Journal of Hydrology 528 (2015) 490–502

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

# Flood frequency analysis and discussion of non-stationarity of the Lower Rhine flooding regime (AD 1350–2011): Using discharge data, water level measurements, and historical records

W.H.J. Toonen *

*Dept. of Physical Geography, Utrecht University, Utrecht, The Netherlands*
*Dept. of Applied Geology and Geophysics, Deltares BGS, Utrecht, The Netherlands*

ARTICLE INFO

*Article history:*
Received 10 October 2013
Received in revised form 3 June 2015
Accepted 6 June 2015
Available online 16 June 2015
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Bruno Merz, Associate Editor

*Keywords:*
Generalised Extreme Value (GEV) distribution
Non-stationarity
Historical records
Flooding regime
Recurrence time
Rhine

SUMMARY

Accurate estimates of the recurrence time of extreme floods are essential to assess flood safety in flood-prone regions, such as the Lower Rhine in The Netherlands. Measured discharge records have a limited length and are, in general, poorly representing extremes, which results in considerable uncertainties when used for flood frequency analysis. In this paper, it is shown how alternative discharge monitoring stations along the Rhine, measurements of water levels, and historical records can be used to increase data availability. Although pre-processing and the conversion of data types into discharge estimates introduces extra uncertainty, the added value of this data in flood frequency analysis is considerable, because extending record length by including slightly less-precise data results in much better constrained estimates for the discharges and recurrence intervals of extreme events. Based on results obtained with the Generalised Extreme Value (GEV) distribution, it was concluded that large floods of the last century are presumably rarer than previously considered using shorter data series. Moreover, the combined effect of climatic and anthropogenic-induced non-stationarities of the flooding regime is more easily recognised in extended records. It is shown that non-stationarities have a significant effect on the outcomes of flood frequency analysis using both short and long input data series. Effects on outcomes of dominant multi-decadal variability are, however, largely subdued in the longer 240-year series.


## 1. Introduction

Floodplains of the Lower Rhine river in the Netherlands have been embanked completely since the 14th century to protect against floods (Hesselink, 2002). A national flood protection act (Waterwet, 2009) currently enforces protection against a flood with a statistical recurrence time of 1250 years: the design flood ($Q_{1250}$ at the Lobith monitoring station, estimated at ~16,000 $m^3 s^{-1}$). To counter possible effects of climate change, it has recently been proposed to further increase safety levels to protect against a discharge of ~18,000 $m^3 s^{-1}$ (Delta Programma, 2014). The value of $Q_{1250}$ is periodically assessed from statistical extrapolation of observational discharge data. Two broadly acknowledged problems in the accuracy and validity of such design flood estimations are (i) statistical uncertainties in estimated recurrence intervals of extreme events (Klemeš, 2000), and (ii) non-stationarities of the flooding regime (Knox, 1993; Redmond et al., 2002; Milly et al., 2008).

Present estimates of the magnitude of $Q_{1250}$ of the Lower Rhine yield relatively large uncertainties (Chbab et al., 2006; te Linde et al., 2010). Estimates of $Q_{1250}$ are based on extrapolation of discharge measurements since AD 1901. This ~110 years of measurements is a limited interval for the prediction of extreme events, as it is likely that a short period lacks a proper representation of the distribution of high-magnitude flood events. This complicates accurate estimation of flood probabilities in the millennial recurrence domain (Klemeš, 2000). The sensitivity of design flood estimates to the timing of large events was demonstrated in the $Q_{1250}$ re-assessment that followed the large AD 1993 and 1995 Rhine floods with estimated recurrence times of respectively 35 and 80 years (Chbab, 1999). The addition of these discharges as data-points to the extrapolation resulted in a ~7% increase in the estimate of $Q_{1250}$ (from 15,000 to 16,050 $m^3 s^{-1}$; Chbab, 1996).

Previous exercises to assess $Q_{1250}$ and its uncertainty mainly depended on increasingly advanced numeric simulation techniques. This included analysis of stochastically resampled 20th

* Address: Heidelberglaan 2, 3584 CS Utrecht, The Netherlands. Tel.: +31 643085070.
*E-mail address:* w.h.j.toonen@gmail.com

http://dx.doi.org/10.1016/j.jhydrol.2015.06.014
0022-1694/

century datasets, the use of precipitation generators in combination with run-off models, and Bayesian weighting of different extrapolation methods (Chbab et al., 2006; de Wit and Buishand, 2007; te Linde et al., 2010). Although this maximised the use of existing datasets and created useful scenarios for engineering, these approaches still suffered from difficulties in fundamentally improving the reliability of predicted discharges of rare high-magnitude events, as they still depended on possibly biased datasets with a too limited length to represent the natural system. Moreover, functions to describe extreme value distributions that are routinely used in flood frequency analysis are mainly chosen based on their goodness of fit on gauged data and lack specific hydrological justification (e.g., Kidson and Richards, 2005). This implies that statistically more-sophisticated approaches do not necessarily lead to better characterisation of the hydrological system and may be poor predictors of rare extreme events despite their goodness of fit on more frequently occurring low-magnitude events.

Non-stationarity of the flooding regime further complicates flood frequency analysis. Non-stationarity includes changes in the relationship between absolute discharges and probability over time. In general it is assumed that small changes in climate and meteorological patterns can have large effects on the occurrence of extremes and general flood probabilities (Knox, 1993). The combined effect of human influence and climatic variability is the main driver of non-stationarity in flooding regimes. Human-induced non-stationarity is the result of deforestation and other land use changes in the catchment, ongoing river management, and in recent decades anthropogenically-induced global warming which is thought to increase the frequency of extreme precipitation events. Deforestation of the Rhine hinterland, initiating in the Neolithic Period and reaching an optimum in the Medieval Period (Kalis et al., 2003; Lang et al., 2003), has influenced effective runoff, which increased the probability of large discharges downstream in the catchment (e.g., Hundecha and Bárdossy, 2004). River engineering works along the Rhine have intensified since the early 19$^{th}$ century (Silva et al., 2001; Lammersen et al., 2002). Embankment, fixation and deepening of the channel bed, shortening the river flow path by artificial cut-offs, and the construction of retention basins, have changed effective downstream flood pulse propagation and thus altered the flood regime. Several studies (e.g., Engel, 1997; Pinter et al., 2006) suggest that both the occurrence of extreme events and common floods of the Rhine have increased significantly in the 20$^{th}$ century, partly due to human influence on land use and channel morphology. Available discharge data does, however, not cover the pre-management era, making comparison and quantification of this effect difficult.

Climate-induced non-stationarity of flooding regimes is widely recognised, especially during climatic anomalies such as the Little Ice Age (LIA; e.g., Barriendos and Martin-Vide, 1998; Swierczynski et al., 2012; Toonen, 2013). Various studies have shown that decadal to millennial periodicities can be identified in Holocene flood records (e.g., Macklin and Lewin, 2008; Brázdil et al., 2011), and that these periodicities correlate with trends in meteorological patterns or solar activity (e.g., Redmond et al., 2002; Czymzik et al., 2010, 2013). The existence of periodicities or anomalous episodes in the occurrence of floods is problematic for flood frequency analysis, as estimates become biased to a certain mode of flooding when only part of the entire period is represented in the sample data sets or when much of the dataset was gathered in a situation that differs from the present. For characterising the present situation, it is important to correct (i.e. normalise) data from periods that differ from the current state of the flood regime. In this respect, generally only non-natural and non-periodic changes (e.g., caused by engineering projects), or changes that are caused by climatic cycles that act on longer time scales than the used data series, need to be assessed. Decadal and even centennial natural variability can be incorporated in flood frequency analysis provided that they are an intrinsic part of the natural fluvial system, and sufficiently covered in data series (Macdonald and Black, 2010).

To achieve better estimates of the discharge of the design flood, and to be able to recognise the effect of non-stationarity, the most straightforward approach is to extend data series back in time. The current reference dataset for the Lower Rhine in the Netherlands is the discharge measurement series since AD 1901 at Lobith ($Q_{Lob}$; Fig. 1). Systematic discharge measurements at Cologne (Germany; Fig. 1) go back to AD 1817, continuous water level records go back to AD 1772 at several gauging stations in the direct vicinity of Lobith (Fig. 1), and reliable historical records of Rhine floods are available from c. AD 1350 (Glaser et al., 2010). At present, the potential of these additional records has not systematically been explored for estimating the $Q_{1250}$ of the Lower Rhine in the Netherlands. Inclusion of discharge data and water level measurements from alternative locations would roughly double the length of the dataset for calculation of $Q_{1250}$. This provides, together with information from historical records, valuable information to assess non-stationarity over multi-decadal to centennial time-scales.

The goals of this paper are (i) to reduce uncertainty in the estimation of $Q_{1250}$ of the Rhine at Lobith by extending data series using alternative data types, and (ii) to analyse the possible effects of non-stationarity of the flooding regime on the distribution of extreme events and the calculated discharge of the design flood. Alternative data types are combined with existing discharge series in order to extend the reference period for estimating $Q_{1250}$. The Generalised Extreme Value (GEV) distribution was applied in a frequency-magnitude analysis on the compiled discharge series. Uncertainties in the original and converted data are carefully

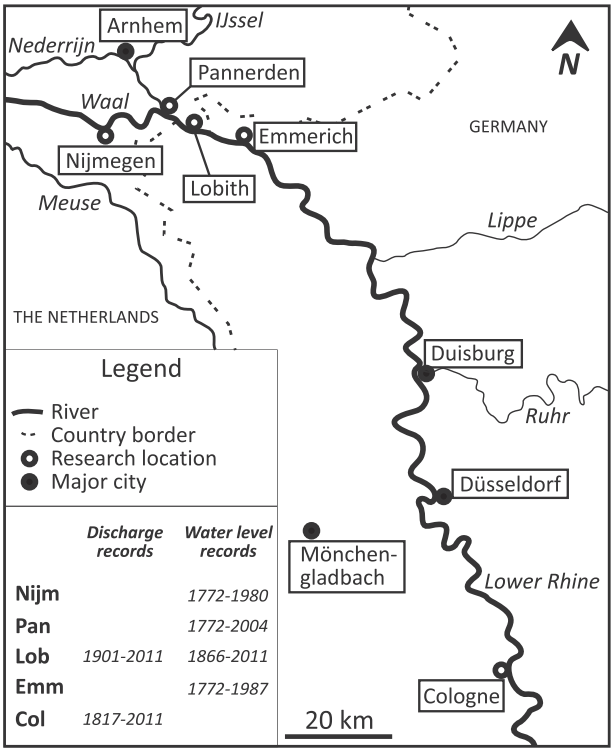


**Fig. 1.** Map of the research area between Cologne (Germany) and the apex of the Rhine-Meuse delta (The Netherlands). The inset shows the length (years AD) of discharge and water level data records used in this paper. Abbreviations in the inset relate to research locations on the map.

assessed and quantified, and included in the extrapolation exercise. In an *explorative* attempt to further increase discharge data availability, the use of intensity classes to generate discharges from historical data was explored (Glaser and Stangl, 2003).

