Research Article

A Cluster-Based Method for Improving Analysis of Polydisperse Particle Size Distributions Obtained by Nanoparticle Tracking

Thorsten Wagner, 1 Martin Wiemann, 2 Inge Schmitz, 3 and Hans-Gerd Lipinski 1

1 Biomedical Imaging Group, Department of Informatics, Dortmund University of Applied Sciences and Arts, Emil-Figge-Straße 42, 44227 Dortmund, Germany
2 Institute for Lung Health (IBE Re-D gGmbH) Münster, Mendelstraße 11, 48149 Münster, Germany
3 Institute for Pathology, Ruhr University Bochum, Bärkë-de-la-Camp Platz 1, 44789 Bochum, Germany

Correspondence should be addressed to Thorsten Wagner; wagner@biomedical-imaging.de

Received 15 December 2012; Accepted 11 February 2013

Academic Editor: Frank Hubenthal

Copyright © 2013 Thorsten Wagner et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Optical tracking methods are increasingly employed to characterize the size of nanoparticles in suspensions. However, the sufficient separation of different particle populations in polydisperse suspension is still difficult. In this work, Nanosight measurements of well-defined particle populations and Monte-Carlo simulations showed that the analysis of polydisperse particle dispersion could be improved with mathematical methods. Logarithmic transform of measured hydrodynamic diameters led to improved comparability between different modal values of multimodal size distributions. Furthermore, an automatic cluster analysis of transformed particle diameters could uncover otherwise hidden particle populations. In summary, the combination of logarithmically transformed hydrodynamic particle diameters with cluster analysis markedly improved the interpretability of multimodal particle size distributions as delivered by particle tracking measurements.

1. Introduction

It has often been shown that the size of nanoparticles determines, among other factors, its biologic or even toxic effects [1]. However, the exact description of a nanoparticle suspension is a challenging issue, for example, during toxicological in vitro testing of nanoparticles [2–4].

During the past 5 years, the Nanoparticle Tracking Analysis (NTA) became increasingly important in nanotoxicology to describe the size distribution of nanoparticle suspensions [5]. As a basic principle, the Brownian motion of laser illuminated NPs is captured by a CCD camera mounted on a conventional light microscope and particle trajectories are tracked by image processing software. Particle size distribution is then obtained via the Stokes-Einstein relation [6]. The variance of the size distribution depends on duration of the observed particle tracks [7] and, in particular, on the mean particle size. Thus, a broadening of the size distribution is to be expected if mean diameter increases. Owing to this broadening effect the proportions of different particle populations are hard to assess by the modal values of a polydisperse suspension. From this consideration it, appears intuitively clear that the larger the broadening effect is, the more difficult it becomes to separate populations of particles with a small difference in size.

The purpose of this paper is to demonstrate these features by means of Monte-Carlo simulations of polydisperse suspensions. We furthermore will show methods useful for the analysis and improved interpretation of polydisperse particle size distributions (PSDs). Therefore, different proportions of size-defined particle populations were simulated, and size distributions were analysed. To effectively attenuate the broadening effect and to increase the interpretability of polydisperse distribution, calculated particle diameters will be logarithmised to normalize their variance. This means that the logarithm of each diameter is calculated and used for further analysis. To exploit the logarithmised diameters, a cluster analysis will be used. The efficacy of these procedures to improve the analysis of size distributions will be validated by simulations and verified by experimental results.
2. Theory: Nanoparticle Size Distributions via Particle Tracking

To track the Brownian motion of NP in suspension, we used a Nanosight instrument (LM10) which combines a conventional light microscope and a laser illumination device. The laser light is guided approximately perpendicular to the optical axis and scattered by NPs which, therefore, can be viewed according to Huygens principle. A CCD camera captures the defraction patterns of diffusing particles at 30 frames per second. Then, the Nanosight software detects the center of each single diffraction pattern and measures the length of the trajectory [6]. Based on the trajectory data, the mean square displacement \( r^2 \) of a particle \( k \) is calculable in several ways [8, 9]. Given a trajectory consisting of \( N \) steps, a formula computes the mean squared distance between two successive particle positions \( p_{k,i} \):

\[
\frac{1}{N-1} \sum_{i=1}^{N-1} (p_{k,i+1} - p_{k,i})^2.
\]

