7th Machine Learning and AI in Bio(Chemical) Engineering Conference

02-03rd July 2024

Conference Booklet

Time	Day 1 - Tuesday 2 July 2024	Location
10:00 – 10:45	Registration + coffee	CEB Lounge
10:45 - 11:00	Welcome & opening remarks **Alexei Lapkin**	LT2
11:00 – 12:00	Keynote #1 **Andreas Bender** Using Chemical and Biological Data for Drug Discovery – Methods, Applications, and Pitfalls	LT2
12:00 – 12:25	**Jiayun Pang** Enhancing Drug Discovery with Contrastive Finetuned Sentence-Transformers	LT2
12:25 – 12:50	**Wenyao Lyu** DoE-SINDy: an automated framework for model generation and selection in kinetic studies	LT2
12:50 – 14:00	**Lunch**	LT4
14:00 – 14:35	**Adam Clayton** Bayesian Self-Optimisation for Multistep Flow Processes and Mixed Variable Reactions	LT2
14:35 – 15:00	**Johannes Seiffarth** Beyond observation in microbial live-cell imaging: Exerting control on microbial population using real time AI image analysis and response triggering	LT2
15:00 – 15:25	**Sebastian Mosbach** Twa: A dynamic knowledge graph Python package for interoperable chemistry	LT2
15:25 – 15:45	**Coffee Break**	CEB Lounge
15:45 – 16:10	**Maximilian Bloor** PC-Gym: Reinforcement Learning Environments for Process Control	LT2
16:10 – 16:35	**Arun Pankajakshan** Bayesian Classification with Active Learning for Closed-loop Identification of Feasible Operating Region in Continious Flow Crystallization	LT2
16:35 – 17:00	**Henrique Magri Marçon** AI-driven site selectivity in halogenation chemistry	LT2
17:00 – 19:00	Networking & conference dinner	Atrium

DAY 2	Day 2 - Wednesday 3 July 2024	Location
09:00	Coffee reception	CEB Lounge
09:15 – 09:50	**Michele Assante** Automation of ab-initio calculations for data-driven reaction models: integrating mechanistic DFT calculations into reaction feasibility routines	LT2
09:50 – 10:15	**Hugo Bellamy** Incorporating uncertainty information into drug design problems	LT2
10:15 – 11:15	Keynote #2 **Fernanda Duarte Gonzales** Bridging the Gap: Enhancing Retrosynthesis Prediction for Heterocycle compounds	LT2
11:15 – 12:45	Poster Session (incl. coffee break)	LT3
12:45 – 14:00	**Lunch**	LT4
14:00 – 14:25	**Thomas Andrews** A Self-Optimizing Platform for Continuous Flow Transfer Hydrogenations Using Catalytic Static Mixer Technology	LT2
14:25 – 14:50	**Aniket Chitre** Accelerating Liquid Formulations Design using Lab Automation and Machine Learning	LT2
14:50 – 15:15	**Emmanuel Agunloye** Application of Artificial Neural Networks Classifier for Rapid Identification of Chemical Reactor Models	LT2
15:15 – 15:30	**Coffee break**	CEB Lounge
15:30 – 16:10	*Workshop* - ReactWise **Henrique Magri Marçon**	LT2
16:10 – 16:30	Closing Remarks & Prizes **Alexei Lapkin**	LT2
16:30	End of Day 2	

Keynote Speakers

Professor Andreas Bender

Andreas Bender is a Professor of Molecular Informatics at the Yusuf Hamied Department of Chemistry, University of Cambridge. He is committed to developing new life science data analysis methods (AI/ML/data science) and their application, primarily related to chemical biology, drug discovery and in silico toxicology. Andreas also holds positions as the Chief Technology & Informatics Officer of PangeAI, and as co-founder of both Healx Ltd and PharmEnable Ltd.

Professor Fernanda Duarte Gonzales

Fernanda Duarte Gonzales is an associate Professor of Computational Organic Chemistry at the University of Oxford. Her group's research focuses on developing computational methodologies that help to elucidate complex (bio)chemical mechanisms and guide the design of novel catalysts and synthetic approaches. The group strives to break down some of the traditional barriers within chemistry, combining expertise in physical organic, supramolecular, and computational chemistry. Current research interests span reaction mechanism elucidation and automation, machine Learning and chemical reactivity, supramolecular design for catalysis and sensing, and enzyme catalysis.

Invited Speakers

Dr Adam Clayton

Adam Clayton is a University Academic Fellow at the University of Leeds, where he is working on the autonomous development of multistep continuous flow syntheses and telescoped catalytic reactions. His research inte rests focus on designing and applying new digital technologies to accelerate the development of more sustainable chemical processes, with particular interest in multistep continuous flow synthesis and telescoped catalytic reactions.

Dr Michele Assante

Michele Assante is a postdoctoral researcher for AstraZeneca based at the University of Cambridge in the Innovation Centre in Digital Molecular Technologies (iDMT). Here, he is working on the digitalisation of chemistry and on the merger of first principle calculations with AI. He has developed ESPlace; a python package for the treatment of solvent effects.

Using Chemical and Biological Data for Drug Discovery – Methods, Applications, and Pitfalls

Andreas Bender*

*Yusuf Hamied Department of Chemistry, University of Cambridge, UBB Cluj, UMF Cluj, and Pangea Bio
andreas@bio.bi

The amount of chemical and biological data available has increased in the public as well as the private domain, and both on the algorithmic and hardware side progress has been tremendous in machine learning. Press releases describe the design of functional proteins and antibodies from scratch, and several 'first AI-designed drugs' have already entered clinical phases.

However, all is not well when it comes to the marriage of algorithms with drug discovery, in particular when it comes to the in vivo relevance of what we are able to do with chemical and biological data at this point in time. Reasons for this are that the field is still stuck in reductionist thinking, in combination with a lack of relevant data and our ability to handle it computationally to support decision making.

This contribution will review the current status of the field, as well as provide case studies where data and computational methods have been able to select compounds with the desired effects on a biological system in various therapeutic areas, and explain what currently still hampers further progress.

Further Reading (for easy access go to: http://www.drugdiscovery.net/AIReview)

Bender A, Cortés-Ciriano I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: Ways to make an impact, and why we are not there yet. Drug Discov Today. 2021 Feb;26(2):511-524. doi: 10.1016/j.drudis.2020.12.009. (open access)

Bender A, Cortes-Ciriano I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 2: a discussion of chemical and biological data. Drug Discov Today. 2021 Apr;26(4):1040-1052. doi: 10.1016/j.drudis.2020.11.037. (open access)

Enhancing Drug Discovery with Contrastive-Finetuned Sentence-Transformers

Jiayun Pang*[1], Ivan Vulic[2]

* Corresponding author
[1] School of Science, Faculty of Engineering and Science, University of Greenwich, Medway Campus, Central Avenue, Chatham Maritime, ME4 3RL, UK. E-mail: j.pang@gre.ac.uk.
[2] Language Technology Lab, University of Cambridge, 9 West Road, Cambridge CB3 9DA, UK. iv250@cam.ac.uk

Keywords: contrastive learning, sentence-BERT, molecular embedding, few-shot learning, drug discovery.

Abstract:

In recent years, Transformer-based deep learning techniques have revolutionised the field of Natural Language Processing (NLP). These methods are increasingly being applied to chemical sciences where the sequence representation of molecular structure (such as SMILES and SELFIES) exhibits similarities with language sequence, therefore making it possible to adopt NLP algorithms to analyse molecules in a manner similar to how text is analysed. This approach can be used for a wide range of tasks, including molecular property prediction and data-driven molecular structure generation. A central task in computer-assisted drug discovery involves constructing models based on known bioactive molecules to identify other promising molecules for further activity screening. This process typically relies on tens of active molecules. When dealing with a limited dataset of bioactive molecule, the standard "pre-training and finetuning" approach in deep learning often proves no more effective than supervised machine learning models based on molecular fingerprint and physicochemical features. To address the challenges posed by the scarcity of bioactive molecules, our research explores a contrastive finetuning technique in conjunction with the Sentence-BERT (Bidirectional Encoder Representations from Transformers) framework.[1] Contrastive learning is a machine learning paradigm where data points are juxtaposed against each other to teach a model which points are similar (positive pairs), and which are dissimilar (negative pairs). Our approach consists of four stages:

1. Pretraining a BERT model using 10 million SMILES and deriving molecular representation with NLP-inspired embedding using Sentence-BERT.
2. General contrastive-finetuning using a framework called Simple Contrastive Learning of Sentence Embeddings (SimCSE) [2] and an unlabelled dataset of 10,0000 molecules.
3. Further contrastive finetuning using active and inactive molecules related to the bio-target in the training dataset.
4. Classification of molecules in the test dataset to predict whether they are active or not using the model finetuned in Stage 3.

