TOPICAL REVIEW • OPEN ACCESS

Soft metrology based on machine learning: a review

To cite this article: Marcela Vallejo et al 2020 Meas. Sci. Technol. 31 032001

View the article online for updates and enhancements.
Topical Review

Soft metrology based on machine learning: a review

Marcela Vallejo¹©, Carolina de la Espriella², Juliana Gómez-Santamaría², Andrés Felipe Ramírez-Barrera¹© and Edilson Delgado-Trejos²©

¹ Department of Electronics and Telecommunications, Instituto Tecnológico Metropolitano ITM, Medellín, Colombia
² AMYSOD Lab, Parque i, Instituto Tecnológico Metropolitano ITM, Medellín, Colombia

E-mail: marcelavallejo@itm.edu.co, carolinadelaesprilla@itm.edu.co, julianagomez@itm.edu.co, andresramirez@itm.edu.co and edilsondelgado@itm.edu.co

Received 6 September 2018, revised 26 August 2019
Accepted for publication 4 October 2019
Published 2 December 2019

Abstract

Soft metrology has been defined as a set of measurement techniques and models that allow the objective quantification of properties usually determined by human perception such as smell, sound or taste. The development of a soft metrology system requires the measurement of physical parameters and the construction of a model to correlate them with the variables that need to be quantified. This paper presents a review of indirect measurement with the aim of understanding the state of development in this area, as well as the current challenges and opportunities; and proposes to gather all the different designations under the term soft metrology, broadening its definition. For this purpose, the literature on indirect measurement techniques and systems has been reviewed, encompassing recent as well as a few older key documents to present a time line of development and map out application contexts and designations. As machine learning techniques have been extensively used in indirect measurement strategies, this review highlights them, and also makes an effort to describe the state of the art regarding the determination of uncertainty. This study does not delve into developments and applications for human and social sciences, although the proposed definition considers the use that this term has had in these areas.

Keywords: soft metrology, soft sensor, virtual sensor, virtual metrology, machine learning, uncertainty analysis

1. Introduction

The purpose of a sensor within a process is to determine key variables needed for monitoring and control procedures. However, there is a wide variety of industrial or perceptual (physiological, psychological and operational) processes in which the important variables are difficult or expensive to measure, making the on-line determination of the value unfeasible. In this sense, technologies and strategies have been developed to use the available information of other variables in the process and computational models infer the value of those that cannot be directly measured [1–5]. These strategies have been used in several contexts, such as oil refinery [6], chemical industry [7], air conditioning systems [8], automobiles [9], semiconductor manufacturing [10], and the quantification of properties determined by human perception (such as smell, taste, color, etc), among others [11–17]. Depending on the context, the type of indirect measurement strategy addressed in this paper can be referred to as soft sensor, virtual sensor, estimator, inferential sensor, virtual metrology,
soft metrology, among others. In order to avoid confusion, in this paper we will use the term soft metrology regardless of the application context.

An important element in the performance of a soft metrology system is the selection of the model used to infer the values and for this many different approaches have been proposed [1]. The ability of machine learning to extract information and create models from an existing database makes it an ideal approach for soft metrology applications and for this reason concepts regarding machine learning routines have been included in this review. Despite existing advances and the success of indirect measurement strategies, there is not much information in the literature concerning the estimation of the error and uncertainty of measurement in these systems from a metrological point of view [8, 18, 19]. This is an important challenge that should be faced in order to promote a more extensive use of soft metrology in real world applications.

This paper presents a review on soft metrology systems based on machine learning techniques with the aim of defining current tendencies, challenges, needs and opportunities in the development of this type of measurement systems. This review is organized as follows: section 2 presents a general overview of the implementation of indirect measurement strategies in different contexts of application, section 3 discusses general considerations for the implementation of soft metrology and machine learning algorithms that are used to build a measurement model, section 4 presents a recollection of the methods that have been proposed in reference to the uncertainty analysis of such measurement systems, and section 5 presents conclusions and closing remarks.

2. Soft metrology: concepts and applications

Since the first appearance in the literature of the term soft metrology in 2003 only a few systems have been proposed using this denomination [20–22]. However, in different contexts, similar indirect measurement techniques have been developed since the 1970s using other terms to designate them such as soft sensors, virtual sensors and virtual metrology, among others [16, 23–25]. Regardless of the term used, in all these techniques the value of a variable that is difficult or expensive to measure is generally inferred using information of other variables from the same process that are easier to measure. This type of indirect measurement system has become popular in contexts such as the chemical industry, where similar development methodologies and methods are used. Among the approaches found in the literature, machine learning techniques are a very popular choice for generating models that correlates the measured and inferred variables. Nevertheless, despite the widespread use of such systems, only a few cases discuss the determination of uncertainty of the estimated variables [8, 18, 19], and there is still a need for standardized prediction quality metrics for this type of measurement systems [26]. Thus, indirect measurement strategies are based on the idea that the important variables in a process that are difficult or expensive to measure online have a functional relationship with other variables that are more feasible to measure and, therefore, the available process data can be used with a computational model to infer the real value [1]. Figure 1 illustrates how the concept of soft metrology has been designated in the literature using different terms, and some of the contexts of application where each term has been used. Consequently, it would be suitable to redefine this term as wider concept that can be applied in all these fields or contexts.

Since the early 1970s, the idea of inferring the value of difficult to measure variables, specifically with the aim of controlling them, was studied under the name of inferential control [23]. This concept was popular in the area of chemical processes, where many variables that describe the quality of the final product are difficult to measure, have important delays or can only be measured at a low frequency [27, 28]. This continues to be an important trend in processes control [6, 11, 29].

There are two approaches in inferential control: the direct approach, in which an estimator is constructed to infer the target variable from secondary variables and this estimator is integrated to a traditional feedback controller; and the indirect approach, in which the model is designed for the overall process without discriminating the measurement stage [28].

Another term that has been used to refer to indirect measurement devices is virtual sensor [16]. This term originated in the context of virtual instruments (software implementations that perform the functions of an instrument in a process) and it has been used in reference to indirect flow measurement [8, 17, 18] and other applications like control systems, automobiles and buildings [13]. However, this term has also been used with a different connotation, to denote the information generated by a computer program that replaces real world values in a simulation [30].

In the 1990s the term soft sensor became popular [24] and nowadays is widely used, especially in the chemical industry [31, 32] and in a wide variety of applications, such as the manufacturing of bio-therapeutics [33], the oil refinery industry [2, 34, 35], in circulating fluidized beds [35], to predict the melt index of thermoplastic polymers [12], ammonia synthesis process [7, 34], among others. Soft sensors are currently a technology that is implemented in real life processes. In 2010, Kano and Ogawa presented the results of a survey made among control engineers from petrochemical, chemical, engineering, petroleum refinery and other companies in Japan. The results show a wide use of soft sensors in application such as distillation, reaction and polymerization, among others [36, 37].

In semiconductor manufacturing these systems have been denoted with the name of virtual metrology [25], and many developments have been made in this area. The typical variables that virtual metrology instruments infer are thickness uniformity for chemical vapor deposition, etch depth for etching, and post-chemical mechanical planarization surface thickness [3, 26, 38, 39].

In 2014, an industry association in the supply chain for manufacturing semiconductors, photovoltaic panels, high-brightness LEDs, micro-electro mechanical systems (MEMS) and other micro and nano-technologies (called SEMI) included the term virtual metrology in the standard SEMI
E1333, which contains specifications for automated process control systems interface. In this document, virtual metrology is defined as a technology for predicting post process metrology variables using process and wafer state information [40].

In 2017, the Institute of Electrical and Electronics Engineers (IEEE) published in the International Roadmap for Devices and Systems (IRDS) a white paper on virtual metrology describing the benefits that its implementation can have in the semiconductor industry, and also identified the main challenges faced in the broader implementation of this technology. One of the main conclusions is the need to develop a standardized prediction quality metric for these systems [26].

