Highlights:

- A data-driven soft-sensing method is presented to deal with batch processes operated under varying initial conditions.
- The method presents high prediction capabilities at very low sampling rates even with scarce training information.
- The method shows very high applicability and robustness through a wide range of metamodeling techniques and case studies.
- Applications based on Ordinary Kriging show extreme competiveness respect to other common data-driven modeling techniques for soft-sensing

Data-driven Soft-sensors for Online Monitoring of Batch Processes with Different Initial Conditions

Ahmed Shokry[a1,c], Patricia Vicente[a1], Gerard Escudero[b1], Montserrat Pérez-Moya[a1], Moisès Graells[a1] , and Antonio Espuña[a1*]

[a] *Department of Chemical Engineering,* [b]*Department of Computer Science.*

[1]*EEBE – Eduard Maristany, 14, 08019 , Universitat Politècnica de Catalunya, Barcelona, Spain.*

[c] *Department of Mechanical Design and Production Engineering, Faculty of Engineering, Zagazig University, Zagazig, Egypt*

[*]*antonio.espuna@upc.edu*

Abstract

A soft-sensing methodology applicable to batch processes operated under changeable initial conditions is presented. These cases appear when the raw materials specifications differ from batch to batch, different production scenarios should be managed, etc. The proposal exploits the capabilities of the machine learning techniques to provide practical soft-sensing approach with minimum tuning effort in spite of the fact that the inherent dynamic behavior of batch systems are tracked through other online indirect measurements. Current data modelling techniques have been also tested within the proposed methodology to demonstrate its advantages. Two simulation case-studies and a pilot-plant case-study involving a complex batch process for wastewater treatment are used to illustrate the problem, to assess the modelling approach and to compare the modelling techniques. The results reflect a promising accuracy even when the training information is scarce, allowing significant reductions in the cost associated to batch processes monitoring and control.

Keywords: Soft-sensors, batch processes, ordinary Kriging, support vector machines, artificial neural networks, photo-Fenton.

1. Introduction

Competitive and rapidly changing market environments bear many sources of uncertainty and variability such as product demands, material availability, prices, product specifications and environmental restrictions. This has favored a continuous and growing interest in batch processes due to their high flexibility and adaptability, which also allow a quicker development of new products. These abilities stem from the relative independence of each equipment/unit and the possibility to reassign them and develop new production schemes. Thus, a wide range of important low-volume and high-value-added products are manufactured in a batch mode, including specialty chemicals, materials

for microelectronics, pharmaceutical, agricultural and biochemical product, etc. (Jin, et al., 2014; Moreno-Benito, 2014).

However, batch processes typically exhibit challenging operational problems (high inherent nonlinearity, transient dynamic behavior with no steady-state operating point, complex reaction kinetics, mechanisms and stoichiometry, etc.) that hamper their optimal management (Bonne & Jorgensen, 2004). A key challenge that commonly complicates their monitoring, supervision and/or control is the unavailability of online measurements of the process Quality Indicator Variables (QIV), which are often obtained through expensive and time consuming offline sampling and processing (Zamprogna et al., 2005; Desai et al., 2006). Moreover, a large laboratory delay also hinders a reliable process monitoring and supervision (Liu et al., 2012). Thus, soft-sensor techniques have been proposed as a promising solution that has proven its effectiveness in many situations (Hoskins & Himmelblau, 1988; Kadlec et al., 2009).

Soft-sensors are computational techniques that provide online estimations of process variables (including QIVs) that cannot be measured online in a continuous and/or reliable way due to technological and/or economic reasons (Kadlec et al., 2009; Lin et al., 2007). These techniques exploit the process variables that are reliably measured and recorded online with minimum cost by means of available physical sensors. Soft-sensors can be used for different purposes, but their basic application field is the online prediction of QIVs, so they could be further integrated in a monitoring and/or a control system (Kadlec et al., 2009; Jin et al., 2015).

Soft-sensor techniques can be categorized in two main classes: analytical model-based soft-sensors and data-based soft-sensors. Analytical model-based soft-sensors rely on First Principle Models (FPMs) that provide a detailed process description based on phenomenological knowledge (Lin et al., 2007; Jin et al., 2014); these FPMs are used to predict/monitor the process behavior, either solely or using the information provided by physical sensors (e.g. for continuously adjusting their parameters). However, accurate and reliable FPMs of chemical processes are often unobtainable, especially for complex highly nonlinear ones (Jain et al., 2007; Jin et al., 2015; Shokry et al., 2014): in many cases, the details required to build the models needed to describe such processes are limited, because of the involved highly nonlinear behaviors, sophisticated mechanisms and complex phenomena as reaction kinetics, thermodynamics etc. (Shokry et al., 2016; Shokry et al., 2017a). Even more, the existing FPMs of many processes have been developed under the assumption of the most favorable/ideal experimental and laboratory conditions, which make them sensitive to parameter variations, uncontrolled disturbances and distinct reactors geometries (Kadlec et al., 2009; Jin et al., 2014). Moreover, in real industrial or pilot plant scale applications, many other factors that affect or interact the process (e.g. mechanical and electrical components and connections, etc.), usually are not taken

into account by the FPMs, which badly affects their prediction accuracy (Jin et al., 2014; Kadlec et al., 2011; Shokry et al., 2015).

As an alternative, data-based soft-sensors are gaining wide interest in the process industry, because of its practicability, robustness and flexibility to be developed and applied to a wide range of processes, in addition to their independence from a process mathematical model (Hoskins & Himmelblau, 1988; Shokry et al., 2017a). They are based on the construction of a data-driven model able to accurately approximate the relation between the QIV and other online variables (Bonne & Jorgensen, 2004; Facco et al., 2009). In this line, many Machine Learning Techniques (MLT) for regression have been used to identify this relation, as Artificial Neural Networks (ANNs) (Gonzaga et al., 2009; Banu & Umab, 2011), Support Vector Regression (SVR) (Yan et al., 2004; Desai et al., 2006), and recently Gaussian Process (GP) models (Jin et al., 2015). These MLT are based on the use of input (online variables)-output (QIV) data from the process history, in order to find a black-box relations that describe the underlying mapping between the inputs and the output (training step). Then the constructed soft-sensor is used for the online prediction of the QIVs once the other online variables (online inputs) are measured (Jin et al., 2015). In this framework, measures of the QIV, when available, can be easily integrated in the system, either offline (e.g.: Kaneko and Funatsu, 2013, to update/improve the soft-sensor) or even online (e.g.: Shokry et al., 2017b, to confirm estimations and eventually reset them to better match the current situation).

Data-based soft-sensors have been vastly applied to continuous processes, in order to predict the process steady sate behavior, although they have shown limitations dealing with the transient states of the process (e.g. start-up and shut-down) (Facco et al., 2009; Wang et al., 2016). Comparatively, the development and application of data-based soft-sensors to batch processes, which are always in transient state, have been found to be relatively more complicated (Bonne & Jorgensen, 2004; Liu et al., 2012).

The combination of Principal Component Regression (PCR) and Partial Least-Squares (PLS) techniques is the most common method for building data-based soft-sensors. PCR is used to reduce the input space via transforming it into a lower dimensional space in which the importance of each new feature/input is determined (Jin et al., 2014; Kadlec et al., 2009; Zamprogna et al., 2005). PLS is then used to find a regression model that correlates the transformed input variables or features with the output variables (QIVs) (Lin et al., 2007; Facco et al., 2009; Zamprogna et al., 2005). However, these methods are basically developed for linear modeling and they may oversimplify the description of complex nonlinear behaviors (Jin et al., 2015; Nagy, 2007). Although some versions of these techniques have been developed to inexpensively handle nonlinearities (as kernel PCR and kernel PLS), more advanced and efficient techniques have been proposed for soft-sensing highly nonlinear processes (Kadlec et al., 2009; Nagy, 2007; Shokry et al., 2015).

In this scope, Artificial Neural Networks (ANNs) approaches (Masters, 1993) have been often selected for soft-sensing, due to their universal approximation and efficient generalization performance, additional to their flexible structure of nonlinear neurons that enable ANNs to capture sophisticated behaviors (Yan et al., 2004; Kadlec et al., 2009). Several types of ANNs have been efficiently used for soft-sensing, as Multi-Layer Perceptron (MLP-ANNs), Radial Basis ANNs (RB-ANNs) and Fuzzy ANNs (Masters, 1993; Nelles, 2001; Nagy, 2007). Nevertheless, ANNs particularly suffer from the curse of dimensionality and require laborious tuning (selection of the network structure and configuration; as number of layers, number neurons in each layer, transfer function type, training method etc.) to achieve a reliable model fitting (Azman & Kocijan, 2007; Davis & Ierapetritou, 2007; Caballero & Grossmann, 2008; Shokry et al., 2015).

Models based on Support Vector Regressions (SVR) have been also proposed for soft-sensing in batch processes (Yan et al., 2004; Desai et al., 2006; Kadlec et al., 2009). The SVR model (Vapnik, 1995) consists of a subset of specific training data (support vector) that compose the margins of the simplest functional shape within which the prediction error for all training data is acceptable. SVR techniques have very good generalization properties, and rapidity of tuning (associated to the optimization problem solution time for the support vectors selection) (Yan et al., 2004; Jain et al., 2007; Kadlec et al., 2009; Shokry et al., 2015). However, the effort and the time required to select the parameters of the SVR model –prior to the optimization-, as the penalty cost, the error margin and the variance become a major limitation (Forrester et al., 2008; Shokry et al., 2015).

In the Bayesian statistics and inference area, the Gaussian Process (GP) models are initially proposed by O'Hagan et al. (1978; 1999) in order to represent a general class of non-parametric probabilistic models. Later on, they have gained a wide popularity within the machine learning community through the works of Rasmussen and Williams (2006). Recently, Gaussian Process (GP) models are attracting huge attention in the soft-sensing area, and have been applied either for continuous (Grbić et al., 2013; Wang et al., 2016; Liu et al., 2016) or batch processes (Jin et al., 2015), offering high prediction accuracy and tuning flexibility while requiring a relatively small set of the training data. Pioneering GP metamodels were developed in the nineties (O'Hagan et al., 1999), mainly for complex computer code emulation (O'Hagan, 2001), and afterward they become popular for dynamic modeling (Azman & Kocijan, 2007). But the computational effort and capabilities required for its parameter tuning could be a serious shortage, especially for high dimensional cases and/or large training datasets.

The techniques based on Ordinary Kriging (OK) (Krige, 1951; Kleijnen, 2017) may be considered as specific forms/applications of the GP models. Their use was first proposed in the field of mining industry and geo-statistics by Dr. Danie Krige (1951) and Matheron (1963), where their predictor derivations and parameters estimations are sought as the "Best linear Unbiased Predictor" of the realization of a spatial process. After the works of Sacks et al. (1989) and Jones et al. (1998), the OK

became very popular in the context of the modelling, simulation and optimization of complex nonlinear systems in many engineering fields (Queipo et al., 2005; Forrester & Keane, 2009; Pasquale et al., 2016). Ordinary Kriging has shown outperforming characteristics, as high prediction accuracy with relatively small number of training data, and relatively high tuning flexibility (Forrester et al., 2008). The use of OK metamodels was introduced to the chemical process engineering area by Davis and Ierapetritou (2007), and Caballero and Grossmann (2008), and since that time it is gaining growing interest mainly for surrogate based optimization and analysis of complex nonlinear chemical systems (Shokry et al., 2014; Quirante et al., 2015; Rogers & Ierapetritou, 2015), and later on for multivariate dynamic modelling (Shokry et al., 2016; Shokry & Espuña, 2017a). However, the use of OK metamodels - as a specific implantation/formulation of Gaussian Process models - has never been introduced to the area of the soft-sensing of batch chemical processes yet.

In the literature, the studies that have addressed data-based soft-sensing of batch processes consider a single operating phase/mode along the batch run, which is also the scope of this paper. Consequently, the MLTs are harnessed to provide a single global data-based model that describes the relation between the online variables and the QIV along the whole batch (Yan et al., 2004; Desai et al., 2006; Jain et al., 2007; Banu & Umab, 2011; Grbić et al., 2013; Gustavsson et al., 2015). However, for alternate batch process conditions, these global models may produce inaccurate predictions in specific input regions, and in these cases, the process should be described through a multiphase/multimode model, based on the construction of several local models, each one responsible for predicting the process behavior over a specific local input domain (Liu et al., 2012; Jin et al., 2014; Wang et al., 2016). Further research addresses the development of adaptive data-based soft-sensors to be used online in processes showing time-varying behavior, e.g. due to process fouling, aging etc. (Grbić et al., 2013; Kaneko & Funatsu, 2013). Different approaches have been proposed as the moving window approach and the recursive adaptation methods that automatically manage online data to update the soft-sensor in order to maintain its prediction performance (Jin et al., 2015).

Most of the data-driven soft-sensing approaches for batch processes proposed in the literature have not considered the initial conditions of the batches in their design, because they have been tailored for batch processes operated under fixed initial conditions. In these cases, the data-based soft-sensing approach have addressed the batch-to-batch data variability -due to a very slight change in the initial condition- from the uncertainty and noise perspectives: input-output training data from different batch runs are assumed to have random errors due to undesired disturbances, which are expected to be representative of a population of batches that are swarming around the mean behavior of the process or what is called the “reference batch” or the “golden batch” (Kadlec et al., 2009). Then, the correct underlying process behavior can be identified, thanks to the regularization abilities of the employed

MLTs, which enable them to learn from this uncertain and perturbed data, and to filter out the assumed noise.

Considering this overview, the novelty of this work relies on three main contributions:

1. To address the development of a soft-sensing approach for a special type of batch processes that is rarely explored in the area of soft-sensing: those batch processes that show a characterized variability in their initial settings or conditions, as a result of a specific purpose decision (i.e.: different from a small uncertain noise or an uncontrolled variation). These cases are usual in such batch processes aiming to manage raw materials whose specifications or properties differ from one batch to another (e.g. waste treatment systems), or when different product qualities/quantities are to be generated. Hence, the objective is to develop a soft-sensor able to estimate the QIVs along the batch run under any set of initial conditions in the expected operating range.

2. To explore the advantages of the OK metamodeling technique –as a kind of GP metamodels- for soft-sensing in the chemical engineering area: the Kriging method is compared to the most common data-based modeling techniques used for soft-sensing as SVR and ANN, in order to assess and compare its capabilities.

3. To provide an efficient soft-sensor able to track an advanced oxidation process (AOPs) based on the photo-Fenton reaction, which in the same time can be used as data-based process model to help understanding the complex behavior usually associated to these processes. Nowadays, AOPs are receiving a huge interest, because of their capacity to manage Contaminants of Emerging Concern (CECs) present in the water, which cannot be degraded using conventional water treatments technologies (biological, physical or chemical). However, due to the complexity and high nonlinearity of these processes, the best way to address their analytical or phenomenological modeling is still under debate in the scientific and research community; while many data-based modelling studies of these processes have been accomplished from the point of view of experimental design in laboratory scale, their monitoring and control have been never addressed from a soft-sensing perspective, i.e. at industrial or pilot plant scale.

The paper is organized as follows: Section 2 presents the theoretical basis of the modeling techniques utilized and the proposed training and validation procedures. Part 3 illustrates the proposed modelling approach associated to the building of the soft-sensor, and its application to two simulation case-studies: a simple first order reaction and a fed-batch fermenter, which allows to assess the modelling approach robustness and to compare the modeling techniques. Part 4 describes the case-study used for experimental application and the results obtained. Finally the main conclusions are summarized.

2. Modelling techniques (data-based models)

Data-based soft-sensors rely on harnessing metamodels or data-driven models defining accurate black-box relations between the QIV and the online variables. This part spots the light on the most common nonlinear data-driven modeling tools that have used in the soft-sensing area as OK, ANN and SVR; including their mathematical basis, software available and details of implantations (training, validation etc.).

2.1. Ordinary Kriging (OK)

The OK is a nonparametric model that has shown potential capabilities to approximate highly nonlinear, multimodal and complex systems (Queipo et al., 2005; Fang et al., 2006; Forrester et al., 2008). These capabilities stem from the ability of the OK to combine global modeling through estimating a general trend of the system to be approximated, and local modeling through a spatial correlation function. Besides, this method is able to estimate a prediction variance or error, which represents an uncertainty measure about the prediction.

Given a set of n input-output training data $[x_i,y_i]$, $i=1,2,..n$, $x \in R^k, y \in R$, the OK assumes the predictor $\hat{y}(x)= \mu_{ok} + Z(x)$, where the constant term μ_{ok} represents the main trend of the system to be approximated, and $Z(x)$ is a deviation/residual from that trend, which accounts for detailed complex behavior of the system that could not be captured via the main trend μ_{ok}. The residual $Z(x)$ is modeled as a stochastic Gaussian process with expected value of zero $E(Z(x))= 0$, and a covariance between two residuals $cov(Z(x_i),Z(x_j))$ that only depends on their corresponding inputs x_i, x_j. Thus it can be calculated as: $cov(Z(x_i),Z(x_j)) = \sigma_{ok}^2 R(x_i,x_j)$, being σ_{ok}^2 the process variance, and $R(x_i,x_j)$ a correlation function, $R(x_i, x_j) = exp\left(-\sum_{l=1}^{k} \xi_l \left|x_{i,l} - x_{j,l}\right|^{p_l}\right) + \delta_{ij}\,\lambda$, where, ξ_l are the model hyper-parameters, δ_{ij} is the Kronecker delta, p_l are smoothing parameters and λ is a regularization constant that enables the Kriging predictor to regress noisy data (Forrester & Keane, 2009).

