



Predicting the concentration of hazardous phenolic compounds in refinery wastewater—a multivariate data analysis approach

Pedro D. A. Bastos^{1,2,3} · Claudia F. Galinha¹ · Maria António Santos² · Pedro Jorge Carvalho³ · João G. Crespo¹

Received: 14 January 2021 / Accepted: 29 July 2021 / Published online: 5 August 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

The present study focused on the methodology for identification of the wastewater stream that presents the highest phenolic impact at a large oil refinery. As a case-study, the oil refinery, Petrogal S.A., in Sines, Portugal, was selected. Firstly, stripped sour water from the cracking complex was identified as the most relevant wastewater stream concerning phenolic emission. Secondly, multivariate data analysis was used, through projection to latent structures (PLS) regression, to find existing correlations between process parameters and phenols content in stripped sour water. The models developed allowed the prediction of phenols concentration with predictive errors down to 20.16 mg/L (corresponding to 8.2% average error), depending on the complexity of the correlation used, and R^2 values as high as 0.85. Models were based in input parameters related to fluid catalytic crackers (FCC) feedstock quality, crudemix and steam injected in the catalyst stripper. The studied data analysis approach showed to be useful as a tool to predict the phenolic content in stripped sour water. Such prediction would help improve the wastewater management system, especially the units responsible for phenol degradation. The methodology shown in this work can be used in other refineries containing catalytic cracking complexes, providing a tool which allows the online prediction of phenols in stripped sour water and the identification of the most relevant process parameters. An optimised system at any refinery leads to an improvement in the wastewater quality and costs associated with pollutant discharge; thus, the development of monitoring online tools, as proposed in this work, is essential.

Keywords Wastewater pollution · Wastewater monitoring · Stripped sour water · Phenols · Projection to latent structures regression · Online prediction

Introduction

Phenols, often referred to as phenolics, are organic compounds that contain phenol in their chemical structure. Such compounds are present in a wide range of industries, including resin manufacturing, oil refineries and plastics. Regarding the composition of wastewaters, phenols play a relevant role in environmental impact. Phenolics are among the several

compounds in industrial wastewater that are known to be harmful especially among the aquatic environment, thanks to their low biodegradability and high toxicity (Villegas et al. 2016). Therefore, national and international environmental entities aimed to set severe legal discharge limits to comply with (Sanches et al. 2017), which have been stimulating the development and implementation of processes that focus on an improvement in the efficiency of current treatments without compromising sustainability. A chart of specific phenolic compounds has emerged as priority pollutants according to the Environmental Protection Agency (EPA) and the EU under the Directive 2008/105/EC, where limits for wastewater have been set regarding discharge, pointing at 0.5 mg/L for surface waters while 1 mg/L being established for sewage systems (ECB 2006; Datta and Uslu 2014).

Refineries use great amounts of raw water with different quality specifications depending on the purpose of the use. Process water, for example, is defined as water that has come in close contact with hydrocarbons in refining processes,

Responsible Editor: Ta Yeong Wu

✉ Claudia F. Galinha
cf.galinha@fct.unl.pt

¹ LAQV-REQUIMTE, DQ, FCT, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal

² Sines Refinery, Petrogal S.A., 7520-952 Sines, Portugal

³ CICECO-Aveiro Institute of Materials, Universidade de Aveiro, Aveiro, Portugal

being subdivided into the following forms of water (IPIECA 2010): sour waters; effluent resulting from desalting processes; storage tank bottoms and spent caustic. Due to a great fraction of the water used in refining that is in direct contact with the processed oil, pollutants including hydrocarbons, sulphides, ammonia and phenolics are carried over and solubilised into the water due to diffusion mechanisms that take place at the oil-water interface. Refinery wastewater contamination is primarily associated with the quality of the feedstock processes combined with the technological level of the units involved in refining.

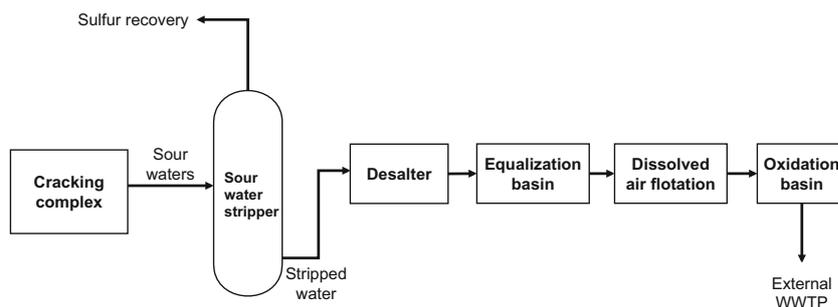
Oily wastewaters that result from water use in refining units are generally referred to as sour waters. These forms of wastewater are treated in sour water stripping units, where stripping steam is injected in a counter current mode with the aim of extracting soluble ammonia, sulphides and volatile compounds in general. It has been reported that there are actually more sources of sour water from refining than in gas processing applications (Addington et al. 2011). Such sources include distillation, wash systems, water from knockouts and amine systems. The most abundant form of sour water has been identified as being from the steam stripping of crude in the atmospheric and vacuum tower processing units. Fluid catalytic crackers (FCC), hydrodesulphurisation units (HDS) and hydrocrackers are also among great generators of sour waters. Sour water comes from the steam, the demineralised water used for thermal cracking and hydrotreating. As heavy products are cracked, sulphur is converted into hydrogen sulphide (H_2S) and mercaptans, and nitrogen compounds are converted into ammonia (NH_3). The formation of phenol results from chemical reactions that take place which involve steam and cyclic hydrocarbons (Leprince 2001). By-products are present in the water such as H_2S , NH_3 , phenolics and hydrogen cyanide (Armstrong et al. 1996). It has been reported (Addington et al. 2011; Stevens and Mosher 2008; Armstrong et al. 1996) that sour water stripping units lack the ability of removing phenolic compounds. This arises as a typical issue present in sour water stripping processes as an efficient form of wastewater treatment in certain refining units, as phenols tend to persist. Units such as cokers, distillation columns and fluid catalytic crackers undergo this particular challenge (Stevens and Mosher 2008); however, in hydrotreaters and desulphurisation units, the phenolic compounds generated in the wastewaters are low enough to not be considered a problem.

