Data-Based Design of Multi-Model Inferential Sensors

Martin Mojto[a,1], Karol Ľubušký[b], Miroslav Fikar[a], Radoslav Paulen[a]

[a]*Faculty of Chemical and Food Technology, Slovak University of Technology in Bratislava, Radlinského 9, 812 37 Bratislava, Slovakia*
[b]*Slovnaft, a.s., Vlčie hrdlo 1, 824 12 Bratislava, Slovakia*

Abstract

This paper deals with the problem of inferential (soft) sensor design. The nonlinear character of industrial processes is usually the main limitation to designing simple linear inferential sensors with sufficient accuracy. In order to increase the inferential sensor predictive performance and yet to maintain its linear structure, multi-model inferential sensors represent a straightforward option. In this contribution, we propose two novel approaches for the design of multi-model inferential sensors aiming to mitigate some drawbacks of the state-of-the-art approaches. For a demonstration of the developed techniques, we design inferential sensors for a Vacuum Gasoil Hydrogenation unit, which is a real-world petrochemical refinery unit. The performance of the multi-model inferential sensor is compared against various single-model inferential sensors and the current (referential) inferential sensor used in the refinery. The results show substantial improvements over the state-of-the-art design techniques for single-/multi-model inferential sensors.

Keywords: Inferential Sensors, Petrochemical Industry, Process Monitoring

---

*Tel.: +421 (0)2 5932 5349, Mail: martin.mojto@stuba.sk (M. Mojto)

Nomenclature

List of Acronyms

GF   Gasoline Fraction

HGO   Hydrogenated Gasoil

LASSO Least Absolute Shrinkage and Selection Operator

MILP   Mixed-Integer Linear Program

MINLP Mixed-Integer NonLinear Program

MIS   Multi-Model Inferential Sensor

NN   Neural Network

OLSR   Ordinary Least Squares Regression

PCR   Principal Component Regression

PLSR   Partial Least Squares Regression

RMSE   Root Mean Squared Error

SAE   Sum of Absolute Errors

SIS   Single-Model Inferential Sensor

SS   Subset Selection

SSE   Sum of Squared Errors

SVM   Support Vector Machine

VGH   Vacuum Gasoil Hydrogenation

List of Symbols and Indices

$\boldsymbol{a}$ Vector of inferential sensor parameters, $a \in \mathbb{R}^{n_\mathrm{p}}$

a_0 Constant inferential sensor off-set (bias)

α Weighting parameter for normal vector of the separation plane

β Weighting parameter for vector of the slack variables

col   Distillation column section

$\mathcal{D}$ Available dataset

$\boldsymbol{e}$ Vector of the slack variables, $\boldsymbol{e} \in \mathbb{R}^n$

F   Feed stream

F Flowrate

h   Heating medium to a reboiler

H_v Heat of vaporization

L Liquid level

λ Weighting parameter of LASSO

$\boldsymbol{m}_i$ Vector of input variables for i^{th} measurement, $\boldsymbol{m}_i \in \mathbb{R}^{n_\mathrm{p}}$

μ Mean value

n Number of measurements

n_p	Number of input variables (candidates)
n_pc	Number of principal components
p	Pressure
PCT	Pressure compensated temperature
R	Universal gas constant
$\mathcal{R}$	Convex polyhedra
reb	Reboiler
rt	Ratio
$\mathcal{S}$	Testing dataset
σ	Standard deviation
t	Top section (location)
T	Termodynamic temperature
$\mathcal{T}$	Training dataset
vo	Valve opening
vap	Vapor phase
$\boldsymbol{w}$	Normal vector of the separation hyperplane, $w \in \mathbb{R}^{n_\mathrm{p}}$
w_0	Constant separation hyperplane off-set (bias)
x	Concentration (composition)

y Output (desired) variable, $y \in \mathbb{R}$

$\hat{y}$ Estimated output (desired) variable, $\hat{y} \in \mathbb{R}$

$\boldsymbol{z}$ Vector of the binary variables, $\boldsymbol{z} \in \mathbb{R}^n$ or $\boldsymbol{z} \in \mathbb{R}^{n_\mathrm{P}}$

1. Introduction

Increasing need for automation triggers a growing demand in the industry for accurate inferential (soft) sensors. The purpose of the inferential sensor is to estimate desired hard-to-measure variables (e.g., product concentration) using measurements from the online process sensors (e.g., temperatures, pressures, flow rates). The use of an inferential sensor represents a cheap alternative that can provide more frequent sensing of the desired variables compared to other ways of process monitoring, e.g., physical sensing devices or lab analysis.

In general, the production processes represent complex systems with many variables and interactions between these variables (Santander et al., 2022) and they usually exhibit nonlinear behavior resulting from the rich interactions of involved physical phenomena. One would conjecture that a nonlinear inferential sensor design is necessary. However, a typical industrial process is usually operated in some target operating range to achieve desired product specifications. Therefore, the nonlinear behavior of the process variable can often be neglected, and the linear inferential sensor can provide an accurate estimate of the desired variable. The advantage of a linear inferential sensor over some nonlinear counterpart lies foremost in lower maintenance expenses, higher transparency, possibility of physical insight, lower computational effort for sensor training, validation, online evaluation, and further calculations (e.g., for optimization and control). The later aspect can be significant mainly when the estimated variable is an input for an advanced process controller (Botha and Craig, 2021) or it is involved in a (complex) plant-wide optimization (Ge, 2017).

6

The structure of the linear inferential sensor usually involves only a single model (hence, single-model inferential sensor (SIS)). According to the SIS design approach, one can distinguish between model-based (Doraiswami and Cheded, 2014), data-driven (King, 2011; Mojto et al., 2021; Sun and Braatz, 2021) and hybrid approaches (Tahir et al., 2019; Zhuang et al., 2022). The complexity and size of industrial processes significantly limits the use of the model-based SIS modeling techniques. Therefore, the data-driven SIS approaches are predominantly used in industrial practice. The most popular advanced design techniques include Principal Components Regression (PCR) (Kendall, 1957), Partial Least Squares regression (PLSR) (Wold et al., 1984, 2001), the Least Absolute Shrinkage and Selection Operator (LASSO) method (Santosa and Symes, 1986; Tibshirani, 2011), and the Subset Selection (SS) approaches (Konno and Yamamoto, 2009; Miyashiro and Takano, 2015; Takano and Miyashiro, 2020). The usage of these techniques to provide an automatic modeling framework is presented in Sun and Braatz (2021). Moreover, a comprehensive analysis of the data-driven approaches can be found in Mojto et al. (2021); Sun et al. (2022). These contributions compare various performance criteria (RMSE, R^2, model complexity) of these approaches considering the real data from complex industrial case studies.

Multi-model linear inferential sensors bridge the gap between linear and nonlinear inferential sensors. To retain the advantages of the linear inferential sensor yet to increase the prediction performance, it is possible to consider multiple linear models within the sensor structure (hence, multi-model inferential sensor (MIS)). Each model within the MIS structure can explain a particular part of the operating region, e.g. a cluster of possible steady

7

states. Therefore, these sensors find applications in complex industrial processes with multitude of operating modes (Khatibisepehr et al., 2012; Jin et al., 2015; Wang et al., 2021).

The state-of-the-art MIS design consists of three sequential steps: (1) a priori labeling, (2) data classification, and (3) individual training of the inferential sensor models. In the first step, a modeler searches for an appropriate number of models and assigns tags to available data to distinguish the models (classes) discovered. The popular approach for a priori labeling is k-means clustering (Forgy, 1965). A comparison of several other techniques for a priori labeling is shown in Lü and Yang (2014). The classification step employs an appropriate data-based (machine-learning) approach to draw model-validity regions, i.e., the boundaries between the classes (models) that would later serve as switching conditions for using predictions from a particular model. A frequently used and well-known classification learner is Support Vector Machine (SVM) (Boser et al., 1992). The method designs classification hyperplanes in the context studied in this work. Lastly, the constituent models of MIS are individually trained for each class by using a suitable regression technique (Mojto et al., 2021). One of the recent examples of learning MIS is given in Bemporad (2022), where piecewise linear regression is considered together with classification based on softmax regression and labeling by the k-means algorithm. A similar approach (Ferrari-Trecate and Muselli, 2002) uses a neural network as a classifier. While the state-of-the-art approaches train MIS effectively, there are still a few drawbacks hindering the overall potential of MIS. The first drawback is that the continuity at the switch between the different MIS models is not guaranteed. This can have a negative

8

impact on the plant production. For example, a common issue might arise that advanced process controllers with MIS implemented might face stability issues because of inferential sensor discontinuity. The second drawback originates from the a priori labeling that is unaware — like any other unsupervised learning approach — of its impact on MIS prediction accuracy. It is, therefore, not likely that the optimal allocation of the model-validity regions is achieved.

In this paper, we propose two novel approaches that prune the MIS design of the negative effects of the drawbacks mentioned above. Firstly, both the proposed approaches ensure continuity when switching between the designed MIS models. This is achieved by merging classification and model training into one decision problem and training an SVM separation hyperplane to act as a switching boundary between the MIS models. Secondly, we propose an optimization-based labeling approach, effectively conducting all the sequential steps of the state-of-the-art MIS design procedure simultaneously. The performance of the proposed approaches is compared on the synthetic dataset from a model of pressure compensated temperature PCT and subsequently on the industrial dataset from the Vacuum Gasoil Hydrogenation (VGH) unit. The VGH unit represents a crucial section within the oil refinery Slovnaft, a.s. in Bratislava, Slovakia. The performance of MIS is compared (a) with the reference SIS currently used in the refinery and (b) with SIS designed by Ordinary Least Squares Regression (OLSR), PCR, PLSR, LASSO, and SS. The conclusions in this study follow from the standard indicators, such as prediction accuracy (RMSE) and complexity (i.e., number of model input variables or principal components).

9

This paper is organized as follows. In Section 2, the basic description of SIS and MIS is introduced. Section 3 shows the development of a novel approach for the MIS design. Section 4 shows the design of SIS and MIS for case studies from the petrochemical industry. In Section 5, the results are discussed. Finally, in Section 6, we summarize the main conclusions and remarks of this work.

2. Problem Description

We aim at designing an inferential sensor for predicting a hard-to-measure process variable $y \in \mathbb{R}$ based on easy-to-measure process variables $\boldsymbol{m} \in \mathbb{R}^{n_{\mathrm{P}}}$, which are also referred to as inputs of the sensor. A mathematical representation of the sensor reads as:

$$\hat{y} = f\left(\boldsymbol{m}\right), \tag{1}$$

where $\hat{y} \in \mathbb{R}$ is the prediction of y, which is also referred to as output of the sensor. We consider just scalar output but inclusion of multiple inputs is a straightforward extension of this work. We focus on static sensors as they are typically desired in industry and because they often represent the appropriate model structure. For example, in the petrochemical industry, a concentration of distillation products fits a nonlinear static thermodynamic function of process variables (e.g., temperatures and pressures).

2.1. Design of a Single-Model Inferential Sensor

We consider a linear SIS that can be represented as (Mojto et al., 2021):

$$\hat{y} = \boldsymbol{m}^{\mathsf{T}}\boldsymbol{a} + a_0, \tag{2}$$

10

where $\boldsymbol{a} := \left(a_1, a_2, \ldots, a_{n_\mathrm{p}}\right)^{\mathsf{T}} \in \mathbb{R}^{n_\mathrm{p}}$ represents the vector of sensor parameters and a_0 is a constant off-set (bias).