The maximization of the likelihood function (Eq.(1)) of the observed data $[Y]_{n\times1}$ yields the closed form mathematical expressions for the optimal values of μ_{ok} and σ_{ok}^2 that are shown in Eq. (2) and Eq. (3) respectively (Caballero & Grossmann, 2008), where $[X]_{n\times k}$ is the matrix of training inputs, $[Y]_{n\times1}$ is the vector of the training outputs, $[R]_{n\times n}$ is the correlation matrix between the training inputs and $[\mathbf{1}]_{n\times1}$ is the identity vector- it highlighted with bold fond to differentiate it from the normal number 1 in the equations-.

$$Lik(\mu_{ok}, \sigma_{ok}^2, X) = \frac{\mathbf{1}}{(2\,\pi\,\sigma_{ok}^2)^{\left(\frac{n}{2}\right)} |R|^{\left(\frac{1}{2}\right)}}\; exp\left(-\frac{(Y - \mathbf{1}\,\mu_{ok})^T\, R^{-1} (Y - \mathbf{1}\,\mu_{ok})}{2\,\sigma_{ok}^2}\right) \qquad (1)$$

$$\mu_{ok} = \frac{\mathbf{1}^{T} R^{-1} Y}{\mathbf{1}^{T} R^{-1} \mathbf{1}} \tag{2}$$

$$\sigma_{ok}^{2} = \frac{(Y - \mathbf{1}\,\mu_{ok})^{T} R^{-1} (Y - \mathbf{1}\,\mu_{ok})^{T}}{n} \tag{3}$$

The substitution of the optimal values of μ_{ok} and σ_{ok}^{2} in the likelihood function leads to the maximization of the concentrated log-likelihood function, which is given by Eq.(4).

$$Max_{\,(\xi_l, p_l)} \left[-\frac{n}{2} ln(\sigma_{ok}^{2}) - \frac{1}{2} ln(|R|) \right] \tag{4}$$

The Kriging predictor (Eq.(5)) and its estimated error (Eq.(6)) are obtained by deriving the augmented likelihood function of the original training data set and a new interpolating point (x^{*},y^{*}). In Eq.(5), $[r]_{n\times 1}$ is the vector of correlations between the point to be predicted x^{*} and the original training data points, and calculated as $R\,(x_i,\, x^{*})$ (Jones et al., 1998; Caballero & Grossmann, 2008; Forrester et al., 2008).

$$\hat{y}(x^{*}) = \mu_{ok} + r^{T} R^{-1} (Y - \mathbf{1}\mu_{ok}) \tag{5}$$

$$\hat{s}^{2}(x^{*}) = \sigma_{ok}^{2} (1 + \lambda - r^{T} R^{-1} r + (1 - \mathbf{1}^{T} R^{-1} r)^{-1} / (\mathbf{1}^{T} R^{-1} \mathbf{1})) \tag{6}$$

The fitting of a Kriging metamodel is achieved by obtaining the optimal parameters [ξ_l, p_l, λ] through the maximization of the concentrated log-likelihood function. In practice, this optimization problem is computationally challenging, because of the high effort associated to the repetitive calculation of the correlation matrix $[R]_{n\times n}$ inverse during the optimization iterations, so this effort quickly grows with the size of the training data set and/or the model input dimensionality. Besides, the nature of the concentrated log-likelihood function itself is quite complicated, because it is flat near the optimum (Fang et al., 2006; Shokry et al., 2014). More details about these computational challenges and the numerical methods and optimization techniques to overcome or reduce these obstacles can be found in (Fang et al., 2006).

Different software implementations of the OK approach have been developed (Lataniotis et al., 2017; Erickson et al., 2018), each one attracting different audiences. This work considers the OK implementation developed by Forrester et al. (2008), because of its high efficiency, generality and applicability. Besides, the proposed approach uses the "*fmincon*" algorithm included in the Matlab-2015b optimization toolbox library for the maximization (nonlinear optimization) of the concentrated likelihood function (Eq.(4)); the Cholesky factorization has been used to find the inverse of $[R]_{n\times n}$ to

avoid ill-conditioning. The smoothness parameters p_l are often kept to the value of 2, which provide smooth infinitely-differentiable correlation functions (Forrester et al., 2008).

This work also considers another different software implementation for the GP model construction: the GP-Regression (GPR) algorithm based on the function "*fitrgp*" included in the Matlab statistical toolbox. Here, it is worthy to emphasize that the objective of this work is not to compare different specific implementations of the GP models but to explore the robustness and flexibility of the proposed soft-sensing methodology by handling different data-based modelling techniques and software.

2.2. Artificial Neural Networks (ANNs)

The ANNs is a well-known efficient method widely used for nonlinear system modelling and approximation is inspired from the biological neural networks of the brain nervous system (Masters, 1993; Banu & Umab, 2011). Basically, an ANN is a network or a structure of simple nonlinear processing units called neurons. The neurons are located/arranged in this structure through specific number of layers (an input layer, a minimum of one hidden layer and an output layer), and are interconnected in order to be able to process the information among them. A certain value of weight w_{ij} is assigned to each connection linking the i^{th} neuron in the previous layer to the j^{th} neuron in the current layer, besides a bias b_j is introduced as an independent input to each neuron. Every j^{th} neuron calculates the sum of its inputs x_{ij} from preceding neurons, additional to the bias input, see Eq.(7), where K and Q are the number of neurons in the previous and current layer, respectively. This summation is calculated taking into account the current weight of each neuron-to-neuron connection. The resulting value is then passed through a transfer function f (also called activation or threshold function), and sent to the next layer in the network and so, on until the output layer (Masters, 1993; Nagy, 2007).

Given a set of input-output training data, the network is trained by manipulating the weights and the biases values, in order to minimize a cost function, which is usually related to the sum of the square errors between the network predicted outputs and the target outputs (Masters, 1993; Nagy, 2007).

$$a_j = f(b_j + \Sigma\, x_{ij}\ w_{ij}), \quad i = 1,2,..K, \qquad j = 1,2,..Q \tag{7}$$

Among the different types of ANNs, the MLP-ANN kind is the most common one in engineering applications, because its efficiency and accuracy is combined with a simple and straightforward applicability (Fang et al., 2006; Nagy, 2007; Jin et al., 2015). For this reason, the MLP-ANN is the

type considered in this work. The Matlab ANN toolbox and the function "*feedforwardnet*" have been used to create a feed forward ANN; the number of neurons and layers, and the training algorithm are selected through a traditional try and cut procedure to balance the structure simplicity and the prediction accuracy.

2.3. Support Vector Regression (*SVR*)

Given the set of *n* input-output training data, SVR (Vapnik, 1995) map the input data original space into a high-dimensional feature space, often through a basis or kernel function $\Phi\ (x_i, x_j)$ that may present different styles as linear, polynomial, Gaussian, etc. Then, the modeling problem becomes the determinations of the optimal (flattest) linear surface $\hat{y}(x) = \mu_{svr} + \sum_{i=1}^{n} w_i \Phi\ (x_i, x_j)$ in this feature space that fits the data, through the minimization of the weights vector norm $|w|^2, w \in R^n$, where μ_{svr} is a base or bias (Forrester & Keane, 2009). In order to ensure better generalization performance, SVR allows specifying margins or a tube around the training data with a radius $\pm\varepsilon$, within which prediction errors in the training data are accepted or tolerable (constraints of the optimization problem). To allow/tolerate outliers, the data that presents a prediction error bigger than $\pm\ \varepsilon$ is penalized using the so called ε-sensitive loss function (Forrester et al., 2008). Then the model fitting problem can be expressed as:

$$Min\ \frac{1}{2}|w|^2 + \frac{C_{svr}}{n}\sum_{i=1}^{n}(\xi_i^+ + \xi_i^-) \qquad (8)$$

Subjected to

$$y_i - \mu_{svr} - w\ x_i \leq \varepsilon + \xi_i^+$$

$$\mu_{svr} + w\ x_i - y_i \leq \varepsilon + \xi_i^- \qquad (9)$$

$$\xi_i^+\ ;\ \xi_i^-\ \geq 0$$

Where ξ_i^+ ; ξ_i^- are the slack variables that describe the size of the positive and negative violation or excess than the tube radius ε for each training data, and $C_{svr} > 0$ is a penalty factor that controls the trade-off between the model complexity (the flatness of $\hat{y}$) and the degree to which errors larger than ε are tolerated.

The constrained optimization problem can be reformulated into a dual problem form by introducing Lagrange multipliers $\eta_i^+, \eta_i^-, \alpha_i^+, \alpha_i^-$ to the constraints in Eq. (9), in order to combine them with the objective forming at the end the Lagrangian function:

$$L = \text{Min}\ \frac{1}{2}|\mathrm{w}|^2 + \frac{\mathrm{C_{svr}}}{\mathrm{n}}\sum_{\mathrm{i=1}}^{\mathrm{n}}(\xi_{\mathrm{i}}^{+} + \xi_{\mathrm{i}}^{-}) - \sum_{i=1}^{n}\alpha_i^{+}(\varepsilon + \xi_i^{+} - y_i + \mu_{svr} + w\,x_i) + \sum_{i=1}^{n}\alpha_i^{-}(\varepsilon + \xi_i^{-} + y_i - \mu_{svr} - w\,x_i) - \sum_{i=1}^{n}(\eta_i^{+}\,\xi_i^{+} + \eta_i^{-}\,\xi_i^{-}) \quad (10)$$

The resulting objective L is then minimized in front of w, μ_{svr} and the primal variables $\xi_i^{\pm}$, and also it is maximized with respect to the dual variables $\eta_i^{\pm}$, $\alpha_i^{\pm}$, where $\eta_i^{\pm}$, $\alpha_i^{\pm} \geq 0$. For the active constraints $(\alpha_i^{+} + \alpha_i^{-}) \geq 0$, the corresponding y_i will become the support vectors, while for inactive constraints $(\alpha_i^{+} + \alpha_i^{-}) = 0$, the corresponding y_i will be excluded from the prediction.

The values of these Lagrange multipliers $\alpha_i^{+}, \alpha_i^{-}$ are determined by solving the dual optimization problem. The training vectors (samples) with non-zero Lagrange multipliers are called support vectors, which represent/construct the margins or the borders of the tube.

Finally, the optimal weights *w* and the constant bias μ_{svr} can be calculated from the relations in Eq.(11) and Eq. (12), and the final predictor is expressed by Eq.(13).

$$w = \sum_{i=1}^{n}(\alpha_i^{+} - \alpha_i^{-})\Phi(x_i) \quad (11)$$

$$\mu = y_i - \sum_{j=1}^{n}(\alpha_j^{+} - \alpha_j^{-})\Phi(x_i, x_j) \quad (12)$$

$$\hat{y}(x^{*}) = \mu_{svr} + \sum_{i=1}^{n}(\alpha_j^{+} - \alpha_j^{-})\,\Phi(x^{*}, x_i) \quad (13)$$

A serious draw back of the SVR is the huge time and effort required to select the kernel function type and the values of its parameters (e.g.: σ_{svr}), which are case dependent. The detailed mathematical description and derivations can be found in (Vapnik, 1995; Forrester et al., 2008; Forrester & Keane, 2009). This work uses the support vector regression algorithm based on the function "*fitrsvm*" included in the Matlab statistical toolbox library.

2.4. Metamodel validation

A common way to validate a metamodel is to use a different input-output dataset, usually known as validation set: $[x^{v}, y^{v}]_{n_v}$. Then, the metamodel is used to predict the outputs $\hat{y}_i^{v}$, and the prediction is compared by the known output y_i^{v}. The Root Mean Square Error (RMSE), the Normalized Root Mean

Square Error (NRMSE), and the Correlation Coefficient (CC) are calculated as accuracy measures of the prediction. The RMSE and the NRMSE are direct measures of accuracy: they are reporting an average deviation measure of predicted values from the actual values, where the RMSE is an absolute quantity, and the NRMSE is a relative quantity to the variability range ($y_{i,max}^{v} - y_{i,min}^{v}$) of the output *y*. However, the *CC* is an indicator of the matching or the trend between the overall predictions and the real output values.

$$RMSE = \sqrt{\frac{1}{n_v}\sum_{i=1}^{n_v}(y_i^v - \hat{y}_i^v)^2} \tag{14}$$

$$NRMSE = 100\frac{RMSE}{(y_{i,max}^{v} - y_{i,min}^{v})} \tag{15}$$

$$CC = r_{y^v,\hat{y}^v} = \frac{\sum_{i=1}^{n_v}(\hat{y}_i^v - \bar{\hat{y}}^v)(y_i^v - \bar{y}^v)}{\sqrt{\sum_{i=1}^{n_v}(\hat{y}_i^v - \bar{\hat{y}}^v)^2}\sqrt{\sum_{i=1}^{n_v}(y_i^v - \bar{y}^v)^2}}, \quad \bar{y}^v = \frac{\sum_{i=1}^{n_v} y_i^v}{n_v}, \quad \bar{\hat{y}}^v = \frac{\sum_{i=1}^{n_v} \hat{y}_i^v}{n_v} \tag{16}$$

3. Modelling approach

Consider a batch process in which a significant QIV *y*(*t*) (e.g.: reaction progress) may be only expensively measured and/or requires offline sampling and analysis over relatively large sampling time periods. This variable coexist along with other variables *x*(*t*) that are measured and recorded online in an automatic and continuous way (over very small sampling periods) probably with a more reduced cost. Additionally and more importantly, the process may run under different combination of initial conditions $[x(t=0)\ y(t=0)]$, which vary within a known range or bounds $[x_{min}^{t=0}{:}\,x_{max}^{t=0}\,, y_{min}^{t=0}{:}\,y_{max}^{t=0})]$.

The objective is to develop a soft-sensor that is able to predict the QIV as a function of the cheaply measured online variables, over the whole batch run and departing from any combination of the initial conditions over their known variation range. This soft-sensor might be further used to monitor the process and to predict the reaction progress at any time along the batch run, saving the cost and the time of the offline experimental sampling and analysis. The proposed design/modeling approach of such soft-sensor is based on the approximation of the current measurements of the offline QIV, *y*(*t*), as a function of the current values of the online variables *x*(*t*) and the initial values of both of the offline *y* (*t=0*) and the online *x*(*t=0*) variables (Eq.(17)). Thus, the initial conditions [*x* (*t*=0), *y* (*t*=0)] provide or identify the overall effect on the main path or trajectory of the batch QIV, while the current values of the online variables *x*(*t*) provide the instantaneous or temporal effect.

$$y(t) = f[x(t), y(t=0), x(t=0)] \quad (17)$$

The methodology steps include:

1. The selection of a proper set of training batches whose ICs $[x(t=0)\ y(t=0)]$ cover or span –as much as possible- the known variation domain of the ICs $[x_{min}^{t=0}: x_{max}^{t=0}, y_{min}^{t=0}: y_{max}^{t=0})]$. In this way, the training data includes more information about the system different dynamic behaviors corresponding to different ICs.

    a. In the case of data generation from a complex FPM simulation, different design of computer experiments techniques can be efficiently used for the definition of a representative set of the training data. Typical situations are when the complexity of the FPM hinders its practical usage for the online monitoring of the process.

    b. In the case of a real process without an available accurate FPM, the process history data should be exploited as much as possible

2. The online data $x(t)$ of each training batch is smoothed using a moving average technique –a proper time window should be selected- to reduce Gaussian noise of the sensors.

3. The input-output (online-offline) data are collected from each training batch. Thus, the number of the input-output training data extracted from each batch equals to the number of the QIV measurements along the batch run.

4. The data collected from each training batch is unfolded to compose the overall training set, noticing that the overall training dataset does not include any information neither about the batches identifications nor about the temporal sequence of the measurements. Then the soft-sensor (Eq.(17)) is trained according to the requirements of the selected modelling technique (OK, GPR, SVR or ANN).

5. The trained soft-sensor is then used along a series of validation batches, in order to predict the values of the offline QIV($y(t)$) along the whole run of each batch, using only the initial values $[x_{t=0};\ y_{t=0}]$ and the online measured variables $x(t)$.

    a. The accuracy measures (Eqs.(14), (15), (16)) can be calculated through comparing the estimated QIV values by the soft-sensor to their known real values.

It is worth noting that the data used to train the soft sensor will be obviously affected by random or Gaussian errors, due to the nature of the real-sensor(s) and the experimental instruments providing these data. Thus, the target accuracy of a soft-sensor cannot be higher than the accuracy offered by the

real sensors and instruments providing the training data. In addition, this maximum accuracy will be also affected (reduced) by the accuracy of the sensors (real-sensors) used to provide the online information required by the already fitted soft-sensor to perform the required online estimations. As a result, it can be concluded that the accuracy of a soft-sensor is expected to be within the same order of magnitude of (but lower than) the accuracy offered by the real-sensor(s) and instruments providing the information required to train and also to use the soft-sensor.

In the next subsections, the application of the proposed soft-sensing approach to two simulated case-studies based on the modeling techniques previously introduced in part 2 (OK, SVR and ANN), will be illustrated.

