

Proceedings of the 3rd

**Foundations of
Process/Product Analytics
and Machine Learning**

FOPAM 2026



**Leo Chiang
AJ Medford
Jean Tom**

Proceedings
of the
**3rd Foundations of Process/Product
Analytics and Machine Learning**
(FOPAM 2026)

edited by
Leo Chiang (Lubrizol)
AJ Medford (Georgia Tech)
Jean Tom (Princeton University)



PSEcommunity.org

PSE Press • Hamilton • Notre Dame

Proceedings of the 3rd Foundations of Process/Product Analytics and Machine Learning (FOPAM 2026)
PSE Press
Hamilton, Ontario, Canada

Copyright © 2026 by the authors and licensed to PSEcommunity.org and PSE Press.
Remaining text © 2026 by the PSE Press. All rights reserved.
Photo credits: Chemical Factory © Salvador Ceja | Dreamstime.com

This book is released under the Creative Commons CC-BY-SA 4.0 License
Credit must be given to creator and adaptations must be shared under the same terms.
See <https://creativecommons.org/licenses/by-sa/4.0/> for license details.



Living Archive for Process Systems Engineering (LAPSE) Archive ID: LAPSE:2026.1200
Digital Object Identifier (DOI): 10.69997/pse.149305

Library and Archives Canada Catalogue

A catalogue record and copy of this book is available from Library and Archives Canada

Archival Permissions

This book as well as individual articles may be deposited in digital research archives, personal archives, or institutional archives.

Electronic Book
ISBN: 978-1-7779403-6-2

10 9 8 7 6 5 4 3 2 1

The information presented in this book was provided by individual contributors in the form of new historical or scientific research contributions, data, and findings that were believed to be reliable and provided in good faith. Neither the PSE Press, PSEcommunity.org, the CACHE Foundation, Inc., the editors, nor authors guarantee the accuracy or completeness of any information provided, nor shall they be responsible for errors, omissions, incomplete information, or damages arising from its use. The information within does not consist of engineering or professional services.

Typeset in Playfair Display

Suggested Citation for Individual Articles

Lastname MC, Otherauthor QR. Title of the article. In: *Proceedings of the 3rd Foundations of Process/Product Analytics and Machine Learning (FOPAM 2026)* pp. startpage-endpage. PSE Press (2026). <https://doi.org/10.69997/pse.restofarticledoi>

Suggested Citation for Entire Book

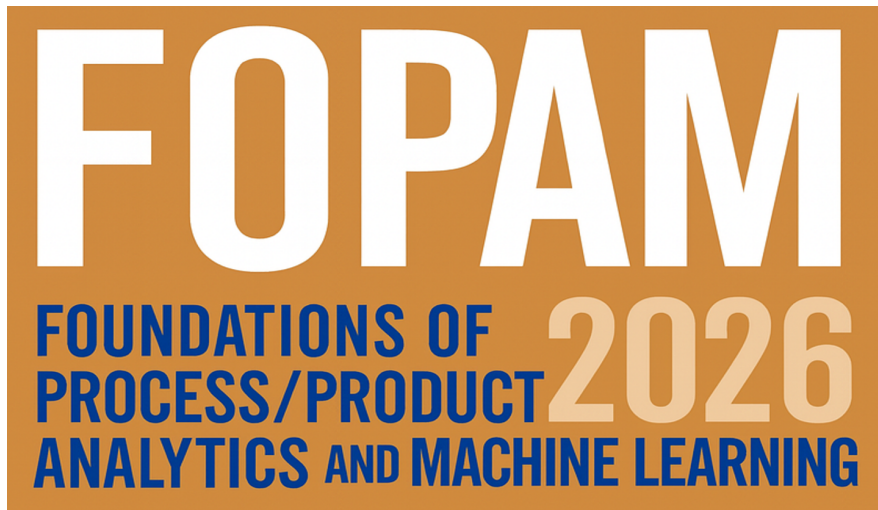
Chiang L, Medford AJ, Tom J (ed). *Proceedings of the 3rd Foundations of Process/Product Analytics and Machine Learning (FOPAM 2026)*, Decatur, USA, July 26-30 (2026). ISBN: 978-1-7779403-6-2

Scientific Editors

Leo Chiang (Lubrizol), AJ Medford (Georgia Tech), Jean Tom (Princeton University)

PSE Press Production Team

Thomas A. Adams II (NTNU)



////////////////////////////////////

FOPAM is intended to be the premier forum for researchers from industry and academia to discuss current status and future directions of data analytics and machine learning in the process industries.

////////////////////////////////////

Organizing Committee

Meeting Program Chairs

Leo Chiang

AJ Medford

Jean Tom

Lubrizol

Georgia Tech

Princeton

Programming Committee

Tom Badgwell

Richard Braatz

Bea Braun

Paulette Clancy

Benjamin Comer

Selen Cremaschi

Rolf Findeisen

Michael Lovette

Ahmet Palazoglu

Rohit Patwardhan

Jose Tabora

Phil Westmoreland

UT Austin

MIT

Dow

Johns Hopkins University

Shell

Auburn University

TU Darmstadt

Merck

University of California at Davis

Saudi Aramco

Bristol Myers Squibb

North Carolina State University

Sponsors

The CAChe Foundation, Inc.

AICHe Computing and Systems Technology Division (CAST)

AICHe Computational Molecular Science & Engineering Forum (CoMSEF)

Shell Corporation

Georgia Institute of Technology (ChBE, GTMI, SEI, IDEaS)

Lubrizol Corporation

Merck & Co. Foundation

Auburn University, Department of Chemical Engineering

Princeton University, Department of Chemical & Biological Engineering

Table of Contents

PREFACE

Organizing Committee	ii
----------------------------	----

ABSTRACTS

Oral Presentations

The Enterprise AI Revolution: How AI Technologies are Unlocking Value Across the Med Tech Value Chain	1
Brenda Remy	
Ensuring GenAI Works for Chemical Process Systems: Perspectives on Use Cases and Alignment -	2
Andrew Allman	
Learning Process Models When Data Are Scarce: Transferable Knowledge for Process Monitoring and Optimization	3
Manabu Kano	
From Insight to Action: AI-Powered Decisions in the Chemical Industry	4
Ivan Castillo	
Grounded Multi-Agent Systems for Decision Support in Industrial Operations	5
Samyakh Tukra	
Machine Learning the Excited State Properties of Crystalline Organic Semiconductors	6
Noa Marom	
Large Scale Datasets and Machine Learning for Direct Air Capture: The Open DAC Project and Beyond	8
Andrew J. Medford	
A Decade of Digitizing Pharmaceutical Manufacturing at JnJ: Lessons Learned and Future Directions	9
Olav Lyngberg	
Beyond Data Science: Driving Industrial Value through Data-Driven Decisions	10
Zhenyu Wang	
Data-driven optimization: efficient adaptive learning for self-driving laboratories	11
Nick Sahinidis	
Local-Global Learning of Interpretable Control Policies: The Interface between MPC and Reinforcement Learning	12
Ali Mesbah	
Designing the Future Engineer: How AI Is Transforming Learning, Work, and Discovery	13
John Kitchen	
Expanding the Science-Guided Machine Learning Applications for Process Industries: Advances, Education, and Workforce Development	14
Y. A. Liu	

Poster Presentations

Leveraging Machine Learning for Multi-Level Optimization In Energy-Water Nexus Systems	15
Elizabeth Abraham	
Digital AI-Driven Methodologies to Support and Accelerate Mabs Development In the Biopharmaceutical Industry	17
Gianmarco Barberi	
From Mechanistic Model to Digital Twin: A Framework for Real-Time Optimization of Ethanol Production In S.Cerevisiae.....	19
Omar Bayomie	
Data Driven Experimental Design of Cellulose and Chitin-Based Sustainable Barrier Films	20
Jessica Bonsu	
A Novel Uncertainty-Aware Computer Vision Framework for Automated Process Optimization In Additive Manufacturing	21
Ronald Borja-Roman	
Multivariate PAT Monitoring of Sodium Phosphate Solubility and Crystallization In Alkaline Media	23
Viviana Cardenas Ocampo	
DEM-CFD Modeling of a Packed Bed Reactor: Analysis of Local Transport & Deactivation Dynamics across Aspect Ratios	24
Raj Chapagain	
How Archimedis Operational Reasoning System Detected a Hidden Furnace Failure In Under an Hour	26
Charles Crowell	
Stitching Misoriented and Misaligned Non-Overlapping Images	27
Michael Fokuo	
Role of Multivariate Data Analyses In Formulation and Process Development of Oral Solid Drug Products: Encapsulation Case Studies	29
Shashwat Gupta	
Science Guided Machine Learning for Conceptual Process Development: A Novel Heterogeneous Azeotropic Separation Case Study	30
Troy Gustke	
How to Teach Programming to Chemists In the Age of AI	31
Robert Hesketh	
A Streamlit-Based Platform for Ternary Solvent Solubility Modeling and Crystallization Process Design	33
Marko Ivancevic	
Orchestrating Modelling & Simulation of Pharmaceutical Production Processes Via Large Language Models.....	34
Christoph Kloss	
Why Transfer Learning Fails Under Target Non-Identifiability.....	36
Yuki Kobayashi	
Domain-Decomposition Pinns for Rapid Prediction of Stirred-Tank Mixing Flows across Geometric Scales.....	37
Yohei Kono	

Efficient Parameter Estimation In Agent-Based Models of Collective Cell Invasion Via Gaussian Process Surrogates and Bayesian Optimization.....	38
Aneesh Krishna	
Empowering Automated Process Analysis through LLM-Based Literature Mining, Flowsheet Digitization, and Simulation.....	40
Jan-Frederic Laub	
From Disparate Data to Accelerating Innovation: A Practical Framework for R&D Digitalization.....	42
Zifeng Li	
Tennet-SAC: A Physics-Embedded Machine Learning Model for Activity Coefficient of Multicomponent Liquid Mixtures	43
Shiang-Tai Lin	
Real-Time Inline Nitric Acid Quantification In Purex Systems Using Raman, ATR-FTIR, and Machine Learning	44
Nischal Maharjan	
Machine Learning-Based Prediction of Heavy Metal Exposure In Spatially Heterogeneous Urban Environments.....	45
Paromita Nath	
Renewable-Driven Microgrid Design, Planning, and Operation of Integrated Gasification Fuel Cell for Biomass Upgradation to Biofuels	46
Oluwatimileyin Ogunsola	
Optimization of Biogas Steam Reforming Toward Low Carbon Hydrogen Production Using Integrated Artificial Neural Network and Genetic Algorithm	48
Ikechukwu Okwuosa	
Sketch2Simulation: Automating Flowsheet Generation Via Multi-Agent Large Language Models	49
Emma Pajak	
Safety System Complexity: Ontology-Grounded Llms That Cross-Link Plant Records to Reduce Spurious Trips and Surface Hidden Process Risk	51
David Parham	
Reinforcement Learning for Nonlinear Optimization In Process Industry.....	52
Kalpesh Patel	
Topology-Guided Response Surface Characterization for ML Model Selection	53
Shenbageshwaran Rajendiran	
Identifiability of Microkinetic Parameters from Multimodal Operando Data.....	54
Gabriel SabenÃ§a GusmÃ£o	
A Machine Learning Framework for Short Peptide Sequence Optimization	56
Anh Trinh	
Generalized Physics-Informed Deep Learning Framework for Chemical Process Modeling	57
Harshit Verma	
Optimal Solvent Mixture Screening with Graph Neural Networks	58
Yipei Zhao	

A Deepsets-Guided Framework for Learning Job Priorities In Single-Machine Scheduling.....	59
Daniel Zhu	

The Enterprise AI Revolution: How AI Technologies are Unlocking Value Across the Med Tech Value Chain

Brenda Remy

General Electric Healthcare

ABSTRACT

The med-tech value chain is being re-shaped by enterprise-grade AI technologies that deliver measurable gains while meeting the bar for safety, quality, and compliance. In this presentation, I'll share how GE HealthCare applies machine learning and GenAI—end to end—to improve manufacturing performance, reduce cost and improve customer service. We'll walk through end-to-end use cases including computer-vision applications to reduce scrap and variation; touchless demand and revenue forecasting that has improved accuracy with corresponding gains in on-time delivery and working-capital reductions; scheduling and logistics optimization; and GenAI-assisted quality applications. Beyond the algorithms, I'll

highlight the socio-technical pieces that make these systems stick: robust MLOps, human-in-the-loop workflows, and alignment to regulatory and safety expectations. The session is designed for a diverse audience—researchers, practitioners, and leaders across industries—offering concrete examples of AI applications across the med-tech value chain, with measurable outcomes when scaling AI across the enterprise.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Ensuring GenAI Works for Chemical Process Systems: Perspectives on Use Cases and Alignment

Andrew Allman

University of Michigan, Department of Chemical Engineering

ABSTRACT

Recent years have seen the rapid proliferation of generative AI tools, from chatbots to autonomous “agents.” These tools have transformative potential in all parts of society, including the chemical process industry. However, deploying these tools to their full effect will require careful consideration of the areas where generative AI can enhance, rather than replace existing chemical engineering knowledge, and how it can be deployed in a safe manner given the current limitations in generative AI technology. The first half of this talk reviews some promising use cases of generative AI in process systems engineering emphasizing areas where first principles knowledge is insufficient for the task. In particular, efforts utilizing AI for generating new decompositions of large-scale optimization problems, explaining optimization outcomes, and incorporating hard-to-quantify sustainability objectives into decision making frameworks are discussed. In the second half of the talk, I transition to considering the alignment of generative AI models, an

essential quality of AI systems to ensure they support real industry objectives and are deployed in a safe manner. I note that the process systems community is particularly well suited for addressing these issues given its expertise in constrained optimization, model predictive control, multi-objective optimization, and dynamic systems. The need to move from soft-constrained to hard-constrained AI models to ensure alignment is emphasized, and efforts in the chemical process systems community towards constrained AI are highlighted, including new methods for constrained network structure detection and neural network training.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Learning Process Models When Data Are Scarce: Transferable Knowledge for Process Monitoring and Optimization

