Simulation of the Estimation of the Maximum Inclusion Size from 2-Dimensional Observation Data on the Basis of the Extreme Value of Statistics

Junichi TAKAHASHI

Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577 Japan. E-mail: junichi@tagen.tohoku.ac.jp

(Received on November 17, 2008; accepted on February 26, 2009)

It is well known that the large-scale inclusions in steel act as the destruction starting point of the material. The labor is extremely necessary for investigating the maximum inclusions that have been distributed in the matrix of a large volume. The statistics of extreme value method is an excellent technique that can be applied for that case. When a material such as steel is opaque to visible light, the electron beam, and so on, the evaluation of the internal inclusions as three-dimensional (3D) information is presumed from the two-dimensional (2D) information observed under a microscope in respect of sample cross section. It is expected that this presumption error margin is large. It is almost impossible to verify 2D data by 3D one in a real material because the true value is uncertain, especially concerning the information of the tail of the size distribution. The purpose of the present work is to offer statistical information necessary to presume 3D characteristic value from 2D measurement data with respect to the maximum extreme value. The simulation whose 3D characteristic value is already-known is effective to this. Some 3D size distributions such as the exponential, log-normal, pseudo-normal, and Rayleigh distribution is set here, and how the 2D maximum extreme-value distribution (MED) changes into the measured number of sections is shown. The result gives a number of sections necessary to gather data with few error margins directly. Further, some findings of the relation between 3D-MED and 2D-MED are given. The overarching point is a type of the MED and conversion of the dimension.

KEY WORDS: extreme statistics; computer simulation; maximum size distribution; non-destructive evaluation; microscopy; inclusion.

1. Introduction

The quality and mechanical properties of steel are influenced by number and size distribution of inclusions included in the volume. Especially large size inclusions affect drastically on fatigue characteristics. In order to assess the large size inclusion, it is available to use a statistical treatment such as the extreme value of statistics. For applying the extreme statistics to the steel or related materials, an issue arises from a state of the target inclusions: small and dispersion in a space of the opaque specimen. In addition, it is important to know a relation between the inclusions and grain structure, as well as composition. Now therefore, a microscopic observation on the cross section of the specimen is still the most popular technique for the metallurgical analysis. In such case another problem arises: the relationship between 2-dimensional (2D) and 3-dimensional (3D) data. Wicksell tried this problem as an estimation of the 3-dimensional size distribution (3DSD) of the spherical particles dispersed in a matrix from the 2-dimensional size distribution (2DSD) of the sections observed on a planar cross-section of the specimen. Since this estimation includes so many assumptions, a large number of analytical methods have still been proposed by now.

The issue on the estimation of the maximum inclusion size distribution (MED) is a theme tried comparatively recently in this couple of decade. Murakami and coauthors used the Gumbel plot to estimate a large size inclusion in steel. Anderson et al. proposed a threshold method using a generalized Pareto or other distributions as a 3DSD. These methods are useful in the point of analyzing the measurement data graphical. It is important whether to have to measure a standard area and the number of view of which extent from the relation to the labor in the analysis that uses an actual sample. Setting this observation condition beforehand is very attended with the difficulty, because the maximum diameter distribution of the truth of a real sample is uncertain. The relation between 2D observed characteristic (properties) and 3D characteristic is clarified by simulating fractography in the sample cross-section by using the particle decentralization model to understand the grain size distribution beforehand in this research. A necessary condition for the extreme value plot of 2D observation data was presented, and 3D extreme value presumption from 2D was
2. Simulation Method

The cutting plane of the specimen involving second phase particles are modeled by a method proposed in Ref. 11. The following assumptions are made for the cutting-plane model: (1) particles are spherical and the sections on the plane are circular; (2) particles are randomly dispersed in space, namely the sections are randomly dispersed on the plane; (3) the size of particles or sections can be measured even when they overlap in space or on the plane.

