Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

On Data Science for Process Systems Modeling, Control and Operations

S. Joe Qin * **Yining Dong** **

* *School of Data Science and Hong Kong Institute for Data Science City University of Hong Kong, 83 Tat Chee Ave. Kowloon, Hong Kong (e-mail: joe.qin@cityu.edu.hk).*
** *Department of Electrical Engineering Stanford University, Stanford, CA 94305, USA*

Abstract:
Data science is emerging as a multidisciplinary field with tremendous recent development in theoretical foundations and expanded applications in both science and engineering. Engineering applications include industrial data analytics, autonomous systems, energy analytics, environmental applications, economic data modeling, image sequence modeling, and other high dimensional time-series data analytics. This paper is concerned with the integration of data science with dynamic systems, monitoring and control. The development of machine learning is reviewed in both a neural-mimic learning route and a learning control route, which deals with intrinsically uncertain dynamic systems. The paper then reviews the interaction of data with process manufacturing systems modeling and control, involving both data and first principles models with varying proportions. Problems include data reconciliation, state and disturbance estimation, system identification, process monitoring, and inferential property estimation. For time series data in process manufacturing systems, we present latent dynamic variable modeling methods to extract the principal dynamics in a low dimensional subspace of the data. The approaches effectively distill latent dynamic features from the data for easy interpretation, prediction, and visualization. Case studies are presented to illustrate how these latent dynamic analytics extract important features for process interpretation, troubleshooting, and monitoring.

Keywords: data science, machine learning, process data analytics, first principles vs. data, system data modeling

1. INTRODUCTION

Data science is emerging as a multidisciplinary field with tremendous recent development in theoretical foundations and expanded applications in science and engineering. Many universities and research institutions have established institutes, departments, and even schools in data sciences (Haas et al. (2019); Rappa (2019)). Harvard University has launched a new journal called Harvard Data Science Review with an inauguration editorial by Meng (2019), parallel to the Harvard Business Review. As stated in the editorial, the central mission of the journal is to help define and shape what data science is or should be.

With the development of internet of things, smart and wireless sensors, wireless communications, mobile devices, smart devices, e-commerce, and smart manufacturing, the amount of data collected and stored are growing exponentially. On the other hand, along with the ever increasing power of data analytics, many fields of science, engineering, and industries, including chemicals, petrochemicals, energy, power grids, and pharmaceuticals, have turned their attention to deriving systematic understanding and insight from various sources of data. The scientific challenges brought by the rich data, in turn have propelled tremendous development in data analytics, or broadly, data science. Statistics, for instance, is embracing the challenges to revitalize the century-old discipline, as expressed by Hastie et al. (Hastie et al. (2009)), *The field of Statistics is constantly challenged by the problems that science and industry brings to its door. . . . With the advent of computers and the information age, statistical problems have exploded both in size and complexity.* The tasks of extracting important patterns and trends, and understanding "what the data says" *have led to a revolution in the statistical sciences.*

Owing to advanced data acquisition and instrumentation in engineering systems, data from process operations and manufacturing are often high dimensional with high frequency features, multiple sampling rates, and a mix of continuous and categorical quantities. Since there is a plethora of existing principles in engineering, the first challenge in applying data science to engineering systems is how to incorporate mechanistic models with data analytics. While data analytics and machine learning provide amazing ability to represent complex relationships embedded in the data, engineering systems such as industrial processes and equipment are designed with well-defined purposes and operated under designed conditions. In these cases mechanistic or first principles models are available and dependable. However, for emerging circumstances that deviate from the design conditions, historical and real time data become valuable assets for troubleshooting and decision-making in real time operations.

While data analytics and machine learning tools have enjoyed fast development in the last couple of decades, model interpretability has become one of the desired features since human decision-makers must understand and feel convinced about what the machine learning models suggest. This is because the decision-making and operation personnel have to shoulder the responsibility in the case that the models produce erroneous suggestions. When the data-driven decision process fails, the human decision-maker must be the fallback.

Automation and control systems have always depended on measurement data for estimation, control, and optimization. This dependance is likely much enhanced with the new development of Industry 4.0. Therefore, this paper focuses on the connection between data and process systems engineering, in particular, with regard to modeling, prediction, control and monitoring. We consider the emerging field of data science as a promising source of new tools or instruments to help engineers improve and optimize the operation and control of engineering systems. Specialized analytics are expected to integrate new data science methods with the established framework of systems and control principles.

