

Developing predictive models for batch cooling crystallization of APIs with limited data availability

ESCAPE36

Sheffield, 23/06/2026

Mauro Davanzo^a, Emanuele Tomba^b, Enrico Carlassare^b, Riccardo Motterle^b, Massimiliano Barolo^a, Zoltan K. Nagy^c, Fabrizio Bezzo^a

^a University of Padova, Padova PD, Italy

^b F.I.S. Fabbrica Italiana Sintetici S.p.a., Montecchio Maggiore VI, Italy

^c Purdue University, West Lafayette, Indiana, USA

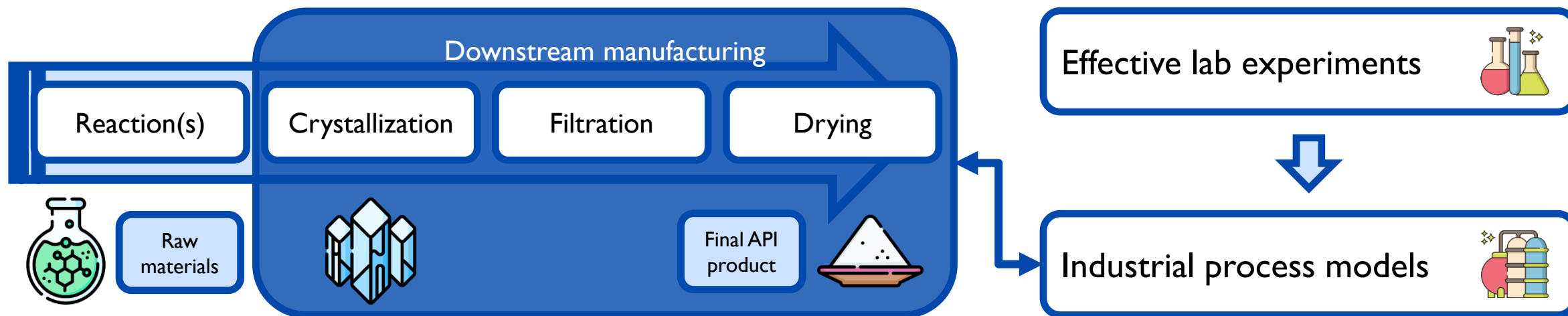
The big picture

Predictive models can accelerate the development and scale-up of API downstream manufacturing processes



Ensure quality

Save resources



From the objective to the starting point

OBJECTIVE



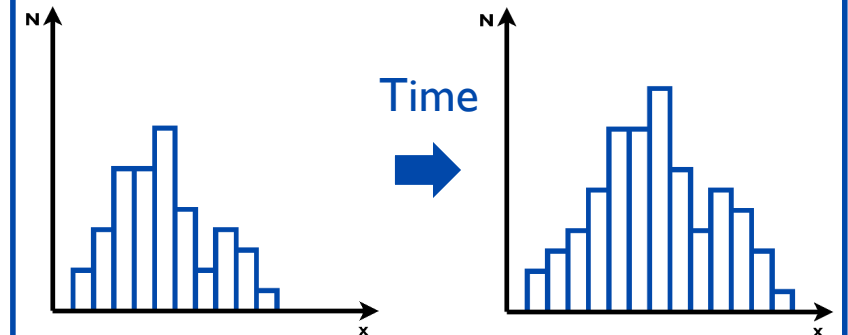
Develop a **strategy** to **calibrate predictive models** for batch cooling **crystallization units**, based on **limited information** from industrial lab experiments

Population balance models

How far can they predict final product attributes under limited measurements?

$$\frac{\partial n}{\partial t} + \nabla \cdot (v_e n) + \nabla \cdot (v_i n) + D - B = 0$$

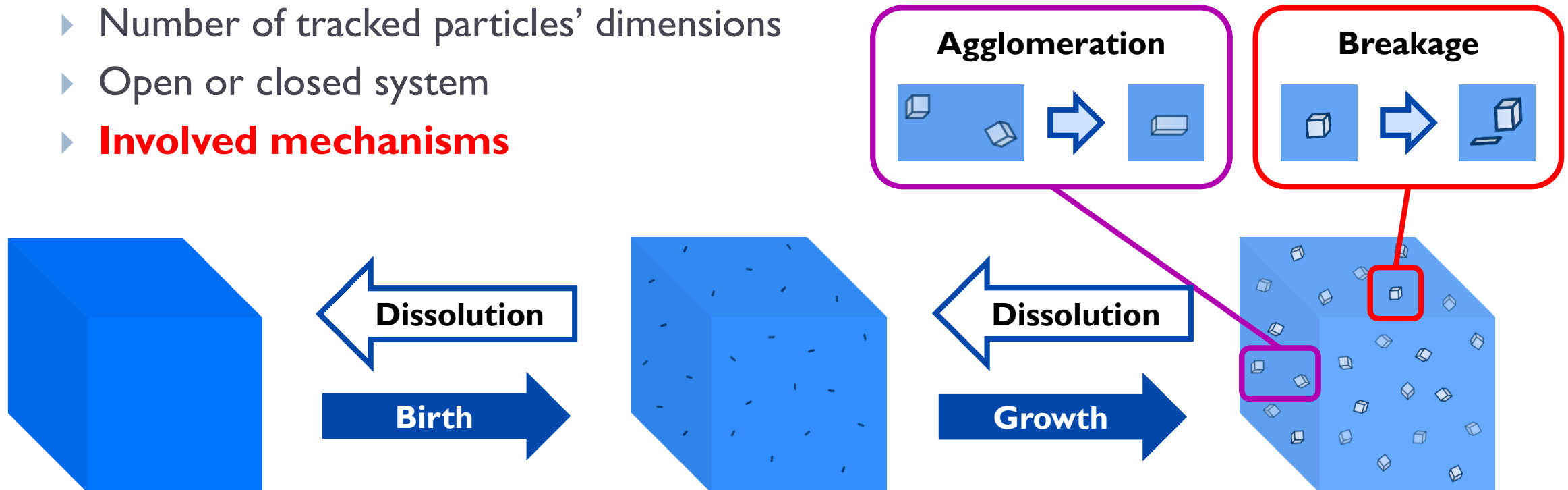
Particle Size Distributions (PSDs)



Critical Quality Attribute (CQA)

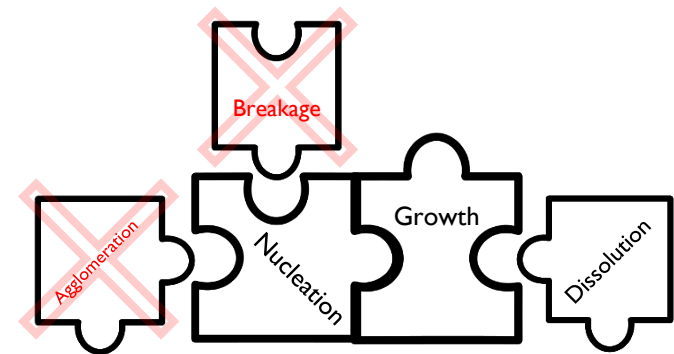
First main challenge: model formulation

- ▶ Crystallization is a complex phenomenon and formulation of population balance **equations** can **vary** significantly, depending on the specific system
 - ▶ Ideal or non-ideal mixing
 - ▶ Number of tracked particles' dimensions
 - ▶ Open or closed system
 - ▶ **Involved mechanisms**



First main challenge: model formulation

- ▶ Crystallization is a complex phenomenon and formulation of population balance **equations** can **vary** significantly, depending on the specific system
 - ▶ Ideal or non-ideal mixing
 - ▶ Number of tracked particles' dimensions
 - ▶ Open or closed system
 - ▶ **Involved mechanisms**

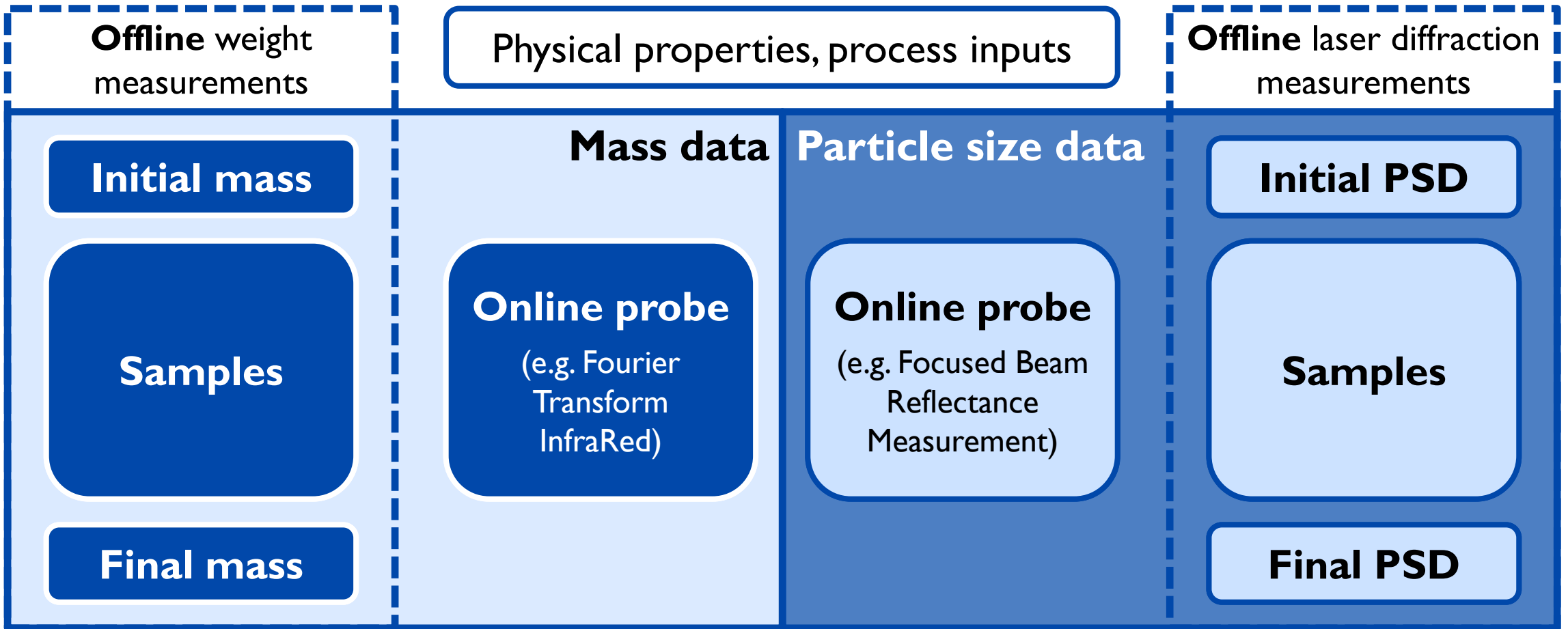


Calibration of population balance models is a **nontrivial** task, due to:

- ▶ Model structure and **numerical aspects**
- ▶ Availability and reliability of **measurements**



Second main challenge: data for model calibration

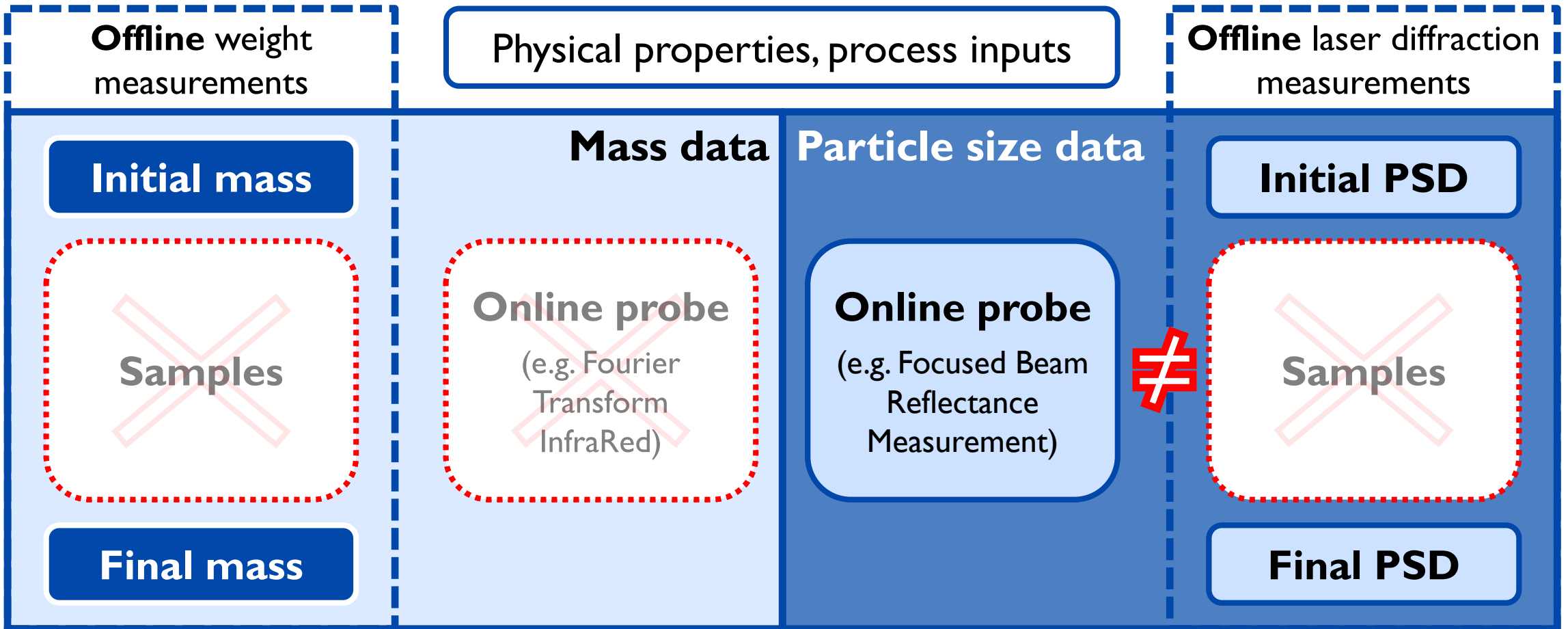


Ref.: Biri, D., Jaeggi, A., Prykhodko, O., Hicks, W., Perini, G., Mazzotti, M., Rajagopalan, A. K. (2025). *Powder Technology*. (466), 121405.

Ref.: Barhate Y., Kilari H., Wu W., Nagy Z. K. (2024). *Chemical Engineering Science*. (287), 119688.



Second main challenge: data for model calibration



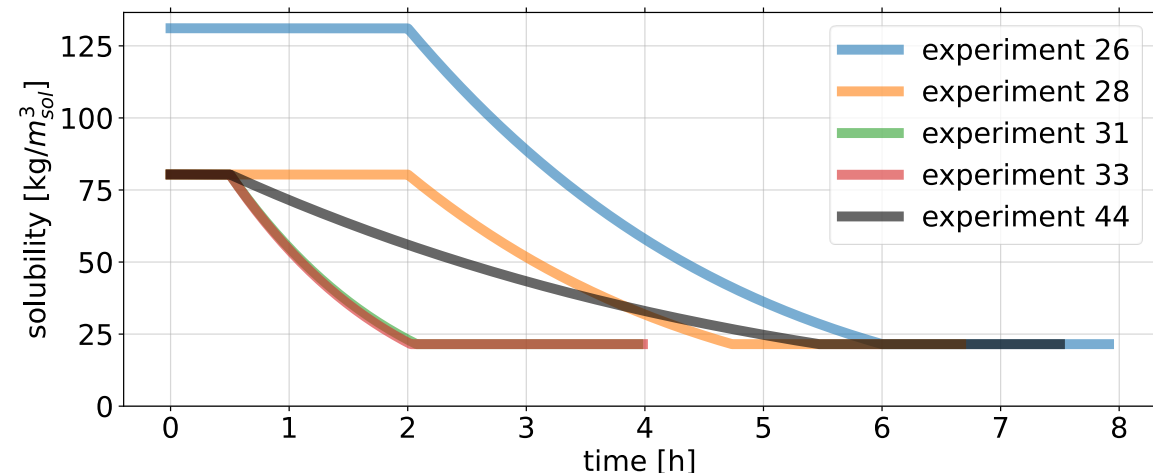
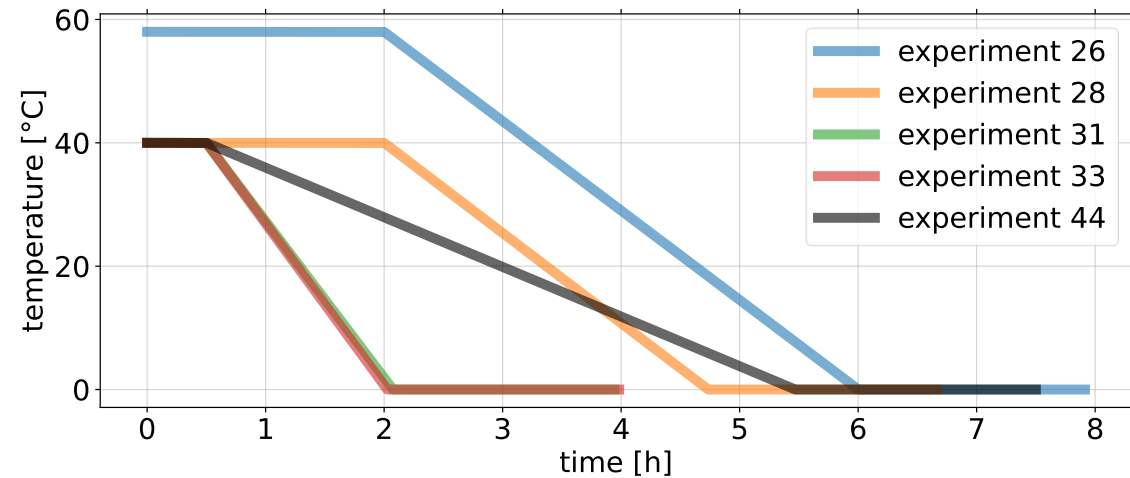
Ref.: Biri, D., Jaeggi, A., Prykhodko, O., Hicks, W., Perini, G., Mazzotti, M., Rajagopalan, A. K. (2025). *Powder Technology*. (466), 121405.

Ref.: Barhate Y., Kilari H., Wu W., Nagy Z. K. (2024). *Chemical Engineering Science*. (287), 119688.



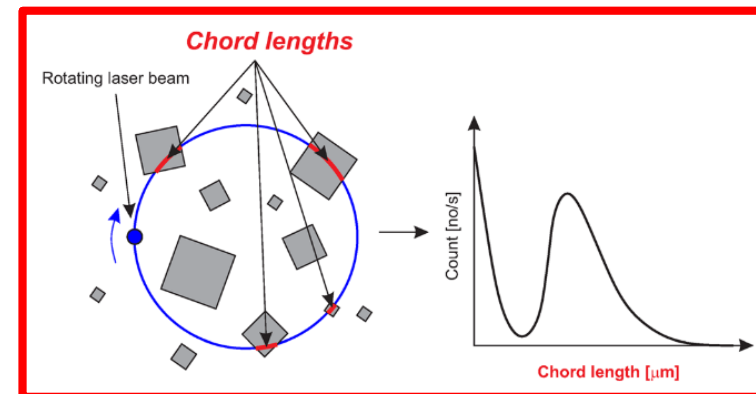
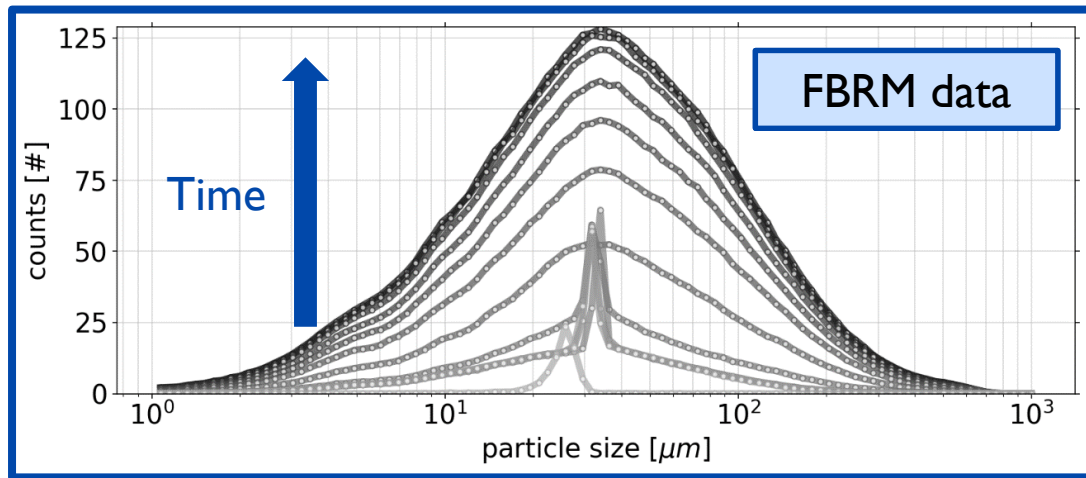
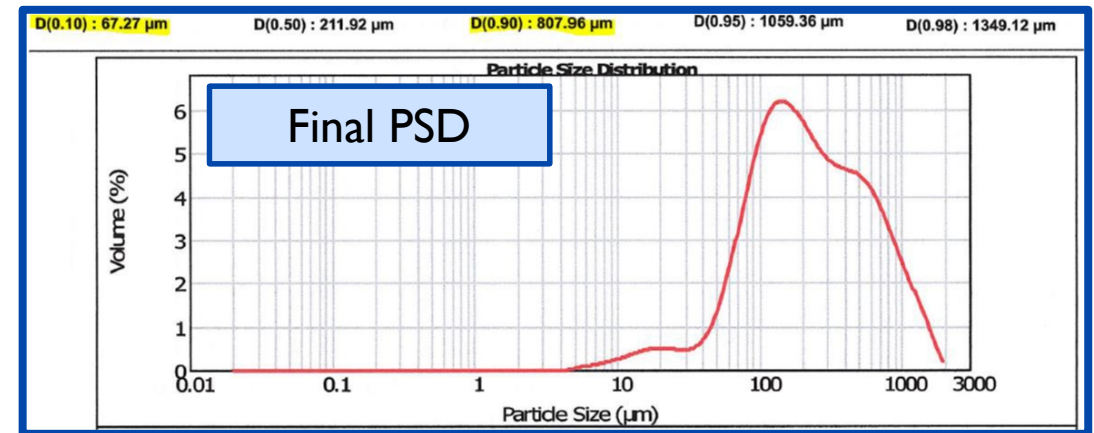
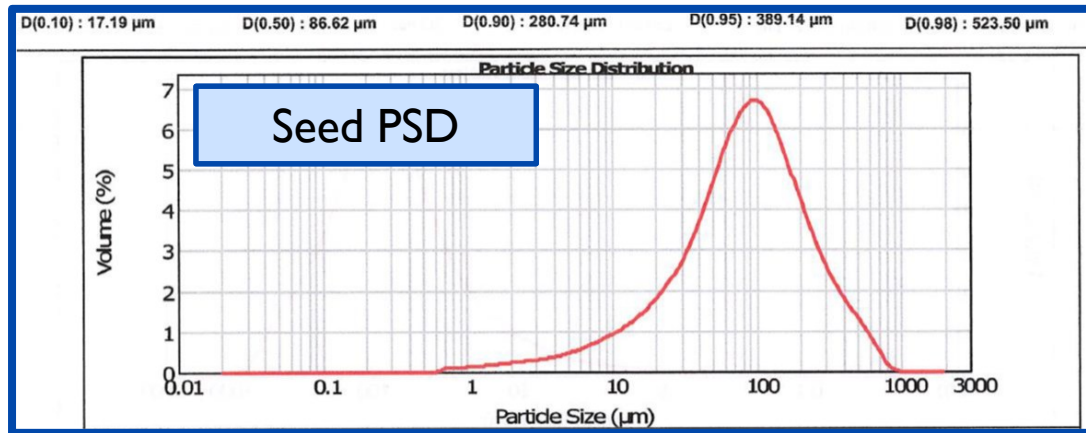
Industrial case study description

