Nat. Hazards Earth Syst. Sci. Discuss., 2, C757–C772, 2014
www.nat-hazards-earth-syst-sci-discuss.net/2/C757/2014/

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "Stochastic daily precipitation model with a heavy-tailed component" *by* N. M. Neykov et al.

N. M. Neykov et al.

neyko.neykov@meteo.bg

Received and published: 20 May 2014

First of all, we would like to thank the anonymous Reviewers for their positive evaluation, encouraging and constructive comments and suggestions. We are convinced that our paper notably benefits from their suggestions. The next points explain in more detail how all the comments and suggestions made by them have been addressed.

Point to point authors reply to Reviewer 1 report

Specific comments

Comment 1. My major criticism against the paper is the lack of its generality. The paper presents indeed a quite comprehensive analysis and model validation for only one station. To what extent those results can be easily generalized is uncertain. For example the authors chose to reproduce the tails of the precipitation distribution using a GP model. For example, in a recent study, various distributions have been found as adequate for describing the behavior of the tail from daily precipitation records (Papalexiou et al., 2012). I believe that this should be discussed in more detail in the manuscript. The question that is still open is the general applicability of the model (e.g. for precipitation records from different climates etc).

Answer 1. The reason to reproduce the tails of the precipitation intensity distribution using the Generalized Pareto (GP) model rely on the well known theoretical result based on the extreme value theory which implies that the upper tail of essentially any distribution can be approximated by the GP tail. This is an asymptotic result and in practice likelihood estimation problem might arise as a consequence of small sample size. The distribution of the extreme precipitation intensity at station Ihtiman is identified as heavy tailed because the GP distribution shape parameter estimate $\xi = 0.127$ is significantly different from zero which is supported by the likelihood ratio test: $\xi = 0$ (exponential distribution) versus $\xi > 0$ (GP distribution).

We agree with the main findings and conclusions presented in Papalexiou et al. (2012). We can confirm based on our experience about the territory of Bulgaria that not every precipitation data series exhibits a heavy upper tail especially sites with relatively short records. As pointed Papalexiou et al. (2012) in such cases the lognormal distribution which is also a heavy tail might be used as an alternative of the gamma and Weibull distributions in modeling the intensity precipitation component. Perhaps, the usage of the gamma and Weibull distributions within the weather generators framework has been dictated due to widely available and easy to use software. It seems to us the at-site hybrid daily precipitation models will be widely although these models are not a panacea in order to fit all the precipitation data well. One could expect only a substan-

tial improvement provided the Weibull and gamma distributions are not capable to fit and produce a heavy enough upper tail for the daily precipitation intensity distribution.

Comment 2. Moreover, the study focuses on the reproduction of precipitation at the daily scale. I believe that this choice of the authors should be further justified. The last two decades, the majority of hydrological and ecological applications, have focused on describing and resolving the relevant processes on much finer scales (e.g. hourly or even finer for urban hydrological applications). Taking this into account, the potential of the suggested model can be limiting. To my belief a comprehensive and complete stochastic model for precipitation should target on the reproduction of all the essential precipitation statistics for a wide range of scales relevant for hydrological and ecological applications. I would encourage the authors to discuss this in further detail in the manuscript.

Answer 2. The hybrid distribution daily precipitation models can be further develop to model hourly precipitation data. For this purpose the precipitation model occurrence and intensity link functions have to incorporate some additional finite Fourier series terms and appropriate autoregressive covariates varying through the daily hours in order to account various dependencies. Also the threshold selection have to be on an hourly time-scale. Essentially, this would be an adaptation of the methodology proposed by Katz and Parlange (1995), see also Chappell et al. (2009). However, estimation problems might arise in case of a short hourly precipitation data series.

The at-site hybrid daily precipitation generalized linear models (GLMs) can be combined with a single-site disaggregation model based on Poisson cluster processes in order to achieve finer temporal resolution. For instance, simulations of long sequences of sub-daily precipitation data can be obtained from hybrid GLMs simulated daily precipitation totals using the approach and HYETOS software discussed by Koutsoyiannis and Onof (2001). In this way a comprehensive stochastic model could be developed that would satisfy various requirements.

Comment 3. One final remark I have for the manuscript concerns its novelty and originality. From my point of view the paper follows closely the work of [Furrer and Katz, 2008]. I recognize that there are differences in the methodology (e.g. In this paper both a gamma and a Weibull distribution are used for the simulation of the body of the precipitation depths), but the overall concept is very similar. I would encourage the authors to put more emphasis on the novelties they introduce in the paper, since I believe this should be the added value of the present (and in general of any) study.

