Hydrol. Earth Syst. Sci. Discuss., 5, S1608–S1616, 2008
www.hydrol-earth-syst-sci-discuss.net/5/S1608/2008/

Hydrology and Earth System Sciences Discussions

Interactive comment on "A space-time hybrid hourly rainfall model for derived flood frequency analysis" *by* U. Haberlandt et al.

Anonymous Referee #1

Received and published: 25 October 2008

by Bliefernicht, Beck and Haslauer

The paper by Haberlandt et al. presents a novel rainfall generation which is used to simulate hourly precipitation time series simultaneously at multiple sites. A detailed description of the methodology is given and the results are well presented. Furthermore, it is shown how the methodology can be applied in the field of derived flood frequency analysis. Nevertheless, there are still corrections necessary and questions left, which have to be comment by the authors.
The abstract is short and clearly formulated. It summarizes the methodology and the main results well, but the purpose of the rainfall model could be pointed out more clearly.

The introduction gives a good review of the existing methodologies and their problems. The objective of the study is also clearly formulated. In this chapter only minor corrections are considered necessary.

The description of the methodology (chapter 2) and the case study (chapter 3) were sometimes difficult to understand for the reviewers and should be carefully revised (see our specific comments). Specifically in this part of the paper the authors should double check for English wording and grammar. There are some missing words and a lot of German sounding phrases and it is fairly time-consuming to do an exhaustive correction (see technical corrections). The main issues of our revision for this chapter can be summarized as follows:

1. The models used for the single-site rainfall generation are explained in detail, but the time series generation itself remains fairly unclear. A few sentences describing the times series generation should be included.
2. What is the criterion for a rain event? The parameters of the distribution functions and the spatial dependence criteria depend strongly on this criterion.
3. It is quite difficult to understand which kind of observation station network is used for the different steps of the methodology, because the number of station seems to be inconsistent and the description of the station network is sometimes confusing.
4. What is the reason for the selection of the test sites? Is it possible to extend the catchment area of Selke subcatchment that the resampling can be also used for the winter period?.

The last chapter (chapter 4) is clear and in this part only some minor corrections are considered necessary. However, two-thirds of this chapter represents a rather a summary than a conclusion. We propose to either rename this chapter from *"conclusion"* to *"summary and conclusion"*, or to write some more concluding sentences.

In general, our revision might be in the some cases quite detailed. However, we would like to emphasize that this detailed revision was only possible because the authors described the methodology quite clearly. Finally, we hope that we can improve the quality of the work with our comments, questions and suggestions. The paper was

HESSD

5, S1608–S1616, 2008

Interactive Comment

revised by Claus Haslauer, Ferdinand Beck and Jan Bliefernicht.

Specific comments

Abstract:
Generally, the purpose of the model could be pointed out more clearly. In most cases, rainfall models are used to generate synthetic time-series in areas without available observations. Here, the aim is different: The model is used to extend given time-series over a long time period with the goal to estimate discharge annuities based on the hydrological model more accurately. From the title one might expect a spatial extrapolation to areas without observations which is not (yet) possible with this model.

Step1: rainfall generation pp. 2463/2464:
1: The paper makes clear how three probability density functions are estimated in what is referred to as *"external structure"*. These three functions are for the three parameters dsd, wsd, and wsi. It is also clear, how for the *"internal structure"* the parameter wsp is estimated based on the parameter wsi as well as on the peak of the precipitation event. Finally, the authors make it also clear how a bi-variate Copula is used to model the dependence between wsd and wsi. It is well described how the parameters are derived from the data but it remains unclear for the reviewers in which way the parameters are used in the time-series generation. Especially, how does the relation between wsd and wsi, which is described by a Copula, effect the generation of the random variables? There are three random variables dsd, wsd, and wsi, each of which is described by a distribution function. In an ordinary Monte-Carlo approach one would pick a random value from each of the distributions. But in that way the interdependence that was examined by the Copula can not be taken into account. We assume that a more sophisticated approach was used to condition the values of wsd and wsi on the Copula and that approach should be explained briefly in this section.
2: What is the meaning of *'identical distribution of the durations of the wet and dry*

