An optimal experiment design strategy for improving parameter estimation in stochastic models

Chunbing Huang

Federica Cattani (Syngenta)
Federico Galvanin (Supervisor)

07/04/2022

Outline

Introduction & Objectives

Methodology & Results

Summary & Future work

Introduction

Stochastic Models

- Commonly used for the characterisation of dynamic systems with uncertainty
  - Thermodynamics[1]: random energy exchange → a thermal conduction
  - Epidemiology[2]: random people contacts → spread of diseases → an outbreak of pandemic (i.e. COVID-19)
  - Chemicals[3]: random molecule collisions → chemical reactions
  - Industrial seed coating[4]: random particle collision → material distribution over particles

1. Gaspard, Pierre, Gilbert, Thomas. On the derivation of fourier's law in stochastic energy exchange systems. J. Stat. Mech: Theory Exp. 2008 (11), P11021.
2. Babacar Mbaye Ndiaye, Lena Tendeng, and Diaraf Seck. Comparative prediction of confirmed cases with covid-19 pandemic by machine learning, deterministic and stochastic sir models. arXiv preprint arXiv:2004.13489, 2020.
3. Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. The journal of physical chemistry, 81 (25):2340–2361, 1977.
4. Deliang Shi and Joseph J McCarthy. Numerical simulation of liquid transfer between particles. Powder Technology, 184(1):64–75, 2008.

Introduction

Stochastic Models

Parameter Estimation[5]

- Key step in stochastic model identification
- Employ experimental observations to estimate parameter values and the corresponding variability
- Improve the predictive capability of a model for a system

5. Bard, Yonathan. Nonlinear parameter estimation. Academic Press; 1974.

Introduction

Stochastic Models

Parameter Estimation

Model-Based Design of Experiments (MBDoE)[6,7,8,9]

- Improve the quality of estimated parameters
- Use Fisher Information Matrix (FIM) to characterise the information quantity given by observed data
- Aim at maximising the Fisher information to determine the best experimental conditions and sampling points

6. Galvanin, Federico, Sandro Macchietto, and Fabrizio Bezzo. "Model-based design of parallel experiments." Industrial & engineering chemistry research 46.3 (2007): 871-882.
7. Galvanin, Federico, Massimiliano Barolo, and Fabrizio Bezzo. "Online model-based redesign of experiments for parameter estimation in dynamic systems." Industrial & Engineering Chemistry Research 48.9 (2009): 4415-4427.
8. Quaglio, Marco, Eric S. Fraga, and Federico Galvanin. "Model-based design of experiments in the presence of structural model uncertainty: an extended information matrix approach." Chemical Engineering Research and Design 136 (2018): 129-143.
9. Pankajakshan, Arun, et al. "A Multi-Objective Optimal Experimental Design Framework for Enhancing the Efficiency of Online Model Identification Platforms." Engineering 5.6 (2019): 1049-1059.

Introduction

Stochastic Models

Parameter Estimation

Model-Based Design of Experiments (MBDoE)

Model parameters $\boldsymbol{\theta}$, Operating conditions $\boldsymbol{\varphi}$, Sampling intervals $\mathbf{t}_{\mathrm{sp}} = \{t_1, t_2, \dots, t_i, \dots\}$

$$\tilde{\mathbf{y}}_i = f(\boldsymbol{\theta}, \boldsymbol{\varphi}) \Rightarrow \tilde{\mathbf{Q}}_i(\boldsymbol{\theta}, \boldsymbol{\varphi}) = \frac{\partial \tilde{\mathbf{y}}_i}{\partial \boldsymbol{\theta}} \Rightarrow \tilde{\mathbf{H}}_i(\boldsymbol{\theta}, \boldsymbol{\varphi}) = \tilde{\mathbf{Q}}_i^{\,T} \boldsymbol{\Sigma}_y^{-1} \tilde{\mathbf{Q}}_i + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \Rightarrow \tilde{\mathbf{V}}_i = \tilde{\mathbf{H}}_i(\boldsymbol{\theta}, \boldsymbol{\varphi})^{-1} \Rightarrow \tilde{J}_{\boldsymbol{i}} = \mathrm{Det}\left(\tilde{\mathbf{V}}_i\right)$$

Stochastic simulation | Sensitivity computation | Stochastic Fisher information matrix | Inverse of matrix | Information quantity

Introduction

Stochastic Models

Parameter Estimation

MBDoE

Objectives

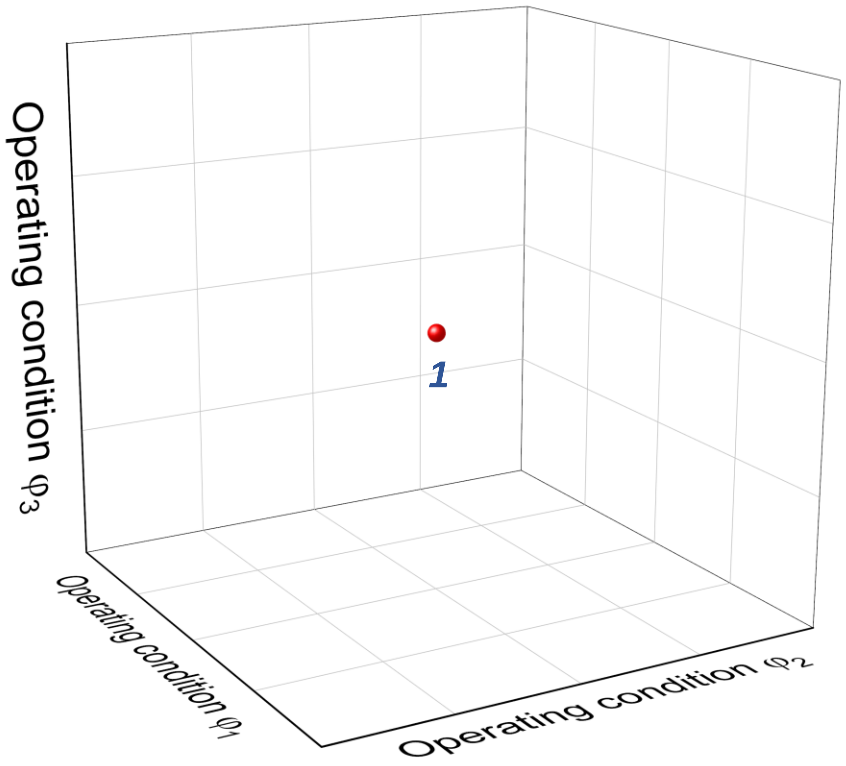
- To present a new method for stochastic model-based design of experiments (SMBDoE) to simultaneously identify the optimal experimental conditions and sampling positions.

