KNMI publication; 196 - IV

Rainfall generator for the Meuse basin

Description of 20 000-year simulations

R. Leander and T.A. Buishand

De Bilt, 2008

KNMI publication = KNMI publicatie; 196 - IV

De Bilt, 2008

PO Box 201
3730 AE De Bilt
Wilhelminalaan 10
De Bilt
The Netherlands
http://www.knmi.nl
Telephone +31(0)30-220 69 11
Telefax +31(0)30-221 04 07

Authors: Leander, R.
Buishand, T.A.

Rainfall generator for the Meuse basin

Description of 20 000-year simulations

R. Leander, T.A. Buishand

KNMI-publication; 196-IV
Work performed under contracts RI-2726A and RI-4750 to Ministry of Transport, Public Works and Water Management, Institute for Inland Water Management and Waste Water Treatment RIZA, P.O. Box 17, 8200 AA Lelystad (The Netherlands).
Telephone + 31.320.298411; Telefax +31.320.294218

Summary : The sensitivity of simulated maxima of 10-day precipitation to the set of historical base years used for resampling has been assessed. Eight subsets of 33 years were extracted from the base period 1930-1998 (with the exclusion of 1940). The fraction of days with 10 mm of precipitation or more in the winter half-year and the mean winter precipitation in these subsets were both either high or low, near the edge of the 95%-confidence region of their expected values. Furthermore, the subsets differed with respect to the inclusion or exclusion of the years 1995 or 1984. With each subset a 20 000-year simulation was performed and the quantiles of the simulated 10-day winter maxima were considered. In addition to this, a 20 000-year simulation was run in which the entire base period was used. This simulation serves as a reference and is assumed to be comparable with earlier simulations. It was found that for return periods up to five years the influence of 1995 or 1984 on the 10-day maxima was negligible compared to that of the mean winter precipitation and the fraction of winter days with 10 mm of precipitation or more. For longer return periods, especially the influence of the winter of 1995 was noticeable. The 10-day precipitation with a return period of 1250 year in the eight simulations spanned a range from 165 to 210 mm (≈ 24%). The influence of the repetition of historical days in the simulation was investigated, using a modified algorithm in which such repetition was excluded. This resulted in slightly lower 10-day maxima (for a return period of 1250 years, the differences are within 5%).

1 Introduction

The GRaDE instrument (Generator of Rainfall and Discharge Extremes), currently under development for the river Meuse, aims at the estimation of the quantiles of extreme river discharges for return periods in the order of 1000 years. This is achieved by simulating long sequences of daily discharge. The GRaDE instrument comprises a weather generator, producing long sequences of daily precipitation and temperature for the different subbasins and a rainfall-runoff model for the entire basin. Both components have a number of uncertain elements affecting the reliability of the estimated discharge quantiles.

This study focusses on the weather generator, which is based on k-NN resampling of historical daily records of precipitation and temperature. An important source of uncertainty is related to the selection of the historical period from which data are resampled. It was suspected that the simulation of extreme multi-day amounts was sensitive to certain historical years containing consecutive periods of very wet days. The influence of these years on the generated precipitation sequences could in turn have consequences for the simulation of extreme discharges. Therefore, 20 000-year simulations were performed using different subsets of the historical data for resampling. Section 2 of this report describes the setup and the choice of the simulations. In Section 3 the results are discussed. A summary is presented in the final Section 4.

2 Setup of the simulations

The simulations conducted in this study are essentially based on the same model and data as those discussed in Leander and Buishand (2004) and Leander et al. (2005). In these studies two different base periods for resampling were considered. The longer of the two spanned the years 1930-1998, (except for the year 1940 for which insufficient data were available) In the simulations based on this period the resampling algorithm was driven by data from a set of stations in the basin (7 for precipitation and 2 for temperature). In a second stage the areal precipitation of the subbasins was resampled from the available records of areal precipitation A secondary nearest-neighbour step was used to substitute resampled days from *before* 1961 by their closest analogue in the period from 1961 onward. The analogy between historical days was based on the station data. The same procedure was followed for the temperature of 11 stations in and around the Meuse basin, using data for the period 1968-1998.