- ▶ **Antitumoral API** in organic solvent
- ▶ **Seeded batch cooling** crystallization
- ▶ **5 experiments** in different conditions
 - ▶ **Temperature ramps**
 - ▶ Initial concentration
 - ▶ Batch size
 - ▶ Seed amount



Available information on particle size

An example of comparison between different distributions



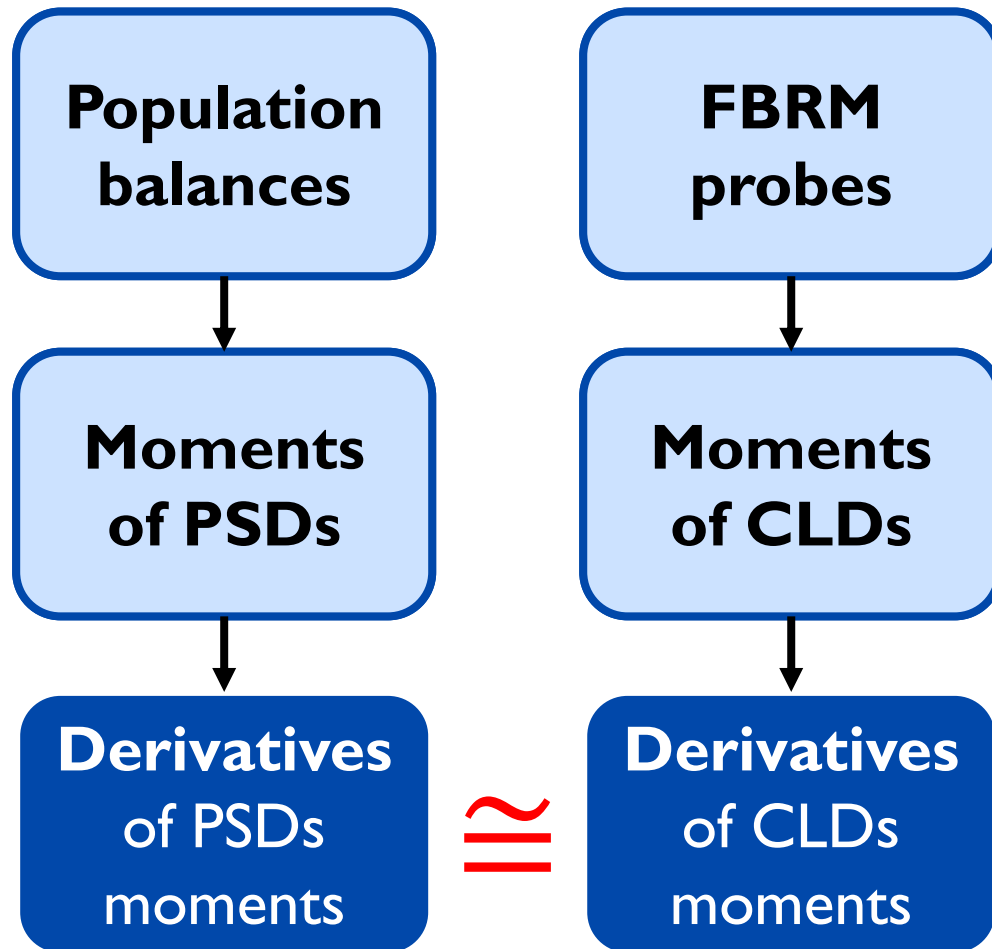
Particle Size
Distributions
(PSDs)



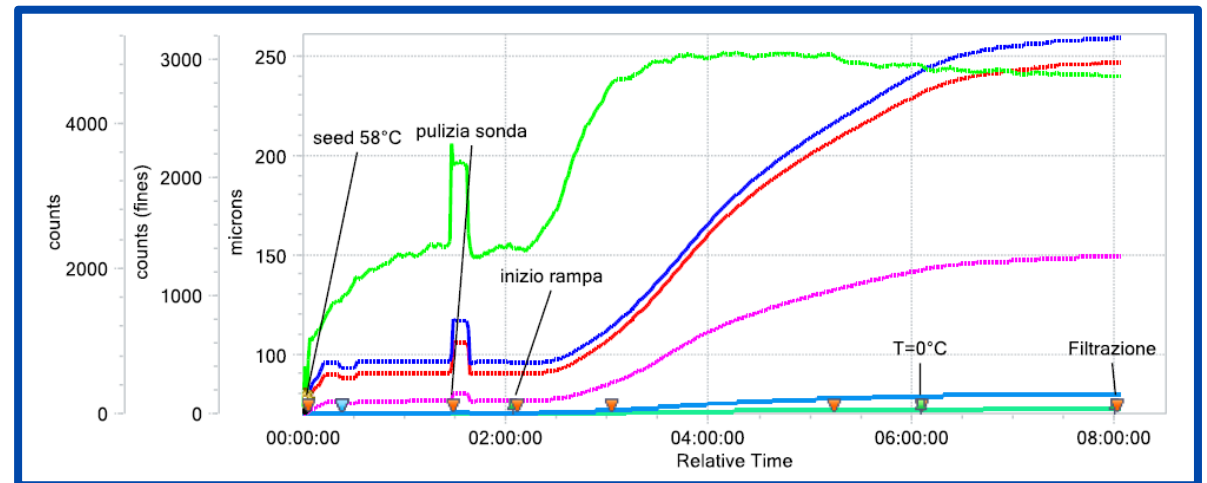
Chord
Length
Distributions
(CLDs)

Ref.: Szilágyi, B., Agachi, P. S., Nagy, Z. K. (2018). *Ind. Eng. Chem. Res.* (57), 3320-3332.

Proposed solution



- ▶ **Assuming** equality between **trends** of PSDs and CLDs moments, to exploit:
 - ▶ **Indirect information** on system's evolution
 - ▶ **Large amount** of data points



Model structure

Mathematical formulation

Population balance equations after applying method of moments

$$\begin{aligned} \triangleright \frac{d\mu_0}{dt} &= \mathbf{B} \cdot V_{slurry} \\ \triangleright \frac{d\mu_1}{dt} &= \mathbf{G} \cdot \mu_0 \\ \triangleright \frac{d\mu_2}{dt} &= 2\mathbf{G} \cdot \mu_1 \\ \triangleright \frac{d\mu_3}{dt} &= 3\mathbf{G} \cdot \mu_2 \end{aligned}$$

Mass balances for each component and variation in liquid volume

$$\begin{aligned} \triangleright \frac{dc_{prod}}{dt} &= -\frac{\rho_{crys} \cdot k_v \cdot (3\mathbf{G} \cdot \mu_2)}{V_{liq}} \cdot \left(1 - \frac{c_{prod}}{\rho_{liq}}\right) \\ \triangleright \frac{dc_{solv}}{dt} &= -\frac{\rho_{crys} \cdot k_v \cdot (3\mathbf{G} \cdot \mu_2)}{V_{liq}} \cdot \left(-\frac{c_{solv}}{\rho_{liq}}\right) \\ \triangleright \frac{dV_{liq}}{dt} &= -\frac{\rho_{crys} \cdot k_v \cdot (3\mathbf{G} \cdot \mu_2)}{\rho_{liq}} \end{aligned}$$

