Method to Determine Number and Size of Samples Taken from Zinc Roof to Analyze Pitting Corrosion

Rongjin Zheng

Jan E. Carmeliet, Ph.D.

Hugo Hens, Ph.D.
Associate Member ASHRAE

Walter F. Bogaerts, Ph.D.

ABSTRACT

Pitting corrosion is one of the most dangerous corrosion types for zinc roofs. The theory of extreme-value statistics is a powerful tool to quantitatively analyze pitting corrosion. For obtaining reliable results, it is very important to know how to collect small samples correctly and effectively. The method proposed by JSCE (1988) is discussed. This method considers only the influence of the number of small samples on the estimated maximum pit depth over a large area and not the influence of the number of pits per small sample. An improved method, which takes both influences into account, is presented in this paper. Two examples are given to demonstrate the application of the improved method in zinc roofs.

INTRODUCTION

Zinc roofs are widely used all over the world. The advantages of the system are many. One of the main problems, however, is the degradation of zinc sheets due to corrosion. Service life of zinc roofs may vary from some decades to longer than a century depending upon the type of zinc, the structure of the roof, and the surrounding environment. Mostly, the maintenance and replacement of zinc roofs are induced by the corrosion of the zinc sheeting.

In the last decades, the growing concern about global warming and energy efficiency has pushed the U-factor of building elements toward lower threshold values. This trend to lower U-factors may greatly affect the performance of zinc sheeting due to a worse hygrothermal environment at the underside of the zinc sheeting.

In order to investigate the performance of highly insulated zinc roofs, a field research program has been carried out at the Laboratory of Building Physics of the KU Leuven since November 1996. In the program, a total of eight zinc roofs (four cold roofs and four warm roofs) with high insulation quality ($U < 4.40 \times 10^{-2}$ Btu/h·ft^2·°F (0.25 W/m²·K) were constructed. The dimensions of the zinc roofs were about 1.8 by 3 m. The roof slope was 5%. A commercial zinc sheet of 0.65 mm thickness (VM zinc titanium) was used in the zinc roofs. The orientation of four zinc roofs was southwest; the orientation of the other four zinc roofs was northeast.

For examining the corrosion of the zinc sheets, one large zinc specimen was taken from each zinc roof in 1999 and in 2000, respectively. The dimensions of the zinc specimens were 0.50 by 0.70 m. We found that, for both the highly insulated cold zinc roofs and the highly insulated warm zinc roofs, severe pitting corrosion at the underside of the zinc sheeting developed. Figures 1 through 4 show the typical appearance of the underside of the zinc sheeting before and after cleaning.

Pitting corrosion is a deadly dangerous corrosion type for zinc roofs. If the depth of the pits continues to increase with time, local leakage may occur. In evaluation of pitting corrosion, both the average corrosion rate based on mass loss and the average pit depths cannot readily be used. The corrosion estimation based on mass loss can seriously underestimate the corrosion penetration caused by pitting corrosion. As for the corrosion estimation based on the average pit depths, first, many smaller pits become too indistinguishable to be measured due to stopping growth very soon after it is initiated, and second, the deeper pits are of the most practical impor-

Rongjin Zheng is a student and **Jan Carmeliet** and **Hugo Hens** are professors in the Department of Civil Engineering and **Walter Bogaerts** is a professor in the Department of Metallurgy and Materials Engineering, University of Leuven, Belgium.

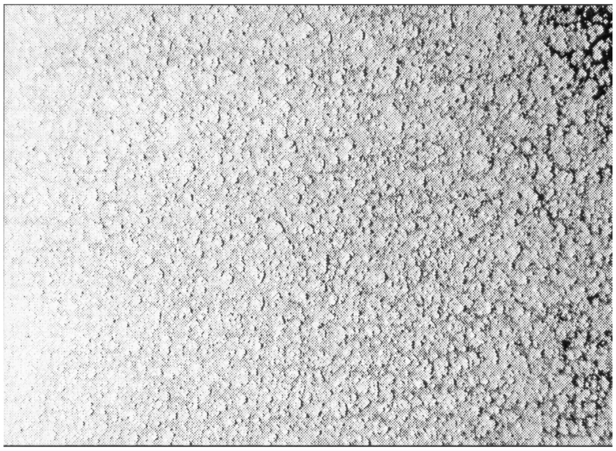

Figure 1 *Typical appearance of the zinc sheeting not contacting the wood deck in the cold zinc roofs before cleaning.*