The term soft metrology appeared for the first time in a report to the National Measurement System Directorate published in 2003 by the National Physical Laboratory (NPL). This document defines soft metrology as ‘measurement techniques and models which enable the objective quantification of properties which are determined by human perception. The human response may be in any of the five senses: sight, smell, sound, taste and touch’ [41]. In 2016, the book ‘New Trends and Developments in Metrology’ included a chapter about soft metrology, in which the author redefines this term as ‘a set of models and techniques that allow subjective measurements intended as the process of experimentally obtaining one or more quantity values that can be attributed to human physiological, operational and psychological quantity—as defined by soft metrology—directly involving one human being’ [20]. This document also states that ‘soft metrology requires the measurement of proper physical parameters and the development of models to correlate them to perceptual quantities. Traceable soft metrology is achievable both through the traceable measurement of the physical parameters and the development of accurate correlation models’ [20]. Thereby, the concept of soft metrology takes the idea of using indirect measurement strategies to infer the value of difficult to measure variables, and expands it to include properties determined

Figure 1. Soft metrology: different terms used and main application contexts.
by human perception such as color, gloss, noise, odor and taste.

There are a few publications in the literature using the term soft metrology, examples of applications are the determination of surface gloss [42], the influence of noise in human performance [21] and the measurement of the frictional component of touch [22]. However, there are other systems that have been developed to infer the value of variables related to human senses but that have not used this term. In 2004, the European Commission developed a program called ‘measuring the impossible’ to fund projects related to the measurement of human perceptions [43]. One of the financed projects was SYSPAQ (Innovative Sensor System for Measuring Perceived Air Quality and Brand Specific Odours), which had the objective of developing an innovative sensor system to measure indoor air quality as it is perceived by humans based on perception modelling [44]. Another example of systems that correlate measured variables with human perception are electronics noses, which propose a relation between the sensorial response of the olfactory organ and measurements from gas sensors and other variables like temperature and humidity [45]. This type of system is commercially available and used in contexts such as wastewater treatment plants [45], the food industry [46, 47] and the detection of air contaminants [48], among others. Another sense for which similar systems have been proposed is taste. Electronic tongues have also been developed since the 1980s [49] and, although they were initially intended to distinguish five basic tastes, they have evolved and have become an important tool in food and drug analysis [49–51]. The term soft metrology has also been used in the context of studies that investigate the problem of measurability, analyzing the problem of what can and cannot be measured [52, 53].

Regardless of the specific application context and designation used, all the indirect measurement strategies mentioned above have a common goal, similar methods and challenges in their implementation. In this sense, they can all be considered as a part of a single measurement strategy and gathered under a common term: soft metrology.

2.1. Wider definition of soft metrology

In this paper, we gather all these concepts and terms, and establish a wider definition of the term soft metrology, as follows:

A set of models and techniques that allows the objective quantification of magnitudes that are subjective, difficult or expensive to measure, such as those related to human perception and/or to process dynamics. This quantification is intended as the process of experimentally obtaining one or more quantity values that can be attributed to the variable that needs to be measured by building a model that correlates it with other quantities, such as available measured variables from the same process, or physiological, operational and psychological response in the case of human perception.

3. Machine learning for soft metrology: applied methods

There are two main approaches used to develop a soft metrology system: model-driven approaches, which develop a model using phenomenological knowledge of the process [17, 54–56], and data-driven approaches, that use observed information of the process (a database) to develop the model [32]. Due to its flexibility, and the possibility of creating a model without precise knowledge of the process dynamics, the second approach is widely used [35, 57, 58]. However, soft metrology systems derived from historical process data have various challenges to face in order to construct a good measurement model [59]. Some of the main challenges are:

- Small and/or incomplete datasets: the construction of a data-driven measurement model usually requires a labeled dataset that contains information about the measured and the inferred variables, but the available information about the inferred variable is usually strongly-constrained (precisely because it is a difficult-to-measure variable). For this reason, available datasets are typically small or have a large proportion of unlabeled data points.
- High dimensional data: generally, there are numerous measurable variables in the process, of which several statistics can be derived from. Nevertheless, many of these available variables can be irrelevant for the inference of the target variable or those variables can have a high-correlation among them.
- Process nonlinearity: Most real-life processes are nonlinear and, thus, the use of simple linear models can result in a poor performance.
- Process delays and dynamics: Many real-life processes exhibit important delays and dynamics in their behavior.
As shown in figure 2, the development process of data-driven soft metrology systems can be summarized in 4 stages: (1) database construction and pre-processing, (2) generation of an effective representation space, (3) model choice, training and validation, and (4) model maintenance. In the following subsections, the purpose of each one of these stages will be explained, along with the main methods used and some of the strategies implemented to face the challenges mentioned above. Also for each stage the way forward in the area and future research directions have been stated.

3.1. Database construction and pre-processing

In this stage, information from process variables is collected to build a database that will be used to train and validate the soft metrology model. This part of the process also includes processing with missing data and outliers (due to sensor failure or malfunction) and data normalization [32, 57, 60, 61]. In some cases, statistics of the measured variables (like the mean, median, maximum, minimum, range, standard deviation, integral, differential and count) can be calculated in order to be used as system features [62, 63].

The database structure depends on the type of learning algorithm used in the following development stages. If a supervised learning approach is going to be used, each instance in this database must contain information about the values of both: the measured variables and variable that is going to be inferred. However, it is not easy to obtain these labeled points, and for this reason some authors propose semi-supervised approaches [12, 57, 64, 65] where the unlabeled information in the database can be used to extract useful information. On the other hand, some authors propose strategies to reduce labeling costs while constructing the database, such as active learning (AL) [35].

Specifically, the value of a database depends on its size and how well the data can be comprehended [66]. In this sense, when facing the problem of a small sample size dataset it is necessary to rely on the generalization capabilities of the technique used in the construction of the model (such as in the case of artificial neural networks (ANN) or Embedding Techniques [67]) and to use techniques based on the processing of the training data set such as noise injection and bootstrap resampling [68]. However, regardless of the strategy used, the size of the database is one of the factors that constrains the quality of the model that will be constructed in the following phases, as it establishes a lower limit to the model representation quality and the system training will determine the success in the inference process. On the other hand, the dataset comprehension influences the representation quality in terms of the intrinsic structure of the data that the phenomenological dynamics provides. That is, the capability that the metrics in the system have to represent the dynamics in the process that is being studied. This comprehension establishes an upper limit to the representation quality, obtaining a quality range between sample size (lower limit) and data comprehension (upper limit), as depicted in figure 3 [66]. Therefore, an appropriate quality range in terms of representation makes the construction of the database a critical issue in the development of a soft metrology system.

The schematic diagram in figure 3 indicates trends derived from equation (1), which is extracted and adapted from [69], where the development of training routines requires the selection of the most relevant features for achieving a more accurate representation result for inference or classification purposes. Thus, representation quality ($Q_R$) can be understood as:

$$Q_R = n_f \ln \left( N_s \right) \left[ \frac{|R_e \cap R_o|}{1 + |R_o|} \right]. \quad (1)$$

Where, $n_f$ is the number of representation features, $N_s$ is the number of samples in the database, $R_e$ is the real representation given by the natural dynamics of the study objects, $R_o$ is the effective representation after feature selection or extraction methods and $R_o$ is the representation noise. It is important to note that the two components of equation (1) are related to the two rigorous constraints for representation quality, mentioned above. That is, the first is related to the sample size and has a logarithmic behavior directly proportional to $Q_R$. The second is related to data comprehension and the intuitive behavior is similar to a logistic curve, even though this continues to be an open issue in the Literature. In this stage there is an open area...
for future researches in the identification of a relation between equation (1) and an uncertainty model derived from the sample size and the data comprehension.

