The Use of the Weibull Distribution for Thunderstorm Parameters

T. SCHÜTTE, O. SALKA AND S. ISRAELSSON

Institute for High Voltage Research, Husbyborg, S-755 90 Uppsala, Sweden

(Manuscript received 11 January 1986, in final form 1 November 1986)

ABSTRACT

Most observations in thunderstorm statistics are positive, having values between 0 and ∞. Since they are of a "duration-length" type, the normal distribution is not, in most cases, a satisfactory approximation. A widely used model is the lognormal distribution. It has the advantage of easy calculation of the distribution parameters, but, as most statistical distributions, it has the disadvantage of a nonanalytic cumulative distribution function.

In the present paper, we describe how the Weibull distribution can be used to estimate some characteristics related to thunderstorms. Examples of the calculation of some further properties assuming Weibull-distributed observations are also given.

From a theoretical point of view, the use of the Weibull distribution seems to be reasonable. It is a good model for describing electrical breakdown in insulating materials, and a lightning discharge is, indeed, such a breakdown.

1. Introduction

In thunderstorm statistics, many parameters, such as the radiation field strengths, are positive and markedly skewed, and therefore a normal distribution cannot be applied in most cases. The most suitable distributions for these kinds of data are the lognormal, the gamma and the Weibull distributions. An advantage of the lognormal distribution is the easy calculation of the distribution parameters, viz. the mean and the variance of the natural logarithms of the values. A disadvantage both of the gamma and the lognormal distribution is the fact that their cumulative distributions are nonanalytic. This means that one needs numerical calculations for the determination of the cumulative distribution. The cumulative Weibull distribution is a fairly simple expression, and even integrals over the density function times another function can often be solved analytically.

The Weibull distribution model has been tested on different kinds of thunderstorm and lightning data, and the fit is satisfactory in most cases. In order to test this, first a Weibull paper and later a computer program for calculation of the distribution parameters were used. This program has also been adopted for electrical breakdown in transformer oil and used successfully for the description of this phenomenon.

As a result of a Weibull distribution for the signal strength of the lightning, analytical expressions for acceptance and effective radius of lightning counters or direction finders and the damping of lightning signals can be derived.

2. Weibull distribution

The probability density function of the Weibull distribution can be written [see Weibull (1939); Hahn and Shapiro (1967)] as

$$f(x) = \begin{cases} 0 & \text{for} \quad x \leqslant c \\ \dfrac{b}{a}\left(\dfrac{x-c}{a}\right)^{b-1} \exp\left[-\left(\dfrac{x-c}{a}\right)^{b}\right] & \text{for} \quad x > c \end{cases}$$

and the cumulative distribution is given by

$$F(x) = \begin{cases} 0 & \text{for} \quad x \leqslant c \\ 1 - \exp\left[-\left(\dfrac{x-c}{a}\right)^{b}\right] & \text{for} \quad x > c \end{cases}$$

where a, b and c are the scale, the shape and the location parameter, respectively. In many practical applications $c = 0$, but sometimes this parameter becomes very important as a lowest possible value.

For $b \leqslant 3.6$ the distribution is skewed to the left, and for $b > 3.6$ to the right. Some special cases are represented by $b = 1$ (exponential distribution) and $b = 2$ (Rayleigh distribution); see Figs. 1 and 2. The Weibull distribution can be considered as a generalization of the exponential distribution.

The hazard function $\dot{n}/n = f(t)/(1 - F)$ is a constant for the exponential distribution (for example, the radioactive decay). Here $\dot{n} = dn/dt$. On the other hand, when the hazard function is a power law (e.g., $\dot{n}/n = (b/a)[(t - c)/a]^{b-1}$ for $x > c$; $\dot{n}/n = 0$ elsewhere), the

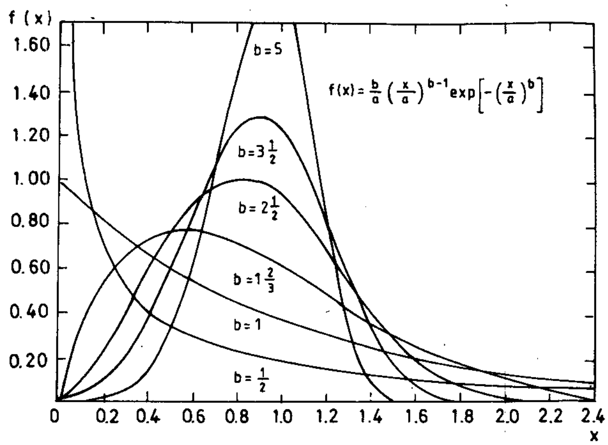

FIG. 1. Weibull distribution for different shape parameters b, $c = 0$; scale parameter $a = 1$.

resulting distribution will be a Weibull distribution with parameters a, b and c (Hahn and Shapiro, 1967).

