



Cite this: *Digital Discovery*, 2025, 4, 2025

## Automated scale-up crystallisation DataFactory for model-based pharmaceutical process development: a Bayesian case study†

Thomas Pickles, <sup>a</sup> Youcef Leghrif, <sup>a</sup> Matt Weisshaar, <sup>b</sup> Mikhail Goncharuk, <sup>b</sup> Peter Timperman, <sup>b</sup> Timothy Doherty, <sup>b</sup> David D. Ford, <sup>b</sup> Jonathan Moores, <sup>a</sup> Alastair J. Florence<sup>ac</sup> and Cameron J. Brown <sup>ac</sup>

Automated model-based design of experiments (MB-DoE) play an important role in enhancing process development efficiencies by minimising material usage and saving significant human labour time. This study describes the conception, installation and application of an automated platform and a model-based design of experiments approach to both plan and automate the experimental load for scale-up crystallisation process development. The platform hardware in detail is a multi-vessel configuration equipped with peristaltic pump transfer, integrated HPLC, image-based process analytical technology and single board computer control based IoT system. To demonstrate the DataFactory's experimental capabilities a 5-point Latin hypercube design was employed to investigate the effects of cooling rate, seed mass, and seed point supersaturation on nucleation, growth, and yield during the cooling crystallisation of lamivudine in ethanol. This initial screening data served as inputs for Bayesian optimisation to determine the optimal next experiment aimed at achieving the target process parameters and reducing uncertainty. This data-driven MB-DoE approach simplifies application, provides flexibility, and accelerates experimental design, achieving a ~10% improvement in the objective function value within just 1 iteration. This study will inform future research comparing the suitability of data-driven, mechanistic, and hybrid models across various crystallisation modes.

Received 23rd December 2024  
Accepted 1st June 2025

DOI: 10.1039/d4dd00406j  
rsc.li/digitaldiscovery

## 1. Introduction

The pharmaceutical industry is challenged by rising costs, fragile and inflexible global supply chains, and the climate crisis, whilst needing to speed up the delivery of new and increasingly complex drug products. Accelerated drug discovery and diverse drug portfolios, coupled with the demands for faster, adaptive clinical trials as seen during the pandemic, pose the risk of manufacturing process development becoming a bottleneck in new drug delivery.<sup>1</sup> Consequently, chemistry, manufacturing and control (CMC) process development must evolve to be quicker, more efficient and sustainable. Embracing industrial digital technologies to automate experiments and simulate manufacturing processes is therefore a strategic move for the industry towards adoption of Quality by Design methodologies.<sup>2,3</sup> This includes approaches such as, increased use of modelling,<sup>4–6</sup>

model-based design of experiments (MB-DoE),<sup>7</sup> laboratory automation,<sup>8</sup> and the integration of these tools into an autonomous closed-loop laboratory referred to as self-driving labs (SDL)<sup>9</sup> or DataFactories.<sup>3</sup> These approaches are seeing increasing application in chemistry and materials science, where they are used to discover new materials<sup>9</sup> and have considerable potential for application in manufacturing process development.

MB-DoE<sup>10,11</sup> integrates mathematical models to optimise experimental planning, particularly in chemical<sup>12–14</sup> and pharmaceutical processes.<sup>15,16</sup> These models can be mechanistic *i.e.*, grounded in physical laws (white-box), statistical *i.e.*, based on experimental data (black-box) and a combination of both (grey-box). Mechanistic models, such as computational fluid dynamics (CFD)<sup>17–19</sup> and population balance modelling<sup>20</sup> (PBM), provide transparent insights into process dynamics, allowing for detailed understanding and simulation. Statistical models vary in interpretability: linear regression<sup>21</sup> and decision trees<sup>22</sup> offer clear relationships between factors and responses, while complex machine learning algorithms like neural networks,<sup>23</sup> support vector machines<sup>24</sup> and Bayesian optimisation (BO)<sup>25–28</sup> excel in prediction without explicit reasoning. MB-DoE enhances experimental efficiency by using these models to simulate and predict system behaviour under various conditions, reducing resource consumption and refining the understanding of critical process parameters.

<sup>a</sup>Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow G4 0RE, UK. E-mail: cameron.brown.100@strath.ac.uk

<sup>b</sup>Snapdragon Chemistry, A Cambrex Company, 360 2nd Ave., Suite C, Waltham, MA 02451, USA

<sup>c</sup>EPSRC Future Manufacturing Hub in Continuous Manufacturing and Advanced Crystallisation, University of Strathclyde, Glasgow, G1 1RD, UK

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d4dd00406j>



Crystallisation presents distinct challenges that set it apart from other chemical and pharmaceutical processes. It is a multiphase process involving the formation of one or more solid phases from a liquid, creating a solute–solvent suspension known as a slurry. This introduces the challenge of designing equipment capable of managing phases with differing compositions, mixing behaviours, and physical properties.<sup>29</sup> Despite efforts to ensure uniform temperature and mixing, local variations between phases can influence process kinetics, where the balance between thermodynamics and kinetics<sup>30</sup> dictates crystal formation. This makes crystallisation highly sensitive to small variations in process conditions including temperature, supersaturation, impurities, solvent composition, tip speed, heat transfer and energy input per mass.<sup>31,32</sup> This sensitivity can lead to issues in reproducibility and scale-up,<sup>33</sup> where minor deviations may result in significant differences in crystal size, shape and purity. Furthermore, the polymorphic nature of many compounds adds another layer of complexity, as different crystal forms can exhibit varying physical and chemical properties, impacting product stability and efficacy.<sup>34,35</sup> The stochastic and often unpredictable nature of nucleation further complicates process optimisation, necessitating process monitoring and control strategies.<sup>36</sup> In order to model and simulate crystallisation processes usefully there is therefore a need for a deep, quantitative understanding of crystallisation mechanisms and the development of robust, model-driven methodologies to ensure consistent and reliable production of high-quality pharmaceuticals.

