Pyomo.DoE: Facilitating collaborations using model-based design of experiments in the Pyomo ecosystem

Alexander W. Dowling

Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, USA

adowling@nd.edu

Abstract

Identifying impactful solutions for global grand challenges often requires concerted research efforts that span molecular, material, device, systems, and infrastructure length-scales and transcend disciplines. In this contribution, we argue predictive multiscale mathematical models, often grounded in scientific theories, provide principled approaches to realize molecular-to-systems engineering. Using a membrane example, we present a tutorial on science-based data analytics including nonlinear regression, practical identifiability, parameter uncertainty quantification, and model-based design of experiments (MBDoE). Next, we provide a tutorial on how to use Pyomo.DoE to perform MBDoE for parameter precision optimization in the open-source Pyomo ecosystem. We conclude the best practices for using models at advance collaborations across disciplines (i.e., outside process systems engineering).

**Keywords**: data science, nonlinear regression, digital twins, membrane science, software engineering

* 1. Mathematical Modeling Facilitates Collaboration

Sustainability is considered a “wicked problem” (Lönngren and Van Poeck 2021) because of the complex interdependencies between social, natural, and engineered systems. These problems thus require interdisciplinary teams and new modeling strategies to consider complex interactions across molecular, material, device, process, and infrastructure scales (Eugene, Phillip, and Dowling 2019). Informed by our recent in collaborations ranging from water treatment (Eugene, Phillip, and Dowling 2021), refrigerant recycling (Befort et al. 2023), and additive manufacturing (K. Wang et al. 2023), we argue predictive mathematical models are a critical important tool for bridging length and timescales as well as disciplines.

* 1. Tutorial on Science-based Data Analytics

Through a short tutorial on membrane modeling, we show systematic steps for building, training, and validating mathematical models, illustrated in Figure 1.



**Figure 1.** Science-based modeling workflow adapted from literature (Franceschini and Macchietto 2008; J. Wang and Dowling 2022) in the Pyomo ecosystem. Starting with preliminary data and prior knowledge, the modeler postulates one or more mathematical models grounded in engineering science. Physically meaningful parameters in these models are estimated via nonlinear optimization. Sensitivity and uncertainty analyses determine which parameters are estimable from the available data. Optionally, statistical information criteria can facilitate model selection. Finally, model-based design of experiments recommends the next most valuable measurements to discriminate between candidate models or increase parameter precision or both.

Diafiltration is a membrane staging technique to efficiently separate charged molecules, with current applications focused on valuable products such as proteins and buffer exchange (Ouimet et al. 2022). Motivated by the ongoing global transportation electrification, we contemplate optimizing diafiltration cascades for lithium-ion battery recycling (Harper et al. 2019). We previously showed how molecular-to-systems optimization using science-based mathematical models facilitates systematic evaluation of trade-offs in membrane system design and sets quantitative materials property targets for new applications (Wamble et al. 2022). However, this analysis was predicated on mathematical models for membrane transport.

Motivated by this goal of designing new membrane materials and systems using predictive science-based models, we share how the synergies between data analytics (e.g., dynamic modeling, nonlinear parameter, Fisher information analysis) and transport experiments (e.g., sensor design for time-series measurements) led to the new Diafiltration Apparatus for high-Throughput Analysis (DATA) technique for higher throughput membrane characterization (Ouimet et al. 2022). Using the DATA framework as a motivating example, we present a tutorial on science-based data analytics:

1. Postulating a dynamic model to describe changing concentrations as rejections as a function of time,
2. Using weighted nonlinear regression to balance prediction errors across three experimentally measured quantities,
3. Performing estimability analysis to demonstrate the importance of diafiltration experiments which sweep a large concentration range,
4. Using model-based design of experiments (MBDoE) to identify the operating regions with the most information (Liu et al. 2022).

Our tutorial shows how these data science tools directly inform apparatus design and experiment optimization, ultimately leading to predictive science-based mathematical models with quantified parametric uncertainty.

* 1. Pyomo.DoE and Model-based Design of Experiments (MBDoE)

Open-source tools, including Pyomo and Pyomo.DoE (J. Wang and Dowling 2022), enable model-based data analysis in Python while exploiting state-of-the-art nonlinear optimization solvers. In summary, MBDoE first sequentially determines the best experiments to discriminate between a set of candidate mathematical models (Olofsson et al. 2019). Then, in parameter precision mode, MBDoE sequentially recommends experiments to reduce parameter uncertainties (Franceschini and Macchietto 2008). Recently MBDoE algorithms have been combined with automated experiments (Pankajakshan et al. 2023) to establish digital twins (Kuchemüller, Pörtner, and Möller 2021). Pyomo.DoE helps Pyomo users automatically formulate and solve MBDoE for parameter precision optimization problems. In this contribution, we provide a tutorial for specifying models in Pyomo.DoE. We conclude by summarizing ongoing development activities for Pyomo.DoE.

* 1. Conclusions and Key Contributions

Through the membrane example, we show how science-based data analystics, especially MBDoE, provides a framework to characterize transport properties of membranes 10x faster than conventional experiments. Moreover, this example provides a tutorial on key data analytics tools.

While MBDoE is a powerful paradigm, it requires expertise in mathematical modeling, optimization, and statitics. To reduce these barriers, we are developing Pyomo.DoE, a package for MBDoE within the popular Pyomo modeling ecosystem.