In the current study the same historical period (1930-1998) is considered. For the purpose of hydrological modelling, the area-average temperature for each subbasin was derived from the simulated station data by means of Thiessen interpolation. The simulated daily potential evapotranspiration (PET) for each subbasin was derived from the simulated daily temperature using a linear relation between PET and daily temperature (Leander and Buishand, 2007). The secondary nearest-neighbour step was not only used for the simulation of new sequences, but also for the artificial extension of the sequence of historical basin-average precipitation backwards in time from 1961 to 1930 (except for 1940). That constructed sequence is used to estimate two characteristics of winter precipitation and for comparison of the simulated 10-day maxima.

2.1 Construction of subsets of the historical period

The effect of the historical period was investigated by resampling from subsets of the historical period. In the context of this study it was desirable to redefine a year as a 365-day period centred around the winter half-year, because of the emphasis on extreme multi-day precipitation in winter. In this definition consecutive winter months are kept together within a year. Thence, year i consists of the last 182 days of the calendar year $i-1$ and the first 183 days of calendar year i (enclosing the entire winter of year i). The historical period of the calendar years 1930-1998, except for 1940, thus contains 66 *complete* winters (1931-1939 and 1942-1998). From these, subsets of 33 (winter half-)years were selected randomly without replacement, 14 from the period 1931-1960 and 19 from the period 1961-1998. This procedure was repeated many times, resulting in an ensemble of 50 000 subsets of historical years. It must be noted that this ensemble does not represent all possible combinations of historical years, since the number of years from before and after 1961 is fixed. These numbers are chosen such that the fraction of years for which the secondary nearest-neighbour step has to be applied is the same as in the entire data set. This setup is similar to that in applications of the bootstrap to time series of unequal length (GREHYS, 1996).

2.2 Selection of eight subsets for simulation

For each subset of 33 years in the ensemble, the fraction of winter days f with 10 mm of precipitation or more and the mean daily winter amount of precipitation P were determined, shown in Fig. 2.1. The dashed ellipse in this figure delineates a 95%-confidence region for the expectations of f and P, derived from the entire record of 66 years, under the assumption that the pairs (f, P) have a bivariate normal distribution. The univariate normality of the individual variables was confirmed by Q-Q plots. Details on the construction of the confidence region can be found in Appendix A.

Based on earlier simulations, it was suspected that particular days in January 1995 had a considerable influence on the simulated 10-day extremes (though this influence did not appear in the modelled extreme discharges from the simulations). Furthermore, it emerged from the simulations without 1995 that the winter of 1984 has some influence. The ensemble of subsets was therefore sorted into four groups. The subsets in group 1 included both 1995 and 1984, those in group 2 and 3 respectively contained either 1995 or 1984 and group 4, finally, consisted of all subsets which had neither 1995 nor 1984. From each group two subsets were selected, one with low values of f and P (denoted as '`a`') and one with high values of f and P (denoted as '`b`'). The relevant locations in the (f, P)-plane relative to the 95%-confidence region are shown in Fig. 2.1. The eight cases are clustered four by four near the edge of the region and will be referred to as `1a`, `1b`, `2a`,...,`4b`.

