Hydrol. Earth Syst. Sci. Discuss., 11, C5366–C5371, 2014
www.hydrol-earth-syst-sci-discuss.net/11/C5366/2014/

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Extreme value statistics of scalable data exemplified by neutron porosities in deep boreholes" *by* A. Guadagnini et al.

Anonymous Referee #1

Received and published: 28 November 2014

This paper explores for the first time the statistical behavior of peaks over thresholds (POTs) associated with samples from scale-mixtures of truncated fractional Brownian motion (tfBm) or truncated fractional Gaussian noise (tfGn). The samples are neutron porosities from six deep boreholes in three diverse depositional environments.

The statistics of scale-dependent geological data is a relevant scientific question for the readers of HESS. The statistical tools used here were already applied to other data sets by the same authors (e.g. Guadagnini et al., PRE 2014) but not to neutron porosity data. Hence, the carefully worked-out results in Secs. 2-5 concerning the scale-dependent statistics of the data (frequency distribution of porosities and porosity increments, scaling of increments, and estimation of variogram parameters) are novel and interesting.

The statistics of data extremes is also a relevant question for HESS readers. However, in contrast to the expectations raised by the title of the manuscript, only the last two sections deal with the results of extreme value statistics of the porosity data (Secs. 6-7). And I find difficult to raise general conclusions about extreme value statistics of scalable data based on these particular results.

I have no doubt that this manuscript contains enough interesting material to be published in HESS. However, I found it hard to read and understand. This is partly because I come from the statistical and nonlinear physics community, and I am not familiar with the terminology employed in hydrology and earth sciences. But it is so also because the manuscript needs an extensive revision. While the title focuses only on extreme value statistics, the manuscript actually pursues three different goals: - First, to show that the neutron porosity data possess indeed the properties of scale-mixtures of tfBm or tfGn. - Second, to provide further support for the authors' unified theoretical framework which captures all common manifestations of scale-dependent statistics without having to associate this behavior with multifractals. - Third, the goal acknowledged in the title of the paper, namely to explore the behavior of POTs in this kind of samples. On the other hand, the organization of the manuscript does not contribute much to its readability.

In order to support these criticisms, and guide the authors in the necessary revision of their manuscript, I will comment on the contents of the different sections.

Title: It refers only to one aspect of the paper contents (extreme value statistics). It should be made more general, to be representative of the actual contents.

Abstract: It does not correspond to what it is expected from the abstract of a scientific paper (namely, a concise and complete summary of the research reported in the manuscript). It should be written again, to provide the following information: the objectives of the research, the procedure followed (what was done), the results obtained, and the conclusions reached.

HESSD
11, C5366–C5371, 2014

Sec. 1. Introduction: I recommend splitting the first paragraph into three different paragraphs, on (i) extreme value statistics, (ii) the key question of scale dependence, and (iii) the interest of neutron porosity data (for clarity).

The paragraph in p. 11640 about research on spatial correlations between large values of transmissivity in subsurface hydrology, and related issues, is distracting. I found this discussion somehow unrelated to the subject of the present paper. Please, make an effort to link this information to the research reported.

The description provided in the last paragraph of the Introduction, starting in p. 11642, could be linked to the actual organization of the paper, now missing. E.g. "... tendency of increments to have symmetric, non-Gaussian frequency distributions characterized by heavy tails that often decay with separation distance or lag, as shown in Sec. __; power-law scaling of sample structure functions (...), presented in Sec. __; etc."

Sec. 2. Source and frequency distributions of neutron porosity data: The authors explain that the data is part of a broader set, previously analyzed within a multifractal formalism by Dashtian et al. (2011). So a comparison of the results reported there and those of the present paper (based on the authors' novel method) would be desirable.

The paragraph in line 17 of p. 11643, on ML fits, could be made a separate paragraph, for readability.

Is there a physical basis for the different models fitted (gaussian, α-stable subgaussian, and NLN subgaussian)? Please, discuss.

Sec. 3. Frequency distributions of neutron porosity increments: The same question applies to the paragraph in line 13 of p. 11644. Is there a physical argument for fitting α-stable and NLN models to the empirical frequency distributions of increments?

Sec. 4. Statistical scaling of neutron porosity increments: - In line 17 of p. 11645 the authors report 'a break in power-law regime'; I would find the expression 'a crossover between two different power-law regimes' more appropriate. The authors interpret it as one regime representing variability within (w) sedimentary layers, and the other variability between (b) them. What are those layers? Is this interpretation based on some other information available, which for instance could match the characteristic vertical separation at which the crossover occurs? Results of the data analysis are followed here by interpretation. I think that taking the interpretation of the results to a (new) Discussion section, just before the Conclusions, would benefit the manuscript.