3.2. Construction of an effective representation space

Taking into account the challenges mentioned at the beginning of this section, before creating a measurement model that relates the sensor readings with the variable that needs to be inferred, such readings must be mapped into an effective representation space where the features do not necessarily have a direct physical meaning but, instead, they capture the majority of intrinsic information embedded in the sensor’s measurements [3]. This procedure not only generates a representation of the information that better characterizes the available information, but also helps to reduce the high dimensionality of the data, especially in processes where the large amount of sensors may lead to a dataset that contains redundant variables. This stage involves feature selection/extraction procedures [7, 12, 70]. It is also helpful to exclude the variables that do not contain useful information and, therefore, improve the prediction and inference performance [71].

This stage is also known as relevance analysis or dimensionality reduction and there are different approaches for obtaining an effective representation space [71, 72]: filter, wrapper, embedded and transformation. In the expressions of table 1, \( X \) is a representation space, \( Z \) is the corresponding effective representation space and \( Y \) is the effective representation in a new space obtained by data transformation.

- **Filter approaches**

Taking into account table 1, these approaches are based on ranking a variable according to its relevance in order to optimize a measure \( \mu \) from an initial dataset. A common method for determining the relevance of the variables in the system is to use a correlation coefficient, such as Pearson correlation coefficient, in order to use the modelling process variables that are highly correlated with the output but weakly correlated with each other [34]. This type of analysis can only detect linear dependencies [72, 73], but its low computational cost makes it a popular method. Another method that uses correlation criteria is nearest correlation spectral clustering based on variable selection (NCSC-VS) [37].

Another option is mutual information (MI), in which the relevance of a given variable is determined by calculating how much the uncertainty in the output of the system decreases by observing that given variable. In this case, Shannon entropy definition of entropy is used [74, 75].

Another method to rank variables is variable importance in the projection (VIP), a score that is calculated in a partial least squares (PLS) model to express the contribution that one input variable makes to the output variable [37, 62, 71, 76].

Also in the case of PLS modelling, another approach is PLS with regression coefficients (PLS-BETA), a method that uses the magnitude of the coefficients in a PLS regression to determine the relevance of the variable [37, 71]. Some authors also consider VIP and BETA between the wrapper approaches, since both methods are tuned to improve the prediction performance [37, 71].

For filter approaches the way forward in research includes the establishment of new metrics for the identification of relevant variables in a multivariate space and the mathematical and computational analysis of convexity, compactness and discriminnability associated to the estimation uncertainty for each technique.

- **Wrapper approaches**

According table 1, in these approaches the construction of the effective representation space is mixed with the modeling phase, because the inference error of the model computed from a target set \( k \) is used as a criterion in the variable selection procedure [59, 71]. A simple popular option between wrapper approach is to use sequential selection algorithms, which construct the representation space by iteratively adding or removing variables, according to a performance criterion. Sequential methods like stepwise regression (SR), sequential forward selection and sequential backward elimination have been used [71, 77-79].

Another option between the wrapper approaches are methods that favor simple models. An example of these is least absolute shrinkage selection operator (LASSO), a method that penalizes the complexity of the model, reducing to zero the coefficients of irrelevant variables [59]. Also some variations of LASSO have been proposed, such as fused least absolute shrinkage selection operator [80]; removing irrelevant variables amidst least absolute shrinkage selection operator iterations (RIVAL), an improved version of LASSO where all the coefficients are greater than zero [71]; and group LASSO, designed to select groups consisting of multiple variables [37].

The same penalization concept used in LASSO is applied in sparse partial least squares (SPLS), which tries to find a trade-off between the performance and the sparsity of the

---

**Table 1. Approaches for obtaining an effective representation space.**

| Approach     | Expression   | Observation |
|--------------|--------------|-------------|
| Filter       | \( Z = f \{ X, \mu \} \) | Where, \( \mu \) is a statistical, geometric, consistency or information measure associated to the objective function \( f \{ \cdot \} \) |
| Wrapper      | \( Z = f \{ X, k \} \) | Where, \( k \) is a target set with which an inference error must be minimized in the objective function \( f \{ \cdot \} \) |
| Embedded     | \( Z = f \{ X, k, \mu \} \) | Where, \( k \) is a target set with which some statistical, geometric, consistency or information measure \( \mu \) must be minimized as the inference error is also minimized |
| Transformation | \( Y = f \{ X \} \) | Here, \( f \{ \cdot \} \) is a transformation function where data in a new space are ordered by some criterion, achieving dimensionality reduction |
model [76]. Another wrapper method that has been used in soft metrology is uninformative variable elimination (UVE) is a method where the most informative variables are chosen according to their reliability index, which are compared to those of artificial random variables [71]. Finally, a popular option in soft metrology is the use of heuristic search strategies like genetic algorithms [1, 77], which select the variable combination that generates a model with an improved performance by heuristically exploring many combinations of the available variables [81].

For wrapper approaches, the open research questions include the relation between the rejection of irrelevant and redundant variables and the information of the classification robustness.

- **Embedded approaches**

Table 1 shows for these approaches that the generation of the effective representation space is performed optimizing both the measure $\mu$ and the inference error of the measurement model, simultaneously [71]. Therefore, there is no need for selection or extraction procedures. An example, and possibly the most commonly used of these approaches, are the ANN [81].

Deep learning has also been proposed as an alternative to avoid the need for a separated feature extraction stage, where the database can be used directly in the modeling phase. Despite the fact that deep learning approaches are proposed as a means to bypass the feature extraction procedure, sometimes it is assumed linear, let $X$ be a phenomenological dynamics $Y$ and can be understood as applied to a space $\beta$ with $\sim f(X, \beta, u_i) = \beta$ and a disturbance variable $u_i$. Likewise, a nonlinear regression is associated to a function $f$ applied to a space $X$ and can be understood as

$$y_i = X\beta + u_i, \quad i = 1, \ldots, q.$$ (2)

Which allow to optimally foresee the $y_i$ variables, by means of the regression coefficients $\beta$ and a disturbance variable $u_i$. Likewise, a nonlinear regression is associated to a function $f$ applied to a space $X$ and can be understood as

$$y_i \sim f(X, \beta, u_i), \quad i = 1, \ldots, q.$$ (3)

Although the regression models have been extensively studied, there is still an open research area regarding the study of the relation between the search space of the representative point of a data set for a specific time and the uncertainty associated to the previous estimations of the input data to the regression model.

- **Transformation approaches**

As expressed in table 1, for these approaches, a transformation is performed on the original variables to generate a new set of variables (new representation space), usually called latent features, where a transformation criterion is used for achieving dimensionality reduction. These types of methods are also called feature extraction. Probably the most popular method used in soft metrology for dimensionality reduction is PCA [70, 76, 82, 83]. Variations of this method have been proposed in order to cope with nonlinear processes, like weighted probabilistic principal component analysis (WPPCA) [70] and kernel principal component analysis (KPCA) [77]. Also, variations that aim for a probabilistic interpretation of the solution have been implemented, such as probabilistic principal component analysis (PPCA) [84]. Methods based on partial least squares (PLS) are used to extract latent variables [85], such as partial constrained least squares (PCLS) [86]. A different alternative for feature extraction is the use of deep learning-based methods, such as the work by Yao and Ge in which, before the use of extreme learning machines (ELM) for regression, autoencoders were used in unsupervised feature extraction [57]. A similar method is the use of stacked autoencoders (SAE), in which representative features are derived from input data with the objective of minimizing reconstruction errors [87]. Also, methods based on modal reconstruction have a place in feature extraction techniques [88].

The research opportunities for transformation approaches include the residual analysis between the lower dimensional space and the input space in relation to the uncertainty derived from the transformation for optimal or robust solutions.

- **Approach considerations**

Table 2 presents advantages and disadvantages of previous approaches.

### 3.3. Model choice, training and validation

In this stage, a learning algorithm is chosen to model the process dynamics. A subset of the database is used to train the model and the remaining data is used to validate the trained model and determine accuracy [89]. Methods like 10-fold cross validation are frequently used in the validation process [32].