2. LEARNING AS DEVELOPED IN NEURAL NETWORKS AND CONTROL

Representative milestones in artificial neural networks are often recognized for the recent breakthroughs in the machine learning and data science field. These milestones include,

- Hebbian learning in Hebb and Hebb (1949)
- The perceptron by Rosenblatt (1958)
- The backpropagation algorithm originated from Kelley (1960)
- Minsky and Papert (1969) recognize that the perceptron is incapable of learning XOR
- Backpropagation through multi-layer neural networks by Werbos (1974)
- Learning in deep neural networks by Hinton and Salakhutdinov (2006)

On the other hand, research efforts along the line of learning control or reinforcement learning control have been developed in parallel. The work by Kelley (1960) in gradient theory of optimal flight paths, which is recognized as one of the earliest appearance of backpropagation learning, is indeed an optimal control problem. The following list of literature summarizes the efforts in learning for the purpose of control.

- Gradient method of backpropagation in Kelley (1960)
- Dynamic programming and policy iteration by Howard (1960), which applies dynamic programming Bellman (1957)
- Reinforcement learning control by Fu (1970), which clearly states that learning is for the case of incomplete knowledge
- Reinforcement learning by trial-and-error by Barto and Sutton (1982)
- The Q-learning by Watkins (1989)
- The backgammon game, applying reinforcement learning by Tesauro (1992)
- AlphaGo, applying deep reinforcement learning by Silver et al. (2016)

While the success of AlphaGo shocked the world in many ways by beating the world champion of Go with an artificial-intelligence program, it is critical to think about what is special about the game of Go with the perspective of real world engineering problems. The game of Go possesses the following characteristics.

- The model is perfectly known, i.e., white box
- Disturbances i.e., the opponent's moves, are 100% observed and measured
- Rules are perfectly clear and time-invariant
- The objective function is clear and time-invariant
- Risk tolerance, i.e., consequences in losing a game are just losing a game.

However, these characteristics do not exist in real engineering control and optimization problems, as illustrated in Qin and Chiang (2019). For example, consequences of making a bad decision in engineering systems can trigger serious economic loss and even catastrophic incidents. Therefore, significant effort is needed to translate the success of AlphaGo to engineering systems control and optimization.

Machine learning today focuses on partial intelligence such as self-driving cars, image analysis and recognition, natural language processing, and Industry 4.0 problems for smart manufacturing. A fundamental question is how to derive intelligence from messy, real world data. It is worth noting that many successful applications do not use deep neural nets, but they benefit from employing new tools from the broad field of statistical learning and data science.

3. DATA AND MODELS IN PROCESS SYSTEMS ENGINEERING

3.1 A brief history of data in PSE

The process systems engineering (PSE) has been dominated by model based control, estimation, and optimization strategies with the adoption of computers for over five decades. The successes in industrial process operation and manufacturing have reached nearly all process industries including chemicals, petroleum, refining, pharmaceutical, semiconductor manufacturing, power systems, pulp and paper, iron and steel, and energy production and supplies. However, since the very beginning of computer based optimization and control, researchers have recognized the need to rely on accurate plant data to make the optimization and control algorithms work in real time. Therefore, the use of data in PSE goes back for about six decades, as shown in the following list.

(1) Data reconciliation first studied by Kuehn and Davidson (1961); Stanley and Mah (1977)
(2) State and disturbance estimation using Kalman and particle filtering (e.g., Kalman and Bucy (1961); Arulampalam et al. (2002); Allan and Rawlings (2019))
(3) Process system identification in Otomo et al. (1972); Eykhoff (1974); Richalet et al. (1978)

(4) Inferential sensors for difficult to measure properties in McAuley and MacGregor (1991); Tham et al. (1991); Qin and McAvoy (1992)
(5) Process and control performance monitoring in Harris (1989); Piovoso et al. (1992); Nomikos and MacGregor (1994)

The above list of methods are organized in the order of increasing reliance on data and decreasing reliance on models. The task of data reconciliation by Kuehn and Davidson (1961) assumes the models are accurate and data are erroneous if they do not agree. Methods in state and disturbance estimation recognize that plant models are subject to disturbances and uncertainty, and therefore are necessarily incomplete. First principles models are combined with disturbance models that are estimated or learned from data. In system identification the models are estimated completely from data, but the data must be collected from designed experiments. It is worth noting that Otomo et al. (1972); Åström (1976) were among the first to apply system identification to real world applications.