Manabu Kano

Kyoto University, Department of Informatics

ABSTRACT

Industrial process data analytics is fundamentally constrained by data scarcity. Process data are strongly dependent on operating regimes, product grades, control policies, equipment, and scale. These characteristics make it difficult to build reliable models for new plants, new products, abnormal conditions, or future operating regimes. This talk discusses emerging approaches for learning process models under such data-scarce conditions. Relevant sources of transferable knowledge include mechanistic models, laboratory experiments, previous products and plants, and pretrained time-series representations. I will discuss how such knowledge can be exploited

through grey-box or hybrid modeling, physics-informed machine learning, transfer learning, and time-series foundation models. The focus will be on these methodologies and their applications to industrial process monitoring and optimization. Particular attention will be paid to source selection and negative transfer.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



From Insight to Action: AI-Powered Decisions in the Chemical Industry

Ivan Castillo

Dow

ABSTRACT

Decision-making in the chemical industry is difficult because operators must choose actions under uncertainty: plant behavior is nonlinear and time-varying, measurements are noisy and incomplete, operating constraints and safety margins are strict, and economic objectives often conflict with each other. Data-driven AI methods can help turn high-frequency process data into predictions and recommendations, but in highstakes settings they are most useful when they leverage prior domain knowledge rather than treating the plant as a black box. In this work, prior knowledge is incorporated in two general ways. First, we use hybrid modeling, where deterministic structure from first principles (e.g., balances, thermodynamics, kinetics) is combined with data-driven learning to improve extrapolation, robustness, and interpretability. Second, we use preference-based learning, where expert judgment is captured qualitatively through pairwise comparisons or rankings of candidate options, enabling optimization even when a single numerical objective is hard to define. We illustrate these ideas with industrial decision-making examples that span physics-grounded monitoring of gradual degradation and preference-driven tuning of operational and control choices. In the hybrid modeling examples, mechanistic knowledge such as mass and energy balances and reaction kinetics is embedded alongside data-driven models to yield physically consistent predictions that remain reliable even when data are sparse

or subject to drift. This work focuses on industrial systems that undergo gradual deterioration over time. Three representative case studies are considered: catalyst deactivation monitoring in continuous reactors, where an adiabatic reactor model is used to compute catalyst activity as a health indicator that is integrated into an empirical lifetime model; heat exchanger fouling forecasting in an ethylene oxide plant, where physics-informed features enable month-ahead prediction of fouling surrogates to support cleaning decisions; and remaining useful life prediction for pollutant scrubbers, where hybrid models capture slow degradation trends to inform maintenance planning. Together, these examples show that combining first-principles understanding with datadriven methods enables accurate characterization of degradation dynamics using limited data, improves extrapolation to unseen operating conditions, and increases confidence in maintenance-related decisions (Sansana et al., 2024; Venegas et al., 2024; Bui et al., 2022).

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Grounded Multi-Agent Systems for Decision Support in Industrial Operations

Samyakh Tukra

Applied Computing

ABSTRACT

Industrial operations need AI systems that can reason across live process data, engineering knowledge, and operator workflows. Yet conventional machine learning models often remain narrow predictors, while large language models lack grounding in plant behaviour, constraints, and real-time operating context. This talk presents Orbital, a grounded multi-agent system for decision support in industrial operations.

Orbital combines three complementary layers: a time-series model for multivariable process dynamics and uncertainty-aware forecasting; a constraint-learning layer that extracts engineering relationships from plant documentation, including P&IDs, datasheets, mass and energy balances, and operating manuals; and a language-fusion layer that aligns process behaviour with engineering descriptions. These components are coordinated through specialist

agents for planning, tool execution, verification, memory, and response composition.

The system moves beyond prediction toward interpretable decision support: detecting abnormal behaviour, retrieving relevant historical events, explaining likely root causes, and grounding recommendations in both data and engineering constraints. More broadly, this work argues that the next generation of industrial AI must be grounded, multi-modal, and operationally trustworthy; connecting data, domain knowledge, and human decision-making in high-consequence environments.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Machine Learning the Excited State Properties of Crystalline Organic Semiconductors

Noa Marom

Carnegie Mellon University, Department of Materials Science and Engineering

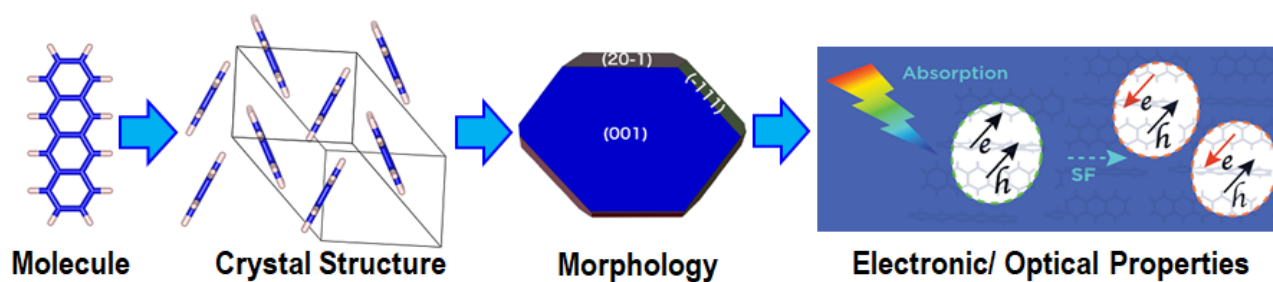
ABSTRACT

Molecular crystals are bound by dispersion (van der Waals) interactions, whose weak nature gives rise to polymorphism, the ability of a compound to crystallize in different structures. Crystal structure profoundly influences the physical and chemical properties, and hence the functionality of molecular solids in applications including pharmaceuticals, electronic devices, and energetic materials. Therefore, the ability to predict the structure and properties of molecular crystals is of paramount importance. To this end, we combine first principles simulations with machine learning.

Molecular crystal structure prediction (CSP) is challenging because it requires searching a high-dimensional configuration space with high accuracy. CSP workflow have two main components, structure generation and stability ranking. For structure generation, we develop the Genarris code [1], which generates random structures in all compatible space groups with physical constraints on intermolecular distances. Machine learned interatomic potentials (MLIPs) are trained on large data sets of first principles

simulations [2], typically density functional theory (DFT) to achieve DFT-level accuracy at the computational cost of classical force fields. We have interfaced Genarris with several types of MLIPs for geometry optimization and stability ranking [1,3,4]. We have shown that MLIPs can completely replace both early-stage screening with classical force fields and final ranking with DFT, paving the way to high-throughput CSP [4].

One of the optoelectronic applications of molecular crystals is singlet fission (SF), the conversion of one photogenerated singlet exciton into two triplet excitons. SF has the potential to increase the efficiency of solar cells by harvesting two charge carriers from one high-energy photon, whose excess energy would otherwise be lost to heat. The realization of SF-based solar cells is hindered by the dearth of suitable materials. The excited-state properties of molecular crystals can be calculated using many-body perturbation theory (MBPT) within in the GW approximation and the Bethe–Salpeter equation (BSE) [5]. The computational cost of GW+BSE is prohibitive for large-scale exploration of the chemical space, and also for generating



large amounts of training data. This calls for ML approaches that work well with small data.

REFERENCES

1. Yang Y, Liu Z, Zheng F, Zhang P, He H, Jha A, Duan HG. Diverse transient chiral dynamics in evolutionary distinct photosynthetic reaction centers. *J. Chem. Theory Comput.* 21:321–332 (2025). <https://doi.org/10.1021/acs.jctc.4c01469>
2. Gharakhanyan V, Barroso-Luque L, Yang Y, Shuaibi M, Michel K, Levine DS, Dzamba M, Fu X, Gao M, Liu X, Ni H, Noori K, Wood BM, Uyttendaele M, Boromand A, Zitnick CL, Marom N, Ulissi ZW, Sriram A. *Scientific Data* 13:354 (2026). <https://doi.org/10.1038/s41597-026-06628-2>
3. Noyal KS, O'Connor D, Zubatyuk R, Anstine DM, Yang Y, Tom R, Deng W, Tang K, Marom N, Isayev O. Efficient Molecular Crystal Structure Prediction and Stability Assessment with AIMNet2 Neural Network Potentials. *Crystal Growth & Design* 25:9092–9106 (2025). <https://doi.org/10.1021/acs.cgd.5c01001>
4. Gharakhanyan V, Yang Y, Barroso-Luque L, Shuaibi M, Levine DS, Michel K, Bernat V, Dzamba M, Fu X, Gao M, Liu X, Noori K, Purvis LJ, Rao T, Wood BM, Rizvi A, Uyttendaele M, Ouderkirk AJ, Daraio C, Zitnick CL, Boromand A, Marom N, Ulissi ZW, Sriram A. FastCSP: Accelerated Molecular Crystal Structure Prediction with Universal Model for Atoms. *arXiv* 2508.02641 (2025)
5. Wang X, Gao S, Luo Y, Liu X, Tom R, Zhao K, Chang V, Marom N. Computational Discovery of Intermolecular Singlet Fission Materials Using Many-Body Perturbation Theory. *J. Phys. Chem. C* 128:7841–7864 (2024). <https://doi.org/10.1021/acs.jpcc.4c01340>
6. Gao S, Liu X, Luo Y, Wang X, Zhao K, Chang V, Schatschneider B, Marom N. PAH101: A GW+BSE Dataset of 101 Polycyclic Aromatic Hydrocarbon (PAH) Molecular Crystals. *Scientific Data* 12:679 (2025). <https://doi.org/10.1038/s41597-025-04959-0>
7. Liu X, Wang X, Gao S, Chang V, Tom R, Yu M, Ghiringhelli LM, Marom N. Finding Predictive Models for Singlet Fission by Machine Learning. *npj Comput. Mater.*, 8, 70 (2022). <https://doi.org/10.1038/s41524-022-00758-y>
8. Gao S, Luo Y, Liu X, Marom N. Predicting the excited-state properties of crystalline organic semiconductors using GW+BSE and machine learning. *Digital Discovery* 4, 1306–1322 (2025). <https://doi.org/10.1039/d4dd00396a>

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Large Scale Datasets and Machine Learning for Direct Air Capture: The Open DAC Project and Beyond

Andrew J. Medford

Georgia Institute of Technology

ABSTRACT

Direct air capture (DAC) with porous adsorbents has the potential to aid large-scale decarbonization, but identifying useful sorbents for capturing CO₂ from humid air remains a formidable challenge given the vast chemical space of candidate materials such as metal–organic frameworks (MOFs). This talk surveys how AI and machine learning, powered by large, high-fidelity computational datasets, are reshaping the discovery pipeline for DAC sorbents using the Open DAC (ODAC) project as a central example. The earlier Open DAC 2023 (ODAC23) dataset established the approach with roughly 38 million density functional theory (DFT) calculations of CO₂ and H₂O adsorption across more than 8,000 MOFs. The new Open DAC 2025 (ODAC25) dataset comprises nearly 60 million DFT single-point calculations for CO₂, H₂O, N₂, and O₂ adsorption in more than 15,000 sorbent structures, introducing chemical and configurational diversity through functionalized MOFs, high-energy GCMC-

derived placements, and synthetically generated frameworks. ODAC25 substantially improves both the accuracy of DFT calculations and the treatment of flexible MOFs relative to ODAC23. Alongside the datasets, we release state-of-the-art machine learning interatomic potentials, evaluate them on adsorption energy and Henry's law coefficient predictions, and show how these models can be integrated into classical molecular simulation workflows to move toward economical and scalable DAC processes. The talk will cover these developments and briefly outline challenges and opportunities in the path from atomistic predictions to process-relevant sorbent screening.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



A Decade of Digitizing Pharmaceutical Manufacturing at JnJ: Lessons Learned and Future Directions

Olav Lyngberg

Janssen Pharmaceuticals

ABSTRACT

In the last decade, the adoption of sophisticated modeling, machine learning, and artificial intelligence, combined with process analytical technologies like Raman and NIR, has become commonplace in Johnson & Johnson's manufacturing landscape. This transformation has driven groundbreaking achievements once considered pure science fiction—such as real-time release of oral solid dosage forms, central & automated process control, and remote fault detection that surpasses human capabilities, demonstrating tangible value and opening new horizons for innovation.