There is not a precise or unique criterion to select the specific algorithm used in the model, but there are a series of considerations that can be have in to an account in this process. The first consideration is related to the nature of the value that needs to be inferred; that is, if the variable can be identified by a numerical value (such as temperature, flowrate, etc) or if it is more likely to be identified as a class, a state or a discrete value (such as the case of electronic tongues that can differentiate the geographical origin of a specific product). In the first case the modeling stage will consist on a regression problem, whereas in the second case it is a classification task.

#### 3.3.1. Regression models

The main objective in this case will be to infer a specific value of the output variable using the available features. Both, linear and non-linear regression methods are commonly used in this stage [90]. If the process is assumed linear, let $X$ be a phenomenological regression composed by $n$ instances of $p$ variables, where the regression problem consists of finding $q$ relations in the following way

$$y_i = X\beta_i + u_i, \quad i = 1, \ldots, q.$$ (2)

Which allow to optimally foresee the $y_i$ variables, by means of the regression coefficients $\beta_i$ and a disturbance variable $u_i$. Likewise, a nonlinear regression is associated to a function $f$ applied to a space $X$ and can be understood as

$$y_i \sim f(X, \beta_i) = \beta_i, \quad i = 1, \ldots, q.$$ (3)

Many proposed methods use simple linear regression approaches [70] such as multiple linear regression (MLR)
However, this method does not perform well with highly collinear data [63]. An alternative for dealing with collinearity is the use of principal component regression (PCR) [60, 92], where principal components (the covariance of the input data is maximized by PCA) are used in the regression process. A similar approach is the use of PLS [61, 63, 93–95], where both the outputs and process variables are decomposed in principal components, trying to find the representation that best explains the covariance between the inputs and the predicted variables. Also, some variations of PLS, such as dynamic partial least squares (DPLS) [60 76], Locally weighted partial least squares (LW-PLS) [96] and co-training PLS [64] have also been proposed.

Likewise, regularized regression methods have been used with collinear data. A typical example of regularized methods is ridge regression (RR) [59], which introduces on the minimization problem a regularization parameter that penalizes bias. This penalization aims for a tradeoff between complexity and prediction accuracy in order to reduce variance and avoid overfitting. Another example is LASSO [59, 78], which is similar to RR, but has a different penalization that allows the coefficients of irrelevant inputs to reach zero, thus favoring sparse solutions. For this reason, LASSO was also mentioned in the previous section as a wrapper approach in the construction of the effective representation space. Elastic nets have been proposed as an alternative that combine the benefits and overcome the drawbacks of LASSO and RR by combining the penalties in these two methods in its objective function [39].

A different approach is the use of Gaussian process regression (GPR), a probabilistic technique in which instead of assuming a specific model function, a Gaussian process (GP) in used in the representation on the possible relations between observed and inferred variables [78, 94]. Despite the popularity of linear regression models, as mentioned before, the nonlinearity of most processes poses a challenge that these methods cannot solve on their own. One possible solution is the use of nonlinear regression models or linear local models using neighbors of the query sample [70].

**Nonlinear regression**

Given the challenge derived from the nonlinearity of real-life processes, one of the well-known options for creating a soft

| Approach | Main strategy | Advantages | Disadvantages |
|----------|--------------|------------|---------------|
| Filter   | Variable ranking according to some relevance criterion | • Low complexity  
• Avoids overfitting  
• Discriminate the variables according to its correlation with other variables  
• Less informative features but informative combined with others could be included | • The variable selection is not associated with the performance of inference methods  
• The selected subset may not be ideal  
• It is not ideal method for choosing the dimension of the representation space |
| Wrapper  | Inference error as principal criterion in the objective function to evaluate the variable subset in the feature selection procedure | • The variable selection is associated with the performance of inference methods  
• The optimality of the feature subset is guaranteed | • High complexity  
• Problems associated to overfitting |
| Embedded | The effective representation space is performed at the same time as the training of the measurement model | • Reduces computational complexity compared with wrapper methods  
• Discriminate the variables according to its correlation with other variables  
• Less informative features but informative combined with others could be included  
• The variable selection is associated with the performance of inference methods | • Problems associated to overfitting  
• The selected subset may not be ideal  
• Problems associated to interpretability |
| Transformation | A feature extraction procedure is performed and, thus, a transformation of the original variables to obtain relevant features in a new space | • Dimensionality reduction in the new representation space while maintaining as much information/correlation as possible  
• Reduces computational complexity compared with wrapper methods  
• Avoids overfitting | • Problems associated to interpretability  
• The optimality of the feature subset is not guaranteed  
• The variable selection is not associated with the performance of inference methods |

Table 2. Advantages and disadvantages of dimensionality reduction approaches.
metrology model is K nearest neighbor (KNN) regression used in the construction of the measurement model, were the value of the of the output variable is calculated using the output values of similar labeled data that are selected according to a distance measure [77]. Likewise, ANN [1, 68, 92, 94] due to its generalization capability and ability to model nonlinear behaviors is frequently used when the feature extraction/selection stage needs to be bypassed. Some authors have proposed neuro-fuzzy approaches, such as adaptive neuro-fuzzy inference system (ANFIS), where some of the components of each neuron are replaced by fuzzy logic operations [6].

Support vector machines (SVM) and support vector regression (SVR) have also been extensively used when the nonlinear relationships in the process want to be captured in the model [97]. In these methods, quadratic programing is required and, for this reason, systems that need to operate online have proposed alternatives such as least squares support vector machine (LS-SVM) [98] and least squares support vector regression (LS-SVR) [82, 93], thus avoiding the complex optimization step. Other variations have been proposed, such as twin SVR [99], semi-supervised SVR [100], and combinations of models like finite impulse response (FIR) and SVM [58].

Another important non-linear approach used is deep learning, which has become a popular trend in process modeling due to the interest in developing models with a better representation performance. In this area, methods like extreme learning machine (ELM) [101, 102], hierarchical ELM (HELM) [57], regularized ELM (RELM) [103] and semi-supervised ELM (SELM) [104] have been proposed. Finally, deep learning has turned applications that previously required perceptual expertise into metrology challenges in order to promote new solutions by including soft metrology concepts [3, 105].

As mentioned above, regardless of the specific method chosen, the amount of data available for the model construction is an important factor that determines the quality of the representation obtained. An important strategy to improve the performance is the use of ensemble methods, where multiple predictors are generated using one or several modeling techniques and processing the training data with bootstrap resampling or noise injection. The different predictors obtained are then combined in into a single model either by averaging or by majority vote strategies [106, 107].

- **Technique considerations**

Table 3 reports principal advantages and disadvantages of linear and nonlinear regression techniques.

### 3.3.2. Classification models

In this case, the objective is to determine if the variable belongs to one of several predefined classes or states. Some examples of this type of modeling are electronic noses and tongues, were some characteristics of a product can be described in terms of classes (e.g. adulterated/unadulterated coffee, regular/export type coffee, identification of geographical origins of wine or honey [83, 91, 108, 109]), the analysis of bio-medical signals (e.g. healthy/unhealthy electromyography or electroencephalography signals [66, 110]) and the biometric sensors (e.g. fingerprint identification [111]).

