Hydrology and Earth System Sciences

A space-time hybrid hourly rainfall model for derived flood frequency analysis

U. Haberlandt[1], A.-D. Ebner von Eschenbach[2], and I. Buchwald[1]

[1]Institute of Water Resources Management, Hydrology and Agricultural Hydraulic Engineering, Leibniz University of Hannover, Hannover, Germany
[2]Federal Institute of Hydrology, Koblenz, Germany

Abstract. For derived flood frequency analysis based on hydrological modelling long continuous precipitation time series with high temporal resolution are needed. Often, the observation network with recording rainfall gauges is poor, especially regarding the limited length of the available rainfall time series. Stochastic precipitation synthesis is a good alternative either to extend or to regionalise rainfall series to provide adequate input for long-term rainfall-runoff modelling with subsequent estimation of design floods. Here, a new two step procedure for stochastic synthesis of continuous hourly space-time rainfall is proposed and tested for the extension of short observed precipitation time series.

First, a single-site alternating renewal model is presented to simulate independent hourly precipitation time series for several locations. The alternating renewal model describes wet spell durations, dry spell durations and wet spell intensities using univariate frequency distributions separately for two seasons. The dependence between wet spell intensity and duration is accounted for by 2-copulas. For disaggregation of the wet spells into hourly intensities a predefined profile is used. In the second step a multi-site resampling procedure is applied on the synthetic point rainfall event series to reproduce the spatial dependence structure of rainfall. Resampling is carried out successively on all synthetic event series using simulated annealing with an objective function considering three bivariate spatial rainfall characteristics. In a case study synthetic precipitation is generated for some locations with short observation records in two mesoscale catchments of the Bode river basin located in northern Germany. The synthetic rainfall data are then applied for derived

flood frequency analysis using the hydrological model HEC-HMS. The results show good performance in reproducing average and extreme rainfall characteristics as well as in reproducing observed flood frequencies. The presented model has the potential to be used for ungauged locations through regionalisation of the model parameters.

1 Introduction

Efficient flood protection measures require a good knowledge about flood frequencies at different points in a catchment. The classical approach to obtain design flows is to carry out local or regional flood frequency analysis using long records of observed discharge data (e.g. Hosking and Wallis, 1997; Stedinger et al., 1993). If flow data are not available or if impacts of climate or land use change are to be investigated rainfall-runoff modelling is a good alternative either using event based or continuous simulation. Disadvantage of the event based simulation is the required assumption about equal return periods for the design storm and the resulting design flood. This is usually not given considering e.g. the initial soil moisture conditions in the catchment which may lead to different floods for the same storm. With continuous rainfall-runoff simulation this problem can be avoided and the design flood is derived by flood frequency analysis of long series of simulated flows. However, such kind of hydrological modelling requires long continuous rainfall series with high temporal and sufficient spatial resolution. Given the restricted availability of those observed data, synthetic precipitation is used more and more commonly for this purpose (Aronica and Candela, 2007; Blazkova and Beven, 2004; Cameron et al., 1999; Moretti and Montanari, 2008).

Correspondence to: U. Haberlandt
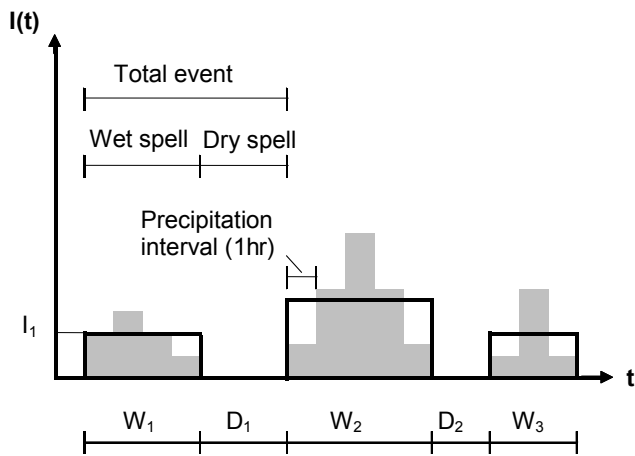(haberlandt@iww.uni-hannover.de)

Fig. 1. Scheme of the precipitation event process.

Over recent years, several stochastic precipitation models for short time step rainfall have been proposed. To the early approaches belong the alternating renewal models which are based on event series of wet-dry spells (Acreman, 1990; Grace and Eagleson, 1966; Haberlandt, 1998; Pegram and Clothier, 2001). Those models have a simple structure, the estimation of parameters from point observations is straightforward and the models can easily be applied to rainfall synthesis at single locations. However, they are usually not able to simulate space-time rainfall for several stations in a catchment. The classical approach of time series models is less important here and more suitable for daily rainfall due to difficulties with modelling the high intermittence of short time step rainfall and the large number of required parameters (Haan et al., 1976; Wilks, 1998). Advanced approaches for rainfall modelling with sub-daily time steps are the point process models like Neyman-Scott or Bartlett-Lewis rectangular pulse models (Cowpertwait, 2006; Onof et al., 2000; Rodríguez-Iturbe et al., 1987), which can also be extended to simulate space-time rainfall. They are based on the physical structure of the rainfall process and describe probabilistically arrival times of storms and cells within storms as well as cell intensities and durations. Often they assume independence between cell intensities and durations. One recent exception is described in Evin and Favre (2008) where cubic copulas are used to model this relationship. A difficulty with point process models arises from parameter estimation which requires usually observed radar data or relies on optimisation which may lead to some kind of parameter equifinality (Beven and Freer, 2001). The latter makes a conditioning of the model parameters on climatic conditions difficult. Other models for synthetic rainfall generation which can be applied for derived flood frequency analysis include different disaggregation approaches (Koutsoyiannis et al., 2003; Lu and Yamamoto, 2008; Olsson, 1998) and various resampling methods (Bárdossy, 1998; Lall and Sharma, 1996).

Objective of this study is to provide a simple parsimonious space-time model for the synthesis of hourly rainfall which can especially be used as data generator for flood frequency analysis. The idea is to use a hybrid or two-step approach. First, an alternating renewal model is used to simulate independently precipitation time series for several locations. In the second step, a resampling procedure is applied on the generated event time series to reproduce the spatial dependence structure of the rainfall process. The approach is validated using observed rainfall characteristics and simulated flood frequencies. The paper is organised as follows. In Sect. 2 the methodologies for the alternating renewal model and the resampling approach are developed. Section 3 discusses a case study with rainfall modelling and derived flood frequency analysis for two mesoscale catchments in northern Germany. Finally, in section 4 a summary, conclusions and an outlook are given.

2 Methodology

2.1 Single-site temporal rainfall synthesis

In the first step of the rainfall generation process single site precipitation synthesis is carried out using an alternating renewal model (ARM). Alternating renewal models describe the precipitation process by dividing the time series into dry and wet spells. The entire precipitation process is separated into an external and an internal structure. The external structure characterises the occurrence and the amount of the precipitation events, as explained by the four random variables dry spell duration (D), wet spell duration (W) and wet spell amount (V) or wet spell intensity (I) with $I=V/W$. Figure 1 shows a scheme of the rainfall event process. The internal structure describes the precipitation distribution within the wet spells. The precipitation occurrence process can be treated as an alternating renewal process if the durations of the spells are independent and the unique spell states are identically distributed. This process is completely determined by establishing probability distribution functions for the variables D and W. The precipitation amount or the intensity can also be modelled using probability distributions. However, V or I cannot be treated as independent of W.

Here, the ARM developed by Haberlandt (1998) for urban hydrologic applications is adapted and modified for the purpose of rainfall generation for flood frequency analysis. Using data from a representative subset of precipitation stations (see Sect. 3) several model structures were evaluated including different probability distribution functions and copulas for the event variables and their relationships. The finally selected components of the alternating renewal model structure are described in the following.

For modelling of the wet spell durations W a generalized extreme value distribution (GEV) is used:

$$F_W(w) = \exp\left[-\exp(-y)\right], \text{ with}$$
$$y = \begin{cases} -\kappa_G^{-1} \ln\left(1 - \frac{\kappa_G(w-\xi_G)}{\alpha_G}\right), & \kappa_G \neq 0, \\ (w-\xi_G)/\alpha_G, & \kappa_G = 0, \end{cases} \quad (1)$$

for a practical range of $w>0$, where $\xi_G \in \Re$ is a location parameter, $\alpha_G>0$ is a scale parameter and κ_G is a shape parameter.