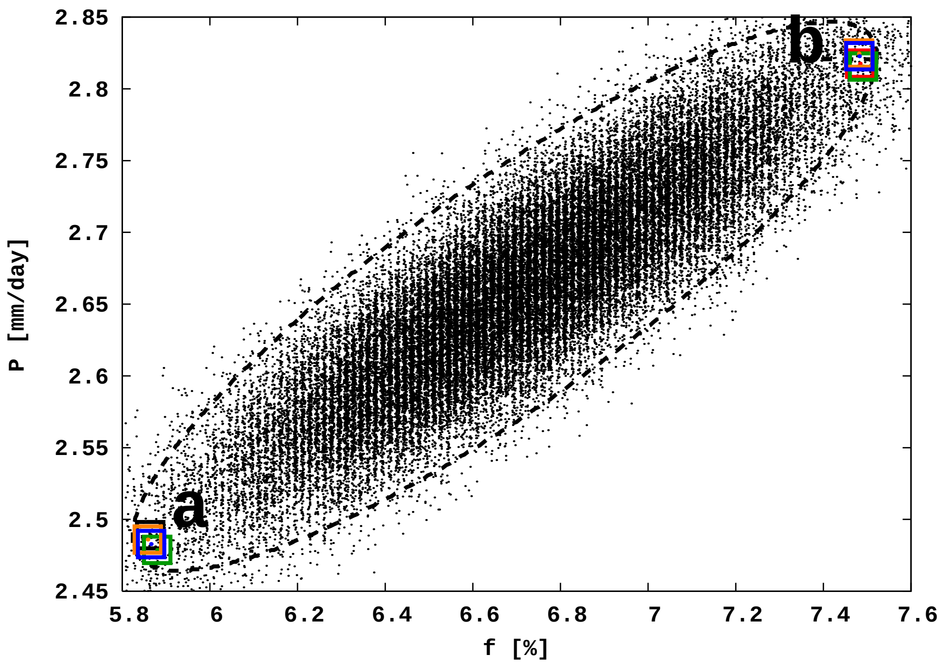

Figure 2.1: Mean daily precipitation amount P versus fraction f of days with 10 mm of precipitation or more in the winter half-year for all 50 000 sets in the ensemble (black dots). The eight sets of historical years selected for simulation are marked with boxes. The dashed ellipse marks the 95%-region for the expectations of f and P, based on the extended 66-year historical record.

3 Results

With each of the eight selected subsets of historical years, a 20 000-year simulation was performed. In addition to this, a 20 000-year simulation based on the complete historical period was available, which is used here as a reference. This simulation is expected to be comparable to the earlier 3000-year `Sim30` simulations with a 121-day moving window (Aalders *et al.*, 2004).

Figure 3.1 compares the Gumbel plots of the simulated winter maxima of 10-day basin-average precipitation for the different cases with the Gumbel plot for the extended historical series. For return periods up to five years, the plots for the simulations with the cases `b` (i.e. those located in the upper right extent of the ellipse in Fig. 2.1) are close together and are systematically above the historical quantiles. For those of the cases `a` (i.e. in the lower left extent of the ellipse) the opposite holds. The difference between `a` and `b` is roughly in the order of 10%. In this regime the winters of 1984 and 1995 are far less influential on the simulated quantiles than the mean precipitation and the occurrence of daily precipitation amounts $\geq$10 mm.

For longer return periods the influence of 1995 (and to a lesser extent 1984) is noticeable as the plots of the cases `a` as well as those of the cases `b` start to diverge (though all the `b`-cases result in higher quantiles than their corresponding `a`-cases). At return periods of approximately 100 years, the plots of cases `1a` and `2a` cross those of cases `3b` and `4b`. This implies that beyond return periods of 100 years the inclusion of certain historical years in (or the exclusion from) a subset has more influence than the aforementioned average characteristics of the subset. Furthermore, the plots of the `b`-cases suggest a larger spread than those of the `a`-cases, which implies that the influence of 1984 and 1995 also depends on the average characteristics of the set. In all simulations 10-day amounts above the largest historical maxima of 164 mm (1995) were found. However, in all simulations (including the reference simulation) the rate of occurrence of this value is less than expected from the historical record, except in those for the cases `1b` and `2b`. These are the two cases that are relatively wet and contain the winter of 1995. Apparently, for that historical event to be simulated with the same frequency as it is associated with in the historical record, the subset of historical years should be relatively wet on average as well as contain the winter of 1995.