This presentation will reflect on our manufacturing & supply chain journey, sharing key lessons learned and focusing on next steps. As we relentlessly pursue new cures, these technological innovations

form a vital part of a comprehensive system dedicated to ensuring the high-quality, reliable supply of complex medicines. With the advent of Cell and Gene Therapies, which introduce substantial patient-specific variability, existing ML and AI tools are being challenged—reminding us that our ongoing mission remains centered on delivering safe, effective, and affordable medicines to patients worldwide, with confidence in outcomes and supply.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Beyond Data Science: Driving Industrial Value through Data-Driven Decisions

Zhenyu Wang

Lubrizol, Decision Science

ABSTRACT

Artificial intelligence is increasingly delivering measurable impact in the chemical industry, moving from isolated pilots toward embedded analytics across operations, supply chain, and Research & Development. At the same time, a clear industry shift is emerging: while developing accurate models remains important, there is a growing emphasis on realizing sustained value from AI investments. This talk presents a practitioner's perspective on applied AI in the chemical industry, focusing on what is working, what is challenging, and where future opportunities lie. We begin with examples of AI applications that have demonstrated value in industrial settings, including process monitoring, demand forecasting, and computer vision-enabled inspection and automation. These successes highlight a critical principle, *i.e.* value is not created by models alone, but by extracting actionable insights from data and enabling data-driven decisions informed by those insights. It is ultimately the decisions that drive the measurable business impact. We then examine key challenges that continue to limit broader adoption. These include inconsistent

data quality, lack of contextualization across systems, and model degradation under changing environment. Additional barriers arise from difficulties in interpreting model outputs, translating insights into decisions, and embedding analytics into real operational workflows. Finally, the talk explores emerging directions in industrial AI, including hybrid modeling that combines physics and data-driven insights, scalable data infrastructures, and tighter integration between analytics, optimization, and automation technologies. The session concludes by outlining a forward-looking path toward more adaptive, integrated, and decision-centric AI systems that can reliably translate data into sustained industrial value at scale.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Data-driven Optimization: Efficient Adaptive Learning for Self-Driving Laboratories

Nick Sahinidis

Georgia Tech, School of Chemical & Biomolecular Engineering

ABSTRACT

Self-driving laboratories promise to compress materials-discovery timelines from years to weeks by replacing trial-and-error experimentation with closed-loop, algorithm-guided campaigns. Yet, despite the rapid proliferation of robotic and automation hardware, today's autonomous labs rely almost exclusively on Bayesian optimization (BO) to decide what experiment to run next. BO is a sensible approach to low-dimensional optimization problems with smooth response surfaces, but it struggles in precisely the regimes that matter most for real materials campaigns: tight experimental budgets, dozens of process parameters, mixed-integer choices, hard physical constraints, and noisy expensive measurements.

In this talk, I will show how moving from BO to partitioning-based algorithms can substantially improve data efficiency, scale gracefully to dozens of process variables, and handle the constraints and noise that characterize realistic experimental campaigns. I will summarize a recently completed large-scale black-box optimization (BBO) benchmark in which we compared 42 solvers across 502 problems ranging from one to 300 dimensions and from smooth and convex to non-smooth and nonconvex. The results overturn several community assumptions: BO solves only about 9% of problems within a 2,500-evaluation budget, while a new branch-and-model (BAM) algorithm reaches an 81% success rate, with GLCCLUSTER, MULTIMIN, MCS, and SNOBFIT also per-

forming strongly. A minimal, irreducible set of eight complementary solvers attains 88% solvability on the full suite.

I will then move from in-silico benchmarks to the wet lab, presenting a recent algorithm-guided experimental campaign on high-performance perovskite solar cells in which a non-BO solver was used to co-optimize six process variables spanning the perovskite, electron-transport, and hole-transport layers. Time permitting, I will also share early results from applying ensembles of BBO algorithms to digital twins of self-driving labs across additional materials systems.

I will close with a forward-looking research vision: accelerating autonomous labs by developing, benchmarking, and experimentally validating data-efficient adaptive algorithms across batteries, semiconductors, catalysts, polymeric membranes, and biomolecules. The benchmarking software will be released as open source, with BAM and most BBO software available free to academic users, so that experimental groups can deploy these tools on their own self-driving platforms.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Local-Global Learning of Interpretable Control Policies: The Interface between MPC and Reinforcement Learning

Ali Mesbah

University of California, Berkeley, Department of Chemical and Biomolecular Engineering

ABSTRACT

Optimal decision-making under uncertainty is a shared challenge across modern chemical, manufacturing, and energy systems that increasingly demand safe, data-driven autonomy. This talk revisits optimal control through the lens of the Bellman equation, emphasizing how optimal control theory and reinforcement learning have developed complementary, yet largely disconnected, perspectives on global optimality. In one view, central to reinforcement learning, the Bellman equation defines a global optimality condition that guides iterative policy learning from interacting with the system, but typically yields opaque control laws that are difficult to interpret, and deploy in safety-critical settings. In another view, widely adopted in model predictive control (MPC), the Bellman equation underpins tractable finite-horizon optimizations that deliver interpretable, constraint-aware, and modular local controllers, yet without explicit guarantees on alignment with global optimality. Building on these ideas, we introduce a local-global paradigm that treats MPC and related optimization-based controllers as structured function approximators designed to approximately satisfy the global

Bellman optimality condition. We discuss algorithmic strategies for learning interpretable local decision makers whose adaptation is guided by Bellman residuals, along with the benefits and practical challenges that arise in terms of stability, constraint satisfaction, and sample efficiency. These concepts are illustrated through case studies that unify reinforcement learning and MPC for safe, high-performance control in complex, uncertain dynamical systems. The talk concludes by outlining open problems and research opportunities in learning interpretable control policies that achieve globally optimal performance while retaining the transparency and reliability required for real-world process control and optimization applications.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Designing the Future Engineer: How AI Is Transforming Learning, Work, and Discovery

John Kitchin

Carnegie Mellon University, Department of Chemical Engineering

ABSTRACT

Artificial intelligence is reshaping what it means to be an engineer and it is changing how we learn, design, and discover. This talk explores the convergence of generative AI, data-driven modeling, and autonomous experimentation, and what that means for engineering education and workforce development. From intelligent tutors and code-generation assistants to self-driving laboratories and agentic research systems, AI is expanding both the cognitive and creative boundaries of the profession. Drawing on examples from open-source educational ecosystems such as pycse, and recent research on agentic science and generative optimization, we will discuss how future engineers can be trained not only to use AI tools, but

to think with them, integrating computation, ethics, and domain expertise into continuous, collaborative learning. We will also discuss challenges of buy-in, resources, resistance from both faculty and students, and the need to maintain a balance of traditional learning approaches with the new opportunities AI offers.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Expanding the Science-Guided Machine Learning Applications for Process Industries: Advances, Education, and Workforce Development

Y. A. Liu

Virginia Tech, Chemical Engineering

ABSTRACT

The chemical engineering (ChE) field is becoming increasingly hybrid digital as new advancements in machine learning (ML) integrate with ChE workflows and industrial plant operations. As data scientists and engineers push artificial intelligence (AI) usage, it is important for the engineering workforce and data science methods to be grounded in ChE fundamentals through methods like science guided machine learning (SGML). SGML is a broad term that includes ML architectures that are informed/embedded with physics, chemistry, and thermodynamics first-principles. This presentation highlights and demonstrates selective SGML advances from 2022 to 2026, including multicomponent phase equilibria, surrogate modeling, uncertainty quantification, and agentic large language models (LLMs). Furthermore, we discuss how each of these advancements accelerates the process design and development workflow, and we demonstrate solving novel process design and separation problems

using this workflow, pushing forward the next era of process design through embedded knowledge and reasoning design tools. Next, we discuss the current state, new developments, and future direction of educational programs. We survey different university approaches to implementing data science in ChE education and give our perspective on how these approaches might change with technological advancements. We also show the evolving landscape of advanced ML for improved workforce training. Finally, we propose our perspective on the future of scalable AI implementation into the ChE workforce.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Leveraging Machine Learning for Multi-Level Optimization In Energy–Water Nexus Systems

Elizabeth Abraham

Texas A&M University, Artie McFerrin Department of Chemical Engineering

ABSTRACT

The energy–water nexus emerged in response to global challenges associated with energy and water resources. Along with their existing residential, commercial, and industrial commitments, the intrinsically complex system now faces additional pressure from the rampant rise in demands from data centers. To factor in the impacts of these new circumstances while accounting for the interdependent and interconnected nature of energy and water supply systems, the nexus holistically manages these resources and their corresponding decisions [1].

However, while these decisions are conventionally modeled from a centralized perspective through representative mathematical programs whose optimal decisions can then be optimized, a more realistic perspective models these decisions sequentially. Here, rather than all involved systems making their decision in a simultaneous fashion, decisions are made one after another in sequential order and characterized using multilevel programming. Bilevel programming, the most well-established multilevel problem, fundamentally consists of an upper-level optimization problem that has a lower-level optimization problem embedded within its constraints. Bilevel programming has been applied to challenges within the energy–water nexus through several traditional solution strategies such as reformulations into a single level, iterative procedures combined with decomposition, and more recently, a more nuanced approach that involves multi-parametric programming [2]. Through this approach, solutions of an optimization problem for all feasible values of uncertain parameters are

determined without iteratively varying parameter values and resolving the problem. The feasible parameter space is demarcated into areas referred to as critical regions, each defined by a specific set of affine functions defining the optimal solution in terms of the uncertain parameter.

The key idea behind the application of this advanced optimization technique to multilevel programs is that decisions of the lower-level problem can be represented in terms of decisions of the upper-level problem or the uncertain parameters in this case [3]. While several works have demonstrated the applicability of this approach to nexus applications and beyond, a major limitation is the complexity that accompanies the scale of multi-parametric problems. The complexity of these problems shows combinatorial growth with the number of uncertain parameters and levels considered, thereby requiring faster approximations that can allow for practical implementations. In this work, seeking inspiration from multi-parametric programming, machine learning strategies will be employed to approximate the behavior of the optimal solution or the decisions of the lower-level problem in terms of the uncertain parameter or the upper-level variables in an energy–water nexus. The approximations determined will then be embedded with the upper-level problem to identify the optimal solution of the bilevel problem. With the increasing number of applications that utilize machine learning, even within nexus studies for the approximation of energy and water system behaviors, it is evident that machine learning strategies provide a unique and innovative

means for approximating the relationship between the upper-level and lower-level decision-makers [4].

REFERENCES

1. Di Martino M, Linke P, Pistikopoulos EN. Overcoming modeling and computational complexity challenges in food-energy-water nexus optimization. *Comput Chem Eng* 195:108902 (2025).
<https://doi.org/10.1016/j.compchemeng.2024.108902>
2. Abraham EJ, Di Martino M, Kenefake D, Al-Mohannadi DM, Pistikopoulos EN. A multi-parametric approach for bi-level optimization within the energy-water nexus. *Comput Chem Eng* 202:109256 (2025).
<https://doi.org/10.1016/j.compchemeng.2025.109256>
3. Avraamidou S, Pistikopoulos EN. Multi-level Mixed-Integer Optimization: Parametric Programming Approach. Walter de Gruyter GmbH & Co KG, 2022.
4. Bagloee SA, Asadi M, Sarvi M, Patriksson M. A hybrid machine-learning and optimization method to solve bi-level problems. *Expert Systems with Applications* 95:142–152 (2018).
<https://doi.org/10.1016/j.eswa.2017.11.039>

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Digital AI-Driven Methodologies to Support and Accelerate Mabs Development In the Biopharmaceutical Industry

Gianmarco Barberi

University of Padova, Department of Industrial Engineering

ABSTRACT

Monoclonal antibodies (mAbs) have become a cornerstone in the treatment of immunological and oncological diseases. Nevertheless, bringing new antibody therapeutics to market remains a costly and time-consuming process, often requiring more than 10 years of development and investments exceeding 2 billion dollars. Critical stages of the development pipeline include the identification of a robust production cell line, capable of ensuring key quality attributes such as productivity, stability, product quality, and production consistency, as well as the optimization of the culture process (Li et al., 2010). These steps typically rely on extensive experimental campaigns and significant resource allocation. As a result, pharmaceutical companies are increasingly investigating digital and AI-driven approaches to streamline development workflows and accelerate drug time-to-market. In this work, we address two key challenges in mAb process development: i) the automated detection of anomalous cell culture experiments and ii) the efficient identification of optimal feeding strategies. First, we developed an assumption-free modeling framework to automatically detect anomalous experimental batches at the Ambr®15 scale and diagnose the underlying causes of abnormal culture behavior (Barberi et al., 2025). This approach supports faster interpretation of experimental campaigns and reduces reliance on manual expert analysis. Second, we introduce a methodology for optimizing glucose and glutamine feeding strategies by partially virtualizing the

experimental campaign through a hybrid semi-parametric model (Barberi et al., 2024). Design of Dynamic Experiments (DoDE) is employed to structure the experimental campaign, enabling the generation of informative data for training the hybrid model and conducting in-silico optimization. Results show that a model trained on only nine experimental batches can identify a feeding strategy that achieves higher antibody titers than DoDE-based campaigns performed with both 9 and 31 experimental batches.