In pharmaceutical process development, accurate scale-up data is critical for translating laboratory results to industrial scales whilst maintaining product quality, optimising yield, and ensuring safety.<sup>37</sup> Traditional methods often rely on empirical correlations, which may not capture the full scope of process dynamics. This highlights the need for precise scale-up data (mL to L to 100 s L) that can guide decision-making and minimise the risk of costly failures. Integrating automation into experimental workflows can address these challenges by enabling high-throughput<sup>38</sup> and reproducible experiments.<sup>5,39,40</sup> Automated systems can efficiently explore a broader parameter space, rapidly generating reliable data<sup>30</sup> that supports effective scale-up. Additionally, automation reduces human error and can accelerate process optimisation, making it a vital tool in modern process development.<sup>37</sup> By combining accurate scale-up data with automation, researchers can achieve more efficient, consistent, and scalable outcomes in chemical manufacturing. Therefore, in this work, we present the Scale-Up Crystallisation DataFactory to act as a stepping stone between screening platforms and pilot plants in the development of crystallisation processes.

## 2. Automated scale-up crystallisation DataFactory platform

### 2.1. Description of the scale-up crystallisation workflow

The automated platform's (Fig. 1a) workflow is governed by five key stages. Setting parameters (Fig. 1b), where variables to

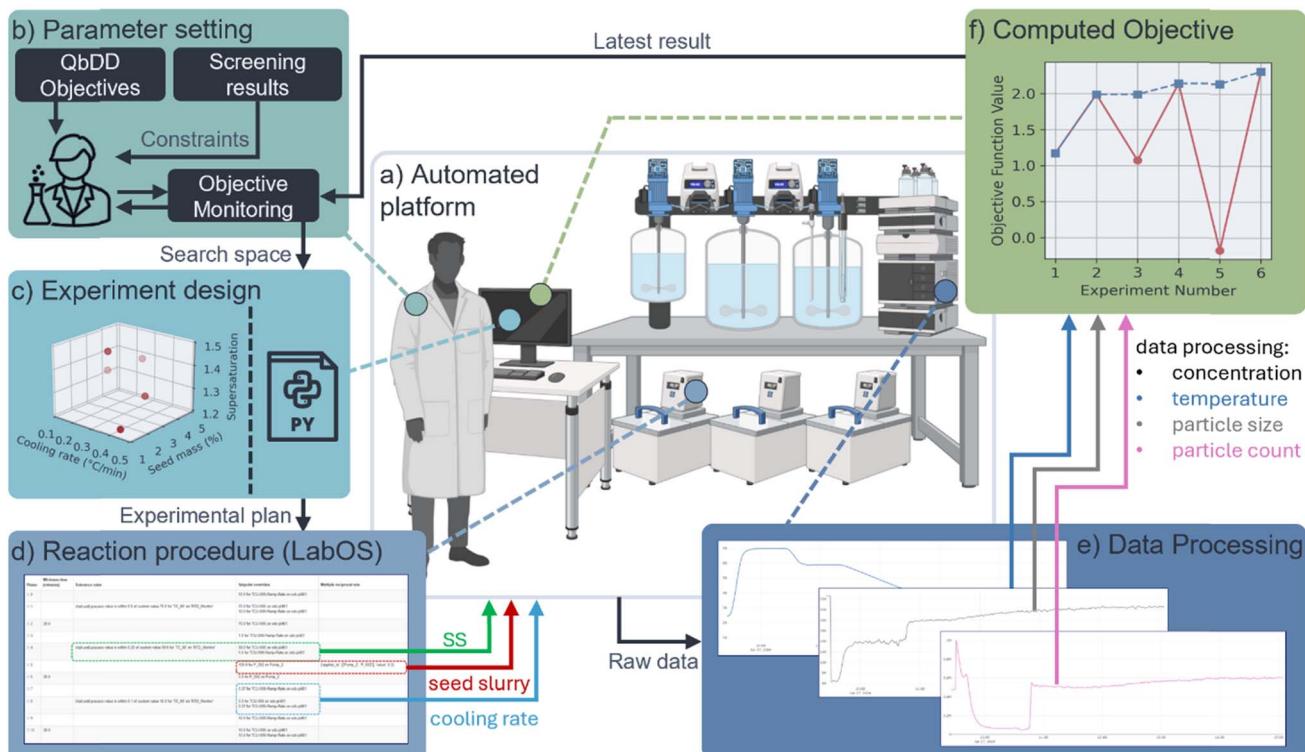


Fig. 1 A schematic showing the flow of logic in the automated crystallisation platform from (a) automated platform hardware, (b) parameter setting, (c) experimental design, (d) reaction procedure generation, (e) data collection and processing to (f) computing the objective function value. Higher resolution sub-figures can be viewed independently in the ESI.†