Regarding the largest oil refinery in Portugal (Petrogal S. A., Sines), with a capacity of over 200,000 bpd of crude oil processing, the main source of phenols emission, which is stripped sour water generated at the catalytic cracking complex, undergoes a sequence of pre-treatment processes (Figure 1) before the final effluent is discharged to the municipal wastewater treatment plant (WWTP). The first step consists of using stripped sour water as wash water useful for

desalters, where a portion of the phenols is absorbed into the crude phase, through a well-known and reported method for phenol recovery in oil refineries worldwide (Metcalf 1957). The portion which is not absorbed is emitted in the desalter effluent, which is sent to an equalisation basin, mixed with other effluents from the refinery. The wastewater treatment circuit then consists of a dissolved air flotation unit (DAF) followed by an oxidation basin, generating the final effluent that is sent to the municipal WWTP. Figure 2 shows the history of phenolic concentration in the final effluent prior to discharge. Although a relatively steady behaviour over the past few years can be seen, the existence of outliers (>20 mg/L) originates from unexplained or unpredictable occurrences, which generate an effluent richer in phenols with higher costs for the refinery associated with pollutant discharge.

A study that shows correlations between the refinery process conditions and concentration of phenols in refinery wastewater arises, thus, as a potential useful anticipation tool to overcome such unpredictable and complex behaviours that could affect the wastewater treatment process downstream. Different modelling approaches can be great statistical resources in order to prompt additional inputs in advanced control systems of refining units (Zhao et al. 2001; Chi et al. 2014). Projection to latent structures (PLS) regression is a multilinear regression tool, able to correlate several descriptors of the process (input variables) in order to predict output parameters. As a regression tool, PLS is based on experimental data to find the correlations existing within the data set. Additionally, PLS regression uses the variance captured in input variables to make the correlation with the variance captured in outputs (as in principal component analysis), making this tool more suitable to be used when input parameters are correlated with each other (Wold et al. 2001). The use of multivariate data analysis allows to reveal patterns between input data (e.g., process conditions) and the output (e.g., phenols contamination in refinery wastewater). Due to the empirical/statistical character of these types of correlations, the resulting equations are valid for the system studied and in the range of conditions assessed. As shown in a previous work, this multivariate data analysis approach allowed a better understanding of the source of polar oil and grease compounds that are emitted into the wastewater stream and affect the quality of the final effluent discharged to the external WWTP and costs associated (Santos et al. 2013). In that study, the developed model was successfully applied to feed a mass balance estimation to the current refinery wastewater circuit and validated through analytical monitoring. Furthermore, the implemented online mass balance has ever since been used by the refinery environmental department to predict polar oil and grease contamination peaks upstream and act on the optimisation of the wastewater treatment units downstream, such as in the dissolved air flotation unit and oxidation basin.

Fig 1 Wastewater treatment route generated from cracking complex of the Sines Refinery (adapted from Bastos et al. 2020)



In the present work, a phenolic emission evaluation was performed throughout the wastewater circuit of the Sines Refinery to identify the effluent stream that contributed the most in terms of phenols. A detailed physico-chemical characterisation of the most relevant stream was conducted over a representative operational period. After identifying the key-stream, multivariate data analysis was used, through projection to latent structures (PLS) regression, to find existing correlations between oil refining process parameters and phenols content in stripped sour water, for the first time at the Sines refinery.

Materials and methods

Characterisation and assessment of Sines Refinery wastewater streams and emission of phenols

An initial evaluation of the phenols present in the wastewater circuit at the refinery was conducted. Figure 3 shows a schematic representation of several sampling collection points based on the process water involved in relevant refining process defined (IPIECA 2010). Samples were collected in 1-L glass bottles and either readily analysed within 4h after sampling or preserved and stored upon acidification with 2 mL H_2SO_4/L , shown to be the most effective method of preserving phenols in wastewaters (Carter and Huston 1978).

