Referee Comments regarding the study "Logit-normal mixed models for Indian Monsoon rainfall extremes"

General Comments

Overall, I see two major issues with the discussion paper. First, there appears to be too little focus on the issue being addressed, extremes of rainfall for the Indian monsoon. In general, identification of extreme events at the station level and of their drivers at a larger-than-station level is of some interest to the researchers studying extremes of rainfall. However, insufficient attention is paid to these important points in both the analysis as well as at the level of motivation of the research (including an incomplete and somewhat incoherent review of the existing results) and choice of the data set (see below for specific details). The overall conclusions of the study are, correspondingly, very vague e.g. "..promise of modeling some of the unobservable physical features within the complicated network of interactions the underly extreme climate patters while maintaining statistical integrity." (pp 206). It is not explained in what sense the method(s) illustrated here are "better than" (or add to) those used in the prior literature, nor is it indicated how larger scale patterns (e.g. regional distinctions in effect of covariates) in rainfall extremes are to be thought of, in the context of the model presented. Finally, there is no discussion of (or motivation for) the possible differences in effects of potential large-scale drivers (e.g. ENSO and the IOD) between "normal" and "extreme" rainfall.

Second, given that the focus of the analysis almost exclusively on the statistical aspects of the GLMM model, and in general statistical analysis, the model details themselves are sketchy and/or incomplete to the point of incoherence. Basic details are sometimes not provided (e.g. the MSIM model in §3.2 is hurriedly and incompletely sketched and in the Bayesian model, in §3.3, virtually no details or equations are provided for the posterior distributions and method of simulation, a key point in Bayesian analysis) while even the basic GLMM model set up, in § 2, suffers from many problems, including unlabeled and unexplained objects ($b(\eta_i)$ and $c(\mathbf{y_i}, \tau)$) and lack of attention to notation (u and U are used (inappropriately) interchangeably and the first equation in the paper, pp 196, is clearly wrong). These issues can leave the reader unfamiliar with GLMM models confused, defeating the very purpose of the detailed exposition of GLMM models provided in the study. This is rather surprising, given the relatively standard nature of these tools in the statistical literature. Finally, §5, a key section of the study, suffers from excessive brevity in both methodology and discussion of results. Too many of the coefficients are not significant, and further, the simulation results (§4.2) indicate that many models are severely biased and do not at all perform well. Given this, the authors do not explain how confident the applied researcher can be with his results, upon using this methodology. Further, model goodness-of-fit measures (e.g. LR tests) are neither reported nor even discussed, rendering the robustness of these models (and, by extension, their applicability) somewhat suspect.

Finally, the writing is in general confused, with far too many clumsy constructions, meaningless (e.g. pp 195, lines 9-11, first sentence, pp 196, sentence beginning on line 14, pp200) or even incomplete (line 13, pp 198, beginning with "and the approximate...") sentences. I suggest the authors carry out a very careful editing of the entire paper in order to eliminate these (and other) errors.

Specific Comments

In addition to the general comments above, I provide a few very specific suggestions for changes.

Extremes for the Indian sub-continent

1. A (very brief) discussion of existing literature beyond Goswami et al. (2006) would be helpful. In particular, the studies Krishnamurthy et al. (2009), Rajeevan et al. (2008), Ajayamohan and A. Rao (2008) and Ajayamohan et al. (2008) deal with different aspects of the issue and their inclusion, if only briefly, will particularly enrich the discussion and motivation.

2. The problems associated with the use of fixed rainfall levels for thresholds has been discussed, for the Indian context, in Krishnamurthy et al. (2009). While recognizing that the use of percentiles at the station level for the current study is probably too time consuming, I feel that at a minimum, a (very brief) discussion of possible problems with the approach used for identifying extreme rainfall in this study is likely very useful.