$$\mathbf{B} = k_{bs} \cdot S(T)^{s_1} \cdot \mu_3^{s_2}$$

$$\mathbf{G} = k_g \cdot S(T)^g$$

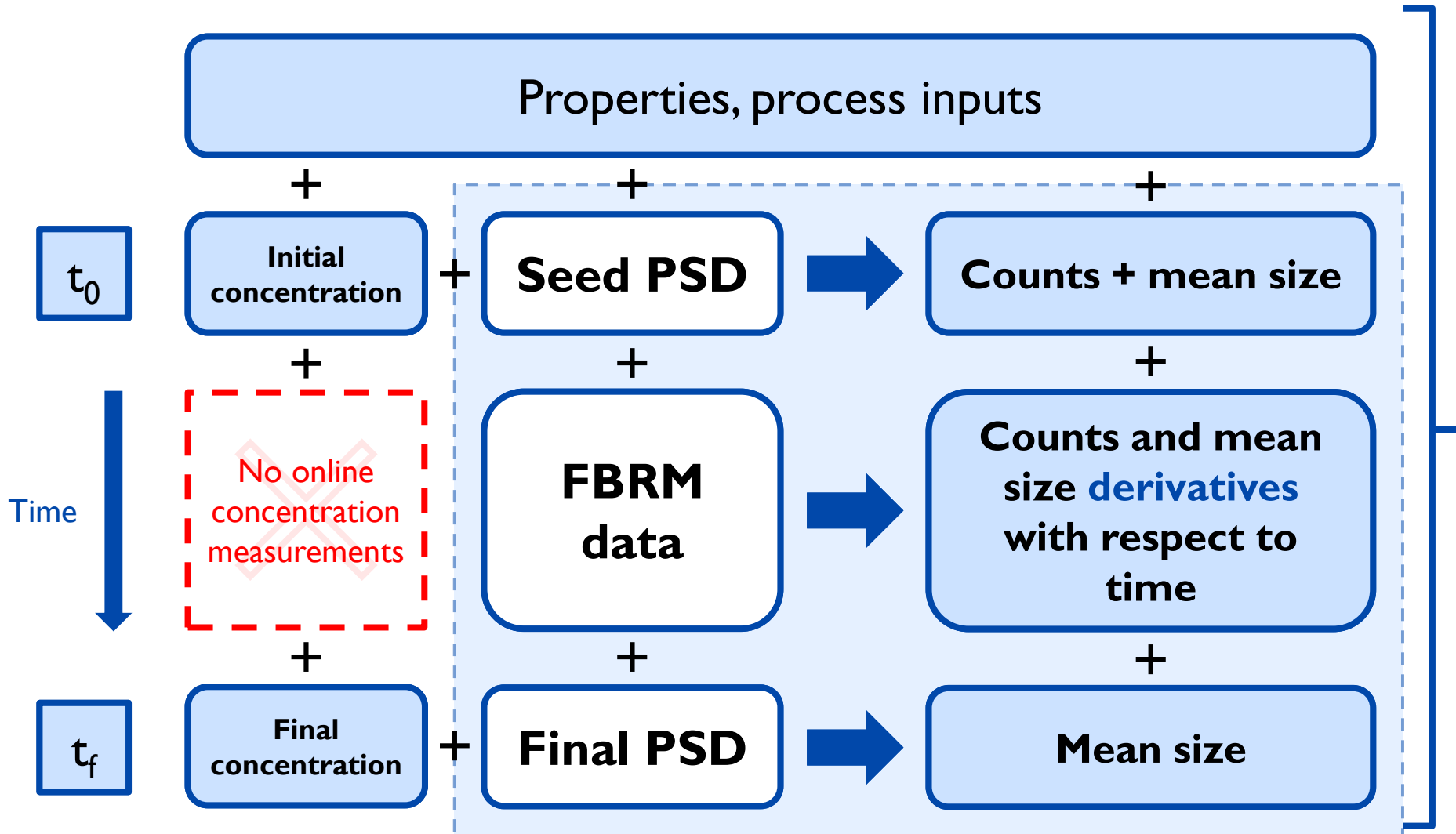
$$S = \frac{c - c_{sat}(T)}{c_{sat}(T)}$$

$$\ln(c_{sat}) = A + B/T$$

+

- ▶ Ideal mixing
- ▶ 1D particle size
- ▶ $r_0 = 0$
- ▶ $V_{slurry} = \text{constant}$
- ▶ No breakage
- ▶ No agglomeration

Strategy for model calibration



Parameter estimation

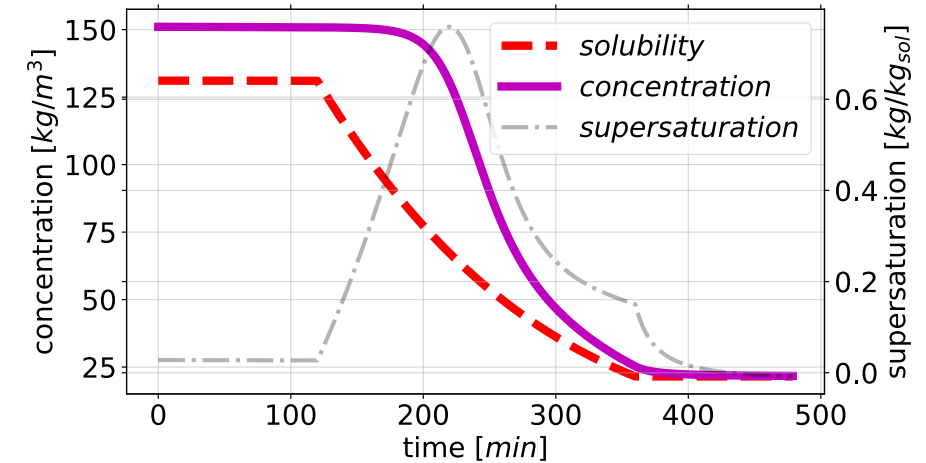
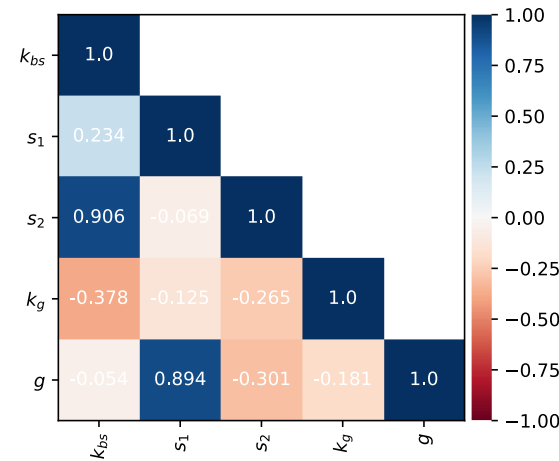
to predict main product features



Model calibration results

Parameter estimation assessment

$$\begin{aligned}
 WMSE = & \frac{w_0}{N_{CLD}} \sum_i^{N_e} \sum_j^{N_{CLD}} (\mu_{0,m,ij}' - \mu_{0,CLD,ij}')^2 \\
 & + \frac{w_1}{N_{CLD}} \sum_i^{N_e} \sum_j^{N_{CLD}} \left(\left(\frac{\mu_{1,m,ij}'}{\mu_{0,m,ij}'} \right) - \left(\frac{\mu_{1,CLD,ij}'}{\mu_{0,CLD,ij}'} \right) \right)^2 \\
 & + \frac{w_2}{N_{PSD}} \sum_i^{N_e} \sum_j^{N_{PSD}} \left(\left(\frac{\mu_{1,m,ij}}{\mu_{0,m,ij}} \right) - \left(\frac{\mu_{1,PSD,ij}}{\mu_{0,PSD,ij}} \right) \right)^2 \\
 & + \frac{w_3}{N_c} \sum_i^{N_e} \sum_j^{N_c} (c_{m,ij} - c_{ij})^2
 \end{aligned}$$



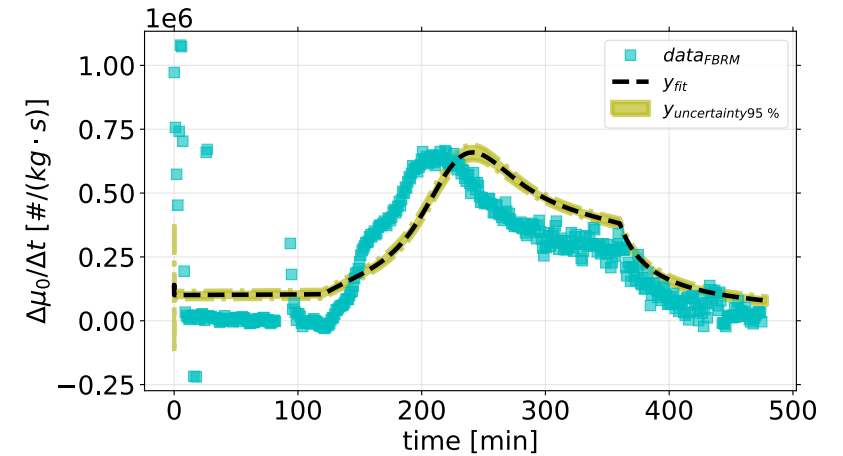
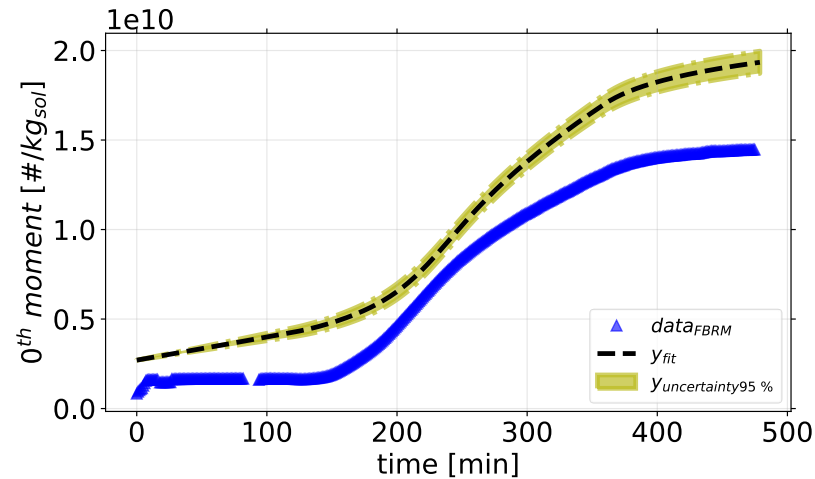
- ▶ Model was calibrated on all **5 experiments**
- ▶ All parameters were estimated with **good precision**
- ▶ $g > b_s$: growth is faster and requires higher levels of supersaturation to be triggered with respect to nucleation

$ t_{ref} = 1.96$	$k_{bs,ln}$	s_1	s_2	$k_{g,ln}$	g
Final value	21.79	0.54	0.57	-5.92	1.61
Std. dev.	0.07	0.05	0.03	0.04	0.09
 t-value 	333.72	11.37	17.96	165.57	17.18
WMSE	1.43E-01				

