1. Getting started

CrySiV, an acronym for Crystallization Simulation and Visualization tool, is a population balance-based crystallization simulation software for cooling, antisolvent, and combined cooling and antisolvent crystallization processes in batch, semi-batch and continuous (MSMPR) platforms involving the primary & secondary nucleation, growth, dissolution, agglomeration and breakage of particles. CrySiV has (limited) 2D crystallization simulation support and optional GPU acceleration.

The software was compiled for 64-bit Windows operating systems and has beta support on Linux systems. It has the minimal and recommended system requirements listed in Table 1.1. In terms of computational load, CrySiV sets adaptively the number of CPU cores at startup if the user confirms the multicore calculation option as a function of available CPU cores and the amount of free physical memory. No more than *N-1* cores are used, where *N* is the number of available CPU cores.

Table 1.1. System requirements for the CrySiV.

	Minimal	Recommended
CPU	2 core CPU, 1.8 GHz clock speed	6 core CPU, 4 GHz clock speed
Memory	2 GB/CPU core, DDR3	4 GB/CPU core, DDR4
Storage	3 GB, HDD	5 GB, SSD
GPU (for GPU acceleration)	nVidia CUDA enabled GPU with compute capability 2.1 or later	nVidia GTX 1060 or higher, nVidia Tesla K20 or higher

Some functionalities of CrySiV require Microsoft Excel for data export and Adobe Reader for viewing the documentation.

CrySiV has five core modules: the experimental data input, the parameter estimation, the simulations, the visualizations, and the optimizations. These modules are described in a nutshell in this documentation.

2. The experimental data input module (EDIM)

The experimental data input module allows to feed the experimental data to the software that will be used for crystallization kinetic parameter estimation. In CrySiV, the experiment dataset is defined in four consecutive steps:

a. Regression of the solubility model,
b. Selection of the PBM dimensionality (1D or 2D),
c. Definition of basic thermodynamic properties,
d. Definition of individual experiments.

These operations will be described in detail below.

2.1. The solubility regression tool

The first step in defining an experiment set is the regression of the solubility model. The rest of the features of the EDIM are locked until a valid solubility model is defined. CrySiV knows whether a pure cooling, pure antisolvent, or combined cooling and antisolvent crystallization is being defined from the solubility model.

The solubility regression tool is opened by clicking the "*Solubility regression*" button presented in Figure 1.1. The solubility regression tool can use up to five solubility datasets. For example, solubility data files can be found in the "Solubility data example" folder in the supporting materials downloaded from the website and in each step-by-step tutorial folder. The units of measure and the relative importance (weight factor) are set individually for each dataset.

! CrySiV transforms all concentration data to the CrySiV concentration units (mass fraction, by default, can be changed under the Preferences/Numerics). Consequently, the figures' values might differ from those in the template files.

A generic solubility model equation is used, where the solubility is related to the temperature and antisolvent mass fraction. In the case of pure cooling crystallization, CrySiV automatically turns off the estimation of terms related to the antisolvent concentration. Conversely, for pure antisolvent crystallization, the temperature-dependent parts are turned off. Terms from the equation can be removed individually to improve the fit quality and/or to decrease the uncertainty, quantified by the confidence intervals of the estimated parameters. After the user obtains a good-fitting solubility, the solubility equation can be moved back to the main CrySiV window by pressing the "*Save and exit*" button ().

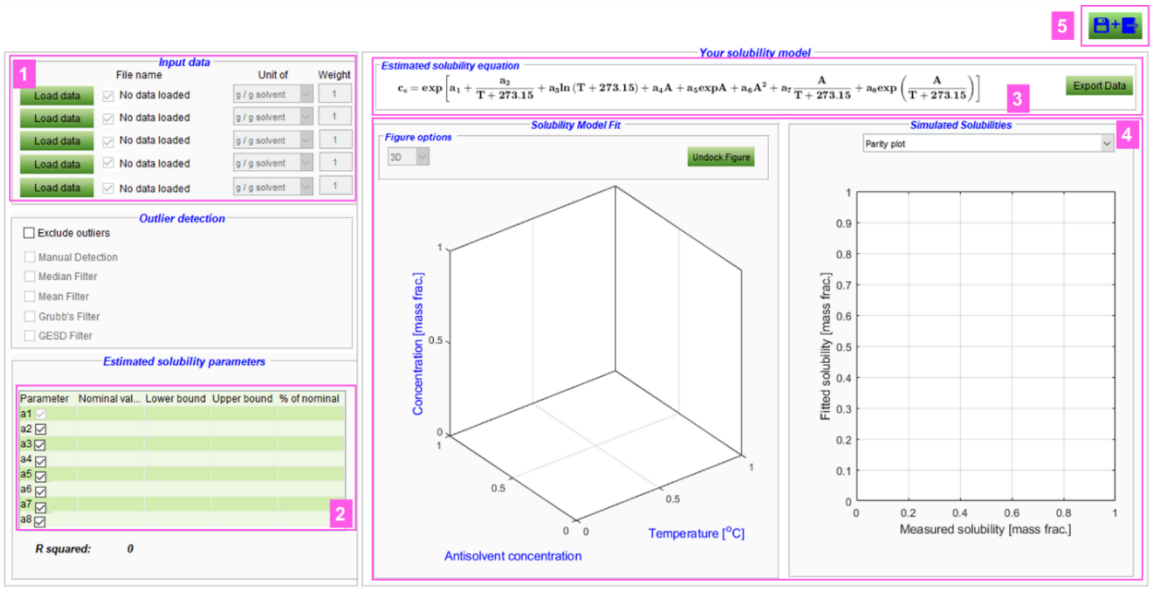

Figure 2.1. The blank solubility regression window has the main parts: 1. Up to five datasets can be loaded; the units of measures and relative importance can be set individually; 2. The terms of the pre-defined equations can be included and excluded by checking/unchecking the boxes; 3. The pre-defined equation; 4. Graphical representations: the phase diagram and the parity plot; 5. Save and exit button: this will move the solubility data to the main window of CrySiV.

After pressing the "*Save and exit*" button (), the definition of PBM dimension becomes available. When the user selects the dimension, the basic crystalline data definition edit boxes and popup menus also become enabled.

2.2. Definition of basic thermodynamic properties

The crystal shape, solvent, antisolvent, and crystal densities must be defined next. The crystal shape is defined through the "*volume shape factor*", but there is a list containing the shape factors for typical crystal shapes.