Dry spell durations D are modelled by a Weibull distribution function (WEI):

$$F_D(d) = 1 - \exp\left[-\left(\frac{d+\zeta}{\beta}\right)^\delta\right], \quad (2)$$

for a practical range of $d>0$, where $\zeta \geq -\min(d)$ is a location parameter, $\beta>0$ is a scale parameter and $\delta>0$ is a shape parameter.

The variable wet spell intensity is modelled here instead of the wet spell amount because I allows better to fit a unique probability distribution for the whole range of values from small to heavy events, which is not feasible for V. The wet spell intensity is modelled using a Kappa distribution function (KAP):

$$F_I(i) = \left\{ 1 - h\left[1 - \frac{\kappa_K(i-\xi_K)}{\alpha_K}\right]^{1/\kappa_K} \right\}^{1/h}, \quad (3)$$

for a practical range of $i>0$, where $\xi_K \in \Re$ is a location parameter, $\alpha_K>0$ is a scale parameter and $\kappa_K \in \Re$ as well as $h \in \Re$ are shape parameters. The cases $\kappa_K=0$ and $h=0$ are included implicitly as continuous limits of Eq. (3) as $\kappa_K \to 0$ and $h \to 0$ (see Hosking, 1994).

The dependence between wet spell intensity and duration is described by a 2-copula. A 2-copula is a bivariate distribution function on the unit square with uniform marginals (see e.g. Nelsen, 2006):

$$C(u,v) = \text{Prob}(U \leq u, \ V \leq v) \quad (4)$$
with $C : [0,1]^2 \to [0,1], \ u \in [0,1], \ v \in [0,1].$

Thus, the 2-copula can be used for describing the dependence structure between the two random variables W and I with the above defined marginal distributions:

$$C(F_W(w), F_I(i)) = F(w,i). \quad (5)$$

Here, after some simple comparisons the Frank copula is chosen, mainly because of practical reasons like its fully covered dependence range, easy parameter estimation and simulation features. Besides, the Frank copula has been successfully applied for linking rainfall duration and intensity before (De Michele and Salvadori, 2003):

$$C(u,v) = -\frac{1}{\alpha} \ln\left[1 + \frac{(e^{-\alpha u}-1)(e^{-\alpha v}-1)}{(e^{-\alpha}-1)}\right] \ \alpha \in \Re - \{0\}, \quad (6)$$
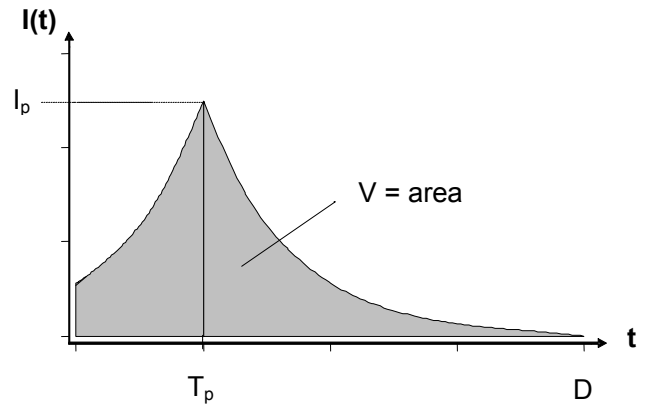

Fig. 2. Scheme of the simple profile model which is used to disaggregate wet spells into precipitation intensities.

with $u=F_W(w)$ and $v=F_I(i)$. The Frank copula has only one parameter α, describing the dependence between u and v. The parameter α can be approximated from Kendall's Tau with:

$$\tau(\alpha) \approx \frac{1}{9}\alpha - \frac{1}{900}\alpha^3 + \frac{1}{52920}\alpha^5 - \frac{1}{2721600}\alpha^7 + \dots. \quad (7)$$

To generate the two dependent random variables W and I using the copula model the procedure outlined by De Michele and Salvadori (2003) was followed. The alternating renewal model part describing the external structure of the rainfall process has 11 station specific parameters in total, which are estimated for summer (May to October) and winter seasons (November to April) separately.

For the simulation of the internal precipitation structure, a simple profile model is adapted (Haberlandt, 1998), disaggregating the wet spell amount into a special predefined pattern. To estimate the temporal distribution of the precipitation intensity $i(t)$ during a wet spell a mixture of two exponential functions is chosen:

$$i(t) = \begin{cases} i_P \times \exp\left[\lambda(t-t_P)\right], & \lambda \geq 0, \ 0 \leq t \leq t_P, \\ i_P \times \exp\left[\lambda(t_P-t)\right], & \lambda \geq 0, \ t > t_P, \end{cases} \quad (8)$$

where i_P is the wet spell peak, t_P is the wet spell peak time and λ is an event specific parameter, which can be calculated from the external variables and the peak. 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 (see Fig. 2). So, the internal rainfall model is completely defined by the four random variables W, V, I_P and T_P. Since W and V are already given by the external model, only the wet spell peak I_P and the wet spell peak time T_P have to be specified here. The wet spell peak is estimated by a simple regression to the mean wet spell intensity of the event using all stations in the study region:

$$i_P = a \times i^b, \ a \geq 1, \ b \geq 1. \quad (9)$$

Realisations of the wet spell peak time t_P are generated from a uniform distribution. The main advantage of this simple profile model is that it needs no station specific parameters. This approach is obviously a strong simplification of the internal rainfall process. However, from the internal structure only a sufficient approximation of the peak intensity is expected. The internal rainfall variability is assumed to be of minor importance compared to the external one for the generation of extreme flow events (see also Haberlandt, 1998).

2.2 Synthesis of the space-time rainfall structure using resampling

In this second step of the rainfall generation process the synthetic precipitation event time series for several locations in the study region will be resampled in order to reproduce the spatial dependence structure of the rainfall process. It is important to note, that for resampling not the hourly data are used but the event series. This is in accordance with the basic idea of the alternating renewal process from step one, which assumes the independence of subsequent events. Thus, the resampling on the event time series preserves the temporal rainfall structure for single time series and does not need to consider any autocorrelation. The resampling is carried out using simulated annealing (Aarts and Korst, 1989), which has been used for rainfall generation before (Bárdossy, 1998). Simulated annealing can be considered as a non-linear discrete optimisation method which minimises a certain objective function.

In our case the objective function should reflect the spatial dependence structure of the rainfall process. Three bivariate criteria are defined for this purpose and can be calculated from the hourly rainfall time series z. The first criterion describes the probability of bivariate rainfall occurrence at two stations k and l:

$$P_{kl}\ (z_k > 0 | z_l > 0) \approx \frac{n_{11}}{n}, \tag{10}$$

where, n_{11} is the number of hours with simultaneously occurring rainfall and n is the total number of non-missing observation hours, at both stations respectively. The second criterion describes the relationship between rainfall intensities at two stations k and l if it rains at both stations. Here, Pearson's coefficient of correlation is applied:

$$\rho_{kl} = \frac{cov\ (z_k, z_l)}{\sqrt{var(z_k) \times var(z_l)}}\ z_k > 0,\ z_l > 0. \tag{11}$$

The last criterion is a continuity measure proposed by Wilks (1998), which compares the expected rainfall amounts at a station k for times with and without rain at a neighbouring station l:

$$C_{kl} = \frac{E\ (z_k | z_k > 0,\ z_k = 0)}{E\ (z_k | z_k > 0,\ z_l > 0)}, \tag{12}$$

where E(.) is the expectation operator. The ratio C_{kl} will become smaller with increasing interrelation between the two stations, and it will become a value of about 1 for independent stations.

The three criteria are chosen to complement each other in describing spatial dependence of the rainfall process. The first one P_{kl} considers only the rainfall occurrence process i.e. the coincidence of wet hours at neighbouring stations. The second one ρ_{kl} measures only the linear association between rainfall intensities for simultaneous rainy hours at two stations. The third one C_{kl} estimates the expectation of the rainfall intensity of one station conditioned on the rainfall status dry or wet of the neighbouring station combining the spatial relation of rainfall occurrence and intensity between the two stations.

