Einführung Extremwertstatistik

Literaturvorschlag:

[1] Dyck S. & Peschke G. (1995): Grundlagen der Hydrologie; Verlag für Bauwesen Berlin
[2] Dyck S. (1980): Angewandte Hydrologie – Teil 1; Verlag für Bauwesen Berlin

Zweck Extremwertstatistischer Auswertungen

Die Kenntnis der Extremwerte von Niederschlag und Durchfluss ist vielfach für Aufgaben der *Bemessung* von Bedeutung. So sind viele wasserwirtschaftliche und verkehrstechnische Anlagen für Starkniederschläge bestimmter Intensität und Dauer ausgelegt oder *„bemessen“*. Ebenso sind viele flussbauliche Anlagen für bestimmte Durchflüsse (Q) oder Wasserstände (W) und ggf. bestimmte Andauern von W und Q bemessen.

Die Extremwertstatistik liefert Aussagen u.a. zu folgende Fragen:

– Wie häufig tritt ein Extremereignis gegebener Stärke statistisch gesehen auf?
– Mit welchem Extremwert muss in einem gegebenen Zeitintervall (z.B. alle 2, 5, 10 ... 500 Jahre) gerechnet werden?

Wichtige Grundlagen für die (Extremwert)-Statistik

Stichprobe und Grundgesamtheit

Für statistische Auswertungen werden zufällige und repräsentative *Stichproben* (hier Stichproben hydrologischer Extremwerte) zugrundegelegt; die Werte der Stichprobe werden als *Realisationen* eines *Zufallsprozesses* betrachtet. Mit den Methoden der Statistik wird versucht, aus den Informationen der Stichprobe auf die Gesetzmäßigkeiten der *Grundgesamtheit* zu schließen. Wesentliche Aufgabe hierbei ist die Schätzung der i.a. unbekannten *Verteilungsfunktion* der Grundgesamtheit (s.u.) aus der *empirischen Verteilung*, welche aus der Stichprobe ermittelbar ist.

Über- und Unterschreitungs-Wahrscheinlichkeiten

Die *Wahrscheinlichkeit* ist allgemein definiert als der Quotient aus der Anzahl der Fälle, in denen ein bestimmtes Ereignis eintritt und der Anzahl der möglichen Fälle. Sie schwankt demnach zwischen Null (das Ereignis tritt unmöglich ein) und Eins (das Ereignis tritt mit Gewissheit ein). In der Hydrologie interessieren v.a. Unter- und Überschreitungswahrscheinlichkeiten.

Die *Unterschreitungswahrscheinlichkeit* P_U gibt die Wahrscheinlichkeit an, dass (innerhalb eines Zeitraumes) die Zufallsgröße x nur Werte unterhalb eines Schwellenwert x_i annimmt oder dieser gerade erreicht wird.

$$P_U = P\,(x \leq x_i) \qquad (1\text{-}1)$$

Dagegen gibt die *Überschreitungswahrscheinlichkeit* $P_Ü$ an, mit welcher Wahrscheinlichkeit (innerhalb eines Zeitraumes) ein bestimmter Schwellenwert x_i überschritten wird.

$$P_Ü = P\,(x > x_i) \qquad (1\text{-}2)$$

Zwischen Über- und Unterschreitungswahrscheinlichkeit besteht die simple Beziehung:

$$P_U + P_Ü = 1 \qquad (1\text{-}3)$$

Häufigkeit, Dichtefunktion und Verteilungsfunktion

In der Hydrologie werden *stetige Zufallsgrößen* betrachtet, d.h. die Größen können innerhalb eines Intervalls beliebige Werte annehmen (im Gegensatz etwa zum Würfel, wo nur die *diskreten* Werte 1...6 möglich sind). Für stetige Zufallsgrößen gilt, dass die Wahrscheinlichkeit für das Eintreten eines beliebigen konkreten Wertes gegen *Null* geht.

In der Regel bildet man daher *Klassen* geeigneter Breite (Intervalle), die jeweils mehrere Werte enthalten. Für jede Klasse lässt sich die *absolute Häufigkeit* (Anzahl der Werte pro Klasse) und die *relative Häufigkeit* (Anzahl der Werte pro Klasse geteilt durch die Anzahl aller vorhandenen Werte) bilden.
Geht die Anzahl der beobachteten Werte (Realisationen des Zufallsprozesses) gegen unendlich, dann repräsentiert die Stichprobe immer mehr die Grundgesamtheit und die relative Häufigkeit ist ein geeigneter Schätzwert für die Wahrscheinlichkeit. Gehen wir von der Betrachtung der Stichprobe auf die Betrachtung der Grundgesamtheit über und verzichten auf die (willkürliche) Diskretisierung der Werte in Klassen – d.h. wir betrachten die Zufallsgröße tatsächlich als stetig – dann tritt an die Stelle der relativen Häufigkeit eine kontinuierliche Funktion: die *Dichtefunktion* oder *Wahrscheinlichkeitsdichte f(x)* (siehe folgende Abb.).

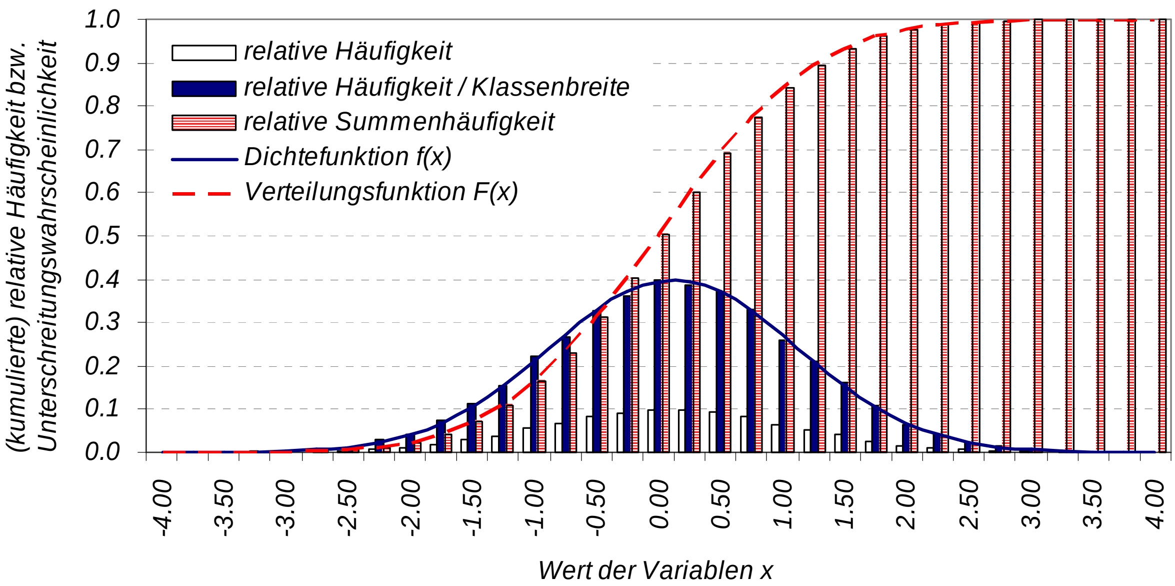

Abb. 1-1 Übergang der relativen Häufigkeit in die Dichtefunktion f(x) und der relativen Summenhäufigkeit in die Verteilungsfunktion F(x) am Beispiel einer normalverteilten Zufallsvariablen x mit dem Mittelwert $\mu=0$ und der Standardabweichung $\sigma=1$ (normierte Normalverteilung NV($\mu=0$; $\sigma=1$)). Zugrunde liegen 20000 Realisationen. Die Klassenbreite für die Darstellung der relativen Häufigkeiten beträgt 0.25.

Summiert man die relativen Häufigkeiten aller Klassen aufsteigend, erhält man die *kumulierte relative Häufigkeit* oder *relative Summenhäufigkeit*. Ihr entspricht beim Übergang von der Stichprobe zur Grundgesamtheit und stetiger Betrachtungsweise wiederum eine Funktion: die *Verteilungsfunktion F(x)*. Die Verteilungsfunktion ist das Integral der Dichtefunktion, d.h.:

$$\int f(x)\,dx = F(x) \qquad \textit{und entsprechend} \qquad \int_a^b f(x)\,dx = F(b) - F(a) \qquad (1\text{-}4)$$

Die Verteilungsfunktion ist deshalb von großer Bedeutung, da sie für jeden Wert der Zufallsvariable x direkt die zugehörige Unterschreitungswahrscheinlichkeit $P_U(x)$ angibt. Im oben gezeigten Beispiel wird der Wert Null mit der Wahrscheinlichkeit von 0.5, d.h. in 50 % aller Fälle, unterschritten. Die Verteilungsfunktion nähert sich für $x \rightarrow -\infty$ dem Wert Null und für $x \rightarrow +\infty$ strebt sie gegen Eins. Das Integral der Dichtefunktion im Intervall $-\infty$ ... $+\infty$ ist also 1; der Wert $-\infty$ wird mit der Wahrscheinlichkeit Null (nie) unterschritten und alle möglichen Werte der Zufallsvariablen sind mit einer Wahrscheinlichkeit von 1 (d.h. 100 %) kleiner als $+\infty$.

$$\int_{-\infty}^{+\infty} f(x)\, dx = F(+\infty) - F(-\infty) = 1 \qquad (1\text{-}5)$$

In der folgenden Abbildung sind die Zusammenhänge zwischen Dichte- und Verteilungsfunktion am Beispiel der Normalverteilung noch einmal dargestellt. Es wird beispielhaft gezeigt, wie sie zur Ermittlung von Wahrscheinlichkeitsaussagen verwendet werden.