Inferential sensors have been a proven application of neural nets and multivariate statistics since 1990s, which are estimated from historical operation data without the need to do designed experiments. Finally, process monitoring and control performance monitoring rely totally on data based models to monitor the health of the processes, control systems, and key performance indices of manufacturing systems. These applications reach beyond the process manufacturing industries.

If one pushes towards the model-based extreme, it would be dynamic programming applications. On the other hand, if one pushes towards the data-driven extreme, it would be reinforcement learning. It is interesting to note that both approaches try to achieve the optimal performance via totally different approaches. The approximate dynamic programming for optimal learning control is one evidence of the shared optimal goals with different approaches.

In chemical engineering, chemistry, biology, and materials science, data-driven methods have become popular for catalyst design, materials design and property predictions, which aims at discovering and designing new materials and formulations with desired properties (Venkatasubramanian (2019)). These methods are hopeful to result in inverse design informatics. They assume massive, high quality data are easy to obtain. Often a neural net representation of the data is realized rather than new principles. Nevertheless, these neural net representation models are useful to accelerate the discovery process for new materials design.

3.2 Necessary attributes for industrial adoptions

Although deep learning is successful in applications such as image processing, it needs to meet certain criteria in order for process industries to adopt. The following are some of the necessary attributes for sustained industrial adoptions, as stated in Qin and Chiang (2019).

- Working compatibly with first principles models or process knowledge as they reflect the laws of physics and chemistry;
- Dealing effectively with uncertainties that are usually time varying; and
- Generating interpretable solutions for the decision-makers.

Since human interventions will always be necessary in the case that decisions based on machine learning fail to work, interpretable solutions are a necessary requirement. The tragic case of the Boeing 737 Max 800 is one that failed to inform the human of the MCAS system's actions, which has led to disasters and tremendous losses.

In many studies results from machine learning can produce higher accuracy predictions than those produced by first principles models. However, *people trust first principles models not because they are always more accurate, but because they are interpretable.* When unexpected situations happen, they can deviate from existing models. In such cases data provide situational knowledge of operations, which can be transformed into real time decisions with proper data analytics. Therefore, industrial systems engineering might see a paradigm shift towards a close interplay between models and data, where models are used for expected situations and data for unexpected situations.

The essence of data science for systems engineering might be summarized as follows.

- Prediction: to predict critical variables from others using predictive analytics;
- Interpretation: to extract features or new knowledge from data to help visualization and diagnostics; and
- Decision-making.

Industrial systems data require specialized analytics combined with domain knowledge to achieve maximum benefits. While deep learning is specially effective for many applications such as image analysis, interpretable machine learning are inevitable for engineering applications.

Data from real time process control and operations are most typically high dimensional time series with complex dynamics and uncertainties. The dynamic contents of the data are useful for prediction, feature analysis, and interpretation. Given that the large dimensional time series data are usually cross-correlated and auto-correlated over time, it is necessary to develop dynamic dimensional reduction techniques or latent variable methods to focus on the modeling of dynamics in the latent subspace. In the next section we present methods in latent dynamics modeling and illustrate their use in process monitoring and troubleshooting from real time operational data. These methods integrate dynamic system theory with data analytics for operational data rather than designed system identification data, which are useful for troubleshooting and process monitoring purposes (Dong and Qin (2020)).

4. DYNAMIC DATA ANALYTICS AND MONITORING

To properly deal with dynamics in data from engineering systems, we present a framework of dynamic latent variable (DLV) methods that include supervised DLV methods and unsupervised DLV methods. Unsupervised DLVs explore dynamic variations in $\mathbf{X}$ only. Supervised DLVs aim to interpret and predict the variations in $\mathbf{Y}$ based on the latent space information in $\mathbf{X}$. Due to page limit, we only illustrate the unsupervised DLV modeling in this paper. For the supervised case we refer to the work in Dong and Qin (2015, 2018d).

Comparing to other existing unsupervised latent variable methods on high-dimensional time series, the recent development in latent dynamics modeling has the following advantages.

- Dimension reduction: the number of DLVs required to exhaust all the dynamics is usually smaller than the number of original variables.
- Descending order of predictability: the DLVs are extracted in descending order of predictability (or dynamics).
- Explicit modeling of the dynamics: auto-regressive models are built to explicitly represent the dynamics in the DLVs.

These advantages allow for easy interpretation, visualization and prediction. Next, we will explain more details about the unsupervised DLV methods.