The simulated 10-day maxima for case `3a` show the largest deviation from the Gumbel plot of the historical maxima. The significance of this deviation was investigated by generating 1500 bootstrap samples of 66 maxima. Pointwise 95%-confidence intervals were determined for the Gumbel plot of these maxima, delineated by the solid lines in Fig. 3.2. Even though this simulation is associated with the lowest plot in Fig. 3.1, the 95%-confidence intervals derived from it enclose most of the historical quantiles.

Table 3.1 lists the empirical quantiles of the 10-day precipitation maxima, obtained from the different simulations corresponding to return periods of respectively 2, 200, 500 and 1250 years. The simulations with cases `b` all result in a 2-year event (median) that is on average 10% higher than those for the cases `a`. The 2-year event from the reference simulation is found right in between cases '`a`' and '`b`'. For return periods of 500 and 1250 years the range of the quantiles for the four `b`-cases is roughly twice as large as the range

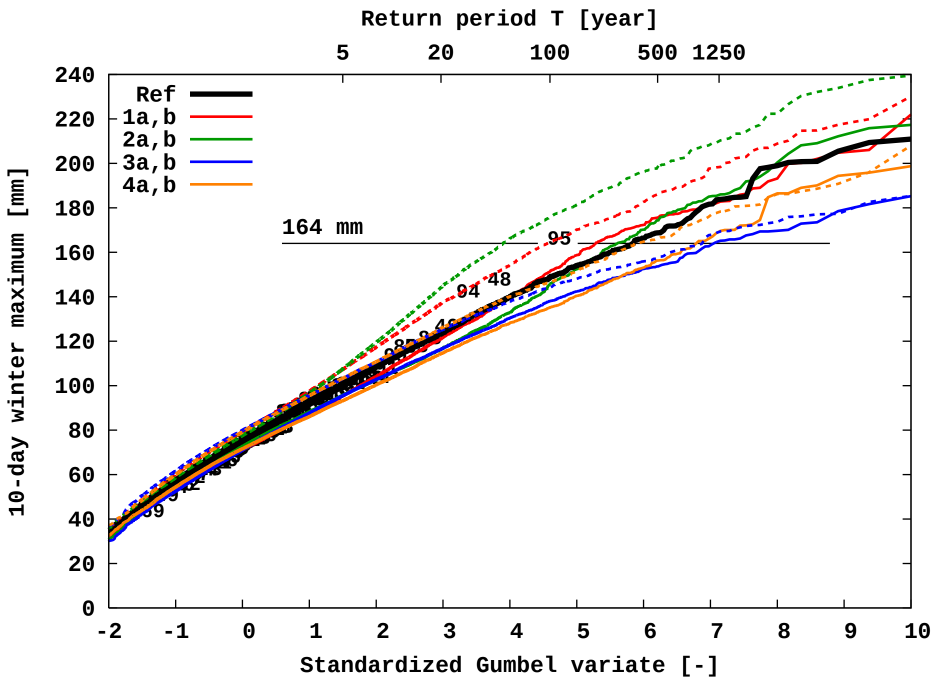

Figure 3.1: Extreme 10-day winter precipitation from resampling subsets `a` (solid), subsets `b` (dashed) and the entire historical period (‘`Ref`’, bold solid). Each of the simulations had a total length of 20 000 years. The ordered 10-day maxima in the extended historical series are indicated by the year of occurrence. The largest 10-day maximum (164 mm in January 1995) is accentuated.