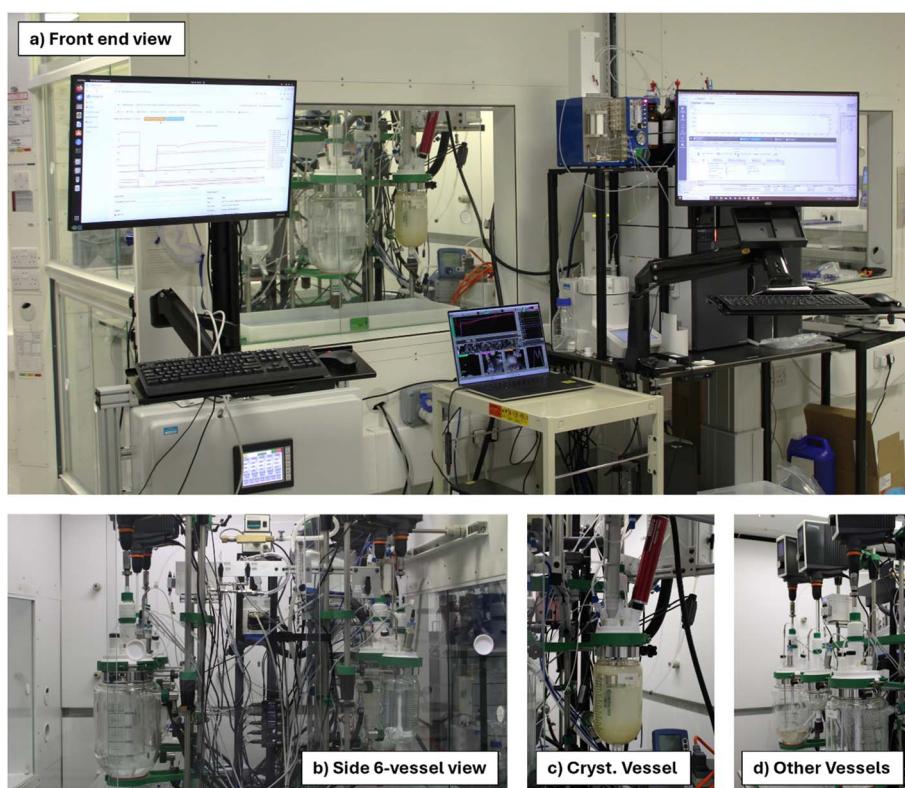


**Table 1** Scale-up crystallisation DataFactory equipment. While each reactor has a primary role defined, the system can be reconfigured by operators to use the vessels in different ways using the LabOST™ software. Each vessel is fitted with a Pt100 RTD and an ultrasonic level sensor

Reactor ID	Size	Primary role	Associated metering system(s)
V-001	5 L	Feed tank with solution of material to be crystallised	Heat-traced pump with ultrasonic flow meter to dose into V-006
V-002	1 L	Feed tank of seed slurry	Peristaltic pump with ultrasonic flow meter to dose into V-006
V-003	5 L	Feed tank of anti-solvent	Diaphragm pump with ultrasonic flow meter to dose into V-006
V-004	300 mL	Feed tank of additive ( <i>e.g.</i> bridging liquid or impurity)	HPLC-style piston pump with ultrasonic flow meter to dose into V-006
V-005	5 L	Feed tank of clean-in-place (CIP) solution	Peristaltic pump with ultrasonic flow meter to dose into V-006
V-006	1 L	Crystalliser	Vacuum/pressure transfer zone system to discharge slurry into collection vessel
S-001	20 mL	Slurry sample grab from V-006	Manual vacuum-based system to draw up a fixed quantity of slurry

explore and measure are selected based on quality by digital design (QbDD)<sup>3</sup> objectives, such as quality attributes (QA), sustainability, manufacturability and model verification, validation and uncertainty quantification (VVUQ).<sup>41</sup> Known data from a material sparing, small scale Crystallisation Screening DataFactory<sup>38,42</sup> is analysed to define constraints, establishing

a feasible design space for scale up study. An experimental design (Fig. 1c) is then constructed using design of experiments (DoE) or machine learning methods to assess the influence of process parameters. This design is translated into a reaction procedure (Fig. 1d), allowing the automated hardware to execute the experiments with a 3-fold round-the-clock time



**Fig. 2** The automated scale-up crystallisation DataFactory platform showing (a) the front end view of the vessels in the fume cupboard with LabOST™ control PC, HPLC, Blaze PC and PLC box, (b) the side view showing all 6 vessels, the manifold and supporting pumps and transfer system, (c) the crystallisation vessel equipped with temperature probe, EasySampler, Blaze Micro, level sensor and transfer lines and (d) other vessels available for clean in place and solvent storage.



saving improvement. The resulting raw data is analysed (Fig. 1e) to access the key crystallisation process parameters of nucleation rates, growth rates and yield. These parameters are optimised against the crystallisation QbDD objectives (Fig. 1f), enabling optimisation in subsequent workflow cycles.

## 2.2. Hardware

The reactor system (Fig. 1a) is composed of a set of vessels with associated metering systems and control systems that can be reconfigured to perform a wide range of crystallisation experiments. The equipment in the system is summarised below in Table 1, Fig. 2 and 3. Hardware capability checks of the system were performed as sources of variation DoE experiments; these can be found in more detail in Tables S.1, S.2, Fig. S.1 and S.2.†

By configuring the system in different ways (Table 2), it is possible to perform batch or continuous, single-MSMPR (mixed suspension, mixed product removal) crystallisations for cooling crystallisations, for anti-solvent crystallisations or for crystallisation processes that use both approaches.

