Sources of Uncertainty and Error Propagation in the Small-Scale Crystallisation DataFactory

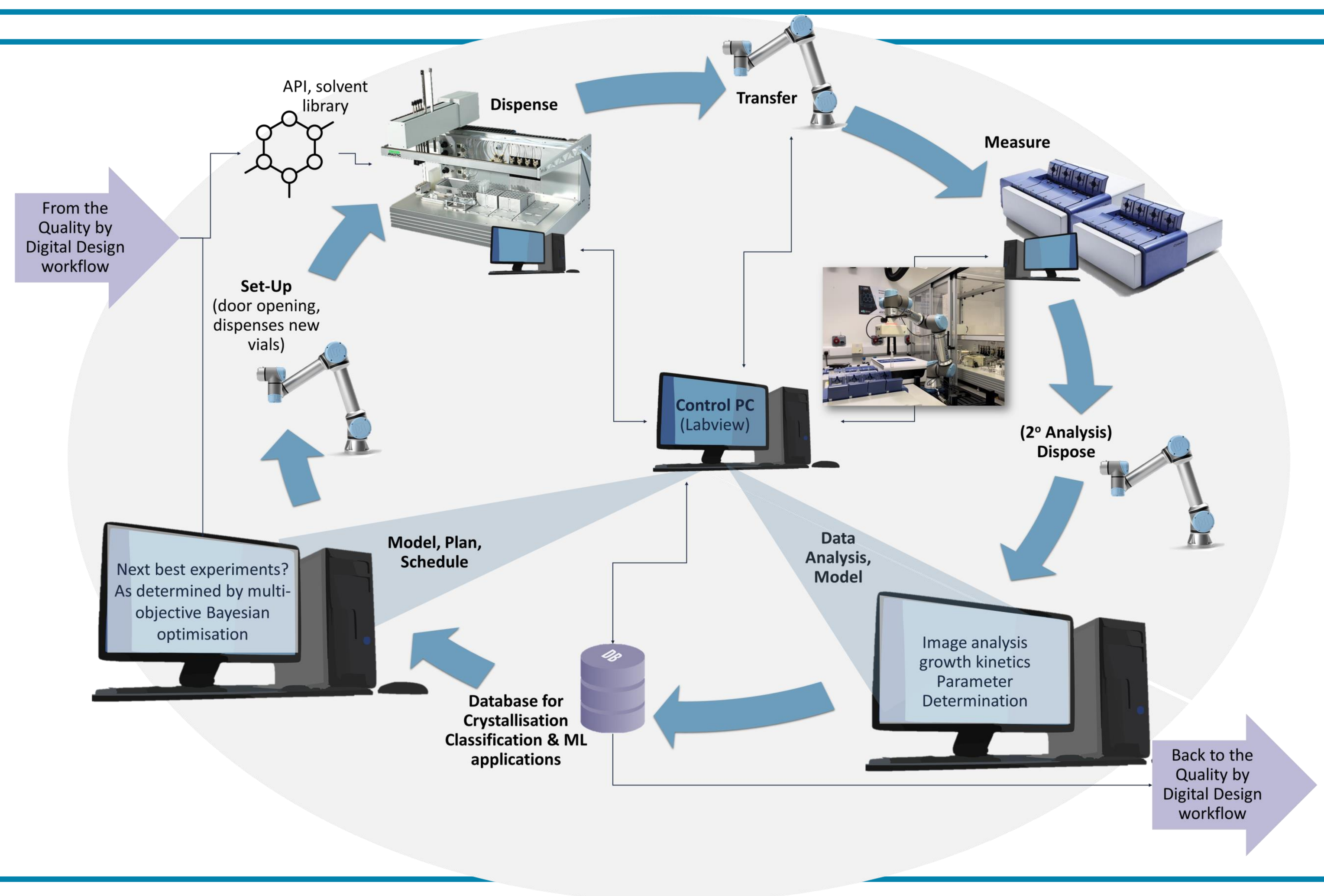
Amal Osman[1,2], Chantal Mustoe[1,2], Murray Robertson[1,2], Michael Chrubasik[3], Paul Duncan[3], Blair Johnston[1,2], Alastair Florence[1,2]
1 Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, UK
2 Continuous Manufacturing and Advanced Crystallisation (CMAC) Future Manufacturing Research Hub, University of Strathclyde, Glasgow, UK
3. National Physical Laboratory, Glasgow, UK

