

KfK 3581
Juli 1983

Simulation von Fehlerverteilungen —

**Untersuchung zur Wiedergewinnung von
Verteilungen aus unvollständigen Daten**

**R. Häberer, A. Brückner
Institut für Reaktorbauelemente
(Zuverlässigkeit und Schadenskunde im Maschinenbau)**

Kernforschungszentrum Karlsruhe



K E R N F O R S C H U N G S Z E N T R U M K A R L S R U H E

Institut für Reaktorbauelemente
(Zuverlässigkeit und Schadenskunde im Maschinenbau)

KfK 3581

Simulation von Fehlerverteilungen-
Untersuchung zur Wiedergewinnung von Verteilungen aus
unvollständigen Daten.

R. Häberer, A. Brückner

Kernforschungszentrum Karlsruhe GmbH, Karlsruhe

Als Manuskript vervielfältigt
Für diesen Bericht behalten wir uns alle Rechte vor

Kernforschungszentrum Karlsruhe GmbH
ISSN 0303-4003

Simulation von Fehlerverteilungen –

Untersuchung zur Wiedergewinnung von Verteilungen aus unvollständigen Daten

Am Beispiel simulierter Fehlerverteilungen wird gezeigt, wie eine an unvollständige Daten angepaßte Verteilung zu bewerten ist. Insbesondere wird untersucht, ob aus den simulierten Daten, die realitätsnah in ein Histogramm eingeteilt sind, die ursprünglichen Parameter und Verteilungen rekonstruiert werden können, welche Methoden bei der Kurvenanpassung zu sinnvollen Ergebnissen führen und welchen Einfluß eine untere Grenze auf die Ergebnisse hat.

Simulation of crack size distributions –

A Study on how to recover distributions from incomplete data.

Simulated crack size distributions are taken as an example to study the goodness of the fit of a distribution function to incomplete data. Special topics are whether the original parameters and distribution functions may be recovered, which are the appropriate methods of curve fitting and how the results are influenced by a lower bound.

Inhaltsverzeichnis

Seite

1.	Einleitung	1
2.	Probleme der Kurvenanpassung	2
2.1	Datenanalyse	2
2.2	Verteilungstyp	3
2.3	Kurvenanpassung	4
2.3.1	Methode der kleinsten Quadrate	4
2.3.2	χ^2 -Fit	4
2.3.3	Maximum-Likelihood-Methode	5
2.4	Beurteilung der Ergebnisse	6
2.4.1	Statistische Tests	6
2.4.2	Konfidenzintervalle	7
3.	Simulierte Fehlertiefenverteilungen	10
4.	Ergebnisse	11
4.1	Beurteilung der Methoden der Kurvenanpassung	11
4.2	Rekonstruktion einer Verteilung aus unvollständigen Daten	12
4.2.1	Exponentialverteilung Gl. (3.1)	12
4.2.2	Lognormalverteilung Gl. (3.2)	13
4.2.3	Weibullverteilung Gl. (3.3)	14
4.2.4	Zusammenfassung	15
4.3	Simulation von Fehlerverteilungen mit unterer Grenze	17
4.4	Einfluß der Simulation	15
4.4.1	Einfluß der Stichprobe	17
4.4.2	Einfluß der Klasseneinteilung	17
5.	Schlußfolgerungen	18
	Literatur	19
	Abbildungen	20
	Tabellen	39

1. Einleitung

Zur Abschätzung der Ausfallwahrscheinlichkeit eines Bauteils mit den Methoden der probabilistischen Bruchmechanik wird die Verteilung der im Bauteil enthaltenen Fehler sowie die Verteilung der Werkstoffkennwerte benötigt. Probabilistische Rechnungen haben gezeigt, daß die Unsicherheiten in den berechneten Ausfallwahrscheinlichkeiten hauptsächlich von der Anfangsrißverteilung und der Verteilung der Rißzähigkeit bestimmt werden [1, 2]. Diese Verteilungen werden durch eine Kurvenanpassung an Meßwerte gewonnen. Dabei tritt das Problem auf, daß vor allem für die Rißtiefe nur unvollständige Daten vorliegen.

In diesem Bericht soll untersucht werden, wie eine an unvollständige Daten angepaßte Fehlerverteilung zu bewerten ist. Dazu werden für vorgegebene Anfangsrißverteilungen Meßwerte mit einem Zufallszahlengenerator simuliert und in Klassen eingeteilt. An das so entstandene Histogramm wird anschließend wieder eine Fehlerverteilung angepaßt. Dabei wird auf folgende Fragen eingegangen:

- Welche Methoden der Kurvenanpassung ergeben die sinnvollsten Werte für die Parameter der rekonstruierten Fehlerverteilung ?
- Kann, ausgehend von dem Histogramm, eindeutig festgestellt werden, welche Verteilungsfunktion die Daten am besten beschreibt ?
- Welchen Einfluß hat die Einführung einer unteren Grenze der Fehlerverteilung ?

2. Probleme der Kurvenanpassung

2.1 Datenanalyse

Daten für eine Zufallsgröße werden in Form einer Stichprobe als Einzeldaten oder in Form von Histogrammen angegeben. Soll die Verteilung der Zufallsvariablen aus diesen Daten bestimmt werden, so ist die Abschätzung der Parameter dieser Verteilung aus den Einzeldaten auf jeden Fall genauer als aus dem Histogramm, da bei der Einteilung in Klassen stets Information verloren geht.

Für die gefundenen Fehler wird in der Literatur [3-5] im allgemeinen ein Histogramm angegeben. Dies liegt hauptsächlich daran, daß für die üblichen Messungen, z.B. mit Ultraschall, nur die Angabe von Fehlerklassen möglich ist.

In diesem Bericht wird, um möglichst realitätsnah zu bleiben, bei der Simulation von Fehlerverteilungen folgendermaßen vorgegangen: Zuerst werden mit Hilfe eines Zufallszahlengenerators Fehler simuliert, wobei eine in der Literatur gängige Anfangsrißverteilung zugrundegelegt wird. Der Umfang N dieser Stichprobe soll in der Größenordnung der Anzahl der Fehler pro Bauteil liegen und wurde auf $N=200$ festgesetzt [3,4]. Die Daten dieser Stichprobe werden in Klassen eingeteilt, wobei als Klassenbreite Werte zwischen 0.5 mm und 1 mm gewählt wurden.

Das so erzeugte Histogramm wird als experimentell gefundene Verteilung der Fehlergrößen betrachtet. Unabhängig von der ursprünglich zugrunde liegenden Verteilung wird jetzt versucht, die Daten mit einer beliebigen Dichtefunktion $f(a)$ zu beschreiben.

Grundsätzlich kann jede Funktion für $f(a)$ eingesetzt werden, sofern sie im Bereich $0 \leq a \leq \infty$ definiert und positiv ist und normiert werden kann. Im folgenden werden nur eine Exponentialverteilung, Weibullverteilung und Lognormalverteilung betrachtet, da typische experimentelle Fehlerverteilungen durch diese Funktionen beschrieben werden können.

2.2 Verteilungstyp

Ein Hinweis auf den zu verwendenden Verteilungstyp ergibt sich aus der Auftragung auf Wahrscheinlichkeitspapier [6]. Wurden in der Klasse (a_j, a_{j+1}) Y_j Werte ($j=1, \dots, K$, K =Anzahl der Klassen) festgestellt, so ist die experimentelle Verteilung gegeben durch

$$F(a_{i+1}) = \frac{1}{N+1} \sum_{j=1}^i Y_j \quad (2.1)$$

wobei N der Stichprobenumfang ist.

Im Wahrscheinlichkeitspapier werden $F(a)$ und a so gegeneinander aufgetragen, daß sich ein linearer Zusammenhang ergibt.

Für die Exponentenverteilung gilt:

$$\ln \left(\frac{1}{1-F(a)} \right) = \lambda a \quad (2.2)$$

Für die Lognormalverteilung erhält man

$$\phi^{-1} (F(a)) = \frac{1}{\sigma} \ln a - \frac{1}{\sigma} \ln \mu \quad (2.3)$$

wobei ϕ die Normalverteilung mit Mittelwert 0 und Standardabweichung 1 bezeichnet. Für die Weibullverteilung gilt:

$$\ln \ln \frac{1}{1-F(a)} = \eta \ln a - \eta \ln \sigma \quad (2.4)$$

Trägt man die experimentell bestimmten Wahrscheinlichkeiten $F(a)$ aus Gl. (2.1) jeweils in geeigneter Form über a bzw. $\ln a$ auf, so sollte sich ein linearer Verlauf ergeben, falls die Daten durch die angenommene Verteilung zu beschreiben sind. Systematische Abweichungen von der Linearität deuten darauf hin, daß die Daten nicht der angenommenen Verteilung folgen.

Eine Ausgleichsgerade wird nach Gl. (2.2) - (2.4) bestimmt, wobei die in Abschnitt 2.3 gefundenen Parameter eingesetzt werden.

Diese Ausgleichsgerade stimmt im allgemeinen nicht mit der Regressionsgeraden überein.

2.3 Kurvenanpassung

Nach der Festlegung des Verteilungstyps müssen die Parameter der Verteilung durch ein geeignetes Schätzverfahren bestimmt werden. In der vorliegenden Arbeit geschieht dies durch Kurvenanpassung.

2.3.1 Methode der kleinsten Quadrate

Bei der Methode der kleinsten Quadrate wird der quadratische Abstand Q^2 zwischen dem normierten Histogramm und der Verteilungsdichte minimiert. Sind in der Klasse (a_i, a_{i+1}) Y_i Meßpunkte aus einer Stichprobe vom Umfang N zu finden, so ist Q^2 gegeben durch

$$Q^2 = \sum_{i=1}^K \left(f\left(\frac{a_i + a_{i+1}}{2}\right) - \frac{Y_i}{\Delta a_i \cdot N} \right)^2 \quad (2.5)$$

wobei $\Delta a_i = a_{i+1} - a_i$ die Klassenbreite bezeichnet.

Bei dieser Methode der Kurvenanpassung wird also eine Ausgleichskurve durch die Klassenmitten des Histogramms gelegt.

2.3.2 χ^2 -Fit [7]

Der χ^2 -Fit ist eigentlich ein Spezialfall der Kurvenanpassung nach der Methode der kleinsten Quadrate. Im Unterschied zu 2.3.1 wird aber die Wahrscheinlichkeit p_i , in der i -ten Klasse einen Meßpunkt zu finden, als die bestimmende Größe betrachtet und nicht die Dichtefunktion $f(a)$. Es gilt

$$p_i = \int_{a_i}^{a_{i+1}} f(a) da \quad (2.6)$$

Die entsprechende Meßgröße ist die relative Häufigkeit Y_i/N . Außerdem wird eine Wichtung mit der zu erwartenden Schwankung σ_i^2 der Meßpunkte vorgenommen. Man kann zeigen, daß die Wahrscheinlichkeit, in der i -ten Klasse Y_i Werte zu finden, durch

eine Multinomialverteilung bestimmt wird [7]. Damit ist die Schwankung σ_i^2 gegeben durch

$$\sigma_i^2 = p_i \cdot N \quad (2.7)$$

Der aufgrund dieser Überlegungen zu minimierende Abstand wird mit χ^2 bezeichnet und berechnet sich nach

$$\chi^2 = \sum_{i=1}^K \frac{(N \cdot p_i - Y_i)^2}{N p_i} \quad (2.8)$$

2.3.3 Maximum - Likelihood-Methode [7]

Die Maximum-Likelihood-Methode beruht darauf, daß die Wahrscheinlichkeit maximiert wird, gerade die Meßpunkte zu finden, die in der Stichprobe gemessen wurden. Sind $a_1 \dots a_N$ die Werte der Stichprobe und ist $f(a)$ die Verteilungsfunktion, so muß das Maximum der Likelihoodfunktion

$$L = f(a_1) \dots f(a_N) \quad (2.9)$$

bestimmt werden.

Für ein Histogramm geht Gl.(2.9) über in

$$L = N! \prod_{i=1}^K \frac{p_i^{Y_i}}{Y_i!} \quad (2.10)$$

wegen der zugrundeliegenden Multinomialverteilung. Dies führt dazu, daß das Maximum der Funktion

$$\ln L' = \sum_{i=1}^K Y_i \ln (N p_i) \quad (2.11)$$

berechnet werden muß, um die Maximum-Likelihood-Schätzwerte der Parameter aus dem Histogramm zu bestimmen.

Die Methoden aus 2.3.2 und 2.3.3 sollten bei genügend hoher Klassen- und Stichprobenzahl zu demselben Ergebnis führen.

2.4 Beurteilung der Ergebnisse

2.4.1 Statistische Tests

Die Hypothese H_0 , daß die Stichprobe aus der gewählten Verteilung mit den nach 2.3 bestimmten Parametern stammt, kann mit Hilfe eines statistischen Tests überprüft werden. Der Grundgedanke dabei ist, aus der Stichprobe und der daraus bestimmten Verteilungsfunktion eine neue Zufallsgröße Z_{test} zu berechnen, deren Verteilung unabhängig von der Hypothese H_0 ist. Damit kann die Wahrscheinlichkeit $1-\alpha$ bestimmt werden, daß Z_{test} kleiner als eine vorgegebene kritische Größe Z_{krit} ist.

$$P(Z_{\text{test}} < Z_{\text{krit}}) = 1 - \alpha \quad (2.12)$$

Ergibt sich aus der Stichprobe und der angepaßten Verteilung ein Wert $Z_{\text{test}} > Z_{\text{krit}}$, so wird die Hypothese H_0 abgelehnt.

Die Signifikanzzahl α wird interpretiert als die Wahrscheinlichkeit, einen Wert Z_{test} größer als Z_{krit} zu finden, obwohl H_0 zutrifft (Fehler 1. Art). In dieser Arbeit wird stets $\alpha = 5\%$ gewählt.

Der bekannteste Test für in Klassen eingeteilte Daten ist der χ^2 -Test, wobei gilt:

$$Z_{\text{test}} = \chi^2 \quad (2.13)$$

nach Gl. (2.8); die kritische Größe $\chi^2_{\text{krit}} = Z_{\text{krit}}$ berechnet sich aus der χ^2 -Verteilung mit $K-N_p-1$ Freiheitsgraden (N_p = Anzahl der zu bestimmenden Parameter).

2.4.2 Konfidenzintervalle

Neben der Punktschätzung von Parametern etwa durch die Methoden aus 2.3 ist in der Statistik auch die Angabe von Konfidenzintervallen für die Parameter üblich. Diese werden dadurch berechnet, daß die Verteilung einer Funktion der Stichprobenwerte, z.B. des Mittelwertes, bekannt ist. Gibt man die Wahrscheinlichkeit an, diese Funktion innerhalb gewisser Grenzen zu finden, so kann nach Auflösen nach dem interessierenden Parameter ein Konfidenzintervall bestimmt werden, das so etwas wie die mögliche Genauigkeit der Parameter angibt. Ergeben die verschiedenen Kurvenanpassungsverfahren Werte der zu schätzenden Parameter, die innerhalb des Konfidenzintervalles liegen, so werden diese Werte als übereinstimmend angesehen; andernfalls werden die Unterschiede als signifikant betrachtet.