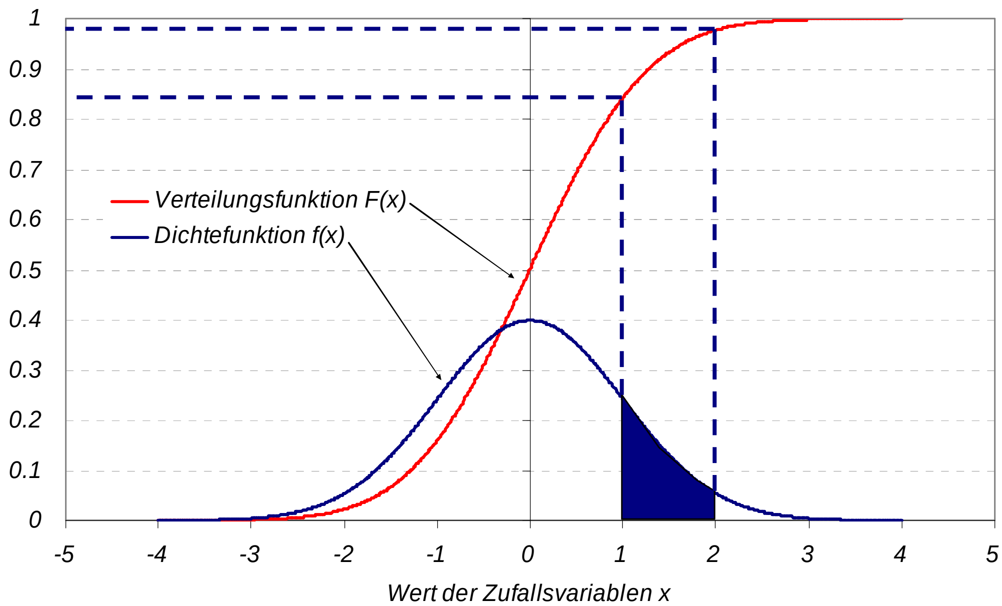

Abb. 1-2 Ableitung von Aussagen zur Auftretenswahrscheinlichkeit aus Dichte- und Verteilungsfunktion am Beispiel der Normalverteilung mit dem Mittelwert μ=0 und der Standardabweichung σ=1

Aus der Abbildung lassen sich folgende Wahrscheinlichkeitsaussagen ableiten:

– Der Wert x=1 wird mit einer Wahrscheinlichkeit von etwa 84 % unterschritten, da F(x=1)= 0.84.
– Für x=2 gilt analog $P_U(x=2) = P(x<=2) = F(x=2) = 0.98$.
– Die Wahrscheinlichkeit, dass die Zufallsvariable x einen Wert im Intervall 1...2 annimmt, berechnet sich aus:

$$P(1 \le x \le 2) = \int_1^2 f(x)\, dx = F(2) - F(1) \approx 0.14 \quad \textit{oder anders ausgedrückt} \qquad (1\text{-}6)$$

$$P(1 \le x \le 2) = P(x \le 2) - P(x \le 1) = P_U(2) - P_U(1)$$

– Die Wahrscheinlichkeit dass die Zufallsvariable x genau den Wert 1 oder 2 annimmt, ist Null.

In den vorangegangenen Beispielen wurde die Normalverteilung nur der Anschaulichkeit halber verwendet. Im Gegensatz zu vielen anderen Messgrößen sind hydrologische Extremwerte in der Regel nicht normalverteilt, weshalb zur Beschreibung andere Verteilungsfunktionen benutzt werden. Auf solche wird weiter unten eingegangen.

Wiederkehrintervall hydrologischer Extremwerte

Das *Wiederkehrintervall T* (auch *Wiederkehrzeit*) ist definiert als durchschnittliche Zeitspanne, in dem ein Ereignis auftritt, dessen Intensität x (z.B. Durchfluss oder Niederschlagsintensität) einen bestimmten Schwellenwert x_i erreicht oder übertrifft. Statistisch gesehen tritt ein Ereignis mit $x >= x_i$ also im zeitlichen Abstand T auf. Häufig betrachtet man jährliche Extremwerte und T wird in Jahren angegeben. Das Wiederkehrintervall wird dann auch als *Jährlichkeit* bezeichnet.

Da das Wiederkehrintervall nur ein durchschnittlicher Wert ist, ist es gut möglich, dass innerhalb der Zeitspanne T kein oder aber mehrere Werte mit $x >= x_i$ auftreten. Wiederkehrintervall und Über-/Unterschreitungswahrscheinlichkeit (vgl. deren Definitionen) stehen in folgender Beziehung:

$$P_Ü = P(x > x_i) = 1/T \tag{1-7}$$
$$P_U = P(x \le x_i) = 1 - P_Ü = 1 - 1/T \tag{1-8}$$

Ein Durchfluss-Ereignis mit einem Wiederkehrintervall von 100 Jahren (100-*jähriges Ereignis oder* HQ_{100}) wird demnach in einem konkreten Jahr mit der Wahrscheinlichkeit von 1/100 überschritten, d.h. $P_Ü = 0.01$. Die zugehörige Unterschreitungswahrscheinlichkeit beträgt $P_U = 0.99$, d.h. in durchschnittlich 99 von 100 Jahren tritt das Ereignis nicht auf.
Ist die Überschreitungswahrscheinlichkeit für einen Durchfluss mit 0.5 a^{-1} gegeben, dann ist im Mittel alle 2 Jahre mit diesem oder einem größeren Durchflusswert zu rechnen.

Ermittlung der Verteilungsfunktion für Extremwerte hydrologischer Variablen

Zielstellung und allgemeines Vorgehen

Um Aussagen über die Über-/Unterschreitungswahrscheinlichkeiten von Ereignissen bestimmter Intensität treffen zu können oder – umgekehrt – die Intensität eines T-jährigen Ereignisses zu bestimmen, stehen gemessene Extremwerte eines begrenzten Zeitraumes zur Verfügung (z.B. Terminwerte von Abfluss und Niederschlag für 50 Jahre). Sie bilden die Stichprobe auf Basis derer eine für die Grundgesamtheit gültige Verteilungsfunktion geschätzt werden soll. Die Ableitung der Verteilungsfunktion aus den Informationen der Stichprobe beinhaltet mehrere Schritte:

1. aus dem Kollektiv der vorhandenen Messwerte muss eine geeignete Extremwert-Stichprobe zusammen gestellt werden und die Werte müssen hinsichtlich verschiedener Voraussetzungen (*Unabhängigkeit, Repräsentativität, Konsistenz, Homogenität*) geprüft werden
2. für die in der Stichprobe enthaltenen Werte müssen *empirische Wahrscheinlichkeiten* bestimmt werden
3. es ist eine von ihrem Typ angemessene *theoretische Verteilungsfunktion* auszuwählen, die geeignet ist, die empirische Verteilung der Werte der Stichprobe zu beschreiben
4. die Funktionsparameter der theoretischen Verteilungsfunktion müssen aus den in der Stichprobe enthaltenen Informationen bestimmt werden (*Anpassung* der Verteilungsfunktion)
5. die Güte der Anpassung ist mit geeigneten statistischen Tests zu prüfen

Die Schritte 1–4 werden im Folgenden näher beschrieben.

Die Stichprobe und deren Prüfung (Unabhängigkeit, Repräsentanz, Konsistenz, Homogenität)

Von großer Bedeutung bei der Zusammenstellung der Datenreihe (Stichprobe) ist, dass die einzelnen Messwerte die Forderung der *Unabhängigkeit* voneinander erfüllen. Dies wird häufig durch die Verwendung *jährlicher Serien* als Stichprobe erreicht. Hierfür wird jeweils nur ein Extremwert pro hydrologischem Jahr in die Stichprobe aufgenommen (z.B. höchster beobachteter Durchflusswert im gesamten Jahr). Es wird also eine lange Messreihe benötigt, um eine Stichprobe geeigneter Größe zu erhalten. Wird als Stichprobe eine *partielle Serie* verwendet (alle Werte ober- bzw. unterhalb eines Schwellenwertes werden als Extremwerte angesehen) ist eine genaue Prüfung der Unabhängigkeit erforderlich.
Darüber hinaus muss die Stichprobe *repräsentativ* sein, sowohl zeitlich (repräsentativer Betrachtungszeitraum) als auch räumlich (z.B. ausreichende Anzahl von Niederschlags-Stationen in einem Gebiet).
Die Prüfung der verwendeten Datenreihe auf *Konsistenz* bedeutet die Identifikation und Entfernung bzw. Korrektur von fehlerhaften Messwerten. Bei Wasserstandsmessungen in Fließgewässern sind z.B. Ablesefehler, Eisstau im Bereich des Pegels, die Verlegung der Messstelle oder die Veränderung des Messquerschnittes mögliche Fehlerquellen.
Weiterhin muss die verwendete Datenreihe das Kriterium *Homogenität* erfüllen, d.h. die Messwerte dürfen nicht durch anthropogene oder natürliche Beeinflussungen während des Messzeitraumes gestört sein. Dieses Kriterium ist oft nicht erfüllt. So sind z.B. langfristige Durchfluss-Messreihen der Flüsse häufig inhomogen, da im Messzeitraum Ausbauten des Gewässers stattgefunden haben, Retentionsflächen entzogen oder Speicher in Betrieb genommen wurden oder durch Meliorationen das Abflussverhalten des Einzugsgebietes verändert wurde.

