Preprints of the 22nd IFAC World Congress
Yokohama, Japan, July 9-14, 2023

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

Martin Mojto * **Karol Ľubušký** ** **Miroslav Fikar** *
Radoslav Paulen *

* *Faculty of Chemical and Food Technology, Slovak University of Technology in Bratislava, Bratislava, Slovakia*
** *Slovnaft, a.s., Bratislava, Slovakia*

Abstract: The data-based design of multi-model linear inferential (soft) sensors (MIS) is studied. These promise increased prediction accuracy yet simplicity of the model structure and training. The state-of-the-art approach to the MIS design consists of three 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 as steps 2) & 3) are separate: (i) discontinuities can occur when switching between the models; (ii) data labeling disregards the quality of the resulting model. Our contribution aims at both the mentioned problems, where, for the problem (i), we introduce a novel support vector method (SVM)-based model training coupled with switching logic identification and, for the problem (ii), we propose a direct optimization of data labeling. The proposed methodology and its benefits are illustrated on an example from the chemical engineering domain.

Keywords: Machine learning and data analytics in process control, Monitoring and performance assessment, Monitoring of product quality and control performance

1. INTRODUCTION

The use potential of advanced inference solutions is rising in many industrial fields. Frequent and accurate estimation of key variables already plays a major role in monitoring and control in various real-world use cases (Li et al., 2021; Qi et al., 2021). One way of dealing with the key-variables estimation is represented by inferential (soft) sensors (ISs) (Joseph and Brosilow, 1978). The principle is to infer the hard-to-measure variables from easy-to-measure variables (e.g., temperatures, pressures), e.g., see (Qin et al., 1997; Zhu et al., 2020). The designed inferential sensor thus usually yields less expensive yet more frequent estimation of the key variable than a physical sensor.

The main IS-design trade-off is that the higher accuracy is provided at the cost of the higher complexity of the model structure and model training. 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 ISs, e.g., nonlinear IS (Park and Han, 2000) or dynamic IS (Wang et al., 2019). However, the design of complex ISs usually involves much greater effort (model selection, data treatment, model validation, etc.) and there are even situations when only a linear IS can be implemented along the actual plant automation solution. In such situations, it is possible to approximate the nonlinear behavior of process by designing a so-called multi-model inferential sensor (MIS) (Khatibisepehr et al., 2012). The MISs found their use, e.g., in the petrochemical industry (Khatibisepehr et al., 2012), in manufacturing (Zhongda et al., 2016), and in the process industry (Hou et al., 2020).

The state-of-the-art approach for the MIS design usually involves three consecutive steps (on the training dataset): a priori labeling, classification, and training of individual models. There are several approaches to a priori labeling (Lü and Yang, 2014), but the most frequent method in use is k-means clustering (Forgy, 1965). Shuang and Gu (2016) claim the support vector machines (Boser et al., 1992) to be suitable to perform classification and to split the whole space into the desired number of model validity regions (classes). The training of individual MIS models can be executed by an appropriate regression method (Mojto et al., 2021). The MIS design is also related to the recent study (Bemporad, 2022) on piece-wise affine (PWA) identification that considers Voronoi partitioning and aims to provide a set of linear models accurately fitting the nonlinear model. The state-of-the-art MIS can be further improved by addressing two drawbacks: 1) continuity of models at MIS switching points, and 2) data labeling unaware of its impact on MIS accuracy.

This contribution aims to overcome the aforementioned drawbacks of the MIS design. We develop a new approach that effectively combines the training of the individual models with SVM. This approach ensures the continuity

⋆ This research is funded by the Slovak Research and Development Agency (projects APVV-21-0019 and APVV SK-FR-2019-0004), by the Scientific Grant Agency of the Slovak Republic (grants VEGA 1/0691/21 and VEGA 1/0297/22), and by the European Commission under the grant no. 101079342 (Fostering Opportunities Towards Slovak Excellence in Advanced Control for Smart Industries). The authors acknowledge discussions with M. Klaučo and K. Kiš from Slovak University of Technology in Bratislava.
Corresponding author: R. Paulen (e-mail: radoslav.paulen@stuba.sk)

of the designed models. We also extend the MIS design with optimal data labeling, thus avoiding the need for pre-labeling of the data. A preliminary version with two-model MIS was investigated in Mojto et al. (2022). We analyze the effectiveness of the studied methods by comparing IS prediction performance and training CPU time on an example from the chemical engineering domain.