3.1. Batch reactor

This simulation case-study was proposed by Ruppen et al. (1995) and involves a common batch reactor model (Eqs.(18)) running the reactions $2A \overset{k1}{\rightarrow} B \overset{k2}{\rightarrow} C$, being A the reactant, B the undesired product and C the desired product. It is assumed that the process depends only on the initial conditions (no perturbations/external inputs are considered along the batch run).

$$\frac{dc_A}{dt} = -2k_1\, c_A^2 \ , \qquad \frac{dc_B}{dt} = k_1 c_A^2 - k_2 c_B c_C \, , \qquad \frac{dc_C}{dt} = k_2 c_B c_C, \quad k_i = k_{i,0} exp\left(\frac{E_i}{RT}\right) \qquad (18)$$

The concentration C_C of product C is considered as the offline QIV, while the concentrations C_A and C_B of products A and B are treated as the online variables. Thus, during each batch run (1 *hr*), C_C is calculated at 7 specific sampling times (every ten minutes), simulating its expensive offline sampling and analysis. On the contrary, C_A, C_B are calculated every second, simulating the automatic recording of data by the physical sensors. Small white or Gaussian noise $\mathcal{N} \approx (\mu = 0, \sigma = 0.018)$ is added to the simulated online variables values (C_A, C_B), while a higher error $\mathcal{N} \approx (\mu = 0, \sigma = 0.16)$ is also added to the simulated offline variable data (C_C) in order to mimic the experimental error. The white noise and the experimental errors in this case-study and the next ones are estimated in a heuristic manner with respect to the variability domain of each variable. Hence, the experimental error variance σ has been set to 1.5 % of the variability range for the offline variables, while the white noise variance σ for the online variables has been set to 0.5 % of their respective variability ranges.

In order to generate the training data, 24 batches are simulated in the previously described way, using different initial concentrations values [$C_A(t=0)$, $C_B(t=0)$, $C_C(t=0)$] selected by a Hammersley sampling procedure within the limits $\left[C_{A,min}^{t=0}: C_{A,max}^{t=0}, C_{B,min}^{t=0}: C_{B,max}^{t=0}, C_{C,min}^{t=0}: C_{C,max}^{t=0},\right]$ = [1 : 4, 0 : 2, 0 : 3] (*Mol*/*L*). Additionally, another set of 100 batches with different initial concentrations (but within the same domain) are simulated and used as the validation set, see Figure 1. The whole online/offline data

set generated for 4 specific training batches is illustrated in Figure 2. The initial conditions of these 4 training batches are highlighted in Figure 1 with their corresponding colors.

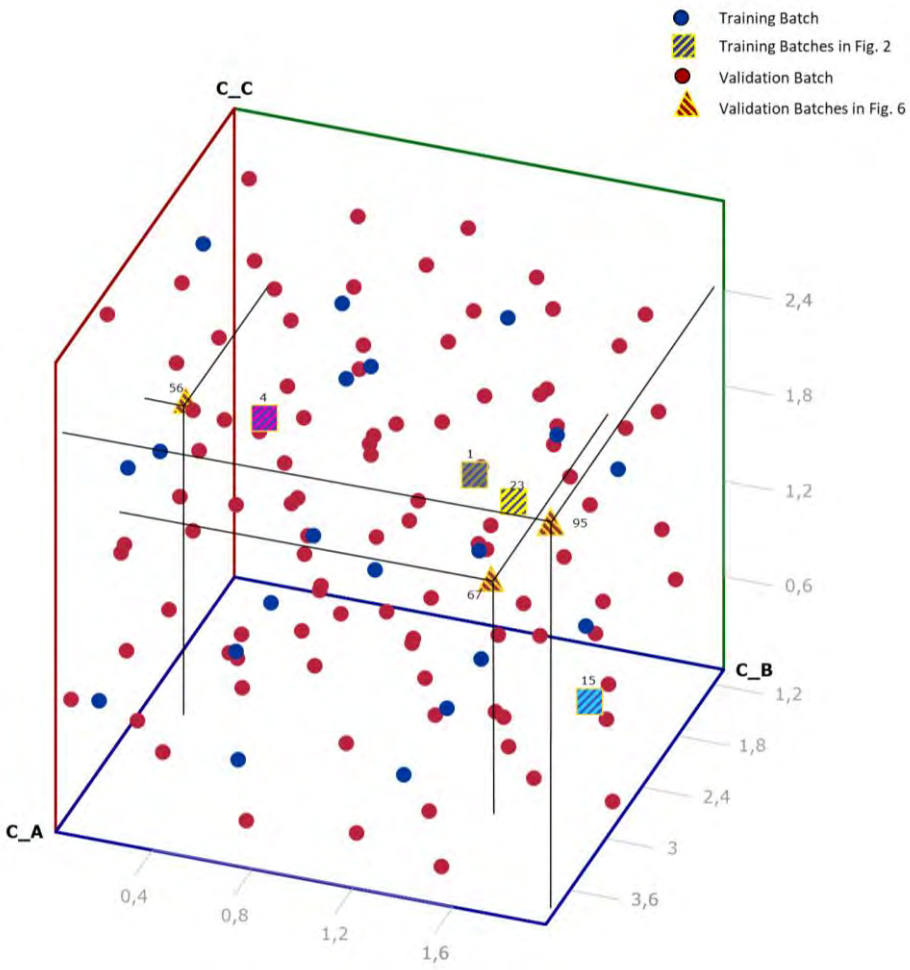

Figure 1. Initial conditions of the training and validation batches.

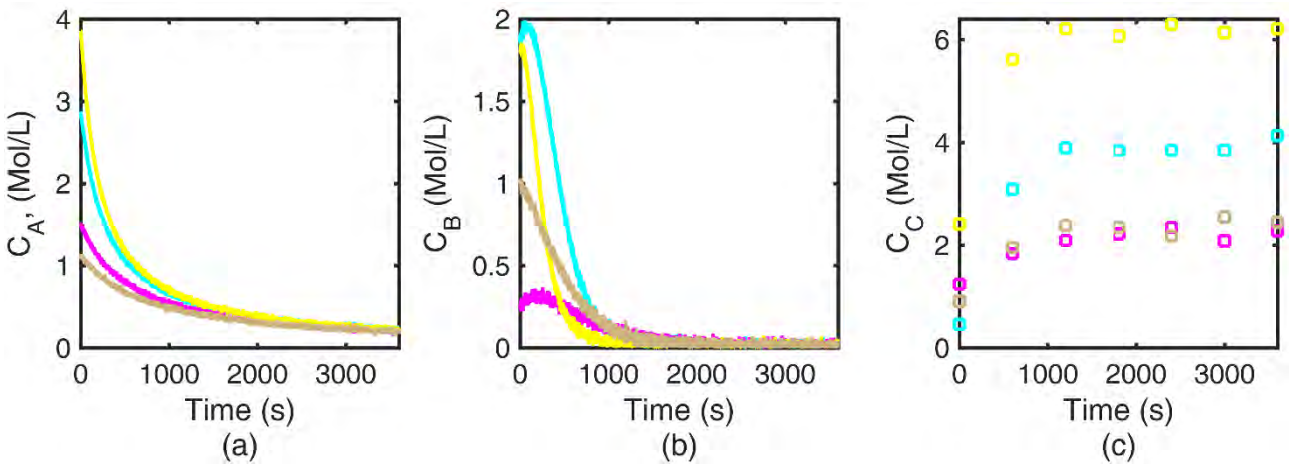

Figure 2. Subset of 4 training batches: (a, b) measured noisy online data, and (c) offline data.

The Hammersley sequence is a sampling technique for computer experiments which is used in this example to find a suitable set of the initial values combinations that uniformly cover the modeling space, in order to collect -as much as possible- information about the process dynamics when it is initiated departing from different initial settings. However, in the real situations (e.g. the pilot plant illustrating, section 4), this welfare of selecting the training data might not be feasible. So, it will be required to take the maximum advantage of the available historical data recorded and measured form the batch process (previous runs).

A moving average technique with a time window of 50 seconds has been used to smooth the online data (C_A, C_B) through all the batches (training and validation) in order to diminish the effect of the artificial (simulated) white noise introduced in the measures, see the zoom-out view in Figure 3(c). After that, 7 input-output (online-offline) training points have been collected from each one of the 24 training batches. Figure 3(a, b) shows the collected input-output (online-offline) data from two training batches: the solid blue and red lines are the measured online data (C_A, C_B) and the dashed black lines correspond to their smoothed values, while the green squares represent the values of the measured offline variable C_C. Hence, the circles, diamonds and squares are the collected input-output data from each batch.

The number of the training batches is selected based on a try and cut procedure, in order to roughly find the minimum number of batches that is able to provide a high prediction accuracy of the soft-sensor. During this try and cut procedure, two main principles are considered: first, the number of the training batches should emulate real situations, where often few information about such kind of processes is available; second, the number of the training batches should be related to the initial conditions variation domain, and also to the complexity of the process behavior. On the contrary, a much higher number of validation batches is considered in this case-study since their generation is feasible (simulation runs) and this allows to get a precise evaluation of the developed soft-sensors, through assessing their prediction accuracy at each sub-region of the initial condition variation domain.

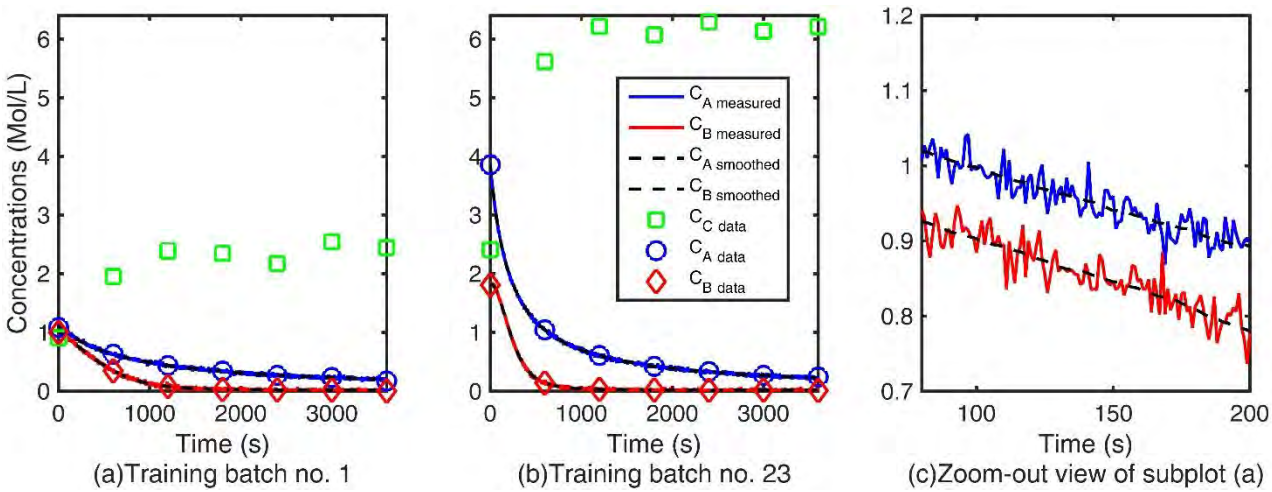

Figure 3. (a ,b) online data smoothing and input-output data collections from training batches no. 1 and no. 23, respectively, (c) Zoom-out view of batch 1.

The collected 168 (7 ×24) input-output data are used to train the soft-sensor in Eq.(19) based on the four metamodel implementations considered (OK, GPR, SVR and ANN). It should be emphasized that the training data does not include any information neither about the batches order/identifications nor the temporal sequence of each specific sample. The training task is accomplished according to the requirements of each metamodel type as previously described in section 2. As previously mentioned in part 2.2., a traditional try and cut procedure has been used to select the best ANN configuration. So, different ANN adjustments have been trained with the 24 training batches and validated with the 100 validation batches, and the adjustment that achieved the best prediction accuracy for the validation batches is selected. The different configurations were in terms of the number of layers (from one to two layers), the number of neurons in each layer (from 3 to 10 neurons) and the training algorithm (Levenberg-Marquardt back-propagation or Bayesian regularization back-propagation). After several trials, an ANN with one hidden layer of six neurons, trained using the Levenberg-Marquardt back-propagation method (via the Matlab algorithm "*trainlm*"), is found to exhibit the best tradeoff between the ANN structure simplicity and prediction accuracy. On another side, a linear kernel function $\Phi\ (x_i, x_j) = x'x$ has been shown to be the best kernel type for the SVR-based soft-sensor.

$$C_C(t) = f[C_A(t), C_B(t), C_A(t=0), C_B(t=0), C_C(t=0)\ ] \qquad (19)$$

After the training of the soft-sensors, they are used to predict the C_C behavior of the 100 validation batches, using only the noisy initial conditions and the values of the smoothed online variables of each batch. It is also worth to emphasize that, although the soft-sensors are trained using noisy measurements of the offline variable C_C (Figure 2(a, b)), their performance assessment is achieved via comparing their predictions with the ideal exact/theoretical behavior of *Cc*.

The quality of the soft-sensors predictions is evaluated numerically through the calculation of several accuracy measures including the RMSE (Eq.(14)), the NRMSE (Eq.(15)) and the CC (Eq.(16)). It's worthy to mention that usually the aforesaid accuracy measures are calculated by comparing the predictions with the available noisy measurements. But then, a conceptual dilemma arises, as these data involve random deviations accumulated from human operation and analysis apparatus errors. Hence, the metamodel prediction is intended to match the experimental data but also to compensate these errors so avoiding over-fitting. In real situations, the aforementioned manner is the only possible way to calculate an accuracy measure, since the real behavior of the system is unknown. But since we are dealing with a simulation case, we may take advantage of our knowledge about the process theoretical/exact behavior. So, the corresponding accuracy measures are calculated by comparing the metamodels prediction with the process theoretical/ideal behavior, as well with the C_C data emulating the noisy measurements.

The calculation of the accuracy measures with respect to theoretical behavior is much more credible to express the soft-sensor accuracy. However, calculating them with respect to the noisy measurement allows extracting some conclusions that can help to evaluate the soft-sensor performance when applied to a real case-study, where only noisy measurements are available. Table 1 shows the accuracy measures of the soft-sensors predictions including the average the RMSE, NRMSE and the CC. In the case when calculating these measures with respect to the data emulating noisy measurements, the number of evaluation data n_v in Eqs. (14) and (15) is 700 (100 batches × 7 measurements). However, when calculating them relative to the system theoretical behavior - that is known from the simulation with one second time step-, so n_v is 360000 data points (100 batches × 3600 measurements).

The results in Table 1 illustrate how the soft-sensor based on any of the metamodels is able to predict the C_C concentration with high accuracy, as it was able to achieve in the worst case (the soft-sensor based on ANN) a NRMSE of 3.89 % of the total variation range of the C_C [0 : 6.3], and a correlation of 0.983. More importantly, the table reflects that the accuracy measures calculated with respect to the system theoretical behavior are better than those calculated with respect to the noisy experimental data. This indicates that the developed soft-sensors are able to identify the process underlying behavior, although they have been trained using noisy measurements, and these soft-sensors are not over-fitting the training data. It is worth to refer that the soft-sensors based on the SVR and the OK have achieved the best prediction accuracy (average NRMSE of 2.56 % and 2.39 % respectively).

Table 1. Average RMSE, NRMSE and CC of the batch reactor soft-sensors.

	W.R.T. the noisy measurements			W.R.T. the known exact behavior		
	RMSE	**NRMSE (%)**	**CC**	**RMSE**	**NRMSE (%)**	**CC**
OK	0.21	3.34	0.988	0.16	2.56	0.992
GPR	0.27	4.39	0.980	0.23	3.80	0.984
SVR	0.20	3.22	0.989	0.15	2.39	0.994
ANN	0.28	4.52	0.978	0.24	3.89	0.983

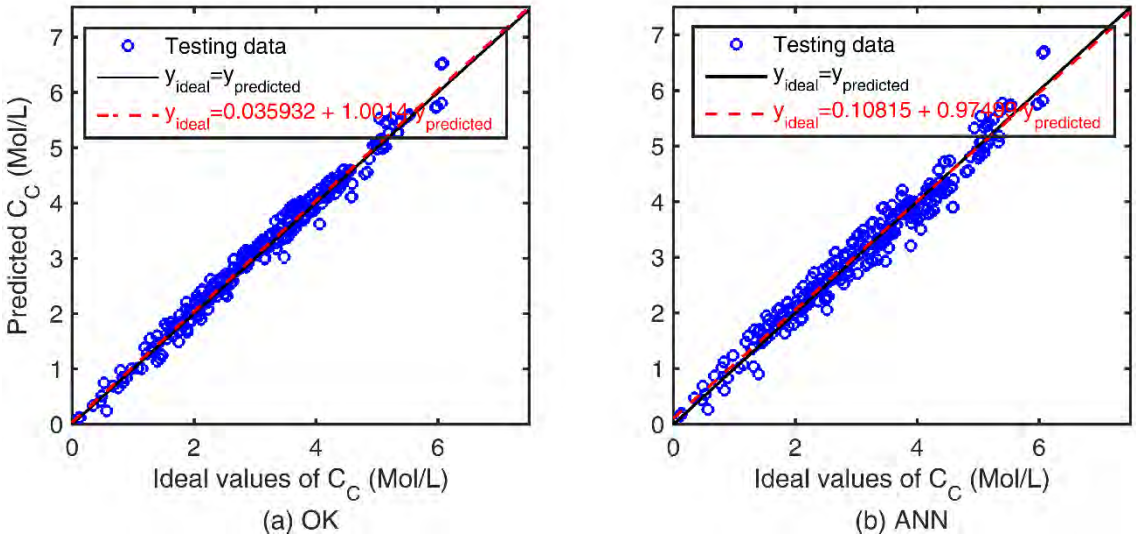

Figure 4. Exact versus predicted values of Cc of a random subset of the 100 validation batches using the soft-sensors based on (a) OK and (b) ANN metamodels.

