ACTA STEREOL 1994; 13/2: 357-362
PROC 6ECS PRAGUE, 1993
ORIGINAL SCIENTIFIC PAPER

ESTIMATION OF FIBER LENGTH AND DIAMETER DISTRIBUTION FROM SEM IMAGES

Hugues Talbot* & Dominique Jeulin†
*Centre de Morphologie Mathématique
†Centre de Géostatistique
Ecole des Mines de Paris, 35 rue St Honoré 77305 Fontainebleau, FRANCE

Abstract

Image analysis often includes a measurement process. It is well known that direct measurements on images may introduce biases that need to be corrected.
In the case of image analysis of man-made vitreous fibers, one needs to measure their diameter and their length, in order to obtain the diameter distribution, which one may want to weight by the length of the fibers, or by their volume. If one is generally able to measure fiber diameters directly after some segmentation steps and correct any measurement bias by Miles-Lantuéjoul-like methods, one cannot access directly the fiber lengths in all cases, for example when both ends of the fibers are not always visible.
In this paper we present three original methods, based on different assumptions, that allow to estimate both the unbiased diameter distribution and the mean length by diameter class in any configuration, which in turn allow to estimate with a high degree of confidence any length, surface or volume-weighted diameter distribution.
These methods were tested on simulated images, and yielded remarquable results.

Keywords : Boolean models, fiber, length distribution, mathematical morphology, measurement bias.

Introduction

In the case of this study, *fibers* are objects appearing as the intersection of elongated rectangles by a field of view, like man-made mineral fibers in SEM, as in figure 1.
In this paper we present three different methods to estimate the mean length by diameter class of a fiber population which requires only the information given by random sampling on the population by fields of view, all at the same magnification, suitable for diameter measurements. These methods allows also to obtain the unbiased diameter distribution.

Preliminary hypothesis

We suppose we need to investigate fibers on a plane support. Additionnaly, we suppose that the fiber concentration is so that the underlying support is visible in places, and

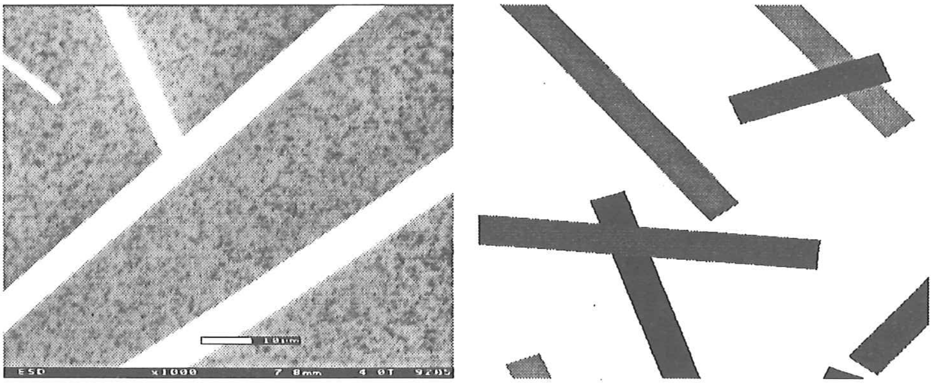


Figure 1: Actual image (left) and simulation (right).

that we are able to segment these fibers (Talbot, 1993), *i.e.* we are able to individualize these fibers along all their potentially visible length in the field of view. (*i.e.*, we can substract the presence of noise or other fibers). We assume that we can measure the diameter of all these fibers and sort them by diameter class. Finally we suppose that the fiber distribution or the field sampling are stationnary processes.

NOTATIONS

In this paper, **D** is a diameter, **L** is a length, $\mathcal{L}$ is a perimeter, **A** is an area, $\overline{X}$ is the mean value of X, Q^* indicates an estimated value of Q.

FIRST METHOD : AREA ACCUMULATION

For a given fiber, we have:

$$\mathcal{L} = 2(D + L)$$

If we calculate the sum of the perimeters of all the fibers of a given diameter class D, we have:

$$\mathcal{L}(D) = 2N(D)[D + \overline{L}(D)]$$

Where $N(D)$ is the number of fibers in the diameter class D, and $\overline{L}(D)$ the mean fiber length in this same class.