Notation: $\mathbf{1}$ denotes a vector of ones, b_p is an off-set (bias) of the inferential sensor, b_w stands for an off-set (bias) of the separation hyperplane, e_w represents a vector of SVM slack variables, γ signifies a weighting parameter of SVM, i denotes a measurement index, I is an index set of training data points with cardinality n, I_{cl} stands for an index set of the classes (models) with cardinality n_{cl}, I_{sp} represents an index set of the separation hyperplanes with cardinality n_{sp}, w signifies a normal vector to the separation hyperplane, and Z denotes a labeling matrix.

2. PROBLEM STATEMENT

The aim is to design a multi-model inferential sensor (MIS) in the piece-wise linear form (Su et al., 2011):

$$\hat{y} = \begin{cases} p_1^\intercal x + b_{p,1}, & \text{if } x \in \mathcal{R}_1, \\ p_2^\intercal x + b_{p,2}, & \text{if } x \in \mathcal{R}_2, \\ \dots \\ p_{n_{\text{cl}}}^\intercal x + b_{p,n_{\text{cl}}}, & \text{if } x \in \mathcal{R}_{n_{\text{cl}}}, \end{cases} \tag{1}$$

where $\hat{y}$ stands for the inferred value of the desired variable, $x \in \mathbb{R}^{n_p}$ is a vector of the input data, $p_r \in \mathbb{R}^{n_p}$ represents a vector of the sensor parameters in region $\mathcal{R}_r$, $b_{p,r}$ is a constant sensor off-set, and n_{cl} is the number of classes/models. The value of n_{cl} is assumed to be fixed throughout this paper. The number of models is a designer's choice. The presented methodology is applicable for the cases with any n_{cl}. Regions of individual model validity denoted as $\mathcal{R}_r$ represent convex polyhedra such that $\mathcal{R}_r \bigcap \mathcal{R}_s = \emptyset, \forall r, s \in I_{\text{cl}} := \{1, 2, \dots, n_{\text{cl}}\}, r \neq s$. Each region considers the same input variables (n_p is constant).

The MIS design is enabled by n measurements of the output variable y being available. The state-of-the-art MIS design (MIS-SotA) consists of three main steps: a priori data labeling, classification, and individual sensor training.

2.1 A priori labeling for data classification

The initial (prior) knowledge about the classes of measurements within a dataset is typically required for designing a classifier. This prior labeling can be obtained directly from the available dataset. Among methods providing accurate and reliable data labeling, k-means clustering (Forgy, 1965) is a simple and intuitive method, which minimizes the average distance between points within the clusters.

2.2 Data classification for switching-logic design

To design a switching logic between the MIS models, i.e., to determine $\mathcal{R}_r, \forall r \in I_{\text{cl}}$, data classification is performed. The commonly used method of linear classifier design is SVM shown in Figure 1. According to the number of the desired classes, it is possible to distinguish a binary (Fig. 1, top plot) or multi-class (Fig. 1, bottom plot) classification. The multi-class SVM-based classification relies on the use of one-vs-rest or one-vs-one approach (Bishop, 2006), which transforms the multi-class classification into several binary classifications.

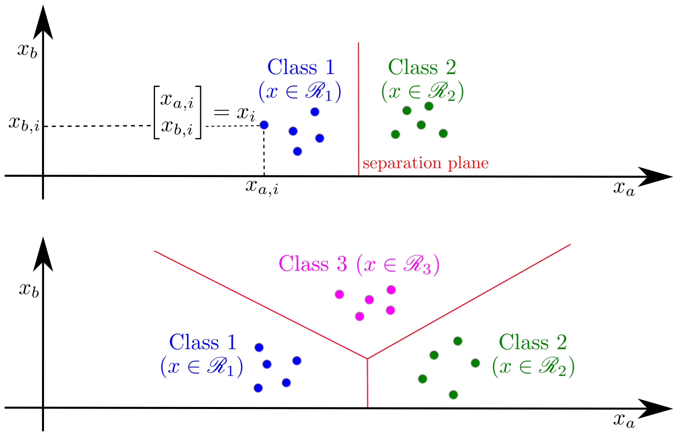

Fig. 1. Binary (top) & multi-class (bottom) classification.

