The University of Tehran Press

Pollution

https://jpoll.ut.ac.ir/

Print ISSN: 2383-451X
Online ISSN: 2383-4501

Enhanced Monitoring of Water Quality in Crude Oil Desalting/ Dehydration Plant (DDP) using Soft Sensing Techniques

Farzaneh Naimi Rad | Mir Mohammad Khalilipour✉ | Bahareh Bidar | Farhad Shahraki | Jafar Sadeghi

Center for Process Integration and Control (CPIC), Department of Chemical Engineering, University of Sistan and Baluchestan, Zahedan, Iran

Article Info

Article type:
Research Article

Article history:
Received: 31 May 2024
Revised: 28 August 2024
Accepted: 07 January 2025

Keywords:
Water quality
Soft sensor
Crude oil desalting dehydration plant (DDP)
State-dependent parameter (SDP)
Local instrumental variable (LIV)

ABSTRACT

The present study introduces a novel soft sensor based on State Dependent Parameter (SDP) models utilizing the Local Instrumental Variables (LIV) method for monitoring a crude oil Desalting and Dehydration Plant (DDP) system. A key advantage of the LIV modeling method is its ability to interpolate directly without necessitating extensive model parameterization. Additionally, the inherent complexity and non-linearity of the process are effectively addressed by LIV-based soft sensors, which require fewer process variables, thereby reducing training time and computational complexity. Two distinct soft sensors were developed to assess the salinity efficiency and water cut efficiency of the DDP system. The efficacy of these soft sensors was evaluated using a dedicated testing dataset, revealing a robust correlation between salinity efficiency, water cut efficiency, and five secondary parameters. Comparisons between SDP-LIV model predictions and real observations of the DDP process show strong agreement. By leveraging these developed soft sensors, continuous evaluation of product properties is possible with minimal delay compared to traditional laboratory analyses. This capability is crucial for pollution control and environmental monitoring, as it allows for real-time detection and mitigation of contaminants in crude oil processing. Lastly, the performance of the proposed soft sensor is benchmarked against other models, such as Multiple Linear Regression (MLR) and Artificial Neural Networks (ANN), demonstrating superior predictive capabilities. This study underscores the potential of SDP-LIV-based soft sensors in enhancing environmental protection and operational efficiency in crude oil processing.

Cite this article: Naimi Rad, F., Khalilipour, M. M., Bidar, B., Shahraki, F., & Sadeghi, J. (2025). Enhanced Monitoring of Water Quality in Crude Oil Desalting/Dehydration Plant (DDP) using Soft Sensing Techniques. *Pollution*, 11(1), 175-190. https://doi.org/10.22059/poll.2024.377336.2401

DOI: https://doi.org/10.22059/poll.2024.377336.2401

INTRODUCTION

In recent decades, there has been a notable surge in the demand for high-quality crude oil. However, the crude oil extracted from most of the world's oil fields often contains substantial impurities, including water in the form of free water or water-in-oil emulsions (suspended water droplets in oil), soluble salts, and various sediments. The elevated levels of salt and water content in crude oil pose significant challenges during processing, refining, and transportation processes, such as corrosion and fouling within operational equipment like pipes, pumps, valves, and tanks. Additionally, these impurities can diminish the effectiveness of catalysts in subsequent processing units, highlighting the critical importance of producing oil with sufficient product purity (Hosseinpour, Ghader, Rahimpour, & Bagheri, 2019; Sotelo, Favela-Contreras, Sotelo, Beltrán-Carbajal, & Cruz, 2018). To address these challenges, crude oil is typically subjected to processing in a DDP before its conveyance to refineries. DDP systems represent

*Corresponding Author Email: *a.khalilipour@eng.usb.ac.ir*

industrial processes aimed at enhancing oil product purity by removing water and soluble salts from crude oils. The necessity of DDP systems is multifaceted: they serve to diminish the flow of salt content to refinery distillation feed stocks, minimize the energy required for pumping and transportation, mitigate corrosion, plugging, and fouling of piping and process equipment, and reduced effectiveness of the catalysts that are used for crude oil refining processes. (Ranaee et al., 2021).

The foundational principles of DDP systems encompass three key steps: initially breaking emulsions, achieved through various methods such as oil heating, chemical injection, or electric field application; subsequently coalescing tiny water droplets into larger ones; and ultimately gravity settling and separating free water. Given the inherent difficulty in separating emulsions from crude oil, DDP systems constitute complex processes whose performance hinges on numerous processing parameters. Extensive studies have been conducted to analyze and understand the factors influencing DDP efficiency (K. Mahdi, R. Gheshlaghi, G. Zahedi, & A. Lohi, 2008; Nasehi, Sarraf, Ilkhani, Mohammadmirzaie, & Fazaelipoor, 2019). Consequently, achieving continuous monitoring and control of DDP systems presents a formidable challenge (Dadari, Rahimi, & Zinadini, 2016; Roodbari, Badiei, Soleimani, & Khaniani, 2016).

The primary or quality variables (e.g., purity, physical or chemical properties) are important process variables, which are often difficult to measure online. They are usually obtained through laboratory measurement with significant delays (time delay of hours) and infrequent (about some samples per day). Soft sensors are developed to overcome measurement problems of primary/quality variables by monitoring suitable secondary variables (Fortuna, Graziani, Rizzo, & Xibilia, 2007; Wang, Liu, & Srinivasan, 2010). Soft sensors are data-driven models, which have used statistical and/or artificial intelligence techniques capable of converting information from measurable secondary variables (temperature, pressure, flow rates, etc.) to estimate primary or quality variables. Use of soft sensors helps to make faster and more appropriate decisions during practical difficulties associated with delays in measurements, unreliable measured variables due to drifts, fouling or accidental damage of hard sensors, and manual errors in laboratory measurements. Many techniques can be used to develop soft sensors based on processing plant data. These kinds of soft sensing models are known as data-driven soft sensors (He, Geng, & Zhu, 2015; Kadlec, Gabrys, & Strandt, 2009). The simplest approach to building data-driven soft sensors is to carry out MLR using the least-squares method, in which the model results can be affected by a number of data issues. There are other techniques that range from linear based on partial least squares (PLS) (Liu, 2014; Zheng & Funatsu, 2018), principal component analysis (PCA) (Jolliffe, 2002; Shi & Xiong, 2018) to nonlinear methods based on neural networks (NN) (Pan, Su, Huang, & Wang, 2021; Sun, Huang, Jang, & Wong, 2016; Zhao, Li, & Cao, 2019), neuro-fuzzy system (NFS) (AbdulJalee & Aparna, 2016; Zhao et al., 2019), support vector regression (SVR) (Herceg, Andrijić, & Bolf, 2019; Zhongda, Shujiang, Yanhong, & Xiangdong, 2016), and Gaussian process regression (GPR) (Kanno & Kaneko, 2020; Li, Xu, Han, Ge, & Wang, 2019).