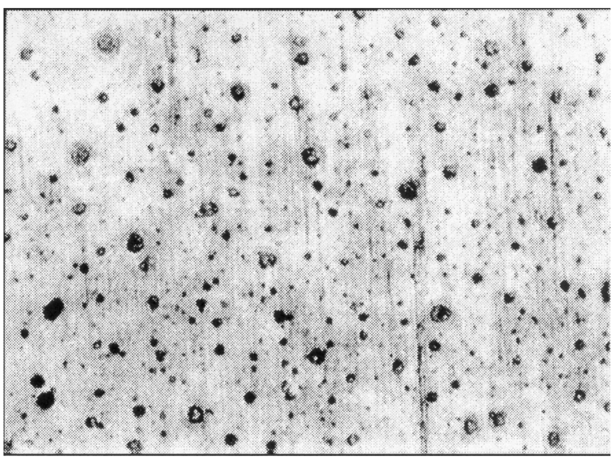

Figure 2 *Typical appearance of the zinc sheeting not contacting the wood deck in the cold zinc roofs after cleaning.*

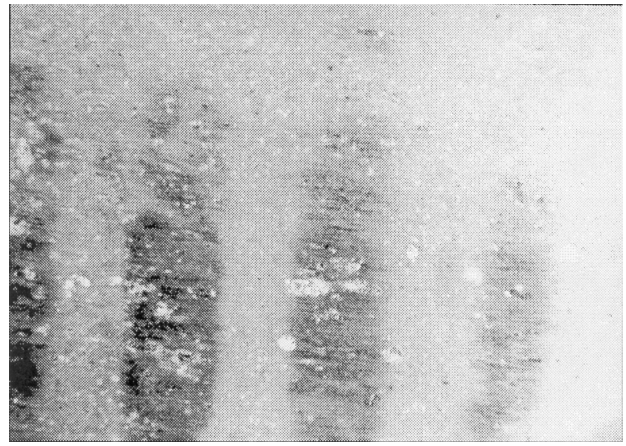

Figure 3 *Typical appearance of the zinc sheeting contacting the mineral wool in the warm zinc roofs before cleaning.*

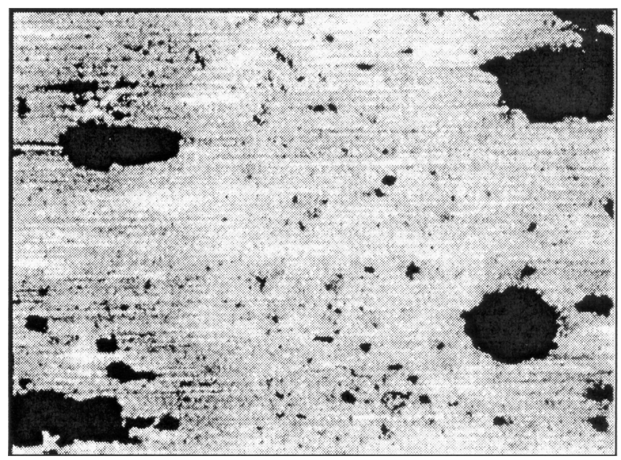

Figure 4 *Typical appearance of the zinc sheeting contacting the mineral wool in the warm zinc roofs after cleaning.*

tance. It is the deepest pits that, if they continue to propagate, will result in leakage. Therefore, it is not the overall corrosion rate or the number of pits that are of interest, but the deepest pits since they cause the failure of zinc roofs.

The theory of extreme value has been proven to be a powerful tool to analyze pitting corrosion. It can be used to estimate the maximum pit depth of a large area of zinc sheeting on the basis of the examination of a small portion of that area. In practical applications, however, it remains a problem of how to take field samples of zinc roofs to collect the data for analysis. Before sampling, we have to answer the following two questions: (1)which number of samples, N, is necessary, and (2) which size, s, should be chosen for the area of the sample for a given large area S, in order to draw reliable conclusions?

Considering these questions, a method (JSCE 60-1 1988) has been recommended in order to select suitable values for s and N based on the variance of the estimation for the maximum pit depth. This method is based on the relation between the number of small samples and the variance of the estimation for the maximum pit depth over a large area. It just requires that s should be chosen so as to contain a plural number of pits (Shibata 1996). This means that the influence of the number of pits per small sample on s and N is not considered quantitatively. However, it is well known that the number of pits on small samples may influence the precision of the distribution of the maximum pit depths predicted by extreme value statistics. Gumbel (1967) pointed out that the number of observations, n (i.e., the number of pits per small sample), from which the extremes (i.e., the maximum pit depths) are taken, must be large while applying the extreme value theory. How large n has to be depends on the initial distribution and on the degree of precision we want to obtain. Galambos (1987) presented formulae to estimate the error of the cumulative probability of extremes due to the passage to the limit as the observations, n, increase indefinitely. So, when selecting small samples, it seems necessary to consider the influence of both the number of pits per small sample and the number of small samples on the estimated maximum pit depth over a large area.