The SIS design is given by three subsequent stages (Khatibisepehr et al., 2013; Botha and Craig, 2021; Mojto et al., 2021):

1. *Data pre-processing stage.* This usually includes standardization for the data (series of y and $\boldsymbol{m}$) to involve zero mean and unit standard deviation, removal of systematic errors and outliers, and selection of the n_p input candidates from the whole set of input variables (e.g., based on linear independence and correlation criteria).

2. *Training stage.* The training dataset (with a corresponding index set $\mathcal{T}$) is being considered in this step, which is a subset of the entire available dataset (index set $\mathcal{D}$). The model parameters are calculated based on a chosen fitting criterion. Sensor structure is selected directly through some measures to prevent overfitting (e.g., cross validation on a validation dataset) or using feature selection, or indirectly by including such measure in the calculation of the model parameters.

3. *Testing stage.* This optional step uses (previously unseen) testing dataset (index set $\mathcal{S}$; $\mathcal{S} := \mathcal{D} \setminus \mathcal{T}$). The purpose of this stage is, for example, to decide between several candidate inferential sensors trained in parallel according to the previous stages. The assessment measure is some accuracy criterion (e.g., root mean squared error (RMSE)).

2.1.1. Training Methods for SIS

We list here some methods of popular choice (Khatibisepehr et al., 2013; Mojto et al., 2021) for SIS training phase.

- Ordinary Least Squares Regression (OLSR)

  The parameters are obtained by solving:

  $$\min_{\boldsymbol{a},a_0} \frac{1}{2} \sum_{\forall i \in \mathcal{T}} (y_i - \boldsymbol{m}_i^\intercal \boldsymbol{a} - a_0)^2, \qquad (3)$$

  where index i stands for a measurement index such that $\mathrm{card}(\mathcal{T}) > n_\mathrm{p}$.

- Principal Component Regression (PCR) (Kendall, 1957)

  This statistical approach reduces the dimensionality of a dataset by linear transformation into a new orthogonal space of principal components $(n_\mathrm{p} \to n_\mathrm{pc}, n_\mathrm{p} \geq n_\mathrm{pc})$. The aim is to preserve most of the information content (variation) within the original dataset by using as few principal components as possible. The principal components are derived by singular value decomposition of a covariance matrix, and therefore, PCR is considered an unsupervised learning approach. However, it can be effectively combined with some of the regression approaches (e.g., OLSR) in order to design a linear SIS.

- Partial Least Squares Regression (PLSR) regression (Wold et al., 1984)

  This statistical approach shares similar characteristics with PCR. The optimization problem (3) is solved over a transformed input space of principal components. The designed principal components are derived by singular value decomposition of a cross-covariance matrix, and therefore, PLSR is considered a supervised learning approach, unlike PCR.

- Least Absolute Shrinkage and Selection Operator (LASSO) regression (Santosa and Symes, 1986)

This method finds the sensor structure by solving:

$$\min_{\boldsymbol{a}, a_0} \frac{1}{2} \sum_{\forall i \in \mathcal{T}} (y_i - \boldsymbol{m}_i^\mathsf{T} \boldsymbol{a} - a_0)^2 + \lambda \|\boldsymbol{a}\|_1, \tag{4}$$

where λ is a weight between the accuracy of the model training and the model overfitting. The resulting model structure is trained using (3) while the inputs corresponding to zero coefficients from the solution to (4) being discarded.

- Subset Selection (SS) (Konno and Yamamoto, 2009) The SS approach seeks the best sensor with a given number of inputs. One solves:

$$\min_{\substack{\boldsymbol{a}, a_0 \\ \boldsymbol{z} \in \{0,1\}^{n_\mathrm{p}}}} \frac{1}{2} \sum_{i \in \mathcal{T}} (y_i - \boldsymbol{m}_i^\mathsf{T} \boldsymbol{a} - a_0)^2 \tag{5a}$$

$$\text{s.t.} \quad -\bar{a} z_k \leq \boldsymbol{a} \leq \bar{a} z_k, \ \forall k \in \{1, \ldots, n_\mathrm{p}\}, \tag{5b}$$

$$\sum_{k=1}^{n_\mathrm{p}} z_k = \tilde{n}_\mathrm{p}, \tag{5c}$$

where $\tilde{n}_\mathrm{p}$ is the number of desired sensor inputs, $\bar{a}$ represents an upper bound on $\|\boldsymbol{a}\|_\infty$ and $\boldsymbol{z}$ denotes a vector with binary entries signifying the selection of the kth input into the sensor structure. A cross-validation approach is commonly used to determine best $\tilde{n}_\mathrm{p}$, a final form of the model.

We note that the SS approach can be also enhanced to directly seek a trade-off between the simplicity and accuracy of the sensor. Here, a chosen cross-validation criterion can be used explicitly in the design (Miyashiro and Takano, 2015; Takano and Miyashiro, 2020; Mojto et al., 2021).

Illustrative example.. To motivate our study, we consider a problem of designing an inferential sensor for the pressure compensated temperature PCT. This variable is frequently used in low-pressure petrochemical distillation columns (Pan et al., 2019). A combination of the Antoine and Clausius-Clapeyron equations forms the following mathematical representation (King, 2011):

$$\frac{1}{PCT} = \frac{R}{H_v} \ln\left(\frac{p}{p_{\text{ref}}}\right) + \frac{1}{T}, \tag{6}$$

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

The ground truth model of the PCT is considered with $R = 8.3 \, \text{J/mol/K}$, $H_v = 55,940.6 \, \text{J/mol}$ and $p_{\text{ref}} = 145.3 \, \text{Pa}$ over the operating region:

$$523.2 \, \text{K} \leq T \leq 573.2 \, \text{K},$$
$$0.4 \, \text{Pa} \leq p \leq 15 \, \text{Pa}, \tag{7}$$
$$635.3 \, \text{K} \leq PCT \leq 1151.4 \, \text{K}.$$

For sensor training and evaluation, all input variables (p, T, PCT) are scaled (normalized) to the interval $[0, 1]$.

The SIS performance is shown in Figure 1 on two training datasets, which simulates the process working in one (one data cluster in Figure 1(a)) and two distinct operating regimes (two data clusters in Figure 1(b)), respectively. Zero-mean Gaussian white noise with a standard deviation $\sigma_{\text{noise}} = 5 \, \text{K}$ is added to the output data. The SIS accuracy, measured by the root mean squared error (RMSE) and tested on fresh data within the training regions, is significantly reduced (more than 2-fold deterioration) when the process
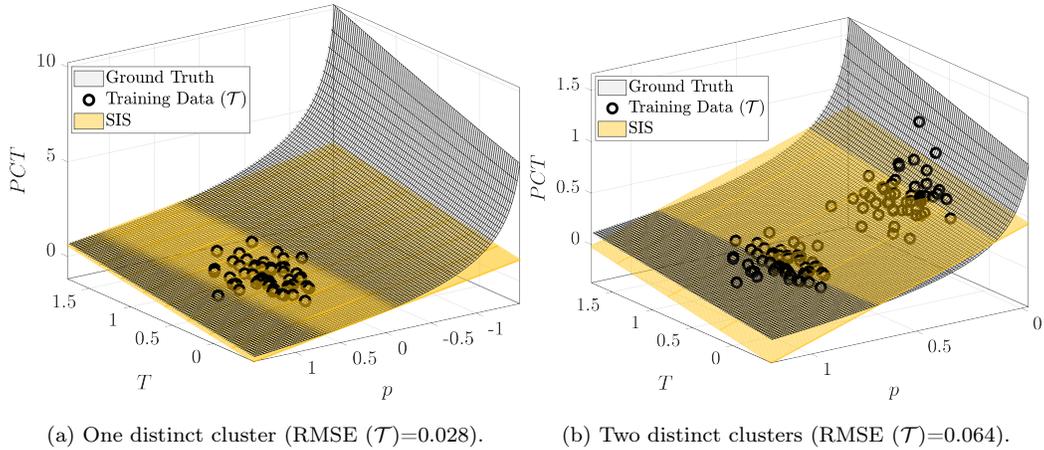
14

(a) One distinct cluster (RMSE ($\mathcal{T}$)=0.028). (b) Two distinct clusters (RMSE ($\mathcal{T}$)=0.064).

Figure 1: The ground truth model of *PCT* with SIS designed on different datasets.

runs in two operating regimes. This stems from inappropriateness of a single linear model to describe a nonlinear behavior of *PCT*.

2.2. Multi-Model Inferential Sensor

Prediction capability of a linear inferential sensor can be improved when considering a multi-model sensor structure. The MIS formulation with two models can be written as follows (Mojto et al., 2022):

$$
\hat{y} = \begin{cases} \boldsymbol{m}^{\intercal}\boldsymbol{a}_1 + a_{0,1}, & \text{if } \boldsymbol{m} \in \mathcal{R}_1, \\ \boldsymbol{m}^{\intercal}\boldsymbol{a}_2 + a_{0,2}, & \text{if } \boldsymbol{m} \in \mathcal{R}_2, \end{cases} \tag{8}
$$

where $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ represent vectors of parameters of the first and second model, respectively, and $a_{0,1}$ and $a_{0,2}$ are constant off-sets of the first and second model, respectively. The regions of individual model validity are denoted as $\mathcal{R}_j$ and represent convex polyhedra such that $\mathcal{R}_1 \bigcap \mathcal{R}_2 = \emptyset$. Consideration of more than two models is possible in a similar setup.

15

Design of MIS follows a three stages as described for SIS. The design phase uses the following state-of-the-art workflow:

1. *Clustering for a priori labeling of the training dataset.* The labeling results from the dataset characteristics (e.g., distinction of operating points). An appropriate clustering approach (e.g., k-means clustering) can be used.

2. *Classifier training based on the labeled training dataset.* The purpose of the classifier is to determine the corresponding model class of a measurement point. In this paper, the support vector machines (SVM) approach is considered using linear separators (Boser et al., 1992). The separators establish the polyhedral partitions $\mathcal{R}_j$ in (8).

3. *Training of the individual MIS models.* The individual MIS models for each class can be fitted using some of the SIS training methods (see in Section 2.1.1).