A literature survey reveals that though quite a few soft sensors for DDP systems have been reported, most of these reported soft sensors are based on machine learning methods. Al-Otaibi et al. (M. B. Al-Otaibi, Elkamel, Al-Sahhaf, & Ahmed, 2003) investigated experimentally the effect of five process variables e.g. gravity settling, chemical treatment, freshwater injection, heating, and mixing on two DDP efficiencies which are defined by salt removal efficiency and water cut dehydration efficiency. The results showed that settling time was the most influential variable while excessive amounts of the demulsifying agent had adverse effects on the performance of the DDP system. In another study, Al-Otaibi et al. (Musleh B Al-Otaibi, Elkamel, Nassehi, & Abdul-Wahab, 2005) simulated and optimized the DDP system by applying the ANN technique. The performance of the DDP system was evaluated by determining the salt removal and water cut efficiencies as quality variables that were expected to depend on five

process variables e.g. the salt concentration, heating, concentration of demulsifying agents, wash water, and the rate of mixing with wash water. The neural network model predictions were shown to be consistent with the experimental data. Abdul-Wahab et al. (Abdul-Wahab, Elkamel, Madhuranthakam, & Al-Otaibi, 2006) developed inferential estimators for the salt removal and water cut efficiencies of the DDP system in terms of five secondary process variables as temperature, settling time, mixing time, chemical dosage, and dilution water rate. The inferential estimators were constructed based on MLR and PCA as well as non-linear regression. The results showed that the performance of the DDP cannot be fully described by linear models and it requires identification of the nonlinear relationship of process variables. Mahdi et al. (K Mahdi, REZA Gheshlaghi, Gholamreza Zahedi, & Ali Lohi, 2008) investigated the effect of five process variables as demulsifying agent concentration, temperature, wash water dilution ratio, settling time, and mixing time with wash water on the performance of the DDP system using statistical analysis. They provided one model with five process variables for the salt removal efficiency and two models with four and five variables for the water cut efficiency, each was valid in a part of the variable domains. The proposed models were successfully tested and all were confirmed with experimental data. Kamari et al. (Kamari, Bahadori, & Mohammadi, 2015) presented a modeling approach based on the least square support vector machine (LS-SVM) and multilayer perceptron artificial neural network (MLP-ANN) model to calculate the salt content in crude oil. The obtained results express the superiority of the LS-SVM model over the MLP-ANN model. The literature reveals the importance of using soft sensors to estimate the product quality of the DDP system. However, constructing soft sensing models with high prediction performance is difficult due to the nonlinear relationship between the efficiency of the DDP system and process variables.

The use of the SDP identification technique as a data-driven soft sensor modeling method, can be referred to the studies done by Gharehbaghi and Sadeghi (Gharehbaghi & Sadeghi, 2016) and Bidar et al. (Bidar, Khalilipour, Shahraki, & Sadeghi, 2018; Bidar, Sadeghi, Shahraki, & Khalilipour, 2017; Bidar, Shahraki, Sadeghi, & Khalilipour, 2018). Results illustrated that SDP-based soft sensors have superiority over the other common data-driven methods like PLS, PCR, SVR, ANN, and so on because they have the significant ability to model the nonlinear system, whilst the obtained model is simple and interpretable using linear paradigms. In the SDP estimations, an Instrumental Variable (IV) is a state that has two specific properties. Each IV must be as highly correlated as possible with correspondent regressors and at the same time have as little correlation as possible with the other regressors. Otherwise, the estimate of SDPs affects each other and since the SDPs are the functions of state variables, distortion occurs in the final estimate of SDP. One possible solution to solving the problems related to the back-fitting algorithm in the SDP estimations can be the local instrumental variable approach. Hence, Bidar et al. (Bidar, Khalilipour, et al., 2018) proposed the SDP modeling approach using the LIV method as a novel approach for the identification and modeling of nonlinear systems to predict the product quality of the industrial crude distillation unit. In the LIV method, after determining IVs there is no need to sort them and also because this method does not consider the effects of other regressors and states on the estimation of the desired parameter, it avoids the use of a back-fitting algorithm. Parvizi Moghadam et al. (Moghadam, Sadeghi, & Shahraki, 2021) proposed soft sensors based on the LIV method for the accurate prediction of isopropyl benzene concentration in an industrial distillation column. The results of prediction models have shown a very low error percentage and supreme agreement with prediction quality from the rigorous model compared with other models.

Previous studies have demonstrated the efficacy of SDP-LIV-based soft sensors in estimating processing variables, showcasing their ability to address the complexity and non-linearity inherent in the process. Notably, the innovative aspect of this work lies in the utilization of a soft sensor based on SDP estimation employing the LIV approach to predict the product quality

of DDP. Unlike existing literature, which predominantly focuses on specific aspects of soft sensing, such as salt removal efficiency (SRE) or water removal efficiency (WRE), this study extends the scope to encompass both, contributing to the relevance of monitoring chemical industrial processes. The designed soft sensing models not only facilitate optimal control of the DDP system but also offer significant improvements in performance prediction indexes while requiring fewer process variables. This reduction in variables not only streamlines training time but also minimizes calculation complexity, enhancing the practical applicability of the approach.

Moreover, through rigorous comparison with real observations, the model prediction results demonstrate the effectiveness of the proposed soft sensor models, which exhibit a simple structure yet deliver robust identification based on the essential process parameters of the DDP system. The manuscript underscores its practical contribution by furnishing detailed real-world industrial data for both SRE and WRE prediction, catering directly to the needs of industrial applications. As a result, the proposed soft sensor based on SDP-LIV stands out as a reliable tool for real-time monitoring and control of DDP systems. In further validation, the proposed soft sensor is rigorously tested against alternative models such as MLR and ANN, highlighting its superiority in practical application. This comparative analysis solidifies the value proposition of the SDP-LIV-based soft sensor as a preferred choice for predictive modeling within the context of DDP systems.

Desalting/Dehydration plant description

A typical DDP removes dissolved salts and water droplets from the oil flow before it can be sold. The process flow diagram (PFD) of the desalting/dehydration plant under study is shown in Fig. 1 (K. Mahdi et al., 2008). The primary goal of a DDP system is to achieve sufficient product purity in terms of salt removal and water cut efficiencies. Based on design specifications, the amount of water and salt in crude oil is reduced to 0.10 volume percent

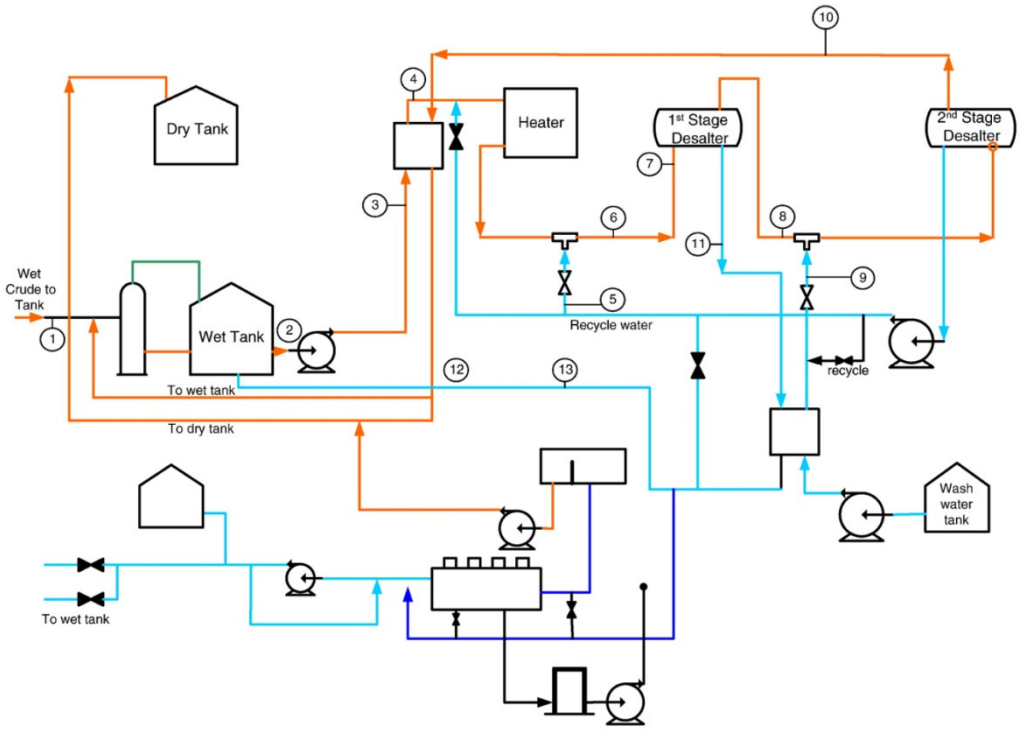

Fig. 1. Schematic of crude oil desalting/dehydration plant (DDP) (Musleh B Al-Otaibi, 2004)