Inkonsistenz und Inhomogenität sind in der Praxis nicht immer sauber voneinander zu Unterscheiden. Im Folgenden sind einige wichtige Methoden der Datenprüfung erläutert.

(A) Prüfung durch Analyse der Datenreihe selbst

- Inhomogenitäten können im einfachsten Fall entdeckt werden, indem statistische Parameter der Datenreihe für separate Zeitabschnitte berechnet und miteinander verglichen; z.B. Mittelwerte (Trendanalyse) oder Varianzen.
- Mittels *Ausreißertests* kann geprüft werden, ob ein Wert einer Datenreihe tatsächlich zum Wertekollektiv gehört oder ob er durch Messfehler oder anthropogene Einflüsse verfälscht ist. Ein Messwert wird als Ausreißer identifiziert, wenn die Abweichung des Messwerts x vom Mittelwert m(x) ein festzulegendes Vielfaches der Standardabweichung s(x) überschreitet, d.h. wenn gilt:

$$abs(x) > m(x) + v(\alpha, n) * s(x) \qquad (1\text{-}9)$$

  Der Faktor v ist von der Anzahl der Messwerte n und dem Signifikanzniveau α (Sicherheit der Aussage) abhängig und kann aus Tabellen diverser Statistikbücher entnommen werden.

(B) Prüfung durch Vergleich der Datenreihe mit einer als homogen bekannten Reihe

- Eine Möglichkeit der Homogenitäts-Prüfung besteht im *Vergleich der Verteilungsfunktion* der zu prüfenden Reihe mit der Verteilungsfunktion einer anderen Datenreihe, die mit ersterer in Zusammenhang steht. So ist es üblich, die Verteilungsfunktionen der Hochwasserdurchflüsse benachbarter Pegel entlang eines Flusses miteinander zu vergleichen. Auch die Verteilungen verschiedener, jedoch kausal zusammenhängender Größen wie Niederschlag und Abfluss können miteinander verglichen werden.
- Einen weitere Methode der Homogenitätsprüfung ist die *Doppelsummenanalyse*, mit der Inhomogenitäten nicht nur erkannt sondern ggf. auch bereinigt werden können. Hierbei

werden die Werte der zu prüfenden Datenreihe sowie einer als unbeeinflusst bekannten und mit ersterer in Zusammenhang stehenden Vergleichsreihe schrittweise aufsummiert:

Summe1= erster Wert der Reihe
Summe2= erster + zweiter Wert der Reihe
Summe3= Summe2 + dritter Wert der Reihe
...

Die aufsummierten Werte beider Reihen werden im Diagramm gegeneinander aufgetragen und auf eine lineare Beziehung überprüft.

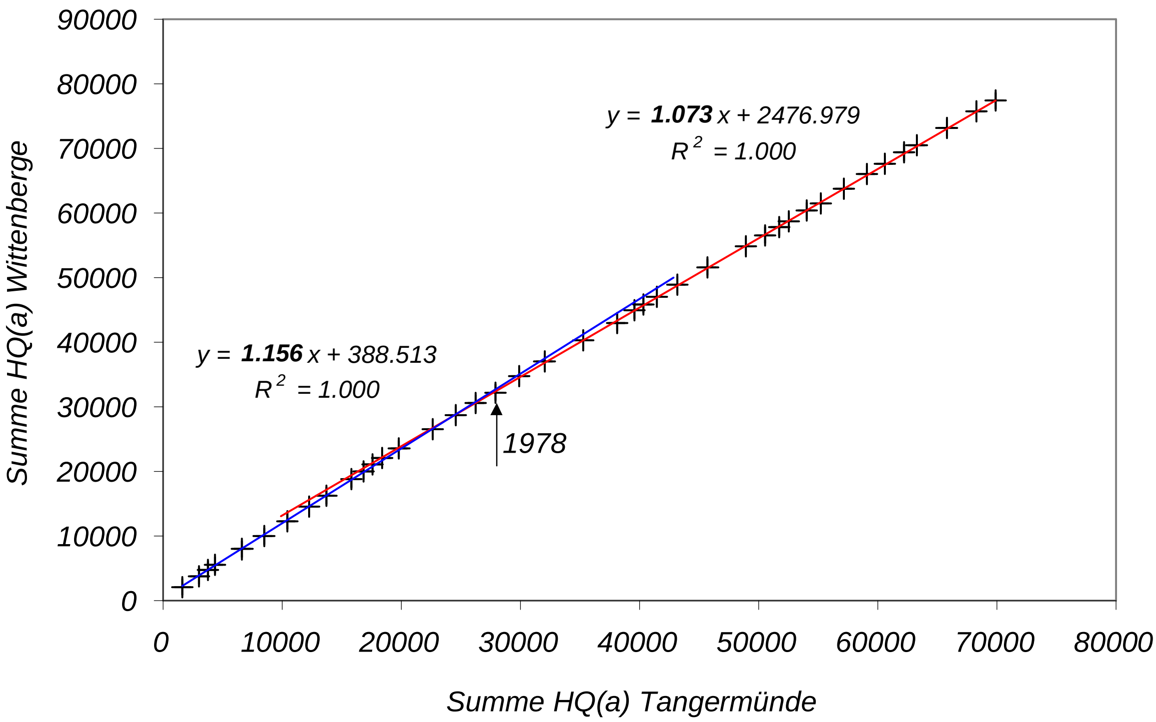

Abb. 1-3 Doppelsummenanalyse für die jährlichen Maxima des Durchflusses HQ(a) an den Elbepegeln Tangermünde und Wittenberge. Zugrunde liegen die hydrologischen Jahre 1960–2001. Etwa 1978 macht sich – verdeutlicht durch die angepassten Geraden – eine Veränderung der Beziehung bemerkbar, was auf die Inhomogenität (mindestens) einer der beiden Datenreihen hinweist. Innerhalb der Zeiträume 1960–1978 und 1978–2001 erscheinen die Reihen homogen. Unter der Annahme, die Werte des Pegels Tangermünde seien unbeeinflusst, könnte man schlussfolgern, dass die HQ(a) am stromab liegenden Pegel Wittenberge etwa ab 1978 systematisch geringer ausfallen, als vor diesem Zeitpunkt, obwohl der stromauf liegende Pegel Tangermünde einen solchen Trend nicht aufweist. Als Ursache kommt aber auch die Verwendung einer veränderten Wasserstands-Durchfluss-Beziehung an einem der beiden Pegel ab etwa 1978 in Frage.

Wurden durch Doppelsummenanalyse Inhomogenitäten festgestellt, kann entweder nur die unbeeinflusste Periode verwendet werden, oder es wird eine Korrektur der beeinflussten Daten vorgenommen. Hierfür werden die Werte der zu korrigierenden Reihe aus den Werten der als homogen bekannten Vergleichsreihe unter Zuhilfenahme des in der Doppelsummenanalyse ermittelten Geradenanstiegs während der unbeeinflussten Periode ermittelt.

Ermittlung von empirischen (Unterschreitungs)-Wahrscheinlichkeiten

Grundlage für die Anpassung einer Verteilungsfunktion sind die aus der Stichprobe berechneten *empirischen Wahrscheinlichkeiten* (korrekt: empirische Unterschreitungswahrscheinlichkeiten). Da der Stichprobenumfang in der Regel relativ klein ist, wäre die Bildung von Klassen mit anschließender Berechnung von relativen Summenhäufigkeiten nicht zielführend. Stattdessen wird für jeden Wert der Stichprobe eine empirische Unterschreitungswahrscheinlichkeit geschätzt. Oft wird hierfür der Ansatz von *Weibull* verwendet, da er für alle relevanten theoretischen Verteilungen einsetzbar ist. Die *Weibull-Formel* lautet:

$$P_U(x_i) = \frac{m_i}{n+1} \tag{1-10}$$

Darin ist:

$P_U(x_i)$ die empirische Unterschreitungswahrscheinlichkeit für den Wert x_i der Stichprobe
m_i der Rang des Wertes x_i in der nach aufsteigenden x-Werten sortierten Stichprobe
n die Anzahl der in der Stichprobe enthaltenen Werte

Demnach gilt für den kleinsten Wert aus einer jährlichen Serie von Hochwasserdurchflüssen $P_U=1/(n+1)$, d.h. es ist $P_U>0$ und für den größten Wert entsprechend $P_U=n/(n+1)$, d.h. es gilt $P_U<1$. Der Ansatz geht somit davon aus, dass der kleinste und größte mögliche Extremwert bisher noch nicht beobachtet wurden.
Die so gewonnenen empirischen Unterschreitungswahrscheinlichkeiten werden oft als *Plotting Positions* bezeichnet, da die ermittelten P_U(x)-Werte über den zugehörigen x-Werten in das *Wahrscheinlichkeitspapier* aufgetragen werden. Als Wahrscheinlichkeitspapier bezeichnet man Koordinatensysteme deren Achsen so skaliert sind, dass sich die Verteilungsfunktion darauf als Gerade darstellt. Jede Verteilungsfunktion erfordert ein spezifisches Wahrscheinlichkeitspapier.