The design phase of MIS is illustrated in Figure 2 (upper part). This phase is performed offline using the available input and output training datasets ($\boldsymbol{m}(\mathcal{T})$ and $\boldsymbol{y}(\mathcal{T})$). Its purpose is to establish or update classification regions ($\mathcal{R}_1$ and $\mathcal{R}_2$) and model parameters ($\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $a_{0,1}$, and $a_{0,2}$) required for the prediction of MIS, as depicted in the bottom section of Figure 2. In contrast to the design phase, the prediction phase is conducted online, while each incoming (actual) measurement of input variables ($\boldsymbol{m}_i$) undergoes a classification to be assigned to one of the considered regions ($\mathcal{R}_1$ or $\mathcal{R}_2$). Subsequently, the actual prediction (y_i) is computed using the individual MIS model corresponding to the determined class (region). These calculations are performed within the time period represented by index i.
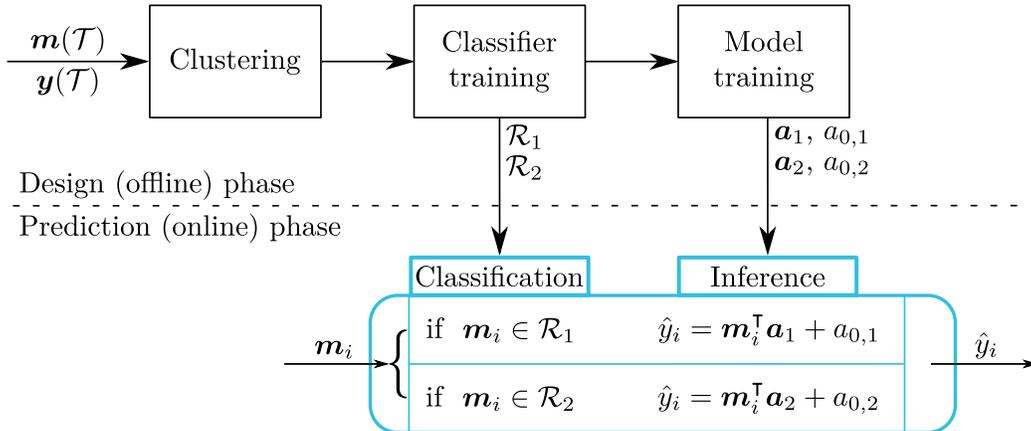
Figure 2: A simplified workflow of design and prediction phases of MIS.

In the following, we abbreviate the sensor designed by this procedure as $\mathrm{MIS_{SotA}}$ and we refer it to as a state-of-the-art approach although the presented procedure is our contribution. This is because there exist no consistent (agreed-upon) technique for the MIS design according to the best of authors' knowledge.

The MIS training assumes an a priori choice of the number of models within the MIS structure. Naturally, more the number of models considered, higher the susceptibility of the resulting sensor to overfitting. The number of models within the MIS structure is thus a tuning parameter of the scheme, whose value can be decide based on cross-validation.

Illustrative example (continued). The usage of the $\mathrm{MIS_{SotA}}$ on the *PCT* dataset is shown in Figure 3(a). The designed models are presented with a yellow surface (Model 1) and a dark green surface (Model 2). The example considers a priori labeling by k-means clustering.

The advantage of using MIS is obvious, as its accuracy (RMSE $(\mathcal{T})$ =

0.032) significantly outperforms the best SIS (Figure 1(b), RMSE ($\mathcal{T}$) = 0.064). This confirms that the MIS models can better explain the nonlinear behavior of PCT compared to SIS. Furthermore, the structure of MIS is flexible as it can involve more models. There are two primary limitations (challenges) of MIS_{SotA}: (a) the designed models are not necessarily continuous, and (b) a priori labeling is unaware of its impact on the accuracy of the resulting inferential sensor. The first drawback can be seen in Figure 3(a). There is a visible discrepancy between the designed models of MIS_{SotA} at the intersection of the surfaces. This behavior can cause issues with the stability of the control strategy if the MIS is involved.

A glimpse of the proposed solution in Figure 3(b) (approach MIS_{con} will be introduced in detail in Section 3.1) reveals that it is possible to achieve continuity when switching between the MIS models, yet potentially, at the expense of model accuracy. In the studied example, the discrepancy between ground truth and the designed models originates from the rotation of Model 1 (the yellow surface in Figure 3(b)) to achieve the desired continuity when switching to Model 2 (the dark green surface). The rotation can be reduced by putting more weight on accuracy and relaxing the continuity constraint when designing a continuous MIS (as discussed in Section 3.1) or optimizing a priori labeling (as discussed in Section 3.2).

3. Solution Approach

We propose a novel approach for the MIS design consisting of two separate developments dealing with the limitations of the MIS_{SotA} approach. The first part deals with the approach for the design of continuous MIS (referred to
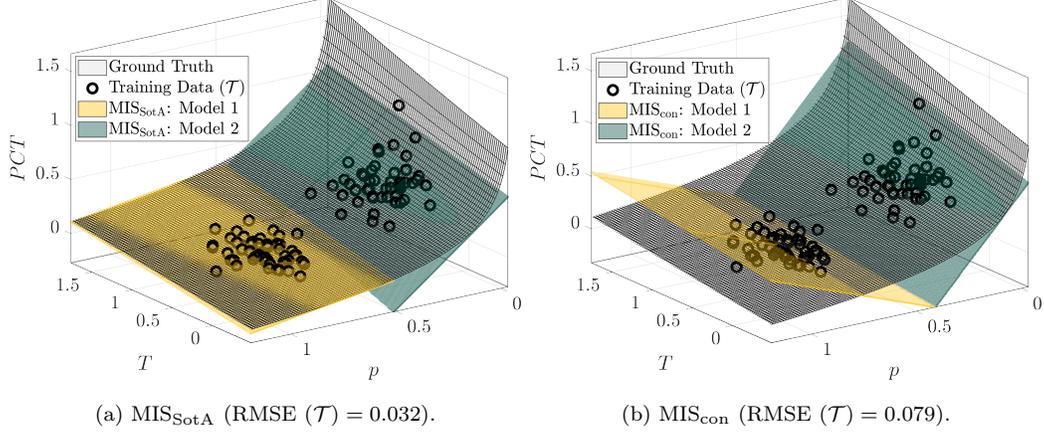
18

(a) MIS_{SotA} (RMSE ($\mathcal{T}$) = 0.032).

(b) MIS_{con} (RMSE ($\mathcal{T}$) = 0.079).

Figure 3: The ground truth model of PCT with MIS_{SotA} and MIS_{con} designed on the dataset with two distinct clusters.

as a MIS_{con} approach). The second part extends the design of MIS with optimized data labeling (referred to as a $\text{MIS}_{\text{con,lab}}$ approach).

3.1. Design of a Continuous MIS

To deal with the limitation of discontinuity of MIS models, we propose a combination of the SVM-based classification of the data with the individual sensor training in the following optimization problem:

$$\min_{\substack{\boldsymbol{w},w_0,\boldsymbol{e}\geq 0 \\ \boldsymbol{a}_1,a_{0,1},\boldsymbol{a}_2,a_{0,2}}} \text{SSE}_1 + \text{SSE}_2 + \alpha\|\boldsymbol{w}\|_2^2 + \beta\|\boldsymbol{e}\|_1 \tag{9a}$$

$$\text{s.t.} \quad (2z_i - 1)\left(\boldsymbol{m}_i^{\mathsf{T}}\boldsymbol{w} + w_0\right) \geq 1 - e_i, \quad \forall i \in \{1, 2, \ldots, n\}, \tag{9b}$$

$$\text{SSE}_1 = \sum_{\forall i \in \mathcal{T}} z_i \left(y_i - \boldsymbol{m}_i^{\mathsf{T}}\boldsymbol{a}_1 - a_{0,1}\right)^2, \tag{9c}$$

$$\text{SSE}_2 = \sum_{\forall i \in \mathcal{T}} (1 - z_i)\left(y_i - \boldsymbol{m}_i^{\mathsf{T}}\boldsymbol{a}_2 - a_{0,2}\right)^2, \tag{9d}$$

$$\boldsymbol{a}_1 - \boldsymbol{a}_2 - \boldsymbol{w} = 0, \quad a_{0,1} - a_{0,2} - w_0 = 0, \tag{9e}$$

19

where $\boldsymbol{w}$ is a normal vector and w_0 constant off-set of the separation hyperplane, respectively, $\boldsymbol{e}$ is a vector of the slack variables, $\boldsymbol{z}$ is a vector of binary parameters that results from the data labeling procedure with $z_i = 1$ if $\boldsymbol{m}_i \in \mathcal{R}_1$ and $z_i = 0$ if $\boldsymbol{m}_i \in \mathcal{R}_2$, SSE is the sum of squared errors, α is a weighting parameter for normal vector of the separation plane and β is a weighting parameter for vector of the slack variables.

The combination of the SVM-based classification of the data with the individual sensor training is represented by (9a)–(9d). The resulting optimization problem is extended with constraints (9e) which ensure the continuity at the switch between the two models. This is achieved by establishing the intersection of model surfaces to coincide with the determined switching hyperplane. To prove this claim, we first notice the following equivalence regarding the models intersection condition:

$$\boldsymbol{m}^\mathsf{T}\boldsymbol{a}_1 + a_{0,1} = \boldsymbol{m}^\mathsf{T}\boldsymbol{a}_2 + a_{0,2} \Leftrightarrow \boldsymbol{m}^\mathsf{T}(\boldsymbol{a}_1 - \boldsymbol{a}_2) + a_{0,1} - a_{0,2} = 0. \tag{10}$$

The mutual intersection of the models and the switching hyperplane is established by the following condition:

$$\boldsymbol{m}^\mathsf{T}(\boldsymbol{a}_1 - \boldsymbol{a}_2) + a_{0,1} - a_{0,2} = \boldsymbol{m}^\mathsf{T}\boldsymbol{w} + w_0 \Leftrightarrow$$

$$\Leftrightarrow \boldsymbol{m}^\mathsf{T}\underbrace{(\boldsymbol{a}_1 - \boldsymbol{a}_2 - \boldsymbol{w})}_{=0} + \underbrace{a_{0,1} - a_{0,2} - w_0}_{=0} = 0, \tag{11}$$

from which (9e) follows and which completes the proof. Note that, we present the formulation of $\mathrm{MIS_{con}}$ in the simplest form (two-model MIS, OLSR setup), for brevity, yet it is possible to extend easily this formulation to multiple models and other training approaches (see Section 2.1.1).

As the a priori data labeling can be inappropriate for the design of a continuous MIS, we allow small violations of the labeling using the slack

variables $\boldsymbol{e}$ in (9b). We also consider that the user can aim at giving up some portion of model (training) accuracy for the better separation by widening the separation band. The latter feature is established by minimizing $\|\boldsymbol{w}\|_2^2$ in (9a). The described features can be enforced/weakened by tuning the positive weights α and β.

Illustrative example (continued). The visualization of $\mathrm{MIS_{con}}$ on the *PCT* dataset can be seen in Figure 3(b). Unlike the $\mathrm{MIS_{SotA}}$ approach (Figure 3(a)), there is no discrepancy at the intercept between the designed models (yellow and dark green surfaces) of $\mathrm{MIS_{con}}$. This confirms that the proposed approach ensures continuity when switching between designed models. The accuracy of $\mathrm{MIS_{con}}$ (RMSE $(\mathcal{T})$=0.079) is significantly decreased compared to $\mathrm{MIS_{SotA}}$ (RMSE $(\mathcal{T})$=0.032). This is a price to pay for MIS continuity and a design trade-off.

Naturally, the continuity constraints (9e) can be relaxed and introduced as soft constraints should one be willing to make the trade-off explicit for the MIS design. The accuracy and continuity of the $\mathrm{MIS_{con}}$ model can then be effectively tuned by varying the weights α and β according to the fidelity of a priori labeling and a desired level of discontinuity. The other way to improve the performance of the $\mathrm{MIS_{con}}$ approach represents an implementation of the optimization of a priori labeling into the MIS design, which is further explored in the following text.