(%Vol.) and 5.0 pounds per thousand barrels (PTB), respectively.

The experiments were used to obtain the data according to the real operation process of the crude oil supplied by Kuwait Oil Company (KOC). The crude oil samples containing the water-oil (W/O) emulsion enter the system and after six major steps of addition of fresh water, heating, chemical injection, mixing, gravity settling, and electrical coalescing, treated crude pass through an analyzer to check the achievement of the design specification of the treated crude. The details of these experiments, the characteristics of the dilution water, the chemical demulsifier, and the instruments used in the laboratory were presented elsewhere (M. Al-Otaibi, 1999; M. B. Al-Otaibi et al., 2003). Experiments are carried out based on KOC standards to study the effect of five process parameters including demulsifier dosage (ppm), crude temperature (°C), dilution water flow rate in ratio to that of the wet crude's quantity (%), mixing/settling time (min) on the performance of the DDP process as listed in Table 1.

The performance of the DDP process was evaluated by the salinity efficiency (η_1) and water cut efficiency (η_2) in various process conditions. The salinity efficiency was calculated from Eq. (1), whereas water cut efficiency was calculated from Eq. (2), respectively,

$$\eta_1 = 1 - \frac{Z_{out}}{Z_{in}} \tag{1}$$

$$\eta_2 = 1 - \frac{X_{out}}{X_{in}} \tag{2}$$

where Z_{out} is outlet salt result (PTB), Z_{in} is inlet salt result (PTB), X_{out} is outlet water cut (%) and X_{in} inlet water cut (%).

MATERIAL & METHODS

The SDP models express nonlinear aspects of the system using a variable parameter model with a linear, simple, and interpretable structure. The parameters of SDP models are functions

Table 1. Description, symbols and values of process parameters of DDP system

	Parameter	Description	Symbol	Values	Number of runs
Output variable	Salinity efficiency (η_1)	Salt removal efficiency	y_1		
	Water cut efficiency (η_2)	Water removal efficiency	y_2		
Input variable	Temperature (°C)	Temperature of the outlet crude	X_1	55°C (low), 70°C (high)	2
	Settling time (min)	Settling time	X_2	1 min (low), 3 min (high)	2
	Mixing time (min)	Mixing time	X_3	1, 3, 5, 7 and 9	5
	Demulsifier dosage (ppm)	Chemical addition	X_4	1, 2, 5, 8, 10, 12 and 15	7
	Dilution water (%)	The amount of fresh water added in ratio to that of the wet crude's quantity	X_5	1, 2, 3, 4, 6, 8 and 10	7
	Total number of runs				**980**

of time or system states, which can be written in the following form (P. Young, 1998):

$$\begin{cases} y_t = \sum_{i=1}^{n} a_{i,t} \cdot z_{i,t} + e_t \\ a_{i,t} = a_i\left(x_{1,i,t}, x_{2,i,t}, \ldots, x_{ns_i,i,t}\right) \end{cases}, \quad \forall t \tag{3}$$

where y_t is the model output, n is the number of SDPs/regressors, $z_{i,t}$ is the i^{th} regressor and $a_i(\cdot)$ is the i^{th} SDP that is a function of ns_i correspondent states ($x_{j,i,t}$, $j = 1, 2, \ldots, ns_i$). When $a_{i,t}$ is assumed to be constant and not state dependent, $ns_i = 0$. $e_t = N\left(0, \sigma^2\right)$ is a zero mean white Gaussian distributed unknown noise with variance σ^2.

The LIV method was derived from the weighted least squares (WLS) method which uses instrumental variables and local polynomial techniques. The LIV method provides models based on the structure of the SDP models as shown in Eq. (3). In this study, the local polynomial modeling method (LPM) was used to estimate each state dependent parameters e.g. $a_{i,t}$ in Eq. (3) (Jianqing Fan, 2018; J Fan & Yao, 2003; Hastie & Tibshirani, 1990; P. C. Young, 2011). So, the functionality of the $a_{i,t}$ is defined by a local polynomial in the state space, hence it is possible to estimate the parameters of these polynomials using the IV method. So, Eq. (3) can be rewritten in the new vector form shown in the following equation:

$$y_t = \mathbf{z}_t \mathbf{A}_t + e_t \tag{4}$$

where, $\boldsymbol{A_t}$ is the vector of the parameters of local polynomials demonstrating SDPs and $\boldsymbol{z_t}$ is the new vector of regressors at time sample *t*. The solution to IV estimation of Eq. (4) at the *k*th sample is calculated as follows:

$$\begin{aligned} \hat{\mathbf{A}}_k &= \mathbf{U}_k^T \mathbf{y} \\ \mathbf{P}_k &= \hat{\sigma}_k^2 \left(\mathbf{U}_k^T \mathbf{U}_k\right) \\ \hat{\sigma}_k^2 &= \mathrm{var}\left(\mathbf{y} - \mathbf{Z}\hat{\mathbf{A}}_k\right) \end{aligned} \tag{5}$$

where $\boldsymbol{P_k}$ is the covariance matrix of SDP estimation. $\hat{\boldsymbol{A}}_{\boldsymbol{k}}$ at the *k*th sample and $\mathbf{U}_k = \left[\mathbf{U}_{1,k} \quad \mathbf{U}_{2,k} \quad \cdots \quad \mathbf{U}_{p,k}\right]$ is IV matrix of proposed approach, which is called LIV (Bidar, Khalilipour, et al., 2018).

The local weighting matrix, $\boldsymbol{W_{m,k}}$, can be considered as a diagonal matrix, which its diagonal elements are the values of kernel function correspondent to the i^{th} SDP at the *k*th sample. $\boldsymbol{W_{m,k}}$ is defined as,

$$\begin{cases} \mathbf{W}_{m,k} = diag\left(K\left(\Delta_{i,k,t}\right)\right) \\ \Delta_{i,k,t} = \sum_{j=1}^{ns_i} \left(\dfrac{x_{j,i,t} - x_{j,i,k}}{\lambda_{j,i}} \right)^2 \end{cases} \quad \begin{cases} \forall t \\ 0 < \sum_{v=1}^{i-1} p_v < m \le \sum_{v=1}^{i} p_v \end{cases} \tag{6}$$

where $k(.)$ is the kernel function. $\boldsymbol{ë_{j,i}}$ is the bandwidth correspondent to $x_{j,i,t}$ that is known as hyper-parameter and it must be obtained through the optimization procedure (Bidar, Khalilipour, et al., 2018). $\boldsymbol{ë_{j,i}}$, which is known as hyper-parameter correspondent to each state of system, $x_{j,i,t}$ in LIV method must be fine-tuned to achieve accurate soft sensing model.