2.3. Defining experimental conditions and outcomes

Each experiment must be defined separately, one after the other. There is no upper limit to the number of experiments that can be used in the CrySiV; a new experiment can be initiated simply by clicking the  button. Each experiment is defined in two consecutive steps:

a. Input the raw experimental data.
b. Manipulate/adjust the experimental data if necessary.

2.3.1. Loading the raw experimental data

The raw data is defined using a dedicated tool, which can be accessed by clicking the "*Data input*" push button. CrySiV divides the experimental data into two groups: inputs and outputs. The inputs are needed to simulate the crystallization process. Hence, all the inputs must be defined. Outputs are used to be compared with the simulated values to regress the kinetic parameters. At least one input must be defined for each experiment.

! It might be a good practice to collect all your data in an Excel file. Then, you can copy the data from Excel and hit "Paste data", which will automatically fill your table. The tables can also be filled out manually cell-by-cell.

Once the experiment is defined, the "*Save and exit*" button () moves the data to the main CrySiV window.

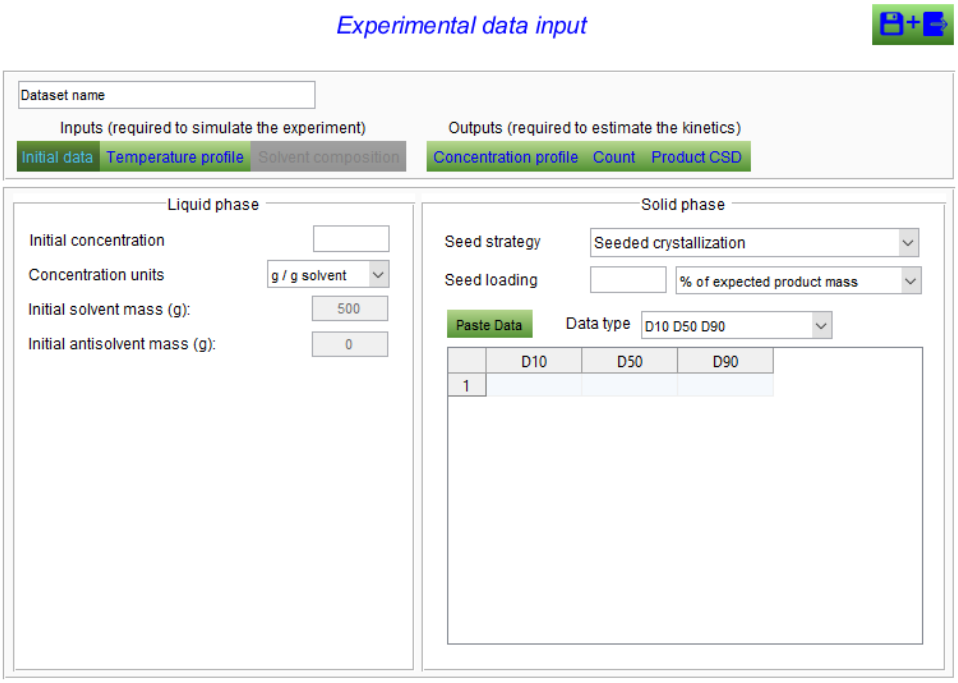

Figure 2.2. UI of the data input tool. The blue text color in the push buttons means that the part is not fully configured. The solvent composition tab remains disabled if a pure cooling crystallization is being defined.

The experimental data often contain artifacts, which should be eliminated before attempting to estimate the kinetic parameters, as these might lead to the estimation of fake model parameters with poor prediction capability. A typical, frequent example of such an artifact is the shift in the concentration measurement, in which case the measured concentration value falls far away from the solubility, even when the equilibrium is reached.

2.3.2. Experimental data manipulation and adjustments

The experimental data manipulation tool of CrySiV was designed to allow for simple adjustments of the experimental data. The tool can be accessed by clicking the "*Data manipulation*" button.

As the experimental outcomes depend on the inputs, the inputs cannot be manipulated, except for the seed size distribution, where a cutoff size can be defined under which all the particles are ignored. This allows us to eliminate the artificial peak in the CSD in the small size domain, which is known to be an artifact of the laser diffraction-based particle size distribution measurement.

When editing the outputs, in addition to adding offset to the measured values, several other changes can also be made. The maximum number of points from the given data can be limited by the "*Number of points*" option. CrySiV will interpolate the defined number of points from the original

data. Reducing the number of points might not hamper the parameter estimation performance because, in numerous cases, the sensor's sampling rate is significantly higher than is necessary to capture the process dynamics. However, the CrySiV solver efficiency is inversely proportional to the number of intermediate sampling instances. Hence, the parameter estimation might be significantly faster with less intermediate measurement data.

! For efficiency reasons, even if the user does not access the Data manipulation tool, CrySiV reduces the number of sampling points to be used automatically.

Practical hint: sometimes, the user might want to use only a limited time range of data. For example, only modest kinetic information can be obtained from the concentration that overlaps with the solubility. Hence, from a kinetic perspective, the most valuable part of the concentration profile is right after seeding when the concentration is relatively far from the solubility. In this case, the experiment is simulated up to a point defined by the "*Include only (time range)*", accelerating the parameter estimation further. If product CSD is also defined, the whole experiment must be simulated anyway. If there is a sensor issue, for example, caused by a stuck particle on the in-line sensor, a sub-part of the data can also be excluded.