Methodology & Results

Iterative identification of operating conditions

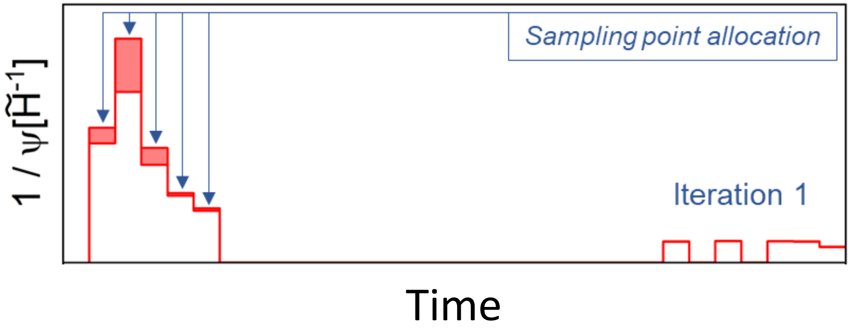
Methodology & Results

Iterative identification of operating conditions



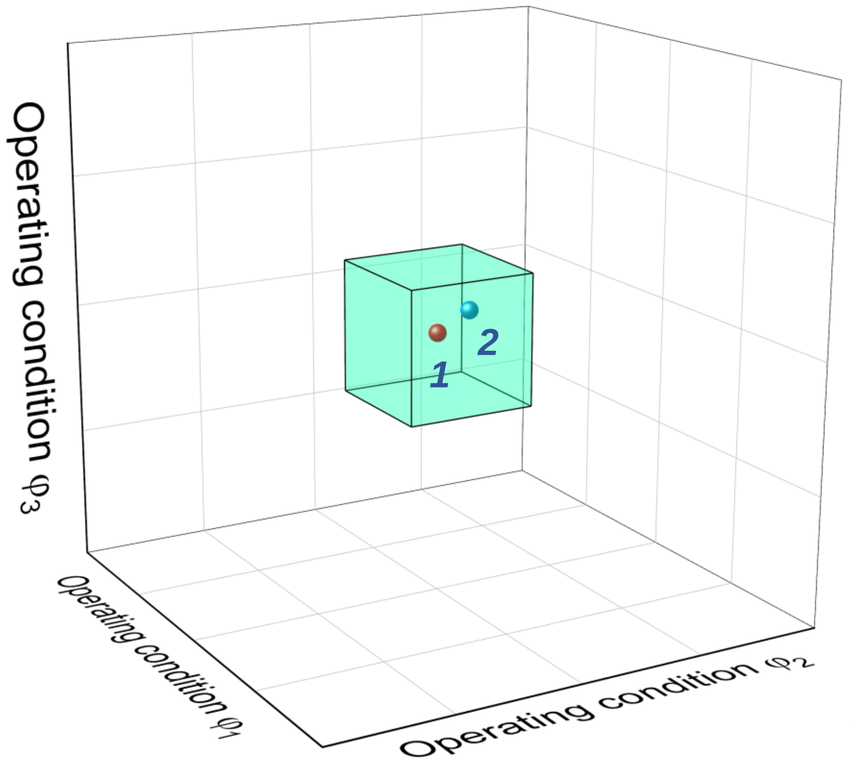
$n_{\mathrm{sp}} = 5$

Stochastic Fisher information update



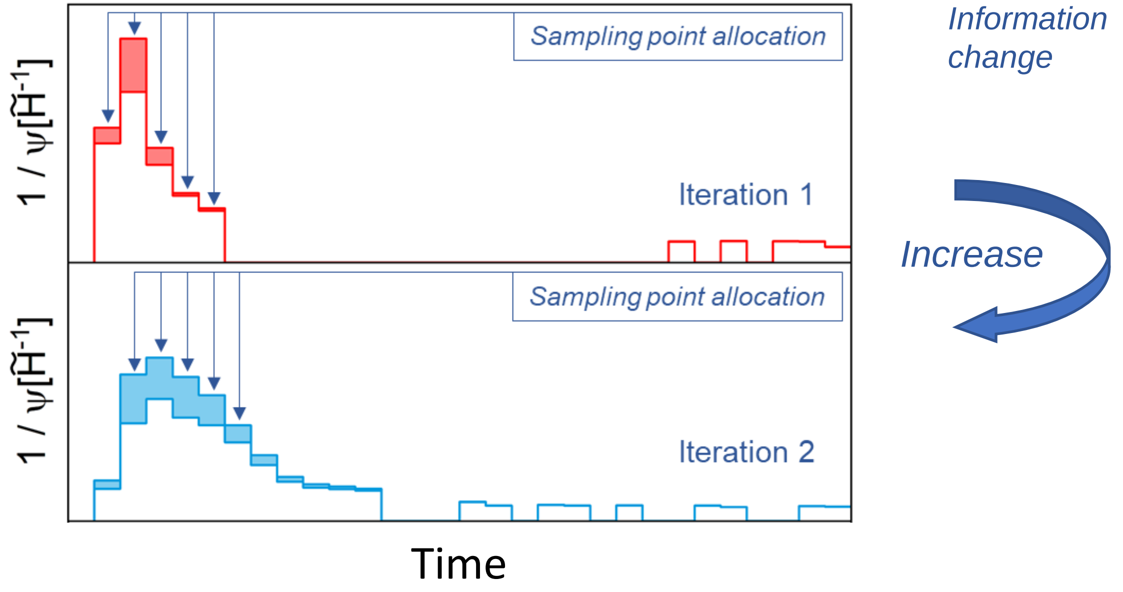
Methodology & Results

