Guest Editorial: Data-Driven Management of Complex Systems Through Plant-Wide Performance Supervision

I. Introduction

TODAY, massive amounts of data are continuously being produced by social and industrial activities. Consequently, data-driven techniques have received considerable attention both in industry and academia in recent years, aiding scientists to manage and interpret the available data. The reasons behind such popularity of data-driven techniques are twofold. On the one hand, advanced data processing and information acquisition technologies have been developed to the extent that large amounts of data in different forms are available for big data analysis from descriptive to prescriptive. On the other hand, with the help of machine learning methodologies, the supervision and management systems can provide effective decisions for plant-wide optimal performance. Compared to the conventional model-based techniques, the data-driven ones can not only save the costly modeling procedures but also extract valuable information from available process data for real-time analysis and management. However, there are many complex and challenging problems in the data-driven supervision and management techniques, such as data-driven supervision on the safety, security, and robustness, as well as the performance-supervised management and their distributed designs.

This special issue targets the recent results, trends, and practical developments in the data-driven methodologies of plant-wide performance supervision and management for complex systems, especially those related to process monitoring and machine learning activities with their industrial applications. To this end, 14 papers have been selected for the special issue after rigorous peer review, in which every submission received a double-blind review by at least three external reviewers. It is expected by the editorial team that this special issue inspires researchers for further theoretical and practical outcomes. The investigations and discussions of the vital issues, challenges, and possible future trends in this subject are very much appreciated.

II. Overview of the Accepted Papers

The developments in computer science and data networking technologies have inspired many studies on data-driven (or model-free) process monitoring and control methodologies, which have become a popular trend both in academic and industrial arenas. Seeking to bridge the gap between model-based and data-driven techniques and to develop a data-driven process monitoring and management framework in an integrated manner, in "Real-time implementation of plug-and-play process monitoring and control on an experimental three-tank system [item 1) in the Appendix]," the authors implemented a co-called plug-and-play process monitoring and control architecture on a three-tank laboratory setup. In general, this implementation not only avoids the modification of the predesigned control system and realizes the process monitoring and performance optimization in a "plug-and-play" manner but also provides architectural design reliability and flexibility for the industrial application of advanced monitoring and control methodologies.

A distributed parallel modeling and monitoring framework has been proposed in "Industrial big data modeling and monitoring framework for plant wide processes [item 2) inthe Appendix]," where the distributed framework contains two layers. One layer is the spatial distributed modeling and hierarchical monitoring for multiple operating units of the plant-wide process, the other is the distributed parallel modeling for big process data with various features. To achieve distributed process monitoring, distributed parallel mixture probabilistic latent variable model is proposed based on the stochastic variational inference algorithm and the parameter server architecture. The authors have constructed the multilevel monitoring indexes and fault contribution indexes for the visualization of the fault detection and diagnosis [item 2) in the Appendix].

Considering the incomplete measurements and the incipient faults in the industrial processes, in "Low rank characteristic and temporal correlation analytics for incipient industrial fault detection with missing data [item 3) in the Appendix]," the authors applied the low-rank matrix decomposition method to recover the missing measurements while reducing the effects of the environmental noise. Since the low-rank matrix decomposition leads to a low-rank component and a sparse component which, respectively, indicate the main variance and the residual information of the industrial process data, the authors developed a canonical variate analysis model to extract the temporal correlation of the process data. Based on the obtained features, three monitoring statistics are established to reflect the operation status of the online sample, in which one statistic is used to measure the static deviation of this sample, and the other two indices evaluate the dissimilarity between the past and future canonical variates.

