Kuba Tatarkiewicz PhD, R&D Global Director

How to present and compare data obtained by particle tracking analysis and other related methods

3-6-2019

Particle tracking analysis

- **Measuring sizes of individual particles**

  hydrodynamic diameter $d_h > d_0$ *due to diluent drag*

- **Counting tracked particles in a given volume**

  investigated volume depends on size and refractive index

- **Testing small volume of a sample**
  - statistical process of estimating particle size distribution
  - volume tested about 100,000x smaller than sample volume

Concentration measurements

- **Definition of C_N: counts per volume** [part/mL]

  other defs: mass C_M [mg/mL] or volume C_V [µL/mL]

- **Typically C_N is given for a range of sizes**

  e.g. between 100 nm and 1 µm diameters

- **Hence use of plots with size bins**
  - bin widths not necessarily equal
  - some software do not support unequal bins (notably Excel)

Statistical considerations

- **Error in counting proportional to $\sqrt{N}$**

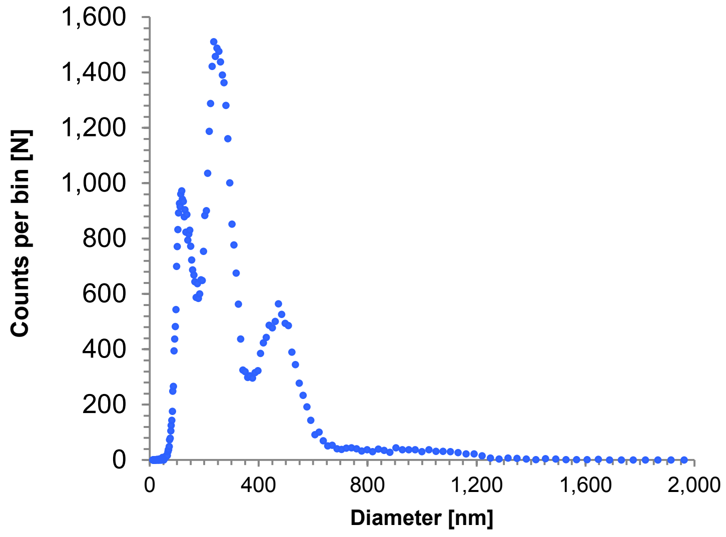
  this applies to total counts and individual bin counts

- **Number of bins depends on expected errors**

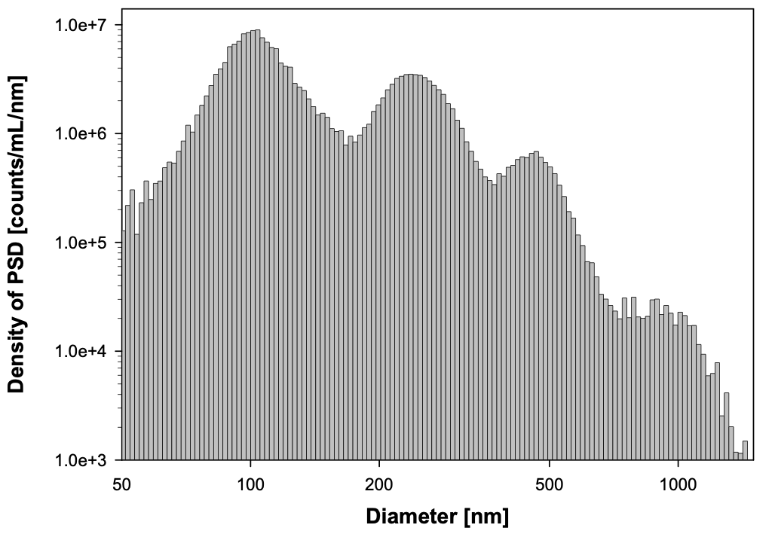
  N=10,000 and 100 bins, average error 10 counts/bin or 10%

- **To obtain "nice", smooth distributions:**
  - decrease number of bins and use wider bins
  - use narrow bins to calculate parameters of a distribution