We use one-vs-one method as it yields a more precise classification. The method results in founding $n_{\text{sp}} = n_{\text{cl}}(n_{\text{cl}} - 1)/2$ separation hyperplanes (classifiers). The binary classification (Kecman, 2005) is resolved as follows:

$$\min_{w, b_w, e \geq 0} \; \|w\|_2^2 + \gamma \|e\|_1 \tag{2}$$

$$\text{s.t. } z_{i,1} \left(w^\intercal x_i + b_w + e_i - 1\right) \geq 0, \quad \forall i \in I, \tag{3}$$

$$- z_{i,2} \left(w^\intercal x_i + b_w - e_i + 1\right) \geq 0, \quad \forall i \in I, \tag{4}$$

where w and b_w are the normal vector and the off-set of the separation hyperplane, respectively, γ is a weighting parameter, and e is a vector of slack variables. The binary parameters $z_{i,j}$ constitute a labeling matrix $Z \in \{0,1\}^{n \times n_{\text{cl}}}$ that follows from the a priori labeling 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 $I_{\text{sp}} := \{1, 2, \dots, n_{\text{sp}}\}$. To ensure the uniqueness of each data point, matrix Z satisfies:

$$\textstyle\sum_{j=1}^{n_{\text{cl}}} z_{i,j} = 1, \quad \forall i \in I. \tag{6}$$

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

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

$$\min_{\{w_k, b_{w,k}, e_k\}_{k \in I_{\text{sp}}}} \; \textstyle\sum_{k=1}^{n_{\text{sp}}} \|w_k\|_2^2 + \gamma \|e_k\|_1 \tag{7a}$$

$$\text{s.t. } \forall i \in I, \forall k \in I_{\text{sp}}, \text{for } (r,s) = \mathfrak{C}(n_{\text{cl}}, k) : e_k \geq 0, \tag{7b}$$

$$z_{i,r} \left(w_k^\intercal x_i + b_{w,k} + e_{k,i} - 1\right) \geq 0, \tag{7c}$$

$$- z_{i,s} \left(w_k^\intercal x_i + b_{w,k} - e_{k,i} + 1\right) \geq 0, \tag{7d}$$

where $\mathfrak{C}(n_{\text{cl}}, k)$ represents k^{th} (unique) combination of two positive integers (r, s) such that $r < s \leq n_{\text{cl}}$. The parameters $\{w_k, b_{w,k}\}_{k \in I_{\text{sp}}}$ define the switching logic.

2.3 Training of the individual models

Once the switching logic is designed, one can train the individual models within the structure of MIS. We use the standard least-squares regression here, for simplicity.

The MIS model structure holds several advantages over fully nonlinear ISs. Firstly, the MIS structure provides high transparency and an intuitive understanding of the model parameters. Secondly, the low complexity of the MIS model structure usually leads to lower computational costs compared to nonlinear ISs. While nonlinear ISs might offer higher flexibility, they are prone to overfitting.

3. PROPOSED DESIGN METHODS FOR MIS

There are two main drawbacks of MIS-SotA. Firstly, the inferential sensor discontinuity can appear at the switching boundaries of the individual models. This can significantly reduce the controlled system efficiency if such a sensor is used in the process monitoring or control. The second issue is that the data labeling is unaware of the prediction accuracy of the resulting sensor and thus construction of optimal model-validity regions is not guaranteed.

3.1 Design of MIS with Continuous Switching

A novel approach (MIS-con) combines the SVM-based classification with the model training. The design optimizes:

$$\min_{\substack{\{w_k,b_{w,k},e_k\}_{k\in I_{\mathrm{sp}}}\\ \{p_j,b_{p,j}\}_{j\in I_{\mathrm{cl}}}}} \sum_{j=1}^{n_{\mathrm{cl}}}\sum_{i=1}^{n} z_{i,j}\left(y_i - p_j^{\intercal}x_i - b_{p,j}\right)^2 + \sum_{k=1}^{n_{\mathrm{sp}}}\gamma\|e_k\|_1 \tag{8a}$$

$$\text{s.t. } \forall i\in I, \forall k\in I_{\mathrm{sp}}, \text{for } (r,s)=\mathfrak{C}(n_{\mathrm{cl}},k): e_k\geq 0, \tag{8b}$$

$$\text{Eqs. (7c)–(7d)}, \tag{8c}$$

$$p_r - p_s = w_k,\ b_{p,r} - b_{p,s} = b_{w,k}. \tag{8d}$$

The constraints in (8d) ensure continuity at the switch of any two models. This matches the intersection of any two model surfaces with the switching hyperplane.

3.2 Design of MIS with Optimized Data Labeling

Performance of the MIS is affected by the initial data labeling of the training dataset. The aim of the labeling is to distinguish measurements within the training dataset. The ability to do so is related to the presence of measurement noise and to the nonlinearity of the output variable. The industrial ISs are usually designed according to data with a significant level of noise and certain variation of operating regimes. We propose the following approach (MIS-con-lab) to reduce the inaccuracies within the a priori labeling.