Given a time series data matrix

$$\mathbf{X} = [\boldsymbol{x}_1 \quad \boldsymbol{x}_2 \quad \cdots \quad \boldsymbol{x}_{N+s}]^T \in \Re^{(N+s)\times M}$$

we wish to extract a dynamic latent variable from them such that its current value can be best predicted from the past. Let

$$\mathbf{Xw} = \mathbf{t} \in \Re^{N+s}$$

denote the latent scores for all observations and $t_k = \boldsymbol{x}_k^T\mathbf{w}$ denote the latent variable at time k. The latent variable is assumed to have the following auto-regressive (AR) dynamics,

$$t_k = \beta_1 t_{k-1} + \cdots + \beta_s t_{k-s} + r_s$$

To build a general auto-regressive moving average (ARMA) model, a high order AR model can be built first followed by a model compaction step (Dong et al. (2020)). The prediction from the dynamic latent model is

$$\begin{aligned}\hat{t}_k &= \beta_1 t_{k-1} + \cdots + \beta_s t_{k-s} \\ &= \boldsymbol{x}_{k-1}^T\mathbf{w}\beta_1 + \cdots + \boldsymbol{x}_{k-s}^T\mathbf{w}\beta_s \\ &= \left[\boldsymbol{x}_{k-1}^T \quad \cdots \quad \boldsymbol{x}_{k-s}^T\right](\boldsymbol{\beta}\otimes\mathbf{w})\end{aligned} \tag{1}$$

where $\boldsymbol{\beta} = [\beta_1 \quad \beta_2 \quad \cdots \quad \beta_s]^T$.

We formulate the following sequence of data matrices and the corresponding latent score vectors,

$$\begin{aligned}\mathbf{X}_i &= [\boldsymbol{x}_i \quad \boldsymbol{x}_{i+1} \quad \cdots \quad \boldsymbol{x}_{N+i-1}]^T\Re^{N\times M} \\ \mathbf{t}_i &= \mathbf{X}_i\mathbf{w} \qquad \text{for} \quad i = 1, 2, \cdots, s+1\end{aligned} \tag{2}$$

Denoting

$$\mathbf{Z}_s = [\mathbf{X}_s \quad \mathbf{X}_{s-1} \quad \cdots \quad \mathbf{X}_1]$$

the predicted score vector for $\mathbf{t}_{s+1}$ is expressed as follows based on (1),

$$\hat{\mathbf{t}}_{s+1} = \mathbf{Z}_s(\boldsymbol{\beta}\otimes\mathbf{w}) \tag{3}$$

The objective of the dynamic inner PCA (DiPCA) algorithm by Dong and Qin (2018c) is to maximize the covariance between the extracted data and the prediction, that is

$$\max_{\mathbf{w},\boldsymbol{\beta}} \ \mathbf{t}_{s+1}^T\hat{\mathbf{t}}_{s+1}$$

subject to $\|\mathbf{w}\| = 1, \|\boldsymbol{\beta}\| = 1$. Subsequently, Dong and Qin (2018b) proposed a dynamic inner canonical correlation analysis (DiCCA) algorithm that maximizes the canonical correlation as follows.

$$\max_{\mathbf{w},\boldsymbol{\beta}} \ \frac{\mathbf{t}_{s+1}^T\hat{\mathbf{t}}_{s+1}}{\|\mathbf{t}_{s+1}^T\|\|\hat{\mathbf{t}}_{s+1}\|}$$

or

$$\max_{\mathbf{w},\boldsymbol{\beta}} \ \mathbf{t}_{s+1}^T\hat{\mathbf{t}}_{s+1} \tag{4}$$

subject to $\|\mathbf{t}_{s+1}\|^2 = 1, \|\hat{\mathbf{t}}_{s+1}\|^2 = 1$. Note that the norm constraints do not affect the optimal solution since they do not affect the objective.

The maximization problem in (4) can be solved by using Lagrange multipliers, which leads to an iterative solution that involve a generalized eigenvector problem. Efficient implementation algorithms via singular value decomposition (SVD) have been proposed to solve the optimization problem in (4) by Dong et al. (2020).

While DiPCA maximizes the covariance between the latent variable and its prediction, DiCCA maximizes the correlation between the latent variable and its prediction, which cares nothing about the magnitude of the extracted latent variable. To make the magnitude of the extracted latent variable relevant, Zhu et al. (2020) modified the DiCCA objective by replacing the constraint $\|\mathbf{t}_{s+1}\|^2 = 1$ with $\|\mathbf{w}\|^2 = 1$ to come up with the latent variable regression algorithm.