A special feature of the Weibull distribution is if $x - c$ is a Weibull-distributed with the parameters a and b, then $y = (x - c)^q$ will also be a Weibull-distributed with the new parameters $\tilde{a} = a^q$, $b = \tilde{b}/q$, viz.,

$$F(x) = 1 - \exp\left[-\left(\frac{x-c}{a}\right)^b\right] = 1 - \exp\left[-\left(\frac{y^{1/q}}{a}\right)^b\right]$$
$$= 1 - \exp\left[-\left(\frac{y}{a^q}\right)^{b/q}\right] = 1 - \exp\left[-\left(\frac{y}{\tilde{a}}\right)^{\tilde{b}}\right].$$

From the parameters a, b and c, the following can be calculated:

Expected value: $\bar{x} = a\Gamma(1 + 1/b) + c$

Modal value: $$\hat{x} = \begin{cases} a\left(\dfrac{b-1}{b}\right)^{1/b} + c & \text{for} \quad b > 1 \\ c & \text{for} \quad b \leqslant 1 \end{cases}$$

Median value: $x_{1/2} = a(\ln 2)^{1/b} + c$

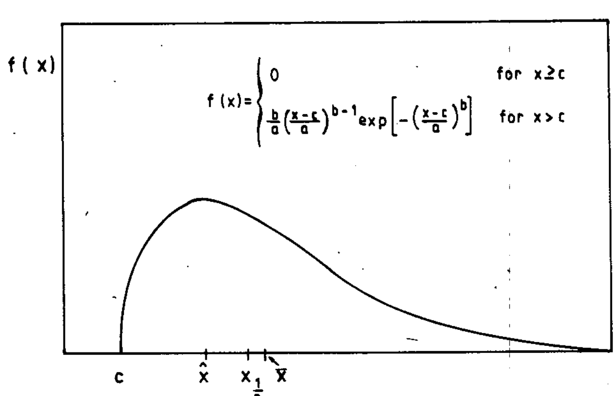

FIG. 2. Weibull distribution for $c \neq 0$, $b = 1.58$. Modal, median and mean value marked.

Variance: $\sigma^2 = a^2[\Gamma(1 + 2/b) - \Gamma^2(1 + 1/b)]$

Skewness: $\beta_1{}^{1/2}$

$$= \frac{\Gamma(1 + 3/b) - 3\Gamma(1 + 2/b)\Gamma[1 + 1(b)] + \Gamma^3(1 + 1/b)}{\Gamma(1 + 2/b) - \Gamma^2(1 + 1/b)^{3/2}}.$$

Kurtosis:

$$\beta_2 = \frac{\Gamma(1 + 4/b) - 4\Gamma(1 + 3/b)\Gamma(1 + 1/b) + 6\Gamma(1 + 2/b)\Gamma^2(1 + 1/b) - 3\Gamma^4(1 + 1/b)}{[\Gamma(1 + 2/b) - \Gamma^2(1 + 1/b)]^2}.$$

The (complete) Gamma function is defined as

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt.$$

Using the hazard function, we get expressions for the distribution of the sum of identical processes, which are Weibull-distributed in time.