HESSD
5, S1608–S1616, 2008

Interactive Comment

spells'? For the estimation of the wet and dry spell duration different distribution functions are used.
3: Instead of writing a symmetric *'double exponential function'* following formulation could be used:
'*To estimate the temporal distribution of the precipitation intensity during a wet spell a mixture of two exponential functions is chosen:*
$PI = wsp * exp[lambda(t - wspt)], for 0 <= t < wspt$
$PI = wsp * exp[-lambda(t - wspt)], for t > wspt$
The first function describes the increase of precipitation intensity from the beginning of the precipitation event to the wet spell peak time. The second function is used for the decrease of the precipitation intensity.'
4: independent events: The proposed method relies heavily on the distinction between times with precipitation and times without precipitation. This distinction raises two issues a.) What is the criterion for a rain event? In other words, what is the threshold amount of precipitation for the event to be included in category *'time with precipitation'*? b.) The authors use a separation time between precipitation events of one hour in the summer and two hours in the winter. Is this time sufficient for two events to be independent? Would not precipitation events really be independent from each other only if they originated from different circulation patterns? Independence would be a prerequisite for distribution functions to be established. If independence is an assumption that works for the proposed model, it should be stressed more clearly.

Step 2: resampling, spatial dependence properties pp. 2465/2466:
1: For counting the number of events, the question of what the threshold for distinguishing a rain-event and a non-rain event plays a critical role, and should be stressed when explaining Equation 7. Equation 7 can be also simplified to following form, which is easier to follow: $Pij \approx n11/(n01 + n10 + n11 + n00) = n11/n$ with n is total number of hours of the observation period and n11 is number of hours, if it rains at both station.
2: When explaining Equation 8, it should be made clearer, that this equation calculates

HESSD

5, S1608–S1616, 2008

the correlation only for those time-steps when it rains at both stations under consideration, simultaneously. Along the same lines, it could be explained with words quite simply that Equation 8 expresses the *'interrelation'* of both stations if it rains at both stations, and Equation 9 expresses the effect of station j on station i, irrespective if it rains at station j or not.

3: Generally, it is not distinguished between empirical and theoretical probabilities. For example, in Equation 7, Pij is the relative frequency and not the probability. We suggest that these should be distinguished more clearly. For example by using $Pij \approx n11/(n01 + n10 + n11 + n00) = n11/n$ instead of $Pij = n11/(n01 + n10 + n11 + n00) = n11/n$

Resampling algorithm pp. 2467/2468:

1: In steps 1 and 2 of the resampling algorith it is difficult to understand, which stations are included in the optimization at what stage. The formulation using language adapted from programming does not really help (e.g. *'with i=2,N'* and *'with for j=1,i-1'*)

2: The optimisation scheme is *'straight forward'* in the sense that the optimisation of station j does not effect the time-series on stations i=1 to j-1 anymore. The first station that is resampled has more "'freedom"' than the last one which depends on all of its antecessors. So the order of the stations surely affects the optimisation result. It would be useful to state clearly why one station is chosen as the reference station and in which way the order of the other stations is defined.

3: In step 8, it does not get clear why a restriction can be put in place why precipitation events should not be *'re-ordered'* if they are longer than a certain time dT away from the event under consideration.

4: In line 4 of page 2468 it is stated that the annealing *'ensures'* that the simulation does not stop *'at any local minimum'*. This is not proven and so the statement is problematic in this form. Experiments with mathematically defined problems show that annealing always finds a very good solution but often not the mathematically optimal one. This should also be seen in context with the comment on the ordering. One might come closer to the global optimum if all the stations were resampled simultaneously -

HESSD

5, S1608–S1616, 2008

Interactive Comment

giving the optimization system more "'freedom of choices"'.

Description of data; pp. 2468/2469:
1: It is quite difficult to follow the authors which set of observation stations is used for the different steps of the rainfall generation and for the hydrological model. For example:
a.) The number of hourly recording stations is inconsistent. The time series of 11 stations (8 *'summer'* +3 *'all year'*) are used for the calibration and for the alternating renewal model, but for the resampling only the time series of 10 stations are used (6 stations Selke catchment + 4 stations Holtemme catchment).
b.) In chapter 3.2 it is mentioned the first time, that in addition to the stations considered so far (see chapter 3.1), additional stations where only daily rainfall measurements were recorded are taken into account for the interpolation. Are these stations marked on a map? Are these stations equal to the 19 non-recording stations? It is not easier to say in the beginning of chapter 3.1, that *'19 stations with daily records and 23 stations with hourly records are used for the robust estimation of the spatial dependence criteria. For the rainfall generation and the hydrological modelling a subset of xx daily and yy hourly recording stations are selected. Only zz of the hourly recording stations are operated the entire year. The station network used for the rainfall generation and the hydrological modeling is shown on Fig. 3'*.
c.) It is also not clear, how the simulated precipitation series are used for the flood frequency analysis. In the calibration procedure the Thiessen polygon method is used for the interpolation of the areal precipitation. Is the same technique applied to interpolate the simulated series?
2: Is the study based on sub-catchments (shaded areas on Figure 3) of the Selke and Holtemme catchment? If so, then this should be indicated on the legend of Figure 3 and in the text (e.g. by including the gauging stations Mahndorf and Silberhütte).
3: It would be interesting to see the entire catchment of the Bode in comparison, as well as indicate somehow the significantly varying geodetic elevations.