The entire system is controlled by Snapdragon Chemistry LabOS™ software, a web-based supervisory control and data

acquisition (SCADA) platform designed to control and monitor laboratory experiments. The platform allows researchers to collect data and control a wide range of scientific instruments through a single unified interface. Instruments are individually connected to LabOS Workers (single board computer IoT devices) with a simple USB plug-and-play architecture. The LabOS workers interface, control, and collect data from a wide range of laboratory equipment (e.g., thermocouples, pumps, stirrers, HPLCs) ensuring flexibility and reconfigurability of the system for different applications. In addition, OPC-UA is also available for the communication of process analytical technology (PAT) instruments (e.g. Blaze-Micro (Blaze Metrics), UV-vis spectroscopy, FTIR spectroscopy, etc.).

Through LabOS's web interface, users can observe and manage their experiments remotely. Users can also define automation procedures and upload sets of experimental conditions to evaluate, referred to as reactions. After configuration, users start data collection and manage and control ongoing experiments with interactive dashboards. After experiments are completed, users can view a full experimental history and export data for further analysis. LabOS also has robust

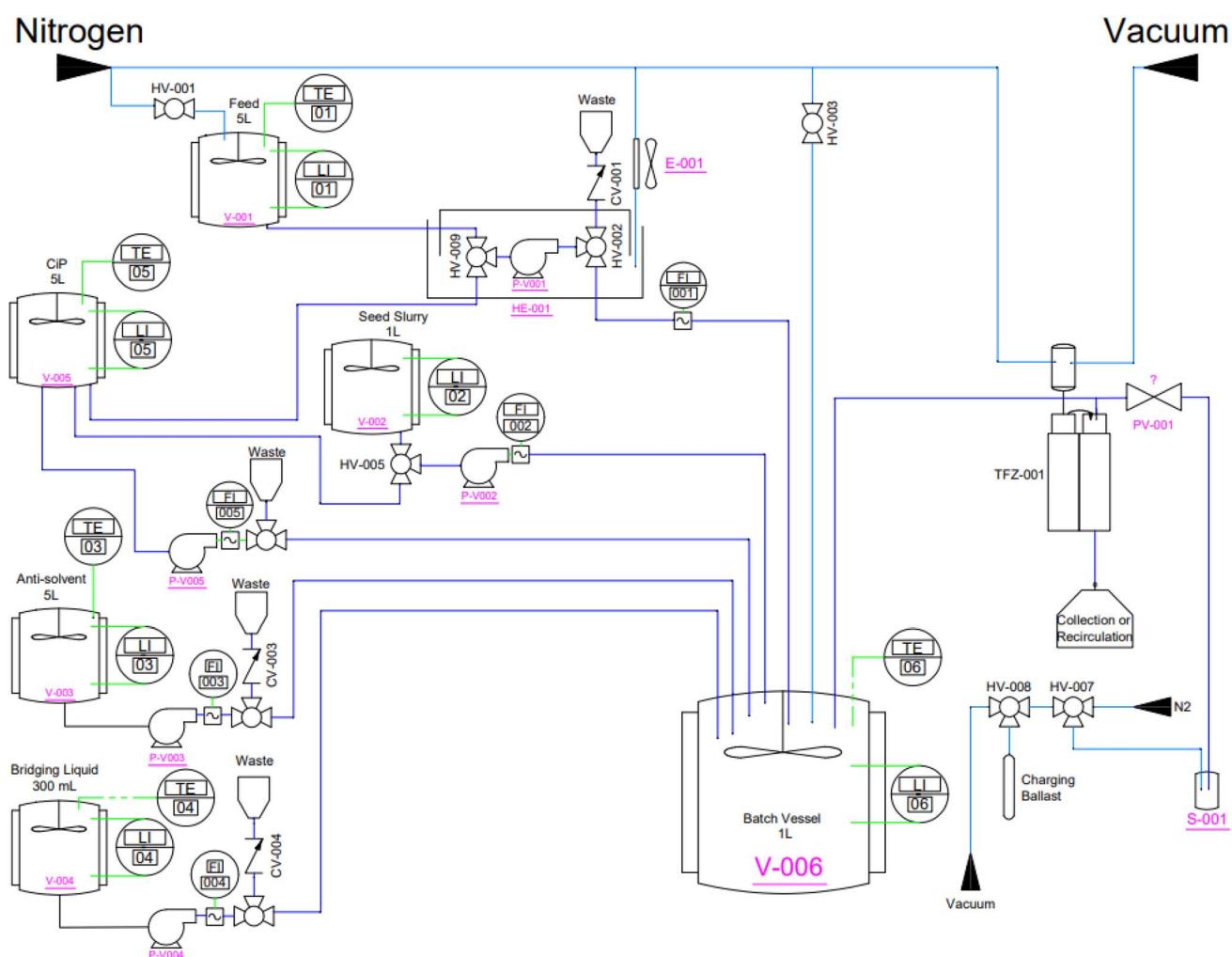


Fig. 3 Piping and instrumentation diagram for the scale-up crystallisation DataFactory equipment.



