Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 54-21 (2021) 180–185

An industrial implementation of a C4 hydrocarbon soft sensor to optimise a debutaniser column

Stefan Botha [*,2] **Ian K. Craig** [**,1]

[*] *Analyte Control, Pretoria, South Africa.*
[**] *Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria, South Africa.*

Abstract: The bottoms product of a debutaniser column in a Fischer-Tropsch refining catpoly unit should be maximised to ensure optimal operation of the downstream units. An accurate estimate of the C4 hydrocarbons in the bottoms product is required to ensure that the specification is not violated. This work demonstrates a practical implementation of a soft sensor to estimate the $\%C_4$ material in the bottoms product of the debutaniser using the General Distillation Shortcut (GDS) method and a random forest (RF) machine learned model. The paper highlights practical challenges when deploying a soft sensor to an industrial plant. It is shown how the GDS method soft sensor had to be refitted after unit maintenance was carried out. In comparison the RF model soft sensor uses more reliable measurements and did not require refitting after unit commissioning. Both soft sensors performed well and the choice of soft sensor depends on the available measurements and measurement reliability.

Keywords: industrial applications of process control, machine learning, petrochemical refining, random forests, soft sensor.

1. INTRODUCTION

A catalytic polymerisation (catpoly) process uses an oligomarisation process to transform butene and propylene into a petrol product (Leprince, 2001). The product has a high research octane number (RON) and is predominantly olefinic. The unhydrogenated petrol product is hydrogenated to produce paraffins which cause a decrease in RON (De Klerk, 2012).

In a Fischer-Tropsch refining catpoly unit the feed composition varies significantly and the feed frequently has heavier components than butene and propylene. The debutaniser considered in this paper has two feed streams. A lighter feed stream (C_4 rich) and a heavier feed (C_3 - C_7) (De Klerk, 2012). To remove the C_5 and heavier material the two feed streams are routed to a debutaniser column. The overheads product of the debutaniser flow to the catpoly reactor feed tanks while the bottoms product goes to an isomerisation unit as shown in Fig 1.

The optimal extraction of C_5 material in the bottoms of the debutaniser is crucial for two reasons. Firstly, in the oligomarisation reaction the C_5 material reacts with C_4 and lighter material which produce hydrocarbon chains that yield lower RON components when hydrogenated (De Klerk et al., 2004). Secondly, the C_5 material can be routed to an isomerisation unit which produces a higher value RON petrol blending component as discussed in De Klerk (2012) and Olivier (2017).

Due to the monetary benefit in maximising the bottoms product flow, an industrial linear model predictive controller (MPC) is used to optimise the debutaniser column. A soft sensor is used for the bottoms $\%C_4$ material since there is no on-line analyser available and lab samples are only taken every 12 hours.

This paper describes the initial $\%C_4$ soft sensor which is designed using a first principle model where the parameters are fit using historical process data. The soft sensor is developed using the general distillation shortcut (GDS) method. The first principle model requires specific measurements (Friedman, 1995) of which some are either unreliable or unavailable and have to be estimated.

One such measurement, infrequently used by production, was calibrated during a unit shutdown. The measurement is an important input for the GDS soft sensor and as a result of a measurement bias the soft sensor accuracy deteriorated and it could not be used in the MPC. Although there are methods to automatically identify erroneous measurements (McCoy and Auret, 2019), in this application the measurement was identified manually and the model refitted.

First principle models have the advantage that they are inherently directionally accurate (Friedman et al., 2002). However, if they rely on measurements that are not actively maintained, the model might need to be frequently refitted, or continuously biased with each lab sample. Black box models have the advantage that they do not require specific inputs and this particular process does have alternative measurements which are more reliable than the ones required in the first principle model.

[1] Corresponding author. E-mail: ian.craig@up.ac.za
[2] The author would like to thank Dries Wiid for his contribution to the measurement estimates used as model inputs.