Nach dem zentralen Grenzwertsatz ist die Größe

$$T = \sqrt{N} \frac{\bar{a} - E(a)}{s} \quad (2.14)$$

mit

$$\bar{a} = \frac{1}{N} \sum_{i=1}^N a_i \quad (2.15)$$

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (a_i - \bar{a})^2 \quad (2.16)$$

$$s = \sqrt{s^2}$$

asymptotisch t - verteilt und die Größe

$$Y = (N - 1) \cdot \frac{s^2}{\text{Var}(a)} \quad (2.17)$$

asymptotisch χ^2 - verteilt mit N-1 Freiheitsgraden. E(a) bezeichnet den Erwartungswert und Var(a) die Varianz der Zufallsvariablen a. Über den Zusammenhang zwischen Erwartungswert bzw. Varianz mit den Parametern der angenommenen Verteilung können Konfidenzintervalle

gewonnen werden. Für große N gilt |8|:

$$\bar{a} - \frac{s \cdot c}{\sqrt{N}} \leq E(a) \leq \bar{a} + \frac{s \cdot c}{\sqrt{N}} \quad (2.18)$$

und

$$\frac{(N-1) \cdot s^2}{c_2} \leq \text{Var}(a) \leq \frac{(N-1) \cdot s^2}{c_1} \quad (2.19)$$

Bei einer Konfidenzzahl $\gamma = 95\%$ ergibt sich aus den asymptotischen Verteilungen:

$$\begin{aligned} c &= 1.97 \\ c_1 &= 239.5 \\ c_2 &= 161.4 \end{aligned} \quad (2.20)$$

Aus Gl. (2.18) folgt für den Parameter λ der Exponentialverteilung:

$$\begin{aligned} \lambda_2 &\leq \lambda \leq \lambda_1 \quad \text{mit} \\ \lambda_1 &= \left(\bar{a} - \frac{s \cdot c}{\sqrt{N}} \right)^{-1} \\ \lambda_2 &= \left(\bar{a} + \frac{s \cdot c}{\sqrt{N}} \right)^{-1} \end{aligned} \quad (2.21)$$

Eine Lognormalverteilung der Zufallsvariablen a mit den Parametern μ, σ ist identisch mit einer Normalverteilung der Zufallsvariablen $\ln a$ und den Parametern $\ln \mu, \sigma$. Aus den Gl. (2.18), (2.19) folgen damit die Konfidenzintervalle

$$\begin{aligned} \mu_1 &\leq \mu \leq \mu_2 \\ \text{mit } \mu_1 &= \exp \left(\overline{\ln a} - \frac{s \ln \cdot c}{\sqrt{N}} \right) \\ \mu_2 &= \exp \left(\overline{\ln a} + \frac{s \ln \cdot c}{\sqrt{N}} \right) \end{aligned} \quad (2.22)$$

$$\overline{\ln a} = \frac{1}{N} \sum_{i=1}^N \ln a_i \quad (2.23)$$

$$s_{\ln}^2 = \frac{1}{N-1} \sum_{i=1}^N (\ln a_i - \overline{\ln a})^2 \quad (2.24)$$

und $\sigma_1 \leq \sigma \leq \sigma_2$

mit $\sigma_1 = \frac{s_{\ln}^2 \cdot (N-1)}{c_1} \quad (2.25)$

$$\sigma_2 = \frac{s_{\ln}^2 \cdot (N-1)}{c_2}$$

Bei der Weibullverteilung ist die Beziehung zwischen $E(a)$, $\text{Var}(a)$ und den Parametern zu kompliziert, um auf einfache Abschätzungen für die Konfidenzintervalle zu führen. Es wird deshalb das in [9] angegebene Verfahren benutzt.

Da die Reißgrößen meist nur in Form eines Histogramms vorliegen, sind der Stichprobenmittelwert \bar{a} und die Stichprobenvarianz s^2 nicht genau berechenbar. Im folgenden werden die Abschätzungen

$$\bar{a} \approx \bar{a}_H = \frac{1}{N} \sum_{i=1}^K Y_i \cdot \frac{1}{2} (a_i + a_{i+1}) \quad (2.26)$$

$$s^2 \approx s_H^2 = \frac{1}{N-1} \sum_{i=1}^K Y_i \cdot \left(\frac{1}{2} (a_i + a_{i+1}) - \bar{a}_H \right)^2 \quad (2.27)$$

$$\overline{\ln a} \approx \overline{\ln a}_H = \frac{1}{N} \sum_{i=1}^K Y_i \cdot \frac{1}{2} (\ln a_i + \ln a_{i+1}) \quad (2.28)$$

$$s_{\ln}^2 \approx s_{\ln H}^2 = \frac{1}{N-1} \sum_{i=1}^K Y_i \cdot \left(\frac{1}{2} (\ln a_i + \ln a_{i+1}) - \overline{\ln a}_H \right)^2 \quad (2.29)$$

verwendet, wobei über die Klassen (a_i, a_{i+1}) summiert wird.

3. Simulierte Fehlertiefenverteilungen

Bei den simulierten Verteilungen handelt es sich um

- eine in [10] angegebene auf einen Druckbehälter eines Leichtwasserreaktors bezogene Exponentialverteilung für die Fehlertiefen nach der Fertigung und vor der Inspektion (Abb. 1).

$$f(a) = \lambda \cdot \exp(-\lambda \cdot a) \quad (3.1)$$

mit $\lambda = 0.161 \text{ 1/mm}$

- eine in [1] angegebene an die Daten von [5] gefittete Lognormalverteilung für eine Stahlhülle nach der Fertigung und Beseitigung einiger Fehler (Abb. 2):

$$f(a) = \frac{1}{\sqrt{2\pi} \cdot \sigma \cdot a} \exp \left[-\frac{1}{2\sigma^2} \ln^2 \left(\frac{a}{\mu} \right) \right] \quad (3.2)$$

mit $\sigma = 0.174$; $\mu = 3.855 \text{ mm}$

- eine Weibullverteilung, die in [11] an die in [4] angegebenen Daten angepaßt wurde (Abb. 3):

$$f(a) = \eta \cdot \sigma (a \cdot \sigma)^{\eta-1} \cdot \exp(-(a \cdot \sigma)^\eta) \quad (3.3)$$

mit $\eta = 0.51$; $\sigma = 2.3 \text{ 1/mm}$

4. Ergebnisse

Aus den in Abschnitt 3 angegebenen Verteilungen wird nach der in 2.1 beschriebenen Vorgehensweise eine Fehlerverteilung simuliert. Die so erhaltenen Histogramme sind zusammen mit den ursprünglichen Verteilungen in Abb. 1-3 eingetragen. Diese Daten sollen jetzt dazu benutzt werden, die in der Einleitung angeschnittenen Probleme näher zu untersuchen.

4.1 Beurteilung der Methoden der Kurvenanpassung

Am Beispiel der Weibullverteilung wird untersucht, wie sich die verschiedenen Methoden zur Bestimmung der Verteilungsparameter auswirken. Dazu wird an das in Abb. 3 dargestellte Histogramm, dem die Weibullverteilung zugrunde liegt, eine Weibullverteilung nach der Methode der kleinsten Quadrate Gl. (2.5), mit dem χ^2 -Fit Gl. (2.8) und der Maximum-Likelihood-Methode Gl. (2.11) angepaßt. Hierfür wurde das von CERN Computer Centre entwickelte Programm MINUIT [12] verwendet. Die Ergebnisse sind in Tabelle 1 und Abb. 4-5 dargestellt. Im einzelnen gilt:

- χ^2 -Fit und Maximum-Likelihood-Fit liefern nahezu übereinstimmende Werte für die Parameter der Weibullverteilung, die in der Größenordnung der ursprünglichen Parameter liegen. Nach der Methode der kleinsten Quadrate erhält man für η einen deutlich abweichenden Wert.
- Berechnet man nach den in [9] angegebenen Methoden aus der Stichprobe Konfidenzintervalle der Konfidenzzahl 95% für die Parameter, so liegen die Ergebnisse aus dem χ^2 -Fit und dem Maximum-Likelihood-Fit innerhalb des Konfidenzintervalles.
- Aus Abb. 4-5 folgt, daß die nach der Methode der kleinsten Quadrate bestimmte Verteilung für große Fehlertiefen, d.h. für kleine relative Häufigkeiten, nicht mit dem Histogramm der gemessenen Werte übereinstimmt.

Insgesamt läßt sich aus den Ergebnissen folgern, daß die Methode der kleinsten Quadrate zur Bestimmung von Parametern einer Verteilung ungeeignet ist. χ^2 -Fit und Maximum-Likelihood-Methode liefern innerhalb der Fehlerschranken übereinstimmende Ergebnisse. Für die folgenden Untersuchungen werden deshalb nur noch diese beiden Methoden verwendet.

4.2 Rekonstruktion einer Verteilung aus unvollständigen Daten

Bei realen Fehlerverteilungen ist unbekannt, welche Verteilungsfunktion zur Beschreibung der Daten ausgewählt werden soll. Mit statistischen Tests ist nicht immer zu entscheiden, welche Verteilungsfunktion am besten geeignet ist [11].

Die simulierten Fehlerverteilungsfunktionen werden deshalb daraufhin untersucht, ob die bei der Simulation zugrundeliegenden Verteilungen gewonnen werden können oder andere Verteilungen ebenfalls mit den Daten verträglich sind. Es werden jeweils die Exponentialverteilung, die Lognormalverteilung und die Weibullverteilung verwendet. Zunächst werden die Daten auf Wahrscheinlichkeitspapier aufgetragen und auf systematische Abweichungen gegenüber dem linearen Verlauf untersucht. Ein χ^2 -Test gibt zusätzlich Auskunft darüber, ob die Daten mit der gewählten Verteilung beschrieben werden können.

4.2.1 Exponentialverteilung Gl. (3.1)

Die Ergebnisse für das Histogramm, das aus der Exponentialverteilung Gl. (3.1) gewonnen wurde, sind in Abb. 6, 7a - 7c und Tab. 2 zusammengefaßt. Daraus folgt:

- Die Auftragung der Daten auf Wahrscheinlichkeitspapier für die Lognormalverteilung zeigt eine leichte systematische Krümmung (Abb. 7c), während Exponentialverteilung und Weibullverteilung gleich gute Ergebnisse liefern (Abb. 7a - 7b).
- Der Parameter η der Weibullverteilung ist ungefähr gleich eins, so daß die angepaßte Weibullverteilung mit der Exponentialverteilung zusammenfällt.
- Alle drei Verteilungen stimmen nach Abb. 6 gut mit den Daten überein.
- Der χ^2 -Test akzeptiert alle drei Verteilungen.
- Berechnet man nach Gl. (2.21), (2.26), (2.27) aus der Stichprobe ein Konfidenzintervall für den Parameter λ , so liegt die ursprüngliche Verteilung innerhalb des Vertrauensbereiches.

4.2.2 Lognormalverteilung Gl. (3.2)

Für die lognormalverteilten Fehler finden sich die Resultate in Abb. 8, 9a, 9b und Tab. 3. Es gilt:

- Die Exponentialverteilung ist zur Beschreibung der Daten ungeeignet (Abb. 8)
- Die Auftragung im Wahrscheinlichkeitspapier ergibt bei der Lognormalverteilung eine leichte Abweichung vom linearen Verlauf für große Werte von a (Abb. 9a). Bei der Weibullverteilung erhält man nach Abb. 9b eine systematische Krümmung.
- Beim χ^2 -Test wird nur die Lognormalverteilung akzeptiert, die Exponentialverteilung und die Weibullverteilung dagegen eindeutig zurückgewiesen.

- Die durch die Kurvenanpassung bestimmten Werte der Parameter der Lognormalverteilung sind mit den ursprünglichen fast identisch. Berechnet man nach Gl. (2.22) und (2.25) Konfidenzintervalle für μ und σ , so liegt für μ der ursprüngliche Wert im Vertrauensbereich, für σ dagegen sowohl der angepaßte als auch der ursprüngliche Wert nicht. Dies ist auf die Näherungen Gl. (2.28) und (2.29) zurückzuführen.

4.2.3 Weibullverteilung Gl. (3.3)

Abb. 10, 11a - 11c und Tab. 4 enthalten die Ergebnisse für die weibullverteilten Fehler. Daraus folgt:

- Die Exponentialverteilung ist aufgrund der Auftragung im Wahrscheinlichkeitspapier Abb. 11b und von Abb. 10 zur Beschreibung der Daten ungeeignet.
- Für die Weibullverteilung und die Lognormalverteilung ergibt sich im Wahrscheinlichkeitspapier ungefähr eine Gerade.
- Beim χ^2 -Test werden Weibull- und Lognormalverteilung akzeptiert.
- Die ursprünglichen Parameter der Weibullverteilung und die Werte aus der Kurvenanpassung unterscheiden sich nur wenig. Berechnet man nach [9] Konfidenzintervalle für η und σ , so stellt sich heraus, daß für σ der ursprüngliche und der gefittete Wert im Vertrauensbereich liegen, während das Konfidenzintervall für η die ursprünglichen Parameter nicht enthält. Dies ist, wie bei der Lognormalverteilung, in Abschnitt 4.2.2 auf Näherungen zurückzuführen, die gemacht wurden, um aus den Histogramm Grenzen für die Parameter zu gewinnen.

4.2.4 Zusammenfassung

Berücksichtigt man die Ergebnisse für alle drei Verteilungen, so können folgende Schlußfolgerungen gezogen werden:

- Beschränkt man sich auf den der Simulation zugrundeliegenden Verteilungstyp, so findet man mit den angegebenen Methoden Parameter, die im Rahmen der vorhandenen Unsicherheiten übereinstimmen; es ist also eine Rekonstruktion der Verteilung aus den unvollständigen Daten möglich.
- Läßt man verschiedene Verteilungsfunktionen zu, so ist die Unterscheidung zwischen Lognormal- und Weibullverteilung aufgrund statistischer Argumente nicht immer möglich.
- In allen betrachteten Fällen konnten die Daten mit einer Lognormalverteilung beschrieben werden.

4.3 Simulation von Fehlerverteilungen mit unterer Grenze

Eine reale Fehlerverteilung, die aus den Ergebnissen der zerstörungsfreien Prüfung eines Bauteiles oder von Proben gewonnen wurde, hat eine "natürliche" untere Grenze bei der Registriergrenze des jeweiligen Prüfverfahrens. Um diese Verhältnisse in den simulierten Fehlerverteilungen nachzubilden, werden die Verteilungen Gl. (3.1) - (3.3) bei einer unteren Grenze a_u abgeschnitten. Für a_u wird bei der Lognormalverteilung Gl. (3.2) ein Wert von 3 mm angenommen, da in [5] die Registriergrenze bei etwa 4 mm lag. Bei der Weibullverteilung Gl. (3.3) wird $a_u = 0.5$ mm gewählt, da in [4] die unterste Klasse der Fehlertiefenverteilung mit (0, 1mm) angegeben wurde. Diese Grenze wird auch bei der Exponentialverteilung Gl. (3.1) eingesetzt. Die abgeschnittenen Verteilungen sind dann:

$$f(a) = \begin{cases} 1,084 \lambda \cdot \exp [- \lambda \cdot a] & \text{für } a \geq 0,5 \text{ mm} \\ 0 & \text{für } a < 0,5 \text{ mm} \end{cases} \quad (4.1)$$

$$\lambda = 0,161 \text{ mm}^{-1}$$

$$f(a) = \begin{cases} \frac{1,081}{\sqrt{2\pi} \cdot a \cdot \sigma} \cdot \exp \left[- \frac{1}{2} \left(\frac{\ln(a/\mu)}{\sigma} \right)^2 \right] & \text{für } a \geq 3,0 \text{ mm} \\ 0 & \text{für } a < 3,0 \text{ mm} \end{cases}$$

$$\mu = 3,855 \text{ mm}, \quad \sigma = 0,174 \quad (4.2)$$

$$f(a) = \begin{cases} 1,519 \cdot \eta \cdot \sigma (a \cdot \sigma)^{\eta-1} \cdot \exp(-(a \cdot \sigma)^\eta) & \text{für } a \geq 0,5 \text{ mm} \\ 0 & \text{für } a < 0,5 \text{ mm} \end{cases}$$

$$\sigma = 2,3 \text{ mm}, \quad \eta = 0,51 \quad (4.3)$$

Es wurden Zufallszahlen nach den Verteilungen Gl. (4.1) - (4.3) erzeugt und in Histogramme mit unterer Grenze a_U eingeteilt. Die Kurvenanpassung erfolgt ebenfalls mit abgeschnittenen Verteilungen. Bei allen drei Histogrammen wird jeweils wieder ein Fit mit der Exponential-, der Lognormal- und der Weibullverteilung versucht. Die Ergebnisse finden sich in Abb. 12 - 14 und in Tab. 5 - 7. Es gilt:

- Alle drei Histogramme lassen sich mit einer Lognormal- und einer Weibullverteilung beschreiben. Die Unterschiede zwischen den Fits mit diesen beiden Verteilungen werden kleiner verglichen mit den Verhältnissen ohne untere Grenze.
- Die ursprünglichen Parameter der zugrundeliegenden Verteilungsfunktionen können nur zum Teil wieder erhalten werden. Insbesondere bei der Weibullverteilung (Tab. 7) ergeben sich große Abweichungen.
- Wie schon bei den Verteilungen ohne untere Grenze, ist die Lognormalverteilung die konservativste der von den Tests akzeptierten Verteilungen, d.h. bei der Extrapolation auf Fehler großer

Tiefe erhält man die höchsten Auftrittswahrscheinlichkeiten.

