Hidden State Observation for Adaptive Process Controls

Melanie Senn, Norbert Link
Institute of Computational Engineering at IAF
Karlsruhe University of Applied Sciences
Moltkestrasse 30, Karlsruhe, Germany
Email: melanie.senn@hs-karlsruhe.de, norbert.link@hs-karlsruhe.de

***Abstract*—In many manufacturing processes it is not possible to measure on-line the state variable values that describe the system state and are essential for process control. Instead, only quantities related to the state variables can be observed. Machine learning approaches are applied to model the relation between observed quantities and state variables. The characterization of a process by its state variables at any point in time can then be used to adequately adjust the process parameters to obtain a desired final state. Also, multiple process controls of a process chain can be linked using standardized transfer state variables between the single processes. This allows the optimization of the entire process chain with respect to the desired properties of the final workpiece. This paper proposes a general method to extract state variables from observable quantities by modeling their relation from experimental data with data mining methods. After transforming the data to a space of de-correlated variables, the relation is estimated via regression methods. Using Principal Component Analysis and Artificial Neural Networks we obtain a system capable of estimating the process state in real time. The feasibility of our approach is shown with data from numerical simulation of a deep drawing process.**

***Keywords*-statistical process model; hidden state prediction; regression analysis; dimension reduction; deep drawing.**

I. INTRODUCTION

Closed-loop controls are capable of reaching desired final states by compensating disturbances in single processes or by adapting to varying input in a process chain. Feedback about the system state is essential for this purpose. The measurement of the real state variables usually requires large efforts and cannot be executed in process real time. Only few process-related quantities can be measured by real production machines during process execution. If these observables can be related to state variables with sufficient unambiguity and accuracy, a state-based closed-loop control can be created. The final state can then be estimated as well and the information be transferred to the control of the next step in a process chain. In deep drawing, observables such as forces and displacements in the tools and the workpiece are accessible during process execution with reasonable measurement effort. Mechanical stress distributions reflecting the state of the sheet material can be used as controlled variable as applied in [1] to optimize the blank holder force for an experimental deep drawing environment. A control system for deep drawing is presented in [2], based on the identification of static material properties as proposed in [3].

Data mining methods for regression analysis such as Artificial Neural Networks (ANNs) or Support Vector Regression (SVR) are widely used in material science for the prediction of static process quantities. In [4], thickness strains have been computed, [3] presents a model to predict material properties from process parameters and conditions. These both affect the final result, however, conditions are constant during execution and can not be used for on-line state control. The texture of cold rolled steels has been predicted from process conditions in [5]. A general overview for the application of ANNs in material science is given in [6] under consideration of model uncertainties and noise.

In our approach a feedforward, completely connected ANN is used due to its capability of modeling the nonlinear relation between observable quantities and the process state. Principal Component Analysis (PCA) is applied for dimension reduction in observables and state variables to decrease the complexity of their relations.

This paper is structured as follows. In Section II the statistical process model and underlying data mining methods are introduced. A proof of concept is given by the application of the statistical model to data from numerical experiments of a deep drawing process in Section III. Results for predicted state quantities are presented and evaluated in Section IV. Section V concludes and outlines future work.

II. MODELING

Numerical models based on first principles have the ability to predict results accurately and reliably after they have been validated by experimental results. However, the high quality comes along with high computational costs. Phenomenological models are based on observations of first principles and normally require less, but still substantial computational resources. Both types of models can be used to describe dynamic process behavior. If it comes to on-line process control, however, high speed models are needed to make fast predictions. Statistical models provide this property and thus can be used to reproduce the relation between observable quantities and process states on the one hand and the relation between state variables and appropriate process parameters on the other hand.