10.1016/j.ifacol.2021.12.031

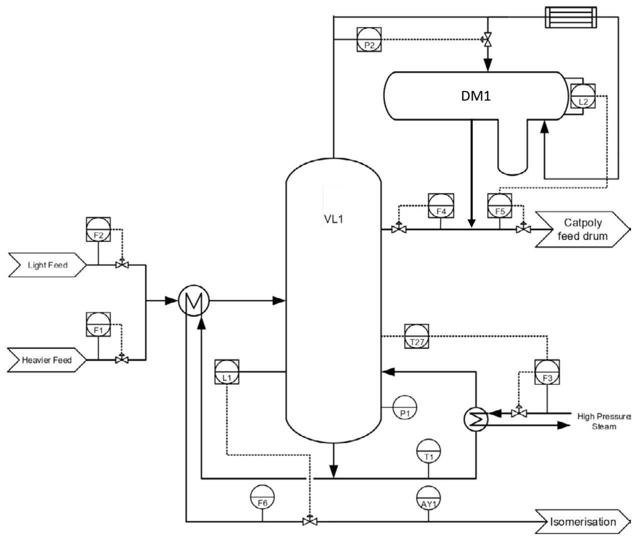

Fig. 1. Debutaniser with two feed streams

The process is described in Section 2. Section 3 details the first principle model used to predict the bottoms $\%C_4$ in the initial implementation. The measurement inaccuracies that led to an inaccurate prediction are highlighted in Section 4. Section 5 then discusses the random forest model used as a soft sensor. The two methods are compared in Section 6. The paper is concluded in Section 7.

2. PROCESS DESCRIPTION

Fig. 1 shows a process flow diagram of the debutaniser column. The lighter feed flow (F_2) and heavier feed flow (F_1) come from separate operating units. The feed streams are then combined, preheated and provided as a feed stream to the column. Temperature T_{27} is controlled using the steam flow rate through a high pressure steam (HPS) reboiler, and a hot vapour bypass is used to control the overheads pressure (P_2).

The distillate flow (F_5) does not have any composition specifications, and is routed to the catpoly feed drum. The bottoms flow (F_6) has a maximum $\%C_4$ material specification as set by the downstream isomerisation unit.

The two feed flows, F_1 and F_2, are both wild flows that can vary by 30% from nominal over a 24 hour period. The upstream units have limited buffering capacity. Therefore, the feed flow rates change with upstream load changes. If the feed debutaniser is decommissioned or hydraulically limited the feed can be bypassed directly to the distillate run-off. However, this is unwanted due to the monetary benefit of increasing the $\%C_5$ and heavier material in the bottoms flow.

The implemented debutaniser MPC has 2 manipulated variables (MVs), 4 controlled variables (CVs), and 2 disturbance variables (DVs) as summarised in Table 1. All MPC variables are measured, except for the bottoms $\%C_4$ material which has no measurement. In Huang and Riggs (2002) a debutaniser column is controlled with an MPC, however in that study almost all of the column variables are controlled by the MPC, some in a band and some at target values. In this case the column PID controls remain intact, and the MPC maintains all of the CVs in a band with a linear programming (LP) objective to maximise the $\%C_4$ material in the bottoms flow. As a result the MPC is implemented with a negative LP, as a negative LP coefficient will maximise a CV (Qin and Badgwell, 2003). The operation of the MPC as well as a description of the variables are given in Table 2.

Table 1. Debutaniser MPC variables

Label	Process tag	Description	Units
Manipulated variables			
MV_1	$F_4{}^{SP}$	Reflux flow setpoint	m^3/h
MV_2	$T_{27}{}^{SP}$	Tray temperature setpoint	$°C$
Controlled variables			
CV_1	$AY_1{}^{PV}$	Bottoms mass $\%C_4$ process value	%
CV_2	$F_3{}^{OP}$	Steam flow controller output	%
CV_3	$L_2{}^{OP}$	Overheads level PID output	%
CV_4	$L_1{}^{OP}$	Bottoms level PID output	%
Disturbance variables			
DV_1	$F_1{}^{PV}$	Heavier feed flow process value	m^3/h
DV_2	$F_2{}^{PV}$	Light feed flow process value	m^3/h

Table 2. Debutaniser MPC operation

Label	Justification/Summary of operation
Manipulated variables	
MV_1	Maximised to improve the overheads product purity. Therefore C_5 material in the overheads is minimised and C_4 material in the bottoms is maximised.
MV_2	Key variable to boil up less heavy material. Maximising the bottoms C_5 material.
Controlled variables	
CV_1	Needs to remain below the maximum specification limit
CV_2	Ensure the reboiler slave loop does not saturate If the PID output goes to 0%, MV_2 should decrease If the PID output goes to 100%, MV_2 should increase
CV_3	Ensure the overheads level controller does not saturate If the PID output goes to 0%, MV_1 should increase If the PID output goes to 100%, MV_1 should decrease
CV_4	Ensure the bottoms level controller does not saturate If the PID output goes to 0%, MV_2 should decrease If the PID output goes to 100%, MV_2 should increase
Disturbance variables	
DV_1	Feed-forward model for the flow of heavier material
DV_2	Feed-forward model for the flow of lighter material

The importance of the bottoms C_4 soft sensor is evident from the MPC operation as it is the primary money-making CV. Degrading soft sensor accuracy may lead to a decrease in MPC monetary benefit. Poor MPC performance may cause operators to constrain the MVs through clamping as described by Olivier (2017). In extreme cases the MPC may even be switched-off.