4.4 Einfluß der Simulation

4.4.1 Einfluß der Stichprobe

Bei einer endlichen Stichprobe ist die Anzahl der Elemente pro Klasse nach Gl.(2.7) statistischen Schwankungen von der Größenordnung $1/\sqrt{N}$ (N = Umfang der Stichprobe) unterworfen. Bei den hier betrachteten Stichproben ist $N = 200$, d.h. die statistische Unsicherheit liegt bei etwa 7 %. Um diesen Einfluß der Stichprobe auf die simulierte Fehlerverteilung zu überprüfen, wurden verschiedene Sätze von Zufallszahlen erzeugt, in Klassen eingeteilt und an das Histogramm die ursprüngliche Verteilungsfunktion angepaßt. Für die Weibullverteilung Gl. (3.3) sind die Ergebnisse in Tab. 8 zu finden.

Die auftretenden Schwankungen der Parameter sind recht beträchtlich; das aus der ersten Stichprobe berechnete Konfidenzintervall erweist sich beim Parameter η als zu eng.

4.4.2 Einfluß der Klasseneinteilung

Aus theoretischen Überlegungen [7] folgt, daß Daten möglichst in Klassen gleicher Wahrscheinlichkeit eingeteilt werden sollten. Bei experimentellen Fehlerverteilungen ist die Klassenbreite meist durch die Genauigkeit des zerstörungsfreien Prüfverfahrens vorgegeben. Deshalb wird im folgenden der Einfluß der Klassenbreite auf die rekonstruierte Fehlerverteilung untersucht.

Ein Beispiel ist in Abb. 15 gegeben. Mit den verwendeten Methoden ergibt sich danach kein bedeutender Effekt, da bei der Kurvenanpassung nach Gl. (2.8) bzw. Gl. (2.11) die Werte in jeder Klasse $|a_i, a_{i+1}|$ mit der Wahrscheinlichkeit p_i , einen Fehler in dieser Klasse zu finden, gewichtet werden.

5. Schlußfolgerungen

Die Simulation von experimentellen Fehlerverteilungen zeigt, daß mit dem χ^2 -Fit und der Maximum-Likelihood-Methode die Parameter der ursprünglichen Verteilungen aus unvollständigen Meßdaten annähernd wiedergewonnen werden können. Selbst in dem hier vorliegenden Fall, wo die zu erwartende Verteilungsfunktion bekannt ist, ist aber aus dem Histogramm keine eindeutige Aussage über die beste Verteilungsfunktion möglich. Weibull- und Lognormalverteilung geben in vielen Fällen gleich gute Ergebnisse. Im Sinne einer konservativen Betrachtungsweise sollte bei der Anpassung von experimentellen Fehlerverteilungen die Lognormalverteilung bevorzugt verwendet werden, da dadurch die Häufigkeit von großen Fehlern, die tiefer als die gefundenen Fehler sind, am höchsten wird.

Literatur

- [1] Häberer, R., Brückner, A., Munz, D., "Zuverlässigkeitsberechnung mit Hilfe der probabilistischen Bruchmechanik am Beispiel des Sicherheitseinschlusses von Druckwasserreaktoren", Teil I, KfK-Bericht Nr. 3458, 1982.
- [2] Brückner, A., Munz, D., "The Effect of Curve Fitting on the Prediction of Failure Probabilities from the Scatter in Crack Geometry and Fracture Toughness", demnächst in Eng. Fract. Mech..
- [3] Becher, P.E., Hansen, B., "Statistical Evaluation of Defects in Welds and Design Implications", Danish Atomic Energy Commission, Risø Denmark, 1974.
- [4] Raussi, J., Tiainen, O., "The Effect of Inspection and Repair on the Size Distributions of the Weld Imperfections in Nuclear Power Plant Pipes", Nuclear Engineering and Design 60, 1980, 395 - 399.
- [5] Wellein, R., "Beurteilung der Zuverlässigkeit der Stahlhülle mit Hilfe der probabilistischen Bruchmechanik", in: Schuëller, G.I. et al. "Zuverlässigkeitsbeurteilung für den Sicherheitseinschluß am Beispiel des Druckwasserreaktors", Anhangband, BMFT 150201C, TU München, 1981.
- [6] Hahn, G., Shapiro, S., "Statistical models in engineering", J. Wiley, New York, 1967.
- [7] Eadie, W.T., Drijard, D., James, F.E., Roos, M. and Sadoulet, B., "Statistical Methods in Experimental Physics", North-Holland, Amsterdam, 1971.
- [8] Kreyszig, E., "Statistische Methoden und ihre Anwendungen", Vandenhoeck & Ruprecht, 7. Auflage 1979.
- [9] Thoman, R.D., Bain, J.L., Antle, E.C., "Inferences on the Parameters of the Weibull Distribution", Technometrics Vol. 11, No. 3, August 1969.
- [10] Marshall, W., "An Assessment of the Integrity of PWR Pressure Vessels", A U.K. Study Group Report, in: Jaeger, T.A., Boley, B.A. (eds), Structural Mechanics in Reactor Technology, Luxembourg 1977, G6/1 S. 1 - 11.
- [11] Brückner, A., Munz, D., "Curve Fitting to Defect Size Distribution for the Calculation of Failure Probabilities" demnächst in Nucl. Eng. Des..
- [12] FORTRAN Programm "MINUIT", CERN COMPUTER CENTRE, 1975.

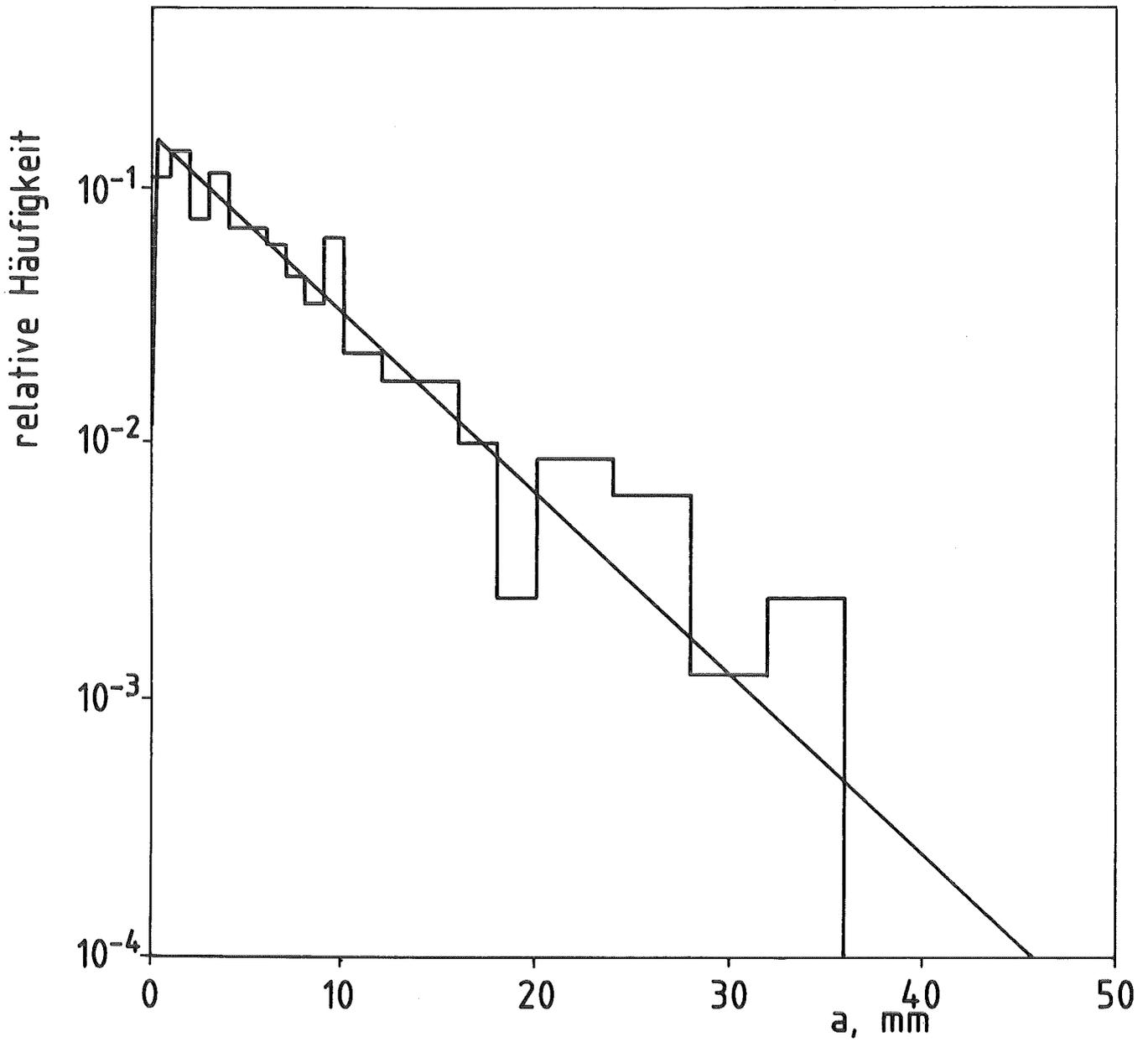


Abb. 1 Simulation von Rissen mit einer exponentiell-
verteilten Tiefe ($\lambda = 0.161$ 1/mm)

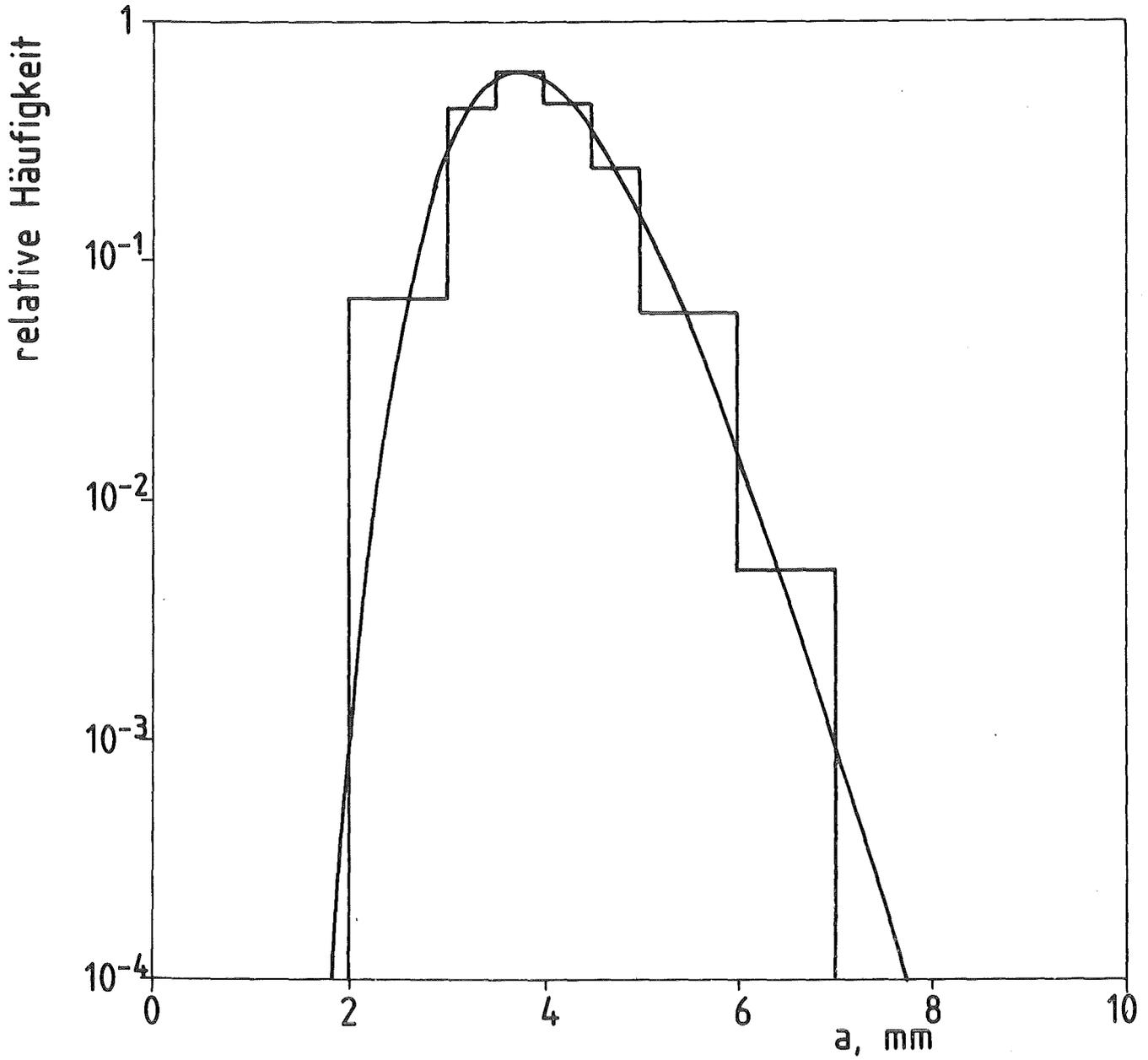


Abb. 2 Simulation von Rissen mit einer lognormalverteilten Tiefe ($\mu=3.855$ mm ; $\sigma=0.174$)