Table 2 Types of crystallisation experiments that can be performed on the platform

Crystallisation type	Reactor mode	Approach
Cooling	Batch	Charge a specified quantity of pre-heated solution of feed from V-001 to V-006. Execute temperature ramp and deliver seed from V-002 at the appropriate time. At end of experiment, pump out mixture with transfer zone, wash with CiP solution followed by other solvent(s) to reset the platform for the next experiment
	Continuous	Continuously charge pre-heated solution of feed from V-001 to V-006. V-006 is held at a lower temperature than V-001. Slurry is removed continuously to maintain constant reactor fill volume
Anti-solvent	Batch	Charge a specified quantity of solution of feed from V-001 to V-006. Dose a specified quantity of anti-solvent at a specified rate from V-003 to V-006. Deliver seed from V-002 at the appropriate time. At end of experiment, pump out mixture with transfer zone, wash with CiP solution followed by other solvent(s) to reset the platform for the next experiment
	Continuous	Continuously charge solution of feed from V-001 to V-006. Continuously charge solution of anti-solvent from V-003 to V-006. Slurry is removed continuously to maintain constant reactor fill volume

process safety features with the ability to detect unsafe conditions and alert users *via* SMS and to safely shutdown experiments. Beyond its SCADA functionality, LabOS also has an integrated electronic lab notebook (eLN) and is compliant with data integrity requirements for use in GMP settings.<sup>43</sup>

The Automated Scale-Up Crystallisation DataFactory Platform achieves Level 3 automation as defined in the self-driving lab framework,<sup>9</sup> enabling a workflow that chains multiple tasks or experiments and can integrate human-defined search spaces with computational selection to guide experimental choices.

### 3. Materials & methods

#### 3.1. Case study

The batch cooling crystallisation of lamivudine in ethanol was selected for this case study due to its favourable solubility-temperature profile, manageable slurry concentration and straightforward polymorphic landscape. Lamivudine, obtained from Phion Ltd, was a white to off-white crystalline powder with a HPLC assay showing over 98% purity and a maximum water content of 0.2%. Ethanol, sourced from VWR, had a purity of more than 99.97%. Neither the active pharmaceutical ingredient nor the solvent underwent further purification.

#### 3.2. Parameter setting

For this campaign, the primary QbDD objective was to determine the relevant process parameters to achieve desirable material and process-quality attributes. To achieve this a model must be developed capable of iterating until a reduction in the objective function, composed of QAs (see Section 3.4 for details), was achieved. The aim was to minimise the number of iterations involving active pharmaceutical ingredients (API) and solvents, aligning with the research centre's sustainability goals.

A thermodynamic model, based on data from the Crystallisation Screening DataFactory,<sup>38,42</sup> identified the maximum operating concentration of lamivudine in ethanol as 47.3 mg mL<sup>-1</sup>, which was fixed as a constant variable. The same model was used to link seed temperature to supersaturation (SS).

Additionally, a kinetic model indicated that primary nucleation occurred at high SS (greater than 2.5) for lamivudine in ethanol, therefore the risk of uncontrolled nucleation was mitigated by operating in a lower SS region.

The following process parameters from a standard cooling crystallisation method were then selected for optimisation (Fig. 1b): cooling rate, SS at seeding, and seed mass percentage, with the high-level goal of enhancing process sustainability.

#### 3.3. Experimental screening design

A five-point Latin hypercube sampling (LHS) approach was used to design the initial screening experiments, with defined bounds for each variable: cooling rate (0.1 to 0.5 °C min<sup>-1</sup>), SS (1.2 to 1.5), and seed mass (1 to 5%). These bounds were chosen to reflect typical conditions in cooling crystallisations,<sup>37</sup> constrained by the equipment's operational limits. As depicted in (Fig. 1c and S.5†), the LHS effectively covers the design space. To simplify the experimental design and avoid a high-dimensional optimisation problem, other variables — stir rate, volume, initial concentration, and initial and final temperatures were held constant (Table S.4†).

The five experiments were conducted as detailed in Table S.4,† following the reaction procedure outlined in (Fig. 1d, S.3 and S.4†). Solution concentrations were quantified by sampling, followed by 80-fold dilution in ethanol, and the area under the peak recorded at 254 nm using HPLC (Shimadzu). Using PAT, particle size and count were analysed with the Blaze-Micro probe (Blaze Metrics), reporting chord length distribution (CLD). Final yield, growth rate, and nucleation rate during the crystallisation period were calculated using the equations provided in Section S.2.3.†

#### 3.4. Bayesian (data-driven) optimisation

In industrial pharmaceutical research, it is typically desirable for the crystallisation process to take place in a growth dominated regime, while also maximising yield. The optimisation problem is therefore a scalarised single-objective problem involving these three process parameters and their associated uncertainties, within a 3-dimensional bounded space



corresponding to the independent variables from the LHS. The objective function, defined in eqn (1), incorporates the thermodynamic and kinetic parameter estimates extracted using the equations in Section S.2.3.† As the goal is to maximise (refer to Item S.1†), the inverse of the nucleation rate ( $-D_X(R_{\text{nuc.}})$ ) is included in the objective function. The extracted yield has no associated uncertainty and as such the error in yield is excluded from eqn (1).

$$f(X) = D_X(\text{yield}) + D_X(R_{\text{growth}}) + D_X(R_{R_{\text{growth}}}^2) - D_X(R_{\text{nuc.}}) + D_X(R_{R_{\text{nuc.}}}^2) \quad (1)$$

where  $X$  refers to cooling rate, seed mass and the SS at the point of seeding,  $R_{\text{growth}}$  is the growth rate,  $R_{\text{nuc.}}$  is the nucleation rate,  $R_{R_{\text{growth}}}^2$  is the  $R^2$  value of the fitted growth parameter and  $R_{R_{\text{nuc.}}}^2$  is the  $R^2$  value of the fitted nucleation parameter. All values within the objective function were normalised and equal weighting applied.

