



Models and inference for corrosion pit depth data

Isogai, Takafumi

Katano, Yukio

Miyata, Kenji

(Citation)

Extremes, 7(3):253-270

(Issue Date)

2004-07

(Resource Type)

journal article

(Version)

Accepted Manuscript

(URL)

<https://hdl.handle.net/20.500.14094/90000136>



Models and inference for corrosion pit depth data

TAKAFUMI ISOGAI*, YUKIO KATANO AND KENJI MIYATA**

*FACULTY OF MARITIME SCIENCES,

KOBE UNIVERSITY,

FUKAE, HIGASHINADA, KOBE 658-0022, JAPAN

E-MAIL:ISOGAI@MARITIME.KOBE-U.AC.JP

AND

*KUBOTA CORPORATION PIPE DIVISION,

SHIKITSUHIGASHI, NANIWA, OSAKA 556-8601, JAPAN

Key words and phrases: maximum value; point process; regression;
Generalized Extreme Value; generalized Gamma; small sample

ABSTRACT A new regression structure is introduced to several families of distributions, including the Generalized Extreme Value (GEV) distribution, which are available to explain the distribution of the maximum pit depth of a small sample. This regression structure depends on sample sizes, but under a simple assumption it is also available in the case of unknown sample sizes. GEV with the proposed regression structure provides us with a new extrapolation model for time. Compared to ordinary GEV models, effectiveness of our GEV model is demonstrated for our real data set. Also, comparisons are made with GEV and other two families of distributions with respect to the proposed regression structure.

1. INTRODUCTION

The soil corrosion is a major cause for troubles of pipelines. The process of soil corrosion should be explained in the context of electrochemical reactions. It is not easy to describe the mechanism of process clearly because of complex soil conditions. Thus we should observe a lot of samples with the soil corrosion in order to find out a statistical law for description of such a complex corrosion phenomenon.

A greater part of our data sets was collected as follows: In each sampling site we first decided an investigation domain of the surface of an underground pipe. After inspection of the domain we only recorded the maximum pit depth of corrosion with no reference of its sample size, and environmental factors around the corrosion giving the maximum pit depth. For these collected data sets we fitted regression models based on some generalized Gamma distribution.

Diagnostics of residuals and the results of goodness-of-fit test provided us with the evidence that the maximum depth of corrosion pits often seems to be distributed as the Lognormal distribution (see Katano et al.(1995) and Katano et al.(1996)). This evidence is against our knowledge that the maximum value is distributed as the Gumbel distribution.

Recently we have had a chance to observe the occurrence of corrosion according to some systematic sampling plan concerning 82 sites, which is referred as Area C in Katano et al.(1996). In this sampling plan we divided an investigation domain of the surface of a pipe into several unit subdomains, and for each unit subdomain we recorded all values of pit depths of the corrosion and environmental factors. This survey enables us not only to obtain a statistical distribution of the depth of a corrosion pit but also to examine a statistical distribution of the maximum value of pit depths.

The most interesting feature of the data in Area C is that more than 80% of 82 sampling sites have their sample sizes less than twenties, which indicates that in most cases of our surveys the maximum value of corrosion pits is not obtained from a large sample. Both this fact and our experience raise a question whether or not the maximum value of pit depths is distributed as the Gumbel distribution under a small sample size.

Now let us denote the depth of a corrosion pit by a positive random variable Y . When we analyzed all observations of pit depths in Area C, we used a model

$$\frac{\log Y - \mu}{\sigma} = \sqrt{k}(W - \log k), \quad (1.1)$$

where μ , σ (> 0) and k (> 0) are respectively unknown location, scale and shape parameters, and μ is expressed as

$$\mu = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \dots + \beta_p z_p + \gamma \log t \quad (1.2)$$

with respect to covariates z_i 's and time t and unknown parameters β_i 's and γ . Also, the random variable W has a density

$$\frac{\exp(kw - e^w)}{\Gamma(k)}. \quad (1.3)$$

Thus Y has the density function $f_\theta(y)$ of the generalized Gamma distribution as follows:

$$f_\theta(y) = \frac{k^k \left(\frac{y}{\eta}\right)^{\frac{\sqrt{k}}{\sigma}} \exp\left\{-k\left(\frac{y}{\eta}\right)^{\frac{1}{\sigma\sqrt{k}}}\right\}}{\sigma\sqrt{k}\Gamma(k)y}, \quad (1.4)$$

where for simplicity we put three parameters η ($= \exp \mu$), σ and k into a vector form $\theta = (\eta, \sigma, k)$. We also denote the distribution function of Y by $F_\theta(y)$.

Note that the distribution of Y includes as special cases the Weibull ($k = 1$) and the Gamma ($\sigma = 1$), and also includes the Lognormal as a limiting case ($k = \infty$).

Though Lawless (1980,1982) has introduced the model (1.1) by reparametrizing the generalized Gamma distribution, we show that the model (1.1) with (1.2) is naturally derived from some compound point process in Section 2. We first introduce the compound point process to describe a statistical distribution of the growth of a pit depth, but we can not deal with it directly due to

our observation process of a pit depth. Therefore, we derive our model (1.1) as an approximation for it.

Let $Y_1, Y_2, \dots, Y_n$ be i.i.d. positive random variables with common distribution function $F_\theta(y)$ and let $Y_{(n)}$ be their maximum order statistic $\max(Y_1, Y_2, \dots, Y_n)$. The distribution function of $Y_{(n)}$ is given by $F_\theta^n(y)$.

In Section 3 we examine which of Lognormal and three types of extreme value distributions (Gumbel, Fréchet and Weibull) can approximate $F_\theta^n(y)$ well for small sample cases ($2 \leq n \leq 50$). The degree of approximation is evaluated by the degree of linearity of configurations with respect to relevant probability plotting methods as well as by the Kullback-Leibler distance between $F_\theta^n(y)$ and each of the other four distributions.

In Section 4, to carry out regression analysis by using Lognormal and three extreme value distributions instead of $F_\theta^n(y)$, we examine the dependence structure of parameters of these four distributions on $\theta = (\eta, \sigma, k)$ and a sample size n with respect to $F_\theta^n(y)$. We solve the problem in a more general setting including the Generalized Extreme Value and other related distributions.

In the final section 5, by using maximum pit depths of unit subdomains in Area C, we examine performance of regression models proposed in the present paper as well as some regression models in a general use.

2. Construction of the distribution of a pit depth

2.1. A descriptive model for a pit depth

For the depth of a corrosion pit on underground pipes, we do not have a concrete electrochemical model to describe its growth-rate. However, when a corrosion pit is not extremely local and its growth-rate is not extremely rapid, from a microscopic point of view, the growth-rate of film thickness with metal oxides seems to approximately reflect the growth-rate of a pit depth.

When the film thickness of metal oxides is not so very thin, it is well known that under some appropriate conditions the rate of growth with thickness follows a simple power-law (for example, the rectilinear law, the parabolic law or the cubic law), which is generally expressed in the form

$$\frac{da}{dt} = C \cdot \frac{s}{a^\rho}, \quad (2.1)$$

where $a = a(t)$ is the film thickness at time t , $s = s(t)$ (≥ 0) is the stress amplitude at time t , and C (> 0) and ρ (≥ 0) are constant terms depending on the metallic material (for the details, see Evans (1960)). We may use the equation (2.1) as a roughly approximation to the rate of growth of a pit depth on pipelines. The assumption of a power-law growth of a pit depth is common in the corrosion literature, and especially this assumption is often made on the growth of the mean maximum pit depth (for example, see Scarf and Laycock (1994)).