According to a study done by Shibata (1994), the pit depth distribution on metal surfaces usually obeys the normal or exponential distribution; both of them belong to the exponential distribution family. Thus, the maximum values of pit depth extracted from the exponential family distribution may reasonably be expected to obey the Gumbel distribution (i.e., $\exp[-e^{-y}]$) where y is the so-called *reduced variate*. Our research (Zheng 1999) also showed that, for both the highly insulated cold zinc roofs and the highly insulated warm zinc roofs, the maximum pit depths of the zinc sheeting on replicate samples were statistically distributed according to the Gumbel distribution.

The objective of this study is to select the optimum combination of area per small sample (s) and number of small samples (N) for obtaining a reasonable prediction of the maximum pit depth over a large area at certain accuracy. In the present work, we assume that the distribution of the maximum pit depths in zinc roofs follows the Gumbel type.

DERIVATION OF THE FORMULAE FOR DETERMINING AREA PER SMALL SAMPLE AND NUMBER OF SMALL SAMPLES

It is well known that the number of pits on small samples influences the precision of the maximum pit depth over a large area estimated by extreme value theory. Suppose $x_1, x_2, x_3, \cdots, x_n$ are independent and identically distributed (i.i.d.) random variables (pit depths) with common distribution function $F(x)$ on a sample of area s, and $Z_n = \mathrm{Max}\,(x_1, x_2, x_3 \cdots, x_n)$, then the probability $P(Z_n < x)$ that x will be the largest among the n pits on the sample is

$$P(Z_n < x) = F^n(x). \tag{1}$$

According to the theory of extreme statistics, there may exist sequences a_n and $b_n > 0$ of constants such that

$$\lim_{n \to \infty} P(Z_n < a_n + b_n x) = \lim_{n \to \infty} F^n(a_n + b_n x) = H(x) \tag{2}$$

where $H(x)$ is a nondegenerating distribution function, which is one of the three types of the asymptotic extreme distributions.

Equation 2 is usually applied to evaluate the parameters a_n and b_n on the basis of data of the maximum pit depths measured in small samples. Then, the maximum pit depth over a large area is estimated by the parameters a_n, b_n, and return period T. Clearly, Equation 2 implies the assumption of an infinite number of pits on small samples. In practice, the number of pits on small samples, however, is not infinite, but a value that depends on the density of pits and on the area of the small sample. Therefore, using Equation 2 may cause some error as a result of the assumption of an infinite number of pits on small samples. For a particular case, when there is only one pit per small sample, according to Equation 1, the exact distribution of extremes is the initial distribution function $F(x)$ itself, which is far from being the distribution of extremes obtained by Equation 2.

Obviously, a larger number of pits on small samples may decrease the error caused by the assumption of an infinite number of pits in Equation 2. The question is how much the error would be due to the replacement of the exact distribution of the extremes by the Gumbel distributions to estimate the maximum pit depth over a large area (i.e., how the pit number n on small samples influences the estimated maximum pit depth over a large area).

Let $H(x)$ be the Gumbel distribution and $F(x)$ be in the domain of attraction of $H(x)$. Then, for given sequences a_n and $b_n > 0$, we have (Galambos 1987)

$$\begin{gathered}\left|P(Z_n < a_n + b_n x) - H(x)\right| \le H(x) \\ \left\lfloor \gamma_{1,n}(x) + \gamma_{2,n}(x) + \gamma_{1,n}(x) \times \gamma_{2,n}(x) \right\rfloor\end{gathered} \tag{3}$$

where

$$P(Z_n < a_n + b_n x) = F^n(a_n + b_n x) \tag{4}$$

$$\gamma_{1,n}(x) = \frac{2z_n^2(x)}{n} + \frac{2z_n^4(x)}{n^2} \times \frac{1}{1-q} \tag{5}$$

$$\gamma_{2,n}(x) = \left|\rho_n(x)\right| + \frac{\rho_n^2(x)}{2} \times \frac{1}{1-p} \tag{6}$$

$$z_n(x) = n[1 - F(a_n + b_n x)] \tag{7}$$

$$\rho_n(x) = z_n(x) + \ln H(x) \tag{8}$$

$$\text{with } \frac{z_n(x)}{n} \le \frac{1}{2}, \ \frac{2z_n^2 x}{3n} \le q < 1, \text{ and } \frac{1}{3}\left|\rho_n(x)\right| \le p < 1 \tag{9}$$

and

n = number of pits per small sample,

q, p = any real numbers satisfying Equation 9.

When a_n and b_n are chosen so that $P[(Z_n - a_n)/b_n < x]$ weakly converges to $H(x)$ as n increases indefinitely, the assumed bounds on $z_n(x)/n$, $z_n^2(x)/n$, and $\left|\rho_n(x)\right|$ are valid for very moderate values of n (usually for n as small as 2 or 3).