3.2. Design of a Continuous MIS with Optimized Data Labeling

In order to mitigate the inaccuracies caused by the a priori labeling of the training dataset, we propose the approach to design MIS with optimized data

labeling ($\text{MIS}_{\text{con,lab}}$). This approach searches directly for the optimal data labeling by adding $\boldsymbol{z}$ among the optimized variables in (9a). The resulting optimization problem is following:

$$\min_{\substack{\boldsymbol{z}\in\{0,1\}^n,\boldsymbol{w},w_0,\boldsymbol{e}\geq 0 \\ \boldsymbol{a}_1,a_{0,1},\boldsymbol{a}_2,a_{0,2}}} \text{SAE}_1 + \text{SAE}_2 + \alpha\|\boldsymbol{w}\|_1 + \beta\|\boldsymbol{e}\|_1 \tag{12a}$$

$$\text{s.t.} \quad (2z_i - 1)\left(\boldsymbol{m}_i^\intercal\boldsymbol{w} + w_0\right) \geq 1 - e_i, \quad \forall i \in \{1, 2, \ldots, n\}, \tag{12b}$$

$$\text{SAE}_1 = \sum_{\forall i \in \mathcal{T}} z_i\left|y_i - \boldsymbol{m}_i^\intercal\boldsymbol{a}_1 - a_{0,1}\right|, \tag{12c}$$

$$\text{SAE}_2 = \sum_{\forall i \in \mathcal{T}} (1 - z_i)\left|y_i - \boldsymbol{m}_i^\intercal\boldsymbol{a}_2 - a_{0,2}\right|, \tag{12d}$$

$$\boldsymbol{a}_1 - \boldsymbol{a}_2 - \boldsymbol{w} = 0, \quad a_{0,1} - a_{0,2} - w_0 = 0, \tag{12e}$$

where SAE is a sum of absolute errors.

Although a formulation similar to (9) with SSE-based objective can be reused here, we adopt the SAE criterion to reduce the complexity. In a similar fashion, 2-norm is replaced for 1-norm to regularize the normal vector of the separating hyperplane. This is a standard approach (Song et al., 2002). The optimization problem (12) can thus be transformed to a mixed-integer linear program (MILP). The transformation uses: (a) the epigraph reformulation (Milano, 2012) of the absolute value, (b) the big-M method (Griva et al., 2008) to linearize the bilinear constraints. As the variables $\boldsymbol{z}$ are binary, the big-M method does not require any new integer variables. If SSE was used in the objective function, the optimization problem would turn into mixed-integer nonlinear program (MINLP), which might be challenging especially when n is high.

The problem (12) is primarily used to determine the data labels, which refer to how the training data is distributed and how the validity regions

22

of the model are established. In principle, this approach parameterizes the choice of data distributions into model classes and thus makes the design independent of the used clustering algorithm. Once the values of z (data labels) are fixed, the final training of MIS models is performed by solving (9) using the SSE criterion. This ensures a fair comparison with other SIS and MIS approaches.

This two-step approach does not require a priori labeling of the training set and can provide an optimal MIS at the expense of increased computational burden. The optimization problem for the $\mathrm{MIS_{con,lab}}$ design increases by one binary optimized variable per training data point. Therefore, the proposed approach is limited to relatively small-scale problems (tens to hundreds of measurements). However, this is typically sufficient for the design of inferential sensors, where only a limited number of measurements is available for the desired (hard-to-measure) variable. Additionally, if a large dataset is available, a smaller size of the training dataset can be selected based on appropriate information criteria, similar to optimal design of experiments or sampling for surrogate model building (Kamath, 2022).

Illustrative example (continued). Figure 4 shows the design of $\mathrm{MIS_{con,lab}}$ on the different PCT datasets. The results in Figure 4(a) show that the designed $\mathrm{MIS_{con,lab}}$ has a high degree of flexibility and precision on the PCT dataset with two distinct clusters. The accuracy of this sensor (RMSE $(\mathcal{T}) = 0.026$) outperforms $\mathrm{MIS_{SotA}}$ (RMSE $(\mathcal{T}) = 0.032$) and $\mathrm{MIS_{con}}$ (RMSE $(\mathcal{T}) = 0.079$) shown in Figure 3. The accuracy improvement of $\mathrm{MIS_{con,lab}}$ is ensured by optimizing a priori labeling instead of using k-means clustering. This can be indicated by comparing classification in Figure 4(a) against Figure 3. The
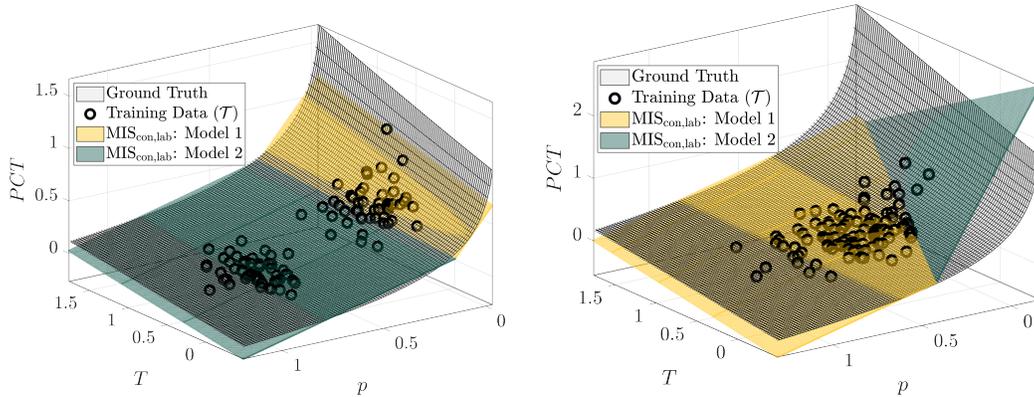
(a) $\text{MIS}_{\text{con,lab}}$ (RMSE $(\mathcal{T}) = 0.023$) designed on the PCT dataset with two distinct clusters.

(b) $\text{MIS}_{\text{con,lab}}$ (RMSE $(\mathcal{T}) = 0.043$) designed on the PCT dataset with indistinguishable clusters.

Figure 4: The ground truth model of PCT with $\text{MIS}_{\text{con,lab}}$ designed on different datasets.

results from Figure 4(b) indicate that the sensor returned by the $\text{MIS}_{\text{con,lab}}$ approach is designed effectively even when the considered dataset has no distinguishable clusters. This is typical for industrial datasets due to the presence of the significant level of noise and multitude of similar operating points.

4. Case Studies

The design of single-model inferential sensors (SIS) and multi-model inferential sensors (MIS) is elaborated on two case studies. Both case studies have an industrial character and practical relevance. The first case study features a pressure compensated temperature PCT, which is briefly explored in Section 2.1.1. The purpose of this case study is to analyze the impact of data quality on the MIS design in multitude of simulations. The second case study involves an industrial dataset from the VGH unit, which is a part of the oil refinery Slovnaft, a.s. in Bratislava, Slovakia. This case study validates the

applicability of the proposed MIS design approaches in practice.

4.1. Implementation Details

The presented design methods are implemented in MATLAB R2022a. To solve the involved optimization problems, we use the Yalmip package (Löfberg, 2004) and Gurobi solver (Gurobi Optimization, LLC, 2023). All the numerical results and graphical representations consider the normalization of variables within the interval $[0, 1]$ in both case studies. The normalization (scaling) parameters are not disclosed for the dataset from the VGH unit due to data confidentiality. For comparison, we also use a single-model sensor represented via neural network. The neural network is trained using Matlab's Neural Network toolbox.

Prior to the inferential sensor design, the entire available dataset is divided into training and testing (unseen) datasets. The information contained within the training and testing datasets is one of the decisive factors directly affecting the performance of the designed inferential sensors. Therefore, the effect of various ways of dividing the data into training and testing datasets on the SIS and MIS performances is further investigated on the PCT dataset in Section 4.2. The inferential sensor design on the industrial dataset from the VGH unit is based on the random distribution of measurements in the training and testing datasets.

The design of MIS_{SotA} performs the a priori labeling by using k-means clustering. Subsequently, the linear classifier of MIS_{SotA} is designed by SVM. Finally, the model parameters within MIS_{SotA} models are calculated by OLSR. The MIS_{con} design is performed according to (9). In order to reduce the computational effort, MIS_{con} is initialized by the results from (9) considering

SAE instead of SSE within the objective function (9a). Subsequently, the $\text{MIS}_{\text{con,lab}}$ design from (12) is initialized by the results from MIS_{con}.

4.2. Design of Inferential Sensors for Pressure-Compensated Temperature

We use the datasets generated by simulating the nonlinear model of PCT represented by (6) with respect to the parameters and specifications introduced in Section 2.1.1. We use this ideal case study to examine the impact of various factors on the performance of the inferential sensor. Specifically, in this section, we analyze the impact of two factors on the SIS and MIS designs: (1) the method of data distribution into the training and testing datasets and (2) the noise variance in the output variable. Overall, the studied PCT dataset involves 620 measurements, which are equally distributed between the training and testing datasets. We use this setup for comparison purpose as it will be used in the second case study due to nature of the considered plant operation.

The considered case study involves two input variables (p and T) and one output variable (PCT), and therefore, it is unnecessary to consider the SIS approaches with advanced input structure selection (i.e., SIS_{PCR}, SIS_{PLSR}, SIS_{LAS}, and SIS_{SS}). We use a soft-sensor designed by a neural network (designated as SIS_{NN}) to compare the linear and multi-linear sensors against a nonlinear SIS. The structure of the network was optimized using cross-validation. The best structure involves fully-connected neural network with two inputs, no hidden layer, and one neuron in the output layer with hyperbolic tangent as an activation function. The set of compared inferential sensors in this section involves SIS_{OLSR}, MIS_{SotA}, MIS_{con}, and $\text{MIS}_{\text{con,lab}}$.

To analyze the impact of the data distribution into the training and test-

(a) The desirable scenario.
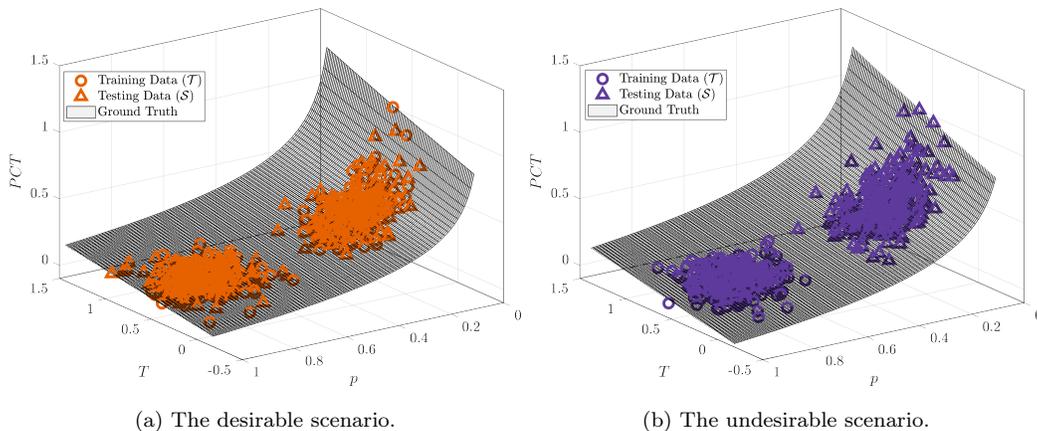
(b) The undesirable scenario.

