Control of an Industrial Distillation Column Using a Hybrid Model with Adaptation of the Range of Validity and an ANN-based Soft Sensor

Mohamed Elsheikh*, Yak Ortmanns, Felix Hecht, Volker Roßmann, Stefan Krämer, and Sebastian Engell

DOI: 10.1002/cite.202200232

Advanced control schemes such as model predictive control can be used to minimize the use of resources while guaranteeing the specified product quality. In this paper, we consider an industrial mother liquor distillation column varying flow rate and composition of the feed. There are specifications of the composition for all product streams. To address this challenging control problem, we employ a nonlinear model-predictive controller using a hybrid model, which consists of a simple phenomenological model augmented by a data-based component to compensate the plant-model mismatch. The trustworthiness of the data-based model is addressed using a domain of validity of the data-based model, which is estimated using a one-class support vector machine. During operation, it may turn out that the model is also reliable in a wider range, therefore, data of recently visited operating points is recorded and the domain of validity is extended if the model is sufficiently accurate. To improve the performance of the controller, an artificial neural network model is used to estimate the product composition from available measurements.

Keywords: Distillation columns, Domain of validity, Hybrid modeling, Model predictive control, Neural networks, Soft sensing

Received: December 16, 2022; *revised:* March 23, 2023; *accepted:* May 02, 2023

1 Introduction

1.1 Hybrid Models for Model Predictive Control of Distillation Columns

Distillation processes are widespread in the chemical industry and major consumers of energy. Their operation is particularly challenging when the feed flow rate and the feed composition change frequently, as is the case in mother liquor distillation where a mixture from batch processes is split into fractions such that most of the material can be recycled. Model predictive control (MPC) schemes have been used for the control of different types of distillation columns, e.g., [1, 2]. Linear MPC is well suited for maintaining a desired operating point or pushing constraints and rejecting process disturbances. For the optimal operation of processes with varying operating points, MPC schemes based on accurate nonlinear models promise a better performance, but to build and implement such models is a bottleneck in the application of nonlinear MPC. Even if high-fidelity process simulators are available, such simulators can typically not be used as the system model within the framework of MPC. As a result, when realizing a model-based controller of a distillation column, one is faced with the problem that first-principles-based modeling requires a significant effort and can result in large-scale differential-algebraic equations (DAE) models that may not be suitable for real-time implementation. Reducing the modeling effort and the computation time while at the same time predicting the future process behavior accurately is therefore highly desired. The availability of large amounts of recorded data motivates the use of data-based models, however, to fit such models, one needs a large number of data points over a broad range of inputs, especially if many manipulated variables are used to control the process [3]. If sufficiently rich data is not available from the real plant, high-fidelity process simulators can be employed to parameterize data-

[1]Mohamed Elsheikh (Mohamed.elsheikh@tu-dortmund.de), [2]Yak Ortmanns, [2]Felix Hecht, [2]Volker Roßmann, [1,2]Stefan Krämer, [1]Sebastian Engell

[1]TU Dortmund, Process Dynamics and Operations Group, Department of Biochemical and Chemical Engineering, Emil-Figge-Straße 70, 44227 Dortmund, Germany.

[2]Bayer AG, Engineering & Technology, 51368 Leverkusen, Germany.

based models as in [4]. However, this does not solve the problem of the effort required to develop sufficiently accurate models, especially for systems with complex thermodynamic properties. A promising approach to overcome these issues is to combine mechanistic models of medium complexity that can be set up relatively easily and data-based models that are trained using operational data with the goal to obtain a better accuracy in the regions where data was available and extrapolation capabilities also in regions with scarce data [5]. Such models are called grey-box or hybrid models. They can be realized by embedding data-based submodels into a mechanistic model, e.g., [6–8], or by using a parallel structure as investigated in [9] for a blast furnace. Hybrid-modeling approaches have gained increasing popularity for separation processes, see [10–12].

Hybrid models, however, inherit the problem of their trustworthiness from the data-based model elements and care has to be taken that the predictions are reliable. Generally, data-based models can be trusted only in regions that are dense in training data, called the domain of validity. The domain of validity can be represented by box constraints of the inputs of the data-based model or by the convex hull of the training data. However, both approaches can lead to an overestimation of the domain of validity. A novel approach was proposed in [13] where the domain of validity is classified by a one-class support vector machine (SVM) overcoming those limitations. In this paper, we apply the one-class SVM to represent the domain of validity of the data-based component of the hybrid model, which is faded out when the current operation is not within this domain.

1.2 Employing Soft Sensors for Missing Measurements

Soft sensors use available measurements to predict variables that cannot be measured online or where the effort to install online measurements is considered too large. Such sensors can be based on structured models or on data-based models such as artificial neural networks (ANN) [14, 15]. For the problem at hand, simulation studies showed that good control with conventional PID-based control structures with feedforward is only possible if measurements of the compositions of the product streams and of the feed are available. In principle, these can be provided by online analytics (PAT), but they are not installed at the plant. Therefore, we employ an ANN-based soft sensor for the compositions of the product streams.

1.3 Proposed Overall Control Concept

In this paper, we propose and demonstrate a nonlinear model-predictive control scheme for a mother liquor distillation problem that is based on a hybrid model with a weighted contribution of the data-based model component as proposed in [16]. An ANN-based soft sensor is trained to estimate the compositions of the product and sump streams of the distillation column. The paper is structured as follows: The process under consideration is presented in Sect. 2. In Sect. 3, the ANN-based soft sensor for the product compositions is explained. The hybrid modeling approach is described in Sect. 4. In Sect. 5, the nonlinear MPC scheme based on the hybrid model with adaptation of the domain of validity is presented and in Sect. 6 the results of applying the ANN-based estimator and the control scheme to a faithful model of the considered column are shown. Finally, we summarize the main aspects and discuss future research directions in Sect. 7.

2 Process Description

The considered plant is a continuously operated mother liquor distillation column with a varying feed rate that is generated by the upstream batch processes. A schematic of the distillation column is shown in Fig. 1. The feed consists of water, a low-boiling solvent, a high-boiling component, and other impurities. The distillate and side streams are considered as the product streams of the column. The top stream consists of the solvent and water, whereby the mass fraction of the low-boiling solvent w_{lb} in the distillate stream should be at a specified level. The side stream mainly consists of water, which should have a mass fraction w_w above a specified value. The sump stream is a waste stream that consists of water, the high-boiling component, and impurities, where the mass fraction of the high-boiling component w_{hb} should be within a certain range to save energy in the downstream process and to avoid fouling in the column sump and in the reboiler. The levels of the

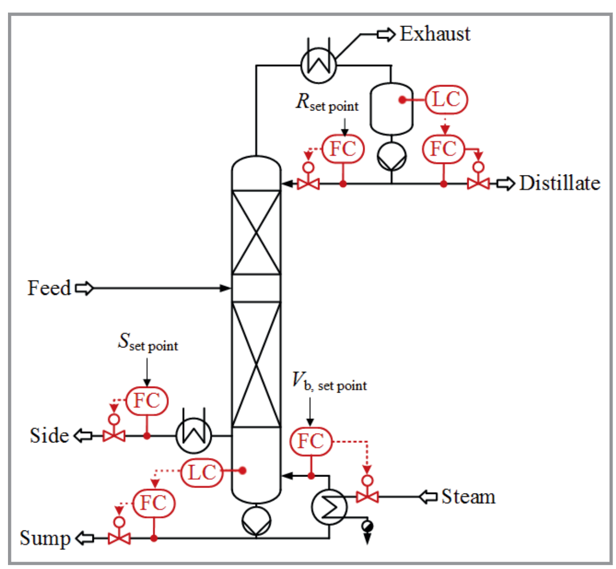

Figure 1. Distillation column schematic.

distillate receiver and of the column sump are maintained by level controllers that act on the distillate and sump flow rates. Flow controllers are employed to control the flowrates of the side stream, the reflux stream, and the steam flow. A detailed model of the column is available in Unisim Design [17] and is used as a virtual plant in this work.

3 ANN-Based Soft Sensor

An ANN-based soft sensor was trained to estimate the controlled variables $\{w_{lb}, w_{hb}, w_w\}$ in the product and sump streams. Online measurements of various temperatures and flowrates are available. Based on process knowledge, some variables are a priori not considered in estimating the product compositions. We used sequential forward selection (SFS) in order to determine the most influential variables among the variables that contain information about the product compositions [18]. SFS finds the best combination of the least number of variables by assessing all possible numbers and combinations of the considered variables in terms of prediction performance. Four different temperatures were selected as the most influential variables for the required compositions using SFS. The current and the past five measurements of these temperatures are used as inputs to the ANN. Tab. 1 contains the inputs and outputs of the ANN-based estimator of the product compositions.

This ANN-based soft sensor has two hidden layers with 40 neurons on each layer and uses rectified linear unit (ReLU) activation functions for the hidden layers and a linear activation function for the output layer. The selected hyper-parameters of the ANN result in accurate predictions of the compositions. Simulation data set consisting of 11 000 data points was used in the identification process of ANN-based soft sensor, the ANN was trained using 75 % of the data and the other 25 % of the data was used for model validation.

4 Hybrid Model with a Weighted Contribution of the Data-Based Model

4.1 Model Description

We use the hybrid modeling approach proposed in [16], which consists of a simplified mechanistic model and a parallel data-driven model to compensate for the steady-state behavior and the dynamics of the plant that are not captured by the simplified model. The data-based component of the hybrid model is weighted based on the domain of validity of the data-based component such that the full model is used in the training region and the contribution of the data-based model fades out smoothly in the regions with insufficient training data. The hybrid process model can formally be written as:

$$x_{k+1} = f_{diff}(x_k, z_k, u_k) \tag{1}$$

$$0 = f_{alg}(x_k, z_k, u_k) \tag{2}$$

$$y_{k+1} = f_{out}(x_{k+1}, z_{k+1}) \tag{3}$$

$$e_{k+1} = f_{ML}(h_k) \tag{4}$$

$$\tilde{y}_{k+1} = y_{k+1} + \gamma_{k+1} e_{k+1} \tag{5}$$

where x_k denotes the vector of states of the mechanistic model, z_k denotes the algebraic states, u_k denotes the vector of the manipulated inputs, y_k represents the output vector of the mechanistic model, e_k is the vector of the errors between the outputs of the full model and the simplified model, $\tilde{y}_k$ denotes the output vector of the hybrid model and h_k represents the input vector of the data-based model. γ_k is the weighting factor of the data-based model. f_{diff}, f_{alg}, f_{out}, and f_{ML} are vectors of nonlinear functions.

4.2 Estimation of the Domain of Validity of the Data-Based Model

To estimate the domain of validity of the data-based model element, a one-class SVM is trained on lower-dimensional projections of the inputs of the data-based submodel f_{ML}. Principal component analysis (PCA) is used to obtain the projections of the training data [19]. Before applying PCA, the training data was centered and scaled. The projection T of the training data is obtained by $T = XW$ where $T \in \mathbb{R}^{N_{TP} \times l}$ is the projection of the data, $X \in \mathbb{R}^{N_{TP} \times i}$ is the centered and scaled training data, $W \in \mathbb{R}^{i \times l}$ is the loading matrix, N_{TP} is the number of training data points, i is the number of inputs of the data-based model in training data and l is the number of the used principal components. The number of principal components (PC) is selected such that

Table 1. Inputs and outputs of the ANN-based estimator of the product compositions.

Inputs	Outputs
The current and the past 5 measurements of the following temperatures: – Temperature at the top of the column – Temperature at the middle of the rectification section – Temperature at the middle of the stripping section – Temperature of the liquid in the reboiler	Mass fraction of water w_w in the side stream Mass fraction of the high boiling component w_{hb} in the bottom stream Mass fraction of the low boiling solvent w_{lb} in the distillate stream

at least 95 % of the variance in the training data is explained.

A one-class SVM is a special type of SVM that is employed when the training data belong only to one class, where it seeks a maximum-margin hyperplane to separate the data from the origin in the feature space [20]. It was proposed in [13] to use the decision function $f_{DF}(h)$ of the one-class SVM to find out if a data point h lies in the domain of validity of a data-based model by checking the inequality $f_{DF}(h) \geq 0$. In this work, the one-class SVM is trained using the projections of the training data on the low-dimensional space defined by the selected principal components, i.e., $f_{DF} = f_{DF}\left(h_k^{s,c} W\right)$ where $h_k^{s,c}$ is the vector of the scaled and centered inputs of the data-based model.

4.3 Weighting of the Contribution of the Data-Based Model

As proposed in [16], the contribution of the data-based element of the hybrid model is weighted based on the decision function f_{DF} of the estimated domain of validity. The idea is to add the contribution of the data-based model to the overall model inside the domain of validity, while the contribution of the data-based model smoothly diminishes outside the domain of validity of the data-based model component. The weighting factor γ_k of the data-based component at time step k is defined by:

$$\gamma_k = \zeta \Gamma\left(h_k^{s,c}\right) + (1-\zeta)\gamma_{k-1} \tag{6}$$

where Γ is a sigmoidal function of the decision function $f_{DF}\left(h_k^{s,c} W\right)$. The weighting factor γ_k is exponentially smoothed with $0<\zeta<1$.

The weighting function Γ is given by:

$$\Gamma\left(h_k^{s,c}\right) = \frac{1}{1+\exp\left(c \times f_{DF}\left(h_k^{s,c} W\right) + \ln\left(\frac{\varepsilon}{1-\varepsilon}\right)\right)}; \quad c = \frac{2\ln\left(\frac{1-\varepsilon}{\varepsilon}\right)}{\mu} \tag{7}$$

where μ is the value of the decision function f_{DF} when Γ is equal to ε. If f_{DF} equals zero, Γ is equal to $1-\varepsilon$. The tuning parameters of Γ are ε and μ. The weighting function Γ is designed to avoid discontinuities in the process model which can result in numerical issues and abrupt control actions.

5 Model Predictive Control with Adaptation of the Weighting Factor

The optimization problem that is solved by the nonlinear model-predictive controller at each time step is a classical tracking error minimization with control move penalization:

$$\min_u \sum_{k=1}^{N_p} \left(\tilde{y}_k - \tilde{y}^{SP}\right)^{\mathrm{T}} Q_{\tilde{y}} \left(\tilde{y}_k - \tilde{y}^{SP}\right) + \sum_{k=1}^{N_p-1} \Delta u_k^{\mathrm{T}} Q_{\Delta u} \Delta u_k \tag{8}$$

s.t. Eq. (1)–(7)

$$\tilde{y}^{ll} \leq \tilde{y}_k \leq \tilde{y}^{ul}, \ \forall k = \left\{1, \cdots, N_p\right\} \tag{9}$$

$$e^{ll} \leq e_k \leq e^{ul}, \ \forall k = \left\{1, \cdots, N_p\right\} \tag{10}$$

$$u^{ll} \leq u_k \leq u^{ul}, \ \forall k = \left\{1, \cdots, N_p - 1\right\} \tag{11}$$

$$\Delta u_k = u_k - u_{k-1}, \ \forall k = \left\{1, \cdots, N_p - 1\right\} \tag{12}$$

where N_p is the length of the prediction horizon, $\tilde{y}^{SP}$ is the vector of desired set points of the controlled variables, $Q_{\tilde{y}}$ and $Q_{\Delta u}$ are tuning matrices of the objective function of the optimization problem. $Q_{\tilde{y}}$ and $Q_{\Delta u}$ are tuned such that the set points are reached as fast as possible while avoiding aggressive changes in the manipulated variables. The vector of the controlled variables is bounded by $\tilde{y}^{ll}$ and $\tilde{y}^{ul}$ to ensure that the controlled variables do not exceed their lower and upper limits, such as the quality limit on the mass fraction of the high boiling component w_{hb} in the sump. e^{ll} and e^{ul} are the vectors of the lower and upper bounds of the predicted error of the mechanistic model. The error bounds are based on the extreme values of the errors in the training data. u^{ll} and u^{ul} represent the vectors of the lower and upper limits of the manipulated variables, while Δu_k is the change in the manipulated variables per time step. After solving the optimization problem, the first computed control input is applied to the plant and the optimization is then repeated using the new measurements.

When the operation of the plant is outside the domain of validity of the data-based element, its contribution is faded out, as described above, and the controller relies only on the simplified mechanistic model. This however may worsen the control performance in favor of a cautious operation. Therefore, the prediction quality of the data-based model is checked in the newly visited regions and the domain of validity is adapted [21]. As a measure of the accuracy of the data-based model, the mean squared error $\mathrm{MSE_{ML}}$ between the measured mismatch of the mechanistic model and the multi-step ahead predictions of the mismatch using the data-based component is computed over the past N_w consecutive steps. $\mathrm{MSE_{ML}}$ is defined as:

$$\mathrm{MSE_{ML}} = \frac{1}{N_{\mathrm{W}}} \sum_{i=k-N_{\mathrm{W}}+1}^{k} \left\| e_i^{meas} - \hat{e}_i \right\|_2 \tag{13}$$

where $\| . \|_2$ is the Euclidean norm, e_i^{meas} is the measured mismatch of the model, $\hat{e}_i$ is the predicted mismatch using the data-based model in a multi-step fashion with the initial error $\hat{e}_{k-N_{\mathrm{W}}}$ set equal to the initial measured error $e_{k-N_{\mathrm{W}}}^{meas}$. The baseline mean squared error $\mathrm{MSE_{bias}}$ is computed in a similar way to Eq. (13) where the mismatch prediction $\hat{e}_i$ is constant along the considered horizon of measurements and equal to $e^{meas}_{k-N_W}$. If the contribution factor γ_i was less than a threshold γ_c in the past N_W steps and MSE_{ML} is less than MSE_{bias}, then the domain of validity is adapted by training a new one-class SVM on the low-dimensional projections of the recorded data in the last N_W steps, the previous online collected data, and the offline data as explained in Sect. 4.2. The adapted decision function f_{DF} is then used to compute the weighting function Γ with ζ and the tuning parameters kept unchanged. The adaptation speed is determined by γ_c and N_W. If the domain of validity was adapted, the adapted function $f_{DF}\left(h_k^{s,c}W\right)$ is passed to the controller at the next time step. Fig. 2 summarizes the adaptation scheme of the domain of validity of the data-based model.

6 Application and Results

6.1 Application of the Hybrid Modeling Approach

The hybrid modeling methodology explained in Sect. 4 was applied to the mother liquor distillation column described in Sect. 2. The basic mechanistic model was set up by introducing several simplifications. The behavior of the flowrate controllers of the manipulated variables, which are the reflux flowrate R, the side flowrate S and the boilup flowrate V_b, are not incorporated in the model and it is assumed that the manipulated variables are perfectly controlled. In a similar fashion, the model does not include the level controllers of the distillate receiver and of the column sump. The hold-up of each equilibrium stage is assumed as being constant. The model neglects the energy balances of the equilibrium stages. Accordingly, only the equilibrium temperature at each stage is computed. Hence the basic model considers only the mass and component balance equations for each equilibrium stage, the condenser, and the reboiler. The vapor phase equilibrium is computed via the modified Raoult's law where the liquid activity coefficient of each component is calculated using the non-random two-liquid (NRTL) model. Using the Antoine equation, the saturated vapor pressure of each component at each stage is computed. The summation equation is solved for each stage in order to obtain the equilibrium temperature of each stage. This leads to a semi-explicit index-1 DAE system that consists of 108 differential equations and 27 algebraic equations. By using orthogonal collocation on finite elements, the DAE system is discretized in time.

As a result of the simplifying assumptions, there is a structural mismatch between the plant and the mechanistic model. This plant model mismatch is compensated by a nonlinear autoregressive network with exogenous inputs (NARX) model. The forward element of the NARX model is a neural network formed of two hidden layers, each consisting of 30 neurons. The outputs of the NARX model are the predicted errors of the mass fraction of the low boiling component in the top product and of the mass fraction of the high boiling component in the bottom stream $\left\{e^{w_{lb}}_{k+1}, e^{w_{hb}}_{k+1}\right\}$. As the mechanistic model prediction of the mass fraction of water w_w in the side flow is accurate enough, this value is not compensated by the NARX model. The FNN uses linear activation functions for the output layer and hyperbolic tangent (tanh) activation functions are employed in the hidden layers to avoid having discontinuous derivatives that result from ReLU activation functions. The numbers of layers and neurons were found by trial and error. The NARX model can be represented as:

$$e_{k+1} = f_{ML}\left(e_k, \cdots, e_{k-d_e}, \bar{y}_k, \cdots, \bar{y}_{k-d_y}, u_k, \cdots, u_{k-d_u}, F_k, \cdots, F_{k-d_F}\right) \tag{14}$$

where e_k is the vector of errors $\left\{e^{w_{lb}}_k, e^{w_{hb}}_k\right\}$, $\bar{y}_k$ is a part of the output vector of the simplified model including the mass fractions $\{w_{lb,k}, w_{hb,k}\}$, u_k is the vector of the control inputs $\{R_k, S_k, V_{b,k}\}$ and F_k is the feed rate. The prediction of the model was found to be most accurate when the numbers of delays d_e, d_y, d_u and d_F are set to be equal to 10, after trying out different numbers of delays.

6.2 Modeling Results

In order to generate a sufficiently broad set of dynamic data covering different dynamics and operating points, a reinforcement learning agent performed simulations with the

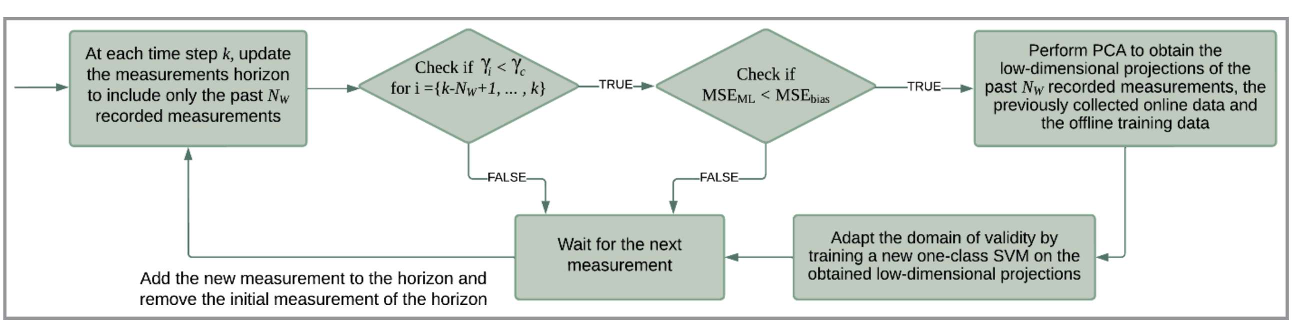

Figure 2. Adaptation scheme of the domain of validity of the data-based model.

detailed Unisim Design model. Simulated measurements were gathered over 14 simulated days of operation using a sampling time of 3 min with a standard deviation of the measurement noise $\sigma_{meas} = [1 \cdot 10^{-3},\ 1 \cdot 10^{-3}]$. The training data was divided into equally sized batches, the training objective function was the sum of the mean squared errors of the single-step ahead predictions and of the multi-step ahead predictions over a training batch. The training algorithm was coded in Python using the Pytorch library [22]. The structure of the hybrid model is depicted in Fig. 3.

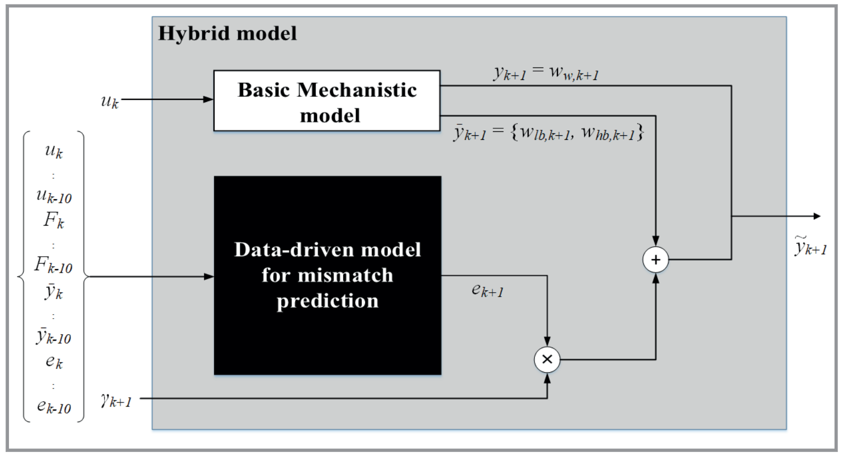

Figure 3. Architecture of the hybrid model developed for the mother liquor distillation column.

Based on an OAT (one-at-a-time) sensitivity analysis, the feed rate and the manipulated variables were determined to be the most influential inputs of the data-based model. The first three principal components are sufficient to explain 95 % of the variance in the training data of the influential input variables as shown in Fig. 4a. A one-class SVM was trained on the resulting three-dimensional projection of the training data to estimate the domain of validity of the data-based model as shown in Fig. 4b. Fig. 4b shows the projection of the training points in red, where positive values of f_{DF} result. The results for a test data set are shown in blue, leading to negative values of f_{DF} as the corresponding feed and boilup flowrates are outside the range of the training data. The relation between the weighting function Γ and the decision function f_{DF} of the one-class SVM is shown in Fig. 4c. The one-class SVM was trained in Python using the Scikit-learn library [23].

Fig. 5 shows the results of comparing the multi-step ahead predictions of the mechanistic model and the hybrid model with the full contribution of the data-based model in the training and validation datasets. The mechanistic model has slower dynamics than the true system especially for w_{lb}, while the data-based model compensates the mismatch in w_{lb} and w_{hb} leading to significantly more accurate multi-step ahead predictions of the hybrid model.

6.3 Estimation and Control Results

The overall proposed approach including the interactions between the components is summarized in Fig. 6. Fig. 7 demonstrates that the estimation of the product compositions is of high quality as the selected column temperatures are highly correlated to the product compositions. The ANN-based soft sensor was trained in Python using the

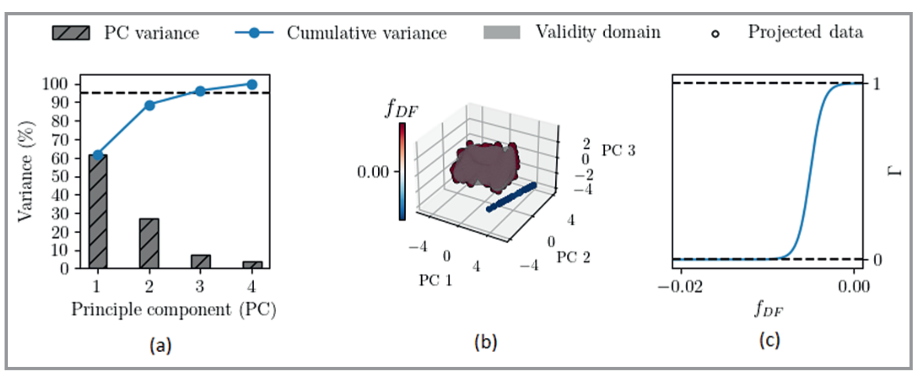

Figure 4. (a) Explained variance by each PC and the cumulative variance. (b) The domain of validity of the data-based model. (c) The weighting function of the data-based model.

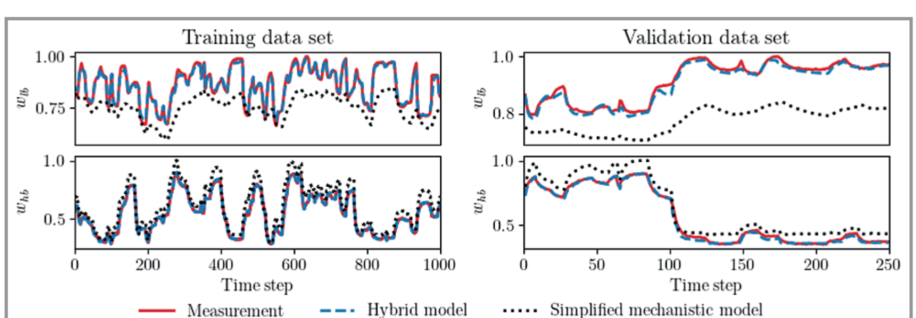

Figure 5. Results of the hybrid model with full contribution of the data-based model.

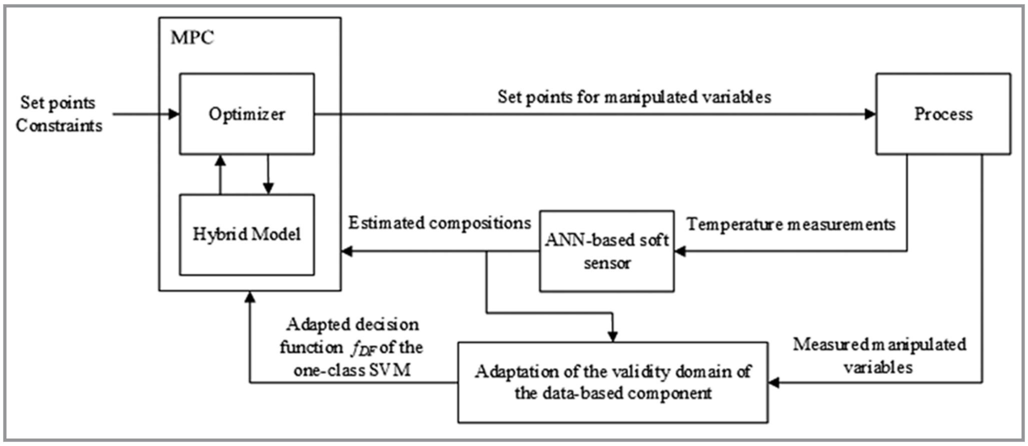

Figure 6. Schematic of the proposed estimation and control approach.

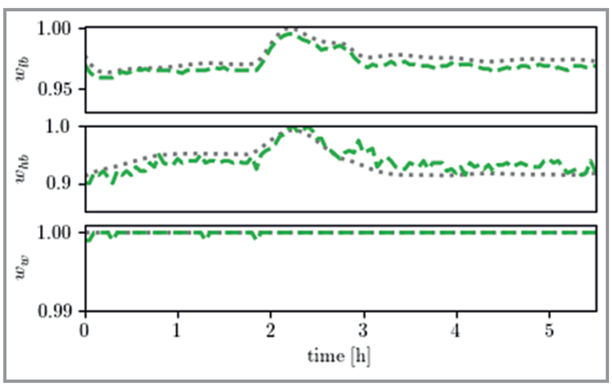

Figure 7. Estimation results of the product compositions for the case of an unmeasured disturbance of the feed composition. Actual mass fraction (· · ·), estimated mass fraction (– –).

Pytorch library [22], the training data was the same as that used for training the data-based component of the hybrid model.

Next, we compare the performance of standard MPC based on the simplified mechanistic model, with MPC based on the full hybrid model, MPC with weighting of the data-based element, and the MPC algorithm with adaptation of the domain of validity explained in Sect. 5. The control task is to achieve a specified purity level of the mass fraction w_{lb} of the low-boiling component in the top product, while the mass fraction w_w of water in the side stream should be as close to 1 as possible, and the mass fraction w_{hb} of the high boiling component should be at a certain level and below a specified threshold. The MPC optimization problems were solved using CasADi [24] and IPOPT [25].

All controllers use a prediction horizon N_p of 4 and assume the same initial conditions of the plant. The tuning is done as following:

$$Q_{\tilde{y}} = \begin{bmatrix} 100 & 0 & 0 \\ 0 & 70 & 0 \\ 0 & 0 & 5 \end{bmatrix} \tag{15}$$

and

$$Q_{\Delta u} = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{16}$$

We use the radial basis function kernel $K(x,x') = \exp\left(\alpha\|x - x'\|^2\right)$ for the one-class SVM with $\alpha = 0.3$ and the upper bound on the fraction of outliers is tuned as 10^{-4} such that f_{DF} is greater than zero for the projected training data. The tuning parameters of the weighting function Γ are $\varepsilon = 10^{-3}$ and $\mu = -0.01$, so the value of Γ is equal to 10^{-3} when $f_{DF} = \mu$, and Γ is equal to 0.999 when $f_{DF} = 0$. The smoothing factor ζ was chosen as 0.2. The ANN-based soft sensor introduced in Sect. 3 is used to estimate the product compositions. The temperature measurements are assumed to be disturbed by noise with a standard deviation equal to 0.1 °C.

Fig. 8 shows the control results of the MPC scheme with weighting of the data-based component and of the MPC scheme using the simplified model when using the estimated compositions provided by the ANN-based soft sensor for the case of an unmeasured feed composition disturbance. $\delta_{w_{lb}}$ is the difference between the set point of w_{lb} and the true value in percent, same for $\delta_{w_{hb}}$. The MPC scheme with weighting of the data-based component results in considerably better tracking of the desired set points than the MPC scheme with the simplified model for most of the simulation time. The MPC scheme with weighting of the data-based component and the MPC based on the full hybrid model behave similarly in the whole simulation period because the data-based model is always operated inside its domain of validity despite of the feed composition disturbance at $t = 2$ h.

Fig. 9 shows the results of three different control approaches for a measured feed rate disturbance scenario, where $\hat{F}$ is

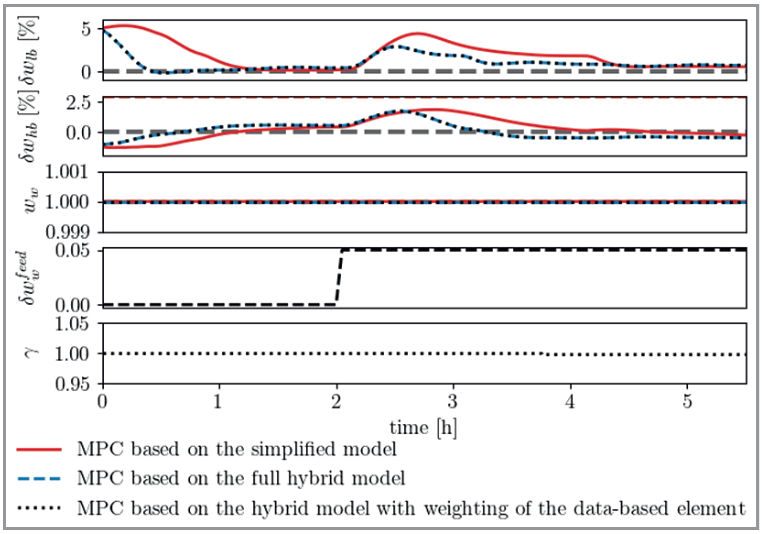

Figure 8. Results of MPC based on the simplified model, MPC based on the full hybrid model and MPC with weighting of the contribution of the data-based component using the ANN-based soft sensor in case of an unmeasured feed composition disturbance.

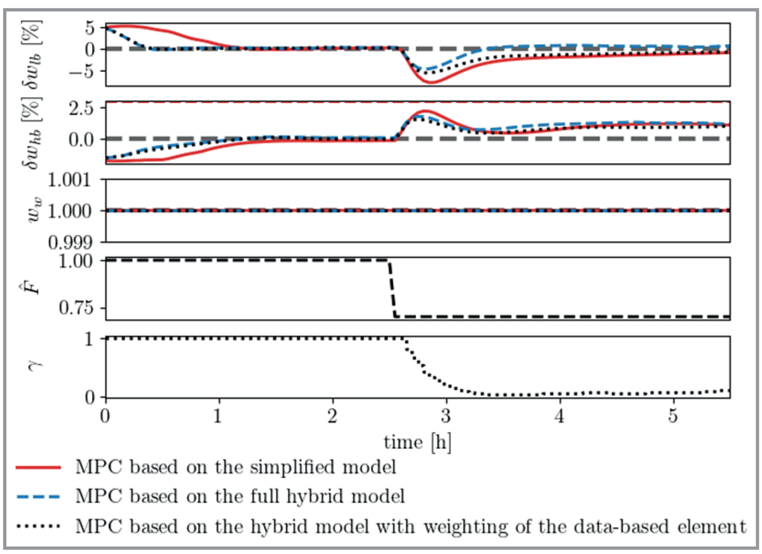

Figure 9. Results of MPC with the simplified model, MPC with the full hybrid model and MPC with weighting of the data-based component using the ANN-based soft sensor in case of measured feed rate disturbance.

the normalized feed flowrate. The MPC scheme with the simplified model results in a good tracking performance for the desired set points of w_{lb} and w_w during the first 2.6 h, satisfying the constraint on the upper limit of w_{hb} but it takes 1.2 h to converge to the set point of w_{lb}. The slow response of the nominal MPC is the result of the inaccuracy of the nominal model, especially the inaccurately estimated values of the holdup on the trays of the column. The MPC scheme without weighting succeeds in converging in less than 0.5 h to the desired set point of w_{lb} as a result of the compensation of the plant model mismatch of the mechanistic model. As the system operates in the domain of validity during the first 2.6 h, the controller with weighting of the data-based element behaves exactly as the controller that uses the full hybrid model all the time. After 2.6 h, the system is moved outside the domain of validity as the feed rate decreases by around 40 % of its nominal value. MPC based on the simplified model needs around 3 h to track the desired set point of w_{lb}, while the controller with the full hybrid model reaches the desired set point much faster. After the disturbance of the feed rate, the use of the ANN-based soft sensor results in an offset in the estimation of w_{hb}, this offset affected the tracking of the desired set point of w_{hb} for all the controllers. The behavior of the controller with weighting of the data-based element is the same as that of the controller with the simplified model because the weight of the data-based model is nearly zero. This behavior motivates to adapt the domain of validity as this would result in faster tracking by relying on the predictions of the full hybrid model.

Fig. 10 shows the results of MPC with adaptation of the domain of validity and of the MPC scheme with fixed weighting of the data-based component for the same case of a feed rate disturbance. After the feed rate dropped at t = 2.6 h, the domain of validity of the data-based model component is adapted after checking its prediction quality in the recently explored region. The adaptation parameters γ_c and N_w were selected as 0.99 and 4 which is the MPC prediction horizon. If γ has been below 0.99 for the last N_w steps, a new one-class SVM is trained on the low-dimensional projections of the online collected data and the offline training data. The MPC scheme with adaptation of the domain of validity outperforms the controller with fixed weighting of the data-based component in tracking the set point of w_{lb} as shown in Fig. 10. Both controllers satisfy the constraint on the upper limit of w_{hb} and maintain the desired purity of the water composition w_w in the side stream well.

7 Conclusion

In this work, we developed nonlinear model predictive control schemes for a mother liquor distillation column based on hybrid model and an ANN-based soft sensor for the missing product composition quality measurements based on a selected set of measured temperatures of the considered column. The nonlinear MPC schemes employ a hybrid model that is formed of a simplified mechanistic model of the column and a parallel data-based model for capturing the mismatch between the plant and the simplified model.

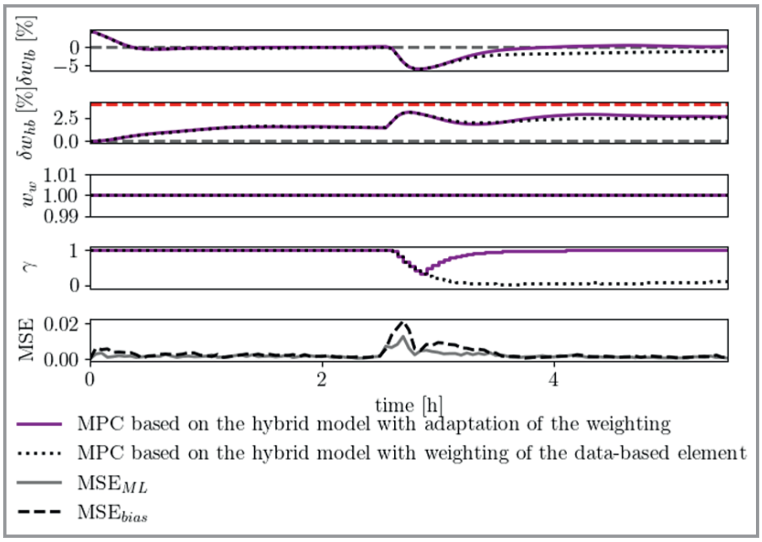

Figure 10. Results of MPC based on the hybrid model with adaptation of the weighting of the data-based element and MPC with weighting of the data-based component using the ANN-based soft sensor in case of measured feed rate disturbance.

As the use of data-based models can be problematic in regions that are poor in training data, the domain of validity of the data-based model is monitored using a one-class SVM that was trained on low-dimensional projections of the training data obtained by selection of variables and PCA. In the hybrid model, the contribution of the data-based model component is weighted based on the closeness to the estimated domain of validity. As this may lead to using an inaccurate model outside the region of validity even though the data-based model does provide good predictions, the prediction quality of the data-based component is assessed online in newly visited regions and the domain of validity is adapted if the prediction quality was found to be high. Simulation studies on a detailed column model demonstrate the good performance of the controller. Future work will focus on retraining the data-based model online in case of predictions of low quality in newly explored regions.

Acknowledgment

This research has been supported by the project "KI-Inkubator-Labore in der Prozessindustrie – KEEN", funded by the Bundesministerium für Wirtschaft und Klimaschutz (BMWK) under grant number 01MK20014T. This support is gratefully acknowledged. Open access funding enabled and organized by Projekt DEAL.

Symbols used

e	[–]	Error vector
F	[kg h^{-1}]	Feed flowrate
$\hat{F}$	[–]	Normalized feed flowrate
f	[–]	Nonlinear function
h	[–]	Input vector of the data-based model
i	[–]	Number of inputs of the data-based model
k	[–]	Current time step
l	[–]	Number of principal components
N	[–]	Number of elements in a matrix or length of a horizon based on the subscript
Q	[–]	Tuning matrix
R	[kg h^{-1}]	Reflux flowrate
S	[kg h^{-1}]	Side flowrate
T	[–]	Projection of the training data
u	[–]	Input vector
V	[kg h^{-1}]	Vapor flowrate
W	[–]	Loading matrix
w	[kg h^{-1}]	Mass fraction
X	[–]	Centered and scaled training data
x	[–]	System state vector
y	[–]	Output vector of the mechanistic model
$\bar{y}$	[–]	Part of the output vector of the mechanistic model
$\tilde{y}$	[–]	Output vector of the hybrid model
z	[–]	Algebraic state vector

Greek letters

α	[–]	Tuning parameter of the radial basis function kernel
Γ	[–]	Weighting function
γ	[–]	Weighting factor of the data-based model
δ	[–]	Percentage difference between the set point and the true value
ε	[–]	Tuning parameter of the weighting function
ζ	[–]	Smoothing factor

μ	[–]	Tuning parameter of the weighting function
σ	[–]	Standard deviation

Sub- and Superscripts

alg	Algebraic
bias	Bias correction
c	Check
DF	Decision function
diff	Difference
feed	Feed mixture
hb	High boiling
k	Current step
lb	Low boiling
ll	Lower limit
ML	Machine learning
meas	Measurement
out	Output
s, c	Scaled and centered
SP	Set point
TP	Training points
ul	Upper limit
W	Measurement horizon
w	Water
w_{lb}	Mass fraction of low boiling component in the distillate
w_{hb}	Mass fraction of high boiling component in the bottom
$\tilde{y}$	Output vector of the hybrid model
Δu	Change of input

Abbreviations

ANN	Artificial neural network
DAE	Differential-algebraic equations
FNN	Feedforward neural network
MPC	Model predictive control
MSE	Mean squared error
NARX	Nonlinear autoregressive network with exogenous inputs
NRTL	Non-random two-liquid model
OAT	One-at-a-time
PCA	Principal component analysis
ReLU	Rectified linear unit
SFS	Sequential forward selection

References

[1] Y. Shin, R. Smith, S. Hwang, Development of model predictive control system using an artificial neural network: A case study with a distillation column, *J. Cleaner Prod.* **2020**, *277*, 124124. DOI: https://doi.org/10.1016/J.JCLEPRO.2020.124124

[2] D. Haßkerl, C. Lindscheid, S. Subramanian, S. Markert, A. Górak, S. Engell, Application of Economics Optimizing Control to a Two-step Transesterification Reaction in a Pilot-Scale Reactive Distillation Column, *IFAC-PapersOnLine* **2018**, *51 (18)*, 67–72. DOI: https://doi.org/10.1016/J.IFACOL.2018.09.252

[3] O. Kahrs, W. Marquardt, The validity domain of hybrid models and its application in process optimization, *Chem. Eng. Process.* **2007**, *46 (11)*, 1054–1066. DOI: https://doi.org/10.1016/j.cep.2007.02.031

[4] G. Brand Rihm, M. Schueler, C. Nentwich, E. Esche, J.-U. Repke, Adaptation of Dynamic Data-Driven Models for Real-Time Applications: From Simulated to Real Batch Distillation Trajectories by Transfer Learning, *Chem. Ing. Tech.* **2023**, *95*, *(7)*, in press. DOI: https://doi.org/10.1002/cite.202200228

[5] M. von Stosch, R. Oliveira, J. Peres, S. Feyo de Azevedo, Hybrid semi-parametric modeling in process systems engineering: Past, present and future, *Comput. Chem. Eng.* **2014**, *60*, 86–101. DOI: https://doi.org/10.1016/j.compchemeng.2013.08.008

[6] B. Bordas, K. Kurt, A. Bamberg, S. Engell, Gray-Box Modeling of the Molecular Weight Distribution in a Batch Polymerization Reactor, *Chem. Ing. Tech.* **2023**, *95*, *(7)*, in press. DOI: https://doi.org/10.1002/cite.202200234

[7] J. Winz, S. Assawajaruwan, S. Engell, Development of a Dynamic Gray-Box Model of a Fermentation Process for Spore Production, *Chem. Ing. Tech.* **2023**, *95*, *(7)*, in press. DOI: https://doi.org/10.1002/cite.202200237

[8] J. Winz, S. Engell, Reliable nonlinear dynamic gray-box modeling by regularized training data estimation and sensitivity analysis, *IFAC-PapersOnLine* **2022**, *55 (7)*, 86–93. DOI: https://doi.org/10.1016/J.IFACOL.2022.07.426

[9] P. Azadi, J. Winz, E. Leo, R. Klock, S. Engell, A hybrid dynamic model for the prediction of molten iron and slag quality indices of a large-scale blast furnace, *Comput. Chem. Eng.* **2022**, *156*, 107573. DOI: https://doi.org/10.1016/j.compchemeng.2021.107573

[10] K. McBride, E. I. Sanchez Medina, K. Sundmacher, Hybrid Semi-parametric Modeling in Separation Processes: A Review, *Chem. Ing. Tech.* **2020**, *92 (7)*, 842–855. DOI: https://doi.org/10.1002/cite.202000025

[11] P. Schäfer, A. Caspari, K. Kleinhans, A. Mhamdi, A. Mitsos, Reduced dynamic modeling approach for rectification columns based on compartmentalization and artificial neural networks, *AIChE J.* **2019**, *65 (5)*, e16568. DOI: https://doi.org/10.1002/aic.16568

[12] V. Krespach, N. Blum, M. Pottmann, S. Rehfeldt, H. Klein, Hybrid Modeling Approaches for Air Separation Unit Control Applications, in *12th International Conference on Distillation & Absorption (D&A)*, Toulouse, September **2022**.

[13] A. M. Schweidtmann, J. M. Weber, C. Wende, L. Netze, A. Mitsos, Obey validity limits of data-driven models through topological data analysis and one-class classification, *Optim. Eng.* **2022**, *23 (2)*, 855–876. DOI: https://doi.org/10.1007/s11081-021-09608-0

[14] P. Nkulikiyinka, Y. Yan, F. Güleç, V. Manovic, P. T. Clough, Prediction of sorption enhanced steam methane reforming products from machine learning based soft-sensor models, *Energy AI* **2020**, *2*, 100037. DOI: https://doi.org/10.1016/J.EGYAI.2020.100037

[15] I. Pisa, I. Santín, J. L. Vicario, A. Morell, R. Vilanova, ANN-based soft sensor to predict effluent violations in wastewater treatment plants, *Sensors* **2019**, *19 (6)*, 1280. DOI: https://doi.org/10.3390/s19061280

[16] M. Elsheikh, Y. Ortmanns, F. Hecht, V. Roßmann, S. Krämer, S. Engell, An Approach to Dependable Hybrid Modeling with

Application to an Industrial Distillation Column, *33rd European Symposium on Computer Aided Process Engineering*, Athens, June **2023**, accepted.
[17] *Unisim design simulation basis*, Technical Report, Honeywell, London **2016**.
[18] D. W. Aha, R. L. Bankert, A Comparative Evaluation of Sequential Feature Selection Algorithms, in *Learning from Data: Artificial Intelligence and Statistics V* (Eds: D. Fisher, H.-J. Lenz), Springer New York, **1996**, 199–206. DOI: https://doi.org/10.1007/978-1-4612-2404-4_19
[19] I. T. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, *Phil. Trans. R. Soc. A* **2016**, *374 (2065)*, 20150202. DOI: https://doi.org/10.1098/rsta.2015.0202
[20] B. Schölkopf, R. C. Williamson, A. Smola, J. Shawe-Taylor, J. Platt, Support Vector Method for Novelty Detection, in *Proc. Advances in Neural Information Processing Systems 12*, **1999**.
[21] M. Elsheikh, Y. Ortmanns, F. Hecht, V. Roßmann, S. Krämer, S. Engell, Model Predictive Control of an Industrial Distillation Column Based on a Hybrid Model: Adapting the Domain of Validity, 22nd IFAC- World Congress, Yokohama, July **2023**, accepted.
[22] A. Paszke, S. Gross, M. Francisco, S. Chintala, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: An Imperative Style, High-Performance Deep Learning Library, in *Proc. of the 33rd Int. Conference on Neural Information Processing Systems*, Curran Associates, Red Hook, NY **2019**, 8024–8035.
[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
[24] J. A. E. Andersson, J. Åkesson, M. Diehl, CasADi – A symbolic package for automatic differentiation and optimal control, *in Recent Advances in Algorithmic Differentiation*, in *Recent Advances in Algorithmic Differentiation* (Eds: S. Forth et al.), Springer, Berlin **2012**, 297–307. DOI: https://doi.org/10.1007/978-3-642-30023-3_27
[25] A. Wächter, L. T. Biegler, On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, *Math. Program.* **2006**, *106*, 25–57.