For N simultaneous two-parameter Weibull processes, the hazard function is given by

$$\frac{\dot{n}}{n} = \frac{Nb}{a}(t/a)^{b-1}.$$

This means that the sum is also Weibull-distributed with the parameters

$$\tilde{a} = N^{-(1/\tilde{b})}a, \quad \tilde{b} = b.$$

For N simultaneous three-parameter Weibull processes with the same parameters a, b and c, the hazard function is approximately

$$\frac{\dot{n}}{n} = \frac{Nb(\bar{t} - c)}{a\bar{t}}(t/a)^{b-1}, \quad \bar{t} = a\Gamma(1 + 1/b) + c.$$

The sum is approximately two-parameter distributed with

$$\tilde{a} = \left[\frac{a^{\tilde{b}}\Gamma(1 + 1/\tilde{b}) + ca^{\tilde{b}-1}}{N\Gamma(1 + 1/\tilde{b})}\right]^{1/\tilde{b}}, \quad \tilde{b} = b.$$

3. The use of Weibull distribution paper

We will first test the Weibull distribution on some thunderstorm data by using special diagram paper. It is possible to design graph paper where the cumulative (two-parameter) Weibull distribution

$$F(x) = 1 - \exp(-(x/a)^b)$$

is transformed into a straight line (Weibull, 1939; Berettoni, 1964):

$$\ln \ln[1/(1 - F(x)] = b \ln x - b \ln a.$$

This means that we have to plot $\ln x$ against $\ln \ln[1/(1 - F)]$. From the plot we can deduce graphical estimates for a and b (Fig. 3).

As a first test, some thunderstorm statistics from Florida were plotted on Weibull paper and the distri-

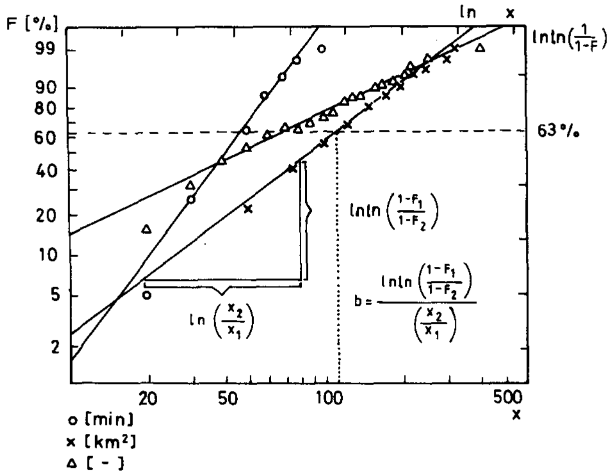

FIG. 3. Weibull plots of some thunderstorm data from Peckham et al. (1984): thunderstorm duration, area and number of lightning strokes.

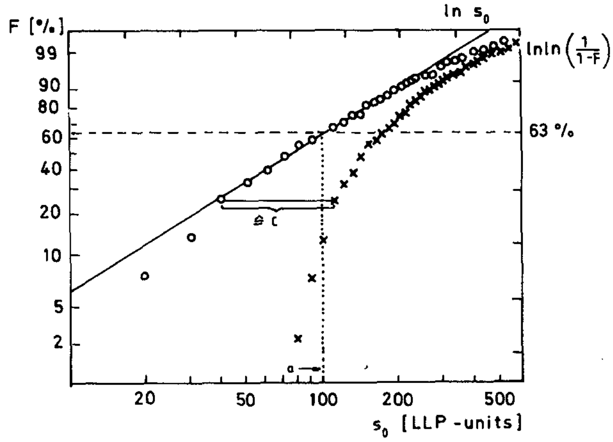

FIG. 5. Weibull plot for three parameter Weibull distribution: normalized (to 100 km) signal strength distribution of negative return strokes over sea, measured by the lightning location system (Pisler, 1984).

bution function, based on the graphic parameter estimates, was compared with the data. The results are satisfactory (Figs. 3 and 4).

The graphical parameters estimate for the three-parameter distribution is more complicated. It will no longer be a straight line. An estimated c has to be subtracted and altered gradually by trial and error until the data points form an approximately straight line (Weibull, 1939; Berettoni, 1964). This was performed for the signal-strength distributions in order to illustrate the procedure (Fig. 5). All signal-strength distributions in this paper were normalized to 100 km distance by the $1/r$ law measured by direction-finding stations of the lightning localization network in Sweden. Here they are presented as arbitrary units, which are proportional to the peak radiation field of the first return stroke in cloud-to-ground lightning flashes.

