

# Advanced Process Data Analytics

WeiKe Sun

Submitted to the Department of Chemical Engineering on May 1, 2020  
In partial fulfillment of the requirements for the degree of  
Doctor of Philosophy

## Abstract

Process data analytics is the application of statistics and related mathematical tools to data in order to understand, develop, and improve manufacturing processes. There have been growing opportunities in process data analytics because of the advances in machine learning as well as the technologies for data collection and storage. However, challenges are encountered because of the complexities of manufacturing processes, which often require advanced analytical methods. In this thesis, two areas of application are considered. One is the construction of predictive models that are useful for process design, optimization, and control. The other area of application is process monitoring to improve process efficiency and safety.

In the first area of study, a robust and automated approach for method selection and model construction is developed for predictive modeling. Two common challenges when building data-driven process models are addressed: the high diversity in data quality and how to select from a wide variety of methods. The proposed approach combines best practices with data interrogation to facilitate consistent application and continuous improvement of tools and decision making.

The second area of study focuses on process monitoring for complex manufacturing systems, which includes fault detection, identification, and classification. Four sets of algorithms are developed to address the limitations of traditional monitoring methods. The first set provides the optimal strategy for Gaussian linear processes, including deep understanding of the process monitoring structure and optimal fault detection based on a probabilistic formulation. The second set aims at building a self-learning fault detection system for changing normal operating conditions. The third set is developed based on information-theoretic learning to address limitations of second-order statistical learning for both fault detection and classification. The fourth set tackles the problem of nonlinear and dynamic process monitoring.

The proposed methodologies and algorithms are tested on several case studies where the value of advanced process data analytics is demonstrated.

Thesis Supervisor: Richard D. Braatz

Title: Edwin R. Gilliland Professor in Chemical Engineering