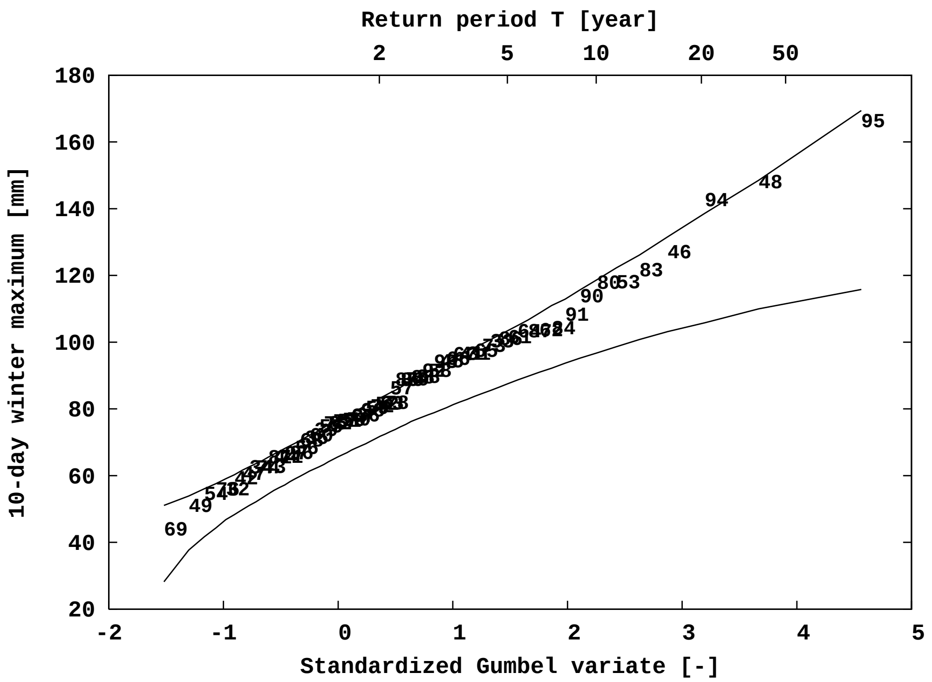

Figure 3.2: Ordered 10-day precipitation maxima in the extended historical series (winter half-year) and pointwise 95%-confidence intervals for the Gumbel plot of 10-day maxima of a 66-year sequence, obtained by bootstrapping 1500 sequences of 66 maxima from the simulation for case `3a`.

Table 3.1: Simulated quantiles corresponding to return periods of resp. 2, 200, 500 and 1250 years for 10-day winter maxima (in mm). The results are shown for the reference simulation and for each of the eight selected cases `1a` to `4a` and `1b` to `4b`. The last two columns indicate whether the winter of '95 or '84 was included in the subsets. In the final row the range of values from respectively the `a`-cases and the `b`-cases is given for each return period.

	2-yr	200-yr	500-yr	1250-yr		2-yr	200-yr	500-yr	1250-yr	'95	'84
`Ref`	81.7	157.3	168.6	183.4							
`1a`	76.9	164.3	175.6	182.6	`1b`	85.8	173.3	186.1	198.4	+	+
`2a`	78.6	157.7	174.4	185.8	`2b`	84.2	186.5	198.6	210.4	+	−
`3a`	77.5	145.5	153.7	165.0	`3b`	85.8	150.9	157.6	169.4	−	+
`4a`	77.2	143.9	156.3	169.3	`4b`	85.4	155.7	166.0	178.3	−	−
	1.4	20.4	21.9	20.8		1.6	35.6	41.0	41.0		

for the four `a`-cases. The difference between the simulations with and without 1995 is larger than the effect of the mean precipitation. For the 1250-year return period the entire range spanned by the eight simulations runs from 165 upto 210 mm (compared to 183 mm for the reference simulation).