The three criteria are coupled into one bivariate objective functions as follows:

$$O_{kl} = w_1 \times \left(P_{kl} - P_{kl}^*\right)^2 + w_2 \times \left(\rho_{kl} - \rho_{kl}^*\right)^2 + \tag{13}$$
$$w_3 \times \left(C_{kl} - C_{kl}^*\right)^2,\ w_1, w_2, w_3 > 0,$$

where the variables marked with * represent the prescribed values and the other ones are the simulated values. The weights w_1, w_2 and w_3 are used to control the importance and to adjust the scale of the different criteria. The prescribed values need to be estimated from observed hourly data before they can be used in the objective function. If it is possible to express those bivariate statistics as functions of the separation distance between two stations k and l, the criteria could be used for resampling of time series between any two points in the study region no matter if observations are directly available at these locations.

Let U be the set of all stations, which are unchanged and R be set of all stations which have already been changed, then the resampling algorithm using simulated annealing works as follows:

1. A rainfall station k with $k = 1, \ldots, N$ is selected at random from the set U to be used for resampling. A backup of the event time series from station k is made and the events are disaggregated into hourly values. If the set U is empty the procedure is finished.

2. All rainfall stations l with $l = 1, \ldots, M$ are selected from the set R which represents the reference stations for resampling of the station k. If the set R is empty then station k is taken from the set U and added unchanged to the set R, and the algorithm returns to step 1. Otherwise it proceeds with step 3.

3. Two events are drawn at random from the resampling station k and swapped. A backup of the modified data from the time period between the two events is made and then the changed period is disaggregated into hourly values.
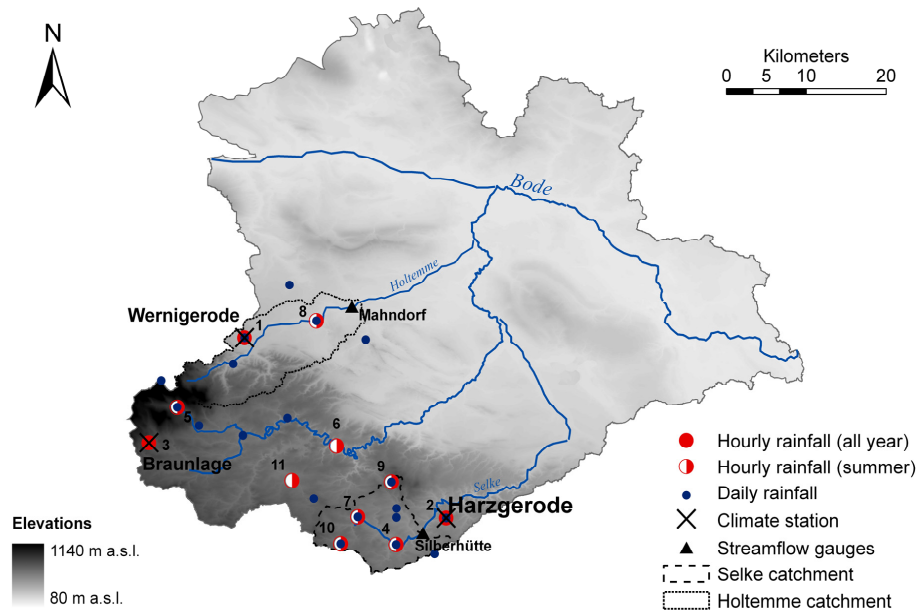
Fig. 3. Study region showing the two mesoscale catchments Holtemme and Selke within the Bode river basin, the topographic structure, the hourly rainfall stations with id's (cf. Table 1), the daily rainfall stations, the climate stations and the two streamflow gauges.

4. The value for the objective function O_{kl} (Eq. 13) is updated using the data from the hourly series. An average objective function value O_k for station k considering all the neighbours l with $l=1,\dots, M$ from set R is calculated:

$$O_k = \frac{1}{k-1} \sum_{l=1}^{M} O_{kl}. \tag{14}$$

5. The new value of the objective function O_{new} is compared with the old value O_{old}, which was calculated before the last swap. If $O_{\text{new}} < O_{\text{old}}$ then the change is accepted.

6. If $O_{\text{new}} \geq O_{\text{old}}$ then the change is accepted with the probability π:

$$\pi = \exp\left(\frac{O_{\text{old}} - O_{\text{new}}}{T_a}\right), \tag{15}$$

where T_a is the annealing temperature.

7. Steps 3 to 6 are repeated K times.

8. The annealing temperature T_a is reduced: $T_a = T_{a-1} \times dT$ with $0 < dT < 1$ and the algorithm proceeds with step 3.

9. Steps 7 and 8 are repeated until the algorithm converges regarding resampling of the station k.

10. Then station k is removed from the set U and added to the set R. The algorithm returns to setp 1 for resampling of the next station.

Step 6 is essential for the optimisation not to stop at any local minimum but to find a good solution near the global minimum. The annealing temperature regulates the probability of negative changes. The lower T_a the less likely is the acceptance of a negative change. The algorithm stops if the objective function O becomes smaller than a prescribed value or if O does not decreases anymore for a certain number of iteration steps. The convergence is ensured in reducing the annealing temperature T_a each time after K iteration steps. A slow temperature decrease rate with values of $dT=0.95$ or 0.90 shows usually a good performance. In order to speed up the algorithm for resampling of long time series the hourly spatial dependence criteria are not recalculated for the whole time series but updated considering only the changed period between the two swapped events. In addition the distance between two randomly selected events for swapping can be restricted, which further speeds up the disaggregation into hourly values and the calculation of the target criteria.

As seen from step 1, the stations are drawn at random. The degrees of freedom for resampling decrease from the first to the last station because the number of reference stations increases. However, the total objective function value over all stations is independent from the order. Note, that the order might have an effect on the hydrological simulation, which will be assessed in future work. It should also be mentioned here, that the first approach was to resample all stations at once. However, the algorithm did not converge to a good solution, probably because in this multivariate case too many degrees of freedom are possible.

Table 1. Recording precipitation stations and their characteristics (see Fig. 3 for locations). MAP is the mean annual precipitation amount for the used observation periods. For better comparison the observed amount for the summer stations is multiplied by two.

| No.	Name	Elevation [m. a.s.l.]	Used observation period	MAP [mm/yr]	Operated all year	Used for Selke basin	Used for Holtemme basin
1	Wernigerode	234	1993–2006	778	1	0	1
2	Harzgerode	404	1993–2006	595	1	1	0
3	Braunlage	607	1993–2006	1568	1	0	1
4	Strassberg	400	1993–2006	732	0	1	0
5	Schierke	609	1993–2001	1266	0	0	1
6	Altenbrak	300	1993–2001	723	0	0	1
7	Günthersberge	420	1993–2003	721	0	1	0
8	Derenburg	161	1993–2001	659	0	0	1
9	Friedrichsbrunn	523	1993–2002	734	0	1	0
10	Breitenstein	466	1993–2002	792	0	1	0
11	Hasselfelde	461	1993–2003	759	0	1	1
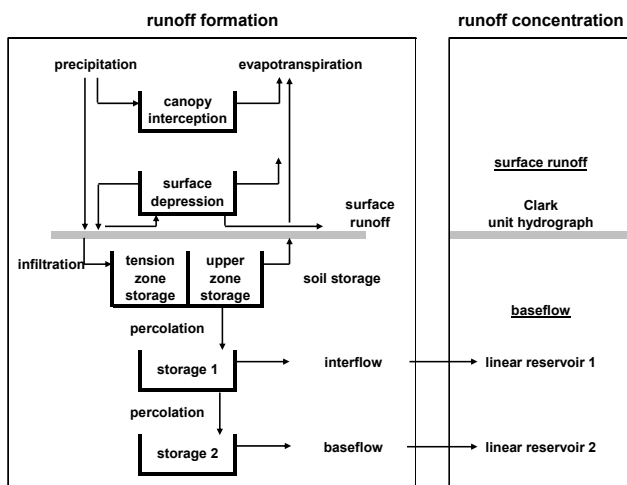
Fig. 4. Selected structure of the hydrological model HEC-HMS.

The second step of the rainfall generation process, which is described above, yields rainfall time series with spatial dependence regarding the criteria in Eqs. (10–12) but preserves the temporal characteristics from the alternating renewal process.

3 Case study

3.1 Data and study region

The hybrid rainfall model is tested and applied for derived flood frequency analysis in two mesoscale catchments within the Bode river basin in northern Germany (Fig. 3). The considered Bode region has elevations between 1140 m a.s.l. at the top of the Brocken Mountain and about 80 m a.s.l. Mean annual rainfall varies between 1700 mm/yr and 500 mm/yr.