3. FIRST PRINCIPLE SOFT SENSOR

A first principle soft sensor model using the General Distillation Shortcut (GDS) method as described in Friedman (1995) is used to predict the bottoms mass $\%C_4$.

3.1 *General Distillation Shortcut soft sensor*

The GDS method uses a short cut simulation of the bottom half of the stripping section of the debutaniser (Friedman et al., 2002). It is derived from the distillation column equations for section performance as given by Colburn (1941). More specifically (1), (4), and (7) are used to describe the ratio between the tray composition and the bottoms composition as a function of the materials volatility, internal reflux, and the number of trays in the column section (Smets et al., 2007). The equations used to implement the bottoms light key inferential are given below:

The bubble point equation is described as,

$$\sum_i K_i x_i = 1, \tag{1}$$

where x_i is the liquid fraction of the i^{th} component and K_i is the equilibrium constant defined by,

$$K_i = \frac{P_i^{vap}}{P_i}. \tag{2}$$

P_i^{vap} is the vapour pressure which is approximated by the Antoine relation (Green and Perry, 2008),

$$P_i^{vap} = e^{A_i - \frac{B_i}{T_i + C_i}} \tag{3}$$

P_i is the absolute pressure, T_i is the temperature, and A_i, B_i and C_i are coefficients for each component i used as fitting parameters in this work.

The Colburn equation is defined as,

$$\sum_i x_i R_i = 1, \tag{4}$$

with the R_i-factor in the Colburn relation for the bottoms section provided in Friedman (1995) as,

$$R_{bottoms,i} = \frac{U_i^{N+1} - 1}{U_i - 1}(K_i - 1) + 1, \tag{5}$$

where N is the theoretical number of trays from the top to the sensitive tray and U_i is defined by,

$$U_i = \frac{K_i F_V}{F_L}, \tag{6}$$

where F_V is the internal vapour stream and F_L is the internal liquid stream for the stripping section. The sum of the fractions of each component is equal to 1 as shown by (7).

$$\sum_i x_i = 1. \tag{7}$$

3.2 Prediction implementation

This section describes the process variables used in the development of the GDS soft sensor. Equations (1) to (6) are used to obtain a practically implementable model.

A binary distillation column is assumed, and a light key and a heavy key are defined as $i \in \{\text{L}, \text{H}\}$.

Vapour liquid traffic calculation The measurements and design parameters required to calculate the internal vapour flow (F_V) and the internal liquid flow (F_L) as specified in Smets et al. (2007) are not available for the process in this study. However, F_V and F_L can still be estimated using energy and mass balances. F_V in the bottoms of the column is approximated using an energy balance as,

$$F_V \approx F_3(\frac{\lambda_s}{\lambda_{hc}}), \tag{8}$$

where F_3 is the mass flow of steam, λ_s is the enthalpy of vaporisation of steam, and λ_{hc} is the enthalpy of vaporisation of the bottoms hydrocarbon product. λ_s and λ_{hc} are used as fitting parameters.

F_L is approximated using a mass balance at the bottoms of the column,

$$F_L \approx F_V + F_6 \rho, \tag{9}$$

where F_6 and ρ is the volumetric flow and density of the bottoms product respectively. ρ is used as a fitting parameter.

