A seasonal mixed-POT model to estimate high flood quantiles from different event types and seasons

Svenja Fischer

No. 1/2018

www.spate-floods.com
A seasonal mixed-POT model to estimate high flood quantiles from different event types and seasons

Svenja Fischer

Institute of Hydrology, Ruhr-University Bochum, 44801 Bochum, Germany

ABSTRACT

Flood events can be caused by several different meteorological circumstances. For example, heavy rain events often lead to short flood events with high peaks, whereas snowmelt normally results in events of very long duration with a high volume. Both event types have to be considered in the design of flood protection systems. Unfortunately, all these different event types are often included in annual maximum series leading to inhomogeneous samples. Moreover, certain event types are underrepresented in the annual maximum series. This is especially unsatisfactory if the most extreme events result from such an event type. Therefore, monthly maximum data are used to enlarge the information spectrum on the different event types. Of course, not all events can be included in the flood statistics because not every monthly maximum can be declared as a flood. To take this into account, a mixture Peak-over-threshold (POT) model is applied, with thresholds specifying flood events of several types that occur in a season of the year. This model is then extended to cover the seasonal type of the data. The applicability is shown in a German case study, where the impact of the single event types in different parts of a year is evaluated.

KEYWORDS

mixed-POT model; flood estimation; event type classification; seasonal statistic

1. Introduction

The estimation of high quantiles is essential in flood statistics since these quantiles are used for the calculation of e.g. flood protection systems. Due to federal regulations, quantiles with a return period of 1000 years or even 10,000 years are needed, which equals the 99.9%- or 99.99%-quantile. Mostly, the annual maximum series of discharges at a gauge form the basis for this estimation. Unfortunately, many discharges series consist of observation periods of less than 30 years. Series with 100 years of observation a rather rare in Germany and also for most of the other countries. Therefore, the estimation of these high quantiles can only by done by extrapolation, for which statistical methods are used. In the case of the annual maximum series (AMS), in Germany but also in other countries like the United States often the Generalized Extreme Value distribution (GEV) is fitted to the data (e.g. [17], [22]), which has theoretical validity due to the Fisher-Tippet-Theorem ([8]). A disadvantage of this approach is the very limited number of considered events. In fact, often more than one flood event can occur in a year but by using the annual maximum series only the largest event is
considered. If all flood events shall be considered in the model, mostly a peak-over-threshold (POT) approach on the monthly maximum discharges is used, where all observations above a certain threshold are included in the estimation ([20]). For these events it can be shown that (under certain assumptions) they converge in distribution to the Generalized Pareto (GPD) distribution ([3]; [18]). To ensure that only flood events are considered in the model and since not every monthly maximum discharge belongs to a flood event, the threshold is chosen as the value, for which an event can be called flood. This choice depends much on the local circumstances but is crucial for the POT model ([4]). In Germany, often triple the long-term mean runoff is used ([6]).

The advantage of the POT-model compared to the classical AMS with fitted GEV is the extension of the considered information. Whereas the AMS limits the number of events to one per year, the POT-approach uses all large flood events over a threshold. Therefore, extraordinary large events can be put into a larger context ([6]).

What remains a problem are the different geneses of the flood events. Flood events can have very different geneses such as short heavy rain events (convective) or rain events of long duration and also snowmelting processes can have a large impact on the flood events.