Different from the traditional feature extraction methods which either consider the common scores or common weightings between different modes of the industrial process, in "A

Digital Object Identifier 10.1109/TII.2020.3023259

novel feature extraction-based process monitoring method for multimode processes with common features and its applications to a rolling process [item 4) in the Appendix]," the authors proposed a two-level feature extraction-based multimode process monitoring method for industrial processes with hidden common features. In the first level, the common features behind multiple operating modes were extracted and presented using a tensor decomposition-based method. In the second level, the specific features of each mode are extracted by the independent component analysis method. Based on the two-level features, the authors developed a moving window Kullback–Leibler divergence-based monitoring statistic to detect the changes in the two kinds of features.

To achieve quality-related root cause diagnosis for nonlinear processes, in "Quality-related root cause diagnosis based on orthogonal kernel principal component regression and transfer entropy [item 5) in the Appendix]," the authors proposed an orthogonal kernel principal component regression model for the orthogonal decomposition of feature space of the process data, so that the quality-related and -unrelated faults can be decoupled and separately detected in the subspaces of the measurements. Comparing the traditional methods, a kernel sample equivalence replacement approach is proposed to reduce the computational complexity of fault diagnosis. Furthermore, the authors developed a transfer entropy algorithm to analyze the causality between the diagnosed candidate faulty variables in order to find the accurate root cause of the fault.

A novel data-driven method named multisubspace orthogonal canonical correlation analysis is proposed in "Multi subspace orthogonal canonical correlation analysis for quality related plant wide process monitoring [item 6) in the Appendix]," which not only indicates whether the fault occurs or not but also can judge if the fault affects product quality in real time. The authors first divided the original process variables space into four subspaces, where the analysis complexity was reduced and the monitoring model was constructed. Then, the developed orthogonal canonical correlation analysis was conducted on process data and quality data for correlation feature extraction. Afterwards, a total of six monitoring statistics were constructed and integrated to four statistics with physical interpretation via a Bayesian fusion strategy.

Due to the inherent dynamics and uncertainty associated with the industrial operations, the chance constrained dynamic optimization problems are hard to cope with in real practice. In "Hybrid intelligence assisted sample average approximation method for chance constrained dynamic optimization [item 7) in the Appendix]," the authors proposed a novel chance constrained dynamic optimization method, where an adaptive sample average approximation method, a control vector parameterization method, and a state constraint handling strategy were integrated. To deal with the problem that a single algorithm is insufficient for an accurate and efficient solution of the chance constrained dynamic optimization problem, a hybrid intelligent optimization algorithm was introduced to realize a global and efficient optimization performance, which combined the global search ability of state transition algorithm and the fast convergence ability of gradient-based method.

In "Double layer distributed monitoring based on sequential correlation information for large-scale industrial processes in dynamic and static states [item 8) in the Appendix]," the authors considered large-scale, static, and dynamic characteristics in industrial processes and proposed a double layer distributed monitoring method based on multiblock slow feature analysis and independent component analysis to handle the static and dynamic characteristics. In the first layer, state characteristics were taken into consideration. To explore the state characteristic of each variable, the sequential mutual information between variables was calculated, which explored the high-order correlations between variables. The correlations between variables were taken as the rule of the second layer division. The k-means clustering method was used to decompose the static and dynamic blocks into subblocks and the monitoring results in each subblock are integrated by Bayesian inference.

To improve the performance of the classic archetypal analysis methods for anomaly detection tasks on nonconvex data sets, in "Anon-convex archetypal analysis for one-class classification based anomaly detection in cyber-physical systems [item 9) in the Appendix]," the authors proposed a nonconvex archetypal analysis approach to one-class classification, which uses the nonconvex hull of normal behavior data to build a natural boundary between normal and abnormal data points. Detecting anomalies is then considered as determining if given new points are inside or outside a nonconvex hull. The proposed algorithm combines the advantages of random projection to dimensional reduction and the advantages of AdaBoost to avoiding overfitting, which does not need expert knowledge to determine thresholds for anomaly detection.