Model calibration results

Total counts fitting assessment

$$\begin{aligned} WMSE = & \frac{w_0}{N_{CLD}} \sum_i \sum_j^{N_e N_{CLD}} (\mu_{0,m,ij}' - \mu_{0,CLD,ij}')^2 \\ & + \frac{w_1}{N_{CLD}} \sum_i \sum_j^{N_e N_{CLD}} \left(\left(\frac{\mu_{1,m,ij}'}{\mu_{0,m,ij}'} \right) - \left(\frac{\mu_{1,CLD,ij}'}{\mu_{0,CLD,ij}'} \right) \right)^2 \\ & + \frac{w_2}{N_{PSD}} \sum_i \sum_j^{N_e N_{PSD}} \left(\left(\frac{\mu_{1,m,ij}}{\mu_{0,m,ij}} \right) - \left(\frac{\mu_{1,PSD,ij}}{\mu_{0,PSD,ij}} \right) \right)^2 \\ & + \frac{w_3}{N_c} \sum_i \sum_j^{N_e N_c} (c_{m,ij} - c_{ij})^2 \end{aligned}$$

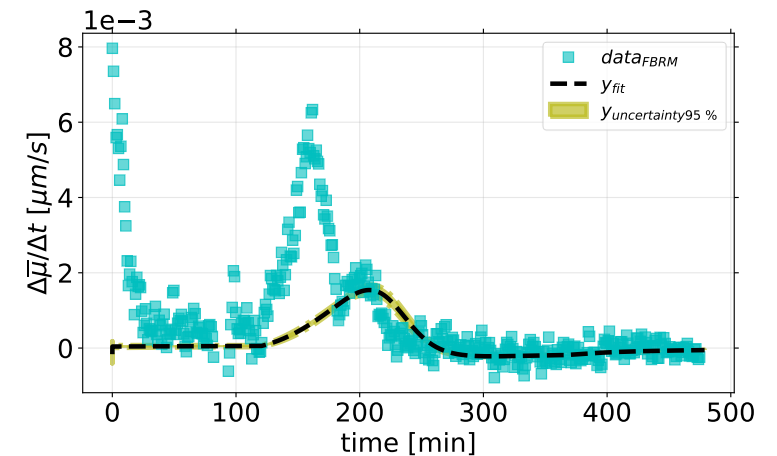
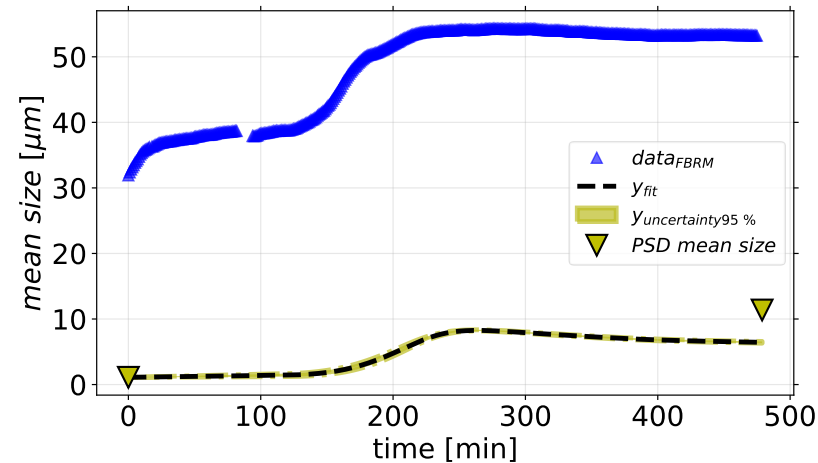


- ▶ Trends in the data were generally **well captured** by the model
 - ▶ especially in the **central part** of the batch
- ▶ **Secondary nucleation** gave the best results among different mechanisms that were tested
- ▶ The result was a sort of **averaged behaviour** among different batches, and different parts of the batches

Model calibration results

Mean size fitting assessment

$$\begin{aligned} WMSE = & \frac{w_0}{N_{CLD}} \sum_i^{N_e} \sum_j^{N_{CLD}} (\mu_{0,m,ij}' - \mu_{0,CLD,ij}')^2 \\ & + \frac{w_1}{N_{CLD}} \sum_i^{N_e} \sum_j^{N_{CLD}} \left(\left(\frac{\mu_{1,m,ij}'}{\mu_{0,m,ij}'} \right) - \left(\frac{\mu_{1,CLD,ij}'}{\mu_{0,CLD,ij}'} \right) \right)^2 \\ & + \frac{w_2}{N_{PSD}} \sum_i^{N_e} \sum_j^{N_{PSD}} \left(\left(\frac{\mu_{1,m,ij}}{\mu_{0,m,ij}} \right) - \left(\frac{\mu_{1,PSD,ij}}{\mu_{0,PSD,ij}} \right) \right)^2 \\ & + \frac{w_3}{N_c} \sum_i^{N_e} \sum_j^{N_c} (c_{m,ij} - c_{ij})^2 \end{aligned}$$



- **Offsets** in the plots of non-differentiated quantities were **expected**
 - Model was trained on **derivatives** of CLD data, but **initial conditions** were set based on laser diffraction measurement
- Prediction of final average size was **underestimated**
 - Increase in average size in the first half of the batch was **not captured**

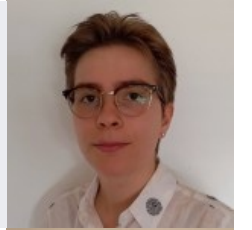
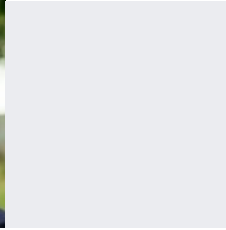
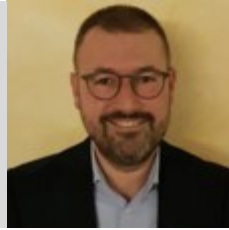
Conclusions and future work

- ▶ The presented strategy could represent an answer to the **challenges** that usually arise in the **crystallization modelling** exercise
 - ▶ **Statistically precise** estimations of kinetic parameters
 - ▶ **Satisfactory predictions** of Critical Quality Attributes
- ▶ Despite its limitations, the model could be useful for **industrial purposes**
 - ▶ to guide **early-phase process development**



- ▶ **Future (and present) work...**
 - ▶ Tests on datasets for **different systems**, with larger information content
 - ▶ Assessment of impact of **measurements availability** on model calibration
 - ▶ **Continuous improvement** of modelling strategy



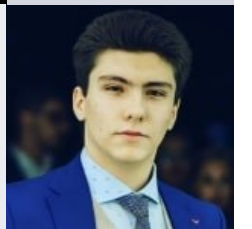
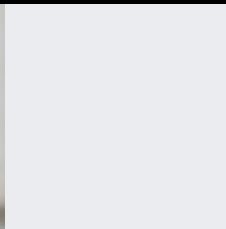
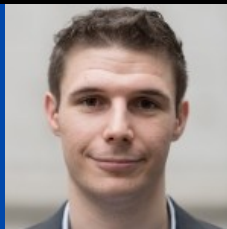
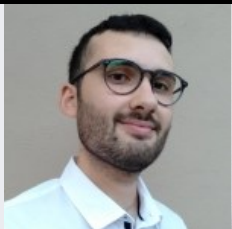
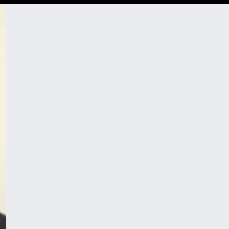
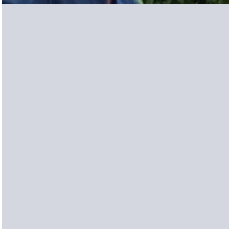
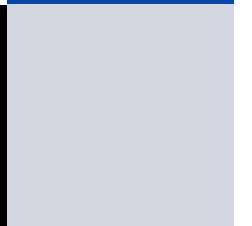


Department of Industrial Engineering
University of Padova - Italy

<https://research.dii.unipd.it/capelab>

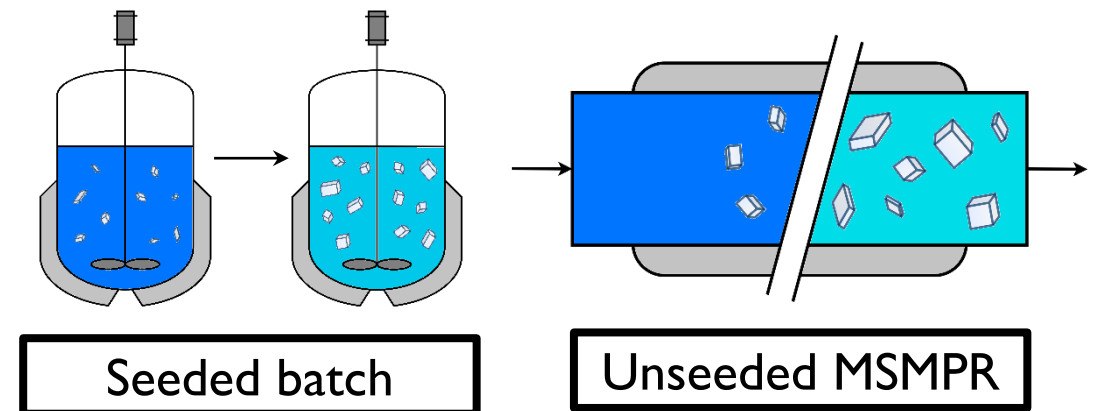
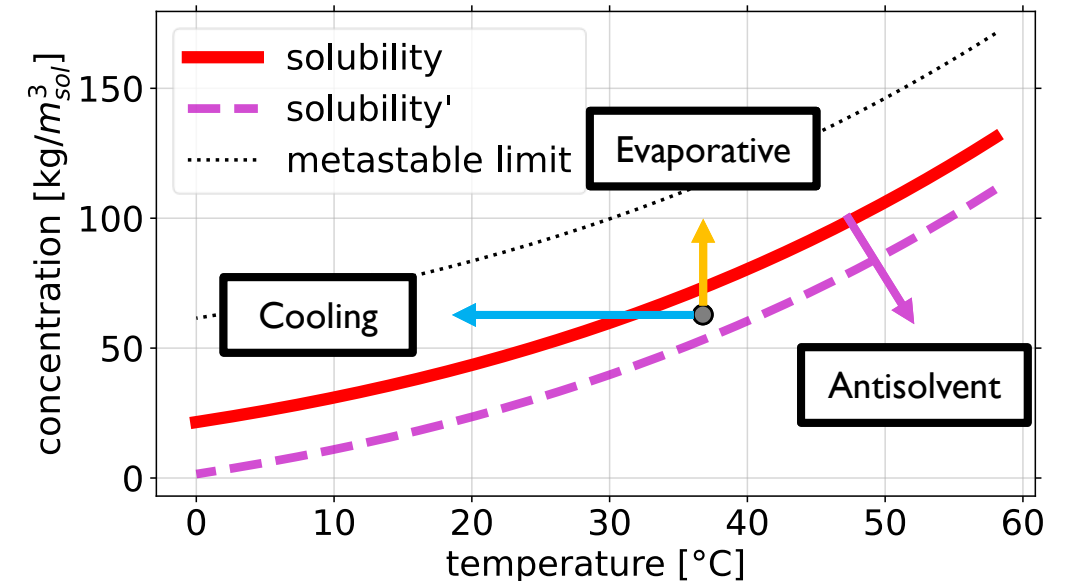


>_ Boosting innovation!



