

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

R. Hut (Referee)

r.w.hut@tudelft.nl

Received and published: 30 January 2012

Authors Vetrova & Bardsley provide a nice short technical note that generalizes previous work from author Bardsley. The method they propose for testing the significance of data sparse areas in scatter plots might prove to be a useful tool for hydrologist (and other scientist) looking for process understanding.

However, as presented, I believe that the manuscript requires some work before it is fit for publication in HESS. Specifically, the issue below need to be addressed (in my opinion).

The authors need to make clear what the hypothesis is that is tested against. They state that "the significant test ... can be defined generally as finding the probability p

C38

that random swapping of data points will give rise to a data-sparse region ...". As I understand it (and please correct me if I'm wrong), this assumes the null-hypothesis that the variables in the scatterplot are not correlated with each other. I would like to ask the authors to elaborate on what "swapping data" means for the assumptions made on the data.

Furthermore, I would invite the authors to spend a few words on data sparse regions when the null-hypothesis is that the variables are correlated, ie. "given that X and Y are correlated according to (ie.) $Y=f(X) + E$, where E is a stochastic variable independent of X, what is the probability p that an observed data sparse area formed by chance".

I believe that for correlated variables, using the authors test to identify data sparse regions in the space spanned by X and E (the determinant and the residual after regression) will help identify processes that are not (yet) part of the regression. This gives modellers an additional tool to identify missing processes.

I look forward to an updated version of this paper,

best regards,

Rolf Hut

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

C39