This is the reason why many models include the approach of seasonality. Different geneses can lead to very inhomogeneous samples and the assumption of independent, identically distributed data is violated ([11]). Whereas the geneses-based distinction of maxima is well-known in the meteorological context (e.g. [21] or [10]), only few models of this kind are used in the context of seasonal discharge maxima ([24]). A subdivision of the year into two periods (winter and summer) is for example proposed by [27] and [24]. For this distinction, the common assumption then is that the respective samples are identically distributed ([1]). For this model two different distributions are fitted to the seasons and then combined in an additive mixture model. But also statistical approaches of a subdivision into homogeneous samples exist (e.g. [5]). Nevertheless, moth of these seasonal models focus on different time periods in the year but do not consider the different geneses within a season. Also within a season very different rain events (heavy rain or frontal rain) or even snow events can occur. This results in very different shapes of the hydrographs, which can have a high peak but low volume (mainly heavy rain events), a moderate peak with high volume or both a high peak with high volume (mainly rain-on-snow events). The different event types also depend much on the catchment characteristics. The higher the elevation of a catchment the shorter and larger (concerning the relative peak discharge) are flood events caused by short and heavy rain events. All these different kinds of flood events can nevertheless stress the flood protection systems heavily and have to be included in the model. Often, individual event types dominate regional flood conditions ([16]). Moreover, because of the different types of flood events, the homogeneity assumption, which is essential for the models described above, is violated. Thus, a mixture model is needed that differentiates the event types and combines them in a statistical model. For a seasonal distinction in combination with different flood types based on winter and summer annual maxima, where the classification is done via the flood time scale, such a model is proposed by [7]. There, a nested maximum mixing model based on winter and summer events as e.g. given in [25] is used, where also the different event types are modelled with the mixing model within the single seasons. The advantage of this model is the separate consideration of the flood types such that the dominating event type can be identified. By using the maximum approach instead of the additive approach of e.g. [24], this information then can be used estimate high quantiles according to this dominating type instead of using a kind of average. Though, for this model the
problem occurs that only one event per season is observed and the remaining events have to be reconstructed statistically.

Here, we want to develop a new model that combines the POT approach with the mixing model. For this purpose, every season is modelled with a mixed-POT approach and then combined in a maximum mixing model to obtain estimations for annual maximum discharges. The mixed-POT model is an extension of the classical POT model and has been proposed by [26]. The combination of POT-model and the event-type distinction shall improve the classical seasonal models and combine the benefits of the well known flood frequency models. Moreover, with this approach, not only estimates for the annual maximum series can be calculated but it has the benefit that also estimations for single seasons can be given. This is important for many practical applications like the operation of flood protection systems and the different seasonal geneses like snowmelt can be taken into account.

This work is organised as follows: in Section 2 the basic methods and models are introduced; in Section 3 the given data are presented and the seasonal mixed-POT is applied, the results are discussed and compared with classical models in flood statistics; Section 4 summarises the results.

2. The seasonal mixed-POT model

Motivated by the need of developing a model that takes into account the different event types as well as the different seasons that have influence on the genesis and magnitude of a flood event, we propose a seasonal mixed-POT model. This model does not only use all monthly maximum discharges and therefore the full spectrum of information on flood peak discharges but it has the advantage that also flood quantiles for single seasons can be estimated under consideration of the frequency of the different event types in the respective season. This can be important for example for the operation of a dam. We want to introduce the model in two steps. First, the mixed-POT model for a single season is defined. In a second step, the single seasonal models are combined in a maximum mixing model to obtain a model for the annual maximum series.

For the distinction and classification of the events we use the flood timescale method proposed by [7]. Here, every monthly maximum discharge is classified as one of three possible event types. The seasons are defined according to the frequency of the single event types. For one basin naturally a homogeneous distinction of seasons occurs, with only small differences for example caused by earlier snowmelt in catchments with higher elevation.

2.1. The mixed-POT model

When modelling the exceedances of a threshold $u$ in a POT model for a sample $Y_1, \ldots, Y_n$ of a series of i.i.d. random variables, the probability distribution of the exceedances $(Y - u)$ conditional on $Y > u$ can be shown to follow the GPD approximately ([3]; [18]). The distribution function of the GPD is given by

$$G(x; u) = \begin{cases} 1 - \left(1 + \kappa \left(\frac{x-u}{\sigma}\right)\right)^{-\frac{1}{\kappa}}, & \kappa \neq 0 \\ 1 - \exp\left(-\frac{x-u}{\sigma}\right), & \kappa = 0 \end{cases}$$
with threshold \( u \), scale-parameter \( \beta > 0 \) and support \( x \geq u \) for \( \kappa \geq 0 \) and \( u \leq x \leq u - \beta / \kappa \) for \( \kappa < 0 \). The special case \( \kappa = 0 \) corresponds to the exponential distribution.

