the extent of enforcing boundaries to data ranges, with inherent unpredictability between the boundaries. For calibrated boundary models for this type of variation, the 1:1 prediction concept no longer applies because only the bound value is predicted. However, the test presented here would still be applicable to the validation data set of the boundary-based model. We will expand the revised paper a little to discuss this aspect.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 1335, 2012.