The DiCCA method tends to be more efficient in extracting the dynamics, but it requires the number of samples to be greater than the number of variables in solving the generalized eigenvector problem. On the other hand, the DiPCA algorithm works when the number of samples is less than the number of variables. Therefore, both methods have their advantages.

After the weight vector $\mathbf{w}$ is solved, the latent score vector $\mathbf{t} = \mathbf{Xw}$ and $\mathbf{X}$ is deflated as

$$\mathbf{X} := \mathbf{X} - \mathbf{tp}^T \tag{5}$$

where the loading vector

$$\mathbf{p} = \mathbf{X}^T\mathbf{t}/\mathbf{t}^T\mathbf{t} \tag{6}$$

The deflated matrix $\mathbf{X}$ is then used to derive the next dynamic latent dimension.

While the extracted dynamic latent variables are convenient for visualization and interpretations, their prediction power can be used to form predictive process monitoring to shrink the normal variations down to the prediction errors as the normal uncertainty. Dong and Qin (2020) proposed several process monitoring schemes using dynamic latent variable models including DiPCA and DiCCA. The predictions of the latent variables are the conditional expectation of future observations, around which the uncertainty bound is defined by the prediction errors, rather than the entire variability of the latent variables. The true uncertainty in the data is composed of the dynamic prediction errors in the DLVs and static residuals that cannot be predicted. The proposed monitoring schemes in Dong and Qin (2020) are based on one-step ahead and multiple-steps ahead predictions. When fault samples are detected, multi-step predictions are made to avoid using faulty data samples for predictions.

5. FEATURE EXTRACTION AND MONITORING CASE STUDIES

In this section, the modeling, visualization, interpretation and diagnosis based on one of the dynamic latent variable method, DiCCA, are demonstrated on an extensively studied plant-wide oscillation dataset from an industrial plant of Eastman Chemical Company. The dataset contains a total of 8640 samples and 60 variables, with 20 seconds sampling frequency. Fig. 1 shows the process diagram.

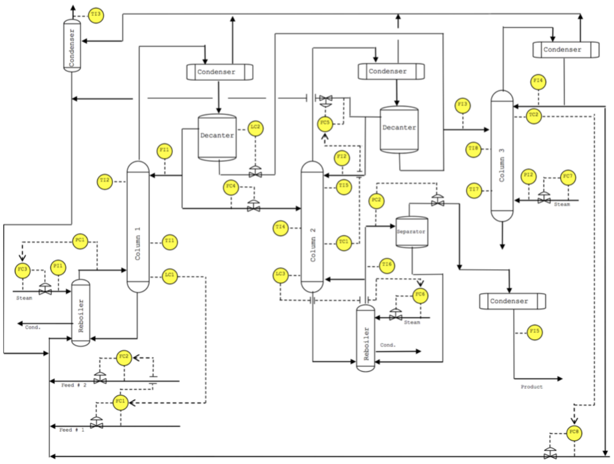

Fig. 1. Process diagram of the plant-wide oscillation dataset

5.1 Extracting features of plant-wide oscillations

DiCCA modeling method is performed on 18 variables: LC1.PV, LC1.OP, FC1.SP, FC1.PV, FC2.SP, FC2.PV, TI4.PV, TC1.PV, TC1.OP, FC5.OP, FC5.PV, LC2.PV, LC2.OP, FC8.SP, FC8.PV, TC2.OP, TI8.PV, FI3.PV. By applying DiCCA method to the 18 variables, we obtained three leading DLVs with clear oscillatory patterns. The plots of the 18 variables and the three leading DLVs are shown in Fig. 2. For the ease of visualization, only the first 1000 samples of the DLVs are plotted.

A closer examination of Fig. 2 shows that the first two DLVs only contain low frequency oscillations with a period of 340 samples, while the third DLV contains not only the low frequency, but also a high frequency oscillation with a period of 18 samples. The low frequency oscillation has been widely studied in literature Thornhill et al. (2003); Yuan and Qin (2014), and the root cause was identified as LC2.OP and LC2.PV. However, the high frequency oscillation has only been recently discovered by DiCCA method, and root cause was identified as PC2.OP and PC2.PV Dong and Qin (2018a).

The two oscillations with different frequencies can also be visualized on the 3D scatter plot of the first three DLVs as shown in Fig. 3, where the first 500 samples are plotted. The big ellipse suggests that the there is a low frequency oscillatory pattern shared by all three DLVs, and the higher frequency oscillations along the big ellipse suggest that there is a high frequency oscillatory pattern in the third DLV.