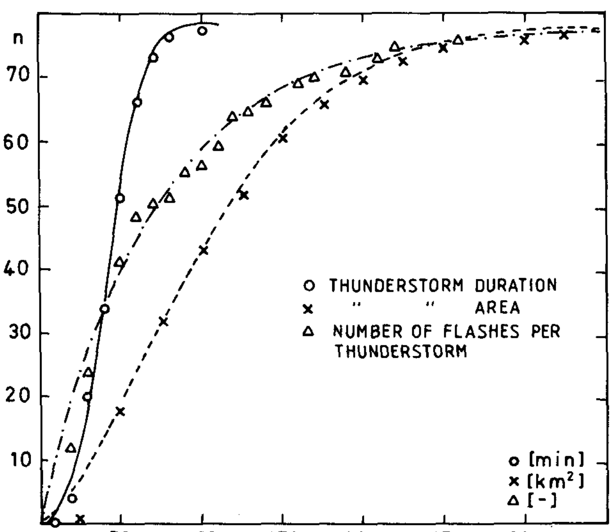

FIG. 4. Cumulative distributions according to Fig. 3.

4. Numerical calculation of the parameters for a three-parameter Weibull distribution

The use of the Weibull paper has many disadvantages. For instance, the strong nonlinearity and, in contrast to the normal distribution, the asymmetric character of this paper strongly exaggerate the region of the low percentiles. The determination of the location parameter is also difficult. This method is too time consuming for a large number of data values. There is therefore a need for a numerical method.

A least-squares fit $\sum_{i=1}^{n} [f(x_i) - y_i]^2 \rightarrow \min$ turns out to be insufficient in this case. A generally good method is the maximum likelihood method successfully used by Lundtang Petersen et al. (1981) for a two-parameter distribution.

The inclusion of the location parameter c gives some complications, however. The set of equations becomes unstable quite often. The reason for this seems to be the different characters of the variables a and b on one side, occurring as factors and powers, and c on the other side, which is added to the arguments. Different ways to circumvent this problem have been used elsewhere (Steward and Essenwanger, 1978). We used a combination of the maximum likelihood for a and b and a least-squares fit for the inverse function $\sum_{i=1}^{n} [x_i - x(F_i)]^2$ for c. This method was stable using proper first guesses derived from calculations analogous to the evaluation using Weibull paper. It was successfully used for the description of the distribution of the breakdown

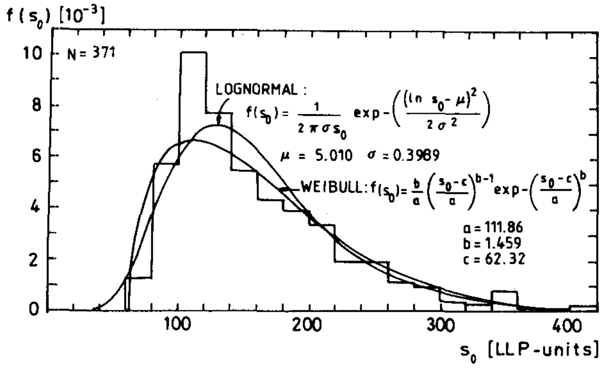

FIG. 6. Histogram of the normalized signal strength distribution of negative return strokes over sea, measured by the LLP system (Pisler, 1984). Calculated lognormal and Weibull distribution inserted.

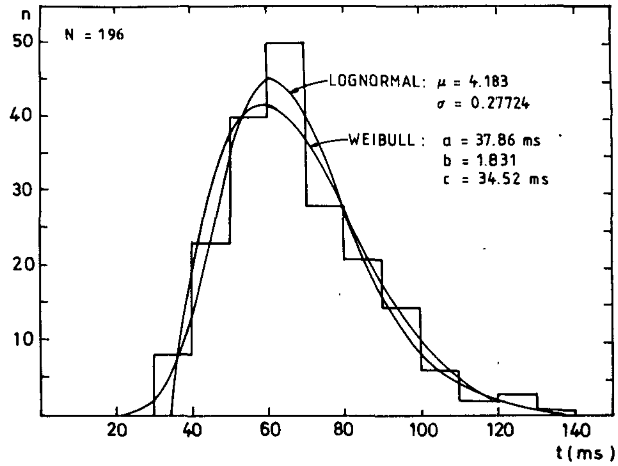

FIG. 8. Histogram of the time difference between first and second return strokes in Sri Lanka (Cooray, personal communication). Weibull and lognormal analysis performed.

voltages in transformer oil, and it will be used for the thunderstorm statistics.