The distance \( (p_{k,i+1} - p_{k,i}) \) is referred to as the step length and easily calculated if magnification and the pixel size of the camera are known. \( \tau \) is the time difference between two subsequent positions. The value \( r^2 \) is associated with the estimated two-dimensional (2D) diffusion coefficient (DC) \( D \) in a more complex manner:

\[
\overline{r^2} = 4D\tau.
\]

Each diffusion coefficient is usually weighted by its track length. For this purpose, a particle diameter estimated from a trajectory composed of \( N \) steps is added \( N \) times to the data set, provided that a minimum number of steps were recorded.

Together, this weighting method countervails a source of error, caused by the fact that smaller particles will diffuse in and out field of view more rapidly and therefore more often than larger particles [10]. Finally, the hydrodynamic diameter \( d_k \) of a particle is determined using the Stokes-Einstein relation:

\[
\overline{\tau^2} = \frac{3K_B}{4\pi\eta d_k}.
\]

in which \( T \) is the temperature in Kelvin, \( \eta \) the viscosity of the suspension, and \( K_B \) the Boltzmann constant. In any case, the accuracy of a particle size distribution measurement depends on having enough particles (>200) observed.

3. Materials and Methods

3.1. Data Acquisition. Particle size measurements were carried out with well-defined polystyrene standard particles sized 50 nm (Thermo Scientific, 3050A NIST), 100 nm (Kisker-Biotech, PPs-0.1), 150 nm (Thermo Scientific, 3150A NIST), and 200 nm (Kisker-Biotech, PPs-0.2). Particle suspensions were adequately diluted with particle-free, double-distilled H₂O to obtain 30–50 particles within one field of view, pipetted on the stage of a Nanosight LM10 laser device (530 nm) and viewed with an intensified CCD camera (Andor-DL-658M-OEM) mounted on a Nanosight LM10. Tracking data were recorded for 160 s using the NTA Software version 2.2. All measurements were repeated at least three fold. In mixed particle populations camera settings were adapted such that the smallest particles could be recognized by the software. The numerical composition of mixed particle suspensions was studied by scanning electron microscopy (SEM). The aqueous suspensions were identical to those measured in Nanosight experiments but were at least 1000-fold less diluted. Suspension was dried on Thermonox slides and sputtered with a thin layer of gold (40 nm, Sputter Coater SI50B, Edwards, North Walsham, UK). Scanning electron microscopic examinations were done with the Gemini DSM 982 (ZEISS, Oberkochen, Germany, m 15 kV). At least 1000 particles from 10 different images were evaluated.

3.2. Virtual Polydisperse Suspensions. NTA measurements strongly depend on instrument and software settings such as camera gain, threshold mode and value, background subtraction, expected particle size, and other parameters. Also the homogeneity of the suspension influences size, distribution, and results are prone to be biased towards larger particles [11].

To eliminate these sources of error, “semireal” tracking data of a polydisperse suspension were generated from monodisperse tracking data. For this purpose, tracks from real monodisperse particle suspensions were gathered within one data file using predefined proportions (see Figure 1). The resulting semireal tracking data were used for statistical purposes and validating the methods described hereafter.

3.3. Simulation of Mono- and Polydisperse Suspensions. To analyse the properties of PSDs, a Monte-Carlo simulation of the Brownian motion of single particles with a specific diffusion coefficient was applied. According to Michalet [9], the step length probability density function (SPDF) of a particle with diffusion coefficient \( D \) can be described:

\[
P(s) = \frac{2s}{4D\tau} \exp\left(-\frac{s^2}{4D\tau}\right).
\]

Using the so-called transformation method [12], it is possible to generate random step lengths with the distribution of (4). Additionally, the length of a track (i.e., the number of steps) can be simulated by resampling NTA measurements. By applying a combined simulation of step length and track length, mono- as well as polydisperse SPDFs are easily simulated. Tracks of particles in monodisperse suspensions were simulated by defining the diffusion coefficient \( D \) of the particles and the total number of steps \( k \). The flow chart in Figure 2 describes the procedure of simulation. The resulting \( k \)-tuple of diameters was used for estimating the PSD.