In "A data-driven health monitoring method using multiobjective optimization and stacked autoencoder based health indicator [item 10) in the Appendix]," the authors proposed an improved nondominated sorting genetic algorithm-II to tackle the problems of multiple oscillations and poor distribution characteristics during convergence. By specifying the target training data of the reconstructed model and optimizing the index, health indicators were used to characterize the specific state and distinguish other states, while limiting the number of selected features to reduce model complexity. Based on that, a data-driven health monitoring method was developed which combined the improved nondominated sorting genetic algorithm-II and a deep neural network.

To handle the limitation of the conventional autoencoder which ignored the relationships among parameters, in "A conditional convolutional autoencoder based method for monitoring wind turbine blade breakages [item 11) in the Appendix]," the authors presented a monitoring method based on conditional convolutional autoencoder for the detection of the impending wind turbine blade breakages. The proposed methods were of twofold to identify the wind turbine blade breakages. First, a novel conditional convolutional autoencoder was developed to derive reconstruction errors, which reflected the changes of system dynamics caused by impending blade breakages. Second, a statistical process control principle was applied to develop the boundaries for triggering blade breakage alarms based on the reconstruction errors, while an exponentially weighted moving average control chart was adopted as a monitoring tool.

In "A data-driven approach of product quality prediction for complex production systems [item 12) in the Appendix]," the authors proposed a data-driven approach based on a parallel deep factorization machine mode as a soft sensor. The deep factorization machine consisted of a factorization machine component and a deep component, which was a supervised neural network-based model in the field of recommendation systems. The proposed method was developed by a label broadcasting method for labeled sample augmentation, a data binning method for data discretization, and a semisupervised deep factorization machine model in parallel. Based on the modified deep factorization machine model, quality information was separately extracted from different components while the high- and low-dimensional features were obtained. Manifold regularization was embedded into the back-propagation algorithm where the unlabeled samples issue was resolved.

Due to the implementation of the attention mechanism, the important information in the process data can be captured. Based on this, in "Gated dual attention unit neural networks for remaining useful life prediction of rolling bearings [item 13) in the Appendix]," the authors designed a new attention gate according to the structural features of the gated recurrent unit. The designed attention gate later combined with another attention gate which was effectively integrated with the interior of the gated recurrent unit. The newly proposed gate had the ability to focus on the output information of the reset and update gates, while the other attention gate was applied to consider the information embedded in the input data and the hidden state at the previous time instant. In this way, the two attention gates enhanced the ability of the gated recurrent unit and reduced the computational load. Using the calculated root mean square value of the bearing vibration signal as the health indicator, the authors proposed a remaining useful life prediction approach for rolling bearings.

In "Deep double supervised embedding neural network enhancing class separation for visual high dimensional industrial process monitoring [item 14) in the Appendix]," the authors proposed a deep double-supervised embedding neural network for visualizing high-dimensional industrial data, which can effectively separate the different fault data in a two-dimensional map for process monitoring. The proposed deep double-supervised embedding neural network consisted of two supervised deep neural networks: A deep class centered uniform distribution neural network which mapped the high-dimensional industrial data to a new feature space and promoted a good and separable situation for subsequent visualization procedures, and a deep supervised *t*-stochastic neighbor embedding neural network which mapped these high-dimensional features onto a two-dimensional space. The authors combined the deep double-supervised embedding neural network with local outlier factor and *k*-nearest neighbor methods for visual process monitoring.

In order to fully integrate the information among process data, in "Multitask-based temporal-channelwise CNN for parameter prediction of two-phase flows[item 15) in the Appendix]," the authors proposed a deep learning-based neural network, namely, the multitask-based temporal-channelwise convolutional neural network, for gas void fraction prediction. The proposed network took into account both temporal dependence and channel connection. The multitask learning methods were adopted in the proposed network to improve the efficiency using redundant obtained features. As a key parameter in the gas-liquid two-phase flow, the gas void fraction was predicted by the proposed deep learning-based soft measure method.