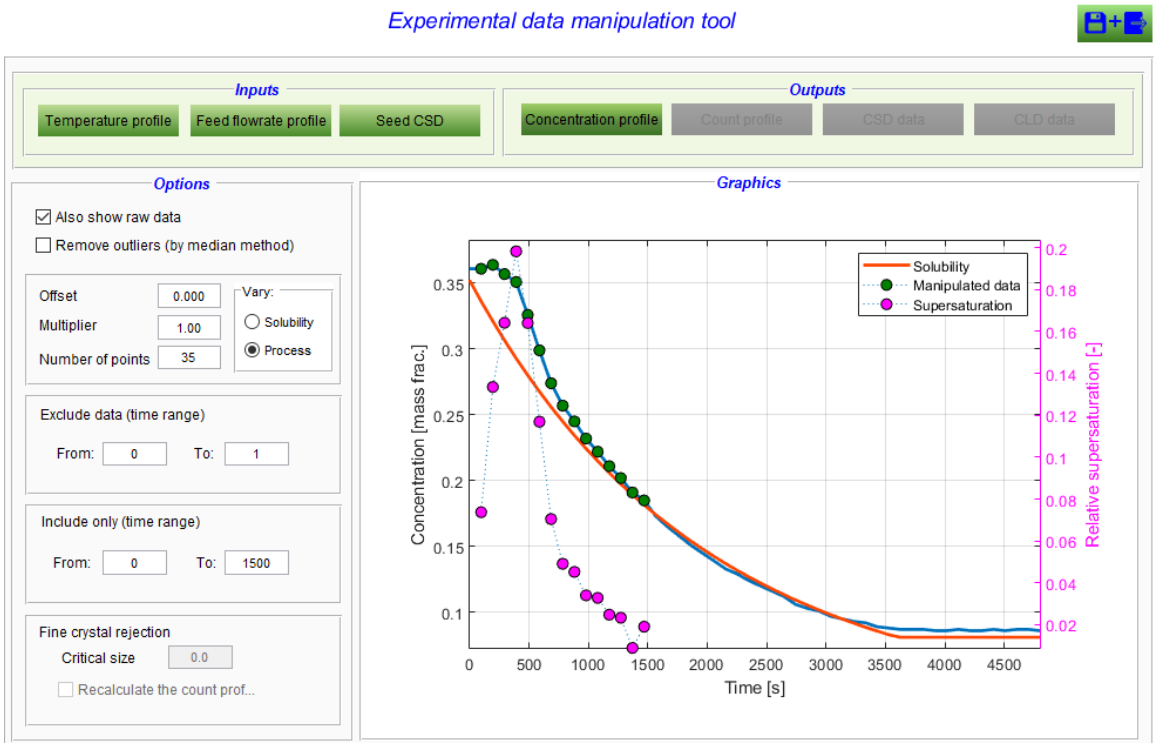

Figure 2.3. UI of the data manipulation tool. The pushbuttons of the outputs not defined for the given experiment are disabled.

3. The parameter estimation module (PE)

The PE module was designed to help identify the best kinetic equations and to regress its unknown parameters based on the experimental data uploaded through the EDIM. PE uses automatic scaling of the decision variables based on sensitivity analyses (described in detail in the paper attached to this documentation), has powerful global optimization algorithms, and helps the user find a good starting point interactively. Displaying the actual fit during the parameter estimation helps the user track the progress better in real-time.

Panel #1 in Figure 3.1 contains the main control and most important information about the PE. The optimization algorithm can also be selected here (the default is the CMA-ES, which was successfully applied in numerous crystallization parameter estimations). Parameter estimation configurations can be saved, and pre-saved configurations can be loaded. Once the parameter estimation is configured, i.e., the initial point and the fitting parameters are defined, the parameter regression process is started by pressing the "*Run*" button ( ).

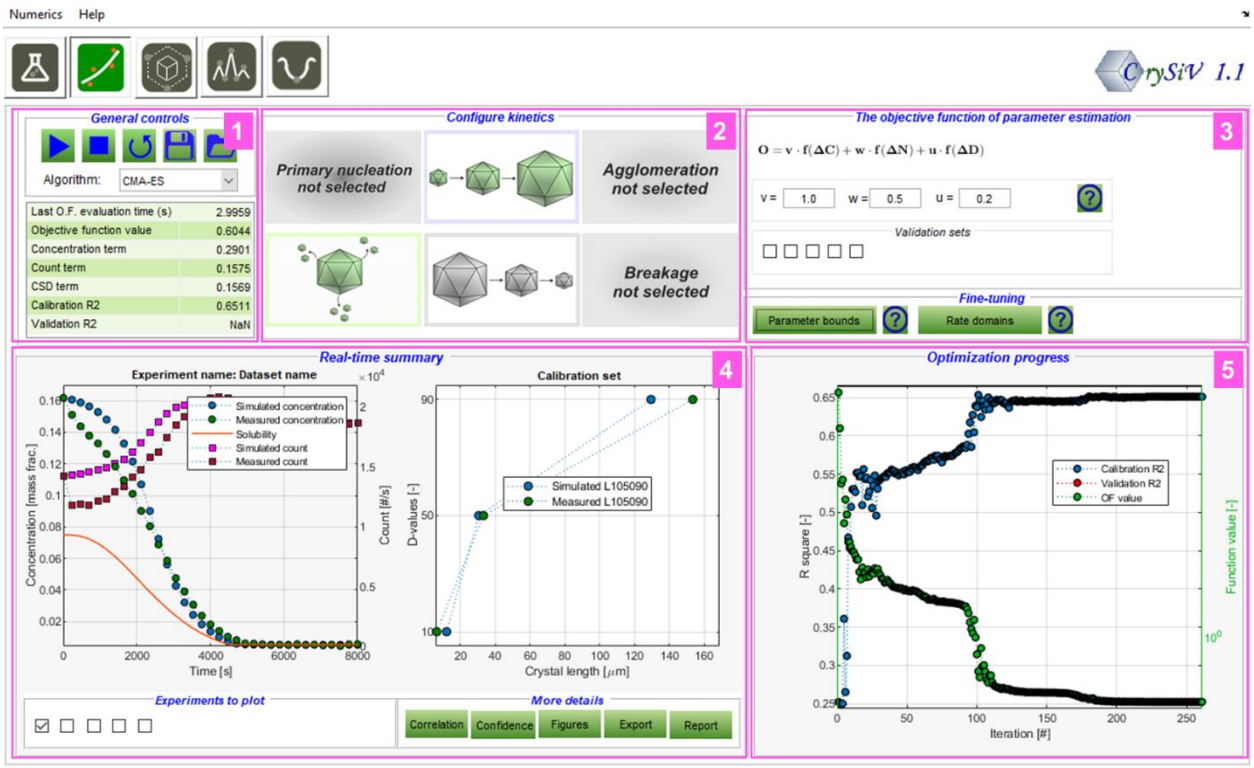

Figure 3.1. UI of the parameter estimation module with the main parts. 1. Main controls for starting, stopping, resetting, saving, and loading parameter estimations; 2. Pushbuttons for the configuration of kinetics and initial parameters; 3. Weight factors for the concentration, relative number density and particle size distribution terms, and selection of the validation experiments; 4. Comparison of simulated and measured quantities, updated iteration-wise, and pushbuttons for statistical interpretations and data export options. 5. Optimization progress: evolution of the equivalent R^2 and objective function value with the iterations.