The two catchments Holtemme and Selke have drainage areas of 168 km^2 and 105 km^2, respectively. Floods are generated either by frontal rainfall, frontal rainfall on snow smelt or convective storms. Large floods in the Selke catchment occur mainly in the winter season while floods in the Holtemme take place mostly in the summer time.

A total number of 23 recording rainfall stations with hourly data and 19 non-recording rainfall stations with daily data are employed for the case study. The daily stations are used only as additional information for hydrological modelling. The 23 stations with hourly records are located in an extended study area within and around the Bode river basin and are used in this total set for the robust estimation of the spatial dependence criteria. Only a subset of 11 hourly stations can be used directly for rainfall synthesis and hydrological modelling regarding the two mesoscale catchments. Figure 3 shows the study area, the selected catchments, the 19 daily stations and the relevant subset of 11 hourly stations. From these hourly stations 3 gauges are operated all year. The remaining 8 hourly stations are operated during the summer season only. Table 1 list these recording hourly rainfall stations with their characteristics and indicates their use for hydrological modelling in the two catchments. The length of the observation periods varies between 9 and 14 years from 1993 to 2006.

3.2 Rainfall-runoff model

For runoff simulations the hydrological model HEC-HMS (Scharffenberg and Fleming, 2005) is used, which has been applied successfully in the past for various purposes (Cunderlik and Simonovic, 2005; Fleming and Neary, 2004; Maskey et al., 2004; Neary et al., 2004). HEC-HMS is a conceptual semi-distributed rainfall-runoff model and offers various tools for the description of the hydrological processes. Figure 4 illustrates the model structure which has
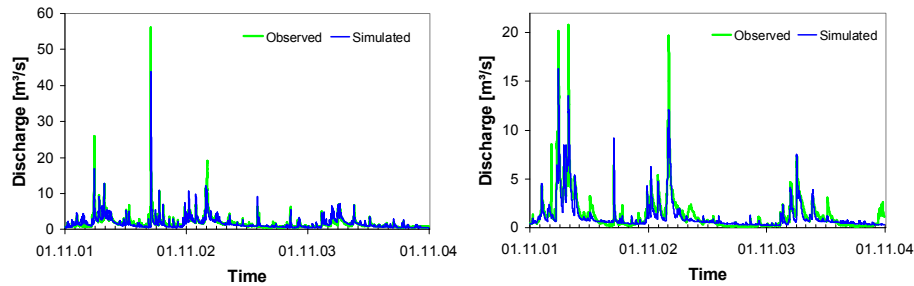
Fig. 5. Observed and simulated flows in the validation period for the Holtemme basin (left) and Selke the basin (right).

been compiled from the available tools for this study. The model is operated continuously on an hourly time step. It uses the soil moisture accounting (SMA) algorithm for runoff generation, the Clark Unit Hydrograph for the transformation of direct runoff, two linear reservoirs to consider interflow and base flow transformation and a simple river routing where the flows are only lagged in time. Snow melt is calculated externally using the degree-day method. Potential evapotranspiration is computed also externally using the method proposed by Turc-Wendling (Wendling et al., 1991) based on observed temperature and global radiation data from the three available climate stations. Potential evapotranspiration is aggregated and averaged to mean monthly values and fed into HEC-HMS. Actual evapotranspiration is calculated hourly depending on water availability from canopy, surface and soil storages. To account for spatial heterogeneity of meteorological and basin characteristics the two catchments are spatially divided into several subcatchments and river reaches.

Because of the sparse network of recording hourly rainfall gauges also daily stations are included in the model calibration. For that purpose daily rainfall totals are disaggregated into hourly data using the intensity profile from the nearest station with high resolution data. Areal rainfall for subcatchments is then calculated by Thiessen interpolation from all daily and hourly station locations.

The model HEC-HMS is calibrated for the period from November 1997 to 2001 and validated for the subsequent period from November 2001 to October 2004 for the streamflow gauge Mahndorf at the Holtemme catchment and the streamflow gauge Silberhütte at the Selke catchment (cf. Fig. 3). The hydrological model was parameterised using iteratively manual and automatic calibration. Five parameters were included in the calibration process comprising mainly capacity values and recession constants of the subsurface storages (cf. Fig. 4). The calibration was focussing on peak flows using e.g. the peak-weighted root mean square error in the objective function (USACE, 1998). Figure 5 shows a comparison of observed and simulated streamflow during the validation period for these two gauges. On average, the obtained model performance is sufficient with Nash-Sutcliffe efficiencies of 0.81 and 0.86 for the Holtemme and Selke

catchments, respectively. However, the hydrographs indicate an underestimation of the peak flows and an overestimation of the low flows, especially for the Selke basin. This is a typical result for simple continuously operating hydrological models, which can provide an unbiased global estimation but often at the cost of smoothing the flow time series (e.g. Götzinger and Bárdossy, 2008, Figs. 14 and 15; Fenicia et al., 2008, Fig. 4).

3.3 Synthesis and application of stochastic rainfall

The hybrid precipitation model is applied and validated for the study region in the following three step procedure:

1. Parameters are estimated for the alternating renewal model (ARM) and single site rainfall is generated independently for all hourly stations in the study region.

2. Spatial dependence criteria are estimated for the study region and the hourly rainfall series are resampled using simulated annealing to generate the spatial rainfall structure.

3. Derived flood frequency analysis is carried out for both catchments using the rainfall-runoff model HEC-HMS and the synthetic rainfall data.

The identification and parameter estimation of the alternating renewal model depends on the definition of a rainfall event. Rainfall events are defined here by a minimum rainfall amount V_{min} and a minimum separation time between events D_{min}. The minimum rainfall amount is set to V_{min}=0.5 mm/event and excludes negligible events which would complicate the fitting of probability distributions. Note, that this leads to a small systematic underestimation of the total rainfall sum, which needs to be corrected before the data can be used for continuous hydrologic modelling. The second criterion D_{min} should theoretically guarantee the statistical independence of subsequent events. In the literature values of D_{min} range from 1 h (Acreman, 1990) to 12 h (De Michele and Salvadori, 2003). If D_{min} is greater than the target time step of 1 h here the generated hourly values underestimate the true precipitation intensity if the model is not

Table 2. Parameter estimates of the alternating renewal model for the three total year hourly rainfall stations Wernigerode, Harzgerode and Braunlage. Kendalls τ is not a model parameter; it is included here for information only.

Model variable	Description	Parameter	Wernigerode		Harzgerode		Braunlage	
			Summer	Winter	Summer	Winter	Summer	Winter
Wet spell duration (W)	GEV	ξ_G [min]	127.1	236.0	126.5	239.1	146.7	323.1
		α_G [min]	77.18	169.6	81.72	168.4	103.7	253.6
		κ_G [-]	-0.335	-0.268	-0.333	-0.232	-0.385	-0.282
Dry spell duration (D)	WEI	ς [-min]	-26.91	-102.9	-35.00	-77.22	-41.77	-117.9
		β [min]	1945.9	1756.7	2031.4	2260.4	1004.5	967.28
		δ [-]	0.6650	0.6370	0.6490	0.6670	0.5470	0.5380
Wet spell intensity (I)	KAP	ξ_K [mm/h]	0.3969	0.1940	0.3486	0.3148	0.4120	0.2778
		α_K [mm/h]	0.5452	0.3795	0.5464	0.2269	0.6131	0.4238
		κ_K [-]	-0.3458	-0.0485	-0.3281	-0.2609	-0.3286	-0.0988
		h [-]	0.6347	0.8170	0.7530	0.1826	0.6658	0.6495
$I=f(W)$	Frank copula	α [-]	-0.7006	-0.6227	-1.0163	-0.9700	-0.3429	1.1841
		τ [-]	-0.078	-0.069	-0.113	-0.108	-0.038	0.132

able to consider clustering within the wet spells (i.e. zero rainfall amounts for certain time steps within a wet spell). To avoid complicated model formulations and underestimation of intensities it was tried to keep D_{min} as small as possible. Analysing event time series based on different values of D_{min} regarding small serial correlation and good fitting performance for the marginal distributions optimal results were found here with minimum dry spell durations of one and two hours for summer and winter seasons, respectively.