 ISBN: 978-1-61208-109-0

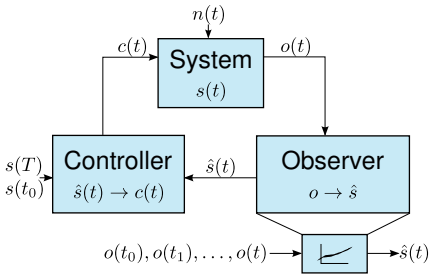


Figure 1. Closed-loop adaptive process control

A. Relating Observables to State Variables

During process execution, the dynamic system moves along in its state space where each state generates observable values related to the respective state variable values. In materials processing, the state variables may be fields of intensive magnitudes such as strains or stresses that are reflected in observables like displacements, forces and temperatures.

A closed-loop adaptive process control based on hidden state observation is shown in Figure 1. The dynamic system is characterized by its state $s(t)$ and is subject to a system noise $n(t)$ that has to be compensated by the controller to reach a defined final state $s(T)$. The observer models the relation between observables and state variables and delivers estimated state variables $\hat{s}(t)$. These are again used by the controller to find appropriate process parameters $c(t)$ considering the reference $s(T)$ as definition for the final state at time T. If multiple process controls are linked to a process chain, the final state of the preceding process serves as initial state of the current process $s(t_0)$. This additionally influences the process parameters determined by the controller during process execution.

The hidden state observer provides state information by deriving the current estimated state $\hat{s}(t)$ at time t from observables between process begin at time t_0 and the current time t. Input and output quantities of the regression analysis are high dimensional, whereas with reasonable measurement effort only a limited number of samples can be provided by experiments. Therefore, we propose to model the complex relation between observables and state variables with an ANN applying dimension reduction to input and output before regression analysis is performed.

B. Regression Analysis

A feedforward, completely connected ANN is used to model the nonlinear relation between observables (input) and state variables (output). We choose a three layer network topology (input, hidden, output), which is sufficient according to the theorem of Kolmogorov [7]. Each of the neurons in the subsequent layer is connected to all neurons of the current layer, where each connection has assigned a certain weight value. A logistic activation function is applied to the superposition of the activations of preceding neurons and the weights added up with a threshold value. The regression analysis by means of ANNs consists of minimizing an error cost function with respect to the weights and thresholds. For the cost function, the sum of squared errors (SSE) between the output values of the network and the output values of the associated input values as given by a sample is selected. The ANN is trained by the backpropagation algorithm [8].

The number of nodes in the hidden layer has been determined according to (1), see [9]. The objective is to retrieve an overdetermined approximation, i.e., the number of training samples must be greater than the number of degrees of freedom, namely the number of connection weights.

$$KN = \alpha(J(I + K) + J + K) \tag{1}$$

Equation (1) reveals the relation between

- the number of input nodes (I)
- the number of hidden nodes (J)
- the number of output nodes (K)
- the number of training samples (N)
- the determinacy factor (α),

which is problem dependent. Starting from a minimum of 1.0 (exact determination), the optimal determinacy factor α has been experimentally identified by evaluation of the network's performance function quantified by the mean squared error (MSE). A first guess of determinacy has been obtained by comparison of network performance results between 1.0 and the maximum determinacy resulting from a network with only one output node with a step width of 10. Successive refinements by step widths of 1 and 0.1 have been performed around the value of the previous step until the optimal determinacy is obtained with respect to the network's performance function.

The number of output nodes is on the one hand predefined by the number of output dimensions of the regression problem itself, but on the other hand the output nodes do not necessarily have to belong to one single network. An extreme configuration is to generate one network per output dimension to reduce the complexity that has to be described by the hidden layer. In our approach we use only one network since the complexity of the regression problem has already been reduced by dimension reduction. The Levenberg-Marquardt algorithm is used to solve the optimization problem of finding optimal connection weights.