Our results indicate that our contrastive-finetuned model significantly outperforms a molecular fingerprint-based classification model when trained with small sets of 16 or 32 active molecules. This highlights its potential for few-shot learning in drug discovery.[3] Additionally, while our approach achieves accuracy comparable to finetuning BERT, it yields 10-15% fewer false positives when screening a large dataset of 70,000 molecules and is significantly faster due to the computational efficiency of the underlying Sentence-BERT framework. We are working towards harnessing the power of molecular embedding and Transformer-base chemistry models for effective search and ranking of molecules for drug discovery.

References

1. Reimers N, Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. 2019, arXiv:1908.10084.
2. Gao T, Yao X, and Chen D. SimCSE: Simple Contrastive Learning of Sentence Embeddings. 2022, arXiv:2104.08821v4.
3. Schimunek J, Seidl P, Friedrich L, Kuhn D, Rippmann F, Hochreiter S, Klambauer G. Context-Enriched Molecular Representation Improve Few-Shot Drug Discovery. 2023, arXiv:2305.09481v1

DoE-SINDy: an automated framework for model generation and selection in kinetic studies

Wenyao Lyu[1], Federico Galvanin*[2]

[1] Department of Chemical Engineering, University College London (UCL), Torrington Place, WC1E 7JE London, United Kingdom

Keywords: nonlinear dynamic system identification, model structure generation, kinetic studies, model selection, design of experiment

Abstract:

Digital twins have revolutionised the manufacturing sector by leveraging robust and reliable kinetic models to accurately predict the behaviour of bio(chemical) reaction systems and explore a wide range of operating conditions. However, identifying these models is challenging, because the set of differentials and algebraic equations representing the reaction system typically involves many state variables and kinetic parameters [1]. Furthermore, the limited observations and inevitable experimental errors when collecting information from the system further complicate the precise identification of complex mechanisms.

To ensure the reliability and precision of the parameters, confirming a correct model structure is prior to the sequential steps of parameter estimation and model validation. Conventional model building approaches require model selection and model modification [2] and necessitate prior knowledge of candidate model structures. Conversely, model generation methods, such as sparse identification of nonlinear dynamics (SINDy) [3] only require defining a library of candidate function terms instead of the full mathematical expression of candidate models. Therefore, we employ SINDy to address situations where there is insufficient theoretical understanding of the system or where the 'true' (i.e., most suitable) model is not among the candidate models.

To model complex systems using the minimal training data, we proposed DoE-SINDy (Figure 1), a modular framework comprising seven key modules: 1) design of experiments (DoE); 2) data collection; 3) model generation; 4) model ranking and preliminary selection; 5) model calibration and reduction; 6) model evaluation and secondary selection; 7) model-based design of experiments (MBDoE) for model discrimination and parameter precision. To enhance robustness against noise and mitigate structural variations across different datasets, DoE-SINDy employs a strategy of generating multiple models from diverse subsets of experiments. Integrated with identifiability checks, parameter calibration, and rigorous evaluation and selection steps, DoE-SINDy improves the reliability of the selected final model.

This framework is tested on a simulated case study: a three-component batch reaction system described by power-law rate expressions [1]. The case study simulates the real-world process of kinetic model identification, demonstrating iterative model refinement with increasing dataset size until achieving the required adequacy or meeting the constraints of the experimental budget. The proposed DoE-SINDy is compared against state-of-the-art model identification approaches, evaluating factors such as required experimental budget, model

identification performance and computational complexity, underlining key strengths and limitations.

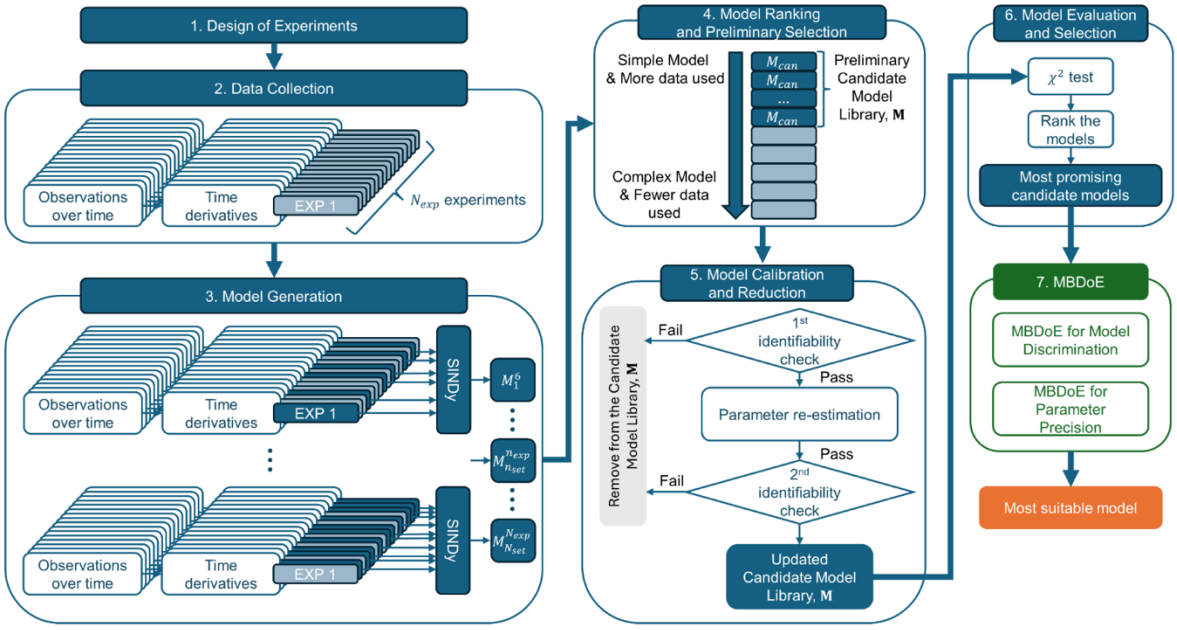


Figure 1. Framework of DoE-SINDy

References

1. Quaglio M, Roberts L, Jaapar MS, Fraga ES, Dua V, Galvanin F. An artificial neural network approach to recognise kinetic models from experimental data. Computers & Chemical Engineering. 2020 Apr 6;135:106759.
2. Asprey SP, Macchietto S. Statistical tools for optimal dynamic model building. Computers & Chemical Engineering. 2000 Jul 15;24(2-7):1261-7.
3. Brunton SL, Proctor JL, Kutz JN. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proceedings of the national academy of sciences. 2016 Apr 12;113(15):3932-7.

Bayesian Self-Optimisation for Multistep Flow Processes and Mixed Variable Reactions

Naser Aldulaijan[1], Joe A. Marsden[1], Jamie A. Manson[1], Edward O. Pyzer-Knapp[2], Mark Purdie[3], Martin F. Jones[3], Alexandre Barthelme[4], John Pavey[4], Nikil Kapur[1], A. John Blacker[1], Thomas W. Chamberlain[1], Richard A. Bourne[1], Adam D. Clayton*[1]

[1] Institute of Process Research and Development, University of Leeds, UK.
[2] IBM Research, Daresbury Laboratory, UK.
[3] AstraZeneca, Pharmaceutical Technology and Development, Macclesfield, UK.
[4] UCB Pharma SA, Brussels, Belgium.

