Computers and Chemical Engineering 151 (2021) 107361

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Refinery production planning optimization under crude oil quality uncertainty

Fupei Li[a,b], Feng Qian[a,*], Wenli Du[a], Minglei Yang[a], Jian Long[a], Vladimir Mahalec[b,*]

[a] Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China
[b] Department of Chemical Engineering, McMaster University, Hamilton, Ontario L8S 4L8, Canada

ARTICLE INFO

Article history:
Received 17 December 2020
Revised 25 March 2021
Accepted 3 May 2021
Available online 11 May 2021

Keywords:
Two-stage stochastic programming
Refinery production planning
Quality uncertainty
Random vector sampling
Product tri-section CDU model

ABSTRACT

Current practice in refinery planning is to assume that the qualities of the crude oil feedstocks are known, even though they often vary. The uncertainty of the quality properties can significantly impact the profit of the refinery and needs to be considered in purchasing decisions. This work employs the product tri-section CDU model (Li et al. 2020) to build an accurate refinery model and determines the optimal crude selection by two-stage stochastic programming. The uncertainty of the crude oil quality properties is defined via the uncertainty of the TBP curves, which is described by the uncertain parameters of the beta functions approximating the TBP curves. The probabilistic scenarios are generated via random vector sampling method, leading to a relatively small number of scenarios required for the two-stage-stochastic programming model convergence. This enables us to determine the best crude oil choice, while requiring acceptable computational times, as illustrated by computational experiments.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

Crude oil purchasing plays a pivotal role in refinery production planning. A plan of crude oil purchasing considers the available crude feedstocks, the expected production slates, and the refinery structure, i.e., capabilities to process given crude. Crude oil purchasing decision is typically made about a couple of weeks to a month prior to the crude oil arriving at the refinery. The crude oils from different fields have distinct quality properties described by the true boiling point (TBP) curves and bulk properties of the oils, which leads to variability in the price on the market. Generally, light and sweet crudes are more valuable and have higher prices than the heavy and sour crudes since the former crudes contain more of the valuable components (e.g., gasoline) and have a lower cost of sulfur removal. Refineries built in the mid to second half of the 20th century were typically designed for a specific type of crude (e.g., light and sweet or heavy and sour). In recent years, the crude oil qualities are becoming poorer (e.g., heavier and sourer), requiring the refineries to revamp and expand their units in order to operate successfully in a volatile crude oil market. Since restructuring a refinery requires significant capital and time, the refineries need to make the best possible use of their existing capabilities (Qian et al., 2017). Consequently, the crude oil selection is the most important step in the decision-making processes from purchasing to the final product delivery.

In the past decade, the refinery production planning problem has been studied by many researchers, focusing on improving the planning models and process models, as well as the optimization algorithms. As for the optimization under uncertainties, the studies in production planning have dealt almost exclusively with the uncertainty of the price or demand in the market. The two-stage stochastic programming model is commonly used to address the optimization under uncertainty. Two-stage stochastic programming requires the determination of the distribution of the uncertainties and then optimizes a number of scenarios that represent potential realizations of the uncertainties. The continuous random variables are commonly described by the probability density function (PDF) or the cumulative distribution function (CDF). The discrete random variables are commonly described by the probable values associated with the probabilities. Statistic characterizations of the random variables are all acquired via the selected samples. The statistical analysis of the samples is used to estimate the characterization of the population, which makes the statistic characterizations (e.g., mean, variance et al.) relevant in stochastic programming. After the characterization of the random variables, the scenario-

* Corresponding authors.
E-mail addresses: fqian@ecust.edu.cn (F. Qian), mahalec@mcmaster.ca (V. Mahalec).

https://doi.org/10.1016/j.compchemeng.2021.107361
0098-1354/© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

Sets and indices

M=$\{m\}$	raw materials m
P=$\{p\}$	final products m
U=$\{u\}$	all units u
T=$\{t\}$	storage tanks m
S=$\{s\}$	probabilistic scenarios s
Q=$\{q\}$	quality properties q ('sg' for specific gravity)
BC=$\{c\}$	blending components c
B=$\{b\}$	blenders b
TI=$\{ti\}$	inlet streams of tanks ti
TO=$\{to\}$	outlet streams of tanks to
UI=$\{ui\}$	inlet streams of units ui
UO=$\{uo\}$	outlet streams of units uo

