Nat. Hazards Earth Syst. Sci. Discuss., 2, C2242–C2248, 2014
www.nat-hazards-earth-syst-sci-discuss.net/2/C2242/2014/

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "A framework for modeling clustering in natural hazard catastrophe risk management and the implications for re/insurance loss perspectives" *by* S. Khare et al.

Anonymous Referee #2

Received and published: 9 October 2014

The manuscript by Kharre et al. describes a procedure for modelling the occurrence of clustered natural phenomena for the purpose of natural risk estimation and re/insurance contract pricing. They used an inhomogeneous Poisson model – a so-called Poisson-Mixtures model having the random and clustered components for generation of event series. Basically, the Poisson-Mixtures model is based on the Negative Binomal distribution with the modulated rate of occurrence to represent the over dispersed point process. The authors investigate the clustering of the European wind storm events over the past 39 years and its relation to NAO which is thought as an underlying modulating process. Finally, the authors compare the return periods of

event counts above a threshold between empirical distribution, homogeneous Poisson and Mixed-Poisson models and discuss implications of using Mixed-Poisson model for re/insurance contract pricing. The manuscript addresses an important emerging question of modelling clustered event occurrences and presents a procedure for using Mixed-Poisson model for risk assessment and contract pricing in the insurance industry. As such, it has a potential to provide a valuable contribution in the risk research community and narrow the gap between statistical theory, geoscience and insurance practitioners. However, I have several concerns with regards to the presented methodology and application which require author's clarification and additional review round before manuscript can potentially be considered for publication. Although the manuscript is generally well-written and logically structured, it is not concise and contains several repetitions and trivial theoretical digressions. It rather presents a model application manual (which is also valuable) rather than a scientific paper. Important issues with regards to the impact of several assumptions in the presented procedure are not tackled.

Major comments:

1. The striking feature of the manuscript is that it completely lacks literature review and does not place the presented work into the contemporary research context. A particular novelty of the manuscript is thus difficult to judge for the audience as it is fully uncoupled from recent developments. The history of using Negative Binomial models to simulate clustered point processes is long, see e.g. Lang (1999), Eastou and Tawn (2010) and references therein. In particular, Eastou and Tawn (2010) described and compared the Negative Binomial, Poisson regression and the Mixed-Poisson model as an extension of regression model capable of considering the random and clustering effects in a single framework. The mixed models seems to be established in the statistical community and has already its access to the geosciences, particularly flood risk research (Eastou and Tawn, 2010). Also, other variations/alternatives for modelling clustered behaviour have been recently presented e.g. Villarini et al. (2013). If

I understand it correctly, the authors here basically apply the Negative Binomial model in the notation of Eastou and Tawn (2010) and controlling the clustering behaviour by modulating gamma variance. When variance tends to zero, the model reduces to a homogeneous Poisson process used to model the random part. The Poisson-Mixtures notation in the presented manuscript means something different as “mixed model” from Eastou and Tawn (2010) which uses covaraites in the Poisson regression model mixed with random component. In the consequence, the manuscript should include a sound literature review and be placed in the contemporary research field paying a due tribute to the previous developments.

2. The major novelty claimed by the authors resides in the introduction and use of the so-called “Super-clusters” within the Poisson-Mixtures framework. These “Super-clusters” pool events above one or more clustering thresholds that are defined by the magnitude of modulating factors and consequently by the magnitude of clustering. The event occurrences within “Super-clusters” are supposed to have similar clustering behaviour. One can define several “super-clusters” having different clustering behaviour and additionally separate the random portion characterized by the homogeneous Poisson process. These “super-clusters” are than parameterized and used to generate synthetic event series in the Poisson-Mixture framework, whereas the correlation of event occurrences is retained within the “super-clusters”. My major concern with the “super-cluster” methodology is that it is a subjective pooling controlled by imposing some hard thresholds or as it says in the manuscript (P5254, L15ff) “apply a statistical clustering algorithm to an archive of historical events. . .”. In the presented case study, neither a description of clustering algorithms is provided, nor is their effect on resulting synthetic time series investigated. Furthermore, it is also not clear how the length of a historical time series affects the parameterization of the “super-clusters”. This is a crucial issue whether the parameterized Poisson-Mixture framework correctly reproduces clustering behaviour only within cluster groups separated by these thresholds or the clustering behaviour of the entire time series. This should be investigated and shown. Finally, pooled groups of occurrences are coined here with the term “super-cluster”.

Truly saying, I do not understand what is really “super” about those clusters. Super in relation to what? Sounds attractive on the first glance, but maybe we should be more modest and name things as they are – clusters characterized by certain degree of dispersion.