Light key calculation To calculate the light key i is substituted with L in (3), which becomes,

$$P_L^{vap} = e^{A_L - \frac{B_L}{T_L + C_L}}. \tag{10}$$

A_L, B_L, and C_L are used as fitting parameters, and T_L is the sensitive tray temperature for the light key component. T_L is not measured. However, the temperature measurement directly above the stripping section (T_{27}) and the bottoms temperature (T_1) can be used to approximate T_L as,

$$T_L = 0.6T_{27} + (1 - 0.6)T_1. \tag{11}$$

The light key equilibrium constant is then calculated as,

$$K_L = \frac{P_L^{vap}}{P_L}, \tag{12}$$

where P_L is the column pressure at the light key sensitive tray. Similar to T_L, P_L does not have a direct measurement, and is estimated using the expected pressure drop per tray from the bottoms pressure measurement (P_1), the sensitive tray location (as a fraction of total trays), and the column differential pressure. Therefore,

$$P_L \approx P_1 - 0.27(P_1 - P_2), \tag{13}$$

where P_2 is the overheads pressure measurement.

U_L is calculated from (8), (9) and (12) as,

$$U_L = \frac{K_L F_V}{F_L}, \tag{14}$$

and correspondingly the Colburn light key R-factor is,

$$R_{bottoms,L} = \frac{U_L^{N+1} - 1}{U_L - 1}(K_L - 1) + 1. \tag{15}$$

Heavy key calculation The heavy key calculation is similar to the light key calculation, however most measurements are directly available. To calculate the heavy key vapour pressure, i is substituted with H in (3),

$$P_H^{vap} = e^{A_H - \frac{B_H}{T_1 + C_H}}. \tag{16}$$

A_H, B_H and C_H are coefficients used as fitting parameters, and T_1 is the bottoms temperature.

The heavy key equilibrium constant is calculated as,

$$K_H = \frac{P_H^{vap}}{P_1}, \tag{17}$$

where P_1 is the column bottoms pressure.

U_H is calculated from (8), (9) and (17) as,

$$U_H = \frac{K_H F_V}{F_L}, \tag{18}$$

subsequently the Colburn heavy key R-factor is,

$$R_{bottoms,H} = \frac{U_H^{N+1} - 1}{U_H - 1}(K_H - 1) + 1. \tag{19}$$

Bottoms fraction calculation From (5), (7), (15) and (19) the bottoms mass fraction of C_4 material is solved as,

$$\begin{aligned} x_L &= \frac{1 - R_{bottoms,H}}{R_{bottoms,L} - R_{bottoms,H}} \\ \therefore AY_1 &= \%C4 = 100 x_L \end{aligned} \tag{20}$$

Although many of the parameters used in the GDS soft sensor can be calculated, or found in literature (Green and Perry, 2008; Reid et al., 1977), it was not possible for this implementation due to a restricted number of available measurements and widely varying operating conditions. To overcome this limitation various estimates were used and all of the fitting parameters were fit from historical data. The model parameters were fit using 6 months of data with lab samples taken every 12 hours. Table 3 shows the fitting constants for the bottoms $\%C_4$ GDS soft sensor.

Table 3. GDS model parameters

Coefficient	Value	Coefficient	Value
λ_s	1424	ρ	668
λ_{hc}	291	N	8.69
A_L	15.76	A_H	15.76
B_L	2131.42	B_H	2405.96
C_L	-33.15	C_H	-39.63

4. PROCESS CHALLENGES

4.1 Bottoms pressure inaccuracy

As shown in Section 6 the initial GDS soft sensor performed well in predicting the bottoms $\%C_4$ material. However, upon the initial data cleaning and model validation it was found that P_1 and P_2 would often overlap, resulting in large prediction errors.

P_1 and P_2 are not suitable for the GDS soft sensor because the measurement error for the pressure transmitters exceeds the difference in pressure between them. The measurement inaccuracy was overcome by inferring the bottoms pressure.

The overheads pressure is controlled with a hot vapour bypass controller and is considered more reliable than the bottoms pressure. This is because the top pressure is used for control and is monitored by process engineers to ensure the column operates within accepted parameters, whereas the bottoms pressure is merely an indication shown on the human machine interface (HMI) graphic. The bottoms pressure input into the soft sensor was changed from the true process value to $P_1 \approx P_2 + 0.35F_4$, where F_4 is the reflux flow and 0.35 is a fitting constant that was used to estimate the pressure drop in the column as the reflux flow rate changes.

4.2 Unit maintenance

After the GDS soft sensor was implemented with the augmented bottoms pressure it was monitored for one month to test the prediction accuracy. Shortly thereafter the debutaniser was decommissioned for routine maintenance.

After the maintenance the GDS soft sensor had a constant large prediction error. The model inaccuracy was traced to a faulty reading on the steam flow transmitter which gave a consistently lower reading compared to the period before the maintenance took place. The lower steam flow measurement did not raise any concerns since it is used in a cascade PID configuration and the T_{27} master temperature loop still maintained setpoint.