3.1. The initial point

CrySiV has a built-in heuristic algorithm, which analyzes the available experimental data and proposes a kinetic model and initial parameter values. This initial model structure and kinetic parameters are often a reasonable starting point but might be far from optimal. CrySiV was designed to carry out parameter estimation for different model structures rapidly, giving the user good decision support for model selection.

The parameter estimation can substantially be improved if the starting point is good. Also, minimizing the number of fitting parameters is a good practice. There are no hard-limits, but it is considered that simultaneous regression of more than ~7-8 parameters should be performed only in specific cases with large and diverse experimental datasets. For example, 3-6 parameters, e.g., nucleation and growth rate constants, the corresponding supersaturation exponents, and a size-dependent growth parameter or agglomeration or breakage rate constant, can deliver a decent fit on the data fit. Figure 3.2 presents the nucleation configuration panel.

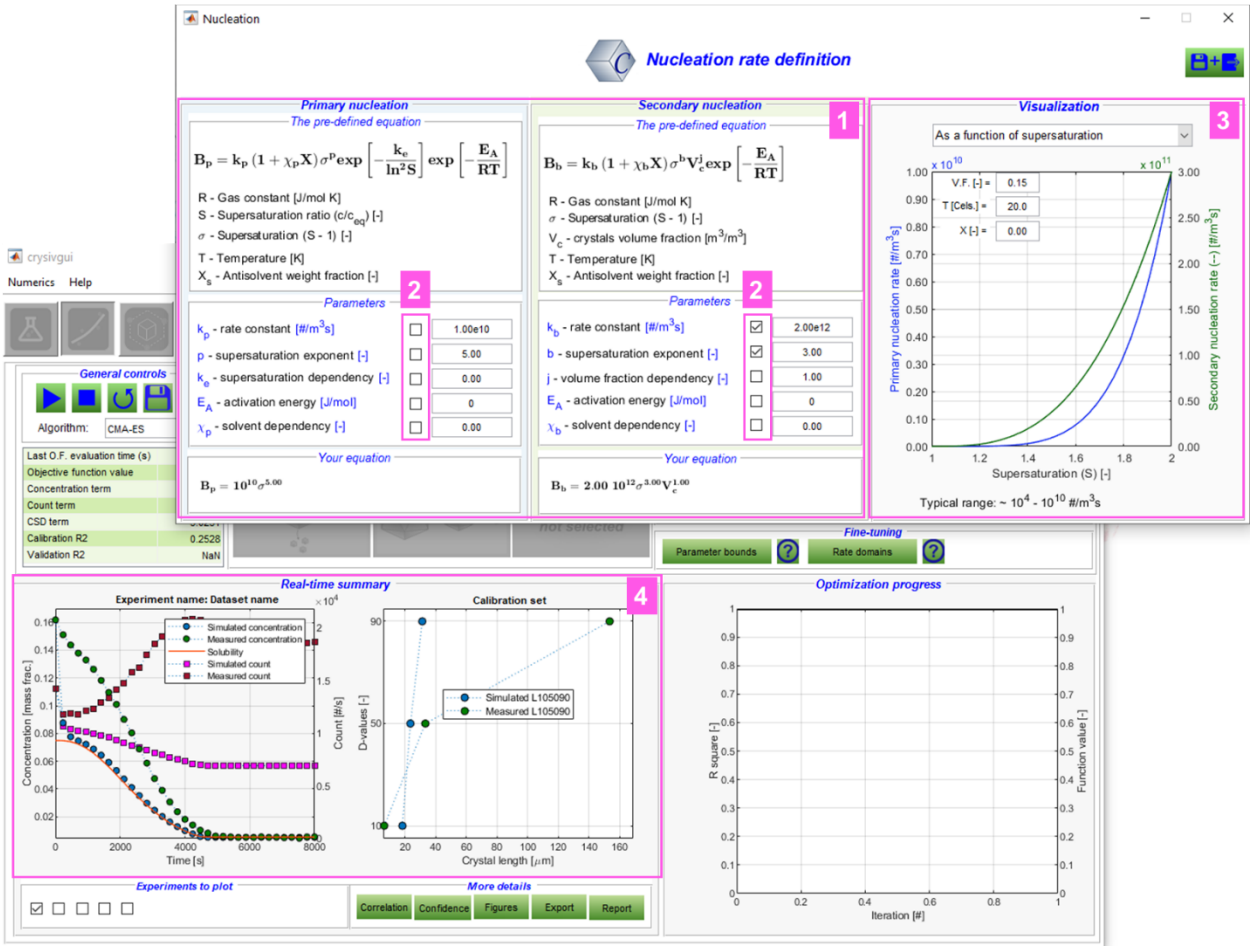

Figure 3.2. UI of the nucleation configuration panel started from the PE. 1. Configuration panel for the pre-defined primary and secondary nucleation rates; 2. Checkboxes for the selection of parameters that will be estimated – the values of unchecked parameters are fixed during the

parameter estimation; 3. Visualization of the nucleation rates as a function of supersaturation, temperature, *etc.* 4. The simulated curves in the main window are updated whenever a parameter is changed in the kinetics panel, which aims to help the user find a good starting point quickly.

3.2. Fine-tuning of the parameter estimation optimization

Box #3 of Figure 3.1 is for the controls of the optimization objective function. The parameter estimation has three possible sub-objectives: matching the measured concentrations, the particle size distributions, and the number densities. In an ideal case, in the optimum solution, all sub-objective values drop to zero; hence, the three terms' weight factors are irrelevant. However, this is very unlikely in practice, and the weight factor values can be used to adjust the relative importance of the three terms. CrySiV has an algorithm that automatically sets the weight factors for the data when entering the PE module to ensure that the objective function is sensitive to all sub-objectives. Therefore, different weight factors will be used by default for different experimental datasets. However, the user may vary the weight factor for better performance, e.g., to improve the concentration fit over the CSD fit.

During the parameter estimation, the optimizer varies the kinetic parameters in a broad domain. There is a typical domain within which the parameters can be for some parameters, such as supersaturation exponents or activation energies. The default bounds for each parameter can be changed by pushing the "*Parameter bounds*" button.