The approach searches directly for the optimal data labeling by adding Z among the optimized variables as:

$$\min_{\substack{\{w_k,b_{w,k},e_k\}_{k\in I_{\mathrm{sp}}}\\ Z,\{p_j,b_{p,j}\}_{j\in I_{\mathrm{cl}}}}} \sum_{j=1}^{n_{\mathrm{cl}}}\sum_{i=1}^{n} z_{i,j}\left|y_i - p_j^{\intercal}x_i - b_{p,j}\right| + \sum_{k=1}^{n_{\mathrm{sp}}}\gamma\|e_k\|_1 \tag{9a}$$

$$\text{s.t. } \forall i\in I, \forall k\in I_{\mathrm{sp}}, \text{for } (r,s)=\mathfrak{C}(n_{\mathrm{cl}},k): e_k\geq 0, \tag{9b}$$

$$\text{Eqs. (6), (7c)–(7d)}, \tag{9c}$$

$$p_r - p_s = w_k,\ b_{p,r} - b_{p,s} = b_{w,k}. \tag{9d}$$

To reduce the problem complexity, we adopt the sum of absolute errors (SAE) criterion in the cost function of (9a), as this can be transformed into a mixed-integer linear program (MILP). The transformation uses the epigraph form (Milano, 2012) of the absolute value to arrive at a mixed-integer quadratically constrained quadratic program. The arising bilinear terms are then transformed to linear ones via the big-M method (Griva et al., 2008). As the variables in Z are binary, the big-M form does not introduce any new integer variables. To ensure that each point belongs to one class only, (6) is involved in (9c).

If the sum of squared errors (SSE) is used, the optimization problem turns into mixed-integer nonlinear program that might be challenging, especially when the number of available training points n is high. SSE can be affected by the outliers. Therefore, the optimization problem benefits from the SAE criterion when significant noise is present within the training dataset. Note that the problem (9) serves primarily to decide about training data labels and, subsequently, about the model validity regions. After fixing Z, the final training can be done via solving (8) so that the resulting models are optimal w.r.t. the SSE criterion.

The MIS-con-lab approach is designed for applications with uncertain prior labeling. Unlike other approaches, it can provide optimal data labeling, albeit at the cost of an increased computational burden. MIS-con-lab should be used when the dataset involves indistinguishable classes and n is reasonably large (i.e., in the order of hundreds). Otherwise, it is advisable to use MIS-con or to select data based on some efficient sampling technique.

4. CASE STUDY

The proposed approach for the MIS design is developed for industrial needs. The industrial practice confirms that the input variables can be effectively combined in appropriate nonlinear structures to provide the maximum estimation accuracy. One of these nonlinear structures is a pressure compensated temperature PCT. This phenomenological variable is very often used in the petrochemical industry, especially in the case of low-pressure distillation columns. It is stated as (King, 2011):

$$PCT = 1/\left(R/H_v \ln\left(P/P_{\mathrm{ref}}\right) + 1/T\right), \tag{10}$$

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

We design the ISs for PCT with the assumption of no knowledge about the model structure. The only information available is given by 90 measurements of PCT at different temperatures and pressures. The ground-truth values of parameters of the PCT model are $R = 8.314\,\mathrm{J/mol/K}$, $H_v = 55,9401\,\mathrm{J/mol}$, $P_{\mathrm{ref}} = 145,325\,\mathrm{Pa}$. The data is generated within the following intervals:

$$0.2\,\mathrm{Pa} \leq 10^{-4}\,P \leq 2\,\mathrm{Pa}, \quad 523.2\,\mathrm{K} \leq T \leq 573.2\,\mathrm{K}. \tag{11}$$

We assume that the training samples are corrupted with a random noise with the standard normal distribution. 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 lie within the interval $[0, 1]$.

We study two scenarios regarding the distribution of the available data: (i) with three (linearly) separable data clusters and (ii) with no apparent clusters (data scattered over the domain). In both, we seek a three-model MIS.

