INTRODUCTION

TO A SPECIAL SECTION ON

Teaching Data Science

Martha A. Grover[1] and Phillip R. Westmoreland[2]

1. Georgia Institute of Technology • Atlanta, GA 30332

2. North Carolina State University • Raleigh, NC 27695

BACKGROUND

Analytics and machine learning are now seen by industry and academics as key knowledge that all students should have. Increasingly, industry feedback to academia highlights the importance of data science in chemical engineering practice, including the need for additional emphasis on data science in the chemical engineering curriculum. There are many ways to do so, and examples help us understand that range. This special section of *Chemical Engineering Education* shares the experiences of chemical engineering educators in teaching data science in a wide variety of contexts and audiences, including the classroom setting for residential students, but also the training of undergraduate researchers and of working professional in the chemical industry. Modes of delivery include both in-person and remote.

TOPICAL CONFERENCE SESSIONS

Teaching Data Science to Students and Teachers was a set of three sessions in the 2021 AIChE Annual Meeting in Boston, featured in the topical conferences "Bridging the Skill Gap in Chemical Engineering," "Next-Gen Manufacturing," and "Molecular and Materials Data Science." There were two related invited sessions on process analytics and machine learning, *Practical Application of Process Data Analytics and Machine Learning* and *Academia-Industry Partnership: The Undergraduate Curriculum*, while the *Teaching Data Science to Students and Teachers* sessions featured contributed papers with the overall description:

> There is plenty of enthusiasm about the future of "data science" being an essential expertise for chemical engineers. Analytics and machine learning build on statistics and use it to identify and quantify correlated behaviors, enabling decision-making. That's powerful, but both students and their instructors need to know about more than means and standard deviations. Participants in this session will share their experiences with teaching statistics and beyond, aiming to benefit education and other educators.

Martha A. Grover *is a Professor in the School of Chemical & Biomolecular Engineering at Georgia Tech. She is the chair of the CAST division of AIChE and the vice-president of the educational nonprofit CACHE Corporation. Her research program is dedicated to understanding, modeling, and engineering the self-assembly of atoms and small molecules to create larger-scale structures and complex functionality. Her approach draws on process systems engineering, combining modeling and experiments in applications dominated by kinetics, including surface deposition, crystal growth, polymer reaction engineering, and colloidal assembly. ORCID: 0000-0002-7036-776X*

Phillip R. Westmoreland *is a Professor in the Department of Chemical and Biomolecular Engineering at North Carolina State University, engaged in fundamental and applied reaction kinetics through experiments and computational chemistry. Major contributions include high-temperature desulfurization, discovery of ROH-catalyzed mechanism for biomass conversion to bio-oils, and flame chemistry of aromatics growth and fluorocarbon chemistry. He is a CACHE trustee, was 2013 President of AIChE, and was founding Chair of AIChE's Computational Molecular Science and Engineering Forum.*

Due to the number of outstanding submissions, three sessions were created on this topic, two in-person sessions and one virtual session, with a total of 16 presentations. In addition to co-chairing these sessions, we have partnered with *Chemical Engineering Education* to create this special section of contributions from the sessions.

PAPERS

Three papers in the issue describe approaches to classroom teaching in the "traditional" campus context. "Chemical Engineering Analysis through Systematic Optimization" by Xie and Davis describe the importance in teaching an undergraduate course of straightforward modeling exercises as well as case studies. "Teaching Process Data Analytics and Machine Learning: Teaching Process Analytics at MIT" by Hong, Sun, and Braatz describes two undergraduate and graduate courses, including students from outside chemical engineering, and emphasizing real data sets from case studies. "Teaching Artificial Intelligence to Chemical Engineers: Experience from a 35-year-old Course" by Venkatasubramanian incorporates a historical perspective on the field and the integration of symbolic artificial intelligence.

The remaining three papers feature teaching in alternative contexts. "From Molecular- to Plant-Scale Computational-Engineering Design: Applied Training Spanning Scales" by English articulates a vision for integration of molecular simulation into the chemical engineering toolkit. "Online Graduate Certificate in Data Science for the Chemical Industry," by Medford, Boukouvala, et al., describe an asynchronous program aimed toward working professionals in the chemical industry. "Effective Laboratory Education with TEXTILE: Tutorials in EXperimentalisT Interactive LEarning" by Helmbrecht and Nance presents a curriculum for teaching image processing skills to undergraduate researchers, which can be accomplished online or in-person.

CONCLUSION

Most chemical engineering students are exposed to probability and statistics in their undergraduate or even high-school educations. Data science can build on that foundation to aid data-based decision-making. Analytics itself is use-oriented statistical analysis, often analyzing very large data sets that first must be cleaned and put into suitable computer-friendly formats. Likewise, machine learning draws on such data sets and traditional distribution models and correlation methods in new ways, creating representations that help interpret complicated information to shape insights and effective actions.

While the data-science curriculum in chemical engineering continues to evolve, the collection of papers in this special issue highlights some key features that can be considered by programs developing their own data science curriculum. Incorporation of real data sets into the curriculum is one cross-cutting feature. However, content will be tailored depending on the student background and preparation, including undergraduate and graduate students and working professionals. Adoption of modern tools for content delivery has been accelerated during the pandemic, which provides new opportunities in the development of a data science curriculum in chemical engineering.

Many more contributions on this topic will appear in future ASEE and AIChE meetings and publications. In the near term, we note FOPAM23, the conference on "Foundations of Process Analytics and Machine learning" that will be held on July 30-August 3, 2023, at the University of California, Davis https://fopam.cache.org/. Researchers and educators from industry and academia will discuss the current status and future directions of data analytics and machine learning in the process industries. The format of the conference is 3 1/2 days of presentations and discussions, preceded by an optional 1 1/2-day workshop.

ACKNOWLEDGMENTS

A special thanks to Allison Godwin, CEE Assistant Editor and Associate Professor of Engineering Education and of Chemical Engineering at Purdue University, for serving as the Special Issues Editor for this section. Her important and timely contributions on the various logistical processes associated with working on this special Teaching Data Science section are greatly appreciated. ❐