It is well known that the Gumbel distribution holds for the initial distribution of an exponential type, which includes the exponential, normal, lognormal, logistic, and chi-square functions. Furthermore, any initial probability functions with an exponential type in the neighborhood of the characteristic largest value for large x may take the following approximation (Gumbel 1967):

$$F(x) = 1 - \frac{\exp\left(-\frac{x-a_n}{b_n}\right)}{n} \tag{10}$$

Equation 10 leads immediately to the Gumbel distribution,

$$\begin{aligned} H(x) &= \lim_{n\to\infty} P(Z_n < a_n + b_n x) = \lim_{n\to\infty} F^n(a_n + b_n x) \\ &= \lim_{n\to\infty}\left[1 - \frac{\exp(-x)}{n}\right]^n = \exp[-\exp(-x)]. \end{aligned} \tag{11}$$

Gnedenko (1943) has shown that the condition in Equation 10 is not only sufficient, but also necessary, for the existence of the asymptotic distribution in Equation 11.

Based on Equations 10 and 11, Equations 7 and 8 may be written as

$$z_n(x) = \exp(-x) \tag{12}$$

$$\rho_n(x) = 0 \tag{13}$$

By inserting Equations 12 and 13 into Equations 5 and 6, we have

$$\gamma_{1,n}(x) = \frac{2\exp(-2x)}{n} + \frac{2\exp(-4x)}{n^2} \times \frac{1}{1-q} \tag{14}$$

$$\gamma_{2,n}(x) = 0 \tag{15}$$

Since $1 > q \geq \frac{2\exp(-2x)}{3n}$ and $\frac{2}{3n} \geq \frac{2\exp(-2x)}{3n}$, we choose $q = \frac{2}{3n}$.

Therefore, from Equations 14 and 15, Equation 3 is simplified to

$$|P(Z_n < a_n + b_n x) - H(x)| \leq H(x) \times \gamma_{1,n}(x). \tag{16}$$

Equation 16 can also be written as

$$\left|P(Z_n < x) - H\left(\frac{x-a_n}{b_n}\right)\right| \leq H\left(\frac{x-a_n}{b_n}\right) \times \gamma_{1,n}\left(\frac{x-a_n}{b_n}\right). \tag{17}$$

Now, let us estimate the value of $\gamma_{1,n}$. Some researchers (Aziz 1956; Shibata 1991, 1996) have shown that b_n/a_n does not change appreciably with time and is usually below, or not much larger than, 0.3. Laycock et al. (1990) also assumed $a_n = a_0 t^c$ and $b_n = b_0 t^c$ (c is a constant and t is time) when they developed a model for data collected on maximum pit depths enabling simultaneous extrapolation into the future over large areas of exposed zinc. The results obtained in the test building (Zheng 1999) showed the variation of b_n/a_n in between 0.12 and 0.35. As an estimation, we assume $b_n/a_n < 0.4$. Since the aim is to evaluate the error of the predicted maximum pit depth over a large area, the variable x in Equation 17 will be estimated with a large return period. We may assume $x_{max} > 2a_n$. On the basis of these assumptions, we can obtain an estimation of $\gamma_{1,n}$ from Equation 14:

$$\gamma_{1,n}\left(\frac{x-a_n}{b_n}\right) < \frac{2e^{-5}}{n} + \frac{6e^{-10}}{n(3n-2)} = k \tag{18}$$

Then, Equation 17 becomes

$$\left|P(Z_n < x) - H\left(\frac{x-a_n}{b_n}\right)\right| < kH\left(\frac{x-a_n}{b_n}\right). \tag{19}$$

Releasing the absolute sign, we have

$$(1-k)H\left(\frac{x-a_n}{b_n}\right) < P(Z_n < x) < (1+k)H\left(\frac{x-a_n}{b_n}\right). \tag{20}$$

Now we can estimate the maximum pit depth over a large area with a return period T using the probability functions in Equation 20 and using the Gumbel distribution, respectively.