Subsets and indices

PU=$\{u\}$	processing units u
CDU=$\{u\}$	crude oil distillation units u
HTU=$\{u\}$	hydrotreating units u
CO=$\{m\}$	crude oils m
VQ=$\{q\}$	quality properties q blended linearly on the volumetric basis
WQ=$\{q\}$	quality properties q blended linearly on the weight basis
MIX=$\{u\}$	mixers u
SPL=$\{u\}$	splitters u
OP=$\{op\}$	operating modes of processing units op

Correlation tuples

(c, p)	blending components c in products p
(p, b)	products p blended in blenders b
(c, p, q)	quality properties q calculated in the components c and products p
(ti, t)	inlet streams ti into tanks t
(t, to)	outlet streams to out of tanks t
(ti, t, to, q)	quality properties q calculated in the inlet streams ti and outlet streams to of tanks t
(ui, u)	inlet streams ui into units u
(u, uo)	outlet streams uo out of units u
(ui, u, uo, q)	quality properties q calculated in the inlet streams ui and outlet streams uo of units u
(op, u)	processing units u associated operating modes op
(ui, op)	inlet streams ui into operating modes op
(op, uo)	outlet streams uo out of operating modes op
(ui, u, uo)	inlet streams ui and outlet streams uo of units u

Parameters

L	length of the planning horizon
Price_p	price values of products p
Cost_m	cost values of materials m
Cost_u	cost values of units u
Cost_t	cost values of inventory in tanks $t\mathrm{V}_t{}^{\mathrm{i}}$ initial inventory in tanks t
$\mathrm{A}_m^{\mathrm{min}}, \mathrm{A}_m^{\mathrm{max}}$	minimum and maximum available amounts of materials m
$\mathrm{Q}_{m,q}^{\mathrm{fix}}$	fixed quality properties q of materials m
$\mathrm{D}_p^{\mathrm{min}}, \mathrm{D}_p^{\mathrm{max}}$	minimum and maximum demands of products p
$\mathrm{R}_b^{\mathrm{min}}, \mathrm{R}_b^{\mathrm{max}}$	minimum and maximum blend rates of blenders b
$\mathrm{R}_{c,p}^{\mathrm{min}}, \mathrm{R}_{c,p}^{\mathrm{max}}$	minimum and maximum recipes (volume fraction) allowed of blend components c in blending products p
$\mathrm{Q}_{p,q}^{\mathrm{min}}, \mathrm{Q}_{p,q}^{\mathrm{max}}$	minimum and maximum quality specifications of quality properties q in products p
$\mathrm{V}_t^{\mathrm{min}}, \mathrm{V}_t^{\mathrm{max}}$	minimum and maximum final inventories in tanks t
$\mathrm{R}_t^{\mathrm{min}}, \mathrm{R}_t^{\mathrm{max}}$	minimum and maximum withdrawal rates of tanks t
$\mathrm{Q}_{t,q}^{\mathrm{i}}$	initial quality properties q in tanks t
$\mathrm{Y}_{uo,u}^{\mathrm{fix}}$	fixed yields of outlet streams uo out of units u
$\mathrm{R}_u^{\mathrm{min}}, \mathrm{R}_u^{\mathrm{max}}$	minimum and maximum processing rates of units u
$\mathrm{Q}_{uo,q}^{\mathrm{fix}}$	fixed quality properties q of outlet streams uo
$\mathrm{Y}_{uo,op}^{\mathrm{fix}}$	fixed yields of outlet streams uo of operating modes op
$\mathrm{R}_{op}^{\mathrm{min}}, \mathrm{R}_{op}^{\mathrm{max}}$	minimum and maximum processing rates of operating modes op
N_u	maximum numbers of operating modes in processing units u
$\mathrm{Q}_q^{\mathrm{profile}}$	profiles of quality properties q

Binary decision variables

$b_{p,s}$	products p blended or not
$b_{op,s}$	operating modes op executed or not

