Nonlin. Processes Geophys. Discuss., 1, C120–C123, 2014
www.nonlin-processes-geophys-discuss.net/1/C120/2014/

Nonlinear Processes in Geophysics
Discussions

Interactive comment on "Instability and change detection in exponential families and generalized linear models, with a study of Atlantic tropical storms" *by* Y. Lu and S. Chatterjee

Anonymous Referee #2

Received and published: 8 May 2014

The title is good and understandable for a wide audience, and the topic they cover is interesting. I'm not sure whether the paper contains new and significant results. Compared to "old" literature on change-point analyses the paper is on a "new" level. The authors do not flag that they have invented something genuine new. The topic is interesting and important. The authors use a technical language – but they can write less technical and be read and understood by a broader group of people. And it is possible without reducing the level of intention and precision. The inclusion of the figures is important. Figure 1 could be sharper and a little bit bigger. Figure 2 is OK, but the authors could have included a filter (for example a running mean) to show the central tendency of the process over time. The authors can also consider including

in figure 2 the time points where they "believe" the time series have changed. To a large extent the paper obtained results into the context, with relevant references. The authors give a short summary of the literature which describes the derivation and application of for example control chart and change-point analyses. In the applied part of the analysis/paper the authors compare the results based the "traditional" approach which is based on the Gaussian-CUSUM and the authors' approach where they use the so-called Exponential Family-CUSUM. The authors argue, if I understand them correctly, that the precondition to obtain a valid change-point analysis is to apply the correct probability density function. If the phenomenon is non-Gaussian, then it is valid to use a non-Gaussian distribution – and then they refer to the exponential family (Poisson, logistic, gamma etc.). However, the authors do not compare the result of the hurricane-analysis (the applied part) with the climate research-literature. Most of the presentation is clear and concise. I have the following remarks: (a) With regard to the application of the EF-CUSUM on the Atlantic tropical storm, the authors apply the Poisson-distribution in the analysis, but they do not "prove" or show that the time series follows a Poisson-distribution. Why not include a figure which shows the frequency distribution of the time series they are focusing on and the authors can as well include a test which shows that the observations follow a Poisson-distribution or process – or at least show that the phenomenon is not normally distributed. Later in the analysis they argue that they detect changes in the parameter. Why not make a figure which show how lambda (the Poisson-lambda) varies over time and also plot the CUSUM. (b) The authors could have described the time series according to whether the processes are autoregressive (AR) or moving average (MA). (c) The authors do not discuss what kind of changes the time series are undergoing. Are we talking about an overall trend, short-term trends, short-term or long-term shifts in the mean and variances? Figure 2 shows clearly that volatility in the central pressure probably changes "significantly" short after the 1950s. The volatility is probably also changing for the maximum sustainable wind and the number of hurricanes. There is no discussion about these things which are important in a modelling context. (d) The method applied requires selection of average run length. The authors have chosen a fixed number 200. To me the choice of run-length is not clear. (e) The method applied requires estimate of the coefficients characterizing the distribution of the variable. In the empirical analysis the authors use the Poisson-distribution and base the lambda on the first observations, i.e. observation from 1850 to 1900 for the longest series and 1950 to 1970 for the short series. The estimates function as benchmarks. My impression is that these choices are ad-hoc and not supported by scientific arguments. They impose restrictions on the model/analysis. What will happen if they choose other time intervals as benchmarks? (f) Please look at table 6 and 7: the constants c = $\frac{1}{4}$, $\frac{1}{2}$ and 1 are unclear for me the role they play in the analysis. I suggest the authors describe more clearly how they are included in the equations and the role they play. (g) In the abstract the authors say: “We derive the related likelihood ratio test statistic”. In the analysis of change-points I cannot see any statistical tests of the potential breakpoints the authors argue that they have identified. Where are the critical values and the estimates? I suggest they show more clearly how critical test-values are derived and applied. (h) Please, look at pp. 392-383, Example 3.0.1, points 1 to 5: The authors derive the CUSUM statistic conditioned on the properties of the variance-covariance matrix. According to what I can see the authors do not use these derivations in the empirical part of the analysis. I'm therefore not certain what role or implication these deductions on pp 382-383 have in the paper. (i) As far as I understand, the method derived in the paper is based on the condition that the variables applied in the analysis independently distributed which means that the observations are not correlated over time. Application of the Poisson-distribution requires that the observations are not correlated in time (Please, see first and second line under section 3 “Distributional stability in exponential families”, pp 377-378. In the empirical analysis the authors do not say anything about the whether the observations are correlated or not. (j) The authors do not discuss or present in the concluding/summary section any critical remarks on the method. What is the strength and what are the weak parts of the method which challenge the validity?

Interactive comment on Nonlin. Processes Geophys. Discuss., 1, 371, 2014.