5. Applying the Weibull distribution to data on lightning signal strengths and times between first and second return strokes

Weibull analysis was performed on two sets of signal-strength data of lightning with propagation over sea and land (Pisler, 1984; Cooray and Lundquist, 1982), and the time calculated between first- and second-return strokes of the lightnings. (Cooray, personal communication). Lognormal analysis was performed as a comparison. In all cases both types of distributions fit the data well (Figs. 6–8). An interesting difference occurs if we put together the signal strengths from the distributions over sea with those over land. Because the first set consists of undamped, and the second of damped signals, we have done something which is not allowed; namely, joining two statistically different samples. This error cannot be detected by using the lognormal distribution because the distribution parameters lie between those of the samples treated separately. By using the Weibull distribution, the shape parameter b can be seen to increase to values greater than those of both samples and much more than the difference between them. This indicates that something has happened (Fig. 9).

The distribution of the time duration between the first and second return stroke is fitted well both by the Weibull and the lognormal distributions. If both the lognormal and the Weibull distributions are describing a distribution quite well, the Weibull distribution is to be preferred, because it offers a simpler mathematical treatment. The location parameter c can be considered as a threshold or lowest possible value for a physical parameter, which is quite reasonable in many cases.

6. Some examples of the use of the simple structure of the Weibull distribution

a. The acceptance of a lightning counter or a direction finder

Consider a normalized (for example, to the distance of 100 km), undamped reference signal-strength dis-

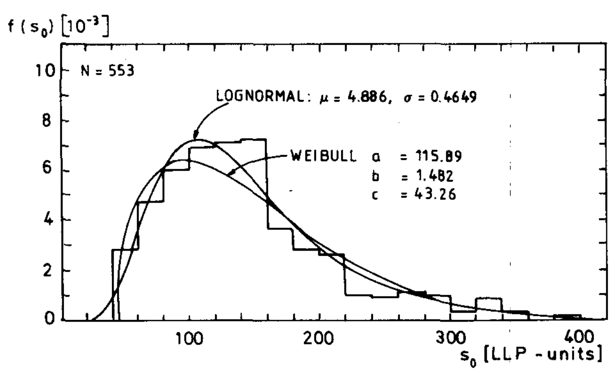

FIG. 7. As in Fig. 6 except for negative return strokes over land (Cooray and Lundquist, 1982).

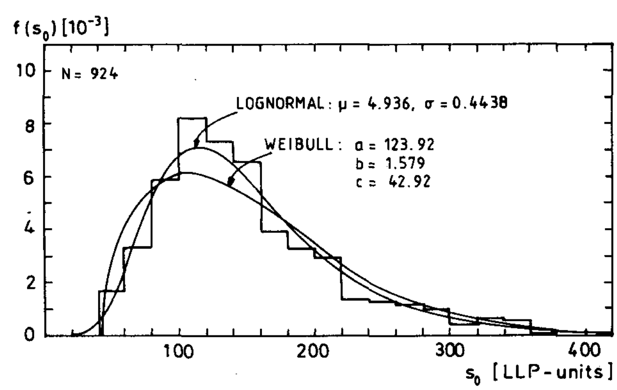

FIG. 9. Histograms from Figs. 6 and 7 added; Weibull and lognormal analysis performed.

tribution with the frequency function $f(s_0)$, and an expression for the signal strength dependence on the distance $s = D(s_0)[s_0 = D^{-1}(s)]$. The simplest case for undamped radiation fields is the inverse power law $s = s_0 r_0/r$, where $s_0 = sr/r_0$. The acceptance can be calculated as a function of distance, lower and higher threshold $s_{\min}$ and $s_{\max}$ according to

$$A(r) = \int_{s_{0_{\min}}}^{s_{0_{\max}}} f(s_0)ds_0, \quad s_0 = D^{-1}(s), \quad ds_0 = D^{-1'}(s)ds$$

$$= \int_{s_{\min}}^{s_{\max}} f[D^{-1}(s)]D^{-1'} \quad ds = F[D^{-1}(s)]|_{s_{\min}}^{s_{\max}}$$

