Image Anal Stereol 2025;43:99-103 doi: 10.5566/ias.3593
Original Research Paper

MONTE CARLO SIMULATION OF THE TOMATO SALAD PROBLEM

JOHANNES A. ÖSTERREICHER

LKR Light Metals Technologies Ranshofen, Austrian Institute of Technology,
Lamprechtshausenerstr. 61, 5282 Ranshofen, Austria
e-mail: johannes.oesterreicher@ait.ac.at
(Received March 20, 2025; revised April 25, 2025; accepted April 28, 2025)

ABSTRACT

The tomato salad problem describes a stereological bias in the microscopic characterization of particulate systems, particularly in transmission electron microscopy (TEM). When a thin section is prepared from a material containing dispersed particles, the observed particle size distribution in micrographs may differ from the true distribution due to truncation effects and sampling bias. Depending on the initial size distribution, the observed mean particle size may appear smaller or larger than the actual mean. This work presents a Monte Carlo simulation of the tomato salad problem, implemented in R, to study the effects of foil thickness and particle size distribution on observed size measurements. Simulated results are compared with analytical predictions, showing good agreement. The study also highlights the impact of stochastic sampling errors, which can exceed the bias introduced by the tomato salad problem, emphasizing the need for sufficient sampling in microscopic analysis. The developed simulation may serve as an educational tool and could be extended in future work to analyze non-spherical particles and sample preparation artifacts.

Keywords: Dispersoids, transmission electron microscopy, particle size distribution, aluminum alloys, spherical, corpuscle problem.

INTRODUCTION

The corpuscle problem is one of the classic problems of spherical stereology. It arises when a planar section is taken from a solid medium containing spherical particles, a common step in sample preparation for reflected light microscopy or scanning electron microscopy (SEM). Because not all particles are sectioned through their center, some will appear with a smaller diameter in the micrograph. This can result in an observed size distribution curve which is skewed to the left (i.e., to smaller diameters) when compared to the actual size distribution. However, small particles have a lower probability of being cut by the section plane; this can result in an observed size distribution which is skewed to the right. The second effect counteracts the former and the size distribution of the particles determines which effect dominates. In some cases, the effects cancel each other out and the observed size distribution is similar to the real size distribution. The corpuscle problem has been analytically solved by Wicksell (1925).

The tomato salad problem is a similar issue encountered in optical bright field microscopy, transmission electron microscopy (TEM), and other microscopic methods. In this case, a thin slice is prepared from a medium containing particles, which can result in particle truncation. Spherical particles whose centers lie within the foil will appear in the micrograph with their actual diameter, regardless of whether they are cut by the foil surface. However, particles with centers outside the foil are truncated and will appear with a reduced diameter. As a result, the observed mean diameter of particles, derived from micrographs, may be smaller than the true mean diameter of the particle distribution. Paradoxically, depending on the shape of the particle distribution, the observed mean diameter can also be larger than the true mean, as larger particles are more likely to have portions of their volume within the foil. In some cases, these effects may cancel each other out, and the observed mean particle diameter will closely match the actual mean diameter, provided a sufficiently large population is studied.

Study of the tomato salad problem has been pioneered by Bach (1958) and an analytical solution has been presented by Goldsmith (1967). The relationship between the observed distribution $G(r)$ and the real, typically unknown distribution $F(s)$ is given by (Fleischer, 1994):

$$G(r) = \frac{t \cdot F(r) + 2 \cdot r \cdot \int_r^\infty \frac{F(s) \cdot s}{\sqrt{s^2 - r^2}} ds}{d + 2 \cdot \int_0^\infty F(s)\, ds}, \quad r \geq 0 \qquad (1)$$

Where t is the thickness of the slice and r and s are the observed and real radii.

In practice, the observed size distribution is derived by measuring the size of a finite number of particles in TEM micrographs. Therefore, the size distribution data is noisy. Gorenflo (2002) showed that noise amplification can become a critical issue, as small errors in the measured data can lead to large deviations in the reconstructed distribution. As a result, the problem transitions from well-posed to ill-posed, making direct inversion infeasible (Gorenflo, 2002).

Here, a Monte Carlo simulation of the tomato salad problem in R is presented, the results of which are in good agreement with the mathematical model (Eq. 1). The simulation may be useful for educational or illustrative purposes. In addition, it allows for the simulation of TEM images and for studying stochastic effects associated with practical limits of microscopic materials characterization, such as limited numbers of micrographs and measured particles.

