Nat. Hazards Earth Syst. Sci. Discuss., 3, C386–C390, 2015
www.nat-hazards-earth-syst-sci-discuss.net/3/C386/2015/

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "Brief Communication: Statistical detection and modeling of the over-dispersion of winter storm occurrence" *by* M. Raschke

F. Serinaldi (Referee)

francesco.serinaldi@ncl.ac.uk

Received and published: 7 April 2015

General comments

In this note the Author suggests the use of a generalization of the Poisson distribution (GPD) introduced by Consul and Jain (1973) as an alternative to the standard Poisson and negative binomial. The main advantages of GPD are the capability of modeling all possible types of dispersion (under-, equi-, and over-dispersion) and a parameterization which allows for rescaling the model parameters in order to represent thinned mixed/compound Poisson processes. Even though the manuscript is a brief communi-

cation, in my opinion, the presentation is too dry and not very clear, taking for granted some concepts, while overlooking some others, which in turn can be important. Moreover, even though the data analyzed are already studied in a previous paper, they should be briefly presented, at least, to make the paper self-standing. Specific comments are reported below.

Specific comments

Section 2: In my opinion, a first point that should be accounted for is the difference between inhomogeneous, compound/mixed Poisson processes. The first involve a dynamic rate of occurrence λ which depends on exogenous parameters (time or other covariate, in a regression/conditional set up), while the latter describe a mixture of processes by distributions with fixed parameters (rate of occurrence and possible additional parameters) resulting from integration or convolution procedures. In this respect, GPD describes the second type of processes, and can be used for the first type only allowing for dynamically-varying parameters, in the spirit of the extension proposed by e.g. Consul and Famoye (1992).

P1777L22: "The likelihood estimation is asymptotically the best estimation" should be "The likelihood estimator is asymptotically the minimum variance estimator"

P1778L5-6: I understand "unbiased" (perhaps, better than "bias-free") and "consistent", but "well established in mathematical statistics" is not clear to me. Please, clarify.

Section 3: I agree with the discussion about the possible different definition of "clustering". However, please reword P1778L12-14: "...partial series of an exceedance of a threshold..."?.

P1778L19: Please, be careful when using engineering concepts. Return level refers to the magnitude of a variable corresponding to a given return period (or better, to an annual probability of exceedance). As the Author refers to the work of Karremann et al. (2014) (please, fix "Karlemann" in L20), in that context, the return levels are spec-

ified thresholds of storm intensity corresponding to 1-, 2-, and 5-year return periods. Therefore, the notation "return level RL=1" should be "1-year RL" meaning "event with intensity equal to the return level corresponding with return period equal to 1 year". Measurement units are fundamental in this context and cannot be overlooked.

P1778L22-25: In my opinion, GPD and negative binomial involve directly or indirectly the integration over the distribution of the random fluctuations of the rate of occurrence. Thus, these models describe a mixed/compound Poisson process rather than an inhomogeneous Poisson process.

Section 4: In my opinion, this section suffers from a widespread practice, which is however questionable. Tests for departures from equi-dispersion based on the index of dispersion or similar provide a formal check resulting in rejection or not of a null hypothesis, while information criteria provide only a ranking of different options with no information about the significance of the differences. In this respect, without a formal test (likelihood ratio tests, or something else), there is no way to reach a reasonably objective conclusion about the actual effectiveness of using one model rather than another. In other words, is GPD better than Poisson or others? Or, can differences be ascribed to simple random fluctuations? Ranking the scores is not sufficient.

P1782L3-14: I have to say that I realized the possible usefulness of GPD only after reading this paragraph, which however should be expanded and made clearer. Even though the data analysis is just a proof of concept, it cannot be reduced to a paragraph. In fact, there is no way to understand this application without reading Karremann et al. (2014), especially because that study uses different products (observations, reanalysis data, and an ensemble of GCM (not GNC!) simulations), and quite an elaborate preliminary analysis and event extraction, which the reader needs to know to assess the results. Note that even acronyms of the different datasets are not defined in the present version.

Section 5: I think that GPD is just one of many options available to model under-, over-

dispersed count data. The same GPD was already further extended, and other more refined methods are available. The conclusion "It is also very helpful for detecting over-dispersion of German winter storms in combination with the introduced criterions of model selection... a simple analysis of the sample of historical storms could not clearly detect the over-dispersion; the sample size is too small." seems to me not supported by results. In fact, owing its flexibility, GPD is expected to be close to negative binomial in the case of over-dispersion and to Poisson in the case of equi-dispersion. My feeling is that GPD with rescaled parameters simply reproduces the Poissonian model for e.g. (NCEP, RL=1), (NCEP, RL=2), and (DWD, RL=5). To conclude that storms are actually over-dispersed in these cases, it should be shown that θ is significantly different from zero, and, if it is so, it should be shown that the results corresponding to small sample sizes (i.e. apparent equi-dispersion) can be obtained by simulating samples with the same size from the over-dispersed GPD. AIC and SBC per se do not give information about the nature of the fitted model, and Fig. 1 is not sufficient as it does not show the uncertainty related of the estimation of β and $E[X]$.

Editing notes

P1779L1: "If over-dispersion results"

P1779L7: "$P_{thinning}$"

P1781L15 and L17: "criteria"

P1781L22: "storms with return period of one year ($x \geq$ 1-year RL)"

P1782L3: "GCM", please fix throughout the text and Fig. 2

Figures: Please, consider to improve the captions making the figures more self-standing

References

Consul , P.C. and Famoye, F.: Generalized Poisson regression model, Communications in Statistics - Theory and Methods 21(1), 89-109, 1992

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., 3, 1775, 2015.