- The contents of lines 21-26 of p. 11645, about a similar dual-scaling phenomenon reported by Siena et al. (2014), would also fit better in a Discussion section.

- Researchers working on multifractality in the statistics of increments (e.g. in fully-developed turbulence) often make use of normalized p-root structure functions $C_p^N = C_p/R_p^G$, where $C_p = (S_p)^{1/p}$ and the normalizing factor $R_p^G = (S_p^G/S_2^G)^{1/p}$ is the ratio of structure functions for a Gaussian distribution, which depends only on p. This is useful to unveil deviations from monofractality and Gaussian statistics. While at short lags C_p^N follows a power law of the lag, with exponent ξ_p/p (and ξ_p depends nonlinearly on p if the signal is multifractal), at large lags C_p^N collapses onto a single curve for all p as expected for Gaussian statistics.

This could be an alternative way of representing the results in Figs. 7 and 16.

- The authors find Hurst scaling exponents $H_w \gt 1/\hat{\alpha}$ and $H_b \ll 1/\hat{\alpha}$, and they associate these results respectively to persistent and antipersistent variability. It would be helpful to explain why, or to provide a reference. This material could also be moved to the new Discussion section.

- Regarding ESS (extended self-similarity), straight line fits have indeed high confidence values R > 0.9; however, to be fair, it should be pointed out also that there is less than one decade of scaling in most cases.

In lines 14-17 of p. 11646, and lines 10-15 of p. 11648, the authors argue that finding

HESSD

11, C5366–C5371, 2014

Interactive Comment

the distribution of increments to satisfy ESS is akin to verifying that the data conforms to the new theoretical scaling framework proposed by them. This is true, but it should not be mistaken with the concept that multifractality –as a framework that would also explain the present resultsâĂŤ should be ruled out.

Is there any statistical analysis that could be applied to the porosity data and would allow a clear-cut discrimination of the two theoretical frameworks? And, also interestingly, in which respect the authors' theory provides a better description or clearer physical insight of the statistical properties of the porosity data analyzed here? The interested reader will find this discussion illuminating. Of course, it should be placed in a section devoted to Discussion, not mixed with the results of the data analysis.

Sec. 5. Estimation of variogram parameters: This section quickly becomes very technical. The first half, devoted to theoretical concepts, could complement Appendices A and B and make together a new section of the manuscript (Theoretical framework) that could go right after the Introduction. In this way the reader would have immediate access to the concepts and terminology employed in the data analysis.

The results for Lévy-stable subordinators (line 22 p. 11649 to line 14 p. 11650), summarized in Fig. 10, and the results for log-normal subordinators (line 15 to 27 p. 11650) show the remarkable fitting power of the authors' theoretical approach. I did not find reference to previous works. Is this methodology applied here for the first time?

The material up to here would already make an interesting and consistent paper, in which a complex description of scale-dependent statistics is successfully applied in its full power to actual neutron porosity data in several deposition environments.

The subject conveyed by the current title of the manuscript, the analysis of extreme value statistics, starts here.

Sec. 6. Frequency distributions of peaks over thresholds: I think that Fig. 13 is not essential. It could be removed. Alternatively it could be complemented with a figure

HESSD

11, C5366–C5371, 2014

of autocorrelations between POTs of neutron porosity increments, which is needed to justify the use of generalized Pareto distributions (GPDs) to represent frequency distributions of POTs.

The sentence about $p > 0.05$ (lines 24-26 p. 11651) should be rewritten in a less technical way –for clarity.

Sec. 7. Statistical scaling of POTs: POTs of absolute increments exhibit similar scaling behavior and Hurst exponents than unfiltered porosity increments, and also verify ESS. These results are summarized in Figs. 16 and 17. Being supposed to be the core of the paper, I found surprising that they are not discussed in depth.

I encourage the authors to provide a more elaborate discussion (in the new Discussion section) of the results on POTs of absolute increments. A question of interest would be to what extent are POTs of increments (for signals exhibiting the kind of scale-dependent statistics analyzed in the present paper) expected to follow the same statistical trends than the original increments. I missed also a discussion of the physical meaning of GPD shape and scale parameters.

Sec. 8: Conclusions I found the conclusions concise and well written. Several things that are said here can be found almost literally also in other sections of the manuscript. Maybe they could be rephrased there or removed.

I hope that the authors will find the previous comments and suggestions worth of consideration. They are intended to improve the clarity and readability of the original manuscript, thereby improving the impact of the research reported.

I will be glad to review the manuscript again after revision.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 11, 11637, 2014.