METHODS

In many physical systems, particle size distributions follow a lognormal distribution, $\ln(s) \sim \mathcal{N}(\mu, \sigma^2)$, due to the multiplicative nature of nucleation and growth processes. μ determines the scale of the distribution. The mean radius (i.e., the expected value of the lognormal distribution) is $\bar{s}$. The relation between μ and $\bar{s}$ is given by:

$$\mu = \ln(\bar{s}) - \frac{\sigma^2}{2} \quad (2)$$

μ was chosen such that $\bar{s}$ remains 50 nm, representing a typical dispersoid size in aluminum alloys.

In Fig. 1, lognormal distributions with different values of σ are presented. A low σ corresponds to a narrower distribution, where most particles have sizes close to the mean. Conversely, a high σ leads to a wider distribution, with some particles significantly smaller or larger than the mean.

For Monte Carlo (MC) simulation of the tomato salad problem in TEM, code was written in R, version 4.4.2 (R Core Team, 2024). It is given in the Supplementary Materials. The following procedure explains how it works:

1. The parameters σ and μ of the log-normal size distribution of the particles are defined.
2. The edge lengths of micrographs to be simulated are defined (here: 5×5 µm^2)
3. The simulation volume is defined by the area of the micrographs (see step 2) and a height that should be much larger than both the mean particle diameter and the TEM foil thicknesses under investigation (e.g., 50 µm).
4. Poisson point process: While the volume fraction of particles is smaller than the desired volume fraction (here: 0.01), particles are generated with a random diameter from the size distribution defined in (1.) and randomly placed within the volume defined in (3.) so that they do not cut the surface of the simulation volume. Overlapping with other particles is not prevented, as this has no influence on the numerical results.
5. A slice (TEM foil) of thickness t is placed at a random height in the simulation volume.
6. Particles that lie entirely within the foil are identified.
7. Particles that are cut by the upper or lower surface of the foil are identified.
8. Particles from (7.) are sorted into particles whose centers lie within the foil and those whose centers lie outside of the foil (these are the particles which appear with reduced diameter on a micrograph).
9. Apparent radii of all particles identified in (6.–8.) are calculated and stored.
10. Visualization of simulated TEM image.
11. Go to (5.) until the desired number of simulations is achieved.

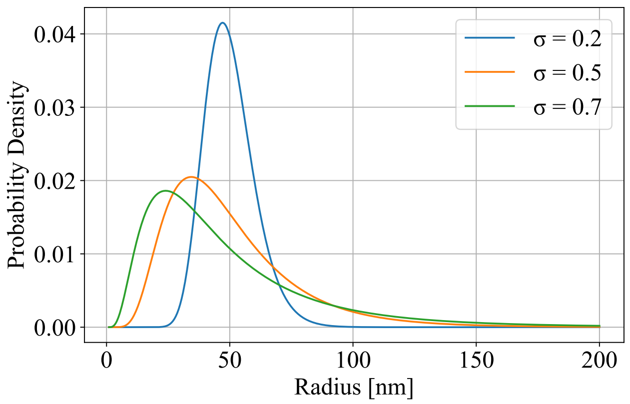

Fig. 1: *Lognormal distributions for different σ values (0.2, 0.5, and 0.7), where the mean radius $\bar{s}$ is 50 nm.*

RESULTS AND DISCUSSION

Fig. 2 presents a simulation of the tomato salad problem in TEM, illustrating how particle positioning relative to the foil influences their apparent sizes in TEM images. Purple and magenta particles (panels A and C) have their centers inside the foil, appearing in the micrograph (panel B) with their actual diameters; while purple particles are partially cut by the foil, magenta ones remain fully intact. In contrast, particles with centers outside the foil (orange in panels A and C) appear with reduced apparent diameters in the micrograph. Fig. 2B presents the simulated TEM bright-field image, where it is typically not possible to distinguish whether an observed particle retains its original radius or appears reduced due to foil intersection. Finally, Fig. 2C visualizes and identifies the particles in the TEM image, explicitly marking their apparent diameters (solid circles) and real diameters (dashed lines).

At first glance, Fig. 2 suggests that the effect of the tomato salad problem is relatively minor. Although many particles are intersected by the foil, those with their center inside the foil retain their full radius. Even for particles with centers outside the foil, the apparent radius in the projection is often close to the original diameter. Only in a few cases does the apparent diameter deviate significantly, appearing much smaller than the true diameter.