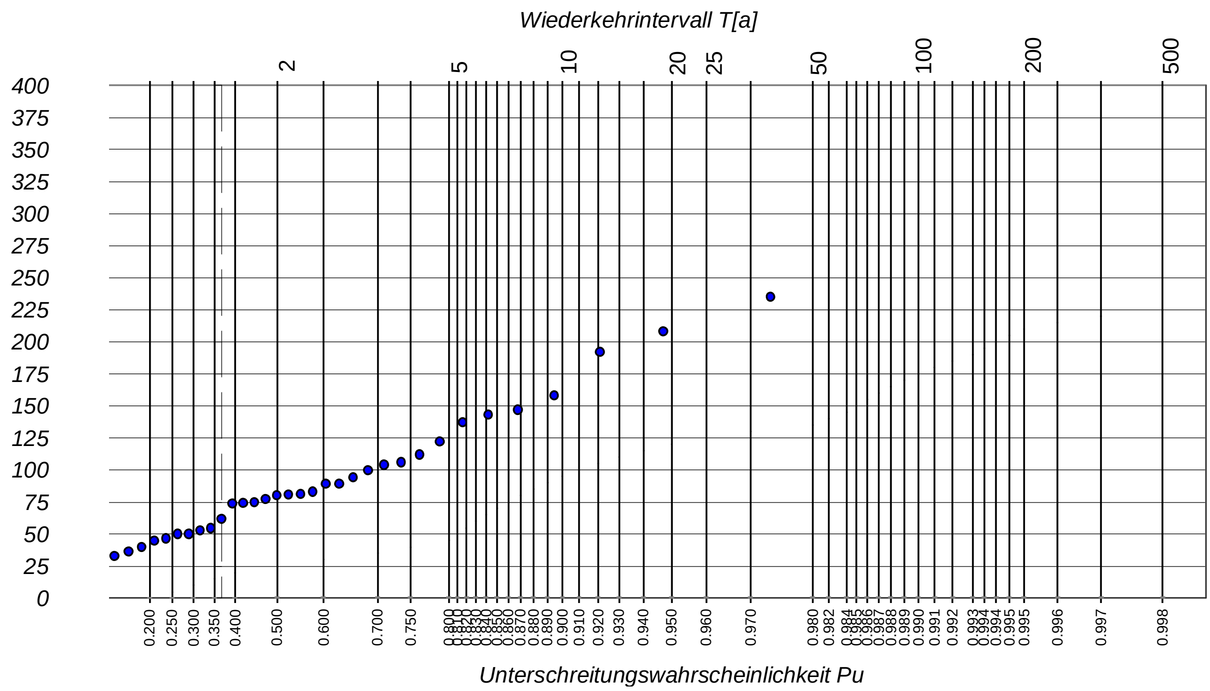

Abb. 1-4 Plotting-Positions im Wahrscheinlichkeitspapier der Extremwertverteilung Typ 1 (Gumbel-Verteilung). Bei einer Auswertung von Hochwasser-Durchflüssen werden die HQ_i -Werte auf der linearen y-Achse über ihren zuvor berechneten empirischen Unterschreitungswahrscheinlichkeiten $P_U(HQ_i)$ aufgetragen. Ist die Verteilungsfunktion zur Beschreibung der Stichprobenwerte geeignet, dann lassen sich die Plotting-Positions durch eine Gerade verbinden. [Beispieldaten aus Dyck & Peschke (1995)]

Auswahl der Verteilungsfunktion

Die bekannteste Verteilungsfunktion ist die Normalverteilung (Gauss-Verteilung). Sie ist eine *zweiparametrige Verteilung*, da sie durch die zwei Parameter Mittelwert (μ) und Standardabweichung

(σ) vollständig beschrieben wird. Dichtefunktion f(x) und Verteilungsfunktion F(x) der Normalverteilung lauten:

$$f(x) = \frac{1}{\sqrt{2\pi}} \ exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right) \tag{1-11}$$

$$F(x) = \int_{-\infty}^{x} f(t)\,dt \ = \ \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2 \right) dt \tag{1-12}$$

Zur Beschreibung der Verteilung hydrologischer Extremwerte wird eine größere Zahl unterschiedlicher Verteilungsfunktionen benutzt. Im Gegensatz zur Normalverteilung, deren Dichtefunktion um den Mittelwert symmetrisch ist, lassen sich mit diesen auch schiefe Werteverteilungen beschreiben. Sie besitzen meist 2 oder 3 Parameter. Besonders häufig sind die *Pearson-Verteilung Typ III (Gamma-III-Verteilung)* und die *Extremwertverteilung Typ I (Gumbel-Verteilung)* sowie die *Log-Normalverteilung* anzutreffen. Dichte- und Verteilungsfunktion der später näher behandelten Extremwert-Verteilung Typ 1 mit den zwei Parametern a und b lauten z.B.:

$$f(x) = a * exp(-a * (x-b) - exp(-a * (x-b))) \quad mit -\infty < x < \infty \tag{1-13}$$

$$F(x) = exp(-exp(-a * (x-b))) \quad mit\ a > 0 \tag{1-14}$$

Ob eine bestimmte Verteilungsfunktion zur Beschreibung der Charakteristik einer Stichprobe geeignet ist, kann a priori oft schwer beantwortet werden. Es werden daher meist verschiedene Verteilungsfunktionen – wie nachfolgend beschrieben – angepasst und die nach statistischen Gütekriterien (Anpassungstests) „beste“ ausgewählt.

Anpassung der Verteilungsfunktion an die empirische Verteilung

Bei der *Anpassung* geht es darum, die Parameter der Verteilungsfunktion aus den Informationen der Stichprobe zu bestimmen. Die angepasste theoretische Verteilungsfunktion soll die empirischen Wahrscheinlichkeiten für die Werte der Stichprobe (plotting positions) möglichst gut widerspiegeln. Drei Methoden der Anpassung werden häufig verwendet:

(A) Freihand-Anpassung
Hierbei werden die plotting positions in das Wahrscheinlichkeitspapier der entsprechenden Verteilung eingetragen und anschließen nach Augenmaß durch eine Gerade mit möglichst geringen Abweichungen verbunden. Die Methode ist wenig objektiv und kann daher nur für erste Abschätzungen oder mit sehr viel Erfahrung sinnvoll eingesetzt werden. Sie bietet den Vorteil, dass kaum Rechenaufwand nötig ist.

(B) Momentenmethode
Die Momentenmethode wird häufig eingesetzt, da sie bei geringem Rechenaufwand reproduzierbare Ergebnisse liefert. Hierbei verwendet man aus der Stichprobe berechenbare Kenngrößen wie Mittelwert, Varianz, Schiefe als Schätzwerte für die Parameter der theoretischen Verteilungsfunktion. Man nimmt also an, dass diese an der Stichprobe beobachteten Parameter auch für die Grundgesamtheit gültig sind. Von „Momentenmethode“ spricht man, weil der Mittelwert auch als *erstes Moment* und Varianz und Schiefe als *2. und 3. Zentralmoment* einer Stichprobe bezeichnet werden.
Während für die Anpassung einer Normalverteilung der aus der Stichprobe berechnete Mittelwert und die Standardabweichung direkt verwendet werden, sind bei anderen

Verteilungsfunktionen erst Umrechnungen vorzunehmen. So stehen z.B. die Parameter a und b der Extremwertverteilung Typ 1 (s.o.) mit Mittelwert (μ) und Standardabweichung (σ) der Stichprobe in folgender Beziehung:

$$a = \frac{\pi}{\sigma\sqrt{6}} \qquad (1\text{-}15)$$

$$b = \mu - \frac{\gamma}{a} \quad mit \;\; \gamma = 0.5772 \qquad (1\text{-}16)$$

(C) Maximum-Likelihood-Methode

Diese Methode wird in der Regel verwendet, wenn eine Auswertung am Rechner mittels hydrologischer Software vorgenommen wird. Die Parameter der Verteilungsfunktion werden so bestimmt, dass die Wahrscheinlichkeiten für die in der Stichprobe enthaltenen Beobachtungswerte möglichst groß sind oder – anders ausgedrückt – dass die angepasste Verteilungsfunktion die beobachteten empirischen Wahrscheinlichkeiten der Messwerte bestmöglich erklärt. Die Parameterbestimmung lässt sich als Optimierungsaufgabe mit der Zielfunktion L formulieren, die als *Likelihood-Funktion* bezeichnet wird:

$$L(u_1,\ldots,u_m) = f(\mathbf{x_1}, u_1,\ldots,u_m) * f(\mathbf{x_2}, u_1,\ldots,u_m) * \ldots * f(\mathbf{x_n}, u_1,\ldots,u_m) \qquad (1\text{-}17)$$

L:	Likelihood-Funktion
$u_1,\ldots,u_m$:	die „m“ zu bestimmenden Parameter der Dichte- bzw. Verteilungsfunktion
$f(\mathbf{x_i}, u_1,\ldots,u_m)$:	Funktionswert der Dichtefunktion für den i-ten Beobachtungswert x von insgesamt „n“ Beobachtungswerten

Die gesuchten Parameter $u_1,\ldots,u_m$ sind so zu wählen, dass $L(u_1,\ldots,u_m)$ maximal wird. Hierfür ist das Gleichungssystems der partiellen Ableitungen von $L(u_1,\ldots,u_m)$ nach allen „u“ zu lösen, was i.A. eine numerische Rechnung erfordert. In der Praxis wählt man als zu maximierende Größe statt L besser log(L), da sich die Likelihood-Funktion dann als Summe darstellen lässt.