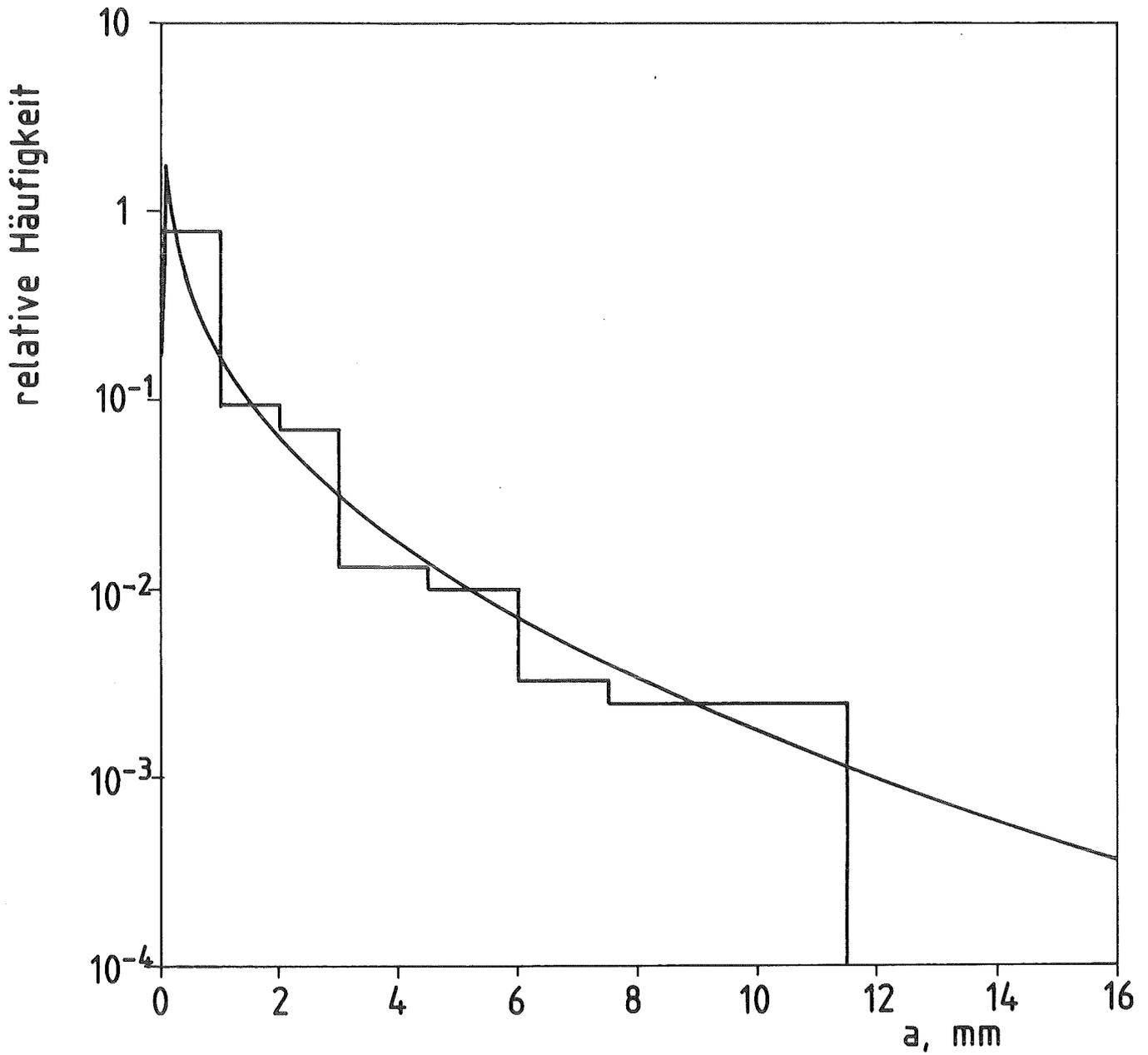


Abb. 3 Simulation von Rissen mit einer weibullverteilten Rißtiefe ($\sigma=2.3$ 1/mm ; $\eta=0.51$)

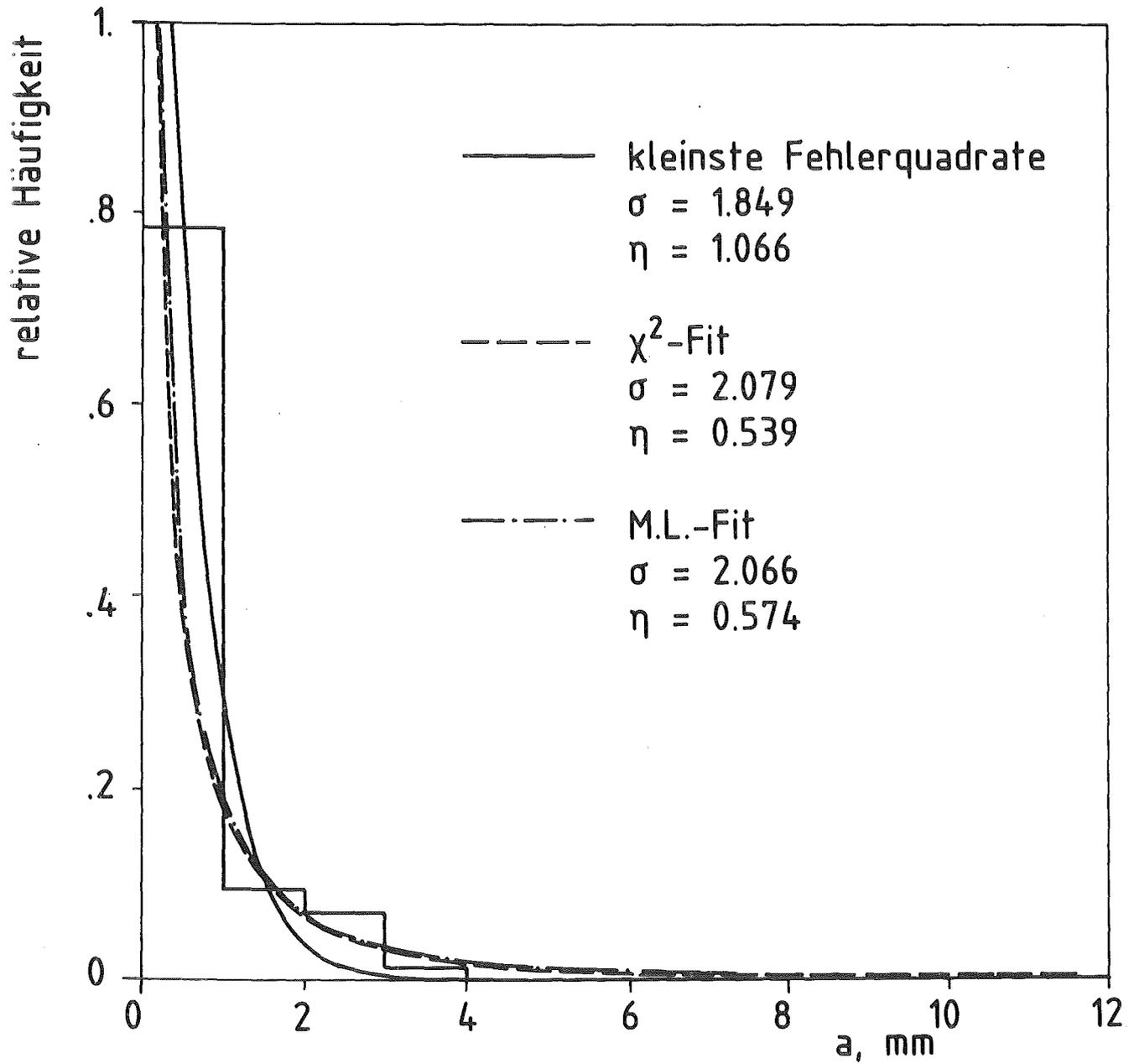


Abb. 4 Anpassung einer Weibullverteilung an weibullverteilte Zufallszahlen ($\sigma=2.3$ 1/mm; $\eta=0.51$) mit unterschiedlichen Methoden der Kurvenanpassung

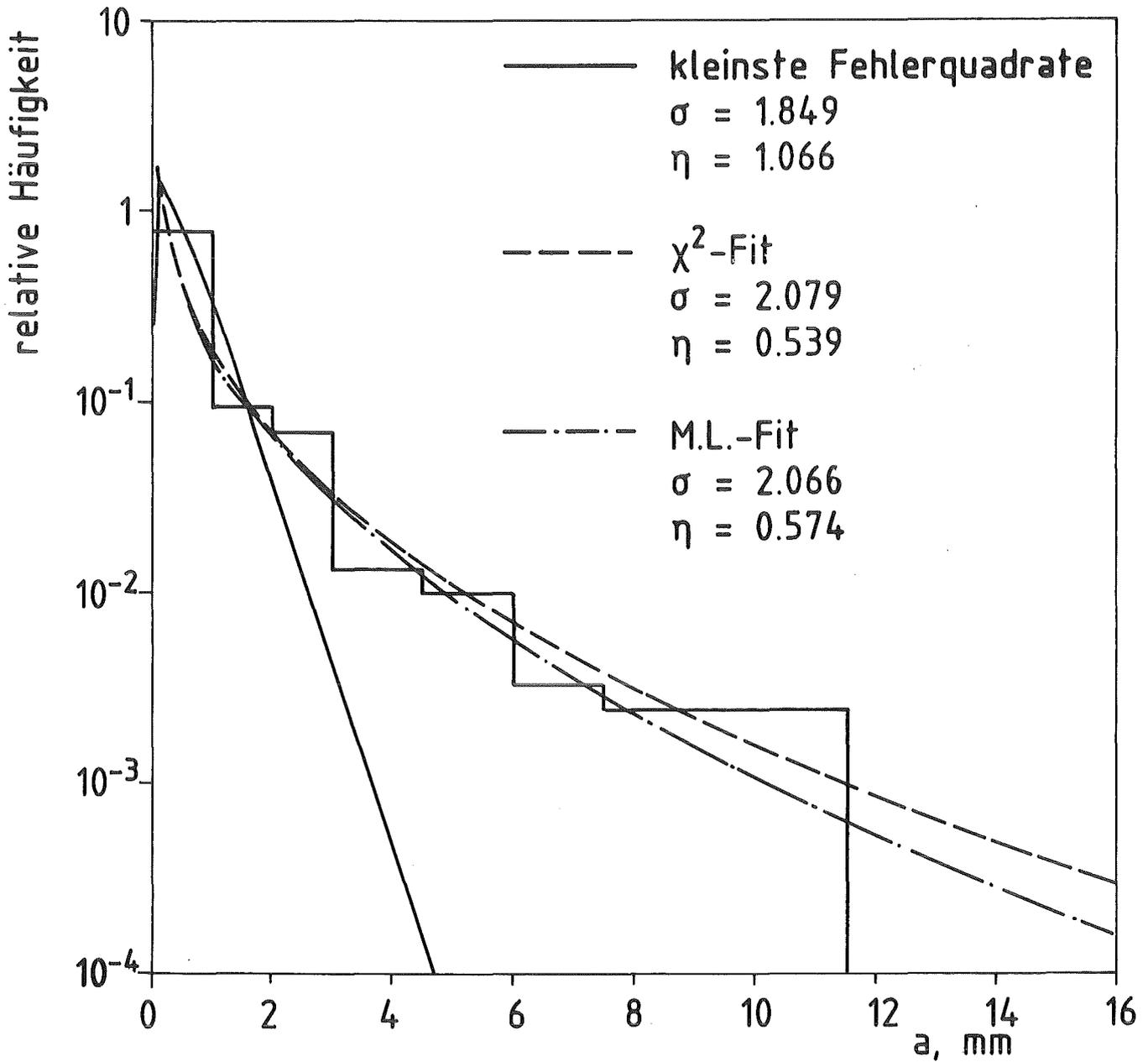


Abb. 5 Anpassung einer Weibullverteilung an weibullverteilte Zufallszahlen ($\sigma=2.3$ 1/mm, $\eta=0.51$) mit unterschiedlichen Methoden der Kurvenanpassung

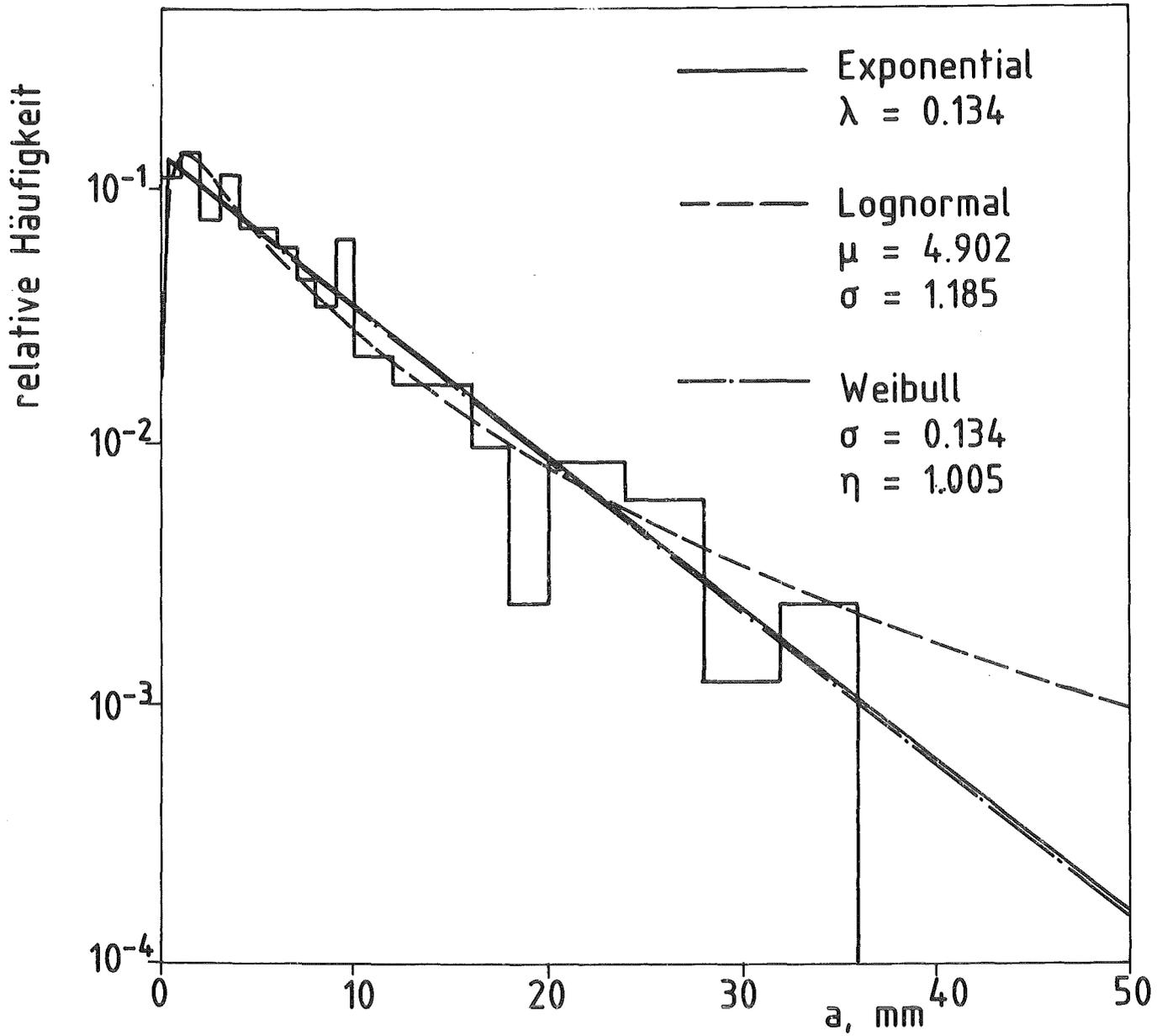


Abb.6 χ^2 -Fit verschiedener Verteilungstypen an exponentiellverteilte Zufallszahlen ($\lambda=0.161$ 1/mm)