However, the extent of distortion in particle size distribution due to the tomato salad problem is influenced by the parameters of the lognormal distribution (σ and μ). In Fig. 3, the observed mean diameters are presented for various lognormal distribution shapes across different foil thicknesses. The data reveal that, depending on the particle size distribution, either a reduction or an increase in the observed mean diameter can occur. While a reduction is anticipated due to particles being partially intersected by the foil, an increase may be less intuitive. This increase occurs because larger particles are more likely to be intersected by the foil, leading to a higher probability of their partial inclusion in the observed sample (i.e., a sampling bias).

Fig. 3 presents results from the Monte Carlo simulation and the analytical solution (Eq. 1), demonstrating good agreement between the two approaches. Moreover, the Monte Carlo method allows for the analysis of stochastic errors, which tend to increase with wider particle size distributions. This is because wider distributions encompass a greater range of particle sizes, leading to increased variability in the simulation outcomes.

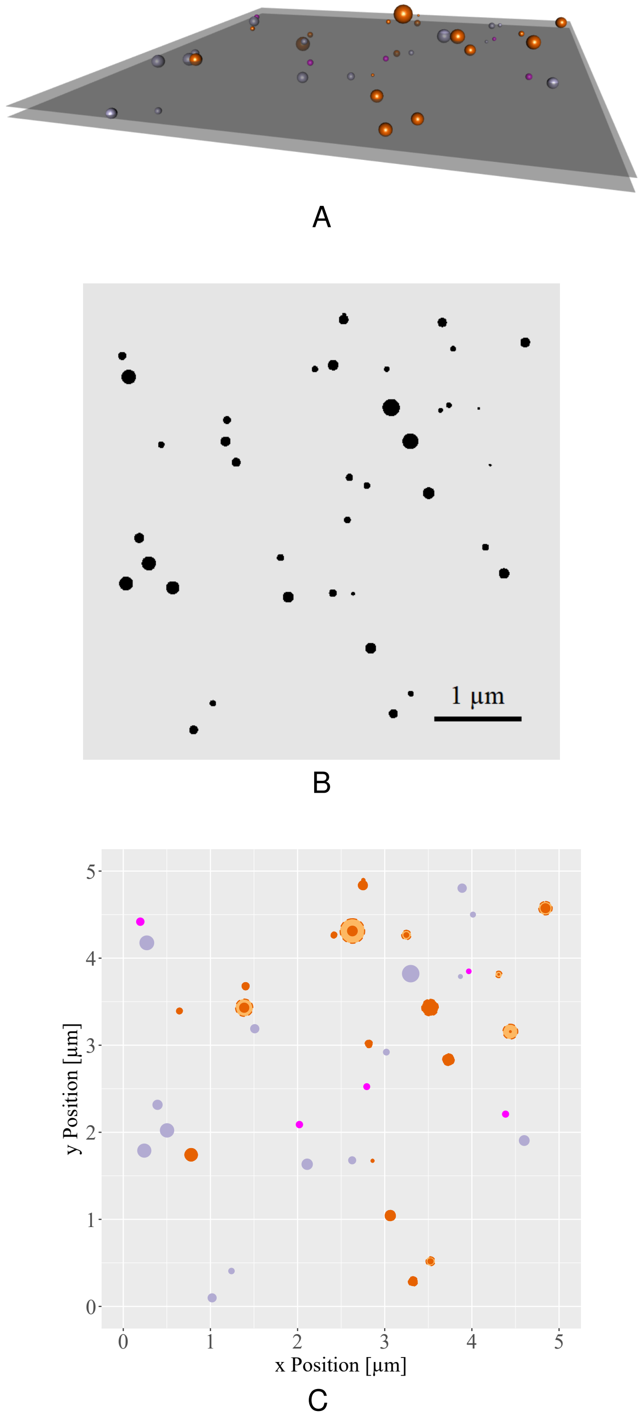

Fig. 2: *Simulation of the tomato salad problem in TEM for lognormally distributed particles (σ = 0.5, $\bar{s}$ = 50 nm, foil thickness 100 nm). (A) Three-dimensional visualization of the simulated TEM foil. (B) Simulated resulting TEM bright-field (BF) image. (C) Visualization and identification of particles visible in the TEM image, showing their apparent diameters (solid circles) and real diameters (dashed lines).*