1. x_{1-k} (the maximum pit depth estimated using the probability of low bound in Equation 20):

$$(1-k)\exp\left[-\exp\left(-\frac{x_{1-k}-a_n}{b_n}\right)\right] = 1 - \frac{1}{T} \tag{21}$$

$$x_{1-k} = a_n - b_n \ln\left(\ln\frac{T(1-k)}{T-1}\right) \tag{22}$$

2. x_p (the exact maximum pit depth):

$$P(Z_n < x_p) = 1 - \frac{1}{T} \tag{23}$$

3. x_{1+k} (the maximum pit depth estimated using the probability of high bound in Equation 20):

$$(1+k)\exp\left[-\exp\left(-\frac{x_{1+k}-a_n}{b_n}\right)\right] = 1 - \frac{1}{T} \tag{24}$$

$$x_{1+k} = a_n - b_n \ln\left(\ln\frac{T(1+k)}{T-1}\right) \tag{25}$$

4. x_h (the maximum pit depth estimated using the Gumbel distribution):

$$\exp\left[-\exp\left(-\frac{x_h-a_n}{b_n}\right)\right] = 1 - \frac{1}{T} \tag{26}$$

$$x_h = a_n - b_n \ln\left(\ln\frac{T}{T-1}\right) \tag{27}$$

Due to the nondegenerating property of the probability distribution functions with x in Equation 20, we may have

$$x_{1-k} > x_p > x_{1+k}. \tag{28}$$

Then, subtraction and division of Equation 28 by a_n and $(x_h - a_n)$ lead to

$$\frac{x_{1-k}-a_n}{x_h-a_n} > \frac{x_p-a_n}{x_h-a_n} > \frac{x_{1+k}-a_n}{x_h-a_n}. \tag{29}$$

By inserting Equations 22, 25, and 27 into Equation 29, we obtain

$$\frac{\ln\left(\ln\frac{T(1-k)}{T-1}\right)}{\ln\left(\ln\frac{T}{T-1}\right)} > \frac{x_p-a_n}{x_h-a_n} > \frac{\ln\left(\ln\frac{T(1+k)}{T-1}\right)}{\ln\left(\ln\frac{T}{T-1}\right)}. \tag{30}$$

Equation 30 shows that the ratio of the difference between the exact maximum pit depth over a large area and the location parameter based on small samples to the difference between the maximum pit depth estimated by the Gumbel distribution over a large area and the location parameter based on small samples varies within certain bounds. Hereafter, we use R to denote this ratio. Certainly, R must converge to unity as n increases indefinitely (i.e., there is no error as n is indefinite). Thus, the degree of R deviating from unity can be used as an indicator to control the error. The bounds of R depend upon the number of pits on small samples and the return period. From Equation 30, we can easily obtain the values of both bounds of R.

In practice, we usually hope to control the degree of R deviating from unity within some accepted levels. For this purpose, we may simply solve the following equations:

$$\frac{\ln\left(\ln\frac{T(1-k)}{T-1}\right)}{\ln\left(\ln\frac{T}{T-1}\right)} \le R_1 \tag{31}$$

and

$$\frac{\ln\left(\ln\frac{T(1+k)}{T-1}\right)}{\ln\left(\ln\frac{T}{T-1}\right)} \ge R_2 \tag{32}$$

where

R_1 = high bound of R,

R_2 = low bound of R.

After rearranging Equations 31 and 32, they can be expressed as

$$k \le 1 - \frac{T-1}{T}\exp\left[\left(\ln\frac{T}{T-1}\right)^{R_1}\right] \tag{33}$$

and

$$k \le \frac{T-1}{T}\exp\left[\left(\ln\frac{T}{T-1}\right)^{R_2}\right] - 1. \tag{34}$$

Neglecting the second term of k in Equation 18, we can obtain the following approximate solutions from Equations 33 and 34, respectively.

$$n_1 \ge \frac{2\exp(-5)}{1-\frac{T-1}{T}\exp\left[\left(\ln\frac{T}{T-1}\right)^{R_1}\right]} \tag{35}$$

$$n_2 \ge \frac{2\exp(-5)}{\frac{T-1}{T}\exp\left[\left(\ln\frac{T}{T-1}\right)^{R_2}\right]-1} \tag{36}$$

where

n_1, n_2 = required pit number in small samples corresponding to the higher and lower bounds of the ratio R, respectively.

Hence, if both Equations 35 and 36 hold, we have

$$R_1 > R > R_2. \tag{37}$$

Assuming that the density of pits in small samples is ρ and the area of a zinc roof is S, then $n = \rho \times s$ and $S = T \times s$, and we can obtain

$$s_1 \ge \frac{2\times\exp(-5)}{\rho\times\left[1-\frac{S-s_1}{S}\exp\left(\ln\frac{S}{S-s_1}\right)^{R_1}\right]} \tag{38}$$

$$s_2 \ge \frac{2\times\exp(-5)}{\rho\times\left[\frac{S-s_2}{S}\exp\left(\ln\frac{S}{(S-s_2)}\right)^{R_2}-1\right]} \tag{39}$$

where

s_1, s_2 = required area of small samples for the higher (R_1) and lower (R_2) bounds of the ratio R, respectively.