Introduction

Automated experiments are currently being implemented in assisting the development of the **Crystallisation Screening DataFactory** Database. Our understanding of the **sources of error**, **controlling** these sources of error and **error propagation** is limited in this new platform. This research project will focus on understanding sources of error to **inform best practices** in our **experimental set-up, data collection and data analysis,** and ultimately build confidence in digital tools developed using data collected by this platform

Methods

- **Dispense:** An automated dispensing platform (Zinsser) is used to prepare samples
- **Crystallise:** An analytical instrument (Crystalline) is used to carry out crystallisation processes with varying temperature, stirring speeds and image capture rates
- **Image analysis:** Machine learning is used to interpret crystalline images to produce thermodynamic/kinetic results



Aims

- Understand the capabilities and limitations of the DataFactory in producing reliable and consistent crystallisation data
- Investigate the uncertainty in data produced within the DataFactory and its propagation
- Quantify the possible overall confidence level of the data produced by the DataFactory

Results and Discussion

Crystallisation Data Acquisition Workflow and Identification of Solvent Systems with Nucleation

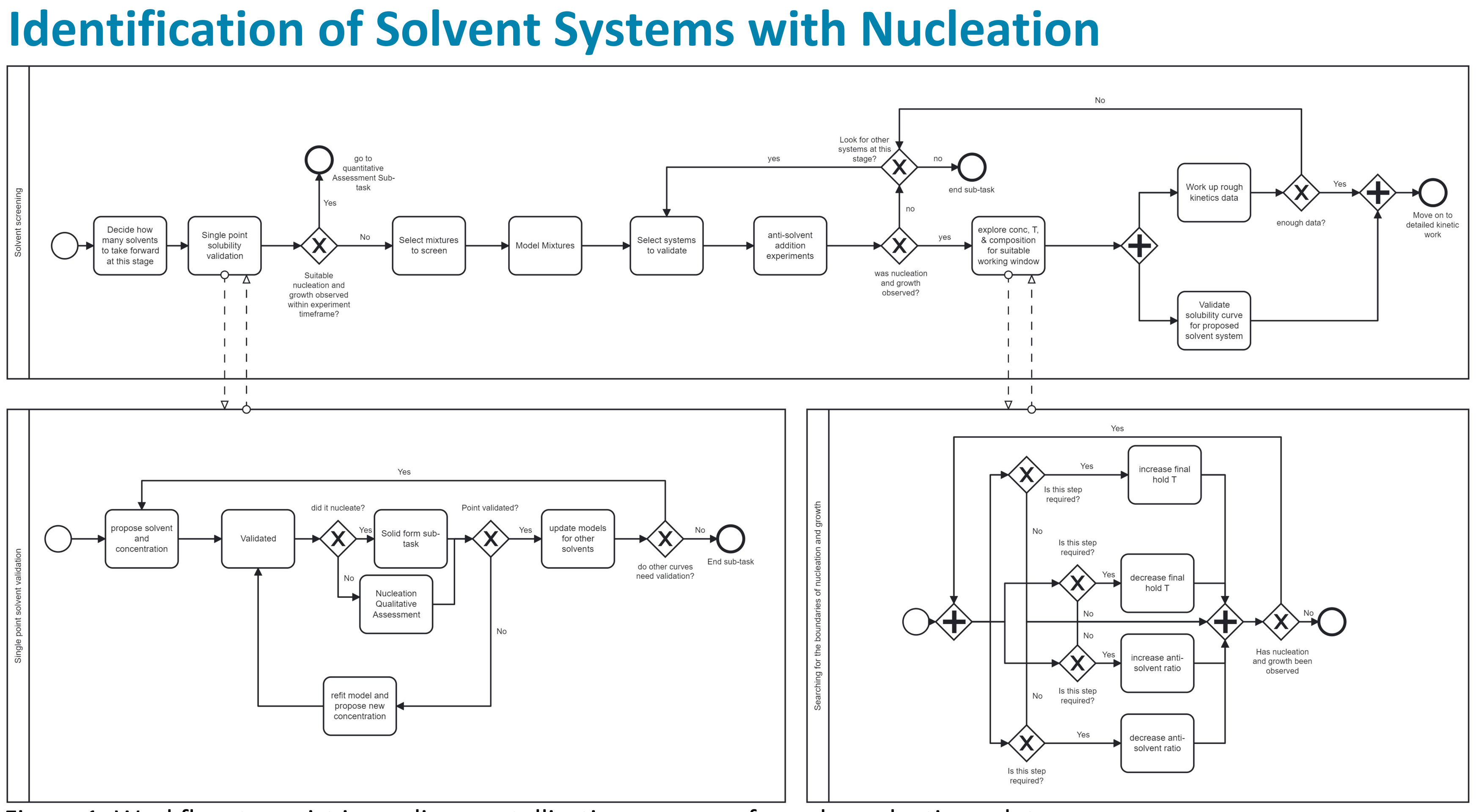


Figure 1: Workflow to assist in cooling crystallization process of poorly nucleating solutes

- Workflow facilitates the move to automated experiments and autonomous decision making in the DataFactory (figure 1)