Extrapolation auf Basis der angepassten Verteilungsfunktion

Die folgende Abbildung zeigt, dass verschiedene Methoden der Anpassung sehr unterschiedliche Ergebnisse liefern können. Besonders wird dies im durch nur wenige Ereignisse belegten Bereich großer Jährlichkeiten sichtbar. Eine visuelle und statistische Bewertung der Güte der Anpassung mittels statistischer Tests (und ggf. die Ablehnung der Verteilungsfunktion) sind unumgänglich.

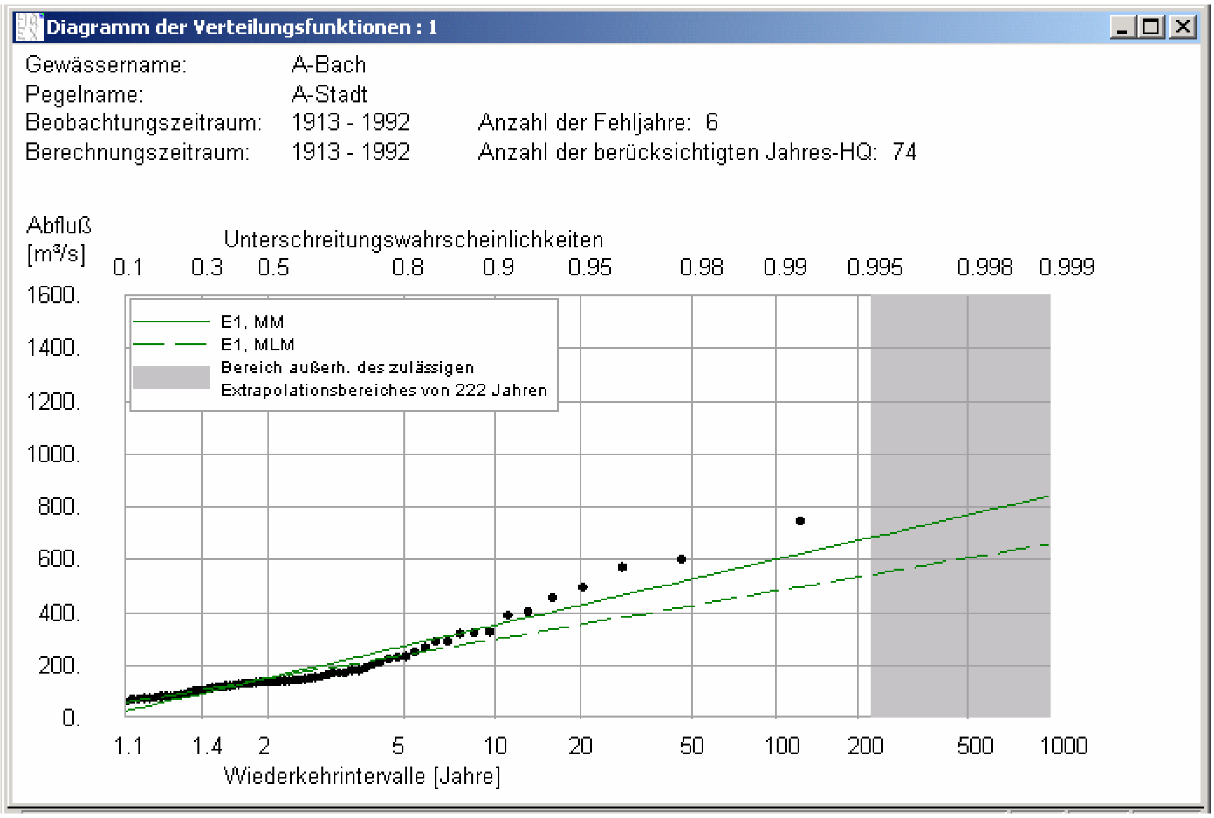

Abb. 1-5 Anpassung einer Extremwertverteilungsfunktion Typ 1 (E1) mittels Momenten-Methode (MM) und Maximum-Likelihood-Methode (MLM). [Beispieldatensatz des Programms HQ-Ex der Firma WASY]

Im gezeigten Beispiel besteht außerdem für HQ(a) großer Jährlichkeit eine offensichtliche und systematische Abweichung zwischen den empirischen Wahrscheinlichkeiten und der mittels Momenten- und Maximum-Likelihood-Methode angepassten Verteilungsfunktion. Es ist deshalb zu prüfen, ob mit einem anderen Typ von Verteilungsfunktion bessere Ergebnisse zu erzielen sind (vgl. nachfolgende Abbildung).

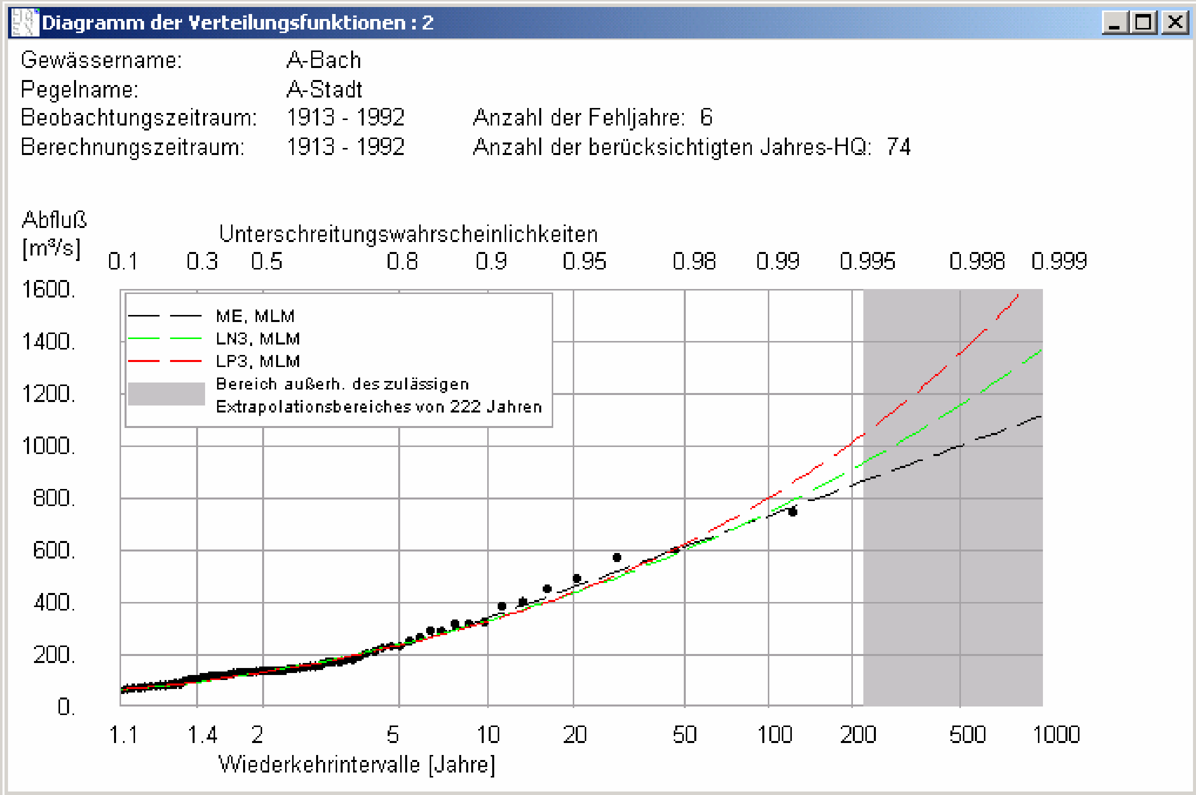

Abb. 1-6 Anpassung von drei unterschiedlichen Verteilungsfunktionen mittels Maximum-Likelihood-Methode. LP3: Logarithmische Pearson-III-Verteilung, LN3: Lognormalverteilung mit 3 Parametern, ME: Gemischte Extremwertverteilung (Rossi-Verteilung). Bereits visuell ist zu erkennen, dass die hier verwendeten Verteilungsfunktionen besser zur Beschreibung der empirischen Verteilung der HQ(a) geeignet sind, als die in der vorhergehenden Abbildung verwendete Extremwertverteilung Typ 1. [Beispieldaten des Programms HQ-Ex der Firma WASY]

Wird die angepasste Verteilungsfunktion benutzt um Aussagen zur Wahrscheinlichkeit sehr großer Extremwerte zu gewinnen oder wird – andersherum – nach der Intensität eines Ereignisses sehr großer Wiederkehrzeit (z.B. HQ_{500}) gefragt, dann muss oft weit über den Bereich der Beobachtungswerte hinaus extrapoliert werden. Eine Faustregel ist, dass die Jährlichkeit bis zu der extrapoliert wird nicht größer sein soll, als die dreifache Anzahl der HQ(a)-Werte der Stichprobe. Die Schätzung des 500-jährigen Durchflusses HQ_{500} setzt also – wenn nicht Zusatzinformationen einbezogen werden – eine über 150-jährige Messreihe voraus.