Continuous decision variables

V_m	volume flows of materials m
$V_{u,s}$	volume throughputs of units u
$V_{t,s}$	volume inventories of tanks t
$TBP_{m,s}$	TBP temperatures of crude oils m
$Q_{m,q,s}$	quality properties q of materials m
$V_{c,p,s}$	volume flows of blend components c to blend product p
$V_{c,s}$	volume flows of blend components c
$V_{p,s}$	volume flows of product p
$Q_{c,q,s}$	quality properties q of blend components c
$QV_{c,p,q,s}$, $QM_{c,p,q,s}$	auxiliary variables representing the quality properties q times volume flows and weight flows of blend components c to blend product p
$V_{ti,s}, V_{to,s}$	volume flows of inlet streams ti and outlet streams to
$Q_{ti,q,s}, Q_{to,q,s}$	quality properties q of inlet streams ti and outlet streams to
$V_{ui,s}, V_{uo,s}$	volume flows of inlet streams ui and outlet streams uo
$Q_{ui,q,s}$, $Q_{uo,q,s}$	quality properties q of inlet streams ui and outlet streams uo
$V_{op,s}$	volume throughputs of operating modes op
$TBP_{ui,s}, TBP_{uo,s}$	TBP temperatures of inlet streams ui and outlet streams uo
$ccw_{uo,s}$	cumulative cut width of outlet streams uo
$ov_{u,s}$	cut point temperatures of outlet streams uo

based (two-stage) stochastic programming approach requires the generation of probabilistic scenarios.

In two-stage stochastic programming, scenario generation for the uncertain parameters aims to accurately represent the uncertain parameter space and to make the best first- and second-stage decisions (Mavromatidis et al., 2018). Monte Carlo sampling method, adopted by the majority of the stochastic programming researches, randomly takes N samples from the distribution function constructing the set of probabilistic scenarios. The probability of each sample is $1/N$, so-called Sample Average Approximation

(SAA). There is a drawback in using SAA in stochastic programming that the number of samples could be very large to mirror the statistic characterizations of the random variables, such as the mean and variance. The scale of the scenario-based stochastic programming depends on the number of samples, which may cause failure or long computational time in solving these problems.

In a refinery, the uncertainty of the crude oil quality can be addressed via different approaches. Since the operations of process units depend on the quality of the feed stream, such as crude oil as the feed to CDUs and distillation cuts as the feed to FCC, the process unit models are improved to handle the uncertainties of the crude oil. Franzoi et al. (Franzoi et al., 2020) built a cut-point temperature surrogate model of CDU to handle the plant-model mismatches and uncertainties in the process, which uses new data to improve the accuracy of the prediction. Durrani et al. (Durrani et al., 2018) used a hybrid framework of the Taguchi method and genetic algorithm to derive a dataset comprised of variations of crude compositions and optimized cut point temperatures, which was used to train an ANN model of CDU. Minh et al. (Minh et al., 2018) proposed a surrogate model of CDU based on Gaussian process regression with a large number of uncertainties in crude oils and operations to get a global sensitivity analysis of the economic revenue and the operating cost. Chen et al. (Chen et al., 2020a) proposed a machine-learning-based FCC model to overcome the uncertainties in crude oil quality. In the refinery planning and scheduling problem, the data-driven method has been used to address the uncertainty of the crude oil quality. Gao et al. (Gao et al., 2014) proposed a multimodel approach for refinery scheduling with various crudes and blend recipes using a deep learning method to classify the mixed crude oil types, which refers to a specific scheduling model. Dai et al. (Dai et al., 2019) built the data-driven uncertainty set of the crude oil property to get the robust optimization of the crude oil scheduling problem. If the uncertainty of the crude oil is considered when the purchase of the crude oil is made, there comes the crude oil selection and refinery planning problem. Yang and Barton (Yang and Barton, 2015) represented the uncertainties in crude oil qualities by using the uncertain yield of residue and uncertain sulfur content of gas oil as a surrogate for the feedstock uncertainty. They employed 120 scenarios generated by the SAA method. The solution time of the planning model using nonconvex generalized Bender's decomposition (NGBD) was above 5 hours even for a simplified refinery model. It is not clear how their method would perform if a more inclusive refinery model was employed.