Unfortunately, this model does not lead to the desired quantiles in flood statistics, since it has the wrong time resolution. We are interested in annualities and related quantiles, such that the probability of exceedance of the threshold within one year is needed to obtain a distribution function \( F \) to calculate these from the monthly maximum discharge POT approach ([6]). Often, the probability of exceeding a threshold \( u \) in a period \( d \) (here \( d = 12 \) months) is estimated with a Poisson distribution ([20];[23]), but also other distributions like the Binomial distribution are possible ([6]). The distribution function \( F \) than can be expressed as

\[
F(x) = \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} (G(x; u))^k,
\]

where \( G \) is the distribution function of the exceedances of the threshold \( u_i \), chosen as GPD, and \( \lambda = \frac{1}{n_i/d} \sum_{i=1}^{n_i} 1_{[Y_i > u]} \) is the parameter of the Poisson distribution representing mean and variance.

A generalisation of this model has been proposed by [26], where non-identically distributed random variables are considered. From now on suppose that we have \( n_s \) seasons and \( n_t \) event types. This leads to \( n_t \cdot n_s \) samples \( X_{i;j} = \left( X_{i;j}^{(1)}, \ldots, X_{i;j}^{(n_{i;j})} \right) \), \( i = 1, \ldots, n_t, \) \( j = 1, \ldots, n_s \), where \( X_{i;j} \) is the sample of events belonging to type \( i \) and season \( j \) with sample size \( n_{i;j} \). Suppose that the events (above and below the threshold) of every season \( j \) follow a mixture distribution

\[
F_j(x) = \sum_{i=1}^{n_t} \omega_{i;j} F_{i;j}(x),
\]

where \( \omega_{i;j} \) is the probability that the event type \( i \) occurs in season \( j \) with \( \sum_{i=1}^{n_t} \omega_{i;j} = 1 \) and \( F_{i;j} \) is the distribution function of the events of type \( i \) in season \( j \). The proposed mixed-POT distribution function for the annual exceedances in season \( j \) is then given by

\[
H_j(x) = 1 - \sum_{i=1}^{n_t} (1 - G_{i;j}(x; u_{i;j}))(1 - F_{i;j}(u_{i;j})) \omega_{i;j}.
\]

This model \( H_j \) estimates the joint distribution of events under consideration of the frequencies of events of certain types.

The distribution function consists of two terms. First, the distribution of the exceedences of a threshold \( u_{i;j} \), \( G_{i;j} \), is calculated for every event type \( i \) in the considered season \( j \). Here again, \( G_{i;j} \) is chosen as the GPD distribution. Please note that a high flexibility is obtained, since the threshold \( u_{i;j} \) can be chosen differently for every event type and season. This is important, since different seasons can lead to different flood events of the same event type. Secondly, the exceedance probability of the threshold \( u_{i;j} \) of the whole sample of the event type (events above and below the threshold), \( 1 - F_{i;j}(u_{i;j}) \), is considered and weighted with \( \omega_{i;j} \). This corresponds to the Poisson distribution in eq. (1).
The distribution functions $F_{i,j}$ are chosen as GEV distribution, since we consider block maxima and thus the Fisher-Tippet Theorem holds (see above). The GEV distribution has the distribution function

$$F(x) = \begin{cases} 
\exp \left( - \left(1 + \xi \left( \frac{x-u}{\sigma} \right) \right) \right), & \xi \neq 0 \\
\exp \left( - \exp \left( - \frac{x-u}{\sigma} \right) \right), & \xi = 0
\end{cases}$$

for $1 + \xi (x - \mu) / \sigma > 0$, where $\xi \in \mathbb{R}$ is the shape parameter, $\sigma > 0$ is the scale parameter and $\mu \in \mathbb{R}$ is the location parameter. Although all distribution functions $F_{i,j}$ respectively $G_{i,j}$ belong to the same family of distributions, their parameters can be chosen very differently to take into account the different nature of the single event types.