Keywords: Automation, Bayesian Optimisation, Catalysis, Flow Chemistry, Machine Learning.

Abstract:

Self-optimisation platforms, which combine reactors, process analytics and machine learning algorithms in a feedback loop, have been shown to accelerate the optimisation of continuous parameters for single step reactions. As the synthesis of most products requires multiple steps, and includes mixtures of continuous and categorical variables, it would be desirable to translate the benefits of self-optimisation platforms to these more complex processes. However, the task of optimising these remains highly challenging, as combining reactions introduces complex interactions between the steps which must be considered holistically. Furthermore, the introduction of categorical variables within the design space can quickly increase the dimensionality of the problem, leading to significantly slower convergence.

In this work, we developed an automated continuous flow platform with a simple method for multipoint sampling, enabling accurate quantification of each reaction in a multistep process, and an in-depth understanding of the reaction pathways.[1] A new Adaptive Latent Bayesian Optimiser (AlaBO) algorithm was also designed to accelerate the development of mixed variable catalytic reactions.[2]

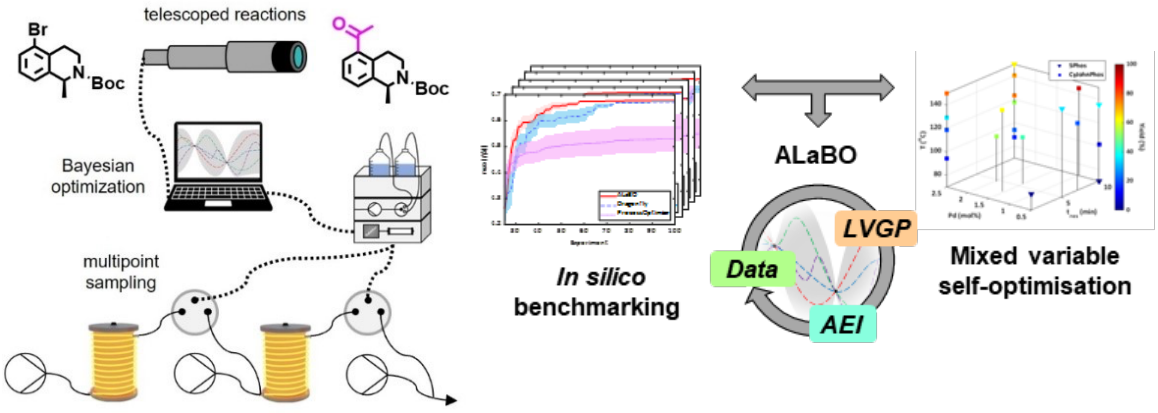


Figure 1. *Multistep self-optimisation platform and mixed variable Bayesian optimisation approach.*

References

1. Clayton AD, Pyzer-Knapp EO, Purdie M, Jones MF, Barthelme A, Pavey J, Kapur N, Chamberlain TW, Blacker AJ, Bourne RA. Bayesian Self-Optimisation for Telescoped Continuous Flow Synthesis. Angew. Chem. Int. Ed. 2023;62(3): e202214511.
2. Aldulaijan N, Marsden JA, Manson JA, Clayton AD. Adaptive mixed variable Bayesian self-optimisation of catalytic reactions. React. Chem. Eng. 2024;9(2): 308-316.

Beyond observation in microbial live-cell imaging: Exerting control on microbial populations using real-time AI image analysis and response triggering

Johannes Seiffarth[1,2], Matthias Pesch[1], Lukas Scholtes[3], Hanno Scharr[3], Dietrich Kohlheyer[1], Katharina Nöh*[1]

[1] Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, Jülich, Germany.
[2] Computational Systems Biology (AVT-CSB), RWTH Aachen University, Aachen, Germany.
[3] Institute for Advanced Simulation, IAS-8, Forschungszentrum Jülich GmbH, Jülich, Germany

Keywords: microbial control, AI image processing, live-cell imaging, real-time control, microfluidics.

Abstract:

Microfluidic live-cell imaging (MLCI) is an emerging technology that provides invaluable insights into the temporal development of living cells to study their behavior at the single-cell level [1,2,3]. Utilizing modern microfluidic devices and advanced automated microscopy, thousands of independent cell populations are monitored in picoliter sized bioreactors in a single experiment [4], turning MLCI into a high-throughput technology for efficient screening of biotechnological process parameters [5]. However, for the last decade, MLCI has been limited by the extensive image processing required for extracting quantitative insights from microscopy time-lapses. The emergence of increasingly powerful AI-driven image analysis tools for cell segmentation [6] and their continuous improvement using new datasets [7] makes automated image processing and data analysis available. Moreover, the fast inference speeds of trained deep neural networks unlocks image processing in real-time providing information during the experiment and raising opportunities to react to and even control cell behaviour [8]. In this talk, we present our new event-driven platform for ultrahigh throughput MLCI experimentation and fine-grained control based on real-time information. We combine real-time AI-driven image processing with software-defined experimentation and high-throughput microfluidic chips. We first show that this combination accelerates MLCI experiments, facilitates ultrahigh-throughput and large-scale data acquisition, and standardizes experimental procedures through software automation. We demonstrate the capabilities of our platform with the example of real-time growth control. During the running experiment triggered light pulses control the growth of photosensitive microbial strains [9]. Utilizing the ultrahigh-throughput capabilities, we demonstrate that individual control parameters can be applied per mini bioreactor to screen a wide range of control parameters in a single experiment. With our platform, we illustrate that the seamless integration of existing live-cell imaging hardware and nouveau real-time AI-driven data processing into a closed feedback loop presents a paradigm shift in MLCI experimentation and leads to a new era of live-cell analysis with unprecedented real-time control over living organisms.

References

1. Helfrich et al. Live cell imaging of SOS and prophage dynamics in isogenic bacterial populations. Molecular Microbiology. 2015;98(4): 636–650.
2. Usaj et al. Single-cell image analysis to explore cell-to-cell heterogeneity in isogenic populations. Cell Systems. 2021;12(6): 608-621.
3. Kasahara et al. Enabling oxygen-controlled microfluidic cultures for spatiotemporal microbial single-cell analysis. Frontiers in Microbiology. 2023;14.
3. Grünberger et al. A disposable picolitre bioreactor for cultivation and investigation of industrially relevant bacteria on the single cell level. Lab on a Chip. 2012;12(11):2060-2068.
4. Ho et al. Microfluidic reproduction of dynamic bioreactor environment based on computational lifelines. Frontiers in Chemical Engineering. 2022;4.
5. Jeckel et al. Advances and opportunities in image analysis of bacterial cells and communities. FEMS Microbiology Reviews. 2021;45(4).
6. Seiffarth et al. ObiWan-Microbi: OMERO-based integrated workflow for annotating microbes in the cloud. SoftwareX. 2024;26.
7. Chiron et al. CyberSco.Py an open-source software for event-based, conditional microscopy. Nature - Scientific Reports. 2022;12(1):1157
8. Hilgers et al. Genetically encoded photosensitizers as light-triggered antimicrobial agents. International Journal of Molecular Sciences. 2019;20(18):460

twa: A dynamic knowledge graph Python package for interoperable chemistry

Jiaru Bai,[1] Simon D. Rihm,[1] Aleksandar Kondinski,[1,2] George Brownbridge,[3] Sebastian Mosbach,[1,2] Jethro Akroyd,[1,2] Markus Kraft[1,2,4,5*]

* Corresponding author: mk306@cam.ac.uk (M.K.)
[1] Department of Chemical Engineering and Biotechnology, University of Cambridge, Philippa Fawcett Drive, Cambridge CB3 0AS, United Kingdom.
[2] Cambridge Centre for Advanced Research and Education in Singapore (CARES), CREATE Tower #05-05, 1 Create Way, Singapore 138602, Singapore.
[3] CMCL Innovations, Sheraton House, Cambridge CB3 0AX, United Kingdom.
[4] School of Chemical and Biomedical Engineering, Nanyang Technological University, 62 Nanyang Drive, Singapore 637459, Singapore.
[5] The Alan Turing Institute, London NW1 2DB, United Kingdom.