## 2. Discharge estimates from alternative sources

Data was gathered from several locations and source types, which asks for different pre-processing to merge them into a single discharge record. Depending on the location, data type, and age of the measurements, discharge data comes with different uncertainties. This section elaborates on the methodology used to generate a (as) uniform flood record (as possible) and specifies the uncertainties present. Some assumptions were, however, necessary to make a sensible comparison and combination of datasets possible (especially in the explorative assessment of the use of historical records in flood frequency analysis); these are listed together with traditional statistical procedures.

### 2.1. Modern discharge data [AD 1901–2011]

Daily discharge measurements at Lobith go back to AD 1901 (Waterbase, 2012). Modern discharge data is often used in flood frequency analysis as a uniform series assuming similar uncertainties in measurements, or without consideration of uncertainties at all. It is, however, important to acknowledge that different methods have been used over time to record discharges, all with different precision. As this paper aims to include uncertainties in the flood frequency analysis, it is important to quantify uncertainty in different periods – also in the modern era.

From AD 1901 until c. 1950, discharge estimates for the Lower Rhine were based on the velocity of floating sticks on the water surface. Because this method only measured flow velocities at the surface, calculation of the total discharge depended on standardised more recent depth-velocity profiles, which results in a ~10% uncertainty (pers. comm. Rijkswaterstaat, Dienst Oost-Nederland, Meet- en Informatiedienst). From AD 1950–2000, current meters were used to produce velocity-depth profiles for each flood. The uncertainty of this method is estimated to be ~5% of the total discharge. Since AD 2000, Acoustic Doppler Current Profilers have been used. The uncertainty of this method is negligible, with exception of discharges of small floods just exceeding bankfull discharge (for the Lower Rhine 4000–8000 $m^3 s^{-1}$; Middelkoop, 1997), when large parts of the inundated floodplains have a limited water depth and discharges are estimated by interpolation. In this range, an uncertainty of 5% applies to the data. Modern discharge data was analysed in its raw form and in a normalised (homogenised) form, which included the effects of river management of the last century. The phasing and magnitude of normalisation was based on conversion formula in Parmet et al., 2001. Extended data series were not normalised, so those results should be mainly compared with non-normalised modern discharge data. In this paper, specified measurement uncertainties were used for Monte Carlo resampling in flood frequency analysis (Section 3.1).

### 2.2. Discharges Cologne [AD 1817–2011]

Annual peak discharges at the German city of Cologne were used to extend measured discharge series for the Lower Rhine in the Netherlands back to AD 1817. The distance along the river between Lobith and Cologne is roughly 160 kilometres. As both sites are located on the same fluvial trunk valley, share the same hinterland, and have only minor tributaries joining in between (Ruhr and Lippe; Fig. 1) discharges at Cologne correlate strongly with $Q_{Lob}$, based on a simple linear regression analysis ($R^2 = 0.96$; Fig. 2) for annual peak discharges in the period AD 1901–2011. Data was screened for multiple annual maxima that originated from a single flood wave; for example the peak discharge of 1994 occurred in the first half of January, but is merely the aftermath of the same flood wave that caused the larger peak discharge of December 1993. In five years, the lower annual maximum of successive years (e.g., 1994) was replaced by the second largest flood maximum of that year (occurring for 1994 in April). Next, the timing of peak discharges at Cologne and Lobith was compared to ensure that the regression was built solely on contemporaneous events; $Q_{Lob}$ occurring at maximum 2 days after the peak discharge at Cologne. Annual maxima at both stations not necessarily relate to the same flood wave, as (especially in years of low peak discharges) local factors, such as discharge from minor tributaries joining the main Rhine between Cologne and Lobith, may have caused a different timing of annual maxima. Lobith is located more than 150 km inland, so the effect of tides on discharge differentiation between stations can be ruled out. Eleven 'mismatched floods', all of a magnitude less than bankfull discharge, were replaced to fit the same discharge wave at both locations; as this study focuses on the Lobith station, data from Cologne was replaced to fit the downstream information.

A linear regression (Fig. 2) was used to convert discharge at Cologne into discharge for Lobith. Measurement imprecision was not incorporated to allow this exercise – including uncertainty for the Cologne measurements at this stage would probably produce a too wide range of possible outcomes for realistic application. In the overlapping period of discharge measurements per location [AD 1901–2011], Cologne-predicted discharges for Lobith ($Q_{Lob\_QCol}$) were compared with truly measured discharges. Despite the good correlation, the uncertainty interval of $Q_{Lob\_QCol}$ amounts ~±12.7% of the measured $Q_{Lob}$ (in this paper two-sided uncertainty intervals are based on the range of twice the standard deviation). This is probably caused by the various ways a flood wave can evolve over short distances – due to contributions of minor tributaries and discharge wave dispersion by floodplain inundation. During the first and second World Wars, there seems to be a particularly poor correlation, likely due to limited precision of measurements (the Cologne measuring station and several bridges were destroyed during World War II). Exclusion of war periods does, however, not significantly increase the correlation between the monitoring sites and does not decrease the variance significantly. Based on the linear regression, annual peak discharges between AD 1817 and 1900, estimated from Cologne data, were added to the Lobith record (Fig. 2).

### 2.3. Water level measurements [AD 1772–1987]

Daily water level measurements at Lobith ($H_{Lob}$) are available from AD 1866 onwards (Waterbase, 2012); fragmental data records are available for Lobith since AD 1824. For the period AD 1772–1866, water levels were reconstructed for Lobith, based on information from the nearby gauging stations at Emmerich, Pannerden and Nijmegen (Fig. 1). To convert water level information from alternative sites into discharge estimates for Lobith, trends in water level due to different local *QH*-relations were removed from all locations. Detrended water levels from alternative sites were correlated with measurements from Lobith in overlapping periods. Based on a linear regression, water level measurements from alternative locations before AD 1866 were converted into a water level estimate for Lobith. The average of predictions from alternative sites was then used to estimate the water level for Lobith ($H_{Lob\text{-}region}$) and calculate the associated uncertainty interval. The estimated $H_{Lob\text{-}region}$ was converted into discharges ($Q_{Lob\_Hregion}$), using a stage-discharge relation at Lobith.

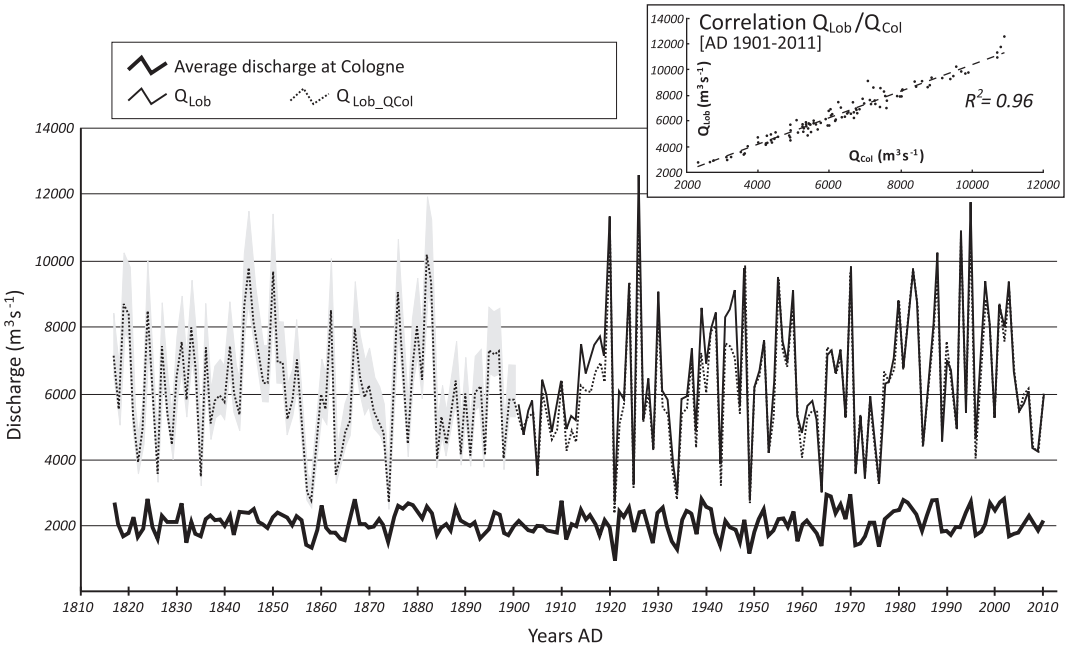


**Fig. 2.** Annual average discharges at Cologne, predicted annual peak discharges for Lobith based on measurements from Cologne [$Q_{Lob\_QCol}$: AD 1817–2011], and annual peak discharges at Lobith [$Q_{Lob}$: AD 1901–2011]. The linear regression (inset at top right; 1901–2011) was used to estimate discharges for Lobith back to 1817. The uncertainty interval for $Q_{Lob\_QCol}$ is indicated with a grey shade.

Trends in local water levels are most likely related to changing bed morphology, as average discharges did not change significantly in the investigated period (Fig. 2). In an attempt to detect and remove the effects of changing bed levels, change point analysis (CPA; Taylor, 2000) was used to identify breaks in annual average water levels (Waterbase, 2012). The change points, detected from changes in the cumulative sum using bootstrapping techniques, divide phases with different trends in water levels. Two important knick-points in water levels were observed (CPA > 95% confidence level) for all investigated sites in 1856 and 1942 (Fig. 3). Intervals divided by these years were detrended separately. After 1850, major works commenced in the Rhine, aimed at reducing flooding risk and improving navigation. Along large sections of the river, groynes were constructed to reduce channel width (and increase depth), main channels were dredged, bars were removed, meanders were cut-off and weirs were constructed (Silva et al., 2001; Vorogushyn and Merz, 2013). Before that period, the channel bed was degrading on average ~7 mm per year at Emmerich, while some aggradation occurred downstream. From 1856 to AD 1942, no significant changes in water levels are distinguished at Lobith (at the other stations some aggradation occurred; Fig. 3). The second knick-point in water levels (AD 1942) is presumably also due to long-term effects of river management (and dredging), although no specific project can be targeted. Since AD 1942, channel bed degradation has occurred at Lobith with an average rate of ~11 mm per year. The Pannerden station also shows some bed degradation, although at a slower rate. Emmerich and Nijmegen show trends of aggradation in the last measured decades. The recent decades of accelerated incision are not documented due to a limited length of measurement series at those stations. Based on the strong resemblance with other sites though (Fig. 3), it is assumed that recently incision has also occurred at Emmerich and Nijmegen.