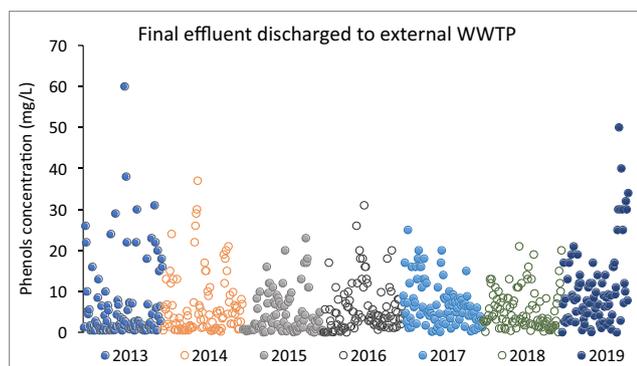


Fig 2 Historical data of phenols discharged to external wastewater treatment plant (WWTP)

Samples of the stripped sour water from the cracking complex of the Sines Refinery were withdrawn and characterised over a 24-month period in terms of the most relevant parameters that compose a key effluent prior to treatment. A continuous characterisation of the wastewater over a representative operational period shows potential advantages in terms of planning, control and treatment of the wastewater, allowing the possibility of witnessing variations in the effluent that could affect treatment performance.

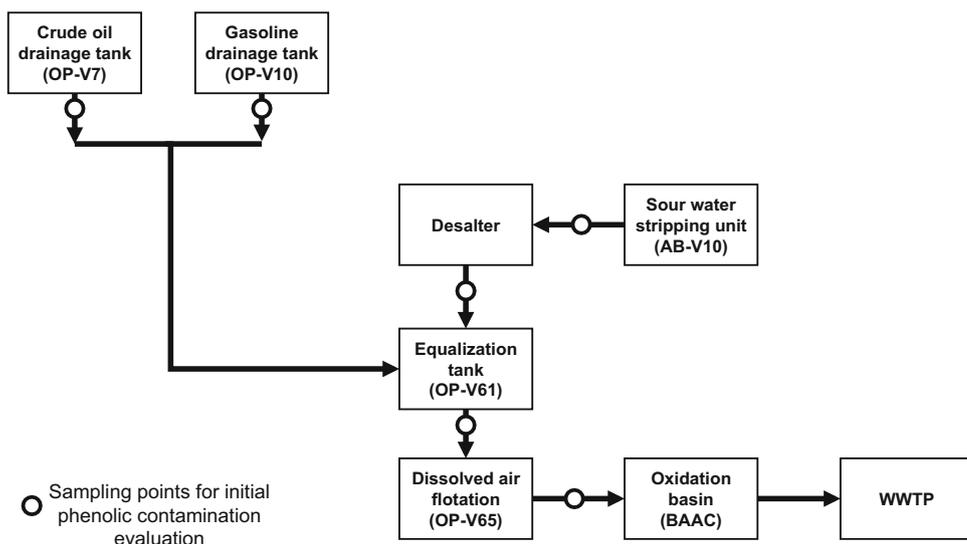
Standard methods were used for determining phenols and oil and grease concentration (O&G) (SMEWW 5530D and SMEWW5520 C/F, respectively), and COD was analysed using a LCI 500 cuvette kit (Hach Lange, GmbH, Germany) in agreement with ISO 15705 using the standard method SWEWW 5220 B. Total dissolved solids (TDS) and conductivity measurements were obtained by means of a portable conductivity meter (WTW, Germany), and pH measurements were conducted using a Basic 20 PH meter (Hach).

The determination of phenolic content was performed using the standard methods for wastewater characterisation, method 5530D (Eaton et al. 1998), upon the general principle that steam-distillable phenolic compounds react with 4-aminoantipyrine at $pH\ 7.9 \pm 0.1$ in the presence of potassium ferricyanide to form a coloured antipyrine dye. The dye is kept in aqueous solution, and the absorbance is measured at 507 nm. Interferences are eliminated or reduced to a minimum by using the distillate from the preliminary distillation procedure (Eaton et al. 1998).

Refinery process data

Acquisition of operational information regarding the refining process was accomplished through the online control system of the refinery, namely the data regarding the mass quantities of steam injected in the FCC catalyst stripper coupled to the riser reactor (VAP_{react}). Data regarding the quality of the crudemix processed, such as crudemix acidity (CM_{acid}) and crudemix density (ρ_{CM}), as well as the FCC feedstock quality, namely density (ρ_{feed}), levels of sulphur (%S) and mass flowrate (Qm_{feed}) were collected from an existing linear model for the refinery that reports (Santos et al. 2013) mass quantities and several properties of the crude oil processed,

Fig 3 Schematic representation of wastewater circuit at the Sines Refinery with sampling points (open circles) for phenols analysis



including its acidity and density, and different feedstock qualities fed to the units of the cracking complex.

The parameters initially assessed as possible contributors (inputs) to phenols contamination in wastewater are summarised in Table 1. Data regarding process was collected for the same refining programme period, in which a sample of treated water of the sour water stripping unit of the cracking complex was withdrawn, for a total of 62 statistical observations. These collections have been reported to be typically in line with the minimum number of observations necessary for adequate industrial data modelling (Forina et al. 2004).

Sour water stripping unit–Development of PLS models

Projection to latent structures (PLS) regression was used to correlate process parameters (the inputs) with phenols

contamination (the output). PLS regression results in multilinear models able to describe one output based on the inputs. In this work, the data collected was divided in two sets, the training set (75% of the observations were used to calibrate the model) and the external validation set (25% of the observations were used to validate the model). These data sets were divided randomly, while it was ensured that the validation set was well distributed and within the training set range of values to avoid extrapolation. The same data sets of training and validation were used to calibrate and optimise the models developed using different input approaches. All data used in PLS models were previously normalised through autoscaling (subtracting the average and dividing by standard deviation).