Clearly, the choice of the hyper-parameter of bandwidth plays an important role in the local polynomial fitting. Too large a bandwidth causes over-smoothing, creating excessive modelling bias.In this study, Cross Validation (CV) and maximum likelihood (ML) approaches are utilized to optimize bandwidths. If $\hat{\boldsymbol{a}}_{\boldsymbol{k}}$ is the residual at the *k*th sample when that sample is removed from the calculation, and $\boldsymbol{R}_{\boldsymbol{k}}$ is the covariance of $\hat{\boldsymbol{a}}_{\boldsymbol{k}}$, then the CV and concentrated likelihood can be defined as,

Cross Validation:

$$CV = \frac{1}{N}\sum_{k=1}^{N}\hat{\varepsilon}_k^2 \tag{7}$$

"Concentrated Likelihood" Function:

$$\log(L_c) = -\frac{1}{2}\left[\frac{1}{N}\sum_{k=1}^{N}\log\left(R_k\right) + \log\left(\frac{1}{N}\sum_{k=1}^{N}\frac{\hat{\varepsilon}_k^2}{R_k}\right)\right] \tag{8}$$

The CV must be minimized with respect to the hyper-parameters and the Likelihood has to be maximized. Since both the Likelihood function and Cross Validation are a nonlinear functions of the unknown hyper-parameters, the minimization needs to be carried out numerically. At the beginning of the optimization, hyper-parameters ($\ddot{\boldsymbol{e}}_{j,i}$) are estimated by either the user or set to default values.

Soft sensor design

The performance of the DDP process depend on the several process parameters, which they can be altered in order to reach an optimum combination of operating conditions. In previous researches (M. Al-Otaibi, 1999; M. B. AL-Otaibi, 2004; M. B. Al-Otaibi et al., 2003), the DDP process has been evaluated to determine the interactions and the combination of process parameters based on a series of experimental runs according to a pre-specified design of experiment. In DDP process, there are five measured process variables were considered in experiments (M. Al-Otaibi, 1999; M. B. Al-Otaibi et al., 2003) included mixing time (min), crude oil temperature (°C), demulsifier dosage (ppm), settling time (min), and amount of dilution water flow rate in ratio to that of the wet crude's quantity (%). Compared with these measured process variables, the determining salinity and water cut efficiencies is more difficult and time-consuming. Thus, it is necessary to modeling these two process efficiencies accurately for product monitoring and performance evaluation of DDP system.

In this study, both salinity and water cut efficiencies and all five process parameters are selected for soft sensor development. Table 1 illustrates the values for each process parameter. The design of the experiment includes all possible combinations of process parameters within the specified range. Crude oil temperature and settling time parameters had the least amount of change in the real process. Accordingly, the settling time and crude oil temperature were considered only in the high and lower values, and the demulsifier dosage, mixing time, and amount of dilution water were tested in different values. Therefore, a total of 980 samples from (2×2×5×7×7 runs) experiments for each process parameters were collected (M. Al-Otaibi, 1999; M. B. AL-Otaibi, 2004; M. B. Al-Otaibi et al., 2003).

Data collected for soft sensor design were randomly divided into two distinct datasets: a training dataset and a testing dataset. Among which 882 samples (90% of the total data) are randomly selected as the training samples, and the remaining 98 samples (10% of the total data) are used as the testing samples.

Variable selection is a critical aspect of soft sensor model development, significantly

influencing its performance by ensuring the inclusion of effective parameter candidates that correlate with efficiencies in the DDP system. The presence of numerous input variables dramatically increases the computational cost of the model and leads to a large number of model parameters to be estimated, generally causing overfitting and diminishing the accuracy of the soft sensor model (F. Curreri et al., 2020; F. A.A. Souza et al., 2013). Hence, the careful selection of variables is paramount to enhance estimation performance. In this study, salinity and water cut efficiencies are identified as the quality variables for soft sensor development. Utilizing correlation analysis, variables with the highest Pearson correlation coefficient (R) are chosen as potential inputs. Additionally, the backward elimination method is employed to iteratively select a subset of explanatory variables for the model. This method involves initially including all inputs from the candidate set X identified through correlation analysis, followed by the systematic removal of the least significant input, one at a time. Subsequently, the soft sensor models are trained using the selected variables, and any inefficient variables are eliminated from the model based on optimized hyper-parameter values, ensuring the robustness and accuracy of the soft sensor predictions.

The regression modeling is performed between the five secondary variables and the salinity efficiency (y_1) and water cut efficiency (y_2) using Eq. (3). A detailed description of the secondary and quality variables for soft sensor design is given in Table 1. Although the DDP process is a multi-output process, it is treated as two single-output processes, so that y_1 and y_2 are modeled separately. The soft sensor models discussed above are coded in MATLAB 7.7 version on Intel Core TM, i7CPU, 2.80GHz, 4GB RAM, 64 bit operating system.

The association between variables can be quantified using a correlation coefficient. In this study, Pearson's Correlation Coefficient (R) is adopted to select the most effective secondary variables of the soft sensor. It is given by following equation.

$$R = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{9}$$

where x_i is the values of the x variable in a sample, $\bar{x}$ mean of the values of the x variable, y_i values of the y variable in a sample, and $\bar{y}$ mean of the values of the y variable. Fig. 2 shows the parameter selection steps and diagram of the soft sensor training based on SDP-LIV method.

The following performance indexes are employed to evaluate the performance of the designed soft sensors. These indicators include root mean square error (RMSE), mean absolute error (MAE), the coefficient of determination (R^2), and also adjusted R^2.

$$\text{RMSE} = \sqrt{\frac{1}{N}(y_i - \hat{y}_i)^2} \tag{10}$$

$$\text{MAE} = \frac{\sum_{i=1}^{N}|y_i - \hat{y}_i|}{N} \tag{11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{12}$$

$$R_{adj}^2 = 1 - \left[\frac{\left(1 - R^2\right)\left(N - 1\right)}{\left(N - k - 1\right)} \right] \tag{13}$$

where N is the number of data, K is the number of predictors and y_i, $\hat{y}_i$, $\overline{y}$ and $\overline{\hat{y}}$ are referred to as the real value, predicted value, mean values of y and $\hat{y}$, respectively.