Our approach for simulating size distributions of polydisperse suspensions is based on the procedure for monodisperse samples. For this purpose, we defined \( n \) diffusion
coefficients $D_1 \cdots D_n$ and the total number of steps $k_1 \cdots k_n$ for each coefficient. Thereafter, each combination $(D_i, k_i)$ was used with the procedure described for monodisperse samples. The resulting $n \times k$-tuples were merged to a single $(k_1 + \cdots + k_n)$ tuple. With this tuple, the PSDs of the polydisperse sample were estimated by a kernel density estimator.

3.4. Cluster Analysis. Finite Mixture Densities Models assume that a population is mixed of $c$ subpopulations (clusters) with various densities. These models are also used in medicine and biology for several purposes (e.g., clustering genes or detection of action potentials) [13–16]. A mixture of $c$ normal densities with different mean values $\mu$, variances $\sigma$, and population proportions $p$ is defined as follows:

$$f(x; p, \mu, \sigma) = \sum_{j=1}^{c} p_j N(x; \mu_j, \sigma_j).$$

All population proportions add up to 1. Also the measured size distributions are construable as a mix of normal densities with unknown parameters. For estimating the number of clusters $c$ and their parameters, the statistic software R 2.15 [17] with the MCLUST 3 [18, 19] package was used. MCLUST utilizes the Bayesian Information Criterion [20] and the Expectation-Maximization (EM) Algorithm [21] for determining the number of clusters and the density parameters, respectively. The initial parameters are estimated and then iteratively optimised by maximizing the log-likelihood up to a certain convergence criterion. If all clusters have the same variance, the number of possible local maxima decreases. With the MCLUST implementation of the EM-algorithm, we were able to account for the constraints of such parameters.

4. Results and Discussion

4.1. Limitations of Conventional PSDs. Normally, PSDs are computed on the basis of untransformed diameter data. To show that the resolution of single particle populations in conventionally calculated PSDs are limited, we used a Monte-Carlo simulation of a polydisperse suspension containing 50 nm, 100 nm, 150 nm, and 200 nm particles. Table 1 lists the absolute and relative counts of steps taken per particle population. The resultant PSD of the untransformed, weighted diameter data is shown in Figure 3. Although there were 1.8 times more 200 nm particles than 100 nm particles in the modelled suspension, the modal values had nearly the same...
Table 1: Composition of the polydisperse simulation shown in Figure 3.

| Diameter | 50 nm | 100 nm | 150 nm | 200 nm | ∑     |
|----------|-------|--------|--------|--------|-------|
| Number   | $1.2 \times 10^4$ | $2.2 \times 10^4$ | $4 \times 10^4$ | $4 \times 10^4$ | $8 \times 10^4$ |
| Proportion | 15% | 30% | 5% | 50% | 100% |

Figure 3: Simulated PSD of a polydisperse suspension.

Figure 4: Standard deviation depending on the mean size diameter of simulated (circles) and measured data (triangles).

4.2. Logarithmised Data and Cluster Analysis. It is a known property of the logarithm that it reduces heteroscedasticity of random variables, so it stabilizes the variance. For this reason, the diameter data were logarithmised to gain more comparability between two particle populations in the PSD of polydisperse suspensions. The logarithmised version of Figure 3 is shown in Figure 5. The modal values of black PSD of Figure 5 compared with the modal values of Figure 3 better represent the true proportions of the particle composition given in Table 1.

Despite the variance stabilizing transform, the 150 nm particle population remains hidden behind the 200 nm population. To uncover such hidden populations and determine quantitative values for the population ratios, an MCLUST cluster analysis was performed. It must be emphasized that the logarithmised diameters are better suited for a cluster analysis, because the variance between particle populations can be assumed as equal.