Figure 5: The comparison of studied distributions of the PCT data into training and testing datasets.

ing datasets, we generate datasets from the PCT model according to two different scenarios. The first scenario, or desirable scenario, considers that the PCT is sampled in two different operating regimes and measurements from both operating regimes are available in the training dataset, as illustrated in Figure 5(a). This scenario occurs relatively frequently in the industry. It assumes that the process operates only within known operating regimes, which is desired for the inferential sensor design. The second scenario, or undesirable scenario, assumes the same operating regimes as the first scenario, but the training dataset involves measurements from one operating regime only, and the testing dataset involves measurements from the other operating regime, as shown in Figure 5(a). This scenario represents an undesirable, yet not unlikely, situation in the industry, when the process operates within a new operating state after the inferential sensor design.

The comparison in Figure 6 shows a statistical evaluation of the results obtained over 100 different datasets for each studied scenario. The datasets

consider two classes of measurements, as shown in Figure 5, with different random distributions of measurements. The noise considered within the output variable is a random variable from $\mathcal{N}(\mu_{\text{noise}}, \sigma^2_{\text{noise}}) = \mathcal{N}(0, 5^2)$, where the value of the standard deviation σ_{noise} represents $0.67\,\%$ ($\sigma_{\text{noise}}/\mu_{PCT} \cdot 100\,\%$) of an averaged value from the original (prior to the normalization) output variable. The boxes in Figure 6 represent the 25th and 75th percentiles of the RMSE reached on the testing test. The red line within the box represents the median value of the considered set of results. The red crosses represent the statistical outliers. For each designed inferential sensor (i.e., SIS_{OLSR}, SIS_{NN}, MIS_{SotA}, MIS_{con}, and $\text{MIS}_{\text{con,lab}}$), there is a pair of orange (desirable scenario) and violet (undesirable scenario) boxes in Figure 6. The trained MIS models involve two models because of the nature of the dataset (in the desirable scenario).

The results presented in Figure 6 demonstrate that the designed inferential sensors exhibit better performance on datasets from the desirable scenario (represented by the orange boxes) compared to those from the undesirable scenario (represented by the violet boxes). These findings support our initial assumption regarding the impact of these scenarios on the performance of inferential sensors.

4.2.1. Analysis of the Desirable Scenario

The results from the desirable scenario further indicate a high degree of robustness (or low variance) in the SIS_{OLSR} and SIS_{NN} accuracy, as seen by the small height of the corresponding orange boxes in Figure 6. On the other hand, the accuracy of SIS_{OLSR} is significantly outperformed by MIS_{NN}, which results from the ability of neural network to represent nonlinear transforma-
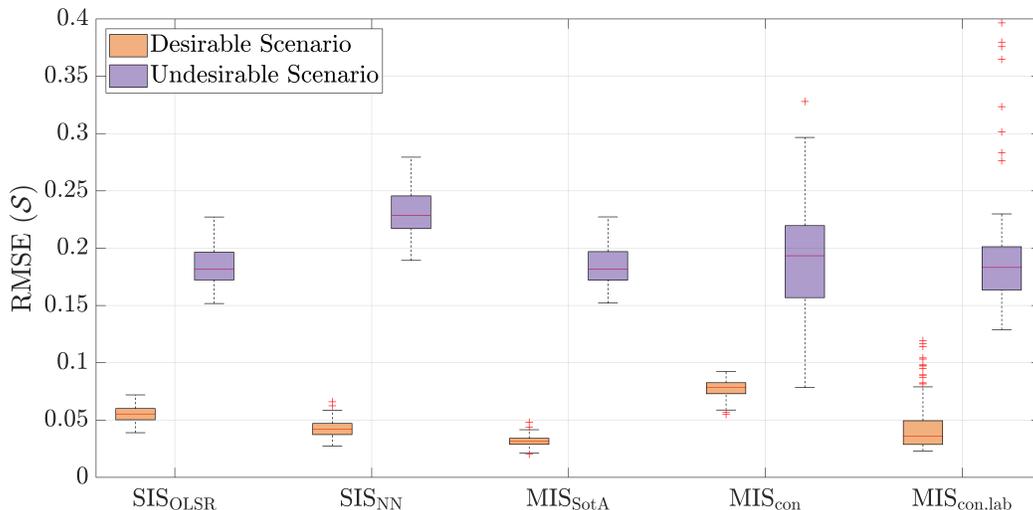
Figure 6: The statistical comparison of RMSE ($\mathcal{S}$) of designed inferential sensors (SIS_{OLSR}, SIS_{NN}, MIS_{SotA}, MIS_{con} and $\text{MIS}_{\text{con,lab}}$) involving 100 different datasets for each studied scenario.

tion from input to output. The accuracy of SIS_{OLSR} is also significantly outperformed by MIS_{SotA}. The performance variance of MIS_{SotA} is even smaller compared to that of SIS_{OLS} and SIS_{NN}. MIS_{SotA} even provides better performance than SS_{NN}, which can be attributed to the number of training data, where neural network cannot identify the nonlinearity in the output to a sufficient extent. The outstanding performance of MIS_{SotA} can be attributed to the nature of the datasets in the desirable scenario (as shown in Figure 5(a)) with distinguishable classes. One class involves measurements that precisely explain the behavior of PCT in the (almost) linear section, while the other class involves mostly measurements from the highly nonlinear section of the PCT model range. Therefore, a priori labeling within the MIS_{SotA} approach (k-means clustering) provides appropriate data labels for the subsequent design of inferential sensors. One can conclude that multi-model linear sensor

29

would be a reasonable choice in this case study.

The results of MIS_{con} (as shown in the corresponding orange box in Figure 6) indicate a small variance in accuracy comparable to that of SIS_{OLSR} and SIS_{NN}. However, MIS_{con} achieves the lowest accuracy compared to other designed inferential sensors on the datasets from the desirable scenario. This suggest that the a priori labeling is not optimal w.r.t. design of a continuous multi-model sensor. Further analysis of the results reveals that the MIS_{con} accuracy (considering the desirable scenario) can be significantly improved by the $MIS_{con,lab}$. The results in the desirable scenario exhibit higher variability in the performance compared to all other designed inferential sensors. The main reason for this is the complexity of the optimization problem that needs to be solved, which can lead to numerical inaccuracies. This assumption is supported by the increased occurrence of outliers in the $MIS_{con,lab}$ results in Figure 6 (represented by the red crosses). Despite the increased variance of the accuracy, $MIS_{con,lab}$ achieves comparable accuracy to MIS_{SotA} in the majority of the cases. This is notable given the similar values of the median (represented by the red lines) of these MISs.

4.2.2. Analysis for the Undesirable Scenario

The results from the undesirable scenario (violet boxes) indicate that SIS_{OLSR} and MIS_{SotA} show similar performance and are the most accurate among the studied sensors. Unlike the observations from the desirable scenario, the accuracy of MIS_{SotA} is only slightly higher than that of SIS_{OLSR}. On the other hand, SIS_{NN} exhibits poor extrapolation properties and performs the worst among all studied sensors. Despite the challenge of making extrapolated predictions, MIS_{SotA} still performs the best and exhibits the

30

lowest variance of accuracy among all the sensors.

We can observe very high variability in the MIS_con accuracy. This stems from the requirement to design continuous models, which restricts the rotation and angle between the models. The results suggest that the variance and accuracy of MIS_con can be improved by using the $\text{MIS}_\text{con,lab}$ approach. However, the possibilities of $\text{MIS}_\text{con,lab}$ are limited due to the nature of the training/testing dataset. As a result, we observe similar accuracy of $\text{MIS}_\text{con,lab}$ compared to SIS_OLSR and MIS_SotA, as indicated by the median values. We also note the occurrence of low-accuracy outliers within the $\text{MIS}_\text{con,lab}$ results, similar (yet more pronounced) to the desirable scenario.

This scenario represents an extreme case where the estimated behavior is completely unknown during the training phase. However, in real-world applications, it is more likely that the studied problem will fall in-between the desirable and undesirable scenarios. Therefore, we can conclude that the results in Figure 6 confirm that MIS_SotA and $\text{MIS}_\text{con,lab}$ have the potential to more accurately capture nonlinear behavior compared to the linear SIS_OLS, as observed in Figures 3(a) and 4(a). The difference in accuracy between linear MIS and SIS would be even more pronounced in more nonlinear case studies.

4.2.3. Analysis of the Measurement Noise

The previous analysis shows that the random distribution of data into the training and testing datasets (the desirable scenario) should provide sufficient informative content for the training dataset. Therefore, we use this distribution in the following analysis focused on the impact of noise (in the output variable) on the performance of the designed inferential sensors. The
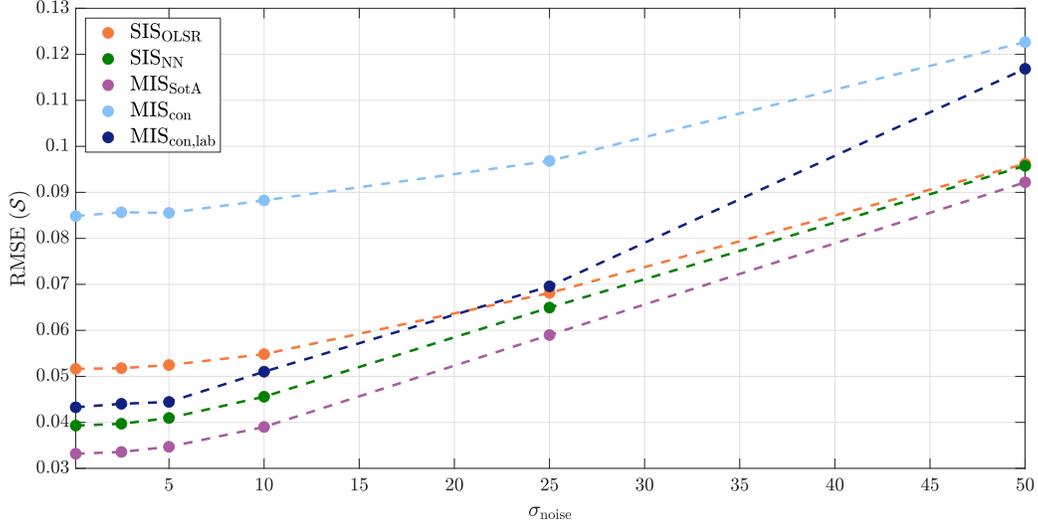
Figure 7: The averaged RMSE ($\mathcal{S}$) value for designed inferential sensors (SIS$_{\text{OLSR}}$, MIS$_{\text{SotA}}$, MIS$_{\text{con}}$, and MIS$_{\text{con,lab}}$) from 100 different realizations of the noise for each studied σ_{noise} within the output variable.

set of noise variances is selected as $\sigma_{\text{noise}} \in \{0.1, 2.5, 5, 10, 25, 50\}$. The minimum, respectively maximum, considered noise variance ($\sigma_{\text{noise}} = 0.1$, respectively $\sigma_{\text{noise}} = 50$) represent approximately $0.1\,\%$, respectively $6.7\,\%$, of the mean value of the output variable from the considered datasets. To provide more representative results, we present the median RMSE ($\mathcal{S}$) value from 100 different realizations of noise for each studied σ_{noise}. The purpose of this analysis is to investigate the practical applicability of the studied MISs concerning data quality, specifically focusing on the influence of noise. This aspect becomes crucial in industrial applications, where the accuracy of measurements is closely linked to the quality of available measurement and sensing devices.

