**Determining particle size distributions from chord length measurements for different particle morphologies**

Jochen Schoell1\*, Roberto Irizarry2, Eric Sirota3, Cameron Mengel3, Lorenzo Codan1, and Aaron Cote3

*1 Process Research &Development, MSD Werthenstein BioPharma, 6105 Schachen, Switzerland
2 Data Science and Applied Mathematics, Merck & Co., Inc., West Point, PA 19486 USA
3 Process Research & Development, Merck & Co., Inc., Rahway, NJ 07065 USA.
\*Corresponding author: jochen.schoell@merck.com*

**Highlights**

* Data-driven model for CLD-to-PSD prediction
* Model required limited data and showed good generalization properties
* Applicable to different crystal morphologies (rod, compact, platelets)
* Comparison with model variations using PCA and geometric models

**1. Introduction**

In pharmaceutical crystallization, Focused Beam Reflectance Measurement (FBRM) has become a standard tool for in situ particle monitoring due to its ease of process implementation and ability to operate at high solids concentrations. However, a major limitation of this technique is that its measurement signal, i.e., the chord length distribution (CLD), differs fundamentally from particle size distribution (PSD) data due to the FBRM measurement principle described in literature [1,2]. Thus, the measured CLD is not only a function of particle number and size, but is generally also affected by particle morphology and various process parameters, e.g., probe positioning, particle concentration, and optical effects of the continuous and dispersed phase. The combination of these factors produces the actual measurement signal, which differs from a theoretically calculated CLD. Thus, a data-driven modeling approach has been proposed recently which allowed the quantitative determination of 1D and 2D PSDs from measured CLDs, suggesting an effective solution to the above-mentioned limitations of the FBRM [1]

**2. Methods**

The data-driven model architecture is based on three sequential steps:

1. CLD data is compressed into a small set of CLD descriptors (low order moments).
2. The CLD descriptors are mapped into a small number of PSD moments using a regression model.
3. Finally, the PSD moments are expanded into a full particle size distribution using a 2-layer network model.

Fundamental for the final step are parameterized functions (herein called generating functions) which reduce the number of model parameters. It is noteworthy that this model architecture has undetermined parameters in step 2 and step 3 which need to be estimated using a training data set to fit these parameters for the system being monitored. The model training was performed with three compounds featuring different particle morphology as shown in Fig. 1.



**Figure 1.** Three compounds featuring different crystal morphologies modeled with the data-driven model:
(a) compact, (b) platelet, and (c) rod morphology.

**3. Results and discussion**

A limited training set of four different PSDs each at four different solid concentrations proved to be sufficient for all three morphologies. Fig. 2 shows as example the comparison of the measured PSD and the PSD estimated from CLD and solids concentration data for the platelet compound:



**Figure 2.** Experimental laser diffraction PSDs for the platelet morphology shown as dots (dotted lines are a guide for the eye) together with corresponding PSD predictions from CLD and solid concentration data.

Model accuracy was assessed using the root mean square deviation between measured and simulated PSDs. The RMSD values were similar not only for all three compounds but also for development batches monitored years before compiling the training data, indicating robust generalization.

**4. Conclusions**

A data-driven CLD-to-PSD model was applied to compounds with different morphologies at slurry concentrations typical for industrial crystallization processes. In all cases, a limited experimental data set which can be generated in less than two days was sufficient to obtain a model which is able to accurately predict particle size distributions from CLD and solids concentration measurements. While the approach has thus far only been tested on three systems, the breadth of morphologies which yielded positive results suggests that this framework has the potential to have a broad application.

**References**

1. R. Irizarry, A. Chen, R. Crawford, L. Codan, J. Schoell. Chem Eng Sci. 164 (2017) 202-218.
2. N. Kail, H. Briesen, W. Marquardt. Part Part Syst Charact. 24 (2007) 184–192.