Answer 3: The focus of our paper was on the adaptation, software implementation and validation of the hybrid GLMs proposed by Furrer and Katz (2008) rather than on it's novelty from a statistical point of view. For instance, the major computations rely heavily on the vector generalized linear models (VGAM) package which encompasses the classical GLMs (not necessarily from the linear exponential family) and generalized additive models, various quantile regression techniques, extreme value models (generalized extreme value, GP, Weibull). Moreover, within VGAM framework all the response distribution parameters can be specified as a parametric or nonparametric functions of the covariates (predictors) allowing a greater flexibility in regression specification.

Reviewer 1 Editorial suggestions:

a) I think Figure 1 is a bit messy and difficult to interpret correctly. The only visual difference a reader can appreciate concerns the very rare extremes. The authors may consider a different option in order option to show those results (e.g. A table with interannual statistics, etc.).

b) In Figure 6 it would be nicer to plot the data points in the graph, rather than as horizontal ticks. Doing that the comparison between model and data would be more straightforward visually.

c) In Figure 9, due to the high skewness it is difficult to interpret the box plots, since all the boxes are very small. It would be probably better e.g. to use a logarithmic

scale for the precipitation depths.

d) Moreover it would be useful to report a table with the basic statistics (mean, standard deviation, skewness, kurtosis etc.) of the observed vs simulated series shown the Figures of the paper.

Answer to Editorial suggestions 1:

a) Fixed, see Fig. 1 of this report, the extremal value is dropped out from the data to get better monthly and annual perception;

b) We are not willing to change the plots in order to make a comparison with those presented by Furrer and Katz (2008);

c) We reproduce the box plots in a logarithmic scale, see Fig 2 in this report;

d) We could report tables of the basic statistics, however, the boxplots contain more information.

Point to point authors reply to Reviewer 2 report

The authors comment: The report of the 2nd Reviewer is based on an early version of the manuscript which was significantly improved and published as NHESS Discuss., 2, 135, 2014, www.nat-hazards-earth-syst-sci-discuss.net/2/1/2014/, doi:10.5194/nhessd-2-1-2014

Reviewer 2 general comments 1. The paper reads well, its formal structure does not need serious comments, the adopted methods are correct and the main conclusion (the novel method performs better than the traditional one) is clear. The results are supported by a number of tables and figures, quality of which is adequate; they only need minor technical improvements (listed below in section Technical comments).

Therefore, I would recommend the manuscript, after minor revisions, for a publication in NHESS.

My main concerns are in line with the opinion of Reviewer 1. First, the study is presented and the stochastic model is evaluated in the case of a single station, and it is not clear whether it is possible to implement the model in other regions of the country or eventually, at sites with considerably different precipitation regime. Second, the authors should at least discuss the possibility of extending the model towards aggregation times below the daily scale (e.g., hourly precipitation intensities).

Answer 1: The results reported in the manuscript are typical of those obtained from other data sets which belongs to different climate regions over the territory of Bulgaria. This suggests that the approach of Furrer and Katz (2008) is widely applicable. Please look at Answer 2 in reply to Reviewer 1.

Reviewer 2 Specific comments:

1. Lines 48-49 I would like to see a short comment on how unique the referred rainfall event was in the light of the climate of the country.
2. Line 112 Can you please explain why exactly the threshold of 87% was selected?
3. Line 162 Can you please explain why the previous day precipitation occurrence was included and not some other covariate? Is it a result of a former analysis?

Answer to Reviewer 2 Specific comments:

1. Ihtiman monthly and annual box-plots of daily precipitation data are presented on the plot of Fig. 1 of this report. The extremal daily value 234 mm recorded on 5 August 2005 is dropped out from the data for better perception.
2. The threshold choice is based on a range (80% - 95%) of thresholds for which

the resulting GP distribution parameter estimates do not change too much and at the same time that provides a reasonable model approximation.

3. Previous day precipitation occurrence, a sine and cosine wave, and lag of NAO index are included as covariates in the model and according to the likelihood ratio test they are significant.

Reviewer 2 Technical comments:

1. Line 127 . . . but there are no strong grounds for applying a varying threshold. Nevertheless, you applied the varying threshold though, did not you? This part of the sentence seems a bit confusing for me. . .
2. Line 132 Mistyping: write Table 2 instead of Table 6.
3. Line 166 Function g(x) is not explained explicitly (Obviously, it is GP, based on the information on Lines 165 as well as 95, but I suppose all variables should appear in the explanation that follows the equation).
4. Lines 171 and 172 The abbreviation GPD is used twice; however, without clarification. I suggest dropping it and using GP distribution instead.
5. Line 187 There are no yearly data in Fig. 5, thus, delete the part . . .and for the entire year (lower line plots).
6. Line 190 Change the word order hybrid and.
7. Line 207 Separate bottomrow.