REFERENCES

1. Barberi G, Giacomuzzi C, Facco P. Bioprocess feeding optimization through in silico dynamic experiments and hybrid digital models—a proof of concept. *Frontiers in Chemical Engineering* 6:1456402 (2024).
<https://doi.org/10.3389/fceng.2024.1456402>
2. Barberi G, Diaz-Fernandez P, Lega D, Kotidis P, Finka G, Facco P. A digital tool for the automatic identification of anomalous cell cultures in biopharmaceutical process development. *IFAC-PapersOnLine*, 59:546–551 (2025)
<https://doi.org/10.1016/j.ifacol.2025.08.002>
3. Li F, Vijayasankaran N, Shen A, Kiss R, Amanullah A. Cell culture processes for monoclonal antibody production. *mAbs* 2:466–479 (2010).
<https://doi.org/10.4161/mabs.2.5.12720>



From Mechanistic Model to Digital Twin: A Framework for Real-Time Optimization of Ethanol Production In *S.Cerevisiae*

Omar Bayomie

University College London

ABSTRACT

The transition to smart bioprocessing requires control strategies capable of managing the nonlinear dynamics and limited observability in industrial fermentation. In this work, we developed a fed-batch digital twin process for ethanol production by *Saccharomyces cerevisiae* via combining a mechanistic model with advanced data assimilation to achieve robust Nonlinear Model Predictive Control (NMPC). Recursive Bayesian state estimators have been developed to overcome nonlinearities of the biological models both anaerobic and aerobic, complex metabolic shifts, batch-to-batch parameters variability, and lack of biomass online measurements. Observers (soft sensors) were constructed and benchmarked for

computational tractability and statistical accuracy, including Extended (EKF), Ensemble (EnKF), and Particle Filters (PF). They achieved best overall MSE improvements over the mechanistic model; The closed-loop framework showed a robust performance tested by parametric model-plant mismatches, and heteroscedastic measurement noise.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Data Driven Experimental Design of Cellulose and Chitin-Based Sustainable Barrier Films

Jessica Bonsu

Georgia Institute of Technology, School of Chemical and Biomolecular Engineering

ABSTRACT

The widespread use of non-degradable petroleum-based plastics in food packaging, favored for their cost-effectiveness and strong barrier properties, has led to severe environmental issues. Cellulose and chitin, the most abundant polysaccharides in nature, offer promising sustainable alternatives. Their nanomaterials exhibit high crystallinity and strong hydrogen bonding, which enable excellent mechanical and barrier performance. This makes cellulose- and chitin-based nanomaterials ideal candidates for developing barrier films to substitute petroleum-based plastics. Developing high-performance barrier films requires an understanding of the process–structure–property (PSP) relationships that govern oxygen and moisture transport as well as mechanical stability. This work applies data-driven materials informatics

to accelerate that understanding. A curated dataset of 104 cellulose- and chitin-based barrier films, compiled from both literature and laboratory experiments, was constructed to link processing parameters with barrier and mechanical performance. Using a customized classification algorithm, a reduced design region of processing parameters associated with improved film performance was identified, providing actionable guidance for experimental design and optimization.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



A Novel Uncertainty-Aware Computer Vision Framework for Automated Process Optimization In Additive Manufacturing

Ronald Borja-Roman

Rowan University, Chemical Engineering

ABSTRACT

The thermomechanical response of materials under high-strain deformation is critical to aerospace, defense, automotive, and additive manufacturing (AM) applications [1–3]. However, capturing this behavior remains challenging due to microsecond timescales of high-velocity impact events and reliance on manual, operator-dependent post-processing [4, 5]. These limitations reduce reproducibility and constrain the generation of high-fidelity datasets for constitutive model development. Current approaches are also limited in accessible strain-rate regimes, restricting applicability to AM processes and introducing uncertainty in predictive modeling [2]. This work presents an autonomous, AI-driven framework that transforms raw high-speed impact videos into structured, model-ready material data. The framework integrates two vision foundation models: Grounding DINO for open-vocabulary object detection and the Segment Anything Model (SAM) for high-resolution segmentation, in a zero-shot configuration, eliminating the need for task-specific labeled datasets [6, 7]. Grounding DINO localizes the deforming specimen, guiding SAM for precise contour extraction over time. These time-resolved contours are processed to extract deformation metrics relevant to constitutive modeling, particularly critical for high-strain experiments where data scarcity limits supervised learning approaches. To ensure robustness, an uncertainty quantification (UQ) strategy is embedded within the pipeline. Leveraging prediction confidence at inference time, the framework

autonomously identifies low-reliability regions and triggers iterative refinement, enabling consistent feature extraction across varying experimental conditions. Experimental validation on a lab-scale impact platform demonstrates accurate extraction of deformation dynamics from high-speed image sequences. This work bridges vision foundation models, materials science, and physics-based modeling, delivering datasets for constitutive calibration, digital-twin integration, and AM optimization. Future efforts target real-time implementation, closed-loop calibration, and physics-informed adaptive control for next-generation AM systems.

REFERENCES

1. Alkhatib SE, Sercombe T. High strain-rate response of additively manufactured light metal alloys. *Mater Des* 217:110664, (2022).
<https://doi.org/10.1016/j.matdes.2022.110664>
2. Alaghmandfard R, Seraj P, Sanjari M, Pirgazi H, Dharmendra C, Odeshi AG, Shalchi Amirkhiz B, Mohammadi M. High strain rate deformation behavior, texture and microstructural evolution, characterization of adiabatic shear bands, and constitutive models in electron beam melted Ti-6Al-4V under dynamic compression loadings. *J Mater Res Technol* 21:4093–4114 (2022).
<https://doi.org/10.1016/j.jmrt.2022.11.016>
3. Su X, Wang G, Li J, Rong Y. Dynamic mechanical

response and a constitutive model of Fe-based high temperature alloy at high temperatures and strain rates. *SpringerPlus* 5:504 (2016).

<https://doi.org/10.1186/s40064-016-2169-6>

4. Liu T, Burner AW, Jones TW, Barrows DA. Photogrammetric techniques for aerospace applications. *Prog Aerosp Sci* 54:1-58, (2012).
<https://doi.org/10.3917/phimag.058.0039>
5. P. Wankhede et al., Measurement, 187, 110273, 2022.

<https://doi.org/10.1016/j.paerosci.2012.03.002>

6. S. Liu et al., arXiv:2303.05499, 2024.

7. N. Ravi et al., arXiv:2408.00714, 2024.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Multivariate PAT Monitoring of Sodium Phosphate Solubility and Crystallization In Alkaline Media

Viviana Cardenas Ocampo

Georgia Institute of Technology

ABSTRACT

Crystallization in alkaline process streams is challenging to monitor because solubility, aqueous speciation, hydrate form, and crystal morphology can evolve simultaneously with temperature, and pH. Motivated by phosphate-bearing alkaline waste streams relevant to nuclear waste processing, this work contributes to the development of real-time analytical and modeling tools for crystallization-prone process systems. We develop a data-rich Process Analytical Technology framework that integrates in situ spectroscopy, particle monitoring, temperature and pH measurements with multivariate calibration to quantify phosphate species and track solubility and crystallization behavior in real time. In situ experiments were conducted to characterize sodium phosphate solubility and crystallization under two chemical regimes: strongly alkaline conditions, using 3 molal NaOH, and unadjusted-pH conditions without added NaOH. These conditions provide access to different phosphate speciation regimes. Online ATR-FTIR spectroscopy was used to monitor solution-phase phosphate concentration, while Raman spectroscopy provided complementary information on phosphate speciation and solid-phase formation. Focused Beam Reflectance Measurement was used to track particle chord-length distributions, and EasyViewer imaging provided direct visualization of crystal morphology during heating and cooling. ATR-FTIR spectra were collected during controlled heat-cool experiments for

sodium phosphate systems in water and alkaline media. The 910–1155 cm^{-1} spectral region was selected because it contains the main phosphate vibrational bands, including overlapping contributions from HPO_4^{2-} and PO_4^{3-} . Focused beam reflectance measurement was used to distinguish clear-solution and slurry regions, enabling spectra from fully dissolved conditions to be selected for calibration and validation. Partial least squares regression was then applied to extract concentration information from the full spectral response rather than relying on a single peak height. The resulting multivariate calibration provides a foundation for real-time estimation of phosphate concentration during dissolution and crystallization. The calibrated models were further used to generate temperature-dependent solubility profiles from continuous in situ measurements. These results demonstrate how data-rich PAT experiments and multivariate modeling can support real-time monitoring of solubility, supersaturation, and crystallization risk in complex alkaline process streams.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



DEM–CFD Modeling of a Packed Bed Reactor: Analysis of Local Transport & Deactivation Dynamics across Aspect Ratios

Raj Chapagain

Tribhuvan University, Institute of Engineering, Department of Chemical Engineering

Keywords: Packed Bed Reactor, CFD–DEM, Discrete Element Method, Low Aspect Ratio, Catalyst Deactivation, ANSYS Fluent, Rocky DEM, Heat Transfer, Pressure Drop, Turbulence Modelling, Nusselt Number, Porosity, Thermal Runaway, Reynolds Stress Anisotropy.

ABSTRACT

Our thesis presents a three-dimensional computational investigation of coupled fluid flow (in an ANSYS 2-Way Fluent Coupling, student version), heat and mass transfer, and catalyst deactivation dynamics within a low aspect ratio packed bed reactor ($D/d_p < 6$), where confining wall effects and packing heterogeneity govern local transport phenomena. The work bridges the gap between molecular-scale reaction kinetics and reactor-scale performance through an integrated Discrete Element Method–Computational Fluid Dynamics (CFD–DEM) modelling framework. The random packing of $M = 243$ spherical catalyst particles inside a cylindrical vessel (diameter $D = 0.15$ m, height $H = 1.2$ m) was generated using DEM with the Hertz–Mindlin contact model and Coulomb friction ($\mu_s = 0.5$), replicating an industrial gravity-driven settling process.

The resulting geometry was transferred via Boolean subtraction to create the interstitial fluid domain, which was discretised using mixed tetrahedral–polyhedral elements (3,705 cells, minimum orthogonal quality 0.584). Particle-resolved Reynolds–Averaged Navier–Stokes (RANS) simulations were performed using the standard k – ϵ turbulence model under a pressure-based transient solver with the SIMPLE pressure–velocity coupling algorithm. Air at standard conditions ($\rho = 1.225$ kg/m³, $\mu = 1.7894 \times 10^{-5}$ Pa·s, $Pr = 0.71$) served as the working fluid, with aluminium ($\rho = 2719$ kg/m³, $\lambda = 202.4$ W/(m·K)) representing the

catalyst pellets. The principal findings of this investigation are as follows.

First, the radial porosity distribution exhibits the characteristic damped oscillatory behaviour, with porosity approaching unity at the confining wall and converging to $\epsilon \approx 0.40$ – 0.42 in the bed interior, in agreement with established correlations by Mueller and de Klerk.

Second, near-wall velocity channelling is observed, with maximum interstitial velocities reaching approximately 3.27 m/s—confirming that the wall’s influence permeates the entire bed cross-section at low aspect ratios.

Third, pressure-drop predictions agree with the Ergun, KTA, and Einfeld–Schnitzlein correlations to within 25%, with the improved Reger formulation providing the closest match through its porosity-dependent form-loss correction.

Fourth, particle-to-fluid heat transfer analysis reveals Nusselt number variations between inner and outer bed regions, with Wakao’s correlation demonstrating superior accuracy; at lower D/d_p ratios, the inner and outer regions exhibit consistent heat transfer behaviour, whereas at higher ratios a significant reduction in outer-region Nusselt number is observed.

Fifth, the theoretical framework for Reynolds stress anisotropy characterisation—including Lumley triangle invariant maps and barycentric mapping—is

developed to quantify the departure from the isotropic Boussinesq assumption inherent in the $k-\epsilon$ model.

Sixth, catalyst deactivation patterns are shown to be markedly asymmetric along both axial and radial directions, with the dimensionless activation energy (γD) and Damköhler number (DaD) controlling the spatial heterogeneity and temporal evolution, respectively.

The “wrong-way” temperature behaviour and local thermal runaway conditions—phenomena entirely invisible to two-dimensional and pseudo-homogeneous formulations—are captured through the three-dimensional pore-level resolution. The comprehensive

analysis demonstrates that particle-resolved CFD-DEM modelling is essential for capturing the local transport heterogeneities and deactivation dynamics that govern packed bed reactor performance, establishing a robust multi-scale framework for the design and safety assessment of industrial catalytic systems.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



How Archimedis Operational Reasoning System Detected a Hidden Furnace Failure In Under an Hour

Charles Crowell

Archimedis

ABSTRACT

A mid-sized refinery (150K BPD) was experiencing repeated thermal cycles and unit trips on its Gasoline Hydrotreater Reactor Furnace. The plant's DCS and APC indicated the furnace was operating within healthy parameters, masking a serious latent failure. This poster presents how Archimedis's Operational Reasoning System (ORS), a multi-agent system that ingests heterogeneous plant data and reasons over it the way an experienced process engineer would identified the root cause in under an hour. Working with plant engineers, the ORS reviewed several months of operating data alongside event data from the cycles and trips. It detected a 3% gap between theoretical and measured stack O₂, indicating an air-deficient, fuel-rich combustion state inconsistent with the air flow transmitter reading. By cross-referencing combustion stoichiometry, draft pressure, control-valve backpressure trends, and radiant-versus-convective duty shifts, the ORS isolated the cause to a major Air Preheater (APH) leak diverting an estimated 34% of measured airflow before combustion. A subsequent bowtie process safety review confirmed that the leak

had compromised the furnace purge cycle and was producing convection-section afterburn creating a credible explosion-during-startup possibility that conventional monitoring had not flagged. The post illustrates three capabilities that distinguish an ORS from conventional APC, anomaly detection, or first-principles models: (1) automated cross-referencing of disparate, weakly-coupled data streams; (2) hypothesis generation and falsification against process physics; and (3) production of engineer-auditable evidence chains. End-to-end diagnosis took approximately one hour versus an estimated week of manual investigation, avoiding ~\$200K/day in further downtime and a portion of the turnaround budget. A realized value of \$1.5–2M on a single event.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Stitching Misoriented and Misaligned Non-Overlapping Images

Michael Fokuo

Auburn University, Department of Chemical Engineering

ABSTRACT

Image stitching is a fundamental task in biomedical imaging, enabling reconstruction of large specimens that exceed the field of view of a single acquisition. Conventional stitching methods rely on overlap between neighboring tiles and known acquisition geometry. These methods typically use feature matching registration to estimate spatial transformations and alignment [1–3]. Although effective under controlled conditions, these assumptions do not hold when tiles do not overlap or when their orientations are unknown. This situation often occurs during the physical handling of samples [4]. We address this challenge by proposing a framework for stitching non-overlapping biomedical image tiles with unknown orientations.

We conduct the study using a single biomedical dataset that is divided into two subsets: (1) a translation-only (alignment) subset, and (2) a translation-and-orientation subset. In the first subset, orientations are known, but both horizontal and vertical offsets between tiles are unknown. In the second subset, tile orientations are unknown, and only horizontal offsets between tiles are considered. The proposed framework follows a two-stage strategy that decouples orientation determination from alignment. First, tile orientation is inferred using a compatibility measure (CM). Next, tile translation is estimated using phase correlation applied to tile boundaries, rather than relying on shared pixel content. This approach achieves approximately 91% correct reconstructions (182 out of 200) on the translation-and-orientation subset.