The calibration of the models and the selection of the latent variables (LV) were done through leave-one-out

Table 1 Variables studied as inputs for PLS modelling of phenolic concentration in stripped sour water (output)

Model role	Variables	Parameter abbreviation	Range	Average	Standard deviation
Inputs	FCC feedstock density (kg/m ³)	ρ_{feed}	[0.913; 0.927]	0.92	0.0034
	FCC feedstock mass flowrate (ton/d)	Q_{mfeed}	[6083; 6186]	6135	22.26
	Steam injected in FCC catalyst stripper (ton/h)	VAP_{react}	[0.56; 5]	2.4	1.19
	RAT in FCC feedstock (ton/d)	RAT_{feed}	[617; 4390]	3265	790
	VGO in FCC feedstock (ton/d)	VGO_{feed}	[1554; 5511]	2676	776
	UCO in FCC feedstock (ton/d)	UCO_{feed}	[72; 431]	109	63
	Crudemix acidity (mg KOH/g ¹)	CM_{acid}	[0.05; 0.42]	0.19	0.09
	Naphtha in FCC feedstock (ton/d)	$NAPH_{\text{feed}}$	[56; 88]	72	9.8
	Naphtha density fed to FCC (kg/m ³)	ρ_{NAPH}	[0.67; 0.675]	0.67	0.0012
	Crudemix density (kg/m ³)	ρ_{CM}	[0.82; 0.87]	0.85	0.014
Output	Phenols concentration in stripped sour water (mg/L)	$\text{Phenols}_{\text{SWS}}$	[112; 376]	196	59.8

RAT, Atmospheric residue; VGO, vacuum gas oil; UCO, unconverted oil

cross-validation, using only the training data set. To assess the quality of the models, the root mean squared error of cross-validation (RMSECV) was used:

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^N (y_{\text{mod}} - y_{\text{exp}})^2}{N}}$$

where y_{mod} was the output value from the model and y_{exp} the output experimental value for each observation leaved out during cross-validation, and N the total number of experimental observations in the training data set. Cross-validation also allowed the calculation of the uncertainty associated with each input, for each model, through Jackknife standard deviations (Duchesne and MacGregor 2001).

External validation of the models was done through the assessment of the root mean squared error of prediction (RMSEP), calculated as the RMSECV, but for the external validation data set (where N is the number of observations in the validation data set). Optimisation of the mathematical models was done through the selection of inputs. This was done aiming at improving the quality of the model and to assess which inputs have a real impact for the output prediction. Selection of useful inputs was performed initially based on the regression coefficient of the model (inputs with higher impact on the output prediction have higher regression coefficient). Subsequently, quadratic and interaction terms were also added to the models to improve prediction. Therefore, instead of a multilinear regression, the models obtained were in the form of the following equation:

$$y = ax_1 + bx_2 + cx_3 + \dots + dx_1^2 + ex_2^2 + fx_3^2 + \dots + gx_1x_2 + hx_1x_3 + ix_2x_3 + \dots$$

where y is the output and x_i the inputs. Selection of quadratic and interaction terms was done by running all combinations of inputs with 5 to 12 inputs, and the models selected based on RMSECV and RMSEP.

All computational routines and algorithms were implemented using GNU Octave.

Results and discussion

Identification of most relevant wastewater streams containing phenol

In order to determine the wastewater stream representing the major source of phenols emission within the refinery, different relevant sampling points were withdrawn throughout the wastewater treatment circuit and were analysed in terms of their phenolic content. Figure 4 shows the mass contribution of the phenols (kg/h) emitted in each of the referred sampling

points, calculated resorting to the experimentally determined concentration of phenols in wastewater (mg/L) and the corresponding flowrates (m^3/h), using average values over a 6-month period. This allows a comparison of different contributions of phenols in terms of mass within relevant points of the wastewater treatment circuit, in order to identify the most relevant wastewater stream.

It can be observed that the effluent that has the greatest phenolic contribution of the sampling points studied is the treated water of the sour water stripping unit from the cracking complex, with mass flow rates reaching over 17.4 kg/h. These large phenol contributions are in line with the refining processes that generate sour waters rich in phenols (Addington et al. 2011; Leprince 2001; Armstrong et al. 1996) which are all routed to a stripping process that has a negligible effect on phenol removal.

Table 2 shows the physico-chemical characterisation of the Sines Refinery stripped sour water from the cracking complex in terms of pH, TDS, conductivity, phenols and COD over a 24-month sampling period from 2017 to 2019. It is possible to observe, from the lower standard deviations and standard errors that pH, TDS and conductivity were the parameters that showed the most consistent behaviour over the 2 years, compared with phenols and COD, with pH values hardly dropping below 7 or exceeding 9, giving the wastewater neutral characteristics. This may be due to the amount of residual ammonia (basic) and sulphides (more acidic) that remain in the wastewater, after stripping, cancelling each other out in terms of pH. The constantly low value of dissolved solids and conductivity compared with the other types of wastewater in the refinery are due to the fact that the water that constitutes sour water generated in the cracking complex, is originated from the demineralised water used in the overhead column of the FCC unit. Demineralised water is always used to avoid fouling due to deposition of ammonia salts on the top of the column. Therefore, the amount of salts in the stripped sour water is practically non-existent, and the dissolved solids are mainly due to dissolved phenols present. Phenols concentration ranged between 112 and 376 mg/L, presenting an average concentration of 195.7 mg/L and a standard deviation of