Keywords: dynamic knowledge graph, laboratory automation, interoperability, the world avatar, software agents

Abstract:
The rapid advancement of digitalisation, automation, and high-throughput experimentation has revolutionised data generation in chemistry, but this progress has not been matched by standardised practices in data reporting and experiment workflow documentation. This leads to poor interoperability and reproducibility among experiments conducted by different researchers. Moreover, the necessity of laborious data mining from literature obstructs the efficiency of data-driven discoveries. Addressing these challenges, we propose leveraging knowledge graph technology. Introducing "The World Avatar," a dynamic knowledge graph designed to comprehensively represent physical and abstract entities, computational agents, and software, we offer a scalable semantic solution. While The World Avatar has historically developed Java libraries, democratising access to such technology necessitates an open-source, user-friendly solution to non-experts. The Python package “twa” expands upon the World Avatar's capabilities in Java with Python-native features. Through illustrative case studies, we showcase twa's efficacy in facilitating interoperable and reproducible research in chemical science.

PC-Gym: Reinforcement Learning Environments for Process Control

Maximilian Bloor[1] , Max Mowbray[1] , Jose Torraca[1], Ilya Orson Sandoval[1], Akhil Ahmed[1], Mehmet Mercangöz[1], Calvin Tsay[1], Ehecatl Antonio Del Rio Chanona[1]*

* Corresponding author
[1]Sargent Centre for Process Systems Engineering, Imperial College London, UK.

Keywords: Reinforcement Learning, Process Control, Machine Learning Benchmarks

Abstract:
While reinforcement learning (RL) shows potential for learning control policies for complex industrial processes, further research and industry understanding has been limited due to the lack of standardized benchmarks and easy-to-use environments. To help overcome these barriers, we present pc-gym [1] - an open-source Python package providing simulation environments specifically designed for developing, evaluating, and benchmarking RL control agents for process systems. The pc-gym package is intended to serve both an educational purpose by introducing process control concepts and RL algorithms to students and industrial practitioners, as well as provide a standardized research platform for academics focused on data-driven process control algorithms such as reinforcement learning. A key focus is making the development and evaluation of RL algorithms more accessible by mitigating the complexities of process modeling and control implementation. The package builds on the OpenAI Gymnasium interface [2], implementing dynamic models of common process units like reactors, distillation columns, and heat exchangers. The modular design simplifies creating new custom environments to facilitate diverse process control studies. All environments adhere to the standard Gymnasium API for seamless integration with pre- existing RL libraries or discrete-time control algorithms.

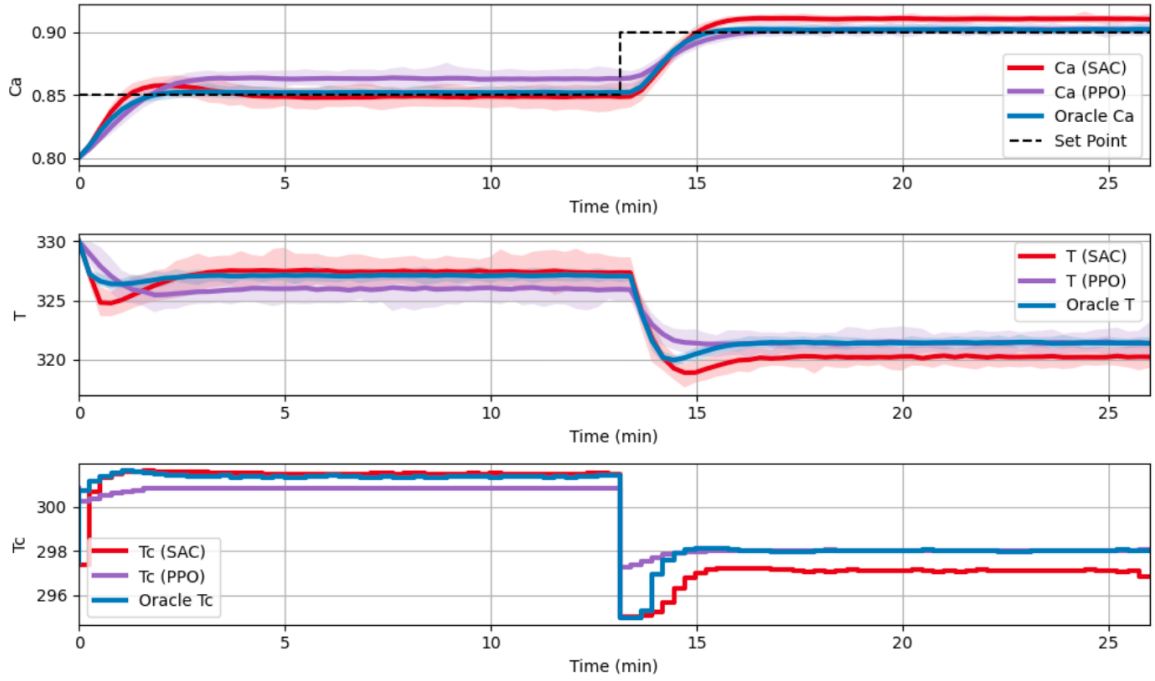


Figure 1. *Example comparison between SAC and PPO RL policies, and the oracle in the CSTR environment using the pc-gym python package.*

Pc-gym enables the creation of realistic discrete-time process control problems by allowing specification of state and control constraints, introduction of process disturbances, and structuring of observations to match industrial scenarios. The library also provides utilities for benchmarking RL policies by visualizing dynamic responses, reward distributions and comparing to optimal control with the use of model predictive control with a perfect model (oracle in Figure 1). With its educational value, standardized control benchmarks, and evaluation capabilities, pc-gym offers a unified platform to accelerate industrial adoption and academic progress in data-driven control techniques for process control.

References

1. Bloor M, Torraca J, Sandoval I, Mowbray M, Ahmed A, Mercangoz M, et al. pc-gym: Reinforcement Learning Environments for Process Control. Software. 2024;Version 0.1.6. Available from: https://github.com/MaximilianB2/pc-gym
2. Brockman G, Cheung V, Pettersson L, Schneider J, Schulman J, Tang J, et al. OpenAI Gym. arXiv:160101540. 2016; Available from: http://arxiv.org/abs/1606.01540

Bayesian Classification with Active Learning for Closed-loop Identification of Feasible Operating Region in Continuous Flow Crystallization

Arun Pankajakshan[1], Ishaa Mane[1], Sayan Pal[1] , Maximilian O. Besenhard[1], Asterios Gavriilidis[1], Luca Mazzei[1] , Federico Galvanin*[1]

* Corresponding author
[1] Department of Chemical Engineering, University College London, London, WC1E 7JE, United Kingdom.

Keywords: Bayesian classification, active learning, closed-loop, continuous flow, crystallization.

Abstract

Constrained process design is a challenging problem, particularly when the constraints that define the feasible operating region in the design space are unknown (1). Without properly identifying this region, optimal process design solutions can become infeasible.