5.2 Predictive process monitoring via DiCCA

Simulated data from the Tennessee Eastman challenge process by Downs and Vogel (1993) that are generated by Chiang et al. (2001) are used to illustrate the process monitoring based on DiCCA models. The motivation is to illustrate the use of the predictions of the dynamic latent variables to shrink the uncertainties for process monitoring. Dong and Qin (2020) noticed that the normal data are simulated in a narrow range of the process and thus do not reflect typical dynamic and stochastic characteristics of real data. Therefore, the stripper unit shown in Fig. 4 is chosen to test the predictive monitoring scheme, since this unit has reasonable dynamic responses under normal conditions. The stripper unit contains 8 variables, which are XMEAS(14-19) and XMV(8,9).

A DiCCA model is built using the normal dataset containing 960 samples with a sampling rate of 3 minutes. The first 2 DLVs are able to exhaust the predictable variability in the data. Fig. 5 compares the process monitoring using standard PCA which has no prediction power vs. the predictive monitoring result using the DiCCA model. When the DiCCA predictions are used, the control regions are cetered around the predictions. Monitoring the prediction errors leads to a much smaller control region. The right chart in the figure shows the prediction error sequence for the first 200 data points, where the origin corresponds to the predicted values. By monitoring the prediction errors, we obtain more compact control regions.

6. CONCLUSIONS

With the recent development in the theoretical foundations and applications of data science, dynamic control, estimation, and monitoring can benefit greatly from these developments. Engineering system data require specialized analytics and knowledge. Both data and first principles models should be integrated with varying proportions of each part to yield reliable solutions. While physical and chemical sciences develop principles based on which mechanistic models are established, data analytics provide real-time information that reflects changes in the processes, characterizes uncertainty, and indicates emerging situations.

While deep learning has demonstrated convincing successes for image analysis and other applications, interpretable machine learning is needed for engineering applications. One common type of engineering data is high dimensional time series data, which is a promising area to benefit from the integration of data science and dynamic system theory. For time series data often found in process systems, latent dynamic variable modeling methods are effective to extract the principal variations in the data in a low dimensional subspace. The approaches effectively distill latent features from the data for easy interpretation, prediction, and visualization. Possible applications include industrial data analytics, energy analytics, environmental applications, economic data modeling, image sequence modeling, and other high dimensional time-series problems.

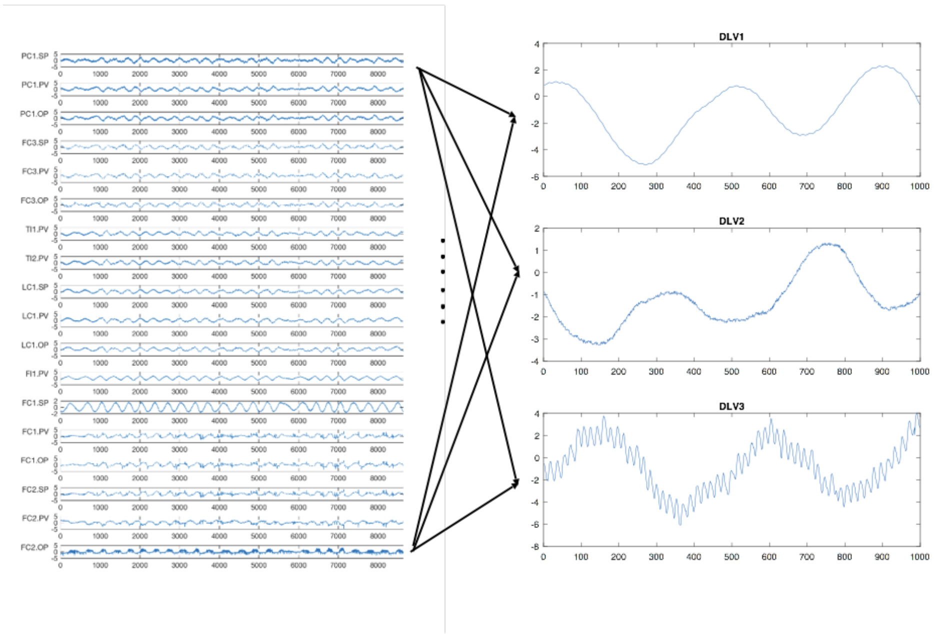

Fig. 2. Left: plots of the 18 variables for DiCCA modeling. Right: plots of three leading DLVs of DiCCA.