The simplicity of linear discriminant analysis (LDA), given by equation (4), has made it a popular method in many soft metrology applications that require classification [91, 102].

\[ \varphi = w^T x. \] (4)

Where \( \varphi \) is the discriminant indicator variable and \( w \) is the boundary coefficient matrix for the instance \( x \). However, LDA assumes that the different classes have identical covariance matrices, and for this reason alternatives that do not make assumption on the covariance of the classes have been proposed, such as quadratic discriminant analysis (QDA), which uses quadratic decision boundaries [110]. However, these two methods assume a Gaussian distribution in the classes, and methods that do not make assumptions on the distribution of the data such as ANN (which is also applied in regression) have been used for classification in soft metrology systems [66, 109, 110, 112]. ANN uses a nonlinear function \( g \) for adjusting the boundary coefficient matrix \( w \), as follows

\[ \varphi = g(w^T x). \] (5)

Where \( x \) includes an additional variable, \( x_0 = 1 \), that provides a bias term \( w_0 \), included in the \( w \) matrix. Some other of the methods mentioned in the previous section for regression have also application or variation for classification purposes such as SVM [109, 111–114] and KNN [115]. Another method that has been implemented in soft metrology models is the use of decision trees, which have de advantage of being able to deal with categorical as well as a numerical values [116].

An important consideration that must be made when choosing a specific model for classification is the number of classes that describe the behavior of the analyzed variable, because several popular algorithms (such as SVM) are binary in nature, and if the structure of the information to be represented implies more than two classes, strategies such as one-versus-one and one-versus-all must be considered [111, 114]. Also an strategy must be used to combine the outputs of the different binary classifiers, such as majority voting or the-winner-takes-all method [111, 114]. Besides the need to deal with multiple classes by means of binary classifiers, the combination of several models is also used in classification to improve the performance by applying ensemble methods, in the same way mentioned above for regression methods and using similar strategies for processing the data set and combining the outputs. Also random forest have been used in the combination of several models in soft metrology applications that require classification [109, 112].

In order to present a comparison of the most common classification techniques in soft metrology applications, table 4 shows a set of computational advantages and disadvantages. For KNN/SVM/ANN techniques, considerations presented in table 3 can be applied.

For classification models, the way forward in research includes efforts to relate the uncertainty of the classifier with the uncertainty associated to feature selection and the embedded uncertainty in the perceptual representation or qualitative
| Method | Technique | Advantages | Disadvantages |
|--------|-----------|------------|---------------|
| Linear | MLR       | • Low complexity<br>• Residual variance reduction<br>• Good performance when variables independently and effectively impact the output | • Low performance with highly collinear data<br>• Sensitive to outliers |
|        | PCR       | • Low complexity<br>• Good performance with highly correlated variables or even collinear ones<br>• Usefulness where the number of descriptors exceeds the number of observations | • From output results, the original space interpretability could be very difficult<br>• Sensitive to which principal components should be discarded |
|        | PLS       | • Useful when there is multicollinearity in data<br>• Useful when the number of descriptors exceeds the number of observations<br>• Predictive relationships can be examined | • Interpretability problems<br>• Risk of omitting real correlations |
|        | RR        | • Use useful when there is multicollinearity in data<br>• Useful when the number of descriptors exceeds the number of observations<br>• Avoids overfitting | • The output is not unbiased<br>• The coefficient of irrelevant variables does not reach zero<br>• Outlying values in any input parameter can result in wrong predictions |
|        | LASSO     | • Computational efficiency<br>• Can eliminate irrelevant variables<br>• Preferred when the solution has sparse features<br>• Dimensionality reduction | • Low performance with highly correlated variables<br>• Low performance if the number of descriptors exceeds the number of observations |
|        | Elastic nets | • Combines the strengths of RR and LASSO<br>• Flexibility of the estimator | • Greater computational complexity than RR and LASSO<br>• Greater flexibility means greater probability of overfitting |
|        | GPR       | • Does not impose a specific structure on the underlying function<br>• Has a very neat way to tune hyper-parameters by maximizing the marginal likelihood<br>• Very intuitive | • High constraints associated to computational demand<br>• The use of relatively small datasets is imperative |
| Nonlinear | KNN       | • Robust to noisy training data<br>• Effective for large datasets | • Estimation of a proper $K$-parameter<br>• Finding the type of distance for obtaining effective results<br>• Need to compute distance of each query instance for all training samples |
|        | ANN/ANFIS | • Bypasses the feature selection/extraction procedure<br>• Good performance for highly nonlinear processes<br>• Capabilities of generalization and nonlinear mapping in large datasets<br>• Possibilities for probabilistic assignment | • Training is computationally demanding<br>• Latent probability of overfitting<br>• A lot of parameters to be adjusted<br>• Low performance if the number of descriptors exceeds the number of observations<br>• Can be very slow to predict |
|        | SVM/SVR   | • Computational complexity does not depend on the dimensionality of the input space<br>• Reach the global optimum due to quadratic programming<br>• Less memory to store the predictive model<br>• Less prone to overfitting<br>• High prediction accuracy | • Finding the right kernel and optimizing it may be very difficult<br>• Difficult to understand and interpret the final model<br>• Difficult to do small calibrations to the model taking advantage of expert knowledge<br>• It is not easy to fine-tune the hyper-parameters, visualizing their impact |

(Continued)
capture of features (measurements) for classes, states or discrete values (i.e. tastes, fragrances, etc).

3.3.3. Model validation. Once the model is trained different indexes can be used to determine its performance. The most commonly used are root mean square error (RMSE) [57–59, 64, 70], mean absolute error (MAE) [34], mean absolute percentage error (MAPE) [71], coefficient of determination ($R^2$) [34, 70, 71], normalized mean squared error (NMSE) [78], and consistency index [71]. This validation is usually performed with a proportion of the labeled data set that has not been used in the model training stage, comparing the labeled values with those inferred by the trained model. Only a few authors have considered the uncertainty of the type of measurement systems addressed in this paper [8, 18, 19].

### 3.4. Model maintenance

Many processes have a highly changeable nature, so there may be either abrupt or progressive changes in the process dynamics. For this reason the quality of a model developed for a process in a specific moment in time can degrade over time [32]. Frequently, authors propose a model maintenance procedure whenever new data is available, in order to adapt the soft sensor model. The objective is to prevent errors due to changes in the process dynamics over time [2, 117].

It must be considered that the data used in the model adaptation must be carefully processed to ensure that the model in not adapting to data disturbances [118]. Some of the approaches used to overcome model degradation are just-in-time learning (JITL), time difference, moving window, and recursive methods [119].

In a just-in-time modeling approach, the new information available is stored on a database, a global model is built off line and local models are constructed on line and discarded after use [62 93]. Some of the methods mentioned in the previous section have a just-in-time modeling version, such as LW-PLS [62] and just-in-time semi-supervised extreme learning machine (JSELM) [120].

The time difference modelling approach is based on the difference of an output variable in two moments in time ($\Delta y$) according to the difference of the input values for those time moments ($\Delta x$). This time difference representation can avoid process linear changes in time, which causes degradation in the accuracy model [121, 122].

The time-window modelling approach is also popular for coping with changes over time in the process characteristics. The main idea is to update the model over time, discarding older data points and including new ones. Many of the previously mentioned methods, such as PLS, ANN and GPR have been used in combination with a time window scheme [63, 94, 123], where a method based on PLS that uses a moving window to recursively update the model were developed.

In the recursive approach, the current model is updated using new data points. A forgetting factor is introduced to ensure that, in the adaptation process, older data samples have smaller weights and newer ones have larger weights. Popular modelling methods like PLS and PCA have recursive variations [118].

### 4. Uncertainty analysis in soft metrology

The uncertainty of a measurement is defined as a non-negative parameter that characterizes the dispersion of the values attributed to a measure, based on the information used, and depending on how this dispersion is characterized, including components of systematic effects, such as those associated with corrections and values assigned to a [124].

To estimate uncertainty, there are non-stochastic tools, such as the Guide to the Expression of Uncertainty in Measurement (GUM) [125], and stochastic tools, such as random-fuzzy [126], Monte Carlo method [127], Bayesian methods [128], and generalized interval [129], among others.