We can write:

$$\overline{L}(D) = \frac{\mathcal{L}(D)}{2N(D)} - D \qquad (1)$$

We need now to get $N(D)$. We have:

$$\begin{aligned} A(D) &= \sum_i A_i(D) \\ &= N(D)D\overline{L}(D) \end{aligned}$$

If we substitute equation (1) in this formula, then we end up with

$$A(D) = \frac{D\mathcal{L}(D)}{2} - D^2 N(D)$$

if we extract N(D), we have:

$$N(D) = -\frac{A(D)}{D^2} + \frac{\mathcal{L}}{2D} \qquad (2)$$

The above equations hold if we have the entiere knowledge of the distribution, however, the variables $N(D)$ and $A(D)$ are summable. It means that we can estimate $N(D)$ and $\overline{L}(D)$ by measuring and summing the visible area of each fiber of diameter D, and their perimeter on a large enough number of images, which looks simple enough. However, equation (2) is numerically unstable. It is a difference between two potentially large numbers. In addition, one of these depends on the perimeter, a term that is not easy to compute accurately. Nevertheless, we can compute theoretically the value of $N(D)$ for all relevant situations, indeed, for a single fiber on a single image, from equation (2) we have:

- For an entirely visible fiber, $N(D) = 1$.
- For a fiber of which we can see one end, $N(D) = \frac{1}{2}$.
- For a fiber of which we can see no end, $N(D) = 0$.

And we will rather estimate $\overline{L}(D)$ by:

$$\overline{L}(D) = \frac{A(D)}{DN(D)} \qquad (3)$$

Therefore, all we need to do is segment all the fibers, reconstruct all these which are crossed over (by other fibers), count their visible extremities, and measure the visible area of the fibers, all for each diameter class. The number of fibers of that diameter is estimated by half the number of fiber extremities (which is intuitive enough).
This method is very simple and very robust, namely it does not depend on the kind of distribution (Poisson process or not) we are working with, but for accurate measurement, $N(D)$ must be large, and in the case of long fibers, extremities do not appear very often.

SECOND METHOD : INTERSECTION WITH THE FIELD OF VIEW

The number of fibers that can be counted on a field of view depends both on the density of the distribution, on the characteristics of the fibers and of the field of view. Indeed, the longer a fiber is, the more likely it is to cross a field of view. The average number of visible fibers in a field of view is therefore :

$$\overline{N}(D, L) = \theta f(D, L)\overline{A}[F \oplus \mathcal{Z}]$$

where $\overline{N}(D, L)$ is the number of fibers on the field of view of diameter D and length L, θ the global density of the distribution, $f(D, L)$ the frequency of the class (D, L) and $\overline{A}[F \oplus \mathcal{Z}]$ the mean area of the fibers dilated (Serra, 1982) by the field of view.

If we expand this expression using Steiner's formula, and if we integrate over all the possible L, we have:

$$\overline{N}(D) = \int_L \overline{N}(D,l)dl = \theta f(D)[A(\mathcal{Z}) + D\overline{L}(D) + \frac{1}{\pi}\mathcal{L}(\mathcal{Z})[D + \overline{L}(D)]] \quad (4)$$

where $\mathcal{L}(\mathcal{Z})$ is the perimeter of the field of view, $f(D)$ the frequency of the diameter D and $A(\mathcal{Z})$ the area of the field of view.
We can estimate θ by summing all the $\theta f(D)$ for all diameter classes, knowing that $\sum_D f(D) = 1$.
Using equation (4), we can estimate $\overline{L}(D)$ and $\theta f(D)$ by varying the dimensions of field of view. For example, we can consider the whole field of view from the images we have, and one fourth or one half of it on the same images. This will yield a system of two equations that we can solve easily :

$$\overline{L}(D) = \frac{N_1[A(\mathcal{Z}_2) + \frac{\mathcal{L}(\mathcal{Z}_2)D}{\pi}] - N_2[A(\mathcal{Z}_1) + \frac{\mathcal{L}(\mathcal{Z}_1)D}{\pi}]}{N_2[D + \frac{\mathcal{L}(\mathcal{Z}_1)}{\pi}] - N_1[D + \frac{\mathcal{L}(\mathcal{Z}_2)}{\pi}]} \quad (5)$$