Figure 4 shows a randomly selected validation subset (300 data points) of the estimated offline variable (C_C(using the OK and the ANN based soft-sensors (best and worst cases), compared to their corresponding theoretical/ideal behavior. The displayed qualitative assessment demonstrates the high quality of the soft-sensors predictions: the prediction linear fit (red dashed line calculated considering the whole validation set - 360000 data points) is very close to the ideal fit (y_{ideal}=$y_{predicted}$; solid black line).

Another qualitative evaluation of the soft-sensors performance is presented in Figure 5, which shows the predictions normalized error distribution for the best (OK) and worst (ANN) soft-sensors; it is clear that the soft-sensors based on the four metamodels behave normally without any biased behavior/ predictions, since the majority (mean) of the predictions exhibit very small error values along with some outliers.

The values in Table 1 represent average or gross measures of the soft-sensors accuracy, so it is necessary to give more deep or detailed sight to the performance of the soft-sensors. Figure 6 shows the estimations of the offline variable (C_C) for the batches with the highest (Figure 6 (a)), average (Figure 6 (b)) and lowest (Figure 6 (c)) prediction accuracy, using the OK (dashed red line) and the ANN (dashed mauve line) soft-sensors. The three batches are selected through the calculation of the NRMSE for each batch independently, using average values with respect to the four soft-sensor types.

Figure 6 demonstrates that the four soft-sensors are able to capture the ideal/theoretical process behavior (continuous black line), without over-fitting the noisy measured behavior/data (green squares), in spite of the deviations introduced to the training data in order to emulate the measurement errors. Much more significant, the figure reflects the ability of the proposed soft-sensors to predict the offline variable (C_C) with high accuracy along different batches operated under different initial settings within the known bounds: the initial conditions [$C_A(t=0)$, $C_B(t=0)$, $C_C(t=0)$] of the selected three batches are [3.01, 1.52, 1.52], [2.68, 0.22, 2.07] and [3.85, 1.95, 2.46] for batch no. 67 (Figure 6. (a)), batch no. 56 (Figure 6 (b)) and batch no. 95 (Figure 6 (c)), respectively.

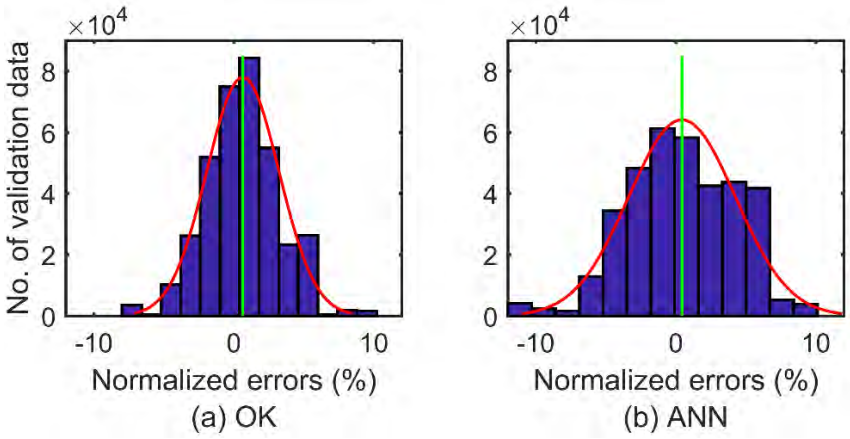

Figure 5. Normalized errors distributions of the OK and ANN based soft-sensors predictions.

Correlating Figure 1 and Figure 6 highlights the effect of the location of the batches initial conditions on the soft-sensor accuracy. For example, the initial conditions of the most accurate predicted batch (no. 67, Figure 1) lay within the middle of the training batches initial conditions, so soft-sensors have sufficient knowledge about the behavior of the process dynamics associated to this local area of the initial conditions. On the contrary, the initial conditions of the batch with least accurate predictions (no. 95, Figure 1) fall far from the main bulk of initial conditions of the training batches (i.e. on the limits of the initial conditions domain). Hence, the soft-sensor includes less information about the process behavior when it departs from this local area of the initial conditions. These conclusions and remarks match with the OK and GPR metamodels main assumptions that distinguish them from any

other metamodel type. This assumption considers that the prediction uncertainty/error increases as the point to be predicted moves far from the training data, and vice versa.

Additionally, the OK (and GPR) estimated errors can be exploited to construct a confidence area that can be used as uncertainty measure about the predictions (Figure 6, grey lines). So, in the cases where the OK predictions behave a significant deviation from the theoretical behavior, these deviations are often accommodated within an estimated confidence area (-/+ 5 σ). The reliability of this estimation (which would correspond to a confidence interval of 99.999% in the case of a normal distribution) depends on the quality of the approach and the application, so other elections are also feasible according to the case characteristics and modeler preferences (consequences of a wrong estimation, risk aversion, knowledge about the system, etc.). In very few batches, the theoretical behavior may fall outside this uncertainty area (Figure 6 (c)).

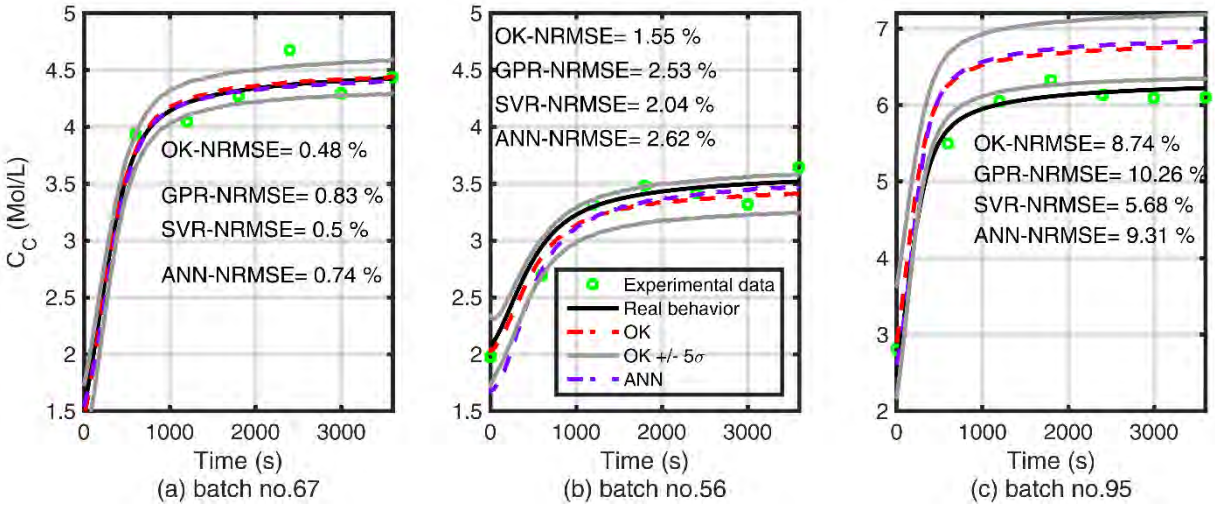

Figure 6. Predictions of Cc for three validation batches.

Finally, it is worth to note that, in this case, both of the online variables and the QIV are concentrations closing the mass balance, so one straightforward procedure to estimate the missing information ($C_C(t)$) in this case would be to exploit this knowledge about the system (analytical model-based soft-sensor). Since the proposed soft-sensors have been trained in a behavior which accomplishes the mass conservation principle, the results would be finally equivalent, but it should be emphasized that the proposed soft-sensors are not based on the exploitation of any "a priory" knowledge, but on the approximation of the latent relationships between the offline variable and other online variables regardless of the nature of this knowledge and the involved variables (i.e. concentrations, temperatures, pressures, etc.).

Hence, a modified version of this case-study is addressed, which is exactly the same as the original one but assuming that no information is available about the online concentration $C_B(t)$. Thus, the soft-

sensors should infer the offline values $C_C(t)$ only from the online variable $C_A(t)$ and the initial conditions of all the variables, i.e. $C_C(t) = f[C_A(t), C_A(t=0), C_B(t=0), C_C(t=0)]$. In this way, no mass balance can be closed to predict/validate the required offline concentration $C_C(t)$.

Table 2 displays the validation results of the soft-sensors resulting from this modified version of the case-study. Although the comparison with Table 1 reveals a slight reduction in the soft-sensors accuracy (reduced information about the process behavior), the results still indicate a very good accuracy even when the concentration $C_B(t)$ is not available. The same subset of the off-line variable predictions previously represented in Figure 4 are now displayed in Figure 7 for the new soft-sensors, confirming the same previously drawn conclusions (again, the liner fit equation has been calculated considering the entire validation set).

Table 2. Average RMSE, NRMSE and CC of the batch reactor soft-sensors (modified case).

	W.R.T. the noisy measurements			**W.R.T. the known exact behavior**		
	RMSE	**NRMSE (%)**	**CC**	**RMSE**	**NRMSE (%)**	**CC**
OK	0.33	5.27	0.9700	0.31	5.09	0.9701
GPR	0.33	5.28	0.9699	0.31	5.09	0.9701
SVR	0.45	7.23	0.9458	0.37	7.26	0.9419
ANN	0.34	5.48	0.9690	0.32	5.18	0.9706

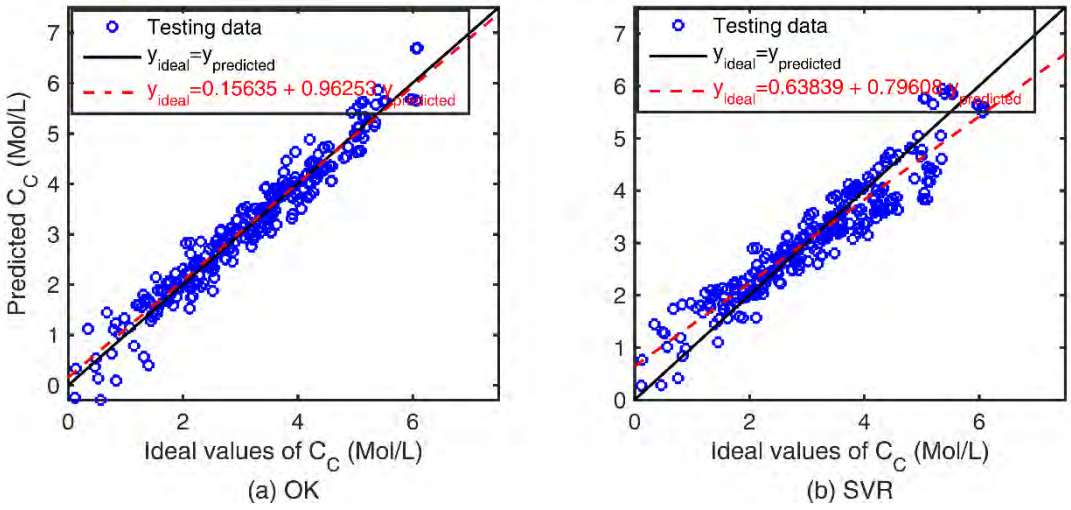

Figure 7. Exact versus predicted values of Cc of a random subset of the 100 validation batches using the soft-sensors based on (a) OK and (b) SVR metamodels (modified case).

The application of the proposed soft-sensing methodology to the modified version of the case-study also highlights the robustness and tuning flexibility of the OK compared to the other considered

techniques. The comparison between Table 1 and Table 2 reveals that the OK has been able to maintain its high prediction accuracy level, while this is not the case of the SVR soft-sensors. This is due to the challenging task of selecting the best SVR kernel through a try and error procedure each application time, which does not guaranty the optimal selection, while in the case of using OK –as well as the GPR- parameters are automatically optimized, which ensures an optimal fitting through repetitive applications considering different soft-sensor structures and also different training data.

3.2. Fed-Batch Fermenter for Penicillin Production

A fed-batch fermentation process for the production of Penicillin is used as a second simulation case-study. The process mechanistic model had been initially developed by Bajpai and Reul (1980), and since this time, it has become a very popular benchmark case-study for the dynamic optimization and control studies, due to its high nonlinearity (Cuthrell & Biegler, 1989; Dadebo & Mcauley, 1995; Banga et al., 2005; Wang et al., 2017). The analytical model in Eqs.(20) describes the relations among the process variables, including the concentrations (g/L) of the biomass B, penicillin P and substrate S inside the reactor, the reactants volume V (L) and the substrate inlet flowrate F (g/h).

This case represents relative further challenges to the proposed soft-sensing method in terms of the higher dimensionality, higher nonlinearity (see the process model in Eqs.(20)) and, more important, the process kind (fed-batch), since the existence of the external input (F) generates a more complicated dynamic behavior of the system.

$$\frac{dB}{dt} = h_1 B - F\left(\frac{B}{500\,V}\right)$$

$$\frac{dP}{dt} = h_2 B - 0.01P - F\left(\frac{P}{500V}\right)$$

$$\frac{dS}{dt} = -\frac{h_1 B}{0.47} - h_2\frac{B}{1.2} - B\left(\frac{0.029S}{0.0001+S}\right) + \frac{F}{V}\left(1 - \frac{S}{500}\right) \qquad (20)$$

$$\frac{dV}{dt} = \frac{F}{500}$$

Where $h_1 = 0.11\left(\frac{S}{0.006B+S}\right)$, $h_2 = 0.0055\left(\frac{S}{0.0001+S(1+10S)}\right)$

During each batch run of 150 minutes, the product (penicillin) concentration P is treated as the offline variable that is measured via expensive offline sampling and analysis (one sample every 15 minutes). Meanwhile, the biomass concentration B, substrate concentration S, the reactants volume V, and the

substrate inlet flowrate F are assumed to be the online variables, measured and registered by the sensors every second. Also white noise distributed according $\mathcal{N}(\mu = 0, \sigma = 0.2)$, $\mathcal{N}(\mu = 0, \sigma = 0.15)$, $\mathcal{N}(\mu = 0, \sigma = 0.05)$ and $\mathcal{N}(\mu = 0, \sigma = 0.05)$ are added to the simulated measurements of the online variables B, S, V, and F, respectively. Besides, a higher experimental error of $\mathcal{N}(\mu = 0, \sigma = 0.2)$ is added to the simulated measurement of the offline variable P. The noise amounts are calculated as mentioned in the previous case.

The nominal initial conditions often used in the literature for this case are $B(t = 0), P(t = 0), S(t = 0), B(t = 0) = [1.5,\ 0,\ 0,\ 7]$ but, for the purpose of this study, the initial conditions are allowed to vary within the limits $[B_{min}^{t=0}: B_{max}^{t=0},\ P_{min}^{t=0}: P_{max}^{t=0},\ S_{min}^{t=0}: S_{max}^{t=0},\ V_{min}^{t=0}: V_{max}^{t=0}] = [0.5\ \ 10;\ 0\ \ 3;\ 0\ \ 10;\ 5.0\ \ 8.5]$ to simulate a situation in which several settings of the batch are feasible, assuming the unavailability of a first principle model.

A set of 24 noisy batches are simulated for training of the soft-sensors, while other 100 batches are also simulated for validation purposes. As in the first studied case (section 3.1), the noisy data (a subset of them are shown in Figure 9) are used to train the soft-sensor, which is then used to predict the Penicillin concentration P of the validation batches. Then, the predicted P values are compared to their corresponding ideal ones.

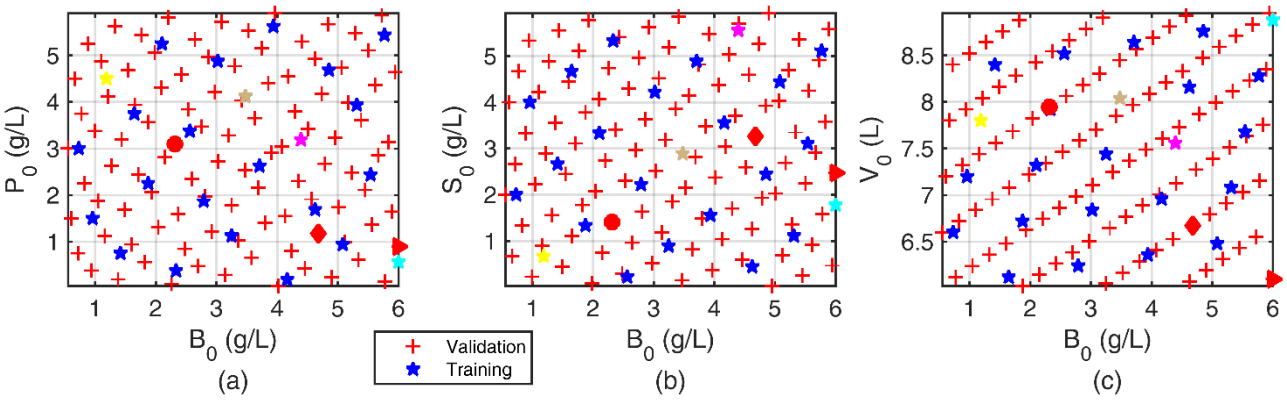

Figure 8. Initial conditions of the training (stars) and validation (crosses) batches of the penicillin production case-study.