Solving equation (2.1) yields

$$\frac{1}{\rho + 1} a^{\rho+1} = C \int_0^t s(u) du, \quad (2.2)$$

where we put $a(0) = 0$ for the initial value. What type of the stress function $s(t)$ can we use in (2.2)? We need a stress function such as describes small and large pit depths. It often happens that there is no pit after many years have passed, which suggests $s(t) = 0$. As it is difficult for us to specify the functional form of $s(t)$ on the whole time interval $[0, t)$, we shall consider a situation where stress happens intermittently. Thus, $s(t)$ may be expressed as

$$s(t) = \sum_{i=1}^{\infty} u_i \delta(t - t_i) \quad (2.3)$$

for $t_i > 0$ ($i = 1, 2, \dots$), where $\delta(t)$ is the Dirac delta function and t_i ($i = 1, 2, \dots$) indicates the time when stress u_i (≥ 0) occurs. Inserting (2.3) into (2.2), the value of integral is $\sum_{i:t_i < t} u_i$, where the summation is carried out with respect to the index i satisfying $t_i < t$.

Here we introduce randomness with respect to u_i and t_i , and suppose that u_i 's are realized values of random variables U_i 's, which are independently and identically distributed as an exponential distribution with unit mean (or unit rate), and that t_i 's are realized occurrence time of some counting process $N(t)$ independent of U_i 's. Then, (2.2) is expressed as

$$a^{\rho+1} = \begin{cases} C' \sum_{i=1}^{N(t)} U_i & \text{for } N(t) \geq 1, \\ 0 & \text{for } N(t) = 0, \end{cases} \quad (2.4)$$

where we put $\Pr(N(0) = 0) = 1$ and $C' = (\rho + 1)C$.

Though the equation (2.4) is a basic model for our corrosion problem, this equation does not reflect our actual observation process. To prevent an unexpected penetration accident due to pits on pipelines, our inspection is carried out once after many years since pipes were laid under the ground. Thus our observation consists of a pit depth $a(t)$ at some time point t , which usually means a long time interval $[0, t)$. Thus t_i 's and u_i 's are unobservable.

To adjust the equation (2.4) to our situation, we rewrite it as follows.

$$a^{\rho+1} = C' \cdot t \cdot \frac{N(t)}{t} \cdot \frac{\sum_{i=1}^{N(t)} U_i}{N(t)}. \quad (2.5)$$

Here, we suppose that for large t

$$\frac{N(t)}{t} \rightarrow \text{some } \xi (> 0) \text{ as } t \rightarrow \infty. \quad (2.6)$$

Note that ξ has an interpretation as an average occurrence rate of stress u , which seems to explain effects of environmental factors around pipes. Furthermore, due to unobservable t_i 's and u_i 's the process $(\sum_{i=1}^{N(t)} U_i / N(t))$ has no its history in the interval $[0, t)$, and the value of $(\sum_{i=1}^{N(t)} U_i / N(t))$ we can use is only one observed value obtained at time t when our inspection was carried out. This means that we can only use the marginal distribution of $(\sum_{i=1}^{N(t)} U_i / N(t))$ at time t .

Taking logarithms of both sides of (2.5) and using (2.6), we have

$$\log Y = \varphi + \gamma \log t + \gamma \log \frac{\sum_{i=1}^{N^*} U_i}{N^*}, \quad (2.7)$$

where we put

$$\begin{aligned} Y &= a(t), \quad N^* = N(t), \\ \varphi &= (\log C' + \log \xi)/(\rho + 1), \\ \gamma &= 1/(\rho + 1). \end{aligned} \quad (2.8)$$

Note that γ has a restriction $0 < \gamma \leq 1$ because of $\rho \geq 0$.

We shall consider the equation (2.7) as a descriptive model for the depth of a corrosion pit. The random variable Y denotes a pit depth at time t , and N^* denotes the total number of times that stress occurred in the interval $[0, t)$.

2.2. An approximation model for the distribution of a pit depth

Due to an unobservable nature of N^* it is difficult to deal with the distribution of the model (2.7) directly. We shall try to approximate the distribution of the model (2.7). Let us denote the distribution of $\log(\sum_{i=1}^{N^*} U_i/N^*)$ by $\mathfrak{L}(\log(\sum_{i=1}^{N^*} U_i/N^*))$, and also denote the conditional distribution of $\log(\sum_{i=1}^{N^*} U_i/N^*)$ given $N^* = m$ by $\mathfrak{L}(\log(\sum_{i=1}^{N^*} U_i/N^*) \mid N^* = m)$. Then, we have

$$\mathfrak{L}(\log \frac{\sum_{i=1}^{N^*} U_i}{N^*}) = \sum_{m=0}^{\infty} \mathfrak{L}(\log \frac{\sum_{i=1}^{N^*} U_i}{N^*} \mid N^* = m) P(N^* = m), \quad (2.9)$$

where we put $\mathfrak{L}(\log(\sum_{i=1}^{N^*} U_i/N^*) \mid N^* = 0) = 1$ for convenience.

Given $N^* = m$ ($m \geq 1$), $\mathfrak{L}(\log(\sum_{i=1}^{N^*} U_i/N^*) \mid N^* = m)$ in (2.9) is expressed as $\tau_m \sqrt{m} \log(\sum_{i=1}^m U_i/m)$ with $\tau_m = 1/\sqrt{m}$. When m is large, $\sqrt{m} \log(\sum_{i=1}^m U_i/m)$ is asymptotically distributed as the standard normal distribution. On the other hand, when m is small, the distribution of $\sqrt{m} \log(\sum_{i=1}^m U_i/m)$ is positively skewed. The distributional shape of $\mathfrak{L}(\log(\sum_{i=1}^{N^*} U_i/N^*))$ depends on a probability that $N^* = m$ concentrates. Thus a simple way to approximate the distribution $\mathfrak{L}(\log(\sum_{i=1}^{N^*} U_i/N^*))$ is to approximate the distribution of $\mathfrak{L}(\log(\sum_{i=1}^{N^*} U_i/N^*) \mid N^* = m)$, that is, $\tau_m \sqrt{m} \log(\sum_{i=1}^m U_i/m)$, of which event $\{N^* = m\}$ has a large probability.

For this purpose we shall consider the distribution of $\tau \sqrt{k} \log(\sum_{i=1}^k U_i/k)$, denoted by $\mathfrak{L}(\tau \sqrt{k} \log(\sum_{i=1}^k U_i/k))$, with free parameters k (> 0) and τ (> 0), and approximate $\mathfrak{L}(\log(\sum_{i=1}^{N^*} U_i/N^*))$ by $\mathfrak{L}(\tau \sqrt{k} \log(\sum_{i=1}^k U_i/k))$. We expect that the parameter k will be able to detect m such that the event $\{N^* = m\}$ has a large probability, and that the parameter τ will adjust τ_m when several events close to m have relatively large probabilities. In practice k and τ are estimated by maximum likelihood principle.