Example plots of size distributions

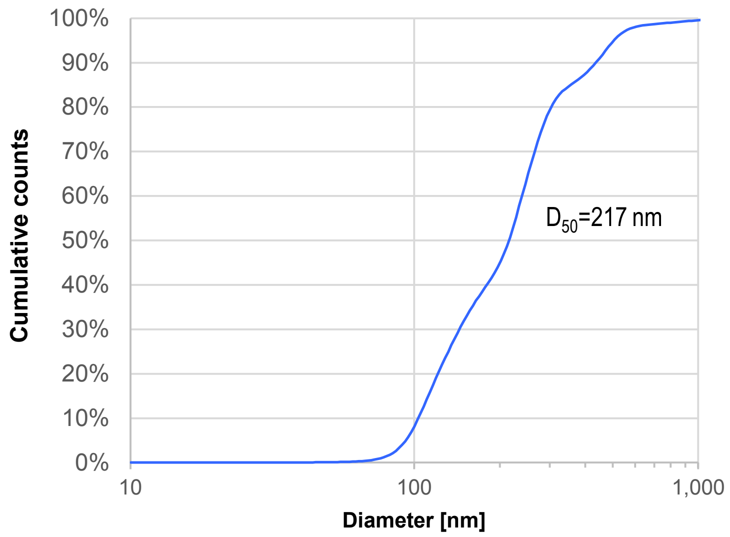


Counts



Density of PSD

logarithmic bins



Cumulative

Binning

- **Definition of a bin:**
  - $d \in [b_i, b_{i+1}[$ *where* $i=1,\ldots,N$
  - *equivalent definition:* $b_i \leq d < b_{i+1}$ *(non-overlapping bins)*
  - *typically* $b_1 = 0$ *and* $b_N = \text{max size to be measured}$
- Strange bins definition encountered:

		Average particles per mL	CV, %
29	particle diameter bin		
30	(50 to 59) nm		
31	(60 to 69) nm		
32	(70 to 79) nm		
33	(80 to 89) nm		
34	(90 to 99) nm		
35	(100 to 109) nm		

Define the particle diameter bin range, for example: 100 nm ≤ *d* ≤ 109 nm. The preferred bin width is 10 nm. The total range should cover from > 100 nm to < 2 µm. The range may be narrower, depending the range of the instrument being used. An example range is given here, but may be modified as appropriate for the measurement/instrument.

What is really plotted

- **Concentration is just a rational number, also per bin**
- **Histogram uses area to represent a distribution**
- **With unequal bins natural measurable is:**

density of particle size distribution (PSD)

units: number of particles per bin width and per investigated volume [part/nm/mL]

total concentration ≡ area of a histogram

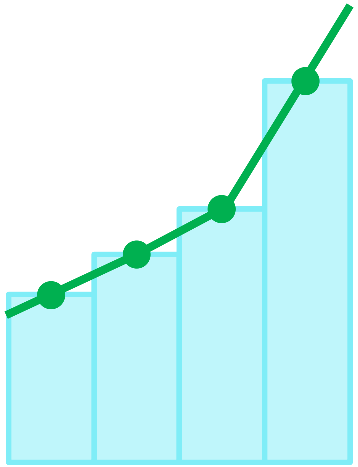
Dimensional analysis

- **When plotting concentration vs. size,**

  *DO NOT connect points**

  - **Total concentration is a sum of numbers, not an integral**
- **When plotting density of PSD, one can connect middle of bins** ☛ *linear approximation*
  - **Total concentration** C_N **is then**

    an area under the curve



**There's no data between points – by definition of a bin*

Practical procedure

- **Create a list of measured sizes**

  typically diameters of nanoparticles are given in [nm]