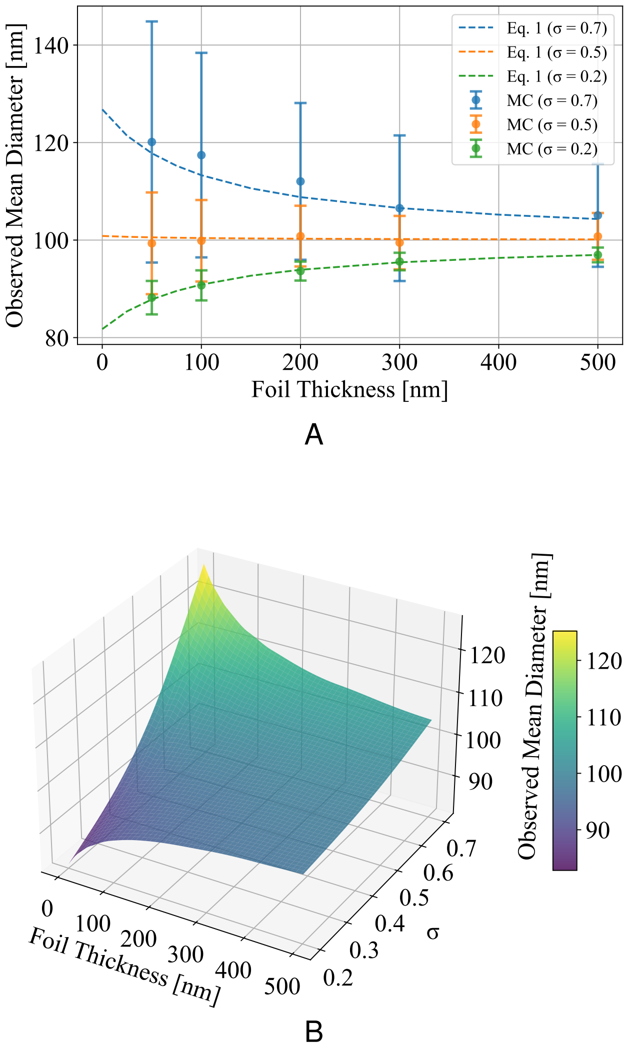

Fig. 3: *(A) Comparison of observed mean diameters from Monte Carlo simulations (100 simulations per data point, with standard deviation error bars) compared to the analytical solution (Eq. 1). (B) Analytical solution over a broader parameter space.*

In practice, a broader distribution of particle sizes—and consequently, a wider range of observed diameters—requires measuring more particles to obtain a reliable observed mean diameter. The standard deviation of individual particle sizes could be used to determine the required sample size. However, since our data consists only of mean particle sizes per micrograph, as individual particle measurements were not stored, we rely on the spread of these mean values across micrographs as an indirect measure of variability. This approach accounts for both the intrinsic particle size distribution and the statistical fluctuations due to limited sampling in each micrograph.

The required sample size n (number of micrographs) can be calculated using the formula:

$$n = \left(\frac{Z_{\alpha/2} \times \mathrm{SD}}{\mathrm{MoE}} \right)^2 \tag{3}$$

where:

- $Z_{\alpha/2}$ is the Z-score corresponding to the desired confidence level,
- SD is the standard deviation of the mean particle sizes across micrographs,
- MoE is the desired margin of error.

Using $Z_{\alpha/2} = 1.96$ and a margin of error (MoE) of 5 nm , we can calculate the number of micrographs that should be analyzed to achieve a 95% confidence level of being within 5 nm of $2\bar{r}$. Rounding up, we obtain $n = 2$ for $\sigma = 0.2$, $n = 11$ for $\sigma = 0.5$, and $n = 68$ for $\sigma = 0.7$.

Our results highlight that the stochastic error introduced by sampling can be significantly larger than the error due to the tomato salad problem, especially when only a few micrographs are taken or analyzed.

For distributions of non-spherical particles, the analysis becomes more complex (Andersen *et al.*, 2008) and Monte Carlo simulation may be an attractive alternative to purely mathematical approaches. Future research could therefore focus on expanding the simulation to non-spherical particles.

While the preceding analysis focused on the theoretical treatment of the tomato salad problem, practical factors such as sample preparation must also be considered. For instance, in aluminum alloys containing dispersoid particles, ion polishing may lead to mechanical capping of these particles. In contrast, electropolishing is less likely to cause such effects, as dispersoids are typically more noble than the surrounding aluminum matrix. This prevents partial dissolution of the dispersoid itself. Rather, preferential dissolution of the matrix (micro-galvanic corrosion) can lead to particle detachment (Kosari *et al.*, 2020; Wu *et al.*, 2024). Conversely, preferential dissolution of second-phase particles leads to pitting (dealloying) (Österreicher *et al.*, 2016). Both effects affect the measured dispersoid size distribution. These mechanisms, including mechanical attack, microgalvanic corrosion, and pitting corrosion, are illustrated in Fig. 4. To account for such effects, the Monte Carlo simulation could be adapted to incorporate the probability of particle loss due to selective matrix dissolution or dealloying, providing a more realistic representation of the observed microstructure.