- Following workflow, the solubility and kinetic parameters of AZD0837 in 8 solvents collected using approximately 60g of API (figure 2)
- Demonstrates COSMO solubility modelling* + experiments provides crystallisation data quickly
- Generalisable approach to other APIs that is automatable with CMAC DataFactory

Temperature		Methanol mass fraction (g/g) 0.9	0.8	0.7	0.6	0.5	0.4
-5	Concentration (MeOH:H2O)			0.1817	0.125-0.198		
	Nucleated (yes/no)			yes	some oiled out		
	clear point			31.1			
	Note			nuc started 1 hour into rerun	some nucleated		
5	Concentration (MeOH:H2O)	0.152	0.153	0.149	0.115-0.165	0.091	0.074
	Nucleated (yes/no)	No	No	yes	Yes	yes but oiling out even at 10C	no
	clear point	No	No	32.4	sample 12 45	50.9	no
	Note	12 hours duration	12 hours duration	nuc started at 4.5 hours	nuc started 1h 45	oiled out, increased temp to 10C	crashed out
15	Concentration (MeOH:H2O)	0.226	0.138	0.136	0.108	0.095	0.081
	Nucleated (yes/no)	No	No	yes	Yes	yes, after oiling out	no
	clear point	No	No	30	42	no	no
	Note	12 hours duration	12 hours duration	some particles after 5.5 hours	nuc started about 1 hour	oiled out	oiled out
25	Concentration (MeOH:H2O)		0.194	0.186	0.34-0.117	0.095	0.074
	Nucleated (yes/no)		No	No	yes	yes	no
	clear point		No	No	38	49	no
	Note		only a few particles	only a few particles	nuc started about 5 hours	nuc started about 70 mins	oiled out
35	Concentration (MeOH:H2O)		0.18	0.174	0.065	0.091	0.084
	Nucleated (yes/no)		No	No	No	No	no
	clear point		No	No	No	No	no
	Note		12 hours duration	12 hours duration	16 hours duration	16 hours duration	oiled out

Figure 2: Grid diagram summarizing kinetic results of AZD0837 antisolvent experiments carried out in methanol/water