The set of inferential sensors studied in this work comprises single-model inferential sensors (SIS), MIS-SotA, MIS-con, and MIS-con-lab. We distinguish three different

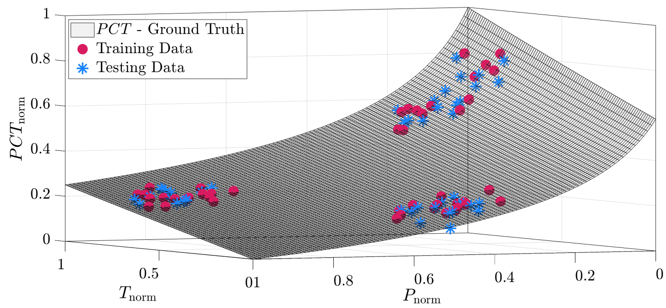

Fig. 2. The PCT model with clustered data.

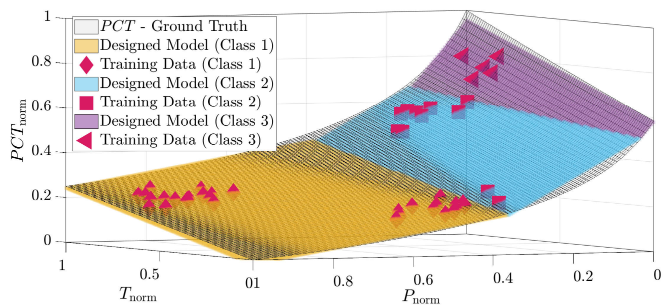

Fig. 3. The approximated PCT model on the clustered training dataset by using MIS-con-lab (gur).

approaches for solving MIS-con-lab: (1) MIS-con-lab (bar) which uses BARON (Sahinidis, 2017) to solve (9) with SSE (MINLP), (2) MIS-con-lab (bar,ht) which is the previous approach with solver's heuristic termination, and (3) MIS-con-lab (gur) which uses Gurobi (Gurobi Optimization, 2021) to solve (9) (MILP). We selected the solvers as the best state-of-the-art solvers available for the problem class.

All solvers are run with their default options (unless mentioned otherwise), with their run times limited to one hour. The time limitation is significant for the MIS-con-lab approach only, whose CPU time increases exponentially with the number of measurements. MIS-con-lab can cope only with approximately 100 measurements in our design settings. This limit can vary according to the number of desired models or the noise variance. While this feature might appear restrictive, it is not rare in industrial applications, where laboratory samples are scarce. To provide an insight into the studied inferential sensors, several tests (clustered and random scenarios) are conducted.

The performance analysis of the studied inferential sensor (SIS and MISs) includes measuring sensor prediction accuracy using the root mean square error (RMSE), separately for the training and testing datasets. Additionally, the complexity of the designed inferential sensors is evaluated based on the training CPU time. As the MIS design is based on classification, it is important to assess the reliability of the classification using appropriate performance metrics such as accuracy, recall, or various misclassification rates (Onel et al., 2019). An inaccurate classification could result in imprecise model regions, leading to suboptimal MISs. Due to low complexity of the studied PCT model and the low noise levels, we did not directly evaluate these metrics but instead relied on visual inspection.

Table 1. The accuracy (RMSE) and CPU time of the studied methods on the clustered data.

Approach	RMSE$\times 10^2$ (training)	RMSE$\times 10^2$ (testing)	CPU time [s]
SIS	4.1	3.5	0.1
MIS-SotA	0.9	1.0	4.3
MIS-con	0.9	1.1	4.1
MIS-con-lab (bar)	0.5	1.6	3,600.0
MIS-con-lab (bar,ht)	0.5	1.6	325.9
MIS-con-lab (gur)	0.4	0.5	2,284.0

4.1 Clustered Dataset

The design of MIS over a clustered dataset mimics the sensor training when industrial data is well treated and distinct operating points are evident. Fig. 2 reveals that the data points are concentrated into three clusters with 30 data points each. We randomly assign 50 % of the data to the training set (magenta points) and the remaining data to the testing set (blue stars).

The performance of MIS-con-lab (gur) can be observed in Figure 3, where the designed classes and models are presented (class 1: diamonds, purple model surface; class 2: squares, blue model surface; class 3: triangles, yellow model surface). Models that were trained with non-optimized labeling faced difficulty in establishing model validity regions based on the visible clusters (a priori). This is reflected in the highest accuracy achieved by MIS-con-lab (gur) on the training dataset, as shown in Table 1.

The resulting values of RMSE and CPU time for each studied approach are presented in Tab.1. The accuracy analysis confirms the necessity of the MIS design, which significantly outperforms the SIS design. The comparison of MISs indicates that MIS-con-lab (gur) provides the best performance. This approach is approximately 50 % more accurate than MIS-SotA. The most significant discrepancy between these approaches arises from the section with the highest values of PCT_{norm}, located in the region of the most pronounced nonlinearity of the PCT model (see in Fig. 2). The reason for this discrepancy is that MIS-con-lab (gur) effectively fits the nonlinear section of PCT with two models (as seen in Fig. 3), whereas the MIS-SotA approach considers only one model. The MIS-con-lab (bar) approach is constrained by the time limit (3,600 seconds), and MIS-con-lab (bar,ht) is heuristically terminated due to slow convergence, resulting in poorer accuracy on the testing dataset than other MIS approaches.