For undamped propagation we obtain

$$A(r) = F(sr/r_0)|_{s_{\min}}^{s_{\max}}.$$

In the case of the Weibull distribution, the acceptance is given by

$$A(r) = 0 \quad \text{for} \quad r \leqslant cr_0/s_{\max}$$

$$A(r) = 1 - \exp\left[-\left(\frac{s_{\max} r/r_0 - c}{a}\right)^b\right] \quad \text{for} \quad cr_0/s_{\max} < r \leqslant cr_0/s_{\min}$$

$$A(r) = \exp\left[-\left(\frac{s_{\min} r/r_0 - c}{a}\right)^b\right] - \exp\left[\left(\frac{s_{\max} r/r_0 - c}{a}\right)^b\right] \quad \text{for} \quad r > cr_0/s_{\min}$$

This gives acceptance versus distance according to Fig. 10 for our reference distribution. The use of different damping models results in different functions for the acceptance level (Schütte et al., 1986).

Another parameter describing the properties of a lightning counter or direction finder is the effective radius ρ defined as

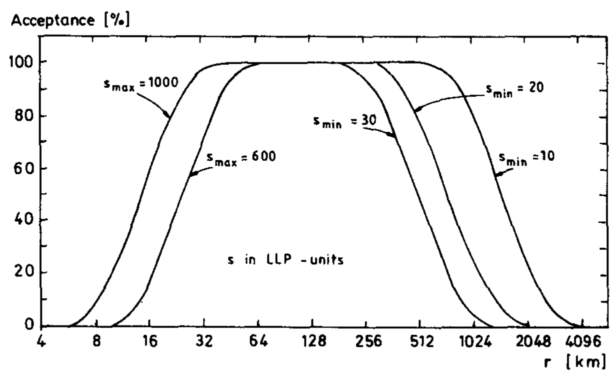

FIG. 10. Acceptance of an ideal lightning counter/direction finder for different low and high threshold values as a function of distance using the signal distribution from Pisler (1984).

$$\pi\rho^2 = 2\pi \int_0^\infty A(r)rdr.$$

That means that an ideal lightning counter, which accepts all lightnings with $r \leqslant \rho$ and none with $r > \rho$, would count the same number of lightnings assuming a homogeneous lightning distribution in space.

When the signal strength is Weibull-distributed and the counter has a sufficient dynamic range ($s_{\min} \ll s_{\max}$), this can be calculated analytically, because some expressions become nearly 0 and the integrations can be performed to infinity; viz.,

$$\rho^2 = 2\int_{cr_0/s_{\max}}^{cr_0/s_{\min}} rdr - 2\int_{cr_0/s_{\max}}^{\infty} r\exp\left[-\left(\frac{s_{\max} r/r_0 - c}{a}\right)^b\right]dr + 2\int_{cr_0/s_{\min}}^{\infty} r\exp\left[\left(\frac{s_{\min} r/r_0 - c}{a}\right)^b\right]dr.$$

By the substitution

$$z = \left(\frac{\frac{sr}{r_0} - c}{a}\right)^b,$$

the integral

$$\int_{cr_0/s}^{\infty} r\exp\left[\left(\frac{\frac{sr}{r_0} - c}{a}\right)^b\right]dr$$

becomes

$$\frac{r_0{}^2 a}{bs^2}\int_0^\infty (az^{(2/b)-1} + cz^{(1/b)-1})e^{-z}dz = \frac{r_0{}^2 a}{bs^2}[a\Gamma(2/b) + c\Gamma(1/b)]$$

and the complete expression becomes

$$\rho^2 = r_0{}^2(1/s_{\min}^2 - 1/s_{\max}^2) \times \left\{c^2 + \frac{2a}{b}[a\Gamma(2/b)] + c\Gamma(1/b)\right\}.$$

The effective radius with undamped propagation as a function of $s_{\min}$ is shown in Fig. 11. Here, $s_{\max}$ has little influence.