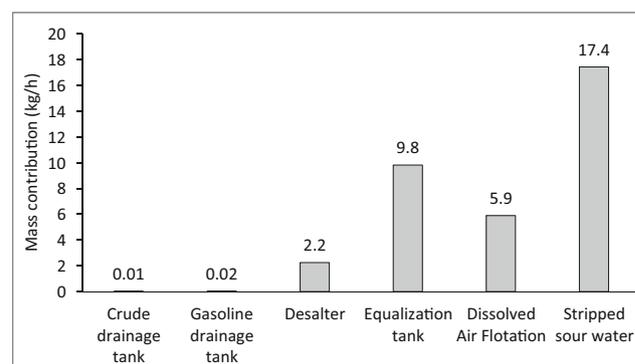


Fig 4 Mass contribution of phenols emitted in the studied wastewater circuit sampling points

Table 2 Experimental characterisation of the Sines Refinery stripped sour water from the cracking complex for sampling from 2017–2019

	pH	TDS (mg/L)	Conductivity ($\mu\text{S}/\text{cm}^1$)	COD (mgO_2/L)	Phenols (mg/L)
Min	6.70	38	85.8	525	112
Max	10.42	144	335.0	6067	376
Average	8.03	65.45	159.4	1241.9	195.7
Standard deviation	0.71	17.29	56.4	1029.4	59.8
Standard error	0.10	2.33	7.61	176.54	7.59

59.8 mg/L. This low variability, in comparison for instance with the amount of polar oil and grease compounds reported (Santos et al. 2013) in the kerosene spent caustic of the Sines Refinery, could show a potential advantage of a trustworthy wastewater treatment process which can operate at steady performance without fluctuations in feedstock that could eventually affect optimum treatment. The behaviour of the experimentally determined phenols concentration in stripped sour water throughout the years of 2017–2019 can be found in Figure S1, Supporting Information.

Sour water stripping unit–PLS modelling for phenolic content

Once the stripped sour water stream was identified as having the greatest phenolic contribution of the refinery, a PLS model was developed in order to predict the content of phenolic compounds in the effluent of the sour water stripping unit. A multilinear combination of the inputs chosen was attempted in the development of a PLS model. These inputs were selected based on the process variables related to upstream units that are directly related to the stream fed to the sour water stripping unit of the cracking complex. Most of the inputs focus on the FCC feedstock quality and steam injected in the riser reactor stripper, where phenols are primarily formed from the reaction between cyclic hydrocarbons and steam (Leprince 2001; Armstrong et al. 1996). In a first approach, over 50 different input combinations were selected as screening in the study of their relevance and usefulness to develop a PLS model based only on multilinear combinations. The best model obtained through these combinations is shown in Table 3, where it is possible to observe that the linear model only captured 55% of the data variance and the poor predictive ability of the model for the training and validation set. However, the screening study allowed to narrow down some of the inputs initially selected, where variables related to the FCC feedstock quality and steam injected in the catalyst stripper indeed ended up playing a major role. A graph of the best linear PLS model obtained with the variables chosen, and their weights in the prediction, can be found in Figure S2, Supporting Information.

Considering these results, in a second approach, it was decided to include quadratic and interaction terms of these

operational input data values aiming to improve the prediction correlation in the development of the PLS model. Additionally, two more inputs were added to the ones selected by the linear model: FCC feedstock density and the atmospheric residue (RAT) of FCC feedstock. Despite these two inputs not being selected by the multilinear modelling approach, they were found to be of high relevance for output estimation when used in other input combinations (data not shown); thus, they were also used as inputs in non-linear models. Table 4 shows the best two non-linear models achieved through inputs selection, where it is possible to clearly observe an improvement in the captured variance and coefficients of determination in training and validation sets. Although the number of terms increased compared to the multilinear model, they resulted from the multiplication or squared of the existing variables (7 input variables), not by new data inputs. Such increase in the number and complexity of terms used in the model allowed the extraction of the information (relative to the output) contained in the inputs. One of the inputs that was initially included in the linear model (crudemix density) ended up being withdrawn from the multilinear model, as it showed to have a minor impact in the improvement of model and is not directly related to the FCC process but rather the very first stream fed to the refining process, that could affect the feedstock quality downstream.

Figure 5 shows the comparison between the PLS model estimations and the experimental data for the concentration of phenols in stripped sour water of the cracking complex, for the most accurate PLS model obtained using quadratic terms (the model with 11 terms shown in Table 4). A graph of the non-linear PLS model with 9 terms obtained can be found in Figure S3, Supporting Information.