Formulae to Determine the Required Number of Small Samples (*N*) and the Area per Small Sample (*s*) for Controlling the Variance of the Predicted Maximum Pit Depth for a Given Large Area Due to the Finite Number of Small Samples

There are several methods that can be used to estimate the location parameter a_n and the scale parameter b_n, such as graphical, MVLUE (minimum variance unbiased estimator), maximum likelihood, and moment methods. Among them, the MVLUE method is found to be more efficient and unbiased for small size samples. Tsuge (1987) had shown that parameters estimated by the MVLUE method are unbiased and efficient and are consistent with values estimated by the method of moments or maximum likelihood when the sample size is more than 30. The distribution parameters can be calculated by the MVLUE method with the following formulae:

$$a_n = \sum a_i(N,j)x_i \text{ and } b_n = \sum b_i(N,j)x_i \tag{40}$$

where

$a_i(N,j)$, $b_i(N,j)$ = weights for each sample, respectively,

N, j = the total number of samples and truncated number, respectively.

The variance $V(x)$ of the estimation x_{max} using the MVLUE method can be calculated by

$$V(x) = b_n^2 \lfloor A(N,j)y^2 + B(N,j)y + C(N,j) \rfloor \quad (41)$$

where

b_n = scale parameter,

y = reduced variant,

$A(N,j)$, $B(N,j)$, $C(N,j)$= weights in MVLUE estimation procedure.

Lieblein (1954) devised these estimation procedures, which are especially adapted to very small samples of extremes, in order to obtain a great amount of information from costly data. He has given the values of $a_i(N,j)$ and $b_i(N,j)$ with $A(N,j)$, $B(N,j)$, and $C(N,j)$ up to $N = 6$. For those samples larger than 6, he proposed a procedure taken from quality control by making subgroups (five or six) from the entire set of samples. Later, White (1964) and Tsuge (1987) calculated the values up to $N = 20$ and $N = 45$, respectively.

If controlling the variance of estimated x_{max} within $(a_n/m)^2$, we may have the following relation from Equation 41:

$$\left(\frac{a_n}{mb_n}\right)^2 \geq A(N,j)y^2 + B(N,j)y + C(N,j) \quad (42)$$

The reduced variant y can be calculated with T as follows:

$$y = -\ln\left[-\ln\left(\frac{T-1}{T}\right)\right] \approx \ln(T)\text{; when } T > 18 \quad (43)$$

$$T = \frac{S}{s_N}. \quad (44)$$

By inserting Equations 43 and 44 into Equation 42, we obtain

$$\left(\frac{a_n}{mb_n}\right)^2 \geq A(N,j)\left(\ln\frac{S}{s_N}\right)^2 + B(N,j)\left(\ln\frac{S}{s_N}\right) + C(N,j) \quad (45)$$

where

m = the assumed number, such as 1, 2, 3, etc.,

s_N = the required area per small sample.

Combining Equations 38 and 39 with Equation 45, we can easily get sequences of s and N such that the precision of the estimated maximum pit depth over a large area meets the requirement we seek. If the ratio a_n/b_n is known from experience or a database, the values for s and N may be estimated easily. Fortunately, the analysis of the accumulated data of the parameter of a_n and b_n showed (Shibata 1991) that the ratio of b_n/a_n for localized corrosion is below, or not much larger than, 0.3.

In theory, for enhancement of estimating reliability of the maximum pit depth over a large area, large N or s is desirable. In consideration of economy, however, there must be a natural limitation to the increase of N or s for the extreme value survey. For a given large area and precision, it is also obvious that numerous combinations of s and N exist, which meet Equations 38, 39, and 45. Naturally, we hope to find an optimum combination of s and N among them such that (1) the total sampling area ($N \times s$) is minimum under the condition of guaranteeing the precision of the predicted maximum pit depths over a large area, or (2) the precision of the predicted maximum pit depth over a large area is highest for a given total sampling area.

Assuming the total sampling area to be A_s (= $N \times s$), then Equations 38, 39, and 45 may be expressed as

$$A_s \geq N \times \frac{2 \times \exp(-5)}{\rho \times \left[1 - \frac{S - \frac{A_s}{N}}{S}\exp\left(\ln\frac{S}{S - \frac{A_s}{N}}\right)^{R_1}\right]} \quad (46)$$

$$A_s \geq N \times \frac{2 \times \exp(-5)}{\rho \times \left[\frac{S - \frac{A_s}{N}}{S}\exp\left(\ln\frac{S}{S - \frac{A_s}{N}}\right)^{R_2} - 1\right]} \quad (47)$$

$$\left(\frac{a_n}{mb_n}\right)^2 \geq A(N,j)\left(\ln\frac{N \times S}{A_s}\right)^2 + B(N,j)\left(\ln\frac{N \times S}{A_s}\right) + C(N,j) \quad (48)$$

Solving Equations 46, 47, and 48, we can obtain the optimum selection of N and s under the given conditions.