RESULTS AND DISCUSSION

The correlation between input and output variables are determined based on Eq. (9) and results are shown in Table 2. Regarding the presented correlation coefficients, all five secondary variables are selected for both quality variables. Therefore, all input variables are considered as system states, and the respective regressor as one is selected. The SDP-LIV model structure for each product quality (salinity and water cut efficiencies) according to Eq. (3) are expressed in the following forms:

$$y_{1,t} = a_{1,t}\left\{X_1, X_2, X_3, X_4, X_5\right\} \times 1 + e_t \tag{14}$$

$$y_{2,t} = a_{2,t}\left\{X_1, X_2, X_3, X_4, X_5\right\} \times 1 + e_t \tag{15}$$

Following the determination of effective variables based on correlation analysis, the backward elimination method and optimized bandwidth criterion are employed to eliminate unnecessary

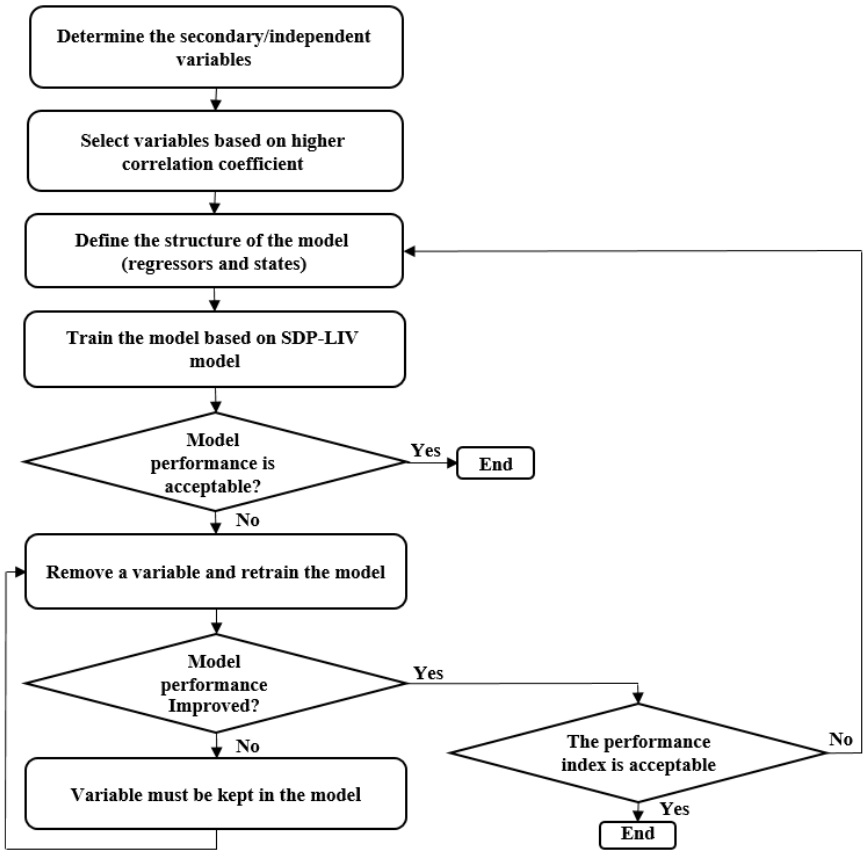

Fig. 2. Schematic diagram of soft sensor training based on the SDP-LIV method

variables from the model.

The model identification steps for each output variable, according to the proposed strategy, are outlined in Fig. 2. Local polynomial regression with a zero-order polynomial is adopted for each state variable. In identifying the output of soft sensor model in each step, the bandwidth ($\lambda_{j,i}$) corresponding to each state variable was determined by the CV optimization method. Then, any variable that had more bandwidth than other variables was removed from the model, and the model was trained again based on the new set of state variables. If the performance indexes of the model are not improved, the removed variable is returned to the model. Ultimately, model structures yielding the best performance indexes for both product qualities are obtained through repeated training iterations using optimized bandwidths. The optimized bandwidths and corresponding performance indexes for the training dataset are tabulated in Tables 3 and 4, with the best model performance highlighted in bold.

The analysis reveals that the removal of any parameter results in a significant escalation in error and a subsequent decline in prediction accuracy. Thus, it is evident that all five input variables exert a strong influence on both product qualities and are indispensable to the model.

Evaluation of the performance indexes of the two soft sensors on the training dataset demonstrates exemplary predictive capabilities. For salinity efficiency, the soft sensor yields impressive values: R = 0.9984, R^2 = 1, R^2adj = 0.9999, RMSE = 0.6663, and MAE = 0.4640. Similarly, the soft sensor's performance for water cut efficiency is remarkable, with R = 1, R^2 = 1, R^2adj = 0.9999, RMSE = 0.1551, and MAE = 0.0674. These metrics underscore the efficacy and reliability of the proposed soft sensor models in accurately predicting product qualities within the DDP system.

Table 2. Pearson correlation coefficients of input variables and between input variables and output variables

	X_1	X_2	X_3	X_4	X_5
X_1	1	0.009	-0.014	0.008	0.008
X_2	-	1	0.001	-0.005	0.001
X_3	-	-	1	-0.001	0.001
X_4	-	-	-	1	0.004
X_5	-	-	-	-	1
y_1	-0.043	0.695	0.456	0.051	0.181
y_2	0.148	0.421	-0.132	0.240	0.261

Table 3. Performance indexes and optimized bandwidths for different sets of input variables in the SDP-LIV model for salinity efficiency (y_1)

Case	Selected states	No. of states	R	RMSE	MAE	R^2	Optimized bandwidths for each state				
1	**X_1, X_2, X_3, X_4, X_5**	**5**	**0.9984**	**0.6663**	**0.4640**	**1**	**0.0000**	**0.1987**	**0.0715**	**0.0563**	**0.1132**
2	X_1, X_3, X_4, X_5	4	0.5795	9.4639	8.1950	0.4323	0.0000	0.0000	1.105×10^6	0.0000	
3	X_1, X_2, X_3, X_4	4	0.9737	2.6429	2.0923	0.9599	0.2615	0.0651	0.0790	0.1272	
4	X_1, X_2, X_4, X_5	4	0.7374	7.8425	5.9566	0.6119	0.5140	0.2870	0.2190	0.2972	
5	X_1, X_2, X_3, X_5	4	0.9766	2.4958	1.9092	0.9888	0.2937	0.2673	0.0770	0.1739	
6	X_2, X_3, X_4, X_5	4	0.8878	5.3398	4.1590	0.9647	0.0000	0.0000	1.038×10^5	0.0000	

Table 4. Performance indexes and optimized bandwidths for different sets of input variables in the SDP-LIV model for water cut efficiency (y_2)

Case	Selected states	No. of states	R	RMSE	MAE	R^2	Optimized bandwidths for each state				
1	**X_1, X_2, X_3, X_4, X_5**	**5**	**1**	**0.1551**	**0.0674**	**1**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0412**
2	X_1, X_2, X_3, X_4	4	0.8670	7.9967	5.7338	0.9663	0.3178	0.1356	0.0008	0.1524	
3	X_2, X_3, X_4, X_5	4	0.8136	9.4249	7.4174	0.9339	0.1292	0.1037	0.2190	0.1967	
4	X_1, X_3, X_4, X_5	4	0.7091	11.5186	8.9757	0.9624	0.3146	0.1185	0.2456	0.2255	
5	X_1, X_2, X_4, X_5	4	0.6889	11.6968	9.4970	0.2211	0.1299	0.3245	0.2149	0.1380	
6	X_1, X_2, X_3, X_5	4	0.8619	8.1616	5.7337	0.9325	0.1336	0.3018	0.0948	0.1196	