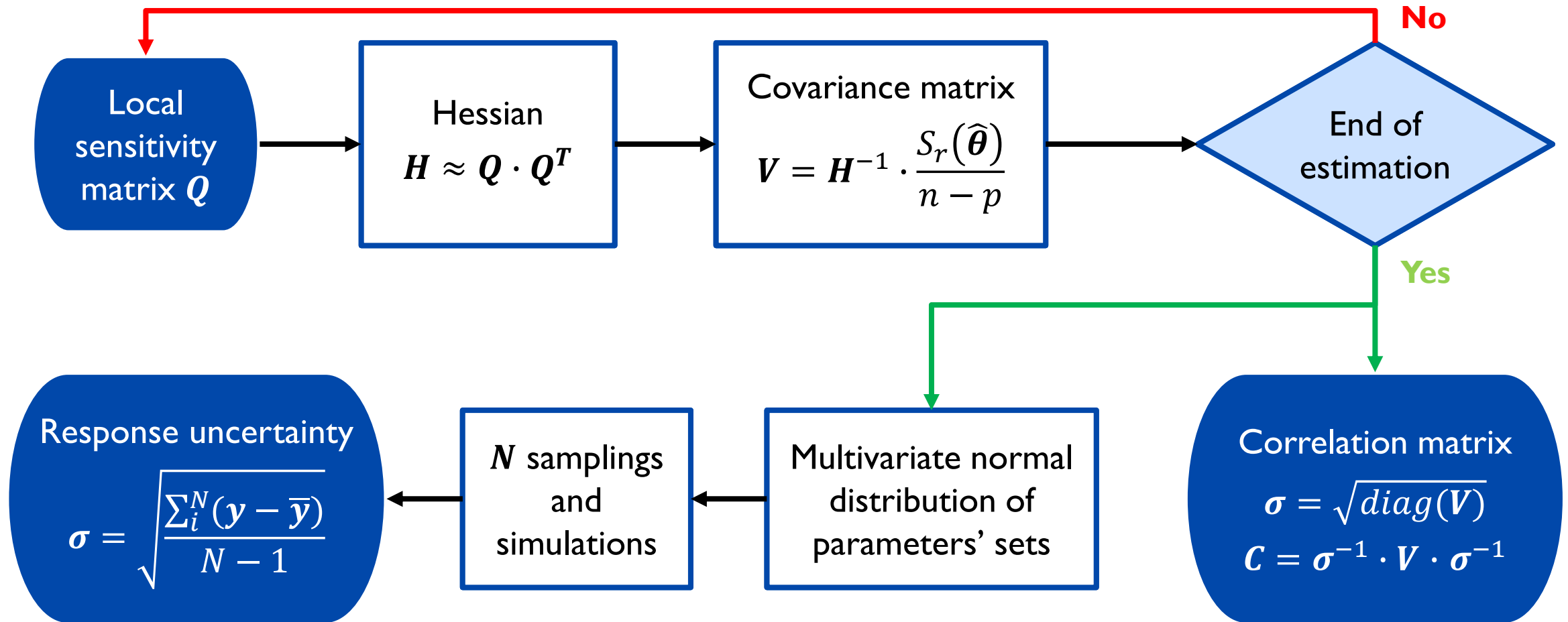
The crystallization process

- ▶ **Separation** unit operation
 - ▶ Form stable **solid phase**
 - ▶ Reject impurities
 - ▶ Control **Critical Quality Attributes**
 - ▶ Particle Size Distribution (PSD), purity
- ▶ Different **operating conditions**
 - ▶ **Supersaturation** creation
 - ▶ Cooling, antisolvent, evaporative, reactive
 - ▶ Seeded vs unseeded
 - ▶ Batch vs continuous (MSMPR)



Computation of relevant statistical properties

Step by step procedure

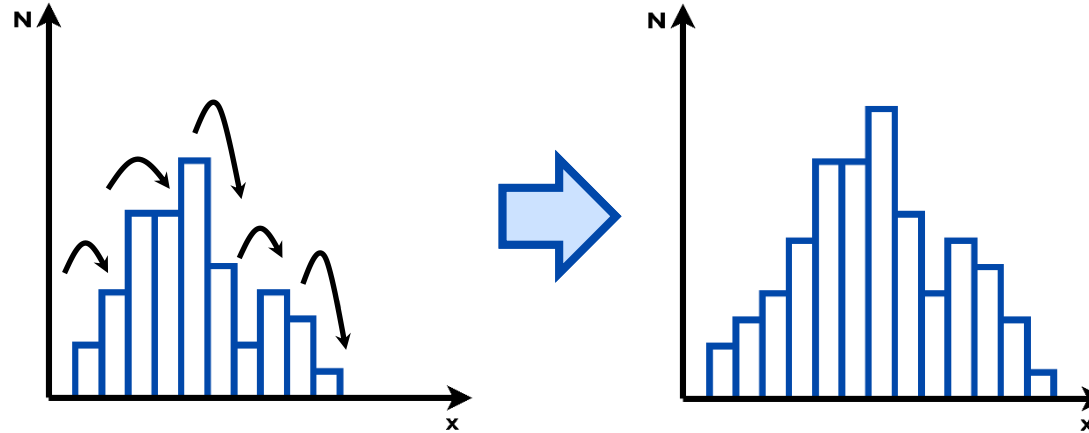


Ref.: Franceschini, G. and Macchietto, S. (2008). *Chemical Engineering Science* (**63**), 4846-4872.

Population balance models

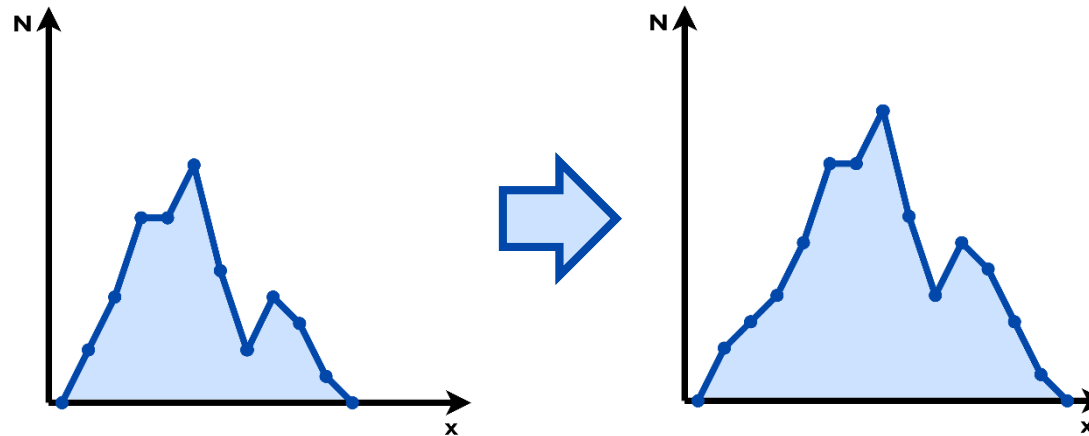
The application

Method of classes



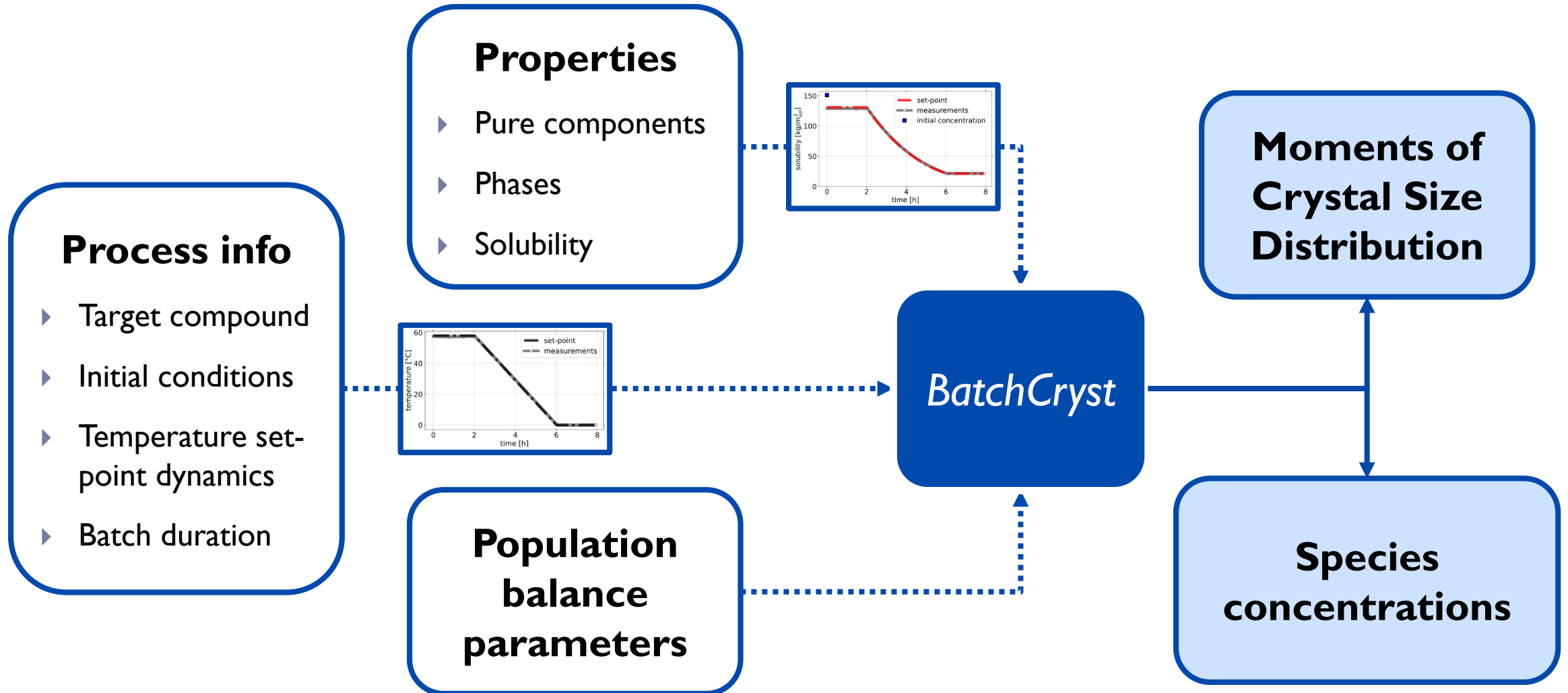
Tracks the evolution of number of particles in each size class