Verwendung der k-T-Beziehung

Den Zusammenhang zwischen dem Wiederkehrintervall T und dem zugehörigen Extremwert der hydrologischen Größe x(T) – der durch die angepasste Verteilungsfunktion hergestellt wird – lässt sich oft bequem mittels der sogenannten k-T-Beziehung ausdrücken. Dahinter steht die Überlegung, dass man die Funktionswerte der Verteilungsfunktion und deren Umkehrfunktion nicht für jede neuen Stichprobe neu berechnen möchte. Deshalb werden die Funktionswerte der Verteilungsfunktion $P_U(k)$ und deren Umkehrfunktion $k(P_U)$ einmalig für die sogenannte *normierte Variable* k berechnet und meist tabelliert (z.B. auch für die Normalverteilung). Diese Funktionstabellen lassen sich dann universell für jede Stichprobe verwenden, indem man die Beobachtungswerte x zuvor in die normierte Variable k transformiert.

$$k = \frac{x - \mu}{\sigma} \tag{1-18}$$

x: Beobachtungswert der Stichprobe
μ: Mittelwert der Stichprobe (1. Moment)
σ: Standardabweichung der Stichprobe (2. Zentralmoment)

Mittels der Tabelle oder Gleichung, die die Beziehung zwischen normierter Variabler k und Wiederkehrintervall k(T) bzw. Unterschreitungswahrscheinlichkeit $k(P_U)$ beschreibt, lässt sich dann leicht das Ereignis x(T) mit T-jähriger Wiederkehrzeit bzw. der Unterschreitungswahrscheinlichkeit P_U angeben (Rücktransformation).

$$x(T) = \mu(x) + k(T) * s(x) \tag{1-19}$$

x(T): Extremwert der Jährlichkeit T
$\mu(x)$: Mittelwert der Stichprobe (1. Moment)
s(x): Standardabweichung der Stichprobe (2. Zentralmoment)
k(T): Wert der normierten Variablen k für die Jährlichkeit T, wobei $k(T) = k(1-(1/P_U))$

Typische Fragen der Bemessung (z.B. „Was ist der 100-jährige Durchfluss?“) lassen sich mit der k-T-Beziehung bequem beantworten. Beachte: Formal bedeutet die Nutzung der k-T-Beziehung die Anwendung der Momentenmethode!

Besitzt die verwendetet Verteilungsfunktion neben Mittelwert und Standardabweichung weitere Parameter, dann ist die normierte Variable k nicht nur von der Jährlichkeit T sondern zusätzlich von diesen weiteren Parametern (z.B. dem Schiefekoeffizienten C_S; $k = k(T,C_S)$) abhängig.

Anwendungsbeispiel: Berechnung des HQ_{200} am Pegel Tangermünde (Elbe)

Aufgabe

Am Elbufer soll nahe Tangermünde eine Online-Wassergüte-Messstation mit teuren Geräten eingerichtet werden. Die Anlage soll für Hochwasser mit einer Jährlichkeit bis zu 200 Jahren überflutungssicher gebaut werden. Die Wasserstands-Durchfluss-Beziehung für den Querschnitt ist bekannt. Um zu berechnen, welcher Wasserstand mit dem 200-jährigen Hochwasser verbunden ist, muss HQ_{200} ermittelt werden.
Für die Analyse stehen Durchflüsse des Elbpegels Tangermünde für den Zeitraum 1961–2001 zur Verfügung. Die höchsten Durchflusswerte HQ(a) wurden für die hydrologischen Jahre bereits ermittelt und hinsichtlich Repräsentativität, Konsistenz und Homogenität geprüft wurde. Von den benachbarten Elbpegeln sei bekannt, dass die Extremwertverteilung Typ 1 zur Beschreibung der HQ(a) geeignet ist.

Ausgangsdaten

Tab. 1-2 HQ(a) am Pegel Tangermünde (Elbe) für die hydrologischen Jahre 1961–2001

Jahr	x=0	x=1	x=2	x=3	x=4	x=5	x=6	x=7	x=8	x=9
196x	–	1617	1374	775	572	2299	1841	1940	1850	1417
197x	2099	1051	757	800	1417	2820	1950	1670	1638	2020
198x	2111	3259	2836	1455	762	1121	1717	2560	3203	1620
199x	1160	832	1496	1175	1920	1950	1520	1610	1060	2530
200x	2490	1600	–	–	–	–	–	–	–	–

Schritt 1: Berechnen empirischer Unterschreitungswahrscheinlichkeiten

Die n HQ(a)-Werte werden aufsteigend nach Größe sortiert und die empirischen Unterschreitungswahrscheinlichkeiten (plotting positions) nach der Weibull-Formel berechnet.

Rang (m)	HQ-Wert	$P_U = m/(n+1)$
1	572	0.0238
2	757	0.0476
3	762	0.0714
4	775	0.0952
5	800	0.1190
6	832	0.1429
7	1051	0.1667
8	1060	0.1905
9	1121	0.2143
10	1160	0.2381
11	1175	0.2619
12	1374	0.2857
13	1417	0.3095
14	1417	0.3333
15	1455	0.3571
16	1496	0.3810
17	1520	0.4048
18	1600	0.4286
19	1610	0.4524
20	1617	0.4762
21	1620	0.5000

Rang (m)	HQ-Wert	$P_U = m/(n+1)$
22	1638	0.5238
23	1670	0.5476
24	1717	0.5714
25	1841	0.5952
26	1850	0.6190
27	1920	0.6429
28	1940	0.6667
29	1950	0.6905
30	1950	0.7143
31	2020	0.7381
32	2099	0.7619
33	2111	0.7857
34	2299	0.8095
35	2490	0.8333
36	2530	0.8571
37	2560	0.8810
38	2820	0.9048
39	2836	0.9286
40	3203	0.9524
41	3259	0.9762

Schritt 2: Eintragen der plotting positions in das Wahrscheinlichkeitspapier

Bevor die plotting positions eingetragen werden können, ist die lineare Achse so zu beschriften, dass die beobachteten HQ(a)-Werte sinnvoll dargestellt werden können.