After data series screening (similar to the Cologne discharge series), annual maximum water levels from Emmerich ($H_{Emm}$), Pannerden ($H_{Pan}$), and Nijmegen ($H_{Nijm}$) were detrended using the identified change points (Fig. 3) and correlated to $H_{Lob}$ with a regression-analysis in overlapping time periods (1866–1942). Water level data after 1942 were excluded for building this regression, because accelerated bed degradation in that period troubles accurate analysis (Fig. 3). Predicted water levels for Lobith from alternative sites yield only a 2–3% difference with truly observed water levels, demonstrating the quality and consistency of water level measurements after screening. The regressions allow the

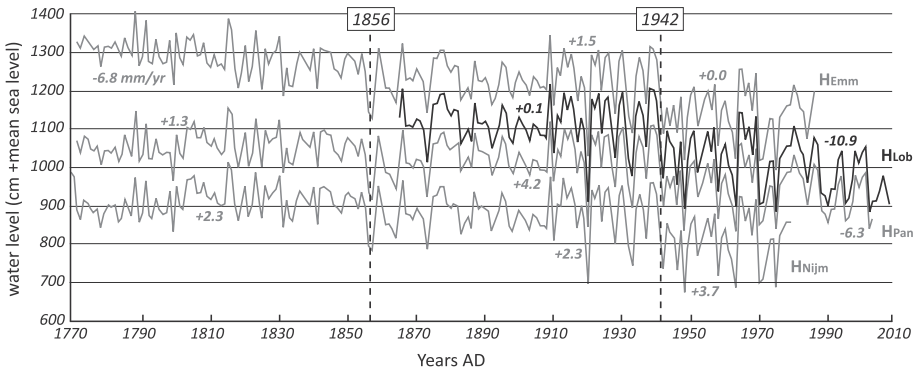


**Fig. 3.** Annual average water levels from all research locations with identified change points (AD 1856 and 1942) and information on trends in channel bed aggradation or degradation (in mm/year).

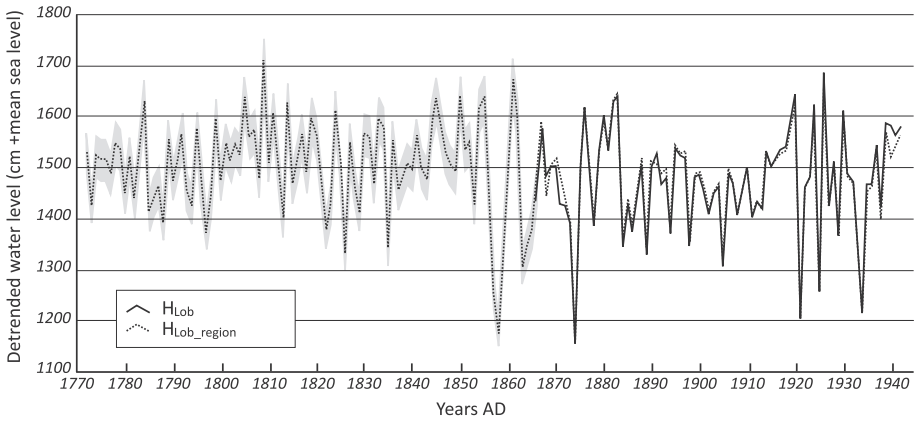


**Fig. 4.** Detrended measured maximum water levels at Lobith [1866–1942] and predicted water levels [1772–1942] with the uncertainty interval in grey.

extension of water level data for Lobith from 1866 back to 1772 (Fig. 4), based on the longer time series at the alternative sites. Predictions for Lobith were based on the average from three alternative sites. Based on cross-correlations of all sites, deviating predictions were further investigated. If a measurement from only a single location resulted in an outlier, this location was excluded from the average (Table 1). Most outliers match severe winters with documented local ice jamming (IJnsen, 1981; Driessen, 1994), and major floods with numerous dike breaches. Ice jamming occurred often in the past (e.g., Buisman, 2000, 2006), but it influenced annual-maximum water levels at the stations in only a few years (1781, 1799, 1809, 1820 and 1830). For these years the highest observed water levels do not correlate among the stations. In years when ice jams are reported, but no large-scale dike breaches occurred, the lowest regionally recorded water level was chosen (Table 1). For the 1799 and 1809 flood peaks, the average water level from all stations was used, because ice jams raised water levels, but at the same time dike breaches in between gauging stations lowered local water levels considerably. Note that for dike breaches to happen on a regional scale, a large discharge must have occurred.

A stage-discharge relation was used to convert predicted $H_{\text{Lob-region}}$ into discharges. The *QH*-relation for annual maxima in the period 1817–1942 was based on discharges (Fig. 2) from Cologne-based predictions (1817–1900) and direct Lobith measurements (1901–1942), and water levels (Fig. 4) from the various alternative research locations (1817–1865) and Lobith (1866–1942). Data after 1942 was excluded, because bed degradation accelerated and channel geomorphology changed significantly, complicating the use of a single *QH*-relation for the entire period (van Vuuren, 2005). Years which fitted poorly to the initial *QH*-relation were further investigated. Depending on the possible cause for outliers (Table 1), several years were excluded (Fig. 5). Years with high discharges were not removed from the correlation as these are natural outliers in *QH*-relations.

The minor uncertainty range in reconstructed water levels (Fig. 4) propagates in the application of the *QH*-relation back to 1772 (Fig. 5). This results in a ~12% uncertainty range for estimated discharges, as it was calculated from both the minimum and maximum estimates of the water level uncertainty range. It was assumed that the 1817–1942 *QH*-relation was also valid for the period before 1817, as water level data suggest no major changes in channel morphology in this period (Fig. 3). Comparison of reconstructed discharges in the period 1817–1900, either deduced from $Q_{\text{Lob\_QCol}}$ or $Q_{\text{Lob\_Hregion}}$, shows both results to correspond well. Major differences only exist in 1855 and 1861 (Fig. 5). Based on regional water level measurements, these are relatively large floods, while the predictions from Cologne suggest smaller discharges. In these years reconstructed water levels deviate only several centimetres among sites in the delta apex, which excludes ice jamming as a major factor. From this it is concluded that for these two floods the measured

**Table 1**
Overview of poorly correlating water level data and years that are excluded for establishing the stage-discharge relation at Lobith.

| Year | Type | Action | Cause (references) |
|---|---|---|---|
| 1773 | $H_{\text{Pan}}$ outlier | $H_{\text{Pan}}$ excluded | Possibly the initiation of the construction of the Bijlands Canal by artificial meander cut-off. |
| 1781 | All sites *H* outliers | Minimum $H_{\text{Emm}}$ adopted | Ice jams (Driessen, 1994) |
| 1795 | $H_{\text{Pan}}$ outlier | $H_{\text{Pan}}$ excluded | Ice jams (Driessen, 1994) |
| 1799 | All sites *H* outliers | – | 12 dike breaches (STIBOKA, 1975), ice jams (Driessen, 1994) |
| 1803 | $H_{\text{Pan}}$ outlier | $H_{\text{Pan}}$ excluded | Probably ice jams (IJnsen, 1981) |
| 1809 | All sites *H* outliers | – | Large *Q*, 9 dike breaches (STIBOKA, 1975), ice jams (Driessen, 1994) |
| 1811 | Error $H_{\text{Emm}}$ | $H_{\text{emm}}$ excluded | Ice jams (Driessen, 1994) |
| 1820 | All sites *H* outliers | Minimum $H_{\text{Emm}}$ adopted | Ice jams, 4 dike breaches (STIBOKA, 1975; Driessen, 1994) |
| 1821 | $H_{\text{Pan}}$ outlier | $H_{\text{Pan}}$ excluded | – |
| 1830 | All sites *H* outliers | Minimum $H_{\text{Emm}}$ adopted | Probably ice jams (IJnsen, 1981) |
| | *QH*-outlier | Excluded for *QH* | |
| 1834 | *QH*-outlier | Excluded for *QH* | – |
| 1841 | $H_{\text{Pan}}$ outlier | $H_{\text{Pan}}$ excluded | Probably ice jams (IJnsen, 1981) |
| 1854 | $H_{\text{Emm}}$ outlier | $H_{\text{Emm}}$ excluded | – |
| | *QH*-outlier | Excluded for *QH* | – |
| 1855 | *QH*-outlier | Excluded for *QH* | Probably ice jams, 1 dike breach (STIBOKA, 1975; IJnsen, 1981) |
| 1861 | *QH*-outlier | Excluded for *QH* | 3 dike breaches (STIBOKA, 1975) |
| 1891 | *QH*-outlier | Excluded for *QH* | Probably ice jams (IJnsen, 1981) |
| 1940 | *QH*-outlier | Excluded for *QH* | Ice jams (Rijkswaterstaat, 1942) |

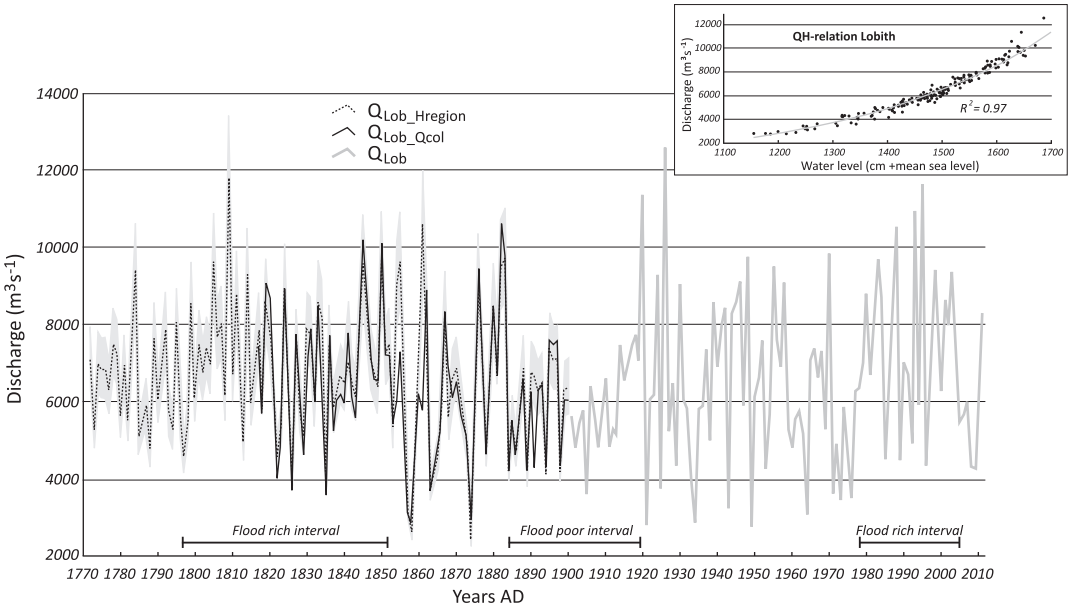


**Fig. 5.** Predicted annual peak discharges for Lobith [AD 1772–1900] with uncertainty interval (grey shade), using reconstructed water levels (conversion by the 1817–1942 stage-discharge relation for Lobith; see inset), and discharge estimates for Lobith, based on discharge data from Cologne.