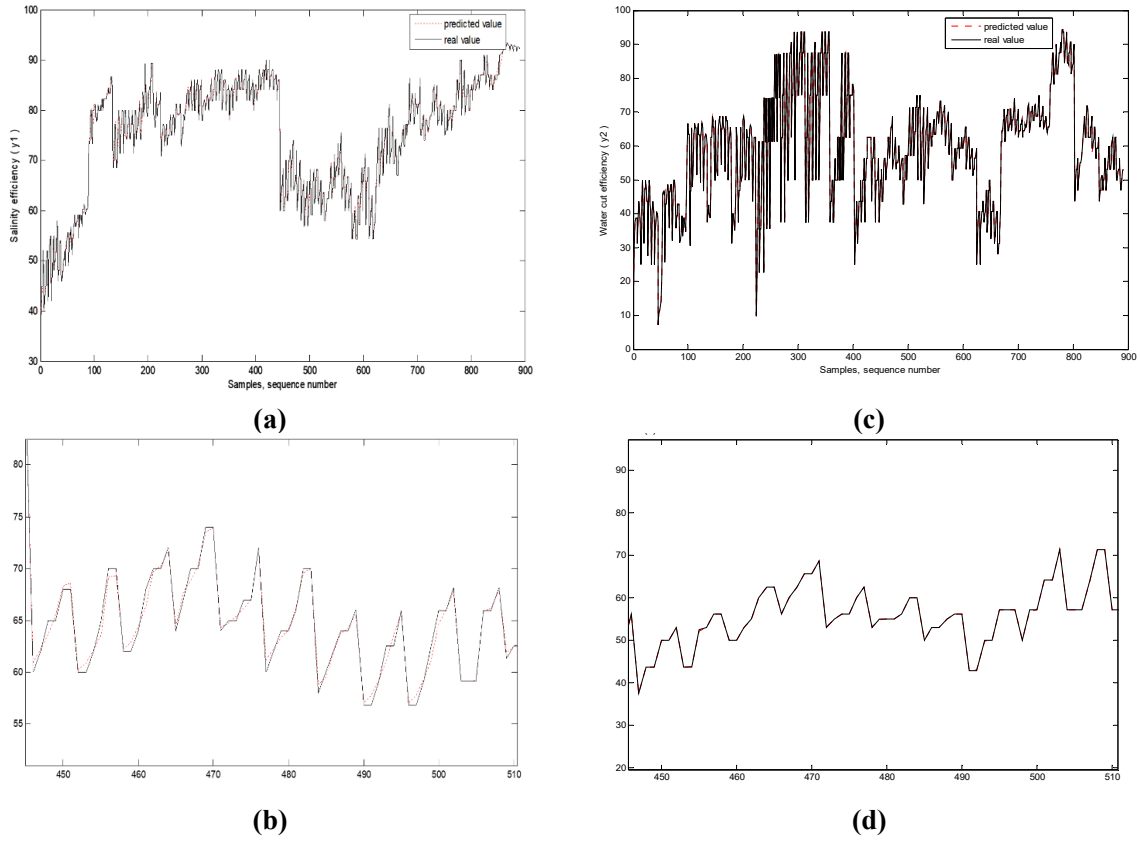

Fig. 3. Prediction results of output variable on the training dataset: (a) Salinity efficiency (y_1), (b) zoom of y_1 for samples 445-510, (c) Water cut efficiency (y_2), (d) zoom of y_2 for samples 445-510

The RMSE and MAE metrics attest to the high accuracy of the soft sensor models, while the R^2 values nearing 1 signify excellent prediction performance. The close alignment between R^2 and adjusted R^2 values indicates a model that perfectly predicts output values. To further scrutinize the prediction performance, plots depicting the model's predictions alongside real data on the training dataset are presented in Fig. 3. These plots illustrate the soft sensor models'

adeptness in tracking the trends of the output variables, underscoring their robust performance.

Moving to the testing dataset, Fig. 4 showcases the prediction results obtained using the SDP-LIV model with optimized bandwidths for each input variable. Additionally, a graphical comparison between the predicted results of the SDP-LIV and ANN soft sensors is provided. Notably, the SDP-LIV model exhibits closer alignment with the real data, particularly evident in the predicted points of water cut efficiency.

Further performance comparisons are depicted in Fig. 5, utilizing scatter plots for the testing dataset. While the prediction results for salinity efficiency demonstrate slight bias within the operating range (Fig. 5a), the data points appear tightly distributed along the diagonal line, indicative of low estimation bias and smaller estimation variance. Conversely, the estimation of water cut efficiency exhibits greater dispersion, indicating comparatively less accurate predictions by the SDP-LIV model (Fig. 5c).

In Fig. 5 (b) and (d), performance comparisons in terms of scatter plots for ANN are depicted alongside those for the SDP-LIV model. Additionally, Table 5 provides a quantitative comparison of the proposed soft sensing model with ANN and MLR soft sensors.

Analysis of the provided performance indexes underscores the stark contrast in prediction performance between the MLR soft sensor and the more complex DDP system. The poor performance of the MLR model can be attributed to the inherently nonlinear behavior of the DDP system. Comparing the results obtained by the ANN model with those of the proposed SDP-LIV technique reveals a notable enhancement in soft sensing performance. Specifically, the RMSE values for salinity and water cut efficiencies show an improvement of approximately 10.5% and 15%, respectively, when utilizing the SDP-LIV method compared to ANN soft sensors. This improvement is further substantiated by the corresponding MAE values for both salinity and water cut efficiencies, with a significant difference observed compared to the MLR method, which operates under a linear model.

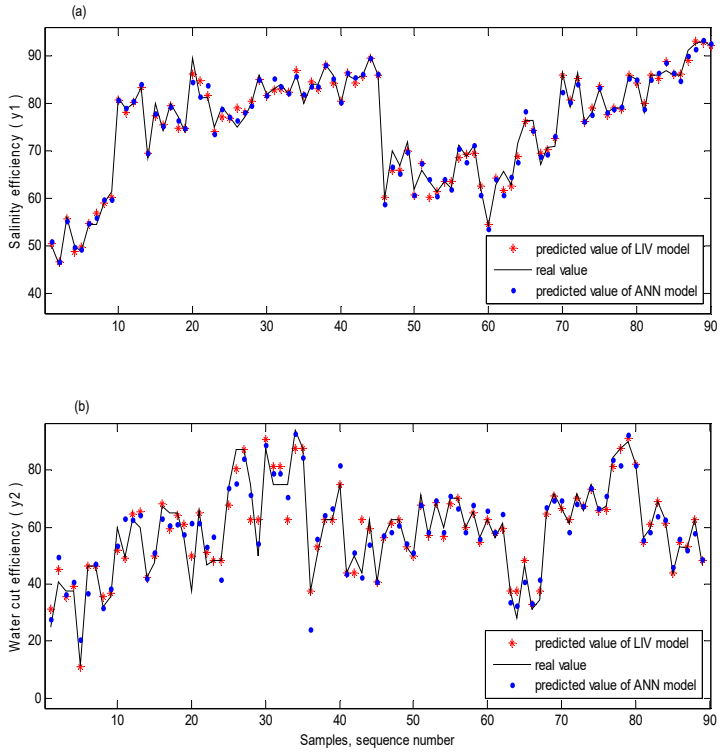

Fig. 4. Prediction results of output variables on the testing dataset: (a) Salinity efficiency (y_1), (b) Water cut efficiency (y_2)