In practice, the weights $\omega_{i,j}$ are estimated by the empirical frequency of the events of type $i$ in season $j$. The thresholds $u_{i,j}$ are chosen according to the frequency of months of the single groups. In detail, for every event type $i$ in season $j$ we calculate the relative frequency for every month of the season. This frequency is multiplied with the long-run mean discharge of this respective month. Then, the results for all months are added and multiplied with the factor 3 (see above).

We are aware that the choice of the threshold has an impact on the estimated results and that there exist several methods to define this threshold. The threshold value should be chosen, such that the model assumptions are valid but also as low as possible ([12], [2]).

In hydrological practice, often a mean number of events per year is defined, according to which a threshold is calculated ([19]). But this methods do not take into account the variability of floods between different years caused by the different climatic conditions and would somehow contradict the idea of the mixing model. Also the use of the minimum annual maximum ([13]) can not be applied here as a threshold, since we want to obtain variable thresholds for the different event types and seasons. Of course, also statistical methods can be used to derive the thresholds. For example, often the mean residual life plot is used as a graphical tool to define the threshold. But as already [26] pointed out, the choice of the threshold using this plot needs some statistical expert knowledge and thus is not suitable for practitioners. These authors then propose a method based on the Goodness-of-Fit of the mixing models to the empirical distribution. In our case this would mean that we would test the Goodness-of-Fit for every season. This procedure has two disadvantages. First, as [26] point out, a non-parametric test is needed. They choose the Kolmogorov-Smirnov test for this purpose. But, as stated e.g. in [14], the theoretical distribution of this test is no longer valid if the parameters of the distribution are estimated. Instead, one would have to simulate the p-values. The same holds true for the Anderson-Darling test. Due to the many parameters of our model, this method is not appropriate. Moreover, the use of the empirical distribution can lead to high uncertainty, because some event types are less represented in certain seasons than others. For example, there are only a few events with long duration and small peak in the summer month. This was the reason for us to use the expert knowledge in hydrology to define the threshold. The definition of a flood as an event three times higher than the average flow is a well known definition in hydrology. It corresponds with the idea of [15] to use statistical characteristics such as the mean or variance to define the threshold.

As a result we have three models taking into account three event types respectively, one for each season. Probabilities for single event types in one season $j$ can simply by
obtained by using the partial POT-model of the mixing distribution $H_j$.

### 2.2. The seasonal extension of the model

To obtain a model for the whole year and therefore be able to estimate quantiles for given annualities, the different seasonal models have to be combined. Since we are interested in the annual maximum event, which is the maximum event of the $n_s$ seasons, we can use the maximum mixing approach similar to [7]. If the annual maximum event in a year does not exceed a certain value, then this holds also for every event in every season. Thus, a multiplicative maximum model, representing the logical AND-structure, can be chosen. In detail, the distribution function $F_a$ of the annual maximum event $X_a = \max_{1 \leq j \leq n_s} \left( \max_{1 \leq i \leq n_t} (X_{i,j}^{(1)}, \ldots, X_{i,j}^{(n_{i,j}))}) \right)$ is given by

$$F_a(x) = \prod_{j=1}^{n_s} H_j(x).$$

This model now allows us to calculate events for given annualities, this means quantiles of the distribution function $F_a$. Please note that because of the complex structure the quantiles can only be calculated numerically and no analytical solution can be given. Still, we are able to assess the influence of single seasons on the different quantiles.

As an overview, the procedure of the model is illustrated in the diagram in Figure 1.