This objective function is constructed in a way to optimise the target process parameters and minimise overall uncertainty, whilst also appropriately weight individual experiments based on their associated confidence levels. The BO model employed (Fig. 1c and Item S.1†) utilises an exploitation-focused strategy to maximise the objective function value with minimal iterations, deliberately avoiding exploration of parameter space regions unlikely to yield optimal results.

## 4. Results & discussion

### 4.1. Latin hypercube sampling screening experiments

The Automated Scale-Up Crystallisation DataFactory platform successfully executed five batch experiments autonomously, requiring human intervention only for setup (feeder vessel content preparation) and post-experiment cleaning. Time series data (Fig. 1e, S.6 and S.7†) from LabOS were processed off-line in Python to extract the responses from the process

parameters. Covariance analysis (Fig. S.8†) of the initial screening data (Table 3) revealed trends consistent with established physical relationships, such as a strong positive correlation between SS and nucleation rate. However, some unexpected trends were observed, such as a decrease in nucleation rate with increasing cooling rate, likely due to the challenges of multivariate analysis on a limited dataset. While further data collection is recommended for developing a low-uncertainty kinetic model of the crystallisation process, the primary objective of this study was to demonstrate the platform's application in crystallisation and MB-DoE. Despite the limited sample size, the data provided reasonable estimates of kinetic parameters. It can also be observed from Table 3, the uncertainty in growth rate remained consistent across the five experiments, indicating a good linear correlation in the extracted data. In contrast, nucleation rate uncertainty varied, ranging from non-linear ( $R^2 < 0.1$ ) to strong linear relationships ( $R^2 > 0.9$ ). Incorporating these uncertainty parameters into the objective function (eqn (1)) was crucial, allowing for reduced influence from poorly extracted nucleation rates in experiments 2 to 5 on the direction of the subsequent experiments.

### 4.2. Optimum experiment *via* Bayesian optimisation

The Bayesian optimisation experimental planner identified the next optimal experiment within 0.15 seconds of computational time (11th Gen Intel® Core™ i5-1145G7 @ 2.60 GHz, 16 GB RAM), with BO parameters detailed in Item S.1† The recommended conditions were a cooling rate of  $0.26 \text{ }^{\circ}\text{C min}^{-1}$ , a seed mass of 2.32%, and a SS at seeding of 1.45. This experiment was subsequently executed autonomously, resulting in a 3% increase in yield over the best LHS experiment, along with a relatively high growth rate (Table 4). Notably, the 'optimum experiment' demonstrated the second-highest nucleation rate among the six trials. High  $R^2$  values for growth and nucleation rates indicate strong confidence in the kinetic parameter extraction, reflected in a significantly improved objective

Table 3 Results of the latin-hypercube sampling experiments with calculated objective function value

Cooling rate ( $^{\circ}\text{C min}^{-1}$ )	Seed mass <sup>a</sup> (%)	SS <sup>b</sup>	Yield (%)	Growth rate ( $\mu\text{m min}^{-1}$ )	$R^2$ growth rate	Nucleation rate (#/s)	$R^2$ nucleation rate	Objective function value
0.27	2.34	1.48	63	0.14	0.93	11.12	0.95	1.17
0.45	1.00	1.23	70	0.32	0.92	3.91	0.11	1.99
0.25	4.84	1.38	61	0.23	0.91	-0.07	0.00	1.07
0.41	3.04	1.30	68	0.37	0.89	-2.32	0.01	2.14
0.14	3.80	1.33	59	0.13	0.83	2.83	0.21	-0.17

<sup>a</sup> Seed mass and the validation of the transfer of seed *via* a slurry can be found in more detail in Section S.1 and is controlled by volume addition.

<sup>b</sup> SS is calculated in respect to the thermodynamic model from the Crystallisation Screening DataFactory and controlled *via* temperature at the point of seed addition.

Table 4 Results of the optimum next-best experiment with calculated objective function value

Cooling rate ( $^{\circ}\text{C min}^{-1}$ )	Seed mass (%)	SS	Yield	Growth rate ( $\mu\text{m min}^{-1}$ )	$R^2$ growth rate	Nucleation rate (#/s)	$R^2$ nucleation rate	Objective function value
0.26	2.32	1.45	73	0.19	0.95	10.23	0.96	2.31



function value. Compared to the LHS experiments, the 'optimum experiment' achieved a 7% higher objective function than the best LHS result, a 46% increase over the LHS average, and a 107% increase over the lowest LHS outcome (Fig. 1f and S.9†). This improvement is attributed to a Pareto balance of yield, growth and nucleation rates and also due to an increased confidence in the kinetic parameters and thus the model.