Because of special interest in the return period of 1250 year, the 10-day event which roughly corresponds with this return period (the 16th largest) was investigated more closely in each of the four simulations `b`. In each of these simulations this event was located and dissected into daily values. These are listed in Table 3.2, along with their historical year and calendar day. In particular in the `2b` simulation a considerable contribution of the winter of 1995 is seen (mainly January 27 and 29), but also an isolated large daily amount (42.3 mm) from a day in February 1962. Repetition of the same historical day on two consecutive days occurs in all four simulations. Especially the degree of repetition in the `3b`- and `4b`-simulations is noteworthy (the slight differences among amounts obtained from the same historical day are the result of the standardisation of historical amounts and the transformation of standardised amounts back to their original scale). The mechanism responsible for this phenomenon in the simulations is not yet entirely clear. To assess its effect on the simulated 1250-year event, the simulations for cases `1b` to `4b` were repeated with a slightly modified resampling algorithm in which the sampling of the same historical day on two consecutive days in the simulation was prohibited. From these simulations the 16th largest 10-day events were extracted in the same way and their daily amounts are listed in Table 3.3. These events are slightly, but systematically, smaller than those obtained with the original algorithm. The decrease is in the range of 2.5-5%. The relative effect of the modification of the algorithm is largest in the `4b`-simulation, probably due to three days with more than 40 mm in the original simulation. Though direct repetition of historical days is prevented, multiple occurrences of certain historical days within a single 10-day stretch, however, are not. Figure 3.3 presents the changes in the Gumbel plots due to the modification of the algorithm.

Table 3.2: Daily values of basin-average precipitation in the 16th largest 10-day event (winter half-year) for each of the simulations for cases `1b`, `2b`, `3b` and `4b`. The historical years (YY) and calendar days (ddd) from which the daily amounts were sampled are also given (notation: $YY.ddd$). The last row reports the 10-day totals for each simulation.

`Sim 1b`		`Sim 2b`		`Sim 3b`		`Sim 4b`	
P [mm]	hist. date	P [mm]	hist. date	P [mm]	hist. date	P [mm]	hist. date
21.5	79.344	12.6	74.334	14.2	86.022	6.8	89.044
36.7	93.354	42.3	62.043	22.7	79.312	6.9	70.024
12.2	80.035	37.0	95.022	13.6	79.348	2.6	70.025
20.7	95.027	3.0	68.051	13.6	79.348	1.0	74.323
25.7	95.029	24.8	95.029	13.6	79.348	11.1	81.338
20.6	95.027	15.5	95.026	13.6	79.348	40.9	62.043
20.6	95.027	19.9	95.027	13.6	79.348	40.8	62.043
0.9	84.040	19.8	95.027	21.5	61.335	40.7	62.043
6.9	74.331	10.9	95.028	21.5	61.335	8.2	66.347
32.6	91.355	24.5	95.029	21.5	61.335	19.1	66.346
198.4		**210.4**		**169.4**		**178.3**	

Table 3.3: Same as in the Table 3.2, except that the resampling algorithm prevents the selection of the same historical day on consecutive days in the four simulations.

`Sim 1b`		`Sim 2b`		`Sim 3b`		`Sim 4b`	
P [mm]	hist. date	P [mm]	hist. date	P [mm]	hist. date	P [mm]	hist. date
12.6	88.042	29.9	95.025	20.4	82.279	18.1	77.306
43.1	62.043	16.0	95.026	31.6	82.280	40.0	61.030
9.9	93.355	20.4	95.027	31.9	86.295	10.1	90.024
3.3	74.023	25.3	95.029	13.3	64.321	0.2	89.358
14.4	95.024	20.4	95.027	15.9	93.356	0.9	90.028
29.6	95.025	11.3	95.028	5.2	93.358	10.0	61.032
15.8	95.026	29.6	95.025	15.9	93.356	16.3	94.005
29.4	95.025	11.9	80.035	6.4	77.317	39.7	61.030
12.9	64.321	13.1	79.348	10.6	72.324	21.3	61.335
20.1	95.027	25.1	95.029	13.6	91.001	12.8	97.345
191.1		**203.0**		**164.8**		**169.4**	