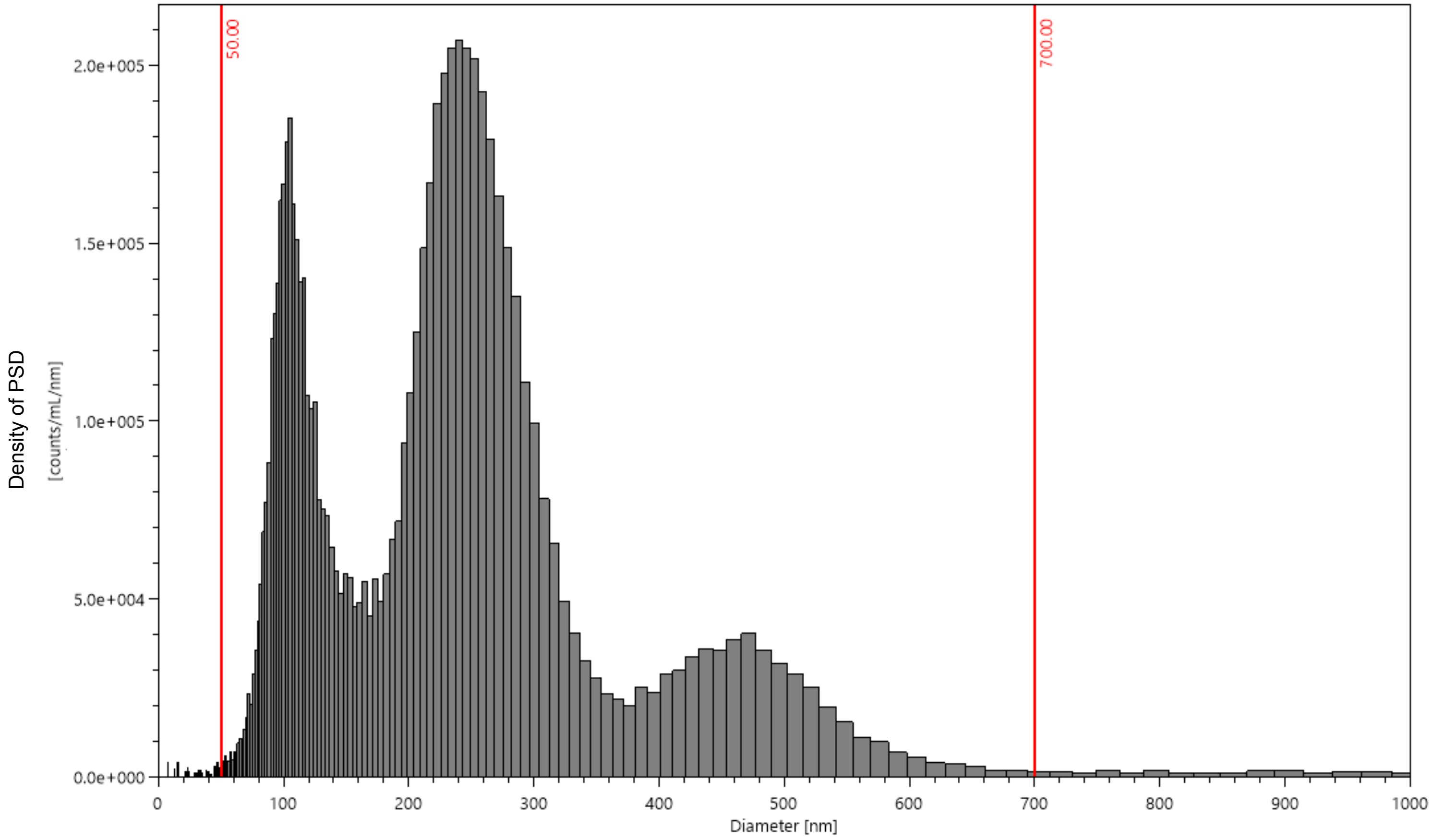
- **Bin those sizes in specific binning**

  typically logarithmic binning

$$(b_{i+2} - b_{i+1}) / (b_{i+1} - b_i) = const$$

- **Calculate density of PSD** [part/nm/mL]; V_0 **needed**
- **Plot as a histogram of density of PSD**
- **Calculate parameters of a distribution (AV, SD, D_{50} ...)**