The impact of the noise variance (within the output variable) on the accu-

racy (RMSE ($\mathcal{S}$)) of the designed inferential sensors is illustrated in Figure 7. In general, an increase in σ_{noise} leads expectedly to an increase in RMSE ($\mathcal{S}$). The performance of SIS_{OLSR} (orange points) appears to be relatively robust as the overall RMSE increase is the smallest among all the studied sensors iver the entire studied noise magnitude interval. Regarding SIS_{NN} and MIS_{SotA} (green and magenta points, respectively), the previously indicated trend is confirmed. The superior performance of MIS_{SotA} is expected due to the nature of the considered datasets (see Figure 5(a)) and nonlinearity of the ground-truth model, as discussed above. The performance of SIS_{NN} sees a deterioration in performance with increasing noise level and even reaches the performance of a linear sensor (SIS_{OLSR}) for the highest value.

The (low) accuracy of MIS_{con} (represented by pale blue points) follows the trends already observed. It is interesting to see that increased noise levels do not have a dramatic impact on performance, so the previously observed large performance variance can be mitigated by a considerate training, involving shuffling of training data and cross-validation. The results of $\text{MIS}_{\text{con,lab}}$ (blue points) indicate that the accuracy of MIS_{con} can be significantly improved by optimizing the data labeling. $\text{MIS}_{\text{con,lab}}$ outperforms SIS_{OLSR} when $\sigma_{\text{noise}} < 20$. The accuracy of $\text{MIS}_{\text{con,lab}}$ decreases steeper compared to other approaches when $\sigma_{\text{noise}} > 20$. The increased flexibility of $\text{MIS}_{\text{con,lab}}$ leads to an increased tendency for this approach to explain the noise within the training dataset, especially for significant noise variances. In other words, the $\text{MIS}_{\text{con,lab}}$ approach has an increased tendency for overfitting for large noise levels.
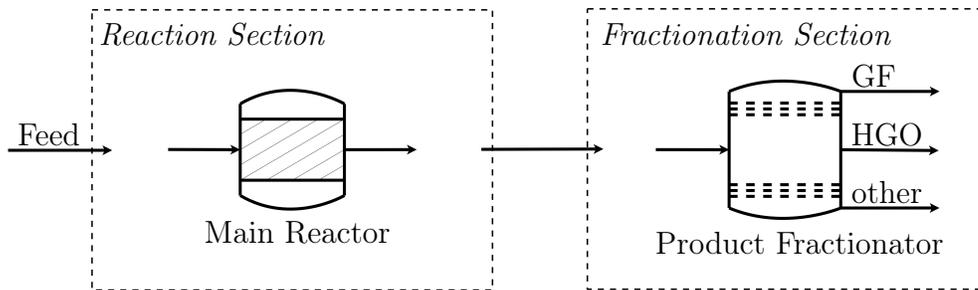
Figure 8: A schematic diagram of the VGH unit.

4.3. Design of Inferential Sensors for Vacuum Gasoil Hydrogenation Unit

The Vacuum Gasoil Hydrogenation (VGH) unit is an essential part of the oil refinery Slovnaft, a.s. in Bratislava, Slovakia. This unit (schematic in Figure 8 processes the vacuum distillates in two consecutive sections. The first section is a high-pressure reaction that significantly reduces the concentration of the impurities (e.g., nitrogen and sulfur) in the feed mixture. The second section is a low-pressure fractionation, where a product fractionator (main distillation column) separates the pre-treated vacuum distillates into a gasoline fraction (GF), a hydrogenated gasoil (HGO), and other products.

The available industrial dataset involves measurements for 24 months of the VGH unit operation. The output (desired) variable to be estimated by inferential sensors is the purity of the HGO product, represented by 95 % point of the distillation curve $T_{95\,\%;\mathrm{HGO}}$. The lab analysis of the HGO product is executed approximately once per day, and therefore, there are 621 measurements of the output variable available for the inferential sensor design. The input variables are measured every minute by online sensors. In order to reduce the impact of the measurement noise, the minutely measurements are replaced by the averaged measurements from 30-minute intervals. The

34

resulting input dataset involves 27,324 measurements. The available output measurements are collocated time-wise within the matching 30-minute interval among the available input data.

The set of input candidates involves the following 35 variables:

$$\boldsymbol{m}_i = (T_{\mathrm{F,5p}}, T_{\mathrm{F,50p}}, T_{\mathrm{F,95p}}, T_{\mathrm{col}}, T_{\mathrm{vap,t}}, T_{\mathrm{ex,1}}, T_{\mathrm{ex,2}},$$
$$T_{\mathrm{ex,3}}, T_{\mathrm{ex,4}}, T_{\mathrm{wb,1}}, T_{\mathrm{wb,2}}, T_{\mathrm{wb,3}}, T_{\mathrm{wb,4}}, T_{\mathrm{wb,5}},$$
$$x_{\mathrm{F,N_2}}, x_{\mathrm{F,S}}, x_{\mathrm{H_2}}, \mathrm{rt}_1, \mathrm{rt}_2, \mathrm{rt}_3, \mathrm{rt}_4, \mathrm{rt}_5, \mathrm{rt}_6, \mathrm{rt}_7, \tag{13}$$
$$F_{\mathrm{F}}, F_{\mathrm{F,rec}}, F_{\mathrm{R}}, F_{\mathrm{h}}, p_{\mathrm{col}}, L_{\mathrm{reb}}, \mathrm{vo}_{\mathrm{R}}, \mathrm{vo}_{\mathrm{h,1}}, \mathrm{vo}_{\mathrm{h,2}},$$
$$PCT_{\mathrm{HGO}}, PCT_{\mathrm{GF}})_i^{\intercal},$$

where $T_{\mathrm{F,5p/50p/95p}}$ is a temperature from the feed stream distillation curve for various points (e.g., 5 %, 50 %, 95 %) of the distillation curve, T_{col} is a temperature within the column section, $T_{\mathrm{vap,t}}$ is a temperature of the vapor phase on the top of the column, $x_{\mathrm{F,N_2/S}}$ is a content of the nitrogen/sulfur in the feed stream, $F_{\mathrm{F/R}}$ is a flowrate of the feed/reflux stream, $F_{\mathrm{F,rec}}$ is reconciled flowrate of the feed stream, F_{h} is a flowrate of heat medium to the reboiler, p_{col} is a pressure within the column section, L_{reb} is a liquid level in the reboiler, vo_{R} is a valve opening of the reflux stream, $\mathrm{vo}_{\mathrm{h,1-2}}$ is a valve opening of the heating medium for the reboiler (1 – input, 2 – output) and $PCT_{\mathrm{HGO/GF}}$ is a pressure-compensated temperature for HGO/GF stream from the product fractionator. Note that $PCT_{\mathrm{HGO/GF}}$ are calculated using (undisclosed) industrial data for the sake of fair comparison with plant current standards. The variables described above are measured directly at or very close to the product fractionator (low-pressure fractionation section). The remaining variables in (13) are located in the high-pressure reaction section, where $T_{\mathrm{ex,1-4}}$ is a temperature at the various locations of the catalytic

reactors monitoring (exothermic) reaction, $T_{\text{wb},1-5}$ is weighted average bed temperature at the different locations of the catalytic reactors, x_{H_2} represents the content of the hydrogen in the reaction section, rt_{1-7} is a ratio of the gas/liquid phases in the different areas of the reaction section.

The supervised learning methods used for inferential sensor design require a paired input-output dataset, where the inputs and outputs correspond to the same measurement time. In this case, the input and output datasets consist of 621 measurements, which are subsequently divided randomly into training (311 measurements) and testing (310 measurements) sets. The equal division between the training and testing datasets stems from the nature of the plant operation that covers a large operating window. We use this setup to guarantee a fair comparison and representativeness of the testing data. It is worth noting that the dataset is comparable in size to the PCT dataset (620 measurements) used in the previous case study (Section 4.2). This allows us to explore any similarities between the results and conclusions of the case studies.

Currently, the refinery has implemented a univariate SIS ($n_{\text{p}}^* = 1$) with the following structure:

$$\hat{y}_i = PCT_{\text{HGO},i} \cdot a_1 + a_0. \tag{14}$$

This structure is based on the expert knowledge of the operators and engineers in the refinery and is used as a reference in the comparison of designed inferential sensors in this study. Therefore, we refer to this sensor as SIS_{Ref}. The OLSR approach is used to evaluate the parameters in (14).

The set of approaches considered for SIS design (see Section 2.1.1) includes OLSR (SIS_{OLSR}), PCR (SIS_{PCR}), PLSR (SIS_{PLSR}), and LASSO (SIS_{LAS}).

These approaches consider the entire set of input candidates ($n_\mathrm{p} = 35$) and search for the optimal input structure (n_p^*) based on their specific objectives. Similar to the previous case study, we use a neural network (designated as $\mathrm{SIS_{NN}}$) to compare the linear and multi-linear sensors against a nonlinear SIS. The structure of the network was optimized using cross-validation. The same structure as in the previous case study (fully-connected neural network with 35 inputs, no hidden layer, and one neuron in the output layer with hyperbolic tangent as an activation function) was concluded to be the best.

The set of compared approaches for MIS design consists of $\mathrm{MIS_{SotA}}$, $\mathrm{MIS_{con}}$, and $\mathrm{MIS_{con,lab}}$. The MIS approaches consider two candidate input structures: (1) the reference input structure (denoted as Ref) given by (14) with $n_\mathrm{p}^* = 1$ (PCT_HGO) and (2) the input structure determined by the subset selection method with $n_\mathrm{p}^* = 2$ (denoted as SS) with $n_\mathrm{p}^* = 2$ (PCT_HGO and $\mathrm{vo_{h,2}}$). The SS input structure is determined using the $\mathrm{MIS_{con}}$ approach for structures selected by the SS method with varying number of non-zero model parameters ($\tilde{n}_\mathrm{p}$). We selected $n_\mathrm{p}^* = 2$ as a compromise between simplicity and performance in the RMSE ($\mathcal{T}$) criterion. The choice of simple structure has two merits. Firstly, expert knowledge from the refinery suggests to consider simple input structures (see reference structure in (14)). Secondly, the $\mathrm{MIS_{con,lab}}$ approach tends to overfit, as indicated in Section 4.2. To provide a fair comparison, the SIS set of approaches is extended by $\mathrm{SIS_{Ref}}$ and $\mathrm{SIS_{SS}}$, considering the same input structures as the MIS approaches. The MIS approaches involve two models as a compromise between simplicity and rate of overfitting. The k-means method suggested the number of clusters to be 2–5.

Table 1: Comparison of the resulting number of input variables n_{p}^* / principal components n_{pc}^* (for PCR and PLSR) and accuracy (RMSE) of designed SISs on the training ($\mathcal{T}$) and testing ($\mathcal{S}$) industrial datasets.