An interesting agreement was found, where a coefficient of determination of 0.85 was obtained for the training and validation set. Furthermore, the training and validation set presented slopes of 0.849 and 0.976, respectively, indicating an interesting correlation even though data showed to be relatively dispersed. As was witnessed for the linear model, practically, all of the most relevant inputs originated from the FCC feedstock quality, with atmospheric residue (RAT) (Abul-Hamayel 2002; Arandes et al. 2008) and vacuum gas oil (VGO) composing over 80% wt of the highly complex stream fed to the FCC unit. Apart from the cyclic

Table 3 Linear PLS model analysis of captured variance, errors and training set and validation set coefficients of determination

Model role	Variables	Captured variance (%)	RMSECV (mg/L)	RMSEP (mg/L)	Training set coefficient of determination	Validation set coefficient of determination
Inputs	VGO _{feed} UCO _{feed} VAP _{react} Q _m _{feed} ρ_{CM} CM _{acid}	55.3	48.33	31.10	0.55	0.62
Output	Phenols _{SWS}					

RMSECV, root mean squared error of cross-validation; *RMSEP*, root mean squared error of prediction; *VGO*, vacuum gas oil; *UCO*, unconverted oil; *VAP*, steam injected; *Q_m*, mass flowrate; ρ_{CM} , crudemix density; *CM_{acid}*, crudemix acidity

hydrocarbons that exist in heavy coke accumulated on the FCC spent catalyst, which in the presence of steam are stripped and form phenols, phenolic compounds may also exist in FCC feedstock (Bai et al. 2019), most likely due to oxygen atoms being naturally bonded to cyclic hydrocarbons that compose the feed. The inputs that showed the most relevance for the model (atmospheric residue (RAT) and vacuum gas oil (VGO)) are part of the FCC feedstock quality and can easily be obtained from the manufacturing programmes of refineries whenever a certain type of crudemix is processed, and therefore be directly used for these types of models, without the need of extra characterisations or lab work.

The usefulness of the model obtained as a predictive anticipation tool arises from the route that stripped sour water takes (Figure 1) before being sent to the municipal WWTP. The first and most decisive step of the route is using the effluent as wash water for the desalters, where 75–88% of the phenols (experimentally determined, based on phenolic concentration in desalter effluent) are absorbed back into the crude phase. However, heavier loads of phenols lead to a higher phenolic content at the final effluent (as it can be seen from the outliers in Figure 2), which can lead to a sensible assumption that a decrease in desalter efficiency occurs whenever a greater emission of phenols takes place in stripped sour water. The model would allow the prediction of the concentration of

phenols in stripped sour water before it is fed to the desalter and in case of a heavier load, alert the wastewater treatment units downstream. Although there is not a unit specifically designed to remove phenols from this wastewater in Sines' refinery, the downstream unit most responsible for decreasing the phenol concentration is the oxidation basin, where hydrogen peroxide is fed to oxidize the pollutants present. Therefore, depending on the amount of phenols emitted, the adjustment of hydrogen peroxide can be optimised, leading to an improved management of such reagent, which from a sustainable point of view, can be worthwhile. This means the existence of such model allows the further development of a control system for hydrogen peroxide to be added in the oxidation basin.

Although other refineries may present other forms of wastewater treatment for the effluents generated, it is safe to affirm that the methodology developed in this work (although not the same equations) can be applied to refineries that contain a catalytic cracking complex, and can provide useful information regarding the phenolic content in stripped sour water, without the need of phenolic determination through lab work, thus avoiding the use of resources while saving time.

Table 4 Non-linear PLS model analysis of captured variance, errors and training set and validation set coefficients of determination

No. of terms	Captured variance (%)	RMSECV (mg/L)	RMSEP (mg/L)	Training set coefficient of determination	Validation set coefficient of determination
9	82.7	30.98	22.74	0.827	0.832
11	84.8	30.89	20.16	0.848	0.845

RMSECV, root mean squared error of cross-validation; *RMSEP*, root mean squared error of prediction

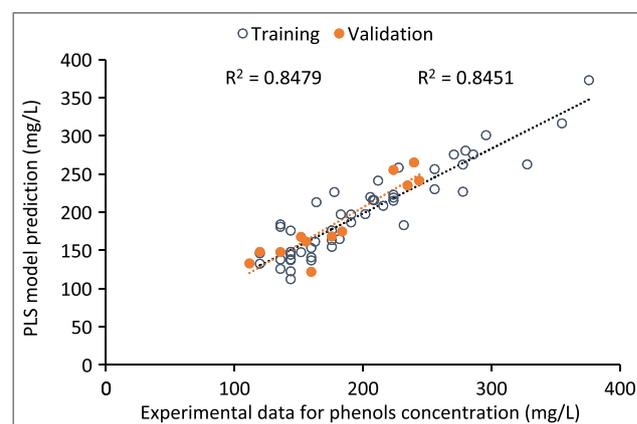
**Fig 5** Non-linear PLS model, with 11 terms, predictions as a function of the experimental data for phenols concentration in stripped sour water from the refinery catalytic cracking complex

Figure 6 a and b show the regression coefficients (for normalised data) of the model inputs for the model with 9 terms and 11 terms, respectively. The low uncertainty found for all inputs (error bars in Figure 6) show low variability of the correlation between each input and the output across training data, which results in robust models. It is possible to observe that, for both models, the mass flowrate of RAT (RAT_{feed}) and its combination with the VGO mass flowrate (VGO_{feed}) that compose the FCC feedstock flowrate showed to be the most relevant, presenting greater regression coefficients than the remaining inputs studied. The mass flowrate of the feedstock that enters the FCC unit itself (Qm_{feed}) also showed a particular relevance for the best model obtained. An interesting observation that can also be withdrawn is that the amount of steam injected in the FCC catalyst stripper (VAP_{react}), that is known to be directly related to the formation of phenol upon stripping of heavy coke, ended up playing a less active role in comparison with the FCC feedstock quality itself. It is therefore possible that most of the phenolic compounds emitted in the stripped sour water actually originated from the cyclic hydrocarbons with oxygen atoms naturally bonded to them that compose the stream fed to the FCC unit, and not so much from the phenols formed from the reaction occurring in the FCC catalyst stripper.