All these calculations are based on an idealized counter, which counts everything between $s_{\min}$ and $s_{\max}$ (Fig. 12). A more realistic counter ought to behave more like the function shown in Fig. 13. If the shape of this function is obtained, and piecewise linear or polynomial functions are used as approximations, the calculation of the acceptance will contain the incomplete gamma function

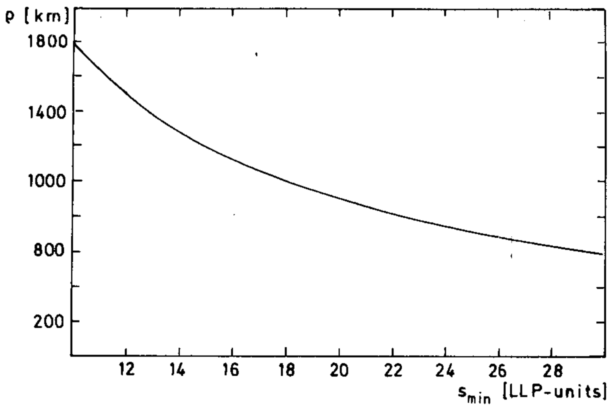

FIG. 11. Effective radius of an ideal lightning counter/direction finder as a function of the lower threshold using the signal distribution from Pisler (1984).

$$\gamma(x, y) = \int_0^y t^{x-1} e^{-t} dt$$

as given in some examples by Lundtang Petersen et al. (1981).

b. Damping

The inverse cumulative Weibull distribution is also analytic, viz.,

$$x(F) = a\{\ln[1/(1-F)]\}^{1/b} + c.$$

This can be used for a test of the hypothesis that the damping is dependent on the signal strength of the lightning, due to different signal shapes for strong and weak lightning. The reference distribution is inverted for undamped signals yielding

$$s = a\{\ln[1/(1-F)]\}^{1/b} + c.$$

The distribution of damped signals is also inverted, viz.

$$\tilde{s} = \tilde{a}\{\ln[1/(1-F)]\}^{1/\tilde{b}} + \tilde{c}.$$

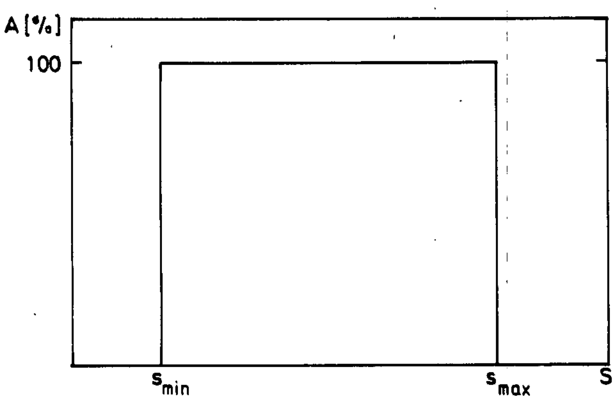

FIG. 12. Acceptance as function of the signal strength for an ideal lightning counter/direction finder.

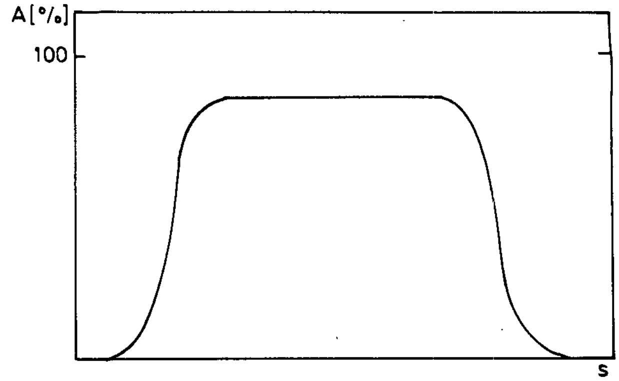

FIG. 13. As Fig. 12 but for a real lightning counter/direction finder.

We can thus define the damping as a function of F, given by

$$d(F) = \tilde{s}/s = \frac{\tilde{a}\{\ln[1/(1-F)]\}^{1/\tilde{b}} + \tilde{c}}{a\{\ln[1/(1-F)]\}^{1/b} + c}.$$

Substituting $1/(1-F) = \exp\left[\left(\frac{s-c}{a}\right)^b\right]$, the damping as a function of s is obtained from

$$d(s) = \frac{\tilde{a}[(s-c)/a]^{b/\tilde{b}} + \tilde{c}}{s}.$$

This was calculated for the two distributions given before (see Fig. 14).