where N_i is the number of fibers effectively encountered in the set of field we are considering, labelled by i (complete field, of half the field...). $\mathcal{L}(\mathcal{Z}_i)$ is the perimeter of that field, and $A(\mathcal{Z}_i)$ its area.
Once we can estimate $\overline{L}(D)$, we have:

$$\theta f(D) = \frac{N_i(D)}{A(\mathcal{Z}_i) + D\overline{L}(D) + \frac{1}{\pi}\mathcal{L}(\mathcal{Z}_i)[D + \overline{L}(D)]} \quad (6)$$

This equation allows to correct the measurement bias we are making by counting visible fibers in the field of view. This method yields relatively good results as far as $\theta f(D)$ is concerned. The equation giving $\overline{L}(D)$ is poorly conditioned and yields unstable results.

Third method : Boolean model

This method is based on the work of G. Matheron (1975), J. Serra (1982) and D. Jeulin (1991) on random Poisson processes.
Let us introduce it rapidly. If X is a binary set in $\mathbb{R}^2$ made of the union of a family of primary grains X' arranged according to a Poisson point process. If B is a compact of $\mathbb{R}^2$, the probability of B to fall entirely outside X is :

$$Q(B) = P(B \subset X^c) = \exp[-\theta\overline{A}(X' \oplus B)] \quad (7)$$

where $\overline{A}(X' \oplus B)]$ is the mean area of X' dilated by B, and θ the intensity of the Poisson point process.
This equation allows us generally to find interesting characteristics inherent to the distribution. For example, if X is generated by a family of fibers of the same diameter class, and B is a family of disks of increasing diameter r, we can say that:

$$\overline{A}(X \oplus B) = \overline{L}D + 2r[\overline{L} + D] + \pi r^2$$

If we expand equation (7), and if use a point for B with: $q = \exp[(-\theta f(D))(\overline{L}D)]$, then

$$\frac{1}{r}\ln\frac{Q(r,D)}{q} = -\theta f(D)[2[\overline{L}+D]+\pi r] \tag{8}$$

which is a linear equation in r. We can estimate $Q(r,D)$ and q by :

$$Q(B(r,D))^* = 1 - A_A(X \oplus B(r)) = A_A(X^c \ominus B(r))$$

Where A_A is an area fraction. If equation (8) does not yeld a linear result, the fiber deposition is not a Poisson process, if the result is linear, both origin and slope allow us to get $\overline{L}$ and $f(D)$. If a is the value at $r=0$ and b the slope, we have:

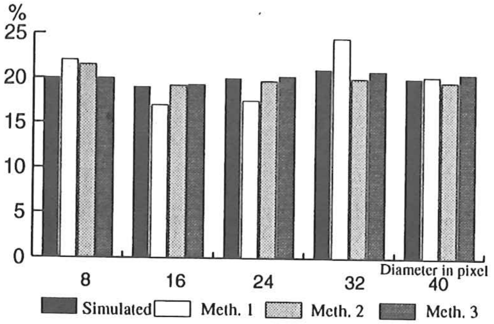
$$\left\| \; \overline{L}(D) = \frac{a\pi}{2b} - D \right. \tag{9}$$

$$\left\| \; \theta f(D) = -\frac{b}{\pi} \right. \tag{10}$$

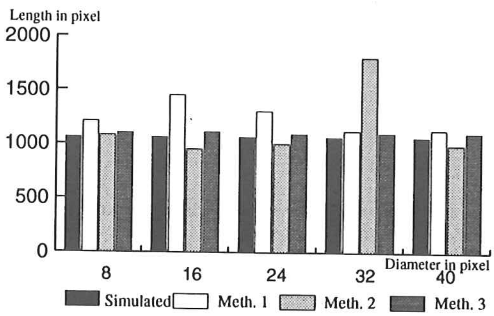
Again, θ can be estimated knowing that $\sum_D f(D) = 1$.