Method of moments

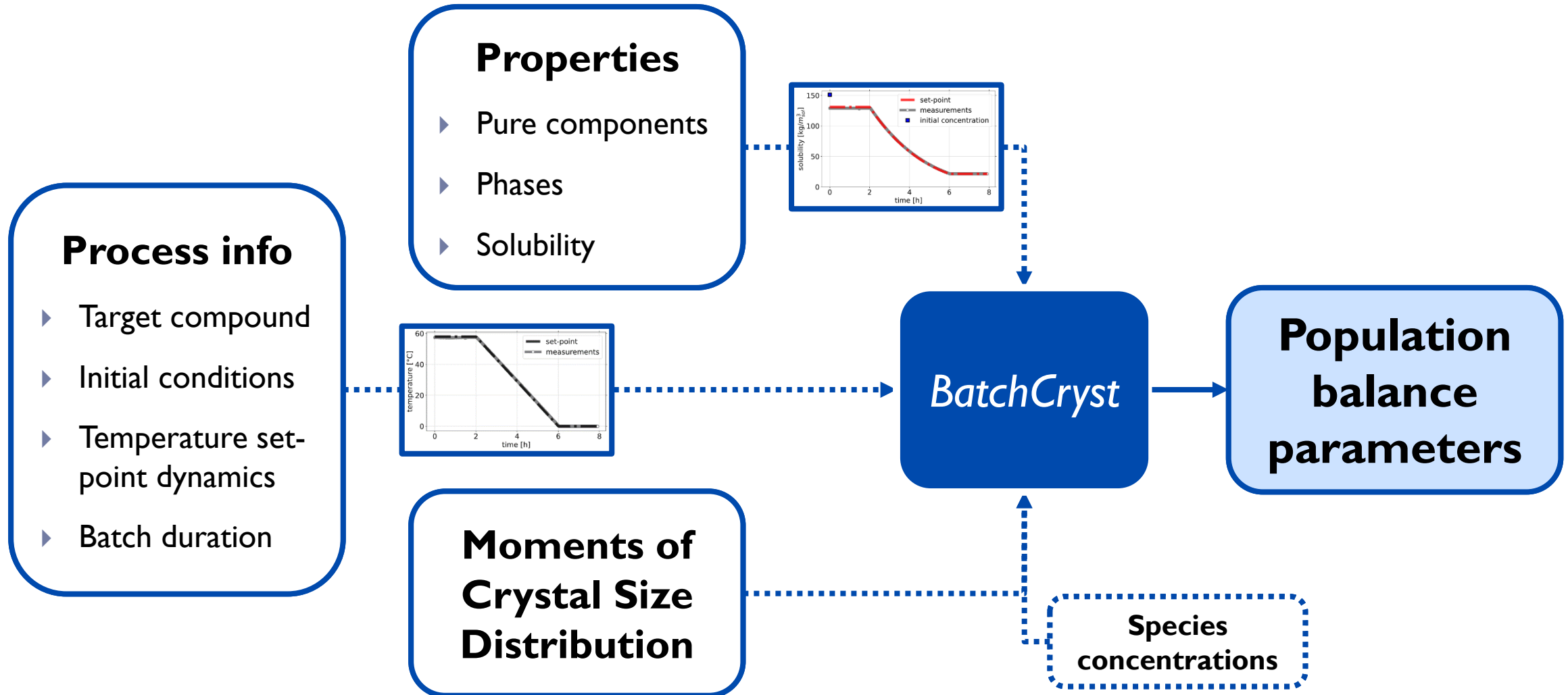


Tracks the evolution of indices that summarize PSD shape

Direct problem: simulation

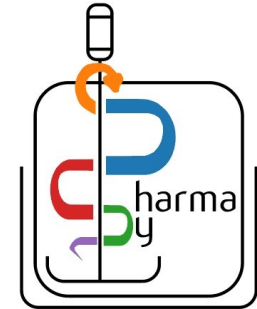


Inverse problem: parameter estimation



What is PharmaPy?

- ▶ Open-source **Python library** to analyze pharmaceutical processes
 - ▶ Fully **dynamic models** (ODEs & DAEs)
 - ▶ Main **unit operations** of pharmaceutical manufacturing
 - ▶ **Sequential-modular** and object-oriented approach



Simulate

flowsheets or
units

Estimate

kinetic
parameters

Optimise

Processes and
operations

Customise

equations and
routines

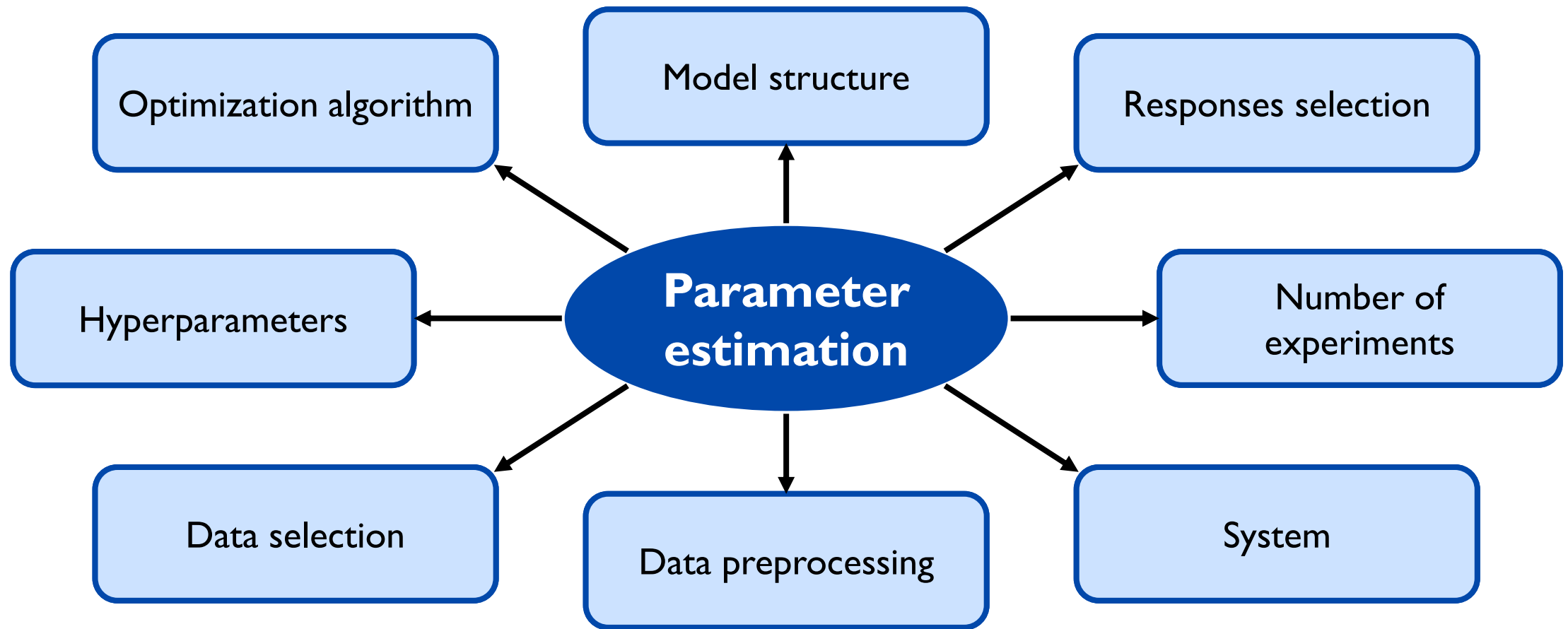
Debugging, adaptations and contributions to the code

Ref.: Casas-Orozco, D. et al. (2021). *Computers and chemical engineering*, **153**, 107406.



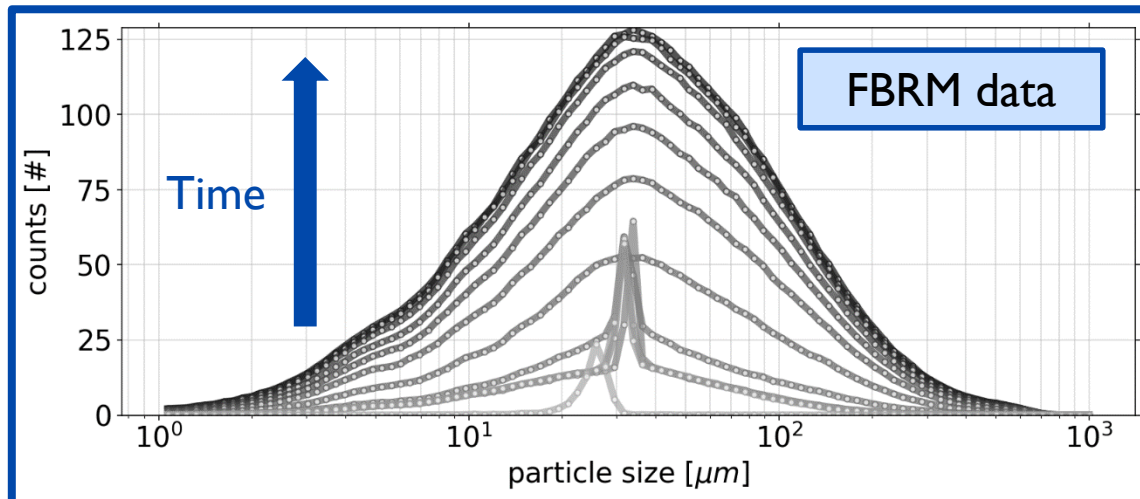
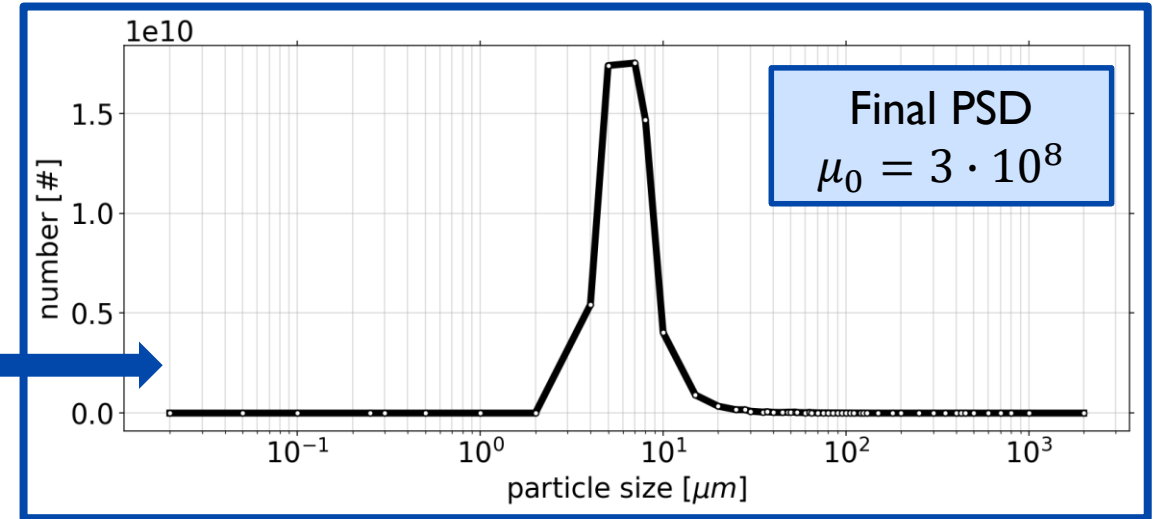
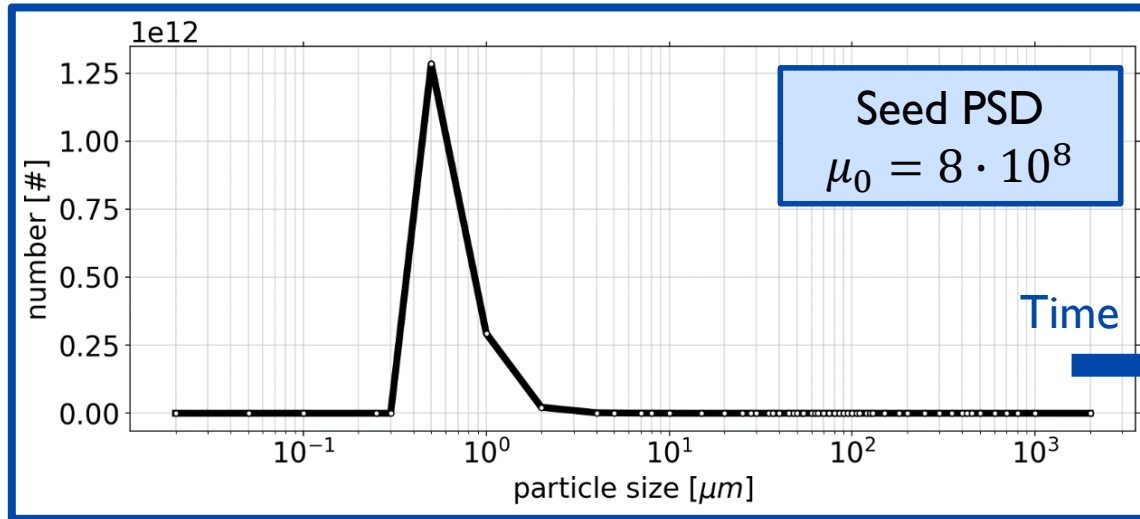
Parameter estimation tests