Example of real data in logarithmic binning



Concentration from 50 nm to 700 nm = area of histogram of density of PSD

Real data vs. standards

- **Most standards of size are mono-disperse**

  there are no standards for concentration…

- **Typically real life samples are poly-disperse**

  e.g. multiple sizes or continuous distributions

- **Natural samples like sea water or blood:**
  - highly poly-disperse with dominating small particles
  - Junge distribution $N(d) \sim d^{-x}$ where $x=3.5 \div 4$

Fitted data problem (NanoSight FTLA)

- **Fitting unknown distribution of sizes**
- **Assumed binomial(s) distribution**
- **Height (or area) of different peaks is NOT conserved during fitting (not invariant of any fitting procedure)**
- **Looks great but lacks numerical accuracy**
- **Artificial peaks can be created** cf. van der Pol et al. J Thrombosis and Haemostasis, 12, 1182 (2014)

Parametric description of distributions

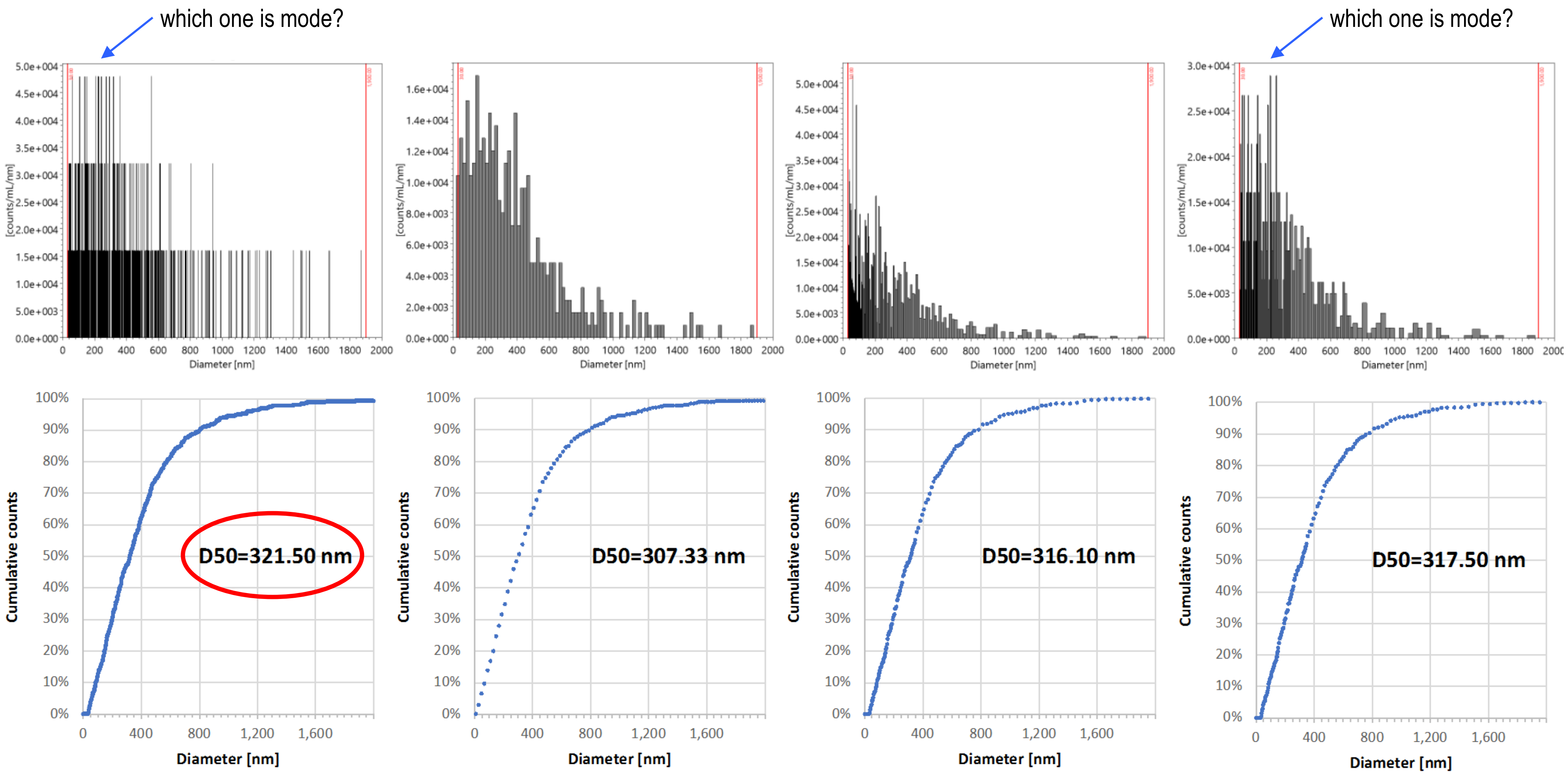
- **Concentration** $B_i = (b_{i+1} - b_i)$ $N_{total} = \sum_{i=1}^{N} \frac{n_i}{B_i} B_i = \sum_{i=1}^{N} n_i$