Methodology

1. Numbering the equations will make it easier for the reader (and the referee) to easily discuss them. The first three equation in §2 must be clearly explained, as it is not now. For instance, the first equation is clearly incorrect, and should read instead

$$\mathbf{Y_i}|\mathbf{u} \overset{ind.}{\sim} f_{n_i}(Y_i|u) = \exp\left\{\frac{\mathbf{Y_i}\eta_i - b(\eta_i)}{\tau^2} - c(\mathbf{Y_i}, \tau)\right\}$$

   In addition, it is never explained to the reader that the expression above is actually the density of a random variable from the exponential family, with $b(.)$ and $c(.)$ having specific relationship to η_i and that τ is a scale parameter. Also, $\mathbf{Y_i}$,$\mathbf{y_i}$ and $\mathbf{u}$, $\mathbf{U}$ are used interchangeably, confusingly and incorrectly, respectively. Finally, there is a typographical error in the first equation, where f_{n_i}reads f_{η_i}.

2. Line 11, pp 197, introduces confusing notation for density function with parameters β, σ, as $f_{\beta,\sigma}(.)$. It might be less confusing (and cumbersome) to instead use $f(.|\beta, \sigma)$.

3. In §3.1, Penalized Quasi-likelihood methods, it is never clearly explained to the reader what is penalized. In the interest of the reader's understanding, it might be worth inserting the equation corresponding to the Laplace approximation and mentioning exactly that the penalty term comes from the approximation process for the likelihood.

4. In §3.2, the MSIM method is very confusingly described. It is not likely clear to the casual reader what is being simulated and where the unknown parameters enter into the second stage Newton-Raphson method. I recommend clarifying these aspects to the reader.

5. I also recommend adding, just before the beginning of §5, a brief summary of the rather confusing §4, to aid the reader in understanding the crux of the issues addressed in §4.

6. In your discussion of the results (e.g. paragraph 3, pp 205) you compare coefficients across time (as in Fig. 6a). However, such a comparison of coefficients over time is carried out without any measure of individual confidence interval (i.e. it is not known how many of those individual fixed effects are significant), leading to difficulties of comparisons. In addition, in the absence of a joint CI for the whole period, the "cyclic patterns" detected can easily be a statistically meaningless artifact. It is important to at least flag these issues to the reader and in general, to not "over interpret" these results.

7. Regarding §5, my concern is also that there are really no "results" to speak of, since many of the "effects" are mentioned as being not significant (although there is no systematic enumerating of how many coefficients are not significant in each of the methods used). As mentioned already, this section is really too brief to allow the reader to judge the utility of the methods illustrated. A clearer, slightly expanded, narrative here will allay such concerns.

8. Finally, given that this study is primarily illustrating the applicability of a methodology to an applied audience, the lack of clear cut recommendations regarding the methods illustrated and an investigation of the robustness of some of the methods must be addressed by the authors. First, some guidelines regarding, or a discussion comparing, the different approaches to GLMM is likely to be very helpful to the applied researcher. For instance, it is never clearly explained which of the methods perform "better" with the data set investigated by the authors. Given this, can the applied researcher wishing to use the method in his own research, draw any broad conclusions? In addition, the issue of goodness-of-fit must be at least discussed. Can reliable tests for the variance components be easily implemented for many of the methods investigated in the study? While a comprehensive

investigation of these issues is likely outside the scope of the current study, a few simulation investigations and some discussion of these issues, possibly drawing on existing statistical literature, is surely pertinent here.

References

Ajayamohan, R. and A. Rao, S. (2008). Indian Ocean Dipole modulates the number of extreme rainfall events over India in a warming environment. *Journal of the Meteorological Society of Japan*, 86(1):245–252.

Ajayamohan, R., Merryfield, W., and Kharin, V. (2008). Increasing trend of synoptic activity and its relationship with extreme rain events over central India. *J. Climate*.

Goswami, B., Venugopal, V., Sengupta, D., Madhusoodanan, M., and Xavier, P. (2006). Increasing trend of extreme rain events over India in a warming environment. *Science*, 314(5804):1442.

Krishnamurthy, C., Lall, U., and Kwon, H. (2009). Changing Frequency and Intensity of Rainfall Extremes over India from 1951 to 2003. *Journal of Climate*, 22:18.

Rajeevan, M., Bhate, J., and Jaswal, A. (2008). Analysis of variability and trends of extreme rainfall events over India using 104 years of gridded daily rainfall data. *Geophys. Res. Lett*, 35.