Even though the kinetic parameters are individually in the physically relevant domain, their combination, under certain operating conditions, might lead to unfeasible kinetic rates. Let us consider the following example, presented through a standard growth rate equation:

$$G = k_g \sigma^g \exp\left(-\frac{E_A}{RT}\right)$$

Let us do the calculations for $\sigma = 0.1$ relative supersaturation and $T = 313$ K temperature and consider the two cases involving kinetic parameters that, if looking at them separately, are within the realistic parameter range:

$$G = 10^3 0.1^{1.5} \exp\left(-\frac{45000}{8.31 \cdot 315}\right) = 9.69 \cdot 10^{-7} \text{ m/s}$$

$$G = 10^3 0.1^{1} \exp\left(-\frac{15000}{8.31 \cdot 315}\right) = 3.13 \cdot 10^{-1} \text{ m/s}$$

The first growth rate is realistic, but it is hard to imagine that a crystal, under normal operating temperature and moderate supersaturation, would grow 0.313 m/s. The problem of this unrealizable kinetics points beyond the fact that it cannot be the solution and adds unnecessary objective function calls. Due to the full discrete implementation of the high-resolution finite volume method employed in the CrySiV solver, the time step size is inversely proportional to the growth rate. Hence, the process simulation using the second growth rate takes significantly longer

than the simulation using the first set of growth parameters. Practically, the user will feel that the CrySiV is stuck in an iteration. To prevent these situations, CrySiV applies nonlinear optimization constraints to delimit the kinetic rates. The "*Rate domains*" button can change the default limits.

Last but not least, a checkbox group permits the selection of the validation experiments. The experiments assigned to validation are not used to calculate the objective function. Hence, as a side effect of assigning experiments for validation, the parameter estimation runs faster. The validation experiments are simulated once per optimization iteration to give a way to track the validation performance during the parameter estimation.

The simplest way to evaluate the parameters' accuracy is by the estimated parameters' confidence interval widths. One wants to see relatively narrow confidence intervals. Since the correlation between the parameters broadens the confidence intervals, it is hard to quantify which confidence interval is too broad or narrow enough. Confidence intervals can be used for two important decisions:

1. Suppose two model structure candidates give comparable nominal fit and involve the same number of fitting parameters. In that case, the best is to go with the model structure whose parameters have the narrowest confidence intervals.
2. If the confidence interval of a parameter varies in a vast range, typically between 0 and ∞, the given parameter might not show sensitivity and might be eliminated. After re-running the parameter estimation, with the given parameter fixed or the associated mechanism neglected (e.g., nucleation or agglomeration), the fit of this newly derived model can be re-evaluated similarly based on the confidence interval widths.

The parameter estimation data can be exported by clicking the "Export" button, and a written report can be generated by clicking the "Report" button.

4. The simulations module (SIM)

The simulations module simulates various crystallization processes in batch, semi-batch, and continuous (MSMPR) systems. The solubility, crystallization kinetics, and parameters can be taken from the literature or regressed using the CrySiV's PE tool. Suppose the SIM is opened from the PE. In that case, all the kinetic and thermodynamic properties defined in the EDIM and calculated in the PE are automatically transferred to the SIM, and a typical batch crystallization is initiated.

Figure 4.1 presents the main window of the SIM module. There are four parts: the main controls, the model-equation panel, the configuration panel, and the graphics. The equation panel only lists the actual kinetic and thermodynamic equations that are used; the model configuration cannot be changed there.

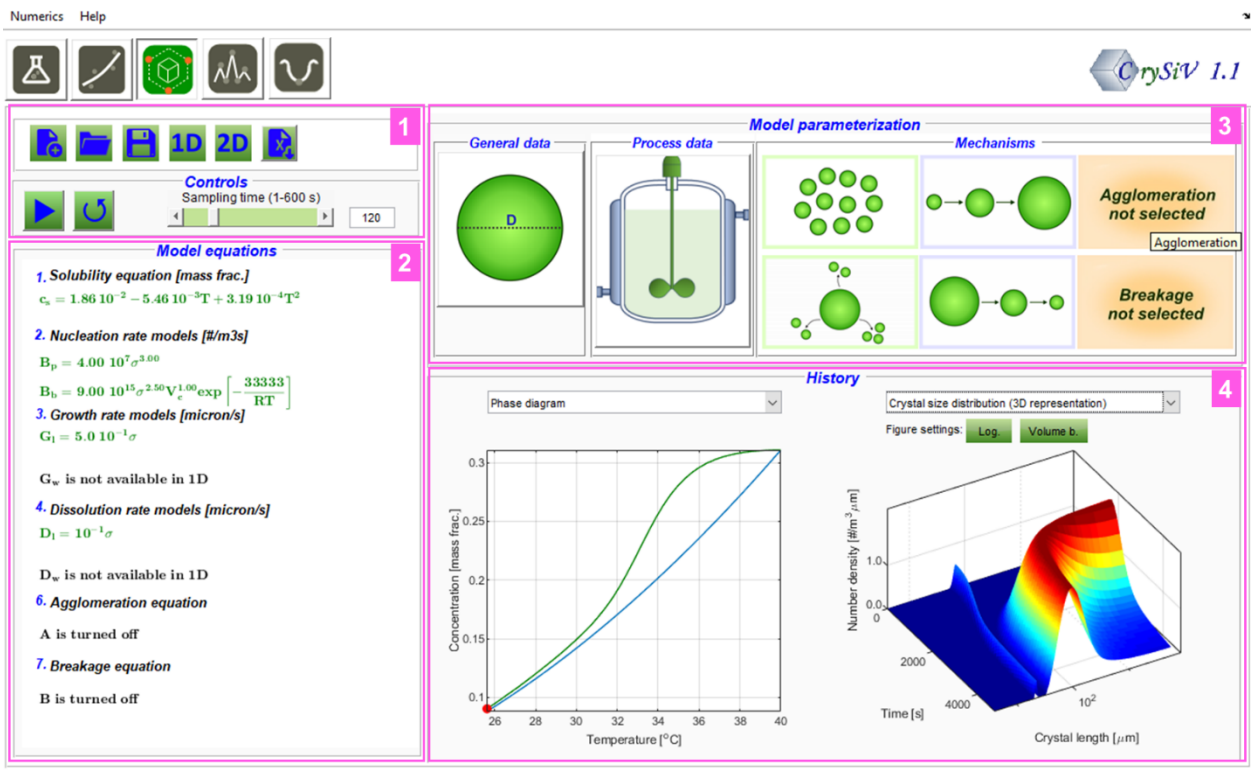

Figure 4.1. UI of the simulations module with the main parts. 1. The main control button group to run simulations, as well as to save, load, or start new simulations and to export data to an Excel file; 2. The model-equations for solubility and crystallization kinetics; 3. Pushbuttons for popup windows to configure the crystallization system-related data, the process-related data, and the crystallization kinetics; 3. Graphical representations were updated during the simulation.