C. Dimension Reduction

PCA is employed to reduce the dimensionality of observables and state variables by removing correlations in space and time. Before applying the PCA algorithm, the data spanned by the three dimensions

- the number of samples (I)
- the number of variables per time frame (J)
- the number of time frames (K)

have to be arranged in two dimensions. Reference [10] states that only two of the six possible unfolding techniques have

 ISBN: 978-1-61208-109-0

practical relevance. In A-unfolding ($KI \times J$) the number of time frames and the number of samples are aggregated in the first dimension, the number of variables per time frame characterizes the second dimension. D-unfolding ($I \times KJ$) uses the number of samples as first dimension and combines the number of time frames and the number of variables per time frame in the second dimension. The latter is therefore more appropriate to remove correlations between different time frames as well as between variables within the same time frame. In [11], dynamic process behavior is monitored by Dynamic Principal Component Analysis (DPCA) considering a limited window of time lagged observations. In our model we use the full history of observables to make use of the complete information available to us.

The data in original dimensions X are subject to a transformation of the principal axes by finding directions of maximum variance. The first new axis points in the direction of largest variance of the data X and is called the first principal component. The second principal component is orthogonal to the first one and points in the direction of second largest variance. Additional components can be found analogously, while higher ones describe less variance.

$$X = \sum_{i=1}^{n} \alpha_i e_i \tag{2}$$

The data can be represented by (2), where n stands for the number of dimensions of X, e represents the basis vectors and α describes the data in the new coordinate system. Dimension reduction can be achieved by removing higher principal components since they do not explain much of the variance in the data.

$$K = \frac{1}{n-1} X^T X \tag{3}$$

Related eigenvectors and eigenvalues can be calculated from the covariance matrix K, see (3), where X has been mean-centered before. Pairs of eigenvalues and eigenvectors are then sorted such that the largest eigenvalue is associated with the first principal component explaining most variance [12]. The covariance matrix can be seen as a description of the rotation in the transformation of the principal axes, the data centroid corresponds to the displacement of the origin of the new coordinate system with respect to the initial one.

If the number of variables is much greater than the number of samples, which might apply to observables, [13] advises to use Singular Value Decomposition (SVD) according to (4) to determine eigenvalues and eigenvectors efficiently.

$$X = USV^T \tag{4}$$

The eigenvalues α can be extracted from the diagonal matrix S by (5) where m corresponds to the number of samples, while the orthonormal matrix V contains associated eigenvectors e.

$$\alpha = \frac{1}{m-1} S^T S \tag{5}$$

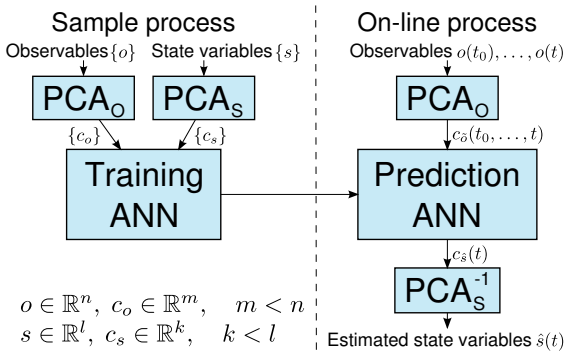


Figure 2. Architecture of the statistical process model

D. Statistical Process Model

The statistical process model for hidden state observation displayed in Figure 2 is divided into a training and a prediction block. For each requested point in time t the system collects previously sampled observables $o(t_0), \ldots, o(t)$ and current state variables $s(t)$. First, PCA is applied to both observables o and state variables s of which a subset is used to train the ANN as input c_o and target c_s, respectively. After successful training the ANN is able to predict state variables in reduced dimensions $c_{\hat{s}}(t)$ from previously unseen observables $o(t_0), \ldots, o(t)$, reduced to $c_{\tilde{o}}(t_0, \ldots, t)$, that have not been included in training. The predicted state quantities $c_{\hat{s}}(t)$ are subject to an inverse dimension transformation to obtain their counterparts $\hat{s}(t)$ in the original, high dimensional space for visualization and validation.

III. Application to Deep Drawing

The feasibility of the proposed approach is tested with an elementary sample process, the cup deep drawing of a metal sheet. In cup deep drawing, a metal sheet is clamped between a die and a blank holder. A punch presses the sheet that undergoes a traction-compression transformation into the die opening to obtain a cup-shaped workpiece. The assembly is displayed in Figure 3. Statistical samples have been generated by experiments performed in a numerical simulation environment. Fur this purpose an axisymmetric finite element deep drawing model has been implemented in ABAQUS (finite element analysis software).

Observable quantities are displacements and forces, temperature behavior has been neglected. Displacements in the cup bottom in direction of the moving punch have been recorded as well as displacements in the sheet edge in orthogonal direction to reflect sheet retraction. Additional displacements in punch and blank holder have been acquired. Reaction forces in the tools have been recorded in both radial and axial direction. Arising partial correlations in observables are removed by the PCA. The state of the deep drawing process is characterized by the von Mises stress distribution within the workpiece.

 ISBN: 978-1-61208-109-0

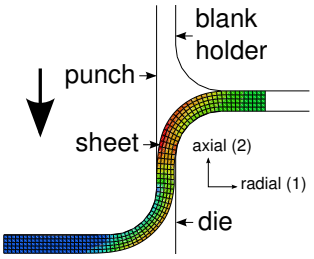


Figure 3. Workpiece and tools in axisymmetric deep drawing

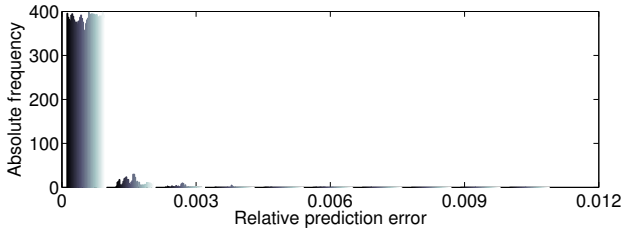


Figure 4. Relative error distribution for predicted state variables

In the performed parametric study, the blank holder force has been varied in the range of [70, 100] KN, process conditions such as drawing height or lubrication have been kept constant. For each sample observables and state variables have been collected for all time frames. A time frame equalization ensures common time frames for all samples. 200 samples have been generated, each consisting of 131 time frames, which in turn contain nine observables and 400 state variables. The extracted data have been randomly partitioned into a training set (80%) and a test set (20%). Dimension reduction has been applied to all samples, where the training data were split again randomly into a training set (80%) and a validation set (20%) for regression analysis. The test set has been used for overall validation of the statistical model. Resampling and subsequent remodeling has been performed to select the best model and to prove independence of specific data sets.

IV. Discussion of the Results

Two use cases were identified for state estimation. The first use case refers to the prediction of the final process state based on the observable values during the entire process execution. This provides a subsequent process with detailed information about its input, allowing it to optimally adjust its parameters. The single process controls of a process chain can be linked by the state information in a standardized way, resulting in an overall quality improvement. This use case is described in Section IV-A. Prediction of the state evolution during process execution is employed in process control as discussed in Section IV-B. The latter can be seen as a generalization of the former.

A. Prediction of the Final Process State

The statistical model for hidden state observation has been validated by a test set of 40 samples (see Section III). The relative prediction error of the 400 state variables never exceeds 0.0110 for all samples, the resulting distribution is shown in Figure 4. The absolute frequency of the number of state variables is high for small errors and drops rapidly with increasing error. Different colors stand for individual samples. The quality of the results shows the principle feasibility of the method. One must be aware that this might be partly due to the simplicity of the experiments: the variance in observables and state variables is not very large since the blank holder force has been the only varied parameter. Investigations with more realistic process models and real experimental data are subject of ongoing work.

The prediction quality was further analyzed as follows. It has been shown that the variation of predicted results is substantially smaller than the variance of the generated data. For this, we have defined the model uncertainty σ by

$$\sigma = \frac{1}{n} \sum_{i=1}^{n} \frac{MSE_i}{Var_i}, \tag{6}$$

which corresponds to the mean value over all dimensions i of the MSE over all samples divided by the corresponding sample variance Var. In (6), n stands for the number of dimensions, i.e., the number of state variables. A low model uncertainty is characterized by a σ value close to zero, whereas values approaching 1.0 indicate high uncertainty. For our experiment σ is 0.0045, which indicates the high accuracy and low uncertainty of the predicted results.

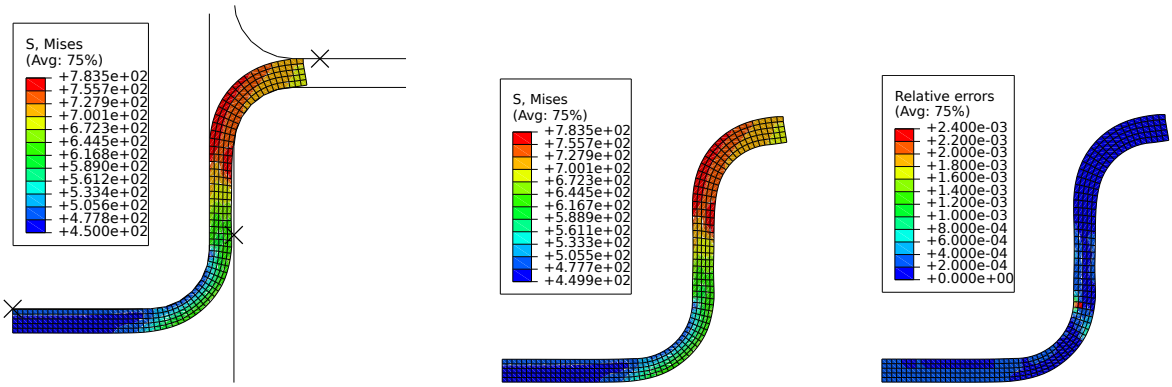
The quality of the statistical model has been quantified by the coefficient of determination R^2, which can be applied to nonlinear regression analysis by (7) according to [14]

$$R^2 = 1 - \frac{SSE}{SST}. \tag{7}$$

The SSE describes the sum of the squared deviations between original data in the test set and associated predicted results. It is divided by the SST, which quantifies the variation in the test set calculated by summed squared deviations of the original state variables from their means. Both quantities are computed over all state variables for all samples in the test set. A resulting R^2 value of 0.9991 confirms the good quality of the statistical model. The MSE of the predicted state variables serves as a base for the determination of a confidence interval for the prediction error. The precision of the estimation amounts to a mean value of 0.0440, which has been calculated over all predicted state variables for a 95% confidence interval.

The overall error of the statistical model is composed of a time frame equalization error, the error from dimension reduction and the ANN prediction error. The MSE of the ANN amounts to 0.0019 at a typical range of [400, 800] MPa of the predicted von Mises stresses. Observables have been reduced from 1179 (9 observables per time frame $\times$ 131 time

 ISBN: 978-1-61208-109-0



(a) Finite element model (original) (b) Statistical model (prediction) (c) Relative prediction errors

Figure 5. Comparison of results of the finite element model and the statistical model

frames) to nine dimensions with a predefined precision of 99.999% and thus a relative error of 0.001%. State variables have been reduced from 400 to seven dimensions with a precision of 99.900%, i.e., a relative error of 0.1%. On the one hand dimension reduction implies information loss that cannot be recovered, but on the other hand it enables the ANN to find correlations in the reduced data. A worse result might have been obtained without the usage of dimension reduction due to the huge number of additional degrees of freedom to be determined by the ANN.

Some results for a representative of the test set visualized in ABAQUS are depicted in Figure 5. It displays the absolute von Mises stress values in MegaPascal predicted by the statistical model in Figure 5b, which are in very good agreement with the results of the finite element model illustrated in Figure 5a. To outline the deviation of the predicted results from the original data the relative error in the range of [0, 0.0024] is presented in Figure 5c. Errors are low in regions with small sheet deformations, while higher errors occur in areas with high deformation gradients.

Robust predictions are characterized by bounded prediction errors despite of model uncertainties and disturbances. In our work we first applied a white noise of 5% to the observables to model a measurement error. State variables have then been predicted with a relative error in the range of [0.0, 0.0569] and a corresponding mean value of 0.0024. The model uncertainty σ amounts to 0.1069, while the model quality is characterized by a R^2 value of 0.9313. Increasing the noise to 10% results in a relative error range of [0.0, 0.1115] with a mean of 0.0033, a model uncertainty σ of 0.1990 and a R^2 value of 0.8412. The size of the error range does not solely represent the quality of prediction, also the model uncertainty affecting the distribution within this range has to be considered. The results indicate that our model is robust to small disturbances and still delivers satisfactory results for small manipulations in observables. However, with increasing noise model quality decreases as uncertainty increases.

B. State Prediction During Process Execution

Process execution time determines the timespan in which process parameters can be adjusted to control the process state. State information is not necessarily needed for each single time frame, since controllers are usually liable to a certain delay in their impact. The implementation of the statistical model offers the selection of time frames that are crucial for control. In this work some representatives have been chosen to demonstrate the feasibility of state evolution prediction. Time frame numbers 1, 45, 90 and 131 have been selected, the results are outlined in Table I.

The parameters of the statistical process model have been set as follows. The respective determinacy of the ANN has been identified incrementally by evaluating the network's performance function (see Section II-B), the precision for dimension reduction has been chosen as 99.999% for observables and 99.900% for state variables. The number of observables and state variables in reduced dimensions each grows with increasing time since their inner relations become more complex. The number of hidden nodes increases as well due to the more complex relation between observables and state variables. Between time frame number 45 and 90 the number of hidden nodes however decreases. At this point the number of input nodes of the ANN given

Table I
PREDICTION CHARACTERISTICS DURING PROCESS EXECUTION

Time frame number	1	45	90	131
# PCA observables	1	1	6	9
# PCA state variables	1	2	3	7
# ANN hidden nodes	33	36	26	38
ANN determinacy	1.3	1.8	1.5	1.4
MSE PCA observables	1.2453	1.3612	51.8063	80.0205
MSE PCA state variables	$4 \cdot 10^{-5}$	$4 \cdot 10^{-4}$	0.0026	0.0667
MSE ANN	$2 \cdot 10^{-5}$	$7 \cdot 10^{-6}$	$3 \cdot 10^{-4}$	0.0019
R^2 statistical model	0.9999	0.9999	0.9998	0.9991
σ statistical model	0.1452	0.0003	0.0028	0.0045
Mean relative error	0.0047	$8 \cdot 10^{-6}$	$3 \cdot 10^{-5}$	$2 \cdot 10^{-4}$
Max relative error	0.1071	0.0021	0.0020	0.0110

 ISBN: 978-1-61208-109-0

by the number of observables in reduced dimensions is for the first time higher than the number of output nodes given by the number of state variables in reduced dimensions. The MSE caused by dimension reduction in observables and state variables rises with increasing time and thus increasing complexity. The performance of the ANN evaluated by its MSE decreases between time frame number 1 and 45 and then increases. This behavior is also reflected in the relative error distribution specified by its mean and maximum value. The explanation is composed of two opposed effects. The model uncertainty σ is on the one hand very high at the beginning of the process, since not much process knowledge by means of observables is available. On the other hand there is not much variance in the state variables at this point, because the impact of different applied blank holder forces is not yet strong, but will play a more important role with increasing time. Although the model uncertainty is very high for time frame number 1, the prediction result is still characterized by a high quality index due to the low variance in the process state. The uncertainty decreases with increasing time, but then also complexity grows and has a stronger impact on the prediction result. The overall model quality R^2 as well as the relative error distribution show that the predictions are in good agreement with the original data.

V. Conclusion and Future Work

It has been shown that the statistical process model for hidden state observation applied to a deep drawing process can be successfully used for process state prediction based on observations. The results outlined in Section IV are very promising and can therefore be taken as a solid base for process control. Process parameters can thereon be adjusted by observing the evolution of the process state implementing a suitable control law. The control of one single process can be extended to process chain optimization by multiple linked process controls. For this purpose workpiece properties are to be deduced from the final state of the final process serving as set value for the optimization procedure. One drawback of the statistical process model is the high uncertainty in state prediction at the beginning of the process. This can be overcome by not considering those early predictions with high uncertainty in process control and by further extensions of the presented approach. Significant time frames for the observation of the process state evolution have to be identified to enable process control. The proposed approach for observation of hidden states for adaptive process controls can be transferred to any process characterized by state variables that can be extracted from related observable quantities.

Acknowledgment

This work has been supported by the DFG Research Training Group 1483 "Process chains in manufacturing". The authors would like to thank the ITM at the KIT for providing the finite element deep drawing model.

References

[1] C. Blaich and M. Liewald, "Detection and closed-loop control of local part wall stresses for optimisation of deep drawing processes," in *Proceedings of the International Conference on New Developments in Sheet Metal Forming Technology*, Fellbach, Germany, 2010, pp. 381–414.

[2] Y. Song and X. Li, "Intelligent control technology for the deep drawing of sheet metal," in *Proceedings of the International Conference on Intelligent Computation Technology and Automation*, Los Alamitos, CA, USA, 2009, pp. 797–801.

[3] J. Zhao and F. Wang, "Parameter identification by neural network for intelligent deep drawing of axisymmetric workpieces," *Journal of Materials Processing Technology*, vol. 166, pp. 387–391, 2005.

[4] S. K. Singh and D. R. Kumar, "Application of a neural network to predict thickness strains and finite element simulation of hydro-mechanical deep drawing," *The International Journal of Advanced Manufacturing Technology*, vol. 25, no. 1, pp. 101–107, 2005.

[5] A. Brahme, M. Winning, and D. Raabe, "Prediction of cold rolling texture of steels using an artificial neural network," *Computational Materials Science*, vol. 46, pp. 800–804, 2009.

[6] H. K. D. H. Bhadeshia, "Neural networks and information in materials science," *Statistical Analysis and Data Mining*, vol. 1, no. 5, pp. 296–305, 2009.

[7] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Proceedings of the International Joint Conference on Neural Networks*, Washington D.C., USA, 1989, pp. 593–605.

[8] M. T. Hagan, H. B. Demuth, and M. H. Beale, *Neural network design*. University of Boulder, Colorado, USA: Campus Publication Service, 2002.

[9] W. C. Carpenter and M. E. Hoffman, "Selecting the architecture of a class of back-propagation neural networks used as approximators," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, vol. 11, pp. 33–44, 1997.

[10] C. Zhao, F. Wang, N. Lu, and M. Jia, "Stage-based soft-transition multiple PCA modeling and on-line monitoring strategy for batch processes," *Journal of Process Control*, vol. 17, no. 9, pp. 728–741, 2007.

[11] J. Chen and K.-C. Liu, "On-line batch process monitoring using dynamic PCA and dynamic PLS models," *Chemical Engineering Science*, vol. 57, no. 1, pp. 63–75, 2002.

[12] C. M. Bishop, *Pattern recognition and machine learning*, 2nd ed. Springer, 2007.

[13] T. Hastie, R. Tibshirani, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, 2nd ed. Springer, 2009.

[14] H. Motulsky and A. Christopoulos, *Fitting Models to Biological Data using Linear and Nonlinear Regression - A practical guide to curve fitting*. GraphPad Software Inc., San Diego CA, 2003.

 ISBN: 978-1-61208-109-0