Probability distribution functions are fitted to the event variables dry spell duration (D), wet spell duration (W) and wet spell intensity (I). The parameters for the distribution functions are estimated using the method of L-moments (Hosking and Wallis, 1997) for winter and summer seasons separately. Figure 6 shows exemplarily the fitting performance of the event variables for the station Wernigerode based on QQ-Plots. A satisfying agreement between empirical and theoretical quantiles can be observed. The Cramer-von Mises test (e.g. Ahmad et al., 1988) was accepted at the 5% significance level in all the cases for D, W and I, respectively.

Table 2 shows the estimates of all ARM parameters for the three all year rainfall stations for both summer and winter seasons. The differences of the values between summer and winter stress the importance of the seasonal parameter estimation. While the parameters for Harzgerode and Wernigerode are often quite similar, the differences to the station Braunlage are more pronounced. This is due to the exposed location of the station Braunlage in the mountains with differing precipitation regime and high annual precipitation amounts.

Several realisations of hourly rainfall time series, each 100 years in length are generated for all stations. Table 3 shows a comparison between observed and simulated event characteristics exemplarily for the three all year rainfall stations. The comparison shows sufficient agreement between observed and simulated statistics with a slight underestimation of mean rainfall and a somewhat larger deviation for the higher order moments, which is typical for this kind of models. Note, that only those features are used here for validations, which do not represent model variables in the precipitation model. In addition a frequency analysis is carried out on the annual maximum precipitation series for different durations. The results are presented in Fig. 7 for the station Harzgerode for both seasons and rainfall durations of 1 and 3 h. It can be seen, that the observed values are plotted mostly within the range of the simulated realisations. For larger return periods and durations a slight overestimation of the observed extreme values occurs. Considering the short observation periods it is difficult to validate the model regarding the synthesis of more extreme rainfall intensities. It is expected, that the hydrological validation, which comprises longer observed flow records, will allow an additional assessment of model performance regarding this issue.

Precondition for resampling of the event time series is the estimation of the spatial dependence criteria needed for the objective function (Eq. 13). Based on an extended data sample (cf. Sect. 3.1) different functions for the summer and winter seasons are fitted to the three statistics in relation to the separation distance between stations. Figure 8 shows the results exemplarily for the summer seasons. The relation between the spatial dependence criteria defined in Eqs. (10) to (12) and the separation distance appear strong enough for
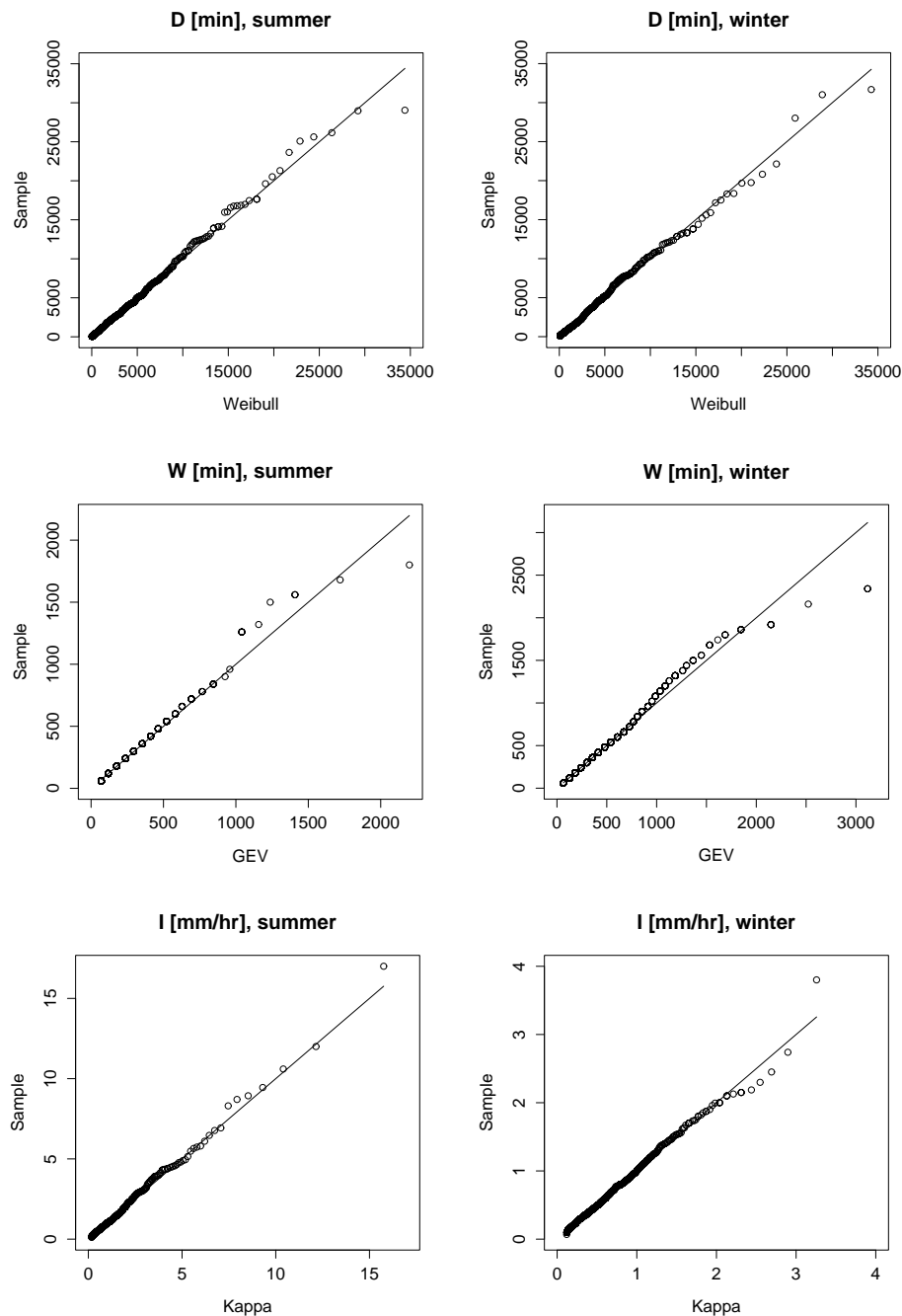
Fig. 6. QQ-Plots illustrating the fitting of distribution functions to the event variables dry spell duration (D), wet spell duration (W) and wet spell intensity (I) for summer and winter rainfall data at the station Wernigerode.

the application of those relationships to derive the required statistics. The functions show the expected behaviour of the spatial rainfall expressing decreasing probability of simultaneous rainfall occurrence, decreasing correlation and increasing continuity with rising separation distance between two stations. From Fig. 8 a range between 150 km and 200 km can be estimated where the dependence between two stations disappears.

The resampling of the synthetic event series is carried out separately for the two catchments and the two seasons using simulated annealing for all available hourly rainfall gauges (see Table 1). Two stations for the winter season and six stations for the summer season are included for this procedure in the Holtemme catchment. Concerning the Selke catchment six stations are used for resampling of the summer data. However, in that basin no resampling for the winter time was
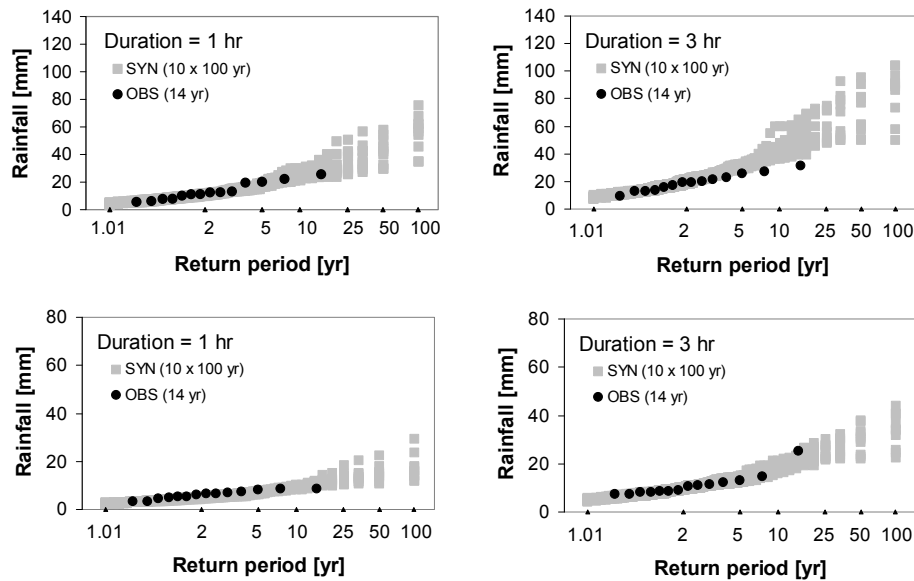
Fig. 7. Empirical probability distribution functions of observed (OBS) and synthetic (SYN) seasonal maximum rainfall for the station Harzgerode and two durations (top: Summer, bottom winter; the x-axis is linear with respect to the Gumbel reduced variate).