In continuous flow crystallization, the feasible operating region can be defined as the design space region where fouling does not occur due to crystal growth and agglomeration, and where, ideally, the formation of stable crystals within the crystallizer is guaranteed. Unfortunately, this feasible region is unknown a-priori, particularly in the first phases of drug development, where resource-intensive experimental trials are carried out at different process conditions and employing different stabilizers. In this work, we propose a systematic data-driven methodology to identify the feasible operating space for the antisolvent crystallization of API drugs in closed-loop using design of experiments (DoE) and machine learning (ML) methods. Specifically, we employ Bayesian classification (2) combined with active learning (AL) (3) as ML methods to respectively identify and iteratively refine the boundaries of the feasible operating region. In Bayesian classification, a Gaussian process (GP) classifier model is adopted as the latent function. During the inference, the predictions of the posterior GP models (one for each class) are converted into relative class probabilities through a softmax (4) transformation. Then, based on a maximum likelihood principle, the class labels are obtained by means of Monte Carlo sampling from the categorical distribution of the relative class probabilities (5). In conventional AL approaches, the classifier models are iteratively updated in order to learn the true boundary separating the classes, for which the prediction of relative class probabilities is used to define uncertainty-based objective functions (such as classification margin, classification uncertainty and classification entropy) (6) to design a sequence of informative experiments. In this work, we propose a novel AL method in which the uncertainty around the predictions of latent function values (predictions of GPs) is propagated to the uncertainty around the predictions of relative class probabilities and the point with the highest value of propagated uncertainty is chosen as the next design point. An advantage of this approach is that it does not concentrate samples around the predicted boundary but explores the region of highly uncertain conditions regardless of its distance from the predicted boundary. This leads to a rapid convergence to the search of the true boundary separating the classes. The effectiveness of the proposed method is demonstrated in a real case study of closed-loop autonomous identification of the feasible operating region for antisolvent crystallization of ketoprofen drug crystals in a continuous flow crystallizer.

References

1. Tian H, Jagana JS, Zhang Q, Ierapetritou M. Feasibility/Flexibility-based optimization for process design and operations. Comput Chem Eng. 2024;180.2. Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning. Gaussian Processes for Machine Learning. 2018.
2. Lewis DD, Catlett J. Heterogeneous Uncertainty Sampling for Supervised Learning. In: Proceedings of the 11th International Conference on Machine Learning, ICML 1994. 1994.
3. Bishop CM. Bishop - Pattern Recognition and Machine Learning - Springer 2006. Antimicrob Agents Chemother. 2014;58(12).
4. Murphy KP. Machine learning: a probabilistic perspective (adaptive computation and machine learning series). Vol. 621485037, Mit Press. ISBN. 2012.
5. Joshi AJ, Porikli F, Papanikolopoulos N. Multi-class active learning for image classification. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009. 2009.

AI-driven site selectivity in halogenation chemistry

Henrique M. Marçon[1,2] , Alexei Lapkin[2,3], Alexandre Barthelme[4], Jonathan M. Goodman[1,2]*

* Corresponding author

[1] Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge – UK.

[2] Innovation Centre in Digital Molecular Technologies, Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge – UK.

[3] Department of Chemical Engineering and Biotechnology University of Cambridge, Cambridge – UK.

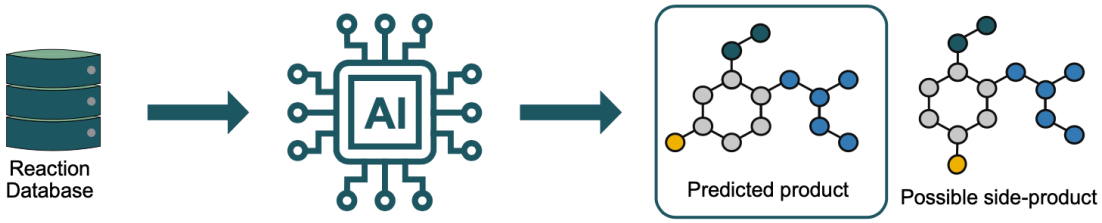
[4] UCB Biopharma SPRL, 1420 Braine l'Alleud, Belgium.

Keywords: selectivity prediction, halogenation, organic chemistry, web-interface for chemistry

Abstract:

Chemical selectivity prediction is a major challenge. Reactions can happen on various sites; predicting them typically requires time and effort from highly trained chemists. The best strategy to obtain a desired outcome is performing multiple experiments and analysing the results individually to determine the conditions which lead to the required regioselectivity. Nonetheless, the wealth of data available in chemical reaction databases should be able to help us accelerate this process. We have developed a workflow for aromatic halogenation, an important reaction class where theoretical background is available. Our pipeline achieves 79.7% accuracy on cross-validation and 79.6% on the test set. We leveraged datasets on different halogenations – fluorination, chlorination, bromination, and iodination – as well as their combination as a superset of over 17,000 reactions to enhance quantity and structural diversity of the datasets. The new workflow does not rely on computationally expensive methods, nor intensive prior knowledge of the transformation and could be quickly reproduced for new transformations where data is available – from literature databases or high-throughput campaigns – to accelerate reaction prediction in complex targets and late stage functionalisation. Our focus on rapid, low-cost models makes possible hosting on a web browser for an accessible user-experience, enabling its use by scientists with no-coding experience.

Currently, we are comparing the workflow with a range of benchmark models available in the literature. They include *ab initio* methods1, graph-based methods2, and hybrid approaches.3 Furthermore, a dataset on human performance is being collected. Currently, AI modelling outperforms human selection by 29%.



References

1. Ree, N., Göller, A. H. & Jensen, J. H. RegioSQM20: improved prediction of the regioselectivity of electrophilic aromatic substitutions. *Journal of Cheminformatics* **13**, 10 (2021).
2. Guan, Y. *et al.* Regio-selectivity prediction with a machine-learned reaction representation and on-the-fly quantum mechanical descriptors. *Chem. Sci.* **12**, 2198–2208 (2021).
3. Ree, N., Göller, A. H. & Jensen, J. H. RegioML: predicting the regioselectivity of electrophilic aromatic substitution reactions using machine learning. *Digital Discovery* **1**, 108–114 (2022).

Automation of ab-initio calculations for data-driven reaction models: integrating mechanistic DFT calculations into reaction feasibility routines

Michele Assante *[1,2]
* Corresponding author

[1] Compound Synthesis & Management, The Discovery Centre, Cambridge Biomedical Campus, 1 Francis Crick Avenue, AstraZeneca, CB2 0AA Cambridge, UK.
[2] Innovation Centre in Digital Molecular Technologies, Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom.

Keywords: Ab-initio Calculations, Automation, Density Functional Theory, Reactivity Prediction, data-driven methods.

Abstract:

Ab-initio calculations, such as Quantum-Mechanics and Density Functional Theory, can offer great insight into chemical systems and can be used to explain and predict reactivity of chemical compounds. However, these methods require significant amount of time and expertise to be set-up, performed and post-processed. A recent promising application of ab-initio calculations sees them in combination with machine-learning models to tackle current challenges such as reaction prediction and optimisation when scarce data is available. [1] However, for an efficient integration with data-driven method, a faster and more reliable way to perform ab-initio calculations is needed. Here, a computational workflow is presented that is able to automate ab-initio calculations, from the SMILES strings of reaction component to the final kinetic and thermodynamic data for the reaction. The workflow is used to describe at DFT level the Nickel dark cycle of a metallaphotoredox sp2-sp3 cross coupling reaction. This approach significantly reduces the time and human intervention required for setting-up and performing such calculations for a small library of reactions (circa 50) and produces computational data ready to be fed to machine learning models.

References

1. Jorner K, Brinck T, Norrby PO, Buttar D. Machine learning meets mechanistic modelling for accurate prediction of experimental activation energies. Chem Sci. 2021 Jan 28;12(3):1163–75.

Incorporating uncertainty information into drug design problems

Hugo Bellamy[1*] Joachim Dickhaut[2] Ross King[1]

* Corresponding author: hpb32@cam.ac.uk

[1] Department of chemical engineering and biotechnology, University of Cambridge, United Kingdom
[2] BASF, Ludwigshafen, Germany

Keywords: Drug design, machine learning, uncertainty

Abstract

Modern drug design uses machine learning to identify possible compounds from databases of millions of candidate molecules. Typically, the property of interest is the EC50 – the concentration at which a molecule has a 50% effectiveness. EC50s are calculated by fitting curves to experimental data. How well the data fits the curve gives an indication of the reliability of the calculated EC50 value. In drug design this source of uncertainty is usually ignored during model fitting because it can be difficult to both identify and incorporate into the model. We describe how to estimate an uncertainty value for each EC50 value and, by adapting a random forest to make use of these noise estimates, we are able to significantly improve the predictive performance of the model. This approach allows for a more efficient use of data in drug design reducing the total number of experiments that must be performed, saving on the time and cost required to design new drugs. Our work demonstrates the importance of having high quality data and properly preparing this data to achieve the best outcomes from machine learning.