The result of such a cluster analysis is shown as colored PSDs in Figure 5. Clusters with a difference of the means less
than 7 nm were merged. Table 2 lists the cluster means and the proportions of the individual clusters. The proportions of the clusters 1, 3, 4, and 5 are in good agreement with the true proportions listed in Table 1. The clusters 2 and 6 are false clusters, which may appear in both simulations and experiments. False clusters represent, however, only small particle populations and could easily be filtered using a threshold proportion.

4.3. Validation by Virtual Polydisperse Suspensions. For the validation of our method described in Section 4.2, a virtual polydisperse suspension was generated from real measurements of four monodisperse suspensions of polystyrene particles with diameters of 50 nm, 100 nm, 150 nm, and 200 nm, respectively. Figure 6 shows the PSD of each measurement. The modal values of the measured PSDs were 54 nm, 99 nm, 142 nm, and 185 nm. The virtual polydisperse suspension

| $i$ | 1  | 2  | 3  | 4  | 5  | 6  |
|-----|----|----|----|----|----|----|
| Cluster mean | 49 | 71 | 99 | 144| 199| 284|
| Proportion | 14.5% | 1.5% | 28.1% | 7% | 47% | 1.7% |

Table 2: Results of the cluster analysis of a simulated polydisperse suspension as shown in Figure 5.

FIGURE 6: The PSDs of the monodisperse polystyrene particle with a diameter of 50 nm (a), 100 nm (b), 150 nm (c), and 200 nm (d) as measured with the Nanosight system.
was then generated according to the proportions given in Table 1. The results of the cluster analysis is shown in Figure 7 and the relative proportions in Table 3. It can be seen that the cluster mean values are in good accordance with the measured modal values of the monodisperse PSDs. Even the hidden cluster ($i = 3$) was detected, although the difference between calculated mean value and measured diameter was somewhat larger than observed for all other peaks. This may be due to the low number of tracks which were integrated for this subpopulation. The false cluster ($i = 5$) had the smallest proportion and might be filtered by a reasonable threshold setting of, for example, 3%. This virtual experiment shows that the properties of the measured PSDs were adequately reflected by the cluster analysis of logarithmised data.

4.4. Verification by Measurement. To verify the method experimentally, we prepared a defined suspension of 100 nm and 150 nm polystyrene particles with equal number concentrations of both particle types. Therefore, an SEM analysis of the mixed suspension was carried out, a representative micrograph of which is shown in Figure 8(b). The result of conventional NTA analysis is shown in Figure 8(a): the modal values were 101 nm and 134 nm. However, the density function shows different peak heights, suggesting a lower particle content of the larger particles. A cluster analysis was carried out for the logarithmised diameter data of Figure 8 and results are shown in Figure 8(c). Cluster means were 104 and 144 nm, respectively. The estimated proportions were nearly equal (1:1.04). This ratio was then used again for a simulation experiment of a bimodal suspension of particles with 101 nm and 134 nm diameter.

The untransformed PSD is shown in Figure 8(d). It can be seen that the ratio of the peaks in Figure 8(d) is nearly the same as in Figure 8(a). Due to the latter result and because the results were in good agreement with the SEM ratio, we conclude that the proportion had been correctly determined by the cluster analysis.

4.5. Applicability and Limitations of the Method. A fundamental assumption of the proposed method is that the standard deviations of the subpopulations in polydisperse suspension are equal after logarithmic transformation. If this is correct, the logarithmic transform is a good method to reduce the impact of the particle size on the broadening effect of the PSD. Another factor influencing the broadening of the PSD is the mean number of steps contributing to the analysed tracks because the standard deviation of $D$ decreases for longer tracks [22]. This is, of course, equivalent to the mean time interval during which a particle is successfully tracked.