APPLICATION OF THE IMPROVED METHOD

Two examples are presented to demonstrate the application of the improved method to the zinc roofs in the laboratory building. The area of each zinc roof is 58.13 ft^2 (5.40 m^2). Since the corrosion of zinc in most atmospheric environments is generally uniform, localized corrosion of zinc has not received much attention. So, little information about the distribution of pits on zinc surfaces is available. The results obtained in the laboratory building (Zheng 1999) showed that the variation of b_n/a_n was between 0.12 and 0.35. We here assume b_n/a_n to be 0.25. The density of pits is assumed to be 3.66×10^5 pits/ft (1.2×10^6 pits/m^2).

Obtaining a Minimum Total Sampling Area

Naturally, we always hope to collect the total sampling area A_s (=$N \times s$) as small as possible in the condition of guaranteeing the precision of the predicted maximum pit depths. Collecting and checking smaller total sampling area may reduce cost and save time.

Assumptions: $R_1 = 1.05$, $R_2 = 0.95$, $m = 1.5$.

Inserting the assumptions into Equations 46, 47, and 48, we have the following:

$$A_s \geq N \times \frac{2 \times \exp(-5)}{1.2 \times \left[1 - \frac{5.4 \times 10^6 - \frac{A_s}{N}}{5.4 \times 10^6}\exp\left(\ln\frac{5.4 \times 10^6}{5.4 \times 10^6 - \frac{A_s}{N}}\right)^{1.05}\right]} \quad (49)$$

$$A_s \geq N \times \frac{2 \times \exp(-5)}{1.2 \times \left[\frac{5.4 \times 10^6 - \frac{A_s}{N}}{5.4 \times 10^6} \exp\left(\ln \frac{5.4 \times 10^6}{5.4 \times 10^6 - \frac{A_s}{N}} \right)^{0.95} - 1 \right]} \tag{50}$$

$$\left(\frac{1}{1.5 \times 0.25}\right)^2 \geq A(N, j)\left(\ln \frac{N \times 5.4 \times 10^6}{A_s}\right)^2 + B(N, \mathrm{j}) \left(\ln \frac{N \times 5.4 \times 10^6}{A_s}\right) + C(N, j) \tag{51}$$

Solving Equations 49, 50, and 51, the minimum total sampling area A_s is 18.09 in.^{2} (11673 mm^2) and the number of the small samples N is 21. The corresponding area per small sample s is 0.862 in.^{2} (556 mm^2).

Obtaining an Optimum Combination of the Area per Small Sample and the Number of Small Samples for a Given Total Sampling Area

In reality, the total sampling area may be restricted for some reason. For this situation, we of course want to choose the optimum combination of s and N, for a given total sampling area, such that the accuracy of estimating the maximum pit depth is highest.

Assumptions: R_1 = 1.05, R_2 = 0.95, A_s = 15.5 in.^{2} (1×10^4 mm^2). The aim is to find the optimum combination of s and N such that the variance of the maximum pit depth is smallest (i.e., m is as large as possible).

Inserting the assumptions into Equations 46, 47, and 48, we obtain the following:

$$10^4 \geq N \times \frac{2 \times \exp(-5)}{1.2 \times \left[1 - \frac{5.4 \times 10^6 - \frac{10^4}{N}}{5.4 \times 10^6} \exp\left(\ln \frac{5.4 \times 10^6}{5.4 \times 10^6 - \frac{10^4}{N}} \right)^{1.05} \right]} \tag{52}$$

$$10^4 \geq N \times \frac{2 \times \exp(-5)}{1.2 \times \left[\frac{5.4 \times 10^6 - \frac{10^4}{N}}{5.4 \times 10^6} \exp\left(\ln \frac{5.4 \times 10^6}{5.4 \times 10^6 - \frac{10^4}{N}} \right)^{0.95} - 1 \right]} \tag{53}$$

$$\left(\frac{1}{m \times 0.25}\right)^2 \geq A(N, j)\left(\ln \frac{N \times 5.4 \times 10^6}{10^4}\right)^2 + B(N, j) \left(\ln \frac{N \times 5.4 \times 10^6}{10^4}\right) + C(N, j) \tag{54}$$

Solving Equations 52, 53, and 54, we get the maximum m to be 1.35. The corresponding optimum combination is N = 20 and s = 0.775 in.^{2} (500 mm^2).