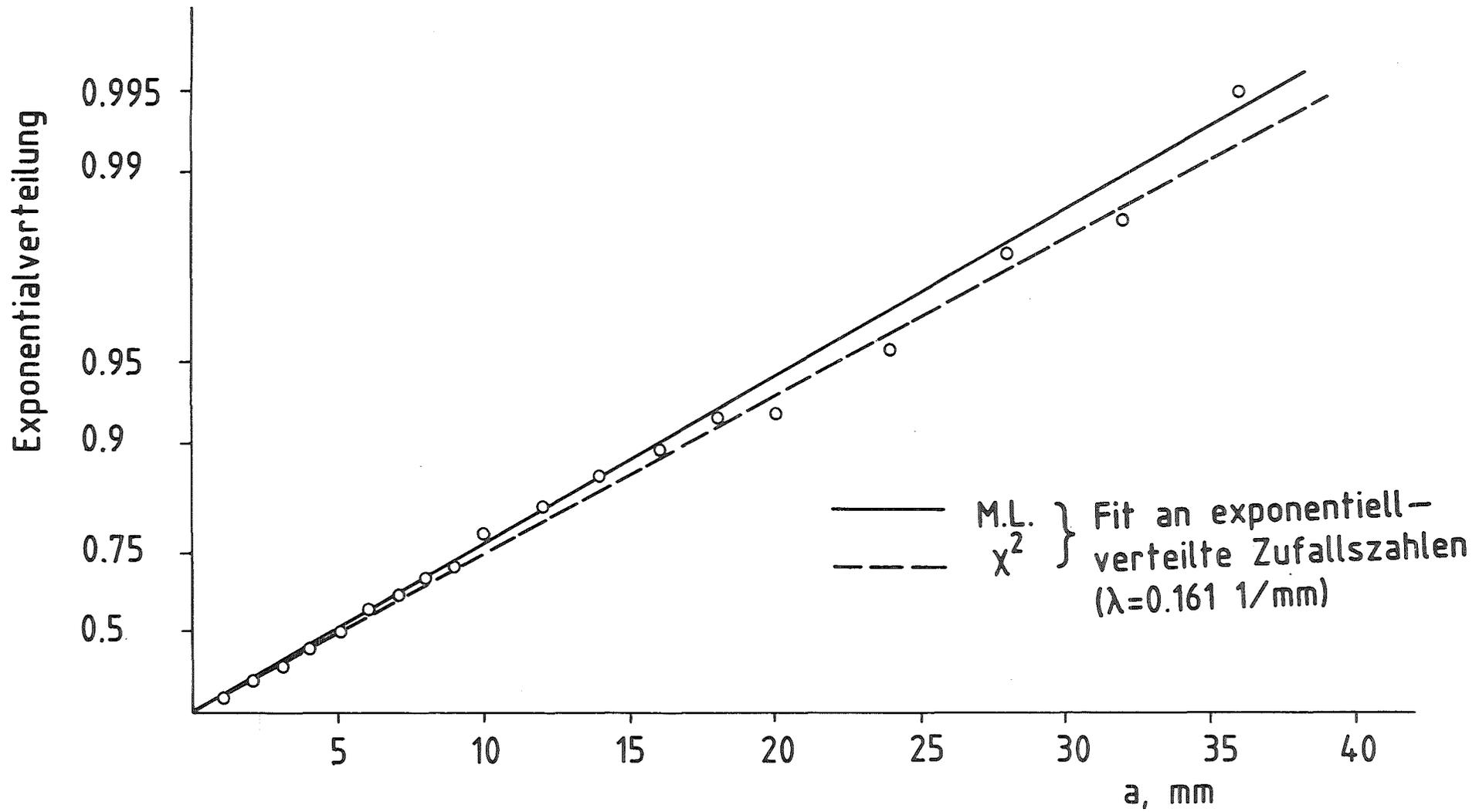


Abb. 7a Wahrscheinlichkeitspapier für Exponentialverteilungen

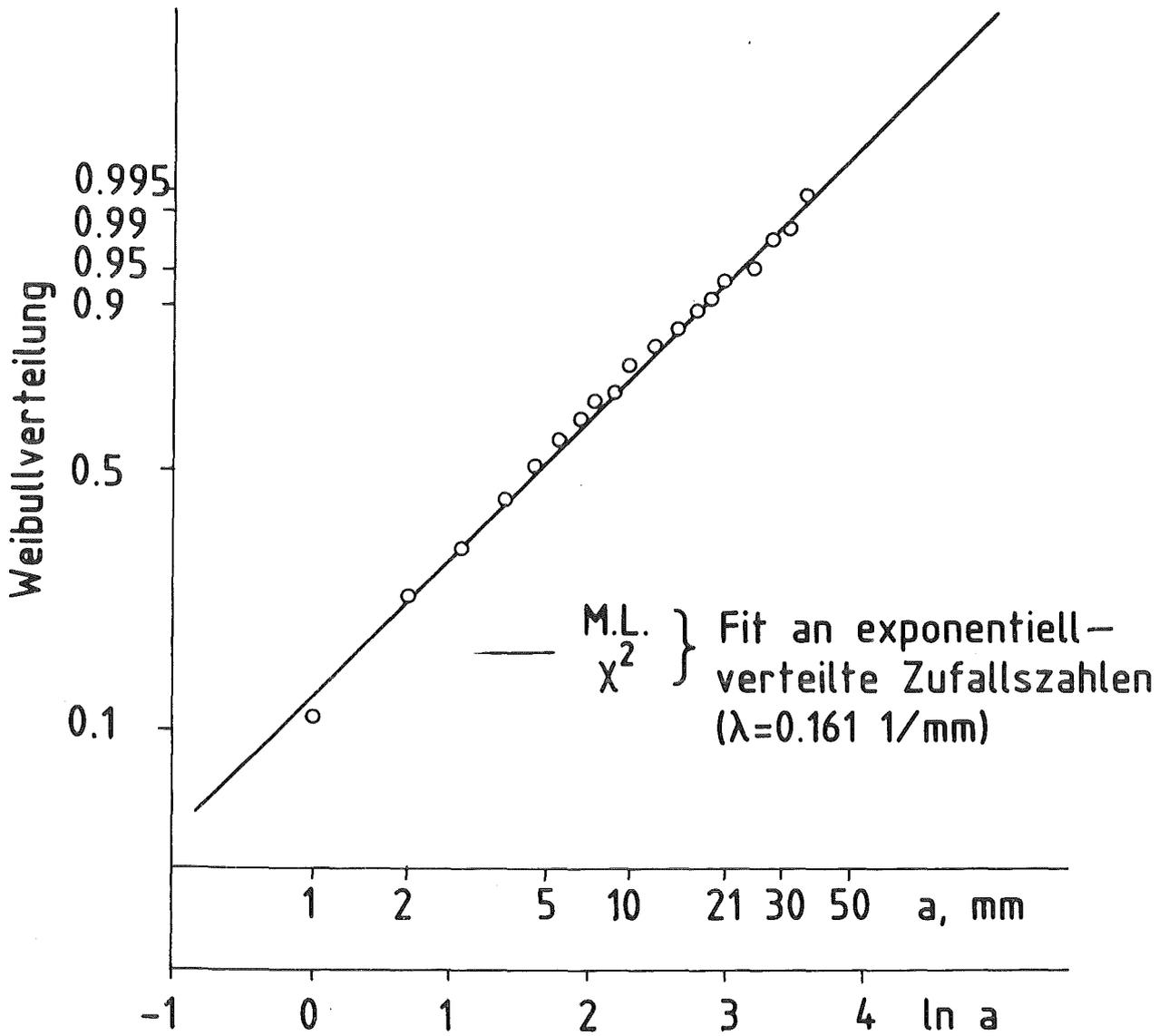


Abb. 7b Wahrscheinlichkeitspapier für Weibullverteilungen

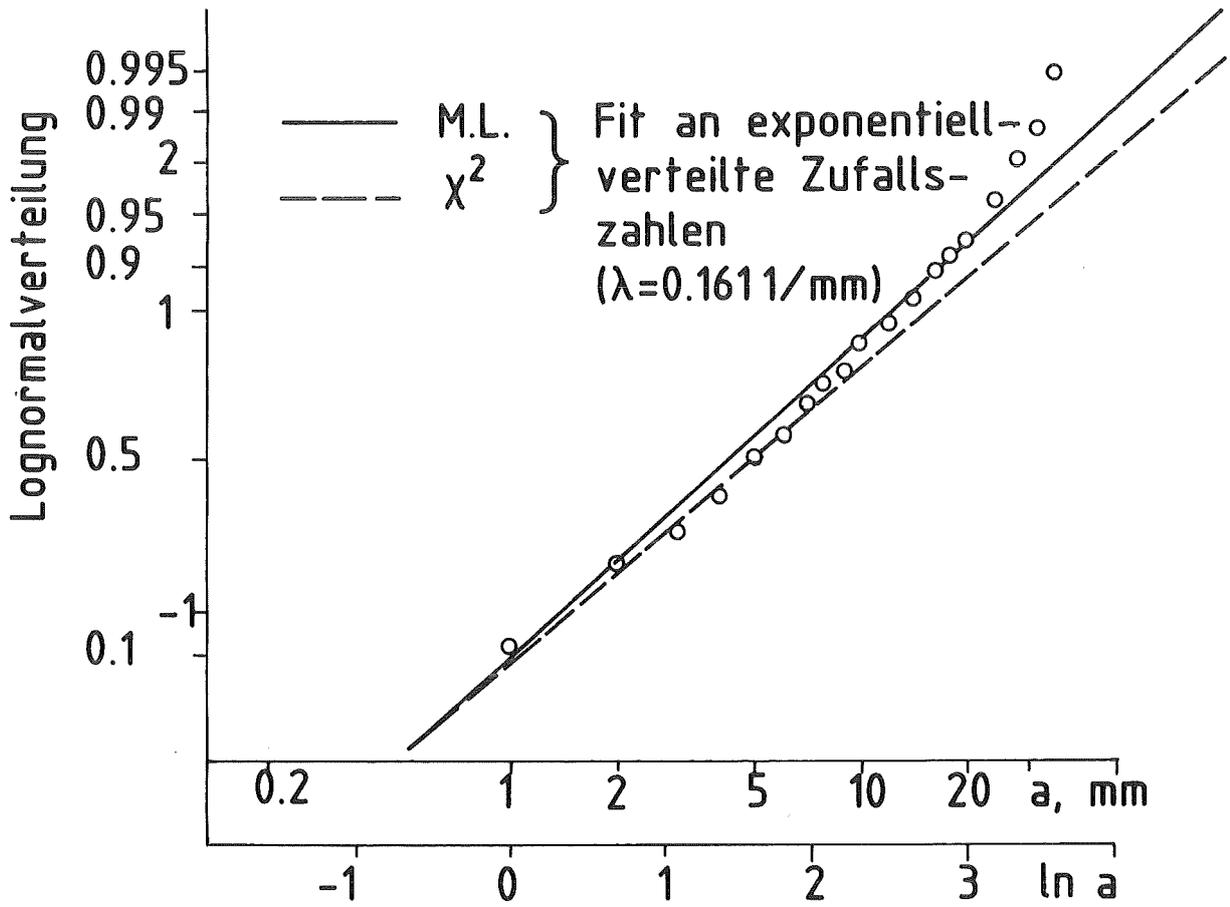
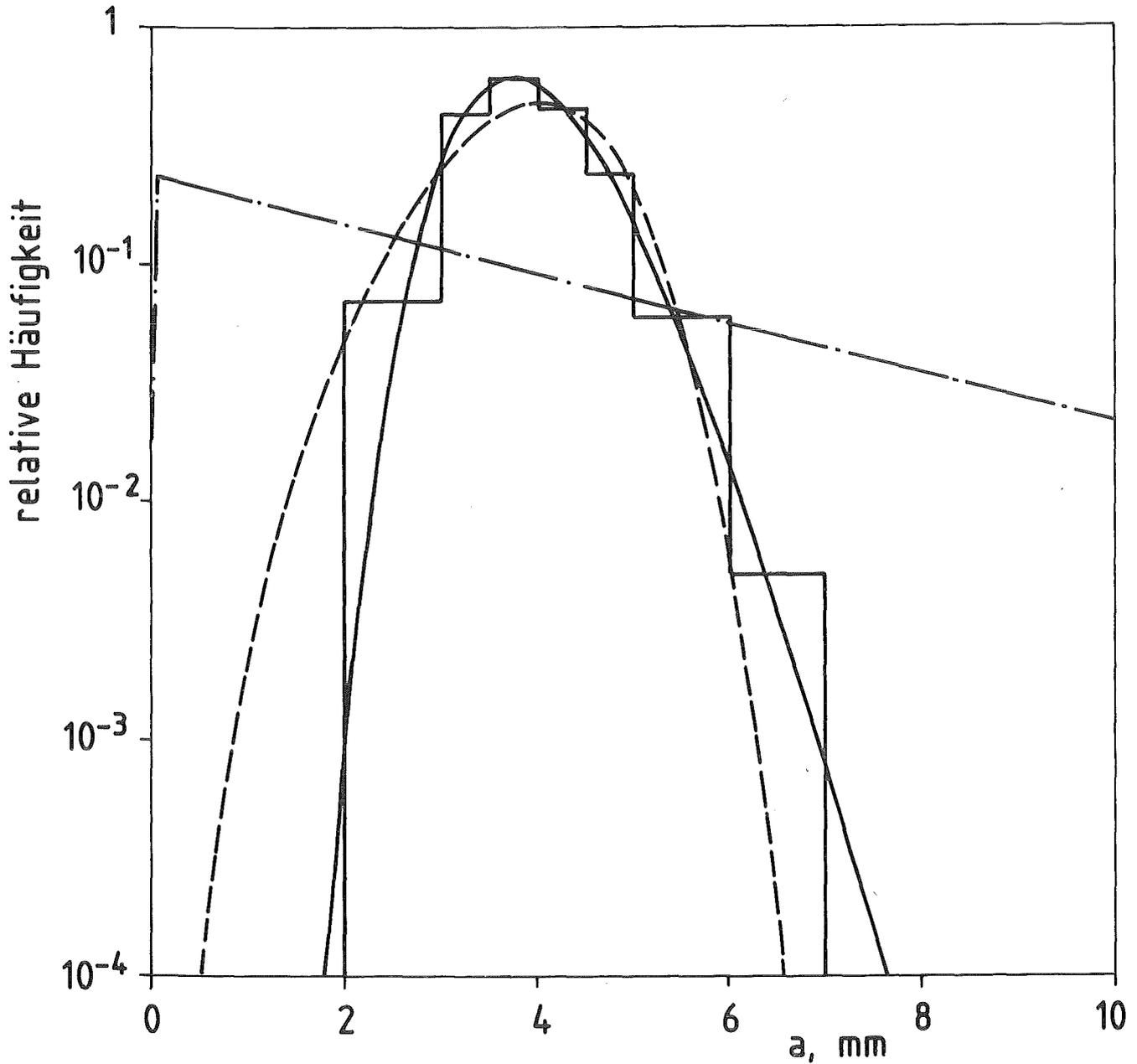


Abb. 7c Wahrscheinlichkeitspapier für Lognormalverteilungen



— Lognormal
 $\mu = 3.856$
 $\sigma = 0.171$

--- Weibull
 $\sigma = 0.238$
 $\eta = 5.419$

-.- Exponential
 $\lambda = 0.239$

Abb. 8 χ^2 -Fit verschiedener Verteilungstypen an lognormalverteilte Zufallszahlen ($\mu=3.855$ mm, $\sigma=0.174$)