The model configuration is carried out by pressing the button in the general data panel. This opens a panel, shown in Figure 4.2. The problem's dimensionality must be selected first (1D or 2D). Crystal shape and density also have to be defined. The user can select four pre-defined solubility models for solubility: the power law and the Apelblat for cooling crystallization, a generic antisolvent, and a combined cooling antisolvent solubility model. The users can also regress a solubility model based on measured solubility data by selecting the "*Estimate from solubility data*" option. The graphics panel gives an insight into the solubility behavior using the actual solubility parameters. If everything is configured, the "*Save and close*" button can be pressed to move the data to the CrySiV's main window.

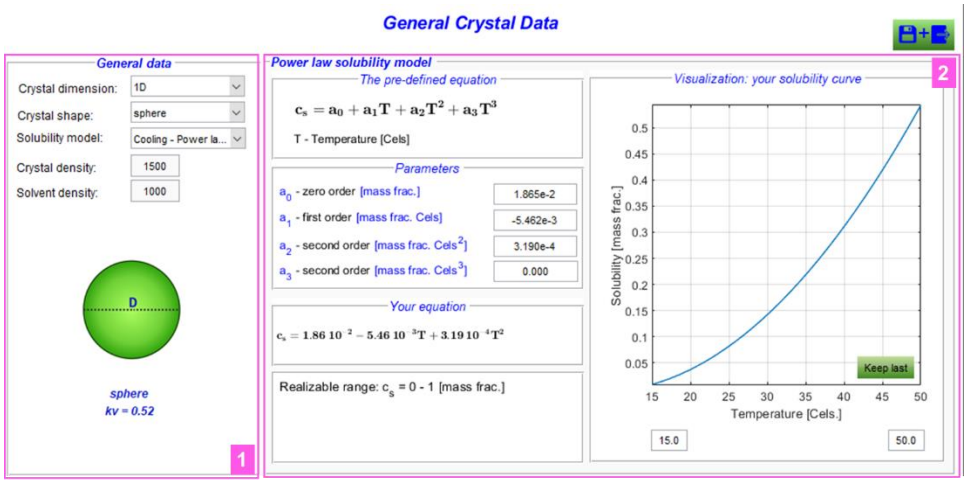

Figure 4.2. UI of the simulations module with the main parts. 1. Definition of crystal dimension (1D or 2D), the crystal shape, solubility model equation, and the densities; 2. Parameterization of the selected solubility model-equation with a quick visualization panel of the solubility as a function of temperature or/and antisolvent composition.

After configuring the general data part, the process data has to be defined by pressing the push button in the "*Process data*" panel. Configuring the process data is much more complex than the general data as all system types (e.g., batch and continuous, cooling or antisolvent) have different inputs. For example, in the case of Figure 4.3, a batch cooling crystallization process is configured. Therefore, the "*Feeding profile*" option in panel #1 is disabled, and the solvent system mass and antisolvent mass fractions in panel #2 are also disabled. Once the system type is selected, all two (or three, for fed systems) input types must be configured, which are briefly described below.

- Initial conditions (seed & composition). The initial solute composition is specified in panel #2 in Figure 4.3. The solvent composition edit boxes are only active for antisolvent-enabled crystallization since these have no impact on pure cooling crystallization. The seed distribution and seed load, in mass fraction, must be specified as well. There are four options to define the seed size distribution: 1. Unseeded crystallization; 2. Lognormal; 3. Custom distribution, and 4. D10, D50, and D90 data (percentiles of the volume-based cumulative distribution). For lognormal distribution, the mean and standard deviations must be defined. There is an example file for custom distribution definition in the documentation support. If the percentiles of the seed distribution are specified, CrySiV fits a lognormal distribution to the mean values, which will be used as seed distribution.
- Temperature profile. Temperature profile can be defined by three methods: 1. Parameterizing a pre-defined equation; 2. Custom profile (an example file is provided in the documentation support); 3. Quick definition: linear batch that only requires the initial and final temperatures and the batch and equilibration times.

- Feeding profile. Two input streams must be configured: the antisolvent & solvent and the solvent & API streams. Three methods can define the flow rates: 1. Constant flowrates; 2. Custom profile (example files are provided in the documentation support). 3. A quick definition, which works in the same way as in the case of temperature profile.

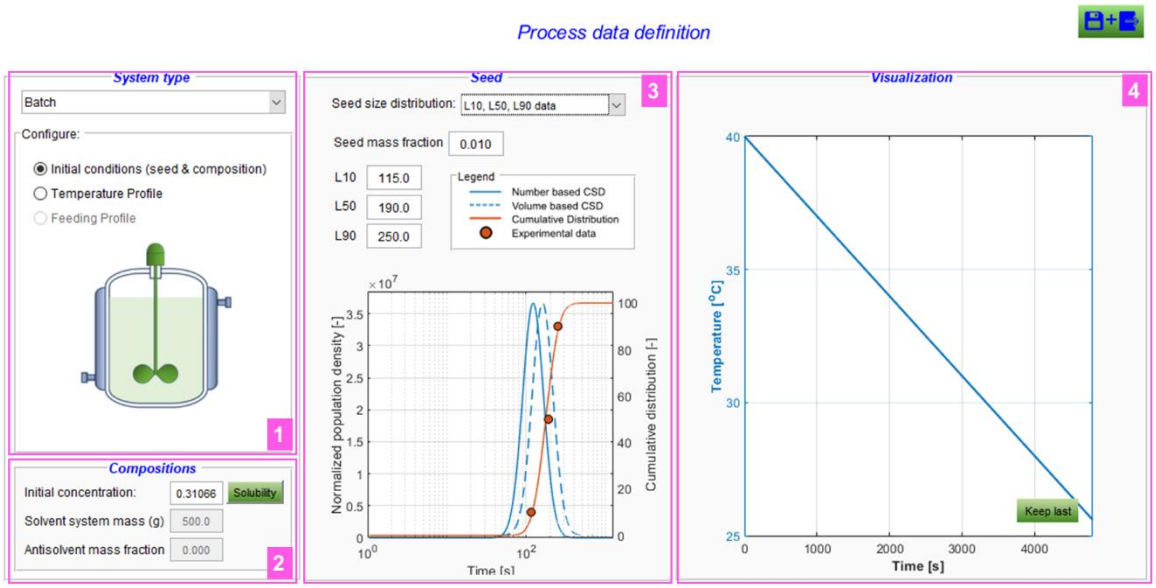

Figure 4.3. UI of the simulations module with the main parts. 1. Master panel with high-level settings, where the system type (batch, semi-batch, continuous) and the selector of input configuration (initial conditions and seed, temperature profile, and feeding profile for semi-batch and continuous systems, which all must be configured) are executed; 2. Initial concentrations; 3. Configuration of the input selected in the panel #1; 4. Graphical representation of the operating profiles.