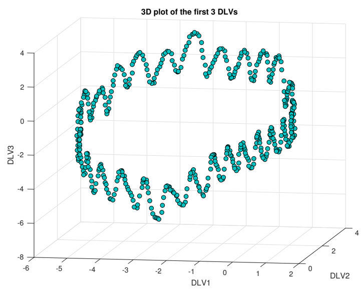

Fig. 3. 3D scatter plot of the first three leading DLVs.

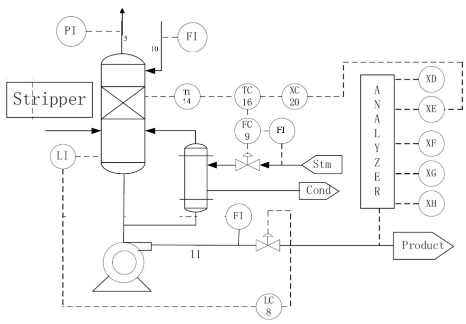

Fig. 4. The stripper unit of the Tennessee Eastman process

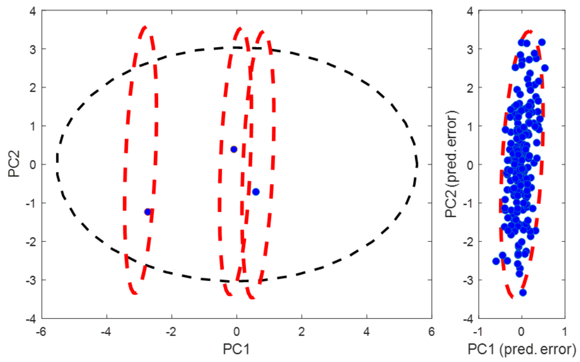

Fig. 5. DiCCA based process monitoring. Left: the large ellipse is the PCA control region without predictions; red ellipses are control regions moving with DiCCA predictions. Right: monitoring of 2-steps ahead prediction errors and the control region.

REFERENCES

Åström, K.J. (1976). State of the art and needs in process identification. In *Proc. of Conf. Process Control-II, AIChE Symposium Series*, 184–194.

Allan, D.A. and Rawlings, J.B. (2019). A lyapunov-like function for full information estimation. In *2019 American Control Conference (ACC)*, 4497–4502. IEEE.

Arulampalam, M.S., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2), 174–188.

Barto, A.G. and Sutton, R.S. (1982). Simulation of anticipatory responses in classical conditioning by a neuron-like adaptive element. *Behavioural Brain Research*, 4(3), 221–235.

Bellman, R.E. (1957). *Dynamic Programming*. Princeton University Press, Princeton, New Jersey.

Chiang, L., Russell, E., and Braatz, R. (2001). *Fault Detection and Diagnosis in Industrial Systems*. Advanced Textbooks in Control and Signal Processing. Springer-Verlag, London, Great Britain.

Dong, Y., Liu, Y., and Qin, S.J. (2020). Efficient dynamic latent variable analysis for high-dimensional time series data. *IEEE Transactions on Industrial Informatics*, 16(6), 4068–4076.

Dong, Y. and Qin, S.J. (2015). Dynamic-inner partial least squares for dynamic data modeling. *IFAC-PapersOnLine*, 48(8), 117–122.

Dong, Y. and Qin, S.J. (2018a). Dynamic-inner canonical correlation and causality analysis for high dimensional time series data. *IFAC-PapersOnLine*, 51(18), 476 – 481.

Dong, Y. and Qin, S.J. (2018b). Dynamic latent variable analytics for process operations and control. *Computers & Chemical Engineering*, 114, 69–80.

Dong, Y. and Qin, S.J. (2018c). A novel dynamic PCA algorithm for dynamic data modeling and process monitoring. *Journal of Process Control*, 67, 1–11.

Dong, Y. and Qin, S.J. (2018d). Regression on dynamic PLS structures for supervised learning of dynamic data. *Journal of Process Control*, 68, 64–72.

Dong, Y. and Qin, S.J. (2020). New dynamic predictive monitoring schemes based on dynamic latent variable models. *Industrial & Engineering Chemistry Research*, 59(6), 2353–2365.

Downs, J.J. and Vogel, E.F. (1993). A plant-wide industrial process control problem. *Computers & Chemical Engineering*, 17(3), 245–255.

Eykhoff, P. (1974). *System Identification*. Wiley, New York.