Key:
No nucleation | Signs of oiling out | Successful nucleation

Identifying Sources of Uncertainty and Preliminary results

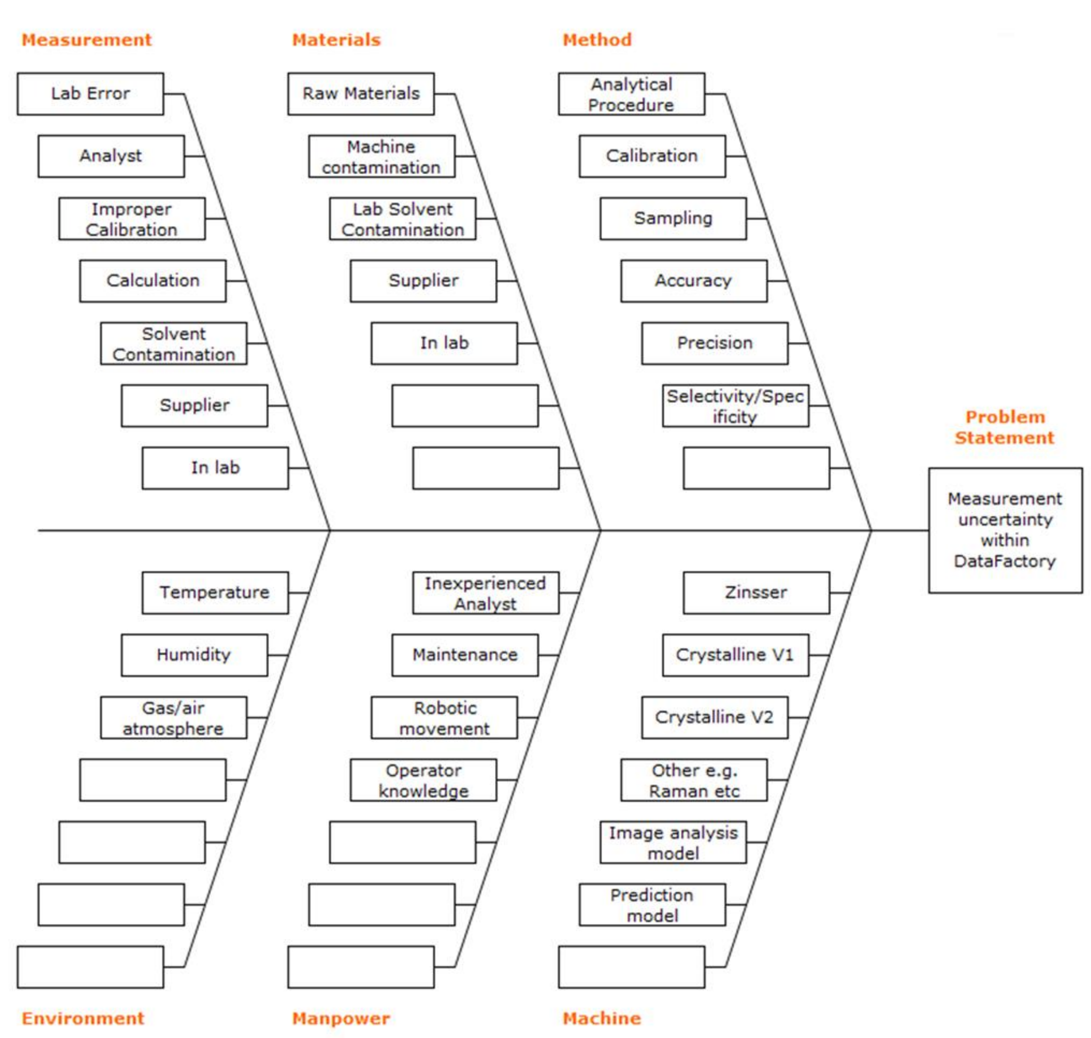


Figure 3: Sources of measurement uncertainty in the crystallisation DataFactory

Key:
Clear points/Cloud points detected by visual image assessment
Clear points/Cloud points detected by turbidity
Clear points/Cloud points detected by ML image analysis