Iterative identification of operating conditions



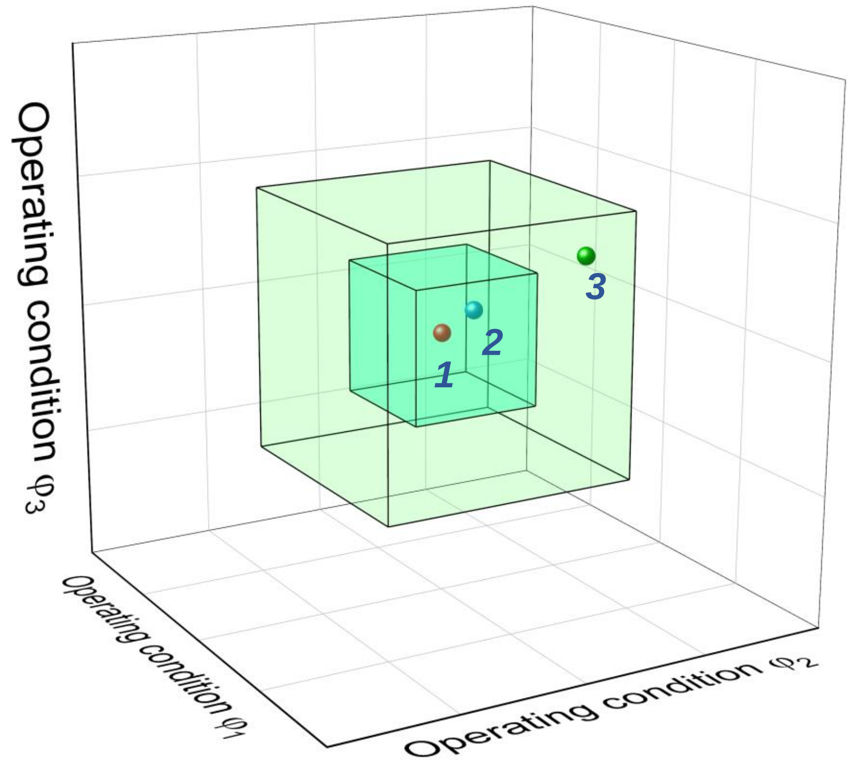
$n_{sp} = 5$

Stochastic Fisher information update



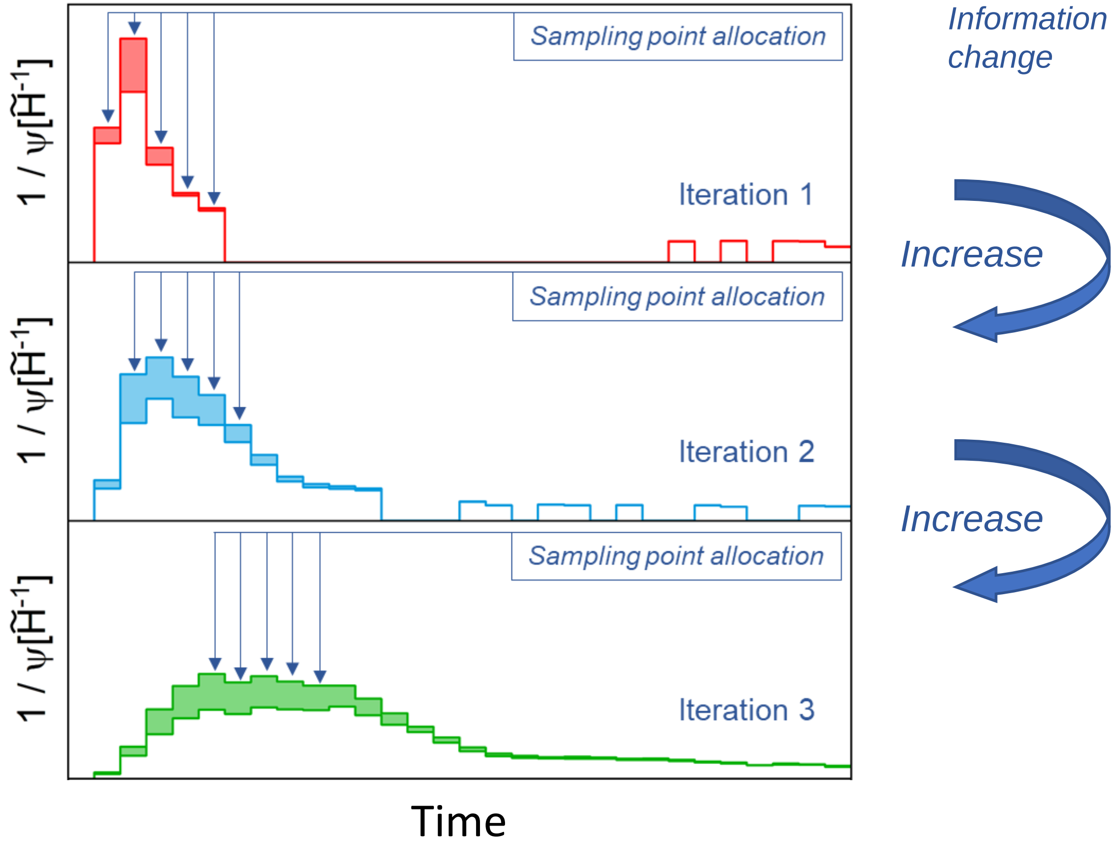
Methodology & Results

Iterative identification of operating conditions



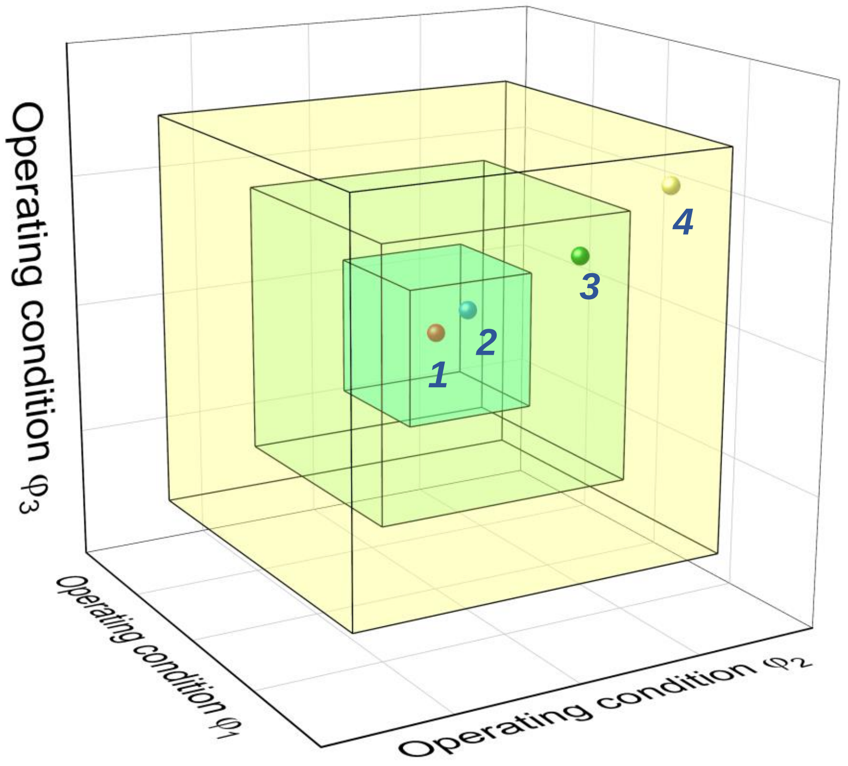
$n_{\mathrm{sp}} = 5$