The model could not be refitted immediately after commissioning because the process changed and was not in operation long enough to provide sufficient training data. This resulted in the MPC being switched off until the GDS soft sensor could be refitted. After gathering enough process data with the new operating conditions the original GDS parameters were refitted in order to use the $\%C_4$ soft sensor again. Section 6 compares the GDS soft sensor prior to ($GDS_{initial}$) and after the maintenance period (GDS_{refit}), and shows how it performed after the refit.

5. RANDOM FORREST SOFT SENSOR

Random forest (RF) regression models are models that combine the outputs of various regression trees through the use of bagging, and they are becoming popular in process control due to their accuracy over conventional regression models (Kneale and Brown, 2018). A random forest model with the initial data set used to fit the GDS parameters was trained to fit the bottoms mass $\%C_4$. The reasons for using a RF model are,

(1) A dataset was already available with features and labels, i.e. the process variables used to predict $\%C_4$ with the actual lab $\%C_4$ values.
(2) Unreliable measurements do not *have* to be included.
(3) Model training is simple with toolboxes such as Scikit-learn in Python (Pedregosa et al., 2011).
(4) Auret and Aldrich (2011) state that expert process knowledge should be applied for data preprocessing and feature selection for tree ensembles. Such knowledge and sufficient data were available from the initial GDS model fitting.

5.1 Feature list for the RF model

As discussed in Section 4, the bottoms pressure (P_1) and the steam flow (F_3) are prone to inaccuracies and are not used as model inputs. All other inputs are used as features, i.e. P_2, F_4,F_6, T_1, and T_{27}.

In processes with sufficient upstream buffering capacity the distillation column feed flows will vary less in relation to the other column measurements to maintain stability. This usually make them bad inputs to use in black box models because it provides little variation with which to train the models. In this work the feed flows, F_1 and F_2, are ideal additional input variables because they vary over a large range in a short span. Additionally, they are reliable as they are used for custody transfer. The ratio of the two feed flows are also a direct indication of the change in feed composition.

Overfitting of the decision tree was prevented by adding F_2 and the ratio of F_2 to F_1 as inputs. F_2 was added since it is the biggest flow, and the input $Feed_{Ratio} = \frac{F_2}{F_1}$ improved the robustness of the model since the range of operation of F_1 and F_2 is no longer a concern, but only their relative movement. By still using F_2 as an independent input the reboiler duty dynamics can inherently be captured.

5.2 *Hyperparameters used in the Scikit-learn training*

A RF model is fitted using the RandomForestClassifier function from Scikit-learn. The RF regressor has a large set of hyperparameters that can be set when training a RF model. To prevent overfitting the depth of each decision tree was limited to 8. Additionally, the number of decision trees in the ensemble was increased to 100. By bootstrapping 100 decisions trees, the different operating range dynamics could be captured by the model. Practically it was observed that with 10 decision trees the outlier values were not predicted, i.e. sudden jumps in $\%C_4$ material could not predicted.

6. RESULTS AND DISCUSSION

In this section the accuracy of the $GDS_{Initial}$, GDS_{Refit} and the RF soft sensors are compared. There are three time periods used for soft sensor development and comparison,

(1) *Training data* - the data used to fit the $GDS_{Initial}$ parameters and then also to train the RF model.
(2) *Deployment testing data* - the period where the $GDS_{Initial}$ soft sensor was used and tested. The RF model is tested with the same process values.
(3) *Post-maintenance* - the process variables after the unit was recommissioned with the new steam flow operating region.

6.1 *Metrics to calculate the soft sensor performance*

The first metric is the model accuracy calculated from the mean absolute percentage error (MAPE),

$$\mathrm{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t - \bar{y}_t}{y_t}\right| \tag{21}$$
$$\therefore \%\mathrm{Accuracy} = 100(1-\mathrm{MAPE}),$$

where n is the total number of sample points, y_t is the actual value at sample point t, and $\bar{y}_t$ the predicted value.