**Table 2**
Class definitions to generate indexed flood intensity series (shown in Fig. 6) from historical flood records and modern discharges.

| Class/ score | Description | Recurrence time (AD 1350–1772) | Discharge (AD 1772–2011 extrapolation – this study) |
|---|---|---|---|
| 0 | +No dike breach recorded | n.a. (in 270 years) | <7160 $m^3 s^{-1}$ |
| 1 | +Single dike breach<br>+Minor damage | 2.8 years (153 events) | >7160 $m^3 s^{-1}$ |
| 2 | +Multiple dike breaches<br>+Moderate damage<br>+Local impact | 5.0 years (84 events) | >8100 $m^3 s^{-1}$ |
| 3 | +Multiple dike breaches<br>+Major damage and life losses<br>+Breaches in Lower Rhine Valley (Germany) and along all deltaic distributaries (the Netherlands) | 11.4 years (37 events) | >9170 $m^3 s^{-1}$ |

discharge at Cologne is not representative for downstream discharges, most likely due to high discharges from the Lippe and Ruhr tributaries or as a result of an error in the measurement.

### 2.4. Historical records [AD 1350–1772]

Historical records of flood events and dike breaches were gathered for the entire Lower Rhine region (compilation from STIBOKA, 1975; Driessen, 1994; Buisman, 1996, 1998, 2000, 2006 and references therein). Information before AD 1350 is not used in this study, because the accuracy of records reduces in this period and source density varies greatly (Glaser and Stangl, 2003). After 1350, sources are assumed to be reliable and complete, because multiple sources verify the occurrence and magnitudes of events; harbour records from multiple towns, and records describing the extent of the damage and the costs associated with dike restoration. Moreover, historical records are continuously available for the region, also in years in which no river flood was documented. The description of other natural events (e.g., storm surges, tempests, droughts, and the general weather) in those years suggests that floods simply did not occur instead of not being documented or records being lost. It is therefore assumed that all relatively large floods that have caused damage are documented in historical records, so variability in specific flood record density is directly related to flood activity.

Historical records reflect relative flood magnitudes, as they basically are damage reports. A by-proxy magnitude of events was obtained following the classification approach of Sturm et al. (2001) and Glaser and Stangl (2003). The geographical spread of reports and extent of damage/life-loss determines class membership and scoring of a flood for the period 1350–1772 (Table 2). To describe trends in flooding, annual scores were averaged over 31 and 101-year periods (arbitrarily chosen and respectively corresponding with a traditional 'climate' window and the length of the original Lobith discharge record), resulting in an indexed flood intensity (Fig. 6). For the period AD 1772-present, a similar flood intensity curve was constructed. Discharges were converted into scores, based on the statistical recurrence of the classes (taken over 1350–1772; Table 2) and associated discharge from flood recurrence-magnitudes relations (1772–2011; this study). The comparison of indexed data from difference periods assumes a degree of stationarity with respect to flood protection (dike strength) and the description of flood damage (discussed in Section 5.3).

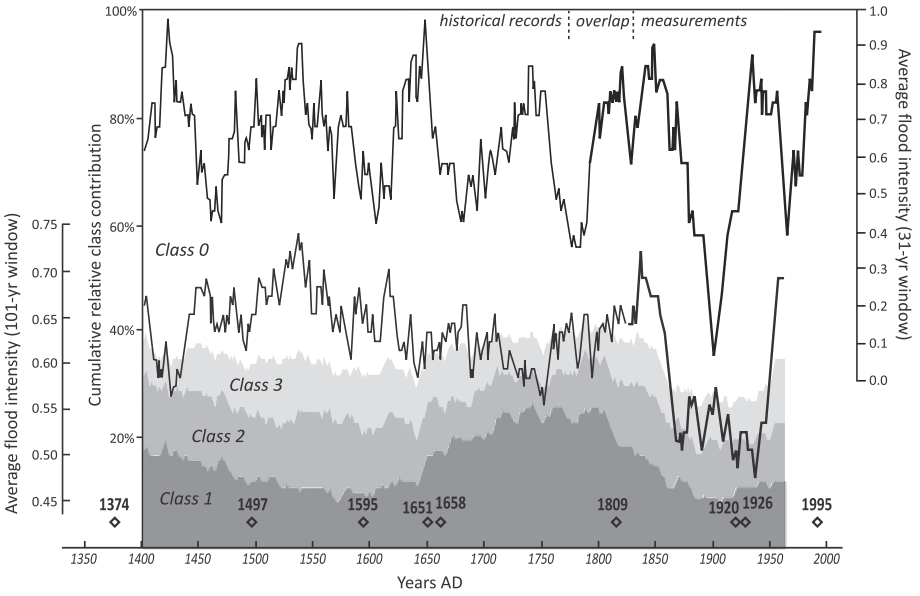


**Fig. 6.** 31-year (upper curve) and 101-year (lower curve) averaged flooding intensity of the Lower Rhine, based on historical records and measured discharges at Lobith (indicated with thick line). Mid-point values of the averaged windows are plotted. Relative contributions of each flood intensity class (Table 2) are plotted for the 101-year window in grey shades in the background. The timing of large events exceeding 11,000 $m^3 s^{-1}$ (see main text) is illustrated by diamonds on the lower axis.

## 3. Flood frequency analysis

### 3.1. Generalised Extreme Value (GEV) approach

A flood frequency analysis was undertaken to estimate discharges of low-probability events. To describe the extreme value distribution, and to estimate the recurrence intervals of extreme events, the cumulative distribution function of the Generalised Extreme Value distribution (GEV; Eq. (1)) was used (Fisher and Tippett, 1928; Hosking et al., 1985). In Eq. (1), $\mu$ is the location parameter (where the origin of the distribution is positioned), $\sigma$ a scaling parameter (describing the spread of data), and $\xi$ a shape parameter (controlling the skewness and kurtosis of the distribution). The shape parameter is essential for describing extreme value distributions, because it allows a 'flexible' fit on the upper tail. When $\xi = 0$, the distribution reflects the type I (Gumbel) distribution, $\xi < 0$ is a type II (Fréchet) distribution, and $\xi > 0$ corresponds with a type III (Weibull) distribution. Parameter values were established using probability weighted moments (pwm; Hosking et al., 1985). Ad-hoc tests indicated that the fit with the pwm method of the GEV on the data was slightly better than using maximum likelihood estimators (as used in Stedinger et al., 1988).

$$F(x) = exp\left\{-\left[\frac{x-\mu}{\sigma}\right]^{\frac{1}{\xi}}\right\} \quad \xi \neq 0 \tag{1a}$$

$$F(x) = -exp\left\{-exp\left[-\frac{x-\mu}{\sigma}\right]\right\} \quad \xi = 0 \tag{1b}$$

Although many different approaches, distributions and fitting methods to describe flood data exist, flood frequency analysis in this paper is restricted to the GEV-approach. This is because focus is not on the comparison of different statistical methods which probably induce as much uncertainty in design discharges as measurements (Chbab et al., 2006; Merz and Thieken, 2009), but on extending data records and to assess non-stationarity with this data. The choice for the GEV distribution was based on previous studies (e.g., Chbab et al., 2006; te Linde et al., 2010) showing a good fit of this distribution on the flooding regime of the Rhine. Moreover, the three-parameter GEV distribution is capable of flattening-off at extreme values (introducing an upper bound; Guse et al., 2010) by having a flexible tail, induced by the shape parameter. This hypothetically corresponds with realistic hydrological bounds in a lowland setting, as the relatively low valley shoulders upstream of Lobith have a limiting effect on discharges that can be conveyed further downstream. This upper bound was calculated from the arbitrarily chosen probability of $1 * 10^{-6}$; tests indicate that in this domain the GEV extrapolation resembles a limit.

To incorporate uncertainties and probability distributions of measured and reconstructed data, Monte Carlo simulations (1000 repetitions) were used. Values for each data point in the period AD 1772–2011 were resampled within the uncertainty interval of the screened measurement data, assuming a normal distribution of the variance. The range of this uncertainty interval differs according to data type (Section 2).

### 3.2. Discharge data generation from historical records

In an explorative and novel attempt to generate discharge data based on historical information, the relationship between the flood intensity index (average 101-year score; Fig. 6) and the shape of the GEV distribution was tested for AD 1772–2011. The idea was that if the flood intensity index is strongly related to the shape of the GEV-distribution, historical flood intensities can be used to generate GEV parameter values and thus the shape of the GEV-distribution. Using these GEV-parameters in combination with a sampled flood recurrence probability, allows then simulating annual peak discharge data. This exercise is mainly designed to test the suitability of historical data for generating continuous discharge data as input for flood frequency analysis with consideration of non-stationarity. Hence a full consideration of uncertainties of this explorative method is not provided and Monte Carlo simulations have not been performed.

Flood frequency analysis was repeatedly undertaken for a 101-year moving window over discharge series in the period AD 1772–2011. GEV-parameters of these 101-year intervals were