## 5. Conclusions

This study demonstrates the successful development and application of a bespoke DataFactory for efficient model-driven process development and scale-up. This platform integrates single-board IoT devices for automated control of laboratory hardware necessary to set-up, operate, monitor and control experiments in real time in order to identify the optimum process parameters to achieve the targeted quality attributes of the bulk crystals. A Bayesian optimisation model with an exploitation-favoured approach yielded a 7% improvement in the objective function value, over the screening data, for a cooling crystallisation of lamivudine problem within a single iteration, indicating the potential for further optimisation. The integrated, automated lab equipment operating system, LabOS enabled the cyber-physical system to effectively develop an improved process in less time than a typical manual approach would take. In addition, it provided access to key process parameters allowing subsequent modelling to be carried out. Full closed loop operation, level 3 to level 4 autonomy, will require further development of hardware to Python to hardware systems integration and incorporation of additional off-line analysis. Looking ahead, the goal of integration of screening and scale-up DataFactory capabilities offers the potential to develop high quality large data sets for model development and training as well as supporting rapid crystallisation process development from mL to L scale.

## Data availability

All data underpinning this publication are openly available from the University of Strathclyde KnowledgeBase at: <https://doi.org/10.15129/e50967b9-1ebf-41e0-bee0-8e7682b89725>. The data supporting this article and code have been included as part of the ESI† and also at [https://github.com/tpicks95/ScaleUpDF\\_DigitalDiscovery](https://github.com/tpicks95/ScaleUpDF_DigitalDiscovery).

## Author contributions

Conceptualization: T. P., M. W., M. G., D. F., A. J. F., C. J. B. methodology: T. P., Y. L., C. J. B. software: P. T. investigation: T. P., Y. L. resources: M. G., T. D. writing – original draft: T. P., D. F., C. J. B. writing – review & editing: A. J. F., C. J. B. supervision: C. J. B. project administration: J. M. funding acquisition: A. J. F.

## Conflicts of interest

There are no conflicts of interest to declare.

## Acknowledgements

The authors would like to acknowledge E. Hadjittofis (UCB), N. Nazemifard (Takeda), J. Merritt (Eli Lilly), K. Nandiwale (Pfizer), O. Watson (AstraZeneca) and Y. Jangjou (Sanofi) for discussion and guidance. The authors would also like to acknowledge Rhys Lloyd (CMAC) for project management and administration. Finally, the authors would like to acknowledge Aaron Bjarnason for their contribution to laboratory training. This work was funded jointly by Astra Zeneca, Chiesi, Eli Lilly, Pfizer, Roche, Sanofi, Takeda, and UCB. It was carried out within the CMAC Future Manufacturing Research Hub (EPSRC Grant ref: EP/P006965/1) using equipment bought through the UKRPIF Net Zero Medicines Manufacturing Research Pilot funded by Research England and the Scottish Funding Council.

## References

- 1 D. Conroy, *Development Velocity: Addressing Longstanding Challenges in Drug Development*, Clinical Researcher, 2022, vol. 36, [https://acrnet.org/2022/12/20/development-velocity-addressing-longstanding-challenges-in-drug-development#:~:text=DevelopmentVelocity\(DV\)isa,towhallyacceleratingdrugdevelopment](https://acrnet.org/2022/12/20/development-velocity-addressing-longstanding-challenges-in-drug-development#:~:text=DevelopmentVelocity(DV)isa,towhallyacceleratingdrugdevelopment).
- 2 P. Kemppainen and S. Liikanen, *Pharma Digitalisation: Challenges and Opportunities in Transforming the Pharma Industry*, European Pharmaceutical Review, 2017, <https://www.europeanpharmaceuticalreview.com/article/51733/pharma-digitalisation-challenges/>.
- 3 C. Mustoe, *Int. J. Pharmaceut. Sci.*, 2024, 125625.
- 4 Y. Barhate, H. Kilari, W.-L. Wu and Z. K. Nagy, *Chem. Eng. Sci.*, 2024, 287, 119688.
- 5 K. Y. Nandiwale, R. P. Pritchard, C. T. Armstrong, S. M. Guinness and K. P. Girard, *React. Chem. Eng.*, 2024, 9, 2460–2468.
- 6 S. Yerdelen, Y. Yang, J. L. Quon, C. D. Papageorgiou, C. Mitchell, I. Houson, J. Sefcik, J. H. Ter Horst, A. J. Florence and C. J. Brown, *Cryst. Growth Des.*, 2023, 23, 681–693.
- 7 H. Lynch, A. Bjarnason, D. Laky, C. Brown and A. Dowling, *Optimizing Batch Crystallization with Model-based Design of Experiments*, National Energy Technology Laboratory (NETL), Pittsburgh, PA, Morgantown, WV..., 2024.
- 8 C. Y. Jong, A. Mittal, G. Tristan, V. Noller, H. L. Chan, Y. Goh, E. W. Q. Yeap, S. R. Dubbaka, H. R. Nagesh and S. Y. Wong, *Org. Process Res. Dev.*, 2024, 28, 1129–1144.
- 9 G. Tom, S. P. Schmid, S. G. Baird, Y. Cao, K. Darvish, H. Hao, S. Lo, S. Pablo-García, E. M. Rajaonson and M. Skreta, *Chem. Rev.*, 2024, 124, 9633–9732.
- 10 J. Wang and A. W. Dowling, *AIChE J.*, 2022, 68, e17813.
- 11 G. Franceschini and S. Macchietto, *Chem. Eng. Sci.*, 2008, 63, 4846–4872.
- 12 M. Asteasuain and A. Brandolin, *Comput. Chem. Eng.*, 2008, 32, 396–408.
- 13 D. Frey, J. H. Shin, C. Musco and M. A. Modestino, *React. Chem. Eng.*, 2022, 7, 855–865.