The results presented in Table 1 indicate that the MIS design is less efficient in terms of CPU time compared to the SIS design. Among the MIS approaches, the proposed MIS-con method has comparable CPU time with the MIS-SotA approach, despite the higher complexity of the optimization problem. It appears that heuristic termination of the BARON solver has the potential to reduce the time burden while maintaining the accuracy of the MIS, as evidenced by the comparison between MIS-con-lab (bar) and MIS-con-lab (bar,ht). The results also suggest that the superior accuracy of MIS-con-lab (gur) comes at the cost of increased CPU time.

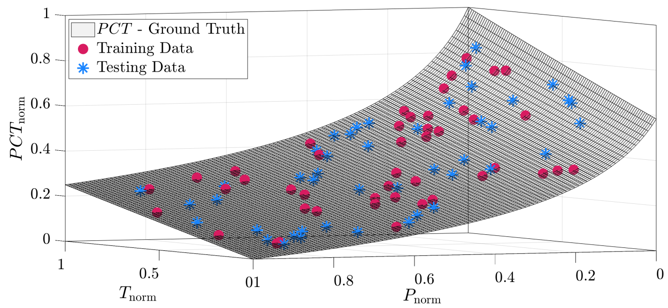

Fig. 4. The PCT model with uniformly distributed data.

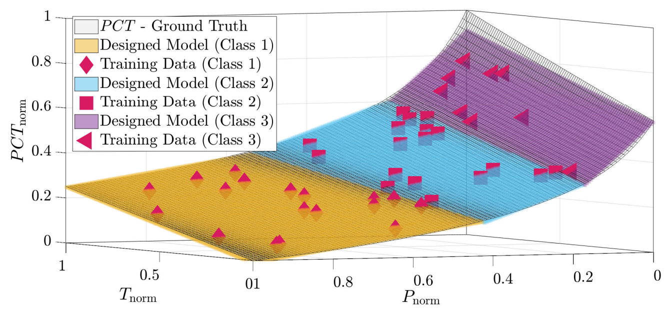

Fig. 5. The PCT model with the uniformly distributed testing dataset approximated by MIS-con-lab (gur).

4.2 Uniformly Distributed Dataset

In this simulated experiment, the data points are not concentrated into clusters but uniformly distributed within the operation interval (see Fig. 4). Unlike the previous case, this situation resembles datasets seen more often in industrial practice. We randomly assign 50 % of the data to the training set (magenta points) and the remaining data to the testing set (blue stars). Due to the stochastic character of this scenario, we perform 100 simulations with different uniformly distributed datasets.

At first, we look at the results from one representative run of the 100 simulations. The approximation of the ground-truth model by MIS-con-lab (gur) approach is shown in Fig. 5. We illustrate this approach because of its superior performance on the training dataset compared to other approaches. The results in Fig. 5 show the designed models, where the same color and symbol code is used as in the case of the clustered dataset. Due to the minimal discrepancy between the designed MIS and the ground-truth model in Fig. 5, we can conclude that the proposed MIS-con-lab (gur) approach can accurately approximate the PCT model within the concerned training dataset.

According to Fig. 6, which shows the accuracy comparison of the chosen designed MISs, we can see that MIS-SotA 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 points (3, 8, 11, 28, 31, and 39) in Fig. 6 which confirm the highest accuracy of the MIS-con-lab (gur) approach. Due to the location of these measurements, we conclude again that MIS-con-lab (gur) is capable of better approximating the nonlinear section of the PCT model compared to other approaches. This is due to optimal distribution of the training points with respect to the learning outcome of the MIS.

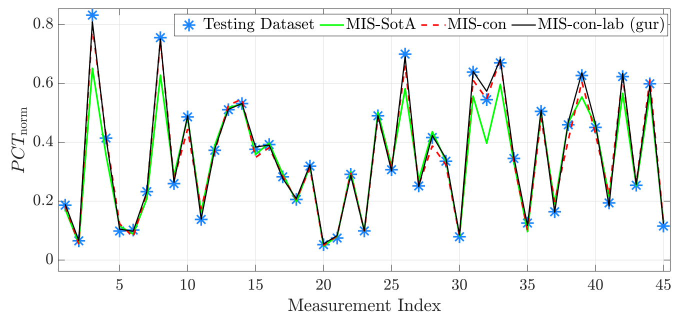

Fig. 6. The prediction performance of the designed ISs on the uniformly distributed testing set.