As mentioned, the GUM methodology is widely used. However, the mathematical models in some systems are complex and applying the GUM methodology is complicated; for this reason, stochastic and non-stochastic methods are
was acceptable and that it provides a low cost alternative showing that the uncertainty achieved by this virtual flow calculating the uncertainty of virtual sensors for packaged air was presented, using an ultrasonic flow meter for comparison, to a database. An experimental validation of this virtual sensor order to estimate the total uncertainty. As it is a model-driven approach. The total uncertainty in the calculated water flow considers the uncertainties in the valve command, the differential pressure sensor and the uncertainty associated with the regression. After identifying the sources of uncertainty, the law of propagation of uncertainty was applied in order to estimate the total uncertainty. As it is a model-driven approach, it does not have to account for uncertainties related to a database. An experimental validation of this virtual sensor was presented, using an ultrasonic flow meter for comparison, showing that the uncertainty achieved by this virtual flow meter was acceptable and that it provides a low cost alternative for air-handling units.

In a similar work, Cheung & Braun presented a method for calculating the uncertainty of virtual sensors for packaged air conditioners [8], using a data-driven approach. They reported that both the measurement (use of physical sensors) and inference (use of analytical methods) of variables introduce uncertainties. They proposed the following uncertainty components:

— **Model output**: It is the uncertainty related to the difference between the model output and the true value of the quantity. It includes model error and covariance.

— **Calibration data**: The dataset used to train the inference model is composed of a series of measurement points that also have an uncertainty.

— **Input measurement**: Is the uncertainty associated with the measurement of the variables that are used as inputs in the inference model.

— **Output measurement**: It includes the effects of output uncertainties outside the calibration data set.

They present a case study with a power consumption virtual sensor for packaged air conditioning systems, and found that an increase in the number of points in the database reduces the uncertainty from the calibration data, but does not have any effect on the uncertainties from the model output deviation and output measurements [8].

Finally, in 2013, Yiqi, Daoping and Zhifu presented a soft sensor method based on a self-calibration model and uncertainty description algorithm [106]. However, these methods do not calculate uncertainty values from a metrological point of view, but rather they focus on the uncertainty related to the inference process by estimating an empirical reliability and the median value of the lengths of all predictive regions obtained for a specific significance level.

Despite a wide variety of publications and industrial implementations of soft metrology systems, there is still a huge lack of information regarding standardized prediction quality metrics, especially in the area of uncertainty quantification for these systems. Therefore, this subject is an open area for

| Technique         | Advantages                                                                 | Disadvantages                                                                 |
|-------------------|----------------------------------------------------------------------------|-------------------------------------------------------------------------------|
| LDA               | • Simple                                                                   | • Assumes Gaussian distribution                                              |
|                   | • Model stability                                                          | • Assumes classes with identical covariance                                   |
|                   | • Preferred when the process requires more than two non-ordinal response classes | • Low performance if the number of features exceeds the number of observations |
|                   |                                                                             | • Sensitive to overfitting                                                     |
|                   |                                                                             | • Does not work well if the dataset is not balanced                           |
| QDA               | • Simple                                                                   | • Assumes Gaussian distribution                                              |
|                   | • Avoids overfitting                                                       | • Low performance if the number of features exceeds the number of observations|
|                   | • Model stability                                                          | • Parameters to be adjusted                                                   |
|                   | • Preferred when the process requires more than two non-ordinal response classes |                                                                             |
| Decision tree     | • Easy to interpret                                                       | • Decision based primarily on expectations                                     |
|                   | • Non parametric routine                                                   | • Unexpected events may alter decisions                                        |
|                   | • Can model nonlinear relations                                             | • High probability of overfitting                                             |
|                   | • Can deal with numerical and categorical attributes                       | • Low prediction accuracy for a dataset                                        |
|                   | • Follows the same approach with which humans make decisions              | • Information gain for categorical variables gives a biased response for attributes with greater number of categories |
|                   | • Can be simplified by visualizations                                      | • Can be computationally demanding for many class labels                       |
|                   | • Hyper-parameters to be tuned is almost null                              |                                                                             |
| KNN/SVM/ANN       | See table 3                                                                | See table 3                                                                   |
future research and constitutes an important way forward in the development of soft metrology systems.

5. Discussion and conclusions

Here we present a general overview of indirect measurement strategies implemented in different contexts of application, taking into account general considerations for the implementation of soft metrology and machine learning algorithms. Additionally, a wider definition of the term soft metrology was established, gathering several concepts and terms commonly found in the literature, such as, soft sensors, virtual sensors, virtual metrology and inferential control, among others. The proposed definition includes several forms of indirect measurement and can promote research questions in this area by collecting the knowledge and developments made in different contexts under a single term. In this sense, it allows a better understanding of the development of indirect measurement systems when the representation space is not directly connected to the phenomenological principle, but an abstract representation is used for training and achieving new perspectives of measurement systems where different concepts and applications of machine learning offer several ways of implementation. This definition does not delve into developments and applications for human and social sciences, although the proposed definition considers the use that this term has had in these areas.

Data-driven soft metrology systems based on machine learning techniques have well-known routines to generate models that correlate the measured and inferred variables in 4 stages: database construction and pre-processing; generation of an effective representation space; model choice, training and validation; and model maintenance. The real problem of constructing an effective representation space is the training sample size, as the available databases usually have a small number of appropriately labeled samples and creating bigger databases is very expensive. Also, appropriate data comprehension is not always easily reachable due to differences between specific technical knowledge and the understanding of the software designer. Despite the widespread use of soft metrology systems, only a few cases present some type of uncertainty analysis of the estimated variables, opening study opportunities in this area of technical assurance in metrology.

The present review has not included the developments made in soft metrology systems using model-driven approaches.

Acknowledgments

This work was supported by the Instituto Tecnologico Metropolitano ITM of Medellin (Grant No. P17202). Additionally, the authors would like to thank the Measurement Analysis and Decision Support Laboratory (AMYSOD Lab) of Parque i, Medellin, Colombia.

ORCID iDs

Marcela Vallejo https://orcid.org/0000-0002-9724-3169
Andres Felipe Ramirez-Barrera https://orcid.org/0000-0003-1431-7116

Edilson Delgado-Trejos https://orcid.org/0000-0002-4840-478X

References

[1] Souza F A A, Araújo R and Mendes J 2016 Review of soft sensor methods for regression applications Chemom. Intell. Lab. Syst. 152 69–79
[2] Liu Z, Ge Z, Chen G and Song Z 2018 Adaptive soft sensors for quality prediction under the framework of Bayesian network Control Eng. Pract. 72 19–28
[3] Terzi M, Masiero C, Beghi A, Maggiopinto M and Susto G A 2017 Deep learning for virtual metrology: modeling with optical emission spectroscopy data 2017 IEEE 3rd Int. Forum on Research and Technologies for Society and Industry (RTSI) pp 1–6
[4] Kullaj J 2017 Bayesian virtual sensing for full-field dynamic response estimation Proc. Eng. 199 2126–31
[5] Pachauri N, Singh V and Rani A 2017 Two degree of freedom PID based inferential control of continuous bioreactor for ethanol production ISA Trans. 68 235–50
[6] Ebnaí M, Shahbazian M and Jazayerirad H 2018 Artificial intelligence for inferential control of crude oil stripping process Iran. J. Oil Gas Sci. Technol. 70 70–92
[7] Yao L and Ge Z 2018 Variable selection for nonlinear soft sensor development with enhanced binary differential evolution algorithm Control Eng. Pract. 72 68–82
[8] Cheung H and Braun J E 2016 A general method for calculating the uncertainty of virtual sensors for packaged air conditioners Int. J. Refrig. 63 225–36
[9] Acosta M, Kanarachos S and Fitzpatrick M E 2018 Robust virtual sensing for vehicle agile manoeuvring: a tyre-model-less approach IEEE Trans. Veh. Technol. 67 1894–908
[10] Kang S 2018 On effectiveness of transfer learning approach for neural network-based virtual metrology modeling IEEE Trans. Semicond. Manuf. 31 149–55
[11] Liang X, Li Y, Wu X and Shen J 2018 Nonlinear modeling and inferential multi-model predictive control of a pulverizing system in a coal-fired power plant based on moving horizon estimation Energies 11 1–27
[12] Liu Y, Yang C, Gao Z and Yao Y 2018 Ensemble deep kernel learning with application to quality prediction in industrial polymerization processes Chemom. Intell. Lab. Syst. 174 15–21
[13] Li H, Yu D and Braun J E 2011 A review of virtual sensing technology and application in building systems HVAC&R Res. 17 619–45
[14] Joseph B 1999 A tutorial on inferential control and its applications Proc. of the 1999 American Control Conf. pp 3106–18
[15] Mari L, Lazzarotti V and Manzini R 2009 Measurement in soft systems: epistemological framework and a case study Measurement 42 241–53
[16] Liu L, Kuo S M and Zhou M 2009 Virtual sensing techniques and their applications 2009 Int. Conf. on Networking, Sensing and Control pp 31–6
[17] Arpaia P et al 2015 Proof-of-principle demonstration of a virtual flow meter-based transducer for gaseous helium monitoring in particle accelerator cryogenics Rev. Sci. Instrum. 86 1–7
[18] Song L, Wang G and Brambley M R 2013 Uncertainty analysis for a virtual flow meter using an air-handling unit chilled water valve HVAC&R Res. 19 335–45
[19] Gubian M and Petri D 2008 Uncertainty analysis of learning-from-examples algorithms Int. Workshop pn Advanced Methods for Uncertainty Estimation in Measurement pp 1–6
[60] Lin B, Recke B, Knudsen J K H and Jorgensen S B 2007 A systematic approach for soft sensor development Comput. Chem. Eng. 31 419–25