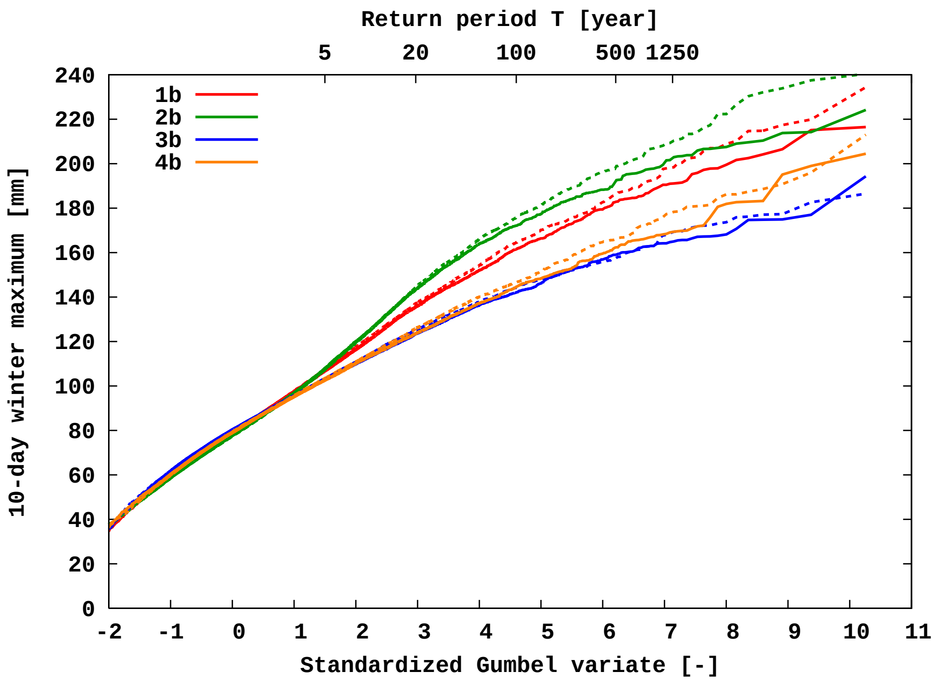

Figure 3.3: Gumbel plots of the winter maxima of 10-day precipitation from the simulations for cases `1b` to `4b` performed with the original (dashed) and the modified algorithm (solid).

4 Conclusions

The sensitivity of the simulation of extreme winter maxima of 10-day precipitation to the selection of historical years was investigated for the Meuse basin. The emphasis was on the basin-average precipitation and the historical period 1930-1998 was considered. Simulations with a length of 20 000 years were performed with various subsets of historical years from this period (the years being centred around the winter half-year by definition). The selection of the subsets for resampling was based on two average properties (the fraction of days with 10 mm of precipitation or more in the winter half-year and the mean winter precipitation) and the inclusion or the exclusion of the winters of 1995 and 1984. The Gumbel plots of the 10-day maxima for the simulations were compared.

For return periods up to five years, where the effect of individual years was negligible, the influence of the average characteristics of the subsets was particularly noticeable. Over the entire range of return periods, a higher mean precipitation and a higher fraction of days with 10 mm of precipitation or more leads to larger simulated 10-day maxima. At long return periods the plots start to diverge, in particular those with and without the winter of 1995. The range of 1250-year return values for the eight simulations runs from 165 to 210 mm. In spite of deviations between the Gumbel plots of the simulations and the empirical quantiles from the historical records, the pointwise 95%-confidence intervals for the most deviant simulation still enclosed most of the historical maxima. Beside the simulations with the subsets, a reference simulation was performed based on the entire historical period. The Gumbel plot of the 10-day precipitation maxima from this simulation approximately falls halfway the range spanned by the subsets, which shows that the influence of 1995 on the maxima reduces as the number of years taken into account increases.

The influence of the winter of 1995 is also seen in the daily amounts from which large 10-day maxima (1250-year events) are formed. In the simulations in which this winter is taken into account, the largest 10-day events generally contain many historical values originating from January 1995. The year 1984 has less influence. It was discovered that in the simulated sequences some historical days are repeatedly drawn on consecutive days. This might be considered less realistic. The influence of this phenomenon on the 10-day maxima was therefore investigated by repeating four simulations with a modified algorithm, in which such repetition of historical days was not allowed. The influence on the composition of the typical 1250-year event as well as the entire Gumbel plot was assessed. It was found that the effect of the altered algorithm on the 1250-year event was within 5%. For short and moderate return periods no effect on the 10-day maxima was found.