Stochastic Fisher information update



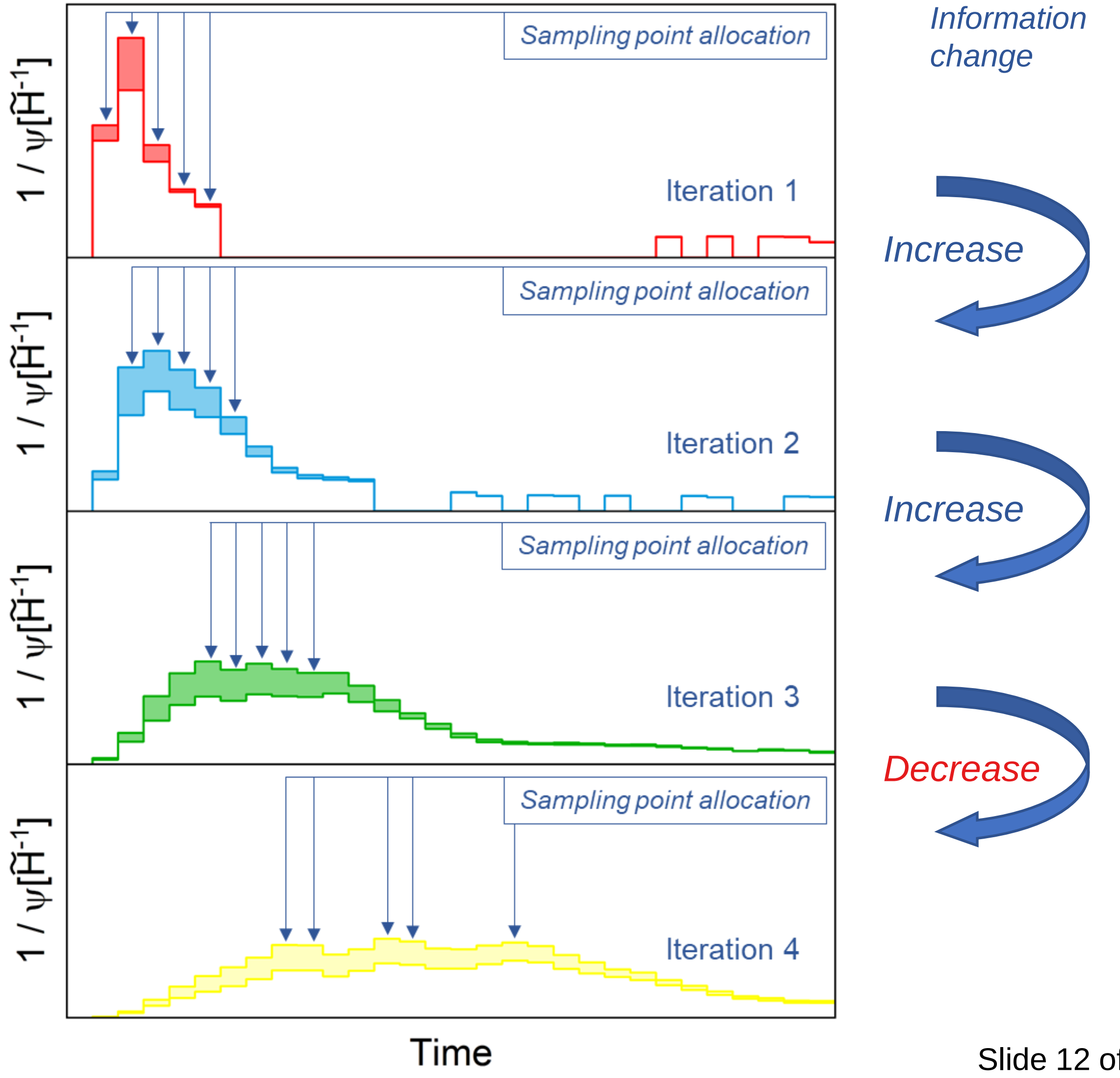
Methodology & Results

Iterative identification of operating conditions



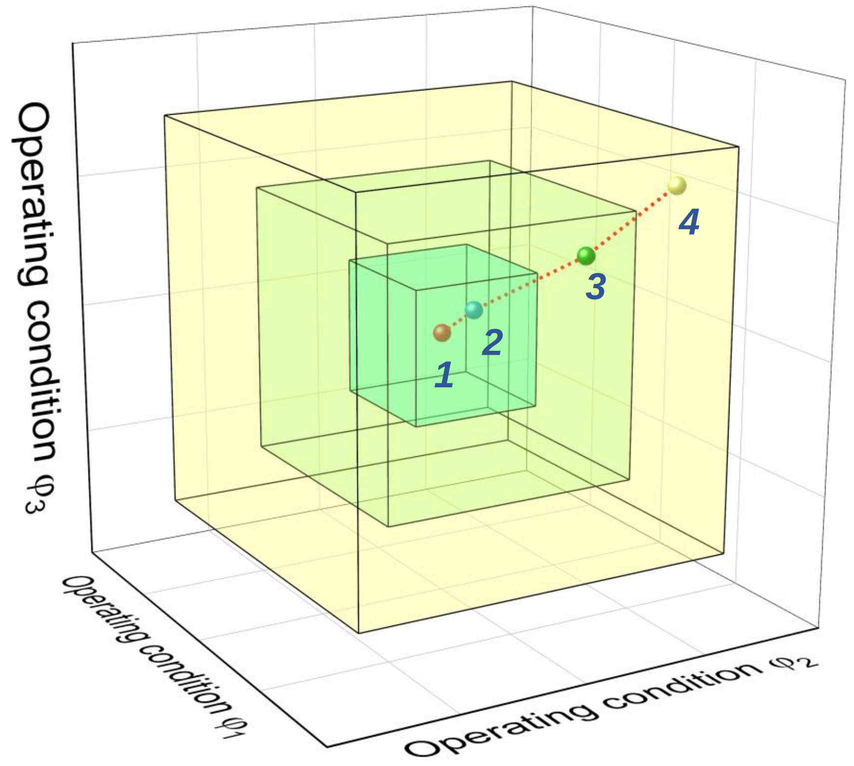
$n_{\text{sp}} = 5$

Stochastic Fisher information update



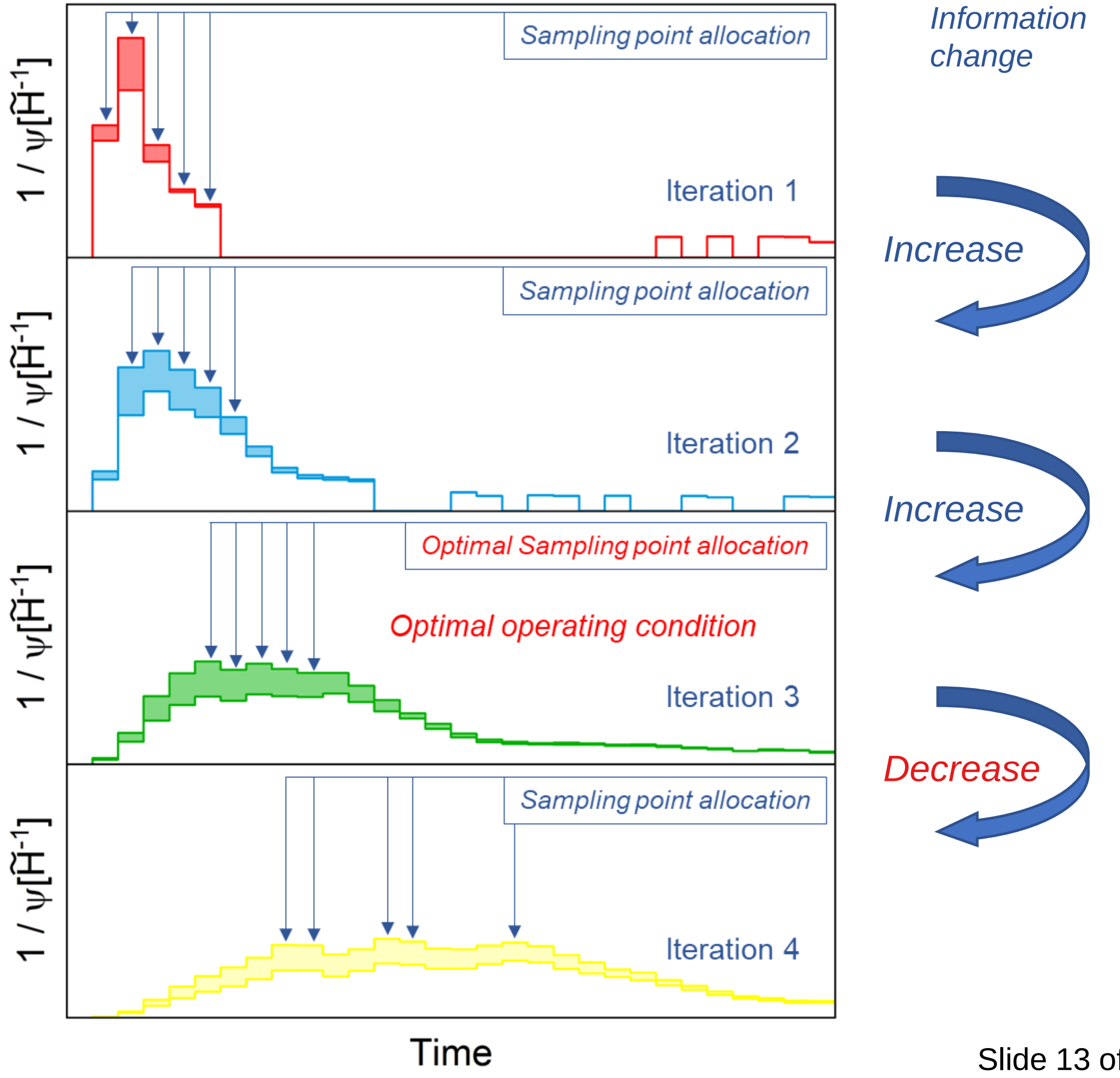
Methodology & Results

Iterative identification of operating conditions