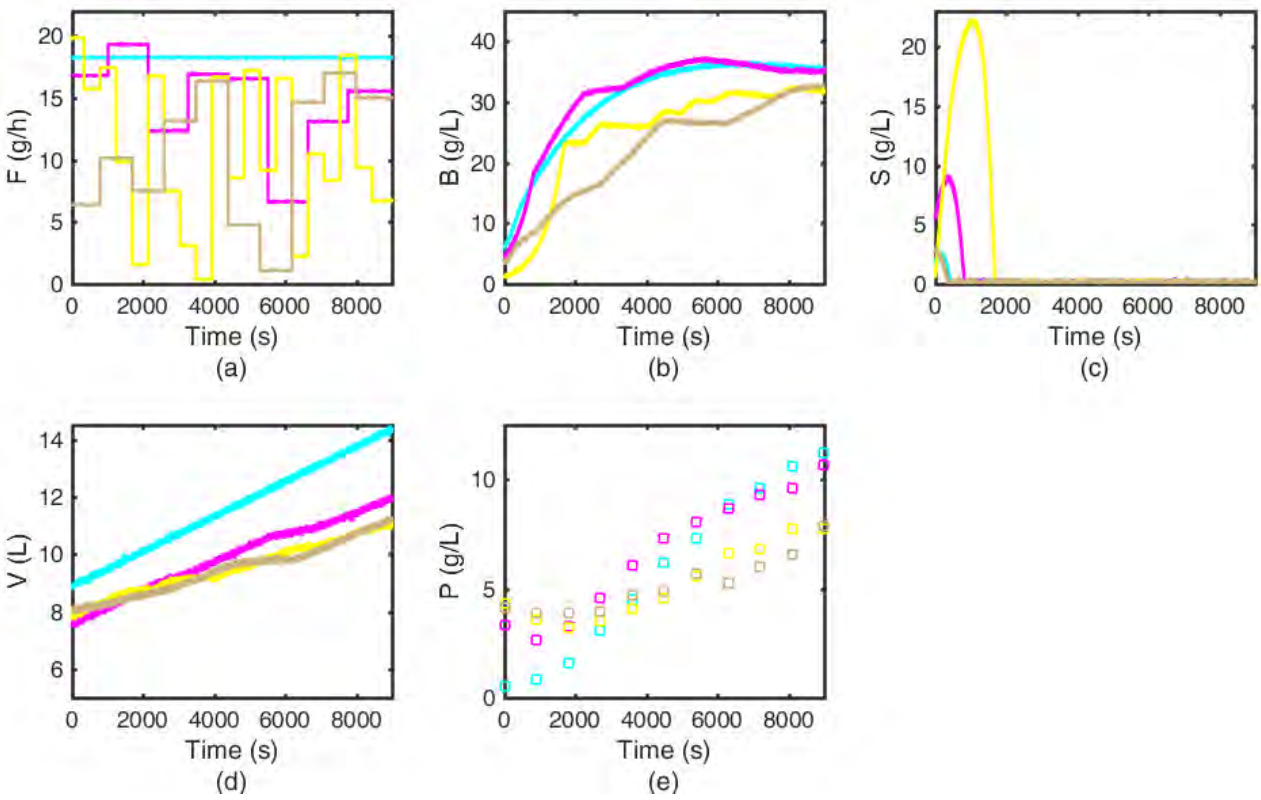

Figure 9. Case 2: Subset of 4 training batches: (a, c, d, e) measured noisy online data, and (b) offline data.

The inlet flowrate profiles (F) associated to the training and validation batches are also synthesized in a random manner in order to ensure that the training will include as much information as possible about the dynamic behavior of the process. Thus, each flowrate profile is characterized by random step changes (piecewise constant) along the batch time. A subset of the training batches are represented in Figure 9, where the effect of the variations in the initial conditions and the inlet flowrate profile F can be observed in terms of significant changes in the process behavior from one batch to another. The initial conditions of this subset of the training batches are also highlighted in Figure 8 with the corresponding colors. Figure 10 (b) illustrates the inlet flowrate profiles associated to the three specific validation batches previously marked in Figure 8 with a red solid circle, a diamond and a triangle, respectivly.

A moving average technique with a time window of 120 seconds is used to manage the (artificial) white noise associated to the inputs (simulated physical sensors measurements), see Figure 11 Then, ten input-output training points have been collected from each one of the 24 training batches (Figure 11). The soft-sensors based on the four metamodel implementations considered (OK, GPR, SVR and ANN) are trained using the collected 264 (24×11) input-output training data. Again, no knowledge about the batch identifications or the temporal sequence of the measurements is introduced during the training. The same SVR kernel type and ANN configurations – but with a Bayesian regularization

back-propagation algorithm, via the Matlab algorithm "*trainbr*"- used in the previous example have been found to be the most suitable customizations in this case too.

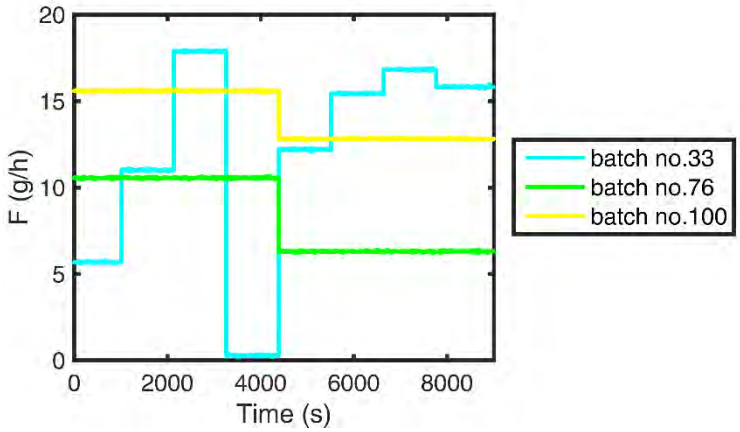

Figure 10. Case 2: Profiles of the inlet flowrate F for the three validation batches.

Table 3 displays the numerical/quantitative assessment of the Penicillin soft-sensors prediction accuracy in terms of the RMSE (Eq.(14)), the NRMSE (Eq.(15)) and the CC (Eq.(16)), as in the previous case (n_v=1100 samples (100 batches × 11 measurements) when compared to noisy measurements, and n_v=900000 samples (100 batches × 9000 measurements) when ideal behavior is used as reference).

$$P(t) = f[B(t), S(t), V(t), F(t), B(t=0), S(t=0), V(t=0), F(t=0), P(t=0)] \qquad (21)$$

Again, the results in Table 3 reflect the high accuracy of the Penicillin soft-sensors, since they were able to achieve in the worst cases (the soft-sensor based on ANN) a NRMSE of 5.9 % of the total variation range of the P (0 : 13.75), and a Pearson coefficient of 0.94. These accuracy measures are not as good as the ones reported in the previous example (Table 1) because of the relatively higher nonlinearity and dimensionality of the Penicillin process/model (Eq.(20)). Besides, this case poses an additional significant challenge to the proposed method compared to the previous example, due to its nature as a fed-batch process with external forcing input (flowrate F).

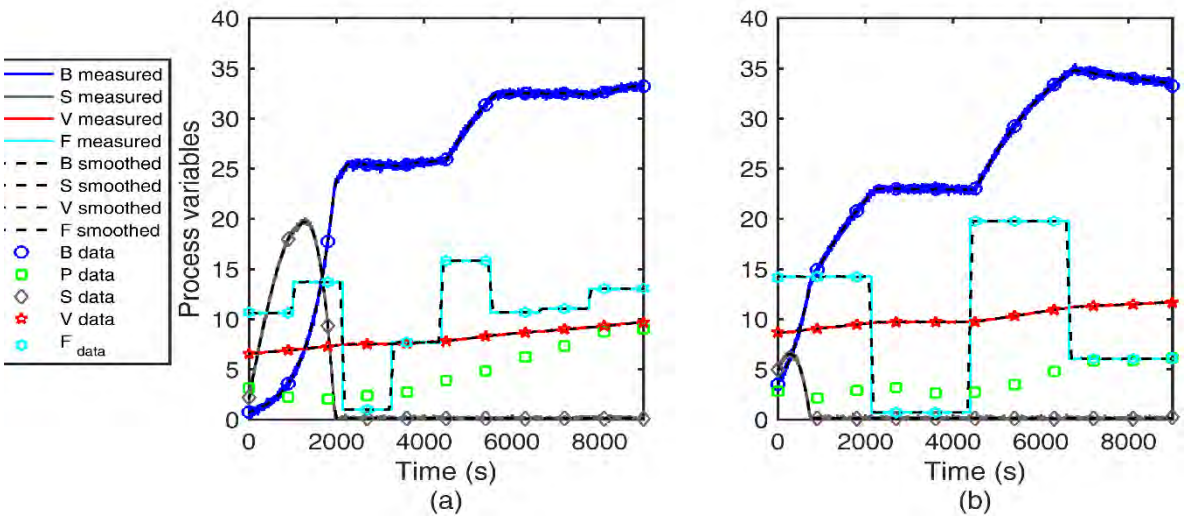

Figure 11. Input-output data collection and smoothing for two different training batches (Penicillin case).

The conclusions drawn from the previous example regarding the capabilities of the proposed soft-sensors to identify the process underlying behavior, even though they have been trained using noisy measurements, are also confirmed in this case. Besides, the results indicate the ability of the proposed soft-sensor to handle fed-batch processes. Table 3 also highlights the performance of the OK based soft-sensor, which is able to achieve highest prediction accuracy than the SVR and ANN based systems.

Figure 12 displays a subset (300 data point) of the estimated Penicillin concentrations compared to their ideal values, using the soft-sensors based on the OK (most accurate soft-sensor) and ANN (least accurate soft-sensor), and represents a qualitative assessment of the soft-sensors accuracy. Also the prediction linear fit is calculated considering the whole validation set (900000 data point). On another side, Figure 13 shows the predictions normalized error distribution of the validation set; again, homogenous normal distributions of the soft-sensors prediction errors can be clearly identified, which reflects the soundness and unbiasedness of the soft-sensors performances.

Table 3. Average RMSE, NRMSE and CC of the Penicillin concentration soft-sensors.

	W.R.T. the noisy measurements			W.R.T. the known exact behavior		
	RMSE	NRMSE (%)	CC	RMSE	NRMSE (%)	CC
OK	0.51	3.91	0.9773	0.49	3.69	0.9778
GPR	0.52	4.00	0.9762	0.50	3.76	0.9771
SVR	0.77	5.84	0.9480	0.71	5.33	0.9527
ANN	0.81	6.16	0.9445	0.81	5.99	0.9417

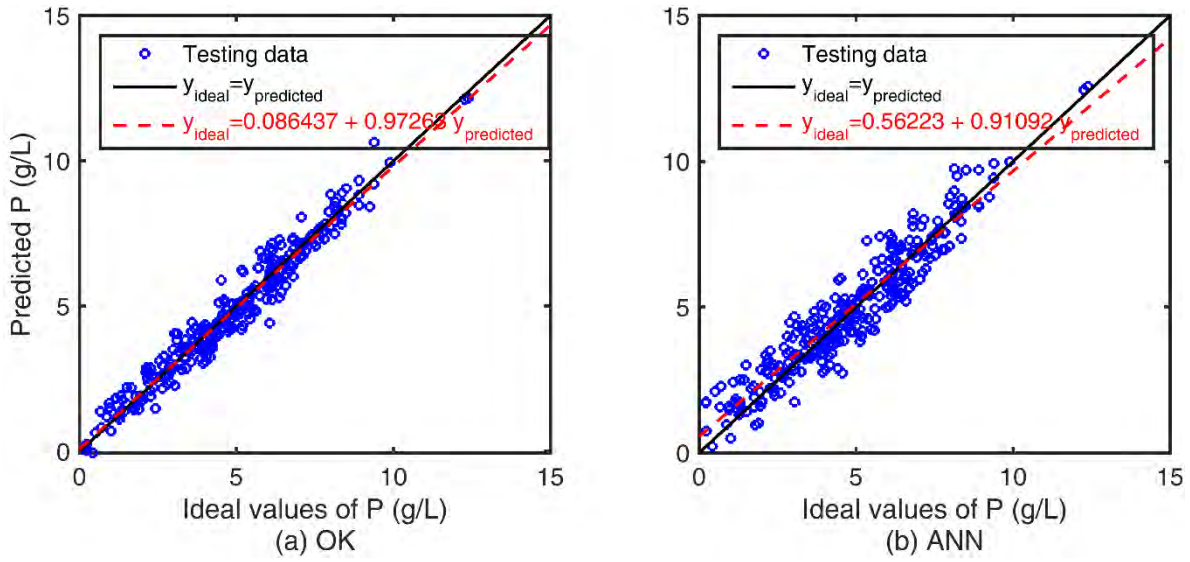

Figure 12. Predictions of a subset of the 100 validation batches (Penicillin case).

Figure 14 shows the predictions of the Penicillin concentration (P) for three validation batches with the highest (a), average (b) and lowest (c) prediction accuracy, using the OK and the ANN soft-sensors. The three batches are selected in the same way as the previous case in section 3.1.

The figure also indicates the high accuracy of the soft-sensors and their ability to continuously (each second) capture the underlying behavior of the Penicillin concentration (continuous black line). Additionally, it confirms the high capability and flexibility of the soft-sensors to predict different dynamic behaviors of the Penicillin concentration, associated to the changes in the initial conditions [$B(t=0)$,$P(t=0)$,$S(t=0)$,$V(t=0)$] (see Figure 8 red solid circle, diamond and triangle, respectively), and - more importantly- different forcing input F profiles, see Figure 10 (b).

Again, when the initial conditions of a validation batch lie very near to or within the main bulk of the training batches initial conditions (batch no. 33, red circle in Figure 8) the results exhibit the highest

prediction accuracy, as the soft-sensor is trained with sufficient information or knowledge about the process dynamics associated to this local area of the initial conditions. In contrary, the initial conditions of batch no. 100 lay far from the main bulk of the training batches initial conditions –on the limits of their domain- (Figure 8, red triangle). Therefore, the soft-sensor has less data/information/knowledge about the process behavior when departing from this local area of the initial conditions. These observations match to the OK and GPR metamodels main principle: as the predicted point moves far from the training data, the prediction error increases.

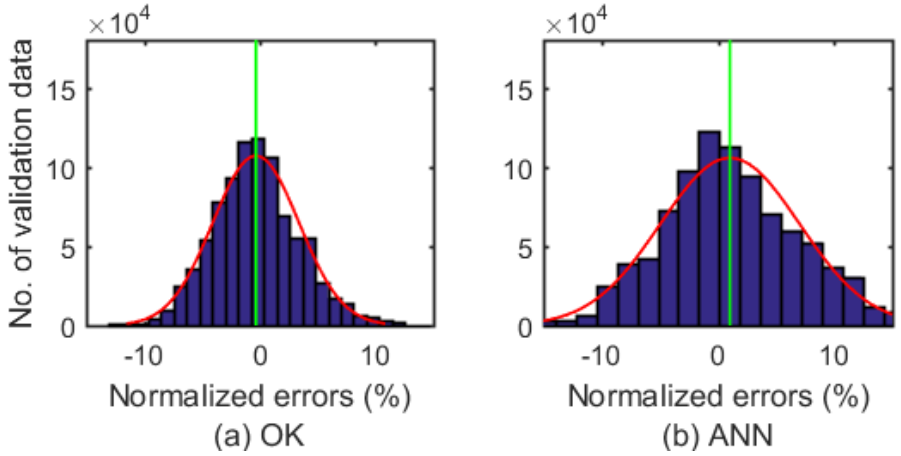

Figure 13. Normalized errors distributions of the Penicillin soft-sensors predictions: (a) OK and (b) ANN.

The figure also emphasizes the merit of the OK confidence area (grey lines) that is established around its prediction thanks to its estimated error. Again, even in the cases where the OK predictions –as well the GPR ones- behave a significant deviation from the theoretical or ideal behavior, these deviations often fall within the confidence area (-/+ 5 σ). Similar to the previous case (part 3.1), Figure 8 and Figure 14 also show that the OK uncertainty area– i.e. distance between the two grey lines– of a batch increases as its initial conditions goes far from the main bulk of the training batches initial conditions.