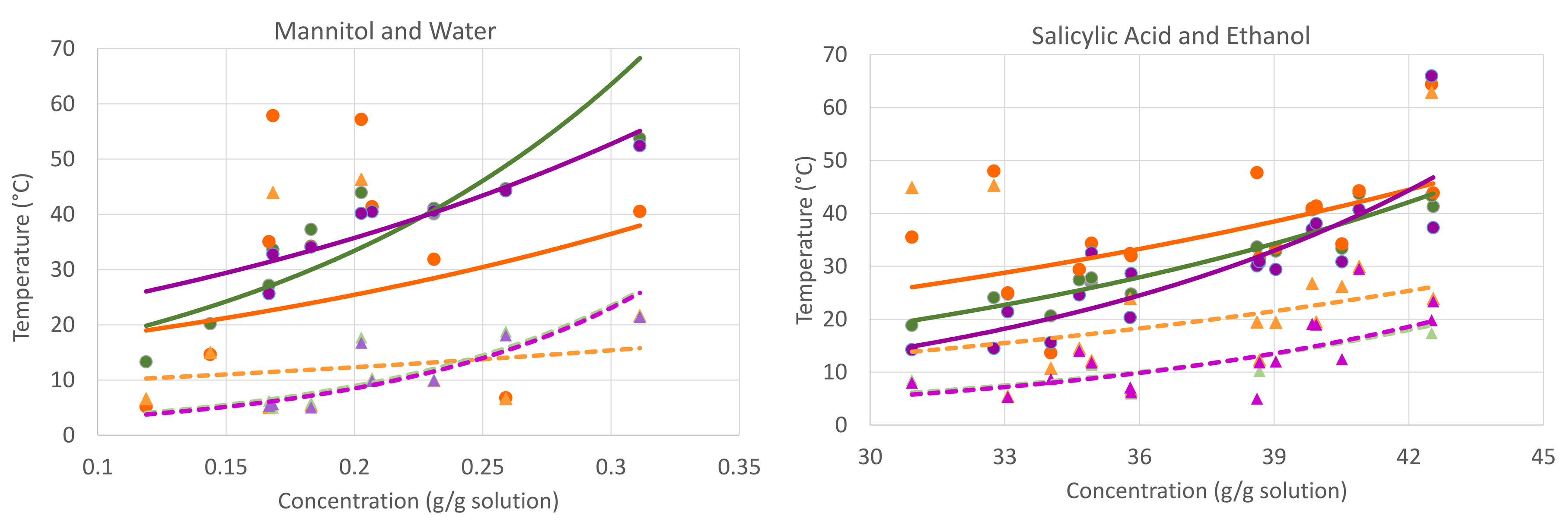


Figure 4: Phase diagrams showing detected clear and cloud points using image features (green), transmissivity (purple) and machine learning image analysis (orange)

- Sources of uncertainty throughout the DataFactory have been identified (Figure 3)
- Quantifying uncertainties in image analysis prioritised due to these results impacting future decisions and experiments. See figure 4 for preliminary results
- In general, turbidity data are reliable with regards to cloud points but not clear points.
- Preliminary image analysis results show potential in using machine learning for automated crystallisation data interpretation
- Further training required to improve its confidence level in detecting clear/cloud points accurately
- Future recommendations: train model to detect/disregard crystallisation phenomena such as non-nucleation, impurities, oiling out, solvent evaporation and other factors

Acknowledgments:
AstraZeneca for the kind donation of AZD0837 to CMAC as a model pharmaceutical compound
*Contact Murray Robertson for details on COSMO solubility models
Many thanks to Thomas Pickles and Christopher Boyle, EPSRC Future Manufacturing Research Hub for Continuous Manufacturing and Advanced Crystallisation (CMAC) for their expertise and support with regards to the DataFactory