Figure 9 illustrates the influence of the track length on standard deviation (SD) for untransformed (a) and logarithmised data (b) of four particle types. In contrast to the untransformed data, the SD of the logarithmised data is nearly identical for all four types of particles if the number of steps per track length is the same.

One may object that the different mean step lengths of different particle populations in a polydisperse suspension lead to different SDs. This would contradict the assumption of equal SDs and, consequently, would impair the cluster analysis. To avoid this difficulty, we removed all tracks with a length below 10 steps in the examples of this work, thus avoiding to track “random noise” in particle videos. The remaining tracks of the 100 nm, 150 nm, and 200 nm particle suspensions had a mean track lengths of 56, 37, and 65 steps, respectively. Regardless of some variations in SD which might be due to this disparity, the cluster analysis worked convincingly well with this data. Nevertheless, a cluster analysis may fail if data with larger variation in SD are being processed.

Accepting the above limitations, the methodological improvements suggested in this study may be helpful to analyse more complex and multimodal systems. For example, the apparent size of nanoparticles suspended in biological fluids such as serum or cell culture fluid increases due to aggregation and/or the formation of a protein corona [23]. These processes are highly dynamic and can be observed during single or repeated Nanosight measurements [24]. As changes of particle size or agglomeration state upon protein coating may be moderate, quantitative evaluation may benefit from our cluster analysis, provided that the number of peaks is not too high (e.g., see [25]). Here, we have used up to five simulated particle populations and found this a reasonable upper limit of complexity.

### Table 3: Results of the cluster analysis of the virtual polydisperse suspension.

| $i$ | 1   | 2   | 3   | 4   | 5   |
|-----|-----|-----|-----|-----|-----|
| Cluster mean | 52  | 98  | 133 | 183 | 277 |
| Proportion    | 14.9% | 27.1% | 9.3% | 47.4% | 1.1% |

[See Figure 7: Cluster analysis of a virtual polydisperse suspension.](#)
5. Conclusion

We presented a method for improving polydisperse particle size distributions based on the logarithmic transform of the estimated diameters to reduce the heteroscedasticity, which is partly due to the constant coefficient of variation of the diameter data. Transformed data were then subjected to a cluster analysis, that shows the ability to uncover hidden populations. Calculated cluster proportions were validated and verified by Monte-Carlo simulations of polydisperse suspensions, NTA measurements, and SEM images. The procedure appears helpful to correctly interpret the composition of polydisperse particle suspensions. The novel method for Monte-Carlo simulation of polydisperse suspension and the concept of virtual polydisperse suspensions seems to be useful for further investigations of the properties of polydisperse size distributions.

Conflict of Interests

The authors declare that they do not have a direct financial relation with the trademarks mentioned in their work that might lead to a conflict of interest.
Acknowledgments

This paper was supported by Grants of the German Federal Ministry of Education and Research (BMBF, NanoGEM Project, FKZ 03X0105G, and 03X0105H).

References

[1] J. G. Teegarden, P. M. Hindeliter, G. Orr, B. D. Thrall, and J. G. Pounds, "Particokinetics in vitro: dosimetry considerations for in vitro nanoparticle toxicity assessments," Toxicological Sciences, vol. 95, no. 2, pp. 300–312, 2007.

[2] C. Buzea, I. I. Pacheco, and K. Robbie, "Nanomaterials and nanoparticles: sources and toxicity," Biointerphases, vol. 2, no. 4, pp. MR17–MR71, 2007.

[3] Z. Yang, Z. W. Liu, R. P. Allaker et al., "A review of nanoparticle functionality and toxicity on the central nervous system," Journal of the Royal Society Interface, vol. 7, no. 4, pp. S411–S422, 2010.

[4] V. Stone, B. Nowack, A. Baun et al., "Nanomaterials for environmental studies: classification, reference material issues, and strategies for physico-chemical characterisation," Science of the Total Environment, vol. 408, no. 7, pp. 1745–1754, 2010.

[5] I. Montes-Burgos, D. Walczyk, P. Hole, J. Smith, I. Lynch, and K. Dawson, "Characterisation of nanoparticle size and state prior to nanotoxicological studies," Journal of Nanoparticle Research, vol. 12, no. 1, pp. 47–53, 2010.