Acknowledgment

The Guest Editors would like to thank all the authors who submitted their valuable contributions and anonymous reviewers who took part in the review process. Also, they would like to extend their appreciation to Prof. R. C. Luo, the Editor-in-Chief of the IEEE Transactions On Industrial Informatics, for providing the opportunity to organize this Special Issue and the Editorial Staff of IEEE Transactions On Industrial Informatics for their outstanding support.

Okyay Kaynak, *Guest Editor*
Department of Electrical and Electronic Engineering
Boğaziçi University
Istanbul 34342, Turkey

Steven Ding, *Guest Editor*
Institute for Automatic Control and Complex Systems
University of Duisburg-Essen
47047 Duisburg, Germany

Ahmet Palazoglu, *Guest Editor*
Department of Chemical Engineering
University of California
Davis, CA 95616 USA

Hao Luo, *Guest Editor*
School of Astronautics
Harbin Institute of Technology
Harbin, 150001, China

Appendix
Related Work

1) M. Huo, H. Luo, X. Wang, Z. Yang, and O. Kaynak, "Real-time implementation of plug-and-play process monitoring and control on an experimental three-tank system," *IEEE Trans. Ind. Informat.*, to be published.
2) L. Yao and Z. Ge, "Industrial big data modeling and monitoring framework for plant-wide processes," *IEEE Trans. Ind. Informat.*, to be published.
3) W. Yu and C. Zhao, "Low rank characteristic and temporal correlation analytics for incipient industrial fault detection with missing data," *IEEE Trans. Ind. Informat.*, to be published.

4) K. Zhang, K. Peng, S. Zhao, and F. Wang, “A novel feature extraction-based process monitoring method for multimode processes with common features and its applications to a rolling process,” *IEEE Trans. Ind. Informat.*, to be published.
5) J. Jiao, W. Zhen, W. Zhu, and G. Wang, “ Quality-related root cause diagnosis based on orthogonal kernel principal component regression and transfer entropy,” *IEEE Trans. Ind. Informat.*, to be published.
6) B. Song, H. Shi, S. Tan, and Y. Tao, “Multi subspace orthogonal canonical correlation analysis for quality related plant wide process monitoring,” *IEEE Trans. Ind. Informat.*, to be published.
7) X. Zhou, X. Wang, T. Huang, and C. Yang, “Hybrid intelligence assisted sample average approximation method for chance constrained dynamic optimization,” *IEEE Trans. Ind. Informat.*, to be published.
8) J. Huang, X. Yang, and K. Peng, “Double layer distributed monitoring based on sequential correlation information for large-scale industrial processes in dynamic and static states,” *IEEE Trans. Ind. Informat.*, to be published.
9) P. Li and O. Niggemann, “A non-convex archetypal analysis for one-class classification based anomaly detection in cyber-physical systems,” *IEEE Trans. Ind. Informat.*, to be published.
10) Z. Chen, R. Guo, Z. Lin, T. Peng, and X. Peng, “A data-driven health monitoring method using multi-objective optimization and stacked autoencoder based health indicator,” *IEEE Trans. Ind. Informat.*, to be published.
11) L. Yang and Z. Zhang, “A conditional convolutional autoencoder based method for monitoring wind turbine blade breakages,” *IEEE Trans. Ind. Informat.*, to be published.
12) L. Ren, Z. Meng, X. Wang, L. Zhang, and L. T. Yang, “ A data-driven approach of product quality prediction for complex production systems,” *IEEE Trans. Ind. Informat.*, to be published.
13) Y. Qi, D. Chen, S. Xiang, and C. Zhu, “Gated dual attention unit neural networks for remaining useful life prediction of rolling bearings,” *IEEE Trans. Ind. Informat.*, to be published.
14) W. Lu and X. Yan, “Deep double supervised embedding neural network enhancing class separation for visual high dimensional industrial process monitoring,” *IEEE Trans. Ind. Informat.*, to be published.
15) Z. Gao, L. Hou, W. Dang, X. Wang, X. Hong, X. Yang, and G. Chen, “Multitask-based temporal-channelwise CNN for parameter prediction of two-phase flows,” *IEEE Trans. Ind. Informat.*, to be published.