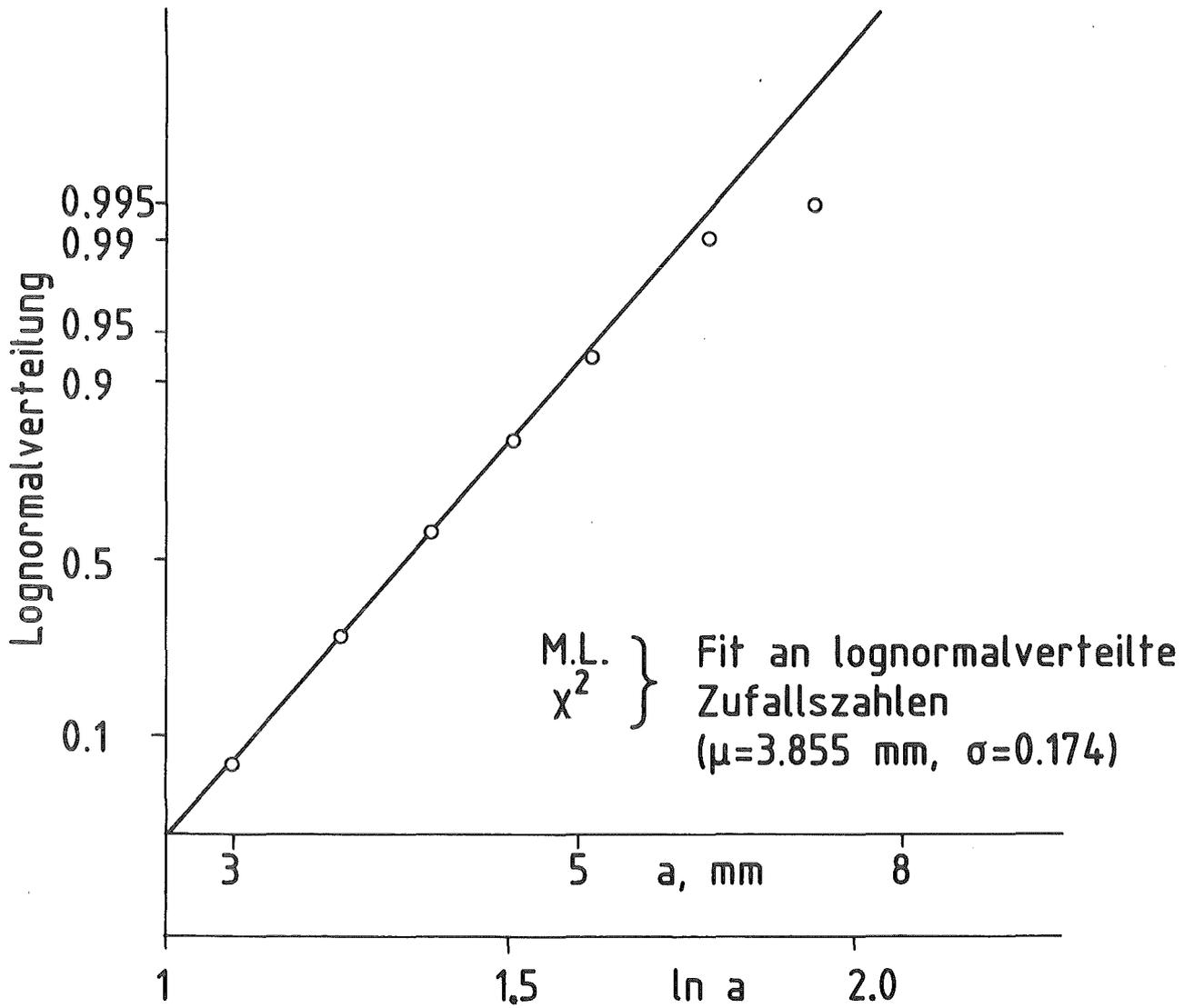


Abb. 9a Wahrscheinlichkeitspapier für Lognormalverteilungen

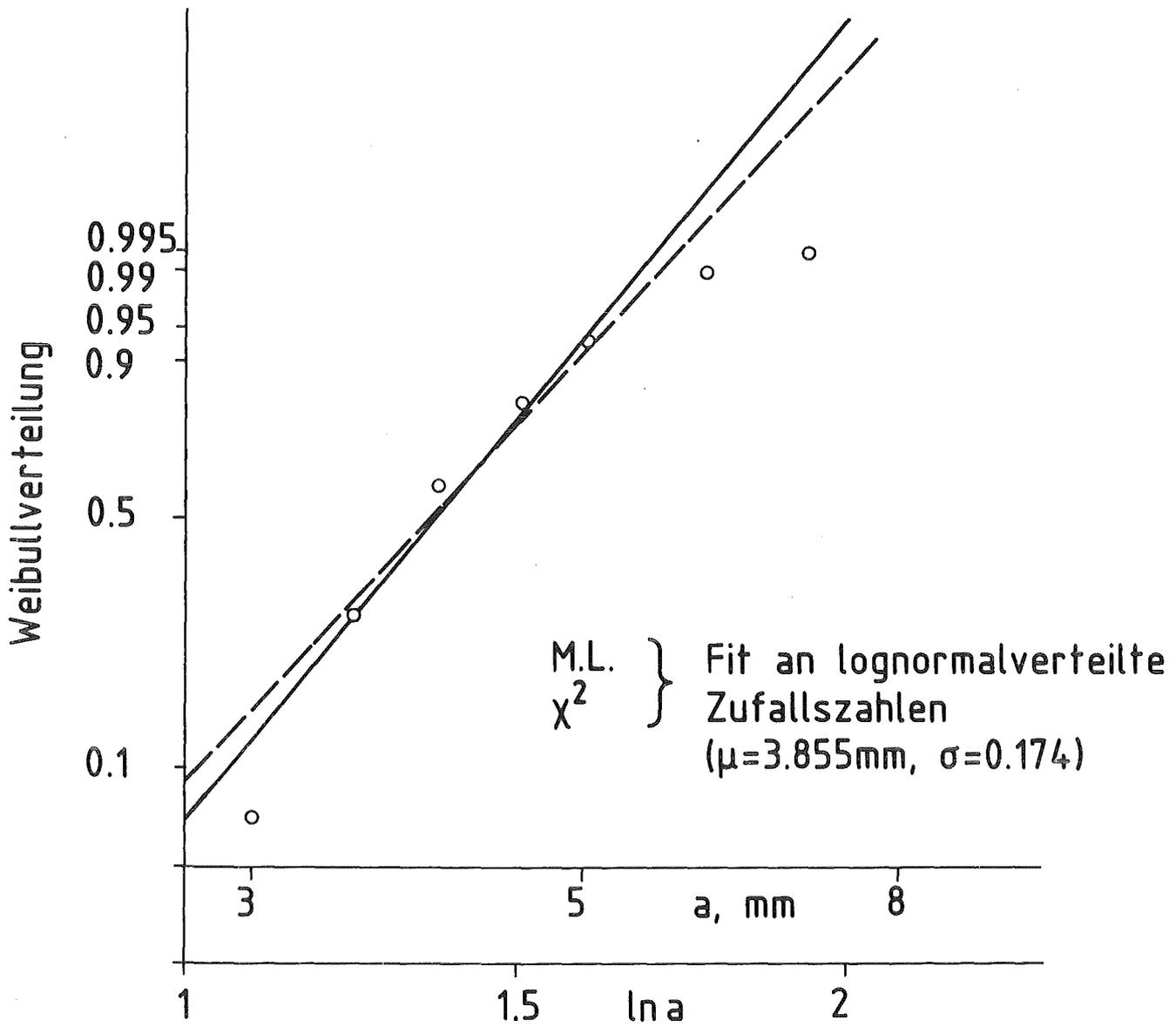


Abb. 9b Wahrscheinlichkeitspapier für Weibullverteilungen

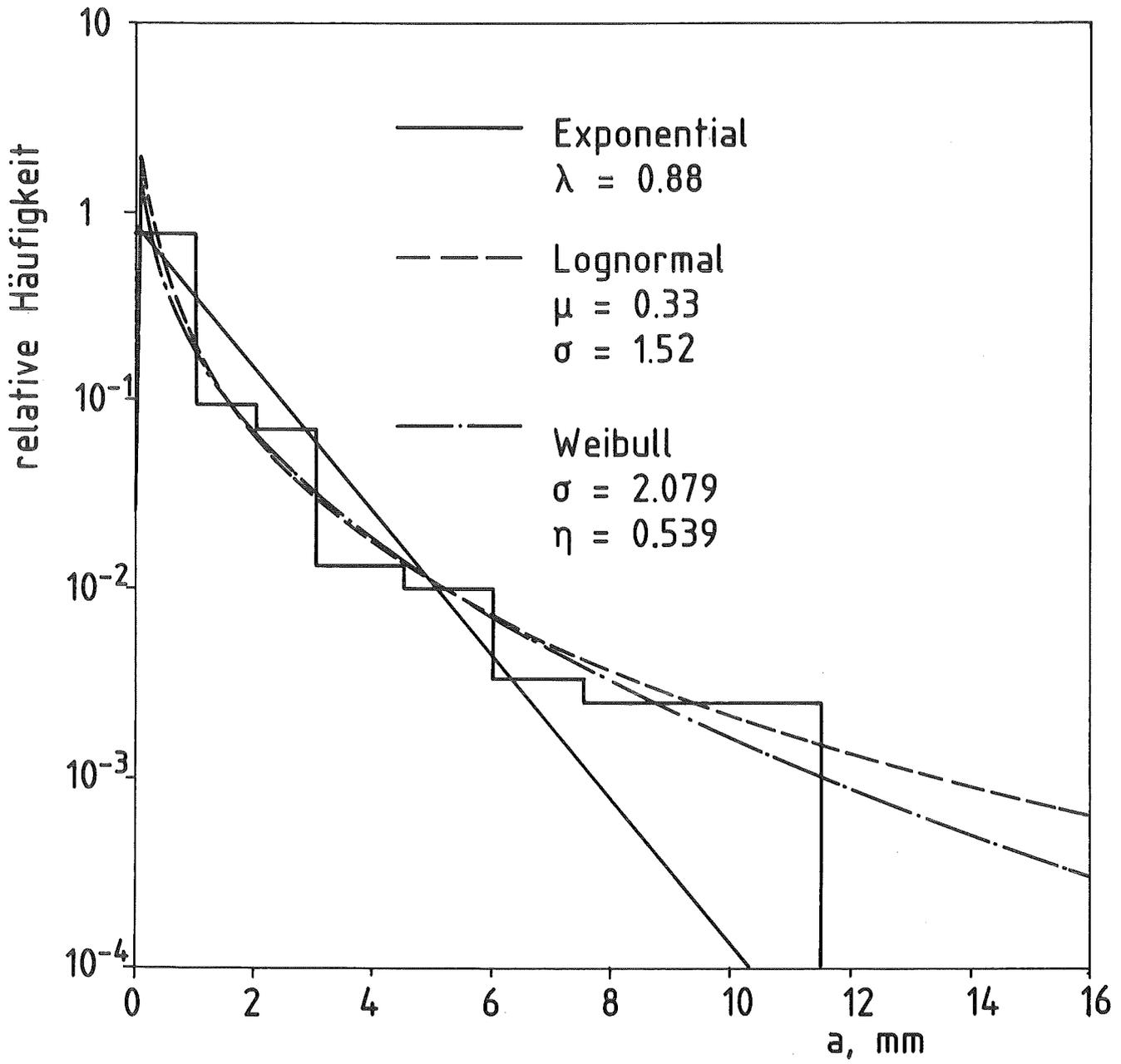


Abb. 10 χ^2 -Fit verschiedener Verteilungstypen an weibullverteilte Zufallszahlen ($\sigma=2.3$ 1/mm, $\eta=0.51$)