Define the following new compound counting process $\widetilde{G}_k$:

$$\widetilde{G}_k = \sum_{m=0}^{\infty} \mathfrak{L}(\tau \sqrt{k} \log(\sum_{i=1}^k U_i/k) \mid N^* = m) P(N^* = m), \quad (2.10)$$

where $\mathfrak{L}(\tau \sqrt{k} \log(\sum_{i=1}^k U_i/k) \mid N^* = m)$ ($m \geq 1$) is $\mathfrak{L}(\tau \sqrt{k} \log(\sum_{i=1}^k U_i/k))$ independent of m , and we put $\mathfrak{L}(\tau \sqrt{k} \log(\sum_{i=1}^k U_i/k) \mid N^* = 0) = 1$ for convenience.

The accuracy of approximation is examined by the variational distance. Let us denote densities of $\mathfrak{L}(\log(\sum_{i=1}^{N^*} U_i/N^*))$ ($= G_{N^*}$, say), $\mathfrak{L}(\log(\sum_{i=1}^{N^*} U_i/N^*) \mid N^* = m)$ ($= G_m$, say), and $\mathfrak{L}(\tau \sqrt{k} \log(\sum_{i=1}^k U_i/k))$ ($= G_k$, say) by g_{N^*} , g_m and g_k respectively. The variational distance between G_{N^*} and $\widetilde{G}_k$ is given by

$$\|G_{N^*} - \widetilde{G}_k\| = \frac{1}{2} \int_{-\infty}^{\infty} \left| \sum_{m=1}^{\infty} (g_m(x) - g_k(x)) P(N^* = m) \right| dx, \quad (2.11)$$

where we put $g_m(x) - g_k(x) = 0$ for $m = 0$. To evaluate the right hand of (2.11), we use the following inequality.

$$\|G_{N^*} - \widetilde{G}_k\|^2 \leq \sum_{m=1}^{\infty} KL(G_m, G_k) P(N^* = m), \quad (2.12)$$

where $KL(G_m, G_k)$ ($m > 0$) means the Kullback-Leibler distance between G_m and G_k , which is given by

$$KL(G_m, G_k) = \int_{-\infty}^{\infty} \log \frac{g_k(x)}{g_m(x)} g_k(x) dx$$

For the details of the inequality (2.12), see Reiss (1993).

As it is very difficult to analytically obtain values of k and τ that reduce $\|G_{N^*} - \widetilde{G}_k\|$ as small as possible, we use the inequality (2.12) to minimize the upper bound with respect to k and τ . Here we have to suppose the distribution of N^* . From our data sets we cannot extract any information about N^* . Thus it would not be unreasonable to suppose that N^* is the Poisson random variable with mean ξ . Put the right hand of (2.11), $(1/2) \int_{-\infty}^{\infty} \left| \sum_{m=1}^{\infty} (g_m(x) - g_k(x)) P(N^* = m) \right| dx = Q_1(k, \tau)$, and put the right hand of (2.12), $\sum_{m=1}^{\infty} KL(G_m, G_k) P(N^* = m) = Q_2(k, \tau)$. We first minimize $Q_2(k, \tau)$ to estimate τ and k . Then, by using those estimated values of τ and k we evaluate $Q_1(k, \tau)$. The results are given in Table 1.

Table 1

The results of Table 1, especially values of $Q_1(k, \tau)$, suggest that our simple approximation $\widetilde{G}_k$ to G_{N^*} is not so bad, because the value of $\min Q_2(k, \tau)$ gives us a fairly rough upper bound. Replacing the term with G_{N^*} by the corresponding term with $\widetilde{G}_k$ in (2.7), we have our model (1.1) again as

$$\log Y = \mu + \sigma \sqrt{k} \log \frac{\sum_{i=1}^k U_i}{k}, \quad (2.13)$$

where we put

$$\mu = \varphi + \gamma \log t,$$

$$\sigma = \tau \gamma.$$

We note that in estimating k we need much iterated computation to obtain a solution of large k . From our experience the convergence is very slow, and when an estimated value of k is greater than 15, our maximum likelihood function has almost no gain. In the present paper, our interest is to classify a distribution of Y rather than to estimate k . Thus we use the parameter k as an indicator to specify the distribution of Y : Weibull ($k = 1$), nearly Lognormal (k is greater than 15) and an intermediate distribution ($15 > k > 1$). For an attempt to remedy a difficulty in estimation of k , see Lawless (1980). Also, for related topics, see Young and Bakir (1987).

3. Distributions for the maximum pit depth

The distribution functions of Gumbel, Fréchet, Weibull and Lognormal considered here are the following:

- (a) Gumbel: $H_G(y) = \exp[-\exp\{-(y - \lambda)/\tau\}] (= K_G[(y - \lambda)/\tau], \text{say});$
 - (b) Fréchet: $H_F(y) = \exp[-\exp\{-(\log y - \lambda)/\tau\}] (= K_F[(\log y - \lambda)/\tau], \text{say});$
 - (c) Weibull: $H_W(y) = 1 - \exp[-\exp\{(\log y - \lambda)/\tau\}] (= K_W[(\log y - \lambda)/\tau], \text{say});$
 - (d) Lognormal: $H_{LN}(y) = \Phi[(\log y - \lambda)/\tau] (= K_{LN}[(\log y - \lambda)/\tau], \text{say}),$
- where λ and $\tau (> 0)$ are location and scale parameters respectively, and Φ denotes the standard normal distribution function.

Our concern is to evaluate the closeness between $F_\theta^n(y)$ and one of the above distributions. For this purpose, by using K functions defined as above, we transform $F_\theta^n(y)$ to $T(F_\theta^n(y))$ as follows:

- (a) for the Gumbel, $T_G(F_\theta^n(y)) = K_G^{-1}(F_\theta^n(y)) = -\log\{-\log F_\theta^n(y)\};$
- (b) for the Fréchet, $T_F(F_\theta^n(y)) = K_F^{-1}(F_\theta^n(y)) = -\log\{-\log F_\theta^n(y)\};$
- (c) for the Weibull, $T_W(F_\theta^n(y)) = K_W^{-1}(F_\theta^n(y)) = \log\{-\log(1 - F_\theta^n(y))\};$
- (d) for the Lognormal, $T_{LN}(F_\theta^n(y)) = K_{LN}^{-1}(F_\theta^n(y)) = \Phi^{-1}(F_\theta^n(y)),$ where Φ^{-1} denotes the inverse of the standard normal distribution function Φ .

If $F_\theta^n(y)$ is equal to the Gumbel distribution $H_G(y)$, then $T_G(F_\theta^n(y)) = (y - \lambda)/\tau$ holds. This is the spirit of the Gumbel probability plotting method, which is to check whether a linear relationship $T_G(F_\theta^n(y)) = Ay + B$ with parameters A and B independent of y holds for the transformation $T_G(F_\theta^n(y))$. Its strength of linearity ensures the closeness of $F_\theta^n(y)$ to $H_G(y)$. Similarly, Lognormal, Weibull and Fréchet probability plotting methods are to check whether relationships $T_{LN}(F_\theta^n(y)) = A \log y + B$, $T_W(F_\theta^n(y)) = A \log y + B$ and $T_F(F_\theta^n(y)) = A \log y + B$ hold for transformations $T_{LN}(F_\theta^n(y))$, $T_W(F_\theta^n(y))$ and $T_F(F_\theta^n(y))$ respectively.