To simulate the model plane one can give the 3D parameters; number of particles per unit volume of the specimen \( N_V \), mean diameter of the particle \( \mu_d \), and probability density function \( f(d_V) \) of the particles, where \( d_V \) refers to the particle diameter. The PDF of the 3DSD used in the present study is shown in Table 1 and the 3D parameters are summarized in Table 2 along with the 2D parameters determined by stereological 2D–3D relationships. 11,12 The reason for having set the average particle diameter of 3DSD as \( 1 \mu m \) is to compare the result of a simulation with the relative distribution which is obtained by dividing the maximum size distribution of an actual sample by their mean diameter. Therefore, the absolute value of the particle diameter in the maximum extreme distribution shown by the below-mentioned results can be read as the maximum value of the relative distribution (not the absolute value) in an actual sample. The discrete sections, which follow a continuous function of the 3DSD, \( f(d_V) \), are obtained by the inverse transform method using a cumulative distribution function (CDF), \( F(d_A) \), using the following equation:

\[
F(d_A) = 1 - \left( \frac{d_A^2 - d_A^0}{\mu_d} \right)^{1/2} f(d_V) d(d_V) \quad \ldots (1)
\]

where \( d_A \) refer to the diameter of section. A section size \( d_A \) is obtained as a solution of Eq. (1) with a given \( f(d_V) \) and \( \mu_d \) by putting a uniform random number into the \( F(d_A) \), which is generated by the multiplicative congruential method. A data set of \( d_A \) \((d_A1, d_A2, \ldots, d_A_n)\) is obtained by repeating the calculation \( n \) times, which is determined from the Poisson value being put the average number of \( N_A \). 11

The data set corresponds to the observation measured within the unit area, \( S_0 \). For the extreme plot, one can choose the maximum size from a data set.

3. Extreme Plot of the Maximum Sections

The MED is dependent on the shape of the large tail of the initial function (PDF and/or CDF), and is classified into three types: type I so-called the Gumbel distribution or the double-exponential distribution, type II the Caucy distribution or Ferechet distribution, type III the Weibull distribution. 4,13

The analysis of the value of the extreme statistics is carried out by plotting the data on the probability paper, or by parameter estimation such as a maximum-likelihood method. It is necessary that any method is a data straight line compared with the standardization variable of the extreme value distribution function. As the type III is utilized

![Extreme Value Plot](image-url)

**Fig. 1.** The extreme value plot of the maximum section size (3DSD: exponential distribution). (a) Type I (Gumbel plot) and (b) type II (Caucy plot).
for the distribution which has a lower and/or upper limit, type I and II plots were applied to the distribution model treated in this work.

The extreme plot of the maximum section size obtained by changing the observation view is shown in Figs. 1 to 4. The extreme plots of type I and II for the exponential distribution as the 3DSD are respectively shown in Figs. 1(a) and 1(b), where \( N \) refers to an average number of sections per unit observation view, \( d_{\text{max}} \) the maximum size in the unit observation view, \( S \) the standardized extreme variate, \( F \) the value of the CDF, and \( T \) the return period. In Fig. 1(a), good linear relationship is seen for a wide range of \( N \) from 20 (\( S_{0}=0.04 \text{ mm}^2 \)) to 200 000 (\( S_{0}=400 \text{ mm}^2 \)). On the other hand, the plot is clearly curved even for the case that the \( N \) is larger that 10 000, as shown in Fig. 1(b). It can be said for the exponential distribution as the 3DSD that the MED follows the Gumbel distribution. For the case of the log-normal distribution as the 3DSD, the MED indicates linear in the small \( N \) range (<20) on the type I plot (Fig. 2(a)), whereas in the large \( N \) range (>5 000) on the type II plot (Fig. 2(b)). For both the pseudo-normal (Fig. 3) and the Rayleigh (Fig. 4) distributions as the 3DSD, the type I plot demonstrated the linearity in the \( N \) range larger than 10 000. Figure 5 shows the relationship between straight
line correlation coefficient (R-factor) and \( N \) obtained by the least square method for the data plotted in Figs. 1–4. It can be read from this figure to which type the extreme distribution belongs statistically.