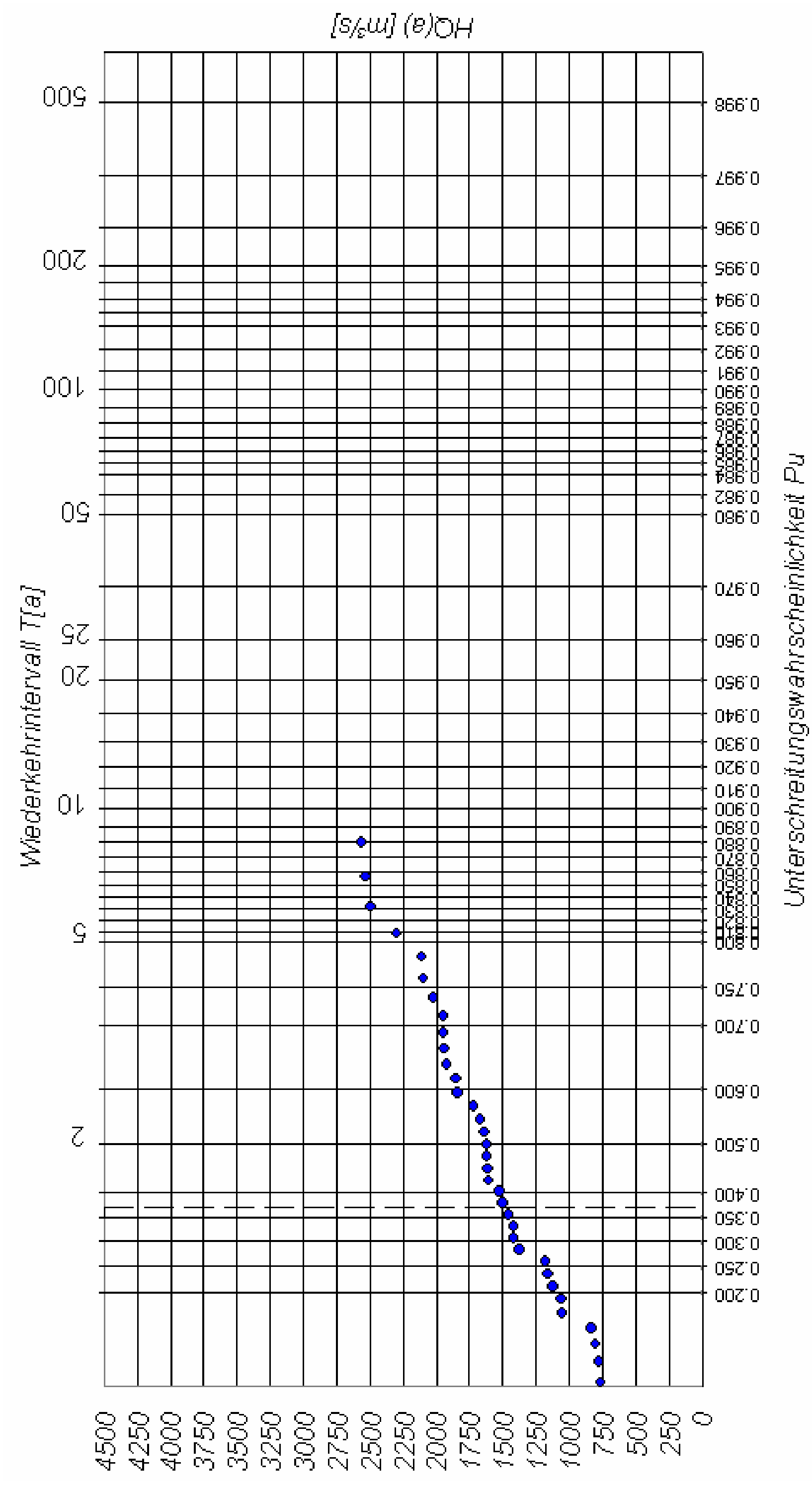

Abb. 1-7 Plotting positions im Wahrscheinlichkeitspapier der Extremwertverteilung Typ 1

Schritt 3: Anpassen der Verteilungsfunktion nach Augenmaß

Nach Augenmaß wird bestmöglich eine Ausgleichsgerade durch die plotting positions gelegt. Obwohl besonders der Verlauf der Gerade im Bereich großer Jährlichkeiten interessiert, muss die Anpassung den gesamten Bereich der Beobachtungswerte angemessen berücksichtigen, da die Verteilungsfunktion das gesamte Kollektiv der HQ(a) beschreiben soll.

Für die gesuchte Jährlichkeit T=200 (P_U=0.995) kann nun ein zugehöriger HQ-Wert von etwa 4550 m³/s abgelesen werden (siehe nachfolgende Abbildung).

Schritt 4: Anpassen der Verteilungsfunktion nach der Momentenmethode

Für die Anpassung der Extremwertverteilung Typ 1 werden als Momente der Stichprobe Mittelwert und Standardabweichung benötigt. Es ergeben sich:

$\mu = 1704.7$

$\sigma = 670.4$

Die Beziehungen zur Berechnung der Parameter a und b der Extremwert-Verteilungsfunktion Typ 1 aus den Momenten wurden bei der Erläuterung der Momentenmethode bereits aufgeführt. Es ergeben sich die Werte:

$a = \pi / \sigma / Wurzel(6) = 0.001913$

$b = \mu - 0.5772 \; / \; a = \quad 1402.995$

Da wir den Wert HQ(T) für eine gegebene Jährlichkeit T bzw. Unterschreitungswahrscheinlichkeit suchen, benötigen wir die Umkehrfunktion der Verteilungsfunktion. Sie lautet für die Extremwertverteilung Typ 1:

$$x = b - 1/a * \ln(-\ln(F(x))) \qquad (1\text{-}20)$$

Setzen wir für F(x) den gegebenen Wert P_U= 1–(1/200)= 0.995 ein, erhalten wir für HQ_{200} einen Wert von 4171 m³/s, der beträchtlich vom Ergebnis der freien Anpassung abweicht. Die mit der Momentenmethode angepasste Verteilungsfunktion lässt sich in das Wahrscheinlichkeitspapier einzeichnen, indem ein weiterer Wert (z.B. HQ(2)) berechnet und mit dem bereits ermittelten durch eine Gerade verbunden wird.

Statt der Umkehrfunktion kann man auch die k-T-Beziehung verwenden. Im Unterschied zu den meisten Verteilungen (z.B. Normalverteilung) existiert bei der Extremwertverteilung Typ 1 eine analytisch lösbare Gleichung für k(T), so dass keine Tabelle verwendet werden muss. Sie lautet:

$$k(T) = -\frac{\sqrt{6}}{\pi}\left(\gamma + \ln\left(\ln\left(\frac{1}{T-1}\right)\right)\right) \qquad (1\text{-}21)$$

Für T=200 ergibt sich für die normierte Variable k(T)= 3.6791. Die Rücktransformation unter Verwendung von Mittelwert und Varianz (siehe Erläuterung zur k-T-Beziehung) führt wie die Umkehrfunktion auf den Wert für HQ_{200} von 4171 m³/s.

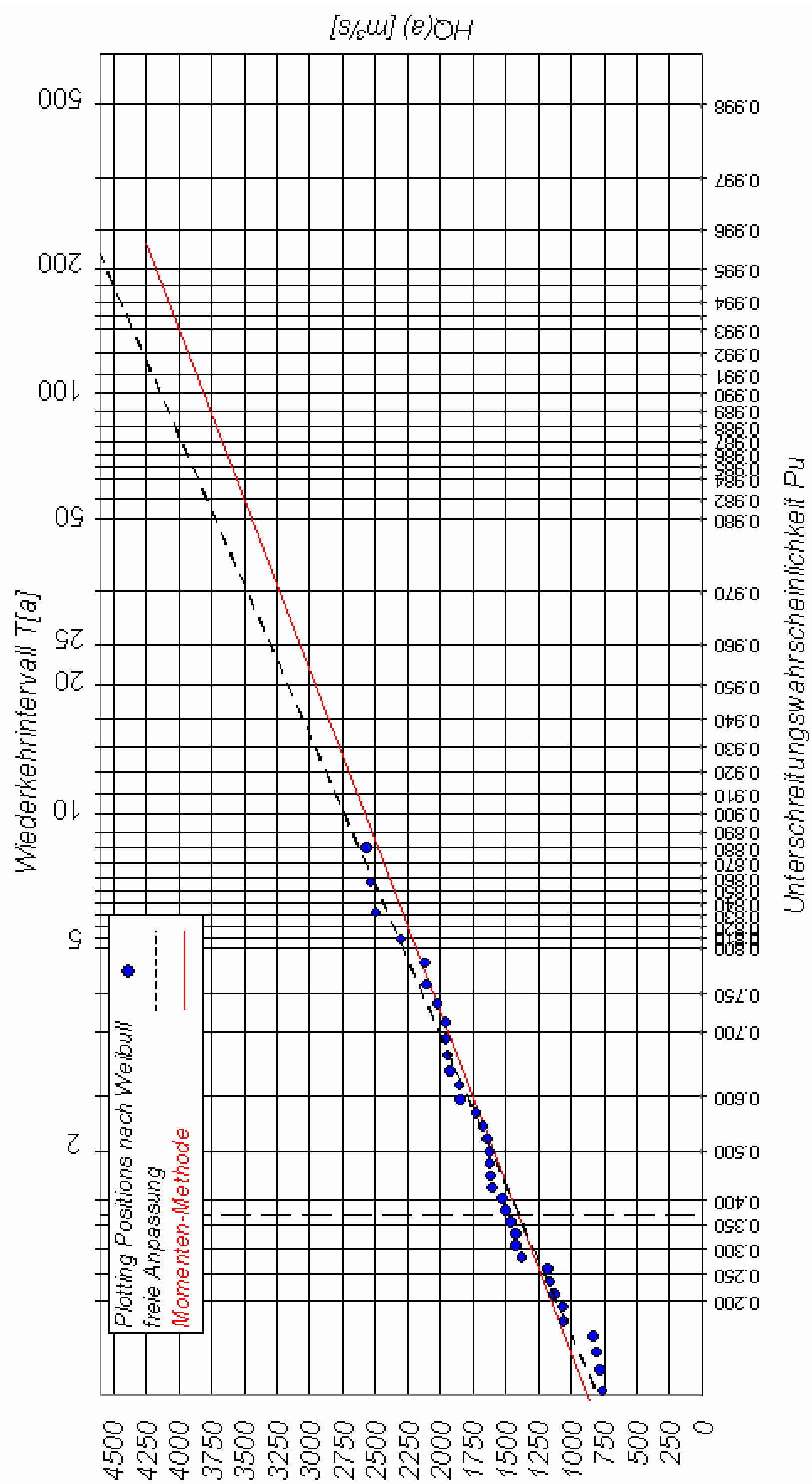

Abb. 1-8 Angepasste Extermwertverteilung Typ 1 an die HQ(a) des Pegels Tangermünde

Stark- und Bemessungsniederschläge

Ähnlich wie für Hoch- und Niedrigwasser-Durchflüsse werden extremwertstatistische Auswertungen auch für Niederschlagsereignisse durchgeführt. Für Bemessungsaufgaben sind Niederschläge mit in Bezug zu ihrer Dauer hohen Intensitäten von Interesse (Starkniederschläge). Seitens der Statistik sind Aussagen zur Häufigkeit von Ereignissen einer bestimmten Dauer und Niederschlagshöhe/Intensität zu treffen. Man spricht von:

– *Höhen-Dauer-Häufigkeits -Beziehungen* und
– *Intensitäts-Dauer-Häufigkeits -Beziehungen*

Bei der Aufstellung dieser Beziehungen kann – eine gute Datengrundlage vorausgesetzt – in zwei Schritten vorgegangen werden:

– Es wird für jede Niederschlagsdauerstufe (z.B. 5 min, 10 min, 15 min, ..., 72 h) eine separate extremwertstatistische Auswertung vorgenommen. Dabei werden die Wahrscheinlichkeiten für das Auftreten einer bestimmten Niederschlagshöhe/-intensität während des Zeitintervalls gegebener Dauer ermittelt (Anpassung einer Verteilungsfunktion).
– Im zweiten Schritt werden an die für die einzelnen Dauerstufen ermittelten Verteilungsfunktions-Parameter Ausgleichsfunktionen angepasst, so dass die Parameter für beliebige Dauerstufen berechnet werden können.