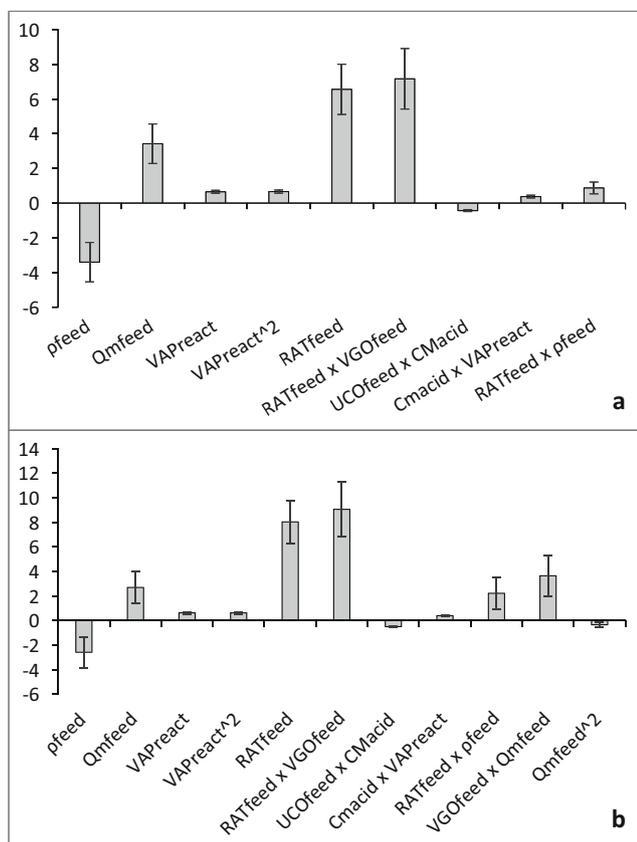


Fig 6 Regression coefficients of the model inputs with the 9-term model (a) and 11-term model (b) for phenols concentration in stripped sour water, for normalised data

The use of non-linear inputs for an insightful PLS model proved to be inevitable, as the inclusion of only linear inputs (such as VGO_{feed} , UCO_{feed} and CM_{acid}) did not improve the model. However, the results achieved with the use of quadratic and interaction terms in non-linear models show that the relationship between the output and the inputs studied is complex and not easily disclosed by linear correlations. One of the inputs that varied the most in terms of relevance and regression coefficient was the mass flow rate of atmospheric residue in the FCC feedstock (RAT_{feed}), which greatly improved the model when added as new input, with coefficients of determination increasing from 0.77 (data not shown) to 0.84.

Conclusion

The first objective of the present study was the evaluation and identification of the most relevant wastewater stream in terms of phenolic impact at the largest Portuguese oil refinery, Petrogal S.A. in Sines, Portugal. It was observed that the effluent presenting the greatest phenolic contribution was the stripped sour water stream of the cracking complex, which is in line with the typical refining processes that generate sour water rich in phenols.

After identifying the most relevant stream, the second objective was to use multivariate data analysis to correlate the process parameters with the concentration of phenols in stripped sour water.

- Seven process parameters, mostly representing the FCC feedstock quality and steam injected in the riser reactor, were identified as necessary to estimate phenols in non-linear correlations.
- These models can predict phenols concentration with a predictive error of 22.74 or 20.16 mg/L, depending on the complexity of the correlation used.
- Above all, an anticipation tool was developed through prediction of phenols in the referred wastewater, which would help control and optimise the efficiency of the current effluent treatment units, leading to an improvement in the quality of the final wastewater stream and to reduced costs associated with the discharge to the external wastewater treatment plant.

Furthermore, the work developed in the present study arises as useful for other refineries which contain catalytic cracking complexes, as it provides the methodology to develop a tool specific to a certain refinery, which allows the determination of phenols in the stripped sour water and the impact of different input process parameters. The usefulness of such information ranges from the possibility of optimising the on-site conditions for an efficient phenol degradation, to the

avoidance of using resources associated with lab work for the determination of phenols in stripped sour water.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11356-021-15785-3>.

Author contribution PDAB did the investigation, visualisation and writing (original draft preparation). CFG developed the methodology, was involved in supervision, resources, investigation, visualisation and writing (review and editing). PJC was involved in writing (review and editing) and supervision. MAS was involved in conceptualisation, resources and supervision. JGC was involved in conceptualisation, methodology, supervision and writing (review and editing). All authors read and approved the final manuscript.

Funding This work was supported by the Associated Laboratory for Sustainable Chemistry-Clean Processes and Technologies-LAQV and by CICECO-Aveiro Institute of Materials, which are financed by Portuguese national funds from Fundação para a Ciência e Tecnologia (Portugal) (UIDB/50006/2020 and UIDP/50011/2020) and co-financed by the ERDF under the PT2020 Partnership Agreement (POCI-01-0145-FEDER-007265). Galp and Fundação para a Ciência e Tecnologia (Portugal) funded the PhD grant PD/BDE/128604/2017 (P. D. A. Bastos) under a PhD project within an industrial environment and Doctoral Program in Refining, Petrochemical and Chemical Engineering (EngIQ). Fundação para a Ciência e Tecnologia (Portugal) funded the contract Investigador FCT 2015, contract number IF/00758/2015 (P. J. Carvalho).