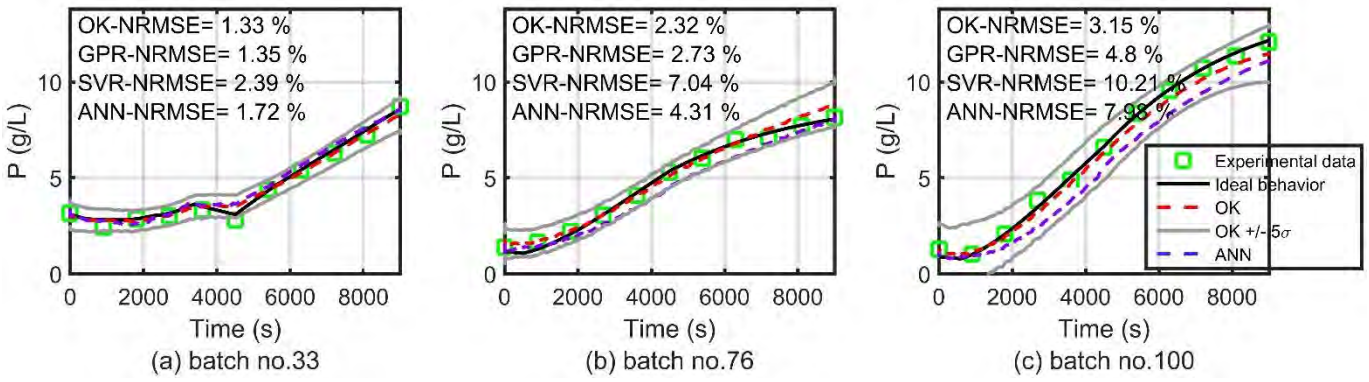

Figure 14. Prediction of the Penicillin concentrations for three validation batches: (a) highest (b) average, and (c) lowest prediction accuracies.

Another very important point that has been raised by the nature of this case-study (fed-batch process) is that the soft-sensor (Eq.(21)) is able to accurately predict the Penicillin concentration $P(t)$ just considering the current values of the control variable -substrate inlet flowrate- $F(t)$ although, in fact, the Penicillin concentration $P(t)$ depends also on the dynamic evolution of the inlet flowrate (i.e. $F(t-1)$, $F(t-2)$, $F(t-3)$ etc.). This can be explained by the fact that the soft-sensor approximates the Penicillin concentration $P(t)$ not only as a function of the control variable $F(t)$, but also as a function of other measured online state variables [$S(t)$, $B(t)$, $V(t)$] whose current values are also dynamically affected by the inlet flowrate. In other words, the values of these online state variables include enough knowledge about the accumulation/history of the inlet flowrate from the beginning of the batch time to build good estimations of the QIV.

In order to better illustrate this aspect, a modified form of this case has been studied, where the online variables $F(t)$ and $S(t)$ have been assumed to be unmeasured and the initial values $F(t=0)$ and $S(t=0)$ are also assumed to be unknown. Consequently, the modified soft-sensors take the form: $P(t) = f[B(t), V(t), B(t=0), V(t=0), P(t=0)]$.

Table 4 and Figure 15 show how the soft-sensors built for the modified case still maintain very high prediction accuracy, even with this dramatic reduction of the available information about the process. Also, the results underline again the consistent and robust performance of the OK approach in front of the oscillating accuracy of the ANN and SVR methods (Table 3 vs. Table 4).

Table 4. Average RMSE, NRMSE and CC of the modified soft-sensors of the Penicillin concentration.

	W.R.T. the noisy measurements			W.R.T. the known exact behavior		
	RMSE	NRMSE (%)	CC	RMSE	NRMSE (%)	CC
OK	0.58	4.46	0.9719	0.55	4.19	0.9733
GPR	0.68	5.24	0.9607	0.66	5.03	0.9604
SVR	0.76	5.82	0.9483	0.69	5.23	0.9542
ANN	0.60	4.62	0.9703	0.58	4.40	0.9708

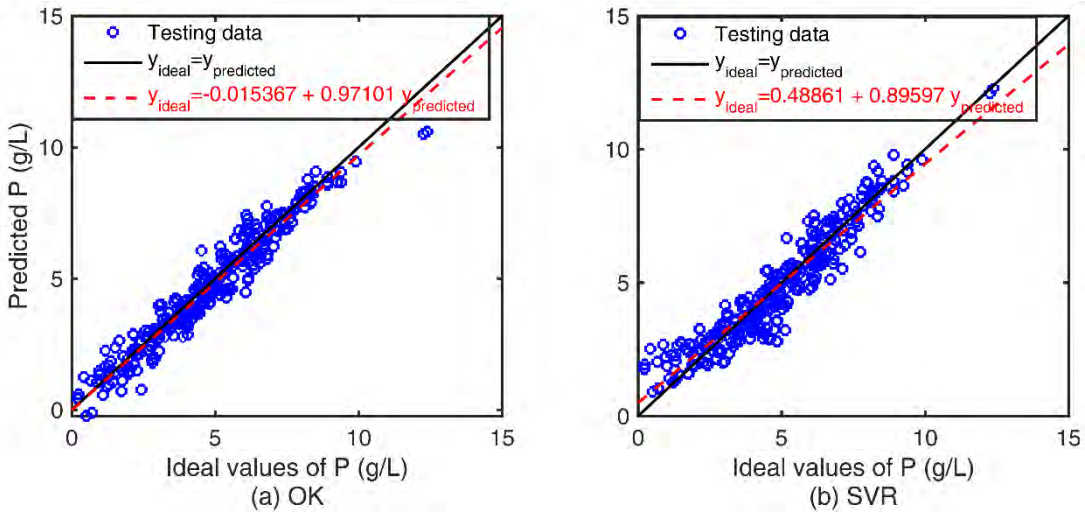

Figure 15. Predictions of a subset of the 100 validation batches (Penicillin case, modified soft-sensors).

3.3. Accuracy assessment

Excluding just one outlier case (SVR soft-sensor in Table 2), the accuracies (NRMSE) of the soft-sensors proposed in this section are in the range [2.5% - 6.0%], and the corresponding Correlation Coefficients are in the range [0.94 - 0.99] (Tables 1, 2, 3, and 4).

As previously indicated, these values are to be compared with the accuracy of the information used to train the soft-sensors, which in these cases is known, since the disturbances in this information have been simulated by introducing white noise in the information provided by the mathematical models: the variance of such noise has been fixed to be 1.5% of the variability range of the offline variables, and 0.5% of the variability range of the online variables. Additionally, the NRMSE values have been calculated in front of 2 references: the exact values (provided by the mathematical model), and the disturbed values, again affected by the corresponding white noise (1.5%).

So the obtained NRMSE values (2.5% - 6.0%) are in accordance with the target accuracy that can be expected from a soft-sensor, as it has been previously discussed in this section: lower but within the same order of magnitude of the accuracy of the offline information used to train these soft-sensors (1.5%), further reduced by the accuracy of the online information available to use them (0.5%).

Besides, the proposed soft-sensors offer better or equivalent accuracies than the ones offered by other soft-sensing approaches recently proposed in the literature for continuous and batch processes: although the available information is presented in different forms and usually the resulting accuracies are not referred to the accuracy of the raw information, common CC results are in the range [0.84 –

0.95] (Grbic et al., 2013; Jin et al., 2015), and values of RMSE of 3.35 (Mota et al., 2014) and average relative errors of 15% (Duran et al., 2006) are reported.

4. Application to a Photo-Fenton batch process

Previous studies have been devoted to model waste water treatments based on the photo-Fenton processes. Some of these works have proposed the use of analytical or First Principle Models (FPMs) for the characterization of these processes in particular experimental conditions, or for specific reactor geometries (Farias et al., 2009; Reina et al., 2012). However, the complex and nonlinear natures of these processes make such FPMs unable to include all the involved mechanisms in a general exhaustive way.

Alternatively, many studies have been carried out focusing on the data-based modeling of different types of treatments based on photo-Fenton processes, mainly using ANNs. However, they have been oriented to address the kinetic modeling from experimental design perspectives: the ANN is used to model a measure of the treatment performance (the total removal of contaminant, the final H_2O_2 concentration, etc.) at the final time of the treatment as a function of the working conditions (Nascimento et al., 1994; Duran et al., 2006; Guimarães et al., 2007; Jaafarzadeh et al., 2012; Khataee et al., 2014; Hassani et al., 2015; De Tuesta et al., 2015; Expósito et al., 2017; Belkacem et al., 2017). Thus, the obtained ANN model can be only used for the identification and analysis of the impact of the experiment initial settings on this final measure of performance.

A different class of data-based modeling studies has been focused on the prediction of the contaminant degradation dynamics during the treatment time (Göb et al., 1999; Salari et al., 2005; Elmolla et al., 2010; Ayodele et al., 2012; Mota et al., 2014; Mustafa et al., 2014; Gazi et al., 2017; Sebti et al., 2017). To do so, the degradation evolution is assumed to be a function of the initial experiment parameters (e.g. initial concentrations of H_2O_2, Fe^{2+}, etc.) and the time. But this approach does not allow monitoring an online process, since any change or disturbance affecting the process cannot be tracked or captured by the model, which only considers the time and not any dynamics nor the rest of the online measurements. Besides, the modeling approaches proposed in the majority of these works have not been systematically validated, neither through the comparison of different modelling techniques nor by using other test case(s). Finally, the data used in most of these works have been collected from the experimentations on a laboratory scale (reactors or a flasks of 500 to 2000 mL), which makes the obtained information/data much more smooth/ideal than in an industrial-size situation (it is easier to control the random error within this scale than within a pilot plant/ industrial scale).

4.1. Materials and methods

In this work, a pilot plant case-study is addressed, consisting on a photochemical pilot plant built to study water treatment processes based on the photo-Fenton reaction, working in a batch mode and considering paracetamol as reference contaminant.

Paracetamol (acetaminophen or 4-amidophenol, PCT from this point) is a widely used analgesic, anti-inflammatory and antipyretic, reported as the most popular non-opioid analgesic sold in Spain in recent years (Martínez Bueno et al. 2012). Due to this, and also due to the difficulties to eliminate it through conventional waste water treatment techniques, it has been widely used to investigate the efficiency of non-conventional waste water treatment processes, like the photo-Fenton technologies.

Experiments were carried out using 98 % purity PCT to prepare samples in distilled water. Fenton reagents of H_2O_2 33 % w/v and $Fe_2SO_4 \cdot 7H_2O$ were used (Yamal-Turbay et al., 2015). The process performance (reaction progress) must be evaluated through an off-line procedure consisting on withdrawing aliquots from the pilot plant reactor and measuring the Total Organic Carbon (TOC) concentration by means of a TOC analyzer, which offers an accuracy of 1% of the measurement range, although the overall analysis procedure includes several manual steps which obviously affect the accuracy of the final obtained information.

4.2. Pilot plant

The photochemical pilot plant (Yamal-Turbay et al., 2015) (Figure 16) consists of an annular photo-reactor equipped with a Philips Actinic BL TL-DK 36W/10 1SL lamp and a pumping system set to keep a constant recirculation. The total volume of the system was 15 L, pumped at 12 $L\ min^{-1}$ to guarantee proper mixing.

The processing conditions that were kept constant for all the experiments were: 90 minutes process time, irradiation, 3±0.2 pH, 10 $mg \cdot L^{-1}$ iron(II) salt .

During every batch run (each of 90 minutes), samples of the reaction mixture were taken out at regular time intervals (every 15 minutes). Then, the reaction progress is measured by expensive offline analysis of the extracted samples, which resulted in only seven measurements (TOC and H_2O_2 evolution) available for each batch run.

Additionally, the SCADA system automatically registers the Temperature (T) and Redox potential (R), which are also expected to be related to the reactions progress. Both variables were measured online and recorded every second at a minimum cost, providing 5400 measurements along each batch run. The accuracy of the respective sensors is of 0.5% of the measuring ranges. It should be noticed that the use of Redox potential as online variable implicitly allows to take into consideration one of the most

significant factors in photo-Fenton process, such as the Fenton reagent ratio with respect to the amount of contaminant ($TOC/H_2O_2/Fe^{2+}$).

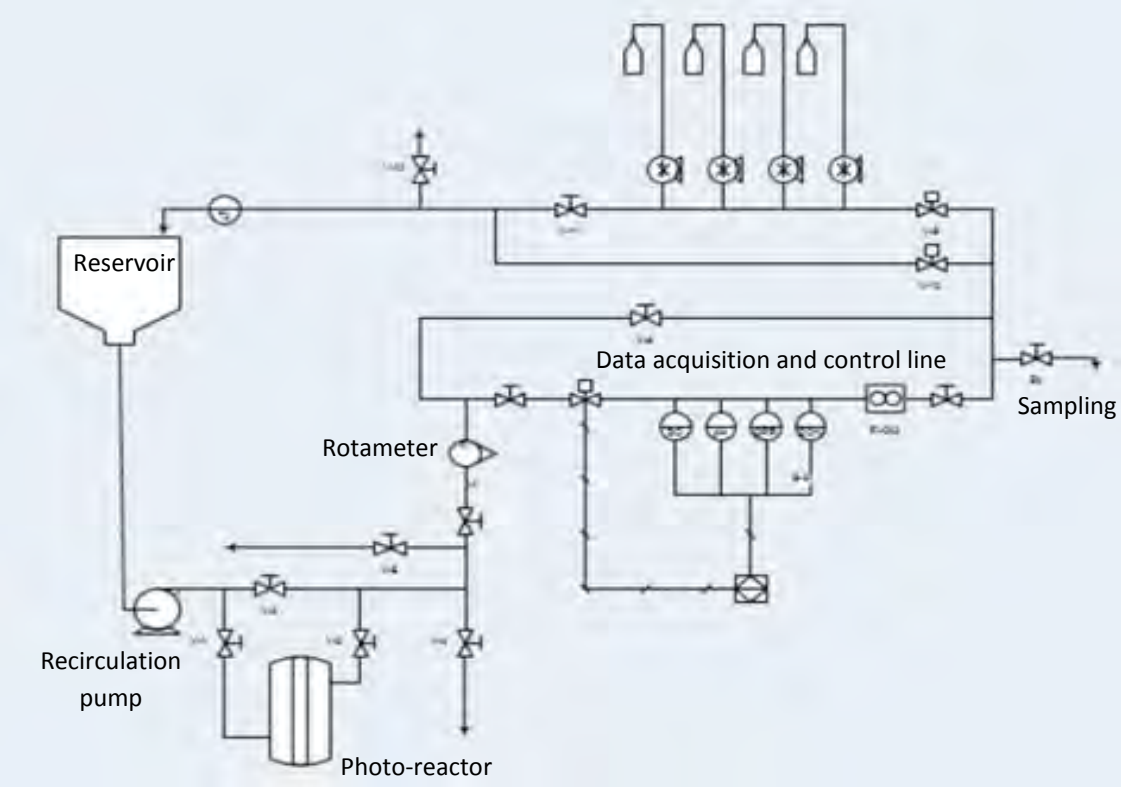

Figure 16: Photo Fenton pilot plant.

Several batch runs have been carried out under different initial concentrations of the contaminant (TOC) and Hydrogen peroxide [H_2O_2], in order to characterize the process behavior and to determine the significant process factors (Figure 17). Some of these batches are carried out considering the same initial conditions (highlighted by dotted gray circles), in order to minimize the noise effects resulted from human errors in the training set, and also to assess the sensitivity /robustness of the metamodel in the validation set for confirmation. 7 batches have been used as training batches, while the soft-sensor performance will be confirmed using other 4 validation batches (Figure 17); the whole data recorded (online) and measured (offline) for the eleven batches are displayed in Figure 18 where each color express a different batch.

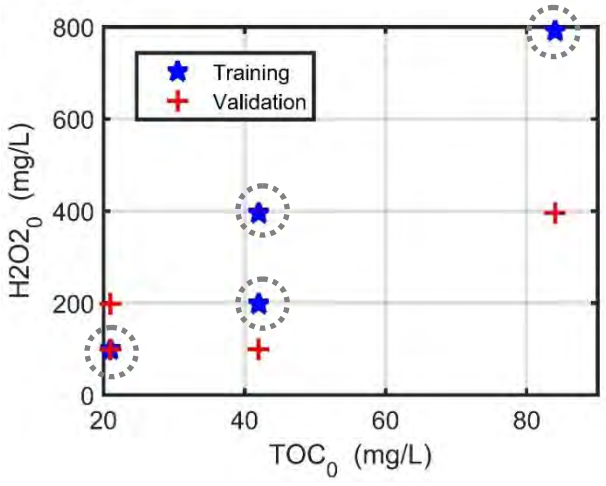

Figure 17. Initial concentrations of the H_2O_2 and TOC of the eleven batches.