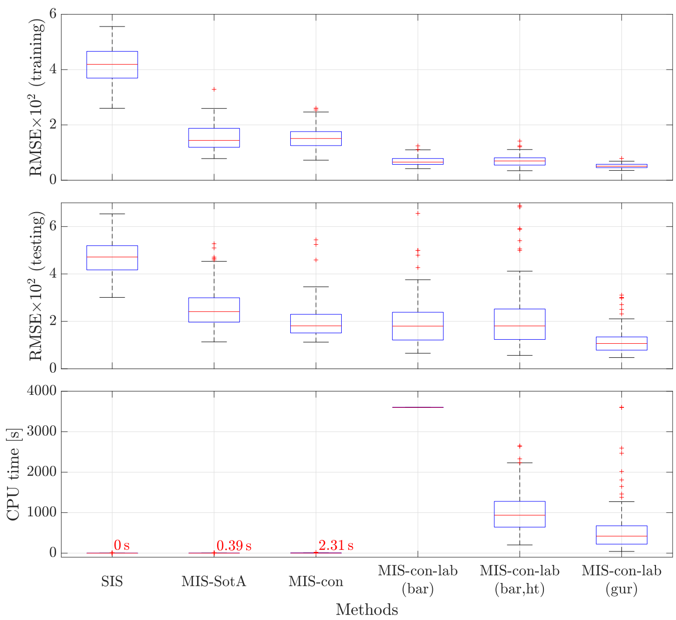

Fig. 7. The statistics of accuracy (RMSE) of the designed ISs on the training (top) and testing (middle) datasets and CPU times of each method (bottom) from 100 simulation runs with uniformly distributed datasets.

We further statistically analyze the results evaluated on 100 runs with different uniformly distributed datasets. The results are shown in Fig. 7, where the blue box represents 25^{th} and 75^{th} percentiles (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. 7, top and middle plots) that SIS is far less accurate than the designed MISs. We can directly conclude that the MIS-con-lab (gur) approach is the most accurate one on both training and testing datasets. This approach outperforms MIS-SotA by about 50 % on the testing dataset. The comparison of the variances (see blue boxes in Fig. 7, 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 show that the MIS-con-lab approach solved with BARON exhibits problems in identifying global optima (there is a tendency of overfitting), which significantly affects the prediction accuracy evaluated on the testing dataset. The improvement of this approach can be achieved by the data treatment analysis or filtering of the dataset available for the MIS design.

Regarding the CPU time (see Fig. 7, bottom plot), we can conclude that the MIS-SotA and MIS-con approaches require lower CPU time than the rest of the methods. Furthermore, we can see that the CPU time of MIS-con-lab (bar,ht) is expectedly reduced compared to MIS-con-lab (bar). Nevertheless, the accuracy of the designed sensors is comparable, as we could see in the previous scenario. Moreover, the results from the bottom plot in Fig. 7 indicate improved CPU time of MIS-con-lab (gur) compared to the previous scenario. This approach seems to be more suited for well distributed measurements over the operational space than the clustered data. The distributed data provide only a tight space for the variation of the MIS-con-lab (gur) results (less local over-fitting).

5. CONCLUSION

We presented novel approaches for multi-model linear inferential sensor (MIS) design to ensure continuity at switching between the individual sensor models and to provide optimized data labeling. The performance of the studied approaches was evaluated in two scenarios: clustered and uniformly distributed datasets. The results indicate the best accuracy of MIS-con-lab (gur). The price to pay is a higher computational burden compared to MIS-SotA, yet the achieved 50 % accuracy improvement is worthy of consideration. Our future work will involve the introduction of advanced training methods (LASSO or subset selection) to the MIS design, effectively allowing for different structures in the individual MIS models.

REFERENCES

Bemporad, A. (2022). A piecewise linear regression and classification algorithm with application to learning and model predictive control of hybrid systems. *IEEE Trans. Autom. Control*, 1–16.

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

Boser, B.E., Guyon, I.M., and Vapnik, V.N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the 5th Annual Workshop on Computational Learning Theory*, COLT '92, 144–152.

Forgy, E. (1965). Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, 21, 768–769.