This result demonstrates that accurate reconstruction is possible even without overlap or

orientation metadata. However, this result is obtained on small grid sizes (single rows), where exhaustive brute-force assembly strategies are still computationally feasible. However, scaling to larger grid sizes makes brute-force assembly computationally infeasible. This motivates the use of derivative-free optimization (DFO)-based assembly strategies, including Bayesian optimization and Tabu search, as scalable alternatives.

Table 1. Comparison of pairwise compatibility measures on the small-grid translation-and-orientation subset (Subset 2).

Compatibility Measure	Assembly Strategy	
	Bayesian Optimization	Tabu Search
New CM	121 out of 166 (72.89%)	182 out of 200 (91%)
MGC [5]	–	200 out of 200 (100%)

A key focus of this study is the comparison of pairwise compatibility measures used in determining the orientation of tiles. Specifically, we compare the compatibility measure we developed with the Mahalanobis Gradient Compatibility (MGC) measure from the image puzzle-solving literature [5,6]. Table 1 presents preliminary results for this comparison using the single row, translation-and-orientation subset (Subset 2). Both measures perform strongly in this setting. The developed CM achieves up to 91% accuracy, while MGC achieves 100% accuracy. These results motivate

the integration of both measures into scalable, optimization-based assembly frameworks for large grids. Overall, this work demonstrates the progression of non-overlapping stitching frameworks from small, controlled experiments to scalable and computationally efficient solutions for large-scale biomedical image reconstruction.

REFERENCES

1. Brown M, Lowe DG. Automatic panoramic image stitching using invariant features. *Int J Comput Vision* 74:59–73 (2007).
<https://doi.org/10.1007/s11263-006-0002-3>
2. Preibisch S, Saalfeld S, Tomancak P. Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* 25: 1463–1465 (2009).
<https://doi.org/10.1093/bioinformatics/btp184>
3. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY. Fiji: an open-source platform for biological-image analysis. *Nature Methods* 9:676–682 (2012).
<https://doi.org/10.1038/nmeth.2019>
4. Chalfoun J, Majurski M, Blattner T, Bhadriraju K, Keyrouz W, Bajcsy P, Brady M. MIST: accurate and scalable microscopy image stitching tool with stage modeling and error minimization. *Scientific Reports* 7:4988 (2017).
<https://doi.org/10.1038/s41598-017-04567-y>
5. Gallagher AC. Jigsaw puzzles with pieces of unknown orientation. In 2012 IEEE Conference on computer vision and pattern recognition (pp. 382–389). (2012)
<https://doi.org/10.1109/CVPR.2012.6247699>
6. Cho TS, Avidan S, Freeman WT. A probabilistic image jigsaw puzzle solver. In 2010 IEEE Computer society conference on computer vision and pattern recognition (pp. 183–190) (2012).
<https://doi.org/10.1109/CVPR.2010.5540212>

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Role of Multivariate Data Analyses In Formulation and Process Development of Oral Solid Drug Products: Encapsulation Case Studies

Shashwat Gupta

Eli Lilly and Company, Synthetic Molecule Design and Development (SMDD)

ABSTRACT

Multivariate data analysis (MVDA) methods are an important tool in a pharmaceutical drug product engineer's toolbox as part of a Quality by Design (QbD) driven framework for formulation and process development of oral solid dosage forms. This work presents such an MVDA application for two commonly used encapsulation processes in the following case studies: Case study 1: Partial least squares regression (PLSR) based enhancement of process efficiency of a vacuum-assisted drum filling encapsulation process A vacuum-assisted drum filling process was used to fill an active pharmaceutical ingredient (API). Intuitively, the drum bore volume is a function of the amount of API to be filled (combination of dose and incoming potency) and API physical properties (particle size distribution, density, etc.). Thus, selection of an appropriate drum bore volume was desired to reduce process setup time. To enable this, a PLSR model was built correlating API physical properties and drum filling process parameters to API plug density. The developed model was then used to identify drum bore volumes for manufacturing campaign support, designing a drum library that encompasses expected variability in API physical properties and potency, and a simple drum lookup table to support commercial manufacturing. Case study 2: Principal component analysis (PCA) to enable formulation development of an

excipient mixture filled with a dosing disk encapsulation process A dosing disk encapsulation process was used to fill an excipient mixture (sodium bicarbonate and dimethicone). During formulation development, the effects of varying excipient mixture properties on the encapsulation process (fill weight variability) and product performance (in-vitro dissolution) were evaluated. Varied excipient mixture properties were generated by manufacturing under a wide range of raw materials (sodium bicarbonate particle size/amount, dimethicone amount, and dimethicone viscosity) and process conditions for different mixing technologies (batch and continuous). PCA was then applied to identify excipient mixtures with the most varied physical properties that encompass all other excipient mixtures. The excipient mixtures extremes were then tested for process and product performance. The analysis demonstrated that the product and process performance was robust across the range of material properties and mixing process technologies that were evaluated for the excipient mixture.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Science Guided Machine Learning for Conceptual Process Development: A Novel Heterogeneous Azeotropic Separation Case Study

Troy Gustke

Virginia Tech, Chemical Engineering

ABSTRACT

Hybrid modeling has stood out as stable method to apply machine learning (ML) to chemical process systems, as it can apply first-principles knowledge and constraints to data-driven methods for increased speed and precision. While substantial work has focused on leveraging operational data, the application of ML to conceptual process design remains limited, despite its strong influence on overall plant economics. Various data-driven methods have improved the speed and accuracy of process development, particularly for optimization; however, they are typically applied to established processes or predefined superstructures, leaving the conceptual process synthesis stage largely unaddressed. In this work, we integrate existing and novel science-guided ML methods into a process development framework to address challenging design problems. This approach integrates with existing process development procedures while benefiting from novel data-driven tools. These methods include modern ML thermodynamic predictive models, Bayesian parameter estimation and uncertainty quantification, science-guided ML/data analysis, and generative AI to improve the speed, depth, and nuance of developed rigorous process models. To appropriately demonstrate this framework with a challenging

design problem, we guide a case study for the heterogeneous azeotropic separation of 2-methyltetrahydrofuran (MeTHF), acetonitrile (ACN), and water; a mixture typically found in pharmaceutical solvent recovery systems. This system is not studied in literature and presents several key challenges including an ill-defined thermodynamic parameter regression, non-trivial process topology, and multi-objective solvent selection. We apply existing ML tools to solve each problem with novel approaches. In addition, we demonstrate the capabilities of modern large language models for process selection when supplied with thermodynamic data and in-the-loop subject matter experts. This case study demonstrates how the integration of data-driven tools enhances the traditional process development framework for conceptual design.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



How to Teach Programming to ChE's In the Age of AI

Robert Hesketh

Rowan University, Chemical Engineering

ABSTRACT

At Rowan University we decided to address a perennial student complaint that the computer science programming class material was never used in later chemical engineering classes. In 2022, we replaced the required programming course with a required chemical engineering course called ChE Modeling. This course introduces students to the modeling of chemical processes using practical simulation tools; the same ones used in industry. Students learn how to build models of complex chemical processes, evaluate the accuracy of models, and use models for process optimization and design decisions. We start this course using the Begin Python with TCLAB [1] modules. This is a unique module in which they learn a programming language to control an Arduino that has 2 heaters and thermistors. This immediately addresses a common complaint that they never used the programming language taught by computer science in a chemical engineering class; they now use python immediately. John Hedengren, the developer of this program, estimates that students can complete the 12 modules in 2 – 3 hours.

In this course they learn how to control a system with 2 heaters and 2 thermistors. In this module they use on/off control, similar to a home heating/cooling system. The TCLAB can be used in several other courses in ChE. The material in the ChE modeling course then focuses on ChE computational tools solving problems introduced in the previous years (1st and 2nd). Based on the Felder, Rousseau and Bullard book they use python jupyter notebook templates, Excel Templates, and Aspen Plus Tutorials to perform rigorous multicomponent flash drum problems using the Rachford-Rice Equation [2], Adiabatic Flame

Temperature, and Transient Material Balances. The energy balances are based on thermodynamic properties from DIPPR. We continue with fluid mechanics and use a solver to solve for the flowrate in pipes and complex pipe networks. Next are transport problems in momentum, mass, and heat transfer ODE's and finite difference approximations. Next are the linear and non-linear regressions with confidence intervals, root finding methods, and plug flow models. Finally of interest to this conference are the Machine Learning Basics that are introduced and the changes required to incorporate AI in teaching programming.

Table: Brief set of problems covered in this course.

Topic / Exercise	Software / Tools
TCLab Learn Python Modules	Python
Felder, Rousseau and Bullard: Flash Drum Calculations (multiple nonlinear equations)	Excel Solver, Aspen Plus, Python template
Felder, Rousseau and Bullard: Adiabatic Flame Temperature	Excel Solver, Python <code>fsolve</code> , Aspen Plus
Felder, Rousseau and Bullard: Transient Material Balance Example Problem	Python ODE solver
Example Adiabatic Flame Temperature	Excel Solver, Python <code>fsolve</code> , Aspen Plus
Flash Drum Interpolations	Excel and Python
Geankoplis: Binary Mass Transfer in a Stefan Tube	Python ODE solver
De Nevers: Calculation of Flow Rate	Excel Solver, Python <code>fsolve</code>

Geankoplis: Flow in Complex Pipe Networks	Excel Solver, Python <code>fsolve</code>
Geankoplis: Unsteady-State Diffusion Using Finite Difference Approximations in the Method of Lines	ODE solver, COMSOL
Process Flow Diagrams	MS Visio
Heat Conduction in a Wire with Electrical Heat Source	Python ODE solver, COMSOL
Linear Regressions with Confidence Intervals	Excel, Python <code>curve_fit</code>
Nonlinear Regressions with Confidence Intervals	Excel, Python <code>curve_fit</code>
Root Finding: Bisection, False Position, Newton-Raphson, Secant, and Modified Secant Methods	Excel and Python
Numerical Methods to Integrate ODEs: Euler and Runge-Kutta	Excel and Python
Data Acquisition Using Microcontrollers and Sensors	Python
Machine Learning Basics: Big Data	https://apmonitor.com/pds/index.php/Main/AutomaticMonitoring
Case Study 1: Catalyst Light-Off	Python

Case Study 2: GIS Map Visualization	Python
Case Study 3: Fuel Efficiency Regression	Python
Geankoplis Plug Flow Mass Transfer Models for Gas Absorption	Python ODE solvers
Plug Flow Models: Double Pipe Heat Exchanger	Python ODE solver
Geankoplis Equilibrium Stage Models for Gas Absorption	Excel and Aspen Plus

REFERENCES

1. Hedengren J, Hedengren E. `begin_python`, https://github.com/APMonitor/begin_python
2. Deriving the Rachford-Rice Procedure for Isothermal Flash Distillation, LearnChemE, https://youtu.be/ACxOiXWqlSQ?si=v01KQtayp_89AGnj

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



A Streamlit-Based Platform for Ternary Solvent Solubility Modeling and Crystallization Process Design

Marko Ivancevic

Bristol Myers Squibb, Chemical Process Development

ABSTRACT

Crystallization design in pharmaceutical development is commonly supported by modeling approaches for single-solvent and binary solvent systems. In some cases, ternary solvent mixtures are desirable due to better impurity purge. However, this benefit can be offset by increased operational and analytical complexity arising from the multicomponent solvent composition, which can limit systematic exploration and broader adoption of ternary crystallization strategies during process development. To address this challenge, a user-accessible analytics platform has been developed using Streamlit to enable ternary solvent solubility modeling and crystallization process design within an integrated workflow. The application allows users to upload experimentally measured solubility data for ternary solvent systems via a web-based interface. Upon data import, the software automatically fits the data to five semi-empirical ternary solubility models using nonlinear regression. Model performance is evaluated using statistical goodness-of-fit metrics. The app outputs parity plots and fitted parameter values for all five candidate models, and an interactive ternary solubility for the best model. This approach enables transparent model comparison and supports informed selection of solubility correlations best suited for a given system, rather than reliance on a single prescribed model. The selected solubility

model parameters are subsequently fed into a crystallization design module tailored to process-relevant decision making. The experimental data and fitted solubility surfaces are visualized on ternary solubility diagrams, facilitating assessment of solvent composition effects that may occur due to process variability. Building on these correlations, a downstream crystallization mapping tool enables users to explore solvent composition trajectories representative of realistic process operations, such as incremental antisolvent addition. By targeting user-specified final solubilities, the platform supports estimation of key performance indicators in pharmaceutical process development such as crystallization yield and mother liquor losses. In summary, this work demonstrates how streamlined process analytics and model selection tools can reduce the practical complexity associated with ternary solvent crystallization, enabling data-driven design decisions that leverage the impurity-purging advantages of multicomponent solvent systems in pharmaceutical manufacturing.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Orchestrating Modelling & Simulation of Pharmaceutical Production Processes Via Large Language Models

Christoph Kloss

DCS Computing GmbH

ABSTRACT

The digitalization of production processes necessitates the transition from siloed simulation tools to integrated, automated, and intelligent workflows. In the domain of particulate processes in pharmaceutical manufacturing, agitated drying is playing an important role. Predicting crystal attrition is critical to maintaining target particle size distributions (PSDs) and bioavailability. Traditionally, detailed 3D discrete element method (DEM) models and reduced-order process models have remained separate due to computational disparities. This work proposes an integrated approach to bridge these scales by leveraging physics-informed models and Large Language Models (LLMs) to orchestrate complex characterization and optimization loops. The first time Machine Learning was used to accelerate the characterization of powder properties for DEM was by Benvenuti et al. (2016). Subsequent innovations introduced physics-informed reduced-order models, such as the two-dimensional population balance equation (2D-PBE) developed by Togni et al. (2025), which links attrition rates to impeller torque, particle aspect ratio, and residual solvent content.