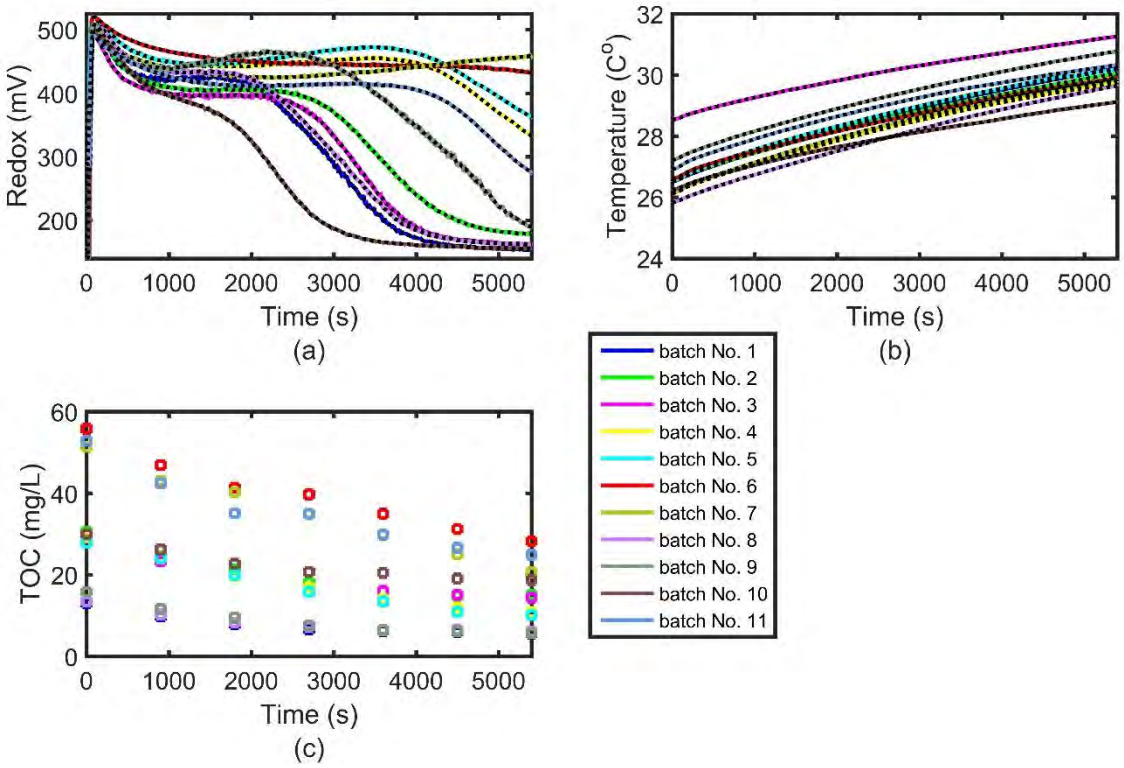

Figure 18. Online and offline data of all the batches.

The same procedure that has been presented and validated in section 3 is directly applied here. Thus, the soft-sensor (Eq.(22)) is designed to model and predict the expensively measured offline TOC for any initial batch settings of the TOC and H_2O_2 as a function of the initial values of the all variables (TOC (t=0), T(t=0), R(t=0)) and the timely measured values of the available online variables (T(t), R(t)). The online data (T, R) are smoothed using a moving average technique with a time window of 60 seconds. Then, seven input-output training points are collected from each batch of the training group, which results in a training set of 49 input-output data. The soft-sensor (Eq.(22)) is then trained

based on each of the four metamodels (OK. GPR, SVR, ANN) according their requirements, see part 2. Using the same try and cut approach previously described, an ANN with one hidden layer of nine neurons, trained using the algorithm "*trainbr*" is selected. Besides, a linear kernel function is selected for the SVR-based soft-sensor.

Each of the constructed soft-sensors is validated by employing it to predict the TOC of the four validation batches. So, with the known initial concentrations [TOC (*t=0*), *T*(*t*=0), *R*(*t*=0)], together with the timely measured values of the online variables, the soft-sensors will enable the TOC monitoring along the whole time of each batch.

$$TOC(t) = f[T(t), R(t), TOC(t=0), T(t=0), R(t=0)\,] \quad (22)$$

The numerical accuracy of the soft-sensors predictions can be assessed through the results shown in Table 5. The NRMSE values are again in the range of the accuracy of the available information used to train the system, and affected by the accuracy of the online information available to make the required predictions. In this case, it must be noted the limited number of data available (49 samples), which probably underestimates the real performance according to the conclusions drawn from the simulation case-studies results (the accuracy measures calculated relative to the noisy measurements often underestimate the real soft-sensor performance –Table 1 and Table 2).

Table 5. Average accuracy measures (RMSE, NRMSE, CC) of the TOC soft-sensors.

	RMSE	NRMSE (%)	CC
OK	1.19	2.37	0.9956
GPR	1.50	2.98	0.9942
SVR	2.84	5.66	0.9730
ANN	1.90	3.84	0.9901

Figure 19 and Figure 20 show the qualitative assessment of the TOC soft-sensors: Figure 19 displays the TOC estimations of the validation batches set using the fitted soft-sensors compared to their corresponding real measurements, resulting in a well distributed correlation along the ideal diagonal of the graphic; Figure 20 shows the distributions of the prediction normalized errors for the soft-sensors based on the OK (a) and the SVR (b), which are quite close to normal type.

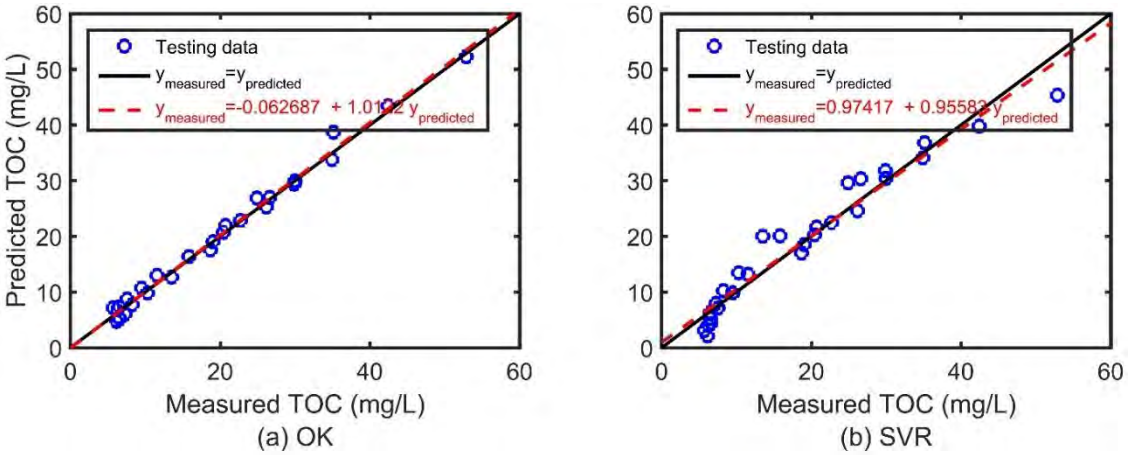

Figure 19. Prediction of the validation batches data versus their experimental measurements.

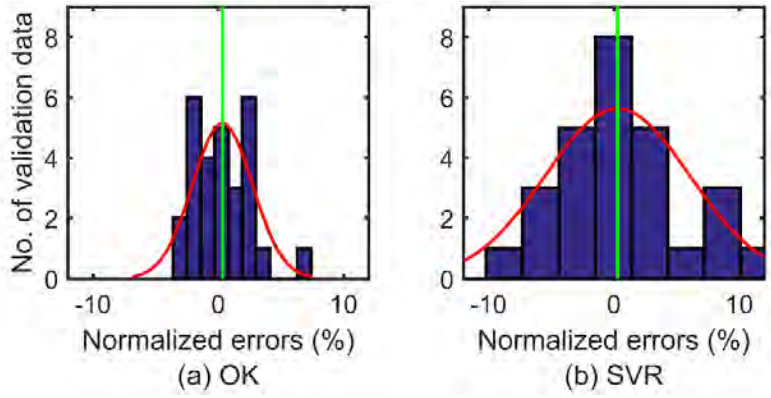

Figure 20. Normalized errors distributions of the TOC soft-sensors predictions: (a) OK and (b) ANN.

Both of the qualitative assessments confirm that the four soft-sensors do not reflect any bias, and also the better performance of the OK based approach. But the distributions of the prediction errors of the photo Fenton soft-sensors (Figure 20) do not show the same very high quality compared to the soft-sensors of the simulation case-studies (Figure 5 and Figure 13). This is explained by three main reasons: first, the very low size of the training and validation datasets in this case (49 data for training and 28 data for validation), while relatively larger datasets are used in the previously presented simulation cases. Second, the uniformity of the training and validation data relative to the initial conditions domain, which in this case was forced by the need to take the maximum advantage of the available expensive data, in front of the use of a well-designed experimentation plan that spans/covers the whole working domain (Hammersley sampling technique - see Figure 1 and Figure 8 vs. Figure 17). Third and finally, the hypothesis that the experimental errors will follow a normal distribution, which may not be true in this case (while this was obviously true in the simulated cases).

Table 5 shows the detailed accuracy measures for each of the four validation batches independently. It should be noticed that batch no. 9 is showing the lowest prediction accuracy (maximum RMSE and NRMSE) in any case. This is associated to the fact that the initial conditions of this batch are far from

the initial conditions of the training batches set (see Figure 17), which makes the trained soft-sensors having relatively less knowledge about the process behavior in this region of the initial operating condition than the other batches (no. 1, 3, 4), whose initial conditions are relatively closer to the initial conditions of the training batches.

Table 6. Accuracy measures for each of the four validation batches.

	Batch No. 1		***Batch No. 3***		***Batch No.4***		***Batch No.9***	
	RMSE	***NRMSE***	***RMSE***	***NRMSE***	***RMSE***	***NRMSE***	***RMSE***	***NRMSE***
OK	*1.05*	*2.09*	*1.10*	*2.20*	*0.74*	*1.48*	*1.68*	*3.35*
GPR	*1.61*	*3.21*	*0.82*	*1.63*	*1.14*	*2.27*	*2.09*	*4.18*
SVR	*3.33*	*6.63*	*2.22*	*4.43*	*1.00*	*1.99*	*3.90*	*7.78*
ANN	*1.38*	*2.76*	*0.73*	*1.45*	*2.12*	*4.23*	*2.81*	*5.60*

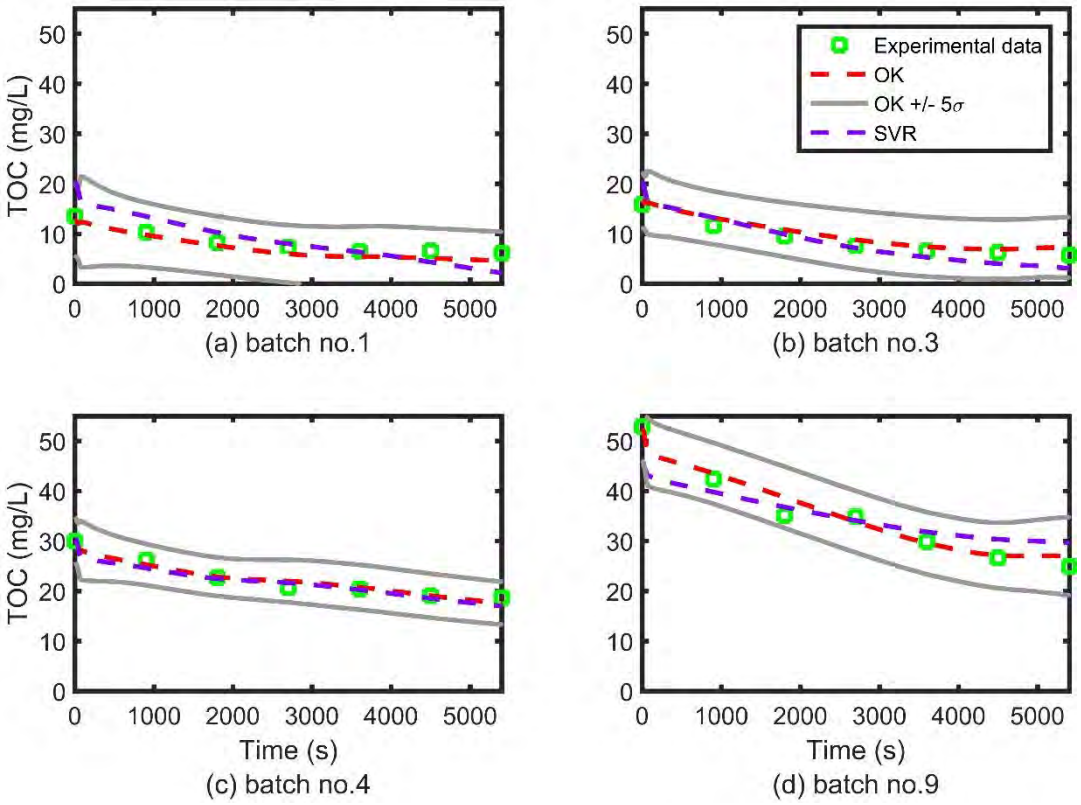

Figure 21. TOC prediction of the four validation batches.

Figure 21 shows the TOC prediction in the four validation batches using the OK (most accurate) and the SVR (least accurate) soft-sensors, proving the high capabilities of the soft-sensors to continuously (each second) predict the TOC along different batch runs with very high accuracy and flexibility in

terms of the initial conditions. The same figure confirms how the OK estimated error can be very useful in practical situations where no knowledge about the exact process behavior (i.e. FPM) is available. Hence, this estimated error is used to establish a confident interval about the soft-sensor prediction (Figure 21, solid gray lines, -/+ 5 sigma). So, even in the cases where the OK –or the GPR- predictions behave a significant deviation from the measured data (green squares), these deviations are still within the confidence area.

5. Conclusions and future work

This work proposes a soft-sensing methodology applicable to a specific class of batch processes, those whose dynamic nature is combined with a significant variability in the initial batch conditions. The presented soft-sensor building approach is found to be appropriate for these cases, that usually appear when the process should manage raw materials whose specifications or properties frequently differ from one batch to another (e.g. waste treatment systems), new process conditions are considered (e.g. to incorporate new production resources), or at the early stages of designing new products, so different alternatives of the initial conditions/settings are explored.

The procedure is tested using the most common data-based modeling architectures in the soft-sensing area, as Ordinary Kriging, Gaussian Processes, Support Vector Regressions and Artificial Neural Networks in order to demonstrate its robustness and potential capabilities. The method is applied to two simulation case-studies and a pilot-plant situation running the photo-Fenton reaction (an advanced oxidation process) for wastewater treatment in batch mode, in order to online predict the reaction progress.

The results produced reveal promising prediction accuracy even when few input-output training data/measurements are available due to large sampling periods. Additionally, they illustrate high capabilities in the approximation of the QIV -that is expensively measured by offline sampling and analyses- along the whole batch, even when the batch initial conditions vary form one batch run to another. Consequently, the application of the proposed methodology may result in huge savings of time and cost consumed by the expensive offline sampling and experimentations, enhancing the online supervision and monitoring of such complex processes difficult to follow through FPMs, and also to allow flexible exploration of many alternative designs (of the initial conditions) with minimum cost.

In quantitative terms, the accuracy shown by the soft-sensors proposed in this work clearly meets the target expectation of being within the same order of magnitude of the accuracy offered by the real-sensor(s) providing the information used to train the soft-sensor, which is additionally affected by the accuracy of the available online information used to make the required estimations. In any case, it is worth noting that, when a soft sensor is needed, this is most likely because no feasible real-sensor is

available to work online and, in this sense, the proposed soft-sensors offer better accuracies than other soft-sensors recently proposed in the literature for continuous and batch processes and, even more, they are applicable to situations not supported by other alternatives (batch processes with changing initial conditions).

Considering the presented application cases in their different versions, the OK and the GPR based models have shown the best average performance among all the considered metamodels in terms of their prediction accuracy. Besides, both exhibit very high flexibility and robustness during their tuning -compared to the SVR and ANN-, since all the model parameters can be easily optimized (the modeler just makes a guess about the parameters initial values and then the optimization/training procedure provides their optimal values). Additionally, the OK and GPR metamodels offer an outperforming characteristic: their ability to estimate a prediction error, which has demonstrated to be very useful to construct a confidence area around the soft-sensor predictions. This estimation can be used for process control when the real behavior of the system is unknown (as in the photo-Fenton case), or when it is hard to estimate the prediction accuracy.

The SVR has also achieved good accuracy, and was very competitive to the OK and GPR in the first simulation cases study. However, the SVR requires higher effort and time to select a suitable kernel function (once the kernel is selected, the SVR training task –the support vectors selection- is usually carried out requiring very low computational effort). Similarly, the ANN requires a significantly higher testing effort to select the best configuration including the number of layers, the number of neurons in each layer and the training algorithm. Nevertheless, it is not the main objective of the work to compare specific metamodeling approaches, but to prove the flexibility and robustness of the soft-sensing approach to work properly with different metamodel types. Thus, any other metamodel type can be employed, and it may achieve better accuracy than the considered ones (e.g. advanced types of ANNs).

In spite of the inherent dynamic behavior of any batch system, in many cases (as the ones studied in this work) it is not required to introduce any lag or delay to the soft-sensor model (i.e.: explicit dynamics) to provide very high prediction accuracy, so the dynamic nature of the process can be correctly tracked just through the available online indirect measurements. In the proposed approach, lag introduction is feasible, but it would result in additional effort in means of the complication of the soft sensor model structure through including additional model inputs (i.e. delayed values of the online variables) and, consequently, additional model parameters which, in turn, would increase the required size of training data –which are often scarce for such processes- in order to identify these extra parameters.

However, a future work line is the investigation of the ability of this soft-sensing methodology to handle batch processes involving delayed behavior, where the lagged values of the online variables can be considered as additional model inputs. Hence, the soft-sensor in equation (17) might be $y(t) = f[x(t), x(t-1), \dots x(t-L), x(t=0), y(t=0)]$, where L=1, 2,... represents the suitable lag to be especifically applied to the model (Espinosa & Vandewalle, 1997; Espinosa & Vandewalle, 1998; Nelles, 2001; Cho et al., 2007).

Another immediate/running extension of the presented approach -which is also related to the pervious point- is the development of dynamic soft-sensors in the form of nonlinear autoregressive models, in order to predict the future values of the QIV over several time steps or intervals (Shokry et al., 2017a). This would improve the chances of effective and flexible process monitoring, supervision and control in the challenging case of scarce measured data along each batch.