Griva, I., Nash, S.G., and Sofer, A. (2008). *Linear and Nonlinear Optimization (2. ed.)*. SIAM.

Gurobi Optimization, L. (2021). Gurobi optimizer reference manual. URL `https://www.gurobi.com`.

Hou, S., Zhang, X., Dai, W., Han, X., and Hua, F. (2020). Multi-model- and soft-transition-based height soft sensor for an air cushion furnace. *Sensors*, 20(3).

Joseph, B. and Brosilow, C.B. (1978). Inferential control of processes: Part i. steady state analysis and design. *AlChE J.*, 24(3), 485–492.

Kecman, V. (2005). *Support Vector Machines – An Introduction*, 1–47. Springer.

Khatibisepehr, S., Huang, B., Xu, F., and Espejo, A. (2012). A bayesian approach to design of adaptive multi-model inferential sensors with application in oil sand industry. *J. Process Control*, 22(10), 1913–1929.

King, M. (2011). *Process Control: A Practical Approach*. John Wiley & Sons Ltd.

Li, M., Pan, T., and Chen, Q. (2021). Estimation of tea quality grade using statistical identification of key variables. *Food Control*, 119, 107485.

Lü, Y. and Yang, H. (2014). A multi-model approach for soft sensor development based on feature extraction using weighted kernel fisher criterion. *Chin. J. Chem. Eng.*, 22(2), 146–152.

Milano, M. (2012). *Principles and Practice of Constraint Programming - CP 2012: 18th International Conference, CP 2012, Québec City, QC, Canada, October 8-12, 2012*. Lecture Notes in Computer Science. Springer.

Mojto, M., Ľubušký, K., Fikar, M., and Paulen, R. (2021). Data-based design of inferential sensors for petrochemical industry. *Comput. Chem. Eng.*, 153, 107437.

Mojto, M., Ľubušký, K., Fikar, M., and Paulen, R. (2022). Support vector machine-based design of multi-model inferential sensors. In L. Montastruc and S. Negny (eds.), *Comp. Aided Process Eng.*, 51, 1045–1050.

Onel, M., Kieslich, C.A., and Pistikopoulos, E.N. (2019). A nonlinear support vector machine-based feature selection approach for fault detection and diagnosis: Application to the tennessee eastman process. *AIChE Journal*, 65(3), 992–1005. doi:https://doi.org/10.1002/aic.16497.

Park, S. and Han, C. (2000). A nonlinear soft sensor based on multivariate smoothing procedure for quality estimation in distillation columns. *Comput. Chem. Eng.*, 24(2), 871–877.

Qi, J., Ding, L., and Lim, S. (2021). Toward cool cities and communities: A sensitivity analysis method to identify the key planning and design variables for urban heat mitigation techniques. *Sust. Cities Soc.*, 75, 103377.

Qin, S.J., Yue, H., and Dunia, R. (1997). Self-validating inferential sensors with application to air emission monitoring. *Ind. Eng. Chem. Res.*, 36(5), 1675–1685.

Sahinidis, N.V. (2017). *BARON 21.1.13: Global Optimization of Mixed-Integer Nonlinear Programs.*

Shuang, Y.F. and Gu, X.S. (2016). Multi-model switching method based on sphere-based SVM classifier selector and its application to hydrogen purity multi-model soft sensor modeling in continuous catalytic reforming. In L. Zhang, X. Song, and Y. Wu (eds.), *Theory, Methodology, Tools and Applications for Modeling and Simulation of Complex Systems*, 57–72. Springer.

Su, Z., hong Wang, P., Shen, J., jun Yu, X., zhong Lv, Z., and Lu, L. (2011). Multi-model strategy based evidential soft sensor model for predicting evaluation of variables with uncertainty. *Appl. Soft Comput.*, 11(2), 2595–2610. The Impact of Soft Computing for the Progress of Artificial Intelligence.

Wang, K., Shang, C., Liu, L., Jiang, Y., Huang, D., and Yang, F. (2019). Dynamic soft sensor development based on convolutional neural networks. *Ind. Eng. Chem. Res.*, 58(26), 11521–11531.

Zhongda, T., Shujiang, L., Yanhong, W., and Xiangdong, W. (2016). A multi-model fusion soft sensor modelling method and its application in rotary kiln calcination zone temperature prediction. *Trans. Inst. Meas. Control*, 38(1), 110–124.

Zhu, Q., Joe Qin, S., and Dong, Y. (2020). Dynamic latent variable regression for inferential sensor modeling and monitoring. *Comput. Chem. Eng.*, 137, 106809.