Ist die Verteilungsfunktion bekannt, kann zur Ermittlung des Bemessungsniederschlags einer bestimmten Dauer D und Jährlichkeit T die k-T-Beziehung (oben erläutert) verwendet werden:

$$P(T,D) = \mu(P(D)) + k(T) * s(P(D)) \qquad (1\text{-}22)$$

$P(T,D)$:	Niederschlagshöhe der Dauer D und Jährlichkeit T
$\mu(P(D))$:	Mittelwert der Niederschlagshöhen für die Dauer D
$s(P(D))$:	Standardabweichung der Niederschlagshöhen für die Dauer D
$k(T)$:	normierten Variable k der entsprechenden Wahrscheinlichkeitsverteilung

Sind bereits erstellte Niederschlagshöhen-Dauer-Häufigkeits-Beziehungen und die erforderlichen Grunddaten nicht vorhanden, kann der *Ansatz von Reinhold* verwendet werden. Dabei wird von einem Basisniederschlag mit der Dauer D= 15 min und der Jährlichkeit T= 1 a ausgegangen. Dabei wird statt der *Niederschlagshöhe* P(T,D) [mm] die *Niederschlagsspende* r(T,D) in der Einheit Liter Sekunde^{-1} ha^{-1} verwendet. Mit Hilfe der (empirischen!) *Reinhold-Formel* lassen sich die Niederschlagsspenden für andere Dauern und Wiederkehrintervalle berechnen:

$$r(T,D) = r(1, 15) * \varphi \qquad mit \quad \varphi = 38\ (T^{0.25} - 0.369)\ /\ (D+9) \qquad (1\text{-}23)$$

Angaben zu Basis-Starkregenspenden r(T=1 a, D=15 min) sind vielerorts verfügbar, da der Ansatz in der Vergangenheit oft für Bemessungsaufgaben eingesetzt wurde. Für Nordost-Deutschland gilt etwa r(1, 15)= 95 l s^{-1} ha^{-1}. Für große Jährlichkeiten sind die berechneten Werte allerdings mit großen Unsicherheiten behaftet. Die Ergebnisse, die mit der Reinhold-Formel berechnet werden, weichen z.T. erheblich von den aus örtlichen Messdaten einer Niederschlagsstation erstellten Höhen-Dauer-Häufigkeits-Beziehungen ab, da die räumliche Variabilität in der Formel unzureichend berücksichtigt wird.

Wo immer möglich, sollten also Stationsdaten ausgewertet werden oder auf die landesweit verfügbare Datengrundlage des *KOSTRA-Atlas (koordinierte Starkniederschlagsregionalisierung)* zurückgegriffen werden.

Übungsaufgabe 1

Mit Hilfe der Reinhold-Formel soll errechnet werden, mit welcher maximalen Niederschlagsmenge innerhalb von 1 Stunde statistisch gesehen alle 5 Jahre gerechnet werden muss. Als Basisregenspende r(1, 15) werden 105 l s^{-1} ha^{-1} angenommen. Das Ergebnis soll in mm Niederschlagshöhe angegeben werden. Lösung:

Mit D= 60 min und T= 5 a ergibt sich die *Niederschlagsspende* r(5, 60)= 65.13 l s^{-1} ha^{-1}.

Für die Umrechnung des Wertes in [mm] folgt aus der einfachen Umrechnungsformel

$$1\ l\ s^{-1}\ ha^{-1} = 0.006\ mm\ /\ min \quad (1\text{-}24)$$

für P(5, 60) eine *Niederschlagshöhe* von 23.4 mm.

Übungsaufgabe 2

Das gegebene Niederschlags-Höhen-Dauer-Häufigkeits-Diagramm soll mit den Ergebnissen der Reinhold-Formel verglichen werden. Der Vergleich soll anhand der Niederschlagshöhen [mm] mit einem Wiederkehrintervall von 5 Jahren und den Dauern 15 min, 60 min und 24 h durchgeführt werden. Die für die Station gültige Basisregenspende r(1, 15) ist aus dem Diagramm zu ermitteln.

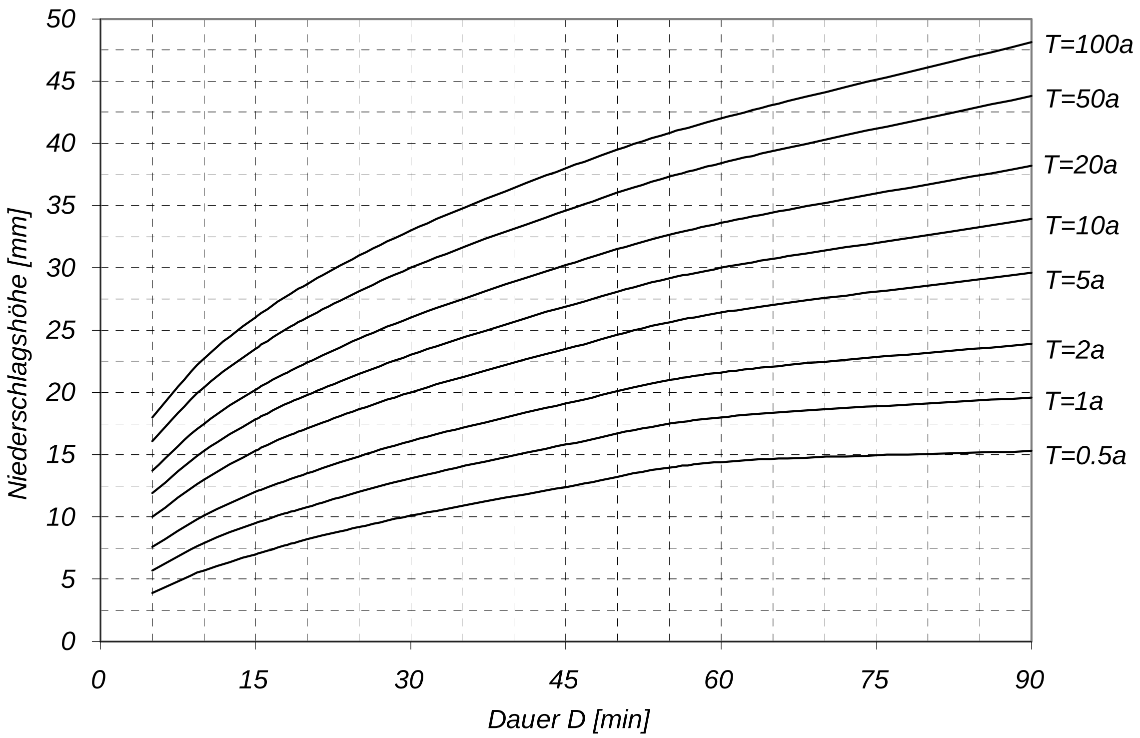

Abb. 1-9 Niederschlags-Höhen-Dauer-Häufigkeitsbeziehung für Dauerstufen von 5–90 Minuten und Jährlichkeiten von 0.5–100 Jahren.

Als Basisniederschlagshöhe P(1, 15) lässt sich ein Wert von 9 mm ablesen. Die für die Anwendung der Reinhold-Formel benötigte Basisregenspende r(1, 15) beträgt somit 100 $l\ s^{-1}\ ha^{-1}$. Die berechneten und aus dem Diagramm abgelesenen Werte für die zum Vergleich herangezogenen Dauern und Wahrscheinlichkeiten sind in der Tabelle zusammengestellt:

Tab. 1-3 Vergleich einiger Werte aus der Höhen-Dauer-Häufigkeits-Beziehung der Beispielstation mit den Berechnungsergebnissen nach der Reinhold-Formel

	aus Höhen-Dauer-Häufigkeits-Diagramm abgelesene Werte [mm]	mittels Reinhold-Formel berechnete Werte [mm]
P(5, 15 min)	15	16.1
P(5, 60 min)	27	22.3
P(5, 90 min)	29	23.3