Type of the MED led from skirt of the preset 3DSD and the result of obtaining from the extreme value plot of 2DSD shown in Figs. 1–4 are summarized in Table 3. From Fig. 5, it is understood well to have to choose view as a unit area including sections of a very large number, more than 10,000. In that case, if the observer pays attention to a large sectioned particle, and the surroundings are measured only by several places, the extreme value distribution that presumes a wide area will extremely part from the true distribution. It is necessary to choose the unit observation area at random to obtain accurate 2D-MED. In addition, it is also a good method that the linearity of the data measured by changing the unit view area is confirmed.

4. Estimation of 3D-MED from 2D-MED

For estimating the 3-dimensional maximum particle size distribution in a volume from the measured 2-dimensional maximum section size distribution on the cross-section of the specimen, it is, at least, required the agreement of the type of their MEDs each other. The distributions summarized in Table 3 satisfy the condition mentioned above. This section discusses the advisability of the estimation by comparing the 3D-MED generated from the 3DSD with the corresponding 2D-MED.

A set of the maximum 3D size within a unit volume of the specimen is generated by the following procedure using Eq. (2).

\[
F(d_V) = 1 - \int_{d_V}^{\infty} f(d_V)d(d_V) \quad \text{..................(2)}
\]

One \( d_V \) value (\( d_{V1} \)) is obtained by solving the Eq. (2), in which the PDF of the 3DSD (see Table 1) and a value generated from the uniform random variable are put. A value of \( d_{V\text{max}} \) is chosen from a data set of \( d_{V1}, d_{V2}, \ldots, d_{Vn} \) obtained by repeating the generation \( n \) times, which is determined from the Poisson value being put the average number of \( N_V \). For the 3D extreme plot, the calculation was iterated 2,000 times.

Figure 6 shows the 3D-MED for the exponential and the log-normal distribution as the 3DSD. As might be expected, the 3D-MED for the exponential distribution shows linearity on the type I plot shown in Fig. 6(a). In this figure, 2D-MED with \( N=10,000 \) is plotted. It can be seen that the 3D-MED shifts to a direction of lower size compared to the 2D one, when the measured number of particles/sections is the same (\( N=10,000 \)). In other word MED chosen from every

<table>
<thead>
<tr>
<th>Table 3. Type of the maximum value of extreme distribution (MED).</th>
</tr>
</thead>
<tbody>
<tr>
<td>3DSD</td>
</tr>
<tr>
<td>Exponential distribution</td>
</tr>
<tr>
<td>Log-normal distribution</td>
</tr>
<tr>
<td>Pseudo-normal distribution</td>
</tr>
<tr>
<td>Rayleigh distribution</td>
</tr>
</tbody>
</table>

![Fig. 5. Correlation coefficient (R-factor) of the 2D-MED as a function of the average number of sections per unit observation view (N).](image)

![Fig. 6. The extreme value plot of the maximum particle size; (a) type I plot for the exponential distribution and (b) type II plot for the log-normal distribution.](image)
10,000 sections on the sample section is correspond to that from 50,000 particles in the volume. Correspondingly, for the case of the log-normal distribution as the 3DSD (type II plot), the 3D-MED plots lower than the 2D-MED, as shown in Fig. 6(b). In this case 2D-MED measured from every 10,000 sections corresponds to 3D-MED from 30,000 particles. These relationships are expressed as \( f(d_{\text{vmax}}) < f(d_{\text{amax}}) \) at the same \( N \). For the Rayleigh distribution, one can express as \( f(d_{\text{vmax}}) = f(d_{\text{amax}}) \) at any \( N \), because PDF of 2DSD and 3DSD completely agree.