[61] Wang D, Jun L and Srivivasan R 2010 Data-driven soft sensor approach for quality prediction in a refining process IEEE Trans. Ind. Inform. 6 11–7

[62] Hirai T and Kano M 2015 Adaptive virtual metrology design for semiconductor dry etching process through locally weighted partial least squares IEEE Trans. Semicond. Manuf. 28 137–44

[63] Khan A A, Moynie J R and Tilbury D M 2008 Virtual metrology and feedback control for semiconductor manufacturing processes using recursive partial least squares J. Process Control 18 961–74

[64] Bao L, Yuan X and Ge Z 2015 Co-training partial least squares model for semi-supervised soft sensor development Chemom. Intell. Lab. Syst. 147 75–85

[65] Ge Z and Song Z 2011 Semisupervised Bayesian method for soft sensor modeling with unlabeled data samples AIChE J. 57 2109–19

[66] Vallejo M, Gallego C J, Duque-Munoz L and Delgado-Trejos E 2018 Neuromuscular disease detection by neural networks and fuzzy entropy on time-frequency analysis of electromyography signals Expert Syst. 35 e12274

[67] Cuesta-Frau D, Murillo-Escobar P J, Orrego A D and Vallejo M, Gallego C J, Duque-Munoz L and Delgado-Trejos E 2019 Embedded dimension and time series length. Practical influence on permutation entropy and its applications Entropy 21 385

[68] Fortuna L, Graziani S and Xibilia M G 2009 Comparison of soft-sensor design methods for industrial plants using small data sets IEEE Trans. Instrum. Meas. 58 2444–51

[69] Delgado-Trejos E 2008 Measurement and feature extraction/selection in the automatic detection of ischemia and valvular pathologies on cardiac signals, Universidad Nacional de Colombia, Manizales (Available: http://bdigital.unal.edu.co/1100/)

[70] Yuan X, Ye L, Bao L, Ge Z and Song Z 2015 Nonlinear feature extraction for soft sensor modeling based on weighted probabilistic PCA Chemom. Intell. Lab. Syst. 147 167–75

[71] Wang Z X, He Q P and Wang J 2015 Comparison of variation selection methods for FLS-based soft sensor modeling J. Process Control 26 56–72

[72] Chandrashkegar G and Sahnin F 2014 A survey on feature selection methods Comput. Electr. Eng. 40 16–28

[73] Lynn S, Ringwood J, Ragnoli E, McLoone S and MacGearailt N 2009 Virtual metrology for plasma etch using tool variables 2009 IEEE/SEMI Advanced Semiconductor Manufacturing Conf. pp 143–8

[74] Li Q, Du X, Liu W and Ba W 2017 Soft sensor modelling based on mutual information variable selection and partial least squares 2017 Chinese Automation Congress (CAC) pp 3649–54

[75] Wang X and Liu H 2017 A new input variable selection method for soft sensor based on stacked auto-encoders 2017 IEEE 56th Annual Conf. on Decision and Control (CDC) pp 3324–9

[76] Liu J 2014 Developing a soft sensor based on sparse partial least squares with variable selection J. Process Control 24 1046–56

[77] Kang P, Kim D, Lee H, Doh S and Cho S 2011 Virtual metrology for run-to-run control in semiconductor manufacturing Expert Syst. Appl. 38 2508–22

[78] Wan J, Pampluri S, O’Harra P G, Johnston A B and McLoone S 2014 On regression methods for virtual metrology in semiconductor manufacturing 25th IET Irish Signals & Systems Conf. 2014 and 2014 China-Ireland Int. Conf. on Information and Communities Technologies (ISSC 2014/CHICT 2014) pp 380–5

[79] Kim M, Kang S, Lee J, Cho H, Cho S and Park J S 2017 Virtual metrology for copper-clad laminate manufacturing Comput. Ind. Eng. 109 280–7

[80] Park C and Kim S B 2016 Virtual metrology modeling of time-dependent spectroscopic signals by a fused lasso algorithm J. Process Control 42 51–8

[81] Zeng D and Spanos C J 2009 Virtual metrology modeling for plasma etch operations IEEE Trans. Semicond. Manuf. 22 419–31

[82] Gao J, Yang X, Shardi Y A W, Zhang Y, He F and Tong C 2017 Soft sensor modeling based on PCA and LS-SVM for strip thickness in cold steel rolling mills 2017 Chinese Automation Congress (CAC) pp 3758–63

[83] Maione C, Barbosa F and Barbosa R M 2019 Predicting the botanical and geographical origin of honey with multivariate data analysis and machine learning techniques: a review Comput. Electron. Agric. 157 436–46

[84] Xiong W and Shi X 2018 Soft sensor modeling with a selective updating strategy for Gaussian process regression based on probabilistic principle component analysis J. Franklin Inst. 355 5336–49

[85] Liu J, Liu T and Chen J 2018 Quality prediction for multi-stage processes by just-in-time latent variable modeling with integration of common and special features Chem. Eng. Sci. 191 31–41

[86] Zheng K and Funatsu K 2018 Partial constrained least squares (PCLS) and application in soft sensor Chemom. Intell. Lab. Syst. 177 64–73

[87] Wang J and Yan X 2019 Mutual information-weighted principle components identified from the depth features of stacked autoencoders and original variables for oil dry point soft sensor IEEE Access 7 1981–90

[88] Avendaño L E, Avendaño-Valencia L D and Delgado-Trejos E 2018 Diagonal time dependent state space models for modal decomposition of non-stationary signals Signal Process. 147 208–23

[89] Duque-Mejía C et al 2019 Cardiac Murmur effects on automatic segmentation of ECG signals for biometric identification: preliminary study Intelligent Information and Database Systems, ACIDS 2019 (Lecture Notes in Computer Science vol 11431) ed N T Nguyen et al (Cham: Springer) pp 269–79

[90] Jebri M, El Adel E, Graton G, Ouladsine M and Pinaton J 2011 Methods for plant valvular pathologies on cardiac signals, Universidad de Valencia, Manizales (Available: http://bdigital.unal.edu.co/1100/)

[91] de Morais T C B, Rodrigues D R, Souto U T D C P and Lemos S G 2019 A simple volumetric electronic tongue for the analysis of coffee adulterations Food Chem. 273 31–8