### 3. Application

In this analysis we use the monthly maximum discharges of 19 gauges located in the Harz region in Central Germany (Fig. 2). The given periods of observation lie between 35 and 85 years with catchment sizes between 21.4 and 456 km$^3$. Besides the peak $Q_p$ [m$^3$/s] of every event also the volume $Vol$ [Mio. m$^3$] and the baseflow are available. The flood timescale ([9])

$$TQ = \frac{Q_p}{Vol},$$

which is the quotient of the peak runoff and the volume, is used to classify the monthly maximum discharges into $n_t = 3$ event types using the method proposed by [7] (Fig. 3). According to the frequency of the events in the single months then $n_s = 3$ seasons are defined (Fig. 4). For the Harz regions three seasons proved to be sufficient for all gauges. Except from small deviations at the beginning or end of a season, which is due to the different elevations and locations of the gauges, we are able to define the three seasons

- summer: May/June-September/October
- winter: October/November-January/February
- spring: February/March- April/May.
Figure 1. Schematic illustration of the seasonal mixed-POT model. The distinct estimation for every event type and every season based on the sample $X_{i,j}$ is shown in a general formula to keep the diagram readable. The methods to obtain the estimates are given in blank boxes.
Figure 2. Location of the 19 considered gauges in the Harz region in Central Germany.

Figure 3. Distinction of flood events into different event types with the direct TQ approach for the Steinerne Renne/Holtemme gauge.

Figure 4. Seasonal deviation of the monthly maximum discharges according to their proportion in the months for the Steinerne Renne/Holtemme gauge.
With these considerations, the models for the three seasons can be fitted.

First, a detailed look on the results is given for one gauge, Steinerne Renne at the river Holtemme. The single seasonal distributions together with the distributions of the single event types for the summer as well as the winter season are given in Figs. 5 and 6. The empirical quantiles for the observed discharges are shown as dots. Nevertheless, since the sample sizes of some event types in certain seasons can be very small, these have to be considered carefully.

It can be seen for the single event types that the estimated seasonal distributions fit well to the empirical distributions. Only the largest event for some event types deviates from the fitted distribution. The importance of the mixture distribution as well as the seasonal differentiation becomes obvious when having a look at the single seasonal distributions. For season 1 (summer) certainly the events of type 1 with high peaks and small volumes are the most relevant and dominate the mixing distribution for all estimated quantiles. This can be explained by their geneses, which are mostly heavy rain events of short duration. These typically occur in the summer months in Germany, corresponding with thunderstorms. The catchment of the Steinerne Renne has a large slope and thus reacts faster, especially if the antecedent moisture is low, which is the case for the summer month. For the winter events also the events of type 1 dominate most of the mixing distribution, although the distribution results in much smaller quantiles. Nevertheless, the high quantiles with return period of 1000 years and more are dominated by events of type 3. This type corresponds to events with small peak and very high volume. Typically, these are snowmelt and rain-on-snow events, where water is induced into the system over a long time the antecedent moisture is high. This is a phenomenon only occurring in the winter months and early spring. In general, the winter floods mainly consist of large amount of small floods with only a few large events. These two examples emphasise the different impact of the event types in the different seasons, mainly resulting from the events’ geneses. A mixture of these in an annual maximum series would not only lead to an inhomogeneous sample but also the different impact factors within a year would vanish. These impact factors can be very important for flood protection, though.

The combination of the single seasons than can be used to obtain flood estimates for certain annualities, which is the main goal of flood frequency analysis. If we compare the seasonal mixed-POT model with the classical annual maximum approach, where a GEV is fitted to the annual maximum series (AMS), the difference between the two
models especially for high quantiles becomes obvious (Fig. 7). Whereas the annual maximum series seeks a compensation between all seasons, the seasonal mixed-POT approach is influenced mostly by the dominating season (and event type). Additionally, we want to compare the results with quantiles estimated by the common POT-approach (eq. 1) applied to the monthly maximum discharges with a threshold of three times the mean discharge. This choice of the threshold makes the approach comparable to the seasonal mixed-POT model. The quantiles of the POT are much lower than the ones of the other two approaches. Moreover, both AMS and POT deviate much from the empirical probabilities for annualities from 20 years on. The seasonal mixed-POT approach instead overestimates the quantiles for annualities of 3-8 years but fits better for quantiles for annualities from 10 years on. Nevertheless, the largest event, corresponding to the extraordinary large event of the year 1994 in this region, is not represented well, too. The results show, that both, the AMS and the POT approach are much influenced by the many small events contained in the sample. In the annual maximum series, a large amount of rather small events from the winter season is mixed up with only a few very large events from summer. The influence of the large events in the distribution fit is thus reduced. For the POT-approach this problem worsens, since the number of small events included in the sample is increased in this case. Only the seasonal mixed-POT approach is able to represent the large flood events, since it is based on the maximum of all seasons and a mixture model of the event types. Thus, the different geneses and seasons are not simply put together in one sample, but the most dominating one is used to determine the quantiles.