Bridging the Gap: Enhancing Retrosynthesis Prediction for Heterocycle compounds

E. Wieczorek,[a] J. W. Sin,[a] M. T. O. Holland,[a,b] Liam Wilbraham,[c] V. S. Perez,[c] A. Bradley,[c] D. Miketa,[c] P. E. Brennan,[b] F. Duarte[a]

* fernanda.duartegonzalez@chem.ox.ac.uk, www.duartegroupchem.org, @duarte_group
[a] Chemistry Research Laboratory, University of Oxford, Mansfield Road, Oxford OX1 3TA, U.K
[b] Alzheimer's Research UK Oxford Drug Discovery Institute, Centre for Artificial Intelligence in Precision Medicine, Centre for Medicines Discovery, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, U.K.}
[c] Exscientia plc, The Schrödinger Building Oxford Science Park, Oxford OX4 4GE, U.K.

Heterocycles are important scaffolds in medicinal chemistry that can be used to modulate binding and pharmacokinetic properties of drugs. Despite their importance, existing datasets on heterocyclic compounds often lack information on how to actually make them, making it challenging to access novel heterocycles. While retrosynthetic prediction models have emerged as promising approaches to assist synthetic chemists, their performance is poor for heterocycle formation reactions due to low data availability.

In this talk, I discuss our efforts to overcome the low data availability problem and improve the performance of retrosynthesis prediction models for ring-breaking disconnections. We explore four different methods to improve these models by leveraging transfer learning techniques, reaching more > 60% are both chemically valid and involve breaking a ring. We illustrate the applicability of this model by successfully recreating the synthesis routes of drug-like compounds recently published.

A Self-Optimizing Platform for Continuous Flow Transfer Hydrogenations Using Catalytic Static Mixer Technology

Thomas Andrews[*1,2], Anastasios Polyzos[1,2], Thomas Kohl[2]

* Corresponding author
[1] School of Chemistry, The University of Melbourne, Parkville, VIC, Australia.
[2] CSIRO Manufacturing, Clayton, VIC, Australia

Keywords: Bayesian optimization, reaction optimization, uncertainty, flow chemistry.

Abstract (<400 words):

Hydrogenation reactions represent an important class of transformation that is critical to the production of pharmaceuticals, agrochemicals, advanced materials and fine chemicals. They are increasingly performed using continuous flow technologies to improve their safety, efficiency and scalability. Catalytic static mixers (CSMs) are a recent invention that promises to significantly improve the practicality and efficiency of performing heterogeneously catalyzed hydrogenations in flow[1]. CSMs immobilize catalyst onto 3D printed static mixer scaffolds (Figure 1) to achieve optimal catalyst exposure, mixing and heat transfer whilst minimizing pressure gradients.

The prevalence of hydrogenations in chemical synthesis mean that there is a constant demand to optimize them for deployment into the industrial settings. Machine learning strategies such as Bayesian optimization have already seen success when integrated into the reaction optimization process[2-5], however these approaches generally focus on small scale reactions with a high degree of experimental certainty.

This talk will discuss the development a self-optimizing system for efficiently optimizing reactions on scalable CSM systems using Bayesian optimization strategies. The algorithm implements bounded length-scales and homoscedastic noise modeling during hyperparameter optimization and a modified version of the expected improvement acquisition function to improve the resilience of the optimizer to noise whilst minimizing time to optima. The modified variant of expected improvement considers experimental spacing when determining value to ensure an adequate distribution of points for reliable fitting of models to noisy data whilst still ensuring optimal exploitation. The performance of the system was demonstrated on a range of transfer hydrogenations over palladium coated CSMs using online ^{1}H NMR to determine reaction performance in real time (Figure 1).

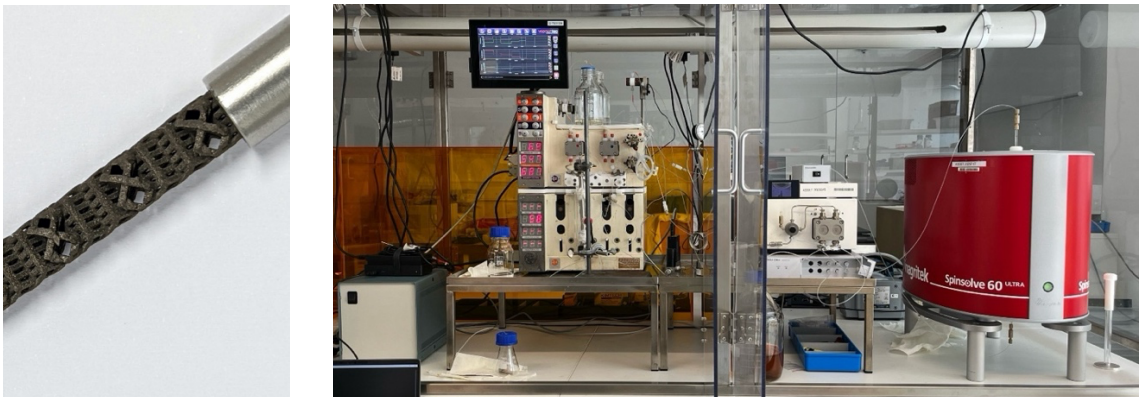


Figure 1: (Left) - A 3D printed CSM. (Right) - System for the automated optimization of reactions over CSMs using online ^{1}H NMR to determine reaction performance.

References

1. Hornung, C.H., Nguyen, X., Carafa, A., Gardiner, J., Urban, A., Fraser, D., Horne, M.D., Gunasegaram, D.R. and Tsanaktsidis, J. Use of catalytic static mixers for continuous flow gas–liquid and transfer hydrogenations in organic synthesis. *Org Process Res Dev.* 2017;21(9):1311-1319. doi:10.1021/acs.oprd.7b00180
2. Slattery, A., Wen, Z., Tenblad, P., Pintossi, D., Sanjose-Orduna, J., den Hartog, T. and Noel, T. An all-in-one multipurpose robotic platform for the self-optimization, intensification and scale-up of photocatalysis in flow. 2023. doi: 10.26434/chemrxiv-2023-r0drq
3. Wagner, F., Sagmeister, P., Jusner, C.E., Tampone, T.G., Manee, V., Buono, F.G., Williams, J.D. and Kappe, C.O. A Droplet Flow Platform with Multiple Process Analytics Facilitates Flexible Reaction Optimization. 2023. doi: 10.26434/chemrxiv-2023-gb117
4. Clayton, A.D., Pyzer-Knapp, E.O., Purdie, M., Jones, M.F., Barthelme, A., Pavey, J., Kapur, N., Chamberlain, T.W., Blacker, A.J. and Bourne, R.A.. Bayesian Self-Optimization for Telescoped Continuous Flow Synthesis. *Angewandte Chemie*. 2022;135(3):202214511. doi:10.1002/anie.202214511
5. Nandiwale, K.Y., Hart, T., Zahrt, A.F., Nambiar, A.M., Mahesh, P.T., Mo, Y., Nieves-Remacha, M.J., Johnson, M.D., García-Losada, P., Mateos, C. and Rincón, J.A. Continuous stirred-tank reactor cascade platform for self-optimization of reactions involving solids. *React Chem Eng.* 2022;7(6):1315-1327. doi:10.1039/D2RE00054G

Application of Artificial Neural Networks Classifier for Rapid Identification of Chemical Reactor Models

Emmanuel Agunloye, Asterios Gavriilidis, Federico Galvanin*
Department of Chemical Engineering, University College London, London WC1E 7JE, Affiliation 1.
* Corresponding author: f.galvanin@ucl.ac.uk

Keywords: Artificial neural networks, chemical reactor models, classification task, differential evolution algorithm, kinetic models.