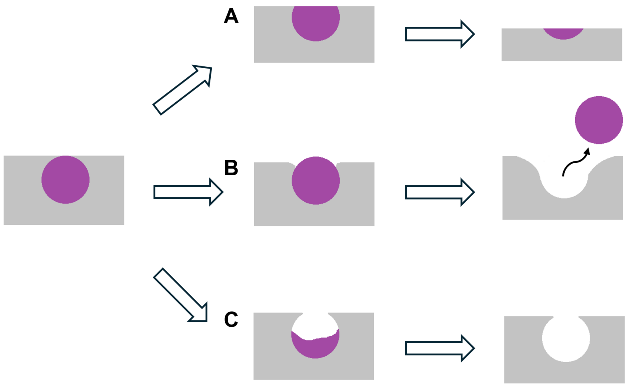

Fig. 4: *Schematic representation of different etching effects: (A) Mechanical attack, (B) Microgalvanic corrosion, and (C) Pitting corrosion (dealloying).*

Similar effects could also occur in other material systems where second-phase particles exhibit different properties than the matrix. Expert judgment is necessary when interpreting microstructural data, especially when sample preparation techniques may introduce biases.

CONCLUSIONS

This paper presents a Monte Carlo simulation of the tomato salad problem in spherical stereology, illustrating how the underlying size distribution influences the observed size distribution in agreement with the analytical solution.

Furthermore, the Monte Carlo approach highlights practical challenges introduced by limited sampling, where stochastic errors can exceed the bias itself. Additionally, the effects of sample preparation are discussed.

These findings reinforce the importance of careful experimental design in microscopic particle analysis. Researchers should be mindful of both systematic biases from sectioning and random errors from finite sampling when interpreting micrographs. Improving accuracy may require not only analytical corrections but also optimized sample preparation techniques to minimize additional artifacts.

ACKNOWLEDGMENTS

The author thanks Alois C. Ott for his valuable feedback.

Funding: The ProMetHeus project is funded under COMET by the Austrian Ministry of Climate Action, Environment, Energy, Mobility, Innovation, and Technology (BMK), the Ministry of Labour and Economy (BMAW), the funding agencies of Upper Austria, Styria, and Lower Austria—conducted by the Austrian Research Promotion Agency (FFG) [grant No. 904919].

REFERENCES

Andersen SJ, Holme B, Marioara CD (2008). Quantification of small, convex particles by TEM. Ultramicroscopy 108(8):750–62.

Bach G (1958). Über der Größenverteilung von Kugelschnitten in durchsichtigen Schnitten endlicher Dicke. ZAMM-J Appl Math Mech 38(7–8):256–8.

Fleischer K (1994). The tomato salad problem reconsidered. Biometrical J 36(2):193–203.

Goldsmith PL (1967). The calculation of true particle size distributions from the sizes observed in a thin slice. Br J Appl Phys 18(6):813.

Gorenflo R (2002). The tomato salad problem in spherical stereology. In: Romanov VG, Kabanikhin SI, Anikonov YE, Bukhgeim AL, eds. Ill-posed and inverse problems. De Gruyter, 117–34.

Kosari A, Zandbergen H, Tichelaar F, Visser P, Taheri P, Terryn H, Mol JMC (2020). In-situ nanoscopic observations of dealloying-driven local corrosion from surface initiation to in-depth propagation. Corros Sci 177:108912.

Österreicher JA, Kumar M, Schiffl A, Schwarz S, Hillebrand D, Bourret GR (2016). Sample preparation methods for scanning electron microscopy of homogenized Al-Mg-Si billets: A comparative study. Mater Charact 122:63–9.

R Core Team (2024). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. `https://www.R-project.org/`

Wu Z, Liu C, Jiang Y, Wang Y (2024). On the formation trajectory of diverse corrosion morphologies associated with T phases in an Al-Cu-Li alloy. Corros Sci 236:112267.

Wicksell SD (1925). The corpuscle problem: A mathematical study of a biometric problem. Biometrika 17(1/2):84–99.