The mean absolute error (MAE) was used to calculate the absolute error that can be expected,

$$MAE = \frac{\sum_{t=1}^{n}|y_t - \bar{y}_t|}{n}. \tag{22}$$

A mean direction accuracy (MDA) metric gives a percentage of how often the sign change for the prediction and true measurement are the same (Blaskowitz and Herwartz, 2011). MDA indicates the qualitative accuracy of the model and is calculated as,

$$MDA = \frac{1}{n}\sum_{t=2}^{n} 1 sign(y_t - y_{t-1}) == sign(\bar{y}_t - y_{t-1}), \tag{23}$$

6.2 *Soft sensor results*

The GDS soft sensor and the RF soft sensor results are shown in Figs. 2 to 4. The performance metrics for the soft sensors are summarised in Table 4.

Table 4. Summary of model results

Training data period			
Soft sensor model	%Accuracy	MAE	MDA
$GDS_{initial}$	66%	0.1	60%
GDS_{refit}	2%	0.54	32%
RF	86%	0.036	73%
Deployment testing period			
Soft sensor model	%Accuracy	MAE	MDA
$GDS_{initial}$	72%	0.085	51%
GDS_{refit}	1.9%	0.52	31%
RF	80%	0.06	62%
Post-maintenance period			
Soft sensor model	%Accuracy	MAE	MDA
$GDS_{initial}$	55%	0.18	50%
GDS_{refit}	73%	0.09	52%
RF	81%	0.06	61%

Fig. 2 shows that the $GDS_{initial}$ was accurate when compared to the lab samples, with a %Accuracy of 72%. After the maintenance period the %Accuracy dropped to 55% with the normalised MAE increasing from 0.085 to 0.18. Fig. 3 shows that the GDS_{refit} model is accurate for the post-maintenance period (%Accuracy is 73%). The MDA for the initial model remained similar in all periods, illustrating the advantage of using a first principle model. Fig. 4 shows that the RF model fits both the deployment and post-maintenance period well with %Accuracy values of 80% and 81% respectively. This shows that with the more reliable measurements the RF model was not impacted by the sudden change in steam flow after the unit shutdown.

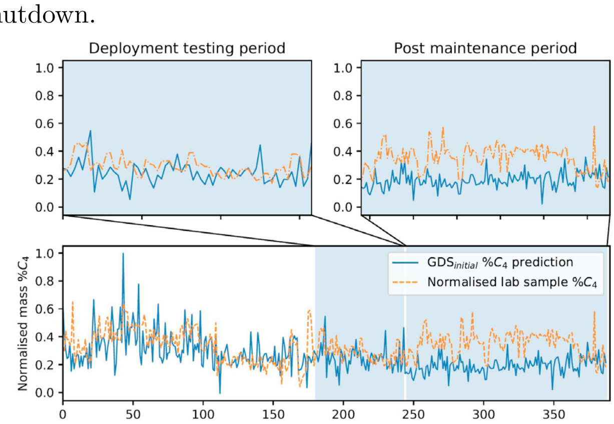

Fig. 2. $GDS_{initial}$ results showing the $\%C_4$ predictions compared to the true lab samples (normalised).

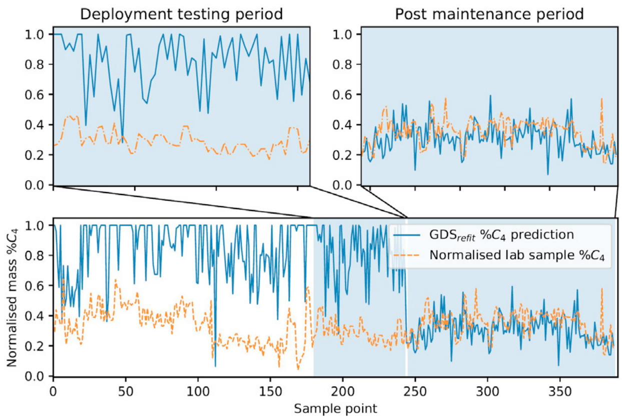

Fig. 3. GDS$_{refit}$ results showing the %C_4 predictions compared to the true lab samples (normalised).

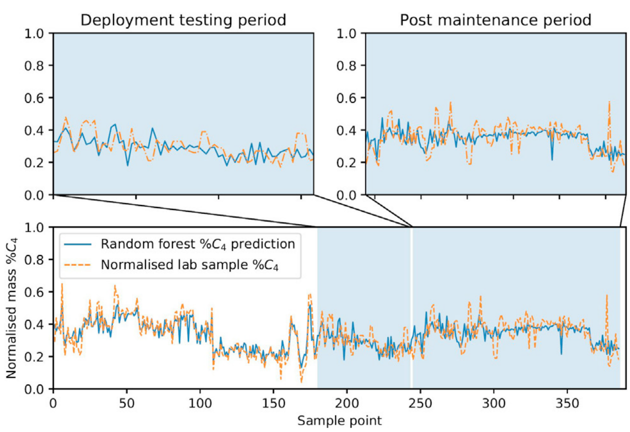

Fig. 4. RF results showing the %C_4 compared to the lab samples (normalised).