3. In the described wind storm example, the authors impose a hard event magnitude threshold on storm severity index (SSI) of 2.5 (without applying any clustering algorithm as mentioned in the methodology description) and assume events below the threshold to follow homogeneous Poisson process and those above to have gamma variance of 1.5 (P5262, L27-28). Is the assumption of Poisson/overdispersed behaviour above and below threshold justified by historical data? How the gamma variance of 1.5 was computed or selected? If these assumptions are not clarified or derived from historical data, all the analysis of different model performance in Figure 5 makes little sense. Models then produce something that can but should not necessarily resemble the empirical exceedance probability given by the red curve and associated confidence interval. I would propose to present an experiment validating the model and assumptions therein which consists of the estimation of the overdispersion from the historical data and generation of several hundreds or thousands realisations of point process of the same length as historical series and comparing it to the empirical occurrence rate exceedance probability using some statistical tests.

4. When reading the description of analysis presented in Figure 5 (P5261, L13ff), it is not clear which quantity or return period of which quantity you analyse. You speak of “exceedance probability of the annual maximum event loss” or related to that “the exceedance probability of the annual maximum SSI” and you term it Occurrence Exceedance Probability (OEP). As becomes clear only on P5264, L2-3, you analyse the exceedance probability of annual occurrence rate of SSI above a certain threshold (how often SSI exceeds a threshold within one year). This is something different from “exceedance probability of the annual maximum loss”. From this perspective, it is not clear to me what OEP2, OEP3, OEP4 actually mean? If Figure 5a shows the return

period of occurrence rate of SSI above threshold, the occurrence rate of e.g. 3 means 3 events per year above threshold. So what is OEP3 then?

In your current analysis you do not show the risk curve for annual maximum loss derived from 10,000 years of synthetically generated windstorm data, but this is what you are actually interested in for contract pricing. You do not need to know the number of exceedances per year, but the cumulative resulting loss, do you? Please, clarify the notation in the text and maybe present the results for SSI magnitude (a proxy for loss) additionally to the exceedance of occurrence rate above threshold.

5. As already criticized by the first reviewer, the analysis of relations between NAO and storm frequency and clustering is rather weak. The R-squared values in Figures 3 and 4 are fairly low and regressions do not say much. You basically use the relationship between storm occurrence and NAO to argue that there is a physical driver behind the over dispersion of storm occurrences. I do not see the rationale behind this argumentation, why do you need it? The methodology proposed does not use any covariate process. It parameterises overdispersion with gamma variance, where this overdispersion comes from is irrelevant, doesn't it? So, I suggest to remove any excursus to the relations with NAO. At most a few sentences on potential drivers could be added and supported by strong references.

6. Although roughly understanding the idea of relevance of clustering for insurance industry and contract pricing, I had no clue at all about the mechanics presented in chapter 4. The terminology and computations used are not well-explained and are absolutely puzzling for me as a geoscientist. It starts with "compute the loss of each event to the "layer" ", and further with "aggregate limit" and "re-instatement". I had no idea what that means, and why those terms computed the way they are. To make it short, either this part should be explain in much greater detail providing definitions of the terms or (and) reviewed by someone from the field of insurance math.

7. Finally, as mentioned in the introduction the manuscript is not concise and should

be significantly condensed on one side and include additional analyses and explanations on the other side. Shorten P5250. Avoid repetitions and lengthy introductions like P5250, L24-25ff or P5251 L8 and at some other places. Please, define homogeneous Poisson process e.g. through the homogeneity in time, independence of successive events and characterization of the waiting or interarrival time as uncorrelated and following exponential distribution. You current description on p5251-5252 is too mechanistic and rather resembles a model application guide than a research paper. One can expect an interested scientific community to be able to draw random variables from the distribution. Delete the paragraphs from P5255, L25 on till the end of chapter. This is mechanistic repetition. Just give the definition of the clustered process instead. Avoid the mechanistic descriptions on P5254 and 5255 accordingly, but keep the presentation of methodology still transparent. Description on pages 5257, L10ff – P5258 is basically a repetition of above-mentioned facts. I disagree with Point 3 on P5257. Why the NB is fundamentally different from you approach? NB just parameterises the bulk of data as overdispersed. In your notation it means setting the threshold just very low. Point 5 on P5258 is rather vague.

Minor remarks:

Suggestion for the alternative title which is maybe more precise "The effect of clustering of natural hazard phenomena for modelling and management of re/insurance risk portfolios".

What is CRESTA?

P5264, L10: The statement "This seems unreasonable given that this year exists in the historical data" is incorrect. You can have in any physical year a 10000-year event, but the probability is small. It may have occurred in the past.

Check description of Figure 5. Colours on graphs, in the legend and elsewhere in the text are mixed.

References

Eastou, E. F., Tawn J. A. (2010): Statistical models for overdispersion in the frequency of peaks over threshold data for a flow series. Water Resources Research, 46, W02510.

Lang, M. (1999): Theoretical discussion and Monte-Carlo simulations for Negative Binomial process paradox. Stochastic Environmental Research and Risk Assessment.

Villarini, G., Smith, J. A., Vitolo, R., Stephenson, D. B. (2013): On the temporal clustering of US floods and its relationship to climate teleconnection patterns. International Journal of Climatology, 33, 629-640.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., 2, 5247, 2014.