Table 3. Event characteristics for the rainfall stations Wernigerode, Harzgerode and Braunlage (for location see Fig. 3) from 14 years observed and 200 years simulated rainfall data each.

Name of rainfall station	Average number of events [-]		Average V [mm]		Std. Dev. of V [mm]		Skewness of V [-]		Rainfall sum [mm]	
	obs	sim	obs	sim	obs	sim	obs	sim	obs	sim
Summer season (May to October):										
Wernigerode	102	92	3.88	3.72	6.02	6.09	4.93	7.51	396	343
Harzgerode	89	86	3.68	3.56	5.36	5.68	4.81	7.73	328	308
Braunlage	134	128	5.33	5.24	8.42	8.96	5.08	6.19	714	672
Winter season (November to April):										
Wernigerode	102	88	3.77	3.53	5.21	4.47	3.85	5.28	382	312
Harzgerode	78	76	3.39	3.25	4.93	4.15	7.49	6.63	266	246
Braunlage	115	110	7.44	7.02	11.5	10.1	3.91	4.21	854	768

necessary, because only one station is available. The performance of the simulated annealing algorithm depends on the number of rainfall stations included and on the specific criterion considered. The results improve with fewer stations. So, it is favourable to process the two neighbouring catchments separately. Comparing the performance regarding the different criteria it is most difficult for the algorithm to simulate the prescribed probability of simultaneous rainfall occurrence P_{kl}^*, while it is easier to mimic the continuity measure C_{kl}^* and it is no problem to reproduce the correlation ρ_{kl}^*. On the whole the reproduction of the spatial dependence criteria was satisfactory for the two mesoscale catchments. Figure 9 shows a comparison of hourly rainfall time series before and after the resampling process for two neighbouring rainfall

stations, which are located about 20 km apart. The effect of the resampling procedure becomes quite clear. Wet and dry spells as well as intensity peaks correspond much better between the two stations after the resampling than before.

In the third step the stochastic rainfall data are used as input for the hydrological model HEC-HMS. The hydrological model was calibrated using hourly and daily rainfall data, but synthetic precipitation is only generated for the hourly stations. It is known, that applying a hydrological model on rainfall data from a different station network as used in the calibration might produce biased runoff and the model parameters might need a recalibration, which is difficult considering synthetic rainfall as input. Especially a model calibrated on denser precipitation information, as it would be the

case here, might fail on sparser information (Bárdossy and Das, 2008). For that reason the generated hourly data are also transferred to the daily station locations, which were used in calibration, applying nearest neighbour interpolation and a correction factor taking into account the different mean seasonal precipitation amounts at the specific locations. Then, synthetic areal rainfall for subcatchments is calculated by Thiessen interpolation from all daily and hourly station locations. No stochastic model for the other climate variables like temperature and global radiation is applied here. To provide long term hourly data of evapotranspiration for continuous modelling a pragmatic approach was chosen. Potential evapotranspiration is estimated from observed temperature and global radiation data using the same methods as for calibration (cf. Sect. 3.2) and applied as upper limits for the calculation of actual evapotranspiration. For snowmelt modelling repetitions of 25 years blocks of observed hourly temperature data are used.

Figure 10 compares observed and simulated flood frequencies from annual series for the two catchments Selke and Holtemme at the streamflow gauges Silberhütte and Mahndorf, respectively. Simulated flows are shown based on hourly rainfall data using the short observed rainfall time series and 10 synthetic rainfall realisations each 100 years in length. In addition observed annual peak flows are plotted, which have notable longer records than the simulated ones using observed precipitation. The observed maximum flows and the simulated ones using observed rainfall data are lying mostly within the range of the simulated flows based on stochastic precipitation data, although located somewhat more in the lower part of the synthetic range. Despite this slight overestimation of flows the overall picture shows the ability of the precipitation model to provide suitable input for derived flood frequency analysis. Note, that in Fig. 10 the difference between the empirical distribution functions of observed and simulated flows using observed precipitation is indicating the model performance of HEC-HMS regarding the reproduction of the flood frequency. So, in the first instance the derived distribution functions using synthetic rainfall should cover the simulation results using observed rainfall.

The results for the Selke catchment (left panel in Fig. 11) also show typical problems with small sample sizes. The largest value from observed flows and simulated flows using observed rainfall each belongs to the same flood and is associated with the maximum possible return period according to the sample size of 56 and 12 years, respectively. Both points are located above the simulated range of flows based on synthetic rainfall. The reason for that is that these values belong to an exceptional flood event occurring in 1994 which can be associated with a much higher return period compared to the length of the observed time series (LAU, 1995).

Comparing Fig. 10 with Fig. 5 it becomes clear, that the underestimation of high flows during calibration of the hydrological model is not apparent in the application any more.
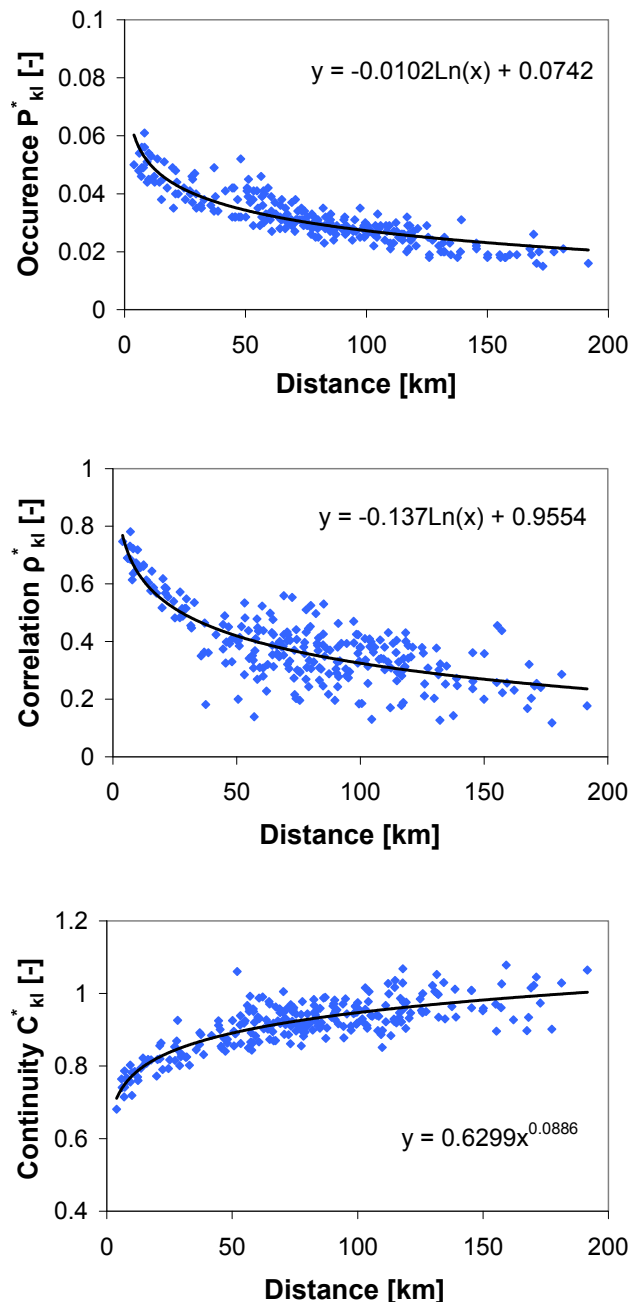

Fig. 8. Relation between the spatial dependence criteria which are used in the objective functions as prescribed values for simulated annealing and the station separation distance as estimated for the summer season.