SOME REMARKS

It is obvious that the variate of the extreme value distribution used is unlimited in both the negative direction and the positive direction. The infinite range of the distribution apparently conflicts with the finite range of the possible maximum pit depths. However, for practical applications, the probability that the maximum pit depth is negative may be neglected even for moderate sample sizes (Gumbel 1967). Consider a variate unlimited in both direction, and let the median be zero. Then the probability H for the maximum pit depth x_n to be negative is

$$H(x_n < 0) = \left(\frac{1}{2}\right)^n. \tag{55}$$

It approaches zero very quickly for increasing n. For n = 10, H < 9.77e-4; n = 20, H < 9.54e-7.

In the positive direction, physical considerations would lead us to believe that the pit depth frequency would close on the x-axis at some finite value x_0 of the variate so that for a pit depth $x > x_0$ the probability is zero. Aziz (1956) pointed out that the possible range of pit depths might be several orders of magnitude greater than the range generally observed so that there was no real problem in applying the theory. Thus, data can be safely extrapolated over at least two or three times the observed range permitting the direct estimation of the probabilities and "return periods" of a given pit depth. Actually, the concept of an infinite range in the mathematical model is implicit in almost all forms of statistical analysis and its use is justified by years of successful application.

According to Gumbel (1967), the essential condition in the analysis is that the distribution from which the extremes are drawn and its parameters remain constant in time (or space), or the influence that time (or space) exercises upon them is taken into account or eliminated. In practical applications, however, conditions may change from one set of samples to the other due to local variations in environment or surface condition of the zinc sheeting. An important measure is to clarify the corrosion condition of the zinc sheeting to allow data to be grouped and collected under the same condition. HAM (heat, air, and moisture) modeling may be of great help in clarifying the corrosion condition of the zinc sheeting.

It should be pointed out that the conditions of Equation 35 and Equation 36, for bounds of the ratio R, are just sufficient. It is possible to get more accurate bounds if more information about a_n and b_n is available.

CONCLUSIONS

1. An improved method has been presented to determine the number of small samples and the size of small samples collected from a large area of a zinc roof to analyzing pitting

corrosion. It should be emphasized that the improved method is based on the assumption that the distribution of the maximum pit depths takes the Gumbel distribution.

2. Before sampling, it is necessary to clarify the corrosion condition of zinc sheets due to local variations in environment at the underside of the zinc sheets. HAM (heat, air, and moisture) modeling may be of great help in clarifying the corrosion condition of the zinc sheets.

3. On the basis of the formulae derived, we can build a set of graphics, from which the optimum combination of number of small samples and area per small sample may easily be obtained.

ACKNOWLEDGMENTS

The authors wish to acknowledge the Flemish government and KU Leuven for the financial support and Union Miniere for the industrial support and help.

REFERENCES

Aziz, P.M. 1956. Application of the statistical theory of extreme values to the analysis of maximum pit depth data for aluminum. *Corrosion* 12: 495-506.

Galambos, J. 1987. *The asymptotic theory of extreme order statistics*. Malabar, Fla.: Roberte Krieger Publishing Company.

Gnedenko, B.V. 1943. Sur la distribution limite du terme maximum d'une serie aléatoire. *Ann. Math.* 44:423. (Cited from Gumbel, E.J. 1967. *Statistics of extremes.* New York and London: Columbia University Press.)

Gumbel, E.J. 1967. *Statistics of extremes*. New York and London: Columbia University Press.

JSCE 60-1 Technical Committee. 1988. *Corrosion Engineering* (Japan) 37:699. (Cited from Shibata, T. 1991. Evaluation of corrosion failure by extreme value statistics. *ISIJ International* 31(2): 115-121.)

Laycock, P., R.A. Cottis, and P.A. Scarf. January 1990. Extrapolation of extreme pit depths in space and time. *J. Electrochem. Soc.* 137(1): 64-69.

Lieblein, J. 1954. A new method of analyzing extreme-value data. National Advisory Committee For Aeronautics, Tech. Note 3053, January.

Shibata, T. 1991. Evaluation of corrosion failure by extreme value statistics. *ISIJ International* 31(2): 115-121.

Shibata, T. 1994. Application of extreme value statistics to corrosion. *Journal of Research of the National Institute of Standards and Technology* 99(4): 327-336.

Shibata, T. 1996. W.R. Whitney Award Lecture: Statistical and stochastic approaches to localized corrosion. *Corrosion* 52(11): 813-830.

Tsuge, H. 1987. Life estimation of corroded material by using the statistics of extreme. *J. Soc. Mater. Sci. Jpn.* 36(4): 35-46.

White, J.S. 1964. Least squares unbiased censored linear estimation for the log Weibull (extreme value) distribution. *Industrial Mathematics* 14: 21-60.

Zheng, R. 1999. Investigation on the corrosion of zinc sheeting of the highly insulated zinc roofs in the VLIET test building. Internal report, Laboratory of Building Physics, KU Leuven, Belgium.