This behaviour can be detected for the whole Harz region. The results for the remaining 18 gauges in the Harz region are given in Figure 9. For all gauges except of one (Elend) it can be seen that the seasonal mixed-POT model delivers higher quantiles for annualities from 100 years on than the AMS approach. For the majority
of the gauges (11) the quantiles estimated with the classical POT-approach lie between the AMS-quantiles and that of the mixed-POT and have a very similar slope as the mixed-POT ones. This can be explained by the events that are taken into account additionally in the POT-approach compared to the AMS. Whereas in the AMS only one event per year is considered and every other event in the same year is neglected, the POT approach only considers events of a certain level (threshold). For most of the gauges this means, that additionally more large events are included in the sample, which would be "overlaid" by large annual maxima in the AMS. Only a few gauges like the example of Steinerne Renne before do not seem to have additional large events that could be included in the sample. Instead, the long-time mean of the discharges is so small that many small events are added to the sample with the POT-approach. Again, this large amount of small events reduces the impact of the very large events to the estimation. Hence, the use of the POT-approach can even worsen the underestimation of large events.

The Goodness-of-Fit of all three approaches is statistically measured with the RMSE applied to the estimated quantiles and the empirical quantiles of the annual maximum discharges. The results are given in Figure 8. As can be seen, for most of the considered gauges (11) the seasonal mixed-POT approach delivers the lowest RMSE values and the second best for seven gauges, where it is very close to the best fitting model. The variance of the RMSE is much smaller for the seasonal mixed-POT model than for the GEV-model and comparable to that of the POT-approach, whereas the mean of the RMSE for all gauges is lowest. Thus we can conclude that the seasonal mixed-POT model results in the overall lowest RMSEs and gives good results for all gauges. The AMS approach instead shows rather large RMSEs for several gauges and thus cannot be assumed to be fitting. The POT-approach is a note-worthy alternative but nevertheless is outperformed for all but four gauges.

Comparing all the approaches it becomes obvious, that the quantiles corresponding
to small annualities are represented best by the POT and AMS approach. This can be explained by the large amount of small event included in the samples that are the basis of this estimation. Nevertheless, the results for the seasonal mixed-POT approach do not deviate much from these quantiles. But for large annualities the mixed-POT approach delivers much better estimates than the other two approaches.

This can be explained, since the right tail of the seasonal mixed-POT approach is almost always dominated by one distinct season. This is the season where the largest events occur. The concept of the seasonal mixture model, which is based on the maximum principle, is constructed in a way, such that always the dominating distribution is the relevant for the respective quantile. This means, a different season and event type can be relevant for small quantiles and for large quantiles. Which season is the relevant one for the large quantiles depends much on the catchment characteristics. Especially height and slope have a large influence on this. Additionally, the dominating season is again defined by a distinct event type for almost all gauges.