The type of precipitation input has a significant impact on the calibration of the parameters of the hydrological model. So, using observed rainfall for calibration and synthetic rainfall for application might be one reason for those results. One possibility to overcome this problem is to utilize stochastic rainfall already in the calibration phase of the hydrological model. This could be achieved if not the hydrograph is used for calibration but the empirical probability distribution
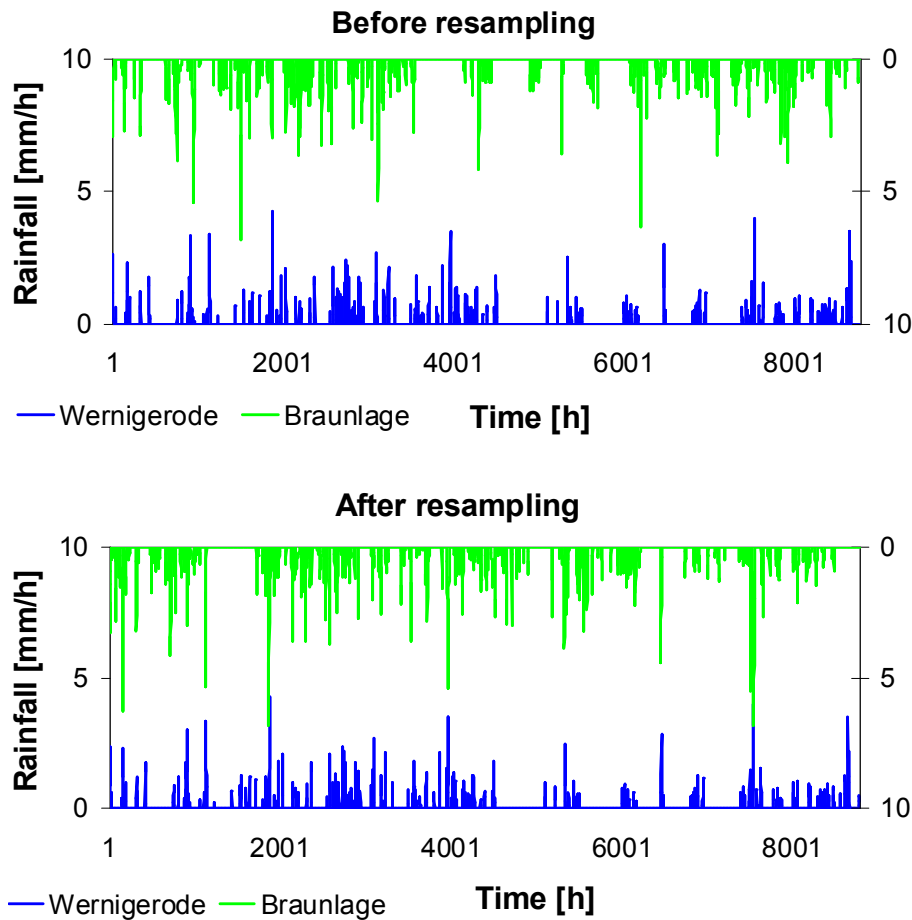
Fig. 9. Comparison of hourly precipitation time series for the stations Wernigerode and Braunlage before and after resampling the events of the latter station using simulated annealing.
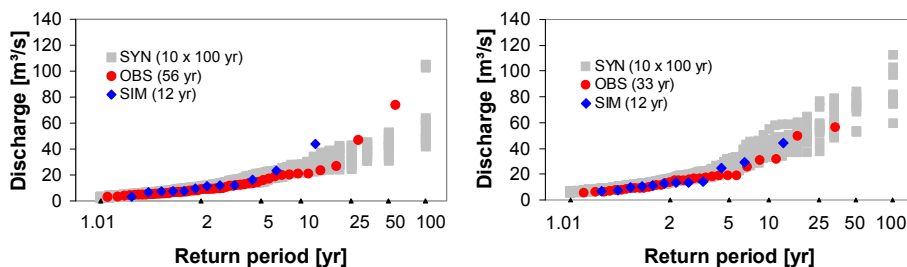

Fig. 10. Empirical probability distribution functions of observed discharge (OBS), simulated discharge using observed rainfall (SIM) and 10 realisations simulated discharge each 100 years in length using synthetic rainfall (SYN) for the Selke (left) and Holtemme (right) catchments (The x-axis is linear with respect to the Gumbel reduced variate).

function of the annual maximum flows. Additional advantage would be to have longer periods with observed peak flows available for model calibration compared to periods with continuous hourly flows and to focus directly on the specific objective of the application. Work is in progress to investigate these ideas for better calibration of the whole framework.

In order to evaluate the importance of the spatial rainfall structure or more precisely the effect of the resampling procedure on the simulated flood frequencies an additional analysis is carried out. For that hydrological simulations with HEC-HMS are made using spatially randomly rainfall, spatially resampled rainfall and spatially uniformly distributed rainfall as input. For the random rainfall case the resampling

procedure was omitted and for the uniform rainfall case only one rainfall station was used as homogeneous areal rainfall. The results are illustrated for the Selke catchment in Fig. 11 considering the summer season only, where the spatial dynamics are higher and more stations for resampling were available. It can be seen, that homogeneous rainfall produces the highest flows, spatial independent rainfall causes the lowest flows and spatially resampled rainfall leads to flows in between these two cases. The differences between the three empirical probability distributions increase with increasing return period. For a return period of 100 years the flood resulting from uniform rainfall is about 3.5 times as high as a flood generated from random rainfall, while the flood from spatially structured rainfall is 2 times as high as in the random case. Comparing the empirical probability distributions with the observed values the random case seems to correspond best to the majority of points. However, for the more important larger return periods, the flows generated by resampled rainfall correspond better to the observed ones. Taking into account the general problem of the model to slightly overestimate the observed flows when using synthetic rainfall (cf. Fig. 10), it can be concluded that the structured rainfall produces the most plausible flood frequency curve. This result encourages the application of the proposed resampling procedure to reproduce important characteristics of the spatial rainfall structure. More research is necessary to investigate the general ability of this procedure to generate sufficient spatially structured rainfall for flood frequency analysis depending on data and local conditions like the number of required stations, the catchment size and the climate.

4 Summary and conclusions

In this study a hybrid hourly rainfall model has been presented, which can be applied for derived flood frequency analysis. The model consists of two parts, an alternating renewal approach for the generation of point rainfall and a resampling procedure based on simulated annealing to reproduce the spatial statistics. Special attention was given to the derivation of a practical applicable parsimonious approach for precipitation synthesis. The performance of the rainfall model has been tested in three stages, for the simulation of temporal rainfall characteristics at single sites, for the reproduction of the spatial rainfall structure and for flood simulations in two mesoscale catchments in northern Germany. The results can be summarised as follows:

1. The single site rainfall model based on the alternating renewal approach allows a good reproduction of average event statistics and extreme value frequencies for short rainfall durations. This result could be achieved despite the parsimonious approach which uses only one probability distribution function for the whole range of wet spell intensities. The dependence between wet spell intensity and duration can easily be considered using
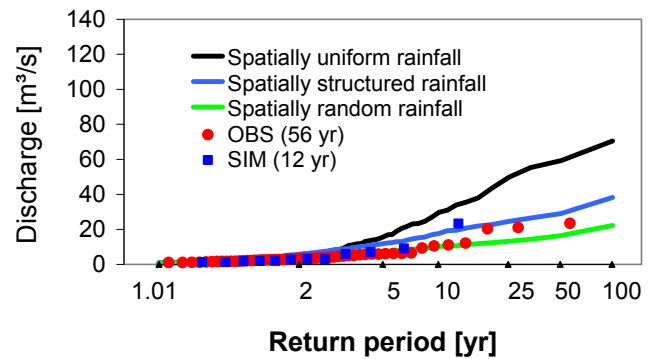

Fig. 11. Empirical probability distribution functions of observed discharge (OBS), simulated discharge using observed rainfall (SIM) and the median of 10 realisations simulated discharge each 100 years in length for three synthetic rainfall scenarios for the Selke catchment in the summer season (The x-axis is linear with respect to the Gumbel reduced variate).

a copula model. Still, improvements of this temporal model are desirable with regard to the overestimation of the extreme values for longer rainfall durations in the summer season.