Acknowledgements

The work was performed in co-operation with the Institute for Inland Water Management and Waste Water Treatment (RIZA), Lelystad. The meteorological data were kindly provided by Météo France and the Royal Meteorological Institute of Belgium (RMIB).

References

Aalders, P., Warmerdam, P. M. M. and Torfs, P. J. J. F., 2004. Rainfall Generator for the Meuse Basin: 3000 year discharge simulations in the Meuse basin. Report No. 124, Sub-department Water Resources, Wageningen University, Wageningen.

GREHYS, 1996. Inter-comparison of regional flood frequency procedures for Canadian rivers. Journal of Hydrology 186, 85–103.

Leander, R. and Buishand, T. A., 2004. Rainfall Generator for the Meuse Basin: Development of a multi-site extension for the entire drainage area. KNMI-publication 196-III, KNMI, De Bilt.

Leander, R., Buishand, T. A., Aalders, P. and de Wit, M. J. M., 2005. Estimation of extreme floods of the river Meuse using a stochastic weather generator and a rainfall-runoff model. Hydrological Sciences Journal 50(6), 1089–1103.

Leander, R. and Buishand, T. A., 2007. Resampling of regional climate model output for the simulation of extreme river flows. Journal of Hydrology 332, 487–496.

Morrison, D. F., 1976. Multivariate statistical methods. McGraw-Hill, New York.

Appendix A Confidence region for the mean of a bivariate normal variate

We assume that the two-dimensional vector $\underset{\sim}{X} = (X_1, X_2)^{\mathrm{T}}$ has a bivariate normal distribution. The density of $\underset{\sim}{X}$ is given by

$$f(\underset{\sim}{x}) = \frac{1}{2\pi|\underset{\sim}{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{\mu})^{\mathrm{T}} \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{x} - \underset{\sim}{\mu})\right) , \tag{A.1}$$

where $\underset{\sim}{\mu} = (\mu_1, \mu_2)^{\mathrm{T}}$ is the vector of means and $\underset{\sim}{\Sigma}$ is the covariance matrix given by

$$\underset{\sim}{\Sigma} = \begin{pmatrix} \sigma_1^2 & \gamma \\ \gamma & \sigma_2^2 \end{pmatrix} . \tag{A.2}$$

It is easy to show that the inverse covariance matrix has the form

$$\underset{\sim}{\Sigma}^{-1} = \frac{1}{\sigma_1^2\sigma_2^2 - \gamma^2} \begin{pmatrix} \sigma_2^2 & -\gamma \\ -\gamma & \sigma_1^2 \end{pmatrix} . \tag{A.3}$$

The α% confidence region of $\underset{\sim}{\mu}$ can be determined from the sample mean $\overline{\underset{\sim}{x}} = (\overline{x}_1, \overline{x}_2)^{\mathrm{T}}$ and the sample covariance matrix $\underset{\sim}{C}$, i.e. the matrix in which the elements of $\underset{\sim}{\Sigma}$ are replaced by their sample estimates. This region is defined by the inequality (Morrison, 1976):

$$N(\underset{\sim}{\mu} - \overline{\underset{\sim}{x}})^{\mathrm{T}} \underset{\sim}{C}^{-1}(\underset{\sim}{\mu} - \overline{\underset{\sim}{x}}) \leq \frac{2(N-1)}{N-2} F_{\alpha,2,N-2} , \tag{A.4}$$

where $F_{\alpha,2,N-2}$ denotes the α%-point of the F-distribution with degrees of freedom 2 and $N-2$. The confidence region appears as a tilted ellipse in the $\underset{\sim}{x}$-plane, the tilt and elongation being determined by the covariance matrix $\underset{\sim}{C}$. In the case of f and P this matrix contains the sample variances and covariance of the amount of precipitation and the fraction of days with 10 mm of precipitation or more in each winter half-year.