Answer to Reviewer 2 Technical comments:

1. Fixed: "... but there are no strong grounds for applying a varying threshold which is the main reason to use a fixed threshold within the hybrid precipitation models development at the next paragraph."
2. Fixed concerning Lines 132, 166, 171 and 172, 187, 190 and 207;

Reviewer 2 From now onward, no line specification in the Tables and Figures section:

1. In a number of Table captions and Figures, the unit of [mm] is missing: Table 1, Figs. 1, 2, 4, 5, 8 and 9.
2. All figures: Please remove the titles placed at the top of individual panels of figures. In some cases they contain redundant information (e.g., location Ihtiman), in other cases there are irrelevant (perhaps internal) codes and abbreviations of your simulation procedures. Please clarify these details rather in the figure captions.
3. Fig. 1 (top panel) seems a bit messy, it does not look like a box-and-whisker plot.
4. Fig. 1 (caption) . . . shown by month is only valid for the bottom panel.
5. Table 4 (caption) Separate gamma(left).
6. Fig. 6 Vertical (threshold) lines cannot be recognized.
7. Fig. 6 (legend) Change gpd to GP, in order to have consistent abbreviations throughout the manuscript.
8. Fig. 7 The dashed lines in the plots are hardly recognizable. Moreover, using grey colour for the individual values of rainfall is not too wise in the upper part of the plots: the highest extremes cannot be really seen, and this evokes the impression that the range of y-axes is not ideally set.

9. Fig. 8 Can you please specify what is the difference between the individual columns of the figure? Please do the same also in the text of the manuscript, lines 209-211.

Answer to Reviewer 2 From now onward, no line specification in the Tables and Figures section:

1. Fixed;
2. Fixed: The following abbreviation code "xxx yyy zz" can be seen at the individual panels of figures. Depending on the covariates included in the hybrid precipitation occurrence and intensity models and the chosen threshold. The occurrence and intensity model components are coded as follow: (1) 000 000 - without covariates; (2) 100 100 with a sine and cosine wave and previous day occurrence; (3) 110 110 with a sine and cosine wave, previous day occurrence and lag of NAO effect. Combining gamma and Weibull intensity precipitation models with GP distribution with threshold values 5mm, 10mm and 15 mm we get the hybrid gamma–GP and Weibull–GP distributions intensity precipitation models. The hybrid gamma–GP and Weibull–GP distributions precipitation amount models is defined occurrence precipitation models with
3. Fixed, see Fig. 1 of this report;
4. Fixed, see Fig. 1 of this report;
5. Fixed;
6. Fixed;
7. Fixed;
8. Fixed, see Fig. 3 of this report;

9. Fixed, the explanation can be found in NHESS Discuss., 2, 135, 2014, www.nat-hazards-earth-syst-sci-discuss.net/2/1/2014/

References

Chappell, N. A., Discenza, A. R., Tych, W., Whittaker, J. and Bidin, K.: Simulating hourly rainfall occurrence within an equatorial rainforest, Borneo Island, Hydrol. Sci. J., 54, 571–581, 2009.

Katz, R. W. and Parlange, M. B.: Generalizations of chain dependent processes: Application to hourly precipitation, Water Resour. Res. 31, 1331–1341, 1995.

Koutsoyiannis, D. and Onof, C.: Rainfall disaggregation using adjusting procedures on a Poisson cluster model, J. Hydrol., 246, 109–122, 2001.

Papalexiou, S. M., Koutsoyiannis, D. and Makropoulos, C.: How extreme is extreme? An assessment of daily rainfall distribution tails, Hydrol. Earth Syst. Sci. Discuss., 9(5), 5757–5778, doi:10.5194/hessd-9-5757-2012.