- **Average size** $d_{average} = \frac{1}{N_{total}} \sum_{i=1}^{N} n_i * (b_i + \frac{B_i}{2})$

- **Standard deviation** $SD = \sqrt{\frac{1}{N_{total}} \sum_{i=1}^{N} n_i * \left[d_{average} - \left(b_i + \frac{B_i}{2}\right)\right]^2}$

- **D_{50}** $D_k = \frac{1}{N_{total}} \sum_{i=1}^{k} n_i$ $definition\ D_k = 0.5$ ☛ k ☛ $b_k\ equal\ D_{50}$

D50 & mode depend on binning!



narrow equal bins | *wide equal bins* | *logarithmic bins* | *variable bins*

most accurate

Lies, damned lies, and statistics

- **Parametric description using AV and SD**

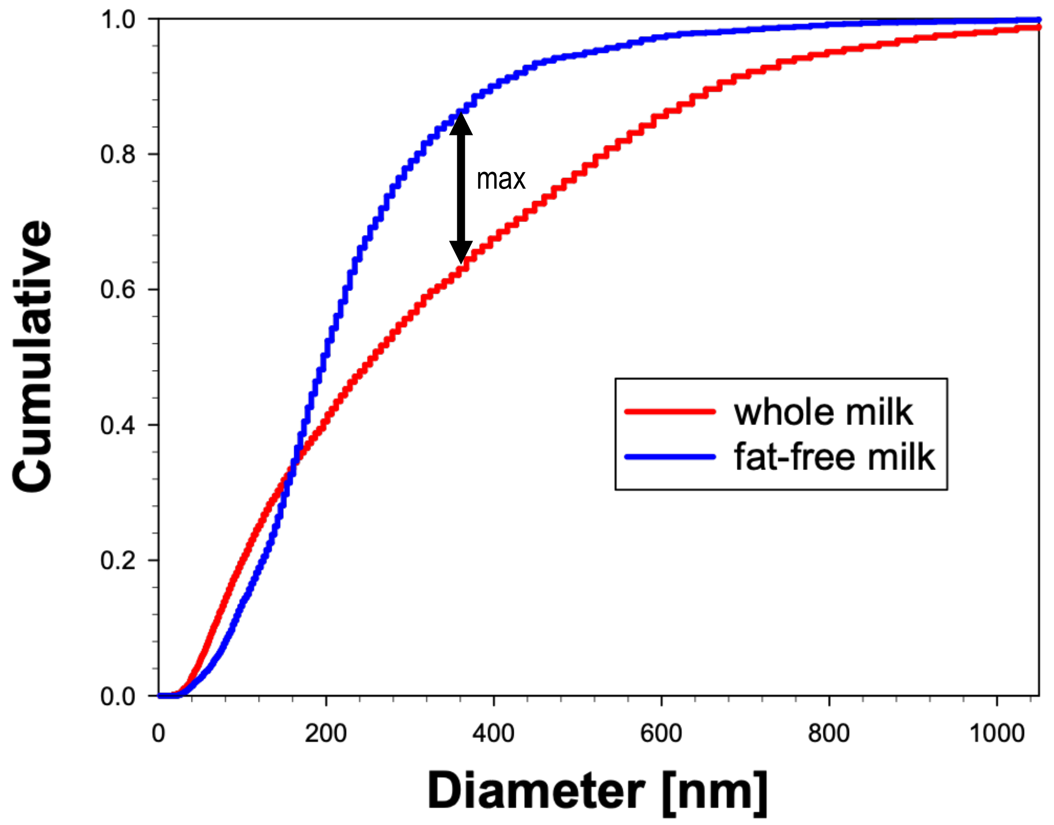
  is not accurate or unique!