Recent studies on refinery planning and scheduling under uncertainty rely on stochastic programming, chance-constrained programming, and robust optimization methods. Gutierrez et al. (Gutierrez et al., 2018) studied the effect of the crude changes on the operation of the hydrogen network. The physical risk related to crude oil import and transportation planning was studied using two-stage stochastic programming, also using conditional value-at-risk (CVaR) constraints on the total cost (Wang et al., 2020). The market and operational uncertainties were considered in the optimal design problem of the refinery hydrogen network using two-stage stochastic programming with flexible constraints modeling framework proposed by Chen et al. (Chen et al., 2020b). Monte Carlo sampling (simulation) regularly accompanies the two-stage stochastic programming for scenario generation. Due to a large number of the scenarios, the stochastic programming model is invariably large scale which leads to challenges in finding the global optimum. Shin and Lee (Shin and Lee, 2018) proposed a multi-time scale stochastic model to solve the price uncertainty in a refinery procurement and production planning problem via mathematical programming and reinforcement learning. Wang et al. (Wang et al., 2009) used the chance-constrained programming model to improve the feasibility and robustness of crude oil operation scheduling with fluctuating product demand and uncertain ship arrival times. Evazabadian et al. (Evazabadian et al., 2019) proposed a short-term crude oil scheduling model using fuzzy programming and the chance-constrained method to address the uncertainty in demand, cost, and time parameters. The uncertainties in the prices, costs, product yields were considered in the refinery operations planning via robust optimization methodology (Leiras et al., 2010). Two-stage adjustable robust optimization was proposed by Lima et al. to address the uncertainty on price and demand for crude oil in the downstream oil supply chain (Lima et al., 2019).

A refinery model which is used to make decisions on crude selection under uncertainties in crude qualities must be more accurate than the errors which are introduced by disregarding the uncertainty in the crude qualities. Arguably, the accuracy of the CDU model is the most important issue in a refinery model since an inaccurate CDU model will predict inaccurate yields of the intermediate streams, and the entire downstream processing will be erroneous. In the previous studies on the refinery planning problem, the simplified CDU models were commonly used considering the convergence and efficiency of the refinery planning models, especially in optimization under uncertainty. The swing-cut CDU model developed by Zhang et al. (Zhang et al., 2001) has been studied in plenty of refinery planning optimization for both deterministic problems and stochastic problems. If the swing-cut model is used to deal with uncertainty, the yields (cut points) of the swing-cut model should be calculated for each crude in every probabilistic scenario, which is computationally expensive. In addition, the swing-cut model does not have enough accuracy in predicting the product yields and bulk properties, which is a critical issue in refinery production planning. In the recent study of the CDU model for planning and scheduling, Fu et al. (Fu et al., 2015) proposed a hybrid CDU model based on both the first principal and the empirical models of CDU, which is better in predicting the product yields and bulk properties than the swing-cut model. Li et al. (Li et al., 2020) derived a simpler version of the hybrid CDU model (product tri-section CDU model) which also has accuracy on par with the rigorous tray-to-tray CDU models, but it has significantly better computational times than Fu et al. hybrid CDU model in an MINLP model of a crude oil selection optimization problem. Product tri-section CDU model is used in this study.

Uncertainty in crude qualities manifests itself as variations in crude distillation curves, which in turn results in variability in CDU product yields. Uncertainties in other crude properties (sulfur, gravity, etc.) are of secondary importance, since downstream units usually have sufficient capacities to accommodate such variations. This work describes variations in crude qualities as deviations from its crude assay TBP distillation curve, which deals with the impact of uncertainty in the crude distillation curve on the yield of CDU product streams. It is assumed that the uncertainty in other properties (e.g., sulfur) results in CDU product streams which can be still be processed by the downstream units. For example, if a yield of a stream that is routed to a hydrotreater changes, then the total amount of sulfur in that stream also changes; the assumption is that the hydrotreater can process that amount of sulfur. The refinery model with the tri-section CDU model is employed as a basis of the two-stage stochastic programming formulation of the refinery production planning problem. Scenario generation employs a random vector sampling (RVS) method, which utilizes the discretization of uncertain parameters (Afzali et al., 2020) to minimize the number of scenarios. The stochastic programming models are solved via a commercial global optimization solver (ANTIGONE).

The remainder of this paper is structured as follows: Section 2 describes the methodology used in this work including (i) the two-stage stochastic programming approach, which formulates the refinery production planning problem under uncertainty; (ii) the crude oil quality uncertainty, which is defined by the uncertain parameters of the beta function approximating the crude oil TBP curve; (iii) the scenarios generation for the uncertain crude oil TBP curve, which uses the discretization of continuous variables and random vector sampling method. Section 3 describes the refinery planning model under uncertainty of crude oil quality properties with the product tri-section CDU model embedded. Section 4 presents case studies of refinery production optimization under quality uncertainty of the crude oil. Section 5 presents the conclusions of this work.

2. Methodology

2.1. Two-stage stochastic programming approach

The decision variables in the two-stage stochastic programming are separated into two sets: the first-stage variables are decided before the realization of the uncertainty parameters (here-and-now decisions) and the second-stage variables are optimized after the observation of the uncertainty parameters (wait-and-see decisions). The expected objective value is optimized covering all the uncertainty realizations.

A general formulation of two-stage stochastic programming is given as follows:

$$\begin{aligned} &\min_{x_1, d_1} f_1(x_1,\ d_1) + \mathrm{E}_\omega[f_2(x_1, d_1, \omega)] \\ &\qquad s.t.\ h_1(x_1, d_1) = 0 \\ &\qquad g_1(x_1, d_1) \le 0 \\ &\qquad x_1 \in \mathbb{R},\ d_1 \in \{0, 1\} \end{aligned} \tag{1}$$

$$\begin{aligned} &f_2(x_1, d_1, \omega) = \min_{x_2 \in X, d_2 \in D} f_2[x_1, d_1, x_2(\omega), d_2(\omega), \omega] \\ &\qquad s.t.\ h_2[x_1, d_1, x_2(\omega), d_2(\omega), \omega] = 0 \\ &\qquad g_2[x_1, d_1, x_2(\omega), d_2(\omega), \omega] \le 0 \\ &\qquad x_2 \in \mathbb{R},\ d_2 \in \{0, 1\} \end{aligned} \tag{2}$$

where x and d are the continuous and binary variables with subscript 1 and 2 representing the first-stage and second-stage decisions, ω represents the uncertainty parameters, and every second-stage variable is referring to a realization of the uncertainty parameters. The objective of the two-stage stochastic programming model consists of the first-stage objective value and the expectation of the second-stage objective value. The equality and inequality constraints $h(\cdot)$ and $g(\cdot)$ are associated with the first-stage decisions and second-stage via subscripts 1 and 2.

In the crude oil selection production planning problem, the first-stage decision variables are the amount of crude oil to purchase. Second-stage decision variables are the amount of the final products to produce and the other operational variables such as the volume of the material processed by the downstream units and the blending recipes. In this work, the uncertainty parameters for the refinery planning model are the TBP temperatures of the crude oil supply, which means the crude oil qualities (TBP curve, specific gravity, surfer content) are unknown and uncertainty for production when the crude oils were bought on the market. Since the crude oil TBP curves are nonlinear and each realization of uncertain TBP curves is different, the crude oil selection production planning problem finding the optimum crude oil recipe is a large-scale mixed-integer nonlinear programming problem and extremely hard to solve. The stochastic programming could address these uncertainties before making the crude oil purchase decisions so that the profit in the refinery production could be optimized. The second-stage decisions, e.g., the operational variables of the refinery production (unit capacity, blending recipe), are determined after the realization of the quality uncertainties. In the following section, the crude oil quality uncertainty is described via crude oil TBP curves and the pseudocomponent method.

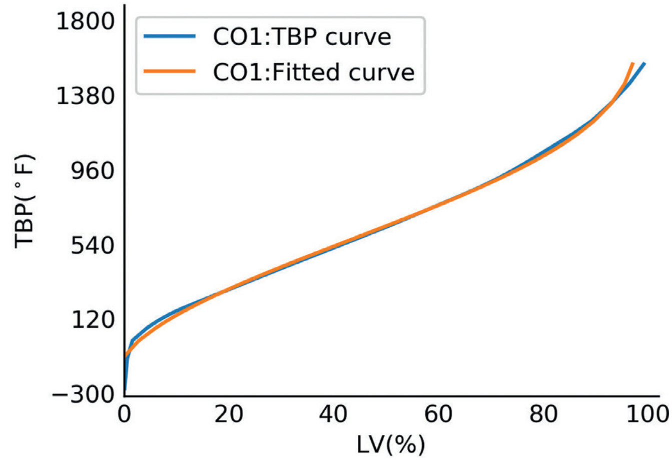

Fig. 1. TBP curve and Beta Function approximation.