Data availability The dataset collected at Galp and analysed during the current study is not publicly available but is available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate Not applicable

Consent for publication Not applicable

Competing interests The authors declare no competing interests.

References

- Abul-Hamayel MA (2002) Atmospheric residue as feedstock to high-severity fluid catalytic cracking. *Pet Sci Technol* 20:497–506. <https://doi.org/10.1081/ft-120003574>
- Addington L, Fitz C, Lunsford K, Lyddon L, Siwek M (2011) Sour water: where it comes from and how to handle it. *Digital Refining*. <https://www.digitalrefining.com/article/1000741/sour-water-where-it-comes-from-and-how-to-handle-it>. Accessed 8 July 2021
- Arandes JM, Torre I, Azkoiti MJ, Ereña J, Bilbao J (2008) Effect of atmospheric residue incorporation in the fluidized catalytic cracking (FCC) feed on product stream yields and composition. *Energ Fuel* 22:2149–2156. <https://doi.org/10.1021/ef800031x>
- Armstrong T, Scott B, Taylor K, Gardner A (1996) Sour water stripping. *Today's Refinery*. In: *Refining Details Notebook*, USA.
- Bai P, Etim UJ, Yan Z, Mintova S, Zhang Z, Zhong Z, Gao X (2019) Fluid catalytic cracking technology: current status and recent discoveries on catalyst contamination. *Catal Rev* 61:333–405. <https://doi.org/10.1080/01614940.2018.1549011>
- Bastos PDA, Santos MA, Carvalho PJ, Crespo JG (2020) Reverse osmosis performance on stripped phenolic sour water treatment – a study on the effect of oil and grease and osmotic pressure. *J Environ Manag* 261:110229. <https://doi.org/10.1016/j.jenvman.2020.110229>
- Carter MJ, Huston MT (1978) Preservation of phenolic compounds in wastewater. *Environ Sci Technol* 12:309–313. <https://doi.org/10.1021/es60139a007>
- Chi Q, Fei Z, Zhao Z, Zhao L, Liang J (2014) A model predictive control approach with relevant identification in dynamic PLS framework. *Control Eng Pract* 22:181–193. <https://doi.org/10.1016/j.conengprac.2013.02.010>
- Datta D, Uslu H (2014) Investigation of extraction of phenol from wastewater using N, N - Didodecyl-1-dodecanamine (Tridodecylamine) in Benzene. *J Chem Eng Data* 59:3858–3862. <https://doi.org/10.1021/je500757n>
- Duchesne C, MacGregor JF (2001) Jackknife and bootstrap methods in the identification of dynamic models. *J Process Control* 11:553–564. [https://doi.org/10.1016/S0959-1524\(00\)00025-1](https://doi.org/10.1016/S0959-1524(00)00025-1)
- Eaton AD, Clesceri LS, Greenberg AE, Franson MAH (1998), Standard methods for the examination of water and wastewater. American Public Health Association, American Water Works Association, Water Environment Federation, Washington DC
- ECB (2006) European Union Risk Assessment Report - Phenol. European Commission, Italy
- Forina M, Lanteri S, Cerrato Oliveros M, Pizarro C (2004) Selection of useful predictors in multivariate calibration. *Anal Bioanal Chem* 380:397–418. <https://doi.org/10.1007/s00216-004-2768-x>
- IPIECA (2010) Petroleum refining water/wastewater use and management. IPIECA Operations Best Practice Series. <https://www.ipieca.org/resources/good-practice/petroleum-refining-water-wastewater-use-and-management/>. Accessed 8 July 2021
- Leprince P (2001) Water treatment. In: *Petroleum refining vol 3 Conversion Processes*. Editions Technip, Paris
- Metcalf RJ (1957) Process for phenol recovery and crude oil desalting. Pittsburg, US Patent, US2785120A
- Sanches S, Fraga M, Silva N, Nunes P, Crespo J, Pereira V (2017) Pilot scale nanofiltration treatment of olive mill wastewater: a technical and economical evaluation. *Environ Sci Pollut Res* 24:3506–3518. <https://doi.org/10.1007/s11356-016-8083-1>
- Santos B, Galinha CF, Crespo JG, Santos MA, Velizarov S (2013) Prediction of polar oil and grease contamination levels in refinery wastewater through multivariate statistical modeling. *Sep Purif Technol* 119:51–57. <https://doi.org/10.1016/j.seppur.2013.09.009>
- Stevens DK, Mosher A (2008) Fundamentals of sour water stripping. KNP Group, Presented at the Brimstone Sulfur Symposium, Vail, Colorado, September 23. <https://www.kps-engr.com/technical-library.html>. Accessed 8 July 2021
- Villegas LGC, Mashhadi N, Chen M, Mukherjee D, Taylor KE, Biswas N (2016) A short review of techniques for phenol removal from wastewater. *Curr Pollut Rep* 2:157–167. <https://doi.org/10.1007/s40726-016-0035-3>
- Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst* 58:109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1)
- Zhao H, Guiver J, Neelakantan R, Biegler LT (2001) A nonlinear industrial model predictive controller using integrated PLS and neural net state-space model. *Control Eng Pract* 9:125–133. [https://doi.org/10.1016/S0967-0661\(00\)00101-5](https://doi.org/10.1016/S0967-0661(00)00101-5)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.