Okyay Kaynak (Fellow, IEEE) received the Ph.D. degree in electronic and electrical engineering from the Department of Electronic and Electrical Engineering, University of Birmingham, Birmingham, U.K., in 1972.

From 1972 to 1979, he held various positions within the industry. In 1979, he joined the Department of Electrical and Electronics Engineering, Boğaziçi University, Istanbul, Turkey, where he is currently a Professor Emeritus. He is also a 1000 People Plan Professor with the University of Science and Technology Beijing, China. He has held long-term (near to or more than a year) Visiting Professor/Scholar positions with various institutions in Japan, Germany, USA, Singapore, and China. He has authored three books, edited five, and authored or coauthored more than 400 papers that have appeared in various journals and conference proceedings. His current research interests include intelligent control and mechatronics.

Prof. Kaynak, currently, holds the UNESCO Chair on Mechatronics. He was the recipient of the Chinese Government’s Friendship Award and the Humboldt Research Prize, in 2016. He has served on the Editorial or Advisory Boards of a number of scholarly journals, as well as the Editor-in-Chief of two IEEE publications.

Steven Ding received the Ph.D. degree in electrical engineering from the Gerhard-Mercator University of Duisburg, Duisburg, Germany, in 1992.

Since 2001, he has been a Professor of Control Engineering and the Head of the Institute for Automatic Control and Complex Systems with the University of Duisburg-Essen, Germany. He has authored three books, Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms and Tools (Springer, 2008) with second edition, in 2013, Data-Driven Design of Fault Diagnosis and Fault-Tolerant Systems (Springer, 2014), and Advanced Methods for Fault Diagnosis and Fault-Tolerant Control (Springer, 2020). His current research interests include model-based and data-driven fault diagnosis, fault-tolerant systems and their application in industry with a focus on automotive systems and chemical processes.

Dr. Ding was the recipient of the Best Paper Award from the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, in 2015.

Ahmet Palazoglu received the B.S. degree from the Middle East Technical University, Ankara, Turkey, in 1978, the M.S. degree from Bogazici University, Istanbul, Turkey, in 1980, and the Ph.D. degree from Rensselaer Polytechnic Institute, Troy, NY, USA, in 1984, all in chemical engineering.

He joined the University of California, Davis, CA, USA, in 1984, where he currently holds the title of Distinguished Professor of Chemical Engineering. He has co-authored two books, Chemical Process Performance Evaluation (with A. Cinar and F. Kayihan) (CRC Press, 2007) and an undergraduate textbook, Introduction to Process Control (with J.A. Romagnoli) (CRC Press, 2006) with a third edition to be published, in 2020. He has authored or coauthored more than 130 refereed journals and offered several short courses to academic and industrial audiences on process monitoring applications. His current research interests include design and operation of demand responsive chemical processes, process monitoring, and statistical modeling.

Prof. Palazoglu is a fellow of the American Institute of Chemical Engineers (AIChE).

Hao Luo (Senior Member, IEEE) received the B.E. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2007, and the M.Sc. and Dr.-Ing. degrees in electrical engineering and information technology from the University of Duisburg-Essen, Duisburg, Germany, in 2012 and 2016, respectively.

He is currently an Associate Professor with the School of Astronautics, Harbin Institute of Technology, Harbin, China. His current research interests include model-based and data-driven fault diagnosis, fault-tolerant systems and their plug-and-play application on industrial systems.

Dr. Luo was the recipient of the Best Paper Award from the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, in 2015. He serves as the Associate Editor for the IEEE TRANSACTIONS ON ARTIFICIAL INTELLIGENCE and IEEE Access. He has also served as the Guest Editor for the Special Issue on *Fault Diagnosis and Adaptive Fault-Tolerant Control for Automatic Control Systems* and IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS.