Design of Multi-model Linear Inferential Sensors with SVM-based Switching Logic

Martin Mojto, Miroslav Fikar and Radoslav Paulen

Abstract—We study the problem of data-based design of multi-model linear inferential (soft) sensors. The multi-model linear inferential sensors promise increased prediction accuracy yet simplicity of the model structure and training. The standard approach to the multi-model inferential sensor design consists in three separate steps: 1) data labeling (establishing training subsets for individual models), 2) data classification (creating a switching logic for the models), and 3) training of individual models. There are two main issues with this concept: a) as steps 2) & 3) are separate, discontinuities can occur when switching between the models; b) as steps 1) & 3) are separate, data labeling disregards the quality of the resulting model. Our contribution aims at both the mentioned problems, where, for the problem a), we introduce a novel SVM-based model training coupled with switching logic identification and, for the problem b), we propose a direct optimization of data labeling. We illustrate the proposed methodology and its benefits on an example from the chemical engineering domain.

I. INTRODUCTION

The use potential of inferential (soft) sensors is rising in many industrial and engineering fields. Frequent and accurate estimation of key variables plays a significant role in monitoring and control in various real-world use cases [1], [2], [3]. One way of dealing with the estimation of such variables is represented by inferential (soft) sensors [4]. The principle of inferential sensing is to infer the desired hard-to-measure variables according to other easy-to-measure variables (e.g., temperatures, pressures) in the process, e.g., see [5], [6]. The designed inferential sensor thus usually yields less expensive yet more frequent estimation of the key process variable than physical sensors.

The main design trade-off of the inferential sensors is that the higher accuracy is provided at the cost of the higher complexity of the model structure and its model training. The industrial processes are usually nonlinear, which often prohibits a use of simple (linear) inferential sensors due to their inaccuracy (i.e., poor extrapolation performance). This aspect can be easily compensated by the design of more complex inferential sensors, e.g., nonlinear inferential sensor [7] or dynamic inferential sensor [8]. However, the design of more complex sensors usually involves much greater effort (model selection, data treatment, model validation, etc.) and there are even situations in practice where only linear inferential sensor can be implemented along with the present control solution. In such situations, it is possible to approximate the nonlinear behavior of process by designing a so-called multi-model inferential sensor (MIS) [9]. The MISs found their use mostly in the same fields as standard inferential sensors, e.g., in the petrochemical industry [9], in manufacturing [10], and in the process industry [11].

The standard approach to the multi-model inferential sensor design consists in three separate steps: data labeling, data classification, and training of individual models. Data labeling is used to establish training subsets for individual models and is usually done using some clustering method, e.g., k-means clustering [12]. Data classification aims at creating a switching logic for the models. This step can be achieved using multi-class (linear) separation. A suitable method here is support vector machines (SVM) [13], which splits the whole space into the desired number of model validity regions (classes). Linear hyperplanes are used for the switching logic due to their simplicity, interpretability, and ease of implementation. The final training of the individual models can be done within the particular regions using standard maximum-likelihood methods or using some dimensionality-reduction methods [14]. In fact, the design of MIS is closely similar to the piece-wise affine identification [15].

As the classification and training are performed separately, discontinuities can occur when switching between the models. Also the performance of the trained MIS depends strongly on the quality of the switching logic between the models, which stems from good data labeling. Simply speaking, data labeling disregards the quality of the resulting model. This contribution aims to overcome these drawbacks of the standard approach. We develop a new approach that effectively combines the training of the individual models with SVM. This approach ensures continuity of the designed models. Subsequently, we extend the MIS design method with direct optimal data labeling, thus avoiding the need for pre-classification of the data. We analyze the effectiveness of the aforementioned methods by comparing sensor prediction performance and the computational time spent in the design phase. We illustrate the proposed methodology and its benefits on an example from chemical engineering domain.

II. NOTATION

The meaning and notation of the symbols used in the contribution is following: 1 denotes a vector of ones, \( b_p \) denotes an off-set (constant part) of the inferential sensor, \( b_w \) denotes an off-set (constant part) of the separation hyperplane, \( c_w \) denotes a vector of slack variables of SVM, \( \gamma \) denotes a weighting parameter of SVM, \( i \) denotes a measurement...
index, \( I \) denotes an index set for the number of measured data points available for training with cardinality \( n \), \( I_q \) denotes an index set for the number of the classes (models) with cardinality \( n_q \), \( I_{sp} \) denotes an index set for the number of the separation hyperplanes with cardinality \( n_{sp} \), \( p \) denotes a vector of the sensor parameters, \( w \) denotes a normal vector to the separation hyperplane, \( x \) denotes a vector of easy-to-measure variables, \( y \) denotes true value of the desired (output, hard-to-measure) variable, \( \hat{y} \) denotes the estimated value of the desired variable, and \( Z \) denotes a labeling matrix.

### III. Problem Statement

Our goal is to identify a multi-model inferential sensor (MIS) in the following piece-wise linear form [16]:

\[
\hat{y} = \begin{cases} 
  p_1^T x + b_{1,1}, & \text{if } x \in \mathcal{R}_1, \\
  p_2^T x + b_{2,2}, & \text{if } x \in \mathcal{R}_2, \\
  \vdots \\
  p_{n_{sp}}^T x + b_{n_{sp},n_{sp}}, & \text{if } x \in \mathcal{R}_{n_{sp}},
\end{cases}
\tag{1}
\]

where \( \hat{y} \) stands for the inferred value of the desired variable, \( x \in \mathbb{R}^p \) is a vector of the input data, \( p_r \in \mathbb{R}^p \) represents a vector of the sensor parameters in region \( \mathcal{R}_r \), \( b_{r,s} \) is a constant sensor off-set, and \( n_{sp} \) is the number of classes/models. We will assume \( n_{cl} \) to be fixed throughout this paper. We note that the number of models is a designer’s choice and that this methodology can be applied for the cases with any \( n_{cl} \). Regions of individual model validity denoted as \( \mathcal{R}_r \) represent convex polyhedra such that \( \mathcal{R}_r \cap \mathcal{R}_s = \emptyset \), \( \forall r, s \in I_{cl} := \{1, 2, \ldots, n_{cl}\}, r \neq s \).

The design of an inferential sensor is possible due to the existence of \( m \) measurements of the output, \( y \in \mathbb{R}^n \). Next, we describe the standard approach of the MIS design (MIS-std). It consists of three principal steps: data labeling, data classification, individual sensor training.

#### A. Data labeling of the training dataset

There are several methods [17] capable of providing accurate data labeling. Here, we use \( k \)-means clustering [12] within MIS-std. This method separates the \( n \) data points to the desired number of clusters while the average distance is minimized of the member point to the center of the cluster.

#### B. Data classification for switching-logic design

To design a switching logic between the models of the inferential sensor, i.e., to determine \( \mathcal{R}_r, \forall r \in I_q \), data classification is performed. We use here the support vector machines (SVM) approach [13] with linear separators.

According to the number of the desired classes, it is possible to distinguish a binary (see in Fig. 1a) or multi-class (see in Fig. 1b) classification problem. The multi-class SVM-based classification relies on the use of one-vs-rest or one-vs-one approach [18], which transforms the multi-class classification into several binary classifications. The one-vs-one method yields more precise classification, yet involves a higher computational burden. It is the method used in this paper, where \( n_{sp} = n_{cl}(n_{cl} - 1)/2 \) separation hyperplanes (binary classifiers) are found.

\[ \text{A binary classification [19] can be performed as:} \]
\[
\min_{w,b} \|w\|_2 + \gamma (1^T e) \tag{2} \\
\text{s.t. } z_{i,1} (w^T x_i + b_w + e_i - 1) \geq 0, \quad \forall i \in I, \tag{3} \\
- z_{i,2} (w^T x_i + b_w - e_i + 1) \geq 0, \quad \forall i \in I. \tag{4}
\]

where \( w \) is the normal vector to the separation hyperplane, \( b_w \) is a constant off-set of the separation hyperplane, \( \gamma \) is a weighting parameter, and \( e \) is a vector of slack variables. The binary parameters \( z_{i,j} \) are elements of a labeling matrix \( Z \in \mathbb{R}^{n \times n_{cl}} \) (clearly, \( n_{cl} = 1 \) for binary classification) that results from the data labeling procedure as:

\[
z_{i,j} = \begin{cases} 
  1, & \text{if } x_i \in \mathcal{R}_j, \\
  0, & \text{if } x_i \notin \mathcal{R}_j.
\end{cases} \tag{5}
\]

We define the index set \( I_{sp} := \{1, 2, \ldots, n_{sp}\} \). To ensure classification uniqueness of each data point, matrix \( Z \) satisfies

\[
\sum_{j=1}^{n_{sp}} z_{i,j} = 1, \quad \forall i \in I. \tag{6}
\]

The result of binary classification establishes the switching logic with the regions defined as \( \mathcal{R}_1 := \{x \in \mathbb{R}^p | w^T x + b_w \geq 0 \} \) and \( \mathcal{R}_2 := \{x \in \mathbb{R}^p | w^T x + b_w < 0 \} \).

The one-vs-one multi-class classification can be solved by:

\[
\min_{\{w_k^T b_k, k \leq n_{sp}\}} \sum_{k=1}^{n_{sp}} \|w_k\|_2 + \gamma (1^T e_k) \tag{7a} \\
\text{s.t. } \forall i \in I, \forall k \in I_{sp}, \text{for } (r,s) = \mathcal{C}(n_{cl}, k) :
\]
\[
z_{i,r} (w_k^T x_i + b_k + e_{k,i} - 1) \geq 0, \tag{7b} \\
- z_{i,s} (w_k^T x_i + b_k - e_{k,i} + 1) \geq 0. \tag{7c}
\]

where \( \mathcal{C}(n_{cl}, k) \) represents \( k \)-th (non-repetitive) combination of two positive integers \( (r,s) \) such that \( r < s \) and \( r, s \leq n_{cl} \). As in
the case of binary classification, the separation hyperplanes parameterized by \( \{ w_k, b_{w,k} \}_k \) define the switching logic.

C. Individual sensor training

Once switching logic is designed, one can calculate the parameters of the linear models within the structure of MIS. While there are many suitable regression methods [14], we use the standard least-squares regression here for simplicity. While there are many suitable regression methods [14], we use the standard least-squares regression here for simplicity.

We adopt the sum of absolute errors criterion in the objective function (9a) in order to reduce the complexity of the optimization problem as this can be transformed to mixed-integer linear program (MILP). If the sum of squared errors is used in the objective function, the optimization problem turns into mixed-integer nonlinear program (MINLP) that might be challenging especially when number of available training points \( n \) is high.

Note that the problem (9) serves primarily to decide about data labels, i.e., distribution of the training data and, subsequently, of model validity regions. After fixing the values of \( Z \), the final training can be performed via solving (8) so that the final models are trained using the same (least-squares) criterion. This approach no longer requires the a priori labeling of the training dataset and can provide optimal MIS. The prize to pay is the increased computational burden.

The problem (9) can be transformed to quadratically constrained quadratic program (MIQCQP) using the epigraph form [20] of the absolute value in (8a). Arising bilinear constraints can be further transformed to linear by using big-M method [21]. As the variables in \( Z \) are already binary, the big-M method does not require any new integer variables. The resulting form of MIS-con-lab is thus an MILP.

IV. PROPOSED DESIGN METHODS FOR MIS

A. Design of MIS with Continuous Switching

We present a novel approach to the MIS design (MIS-con) that effectively combines the SVM-based data classification with the individual sensor training and that ensures continuity of MIS at the switching boundaries of the individual models. The underlying optimization problem reads as:

\[
\begin{align}
\min \ & \sum_{i=1}^{n} \sum_{j=1}^{n} z_{i,j} (y_i - p^T_j x_i - b_{p,j})^2 \\
s.t. \ & \forall i \in I, \forall k \in I_{sp}, \text{ for } (r,s) = \mathcal{C}(n_{cl}, k) : \\
& z_{i,r} (w_k^T x_i + b_{w,k} + e_{k,i} - 1) \geq 0,
\end{align}
\]

where \( R \) is the heat of vaporization, \( \ref \) is the universal gas constant, \( P_{ref} \) is the reference pressure, \( P \) is an absolute pressure and \( T \) is an absolute temperature.

We assume no knowledge about the structure of the \( PCT \) model. The only information available is given by the 90 measurements of \( PCT \) at different temperatures and pressures. The ground-truth values of parameters of the \( PCT \) model are \( R = 8.314 \text{ J/mol/K}, H_r = 55.940 \text{ 550J/mol}, P_{ref} = 145.325 \text{ Pa} \). The data used in the simulations is generated within the following intervals:

\[
2,000 \text{ Pa} \leq P \leq 20,000 \text{ Pa},
\]

\[
523.15 \text{ K} \leq T \leq 573.15 \text{ K}.
\]

We further assume that the training samples of \( PCT \) are corrupted with some noise. This noise is generated as a random variable from the standard normal distribution.

In order to remove the discrepancies in the variables magnitudes (\( P, T, PCT \)), we performed a normalization. Therefore, the normalized variables \( (P_{norm}, T_{norm}, PCT_{norm}) \) used in the further experiments are within the interval \([0, 1]\).

We will study two situations regarding the distribution of the a priori available training data: one with three linearly separable clusters of data points and the other with no apparent clusters (data points randomly scattered in the training area). In both cases, we seek a three-model sensor.
We compare the performance of single-model inferential sensor (SIS), MIS-std, MIS-con, and MIS-con-lab. Except for MIS-con-lab that is identified using BARON [23], all the designs are found using Gurobi [24]. All the solvers are limited to at most one hour of run time. Alternatively to solving MIS-con-lab using BARON (denoted as MIS-con-lab (bar)), we use BARON with its heuristic termination option and denote the approach as MIS-con-lab (bar,ht). When the MINLP problem (9) is transformed into MILP, the resulting problem is solved with Gurobi (denoted as MIS-con-lab (gur)).

A. Clustered Dataset

In this simulated experiment, we design the MIS using clustered data points. This situation mimics the sensor training when industrial data is well treated and distinct operating points are evident. As we can see in Fig. 2, the data points are concentrated into three clusters. Each cluster contains 30 data points. Subsequently, we randomly assign 50% of the data to the training set (black points) and the remaining data to the testing set (blue points).

The approximation of the original PCT model by MIS-con-lab (gur) is shown in Fig. 3. The training dataset was optimally labeled into the different classes and individual models were trained (class 1 - circles, red model surface; class 2 - stars, yellow model surface; class 3 - diamonds, blue model surface). Models with non-optimized sampling fall for establishing the model validity regions based on the (a priori) visible clusters. The MIS-con-lab (gur) seems to be quite accurate on the training dataset.

We further compare the accuracy of the designed MISs on the testing dataset. This comparison is shown in Fig. 4. As we can see, the accuracy of the standard approach MIS-std (solid green line) is lower compared to the MIS-con (dashed red line) approach or the MIS-con-lab (gur) approach (solid black line). The largest discrepancy between the approaches is visible within the interval with highest values of $PCT_{\text{norm}}$ (measurements 1–16) located in the part of most pronounced nonlinearity of the PCT model (see in Fig. 2). The reason for this discrepancy is that the proposed approaches try to effectively fit the nonlinear section of $PCT$ by two models (see in Fig. 3) while the MIS-std approach uses only one model in this area.

The resulting values of the accuracy measured by the root-mean squared error (RMSE) and computational time ($t_{\text{comp}}$) for each studied approach are shown in Tab. I. The values of the accuracy of designed MISs confirm the best performance of the MIS-con-lab (gur) approach that we could indicate in Fig. 4 as well. The MIS-con-lab (bar) approach is constrained by the time limit (3,600 seconds), and therefore this approach achieved worse accuracy on the testing dataset than other MIS approaches. From the computational time perspective (see in Tab. I), the proposed MIS-con approach is comparable with the MIS-std approach. It seems that the heuristic termination of the BARON can reduce this time burden while it provides the MIS with the same accuracy as the MIS-con-lab (bar) approach.

B. Uniformly Distributed Dataset

In this simulated experiment, the data points are not concentrated into clusters but uniformly distributed within the operation interval (see Fig. 5). Unlike the previous
Fig. 5: The PCT model with uniformly distributed dataset.

Fig. 6: The approximated PCT model on the uniformly distributed testing dataset by using MIS-con-lab (gur) experiment, this situation resembles datasets seen more often in industrial practice. We randomly assign 50% of the data to the training set (black points) and the remaining data to the testing set (blue points). Due to the random character of this scenario, we perform 100 simulation runs with different uniformly distributed datasets. Subsequently, we statistically evaluate the achieved results.

Before analyzing the results from 100 simulation runs, we look at the results from one representative run. The approximation of the ground-truth model by MIS-con-lab (gur) approach is shown in Fig. 6. We illustrate this approach because of its better performance (higher accuracy) on the training dataset compared to other approaches. The results in Fig. 6 show designed models, where same color and symbol code is used as in the case of clustered dataset. Due to the minimal discrepancy between the designed MIS and the ground-truth model in Fig. 6, we can conclude that the proposed MIS-con-lab (gur) approach can accurately approximate the PCT model within the concerned training dataset.

According to the results from 100 simulation runs, we see a comparison of the chosen designed MISs, we can see that the MIS-std approach is less accurate on the testing dataset than the MIS-con approach and the MIS-con-lab (gur) approach. The accuracy of the MIS-con and MIS-con-lab (gur) approaches seems to be similar. However, there are testing measurements (3, 8, 11, 28, 31, and 39) in Fig. 7 which confirm the highest accuracy of the MIS-con-lab (gur) approach. Due to the location of these measurements, we conclude again that the MIS-con-lab (gur) approach is capable of better approximating the nonlinear section of the PCT model compared to other approaches.

We further statistically analyze the results evaluated on 100 runs with different uniformly distributed datasets. The results are shown in Fig. 8 where the blue box represents 25th and 75th (indicating variance), the red line inside the blue box is a median, and red crosses outside of the blue box are outliers. We can conclude (see in Fig. 8 top and middle plots) that the single-model inferential sensor (SIS) is far less accurate than the designed multi-model inferential sensors (MIS). According to the accuracy of designed MISs, we can directly conclude that the MIS-con-lab (gur) approach is the most accurate one on both training and testing datasets. The comparison of the variances (see blue boxes in Fig. 8 top and middle plots) confirms the best performance of the MIS-con-lab (gur) approach. The large variance and the amount of the present outliers within the results from the testing dataset confirm that the MIS-con-lab approach solved with BARON exhibits problems in identifying global optima (there is a tendency of over-fitting), which significantly affects the prediction accuracy evaluated on the testing dataset.

According to the results from the computational time (see in Fig. 8 bottom plot), we can conclude that the MIS-std and MIS-con approaches require lower computational time than the rest of the methods. Furthermore, we can see that the computational time of MIS-con-lab (bar,ht) is significantly reduced compared to MIS-con-lab (bar). Nevertheless, the accuracy of designed sensors is comparable. We could make similar conclusions as in the previous scenario. Moreover, the results from the bottom plot in Fig. 8 indicate improvement of the computational time of the MIS-con-lab (gur) compared to the previous scenario (see in Tab. I). It seems that this approach is more suited for well distributed measurements over the operational space than the data concentrated in the clusters. This is caused by the fact that such distributed data points provide only a small space for the variation of the results (less local over-fitting) of the MIS-lab-con (gur) approach.
and computation time of each method (bottom) from 100 sensors on the training (top) and testing datasets (middle).

Fig. 8: The accuracy (RMSE) of the designed inferential (LASSO or subset selection) within the MIS design.

Involves the introduction of more advanced training methods might be worthy of consideration. Our future work will conclude the introduction of more advanced training methods for optimal margin classifiers, in Proceedings of the Fifth Annual Workshop on Computational Learning Theory, ser. COLT ’92. New York, NY, USA: Association for Computing Machinery, 1992, pp. 144–152.

VI. CONCLUSIONS

In this paper, we presented novel approaches for designing multi-model linear inferential sensors that ensure continuity at switching between models and that allow for optimized data labeling of the training dataset. While the optimal data labeling involves computationally intensive problem (MINLP), we provided a transformation that yields an MILP form. The performance of the studied approaches was evaluated according to two scenarios. The first scenario involved the clustered data, while the second scenario involved uniformly distributed data. The results from both scenarios indicate the best accuracy of the MIS-con-lab (gur) approach on the training and testing datasets. The price for the achieved accuracy is a higher computational burden that resulted in significantly longer computational time than the MIS-std approach, yet the achieved 50% accuracy improvement might be worthy of consideration. Our future work will involve the introduction of more advanced training methods (LASSO or subset selection) within the MIS design.

ACKNOWLEDGEMENT

We acknowledge discussions with M. Klaúčo and K. Kiš from STUBA and with K. L’ubušký from Slovnaft, a.s.

REFERENCES

[1] M. Abouzari, P. Pahlavani, F. Izaditame, and B. Bigdeli, “Estimating the chemical oxygen demand of petrochemical wastewater treatment plants using linear and nonlinear statistical models – a case study,” Chemosphere, vol. 270, p. 129465, 2021.

[2] M. Li, T. Pan, and Q. Chen, “Estimation of tea quality grade using statistical identification of key variables,” Food Control, vol. 119, p. 107485, 2021.

[3] J. Qi, L. Ding, and S. Lim, “Toward cool cities and communities: A sensitivity analysis method to identify the key planning and design variables for urban heat mitigation techniques,” Sustainable Cities and Society, vol. 75, p. 103377, 2021.

[4] B. Joseph and C. B. Brossilow, “Inferential control of processes: Part i. steady state analysis and design,” AIChE Journal, vol. 24, no. 3, pp. 485–492, 1978.

[5] S. J. Qin, H. Yue, and R. Dunita, “Self-validating inferential sensors with application to air emission monitoring,” Industrial & Engineering Chemistry Research, vol. 36, no. 5, pp. 1675–1685, 1997.

[6] Q. Zhu, S. Joe Qin, and Y. Dong, “Dynamic latent variable regression for inferential sensor modeling and monitoring,” Computers & Chemical Engineering, vol. 137, p. 106809, 2020.

[7] S. Park and C. Han, “A nonlinear soft sensor based on multivariate smoothing procedure for quality estimation in distillation columns,” Computers & Chemical Engineering, vol. 24, no. 2, pp. 871–877, 2000.

[8] K. Wang, C. Shang, L. Liu, Y. Jiang, D. Huang, and F. Yang, “Dynamic soft sensor development based on convolutional neural networks,” Industrial & Engineering Chemistry Research, vol. 58, no. 26, pp. 11521–11531, 2019.

[9] S. Khatibisephehr, B. Huang, F. Xu, and A. Espejo, “A bayesian approach to design of adaptive multi-model inferential sensors with application in oil sand industry,” Journal of Process Control, vol. 22, no. 10, pp. 1913–1929, 2012.

[10] T. Zhongda, L. Shujiang, W. Yanhong, and W. Xiangdong, “A multi-model fusion soft sensor modelling method and its application in rotary kiln calcination zone temperature prediction,” Transactions of the Institute of Measurement and Control, vol. 38, no. 1, pp. 110–124, 2016.

[11] S. Hou, X. Zhang, W. Dai, X. Han, and F. Hua, “Multi-model- and soft-transition-based height soft sensor for an air cushion furnace,” Sensors, vol. 20, no. 3, 2020.

[12] E. Forgy, “Cluster analysis of multivariate data: efficiency versus interpretability of classifications,” Biometrics, vol. 21, pp. 768–769, 1965.

[13] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “A training algorithm for optimal margin classifiers,” in Proceedings of the Fifth Annual Workshop on Computational Learning Theory, ser. COLT ’92. New York, NY, USA: Association for Computing Machinery, 1992, pp. 144–152.

[14] M. Mojto, K. L’ubušký, M. Fikar, and R. Paulen, “Data-based design of inferential sensors for petrochemical industry,” Computers & Chemical Engineering, vol. 153, p. 107437, 2021.

[15] I. Bako and O. Yahya, “Piecewise affine system identification: A least harmonic mean approach,” in CDC, Nice, France, 2019.

[16] Z. gang Su, P. hong Wang, J. Shen, X. jun Yu, Z. zhong Lv, and L. Lu, “Multi-model strategy based evidential soft sensor model for predicting evaluation of variables with uncertainty,” Applied Soft Computing, vol. 11, no. 2, pp. 2595–2610, 2011, the Impact of Soft Computing for the Progress of Artificial Intelligence.

[17] O. Maimon and L. Rokach, Data mining and knowledge discovery handbook. Springer, 2010, vol. 14.

[18] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics). Berlin, Heidelberg: Springer-Verlag, 2006.

[19] V. Kecman, Support Vector Machines – An Introduction. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 1–47.

[20] M. Milano, Principles and Practice of Constraint Programming - CP 2012: 18th International Conference, CP 2012, Québec City, QC, Canada, October 8-12, 2012, Proceedings, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2012.

[21] J. Griva, S. G. Nash, and A. Sofer, Linear and Nonlinear Optimization (2 ed.). SIAM, 2009.

[22] M. King, Process Control: A Practical Approach. John Wiley & Sons Ltd., 2011.

[23] N. V. Sahinidis, BARON 21.1.13: Global Optimization of Mixed-Integer Nonlinear Programs, User’s Manual, 2017.

[24] L. Gurobi Optimization, “Gurobi optimizer reference manual,” 2021. [Online]. Available: https://www.gurobi.com.