  Anscombe's quartet – same AV and SD, various shapes

- **Even higher moments do not give enough info**
- **Basic experimental question:**

  How similar are two measured distributions?

Kolmogorov-Smirnov statistics



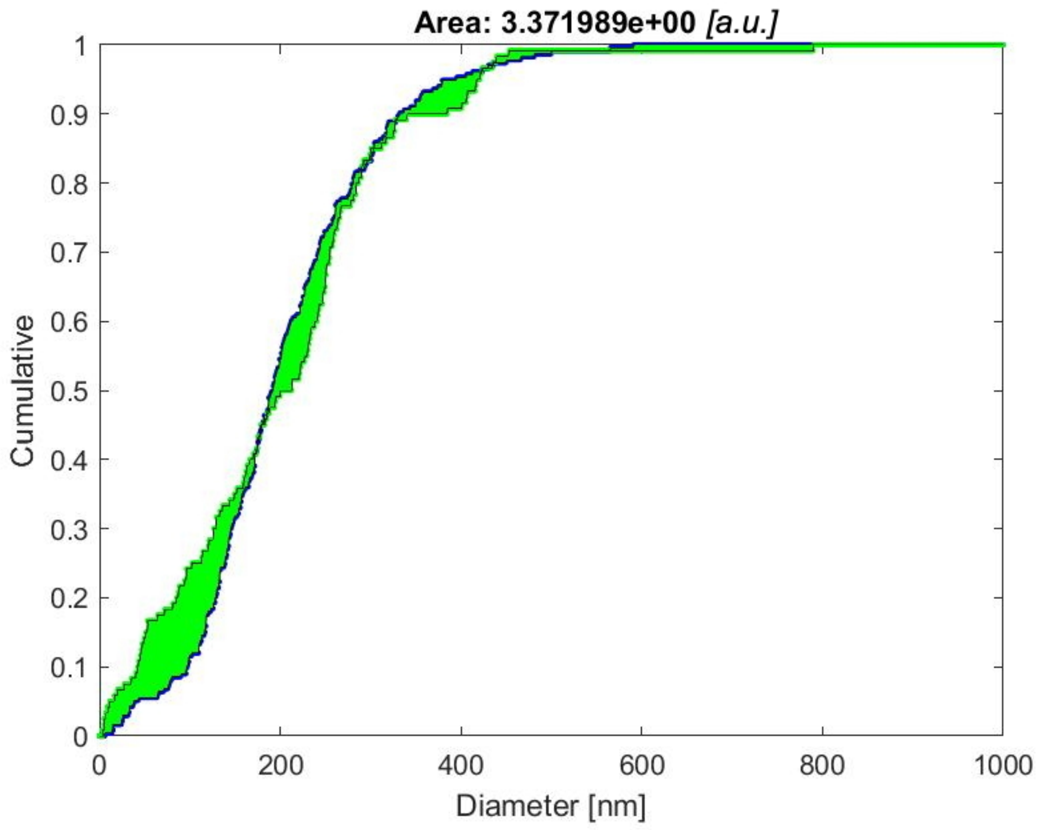
Comparing parameters:

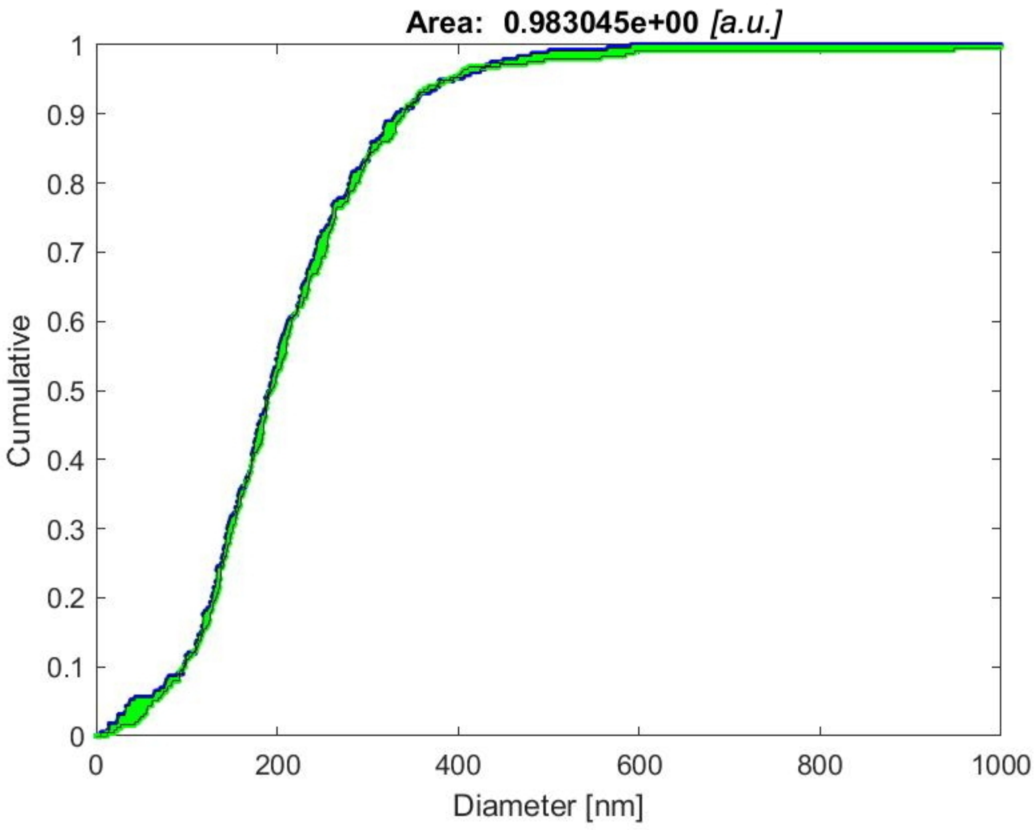
d_{av} = 256 nm, SD = 145 nm, CV = 0.57
d_{av} = 163 nm, SD = 68 nm, CV = 0.42

Non-parametric test:
Kolmogorov-Smirnov statistics

$D_{A,B}$	*alpha*	$D_{A,B,\alpha}$	*Reject?*
0.2335	0.05	0.0338	**yes**

Anderson-Darling statistics





Area between two cumulatives is a good measure of distance between two or more distributions with unknown shape.

Distance between distributions

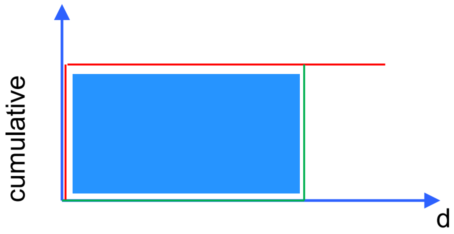
- **Normalize area between cumulatives**

  extreme case: a) particles at 10 nm, b) particles at 1000 nm

- **Distance is a number from the range $[0,1]$**

  same distributions have distance 0

  extreme case – distance 1



- **If areas are calculated between distributions in same normalization, it can be a good measure**

Thank you

Omoshiro-okashiku
Joy and Fun

감사합니다

Cảm ơn

ありがとうございました

Dziękuję

धन्यवाद

Grazie

Merci

谢谢

நன்றி

ขอบคุณครับ

Obrigado

Σας ευχαριστούμε

Tack ska ni ha

شُكْرًا

Большое спасибо

Danke

Gracias