While recent research has utilized Graph Neural Networks to further accelerate 3D granular simulations (Mayr et al, 2021 and 2021), the current challenge lies in the seamless integration of these heterogeneous models, material databases, and experimental data. We demonstrate an approach where an LLM assistant facilitates the re-combination and execution of

python-based calculation elements—modular code units representing specific physical equations or ML methods—into executable workflows. Within our platform *engicloud.ai*, these calculators are derived both from expert knowledge and extracted from research literature, providing a verified foundation for cross-disciplinary collaboration. For the specific use case of predicting crystal attrition, the platform enables researchers to run a physics-based PBE model informed by 3D simulation data (Togni 2025) in an LLM-driven loop to optimize powder properties by dynamically adjusting process conditions based on material-dependent parameters calibrated from ring shear-cell and lab-scale experiments.

The role of the LLM can be manifold, such as complementing traditional optimization methods, e.g. to simulate scenarios in case of supply chain disruption (such as due to a war or pandemic) where the LLM agent would provide a list of materials available in a certain scenario, which in turn provides the material properties for the PBE model. By providing a low-entry-barrier, web-based environment for the exploration and the ability to interoperate with material models and databases via APIs and access to ~20,000 static (verified) models extracted from LLMs, *engicloud.ai* enhances the reproducibility and scalability of digital twins in materials science.

This work highlights how such modular software architectures foster FAIR initiatives and accelerate the transition from initial material conception to late-

stage process validation and optimization in real life scenarios.

REFERENCES

1. Benvenuti L, Kloss C, Pirker S. Identification of DEM simulation parameters by Artificial Neural Networks and bulk experiments. *Powder Technology* 291:456–465 (2016).
<https://doi.org/10.1016/j.powtec.2016.01.003>
2. Mayr A, Lehner S, Mayrhofer A, Kloss C, Hochreiter S, Brandstetter J. Learning 3D Granular Flow Simulations. *arXiv* (2021)
<http://arxiv.org/pdf/2105.01636v1>
3. Mayr A, Lehner S, Mayrhofer A, Kloss C, Hochreiter S, Brandstetter J. Boundary Graph Neural Networks for 3D Simulations. *arXiv* (2021).
<http://arxiv.org/pdf/2106.11299v7>
4. Togni R, Saurer EM, Hicks W, Rao P, Alabanza LM, Clements P, DiPietro A, Derdour L, Engstrom J, Hartmanshenn C, Jayaraman S, Jones–Salkey O, Lamberto DJ, Mustakis J, Ncube G, Kloss C. A two-dimensional population balance model for predicting the attrition of elongated particles during agitated drying, *Powder Technology* 464:121198 (2025).
<https://doi.org/10.1016/j.powtec.2025.121198>

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Why Transfer Learning Fails Under Target Non-Identifiability

Yuki Kobayashi

Kyoto University

ABSTRACT

Transfer learning (TL) improves model performance in a target domain with limited data by leveraging data from a source domain. When the source–target discrepancy is large, TL can degrade target-domain performance, a phenomenon known as negative transfer (NT). In linear regression, the coefficient vector is only partially identifiable when the target design matrix is rank-deficient. Although TL can exploit source information to address such non-identifiability, it may also amplify coefficient estimation error. However, the mechanisms and conditions underlying this type of NT have not been fully characterized. Frustratingly easy domain adaptation (FEDA) is a TL method that has been successfully applied in the process industry. This study derives the mechanism and conditions of NT in FEDA with linear regression. Because the derived NT condition involves the unob-

servable true target coefficient, we further construct a proxy condition that can be evaluated from observed data by assuming an upper bound on the source–target coefficient discrepancy. Synthetic experiments confirm the mechanism of NT and examine the effects of coefficient-shift sparsity, coefficient-shift magnitude, and target sample size. This study contributes to a theoretical understanding of the NT mechanism under target non-identifiability in FEDA with linear regression.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Domain-Decomposition Pinns for Rapid Prediction of Stirred-Tank Mixing Flows across Geometric Scales

Yohei Kono

Hitachi, Ltd., Research & Development Group

ABSTRACT

Stirred-tank mixing is a central operation in batch chemical process design and scale-up. While high-fidelity Computational Fluid Dynamics (CFD) is effective for evaluating internal flow states, its computational cost prohibits its use in many-query tasks such as design-space exploration and parametric screening across varied equipment scales. Recent advances in Physics-Informed Neural Networks (PINNs) offer a promising path toward rapid surrogate modeling; however, conventional PINNs often fail to capture complex boundary conditions near impeller regions, frequently collapsing to trivial zero-velocity solutions due to dominant localized forcing terms. In this work, we present a data-informed, physics-constrained hybrid surrogate model for stirred-tank mixing within a bounded geometric-operating space. We employ a domain-decomposition approach: the complex impeller-region dynamics are represented through data-driven boundary conditions derived from a limited set of initial CFD simulations, while the bulk inertial flow field is reconstructed using physics-based constraints (Navier–Stokes and continuity equations).

The model was evaluated across multiple stirred-tank specifications, including varying scale factors (1x, 1.5x, 2x) within the sampled design space. Results demonstrate that a single proposed model accurately reproduces three-dimensional velocity fields and mixing trends (RMSE = 0.0261), while reducing inference time by over 99% compared to repeated CFD simulations. Rather than functioning as an extrapolative general scale-up model, this framework is positioned as a local surrogate. By amortizing the initial offline CFD costs, it provides rapid, physics-consistent flow predictions for scale-related design analysis and operating-condition screening within an admissible interpolation domain.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Efficient Parameter Estimation In Agent-Based Models of Collective Cell Invasion Via Gaussian Process Surrogates and Bayesian Optimization

Aneesh Krishna

University at Buffalo, The State University of New York, Institute for Artificial Intelligence and Data Science

ABSTRACT

Cell migration and invasion are key processes underlying cancer metastasis, driven by cell–cell adhesion, chemotaxis, and matrix remodeling. Agent-based models (ABMs) are a powerful computational approach for simulating these complex, multicellular behaviors. In an ABM, each cell is represented as an autonomous agent that follows local rules governing its interactions with neighboring cells and the surrounding environment. This bottom-up framework captures the emergent, heterogeneous dynamics of cell populations that continuum models cannot easily reproduce. CompuCell3D is a widely used platform for ABMs in biological cell systems; however, a critical limitation is that it and most other ABM tools do not natively support systematic, black-box parameter estimation for their stochastic simulations. Identifying biologically meaningful parameter sets for ABMs is especially challenging because the stochastic nature of these simulations means that repeated runs with identical parameters produce different outputs, making it difficult to reliably assess how well any given parameter set matches experimental data. Existing calibration strategies, such as Monte Carlo sampling and grid search, address this challenge by simply running the model many times across a broad parameter space; an approach that is computationally expensive and scales poorly as model complexity grows.

What is needed is a more efficient approach that intelligently navigates the parameter space, reducing the number of costly simulations required to achieve

accurate calibration. This work focuses on parameter estimation for agent-based modeling of collective cell invasion for two cancer cell phenotypes: a network (invasive) phenotype and a spheroid (non-invasive) phenotype emerging from growth of tumor spheroids simulated using CompuCell3D. We adopt a data-centric approach based on Gaussian process surrogate models and Bayesian optimization (BO). Unlike traditional sampling methods that treat parameter estimation as exhaustive search, BO constructs a probabilistic surrogate model of the simulation's behavior and uses it to sequentially propose new parameter sets that are expected to provide maximal information gain. This allows the framework to efficiently operate with limited simulation data, dramatically reducing the number of model evaluations needed. Leveraging simulation outputs alongside experimental data for invasion dynamics and circularity of cell clusters, the method iterates through sequential rounds of parameter selection, ABM simulation, and comparison with experimental observations.

The calibrated model successfully reproduces experimental invasion dynamics and circularity across both phenotypes. This framework provides a broadly applicable, efficient solution for parameter estimation in stochastic ABMs, addressing a gap in existing modeling tools for ABM calibration and reducing the need for extensive trial-and-error simulation of expensive models.



Empowering Automated Process Analysis through LLM-Based Literature Mining, Flowsheet Digitization, and Simulation

Jan-Frederic Laub

ETH Zurich, NCCR Catalysis, Energy and Process Systems Engineering

ABSTRACT

The chemical industry needs to transition from predominantly linear, carbon-emitting production routes to circular, carbon-reusing processes. Therefore, every current and future production process needs to be critically evaluated and potentially re-designed. Today, process design and assessment rely on detailed process simulations [1]. However, constructing these simulations remains a bottleneck, demanding a high degree of expertise and manual work.

Here, we present an automated workflow which gathers process knowledge, generates process simulations, and evaluates process performance. The automated workflow consists of two data pipelines: The first pipeline systematically extracts information from literature [2] and prepares a knowledge base of established, industrially relevant chemical processes down to the level of unit operations and thermodynamic properties.

This knowledge is aggregated into one text per process and fed into the second pipeline, “text2flowsheet” [3], which digitizes expert-level flowsheet graphs from natural-language descriptions. Both pipelines leverage LLMs' language comprehension capabilities, while the flowsheet digitization is further grounded in rigorous thermodynamic calculations.

The digitized flowsheet graphs are systemically translated into simulations within an established commercial process simulator. Potential simplifications necessary to achieve convergence are recorded transparently. Missing information on operating

parameters is augmented by systematic, unit-by-unit black-box optimization.

We show that our integrated pipelines can faithfully collect and digitize chemical process information by comparing to expert-curated datasets and manually drawn flowsheets. Furthermore, the generated process simulations are on par with expert interpretations with significantly less manual effort. We present case studies illustrating how the results of the automatically generated process simulations can be used to assess process sustainability and derive optimization potential.

REFERENCES

1. Parvatker AG, Eckelman MJ. Simulation-Based Estimates of Life Cycle Inventory Gate-to-Gate Process Energy Use for 151 Organic Chemical Syntheses. *ACS Sust Chem Eng* 8:8519–8563 (2020). <https://doi.org/10.1021/acssuschemeng.0c00439>
2. Ullmann's Encyclopedia of Industrial Chemistry, John Wiley & Sons, Inc., USA (2011).
3. Laub JF, Bosetti L, Bardow A. Text-to-Flowsheet: An Automated LLM-Pipeline for Digitization and Simulation of Chemical Processes with Expert-Level Accuracy. *ChemRxiv* (2026). <https://doi.org/10.26434/chemrxiv.15001581/v1>

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and

adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



From Disparate Data to Accelerating Innovation: A Practical Framework for R&D Digitalization

Zifeng Li

Qnity Electronics Inc.

ABSTRACT

Manufacturing has benefited from decades of digital standardization, automation, and mature data pipelines. Digitalization in research and development (R&D)—especially in materials and process development—lags significantly due to its heterogeneous and continuously evolving datasets spanning structured measurements, semi-structured metadata, and unstructured content such as lab notes. The lack of flexible, end-to-end digital infrastructure leads to common pain points, including data loss, repeated experiments, long cycle times to derive insight, and barriers to finding and reusing prior knowledge. This work presents a practical digital transformation framework for R&D, centered on knowledge-centric platforms designed and operated as scalable digital products. At Qnity, an R&D digital transformation project focused on four tightly connected areas. First, customer insight, where direct customer feedback is translated into scalable physics-based and AI models for product development, supported by materials databases. Second, lab digitalization, which connects Qnity laboratories, enabling cross-functional teams to access critical experimental information in one place. Third, business analytics, where aggregated innovation data is used to track key performance indicators and monitor portfolio health—allowing leaders to review projects within minutes and make decisions based on evidence, with digital tools embedded directly into decision-making forums. Finally, generative AI ac-

eleration enables faster creation, search, and reuse of technical reports, patents, and project documents through a unified knowledge hub, reducing learning cycles for both new hires navigating steep semiconductor learning curves and experienced engineers seeking to leverage prior work. These platforms are intentionally integrated: lab digitalization is organized on a project basis through the business analytics system, while data from customer insights, lab operations, and business analytics feed into the generative AI platform. Yet technology alone is insufficient. Gaps between scientific intent, business value, and IT execution often emerge as teams speak different technical “languages,” leading to unclear requirements and delays in execution. The final piece is strong end-to-end ownership. Embedding a dedicated digital team into R&D with full lifecycle accountability has proven essential for delivering and sustaining long-term digital value. Qnity’s R&D digitalization has already demonstrated success in accelerating new product development timelines, shaping innovations that will define the future of advanced electronics.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Tennet-SAC: A Physics-Embedded Machine Learning Model for Activity Coefficient of Multicomponent Liquid Mixtures

Shiang-Tai Lin

National Taiwan University, Department of Chemical Engineering

ABSTRACT

We present TeNNet-SAC (Thermodynamics-embedded Neural Network for Segment Activity Coefficient) [1], a physics-embedded machine learning framework for predicting activity coefficients in multicomponent liquid mixtures directly from SMILES. Building upon the concept of segment-based thermodynamic foundation of COSMO-SAC model [2], TeNNet-SAC preserves physical interpretability while eliminating the need for quantum chemical calculations. The model comprises three components: (i) a σ -profile predictor that infers molecular surface charge distributions from SMILES, (ii) a geometry predictor for molecular volume and surface area, and (iii) a Γ predictor that computes segment activity coefficients. The σ -profile and geometry predictors are trained on 39,745 chemically diverse quantum-calculated structures, ensuring broad chemical coverage. The Γ predictor is designed to enforce thermodynamic consistency and is pretrained on one million synthetic data points to reproduce segment activity coefficients from the COSMO-SAC model, followed by fine-tuning using 34,371 curated experimental data points from VLE measurements of 397 binary mixtures. TeNNet-SAC achieves accuracy comparable to—and after fine-tuning, exceeding—the COSMO-SAC model, while inherently satisfying thermodynamic consistency and naturally extending to multicomponent systems. To facilitate adoption, the full implementation is openly available on GitHub [3], along with an easy-to-use web interface for rapid prediction without specialized

software [4]. In addition, both the code and the web platform provide functionality to generate NRTL parameters directly, enabling seamless integration with Aspen Plus for process simulation. This combination of physical rigor, scalability, and accessibility makes TeNNet-SAC a practical tool for phase equilibrium modeling and process design.