Abstract:
Artificial neural networks (ANNs) are deep learning algorithms increasingly applied in different chemical engineering areas [1]. Recent work employed ANNs in a simulated reaction case study to successfully classify and rapidly recognise reaction kinetics models [2] and optimise the design of experiments (DoEs) to improve the ANN performance by using a differential evolution algorithm (DEA) [3]. In this work, we apply this ANN-DEA classification framework to the analysis of a cascade of continuously stirred tank reactor (CSTR) and plug flow reactor (PFR) (equivalent to infinite CSTRs-in-series [4]), considering as a case study the pharmaceutically relevant nucleophilic aromatic substitution comprising series and parallel reaction steps [5].
Following the hybrid modelling ANN-DEA framework, candidate reactor models are first simulated to generate large sets of in-silico data by sampling the uncertain kinetic parameters space specified a priori at fixed DoEs defined by reactants' inlet concentrations, reactor temperature and fluid residence time. Thereafter, the generated labelled data, split in the ratio 60:20:20, is used for ANN training, validation and testing using the classification accuracy metric for monitoring the network performance. DEA searches the experimental design space via Latin hypercube sampling of a population of DoEs to produce a new generation, ranking the resulting experiments based on ANN classification performance and removing poorly performing designs before selecting the optimal one. The in-silico data are generated by adding to each candidate reactor model predictions a normally distributed Gaussian noise with zero mean and known constant relative variance to mimic the errors in concentration measurements affecting the physical experimentation.

Results show that an ANN accuracy of 100% (i.e. perfect classification) can be achieved for a quinary classification involving CSTR, 2, 5, 10 CSTRs-in-series and PFR in absence of measurements noise. With a measurement noise of 0.1%, however, Fig. 1 shows that the ANN overall performance accuracy decreases to 71% even using an optimal DoE designed by DEA. While there is no confusion with CSTR, some data from PFR, 2, 5 and 10 CSTRs-in-series were misclassified. As illustrated in Figure 2, the reactant's concentration (of 2,4-difluoronitrobenzene) in nucleophilic aromatic substitution decreases with the number of CSTRs-in-series, reaching a minimum in a PFR and a maximum in a single CSTR (Fig. 2). Therefore, adjacent reactor model classes can be confused when their predictions are corrupted by measurement noise. Future work will validate the developed ANN-DEA method using real experimental data generated by an autonomous reaction platform [6].

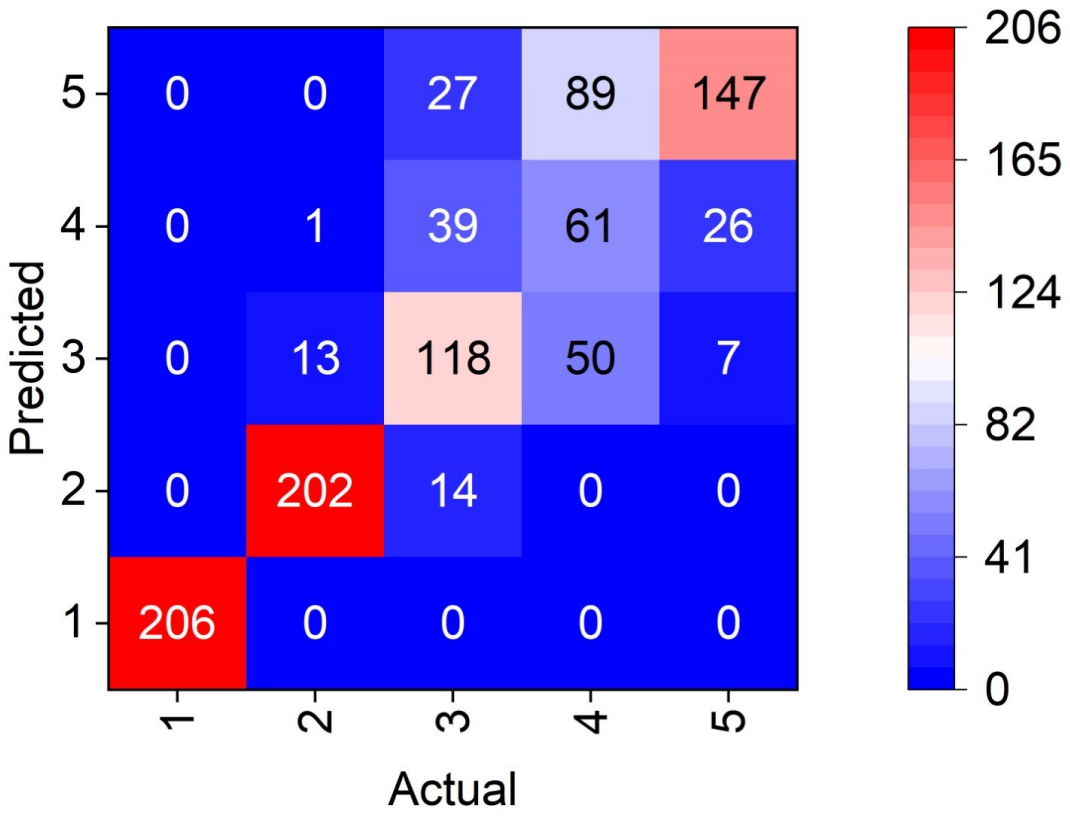


Fig. 1: Confusion matrix for quinary classification of CSTR (1), 2 CSTRs (2), 5 CSTRs (3), 10 CSTRs (4) and PFR (5) at 0.1% noise

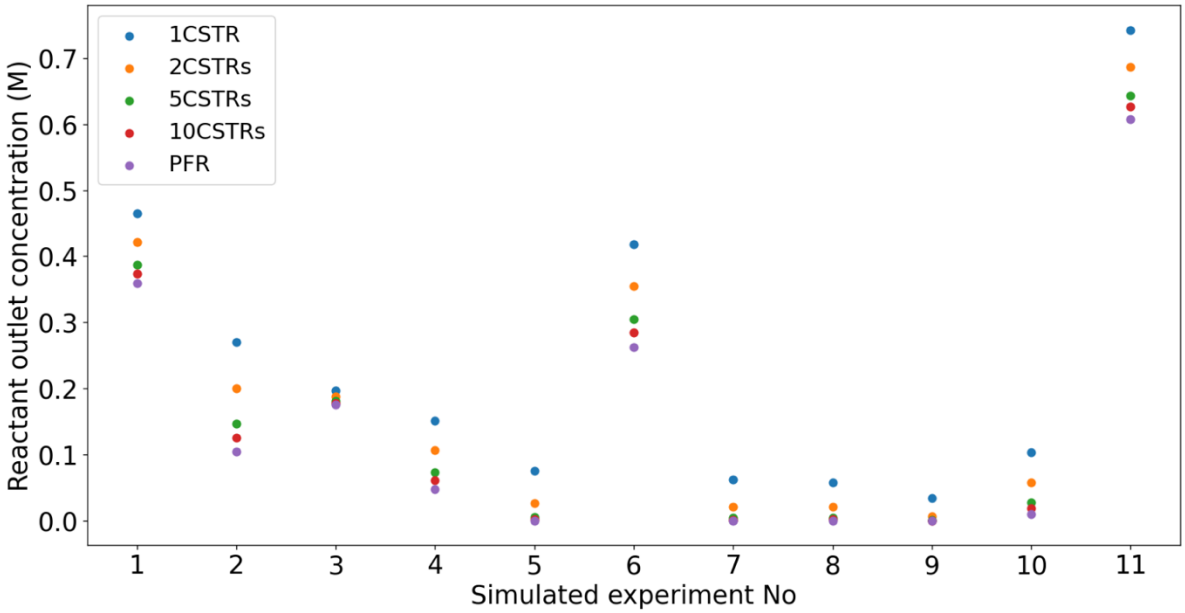


Fig. 2: Reactor model predictions for reactant (2,4-difluoronitrobenzene) outlet concentration in nucleophilic aromatic substitution.