The above probability plotting methods visually examine linearity of the graph $(y, T(F_\theta^n(y)))$ or $(\log y, T(F_\theta^n(y)))$ for a given transformation T . It is well known that for a smooth graph $(\alpha(y), \beta(y))$ the degree of linearity at y is evaluated by the value of its curvature:

$$\kappa(y) = \frac{\dot{\alpha}(y)\ddot{\beta}(y) - \ddot{\alpha}(y)\dot{\beta}(y)}{(\dot{\alpha}(y)^2 + \dot{\beta}(y)^2)^{3/2}}, \quad (3.1)$$

where $\dot{\alpha}, \dot{\beta}$ and $\ddot{\alpha}, \ddot{\beta}$ denote the first and second derivatives of α and β respectively. The smaller the absolute value of $\kappa(y)$ becomes, the more the graph $(\alpha(y), \beta(y))$ approaches a straight line in a neighborhood of y . Note that $\kappa(y)$ is scale invariant under transformation $y \rightarrow y/\eta$.

To evaluate linearity of the graph $(y, T(F_\theta^n(y)))$ or $(\log y, T(F_\theta^n(y)))$ on the whole domain of y , we introduce the following measure of linearity:

$$\Gamma = \int_0^\infty |\kappa(y)| dF_\theta^n(y), \quad (3.2)$$

where $\alpha(y) = y$ or $\log y$ and $\beta(y) = T(F_\theta^n(y))$ in $\kappa(y)$ of (3.1). The numerical integration is carried out over the interval $[y_L, y_U]$ such that $F_\theta^n(y_L) = 0.00001$ and $F_\theta^n(y_U) = 0.99999$.

Because of scale invariant property of $\kappa(y)$, we have only to examine Γ of (3.2) for the case: $\eta = 1$ in (1.4). For the other parameters k and σ in (1.4), we set $k = 0.1, 0.5, 1, 5, 15$ and $\sigma = 0.1, 0.3, 0.5, 0.8$. These upper and lower values for k and σ are selected to include the results of data analysis of several sampling areas as well as Area C. In cases of $2 \leq n \leq 50$ we have obtained the distribution that gives the minimum value of Γ among Gumbel, Lognormal, Fréchet and Weibull distributions.

Apart from the mean curvature Γ , we have also examined the Kullback-Leibler distance $KL(H, F_\theta^n)$ between $F_\theta^n(y)$ and each of four distributions (Gumbel, Fréchet, Weibull and Lognormal), where H means one of those distributions.

Roughly speaking, in small sample cases ($2 \leq n \leq 50$) our results are as follows. (i) When k is large ($k \geq 15$), $F_\theta^n(y)$ is near the Lognormal for the case $\sigma < 0.3$, and $F_\theta^n(y)$ is near the Gumbel for the case $\sigma \geq 0.3$. (ii) When $k = 1$, $F_\theta^n(y)$ is near the Lognormal. (iii) When k is small ($0 < k < 1$), $F_\theta^n(y)$ is near the Weibull or the Lognormal. (iv) When k is intermediate ($1 < k < 15$), $F_\theta^n(y)$ is near the Lognormal for the case $\sigma < 0.5$, and $F_\theta^n(y)$ is near the Gumbel for the case $\sigma \geq 0.5$.

4. Models for the maximum pit depth

The above results indicate that under a small sample case $F_\theta^n(y)$ is often well approximated by Lognormal and Weibull distributions rather than by the Gumbel. Here our concern is to examine the dependence structure of parameters λ and τ on $\theta (= (\eta, \sigma, k))$ and a sample size n . For each of Lognormal and three extreme value distributions, such a relationship is easily and approximately obtained by the idea of regression coefficients through linearity of $T(F_\theta^n(y))$ on y or $\log y$.

The essential point of our approach is to use a K function that defines a distribution function of a special form such as $K[(y - \lambda)/\tau]$ or $K[(\log y - \lambda)/\tau]$. Therefore, in order to deal with our problem in a unified way, we have only to use more systematic families of distributions, which include some of Lognormal and three extreme value distributions as well as have a K function of the special form.

Examples of families of distributions are as follows.

(1) The Generalized Extreme Value (GEV) distribution including Gumbel, Weibull, Fréchet distributions, of which distribution function is given by

$$H_{GEV}(y) = \exp \left\{ - \left[1 - \delta \left(\frac{y - \lambda}{\tau} \right) \right]^{1/\delta} \right\} \quad (4.1)$$

$$(\text{= } K_{GEV}[(y - \lambda)/\tau], \text{ say}),$$

where δ is a shape parameter and $\delta y \leq \tau + \delta \lambda$.

(2) The new power normal family (see Isogai (1999)) including Lognormal, Weibull and Fréchet distributions, of which distribution function is given by

$$H_{NPN}(y) = \Phi \left[\frac{1}{\delta} \left(\exp \left\{ \delta \left(\frac{\log y - \lambda}{\tau} \right) \right\} - \exp(-\delta^2) \right) \right] \quad (4.2)$$

$$(\text{= } K_{NPN}[(\log y - \lambda)/\tau], \text{ say}),$$

where δ ($-1/3 < \delta < 1/3$) is a shape parameter.

The new power normal family is based on the Wilson-Hilferty transformation of the central F distribution, and its random variable Y is defined by

$$(Y / \exp \lambda)^{\delta/\tau} = \left(F_{2\phi_2}^{2\phi_1} \right)^h, \quad (4.3)$$

where $F_{2\phi_2}^{2\phi_1}$ ($\phi_1 > 1, \phi_2 > 1$) is a random variable of the F distribution with $2\phi_1, 2\phi_2$ degrees of freedom and we put

$$\delta = (1/3) (1/\phi_1 - 1/\phi_2) / (1/\phi_1 + 1/\phi_2)^{1/2},$$

$$h = (1/3) (1/\phi_1 - 1/\phi_2) / (1/\phi_1 + 1/\phi_2).$$

The power parameter h works so as to make the third order cumulant of $F_{2\phi_2}^{2\phi_1}$ close to zero. After substituting the right hand of (4.3) into the following Wilson-Hilferty transformation formula of $F_{2\phi_2}^{2\phi_1}$

$$\left[\left(F_{2\phi_2}^{2\phi_1} \right)^h - \exp\{-\delta^2\} \right] / \delta,$$

δ is treated as a free shape parameter in distribution fitting problems. Thus, δ is chosen so as to make the third order cumulant of Y close to zero. We remark that as $\delta \rightarrow 0$, $H_{NPN}(y)$ tends to the Lognormal $H_{LN}(y)$; as $\delta \rightarrow 1/3$ (i.e. $\phi_1 \rightarrow 1, \phi_2 \rightarrow \infty$), $H_{NPN}(y)$ tends to the Weibull $H_W(y)$; and as $\delta \rightarrow -1/3$ (i.e. $\phi_1 \rightarrow \infty, \phi_2 \rightarrow 1$), $H_{NPN}(y)$ tends to the Fréchet $H_F(y)$.