[92] Sliškovsk D, Grbic R and Hocenski Ž 2011 Methods for plant data-based process modeling in soft-sensor development Automatika 52 306–18

[93] Ge Z and Song Z 2010 A comparative study of just-in-time-learning based methods for online soft sensor modeling Chemom. Intell. Lab. Syst. 104 306–17

[94] Lynn S A, Ringwood J and MacGearailt N 2012 Global and local virtual metrology models for a plasma etch process IEEE Trans. Semicond. Manuf. 25 94–103

[95] Saitd I, Moufid M, Zaim O, El Bari N and Bouchikhi B 2018 Volantometric electronic tongue combined with chemometric techniques for direct identification of creatinine level in human urine Measurement 115 178–84

[96] Zhang X, Kano M and Li Y 2017 Locally weighted kernel partial least squares regression based on sparse nonlinear
features for virtual sensing of nonlinear time-varying processes Comput. Chem. Eng. 104 164–71

[97] Popli K, Maries V, Afacan A, Liu Q and Prasad V 2018 Development of a vision-based online soft sensor for oil sands flotation using support vector regression and its application in the dynamic monitoring of bitumen extraction Can. J. Chem. Eng. 96 1532–40

[98] Yu H, Zhang R, Meng Q and Lu S 2018 Research on soft sensing of cement clinker I-CaO based on LS_SVM and burning zone temperature 2018 Chinese Control and Decision Conf. (CCDC) pp 6366–70

[99] Meng Y, Lan Q, Qin J, Yu S, Pang H and Zheng K 2019 Data-driven soft sensor modeling based on twin support vector regression for cane sugar crystallization J. Food Eng. 241 159–65

[100] Kang P, Kim D and Cho S 2016 Semi-supervised support vector regression based on self-training with label uncertainty: an application to virtual metrology in semiconductor manufacturing Expert Syst. Appl. 51 85–106

[101] Puggini L and McLoone S 2015 Extreme learning machines for virtual metrology and etch rate prediction 2015 26th Irish Signals and Systems Conf. (ISSC) pp 1–6

[102] Wang J, Zhu L, Zhang W and Wei Z 2019 Application of the voltammetric electronic tongue based on nanocomposite modified electrodes for identifying rice wines of different geographical origins Anal. Chim. Acta 1050 60–70

[103] Jin W, Liu Y and Gao Z 2017 Fast property prediction in an industrial rubber mixing process with local ELM model J. Appl. Polym. Sci. 134 45391

[104] Zheng W, Gao X, Liu Y, Wang L, Yang J and Gao Z 2017 Industrial Mooney viscosity prediction using fast semi-supervised empirical model Chemom. Intell. Lab. Syst. 171 86–92

[105] Mavigipinto M, Terzi M, Masiero C, Beghi A and Susto G A 2018 A computer vision-inspired deep learning architecture for virtual metrology modeling With 2-dimensional data IEEE Trans. Semicond. Manuf. 31 376–84

[106] Yiqi L, Daoping H and Zhiflu L 2013 A SEVA soft sensor method based on self-calibration model and uncertainty description algorithm Chemom. Intell. Lab. Syst. 126 38–49

[107] Soares S, Araújo R, Soupa S and Souza F 2011 Design and application of soft sensor using ensemble methods EFTA2011 pp 1–8

[108] Rodríguez-Gamboa J C, Albarracin-Estrada E S and Delgado-Trejos E 2011 Quality Control Through Electronic Nose System, in Modern Approaches to Quality Control A. B. Eldín, Ed. (Rijeka: IntechOpen) pp 505–22

[109] Batista B L et al 2012 Multi-element determination in Brazilian honey samples by inductively coupled plasma mass spectrometry and estimation of geographic origin with data mining techniques Food Res. Int. 49 209–15

[110] Ruiz-Gómez J S et al 2018 Automated multiclass classification of spontaneous EEG activity in Alzheimer’s disease and mild cognitive impairment Entropy 20 35

[111] Hong J-H, Min J-K, Cho U-K and Cho S-B 2008 Fingerprint classification using one-versus-all support vector machines dynamically ordered with naïve Bayes classifiers Pattern Recognit. 41 662–71

[112] Leopold J H et al 2015 Comparison of classification methods in breath analysis by electronic nose J. Breath Res. 9 46002

[113] Hag A U and Djurdjanovic D 2016 Virtual metrology concept for predicting defect levels in semiconductor manufacturing 49th CIRP Conf. on Manufacturing Systems vol 57, pp 860–4

[114] Yan K and Zhang D 2015 Improving the transfer ability of prediction models for electronic noses Sens. Actuators B 220 115–24

[115] Rajasekar L and Sharmila D 2019 Performance analysis of soft computing techniques for the automatic classification of fruits dataset Soft Comput. 23 2773–88

[116] Cho J H and Kurup P U 2011 Decision tree approach for classification and dimensionality reduction of electronic nose data Sens. Actuators B 160 542–8

[117] Lu B and Chiang L 2018 Semi-supervised online soft sensor maintenance experiences in the chemical industry J. Process Control 67 23–34

[118] Kadlec P, Grbic R and Gabrys B 2011 Review of adaptation mechanisms for data-driven soft sensors Comput. Chem. Eng. 35 1–24

[119] Yuan X, Zhou J and Wang Y 2018 A comparative study of adaptive soft sensors for quality prediction in an industrial refining hydrocracking process 2018 IEEE 7th Data Driven Control and Learning Systems Conf. (DDCLS) pp 1064–8

[120] Zheng W, Liu Y, Gao Z and Yang J 2018 Just-in-time semi-supervised soft sensor for quality prediction in industrial rubber mixers Chemom. Intell. Lab. Syst. 180 36–41

[121] Kanelo H and Funatsu K 2011 Maintenance-free soft sensor models with time difference of process variables Chemom. Intell. Lab. Syst. 107 312–7

[122] Fu Y, Yang W, Xu O, Zhou L and Wang J 2017 Soft sensor modelling by time difference, recursive partial least squares and adaptive model updating Meas. Sci. Technol. 28 045101

[123] Kneale C and Brown S D 2018 Small moving window calibration models for soft sensing processes with limited history Chemom. Intell. Lab. Syst. 183 36–46

[124] Joint committee for guides in metrology (JCGM) 2012 International vocabulary of metrology—Basic and general concepts and associated terms (VIM) pp 1–104

[125] Joint Committee for Guides in Metrology (JCGM) 2008 Evaluation of measurement data—Guide to the expression of uncertainty in measurement pp 1–134

[126] Ferrero A and Salicone S 2004 The random-fuzzy variables: a new approach to the expression of uncertainty in measurement IEEE Trans. Instrum. Meas. 53 1370–7

[127] Joint Committee for Guides in Metrology (JCGM) 2008 Evaluation of measurement data—Supplement 1 to the “Guide to the expression of uncertainty in measurement”—Propagation of distributions using a Monte Carlo method pp 1–90

[128] Cheng Y et al 2018 Analysis and comparison of Bayesian methods for measurement uncertainty evaluation Math. Probl. Eng. 2018 1–10

[129] Hu Y, Xie F, Wu B and Wang Y 2013 An uncertainty quantification method based on generalized interval 2013 12th Mexican Int. Conf. on Artificial Intelligence, MICAI 2013 pp 145–50

[130] Sediva S and Havlikova M 2013 Comparison of GUM and Monte Carlo method for evaluation measurement uncertainty of indirect measurements Proc. of the 14th Int. Carpathian Control Conf. (ICCC) pp 325–9

[131] Korczynski M J and Hetman A 2005 A calculation of uncertainties in virtual instrument 2005 IEEE Instrumentationand Measurement Technology Conf. Proc. vol 3, pp 1697–701

[132] Xi W, Feng Y, Qin L and Yao Y 2011 Metrological methods for virtual instrument 2011 IEEE Int. Conf. on Computer Science and Automation Engineering pp 558–61