As summarizing the results aforementioned, MED of the 2DSD and 3DSD relates to \( f(d_{\text{vmax}})/H_{11021} f(d_{\text{amax}}) \) at the same \( N \). This relationship can be explained by comparing the PDF of the 2DSD and 3DSD. Figure 7 shows the PDF for the case of the exponential and the Rayleigh distributions. The tail of the PDF of the 3DSD in large size range is the same as that of 2DSD for the Rayleigh distribution (Fig. 7(b)) and is smaller than that of the 2DSD for the other distributions (ex. exponential distribution shown in Fig. 7(a)), so that the MED, which consists of the sampling values extracted from the tail, is predicted to \( f(d_{\text{vmax}}) \leq f(d_{\text{amax}}) \). This relation seems to contradict intuitively that the sectioned diameter \( (d_A) \) of the sphere \( (d_V) \) is always \( d_V \leq d_A \) (Fig. 8).

Murakami et al. investigated the relationship between 2D-MED and one-dimensional (1D) MED from the diameter and the section length of the spherical carbon particles which appeared in the cutting plane of high-C steel, and showed \( f(d_{2\text{Dmax}}) > f(d_{1\text{Dmax}}) \) where the PDF of a high dimension is larger than that of a low dimension. However, it should be noted that this is not a random sampling statistically as the data of the 1D and 2D were obtained from the same sectioned particles.

It is difficult to estimate 3D-MED from 2D-MED directly using Fig. 6, even though the type of the extreme plot is common. For the estimation, dimensionality should be taken into account. Zhou et al. introduced a thickness term \( h \), which convert a 2D-MED to the corresponding 3D volume data. The term \( h \) is defined by a ratio of the 3D standardized volume \( V_0 \) to 2D standardized observation area \( S_0 \), as \( h = V_0 / S_0 \). That is, the 2D data measured from every \( S_0 \) is regarded as the 3D data measured at \( V_0 \). The \( h \) was examined as a mean value of the square root of the sectioned diameters used for the 2D-MED. Table 4 summarizes the \( h \) value when the examined method was applied to the simulation results in this work. On the other hand, the \( h \) value is determined from the ratio of the measured number when the 2D-MED and 3D-MED agreed in each other. The results are also tabulated. For all distributions used in the present work, the \( h \) value as an average square root of the section areas underestimates the 3D-MED. To estimate the 3D-MED precisely from the 2D-MED, it is required to obtain a proper translation term.

### Table 4. Values for translating 2D-MED to 3D-MED.

<table>
<thead>
<tr>
<th>3DSD</th>
<th>N value of the 3D-MED agree with the 2D-MED (n = 10000)</th>
<th>2D-3D translation term, ( N (\mu m) )</th>
<th>The ( h ) calculated from the average [area](^2) of the sections of 2D-MED (( \mu m ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>50000</td>
<td>5</td>
<td>9.97</td>
</tr>
<tr>
<td>Log-normal</td>
<td>c.a. 30000</td>
<td>3</td>
<td>5.53</td>
</tr>
<tr>
<td>Rayleigh</td>
<td>10000</td>
<td>1</td>
<td>3.12</td>
</tr>
</tbody>
</table>

5. Concluding Remarks

The maximum value of the extreme distribution (MED) of the sectioned particles observed on the fractured surface agrees with that of the dispersed particles in the volume when the size distribution of the small spherical particles dispersed in a material follows the exponential, log-normal, pseudo-normal, or Rayleigh distribution. It is noted that in order to obtain the 2-dimensional maximum size data to estimate the 3-dimensional one, at least, one can choose a unit observation area that includes more than 10,000 sections. In such case the extreme plot can be available (linearity of the plot would be satisfied). However, for quantitative estimation, a translation term, so-called \( h \), is necessary to convert the 2D data into 3D ones. An average size of the measured 2D-MED was a candidate for the term \( h \), but was produce underestimation for 3D-MED. An appropriate \( h \) to transform the dimension is required.
Acknowledgement
This work was partially supported by ISIJ Research Promotion Grant.

REFERENCES
5) S. D. Wicksell: Biometrika, 17 (1925), 84.