The formulas for the dependence structure of λ and τ on θ ($= (\eta, \sigma, k)$) and n are given by

$$\tau = \frac{V_\theta(Y_{(n)})}{\text{Cov}_\theta(Y_{(n)}, T(F_\theta^n(Y_{(n)})))}, \quad (4.4)$$

$$\lambda = E_\theta[Y_{(n)}] - \tau E_\theta[T(F_\theta^n(Y_{(n)}))] \quad (4.5)$$

for distributions of the $K[(y - \lambda)/\tau]$ type, and

$$\tau = \frac{V_\theta(\log Y_{(n)})}{Cov_\theta(\log Y_{(n)}, T(F_\theta^n(Y_{(n)})))}, \quad (4.6)$$

$$\lambda = E_\theta[\log Y_{(n)}] - \tau E_\theta[T(F_\theta^n(Y_{(n)}))] \quad (4.7)$$

for distributions of the $K[(\log y - \lambda)/\tau]$ type, where $E_\theta[\cdot]$, $V_\theta(\cdot)$ and $Cov_\theta(\cdot, \cdot)$ denote expectation, variance and covariance operators with respect to $F_\theta^n(y)$ respectively.

To obtain explicit expressions of τ and λ , we utilize asymptotic results concerning $F_\theta^n(y)$, the distribution of $Y_{(n)}$. We can approximate $F_\theta^n(y)$ in the expressions (4.4)-(4.7) by $G((y - a_n)/b_n) = \exp[-\exp\{-(y - a_n)/b_n\}]$, where $G(y)$ is the Gumbel distribution $G(y) = \exp\{-\exp(-y)\}$, and sequences a_n and b_n satisfy $F_\theta(a_n) = 1 - 1/n$ and $b_n = (1 - F_\theta(a_n))/f_\theta(a_n)$ (for example, see Takahashi (1987)).

Put $Z = G((Y_{(n)} - a_n)/b_n)$. Then, expressions (4.4) and (4.5) can be approximately rewritten in terms of the uniform random variable Z as follows.

$$\tau \doteq b_n C_1, \quad (4.8)$$

$$\lambda \doteq a_n + b_n C_2, \quad (4.9)$$

where we put

$$\begin{aligned} C_1 &= \frac{V(-\log(-\log Z))}{Cov(-\log(-\log Z), T(Z))}, \\ C_2 &= E[-\log(-\log Z)] - \frac{V(-\log(-\log Z))}{Cov(-\log(-\log Z), T(Z))} E[T(Z)] \end{aligned}$$

and $E[\cdot]$, $V(\cdot)$ and $Cov(\cdot, \cdot)$ denote expectation, variance and covariance operators with respect to Z respectively. Note that usually $C_1 > 0$, because the pair $(-\log(-\log Z), T(Z))$ has the positive quadrant dependent property (see, Lehmann (1966)).

Following Villasenor's procedure (1976) (also, see Galambos (1978)), we have

$$\log a_n \sim \log \eta + \sigma \sqrt{k} \log(1 + \log n) \quad (4.10)$$

for large n , and similarly,

$$\frac{b_n}{a_n} \sim \frac{\sigma \sqrt{k}}{1 + \log n}. \quad (4.11)$$

Combining (4.10) and (4.11), expressions (4.8) and (4.9) become

$$\tau \sim \eta \sigma \sqrt{k} C_1 (1 + \log n)^{\sigma \sqrt{k} - 1},$$

$$\lambda \sim \eta (1 + \log n)^{\sigma\sqrt{k}} \left(1 + \frac{\sigma\sqrt{k}}{1 + \log n} C_2 \right).$$

Thus, for distributions of the $K[(y - \lambda)/\tau]$ type, we can parametrize

$$\tau = \eta \sigma^* (1 + \log n)^{\psi-1}, \quad (4.12)$$

$$\lambda = \eta (1 + \log n)^\psi. \quad (4.13)$$

where σ^* ($= \sigma\sqrt{k}C_1 > 0$) and ψ ($= \sigma\sqrt{k} > 0$) are new parameters, and a constant term β_0 in $\log \eta$ ($= \mu$) of (1.2) may be changed. Similarly, for distributions of the $K[(\log y - \lambda)/\tau]$ type, we can parametrize

$$\tau = \frac{\sigma^*}{1 + \log n}, \quad (4.14)$$

$$\lambda = \log \eta + \psi \log(1 + \log n), \quad (4.15)$$

Substituting (4.12) and (4.13) into λ and τ of the Generalized Extreme Value distribution (4.1), we have

$$H_{GEV}(y) = \exp \left\{ - \left[1 - \delta \left(\frac{\left(\frac{y}{\eta_n^*} \right) - 1}{\sigma_n^*} \right) \right]^{1/\delta} \right\}, \quad (4.16)$$

$$\delta y \leq \eta_n^* (\sigma_n^* + \delta)$$

where we put $\eta_n^* = \eta(1 + \log n)^\psi$ and $\sigma_n^* = \sigma^*/(1 + \log n)$. Similarly, for each of Lognormal and three extreme value distributions, we have

(a) Gumbel: $H_G(y) = \exp[-\exp\{-\frac{(\frac{y}{\eta_n^*})-1}{\sigma_n^*}\}];$

(b) Fréchet: $H_F(y) = \exp[-\left(\frac{y}{\eta_n^*}\right)^{-1/\sigma_n^*}];$

(c) Weibull: $H_W(y) = 1 - \exp[-\left(\frac{y}{\eta_n^*}\right)^{1/\sigma_n^*}];$

(d) Lognormal: $H_{LN}(y) = \Phi[\frac{1}{\sigma_n^*} \log\left(\frac{y}{\eta_n^*}\right)].$

We remark that from the above functional forms of $H_G(y)$ and $H_F(y)$, we can construct a power Gumbel family of distributions connecting the Gumbel and Fréchet as follows.

$$H_{PG}(y) = \exp[-\exp\{-\frac{1}{\sigma_n^*} \frac{\left(\frac{y}{\eta_n^*}\right)^\delta - 1}{\delta}\}], \quad (4.17)$$

$$y \geq 0$$

where δ is a shape parameter. We are interested in the positive value of δ . The Gumbel corresponds to $\delta = 1$ and as $\delta \rightarrow 0$, $H_{PG}(y)$ tends to the Fréchet $H_F(y)$. When $\delta > 0$, in the case of $\delta > 1$ the corresponding distribution $H_{PG}(y)$ has a shorter tail than the Gumbel and in the case of $0 < \delta < 1$ $H_{PG}(y)$ has a longer tail than the Gumbel.

5. Analysis of the data of Area C

First we check the distribution of pit depths by applying the generalized Gamma (denoted by GG) model (1.1) with (1.2) to all 1487 observed pit depths of Area C. There are 18 covariates (i.e. exposure time and environmental factors around corrosion pits), and also there is no penetration in our data. The result are as follows: the minimum value of the negative log likelihood function (denoted by SSE) = 1506.6, its degrees of freedom (denoted by d.f.) = 1467, $\hat{k} = 23.3(0.95)$, $\hat{\sigma} = 0.33(0.006)$, $\hat{\gamma} = 0.53(0.11)$, AIC = 3053.2, where numbers in brackets denote approximate standard errors of the corresponding estimates, and AIC is the well-known Akaike's Information Criterion.