Possible variations



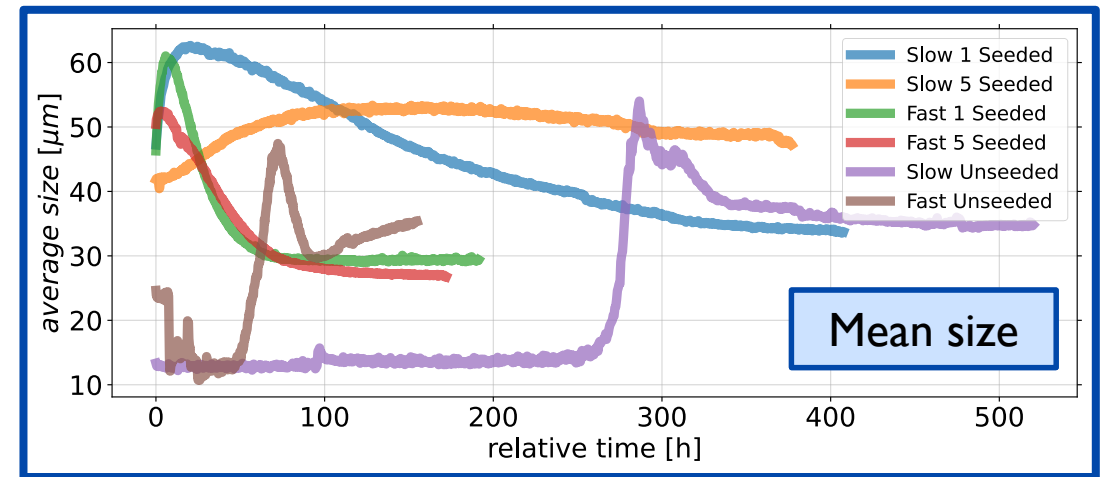
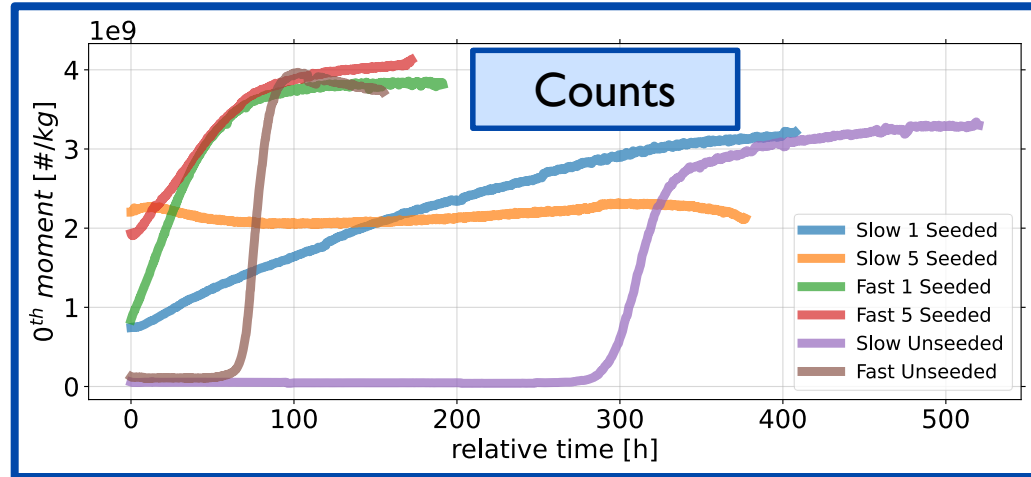
Available information on particle size

Comparison between seed and final number distributions



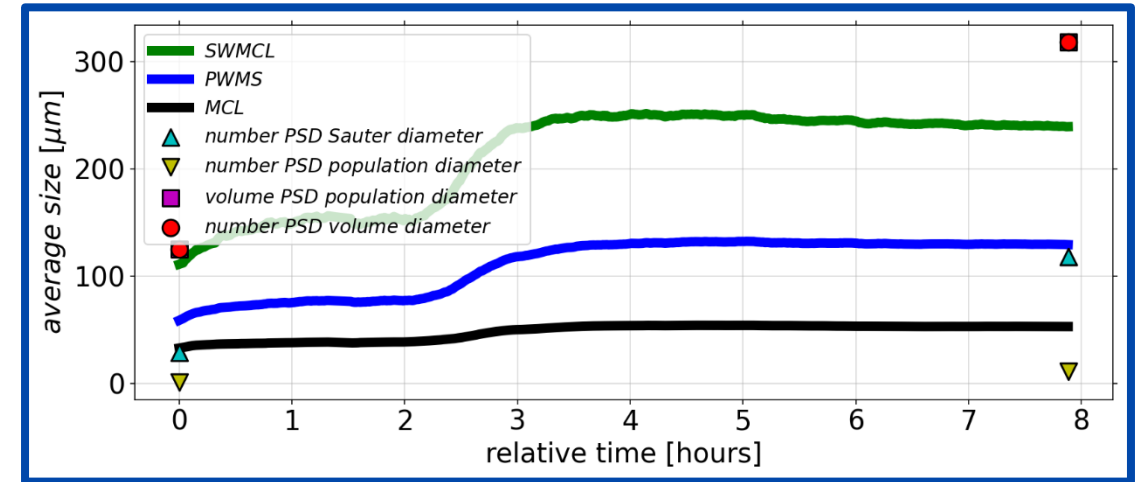
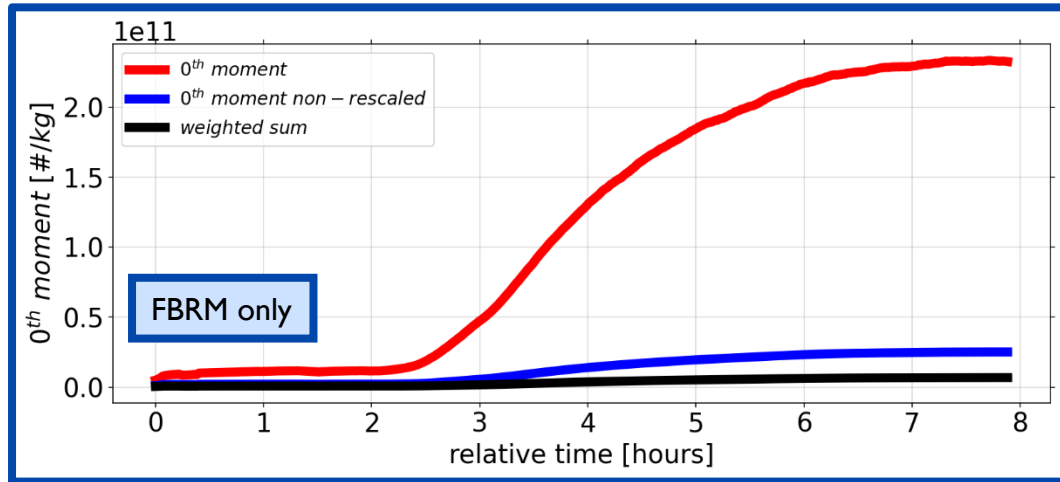
- ▶ Shift towards **right** due to particle **growth** (coherent with images)
- ▶ Particle number **decreases** according to distributions, but **increases** looking at FBRM data

Comparison of CLD data for all experiments



- ▶ Profiles of the total counts and mean size computed from **CLD data** were compared for all the experiments
- ▶ The trends were similar, except for the case of **slow** cooling with **large** seed
 - ▶ The differences in the other cases were mainly in the **intensity** and in the **timing** of the phenomena, in any case aligning with expectations

Different data interpretations and treatment



- ▶ $PWMS = \frac{\mu_1}{\mu_0}$
- ▶ $LWMS = \frac{\mu_2}{\mu_1}$
- ▶ $SWMCL = \frac{\sum_i n_i M_i^3}{\sum_i n_i M_i^2}$
- ▶ $MCL = \frac{\sum_i n_i M_i}{\sum_i n_i}$

- ▶ **Geometric mean** was used as mean size for each size class to compute moments of the initial and final PSDs
- ▶ Different combinations were tested to perform estimations
 - ▶ **MCL** (mean chord length) + number PSD **population** diameter
 - ▶ **PWMS** (population weighted mean size) + PSD **Sauter** diameter
 - ▶ **SWMCL** (square weighted MCL) + PSD **volume** diameter