14 S. Knoll, C. E. Jusner, P. Sagmeister, J. D. Williams, C. A. Hone, M. Horn and C. O. Kappe, *React. Chem. Eng.*, 2022, **7**, 2375–2384.

15 A. Belič, I. Škrjanc, D. Z. Božič and F. Vrečer, *Int. J. Pharm.*, 2010, **389**, 86–93.

16 M. Garg, M. Roy, P. Chokshi and A. S. Rathore, *Cryst. Growth Des.*, 2018, **18**, 3352–3359.

17 G. Aversano, A. Bellemans, Z. Li, A. Coussement, O. Gicquel and A. Parente, *Comput. Chem. Eng.*, 2019, **121**, 422–441.

18 A. Pohar and B. Likozar, *Ind. Eng. Chem. Res.*, 2014, **53**, 10762–10774.

19 X. Li, K. Chen, X. Wei, H. Jin, G. Wang, L. Guo and E. Tsotsas, *Powder Technol.*, 2023, **420**, 118357.

20 Q. Hu, S. Rohani, D. X. Wang and A. Jutan, *Powder Technol.*, 2005, **156**, 170–176.

21 D. C. Montgomery, E. A. Peck and G. G. Vining, *Introduction to Linear Regression Analysis*, John Wiley & Sons, 2021.

22 Y.-Y. Song and L. Ying, *Shanghai Arch. Psychiatry*, 2015, **27**, 130.

23 A. Dongare, R. Kharde and A. D. Kachare, *Int. J. Eng. Innov. Technol.*, 2012, **2**, 189–194.

24 M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt and B. Scholkopf, *IEEE Intell. Syst. Their Appl.*, 1998, **13**, 18–28.

25 J. M. Bernardo, *J. R. Stat. Soc. Ser. B: Stat. Method.*, 1979, **41**, 113–128.

26 B. Song, T. Liu, M. Zhao, Y. Cui, J. Chen, Z. K. Nagy and R. Findeisen, *Ind. Eng. Chem. Res.*, 2025, **64**, 9272–9286.

27 A. D. Clayton, E. O. Pyzer-Knapp, M. Purdie, M. F. Jones, A. Barthelme, J. Pavey, N. Kapur, T. W. Chamberlain, A. J. Blacker and R. A. Bourne, *Angew. Chem., Int. Ed.*, 2023, **62**, e202214511.

28 R. Moriconi, M. P. Deisenroth and K. S. Sesh Kumar, *Mach. Learn.*, 2020, **109**, 1925–1943.

29 C. Brown, T. McGlone and A. Florence, *Continuous Manufacturing of Pharmaceuticals*, 2017, pp. 169–226.

30 T. Pickles, C. Mustoe, C. J. Brown and A. J. Florence, *Cryst. Growth Des.*, 2024, **24**, 1245–1253.

31 S. Vedantam and V. V. Ranade, *Sadhana*, 2013, **38**, 1287–1337.

32 J. McGinty, N. Yazdanpanah, C. Price, J. H. ter Horst and J. Sefcik, Nucleation and Crystal Growth in Continuous Crystallization, *The Handbook of Continuous Crystallization*, 2020, DOI: [10.1039/9781788013581-00001](https://doi.org/10.1039/9781788013581-00001).

33 R. R. E. Steendam, L. Keshavarz, M. A. R. Blijlevens, B. De Souza, D. M. Croker and P. J. Frawley, *Cryst. Growth Des.*, 2018, **18**, 5547–5555.

34 R. K. Harris, R. R. Yeung, R. B. Lamont, R. W. Lancaster, S. M. Lynn and S. E. Staniforth, *Perkin Trans. 2*, 1997, 2653–2660.

35 D. Mangin, F. Puel and S. Veesler, *Org. Process Res. Dev.*, 2009, **13**, 1241–1253.

36 P. Barrett, B. Smith, J. Worlitschek, V. Bracken, B. O'Sullivan and D. O'Grady, *Org. Process Res. Dev.*, 2005, **9**, 348–355.

37 T. Pickles, V. Svoboda, I. Marziano, C. J. Brown and A. J. Florence, *CrystEngComm*, 2024, **26**, 4678–4689.

38 T. Pickles, C. Mustoe, C. Brown and A. Florence, *APS Spec. Issues*, 2022, **2022**, 7.

39 S. Agrawal, S. H. Rawal, V. R. Reddy and J. M. Merritt, *Chem. Eng. Res. Des.*, 2024, **205**, 578.

40 A. M. K. Nambiar, C. P. Breen, T. Hart, T. Kulesza, T. F. Jamison and K. F. Jensen, *ACS Cent. Sci.*, 2022, **8**(6), 825–836.

41 C. J. Roy and W. L. Oberkampf, *Comput. Methods Appl. Mech. Eng.*, 2011, **200**, 2131–2144.

42 T. Pickles, C. Mustoe, C. Boyle, J. Cardona, C. J. Brown and A. J. Florence, *CrystEngComm*, 2024, **26**, 822–834.

43 H. Alosert, J. Savery, J. Rheaume, M. Cheeks, R. Turner, C. Spencer, S. S. Farid and S. Goldrick, *Biotechnol. J.*, 2022, **17**, 2100609.