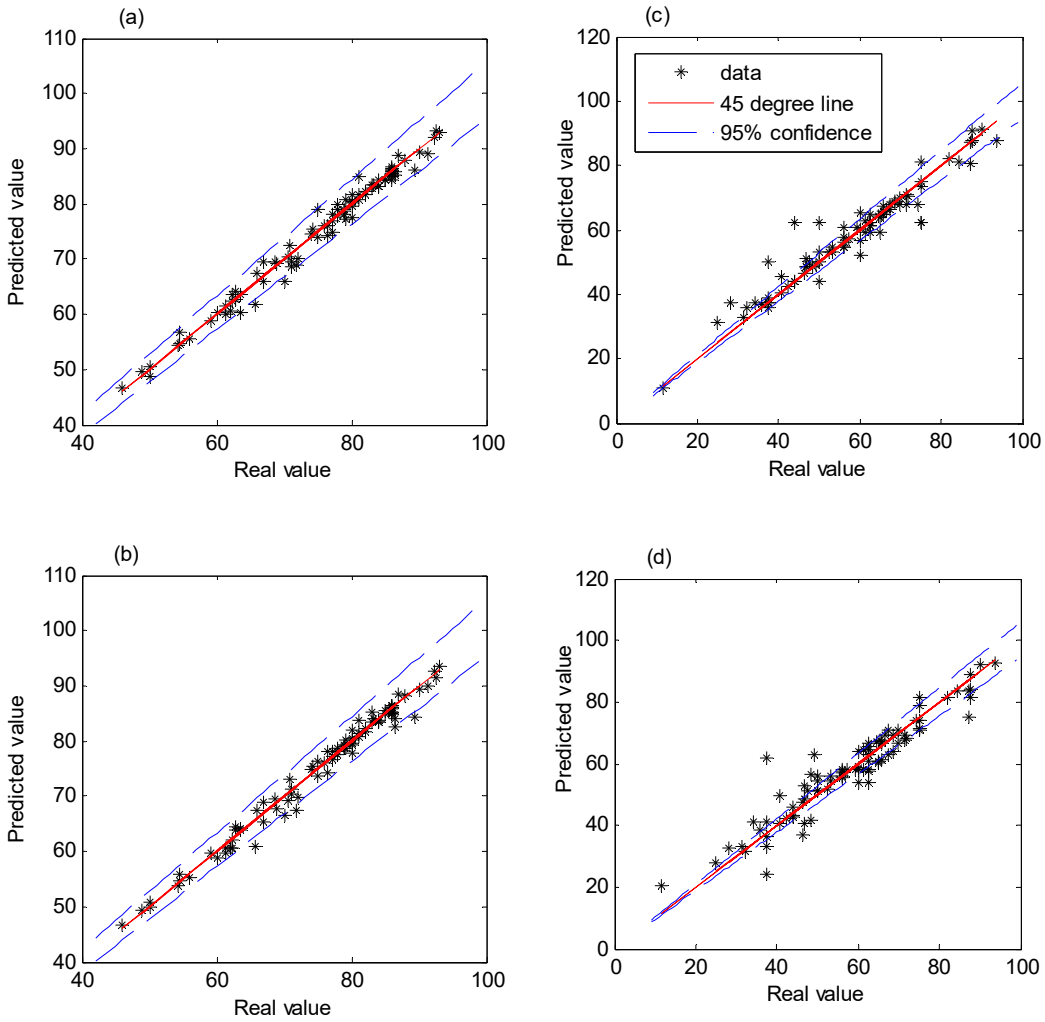

Fig. 5. Predicted and real value of output variables on the testing dataset: (a) and (b) predicted values of salinity efficiency by SDP-LIV model and ANN model respectively, (c) and (d) predicted values of water cut efficiency by SDP-LIV model and ANN model respectively

Table 5. Performance indexes of the DDP system soft sensors on the testing dataset

Publication	Model type	Testing performance indexes									
		R		R^2		R^2_{adj}		RMSE		MAE	
		Salinity Eff	Water cut Eff	Salinity Eff	Water cut Eff	Salinity Eff	Water cut Eff	Salinity Eff	Water cut Eff	Salinity Eff	Water cut Eff
Al-Otaibi et al. (Al-Otaibi et al., 2005)	MLR	0.8523	0.6002	-		-		5.8957	12.9310	4.4645	10.8709
	ANN	Not reported	-	Not reported	-	Not reported	-	1.4638	5.1099	1.0523	3.6175
This work	**SDP-LIV**	**0.9917**	**0.9640**	**1**	**0.9907**	**0.9999**	**0.9901**	**1.3589**	**4.3434**	**0.9424**	**2.4973**

The comprehensive analysis provided by Fig. 4, Fig. 5, and the performance indexes listed in Table 5 collectively illustrate the superior predictive capabilities of the SDP-LIV-based soft sensors compared to ANN soft sensors. This capability is particularly evident in the SDP-LIV model's ability to better track the trend of real values, highlighting its efficacy and accuracy in predicting product qualities within the DDP system.

CONCLUSION

This study introduces a novel soft sensing approach based on the SDP-LIV model, designed to optimize the assessment of desalting and dehydration process efficiencies. Through correlation analyses, we identified five critical secondary parameters—temperature, dilution water percentage, settling/mixing time, and demulsifier dosage—that significantly influence salinity and water cut efficiencies. Using these parameters as predictor variables, our soft sensing analyses culminated in the selection of a robust model incorporating all process variables, yielding the most accurate estimators for salinity and water cut efficiencies. Comparisons with ANN and MLR soft sensors revealed remarkable enhancements. RMSE values for salinity and water cut efficiencies improved by approximately 10.5% and 15%, respectively, when employing our SDP-LIV method compared to ANN soft sensors. Graphical comparisons further highlighted the SDP-LIV model's superior predictive performance, demonstrating its efficacy in outperforming ANN soft sensors. These results underscore the practical significance of our proposed method, offering a streamlined yet highly accurate approach for predicting desalting and dehydration process efficiencies in real-world applications. Furthermore, the ability to accurately monitor and optimize these processes contributes to enhanced pollution control by ensuring more efficient removal of contaminants from crude oil. This study highlights the potential of SDP-LIV based soft sensors to not only improve operational efficiency but also to play a crucial role in environmental protection within the crude oil industry.

GRANT SUPPORT DETAILS

The present research did not receive any financial support.

CONFLICT OF INTEREST

The authors declare that there is not any conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/ or falsification, double publication and/or submission, and redundancy has been completely observed by the authors.

LIFE SCIENCE REPORTING

No life science threat was practiced in this research.

REFERENCES

Abdul-Wahab, S., Elkamel, A., Madhuranthakam, C., & Al-Otaibi, M. (2006). Building inferential estimators for modeling product quality in a crude oil desalting and dehydration process. Chem. Eng. Process. Process Intensif., 45(7), 568-577.

AbdulJalee, E., & Aparna, K. (2016). Neuro-fuzzy Soft Sensor Estimator for Benzene Toluene Distillation Column. Procedia Technol., 25, 92-99.

Al-Otaibi, M. (1999). Experimental investigation of Kuwaiti crude oil desalting/dehydration process. (M.S. Thesis), Kuwait University.

Al-Otaibi, M. B. (2004). Modelling and optimising of crude oil desalting process. (Ph.D. Thesis), Loughborough University

Al-Otaibi, M. B., Elkamel, A., Al-Sahhaf, T., & Ahmed, A. S. (2003). Experimental investigation of crude oil desalting and dehydration. Chem. Eng. Commun., 190(1), 65-82.

Al-Otaibi, M. B., Elkamel, A., Nassehi, V., & Abdul-Wahab, S. A. (2005). A computational intelligence based approach for the analysis and optimization of a crude oil desalting and dehydration process.