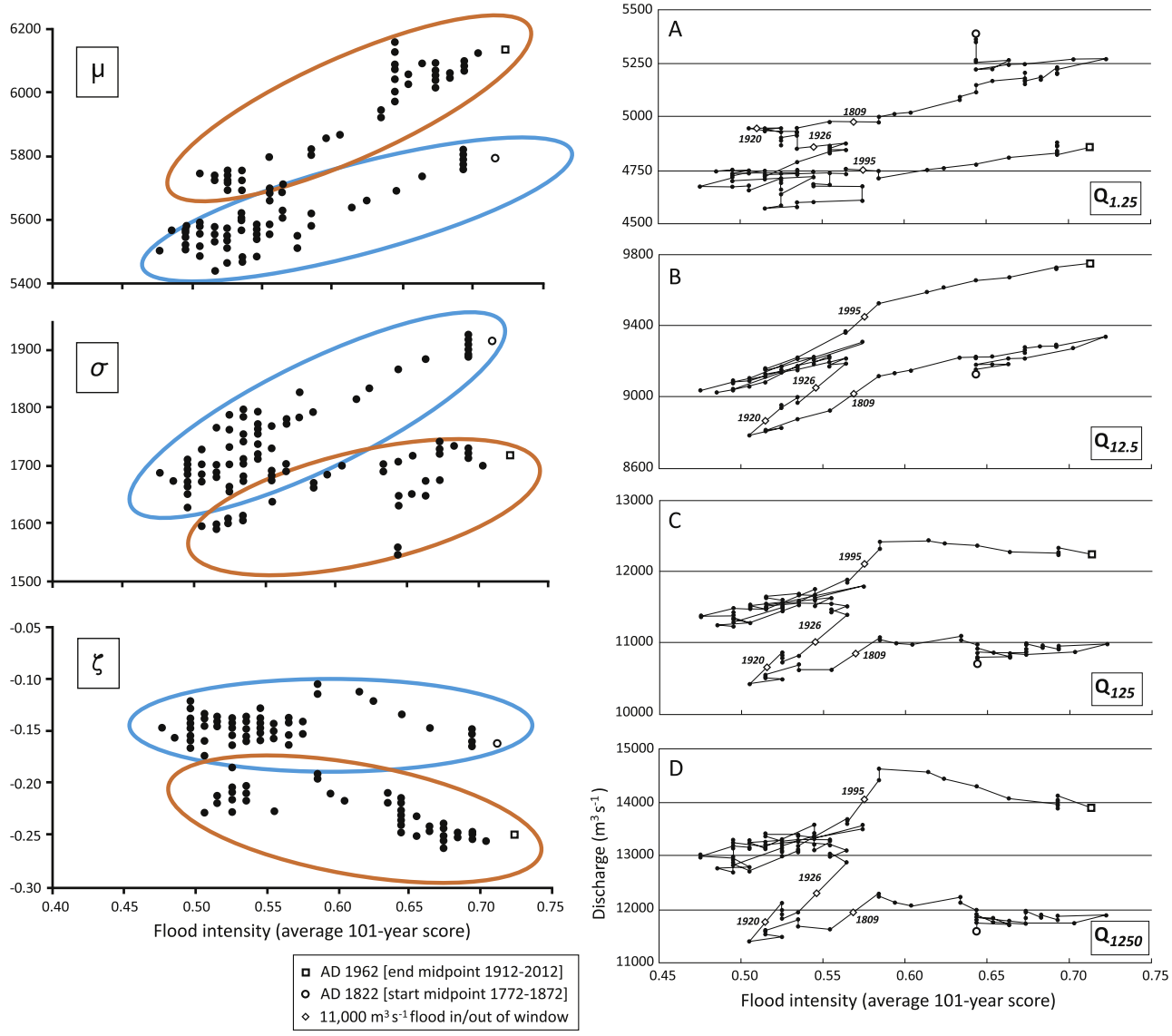


**Fig. 7.** Correlation of GEV-parameters and GEV-estimated discharges (for various recurrence classes; 1.25, 12.5, 125, and 1250 years) with reconstructed flood intensities, calculated over a moving 101-year window from AD 1772–2011. Mid-point values are plotted for the moving window; diamonds indicate important changes in discharge calculations caused by inclusion or exclusion of large floods. The AD 1926 event has been used to separate modes of high/low (respectively grouped in red and blue) GEV-parameter values, used for the estimation of historical GEV-parameters.

plotted versus the reconstructed flood intensity index over the same period (Fig. 7). The results show that (i) GEV-calculated discharges generally increase linearly with flood intensities for low-magnitude floods (Fig. 7a and b), while predicted discharges for large floods show less response to increasing flood intensities (Fig. 7c and d), that (ii) the occurrence of largest floods (exceeding 11,000 $m^3 s^{-1}$ of discharge in 1809, 1920, 1926 and 1995) influences the variance of the dataset and thus the values of GEV-parameters and associated discharges for moderate to extreme floods significantly – the entire linear relation between flood intensity and discharge shifts along the vertical axis (Fig. 7; introduction or removal of the largest flood data points within the window are marked with diamonds), and that (iii) GEV-predicted discharges for a given recurrence interval have increased in recent decades, with exception of low-magnitude floods.

Historical discharges were simulated, based on historical flood intensities and the linear and step-wise relations between flood intensity and GEV-parameters (Fig. 7). To mimic relations between flood intensity and the shape of the GEV-distribution, the timing of major historical floods exceeding 11,000 $m^3 s^{-1}$ was used independently to pre-set the mode of flooding (red or blue cluster in Fig. 7) for the linear relation between flood intensity and GEV-parameter values. Extreme floods occurred in 1374, 1497, 1595, 1651, and 1658; estimates on the magnitudes of these floods were taken from independent historical flood reconstructions (Herget and Euler, 2010; Toonen et al., 2015). Within a 101-year window around these events, historical GEV-parameters were established based on the regression of flood intensities and GEV-parameters from 1876–1962, which represent a similar flood-intense period with the large 1926 flood, and producing high values of $\mu$ and low values of $\sigma$ and $\xi$ (cluster 'high' in Fig. 7). GEV-parameters for years falling outside this 101-year window around largest events were established by using a regression between flood intensity and GEV-parameters of the less flood-intense period before the 1926 flood (cluster 'low' in Fig. 7). Annual flood probabilities were repeatedly sampled ($n = 10$) to simulate historical peak discharge data to AD 1400 (the earliest mid-point value of the 101-year window) from the calculated GEV-distribution for each year.

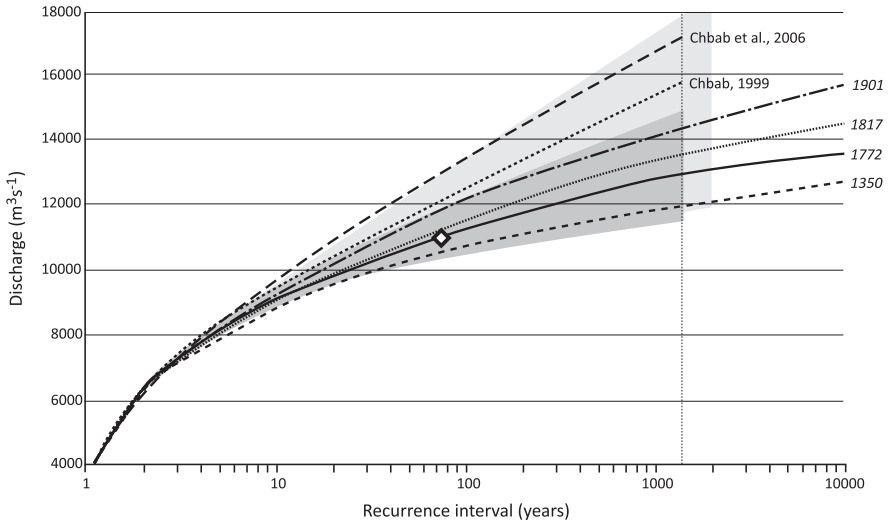


**Fig. 8.** GEV-based flood frequency curves for various data series (Table 3). For the [AD 1901–2011] and [AD 1772–2011] extrapolations the 95% uncertainty range is shown respectively in light and dark grey. The diamond indicates the Peak-Over-Threshold recurrence time of floods larger than 11,000 $m^3 s^{-1}$ in the AD 1350–2011 data series.

**Table 3**
Overview of the GEV-results with lengthened data series ($n$ = sample volume, MC = number of Monte Carlo simulations) and relative difference compared to the standard extrapolation (AD 1901–2011). For the design flood and the upper bound, the lower and upper range of the 95% confidence interval are given.

| Data | $n$ | MC | $Q_{10}$ | $Q_{100}$ | $Q_{1000}$ | Lower = $Q_{1250}$–low | $Q_{1250}$ | Upper = $Q_{1250}$–high | Lower = $Q_{bound}$–low | $Q_{bound}$ | Upper = $Q_{bound}$–high |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1901–2011 | 111 | – | 9250 | 12060 | 14130 | 11990 | **14300** | 17710 | – | 20440 | – |
| 1901–2011 (normalised) | 111 | – | 9320 | 12100 | 14150 | – | **14320** | – | – | 20602 | – |
| 1901–2011 | 111 | 1000 | 9270(+0%) | 12110(+0%) | 14210(+1%) | 11990 | **14380**(+1%) | 17920 | 18190 | 20720(+1%) | 24960 |
| 1817–2011 | 195 | 1000 | 9050(−2%) | 11600(−4%) | 13400(−5%) | 11870 | **13550**(−5%) | 15700 | 16260 | 18060(−12%) | 20720 |
| 1772–2011 | 240 | 1000 | 9060(−2%) | 11330(−6%) | 12850(−9%) | 11500 | **12970**(−9%) | 14960 | 14730 | 16010(−22%) | 17870 |
| 1350–2011 | 612 | 10 | 8880(−4%) | 10760(−11%) | 11850(−16%) | 11180 | **11920**(−17%) | 12800 | – | 13480(−34%) | – |
| Chbab et al. (2006) | 102 | – | – | – | – | – | **15260**(+7%) | – | – | – | – |
| Chbab et al. (2006) | 1000 | – | – | – | – | – | **17504**(+22%) | – | – | – | – |

## 4. GEV-extrapolation results

### 4.1. Modern discharge data [AD 1901–2011]

GEV-predictions using the various data series (including Monte Carlo simulations), produce design flood estimates ($Q_{1250}$) that range between 12,970 and 14,380 $m^3 s^{-1}$ (Fig. 8: 1901, 1817, and 1772 curves; Table 3). These estimates are strikingly lower than results of previous studies (Chbab et al., 2006; Fig. 8; Table 3). The screened dataset results in a 7–22% lower value for $Q_{1250}$ compared to various results of Chbab et al. (2006). A difference already occurs with a reanalysis of the modern dataset (from AD 1901 onwards). Differences originate from (i) a larger sample size by natural growth of the dataset (ii) screening of our record (thereby lowering several annual maxima), and (iii) data-resampling methods used in Chbab et al. (2006). The effect of homogenisation of discharge data is very limited for the Lower Rhine (Table 3), mainly due to the minor changes in absolute discharges it produced and because recent flood retention engineering projects have largely countered earlier projects that increased absolute discharges and effective flood pulse propagation. The combined effect of different pre-treatment and data series length since AD 1901 (from $n$ = 102 to $n$ = 110) is a ~7% decrease in $Q_{1250}$ when compared to GEV-results in Chbab et al. (2006). A larger difference is seen with the data-resampling approach in Chbab et al. (2006); which draws upon 1000 hydrologically simulated discharge peaks generated from random combinations of 35 years of 20$^{th}$ century precipitation data from multiple stations in the Rhine catchment. This procedure raises their GEV-estimate for $Q_{1250}$ by ~13% in comparison to their standard series (Table 3).

Inclusion of measurement uncertainties through Monte Carlo simulations produces comparable results as the initial extrapolation (the $n$ = 111 extrapolation; Table 3). The uncertainty range of the design flood expands only with ~1%. This is due to the relatively small measurement error of recent data. This should, however, not be a reason to neglect measurements uncertainties completely in flood risk assessments, especially when using less precise data. With the GEV-extrapolations, also the range for the upper bound of the system is determined. These bounds, ranging from 18,190 to 24,960 $m^3 s^{-1}$ (Table 3), are comparable to reconstructed discharge estimates for the extraordinary flood of AD 1374 in Cologne (Herget and Meurs, 2010). Meteorological limits to discharge generation in the Rhine catchment and the limited height of the Lower Rhine valley shoulders downstream of Cologne (Lammersen, 2004; Hegnauer et al., 2014) suggest, however, that such discharges are probably unrealistic for the Lobith station.