REFERENCES

1. Yang Y, Lin ST. Physics-Embedded Machine Learning Model for Phase Equilibrium Prediction in Multicomponent Systems. *J Chem Info Mod* 65:10180–10193 (2025).
<https://doi.org/10.1021/acs.jcim.5c01804>
2. Lin ST, Sandler SI. A priori phase equilibrium prediction from a segment contribution solvation model. *Ind Eng Chem Res* 41:899–913 (2002).
<https://doi.org/10.1021/ie001047w>
3. <https://github.com/yueyue2299/TeNNet-SAC>
4. <https://huggingface.co/spaces/stlin/tennetsac>

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Real-Time Inline Nitric Acid Quantification In Purex Systems Using Raman, ATR-FTIR, and Machine Learning

Nischal Maharjan

Georgia Institute of Technology, School of Chemical and Biomolecular Engineering

ABSTRACT

Liquid-liquid extraction (LLE) systems used in spent nuclear fuel reprocessing require reliable, real-time monitoring methods to improve process control, operational efficiency, and material accountancy. Spectroscopic techniques combined with chemometric and machine learning approaches provide a promising pathway for rapid, non-destructive quantification in these complex biphasic environments. In this work, PUREX-relevant solvents, nitric acid solutions (≤ 5 M) and 30% v/v tributyl phosphate (TBP) in n-dodecane, were used as a model LLE system to develop a chemometric workflow for direct quantification of nitric acid extraction from mixed-phase Raman spectra without phase separation. In parallel, single-phase aqueous and organic measurements were collected using both Raman and attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectroscopy to evaluate sensor-specific chemometric performance across relevant concentration ranges. Different machine learning methods were applied to

improve nitric acid quantification from Raman and ATR-FTIR spectra, and the performance of these approaches was compared across relevant concentration ranges. Data fusion strategies combining Raman and ATR-FTIR measurements were also explored to improve quantification in regions where one technique showed reduced sensitivity or higher prediction uncertainty. The results show that good quantification of nitric acid extraction can be obtained directly from mixed biphasic spectra using tailored algorithms. Also, integrating spectroscopy with machine learning and data fusion can improve inline monitoring capabilities in PUREX-relevant solvent extraction systems.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Machine Learning–Based Prediction of Heavy Metal Exposure In Spatially Heterogeneous Urban Environments

Paromita Nath

Rowan University, Department of Mechanical Engineering

ABSTRACT

Machine learning (ML) is increasingly used for predictive analysis in complex environmental systems. However, its effectiveness is often constrained by spatial heterogeneity and resource limitations that restrict data availability. This study evaluates the feasibility and limitations of ML-based predictive modeling for estimating stormwater heavy metal concentrations and associated public health risks in Camden, New Jersey, a historically industrial city with known contamination challenges. In this work, limited stormwater sampling data are integrated with environmental and anthropogenic predictors, including land use, proximity to industries, vegetation, and elevation for machine learning model development. Multiple regression-based ML models, including linear, ridge, lasso, random forest, and support vector regression, are trained and evaluated. In addition, hierarchical modeling approaches are explored to improve predictive performance. Results highlight challenges in predictive generalizability due to spatial heterogeneity and localized contamination hotspots, with models showing limited interpretability across sampling subsets. However, hierarchical modeling

approaches incorporating lead concentrations improve predictive performance for arsenic and cadmium. The findings reveal that limited predictive performance can provide valuable diagnostic insight into system complexity, data limitations, and the need for improved sampling strategies. Geographic information system (GIS)–based mapping of model predictions is used to visualize spatial patterns of contamination and identify potential high-risk areas. This work highlights the need for hybrid, interaction-aware modeling frameworks in data-constrained systems to better understand environmental processes. The proposed ML–GIS framework provides a scalable approach for identifying exposure risk patterns and supporting data-informed decision-making in environmental monitoring, urban planning, and public health.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Renewable-Driven Microgrid Design, Planning, and Operation of Integrated Gasification Fuel Cell for Biomass Upgradation to Biofuels

Oluwatimileyin Ogunsola

Oklahoma State University, School of Chemical Engineering

ABSTRACT

Biofuel production and upgrading offer a promising pathway for reducing the carbon intensity of liquid fuels. However, biomass-to-biofuel systems are energy intensive and require coordinated supplies of electricity, heat, and hydrogen. In a typical process, biomass is converted to bio-oil through pyrolysis, hydrogen is produced using an electrolyzer, and the resulting bio-oil is upgraded to transportation-grade biofuel. When the pyrolyzer is electrically heated and the upgrading unit uses an electrochemical pathway, the overall process imposes a large and time-varying electrical demand. Supplying this demand with renewable energy is attractive, but the variability of solar and wind generation can lead to renewable curtailment, grid dependence, and operational challenges.

These issues motivate the development of integrated microgrid scheduling frameworks that can coordinate renewable generation, storage, grid exchange, and onsite fuel-based power generation. This work develops a renewable-driven microgrid optimization framework for supporting a biofuel production and upgrading plant.

The proposed microgrid includes solar photovoltaic generation, wind turbines, battery energy storage, backup generators, grid interconnection, and an integrated gasification fuel cell (IGFC) subsystem. The IGFC subsystem consists of an electrified biomass gasifier, syngas storage, and a solid oxide fuel cell (SOFC). A fraction of the available biomass is fed to the gasifier to produce syngas for SOFC power

generation, while the remaining biomass is supplied to the main biofuel production plant. The intermediate syngas storage decouples gasifier operation from SOFC power generation, allowing syngas to be produced when energy is available and consumed by the fuel cell when dispatchable electricity is needed. High-fidelity equation-based models are developed for the electrified gasifier and SOFC to capture the key thermochemical and electrochemical behavior of the IGFC subsystem. Directly embedding these nonlinear models in the microgrid scheduling problem would lead to a computationally expensive MINLP formulation. Therefore, ReLU neural-network surrogate models are trained from simulation data generated using the detailed first-principle gasifier and SOFC models. ReLU activation functions are used because they provide piecewise-linear surrogate representations that can be embedded explicitly into an algebraic optimization model via OMLT (Ceccon, et al., 2022).

The trained surrogates are reformulated as mixed-integer linear constraints and incorporated into a mixed-integer linear programming framework for microgrid scheduling. The resulting optimization model coordinates renewable generation, battery charging and discharging, syngas production and storage, SOFC dispatch, generator operation, and grid electricity purchase while meeting the energy demand of the biofuel production plant. Case-study results demonstrate that the IGFC-enabled microgrid can improve renewable utilization by converting

otherwise curtailed renewable electricity into stored syngas and dispatchable SOFC power. The results also show how energy storage provides an additional form of energy flexibility, particularly during periods of low renewable generation and peak demand times for the main grid. By optimally coordinating gasifier operation, SOFC power production, and grid imports, the proposed framework reduces reliance on external electricity purchases and supports more economical operation of the biofuel facility.

This work provides a machine learning-based optimization framework for integrating thermochemical conversion, electrochemical power generation, renewable energy, and biofuel production within a unified optimization architecture. The proposed approach highlights the potential of IGFC-enabled microgrids to support low-carbon fuel production while

improving renewable energy utilization, reducing curtailment, and enhancing the operational flexibility of biomass-to-biofuel systems.

REFERENCES

1. Ceccon F, Jalving J, Haddad J, Thebelt A, Tsay C, Laird CD, Misener R. OMLT: Optimization & Machine Learning Toolkit. *J Machine Learning Res*, 23(349):1-8 (2022).

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Optimization of Biogas Steam Reforming Toward Low Carbon Hydrogen Production Using Integrated Artificial Neural Network and Genetic Algorithm

Ikechukwu Okwuosa

Obafemi Awolowo University

ABSTRACT

Hydrogen has been identified as a versatile energy carrier, offering a viable route to decarbonize and meet escalating global energy demands. Biogas produced from the anaerobic digestion of organic matter can potentially serve as a feedstock for hydrogen production using steam reforming process. This research investigates the optimization of a steam reforming process utilizing biogas feedstock for low-carbon hydrogen production using Artificial Neural Network (ANN) integrated with Genetic Algorithm (GA). An equilibrium based steady-state simulation of the process was developed using Aspen HYSYS to generate data for neural network training, validation and testing. Key process parameters considered for optimization include: biogas flow rate, steam flow rate, reformer temperature and reformer pressure with hydrogen mole fraction at reformer outlet as the response variable. A two-layer feedforward neural network with 4-12-1 architecture was trained on simulation data, achieving a correlation coefficient (R-value) of 0.99. This ANN model was integrated within the fitness function of GA to iteratively optimize process parameters subject to a steam-to-carbon ratio

constraint ≥ 2.5 to maximize hydrogen mole fraction while reducing the risk of catalyst deactivation via coking. The optimal parameters identified were 63 kg/h biogas flow rate, 62.04 kg/h steam flow rate, 1000°C reformer temperature, and 12.34 bar reformer pressure corresponding to a maximum hydrogen mole fraction of 0.5536 at the reformer outlet as predicted by the ANN model. Validation of these optimal parameters against the Aspen HYSYS model showed a relative error of 2.67% and 98.53% hydrogen yield at the reformer outlet. The proposed hybrid ANN-GA framework provides a robust, systematic approach for determining optimal operating conditions that enhance yield while maintaining operational reliability and efficiency.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Sketch2Simulation: Automating Flowsheet Generation Via Multi-Agent Large Language Models

Emma Pajak

*Imperial College London, The Sargent Centre for Process Systems Engineering,
 Department of Chemical Engineering*

ABSTRACT

Converting process flow diagrams into complete simulation models remains a persistent bottleneck in process systems engineering (PSE), requiring significant manual effort and simulator-specific expertise. Although advances in diagram interpretation and automated model generation have been made, these tasks are typically addressed in isolation, limiting the automation of end-to-end workflows. This work introduces Sketch2Simulation, a unified computational framework that automates flowsheet generation directly from raw engineering diagrams using a multi-agent large language model (LLM) architecture. The proposed framework integrates three coordinated layers: (i) Diagram Parsing and Interpretation, (ii) Simulation Model Synthesis, and (iii) Multi-level Validation.

In the first layer, multimodal LLM agents extract process semantics, identify unit operations and stream connectivity, and resolve implicit structural features. This information is encoded into a directed graph-based intermediate representation that captures process topology while enforcing simulator-compatible constraints. This intermediate representation serves as a formal interface between diagram interpretation and simulator execution, enabling consistent translation of unstructured visual inputs into simulator-compatible models.

The second layer translates this representation into a simulation model through sequential agents responsible for thermodynamic specification, object instantiation, and operating condition assignment,

culminating in simulation execution within Aspen HYSYS. The use of a multi-agent architecture enables decomposition of the workflow into specialised reasoning tasks spanning multimodal interpretation, structured model synthesis, and simulator interaction, improving scalability, interpretability, and robustness compared to monolithic LLM approaches.

The final layer introduces validation at multiple stages, including schema enforcement and an execution-and-correction loop that iteratively resolves runtime errors to ensure model validity. The framework is evaluated across four case studies of increasing complexity, including industrial-scale flowsheets with recycle loops.

Results demonstrate consistent generation of simulation models with high structural fidelity, achieving near-complete recovery of process topology (e.g., connection consistency ≥ 0.93 , stream consistency ≥ 0.96). Performance degradation is primarily associated with increased diagram complexity and dense interconnections. This work demonstrates that diagram-to-simulation transformation can be formulated as a unified computational problem, reducing reliance on manual model construction and advancing the digitalisation of PSE workflows. Crucially, this enables faster iteration between conceptual design and simulation, lowering the barrier to deploying high-fidelity models in both research and industrial settings.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and

adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Safety System Complexity: Ontology-Grounded Llms That Cross-Link Plant Records to Reduce Spurious Trips and Surface Hidden Process Risk

David Parham

uc2c.ai

ABSTRACT

Instrumented safety systems have driven incident rates down, but also create new challenges: the plants they protect are now far more complex. Two problems follow. First, true process risk hides in combinations of latent failures, process variability, corrosion, and degradation, so catastrophic incidents persist despite layered protection. Second, spurious safety-system trips have risen, eroding availability and driving unplanned downtime. The challenge is to keep the safety benefits while giving operations the information needed to run both safely and reliably. We present a case study in which large language models, grounded in a domain-specific process-safety ontology, extract and cross-link evidence across P&IDs, asset registers,

PHAs, management-of-change records, and incident reports. The system surfaces discrepancies between sources and answers operators' questions on demand, at the point of decision, with each answer traceable to its evidence. We report how ontology grounding changes extraction quality and trust, where it failed, and what generalizes to other operations.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Reinforcement Learning for Nonlinear Optimization In Process Industry

Kalpesh Patel

Saudi Aramco, Process & Control Systems Department

ABSTRACT

Reinforcement Learning (RL) is a machine learning technique which is capable of generating data and learning from it autonomously by interacting with the environment. RL has been successfully applied for learning and playing various games such as Go, Chess, Atari etc but its application to address process control and optimization problems is not trivial. There is a need for RL implementations in process industry to be safe, fast learning and explainable. A method for achieving such an implementation, for linear systems without disturbance variables, was published by the author in the past. Taking the work further, this paper proposes significant enhancements to the method in terms of ability to address severe process non-linearities, that can't be linearized, and ability to address disturbance variables explicitly in the RL problem formulation. Along with presenting the details on the enhancements, the paper also provides details on actual

implementation of the enhanced method for optimization of NGL fractionation unit. Not only does the enhanced method successfully enable the RL agent to learn the non-monotonic non-linearity representing the tradeoff between production and energy consumption, it also enabled the RL agent to learn how the non-linearity changed with changes in product and utility prices. The authors believe that this will further the potential of intelligent process control and optimization capable of enabling autonomous operation in the process industry.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Topology-Guided Response Surface Characterization for ML Model Selection