Energy fuels, 19(6), 2526-2534.
Bidar, B., Khalilipour, M. M., Shahraki, F., & Sadeghi, J. (2018). A data-driven soft-sensor for monitoring ASTM-D86 of CDU side products using local instrumental variable (LIV) technique. J. Taiwan Inst. Chem. Eng., 84, 49-59.
Bidar, B., Sadeghi, J., Shahraki, F., & Khalilipour, M. M. (2017). Data-driven soft sensor approach for online quality prediction using state dependent parameter models. Chemom. Intell. Lab. Syst., 162, 130-141.
Bidar, B., Shahraki, F., Sadeghi, J., & Khalilipour, M. M. (2018). Soft sensor modeling based on multi-state-dependent parameter models and application for quality monitoring in industrial sulfur recovery process. IEEE Sens. J., 18(11), 4583-4591.
Curreri, F., Fiumara, G., & Xibilia, M. G., (2020). Input Selection Methods for Soft Sensor Design: A Survey. Future Internet, 12(6), 97.
Dadari, S., Rahimi, M., & Zinadini, S. (2016). Crude oil desalter effluent treatment using high flux synthetic nanocomposite NF membrane-optimization by response surface methodology. Desalination, 377, 34-46.
Fan, J. (2018). Local polynomial modelling and its applications: monographs on statistics and applied probability, CRC Press, Routledge.
Fan, J., & Yao, Q. (2003). Nonlinear Time Series: Nonparametric and Parametric Methods Springer-Verlag. New York.
Fortuna, L., Graziani, S., Rizzo, A., & Xibilia, M. G. (2007). Soft sensors for monitoring and control of industrial processes. London: Springer.
Gharehbaghi, H., & Sadeghi, J. (2016). A Novel Approach for Prediction of Industrial Catalyst Deactivation Using Soft Sensor Modeling. Catalysts, 6(7), 93.
Hastie, T. J., & Tibshirani, R. J. (1990). Generalized additive models, volume 43 of Monographs on Statistics and Applied Probability. In: Chapman & Hall, London.
He, Y.-L., Geng, Z., & Zhu, Q.-X. (2015). Data driven soft sensor development for complex chemical processes using extreme learning machine. Chem. Eng. Res. Des., 102, 1-11.
Herceg, S., Andrijić, Ž. U., & Bolf, N. (2019). Development of soft sensors for isomerization process based on support vector machine regression and dynamic polynomial models. Chem. Eng. Res. Des., 149, 95-103.
Hosseinpour, F., Ghader, S., Rahimpour, M. R., & Bagheri, H. (2019). Modification of an industrial crude oil desalting unit by electric mixing to improve the dehydration efficiency. J. Chem. Technol. Metall., 54(1), 124-134.
Jolliffe, I. T. (2002). Principal component analysis (Second ed. ed.): Springer.
Kadlec, P., Gabrys, B., & Strandt, S. (2009). Data-driven soft sensors in the process industry. Comput. Chem. Eng., 33(4), 795-814.
Kamari, A., Bahadori, A., & Mohammadi, A. H. (2015). On the determination of crude oil salt content: Application of robust modeling approaches. J. Taiwan Inst. Chem. Eng., 55, 27-35.
Kanno, Y., & Kaneko, H. (2020). Ensemble just-in-time model based on Gaussian process dynamical models for nonlinear and dynamic processes. Chemom. Intell. Lab. Syst., 203, 104061.
Li, K., Xu, W., Han, Y., Ge, F., & Wang, Y. a. (2019). Soft sensor for the moisture content of crude oil based on multi-kernel Gaussian process regression optimized by an adaptive variable population fruit fly optimization algorithm. Trans. Inst. Meas. Control, 42(4), 770-785.
Liu, J. (2014). Developing a soft sensor based on sparse partial least squares with variable selection. J. Process Control , 24(7), 1046-1056.
Mahdi, K., Gheshlaghi, R., Zahedi, G., & Lohi, A. (2008). Characterization and modeling of a crude oil desalting plant by a statistically designed approach. J. Petrol. Sci. Eng., 61(2–4), 116-123.
Mahdi, K., Gheshlaghi, R., Zahedi, G., & Lohi, A. (2008). Characterization and modeling of a crude oil desalting plant by a statistically designed approach. J. Petrol. Sci. Eng., 61(2-4), 116-123.
Moghadam, R. P., Sadeghi, J., & Shahraki, F. (2021). Soft sensor model for monitoring and online control based on a dynamic model and local instrumental variable technique. Int. J. Modell. Identif. Control, 39(3), 192-203.
Nasehi, S., Sarraf, M. J., Ilkhani, A., Mohammadmirzaie, M. A., & Fazaelipoor, M. H. (2019). Statistical evaluation and Optimization of Crude Oil Desalting Unit: A Case Study of Bandar Abbas oil Refinery. J. Biochem. Technol., 10(2), 59-68.
Pan, H., Su, T., Huang, X., & Wang, Z. (2021). LSTM-based soft sensor design for oxygen content of

flue gas in coal-fired power plant. Trans. Inst. Meas. Control, 43(1), 78-87.
Ranaee, E., Ghorbani, H., Keshavarzian, S., Abarghoei, P. G., Riva, M., Inzoli, F., & Guadagnini, A. (2021). Analysis of the performance of a crude-oil desalting system based on historical data. Fuel, 291, 120046.
Roodbari, N. H., Badiei, A., Soleimani, E., & Khaniani, Y. (2016). Tweens demulsification effects on heavy crude oil/water emulsion. Arabian J. Chem., 9, S806-S811.
Shi, X., & Xiong, W. (2018). LWS based PCA subspace ensemble model for soft sensor development. IFAC-Papers OnLine, 51(18), 649-654.
Sotelo, C., Favela-Contreras, A., Sotelo, D., Beltrán-Carbajal, F., & Cruz, E. (2018). Control Structure Design for Crude Oil Quality Improvement in a Dehydration and Desalting Process. Arabian J. Sci. Eng., 43(11), 6579–6594.
Souza, F. A. A., Araújo, R., Matias, T., & Mendes, J., (2013). A multilayer-perceptron based method for variable selection in soft sensor design. J. Process Control, 23(10), 1371-1378.
Sun, K., Huang, S.-h., Jang, S.-S., & Wong, D. S.-H. (2016). Development of soft sensor with neural network and nonlinear variable selection for crude distillation unit process. Comput. Aided Chem. Eng., 38, 337-342.
Wang, D., Liu, J., & Srinivasan, R. (2010). Data-Driven Soft Sensor Approach for Quality Prediction in a Refining Process. IEEE Trans. Ind. Inf., 6(1), 11 - 17.
Young, P. (1998). Data-based mechanistic modeling of engineering systems. J. Vib. Control, 4(1), 5-28.
Young, P. C. (2011). Recursive estimation and time-series analysis: An introduction for the student and practitioner: Springer Science & Business Media.
Zhao, T., Li, P., & Cao, J. (2019). Soft sensor modeling of chemical process based on self-organizing recurrent interval type-2 fuzzy neural network. ISA Trans., 84, 237-246.
Zheng, K., & Funatsu, K. (2018). Partial constrained least squares (PCLS) and application in soft sensor. Chemom. Intell. Lab. Syst., 177, 64-73.
Zhongda, T., Shujiang, L., Yanhong, W., & Xiangdong, W. (2016). A multi-model fusion soft sensor modelling method and its application in rotary kiln calcination zone temperature prediction. Trans. Inst. Meas. Control, 38(1), 110-124.