### 4.2. Extended discharge records [AD 1772–2011]

Extension of the data series back to AD 1817 and 1772 produces lower estimates for $Q_{1250}$ (Table 3). The added period has relatively high flood intensities (Fig. 6), but lacks very large floods (with exception of 1809). This causes the GEV-estimation to decrease

only slightly for moderate floods, while the estimated size of the design flood decreases with ~5–9% compared to GEV-estimates using 1901–2011 exclusively (Table 3; Fig. 8) to 12,970–13,550 $m^3 s^{-1}$. If these estimates are representative for the current system, ignoring effects of non-stationarity, the 1926 flood (~12,600 $m^3 s^{-1}$) can be considered a very rare event.

Upper bounds of the GEV-extrapolation are affected severely by extending the record; a 22% decrease based on data since 1772 places the upper bound on a similar level as the current design flood (~16,000 $m^3 s^{-1}$). The level of this bound corresponds poorly with the reconstructed discharges of the 1374 AD flood at Cologne (Herget and Meurs, 2010), but it corresponds well with estimated hydrological bounds for the Lower Rhine in the Netherlands; model studies indicate that the maximum amount of precipitation-derived discharge in the catchment amounts ~18,700 $m^3 s^{-1}$, while modelled floodplain inundation and upstream dike breaches presumably limit the maximum Rhine discharge at Lobith to ~15,500 $m^3 s^{-1}$ (Lammersen, 2004; the two-dimensional Delft-FLS model was used to simulate floodwave propagation; discharges were based on a one-dimensional runoff model with input from resampled 30-year meteorological data).

The uncertainty envelope of extreme flood discharges decreases in width despite the limited precision of the added measurements and reconstructions (Fig. 8). Especially the upper limit of the uncertainty interval decreases considerably (~10%). The resulting uncertainty interval is comparable to previous studies (e.g., Chbab, 1999), which indicates that long data series enable the use of fairly inaccurate data; the addition of 129 data points counteracts the low precision of alternative measurements.

An alternative way to compare the extrapolation results with historical information, and to have an independent check on the performance of extrapolation curves, is to estimate recurrence intervals of large floods by counting peaks over a certain discharge threshold (POT). In the last ~660 years, reconstructions of historical floods indicate that at least 9 floods presumably exceeded 11,000 $m^3 s^{-1}$ (Fig. 6; Herget and Euler, 2010; Toonen et al., 2015). The associated recurrence interval of that discharge is approximately 73 years. Comparison with the results of the flood frequency distributions (Fig. 8) indicates that the initial data series (AD 1901–2011) is too short to accurately predict recurrence intervals of large events. The lengthened series (back to AD 1817 and AD 1772) correspond much better with these independent estimates from peak over threshold, and are suggested to be more realistic in the large flood recurrence-magnitude domain.

### 4.3. Historical by-proxy flood series

Addition of synthetically generated flood magnitudes for the period before 1772, results in a further lowered estimate for the design flood (−17% compared to the AD 1901–2011 estimate; Table 3). The design flood is calculated to even lower discharges than actually measured discharges in the 20th century. Although the upper bound is also much lower than previous estimates, it still exceeds any observed discharge of the past century. The addition of historical data also lowers discharge estimates for decadal floods, indicating that the relationship between flood magnitudes and recurrence interval changes for all flood sizes. Obviously, the results of this explorative approach suffer from some important drawbacks, which make it difficult to establish the quality of this data for flood frequency analysis (Section 5.3).

## 5. Discussion

Adding alternative data types to existing discharge series involves two important trade-offs; (i) introducing less precise measurements versus a gain in accuracy of extrapolated results for extreme events: precision of data versus length of series, and (ii) constraining the uncertainty of design flood estimates by using a larger data set versus the representativeness of extended records for the current situation: length of series versus a dynamic flooding regime influenced by anthropogenic and climatic non-stationarity.

### 5.1. Alternative measurements and extrapolation results

In this study, inclusion of discharge measurement uncertainty increased the uncertainty envelope of extrapolated design flood estimates only slightly, because modern measurements are fairly precise for the Rhine. In other catchments with less dense, less precise or shorter records of discharge measurements, the reduction of uncertainty could be larger. Inclusion of alternative data types was straightforward for the Rhine, and a key benefit was that water level records could be verified among multiple sites. Changes to the fluvial system, such as channel bed changes that lead to variable stage-discharge relations, have been monitored and could be quantified for the recent centuries. This restrains the uncertainty of alternative data and conversion methods. Hence, addition of this information improved extrapolation estimates of shorter discharge records (Table 3). The addition of information based on historical records, describing the occurrence and extent of past floods, can be favourable for use in flood frequency analysis (Macdonald, 2012), though the use of flood intensities remains challenging.

### 5.2. Non-stationarity

The second trade-off is more complicated to assess. In our study, the addition of information further back in time resulted in a lower estimate for the 1250-year design flood. This lowering is mainly caused by a lack of extreme floods in the added period, which increases their estimated recurrence time. In case of a stationary flooding regime adding historical data better constrains the estimates of recurrence intervals of extreme events. Previous research, however, has proven that the assumption of stationarity in flooding regimes is not valid (e.g., Knox, 1993).

Variations in average annual discharge (Fig. 2) and indexed by-proxy flood intensity (Fig. 6) are relatively small. Although clear variations can be observed in the 31-year indexed flood intensity, almost no trends remain when a 101-year window is used, which implies that most variability occurs on a (multi-)decadal interval. It is common practice to normalise discharge records of the last century for river management, but according to Parmet et al. (2001), Bronstert et al. (2007) and Vorogushyn and Merz (2013), the effect of recent changes on discharges of extreme floods in the Lower Rhine is small. This is confirmed by the results of this study (Table 3). Gradually decreasing flood intensities can be observed from AD 1400 to 1900 (101-year window; Fig. 6), but part of this decline can be attributed to the historical nature of the data (discussed in next section), with a varying impact of historical floods over time that is not necessarily connected to discharges exclusively. Moreover, this trend is not clearly reflected in the distribution of largest events (Fig. 6), as their occurrence is not corresponding with general variability in flood frequency over time.

Two important contrasting periods (amidst other less pronounced episodes of increased and reduced flooding; Fig. 5), both widely associated with anomalous climate conditions, are the LIA and recent decades. It is generally assumed that the LIA was a period with extensive flood damage caused by many extreme floods in the Netherlands (e.g., Tol and Langen, 2000). The LIA climate anomaly (c. 1550–1850; Glaser and Riemann, 2009) is in our record, however, associated with rather 'normal' flood intensities and a relatively low frequency of large floods (class 3; Fig. 6). Several

**Table 4**
Seasonality of Lower Rhine flooding. Relative occurrence of class 1–3 floods during DJF (Dec., Jan., and Feb.), MAM (Mar., Apr., and May), JJA (June, July, and Aug.), and SON (Sep., Oct., and Nov.).

| | DJF (%) | MAM (%) | JJA (%) | SON (%) |
|---|---|---|---|---|
| AD 1350–1399 | 75.0 | 12.5 | 12.5 | 0.0 |
| AD 1400–1499 | 64.7 | 14.7 | 8.8 | 11.8 |
| AD 1500–1599 | 70.6 | 17.6 | 11.8 | 0.0 |
| AD 1600–1699 | 90.9 | 9.1 | 0.0 | 0.0 |
| AD 1700–1799 | 79.3 | 17.2 | 3.4 | 0.0 |
| AD 1800–1899 | 66.7 | 30.3 | 0.0 | 3.0 |
| AD 1900–2011 | 66.7 | 22.2 | 0.0 | 11.1 |

ice-related catastrophic flooding events were recorded in the Lower Rhine (for example in AD 1784; Demarée, 2006; Brázdil et al., 2010), but most floods of this period have not been associated with extreme discharges (Toonen et al., 2015). The occurrence of extreme floods and the amplitude of flood intensity differ not much from that for the last ~150 years, although floods occurred more often during winter months compared to other periods (Table 4). Yet, an important difference with the last century is the occurrence of many minor events (class 1: Fig. 6). This probably significantly influences results of flood frequency analysis, as the 300-year duration of the LIA biases the extrapolation result (Fig. 8) towards a situation with many minor floods. As this period encompasses nearly half of the dataset, this can explain the relative low design flood and upper bound predictions for the AD 1350 curve (Fig. 8).

It remains unclear how the period of intense flooding (Figs. 5 and 6) of the recent decades fits in the discussion of non-stationarity. The rise of the flooding index during the last century (Fig. 6) could be a response of the flooding regime to an exceptionally warm and wet interval driven by anthropogenic climate change. Based on the reconstructed flood intensities it could also be part of a natural multi-decadal variability in flood intensity, similar to increased flood occurrence in the early 19$^{th}$ century (Fig. 6). Although several studies suggest intensified flooding in recent decades (e.g., Milly et al., 2002) and point to climate change as the main driver, other studies indicate that there are no major changes in recent European flooding (Mudelsee et al., 2003; Brázdil et al., 2006; Glaser et al., 2010). Part of the proposed rising trend may be attributed to a limited observational period: when only discharges of the 20$^{th}$ century are considered in flood frequency analysis (Fig. 6), it is easily assumed that recent changes are significant, and exceeding normal natural variability. Adding alternative measurements and historical records, however, places recent changes in a context of ongoing longer-term variations in flood occurrences with possible changes in flood generating mechanisms (Glaser et al., 2010) and flood seasonality (Krahe and Larina, 2010; Macdonald, 2012; Table 4), and shows that periods with high flood intensities have occurred previously. The extended data series presented in this study thus largely diminishes the influence of short term climate variability and reduces the issue of (multi-decadal) non-stationarity as a larger range of cooler/warmer episodes is considered.

No straightforward relationship has yet been discovered for the river Rhine between the occurrence of extreme floods and periods of varying flood intensity (Glaser, 2001; and Fig. 6 in this paper). As especially extreme events have a profound effect on the results of flood frequency analysis and pose the greatest threat to communities in lower reaches of large rivers, it is recommended to further investigate the specific timing of extremes in relation to climate anomalies, and the relationship between general flood intensities and extreme events. Furthermore, detailed discharge reconstructions of historical events (e.g., Wetter et al., 2011), and flood series derived from additional sources such as sedimentological records (Baker, 2008; Benito et al., 2004; Czymzik et al., 2013; Toonen et al., 2013) can further increase data availability for extreme events and non-stationarity of the flooding regime.

### 5.3. The use of historical records

Flood intensity indexing based on historical information and subsequent translation to discharge series by applying the relations in Fig. 7, allows stretching flood frequency analysis over many centuries. There are, however, some drawbacks and limitations to this approach.