Shenbageshwaran Rajendiran

Auburn University, Department of Chemical Engineering

ABSTRACT

Machine learning (ML) models are widely used as surrogates for complex optimization problems in process systems engineering; however, selecting an appropriate surrogate remains challenging. Current selection approaches often rely on cross-validation, prior experience, or statistical and gradient-based response-surface descriptors. We hypothesize that topological descriptors such as the Euler characteristic curve (ECC), which captures structural changes across thresholds, provide complementary information for structure-aware surrogate model selection. We evaluated this hypothesis using 41 two-dimensional optimization test functions annotated with four landscape characteristics: modality, ruggedness, abruptness, and geometric complexity. For each function, ECCs were first computed on a dense 25×25 (625 samples) grid using sublevel filtration. Summary metrics from ECC, including the number of ECC jumps, cumulative ECC variation, and maximum ECC jump, were extracted from each curve. Thresholds for classifying each landscape characteristic were calibrated from the dense grid using Youden's index. To evaluate performance under sparse sampling, each function was randomly sampled at 10, 25, 50, 100, 300, and 600 points, with 30 replicates per sample size. The ECC computation was performed using a grid-based response surface representation. Therefore, the randomly sampled values were constructed onto 25×25 grids. Since random sampling does not cover every

grid cell, the remaining empty cells were filled using nearest-neighbor cell filling, where each empty grid cell was assigned the value of its nearest occupied cell. ECC summary metrics were then computed for each reconstructed grid, and the dense-grid-derived thresholds were applied to predict landscape characteristics. Prediction performance was evaluated using accuracy and the Matthews correlation coefficient (MCC). Across random-sampling experiments, ECC-based characterization improved with sample size and stabilized at approximately 100 samples. At this sample size, ECC metrics identified landscape characteristics with approximately 0.8 accuracy and 0.8 MCC, compared with approximately 0.9 for both metrics on the dense grid. These results suggest that ECC metrics capture response-surface structure that remains informative under sparse sampling and provide a promising avenue for structure-aware ML model selection. Future work includes extending the current framework to evaluate different sampling techniques, extending to higher dimensions, and relating ECC metrics to the performance of surrogate models with varying levels of complexity.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Identifiability of Microkinetic Parameters from Multimodal Operando Data

Gabriel Sabença Gusmão

Georgia Institute of Technology, School of Chemical & Biomolecular Engineering

ABSTRACT

Chemical kinetics provides the phenomenological framework for the elucidation of reaction mechanisms, in which ab-initio microkinetic models translate density-functional energetics into catalytic rates. Yet the underlying barriers carry uncertainties of 0.1 to 0.3 eV, and the extent to which a given set of measurements can retrieve them remains largely unquantified. Here, we frame the operando inverse problem as a maximum-likelihood estimation over a differentiable microkinetic model. The pseudo-steady-state surface enters as an algebraic constraint, and parameter sensitivities follow by automatic differentiation through its adjoint, the implicit function theorem applied at the converged root rather than through the solver iterations. These sensitivities propagate the measurement covariance into the parameter covariance, and the resulting Fisher information, the information each experiment carries about each barrier, ranks candidate experiments to establish which measurement determines which barrier.

We analyze two mechanisms as synthetic case studies, a CO oxidation model and a reverse water-gas shift submechanism drawn from a larger CO₂-hydrogenation reaction network, and assess the extent to which their barriers can be retrieved across gas chromatography, mass spectrometry, infrared surface spectroscopy, and isotopic transients. Multi-temperature data makes the activation energies and prefactors separately identifiable through the Arrhenius dependence, and gas chromatography alone retrieves the rate-determining barrier to a few meV and its prefactor to a few percent. The remaining steps reside

near equilibrium, in which the gas-phase observables carry vanishing sensitivity to their barriers, placing those directions in the nullspace of the Fisher information, so they are unidentifiable from gas data. For the reverse water-gas shift, isotope-resolved infrared spectroscopy of the adsorbed intermediates measures the unidirectional rates of the surface steps rather than their net rates, whereby this sensitivity is restored. Were the O* and OH* surface bands observable, resolving each adsorbed species' infrared signal by isotope label would retrieve all three of its barriers, including the quasi-equilibrated one, without assumed priors and to within tens of meV (joint multi-temperature Fisher, prefactors co-fit). On Rh(211), however, only the CO* band clears the infrared detection floor, whereas the O* and OH* intermediates that carry the quasi-equilibrated barrier lie orders of magnitude below it, leaving that barrier unidentified.

Identifiability is therefore set by the catalyst's intermediate binding energies rather than by the choice of experiment. More broadly, it emerges as a joint property of timescale, experiment, and catalyst rather than of a parameter alone: a barrier is identifiable only where an experiment resolves the step's intrinsic relaxation time, accesses unidirectional rather than net rates, and populates the carrying intermediate above the detection floor. Stated as a model-based design of experiments, this condition turns the modality, temperature, sampling location, and catalyst into the levers it ranks. The same framework could be extended to larger microkinetic models, including CO₂-hydrogenation routes to methane and alcohols, where

identifiability-guided operando design would point to the measurements and catalysts that pin the barriers governing activity and selectivity.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons

CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



A Machine Learning Framework for Short Peptide Sequence Optimization

Anh Trinh

Georgia Institute of Technology, School of Chemical and Biomolecular Engineering

ABSTRACT

Designing peptides plays an important role in applications ranging from therapeutics and biomaterials to diagnostics. However, due to the large combinatorial sequence space and the high cost and time required for experimental screening, experimental trial and error approaches are prohibitively expensive. Furthermore, peptide design is inherently a multi-objective problem that requires simultaneous optimization of different properties such as biological activity, stability, solubility, and safety. These challenges motivate the use of computational design strategies. Traditional physics-based and sequence-alignment methods often struggle to handle variable length sequences and often rely on structural information that is unavailable for many peptides. More recently, deep learning models such as AlphaFold, ESM, and diffusion-based approaches have transformed protein modeling. However, their large data requirements, high computational cost, and black-box nature reduce their practicality for deterministic multi-objective optimization in limited-data settings. This work proposes a data-driven, multi-objective peptide design framework that integrates sequence-to-feature transformations using Fast Fourier Transform (FFT) – based representations, interpretable feature attribution through GroupSHAPLEY, and metric-learning based optimization strategies. A bidirectional mapping between sequence space and feature space is introduced to identify critical feature contributions and

improve interpretability during peptide optimization. The primary focus of this poster is the optimization component of the framework. Specifically, distance metric learning methods, including Neighborhood Component Analysis (NCA) and Large Margin Nearest Neighbor (LMNN) are investigated to maximize class separation between peptide groups and identify discriminative feature representations. Comparative analyses were performed to evaluate the robustness, tunability, and optimization behavior of these methods on both synthetic and peptide-representative datasets. Using Support Vector Machine (SVM) as the predictive model, the proposed methods demonstrate improved optimization efficiency through dimensionality reduction in synthetic data experiments. Multiple optimization constraints can be incorporated to assess the robustness, scalability, and adaptability of the framework across different design scenarios. The proposed framework aims to provide an interpretable and computationally efficient alternative for peptide design under limited-data constraints.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Generalized Physics-Informed Deep Learning Framework for Chemical Process Modeling

Harshit Verma

Princeton University, Chemical and Biological Engineering

ABSTRACT

The incorporation of mechanistic, first-principles chemical unit operation models into process modeling frameworks remains computationally challenging. Mechanistic models governed by complex nonlinear systems of ordinary and partial differential equations are intractable for modern deterministic global solvers, particularly within large-scale nonlinear and mixed-integer nonlinear programming (MINLP) formulations. As a result, surrogate modeling approaches have gained increasing attention. However, conventional surrogate models typically rely on strong process-level assumptions, simplified physics, or extensive data generation, which limits their extrapolation capability, physical consistency, and reliability across feasible process operating regions. Physics-informed neural networks (PINNs) offer a promising alternative by embedding governing physics directly into the learning objective loss function, thereby reducing dependence on large, supervised training datasets while preserving governing physical laws. Existing PINNs implementations are typically formulated in a unit-specific or problem-specific manner. However, PINNs formulations lack modularity, transferability, and integration readiness for process modeling workflows. Therefore, a critical gap exists in the development of a highly generalizable and customizable PINNs framework applicable for diverse complex process unit operations. In this work, we develop a unified physics-informed deep learning framework for modeling complex process unit operations. The developed

framework ensures governing physical consistency while maintaining flexibility across diverse nonlinear systems. Numerical stability is achieved through structured normalization and consistent dimensional scaling, enabling stable training across wide operating domains. A modular representation-agnostic architecture allows flexible specification of input-output dimensional spaces, systematic enforcement of boundary and operating constraints, and adjustable coupling between physics-based and data-driven loss objectives. Further, the developed framework promotes transferability and scalability across diverse modeling tasks by avoiding problem-specific architectural redesigning. Beyond predictive accuracy, the framework facilitates seamless integration of physics-informed surrogates within broader hybrid modeling workflows. Further, we validate the framework performance through multiple case studies, demonstrating robustness, scalability, and reduced reformulation effort relative to conventional PINNs implementations. This work advances physics-guided deep learning toward a reusable computational infrastructure for AI-enabled process chemical process modeling and simulation.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



Optimal Solvent Mixture Screening with Graph Neural Networks

Yipei Zhao

University of Sheffield, School of Chemical, Materials and Biological Engineering

ABSTRACT

Solubility is a critical parameter governing the efficacy and stability of agrochemical active ingredients (AIs) formulated in solvent mixtures. Traditionally, the screening of efficient solvent mixtures heavily relies on empirical, trial-and-error experiments. The non-linear behaviours of complex solutes in multi-component solvent systems remain difficult to predict. Furthermore, while predictive machine learning models have shown success in predicting the interaction between multi-component solvents, the application of multi-component system solubility prediction is lacking in the agrochemical field. To address this gap, our study employs a Graph Neural Network (GNN) to model the behaviour of dissolving systems from both intramolecular and intermolecular perspectives. Rather than altering the underlying structural architecture, this work focuses on the domain-specific application and transferability of the model to a novel, high-standard experimental dataset of agrochemical

AIs and targeted solvent mixtures. By training the model exclusively on specialised agrochemical data, we bypass the limitations of generalised chemical databases and improve the model's predictive accuracy for our specific use case. This work highlights the critical importance of agrochemical-specific data application in applied machine learning for chemical engineering. Ultimately, we demonstrate how leveraging established GNN architectures alongside specialised datasets can provide a highly scalable, pre-screening tool that can drastically reduce the cost of experiments in the development of novel agrochemical formulations.

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>



A Deepsets-Guided Framework for Learning Job Priorities In Single-Machine Scheduling

Daniel Zhu

Princeton University, Chemical and Biological Engineering

ABSTRACT

Scheduling is a fundamental decision-making problem in chemical engineering as well as numerous other sectors, arising in manufacturing, energy systems, and supply chains. Many scheduling problems are NP-hard [1], meaning that even small problems are computationally hard to solve deterministically. As a result, existing exact and heuristic methods face trade-offs between scalability, solution quality, and generalizability. This work addresses these limitations through a hybrid machine learning–optimization framework for the single-machine total tardiness scheduling problem (SMTTP). We introduce the Deep-Sets-Guided Scheduling Framework (DGSF), a hybrid methodology that integrates data-driven priority learning with structured optimization for single-machine scheduling.

First, we propose a geometric instance classification rule that characterizes scheduling instances through aggregate structural parameters, enabling models trained on small instances to generalize to larger instances within the same structural class.

Second, we develop a modified DeepSets [2] machine learning (ML) architecture that processes variable-sized sets of jobs and produces job-aligned priority scores. The model combines job-level feature transformations with an attention-based aggregation mechanism to incorporate instance-level context, allowing priority estimation to depend jointly on job and instance characteristics. Input features are constructed to retain interpretability and include normalized processing, release and due times, slack-

based measures, and features derived from classical heuristics.

Third, we introduce a two-stage post-processing step. A fast local pairwise-swap heuristic improves the predicted sequence, which is then used to warm-start a neighborhood-restricted continuous-time mixed-integer programming formulation. By explicitly limiting the search space around the learned solution, this formulation achieves high-quality refinement while controlling combinatorial complexity.

Computational experiments indicate that the learned priority structure aligns with classical one-shot dispatching heuristics while improving solution quality. On instances with up to 120 jobs, DGSF consistently outperforms these heuristics, achieving optimality gaps of 3–6% compared to 26–31% for the best-performing one-shot heuristic. Furthermore, DGSF maintains high-quality solutions across different instance structures, time discretizations, and product catalogues.

REFERENCES

1. Du J, Leung JYT. Minimizing Total Tardiness on One Machine is NP-Hard. *Mathematics of Operations Research* 15:483–495 (1990).
<https://doi.org/10.1287/moor.15.3.483>
2. Zaheer M, Kottur S, Ravanbakhsh S, Póczos B, Salakhutdinov R, Smola A. Deep sets. *Cornell University*. (2017)
<https://doi.org/10.48550/arxiv.1703.06114>



Editorial Notes

Follow the DOI of each individual article for the latest updates, links to supplementary material and related works, and high-resolution versions of the article.



PSEcommunity.org

PSE Press • Hamilton • Notre Dame