References

1. Machado Cavalcanti, F., Emilia Kozonoe, C., André Pacheco, K., Maria de Brito Alves, R., 2021. Application of Artificial Neural Networks to Chemical and Process Engineering [Internet]. Deep Learning Applications. IntechOpen, DOI: 10.5772/intechopen.96641
2. Quaglio, M., Roberts, L.R., Jaapar, M.S., Fraga, E.S., Dua, V., & Galvanin, F. 2020. An artificial neural network approach to recognise kinetic models from experimental data. *Comput. Chem. Eng., 135, 106759.* DOI: 10.1016/j.compchemeng.2020.106759.
3. Sangoi, E., Quaglio, M., Bezzo, F., Galvanin, F., 2022. Optimal Design of Experiments Based on Artificial Neural Network Classifiers for Fast Kinetic Model Recognition, in: Yamashita, Y., Kano, M. (Eds.), Computer Aided Chemical Engineering, 14 International Symposium on Process Systems Engineering. Elsevier, DOI: 10.1016/B978-0-323-85159-6.50136-6
4. Fogler, H.S., 2004. Elements of Chemical Reactions Engineering, 3rd Edition, Pearson Education, Inc. ISBN-81-203-2234-7
5. Hone, C.A., Boyd, A., O'Kearney-McMullan, A., Bourne, R.A., and Muller, F.L., 2019. Definitive screening designs for multistep kinetic models in flow, *React. Chem. Eng.*, DOI: 10.1039/C9RE00180H
6. Agunloye, E., Petsagkourakis, P., Yusuf, M., Labes, R., Chamberlain, T., Muller, F.L., Bourne, R.A., and Galvanin, F., 2024. Automated kinetic model identification via cloud services using model-based design of experiments, *React. Chem. Eng.*, DOI: 10.1039/D4RE00047A

Accelerating Liquid Formulations Design using Lab Automation and Machine Learning

Aniket Chitre[1,2,3], Kedar Hippalgaonkar[3], Alexei A. Lapkin*[1,2]

* Corresponding author
[1] Department of Chemical Engineering and Biotechnology, University of Cambridge, Philippa Fawcett Drive, Cambridge CB3 0AS, United Kingdom.
[2] Cambridge Centre for Advanced Research and Education in Singapore, CARES Ltd. 1 CREATE Way, CREATE Tower #05-05, Singapore 138602, Singapore.
[3] Institute of Materials Research and Engineering, Agency for Science, Technology and Research (A*STAR), Singapore 138634, Singapore.

Keywords: Liquid formulations, machine learning, high-throughput/automated experiments.

Abstract:

Liquid formulations are ubiquitous yet have lengthy product development cycles owing to the complex physical interactions between ingredients, making it challenging to tune formulations to customer-defined property targets. We seek methods to accelerate liquid formulation design to address changing customer preferences, supply chain/regulatory pressures, and a drive to develop more sustainable products. This work focuses on using lab automation to develop a high-throughput liquid formulations workflow and machine learning (ML) to build property prediction models for a system of shampoo formulations. The methods developed here are generalisable to other surfactant-based products in the personal care industry.

ML and optimisation have expedited a broad range of product and process development but are critically dependent on the volume and quality of data available. We focused on developing a high-throughput formulation workflow comprised of modular unit operations. Formulation development involves working with viscous materials and often challenging and labourious processing or characterisation steps, e.g., pH adjustment or rheology measurement. Within this work, we present (i) an automated viscous liquid handling protocol using a retrofitted Opentrons OT-2 robot, (ii) a self-driven pHbot for automated titration of viscous liquid formulations, (iii) computer vision for stability prediction, and (iv) a proxy viscometer for Newtonian fluids. We used the developed workflow to collect a dataset of over 800 liquid formulations with phase stability, turbidity, and viscosity measurements.

We prepared formulations with a binary mixture of surfactants, a conditioning polymer, and a thickener, selecting from eighteen industrial ingredients. Formulation design typically results in a high-dimensional, mixed discrete-continuous design problem for which there was no suitable design of experiments (DoE). We developed a weighted space-filling design using Maximum Projection Designs with Quantitative and Qualitative Factors (MaxProQQ). The weighting was from a phase stability classifier trained within an active learning cycle for difficult-to-formulate (unstable) formulation sub-systems to guide them to regions of stability.

Finally, we used the generated dataset to develop phase stability, turbidity, and viscosity models. Previous work developed these models based only on the concentration of ingredients, which would not extrapolate to new ingredients. Therefore, we introduced a featurisation based on the surfactant functional groups as an initial step towards generalisation. We additionally explored a set of surfactant molecular descriptors selected based on our domain knowledge to improve the quality of the viscosity model. However, we concluded that system-level descriptors are required instead.

Poster Sessions:

Name	Poster Title
Layla Hosseini-Gerami	SAFEPATH: Using AI to understand the molecular mechanisms causing safety failures, enabling drug optimisation and turnaround
Jana Mousa	Enhancing Neural Network Predictions in Chemical Engineering Unit Operation Modeling through Physical Constraint Integration for Improved Accuracy
Jinwen Cui	Model-based Design of Transient Flow Experiments for The Identification of Kinetic Models in the Presence of Catalyst Deactivation
Zheqi Jin	Generative Machine Learning for Automating Structure Elucidation in Synthesis
Daniela Kalafatović	Can we accelerate peptide materials discovery using machine learning?
Dingyun Huang	A database of thermally activated delayed fluorescent molecules auto-generated from scientific literature
Florian Dietrich	Machine Learning Order Parameters in Atomistic Systems
Yao Wei	Comparison of Protein Design Tools For Engineering Non-heme Iron Dependent Dioxygenases
Frixos Papadopoulos	Simpler bag-of-words histograms are competitive with sum-of-word2vec representations across multiple protein inference problems

Organisation Committee

Miruna Cretu
University of Cambridge - Syntech CDT

Francesco Ceccarelli
University of Cambridge - Syntech CDT

Ruben Sharma
University of Cambridge - Syntech CDT

Kerstin Enright
University of Cambridge – Innovation Centre on Digital Molecular Technologies (iDMT)

Prof. Alexei Lapkin
University of Cambridge – Department of Chemical Engineering and Biotechnology

Contact: mab@ceb.ac.uk

YOUR FEEDBACK: SCORE THE PRESENTATIONS OF DAYS 1 & 2

YOUR FEEDBACK: SCORE THE POSTERS

The sponsors and exhibitors

Digital Discovery is an open access journal that publishes both theoretical and experimental research at the intersection of chemistry, materials science and biotechnology. We focus on the development and application of machine learning, AI and automation tools to unravel scientific problems, and we put data first to ensure reproducibility and faster progress for everyone.

Reaction Chemistry & Engineering is an interdisciplinary journal reporting cutting-edge research focused on enhancing the understanding and efficiency of reactions. Reaction engineering leverages the interface where fundamental molecular chemistry meets chemical engineering and technology. Challenges in chemistry can be overcome by the application of new technologies, while engineers may find improved solutions for process development from the latest developments in reaction chemistry. Reaction Chemistry & Engineering is a unique forum for researchers whose interests span the broad areas of chemical engineering and chemical sciences to come together in solving problems of importance to wider society.

AstraZeneca is a global, science-led biopharmaceutical company that focuses on the discovery, development, and commercialisation of prescription medicines. Some of their primary pipelines are for the treatment of diseases in three therapy areas – Oncology, Cardiovascular, Renal & Metabolism, and Respiratory & Immunology. AstraZeneca operates in over 100 countries and its innovative medicines are used by millions of patients worldwide.

Vapourtec, the world's leading manufacturer of flow chemistry equipment was founded in 2003 by Duncan Guthrie. Since then, Vapourtec has been at the forefront of the flow chemistry industry ever since. Headquartered in Bury St Edmunds, UK, Vapourtec design and manufacture the R-Series and E-Series flow chemistry systems that have empowered chemists throughout the world to further scientific discovery. Trusted by academics, chemists, and manufacturers around the world, the modular R-Series system has revolutionised the way many deliver the research, chemicals, and products we all rely on. With an installation base of more than 600 systems, resulting in over 970 citations in peer-reviewed publications, they continue to support their customers across the globe with the world-class products and services with which Vapourtec has become synonymous.