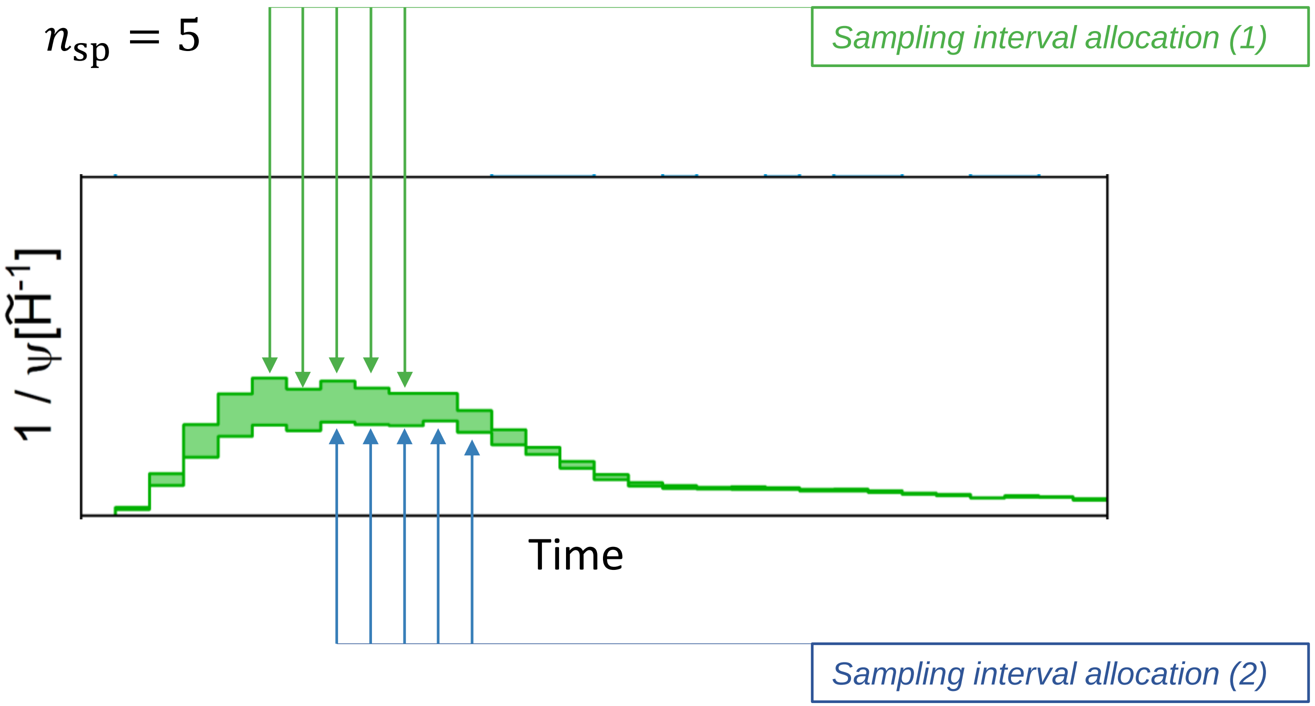
$n_{\mathrm{sp}} = 5$

Stochastic Fisher information update



Methodology & Results

Sampling interval selection criteria



- Selection criterion (1) based only on the **average** of Fisher information

- Selection criterion (2) based on both the **average** and **uncertainty** of Fisher information