To diagnose the degree of goodness-of-fit with respect to GG model, we examined a relevant probability plot. For an easy comparison between GG and GEV models, we transformed observed pit depths to exponential quantiles by using estimated GG. An exponential probability plot of GG is given in Figure 1(a). There seems no peculiarity.

The value of $\hat{k} = 23.3$ affirms lognormality of the distribution of pit depths in Area C. The value of $\hat{\gamma} = 0.53$ is also satisfied with a restriction $0 < \gamma \leq 1$ derived from its definition (2.8). Our experimental and field studies of corrosion pits on underground pipes have confirmed that $\hat{\gamma}$ lies between 0 and 1 (see Katano et al. (1988)).

Then we investigate the data of maximum pit depths of 364 unit subdomains in two aspects. One is to compare regression structures proposed in the previous section with ordinary regression structures in GEV models. The other is to compare performance of our proposed regression structures among different families of distributions. We note that 90% of 364 unit subdomains have sample sizes less than 9. Furthermore, sample sizes of 273 unit subdomains are less than 6. The maximum sample size is 30 (one sample), the second largest is 16, and the minimum is one.

For a simplicity of expressions, we put

$$\varphi = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \dots + \beta_p z_p,$$

where z_i 's denote covariates except for exposure time t .

GEV-O₁: the GEV distribution (4.1) with

$$\begin{aligned} \lambda &= \varphi + \gamma \log t, \\ \delta y &\leq \tau + \delta \lambda \end{aligned}$$

GEV-O₂: the GEV distribution (4.1) with

$$\begin{aligned} \lambda &= t^\gamma \exp(\varphi), \\ \delta y &\leq \tau + \delta \lambda \end{aligned}$$

GEV-O₃: the GEV distribution (4.1) with

$$\begin{aligned}\lambda &= t^\gamma \exp(\varphi), \\ \tau &= \tau_0 t^\gamma \quad (\tau_0 > 0), \\ \delta y &\leq t^\gamma \{\tau_0 + \delta \exp(\varphi)\}\end{aligned}$$

GEV-N: the GEV distribution (4.1) with

$$\begin{aligned}\lambda &= t^\gamma \exp(\varphi), \\ \tau &= \sigma^* t^\gamma \exp(\varphi), \\ \delta y &\leq (\sigma^* + \delta) t^\gamma \exp(\varphi)\end{aligned}$$

GEV-U: the GEV distribution (4.1) with

$$\begin{aligned}\lambda &= (1 + \log n)^\psi t^\gamma \exp(\varphi), \\ \tau &= (1 + \log n)^{\psi-1} \sigma^* t^\gamma \exp(\varphi), \\ \delta y &\leq \{\sigma^*/(1 + \log n) + \delta\} (1 + \log n)^\psi t^\gamma \exp(\varphi)\end{aligned}$$

Here we remark that in the above models an additional notation ‘O’ means the use of an ‘ordinary’ regression model for GEV, and ‘U’ and ‘N’ mean the ‘use’ and ‘no use’ of sample sizes of the corresponding maximum pit depths with respect to our proposed regression models respectively.

GEV-O₁ and -O₂ correspond to models that we have applied to our data sets of maximum pit depths so far in the case of $\delta = 0$. GEV-O₃ means a model that Laycock et al.(1990) have adopted to extrapolate extreme pit depths in time. GEV-N corresponds to our model (4.16) in which the sample size n is set equal to one. This assumption explains our situation lacking in information of sample sizes. We are obliged to suppose that sample sizes are unknown but common. GEV-U is a model of our interest. Note that GEV-N and GEV-U provide us with some generalization of the extrapolation model for time given by Laycock et al. (1990).

Table 2

In Table 2 we show the results of maximum likelihood estimation with respect to the above five regression structures in GEV model. Table 2 displays SSE, AIC, estimates of shape and scale parameters and estimates of γ with their approximate standard errors.

The results in Table 2 are very different in estimates of δ and γ . Especially, values of $\hat{\delta}$ in GEV-O₁, -O₂ and -O₃ indicate that the distribution of maximum pit depths is the Gumbel. On the other hand, values of $\hat{\delta}$ in GEV-N and -U suggest that the distribution of maximum pit depths is the Weibull. Furthermore, values of $\hat{\gamma}$ in GEV-O₁ and -O₃ deviate from its restriction $0 < \gamma \leq 1$, but those of $\hat{\gamma}$ in GEV-O₂, -N and -U do not. Note that all values of $\hat{\delta}$ are positive, which coincides with a corrosion context.

When we carried out maximum likelihood estimation with respect to GEV models, in the case of $\delta > 0$ we had to take account of the left truncated property of GEV models introduced by the nature of our data (i.e. $y \geq 0$), and found that the effect of truncation was very small and negligible in most cases except for GEV-U.

Exponential probability plots of GEV-O₃, -N and -U are given in Figure 2. Evidently, the degree of goodness-of-fit with respect to GEV-N and -U is better than that of GEV-O₃ which have larger values of SSE. Serious peculiarities do not appear in plots of GEV-N and -U.

Figure 1

Figure 2

Next we apply other two different distributions, the new power Normal (denoted by NPN) and the power Gumbel (denoted by PG) families to analyze maximum pit depths of 364 unit subdomains. We consider the following four cases.

NPN-U: the new power Normal family (4.2) with

$$\begin{aligned}\lambda &= \varphi + \gamma \log t + \psi \log(1 + \log n), \\ \tau &= \sigma^* / (1 + \log n)\end{aligned}$$

NPN-N: the new power Normal family (4.2) with

$$\begin{aligned}\lambda &= \varphi + \gamma \log t, \\ \tau &= \sigma^*\end{aligned}$$

PG-U: the power Gumbel family (4.17) with

$$\begin{aligned}\eta_n^* &= (1 + \log n)^\psi t^\gamma \exp(\varphi), \\ \sigma_n^* &= \sigma^* / (1 + \log n)\end{aligned}$$

PG-N: the power Gumbel family (4.17) with

$$\begin{aligned}\eta_n^* &= t^\gamma \exp(\varphi), \\ \sigma_n^* &= \sigma^*\end{aligned}$$

Table 3

The results of the above four cases are given in Table 3. PG-U has the best performance among all models considered in the present paper. In maximum likelihood estimation with respect to PG models we had to take account of the left truncated property (i.e. $y \geq 0$) of $H_{PG}(y)$ for the case of $\delta > 0$. The

effect of truncation was very small and negligible in the case of PG-N, but not in the case of PG-U. An exponential probability plot of PG-U is also given in Figure 2. There seems no peculiarity.

The value of $\hat{\delta}$ in NPN-N indicates that the distribution of maximum pit depths is near the Weibull, and also $\hat{\delta}$ of PG-N suggests that the distribution of maximum pit depths deviates from the Gumbel and further from the Fréchet. Comparisons between ‘U’ and ‘N’ cases of GEV, NPN and PG models tell us that the use of sample sizes of maximum pit depths strengthens evidence for the shape of the distribution of maximum pit depths.

We note that NPN-N and PG-N models have almost the same performance as GEV-N in Table 2. Actually, their estimates for regression coefficients of covariates and their standard errors are almost the same as those of GEV-N. This fact means that we can always obtain similar results in application of GEV-N, NPN-N and PG-N models, because in practical fields we do not record sample sizes with respect to maximum values.