HESSD

5, S1608–S1616, 2008

Interactive Comment

Chapter 3.3 p. 2472:
1: Why is the rainfall sum (Table 2, Winter, Station Wernigerode and Braunlage) estimated so poorly when compared to measurements? Is this an effect describing the dependence of the precipitation parameters using a Copula? If the parameters would have been drawn purely randomly based on the estimated distribution function, then the simulated precipitation sums should fit better to the observed precipitation sums, but the characteristics regarding time should fit worse. Is that right? Could you comment about the trade of and how to quantify a decision which way to go?
2: What is the reason for the selection of the test sites? Is it possible to extend the catchment area of Selke subcatchment that the resampling can be also used for the winter period?

Conclusions p. 2475:
In the conclusion (as well as in the abstract) the results of the investigation are summarized. One result is, that *'it is important to consider the same station network in calibration and application of the hydrological model.'* We cannot really find the link between this statement and the results presented in chapter 3

Technical corrections

p.2460/5: A *'hybrid two step procedure'* seems to be a tautology - *'two step procedure'* seems sufficient
p.2460/21: To *'consider'* the same rainfall station network, or to actually use the same network?
p.2461/4: ...in this approach ...
p.2461/9: ... such kind of hydrological ...
p.2461/12: Replace ... *'utilized with increasing frequency'* with *'used more and more*

HESSD
5, S1608–S1616, 2008

commonly'
p.2461/20: *'applied to'* instead of *'applied for'*
p.2461/22: ...catchment. The classical approach of time series models is less important *'here'* and more suitable due to *'difficulties'* with modeling ...
p.2462/4: ... arises from parameter estimation ...
p.2462/7: ... on climatic conditions...
p.2463/5: ... and wet-spell-amount (wsa) or the wet-spell-intensity (wsi) (wet-spell-amount divided by wet-spell-duration)...
p.2463/10: ... the precipitation amount or the intensity can also be modeled ...
p.2463/25: *'Thus, a 2-copula'* instead of *'the 2-copula'*
p.2463/25: Be careful with wording! A n-copula describes the dependence between n variables, not between n marginal distributions. The dependence is described with a copula no matter what the marginal distributions might be.
p.2465/5: ...expected. The internal rainfall...
p.2465/22: ... can be calculated from the hourly rainfall time series z.
and the formulation in line 25 *'where z is the hourly rainfall intensity'* can be deleted
p.2467/4: ...stations j with j=1 ...
p.2467/17: ... with the probability π
p.2467/18: π =(...)= with Ta as annealing temperature.
p.2468/4:... local minimum...
p.2468/5: Ta is not necessary
p.2468/6: The lower Ta the less likely ...
p.2468/12: The second step of the rainfall generation
p.2469/19: The model is operated on an hourly time step.
p.2469/20: Potential evapotranspiration is aggregated to monthly
p.2470/1: ...the sparse network of recording hourly rainfall stations daily ...
p.2470/23-24: ... are estimated from all hourly time series for the winter and summer season.
p.2470/24: Several realizations of an hourly times series are simulated. The simulation

HESSD

5, S1608–S1616, 2008

period of one realization is 100 years.
p.2471/19-20: The relation between the spatial dependence criteria defined in equ. 7 to 9 and the separation distance
p.2472/1-2: It is better to use *'catchment'* than *'basin'* for both test sites. In our opinion the word basin belong to a larger catchment like Danube basin, Neckar basin, ...
p.2472/5-6: The results improve with fewer stations
p.2472/22: might need a calibration
p.2473/3: ...synthetic rainfall repetitions of
p.2473/7-19: poor English: this part of the text should be checked again
Fig. 6, 9 and 10: The x-axis (gumbel reduced variate) is not necessary. The return period should be the x-axis.
Fig. 1: the indices of the third wet spell must be changed from wsd2 to wsd3
Tab. 2: Is that the average number of events?

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 5, 2459, 2008.