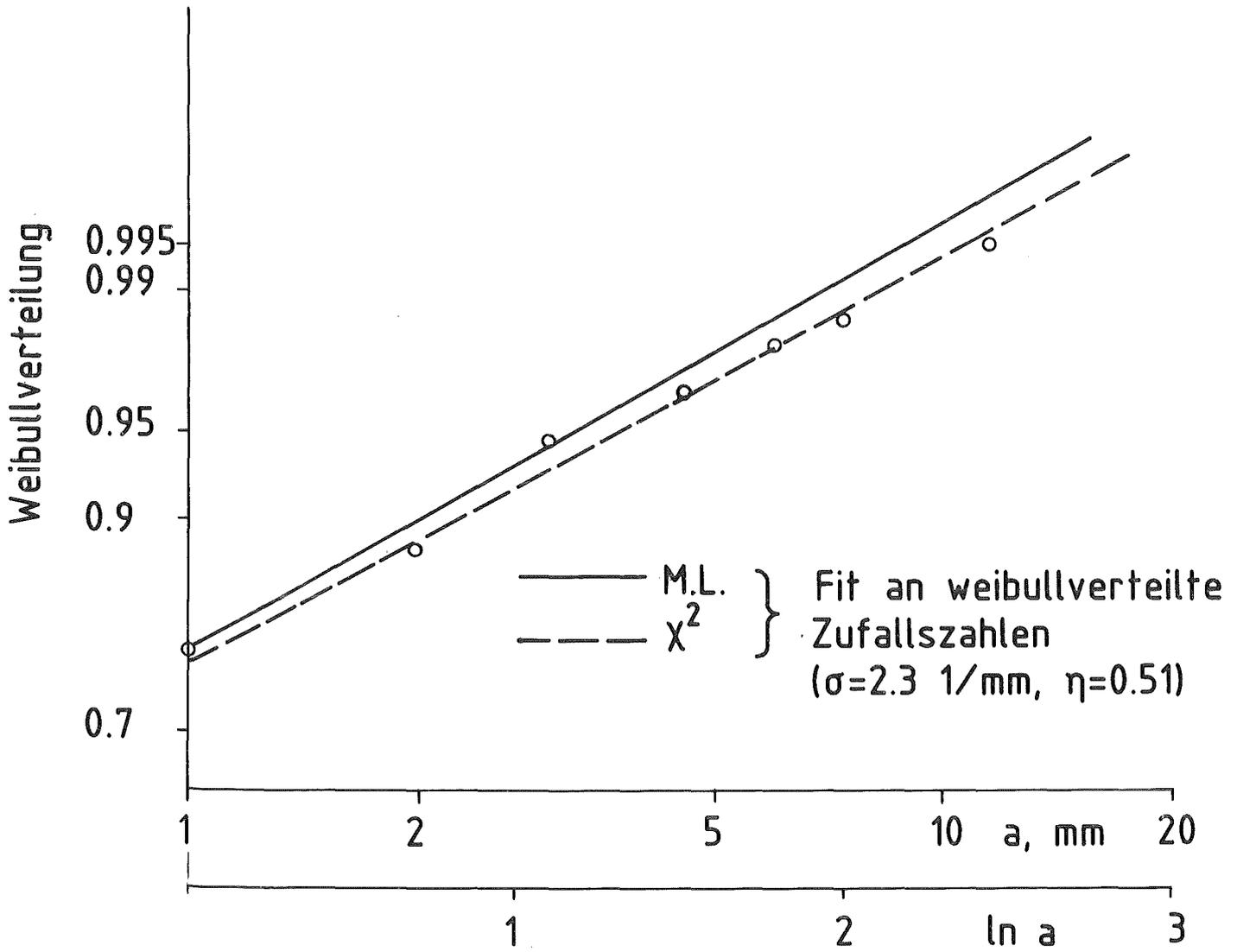


Abb. 11a Wahrscheinlichkeitspapier für Weibullverteilungen

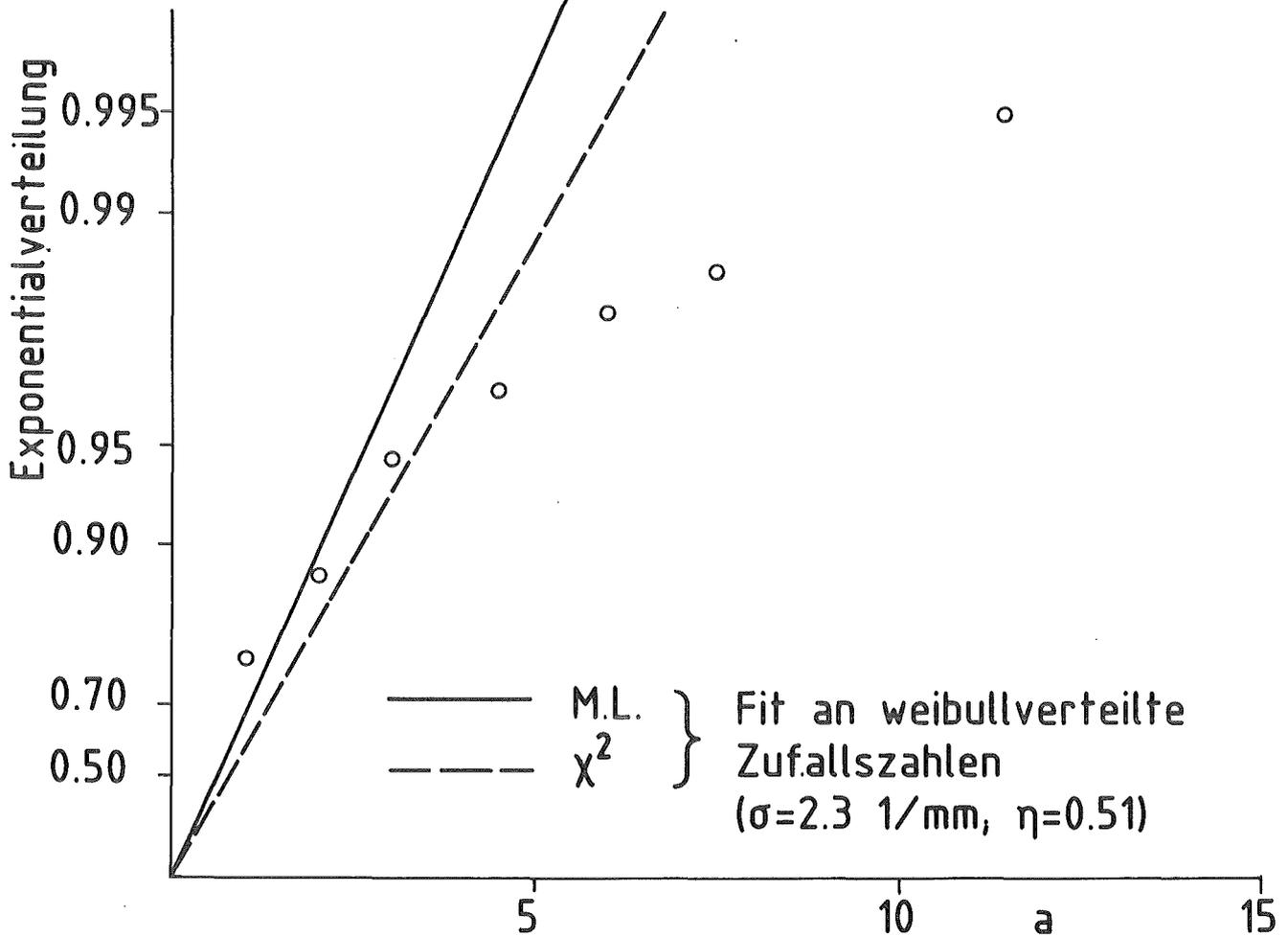


Abb. 11b Wahrscheinlichkeitspapier für Exponentialverteilungen

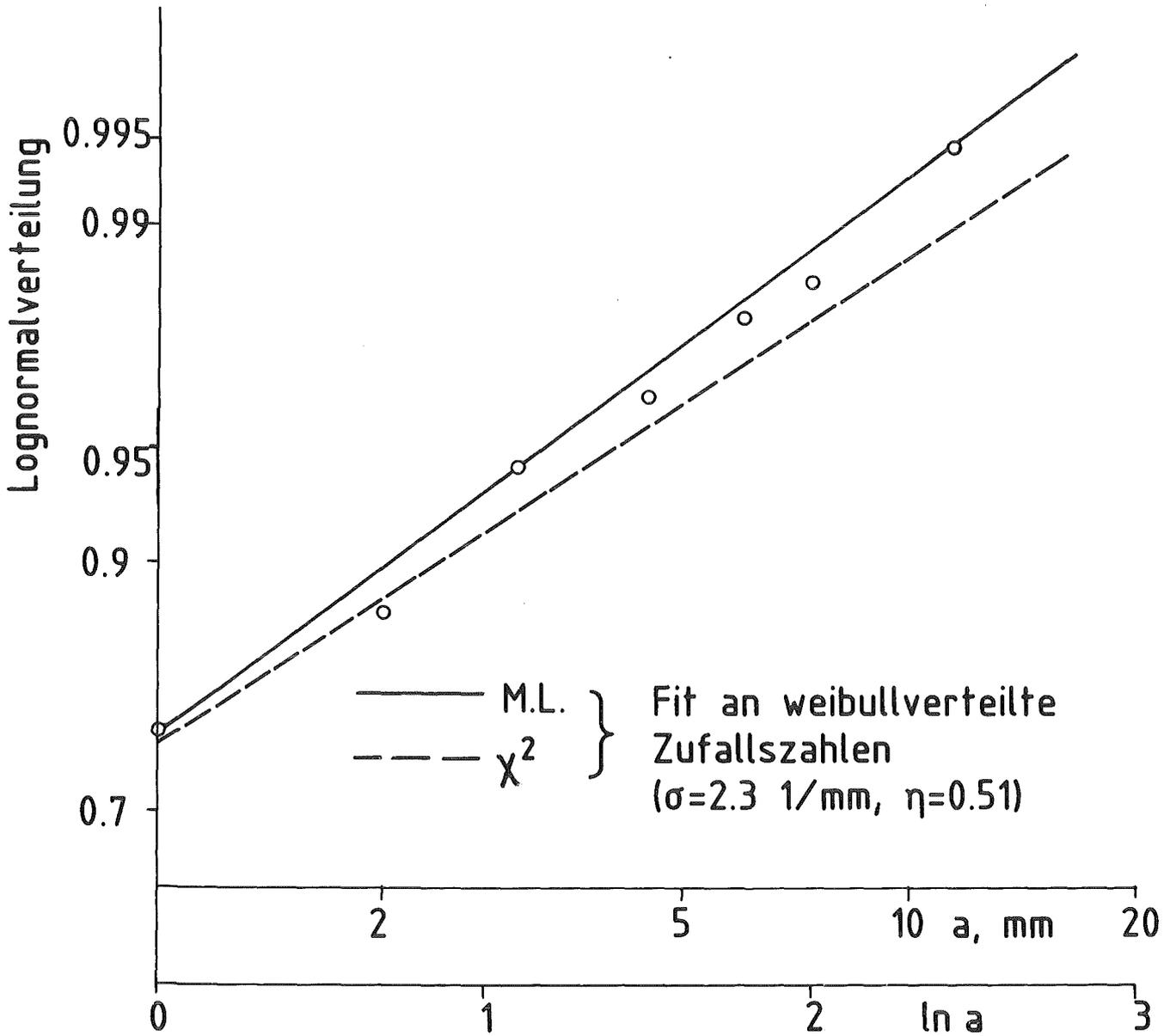


Abb. 11c Wahrscheinlichkeitspapier für Lognormal -
verteilungen

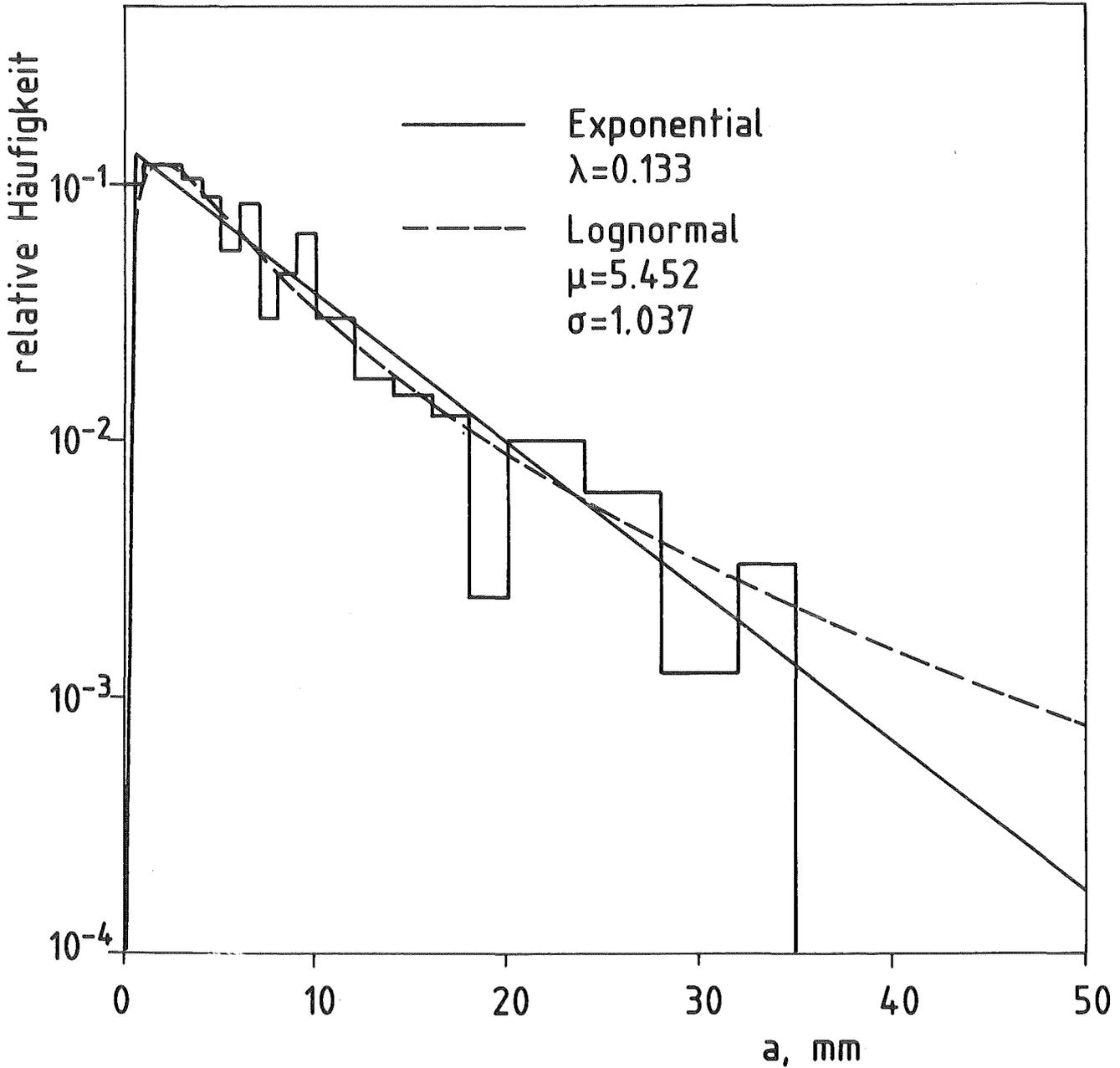


Abb. 12 χ^2 -Fit verschiedener Verteilungstypen an exponentiellverteilte Zufallszahlen ($\lambda=0.161$ 1/mm) mit einer unteren Grenze von 0.5 mm

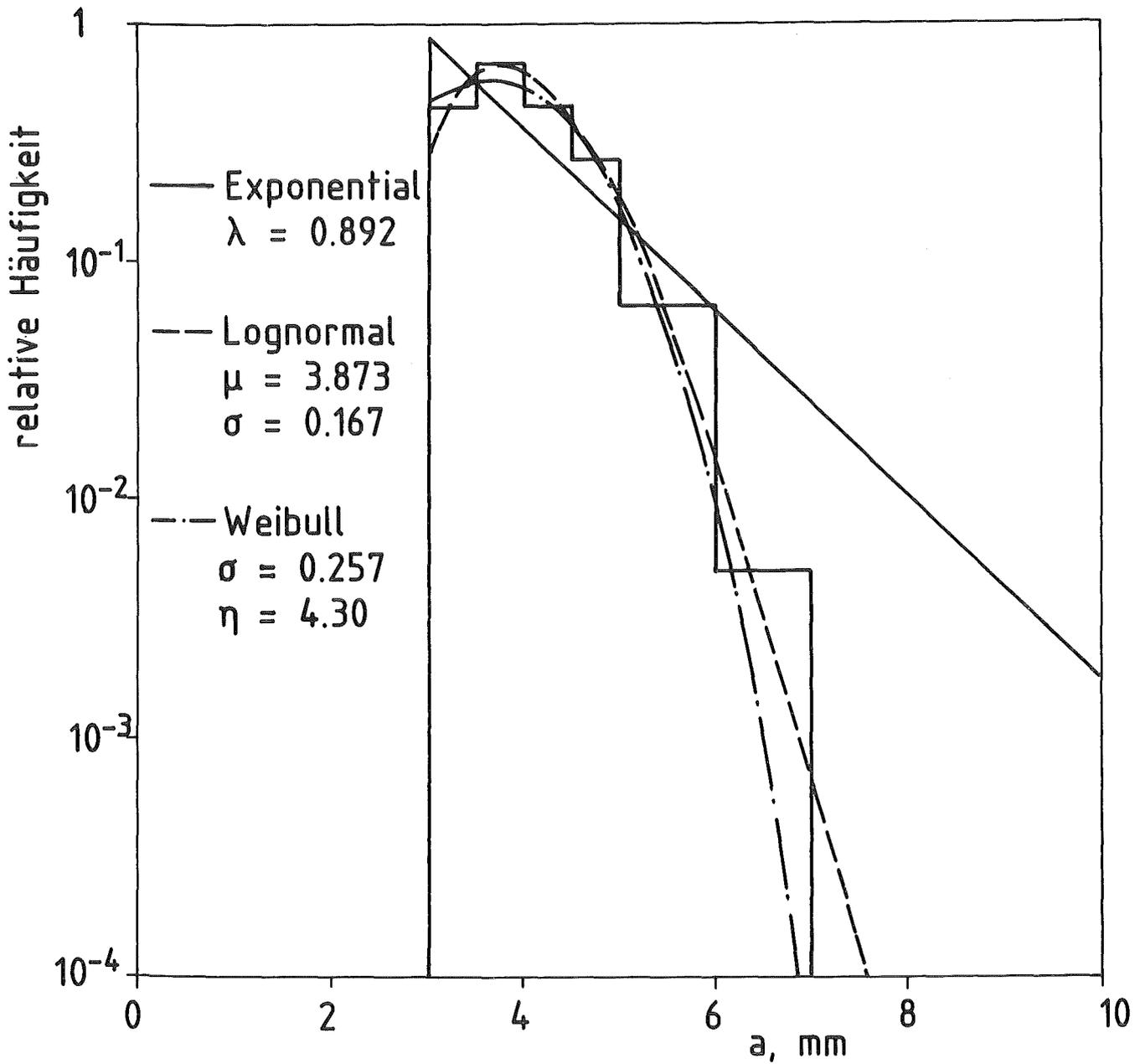


Abb. 13 χ^2 -Fit verschiedener Verteilungstypen an lognormalverteilte Zufallszahlen ($\mu=3.855\text{mm}$, $\sigma=0.174$) mit einer unteren Grenze von 3mm