We also note that GEV-N, NPN-N and PG-N are available to analyze all 1487 observations of pit depths. The results are as follows.

GEV-N: SSE = 1505.2, d.f. = 1467, $\hat{\gamma} = 0.52(0.11)$, $\hat{\delta} = 0.05(0.018)$, $\hat{\sigma}^* = 0.33(0.006)$, AIC = 3050.4

NPN-N: SSE = 1506.4, d.f. = 1467, $\hat{\gamma} = 0.53(0.11)$, $\hat{\delta} = 0.07(0.022)$, $\hat{\sigma}^* = 0.34(0.006)$, AIC = 3052.8

PG-N: SSE = 1504.3, d.f. = 1467, $\hat{\gamma} = 0.53(0.11)$, $\hat{\delta} = 1.22(0.071)$, $\hat{\sigma}^* = 0.33(0.006)$, AIC = 3048.6

Including the result of GG in the first paragraph of the present section, these results are very similar as expected. Especially $\hat{\delta}$ in NPN-N indicates that the population of pit depths is near the Lognormal, and $\hat{\delta}$'s in GEV-N and PG-N suggest that the population is near the Gumbel or a little bit far from it. We give an exponential probability plot with respect to GEV-N in Figure 1(b). There is also no peculiarity.

6. Conclusions

Based on some generalized Gamma distribution proposed to describe the growth of a pit depth, a new regression structure is introduced to several families of distributions, including GEV, which are available to explain the distribution of the maximum pit depth of a small sample.

As for GEV with a new regression structure, GEV-U utilizes the sample size of a maximum pit depth, but GEV-N does not. Unlike ordinary GEV regression models, GEV-U and -N have two scale parameters. This structure naturally leads us to a new extrapolation model for time.

Our case study shows that the degree of goodness-of-fit with respect to GEV-U and -N is clearly superior to that with respect to ordinary GEV regression models. GEV-U has the best performance among all GEV models considered here. GEV-N has a little less performance than GEV-U. The use of sample sizes in GEV-U increases information on the shape of the GEV distribution.

In addition to GEV, other two families of distributions NPN (the new power normal family) and PG (the power Gumbel family) are also proposed to fit the distribution of the maximum pit depth of a small sample. Our case study shows that PG-U has the best performance among all. When we can not use any information about the sample size of a maximum pit depth, PG-N and NPN-N have the same performance as GEV-N.

Our choice of different families of distributions with a new regression structure depends on whether we can use the sample size of a maximum pit depth or whether we need a physical interpretation of real data such as the depth of a corrosion pit on pipelines.

7. Acknowledgement

The authors would like to thank two referees including the associate editor, and the editor for their helpful comments and suggestions, which have considerably improved an early draft of the present paper.

8. References

- Evans, U. R. *The Corrosion and Oxidation of Metals: Scientific Principles and Practical Applications*, Edward Arnold Publishers, 1960.
- Galambos, J. *The Asymptotic Theory of Extreme Order Statistics*, John Wiley & Sons, 1978.
- Isogai, T. "Power transformation of the F distribution and a power normal family", *Journal of Applied Statistics*, 26, 355-371, (1999).
- Katano, Y., Nishizaki, K. and Iwamatsu, J. "Corrosive Environment of Underground Pipes and Prediction of Corrosion", *Boshoku Gijutsu*, 37, 205-211, (1988), (in Japanese).
- Katano, Y., Miyata, K., Shimizu, K. and Isogai, T. "Examination of Statistical Models for Pitting on Underground Pipes and Data Analysis", *Proceedings of International Symposium on Plant Aging and Life Predictions of Corrodible Structures*, NACE International, Sapporo, 161-167, 1995.
- Katano, Y., Miyata, K., Shimizu, K. and Isogai, T. "Statistical Data analysis for Pitting on Underground Pipes", *Proceedings of 13th International Corrosion Congress: Toward Corrosion Prevention*, Session: Underground Pipes, Melbourne, 1996.
- Laycock, P. J., Cottis, R. A., and Scarf, P. A. "Extrapolation of Extreme Pit Depths in Space and Time", *Journal of the Electrochemical Society*, 137, 64-69, (1990).
- Lawless, J.F. "Inference in the Generalized Gamma and Log Gamma Distributions", *Technometrics*, 22, 409-426, (1980).
- Lawless, J.F. *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, 1982.
- Lehmann, E. L. "Some Concepts of Dependence", *Annals of Mathematical Statistics*, 37, 1137-1153, (1966).
- Reiss, R.D. *A Course on Point Processes*, Springer-Verlag, 1993.

Scarf, P.A. and Laycock, P.J. “Applications of Extreme Value Theory in Corrosion Engineering”, *Journal of Research of the National Institute of Standards and Technology*, 99, 313-320, (1994).

Takahashi, R. “Normalizing constants of a distribution which belongs to the domain of attraction of the Gumbel distribution”, *Statistics and Probability Letters*, 5, 197-200, (1987).

Villasenor, J. *On univariate and bivariate extreme value theory*, Ph.D. Thesis, Iowa State University, 1976.

Young, D.H. and Bakir, S.T. ‘Bias Correction for a Generalized Log-Gamma Regression model’, *Technometrics*, 29, 183-191, (1987).

Table 1. The values $Q_1(k, \tau)$ for the variational distance $\|G_{N^*} - \widetilde{G}_k\|$.

$P_o(\xi)$	estimated (τ, k)	$\min Q_2(k, \tau)$	$Q_1(k, \tau)$
$\xi = 1$	(0.795, 1.582)	0.0405	0.0246
$\xi = 5$	(0.445, 5.033)	0.0603	0.0361
$\xi = 10$	(0.316, 10.00)	0.0286	0.0177
$\xi = 15$	(0.258, 14.98)	0.0181	0.0115
$\xi = 20$	(0.223, 19.99)	0.0133	0.0093
$\xi = 30$	(0.182, 29.94)	0.0244	0.0066