2. With multisite resampling of the synthetic event series using simulated annealing the spatial rainfall structure could sufficiently be reproduced. The resampling of the events instead of the single hours preserves the temporal rainfall properties, which has been prescribed in the first part of the model and, besides, it is computationally very favourable. It has been shown, that the consideration of the spatial rainfall structure in hydrological modelling has a significant effect on the derived flood frequencies.

3. The simulation results from rainfall runoff modelling have demonstrated the suitability of the synthetic precipitation data for derived flood frequency analysis in mesoscale catchments. However, the results have also shown that using a different type of precipitation data in calibration and application of the hydrological model might lead to a reversed bias regarding observed flows. In future work the synthetic rainfall data might be included directly for the calibration of the hydrological model e.g. with the objective to minimize deviations to the observed flood frequencies.

Derived flood frequency analysis using synthetic precipitation data and rainfall-runoff modelling is becoming increasingly important in practical applications. This is especially the case for investigating the effect of complex flood protection measures and for analysing impacts of climate and land use changes. The presented model has the potential to be used for statistical downscaling in climate impact studies; because its parameters are observable quantities which are

related to larger scale climate characteristics like circulation patterns which can be obtained from climate models. Also, an application for ungauged locations is possible through regionalisation of the model parameters. With such a method the hourly network density might be increased in the future e.g. by including the daily station locations in a more sophisticated way as by simple data transfer.

References

Aarts, E. and Korst, J.: Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing, John Wiley, New York, 284 pp., 1989.

Acreman, M. C.: Simple Stochastic Model of Hourly Rainfall for Farnborough, England, Hydrol. Sci. J., 35, 119–148, 1990.

Ahmad, M. I., Sinclair, C. D., and Spurr, B. D.: Assessment of Flood Frequency Models Using Empirical Distribution Function Statistics, Water Resour. Res., 24, 1323–1328, 1988.

Aronica, G. T. and Candela, A.: Derivation of flood frequency curves in poorly gauged Mediterranean catchments using a simple stochastic hydrological rainfall-runoff model, J. Hydrol., 347, 132–142, 2007.

Bárdossy, A.: Generating precipitation time series using simulated annealing, Water Resour. Res., 34, 1737–1744, 1998.

Bárdossy, A. and Das, T.: Influence of rainfall observation network on model calibration and application, Hydrol. Earth Syst. Sci., 12, 77–89, 2008,
http://www.hydrol-earth-syst-sci.net/12/77/2008/.

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, J. Hydrol., 249, 11–29, 2001.

Blazkova, S. and Beven, K.: Flood frequency estimation by continuous simulation of subcatchment rainfalls and discharges with the aim of improving dam safety assessment in a large basin in the Czech Republic, J. Hydrol., 292, 153–172, 2004.

Cameron, D. S., Beven, K. J., Tawn, J., Blazkova, S., and Naden, P.: Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty), J. Hydrol., 219, 169–187, 1999.

Cowpertwait, P. S. P.: A spatial-temporal point process model of rainfall for the Thames catchment, UK, J. Hydrol., 330, 586–595, 2006.

Cunderlik, J. M. and Simonovic, S. P.: Hydrological extremes in a southwestern Ontario river basin under future climate conditions, Hydrol. Sci. J., 50, 631–654, 2005.

De Michele, C. and Salvadori, G.: A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-Copulas, J. Geophys. Res., 108(D2), 4067, doi:4010.1029/2002JD002534, 2003.

Evin, G. and Favre, A.-C.: A new rainfall model based on the Neyman-Scott process using cubic copulas, Wat. Resour. Res., 44, W03433, doi:03410.01029/02007WR006054, 2008.

Fenicia, F., Savenije, H. H. G., Matgen, P., and Pfister, L.: Understanding catchment behavior through stepwise model concept improvement, Wat. Resour. Res., 44, W01402, doi:01410.01029/02006WR005563, 2008.

Fleming, M. and Neary, V.: Continuous Hydrologic Modeling Study with the Hydrologic Modeling System, J. Hydrologic Eng., 9, 175–183, 2004.

Götzinger, J. and Bárdossy, A.: Generic error model for calibration and uncertainty estimation of hydrological models, Wat. Resour. Res., 44, W00B07, doi:10.1029/2007WR006691, 2008.

Grace, R. A. and Eagleson, P. S.: The Synthesis of Short-Time-Increment Rainfall Sequences, Hydrodynamics Laboratory, Massachusetts Institute of Technology, Cambridge, USA, Report No. 91, 1966.

Haan, C. T., Allen, D. M., and Street, J. O.: A Markov Chain Model of Daily Rainfall, Water Resour. Res., 12, 443–449, 1976.

Haberlandt, U.: Stochastic rainfall synthesis using regionalized model parameters, J. Hydrol. Eng., 3, 160–168, 1998.

Hosking, J. R. M.: The four parameter kappa distribution, IBM J. Res. Develop., 38, 251–258, 1994.

Hosking, J. R. M. and Wallis, J. R.: Regional frequency analysis: an approach based on L-moments, Cambridge University Press, New York, USA, 1997.

Koutsoyiannis, D., Onof, C., and Wheater, H. S.: Multivariate rainfall disaggregation at a fine timescale, Water Resour. Res., 39, 1173, doi:1110.1029/2002WR001600, 2003.

Lall, U. and Sharma, A.: A nearest neighbor bootstrap for resampling hydrological time series, Water Resour. Res., 32, 679–693, 1996.

LAU: Das Frühjahrshochwasser vom April 1994 in den Flusseinzugsgebieten der Saale und Bode in Sachsen-Anhalt, Berichte des Landesamtes für Umweltschutz Sachsen-Anhalt, Germany, 1995.

Lu, M. and Yamamoto, T.: Application of a Random Cascade Model to Estimation of Design Flood from Rainfall Data, J. Hydrol. Eng., 13, 385–391, 2008.

Maskey, S., Guinot, V., and Price, R. K.: Treatment of precipitation uncertainty in rainfall-runoff modelling: a fuzzy set approach, Adv. Water Res., 27, 889–898, 2004.

Moretti, G. and Montanari, A.: Inferring the flood frequency distribution for an ungauged basin using a spatially distributed rainfall-runoff model, Hydrol. Earth Syst. Sci., 12, 1141–1152, 2008,
http://www.hydrol-earth-syst-sci.net/12/1141/2008/.

Neary, V. S., Habib, E., and Fleming, M.: Hydrologic Modeling with NEXRAD Precipitation in Middle Tennessee, J. Hydrol. Eng., 9, 339–349, 2004.

Nelsen, R. B.: An Introduction to Copulas, 2nd ed., Springer, New York, USA, 2006.

Olsson, J.: Evaluation of a scaling cascade model for temporal rainfall disaggregation, Hydrol. Earth Syst. Sci., 2, 19–30, 1998, http://www.hydrol-earth-syst-sci.net/2/19/1998/.

Onof, C., Chandler, R. E., Kakou, A., Northrop, P., Wheater, H. S., and Isham, V.: Rainfall modelling using Poisson-cluster processes: A review of developments, Stochastic Environ. Res. Risk Assess., 14, 384–411, 2000.

Pegram, G. G. S. and Clothier, A. N.: High resolution space–time modelling of rainfall: the "String of Beads" model, J. Hydrol., 241, 26–41, 2001.

Rodríguez-Iturbe, I., Febres de Power, B., and Valdés, J. B.: Rectangular Pulses Point Process Models for Rainfall: Analysis of Empirical Data, J. Geophys. Res., 92(D8), 9645–9656, 1987.

Scharffenberg, W. A. and Fleming, M. J.: Hydrologic modelling system, HEC-HMS, User's Manual, 248 pp., 2005.

Stedinger, J. R., Vogel, R. M., and Foufoula-Georgiou, E.: Frequency analysis of extreme events, in: Handbook of hydrology, edited by: Maidment, D. R., MacGRAW-HILL, New York, USA, 18.11–18.66, 1993.

USACE: HEC-1 flood hydrograph package. User's manual, US Army Corps of Engineers, Hydrologic Engineering Center, Davis, CA, USA, 1998.

Wendling, U., Schellin, H.-G., and Thomä, M.: Bereitstellung von täglichen Informationen zum Wasserhaushalt des Bodens für die Zwecke der agrarmeteorologischen Beratung, Z. Meteorologie, 41, 468–474, 1991.

Wilks, D. S.: Multisite generalization of a daily stochastic precipitation generation model, J. Hydrol., 210, 178–191, 1998.