All simulations in CrySiV are started from the main window. There is only one exception: the MSMPR cascade simulator. The MSMPR cascade simulator tool is started from the "*Process data*" tool by selecting "MSMPR cascade" as the process type. This will initiate the popup window shown in Figure 4.4.

! In the background of the MSMPR cascade, dynamic simulations are carried out, which are run until the steady state is reached, as determined by the Phillips-Perron test.

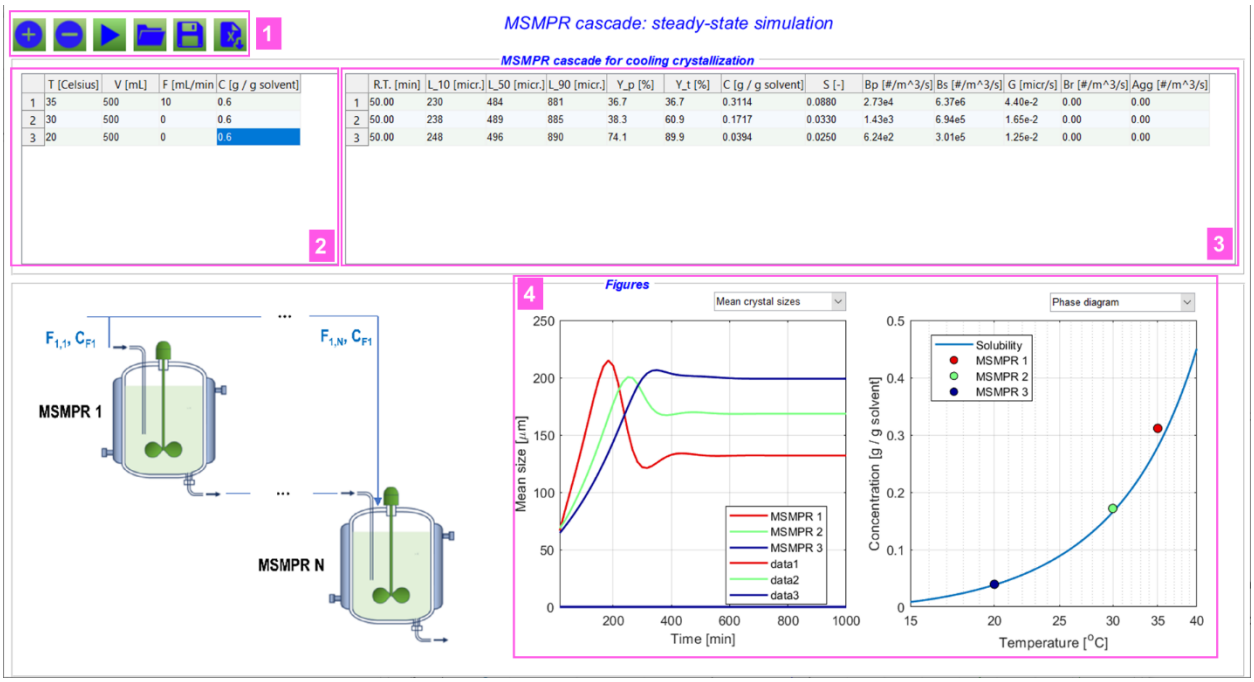

Figure 4.4. UI of the steady-state MSMPR cascade simulation panel with the main parts. 1. Main configuration buttons to add or remove crystallizers, run the simulation, save or load simulation configuration, and export the data to an Excel file; 2. Input configuration table: the operating conditions for each MSMPR must be configured here; 3. Output table: automatically filled with the steady-state simulation results; 4. Graphical representations of the operating points in the phase diagram and the dynamics of the mean crystal sizes.

5. The visualizations module (VIS)

VIS permits loading up to two pre-saved simulations and animate the solutions side by side. High-quality figures and animated gif files can also be generated using VIS.

When entering VIS from the simulations module, the simulation results are automatically transferred to the place of the first simulation (green curves in the figures). A second simulation, pre-saved from the SIM, can be loaded, and the two simulations can be animated side by side. As a limitation, only identical process types can be compared (i.e., batch crystallization to batch crystallization, cooling to cooling, and 1D to 1D). Parametric studies can be carried out easily and interactively to investigate the effects of various processes and kinetic parameters.

There are three options to extract the data from the VIS. Firstly, there is an export option in the control group panel, highlighted with box no #1 in Figure 5.1. Furthermore, there are two pushbuttons in the graph areas. The right-hand side pushbutton undocks the current figure, which becomes editable as any Matlab figure object. The left-hand side pushbutton generates a .gif animation of the given figure from the beginning to the current process time.

Figure 5.1. UI of the visualizations module with the main parts. 1. The main control button group to load or discard a simulation, run a visualization, and export data to an Excel file; 2. High-level comparison of the two simulations; 3. Visual comparison of the process type, particle shape, and mechanisms in the two simulations; 4. Graphical representations are updated during the visualization with options to undock the figure or create an animated .gif file.

6. The optimizations module (OPT)

CrySiV's optimization module was designed to optimize a crystallization process. The thermodynamic and kinetic data is taken from the simulations module (the solubility model, densities, crystal shape factor, and all the kinetic model parameters). The optimizations module permits the change of the kinetic parameters.