7. CONCLUSION

Two different methods were proposed to develop a bottoms mass %C_4 material soft sensor for a debutaniser column. Both methods proved to provide accurate results and the performance metrics indicated both soft sensors are quantitatively and qualitatively accurate making either suitable to implement in an industrial MPC.

A first principle GDS model for predicting the bottoms mass %C_4 material was detailed along with practical examples of how it could be used. This model had to be refitted after a measurement error occurred, and subsequently performed just as well as before. This soft sensor has the advantage that it remains directionally accurate in the presence of such a measurement error.

For a RF machine learned soft sensor it was shown how a proper feature set should be selected. This set contained reliable measurements that directly influence the quality specification, and only a few hyperparameters were required to fit the RF model. The RF model had the advantage that it does not have to conform to a strict input vector and as a result more accurate and reliable measurements can be considered. This was illustrated by the RF soft sensor remaining accurate before and after the unit maintenance period. The RF model was not tested with artificial measurement error on any of its inputs and is recommended for future study.

REFERENCES

Auret, L. and Aldrich, C. (2011). Empirical comparison of tree ensemble variable importance measures. *Chemometrics and Intelligent Laboratory Systems*, 105(2), 157–170.

Blaskowitz, O. and Herwartz, H. (2011). On economic evaluation of directional forecasts. *International Journal of Forecasting*, 27(4), 1058–1065.

Colburn, A.P. (1941). Simplified calculation of diffusional processes. *Industrial & Engineering Chemistry*, 33(4), 459–467.

De Klerk, A. (2012). *Fischer-Tropsch Refining*. John Wiley and Sons.

De Klerk, A., Engelbrecht, D.J., and Boikanyo, H. (2004). Oligomerization of fischer- tropsch olefins: Effect of feed and operating conditions on hydrogenated motor-gasoline quality. *Industrial & Engineering Chemistry Research*, 43(23), 7449–7455.

Friedman, Y.Z. (1995). *Simulation based inferential controls*. American Institute of Chemical Engineers.

Friedman, Y., Neto, E., and Porfirio, C. (2002). First-principles distillation inference models for product quality prediction. *Hydrocarbon Processing*, 81(2), 53–60.

Green, D.W. and Perry, R.H. (2008). *Perry's Chemical Engineers' Handbook, Eighth Edition*. McGraw-Hill Education.

Huang, H. and Riggs, J.B. (2002). Comparison of PI and MPC for control of a gas recovery unit. *Journal of Process Control*, 12(1), 163–173. doi:10.1016/S0959-1524(01)00004-X.

Kneale, C. and Brown, S.D. (2018). Small moving window calibration models for soft sensing processes with limited history. *Chemometrics and Intelligent Laboratory Systems*, 183, 36–46.

Leprince, P. (2001). *Petroleum refining. Vol. 3 conversion processes*, volume 3. Editions Technip.

McCoy, J. and Auret, L. (2019). Machine learning applications in minerals processing: A review. *Minerals Engineering*, 132, 95–109.

Olivier, L.E. (2017). A suggested approach to practical unclamping of model predictive control. *IFAC-PapersOnLine*, 50(2), 115 – 120. doi: https://doi.org/10.1016/j.ifacol.2017.12.021. Control Conference Africa CCA 2017.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Machine learning in Python. *Journal of machine learning research*, 12(Oct), 2825–2830.

Qin, S.J. and Badgwell, T.A. (2003). A survey of industrial model predictive control technology. *Control Engineering Practice*, 11(7), 733–764.

Reid, R., Prausnitz, J., and Poling, B. (1977). *The Properties of Gases and Liquids*. New York, McGraw-Hill Book Company.

Smets, I., Boon, S., Boelen, T., Espinosa, J., and Impe, J.V. (2007). Inferring distillation product composition: a hybrid soft sensor approach. In *8th International IFAC Symposium on Dynamics and Control of Process Systems*, volume 3, 169–174.