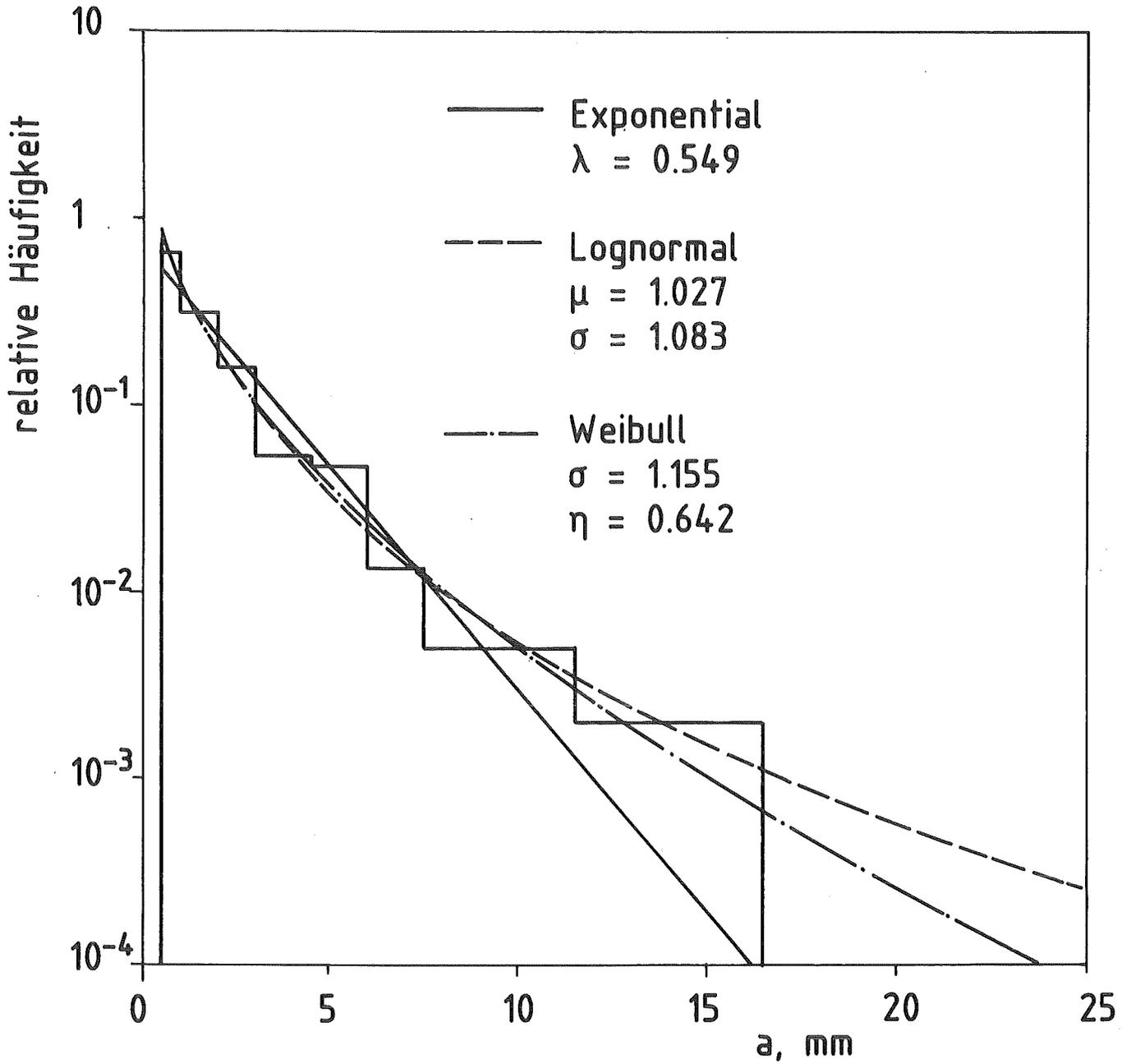
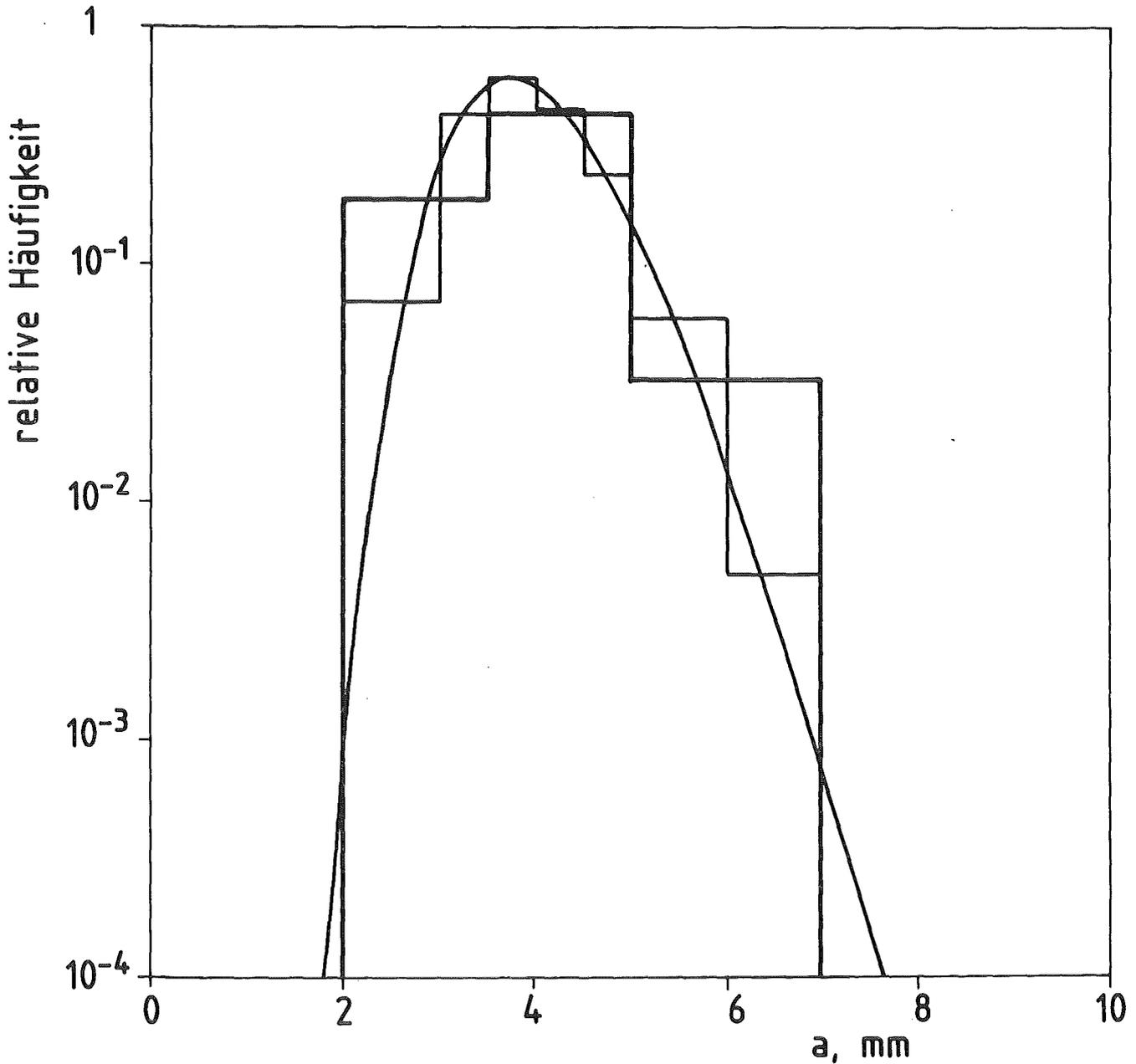


Abb.14 χ^2 -Fit verschiedener Verteilungstypen an weibullverteilte Zufallszahlen ($\sigma=2.3$ 1/mm, $\eta=0.51$) mit einer unteren Grenze von 0.5 mm



feine Klasseneinteilung $\mu = 3.856$
 $\sigma = 0.171$

grobe Klasseneinteilung $\mu = 3.859$
 $\sigma = 0.176$

Abb. 15 χ^2 -Fit einer Lognormalverteilung an lognormalverteilte Zufallszahlen ($\mu=3.855$ mm, $\sigma=0.174$) bei unterschiedlicher Klassifizierung des Histogramms

	Parameter	χ^2 -Test	χ^2_{krit} für $\alpha = 5 \%$
Kleinste Fehlerquadrate	$\sigma = 1.849$ $\eta = 1.066$	292129.2	9.49
χ^2 -Fit	$\sigma = 2.079$ $\eta = 0.539$	3.95	
M.L.-Fit	$\sigma = 2.066$ $\eta = 0.574$	4.40	
Konfidenzintervall	$1.603 < \sigma < 2.664$ $0.518 < \eta < 0.644$		

Tab. 1 Anpassung einer Weibullverteilung an weibullverteilte
Zufallszahlen ($\sigma = 2.3$ 1/mm ; $\eta = 0.51$) mit unterschied-
lichen Methoden der Kurvenanpassung.

	Methode	Parameter	χ^2 -Test	χ^2_{krit} für $\alpha = 5 \%$
Exponential- verteilung	χ^2	$\lambda = 0.134$	16.47	27.59
	M.L.	$\lambda = 0.143$	17.3	
Weibull- verteilung	χ^2	$\eta = 1.005$ $\sigma = 0.134$	16.47	26.3
	M.L.	$\eta = 1.035$ $\sigma = 0.141$	17.58	
Lognormal- verteilung	χ^2	$\mu = 4.902$ $\sigma = 1.185$	17.67	
	M.L.	$\mu = 4.32$ $\sigma = 1.068$	21.59	
Konfidenz- intervall	$0.125 < \lambda < 0.165$			

Tab. 2 Kurvenanpassung an exponentiellverteilte Zufallszahlen
($\lambda = 0.161$ 1/mm)

	Methode	Parameter	χ^2 -Test	χ^2_{krit} für $\alpha = 5 \%$
Exponential- verteilung	χ^2	$\lambda = 0.239$	637.1	11.07
	M.L.	$\lambda = 0.256$	639.1	
Weibull- verteilung	χ^2	$\sigma = 0.238$ $\eta = 5.419$	24.0	9.49
	M.L.	$\sigma = 0.239$ $\eta = 6.066$	48.2	
Lognormal- verteilung	χ^2	$\mu = 3.856$ $\sigma = 0.171$	0.087	
	M.L.	$\mu = 3.856$ $\sigma = 0.171$	0.089	
Konfidenz- intervall	$3.722 < \mu < 3.928$ $0.176 < \sigma < 0.214$			

Tab. 3 Kurvenanpassung an lognormalverteilte Zufallszahlen
($\mu=3.855$ mm; $\sigma= 0.174$)

	Methode	Parameter	χ^2 -Test	χ^2_{krit} für $\alpha = 5 \%$
Exponential- verteilung	χ^2	$\lambda = 0.88$	49.0	11.07
	M.L.	$\lambda = 1.13$	118.0	
Weibull- verteilung	χ^2	$\sigma = 2.079$ $\eta = 0.539$	3.95	9.49
	M.L.	$\sigma = 2.066$ $\eta = 0.574$	4.40	
Lognormal- verteilung	χ^2	$\mu = 0.33$ $\sigma = 1.52$	4.61	
	M.L.	$\mu = 0.35$ $\sigma = 1.348$	5.47	
Konfidenz- intervall	$1.603 < \sigma < 2.664$ $0.518 < \eta < 0.644$			

Tab. 4 Kurvenanpassung an weibullverteilte Zufallszahlen
($\sigma = 2.3$ 1/mm; $\eta = 0.51$)

	Methode	Parameter	χ^2 -Test	χ^2_{krit} für $\alpha = 5 \%$
Exponential- verteilung	χ^2	$\lambda = 0.133$	16.88	27.59
	M.L.	$\lambda = 0.144$	18.08	
Weibull- verteilung	χ^2	$\hat{\eta} = 1.0$ $\sigma = 0.133$	16.88	26.3
	M.L.	$\eta = 1.05$ $\sigma = 0.14$	18.51	
Lognormal- verteilung	χ^2	$\mu = 5.452$ $\sigma = 1.037$	14.97	
	M.L.	$\mu = 4.96$ $\sigma = 0.941$	18.55	

Tab. 5 Kurvenanpassung an exponentiellverteilte Zufallszahlen
($\lambda = 0.161$ 1/mm) mit einer unteren Grenze von 0.5 mm.

	Methode	Parameter	χ^2 -Test	χ^2_{krit} für $\alpha = 5 \%$
Lognormal- verteilung	χ^2	$\mu = 3.873$ $\sigma = 0.167$	1.96	7,81
	M.L.	$\mu = 3.854$ $\sigma = 0.171$	2.09	
Weibull- verteilung	χ^2	$\eta = 4.3$ $\sigma = 0.257$	4.58	
	M.L.	$\eta = 4.43$ $\sigma = 0.255$	4.70	
Exponential- verteilung	χ^2	$\lambda = 0.892$	41.75	9.49
	M.L.	$\lambda = 1.013$	43.71	

Tab. 6 Kurvenanpassung an lognormalverteilte Zufallszahlen
($\mu = 3.855$ mm; $\sigma = 0.174$) mit einer unteren Grenze
von 3 mm

	Methode	Parameter	χ^2 -Test	χ^2_{krit} für $\alpha = 5 \%$	
Weibull- verteilung	χ^2	$\eta = 0.642$ $\sigma = 1.155$	3.88	11.07	
	M.L.	$\eta = 0.672$ $\sigma = 1.098$	4.10		
Lognormal- verteilung	χ^2	$\mu = 1.027$ $\sigma = 1.083$	4.28		
	M.L.	$\mu = 1.081$ $\sigma = 1.006$	4.81		
Exponential- verteilung	χ^2	$\lambda = 0.549$	18.96		12.59
	M.L.	$\lambda = 0.598$	21.97		

Tab. 7 Kurvenanpassung an weibullverteilte Zufallszahlen
($\sigma = 2.3$ 1/mm; $\eta = 0.51$) mit einer unteren Grenze
von 0.5 mm

Stichprobe	Parameter	χ^2 -Test	χ^2_{krit} für $\alpha = 5\%$
1.	$\sigma = 2.066$ $\eta = 0.574$	4.40	9.49
2.	$\sigma = 1.89$ $\eta = 0.519$	7.17	
3.	$\sigma = 2.068$ $\eta = 0.49$	10.82	11.07
4.	$\sigma = 1.938$ $\eta = 0.538$	4.89	9.49
5.	$\sigma = 1.8$ $\eta = 0.631$	4.43	7.81
Konfidenz- intervall für 1. Stichprobe	$0.518 < \eta < 0.644$ $1.603 < \sigma < 2.664$		

Tab. 8 Einfluß der Stichprobe auf die simulierte Fehlerverteilung am Beispiel weibullverteilter Zufallszahlen ($\sigma = 2.31/\text{mm}$; $\eta = 0.51$)