[6] B. Carr, A. Malloy, and J. Warren, "Nanoparticle Tracking Analysis," IPT, pp. 38–40, 2008.

[7] H. Saveyn, B. de Baets, O. Thas, P. Hole, J. Smith, and P. van der Meeren, "Accurate particle size distribution determination by nanoparticle tracking analysis based on 2-D Brownian dynamics simulation," Journal of Colloid and Interface Science, vol. 352, no. 2, pp. 593–600, 2010.

[8] H. Qian, M. P. Sheetz, and E. L. Elson, "Single particle tracking. Analysis of diffusion and flow in two-dimensional systems," Biophysical Journal, vol. 60, no. 4, pp. 910–921, 1991.

[9] X. Michalet, "Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium," Physical Review E, vol. 82, no. 4, Article ID 041914, 2010.

[10] ASTM, "Standard Guide for Measurement of Particle Size Distribution of Nanomaterials in Suspension by Nanoparticle Tracking Analysis," 2012.

[11] R. F. Domingos, M. A. Baalousha, Y. Ju-Nam et al., "Characterizing manufactured nanoparticles in the environment: multimethod determination of particle sizes," Environmental Science and Technology, vol. 43, no. 19, pp. 7277–7284, 2009.

[12] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, 3rd edition, 2007.

[13] R. Benitez and Z. Nenadic, "Robust unsupervised detection of action potentials with probabilistic models," IEEE Transactions on Biomedical Engineering, vol. 55, no. 4, pp. 1344–1354, 2008.

[14] X. Dai, T. Erkkilä, O. Yli-Harja, and H. Lähdesmäki, "A joint finite mixture model for clustering genes from independent Gaussian and beta distributed data," BMC Bioinformatics, vol. 10, article 165, 2009.

[15] B. S. Everitt and E. T. Bullmore, "Mixture model mapping of brain activation in functional magnetic resonance images," Human Brain Mapping, vol. 7, no. 1, pp. 1–14, 1999.

[16] S. H. Meghani, C. S. Lee, A. L. Hanlon, and D. W. Bruner, "Latent class cluster analysis to understand heterogeneity in..."
prostate cancer treatment utilities,” *BMC Medical Informatics and Decision Making*, vol. 9, article 47, 2009.

[17] R Development Core Team, *R: A Language and Environment for Statistical Computing*, vol. 1, R Foundation for Statistical Computing, Vienna, Austria, 2011.

[18] C. Fraley and A. E. Raftery, “Model-based clustering, discriminant analysis, and density estimation,” *Journal of the American Statistical Association*, vol. 97, no. 458, pp. 611–631, 2002.

[19] C. Fraley and A. E. Raftery, *MCLUST Version 3 for R: Normal Mixture Modeling and Model-Based Clustering*, 2006.

[20] G. Schwarz, “Estimating the dimension of a model,” *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[21] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[22] M. J. Saxton, “Single-particle tracking: the distribution of diffusion coefficients,” *Biophysical Journal*, vol. 72, no. 4, pp. 1744–1753, 1997.

[23] M. Monopoli, C. Åberg, A. Salvati, and K. Dawson, “Biomolecular coronas provide the biological identity of nanosized materials,” *Nature Nanotechnology*, vol. 7, pp. 779–786, 2012.

[24] T. Wagner, S. O. Luettmann, D. Swarat, M. Wiemann, and H. G. Lipinski, “Image analysis of free diffusing nanoparticles in vitro,” in *Clinical and Biomedical Spectroscopy and Imaging II*, vol. 8087 of *Proceedings of SPIE*, May 2011, 808726.

[25] A. Pitek, D. O’Connell, E. Mahon, M. Monopoli, F. Bombelli, and K. Dawson, “Transferrin coated nanoparticles: study of the bionano interface in human plasma,” *PloS One*, vol. 7, no. 7, Article ID e40685, 2012.