Configuring the process optimization is performed in four consecutive steps. First, the seed crystal PSD must be defined by clicking the "*Seeds*" push button. Note that the seed PSD cannot be optimized as practically it would be difficult to obtain the optimized seed PSD. Once the seed PSD is defined, the decision variables, constraints, and objectives have to be specified. The definition of objectives for batch cooling crystallization is presented in Figure 6.1. For other types of crystallizations (e.g., antisolvent or combined cooling and antisolvent) or systems (e.g., milling enabled), the configuration panels will differ from that presented in Figure 6.1. Eight potential objectives can be involved, which are logically divided into two groups (product and process-

related objectives). This division does not influence the objective function; it is only to help the user oversee the selections. Only the objectives with enabled checkboxes will be considered in the optimization. Crysiv solves this multiobjective optimization by minimizing the weighted sum of sub-objectives. The user must specify the weight factor for each term. The value of the weight factor substantially impacts the optimal solution. To help the user obtain a more reasonable weighting, the "*Optimization objectives*" panel provides the contribution of each sub-objective to the aggregated objective function value in the starting point. With this, the user can rationally set the weighting factors as, following the example of Figure 6.1, there are 13 orders of magnitude difference between the weight factors of the "*Nucleation rate*" and "*Smooth temperature profile*" objectives. Yet, they yield the same objective function contribution.

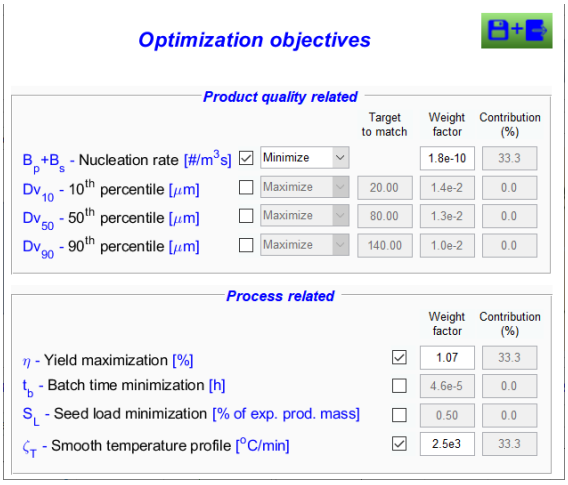

Figure 6.1. The panel for defining the optimization objectives for a batch cooling crystallization. The user can adjust the relative importance of the selected optimization objectives through a weight factor.

The decision variables and constraints are configured similarly to the objectives. Figure 6.2 shows the optimizations panel after a converged optimization run. It can be seen from the graphical representation that the temperature profile is constructed from three linear segments, and under optimal operation, it has a parabolic shape that is known to be the best for nucleation rejection and, subsequently, crystal size maximization in a seeded cooling crystallization. The OPT also displays the product PSD in the actual iteration. The optimization progress graph allows the user to track the evolution of the objective function value. There is an option to switch to a table view as well.

Finally, the optimization results can be exported to an Excel file containing the optimal values of decision variables and the results of a process simulation using the optimized decision variables. These results can be implemented in the lab, where the model-based process design loop closes. Suppose there are deviations between the measurements and simulations. In that case, these can be considered, e.g., in the optimization constraints as an output error correction, and a new optimization can be performed. Suppose the deviations are too large to be manageable in that

simple way. In that case, this new experiment can be added to the dataset used for parameter estimation, and the parameter identification can be repeated to obtain an improved model.

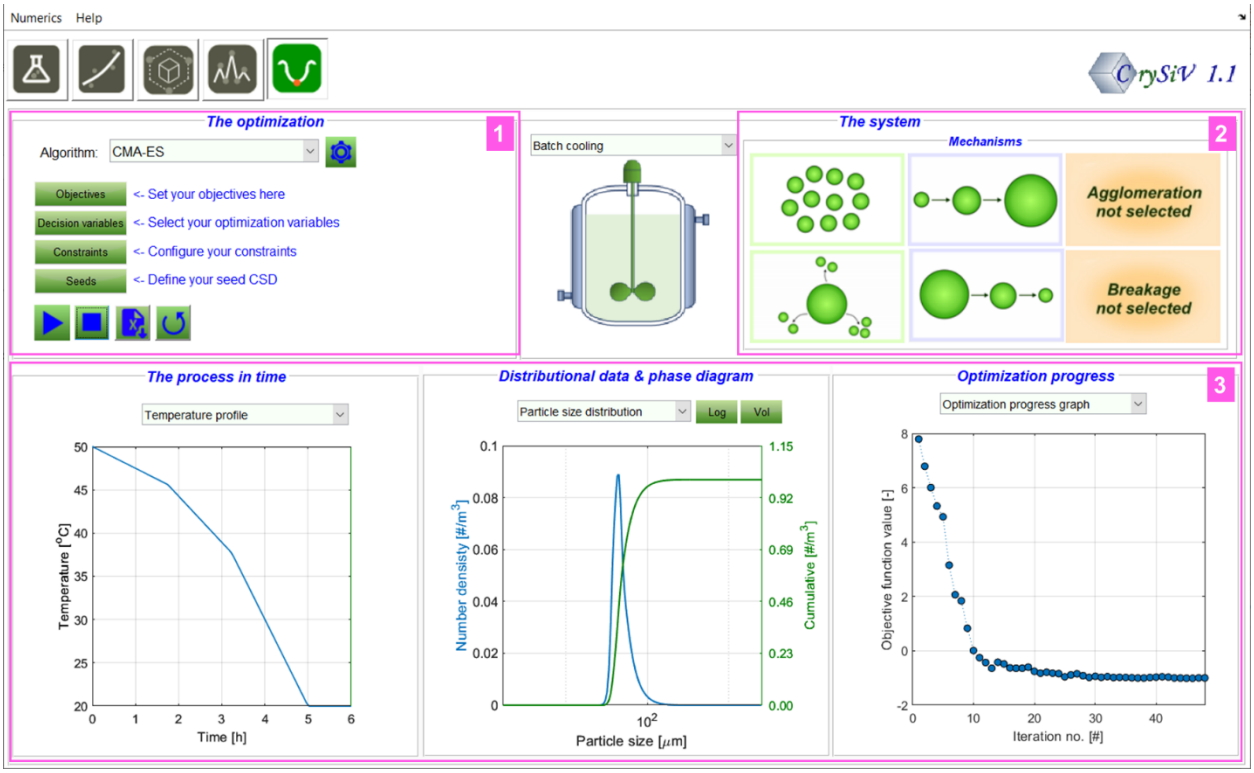

Figure 6.2. UI of the OPT with the main parts. 1. The main control panel of the OPT module to configure an optimization (objectives, decision variables, and constraints), define the seed crystals, run a visualization, and export data to an Excel file; 2. The mechanisms definition panels allow the user to change the kinetics and quickly re-optimize a process with an altered crystallization kinetics; 3. Graphical representations are updated during the optimizations to allow the users to track the optimization convergence from iteration to iteration.