References

Befort, Bridgette J., Alejandro Garciadiego, Jialu Wang, Ke Wang, Gabriela Franco, Edward J. Maginn, and Alexander W. Dowling. 2023. “Data Science for Thermodynamic Modeling: Case Study for Ionic Liquid and Hydrofluorocarbon Refrigerant Mixtures.” *Fluid Phase Equilibria* 572 (September): 113833. https://doi.org/10.1016/j.fluid.2023.113833.

Eugene, Elvis A., William A. Phillip, and Alexander W. Dowling. 2019. “Data Science-Enabled Molecular-to-Systems Engineering for Sustainable Water Treatment.” *Current Opinion in Chemical Engineering*, Energy, Environment & Sustainability: Sustainability Modeling ● Reaction engineering and catalysis: Green Reaction Engineering, 26 (December): 122–30. https://doi.org/10.1016/j.coche.2019.10.002.

———. 2021. “Material Property Targets to Enable Adsorptive Water Treatment and Resource Recovery Systems.” *ACS ES&T Engineering* 1 (8): 1171–82. https://doi.org/10.1021/acsestengg.0c00046.

Franceschini, Gaia, and Sandro Macchietto. 2008. “Model-Based Design of Experiments for Parameter Precision: State of the Art.” *Chemical Engineering Science*, Model-Based Experimental Analysis, 63 (19): 4846–72. https://doi.org/10.1016/j.ces.2007.11.034.

Harper, Gavin, Roberto Sommerville, Emma Kendrick, Laura Driscoll, Peter Slater, Rustam Stolkin, Allan Walton, et al. 2019. “Recycling Lithium-Ion Batteries from Electric Vehicles.” *Nature* 575 (7781): 75–86. https://doi.org/10.1038/s41586-019-1682-5.

Kuchemüller, Kim B., Ralf Pörtner, and Johannes Möller. 2021. “Digital Twins and Their Role in Model-Assisted Design of Experiments.” In *Digital Twins: Applications to the Design and Optimization of Bioprocesses*, edited by Christoph Herwig, Ralf Pörtner, and Johannes Möller, 29–61. Advances in Biochemical Engineering/Biotechnology. Cham: Springer International Publishing. https://doi.org/10.1007/10\_2020\_136.

Liu, Xinhong, Jialu Wang, Jonathan A. Ouimet, William A. Phillip, and Alexander W. Dowling. 2022. “Membrane Characterization with Model-Based Design of Experiments.” In *Computer Aided Chemical Engineering*, edited by Yoshiyuki Yamashita and Manabu Kano, 49:859–64. 14 International Symposium on Process Systems Engineering. Elsevier. https://doi.org/10.1016/B978-0-323-85159-6.50143-3.

Lönngren, Johanna, and Katrien Van Poeck. 2021. “Wicked Problems: A Mapping Review of the Literature.” *International Journal of Sustainable Development & World Ecology* 28 (6): 481–502. https://doi.org/10.1080/13504509.2020.1859415.

Olofsson, Simon, Lukas Hebing, Sebastian Niedenführ, Marc Peter Deisenroth, and Ruth Misener. 2019. “GPdoemd: A Python Package for Design of Experiments for Model Discrimination.” *Computers & Chemical Engineering* 125 (June): 54–70. https://doi.org/10.1016/j.compchemeng.2019.03.010.

Ouimet, Jonathan A., Xinhong Liu, David J. Brown, Elvis A. Eugene, Tylar Popps, Zachary W. Muetzel, Alexander W. Dowling, and William A. Phillip. 2022. “DATA: Diafiltration Apparatus for High-Throughput Analysis.” *Journal of Membrane Science* 641 (January): 119743. https://doi.org/10.1016/j.memsci.2021.119743.

Pankajakshan, Arun, Solomon Gajere Bawa, Asterios Gavriilidis, and Federico Galvanin. 2023. “Autonomous Kinetic Model Identification Using Optimal Experimental Design and Retrospective Data Analysis: Methane Complete Oxidation as a Case Study.” *Reaction Chemistry & Engineering* 8 (12): 3000–3017. https://doi.org/10.1039/D3RE00156C.

Wamble, Noah P., Elvis A. Eugene, William A. Phillip, and Alexander W. Dowling. 2022. “Optimal Diafiltration Membrane Cascades Enable Green Recycling of Spent Lithium-Ion Batteries.” *ACS Sustainable Chemistry & Engineering* 10 (37): 12207–25. https://doi.org/10.1021/acssuschemeng.2c02862.

Wang, Jialu, and Alexander W. Dowling. 2022. “Pyomo.DOE: An Open-Source Package for Model-Based Design of Experiments in Python.” *AIChE Journal* 68 (12): e17813. https://doi.org/10.1002/aic.17813.

Wang, Ke, Minxiang Zeng, Jialu Wang, Wenjie Shang, Yanliang Zhang, Tengfei Luo, and Alexander W. Dowling. 2023. “When Physics-Informed Data Analytics Outperforms Black-Box Machine Learning: A Case Study in Thickness Control for Additive Manufacturing.” *Digital Chemical Engineering* 6 (March): 100076. https://doi.org/10.1016/j.dche.2022.100076.