Results and conclusion

We have extensively tested all three methods on simulated images like shown in Fig 1. Fibers had variable diameter and variable length. Diameter distribution was uniform as well as length distribution. Mean length was 1060 pixels. The field of view was 512 × 512 pixels. Mean fiber length was therefore longer than the diagonal of the field, so no direct method could estimate this mean length. Results are shown in Fig 2. A convergence analysis on the length estimation was also made by sub-sampling a large number of images with increasing samples size, shown in Fig 3 (min and max result among subsets are printed). Both method 1 and 2 were tested on the same images. Method 3 required a different set, hence the small discrepancy in the final estimation.



(a) Diameter distribution



(b) Length estimation

Figure 2: Estimation of $N(D)$ and $\overline{L}(D)$ on a simulation of 1000 images. Distribution is uniform and mean length is 1060 pixels.

Method 1 converges the fastest towards the correct mean length for small samples, but method 3 become better for mid-size and bigger samples. Method 2 is the poorest length estimator. Method 1 yield an unstable diameter bias correction, but a correct length

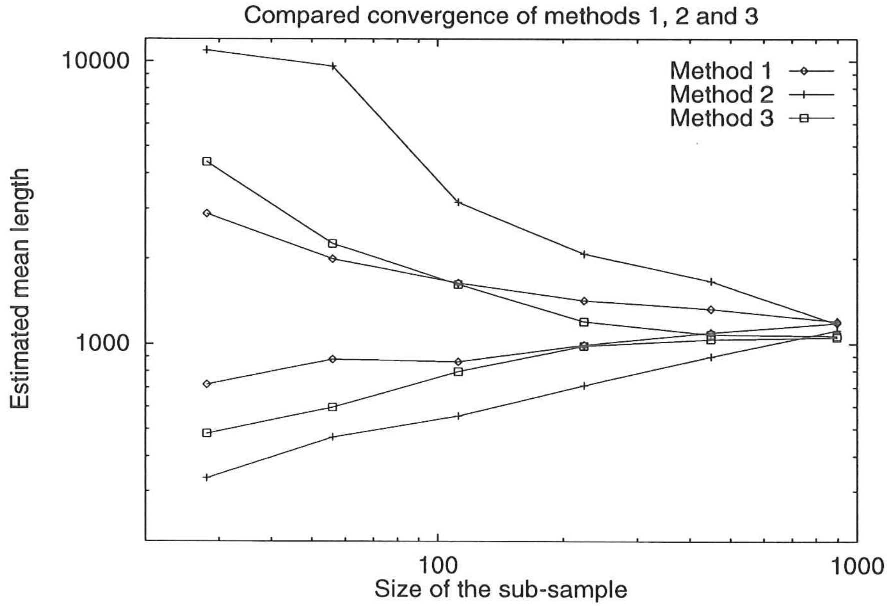


Figure 3: Convergence analysis for all three methods.

estimation. Method 2 gives an unreliable length estimation, but a good diameter bias correction. Method 3 performs well in both diameter measurement bias correction and length estimation.

In conclusion, in cases where method 3 can be used (the Poisson process hypothesis holds), this is the method of choice. When it is not the case, diameter measurement bias correction should be performed by method 2 while length estimation should be obtained by method 1.

These methods have been applied on actual images of insulation mineral fibers (Talbot, 1993).

REFERENCES

Jeulin D. *Modèles Morphologiques de Structures Aléatoires et de Changement d'Echelle*, Doctorat d'état, University of Caen, April 1991.

Lantuéjoul Ch. *La squelettisation et son application aux mesures topologiques des mosaïques polycristallines*, Thèse de doctorat, Ecole des Mines de Paris, 1978.

Matheron G. *Random Sets and Integral Geometry*, Wiley, NY, 1975.

Miles R. in *Stochastic Geometry*, Wiley, NY, 1974.

Serra J. *Image Analysis and Mathematical Morphology*, Academic Press, London, 1982.

Talbot H. *Analyse morphologique de fibres minérales d'isolation*, Thèse de doctorat, Ecole des Mines de Paris, October 1993.