*Proceedings of the 5th European International Conference on Industrial Engineering and Operations Management*
*Rome, Italy, July 26-28, 2022*

# Comparison of the Application of Different Machine Learning Outlier Detection Methods on Actual Chemical Process Plant Data

**Mohammed Alkatheri**
Department of Chemical Engineering,
University of Waterloo, Waterloo, Ontario, N2L 3G, Canada
mohammed.alkatheri@uwaterloo.ca

**Ali Almansoori**
Professor of Chemical Engineering and Associate Provost for Education
Khalifa University of Science and Technology
Abu Dhabi, UAE
ali.almansoori@ku.ac.ae

**Peter L. Douglas**
Professor of Department of Chemical Engineering,
University of Waterloo, Waterloo, Ontario, N2L 3G, Canada
pdouglas@uwaterloo.ca

**Ali Elkamel**
Professor of Department of Chemical Engineering,
University of Waterloo, Waterloo, Ontario, N2L 3G, Canada
aelkamel@uwaterloo.ca

## Abstract

Recent advancements in supervised machine learning tools have proven their capability to act as accurate approximation surrogate models for complex the chemical production processes. In this approach, complex unit models are replaced with surrogate models built from actual chemical plant Nevertheless, real data should be handled with caution as it isn't devoid of missing points, outliers, and faulty measurement, and using them without pre-processing could lead to inaccurate prediction models. Moreover, it is well-known that ideal real data without any outliers is almost nonexistence. Hence, cleaning data from outliers is very important step in data-driven modeling development Therefore, in this study different machine learning outlier detection method are implemented and compared to clean actual plant data before they are introduced to the data-driven surrogate models. Outliers are observations that do not follow bulk pattern of the data points and are unlikely observation of data. it is worth mentioning that identifying outliers by simple inspection and visualizing data set is challenging. There are different methods that can be used to identify outliers some of these methods are based on univariate statistical methods (Interquartile Range Method) and the others are based on unsupervised machine learning methods (Local outlier Factor, Isolation Forest, and One Class Support Vector Machine) The performances of these outlier detection methods on understudy data sets, are evaluated using linear regression that is used to predict certain process variables. Results show that removing outliers using these outlier detection

methods before training the surrogate models can enhance the prediction accuracy of the machine learning approximation models.

## Keywords (12 font)
Unsupervised outlier detection methods, application of machine learning models, data cleaning. (10 font)

## Acknowledgements
Add acknowledgement if need

## Biography / Biographies

**Mohammed Alkatheri** holds a BSc degree in Chemical Engineering from United Arab Emirates University, and MSc degree in Chemical Engineering from the Khalifa University in Abu Dhabi. In December 2021, he received a PhD degree from Chemical Engineering at University of Waterloo. His PhD research is focusing on the application and integration of big-data tools (i.e. Artificial Intelligence and Machine Learning) in chemical process optimization and process system engineering. The scope of his PhD project is to address the challenges associated with chemical engineering process design and operation, namely, uncertainty handling, parameter estimation and unit process equation complexity.

**Ali Almansoori** is a Professor of Chemical Engineering at Khalifa University of Science and Technology in Abu Dhabi. During his profession, Dr. Almansoori held several administrative positions including: Coordinator of President's Duties, Dean of Engineering, and Chair and Deputy Chair of the Chemical Engineering Department. He also was the Interim Senior Vice President for Academic Affairs during the merge between Petroleum Institute, Masdar Institute, and Khalifa University of Science, Technology, and Research. His main research interest is in the area of Process Systems Engineering, with focus on energy systems design, simulation, modeling, and optimization. He also conducts general research in the area of renewable energy and fuel cell technology with applications to the oil and gas industry. Dr. Almansoori has published numerous articles in renowned refereed journals and conference proceedings. He also delivered several presentations in international conferences and is the author of a few book chapters. Furthermore, he serves as a reviewer for reputable international journals in the area of energy and process systems. Dr. Almansoori has received a number of educational and research awards, including the Sheikh Mohammed Bin Rashid Medal for Scientific Excellence in the UAE for 2019, PI 2014–2015 Research & Scholarship Award for Senior Faculty, two Best Track Paper Awards in IEOM'15 & GSR'15 Conferences, ADNOC 2013 R&D Science Lantern Faculty Award, PI 2012 Excellence in Academic Advising Award, and Sheikh Rashid Award for Scientific Outstanding Performance in 2008.

**Peter Douglas** is the Associate Dean of Engineering (Undergraduate Studies) and a Professor of Chemical Engineering at the University of Waterloo. He was previously the Director of the University of Waterloo United Arab Emirates Campus in Dubai from 2009 to 2013, the Associate Dean of Engineering (Computing), and the Associate Dean of Engineering (Graduate Studies). Professor Douglas was a founding member of WISE the Waterloo Institute for Sustainable Energy at UWaterloo. His primary research area of interest is in the development and application of PSE technology to industrial processes including process modelling, simulation, control and optimization. He is currently working on simulation and optimization issues related to the mitigation and capture of carbon dioxide from large scale emitters. Professor Douglas has consulted on a world-wide basis for many clients and has worked in Canada, Australia, Malaysia, Thailand, the UAE. Additionally, he is a co-inventor of the Dryer Master online measurement and control systems for the food processing industry; such systems are finding widespread use in Canada, USA, Europe and Asia. In addition to his research work, Professor Douglas has co-authored more than 200 related research publications and has supervised more than 80 postgraduate students.

**Ali Elkamel** is a Professor in the Department of Chemical Engineering and is cross-appointed to Systems Design Engineering. He holds a BSc in Chemical Engineering and BSc in Mathematics from Colorado School of Mines, MS in Chemical Engineering from the University of Colorado-Boulder, and PhD in Chemical Engineering from Purdue University – West Lafayette, Indiana. The goal of his research program is to develop theory and applications for

process systems engineering. The applications are focused on planning and scheduling of process operations, energy production, pollution monitoring and control, waste minimization, carbon management, sustainable operations, molecular design, and product formulation. Among his accomplishments are the Outstanding Faculty Award, the Best teacher award, the IEOM (Industrial engineering and Operations Management) Outstanding Service and Distinguished Educator Award, UAE MBR Academy of Scientists, The Engineering Research Excellence Award, and Excellence in Graduate Supervision Award. He has been on the program and organization committees of many international conferences. Prof. Elkamel has a strong track record of research excellence. He published over 330 journal articles, 145 proceedings, and 33 book chapters, and has been an invited speaker on numerous occasions at academic institutions throughout the world and at national and international conferences. He has also written 5 books, including "Environmentally Conscious Fossil Energy Production", "Planning of Refinery and Petrochemical Operations", and "Electric Vehicles in Energy Systems: Modelling, Integration, Analysis, and Optimization."