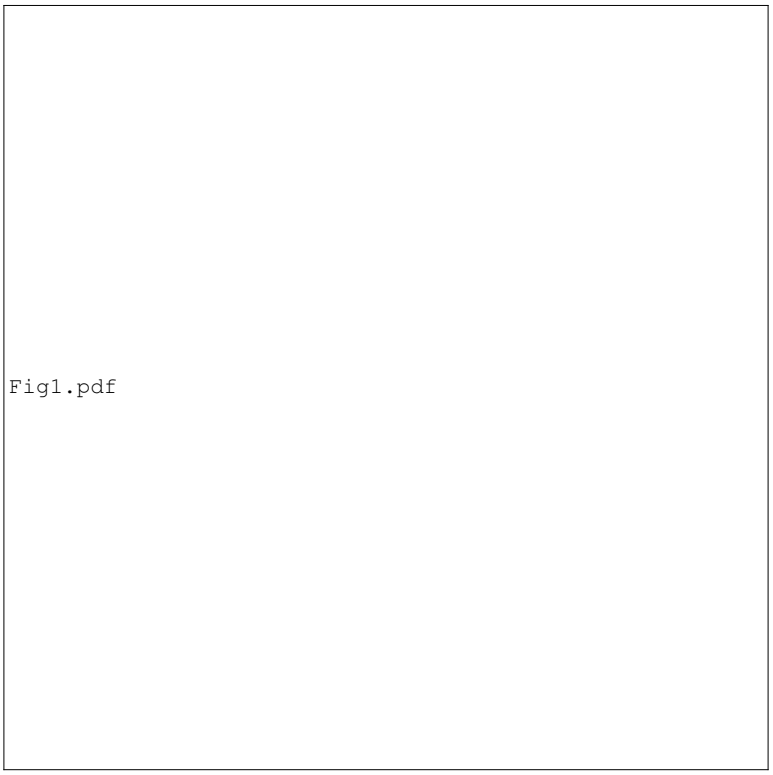


Fig. 1. Station Ihtiman: monthly (left) and annual (middle) box-and-whisker plots of daily precipitation data, and monthly (right) precipitation amounts, the extreme daily value 234 mm recorded on 5 August 2005 is dropped out from the data.

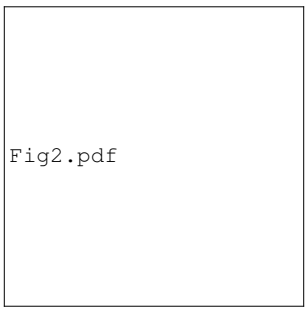


Fig. 2. Monthly box-plots of daily observed and simulated precipitation totals in log-scale for Ihtiman station. The simulated data are generated using seasonal model with a lagged NAO and occurrence covariates; the intensity component is based on: (i) hybrid gamma–GP (top left and middle plots) and Weibull–GP (bottom left and middle plots) distribution with threshold values 5 mm and 15 mm; (ii) standard gamma and Weibull distributions (right column plots).

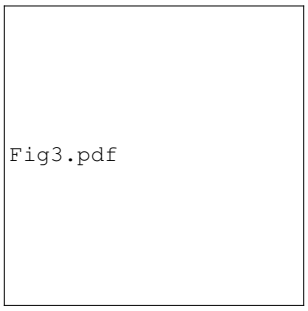


Fig. 3. Top row: High quantiles (95% and 99%) of fitted gamma (solid lines) and hybrid gamma–GP (dashed lines) distributions as functions of the day of the year for Ihtiman precipitation intensity with thresholds 5mm (left) and 15 mm (right). Lower line plots: The same as the top row but based on Weibull and hybrid Weibull–GP distributions. The extreme daily value 234 mm recorded on 5 August 2005 is dropped out from the data for better perception, it is labeled as a reminiscent triangle at the plots.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., 2, 1223, 2014.

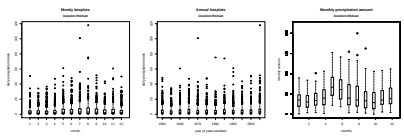
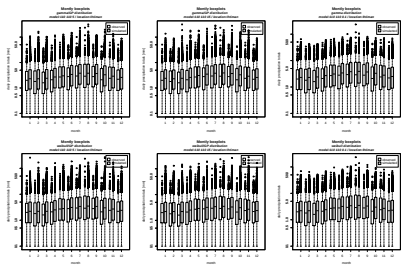
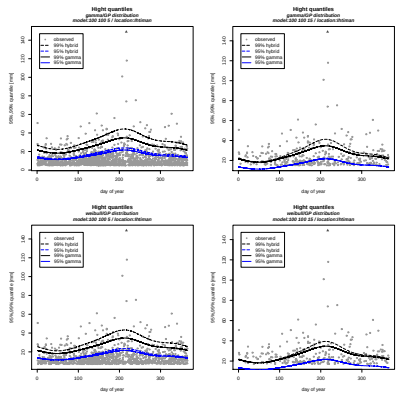


Fig. 4.



1

Fig. 5.


Hight quantiles
gamma/GP-distribution
model 100 100 0 / location:thinman
Hight quantiles
gamma/GP-distribution
model 100 100 10 / location:thinman
Hight quantiles
weibull/GP-distribution
model 100 100 0 / location:thinman
Hight quantiles
weibull/GP-distribution
model 100 100 10 / location:thinman
observed
99% hybrid
95% hybrid
99% gamma
95% gamma
99%,95% quantile [mm]
day of year

1

Fig. 6.