A Universal Model for Hidden State Observation in 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. This paper proposes a general method to extract state variables from observable quantities by modeling their relations 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 general method features a high flexibility in adjusting the complexity of the regression relation by an adaptive history and by a variable determinacy in terms of degrees of freedom in the model parameters. The universal model is applied to data from numerical deep drawing simulations to show the feasibility of our approach. The application to the two sample processes, which are of different complexity confirms the generalizability of the model.

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

I. INTRODUCTION

In our previous work [1], we have presented a statistical model for hidden state observation based on data from an elementary deep drawing process. In this paper, we extend the observation to a complex deep drawing process with anisotropic plastic material behavior. It is shown that the universal statistical process model is capable of generating model instances for both complexity categories of the deep drawing process with good prediction results.

Closed-loop controls are capable of reaching desired final states by compensating disturbances in individual 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. Multiple process controls of a process chain can be linked together 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. Some approaches follow this idea by observing such quantities, which are directly correlated to the controlled variables. This holds usually true only for one specific process.

In deep drawing, observables such as forces and displacements in the tools and in the workpiece are accessible with reasonable measurement effort during process execution. Mechanical stress distributions reflecting the state of the sheet material can be used as the controlled variable as applied in [2] to find optimal values for the blank holder force of an experimental deep drawing environment. A control system for deep drawing is presented in [3], based on the identification of static material properties in [4].

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 time-invariant process quantities. In [5], thickness strains in different directions of the sheet are computed from material parameters, and [4] 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 cannot be used for on-line closed-loop state control. The texture of cold rolled steels is predicted from process conditions in [6]. A general overview for the use of ANNs in material science is given in [7] considering model uncertainties and noise.

In our approach, a generic state estimator is proposed, which represents the functional dependence of state variables on observable quantities, by adapting its structure and parameters to the specific process under consideration. The estimator consists of a feedforward, completely connected ANN, which 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. An adaptive history of observable quantities allows an additional adjustment of the complexity of the regression relation.

This paper is structured as follows. In Section II, the statistical process model and its underlying data mining methods are introduced. A proof of concept is given by the application of the universal statistical model to data from numerical experiments of two deep drawing processes of different complexity in Section III. Results for predicted state quantities for both sample processes are presented and evaluated in Section IV. In particular, the creation of reliable and robust models is achieved by the assessment of various modifications of the input history. 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 model types can be used to describe the dynamic process behavior during its execution. If it comes to on-line process control, however, high speed models are needed to perform 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.

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 $\mathbf{s}(t)$ for each point in time t and it is subject to a system noise $\mathbf{n}(t)$ that has to be

Figure 1.   Closed-loop adaptive process control

compensated by the controller to reach a defined final state $\mathbf{s}(T)$. The observer models the relation between observables and state variables and delivers estimated state variables $\hat{\mathbf{s}}(t)$, or $\hat{\mathbf{s}}(t_c)$ for one particular observation point in time t_c, respectively. The estimated state variables are then used by the controller to find appropriate process parameters $\mathbf{c}(t)$ considering the reference $\mathbf{s}(T)$ as a definition for the final state at time T. If multiple process controls are linked together in a process chain, the final state of the preceding process serves as an initial state of the current process $\mathbf{s}(t_0)$. 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 $\hat{\mathbf{s}}(t_c)$ at time t_c from observables between a defined starting point at time t_0 and the current time t_c. The begin of the process may be chosen as a starting point for observation, but also a limited history of preceding time frames is admissible. The consideration of a sampled history of observable quantities as the basis for state estimation results in a high dimensionality of the estimator input. When fields of physical quantities, which are spatially sampled, represent the state, also the estimator output is high dimensional. The estimator parameters are determined by nonlinear regression, which would require a large number of samples for high dimensional input and output spaces that are usually not available from experiments. Therefore, we propose to model the complex relation between observables and state variables with an ANN applying PCA 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 [8]. Each of the neurons in the subsequent layer is connected to all neurons of the current layer, where each connection is 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 [9].

The number of nodes in the hidden layer is determined according to

$$CN_t = \alpha(B(A+C) + B + C), \qquad (1)$$

see [10] for details. The objective is to retrieve an overdetermined approximation, i.e., the number of training samples
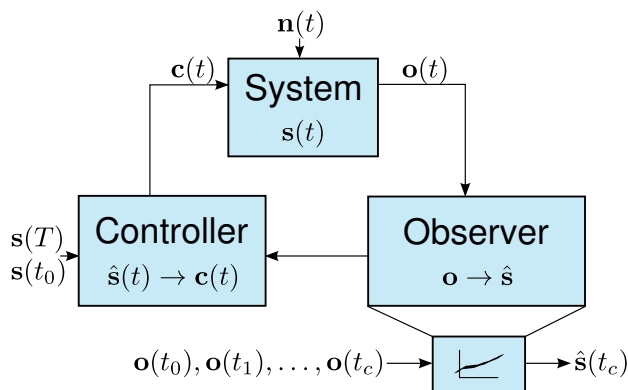
must be greater than the number of degrees of freedom, namely the number of connection weights.

Equation (1) reveals the relation between

- the number of input nodes (A)
- the number of hidden nodes (B)
- the number of output nodes (C)
- the number of training samples (N_t)
- the grade of determinacy (α),

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

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 and the spaces of input and output have been transformed to their decorrelated counterparts. The Levenberg-Marquardt algorithm is used to solve the optimization problem of finding optimal connection weights by a second-order approximation.

C. Dimension Reduction

PCA is applied to reduce the dimensionality of observables and state variables by removing correlations in space (between the variables) and time. Here, we do not perform either regression nor prediction, but a pure dimension reduction. The reduced observables and state variables are then used for regression by an ANN as described in Section II-B.

In the following, $\mathbf{X}$ is a place holder for a set of sequences of variables. In our case, $\mathbf{X}$ stands for the observable history $\mathbf{o}(t_0)_n, \ldots, \mathbf{o}(t_c)_n$ or for the current state variables $\mathbf{s}(t_c)_n$ for all $n = 1 \ldots N$ samples. Before executing the PCA algorithm, the data spanned by the three dimensions

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

have to be arranged in two dimensions. Reference [11] states that only two of the six possible unfolding techniques have practical relevance. In A-unfolding ($KN \times J$), the number of time frames and the number of samples are aggregated

in the first dimension, and the number of variables per time frame characterizes the second dimension. D-unfolding ($N \times KJ$) uses the number of samples as the 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 individual variables within the same time frame. In [12], dynamic process behavior is monitored by Dynamic PCA (DPCA) considering a limited window of time-lagged observations.

We perform a MPCA with D-unfolding where the data $\mathbf{X}$ are arranged in a 2D matrix of dimension $N \times KJ$, in which blocks of variables for each time frame are aligned side by side. The history of the time frames $t_0 \ldots t_c$ is therefore mapped to $1 \ldots K$. We can apply the complete history of observables to make use of the entire information available to us. Alternatively, we can use a reduced observable history of selected preceding time frames to scale down the complexity of the regression relation. In the case of convoluted regression relations as arising from the complex sample process, we prefer an adjusted history. In this case, the number of time frames is reduced to a part of the complete observable history and at the same time, the precision requirement for dimension reduction is increased. This results in a selection of additional principal components in order to extract more input information to explain more variance in the target quantity. The current state variables, which are only extracted at time t_c, are of dimension $N \times J$ ($K = 1$) and do therefore not have to be unfolded.

The data in original dimensions $\mathbf{X}$ 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 $\mathbf{X}$ 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. The data $\mathbf{X}$ can be represented by

$$\mathbf{X} = \sum_{w=1}^{W} \mathbf{t}_w \mathbf{p}_w^T = \mathbf{T}\mathbf{P}^T, \qquad (2)$$

where W stands for the number of principal components, $\mathbf{P}$ represents the basis vectors of the new coordinate system and $\mathbf{T}$ 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.

Related eigenvectors and eigenvalues can be calculated from the empirical covariance matrix (notice the division by $N - 1$) given by

$$\mathbf{K} = \frac{1}{N-1}\mathbf{X}^T\mathbf{X}, \qquad (3)$$

where $\mathbf{X}$ has been mean-centered before and N corresponds to the number of samples in $\mathbf{X}$. Pairs of eigenvalues and eigenvectors are then sorted such that the largest eigenvalue is associated with the first principal component explaining the most variance [13]. 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 higher than the number of samples, which might apply to observables, [14] advises to use Singular Value Decomposition (SVD) according to

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T \qquad (4)$$

to determine the eigenvalues and eigenvectors efficiently. The N eigenvalues of $\mathbf{X}^T\mathbf{X}$ can then be extracted from the diagonal matrix $\mathbf{S}^T\mathbf{S}$ by

$$\lambda_n = \frac{1}{N-1}(\mathbf{S}^T\mathbf{S})_{nn}, \qquad (5)$$

and the orthonormal matrix $\mathbf{V}$ contains the associated eigenvectors $\mathbf{P}$ of $\mathbf{X}^T\mathbf{X}$. The data in the new coordinate system $\mathbf{T}$ can finally be determined by a matrix multiplication of $\mathbf{U}$ and $\mathbf{S}$.

D. Statistical Process Model

For each requested observation point in time t_{cd}, the system collects previously sampled observables $\mathbf{o}(t_{0d}), \ldots, \mathbf{o}(t_{cd})$ and current state variables $\mathbf{s}(t_{cd})$ as shown in Figure 2. Thereby, either the complete history $(t_{01} = t_{02})$ or an adjusted history of selected preceding time frames $(t_{01} \neq t_{02})$ is used as a basis for state estimation.

The statistical process model for hidden state observation at one particular observation point in time t_c is divided into a training and a prediction block as indicated in Figure 3. First, PCA is applied to both observables $\mathbf{o}$ and state variables $\mathbf{s}$, of which a subset is used to train the ANN as input $\mathbf{c}_o$ and target $\mathbf{c}_s$, respectively. After successful training, the ANN can predict state variables in reduced dimensions $\mathbf{c}_{\hat{s}}(t_c)$ from previously unseen observables $\mathbf{o}(t_0), \ldots, \mathbf{o}(t_c)$, reduced to $\mathbf{c}_{\tilde{o}}(t_0, \ldots, t_c)$, that have not been included in training. The

Figure 3. Architecture of the statistical process model

predicted state quantities $\mathbf{c}_{\hat{s}}(t_c)$ are subject to an inverse dimension transformation to obtain their counterparts $\hat{\mathbf{s}}(t_c)$ in the original, high dimensional space for visualization and validation.

III. APPLICATION TO DEEP DRAWING

The feasibility of the proposed approach is tested with two sample processes for 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. An axisymmetric 2D deep drawing model (Figure 4 left) represents an elementary sample process, whereas a 3D deep drawing model with anisotropic plastic material behavior (Figure 4 right) describes a complex sample process. Anisotropic plasticity is expressed by a direction-dependent forming resistance in the material resulting in an earing profile in deep drawing. This material behavior might be induced by a preceding rolling process and is undesired in this context, but may be compensated by process control. Details about the processes of deep drawing and rolling are contained in [15].

Statistical samples are generated by experiments performed in a numerical simulation environment. For this purpose, two finite element deep drawing models have been implemented in ABAQUS (finite element analysis software).

$$\mathbf{s}(t_{c1}) = \mathbf{f}_1(\mathbf{o}(t_{01}), \ldots, \mathbf{o}(t_{c1}))$$
$$\mathbf{s}(t_{c2}) = \mathbf{f}_2(\mathbf{o}(t_{02}), \ldots, \mathbf{o}(t_{c2}))$$

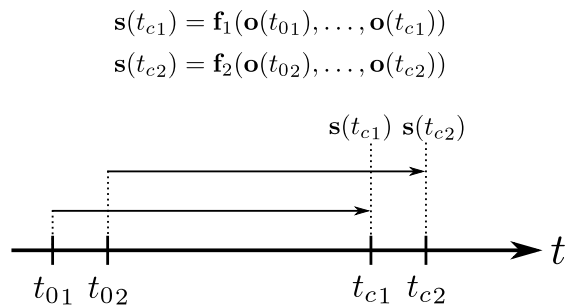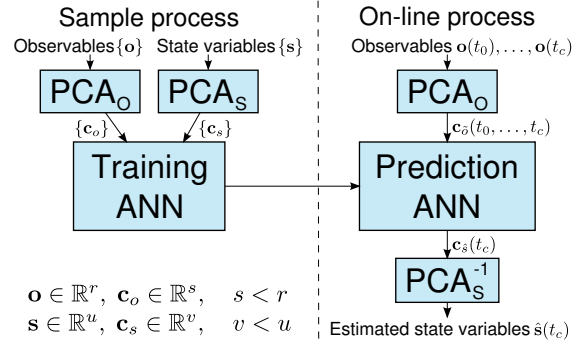
Figure 2. State observation by observable histories for different observation points in time t_{cd} (for $d = 1, 2$)
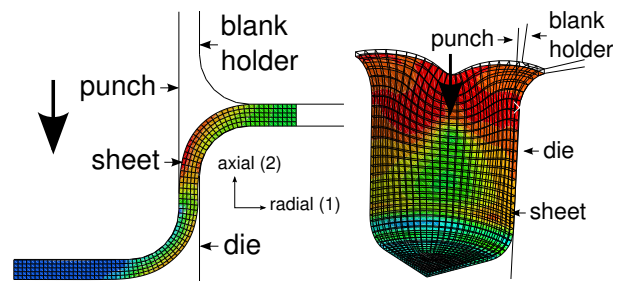

Figure 4. Workpiece and tools in cup deep drawing for the elementary sample process (left) and the complex sample process (right)

A. Elementary Sample Process

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

In the performed parametric study, the blank holder force has been varied in the range of [70, 100] kN, whereas 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 9 observables and 400 state variables. The extracted data are randomly partitioned into a training set (80%) and a test set (20%). Dimension reduction is applied to all samples, where the training data are split again randomly into a training set (80%) and a validation set (20%) for the following regression analysis. The test set is used for an overall validation of the statistical model. Resampling and subsequent remodeling is performed to select the best model and to prove independence of specific data sets.

B. Complex Sample Process

Displacements in the sheet edge in different directions ($0°$, $45°$, $90°$ with reference to the rolling direction) are selected as observable quantities. Also, reaction forces in the punch and logarithmic strains reflecting the minimum and maximum forming grade in selected sheet wall locations are observed during deep drawing. Altogether, 12 observables are recorded for each of the common 368 time frames. The process state is again given by the von Mises stress distribution, this time in 960 elements. 499 experiments have been executed under variation of the blank holder force in the range of [4000, 6000] N with otherwise constant process conditions. The partition of the data into a training set, a validation set and a test set is performed as for the elementary sample process. Again, multiple resampling runs are accomplished and evaluated by their obtained results.

IV. DISCUSSION OF THE RESULTS

The state prediction results of the statistical model, which has been applied to the two sample processes are analyzed. The common underlying prediction characteristics are defined in Section IV-A. Results for the particular instances are given in Section IV-B for the elementary sample process and in Section IV-C for the complex sample process.

A. Prediction Characteristics

The prediction characteristics are computed over all samples from the test set for multiple resampling runs with different random initial conditions to show the independence of specific data sets. Random initial conditions appear in the selection of data (training, validation and test sets) and in the initialization of the ANN weights before training starts. The best results are then presented in the following sections.

The quality of the statistical model is quantified by the coefficient of determination by

$$R^2 = 1 - \frac{SSE}{SST}, \qquad (6)$$

which can be applied to nonlinear regression analysis [16]. The sum of squared errors given by

$$SSE = \sum_{n=1}^{N} \sum_{l=1}^{L} (y_l - \hat{y}_l)_n{}^2 \qquad (7)$$

describes the sum of the squared deviations between the original data in the test set y_l and the associated predicted results $\hat{y}_l$. The SSE is calculated over all dimensions in the predicted quantity l, i.e., the number of state variables, and all samples n. It is divided by the SST, which quantifies the total variation in the test set calculated by the summed squared deviations of the original data from their means. The R^2 lies between 0.0 and 1.0, where a high value indicates a good model fit and and a low value reveals a poor fit.

The root mean square error according to

$$RMSE = \sqrt{MSE} = \sqrt{\frac{SSE}{N}} \qquad (8)$$

is a further measure of model quality, where *MSE* stands for the mean squared error, and *N* corresponds to the number of samples in the test data set. The *RMSE* can be used to compare different models in the same complexity category, i.e.., in order to select the best model from multiple resampling runs.

The relative prediction error for each variable dimension l defined by

$$RE_l = \left| \frac{Target_l - Prediction_l}{Target_l} \right| \text{ for } l = 1 \ldots L, \qquad (9)$$

quantifies the error percentage, where L is the number of variable dimensions in the predicted quantity and the target quantity comes from the training data. The resulting distribution can be characterized by the mean relative error RE_μ and the maximum relative error RE_{max} with respect to L and the number of samples N.

In order to compare the variation of the predicted results to the variance of the generated data, we have defined the model uncertainty by

$$U = \frac{1}{L} \sum_{l=1}^{L} \frac{MSE_l}{Var_l}, \qquad (10)$$

which corresponds to the mean value of the *MSE* of each variable l in relation to its variance Var. In (10), L stands for the number of variable dimensions. A low model uncertainty is characterized by a U value close to zero, whereas values approaching 1.0 indicate a high uncertainty. The objective is to show that the variation of the predicted results is substantially smaller than the variance of the generated data.

B. Elementary Sample Process

Two use cases are identified for the state estimation based on the data from the elementary sample process. The first use case refers to the prediction of the final process state based on the collected observables during process execution. This provides a subsequent process with detailed information about its input, allowing it to optimally adjust its parameters. The individual 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-B1. On the other hand the prediction of the state evolution during process execution can be applied to process control as discussed in Section IV-B2. The latter can be considered as a generalization of the former use case.

1) Prediction of the Final Process State: The statistical model for hidden state observation of the elementary sample process is validated by a test set of 40 samples (see Section III-A). The relative prediction error of the 400 state variables never exceeds 0.0110 for all samples, the resulting distribution is shown in Figure 5. 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 feasibility of the method in principle. 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 only the blank holder force has been varied.

The prediction quality was further analyzed as follows. The model uncertainty U amounts to 0.0045, which indicates a high accuracy and a low uncertainty of the predicted results. A resulting R^2 value of 0.9991 and a corresponding *RMSE* value of 0.2726 confirm the good quality of the statistical model. The MSE of the predicted state variables serves as a base to determine a confidence interval for the prediction error. The precision of the estimation amounts to
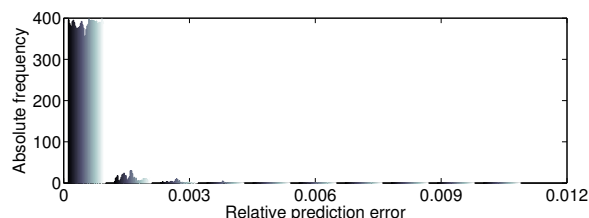

Figure 5. Relative error distribution for predicted state variables of the elementary sample process

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 resulting 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. The observables are reduced from 1179 (9 observables per time frame × 131 time frames) to 9 dimensions with a predefined precision of 99.999% and thus a relative error of 0.001%. The state variables are reduced from 400 to 7 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 and de-correlated data in a more reliable and robust way. A worse result might have been obtained without dimension reduction due to the huge number of additional degrees of freedom of the ANN.

Some results for a representative of the test set visualized in ABAQUS are depicted in Figure 6. It displays the absolute von Mises stress values in MegaPascal units predicted by the statistical model in Figure 6b, which are in very good agreement with the results of the finite element model illustrated in Figure 6a. To outline the deviation of the predictions from the original data, the relative error in the range of [0, 0.0024] is presented in Figure 6c. Errors are low in regions with small deformations, while higher but still small 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 have first applied a white noise of 5% to the observables to model a measurement error. The state variables have then been predicted with a relative error in the range of [0.0, 0.0581] and a corresponding mean value of 0.0029. The model uncertainty U amounts to 0.1610, while the model quality is characterized by a R^2 value of 0.9128 and a *RMSE* value of 2.8156. Increasing the noise to 10% results in a relative error range of [0.0, 0.0943] with a mean of 0.0038, a model uncertainty U of 0.2379, a R^2 value of 0.7830 and a *RMSE* value of 4.1761. The size of the error range does not solely represent the quality of the 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 the observables. However, with increasing noise, the model quality decreases as the uncertainty increases.

2) 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 statistical model offers the selection of time frames that are crucial for control. In this work, some

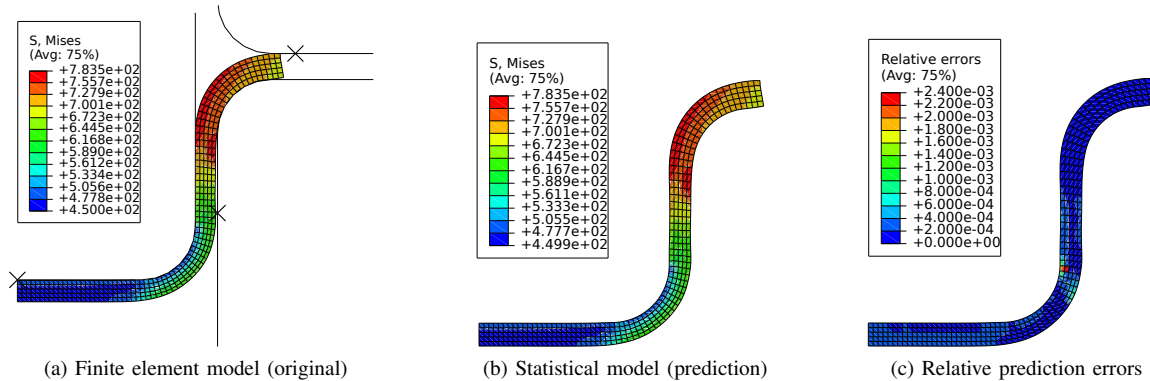(a) Finite element model (original)　　(b) Statistical model (prediction)　　(c) Relative prediction errors

Figure 6.　Comparison of results of the finite element model and the statistical model for the elementary sample process

representatives are chosen to demonstrate the feasibility of the prediction of the state evolution for the elementary sample process. The time frame numbers 1, 45, 90 and 131 are selected, the results are outlined in Table I.

The respective grade of determinacy of the ANN is identified incrementally by evaluating the network's performance function (see Section II-B), the precision for dimension reduction is 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 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 U 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 the complexity grows and has a stronger impact on the prediction. The overall model quality expressed by the R^2, the *RSME* and the *RE* shows that the predictions are in good agreement with the original data.

C. Complex Sample Process

The validation of the statistical model for hidden state observation of the complex sample process is performed by a test set of 99 samples (see Section III-B). We have learnt from the elementary sample process that the observation of early process states can be neglected. In this time, there is only a very short history of observables available and the prediction of the current process state is characterized by a high model uncertainty. The cause is a low variation in the process state, since the impact of different applied values for the process parameters is not yet strong at the very beginning and therefore not crucial for process control. Thus, the time frame numbers 92, 184, 276 and 368 are selected for the state prediction based on a history of observable quantities.

In a first step, the complete history is chosen as for the elementary sample process. This use case is described in Section IV-C1. In order to further reduce the complexity of the regression relation, the history is limited by a fixed number of time frames in terms of a sliding window. In addition, the precision of dimension reduction performed on the observables is increased to explain more variance in the state variables as the result. Details about this use case are

Table I
PREDICTION CHARACTERISTICS DURING EXECUTION OF THE
ELEMENTARY SAMPLE PROCESS

Time frame number	001	045	090	131
# PCA observables	1	1	6	9
# PCA state variables	1	2	3	7
# ANN hidden nodes	33	36	26	38
ANN grade of 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
RMSE statistical model	0.0061	0.0186	0.0469	0.2726
U statistical model	0.1452	0.0003	0.0028	0.0045
RE_μ statistical model	0.0047	$8 \cdot 10^{-6}$	$3 \cdot 10^{-5}$	$2 \cdot 10^{-4}$
RE_{max} statistical model	0.1071	0.0021	0.0020	0.0110

outlined in Section IV-C2. In Section IV-C3, the results of the two history variants are compared to each other with respect to the sensitivity to random initial conditions.

1) State Prediction with Complete History: For the state prediction based on the complete history of observables, the predefined precision of dimension reduction in observables and state variables is chosen as 99.90% and 99.00%, respectively. The results are presented in Table II. The MSE induced by PCA in observables and state variables grows with increasing time. In the case of time frame number 276, both values are greater than the ones of the last time frame with number 368. The number of reduced dimensions in observables does not change between time frame number 276 and 368. This means that a longer history leads to a smaller MSE with the same number of reduced dimensions. The complete history of 368 time frames in fact entirely contains the selected history for the prediction of time frame number 276. As a consequence, the MSE proportion in the remaining 92 time frames must be smaller than the ones of the common history interval. The increasing confidence in the prediction is also indicated by a lower model uncertainty U for time frame number 368. Compared to the elementary sample process, the MSE induced by PCA in observables is much smaller, which is the result of a generally smaller variance in the observables of the complex sample process. The complexity of the regression relation in terms of the hidden ANN nodes increases with time. Time frame number 184 represents an exception, since the number of reduced dimensions in observables is smaller than the number of reduced dimensions in state variables.

The model quality expressed by the R^2 is lower compared to the predictions performed on data from the elementary sample process. Additionally, the upper bounds of the relative error distribution are much higher for predictions based on the complex sample process. A possible reason might be the high complexity of the regression relation in terms of ANN nodes. On the one hand, a more sophisticated model is necessary to describe complex material behavior as plastic anisotropy. But on the other hand, the high complexity

needs to be reduced to the crucial characteristics to obtain a generalizable model. This can be achieved by adapting the history of observables by use of the following procedure.

2) State Prediction with Adjusted History: In this use case, we adjust the history of observables by limiting the number of time frames in terms of a sliding window. A variable history length makes the approach very flexible. At the same time, the question arises what might be the minimum history length to describe the current process state unambiguously? Further efforts may be investigated to realize a self-adaptive history with reference to performance and accuracy criteria. The sliding window is set to a fixed length of 92 times frames, which is one quarter of the complete history length and seems to be sufficient for state prediction in this case. That means that the process state for the currently selected time frame is predicted by a history of the preceding 92 time frames. This adaptation indeed reduces the complexity, but does not yield better prediction results. Furthermore, this reduced history approach is sensitive to random initial conditions as it is also observed for predictions based on the complete history in Section IV-C1.

In order to explain more variance in the state variables as the regression output, the precision of the dimension reduction procedure performed on the observables is increased to 99.99%. The combination of reducing the number of time frames in the history and inflating the remaining observables by a higher precision represents the adjusted history approach. This procedure yields similar or even better prediction results with a lower complexity in terms of number of ANN nodes. It also reduces the sensitivity to random initial conditions as demonstrated in Section IV-C3.

The prediction results with adjusted history and data from the complex sample process are summarized in Table III. Note that the history is not adjusted for early predictions up to time frame number 92. Despite of a shorter history, the number of dimensions in the reduced observable space is about twice the size of the number of principal components with lower precision in the complete history variant. The increase in the precision seems to have a stronger effect

Table II

PREDICTION CHARACTERISTICS DURING EXECUTION OF THE COMPLEX SAMPLE PROCESS WITH COMPLETE HISTORY OF OBSERVABLES

Time frame numbers	001-092	001-184	001-276	001-368
# PCA observables	7	6	11	11
# PCA state variables	3	7	7	10
# ANN hidden nodes	73	114	98	121
ANN grade of determinacy	1.2	1.4	1.2	1.2
MSE PCA observables	0.0337	0.1103	0.5211	0.4539
MSE PCA state variables	0.0066	0.0213	0.0802	0.0713
MSE ANN	$8 \cdot 10^{-6}$	0.0049	0.0044	0.0076
R^2 statistical model	0.9918	0.9914	0.9844	0.9822
RMSE statistical model	0.0829	0.1455	0.3694	0.3699
U statistical model	0.0781	0.0319	0.0656	0.0409
RE_μ statistical model	0.0001	0.0003	0.0013	0.0021
RE_{max} statistical model	0.0130	0.0338	0.3295	0.4194

Table III

PREDICTION CHARACTERISTICS DURING EXECUTION OF THE COMPLEX SAMPLE PROCESS WITH ADJUSTED HISTORY OF OBSERVABLES

Time frame numbers	001-092	092-184	184-276	276-368
# PCA observables	7	10	20	24
# PCA state variables	3	7	7	10
# ANN hidden nodes	73	89	62	76
ANN grade of determinacy	1.2	1.4	1.3	1.2
MSE PCA observables	0.0337	0.0220	0.1472	0.0126
MSE PCA state variables	0.0066	0.0213	0.0802	0.0713
MSE ANN	$8 \cdot 10^{-6}$	0.0004	0.0068	0.0036
R^2 statistical model	0.9918	0.9911	0.9882	0.9840
RMSE statistical model	0.0829	0.1554	0.3524	0.3395
U statistical model	0.0781	0.0350	0.0519	0.0458
RE_μ statistical model	0.0001	0.0003	0.0012	0.0025
RE_{max} statistical model	0.0130	0.0350	0.1342	0.2860

on the number of reduced dimensions than the decrease in the history to one quarter of its entire length. The MSE induced by dimension reduction in observables is much smaller due to the higher requested precision. For time frame number 276, this quantity is still higher compared to the other time frames, even though 20 components are extracted to describe the reduced space. There might be some nonlinear behavior in dominant observables around this point, which cannot be explained adequately by linear dimension reduction methods. The adjustment in the history leads to a kind of linearization by limiting the number of time frames and thus yields more reliable prediction results.

The prediction characteristics of the statistical model, namely the R^2, the *RMSE* and the uncertainty U are similar to the results of the complete history variant. Indeed, the maximum of the relative prediction error RE_{max} is reduced substantially by the adjusted history variant. The relative error distribution based on 960 state variables and 99 test samples with an upper bound of 0.2860 is depicted in Figure 7. The absolute frequency is high for small errors and decreases rapidly for higher errors. The individual samples can be distinguished by the different colors in the histogram.

A representative is selected from the test set for predicted state variables at the end of the complex sample process with an adjusted history of observables. The obtained results are visualized in ABAQUS as presented in Figure 8. The absolute von Mises stress values at a typical range of [1, 300] MPa, which are calculated by the finite element model are depicted in Figure 8a. The predicted von Mises stresses are displayed in 8b. They are in good agreement with the original results. The relative error in the range of [0, 0.0449] is illustrated in Figure 8c to compare the prediction with the original results. As also observed for the elementary sample process, the largest errors occur in the area of high deformation gradients as in the rounding of the sheet wall.

3) Sensitivity to Random Initial Conditions: The two variants with complete and adjusted history are compared with reference to their sensitivity to random initial conditions. Random initial conditions appear in the selection of training, validation and test data and in the initialization

of the ANN weights before training. Figure 9 visualizes the sensitivity to random initial conditions observed on the *RMSE* over 10 resampling runs for both history variants. On the left, the error intervals are illustrated for the complete history, whereas the ones for the adjusted history are depicted on the right. The intervals are the same for the predicted state variables up to time frame number 92, since the history has not been adjusted at the early stage of prediction. For all other time frames, the error intervals decrease substantially in case of the adjusted history variant. This means that the sensitivity to random conditions is much less and yields a higher reliability in the prediction.

The distribution characteristics $RMSE_\mu, RMSE_\sigma$ and $RMSE_{max}$ for the corresponding time frame numbers 184, 276 and 368 are compared in Table IV for the complete and the adjusted history. The $RMSE_\mu$ is minimized up to one half with an adjusted history. An improvement is also achieved in a reduced variation $RMSE_\sigma$ ($\approx$ up to one forth) and a reduced maximum $RMSE_{max}$ ($\approx$ up to one forth). In particular, the *RMSE* for the prediction of time frame number 276 is characterized by much lower error bounds. The adjusted history of observables used for prediction at this point in time has already attracted attention by higher errors in context of its reduced space in Section IV-C2.
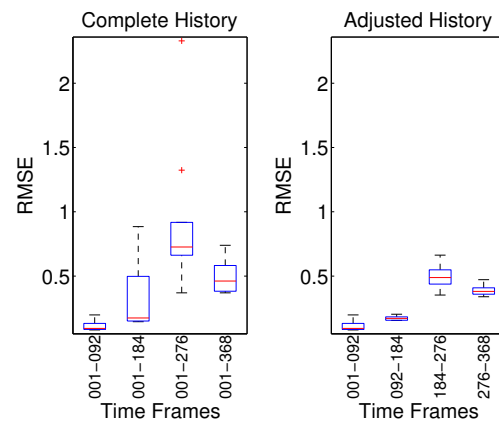

Figure 9.   RMSE sensitivity w.r.t. random initial conditions for different time frames with complete history (left) and adjusted history (right)
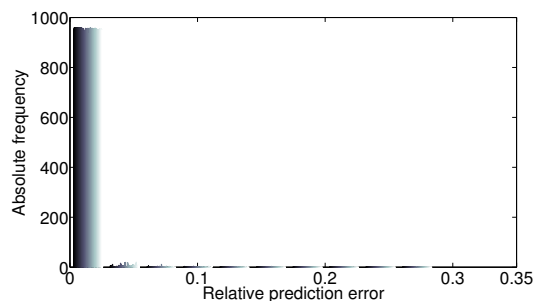

Figure 7.   Relative error distribution for predicted state variables of the complex sample process with adjusted history of observables

Table IV
DISTRIBUTION CHARACTERISTICS OF RMSE SENSITIVITY W.R.T. RANDOM INITIAL CONDITIONS FOR DIFFERENT TIME FRAMES

Distribution characteristics	$RMSE_\mu$	$RMSE_\sigma$	$RMSE_{max}$
Time frames 001-092	0.1156	0.0419	0.1983
Time frames 001-184 (complete)	0.3313	0.2851	0.8845
Time frames 092-184 (adjusted)	0.1727	0.0168	0.2032
Time frames 001-276 (complete)	0.9129	0.5572	2.3296
Time frames 184-276 (adjusted)	0.5034	0.0966	0.6631
Time frames 001-368 (complete)	0.4948	0.1282	0.7395
Time frames 276-368 (adjusted)	0.3857	0.0362	0.4722

(a) Finite element model (original)   (b) Statistical model (prediction)   (c) Relative prediction errors
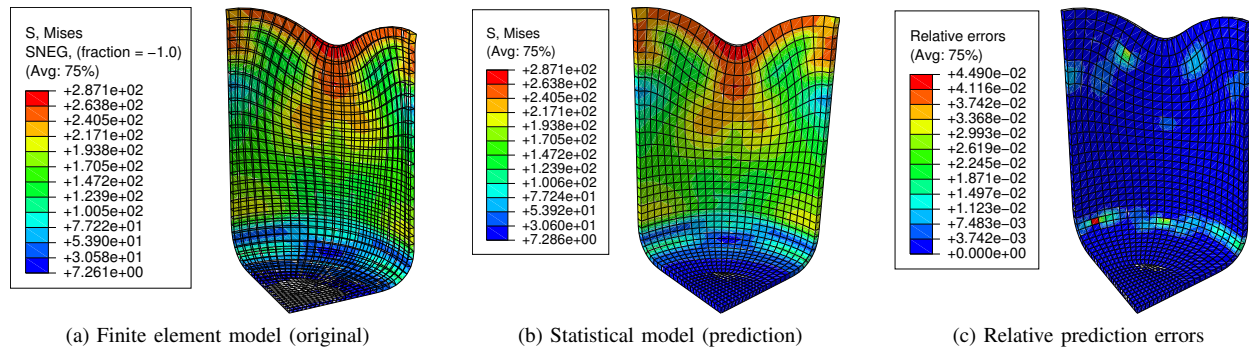
Figure 8. Comparison of results of the finite element model and the statistical model for the complex sample process with adjusted history of observables

V. Conclusion and Future Work

A universal statistical process model for hidden state observation based on the nonlinear regression analysis of a history of observables and the current state by means of an ANN has been created. PCA as a linear dimension reduction method is applied to input and output separately before performing the regression analysis. The complexity of the relation between observables and state variables can be adjusted by the grade of determinacy of the ANN and by an adaptive history of observables. This makes the approach very flexible with reference to performance and accuracy.

The presented statistical process model has been applied to two deep drawing processes of different complexity. It has been shown that the universal model adapts well to the specific data of the different sample processes. The resulting model instances can be successfully used for state prediction based on observations. For the complex sample process, the sensitivity to random initial conditions has been reduced by an adjusted history of observables, such that more reliable prediction results are achieved. 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. The predefined state can then serve as set value for the optimization procedure.

One drawback of the statistical process model for hidden state observation 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. By observing an entire process chain, the final state information of the preceding process can be used as reliable characterization at the begin of the current process. Significant time frames for the observation of the process state evolution have to be identified to enable process control. The proposed universal approach for the observation of hidden states in adaptive process controls may be transferred to any process characterized by state variables that can be derived from related observable quantities.

Future work includes the extension of the statistical process model for hidden state observation to a local process control based on extracted features. For this purpose, dimension reduction shall be integrated into the regression analysis, such that the resulting compact feature space represents the relation between observables and state variables. The extracted features may then serve as a base for an efficient process control. Therefore, nonlinear dimension reduction methods are taken into account to obtain a more general description with a feature space as small as possible.

Acknowledgment

References

[1] M. Senn and N. Link, "Hidden state observation for adaptive process controls," in *Proceedings of the Second International Conference on Adaptive and Self-adaptive Systems and Applications, ADAPTIVE 2010*, pp. 52 – 57.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

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

[10] 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.

[11] 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.

[12] 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.

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

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

[15] W. F. Hosford and R. Caddell, *Metal Forming: Mechanics and Metallurgy*, 4th ed.   Cambridge University Press, 2011.

[16] 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.