7. Discussion

In many practical applications, the scatter and uncertainty in the data is larger than the difference between statistical models fitted to the data. In such cases,

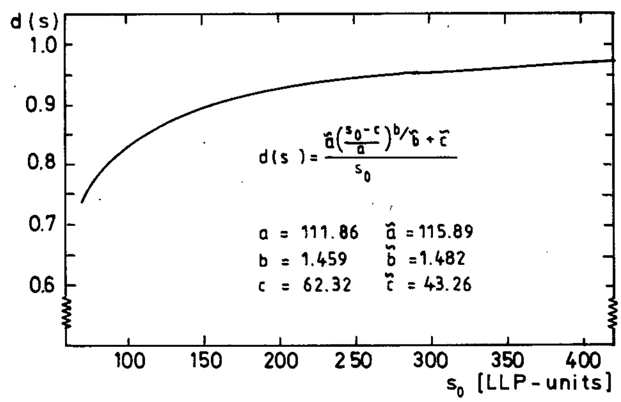

FIG. 14. Damping as a function of the signal strength; reference distribution from Pisler (1984), damped distribution from Cooray and Lundquist (1982).

one wants to choose the distribution model which is most favorable for calculations and further evaluations. From such a point of view, the Weibull distribution has been shown to be a statistical model with wide applicability in thunderstorm research. Its simple mathematical structure enables quantitative calculations from lightning statistical data, extracting information in an effective way. In evaluations of the performance of lightning detection systems, the Weibull model was used as the base for the calculations (Schütte et al., 1986; Schütte and Pisler, 1986). The common use of the Weibull distribution in insulation breakdown research is encouraging from a theoretical point of view, since lightning itself can be considered as a breakdown phenomenon on a larger scale.

A first report on signal-strength distributions calculated by the use of lognormal and Weibull paper for a very large amount of lightning from Orville et al., (1985)—when this paper was nearly completed—shows that the distribution cannot be fitted perfectly by these models. The consequence of different kinds of lightning and the effects of the measuring equipment probably will be that a very exact fit is possible only by using a sum of distributions. But, as already mentioned, for calculations of the parameters in this paper, a simple Weibull fit will be sufficient. The errors arising from other assumptions, such as the threshold values of a lightning counter/distribution finder, will necessarily be larger than those of the use of the Weibull distribution.

Acknowledgments. We wish to thank Professor Stig Lundquist for showing a great interest in this work and its further applications and for encouraging us. We thank Dr. E. Pisler and Dr. V. Cooray for the supply of relevant data and for discussions, especially on the damping problems. Our thanks for the layout of this report are directed to M. Björklund for the typing and A. Ekström for the drawings.

REFERENCES

Berrettoni, J. B., 1964: Practical application of the Weibull distribution. *Ind. Qual. Contr.*, **21**, 71–79.

Cooray, C., and S. Lundquist, 1982: On the characteristics of some radiation fields and their possible origin in positive ground flashes. *J. Geophys. Res.*, **87**, 11 203–11 214.

Hahn, G., and S. Shapiro, 1967: *Statistical models in engineering.* Wiley, 108 pp.

Lundtang Petersen, E., I. Troen, S. Frandsen and K. Hedegaard, 1981: *Wind Atlas for Denmark.* Risø National Laboratory, Roskilde, Denmark.

Orville, R. E., R. B. Pyle, R. W. Henderson, R. E. Orville, Jr. and R. A. Weisman, 1985: Characteristics of cloud-to-ground lightning flashes along the east coast of the United States. *Tenth Int. Aerospace and Ground Conf. on Lightning and Static Electricity,* Paris, 333–336.

Peckham, W., M. Uman and C. Wilcox, 1984: Lightning phenomenology in the Tampa Bay Area. *J. Geophys. Res.*, **D89**, 11 789–11 805.

Pisler, E., 1984: Registration of lightnings during the winter 79/80 using LLP system. Institute of High Voltage Research, Uppsala, UURIE: 153–84.

Schütt, Th., and E. Pisler, 1986: Some improved evaluation of the performance of lightning detection systems. Institute of High Voltage Research, Uppsla, UURIE: 186–86.

——, ——, D. Filipović and S. Israelsson, 1986: The acceptance of lightning detectors and lightning localization systems under different damping conditions. Institute of High Voltage Research, Uppsala, UURIE: 180–86.

Stewart, D. A., and O. M. Essenwanger, 1978: Frequency Distribution of Wind Speed Near the Surface. *J. Appl. Meteor.*, **17**, 1633–1642.

Weibull, W., 1939: A statistical theory of the strength of materials. *Ingenjörsvetenskapsakademiens handlingar,* **151**, Stockholm.