Table 2. Maximum likelihood estimation with respect to GEV models

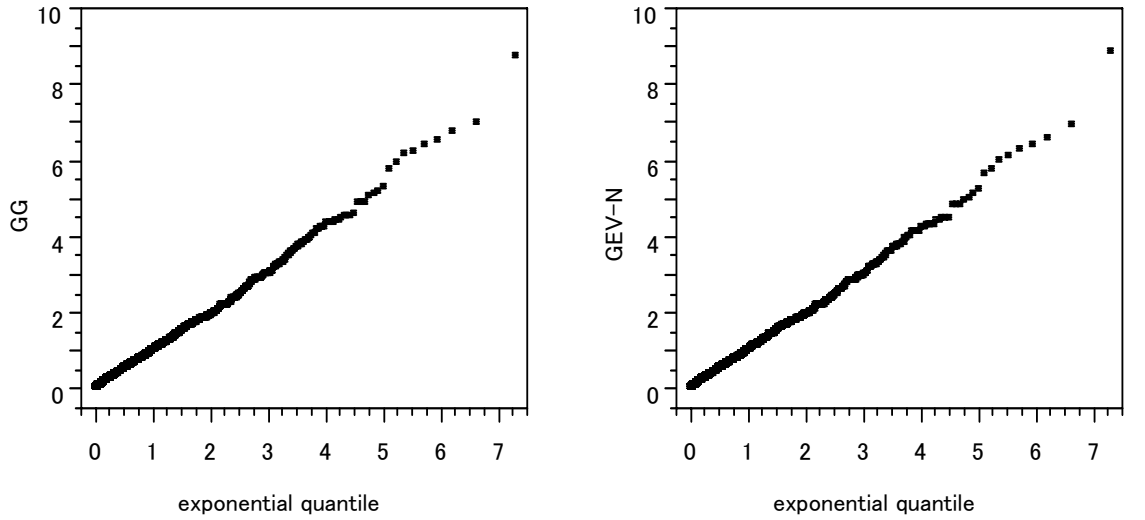
	GEV-O ₁	GEV-O ₂	GEV-O ₃	GEV-N	GEV-U
	498.10	496.65	482.53	440.4	431.7
	$\widehat{\delta} = 0.08$	$\widehat{\delta} = 0.08$	$\widehat{\delta} = 0.05$	$\widehat{\delta} = 0.12$	$\widehat{\delta} = 0.19$
	(0.039)	(0.041)	(0.042)	(0.039)	(0.033)
SSE	$\widehat{\tau} = 0.85$	$\widehat{\tau} = 0.85$	$\widehat{\tau}_0 = 0.008$	$\widehat{\sigma}^* = 0.34$	$\widehat{\sigma}^* = 0.73$
	(0.036)	(0.037)	(0.0057)	(0.014)	(0.035)
					$\widehat{\psi} = 0.63$
					(0.049)
	d.f.=350	d.f.=350	d.f.=350	d.f.=351	d.f.=350
$\widehat{\gamma}$	1.72	0.89	1.39	0.72	0.71
	(0.53)	(0.28)	(0.20)	(0.21)	(0.20)
AIC	1024.2	1021.3	993.06	906.8	891.4

Note: SSE means the minimum value of negative log likelihood function, and d.f. is its degrees of freedom. $(\cdot)$ denotes standard errors of estimates.

Table 3. Maximum likelihood estimation with respect to NPN and PG models

	NPN-U	NPN-N	PG-U	PG-N
	441.1	440.9	421.4	441.2
	$\hat{\delta} = 0.18$	$\hat{\delta} = 0.14$	$\hat{\delta} = 2.10$	$\hat{\delta} = 1.34$
	(0.037)	(0.041)	(0.18)	(0.13)
SSE	$\hat{\sigma}^* = 0.62$	$\hat{\sigma}^* = 0.33$	$\hat{\sigma}^* = 0.77$	$\hat{\sigma}^* = 0.33$
	(0.026)	(0.013)	(0.063)	(0.013)
	$\hat{\psi} = 0.49$		$\hat{\psi} = 0.58$	
	(0.052)		(0.045)	
	d.f.=350	d.f.=351	d.f.=350	d.f.=351
$\hat{\gamma}$	0.73	0.72	0.71	0.74
	(0.19)	(0.20)	(0.19)	(0.21)
AIC	910.2	907.8	870.8	908.4

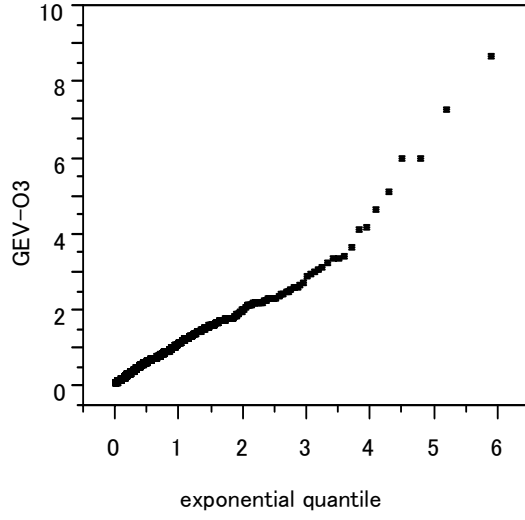
Note: SSE means the minimum value of negative log likelihood function, and d.f. is its degrees of freedom. $(\cdot)$ denotes standard errors of estimates.



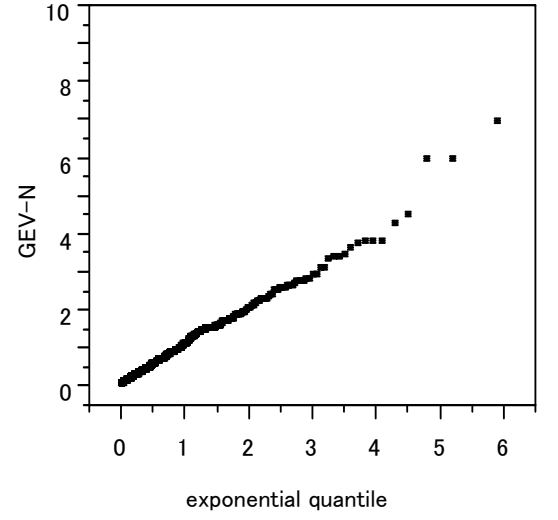
(a) GG model

(b) GEV-N model

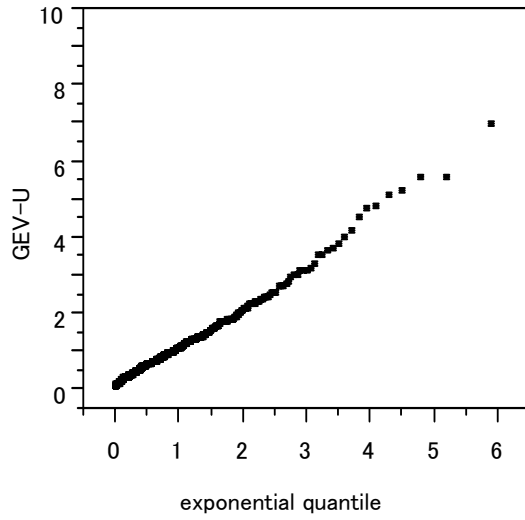
Figure 1. Exponential probability plots for the residuals of generalized Gamma and GEV models in the case of using all observations ($n = 1487$).



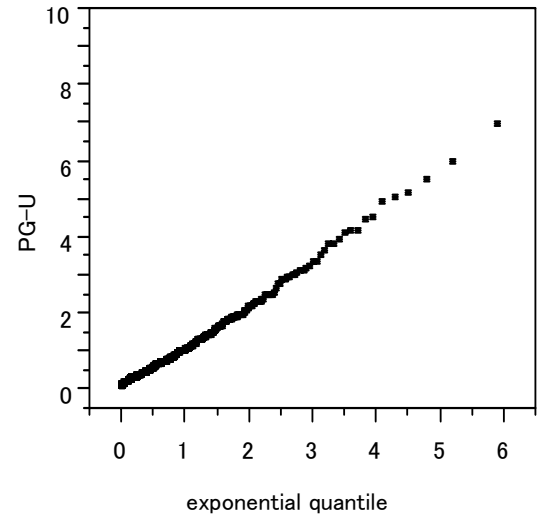
(1) GEV-O₃ model



(2) GEV-N model



(3) GEV-U model



(4) PG-U model

Figure 2. Exponential probability plots of residuals with respect to GEV and PG models in the case of 364 unit subdomains.