First, historical flood magnitudes were classified in four arbitrary classes, so discrete data was used to generate continuous data. More classes would produce higher quality data, but unfortunately historical records do not allow a more detailed classification, especially for minor floods of which documentation is generally less extensive than for catastrophic events. Potentially with additional flood magnitude information from sedimentological records (e.g., Toonen et al., 2015) centennial floods could be assigned to a separate class. Introducing an extra class for large floods would probably strengthen this approach, as it supports simulating large discharges in predefined timeslots. Large events are currently underrepresented in the historically-generated discharge series, but are especially important in flood frequency analysis, and have led to a severely (and probably unrealistic) reduced design flood estimate.

Second, there are several natural and societal factors that influenced flood intensity over time. In the current assessment, the relation between flood damage and associated flood magnitude is assumed to be stable over time. Natural variability in the relationship between floods and dike breaches was caused by the dynamic discharge distribution of flood waters over different branches of the Rhine in historical periods and the occurrence, frequency and severity of ice jamming. Until the artificial fixation of the delta apex main river channel bifurcation in 1707 (De Pannerdense Kop; Van de Ven, 1976), the right-hand bifurcate channel (the Nederrijn; Fig. 1) became gradually abandoned since the end of the Medieval period (Kleinhans et al., 2011). Consequently, up to 90% of the discharge gathered in the Waal (left-hand bifurcate channel), which resulted in increased flooding along this branch (Van de Ven, 1976; Glaser and Stangl, 2003). The effect of such changing discharge distributions on reconstructed flood intensities are difficult to assess in historical datasets. Intensified ice jamming in colder periods has probably also had a profound effect, as there are many reports about severe winters followed by catastrophic flooding. For much of the smaller, local floods it is very well possible that ice jams have raised water levels or created additional loading on dikes, and such contributed to relatively raised flood intensities compared to warmer periods without ice formation. For regional flooding, and particularly for catastrophic floods, ice jams have contributed to the extent of dike breaches, but regional floods are very unlikely to have occurred without large peak discharges. Flood intensities as reconstructed for the LIA are, however, similar or even lower compared to warmer intervals. Moreover, a limited number of extreme events occurred during the LIA (Fig. 6), so although this cold episode is traditionally associated with a many severe floods, this paper shows a general reducing effect of this period on the outcomes of flood frequency analysis.

Especially in the historical timeframe also political and economical factors presumably were an important factor for the magnitude of flood damage (Brázdil et al., 1999), and hence, our flood intensity estimates. Dike strength and chance of failure were largely determined by maintenance. Periods of political and economic instability led to reduced maintenance and fragmentation of governmental areas, leaving embanked areas under different rules with different standards (Van Heiningen, 1978). Occasionally,

dikes were targeted by military actions to deliberately cause extensive flooding (e.g., Buisman, 2000). Although these events are extensively described in historical records, it is difficult to filter 'the human factor' in the occurrence and severity of dike breaches. Until Napoleonic rule, local people were largely responsible for dike maintenance, resulting in regionally variable dike strengths (Van Heiningen, 1978). This explains why even minor floods are well-documented, as there was always a weak spot where a breach could occur. Dike restoration after a breach sometimes took more than a year to completely restore a dike, which may have resulted in repeated flooding in the following years, as dikes were still not fully repaired (Van Heiningen, 1978). Severe flooding also resulted in raised awareness and extra investments in dike strength (similar to the response after recent floods), until the necessity of maintaining dikes waned again. This could explain, in combination with a more centralised government during and after Napoleonic rule, that after the flood of 1809 the frequency of class 1 flood gradually reduced (Fig. 6) by increased flood protection measures.

## 6. Conclusions

Discharge series supplemented with discharge or water level measurements from alternative nearby stations are very useful for improving estimates of the recurrence interval of large floods, despite the lower precision of these alternative measurements, and the uncertainty introduced by conversion methods. Extending measured discharge series back to AD 1772 results in a reduction of the extrapolated design flood ($Q_{1250}$) by ~10% (compared to the 1901–2011 series; Fig. 8) and reduces the upper limit of the uncertainty envelope.

The results suggest that the current design flood discharge for the Lower Rhine in the Netherlands is based on a series that is too short to assess the natural variability of the flooding regime. The current design flood standard is thus biased towards the current episode of increased flooding and is therefore not very suitable for analysis of the recurrence of extremes in the Lower Rhine flooding regime. The results presented in this paper indicate nonetheless that current protection levels along the Rhine (~16,000 $m^3 s^{-1}$) are similar to the upper limits of the system and are thus in agreement with current flood design standards – actually protecting against much larger floods than currently required by law. Furthermore, it should be stressed that the extended data series provides no estimates for the largest discharges that can possibly occur in the future under a 'greenhouse climate' for which no analogue exists in data from the recent past.

The direct use of historical records in flood frequency analysis is complicated, when flood magnitudes are derived from flood intensities, which are based on categorised flood magnitudes. Flood intensities are not exclusively reflecting discharges, but are also influenced by anthropogenic and environmental factors. This analysis demonstrates that the distribution and magnitude of extreme floods is most important in flood frequency analysis, as illustrated by the responses of the GEV-extrapolation to extremes (Fig. 7). The occurrence of extreme floods correlates poorly with general flood intensities, making dependence on flood intensities for generating discharge data at this stage problematic. Specifically targeted discharge reconstructions of large floods in the historical period at multiple locations, combining advanced hydrological modelling and sedimentary flood magnitude reconstructions with historical records may provide important additional information on the magnitude and recurrence intervals of extreme events, and can be used to verify extrapolation results based on discharge data from measurements and historical records only (Toonen et al., submitted for publication).

The reconstructed flood intensities show imprints of non-stationarity, especially as multi-decadal variations in the flooding index curves (Fig. 6). Non-stationarity is often regarded as problematic in flood frequency analysis, as it may bias results when past flooding regimes not similar to the present situation are included. Indeed the results indicate that considering only recent data is deceptive: to cover multi-decadal or centennial variability a longer record should be consulted. Extended records covering anomalous periods of flooding, for example the LIA, should however be used with caution equally, as results indicate that a slightly different flooding regime during several centuries can change outcomes of flood frequency analysis significantly.

## Acknowledgements

The Dutch Ministry of Infrastructure and the Environment is acknowledged for sharing data and providing estimates on measurement uncertainties. Sofia Caires (Deltares, HYE) is thanked for guidance in statistical procedures. Kim Cohen and Hans Middelkoop (Utrecht University), Sofia Caires (Deltares, HYE), and Quirijn Lodder (Rijkswaterstaat) are thanked for comments on previous versions of this manuscript. This research was funded by Deltares BGS, Utrecht University and supported with a NWO Rubicon research grant. Two anonymous reviewers are cordially thanked for constructive comments.

## References

Baker, V.R., 2008. Paleoflood hydrology: origin, progress, prospects. Geomorphology 101, 1–13.

Barriendos, M., Martin-Vide, J., 1998. Secular climatic oscillations as indicated by catastrophic floods in the Spanish Mediterranean coastal area (14th–19th centuries). Clim. Change 38, 473–491.

Benito, G., Lang, M., Barriendos, M., Llasat, M.C., Frances, F., Ouarda, T., Thorndycraft, V.R., Enzel, Y., Bardossy, A., Coeur, D., Bobee, B., 2004. Use of systematic, palaeoflood and historical data for the improvement of flood risk estimation. review of scientific methods. Nat. Hazards 31, 623–643.

Brázdil, R., Glaser, R., Pfister, C., Dobrovolny, P., Antoine, J.-M., Barriendos, M., Camuffo, D., Deutsch, M., Enzi, S., Guidoboni, E., Kotyza, O., Sanchez Rodrigo, F., 1999. Flood events of selected European rivers in the sixteenth century. Clim. Change 43, 239–285.

Brázdil, R., Kundzewicz, Z.W., Benito, G., 2006. Historical hydrology for studying flood risk in Europe. Hydrol. Sci. J. 51, 739–764.

Brázdil, R., Demarée, G.R., Deutsch, M., Garnier, E., Kiss, A., Luterbacher, J., Macdonald, N., Rohr, C., Dobrovolný, P., Kolář, Chromá, K., 2010. European floods during the winter 1783/1784: scenarios of an extreme event during the 'Little Ice Age'. Theor. Appl. Climatol. 100, 163–189.

Brázdil, R., Reznickova, L., Valasek, H., Havlicek, M., Dobrovolny, P., Soukalova, E., Rehanek, T., Skokanova, H., 2011. Fluctuations of floods of the River Morava (Czech Republic) in the 1691–2009 period: interactions of natural and anthropogenic factors. Hydrol. Sci. J. 56, 468–485.

Bronstert, A., Bárdossy, A., Bismuth, C., Buiteveld, H., Disse, M., Engel, H., Fritsch, U., Hundecha, Y., Lammersen, R., Niehoff, D., Ritter, N., 2007. Multi-scale modelling of land-use change and river training effects on floods in the Rhine basin. River Res. Appl. 23, 1102–1125.

Buisman, J., 1996, 1998, 2000, 2006. Duizend jaar weer, wind en water in de Lage Landen, Deel 2–5. Van Wijnen, Franeker.

Chbab, E.H., 1996. How extreme were the 1995 flood waves on the Rivers Rhine and Meuse? Phys. Chem. Earth 20, 455–458.

Chbab, E.H., 1999. Onderzoek 1/1250 jaar afvoer bij Borgharen en Lobith: Bayesiaanse analyse. RIZA werkdocument 2001.183x, Lelystad.

Chbab, E.H., Buiteveld, H., Diermanse, F., 2006. Estimating exceedance frequencies of extreme river discharges using statistical methods and physically based approach. Osterreichische Wasser- und Abfallwirtsch. 58, 35–43.

Czymzik, M., Dulski, P., Plessen, B., von Grafenstein, U., Naumann, R., Brauer, A., 2010. A 450 year record of spring-summer flood layers in annually laminated sediments from Lake Ammersee (southern Germany). Water Resour. Res. 46, W11528. http://dx.doi.org/10.1029/2009WR008360.

Czymzik, M., Brauer, A., Dulski, P., Plessen, B., Naumann, R., von Grafenstein, U., Scheffler, R., 2013. Orbital and solar forcing of shifts in Mid- to Late Holocene flood intensity from varved sediments of pre-alpine Lake Ammersee (southern Germany). Quat. Sci. Rev. 61, 96–110.

De Wit, M., Buishand, A., 2007. Generator of rainfall and discharge extremes (GRADE) for the Rhine and Meuse basins. RWS RIZA report 2007.027, Lelystad. KNMI-publication 218, de Bilt.