	$\mathrm{SIS_{OLSR}}$	$\mathrm{SIS_{PCR}}$	$\mathrm{SIS_{LAS}}$	$\mathrm{SIS_{Ref}}$	$\mathrm{SIS_{SS}}$	$\mathrm{SIS_{NN}}$
$n_{\mathrm{p}}^*/n_{\mathrm{pc}}^*$	31/-	35/6	13/-	1/-	2/-	35/-
RMSE ($\mathcal{T}$)	0.077	0.109	0.083	0.111	0.104	0.077
RMSE ($\mathcal{S}$)	0.145	0.105	0.097	0.1	0.095	0.105

The resulting performance criteria of the studied single-model inferential sensors designed on the VGH dataset are shown in Table 1. The complexity of the inferential sensors is represented by the resulting number of input variables n_{p}^* and the number of principal components n_{pc}^* (for $\mathrm{SIS_{PCR}}$).

The results indicate that $\mathrm{SIS_{OLSR}}$ achieves the lowest RMSE ($\mathcal{T}$) value among all studied SISs, but it produces a relatively high RMSE ($\mathcal{S}$) value and a complex input structure ($n_{\mathrm{p}}^* = 31$), which suggests overfitting. The variance-covariance approach, $\mathrm{SIS_{PCR}}$, indicates also high input structure complexity ($n_{\mathrm{p}}^* = 35, n_{\mathrm{pc}}^* = 6$). Performance of $\mathrm{SIS_{PLSR}}$ is not included as it is practically the same as the performance of $\mathrm{SIS_{LAS}}$. $\mathrm{SIS_{LAS}}$ shows improved accuracy compared to the aforementioned two sensors and its input structure is less complex ($n_{\mathrm{p}}^* = 13$). $\mathrm{SIS_{Ref}}$ ($n_{\mathrm{p}}^* = 1$) and $\mathrm{SIS_{SS}}$ ($n_{\mathrm{p}}^* = 2$) have significantly lower complexity compared to other designed SISs. Based on these results, it is possible to improve the performance of $\mathrm{SIS_{Ref}}$ (currently implemented in the refinery) by about $5\,\%$ using $\mathrm{SIS_{SS}}$. This would only require to maintain one extra online sensor (compared to $\mathrm{SIS_{Ref}}$), to ensure the accuracy and reliability of $\mathrm{SIS_{SS}}$. Performance of $\mathrm{SIS_{NN}}$ is at the lower

Table 2: Comparison of the resulting number of input variables n_p^* and accuracy (RMSE) of designed MISs on the training ($\mathcal{T}$) and testing ($\mathcal{S}$) industrial datasets.

	$\mathrm{MIS_{SotA}}$		$\mathrm{MIS_{con}}$		$\mathrm{MIS_{con,lab}}$	
n_p^*	1 (Ref)	2 (SS)	1 (Ref)	2 (SS)	1 (Ref)	2 (SS)
RMSE ($\mathcal{T}$)	0.098	0.097	0.109	0.130	0.108	0.1
RMSE ($\mathcal{S}$)	0.094	0.092	0.113	0.133	0.105	0.098

end being similar to $\mathrm{SIS_{PCR}}$. This is attributed to the lack of training data and to high level of noise, as indicated in the analysis done for the previous case study.

Table 2 presents a comparison of the performance of MISs ($\mathrm{MIS_{SotA}}$, $\mathrm{MIS_{con}}$, and $\mathrm{MIS_{con,lab}}$) designed for the VGH dataset. Each of these approaches shows the resulting quality of the designed inferential sensors using both input structures (Ref and SS), and their accuracy is evaluated according to the same criteria as in the case of SISs (Table 1). This enables a direct comparison of the results from the studied SIS and MIS approaches.

Table 2 suggests that $\mathrm{MIS_{SotA}}$ achieves the highest accuracy (RMSE ($\mathcal{S}$)) compared to other designed MISs and SISs, taking into account both input structures. These results confirm the excellent accuracy of the $\mathrm{MIS_{SotA}}$ approach on the PCT datasets (Section 4.2). Moreover, the accuracy of $\mathrm{MIS_{SotA}}$ is higher with the SS input structure than with the Ref input structure, suggesting that the SS input structure is more appropriate for the MIS design. The MIS with the Ref input outperforms the currently implemented inferential sensor in the refinery (see $\mathrm{SIS_{Ref}}$ in Table 1) by about $6\,\%$ and with the SS input structure by about $8\,\%$.

The accuracy of MIS_{con} appears worse compared to MIS_{SotA} for both input structures. This aligns with observations made on the *PCT* datasets (Section 4.2). The poor accuracy of MIS_{con} is primarily caused by the requirement to design continuous models. The table suggests that the accuracy of MIS_{con} is decreased with the SS input structure compared to its performance with the Ref input structure. The additional variable within the SS input structure appears to be unhelpful for MIS_{con} accuracy, and it further increases the negative impact of the model continuity constraint on the MIS_{con} performance.

The results further indicate that the optimized data labeling within the $\text{MIS}_{\text{con,lab}}$ approach significantly improves its accuracy, considering both input structures. The accuracy of $\text{MIS}_{\text{con,lab}}$ is not as high as that of MIS_{SotA} but is comparable to that of SIS_{Ref} and SIS_{SS} in Table 1. Moreover, the $\text{MIS}_{\text{con,lab}}$ approach ensures continuity when switching between the designed models, which can be crucial in specific applications, particularly if the inferential sensor is part of a process control strategy. $\text{MIS}_{\text{con,lab}}$ achieves higher accuracy with the SS input structure than with the Ref input structure. Unlike the MIS_{con} approach, the optimized data labeling enables the $\text{MIS}_{\text{con,lab}}$ approach to effectively use the additional variable within the SS input structure with respect to the resulting accuracy of $\text{MIS}_{\text{con,lab}}$.

The designed MISs considering the SS input structure ($n_{\text{p}}^* = 2$) are illustrated in Figure 9. The yellow and dark green surfaces represent the designed models within the MIS structure, and the gray vertical surface is the separation hyperplane. The circles represent the training dataset, and the triangles form the testing dataset. The pink color of circles or triangles represents

40
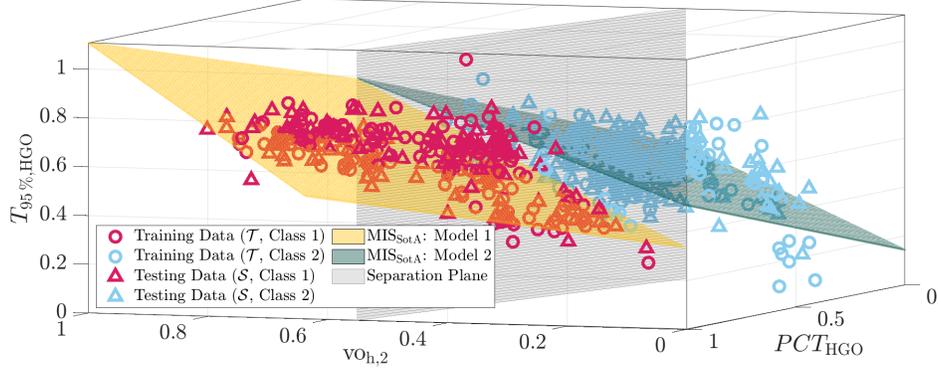
(a) MIS_{SotA}: model 1 (RMSE $= [0.091(\mathcal{T}), 0.091(\mathcal{S})]$), model 2 (RMSE $= [0.102(\mathcal{T}), 0.094(\mathcal{S})]$).
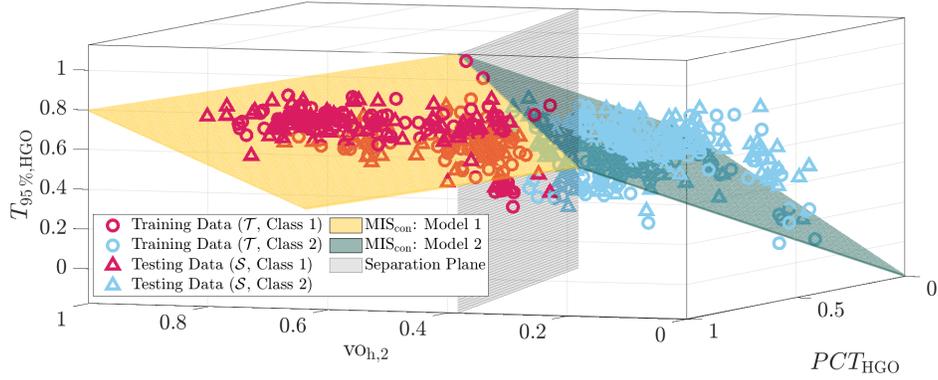
(b) MIS_{con}: model 1 (RMSE $= [0.125(\mathcal{T}), 0.133(\mathcal{S})]$), model 2 (RMSE $= [0.136(\mathcal{T}), 0.134(\mathcal{S})]$).
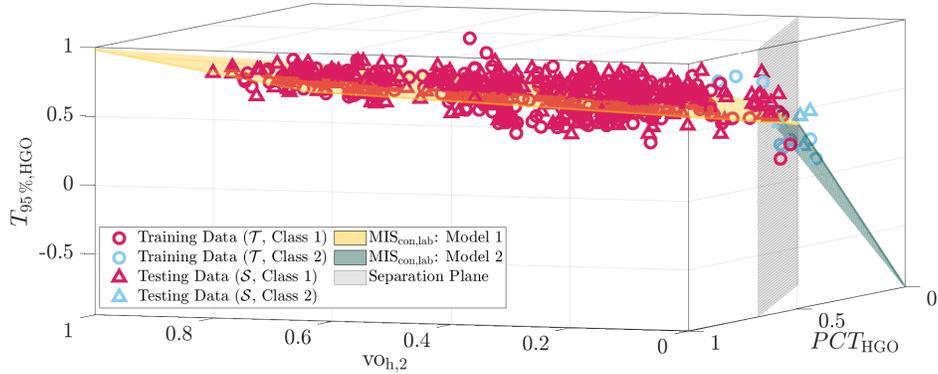
(c) $\text{MIS}_{\text{con,lab}}$: model 1 (RMSE $= [0.099(\mathcal{T}), 0.095(\mathcal{S})]$), model 2 (RMSE $= [0.13(\mathcal{T}), 0.211(\mathcal{S})]$).

Figure 9: The comparison of designed MISs on the industrial dataset from the VGH unit.

measurements from the first class, and the blue color of circles or triangles indicates measurements from the second class.

The models of MIS_{SotA} are shown in Figure 9(a). The model surfaces exhibit an obvious discontinuity at the interface (separation plane) between models. The discontinuity appears to be beneficial for model accuracy. The MIS_{con} models are shown in Figure 9(b). We can see that the designed models are are continuous, which does not allow the arbitrary rotation of the MIS_{con} model surfaces. Therefore, the models of MIS_{con} deviate more from the measurements more than the models of MIS_{SotA}, resulting in lower accuracy of the MIS_{con} model compared to MIS_{SotA}.