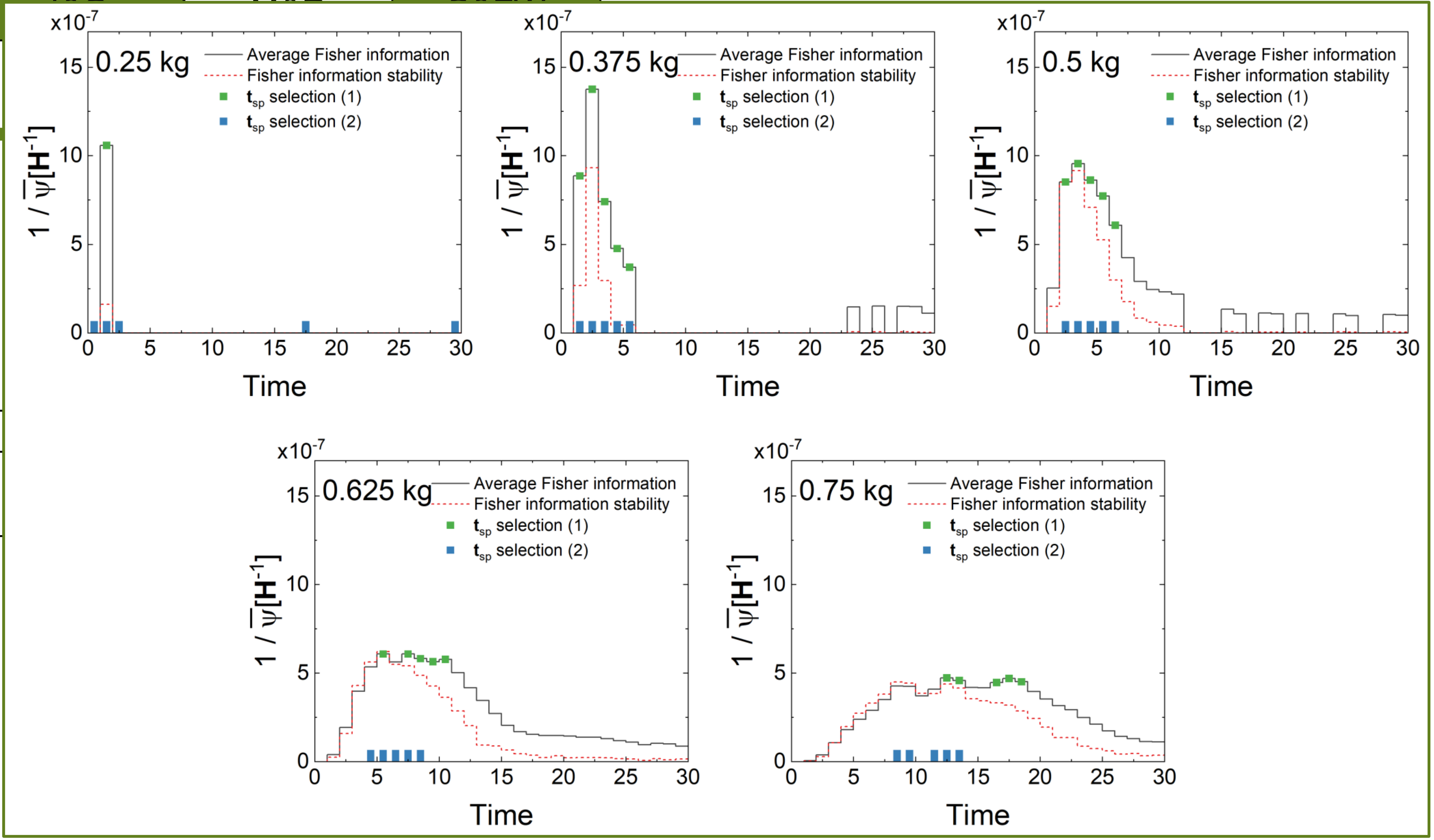
Methodology & Results

Case study

- Determine the best seed loading in the seed coating process

Example: $\boldsymbol{\theta}_{\text{true}} = [\theta_1, \theta_2, \theta_3] = [0.5000, 0.4000, 3000]$

Input variables and conditions							
	Seed loading	m_{avr}	K_x	K_y	K_z	T_{total}	T_{feed}
Range $\boldsymbol{\phi}$	0.25 – 0.75 kg	30	50	5	2-6	0-30	0.5-2



Two sampling interval selection criteria

- *Selection (1) based only on the* ***average*** *of Fisher information*
- *Selection (2) based on both the* ***average*** *and* ***uncertainty*** *of Fisher information*

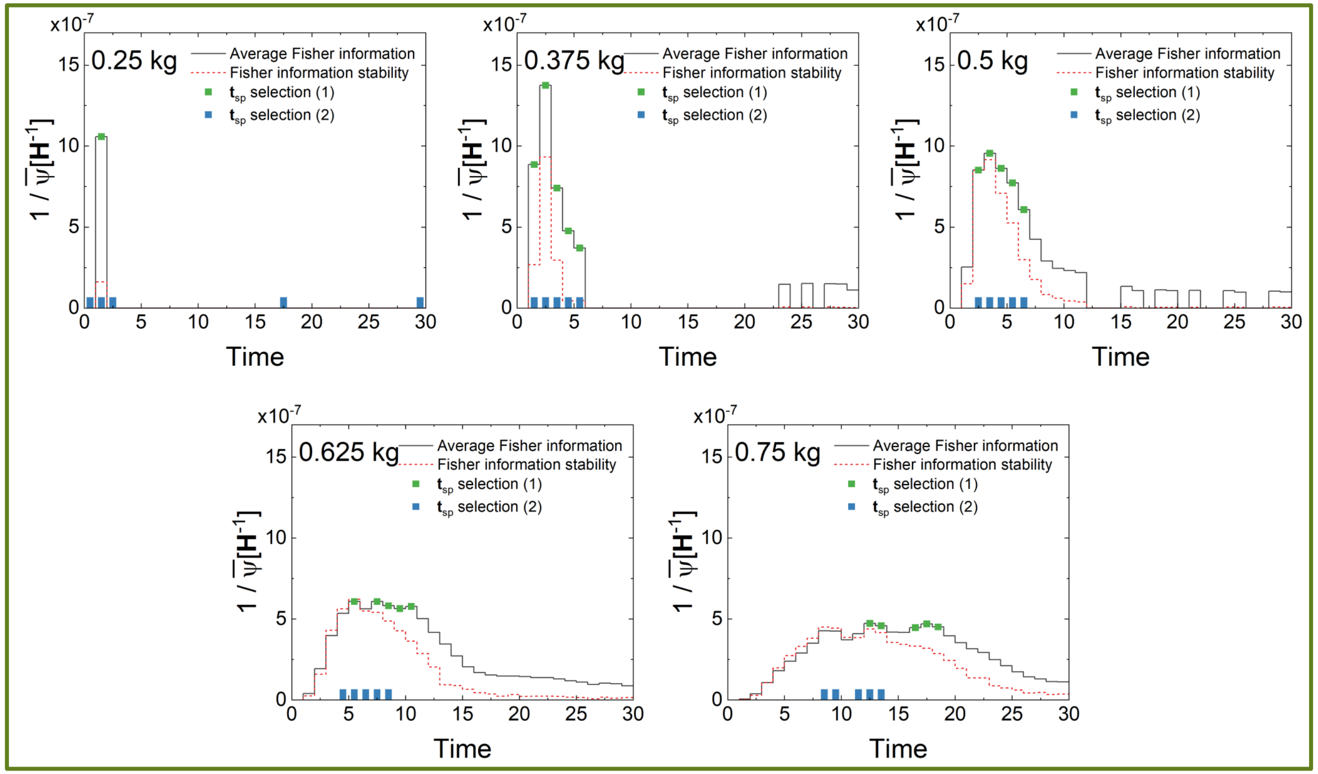
Methodology & Results

Case study

- Determine the best seed loading in the seed coating process

Example: $\boldsymbol{\theta}_{\text{true}} = [\theta_1, \theta_2, \theta_3] = [0.5000, 0.4000, 3000]$

Iteration	Experiment	Estimate	95% C.I.
1	0.25 kg	[0.6415, 0.5505, 3014.2]	[0.0102, 0.0110, 46.4]
2	0.375 kg	[0.5098, 0.4096, 3111.7]	[0.0059, 0.0060, 32.9]
3	0.5 kg	[0.4965, 0.3930, 3128.5]	[0.0032, 0.0034, 24.4]
4	0.625 kg (1)	[0.4858, 0.3878, 3164.8]	[0.0049, 0.0050, 35.2]
	0.625 kg (2)	[0.4905, 0.3920, 3117.2]	[0.0042, 0.0045, 29.5]
5	0.75 kg (1)	[0.5269, 0.4288, 2942.6]	[0.0078, 0.0081, 48.9]
	0.75 kg (2)	[0.4949, 0.3954, 3114.4]	[0.0053, 0.0055, 36.0]