Fu, K.S. (1970). Learning control systems–review and outlook. *IEEE transactions on Automatic Control*, 15(2), 210–221.

Haas, L., Hero, A., and Lue, R.A. (2019). Highlights of the national academies report on "undergraduate data science: Opportunities and options. *Harvard Data Science Review*, 1(1). doi:10.1162/99608f92.38f16b68.

Harris, T.J. (1989). Assessment of control loop performance. *Can. J. Chem. Eng.*, 67(10), 856–861.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. New York, NY: Springer.

Hebb, D.O. and Hebb, D. (1949). *The organization of behavior*, volume 65. Wiley New York.

Hinton, G.E. and Salakhutdinov, R.R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786), 504–507.

Howard, R.A. (1960). *Dynamic Programming and Markov Processes*. NEW YORK: JOHN-WILEY.

Kalman, R.E. and Bucy, R.S. (1961). New results in linear filtering and prediction theory. *Trans. ASME, J. Basic Engineeering*, 95–108.

Kelley, H.J. (1960). Gradient theory of optimal flight paths. *ARS Journal*, 30(10), 947–954.

Kuehn, D. and Davidson, H. (1961). Computer control II: Mathematics of control. *Chem. Eng. Prog*, 57(6), 44 – 47.

McAuley, K. and MacGregor, J. (1991). On-line inference of polymer properties in an industrial polyethylene reactor. *AIChE Journal*, 37(6), 825–835.

Meng, X.L. (2019). Data science: An artificial ecosystem. *Harvard Data Science Review*, 1(1). doi: 10.1162/99608f92.ba20f892.

Minsky, M. and Papert, S. (1969). Perceptron: an introduction to computational geometry. *The MIT Press, Cambridge, expanded edition*, 19(88), 2.

Nomikos, P. and MacGregor, J.F. (1994). Monitoring batch processes using multiway principal component analysis. *AIChE Journal*, 40(8), 1361–1375.

Otomo, T., Nakagawa, T., and Akaike, H. (1972). Statistical approach to computer control of cement rotary kilns. *Automatica*, 8(1), 35–48.

Piovoso, M.J., Kosanovich, K.A., and Yuk, J.P. (1992). Process data chemometrics. *IEEE Trans. Instrum. and Meas.*, 41, 262–268.

Qin, S.J. and McAvoy, T. (1992). Nonlinear PLS modeling using neural networks. *Comput. and Chem. Eng.*, 16(4), 379–391.

Qin, S.J. and Chiang, L.H. (2019). Advances and opportunities in machine learning for process data analytics. *Computers & Chemical Engineering*, 126, 465 – 473.

Rappa, M. (2019). Behold the data scientist! In *Foundations of Process Analytics and Machine learning*. CACHE, Raleigh, NC.

Richalet, J., Rault, A., Testud, J., and Papon, J. (1978). Model predictive heuristic control. *Automatica (Journal of IFAC)*, 14(5), 413–428.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 386.

Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., and Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489.

Stanley, G.M. and Mah, R.S.H. (1977). Estimation of flows and temperatures in process networks. *AIChE Journal*, 23, 642–650.

Tesauro, G. (1992). Temporal difference learning of backgammon strategy. In *Machine Learning Proceedings 1992*, 451–457. Elsevier.

Tham, M.T., Montague, G.A., Morris, A.J., and Lant, P.A. (1991). Soft-sensors for process estimation and inferential control. *Journal of Process Control*, 1(1), 3 – 14.

Thornhill, N.F., Cox, J.W., and Paulonis, M.A. (2003). Diagnosis of plant-wide oscillation through data-driven analysis and process understanding. *Control Engineering Practice*, 11(12), 1481–1490.

Venkatasubramanian, V. (2019). The promise of artificial intelligence in chemical engineering: Is it here, finally? *AIChE Journal*, 65(2), 466–478.

Watkins, C.J.C.H. (1989). *Learning from delayed rewards*. King's College, Cambridge.

Werbos, P. (1974). *Beyond regression: New tools for prediction and analysis in the behavioral sciences*. Ph.D. dissertation, Committee on Appl. Math., Harvard Univ.

Yuan, T. and Qin, S.J. (2014). Root cause diagnosis of plant-wide oscillations using granger causality. *Journal of Process Control*, 24(2), 450–459.

Zhu, Q., Qin, S.J., and Dong, Y. (2020). Dynamic latent variable regression for inferential sensor modeling and monitoring. *Computers & Chemical Engineering*, 137, 106809.