The designed models of $MIS_{con,lab}$ are shown in Figure 9(c). We can observe the continuity at the switch between the $MIS_{con,lab}$ models. We can also see that one data class involves the majority of the measurements, while the remaining points are assigned to another, smaller class. This occurs when there are no discernible classes of measurements in the provided dataset (as seen in Figures 4(b) and 9(c)), although this is not always the case (as seen in Figures 4(a)). In the case of indistinct classes of measurements in the available dataset, $MIS_{con,lab}$ attempts to improve the accuracy of the model designed on the majority of measurements by assigning the most deviated measurements to the smaller class. As shown in Figure 4(b), the measurements from the smaller class have the potential to explain the nonlinear nature of the estimated variable. The results from $MIS_{con,lab}$ indicate that the studied industrial unit is operated primarily an operating range, where one model suffices to capture the behavior of the inferred output. The second model, on the other hand, involves measurements from other operating

regimes, which occurs only occasionally in the plant operation. This observation leads us to conclude that the proposed $\mathrm{MIS_{con,lab}}$ approach effectively captures the different operating conditions of the industrial unit.

5. Discussion

Multi-model linear inferential sensors offer several advantages over single-model (linear or nonlinear) sensors. They can enhance the model performance over the linear SISs without compromising on the model complexity (and implied usability) as nonlinear SISs do. As for any other data-based model, there is a risk of overfitting, which is connected to the number of chosen models within the MIS structure. In this contribution, we used the simplest MIS structure with two concurrent models involved. The presented methodology for the MIS design can be easily extended to involve more than two models. We have already designed MIS considering more than two models for several simulation cases. It seems that the increasing number of designed models can further increase the accuracy of MIS. On the other hand, this type of MIS has higher requirements for the quantity and quality of data compared to the MIS with two models. A practical strategy to decide on the number of models might use cross-validation as commonly exploited in overfit mitigation for SISs.

Additionally, MIS has the potential to reduce the number of input variables and thus simplify the sensor structure. The MIS design on the industrial dataset (see in Section 4.3) considered only the reference input structure given by (14) and the enhanced input structure determined by the subset selection approach exploring all possible structures with one or two input variables.

In the case of the VGH unit, a simple input structure seems to be desired, which is confirmed by the excellent performance of SIS_{Ref} considering only one input variable. Another possible approach to finding a suitable input structure for MIS is to combine cross-validation with some feature selection approach taking into account the objectives of the MIS design. An advanced alternative to determine the MIS input structure would involve extension of the objective function of MIS by an appropriate penalization element, whose purpose would be to reduce the absolute value of the model parameters. Such penalization is considered, e.g., in the LASSO (1-norm) or ridge (2-norm) regression approaches. The enhanced form of the MIS design would directly provide the optimal input structure for the designed models and would allow for a variation of input set among different individual models within the MIS. Establishment of continuity within such MIS is not straightforward and should be explored in the future works.

The presented case studies of the inferential sensor design provide several important insights about how to choose the appropriate MIS approach in a particular situation. The design of MIS_{SotA} should be performed if the studied process requires an inferential sensor with high accuracy, reliable knowledge about different operating regimes is provided, and discontinuous in switching between the models cannot cause any (e.g., stability) issues within the considered process. If all previous specifications remain the same but continuity of the designed models is necessary, then MIS_{con} should be considered. The results indicated that the continuity of the models is achieved at the expense of the inferential sensor accuracy. In cases when reliable knowledge about different operating regimes within the process is not provided, then

$MIS_{con,lab}$ represents the best option. In the case that the previous specification remains but the continuity when switching is not necessary, it is possible to solve problem (12) with a relaxation of continuity constraints (12e).

6. Conclusions

In this paper, novel approaches for multi-model inferential sensor (MIS) design were introduced. These approaches provide (a) continuous switching between models of the inferential sensor and (b) optimized data labeling of the training dataset. The performance of the studied MISs was compared against the performance of the single-model inferential sensors (SIS) on the datasets from two case studies from the petrochemical industry: (1) the model of pressure compensated temperature and (2) the Vacuum Gasoil Hydrogenation unit.

The statistical evaluation of the results from the design of the inferential sensor for the model of pressure compensated temperature shows that MISs outperform SISs, even the nonlinear ones. The results indicate that the performance of MISs is highly affected by the information content within the training dataset. If the data within the training dataset represent different operating regimes than the testing data, then MISs achieved similar performance compared to SIS. Moreover, the analysis of the impact of the output variable noise on the performance of the inferential sensors indicated that MIS achieved higher accuracy than SIS over the entire studied set of noise variances.

The design of inferential sensors on the industrial dataset from the Vacuum Gasoil Hydrogenation unit showed that the MIS design can provide the

45

inferential sensor with higher accuracy compared to the SIS design. It seems that MIS can outperform the accuracy of the currently implemented reference SIS by about 6 % considering the same input structure or by about 8 % considering enhanced input structure.

Acknowledgments

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.

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers, in: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, ACM, New York, USA. pp. 144–152. doi:10.1145/130385.130401.

Botha, S., Craig, I.K., 2021. An industrial implementation of a c4 hydrocarbon soft sensor to optimise a debutaniser column. IFAC-PapersOnLine 54,

180–185. doi:https://doi.org/10.1016/j.ifacol.2021.12.031. control Conference Africa CCA 2021.

Doraiswami, R., Cheded, L., 2014. Robust model-based soft sensor: Design and application. IFAC Proceedings Volumes 47, 5491–5496. doi:https://doi.org/10.3182/20140824-6-ZA-1003.00245. 19th IFAC World Congress.

Ferrari-Trecate, G., Muselli, M., 2002. A new learning method for piecewise linear regression, in: Dorronsoro, J.R. (Ed.), Artificial Neural Networks — ICANN 2002, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 444–449.

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

Ge, Z., 2017. Review on data-driven modeling and monitoring for plant-wide industrial processes. Chemometrics and Intelligent Laboratory Systems 171, 16–25. doi:https://doi.org/10.1016/j.chemolab.2017.09.021.

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

Gurobi Optimization, LLC, 2023. Gurobi Optimizer Reference Manual. URL: https://www.gurobi.com.

Jin, H., Chen, X., Yang, J., Zhang, H., Wang, L., Wu, L., 2015. Multi-model adaptive soft sensor modeling method using local learning and online support vector regression for nonlinear time-variant batch processes. Chemical Engineering Science 131, 282–303. doi:https://doi.org/10.1016/j.ces.2015.03.038.

Kamath, C., 2022. Intelligent sampling for surrogate modeling, hyperparameter optimization, and data analysis. Machine Learning with Applications 9, 100373. doi:https://doi.org/10.1016/j.mlwa.2022.100373.

Kendall, M., 1957. A Course in Multivariate Analysis. Griffin's statistical monographs & courses, Hafner Publishing Company.

Khatibisepehr, S., Huang, B., Khare, S., 2013. Design of inferential sensors in the process industry: A review of bayesian methods. Journal of Process Control 23, 1575–1596. doi:https://doi.org/10.1016/j.jprocont.2013.05.007.

Khatibisepehr, S., Huang, B., Xu, F., Espejo, A., 2012. A bayesian approach to design of adaptive multi-model inferential sensors with application in oil sand industry. Journal of Process Control 22, 1913–1929. doi:https://doi.org/10.1016/j.jprocont.2012.09.006.

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

Konno, H., Yamamoto, R., 2009. Choosing the best set of variables in regression analysis using integer programming. Journal of Global Optimization 44, 273–282.

Löfberg, J., 2004. Yalmip: A toolbox for modeling and optimization in MATLAB, in: Proceedings of the CACSD Conference, Taipei, Taiwan.

Lü, Y., 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, 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, Proceedings. Lecture Notes in Computer Science, Springer.

Miyashiro, R., Takano, Y., 2015. Mixed integer second-order cone programming formulations for variable selection in linear regression. European Journal of Operational Research 247, 721–731.

Mojto, M., Ľubušký, K., Fikar, M., Paulen, R., 2021. Data-based design of inferential sensors for petrochemical industry. Computers & Chemical Engineering 153, 107437. doi:https://doi.org/10.1016/j.compchemeng.2021.107437.

Mojto, M., Ľubušký, K., Fikar, M., Paulen, R., 2022. Support vector machine-based design of multi-model inferential sensors, in: Montastruc, L., Negny, S. (Eds.), 32nd European Symposium on Computer Aided Process Engineering. Elsevier. volume 51 of *Computer Aided Chemical Engineering*, pp. 1045–1050. doi:https://doi.org/10.1016/B978-0-323-95879-0.50175-2.

Pan, H., Wu, X., Qiu, J., He, G., Ling, H., 2019. Pressure compensated temperature control of kaibel divided-wall column. Chemical Engineering Science 203, 321–332. doi:https://doi.org/10.1016/j.ces.2019.03.061.

Santander, O., Kuppuraj, V., Harrison, C.A., Baldea, M., 2022. An open source fluid catalytic cracker - fractionator model to support the de-

velopment and benchmarking of process control, machine learning and operation strategies. Computers & Chemical Engineering 164, 107900. doi:`https://doi.org/10.1016/j.compchemeng.2022.107900`.

Santosa, F., Symes, W.W., 1986. Linear inversion of band-limited reflection seismograms. SIAM Journal on Scientific and Statistical Computing 7, 1307–1330.

Song, M., Breneman, C.M., Bi, J., Sukumar, N., Bennett, K.P., Cramer, S., Tugcu, N., 2002. Prediction of protein retention times in anion-exchange chromatography systems using support vector regression. Journal of Chemical Information and Computer Sciences 42, 1347–1357. doi:`10.1021/ci025580t`.

Sun, W., Braatz, R.D., 2021. Smart process analytics for predictive modeling. Computers & Chemical Engineering 144, 107134. doi:`https://doi.org/10.1016/j.compchemeng.2020.107134`.

Sun, Y.N., Qin, W., Hu, J.H., Xu, H.W., Sun, P.Z., 2022. A causal model-inspired automatic feature-selection method for developing data-driven soft sensors in complex industrial processes. Engineering doi:`https://doi.org/10.1016/j.eng.2022.06.019`.

Tahir, F., Islam, M.T., Mack, J., Robertson, J., Lovett, D., 2019. Process monitoring and fault detection on a hot-melt extrusion process using in-line raman spectroscopy and a hybrid soft sensor. Computers & Chemical Engineering 125, 400–414. doi:`https://doi.org/10.1016/j.compchemeng.2019.03.019`.

Takano, Y., Miyashiro, R., 2020. Best subset selection via cross-validation criterion. TOP 28, 475–488.

Tibshirani, R., 2011. Regression shrinkage and selection via the lasso: a retrospective. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73, 273–282. doi:`10.1111/j.1467-9868.2011.00771.x`.

Wang, B., Wang, X., He, M., Zhu, X., 2021. Study on multi-model soft sensor modeling method and its model optimization for the fermentation process of pichia pastoris. Sensors doi:`10.3390/s21227635`.

Wold, S., Ruhe, A., Wold, H., Dunn, III, W.J., 1984. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. SIAM Journal on Scientific and Statistical Computing 5, 735–743.

Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58, 109–130. doi:`10.1016/S0169-7439(01)00155-1`.

Zhuang, Y., Liu, Y., Ahmed, A., Zhong, Z., del Rio Chanona, E.A., Hale, C.P., Mercangöz, M., 2022. A hybrid data-driven and mechanistic model soft sensor for estimating co2 concentrations for a carbon capture pilot plant. Computers in Industry 143, 103747. doi:`https://doi.org/10.1016/j.compind.2022.103747`.