6. References

Amozeghar, M. & Khorasani, K., 2016. An ensemble of dynamic neural network identifiers for fault detection and isolation of gas turbine engines. *Neural Networks,* Volume 76, p. 106–121.

Ayodele, O. B., Auta, H. S. & Nor, N. M., 2012. Artificial Neural Networks, Optimization and Kinetic Modeling of Amoxicillin Degradation in Photo-Fenton Process Using Aluminum Pillared Montmorillonite-Supported Ferrioxalate Catalyst. *Ind. Eng. Chem. Res.,* Volume 51 , p. 16311–16319.

Azman, K. & Kocijan, J., 2007. Application of Gaussian processes for black-box modelling of biosystems. *ISA Transactions,* Volume 46, p. 443–457.

Bajpai, R. K. & Reul, M., 1980. A Mechanistic Model for Penicillin Production. *J. Chem. Tech. Biotechnol.,* Volume 30, pp. 332-344.

Banga, J. R., Canto, E. B., Moles, C. G. & Alonso, A. A., 2005. Dynamic optimization of bioprocesses: Efficient and robust numerical strategies. *Journal of Biotechnology,* Volume 117, p. 407–419.

Banu, U. S. & Umab, G., 2011. ANFIS based sensor fault detection for continuous stirred tank reactor. *Applied Soft Computing,* Volume 11, p. 2618–2624.

Belkacem, S., Bouafia, S. & Chabani, M., 2017. Study of oxytetracycline degradation by means of anodic oxidation process using platinized titanium (Ti/Pt) anode and modeling by artificial neural networks. *Process Safety and Environmental Protection,* Volume 111, pp. 170-179.

Bonne, D. & Jorgensen, S. B., 2004. *Data-driven modeling of batch processes.* s.l., s.n.

Caballero, J. A. & Grossmann, I. E., 2008. An algorithm for the use of surrogate models in modular flowsheet optimization. *AIChE Journal,* Volume 54, p. 2633–2650.

Cho, J., Principe, J. C., Erdogmus, D. & Motter, M., 2007. Quasi-sliding mode control strategy based on multiple-linear models. *Neurocomputing,* Volume 10, p. 960–974.

Cressi, N. A., 1993. *Statistics for spatial data.* New York: Wiley & Sons.

Cuthrell, J. & Biegler, L., 1989. Simultaneous optimization and solution methods for batch reactor control profiles. *Computers & Chemical Engineering,* Volume 13, pp. 49-62.

Dadebo, S. & Mcauley, K., 1995. Dynamic optimization of constrained chemical engineering problems using dynamic programming. *Computers & Chemical Engineering,* Volume 19, pp. 513-525.

Davis, E. & Ierapetritou, M., 2007. A kriging method for the solution of nonlinear programs with black-box functions. *AIChE,* Volume 53, pp. 2001-2012.

De Tuesta, j. L. D. et al., 2015. Application of high-temperature Fenton oxidation for the treatment of sulfonation plant wastewater. *J Chem Technol Biotechnol,* Volume 90, p. 1839–1846.

Desai, K., Badhe, Y., Tambe, S. S. & Kulkarni, D., 2006. Soft-sensor development for fed-batch bioreactors using support vector regression. *Biochemical Engineering Journal,* Volume 27, p. 225–239.

Duran, A., Monteagudo, J. & Mohedano, M., 2006. Neural networks simulation of photo-Fenton degradation of Reactive Blue 4. *Applied Catalysis B: Environmental,* Volume 65, p. 127–134.

Elmolla, E. S., Chaudhuri, M. & Eltoukhy, M. M., 2010. The use of artificial neural network (ANN) for modeling of COD removal from antibiotic aqueous solution by the Fenton process. *Journal of Hazardous Materials,* Volume 179, p. 127–134.

Erickson, C. B., Ankenman, B. E. & Sanchez, S. M., 2018. Comparison of Gaussian process modeling software. *European Journal of Operational Research,* Volume 266, pp. 179-192.

Espinosa, J. J. & Vandewalle, J., 1997. Fuzzy Modeling and Identification, A guide for the user.

Espinosa, J. & Vandewalle, J., 1998. Predictive control using fuzzy models applied to a steam generating unit. *Proceedings of the 3rd International Workshop on Fuzzy Logic and Intelligent Technologies for Nuclear Science and Industry.*

Expósito, A., Monteagudo, J., Durán, A. & Fernández, A., 2017. Dynamic behavior of hydroxyl radical in sono-photo-Fenton mineralization of synthetic municipal wastewater effluent containing antipyrine. *Ultrasonics Sonochemistry,* Volume 35,A, pp. 185-195.

Facco, P., Doplicher, F., Bezzo, F. & Barolo, M., 2009. Moving average PLS soft sensor for online product quality estimation in an industrial batch polymerization process.. *Journal of Process Control,* Volume 19, p. 520–529.

Fang, K.-T., Li , R. & Sudjianto, A., 2006. *Design and modelling for computer experiment.* New York: Chapman and Hall/CRC.

Farias, J., Albizzati, E. & Alfano, O., 2009. Kinetic study of the photo-Fenton degradation of formic acid: Combined effects of temperature and iron concentration. *Catalysis Today,* Volume Catalysis Today, pp. 117-123.

Forrester, A. I. & Keane, A. J., 2009. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences,* Volume 45, p. 50–79.

Forrester, A., Sobester, D. A. & Keane, A., 2008. *Engineering Design via Surrogate Modelling: A Practical Guide.* Southampton, UK: John Wiley and Sons.

Forrester, A., Sobester, D. A. & Keane, A., 2008. *Engineering Design via Surrogate Modelling: A Practical Guide.* Southampton, UK: John Wiley and Sons.

Gazi, M., Oladipo, A. A., Ojoro, Z. E. & Ozan, H., 2017. High-Performance Nanocatalyst for Adsorptive and Photoassisted Fenton-Like Degradation of Phenol: Modeling Using Artificial Neural Networks. *Chemical Engineering Communications,* Volume 204, pp. 729-738.

Göb, S. et al., 1999. Modeling the kinetics of a photochemical water treatment process by means of artificial neural networks. *Chemical Engineering and Processing: Process Intensification,* Volume 38, p. 373–382.

Gonzaga, J., Meleiro, L., Kiang, C. & Filhoc, R. M., 2009. ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process. *Computers & Chemical Engineering,* Volume 33, p. 43–49.

Grbić, R., Slišković, D. & Kadlec, P., 2013. Adaptive soft sensor for online prediction and process monitoring based on a mixture of Gaussian process models. *Computers and Chemical Engineering,* Volume 58, p. 84–97.

Guimarães, O. L. C. et al., 2007. Prediction via Neural Networks of the Residual Hydrogen Peroxide used in Photo-Fenton Processes for Effluent Treatment. *Chemical Engineering & Technology,* Volume 30, p. 1134–1139.

Gustavsson, R., Lukasser, C. & Mandenius, C.-F., 2015. Control of specific carbon dioxide production in a fed-batch culture producing recombinant protein using a soft sensor. *Journal of Biotechnology,* Volume 200, p. 44–51.

Hassani, A., Khataee, A. & Karaca, S., 2015. Photocatalytic degradation of ciprofloxacin by synthesized TiO2 nanoparticles on montmorillonite: Effect of operation parameters andartificial neural network modeling. *Journal of Molecular Catalysis A: Chemical,* Volume 409, pp. 149-161.

Hoskins, J. & Himmelblau, D., 1988. Artificial neural network models of knowledge representation in chemical engineering. *Computers & Chemical Engineering,* Volume 12, pp. 881-890.

Jaafarzadeh, N. et al., 2012. Predicting Fenton modification of solid waste vegetable oil industry for arsenic removal using artificial neural networks. *Journal of the Taiwan Institute of Chemical Engineers,* Volume 43, p. 873–878.

Jain, P., Rahman, I. & . Kulkarni, B. D., 2007. Development of a Soft Sensor for a Batch Distillation Column Using Support Vector Regression Techniques. *Chem.Eng.Res.Des. 85,* pp. 283-287.

Jin, H., Chen, X., Wang, L. & Yang, K., 2015. Adaptive Soft Sensor Development Based on Online Ensemble Gaussian Process Regression for Nonlinear Time-Varying Batch Processes. *Ind. Eng. Chem. Res.,* Volume 54, p. 7320−7345.

Jin, H., Chen, X., Yang, J. & Wu, L., 2014. Adaptive soft sensor modeling framework based on just-in-time learning and kernel partial least squares regression for nonlinear multiphase batch processes. *Computers & Chemical Engineering,* Volume 71, pp. 77-93.

Jones, D. R., Schonlau, M. & Wel, W. J., 1998. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization,* Volume 13, pp. 455-492.

Kadlec, P., Gabrys, B. & Strandt, S., 2009. Data-driven Soft Sensors in the process industry. *Computers and Chemical Engineering,* Volume 33, p. 795–814.

Kadlec, P., Grbić, R. & Gabrys, B., 2011. Review of adaptation mechanisms for data-driven soft sensors. *Computers & Chemical Engineering,* Volume 35, p. 1–24.

Kaneko, H. & Funatsu, K., 2013. Classification of the Degradation of Soft Sensor Models and Discussion on Adaptive Models. *AIChE Journal,* Volume 59, p. 2339–2347.

Khataee, A. et al., 2014. Modeling and optimization of photocatalytic/photoassisted-electro-Fenton like degradation of phenol using a neural network coupled with genetic algorithm. *Journal of Industrial and Engineering Chemistry,* Volume 20, pp. 1852-1860.

Kleijnen, J. P., 2017. Regression and Kriging metamodels with their experimental designs in simulation: A review. *European Journal of Operational Research,* Volume 256, p. 1–16.

Kocijan, J., 2016. *Modelling and Control of Dynamic Systems Using Gaussian Process Models.* s.l.:Springer International Publishing.

Krige, D., 1951. A statistical approach to some mine valuations and allied problems at the Witwatersrand. *Master's thesis of the University of Witwatersrand.*

Lataniotis, C., Marelli, S. & Sudret, B., 2017. *Gaussian process modelling using UQLab,* Zúrich,Switzerland: Chair of Risk, Safety and Uncertainty Quantification-ETH Zúrich.

Lin, B., Recke, B., Knudsen, J. . K. & Jørgensen, S. B., 2007. A systematic approach for soft sensor development. *Computers and Chemical Engineering,* Volume 31, p. 419–425.

Liu, Y., Gao, Z., Li, P. & Wang, H., 2012. Just-in-Time Kernel Learning with Adaptive Parameter Selection for Soft Sensor Modeling of Batch Processes. *Ind. Eng. Chem. Res.,* Volume 51, p. 4313–4327.

Liu, Y. et al., 2016. Development of multiple-step soft-sensors using a Gaussian process model with application for fault prognosis. *Chemometrics and Intelligent Laboratory Systems,* Volume 157, p. 85–95.

Masters, T., 1993. *Practical. Neural Network Recipes in C++.* San Diego New York: Academic Press.

Matheron, G., 1963. Principles of geostatistics. *Economic Geology,* Volume 58, p. 1246–1266.

Moreno-Benito, M., 2014. *Integrated Batch Process Development based on Mixed-Logic Dynamic Optimization.* Barcelona, Spain: PhD thesis.

Mota, A. L. N. et al., 2014. Application of artificial neural network for modeling of phenol mineralization by photo-Fenton process using a multi-lamp reactor. *Water Sci Technol.,* Issue 69, pp. 768-774.

Mustafa, Y. A., Jaid, G. M., Alwared, A. I. & Ebrahim, M., 2014. The use of artificial neural network (ANN) for the prediction and simulation of oil degradation in wastewater by AOP. *Environ Sci Pollut Res,* Volume 21, p. 7530–7537.

Nagy, Z. K., 2007. Model based control of a yeast fermentation bioreactor using optimally designed artificial neural networks. *Chemical Engineering Journal,* Volume 127, pp. 95-109.

Nascimento, C. A. D., Oliveros, E. & Braun, A. M., 1994. Neural network modelling for photochemical processes. *Chemical Engineering and Processing: Process Intensification,* Volume 33, pp. 319-324.

Nelles, O., 2001. *Nonlinear System Identification From Classical Approaches to Neural Networks and Fuzzy Models.* Berlin,Heidelberg: Springer.

O'Hagan, M. C. K. a. A., 2001. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* Volume 63, p. 425–464.

O'Hagan, A., Kennedy, M. C. & Oakley, J. E., 1999. Uncertainty analysis and other inference tools for complex computer codes. *In Bayesian Statistics 6, (J. M. Bernardo et al (eds.)), Oxford University Press,* pp. 503-524.

O'Hagan, A. & Kingman, J. F. C., 1978. Curve Fitting and Optimal Design for Prediction. *Journal of the Royal Statistical Society. Series B (Methodological) ,* Volume 40, pp. 1-42 .

Pasquale, N. D., Davie, S. J. & Popelier, P. L. A., 2016. Optimization Algorithms in Optimal Predictions of Atomistic Properties by Kriging. *Journal of Chemical Theory and Computation,* Volume 12, pp. 1499-1513.

Queipo, N. V. et al., 2005. Surrogate-based analysis and optimization. *Progress in Aerospace Sciences,* Volume 2005, p. 1–28.

Quirante, N., Javaloyes, J. & A. Caballero, J., 2015. Rigorous design of distillation columns using surrogate models based on Kriging interpolation. *AIChE Journal,* Volume 61, p. 2169–2187.

Rasmussen, C. E. & Williams, C. K. I., 2006. *Gaussian processes for machine learning.* Cambridge, Massachusetts: MIT Press.

Reina, A. C. et al., 2012. Modelling photo-Fenton process for organic matter mineralization, hydrogen peroxide consumption and dissolved oxygen evolution. *Applied Catalysis B: Environmental,* Volume 119–120, p. 132–138.

Rogers, A. & Ierapetritou, M., 2015. Feasibility and flexibility analysis of black-box processes Part 1: Surrogate-based feasibility analysis. *Chemical Engineering Science,* Volume 137, p. 986–1004.

Ruppen, D., Benthack, C. & Bonvin, D., 1995. Optimization of batch reactor operation under parametric uncertainty - computational aspects. *Journal of Process Control,* Volume 5, pp. 235-240.

Sacks, J., Welch, W. J., Mitchell, T. J. & Wynn, H. P., 1989. Design and Analysis of Computer Experiments. *Statistical Science ,* Volume 4, pp. 409-423 .

Salari, D., Daneshvar, N., Aghazadeh, F. & Khataee, A., 2005. Application of artificial neural networks for modeling of the treatment of wastewater contaminated with methyltert-butyl ether (MTBE) by UV/H2O2 process. *Journal of Hazardous Materials B,* pp. 205-210.

Sebti, A., Souahi, F., Mohellebi, F. & Igoud, S., 2017. Experimental study and artificial neural network modeling of tartrazine removal by photo-catalytic process under solar light. *Water Science & Technology,* Volume 76, pp. 311-322.

Shokry, A. et al., 2016. Kriging based Fault Detection and Diagnosis Approach for Nonlinear Noisy Dynamic Processes. *Computer Aided Chemical Engineering,* Volume 38, p. 55–60.

Shokry, A. et al., 2017a. Dynamic Kriging based Fault Detection and Diagnosis Approach for Nonlinear Noisy Dynamic Processes. *Computers & Chemical Engineering,* Volume 758-776, p. 106.

Shokry, A. et al., 2015. Modeling and simulation of complex nonlinear dynamic processes using data based models: application to photo-fenton process. *Computer Aided Process Engineering,* Volume 37, p. 191–196.

Shokry, A., Espuña & A., 2014. Applying Metamodels and Sequential Sampling for Constrained Optimization of Process Operations. *Lecture Notes in Computer Science,* Volume 8468, pp. 396-407.

Shokry, A. et al., 2017b. Data-Driven Dynamic Modeling of Batch Processes Having Different Initial Conditions and Missing Measurements. *Computer Aided Chemical Engineering,* Volume 40, pp. 433-438.

Vapnik, V., 1995. *The Nature of Statistical Learning Theory.* New York.: Springer.

Wang, L. et al., 2016. Soft Sensor Development Based on the Hierarchical Ensemble of Gaussian Process Regression Models for Nonlinear and Non-Gaussian Chemical Processes.. *Ind. Eng. Chem. Res.,* Volume 55, pp. 7704-7719.

Wang, L., Liu, X. & Zhang, Z., 2017. A new sensitivity-based adaptive control vector parameterization approach for dynamic optimization of bioprocesses. *Bioprocess Biosyst Eng,* Volume 40, p. 181–189 .

Yamal-Turbay, E., E., O., O., C. L. & M., G., 2015. Photonic efficiency of the photodegradation of paracetamol in water by the photo-Fenton process. *Environ Sci Pollut Res,* Volume 22, p. 938–945.

Yan, W., Shao, H. & Wang, X., 2004. Soft sensing modeling based on support vector machine and Bayesian model selection. *Computers and Chemical Engineering,* Volume 28, p. 1489–1498.

Zamprogna, E., Barolo, M. & Seborg, D. E., 2005. Optimal Selection of Soft Sensor Inputs For Batch Distillation Columns Using Principal Component Analysis. *Journal of Process Control,* Volume 15, p. 39–52.