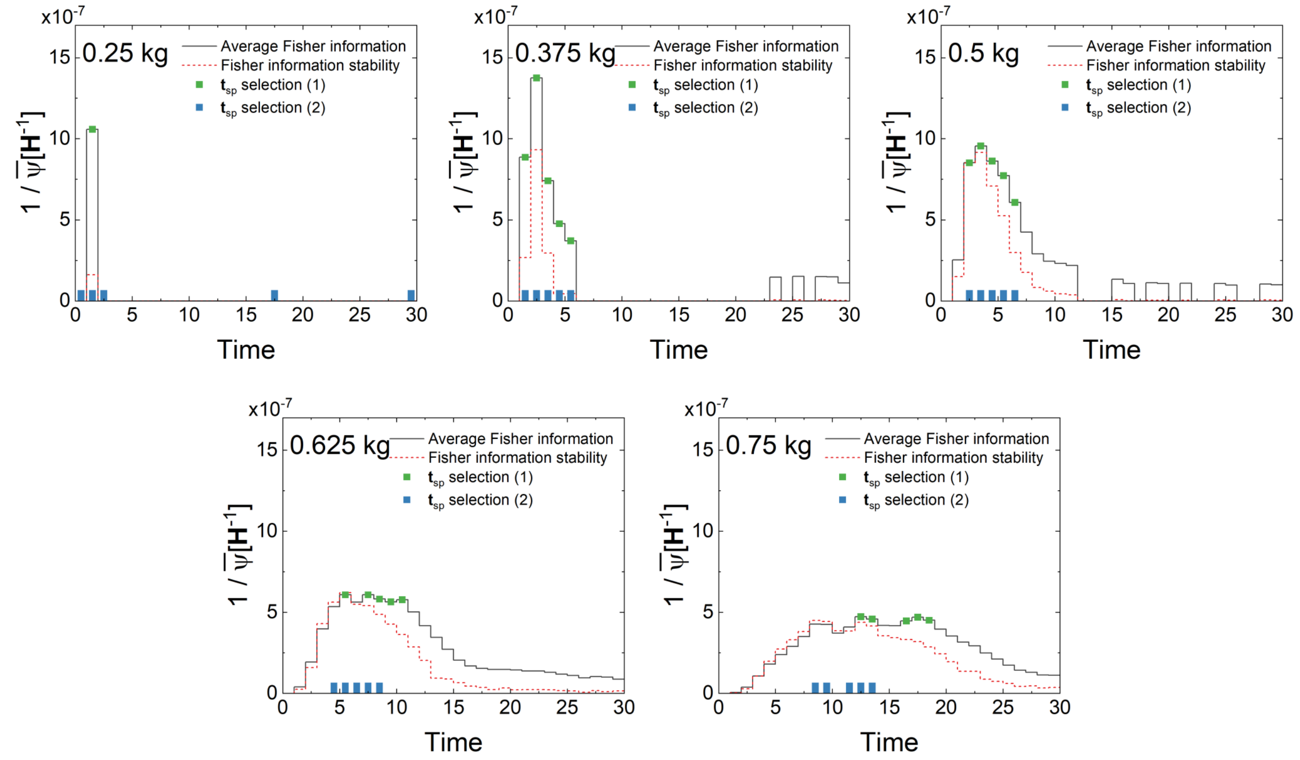
Methodology & Results

Case study

- Determine the best seed loading in the seed coating process

Example: $\boldsymbol{\theta}_{\text{true}} = [\theta_1, \theta_2, \theta_3] = [0.5000, 0.4000, 3000]$

Iteration	Experiment	Estimate	95% C.I.
1	0.25 kg	[0.6415, 0.5505, 3014.2]	[0.0102, 0.0110, 46.4]
2	0.375 kg	[0.5098, 0.4096, 3111.7]	[0.0059, 0.0060, 32.9]
3	0.5 kg	[0.4965, 0.3930, 3128.5]	[0.0032, 0.0034, 24.4]
4	0.625 kg (1)	[0.4858, 0.3878, 3164.8]	[0.0049, 0.0050, 35.2]
	0.625 kg (2)	[0.4905, 0.3920, 3117.2]	[0.0042, 0.0045, 29.5]
5	0.75 kg (1)	[0.5269, 0.4288, 2942.6]	[0.0078, 0.0081, 48.9]
	0.75 kg (2)	[0.4949, 0.3954, 3114.4]	[0.0053, 0.0055, 36.0]



Selection criterion (2) is better !

The experiment with seed loading of 0.5 kg presents the best parameter estimation result, considering the best five sampling intervals of the experiment[10]

10. Huang, C., Cattani, F. and Galvanin, F., An optimal experiment design strategy for improving parameter estimation in stochastic models, under review.

Summary & Future work

- ✓ An iterative experimental design procedure is developed to identify the best experimental conditions and sampling intervals for a better parameter estimation in stochastic models
  - ✓ Stochastic Fisher information
  - ✓ Uncertainty-based sampling interval selection criterion
- ✓ The work is validated by an industrial seed coating process as a case study
  - ✓ Determined the best seed loading condition
  - ✓ Determined the best sampling intervals during the coating process

Summary & Future work

Future work

- Consider the optimisation of overall Fisher information based on variable number of sampling intervals in different domains of operating condition
  - Use unfixed n_{sp} for stochastic Fisher information optimisation

Acknowledgements

The authors gratefully acknowledge the support of the followings

- Department of Chemical Engineering, UCL (Dean's Prize Studentship)
- Syngenta

The authors would like to thank Emily Kynaston from Syngenta for the help and support on the experiments.

Thank you !

Chunbing Huang

Federica Cattani (Syngenta)
Federico Galvanin (Supervisor)

07/04/2022

END OF PRESENTATION

Chunbing Huang

Federica Cattani (Syngenta)
Federico Galvanin (Supervisor)

07/04/2022