Delta Progamma, 2014. Fifth Deltaprogam (2015). Ministry of Finances and Ministry of Infrastructure and Environment, The Netherlands. <www.rijksoverheid.nl/deltaprogramma.
Demarée, G.R., 2006. The catastrophic floods of February 1784 in and around Belgium – a Little Ice Age event of frost, snow, river ice. . .and floods. Hydrol. Sci. J. 51, 878–898.
Driessen, A.M.A.J., 1994. Watersnood tussen Maas en Waal. Walburg Pers, Zutphen.
Engel, H., 1997. The flood events of 1993/1994 and 1995 in the Rhine river basin. In: Destructive water: Water-caused Natural Disasters, their Abatement and Control (Proceedings of the Conference held at Anaheim, California, June 1996). IAHS Publication 239.
Fisher, R.A., Tippett, L.H.C., 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proc. Camb. Philos. Soc. 24, 180–190.
Glaser, R., 2001. Klimageschichte Mitteleuropas. 1000 Jahre Wetter, Klima, Katastrophen. Darmstadt.
Glaser, R., Riemann, D., 2009. A thousand-year record of temperature variations for Germany and Central Europe based on documentary data. J. Quat. Sci. 24, 437–449.
Glaser, R., Stangl, H., 2003. Historical floods in the Dutch Rhine Delta. Nat. Hazards Earth Syst. Sci. 3, 605–613.
Glaser, R., Riemann, D., Schönbein, J., Barriendos, M., Brázdil, R., Bertolin, C., Camuffo, D., Deutsch, M., Dobrovolny, P., van Engelen, A., Enzi, S., Halíčková, M., Koenig, S.J., Kotyza, O., Limanówka, D., Macková, J., Sghedoni, M., Martin, B., Himmelsbach, I., 2010. The variability of European floods since AD 1500. Clim. Change 101, 235–256.
Guse, B., Hofherr, T., Merz, B., 2010. Introducing empirical and probabilistic regional envelope curves into a mixed bounded distribution function. Hydrol. Earth Syst. Sci. 14, 2465–2478.
Hegnauer, M., Beersma, J.J., Van den Boogaard, H.F.P., Buishand, T.A., Passchier, R.H., 2014. Generator of Rainfall and Discharge Extremes (GRADE) for the Rhine and Meuse basins. Final report of GRADE2.0 1209424-004, Deltares, Delft.
Herget, J., Euler, T., 2010. Hoch- und Niedrigwasser in historischer und prähistorischer Zeit. Geograph. Rundsch. 63, 4–10.
Herget, J., Meurs, H., 2010. Reconstructing peak discharges for historic flood levels in the city of Cologne, Germany. Glob. Planet. Change 70, 108–116.
Hesselink, A.W., 2002. History makes a river; morphological changes and human interferences in the river Rhine, the Netherlands. Neth. Geogr. Stud., 292
Hosking, J.R.M., Wallis, J.R., Wood, E.F., 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics 27, 251–261.
Hundecha, Y., Bárdossy, A., 2004. Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. J. Hydrol. 292, 281–295.
IJnsen, F., 1981. Onderzoek naar het optreden van winterweer in Nederland. KNMI, Wetenschappelijk Rapport W.R. 74-2.
Kalis, A.J., Merkt, J., Wunderlich, J., 2003. Environmental changes during the Holocene climatic optimum in central Europe – human impact and natural causes. Quat. Sci. Rev. 22, 33–79.
Kidson, R., Richards, K.S., 2005. Flood frequency analysis: assumptions and alternatives. Prog. Phys. Geogr. 29, 392–410.
Kleinhans, M.G., Cohen, K.M., Hoekstra, J., IJmker, J.M., 2011. Evolution of a bifurcation in a meandering river with adjustable channel widths, Rhine delta apex, the Netherlands. Earth Surf. Proc. Landf. 36, 2011–2027.
Klemeš, V., 2000. Tall tales about tails of hydrological distributions. I. J. Hydrol. Eng. 5, 227–231. http://dx.doi.org/10.1061/(ASCE)1084-0699(2000) 5:3(227).
Knox, J.C., 1993. Large increases in flood magnitude in response to modest changes in climate. Nature 361, 430–432.
Krahe, P., Larina, M., 2010. Hoch- und Niedrigwasser in Köln seit AD 1000. Geographische Rundschau 3, 34–41.
Lammersen, R., 2004. Grensoverschrijdende effecten van extreem hoogwater op de Niederrhein. Eindrapport Duits-Nederlandse Werkgroep Hoogwater, 106pp.
Lammersen, R., Engel, H., van de Langemheen, W., Buiteveld, H., 2002. Impact of River training and retention measures on flood peaks along the Rhine. J. Hydrol. 267, 115–124.
Lang, A., Bork, H.-R., Mäckel, R., Preston, N., Wunderlich, J., Dikau, R., 2003. Changes in sediment flux and storage within a fluvial system: some examples from the Rhine catchment. Hydrol. Proc. 17, 3321–3334.
Macdonald, N., 2012. Trends in flood seasonality of the River Ouse (Northern England) from archive and instrumental sources since AD 1600. Clim. Change 110, 901–923.
Macdonald, N., Black, A.R., 2010. Reassessment of flood frequency using historical information for the River Ouse at York, UK (1200–2000). Hydrol. Sci. J. 55, 1152–1162.
Macklin, M.G., Lewin, J., 2008. Alluvial responses to the changing Earth system. Earth Surf. Proc. Landf. 33, 1374–1395.
Merz, B., Thieken, A.H., 2009. Flood risk curves and uncertainty bounds. Nat. Hazards 51, 437–458.
Middelkoop, H., 1997. Embanked floodplains in the Netherlands. Geomorphological evolution over various time scales. Neth. Geogr. Stud., 102
Milly, P.C.D., Wetherald, R.T., Dunne, K.A., Delworth, T.L., 2002. Increasing risk of great floods in a changing climate. Nature 415, 514–517.
Milly, P.C.D., Betancourt, J., Falkenmark, M., Robert, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., Stouffer, R.J., 2008. Stationarity is dead: whither water management? Science 319, 573–574. http://dx.doi.org/10.1126/science.1151915.
Mudelsee, M., Börngen, M., Tetzlaff, G., Grünewald, U., 2003. No upward trends in the occurence of extreme floods in central Europe. Nature 425, 166–169.
Parmet, B.W.A.H., van de Langemheen, W., Chbab, E.H., Kwadijk, J.C.J., Diermanse, F.L.M., Klopstra, D., 2001. Analyse van de maatgevende afvoer van de Rijn te Lobith. RIZA, Arnhem.
Pinter, N., van der Ploeg, R.R., Schweigert, P., Hoefer, G., 2006. Flood magnification on the river Rhine. Hydrol. Proc. 20, 147–164.
Redmond, K.T., Enzel, Y., House, P.K., Biondi, F., 2002. Climate impact on flood frequency at the decadal to millennial time scales. In: House, P.K., Webb, R.H., Baker, V.R., Levish, D. (Eds.), Ancient Floods, Modern Hazards: Principles and Applications of Palaeoflood Hydrology, Water Science and Application, vol. 5. American Geophysical Union, pp. 21–45.
Rijkswaterstaat, 1942. Ijsverslag winter 1939–1940. RIZA, Dienst Binnenwateren. Algemeene Landsdrukkerij, 's-Gravenhage.
Silva, W., Klijn, F., Dijkman, J., 2001. Room for the Rhine Branches in The Netherlands; what research has taught us. Technical Report 2001.031, WL Delft Hydraulics, Ministry of Transport, Public Works and Water Management, Delft, Lelystad, the Netherlands.
Stedinger, J.R., Surani, R., Therivel, R., 1988. Max Users Guide, a Program for Flood Frequency Analysis Using Systematic-record, Historical, Botanical, Physical Paleohydrologic and Regional Hydrologic Information Using Maximum-likelihood Techniques. Cornell University, Dept. Environmental Engineering, Ithaca, New York.
STIBOKA, 1975. Bodemkaart van Nederland schaal 1:50000. Toelichting bij de kaartbladen 40 West Arnhem en 40 Oost Arnhem. Stichting voor Bodemkartering, Wageningen.
Sturm, K., Glaser, R., Jacobeit, J., Deutsch, M., Brázdil, R., Pfister, C., 2001. Floods in Central Europe since AD 1500 and their relation to the atmospheric circulation. Petermanns Geograph. Mitt. 148, 18–27.
Swierczynski, T., Brauer, A., Lauterbach, S., Martin-Puertas, C., Dulski, P., von Grafenstein, U., Rohr, C., 2012. A 1600 yr seasonally resolved record of decadal-scale flood variability from the Austrian Pre-Alps. Geology 40, 1047–1050.
Taylor, W., 2000. Change-Point Analyzer 2.0, Taylor Enterprises, Libertyville, Illinois. <http://www.variation.com/cpa>.
Te Linde, A.H., Aerts, J.C.J.H., Bakker, A.M.R., Kwadijk, J.C.J., 2010. Simulating low-probability peak discharges for the Rhine basin using resampled climate modeling data. Water Resour. Res. 46, W03512.
Tol, R.S.J., Langen, A., 2000. A concise history of Dutch river floods. Clim. Change 46, 357–369.
Toonen, W.H.J., 2013. A Holocene flood record of the Lower Rhine. Utrecht Studies in Earth Sciences, vol. 41.
Toonen, W.H.J., de Molenaar, M.M., Bunnik, F.P.M., Middelkoop, H., 2013. Middle-Holocene palaeoflood extremes of the Lower Rhine. Hydrol. Res. 44, 248–263.
Toonen, W.H.J., Winkels, T.G., Cohen, K.M., Prins, M.A., Middelkoop, H., 2015. Lower Rhine historical flood magnitudes of the last 450 years reproduced from grain-size measurements of flood deposits using End Member Modelling. Catena 130, 69–81.
Toonen, W.H.J., Middelkoop, H., Konijnendijk, T.Y.M., Macklin, M.G., Cohen, K.M., submitted for publication. The influence of hydroclimatic variability on flood frequency in the Lower Rhine. Earth Surf, Process. Landf. (submitted for publication).
Van de Ven, G., 1976. Aan de wieg van Rijkswaterstaat – wordingsgeschiedenis van het Pannerdens Kanaal. De Walburg Pers, Zutphen.
Van Heiningen, H., 1978. Dijken en Dijkdoorbraken in het Nederlandse rivierengebied. Triangelreeks, 's-Gravenhage.
Van Vuuren, W., 2005. Met balen vol sigma's de Rijn en Maas afzakken. Werkdocument RIZA 2005.145X, 73pp.
Vorogushyn, S., Merz, B., 2013. Flood trends along the Rhine: the role of river training. Hydrol. Earth Syst. Sci. 17, 3871–3884.
Waterbase, 2012. Ministry of Infrastructure and the Environment. <live.waterbase.nl.> (accessed January 2013).
Waterwet, 2009. Ministry of Infrastructure and the Environment. <www.helpdeskwater.nl/onderwerpen/wetgeving-beleid/waterwet> (accessed January 2013).
Wetter, O., Pfister, C., Weingarter, R., Luterbacher, J., Reist, T., Trösch, J., 2011. The largest floods in the High Rhine basin since 1268 assessed from documentary and instrumental evidence. Hydrol. Sci. J. 56, 733–758.