If we apply the model to the whole Harz region, we can detect a pattern in the influence of the single seasons. Here, we want to lay the focus on extreme quantiles, which are the most important for assessing design floods. If we consider the mean relation of the HQ(1000) of one event type to the other two types for each season separately, local coherences of the dominating event type for certain seasons can be seen (Figs. 10 and 11). Especially in spring the catchments located close to the Brocken mountain, except of the two gauges lying at the head of the Bode basin, are dominated by events of type 2. These are floods with moderate peaks and high volumes. These events are mainly caused by snowmelt and rain-on-snow events and thus dominate in those catchments, that have a high elevation. On the other hand, in the summer season almost all catchments have dominating events of type 1, indicating high peaks and small volumes. This corresponds to heavy rain events mainly occurring in the summer period. Only the southern and northern catchments have a dominating type 2. The southern catchments are located in the wind-shadow of the Brocken mountain, though, and thus almost no heavy rain events occur. For the winter season no distinct pattern can be detected.

4. Conclusion

For the estimation of design floods, which are quantiles of high probability, a simple fitting of distribution to the annual series is not sufficient. Instead, the different geneses of flood events have to be taken into account. Moreover, not only maximum values in a year are of interest. These needs in flood statistics can be faced by the seasonal mixed-POT model which is proposed here. By a division of the events in different seasons and event types, the different origins and times of occurrence can be analysed and than can be taken into account in the assessment of design floods. Moreover, the statistical assumption of homogeneous (sub)samples is fulfilled. The single event types are reasoned by the different precipitation conditions but also by the location of the catchment. For example, in the spring season rain-on-snow floods lead to the highest discharges in small catchments of high elevation. This information can be used to develop a season- and precondition-specific handling of flood events. Whereas the classical approach based on the annual maximum series tends to underestimate the most extreme events if they belong to a rare event type, the seasonal mixed-POT model takes the different impacts of event types on the different quantiles into account. Thus, the seasonal mixed-POT model is advantageous when the annual
Figure 9. Comparison of the seasonal mixed-POT model with the AMS approach and influence of the single seasonal models for all gauges in the Harz region with empirical observations of the annual maximum discharges (black dots).
maximum series of flood events is very inhomogeneous and the single event types are not distributed equally in the sample. For example, in Germany often only a few very large summer floods generated by heavy rain are included in the AMS whereas we can find a large amount of rather small winter floods. The use of the AMS approach could lead to serious underestimation of design events since the compensation reduces the impact of the most extreme events in the sample. Also the classical POT-model can lead to an underestimation as was shown in the results. Due to the large amount of small events in the monthly maximum series, here again the impact of large events is reduced in the estimation. Nevertheless, as long as the threshold is not too small, which is the case if the sample consists of many small events and only a few larger events, the POT-approach delivers better results than the AMS. In general, the results of this paper show that the AMS as well as the POT-approach are advantageous for the estimation of quantiles for annualities up to 10 years (20-30% of the sample length). For large annualities, which are the important ones for the assessment of flood protection systems and many other hydrological applications, these approaches tend to underestimate the quantiles. The seasonal mixed-POT model instead fits better to the observed data for these high quantiles and also deviates not much from the observation for small quantiles. This is emphasised by the results of the RMSEs, where the three models are compared to the empirical return periods of the annual maxima series. For almost all gauges the seasonal mixed-POT approach delivers very good results, resulting in small RMSEs with low variance between the gauges.

The distinction into homogeneous subsamples and the classification of dominating seasons and event types can also be used to define classes for regionalisation. For regionalisation, where we want to estimate floods for ungauged basins, a homogeneous group of gauges has to be found, that can be used to represent characteristics of the ungauged basin that are not catchment specific. Here, the coherences between dominating flood types in certain seasons can be used to lead to conclusions on the catchment behaviour.

The model in general cannot only be of interest in the context of hydrology but also whenever a seasonal and mixed sample has to be considered. For example, the normalised difference vegetation index (NDVI), which is a measure for vegetation activity, depends as well on seasonal (climate) conditions as on the type of vegetation. The seasonal mixed-POT model can used to investigate the impact of season and plant...
Funding

This work was supported by the DFG research project SFB 823, ‘Statistic of non-linear dynamical systems’ and the DFG research group FOR 2416 “Space-Time Dynamics of Extreme Floods (SPATE)”.

References