2.2. Crude oil quality uncertainty

Crude oil is a mixture of many organic compounds (mostly hydrocarbons) having a wide range of boiling point temperatures. Pseudocomponent method is a widely used procedure to characterize such complex mixtures, e.g., crude oils and petrochemical products. It is applied to calculate the bulk property of the crude oils and the product cuts. Pseudocomponents describing the crude oil are determined from the true boiling point (TBP) distillation curve. A point on the TBP curve is the temperature at which this component and all components with lower boiling points evaporate, i.e., the temperature at which the entire oil fraction lighter than this temperature is vaporized. An example crude oil TBP curve is drawn as the blue line in Fig. 1. The TBP curve can be fitted by some probability density functions, which have been studied by Sánchez and co-workers (Sánchez et al., 2007). Their comparison of various distribution functions leads to the recommendation that the four-parameter distribution functions, including the beta function, offer the best fitting capability. The four-parameter beta distribution function is defined by Eq. (3).

$$f(x, \alpha, \beta, A, B) = \int_A^{x \le B} \left(\frac{1}{B - A}\right) \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{x - A}{B - A}\right)^{\alpha - 1} \left(\frac{B - x}{B - A}\right)^{\beta - 1} dx \tag{3}$$

where α and β are the positive parameters controlling the shape of the distribution function, A and B are lower and upper bounds on the distribution function, x represents the normalized temperature recovery $x = (T - T_l)/(T_u - T_l)$, and Γ is the standard gamma function. The value of the function is the liquid volume fraction between 0 to 1. The four parameters are estimated via least-squares fitting the crude oil TBP curve. For crude oil CO1 TBP curve in Fig. 1, the optimal parameters of the beta distribution function are $\alpha = 0.4554$, $\beta = 0.3688$, $A = -0.0313$, $B = 0.9782$. The upper and lower bound for the normalized temperature recovery are $T_u = 1800^\circ$F and $T_l = -300^\circ$F. It should be noted that the beta function approximation still has some errors, especially at the front-end and back-end points of the TBP curve. As explained by Sánchez et al. (Sánchez et al., 2007), the discrepancies are likely caused by the low accuracy of the experimental measurements of the front-end and back-end of the TBP curve. Therefore, in this work, the initial boiling point (IBP) at the front-end and final boiling point (FBP) at the back-end are set to be constant, which means whatever the fitted beta functions are like, the IBP and FBP are deterministic as the known values from the crude oil TBP data.

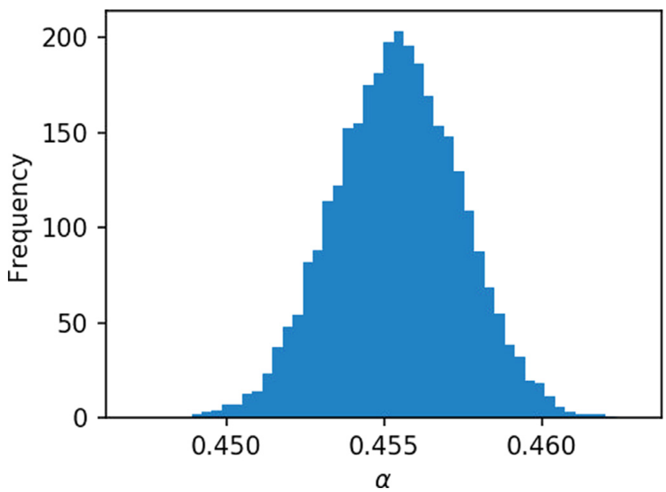

Fig. 2. Histogram of α.

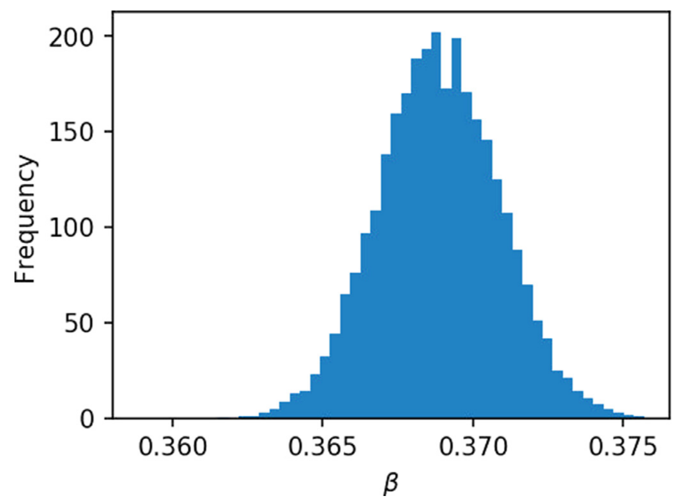

Fig. 3. Histogram of β.

Using probability density functions is an appropriate way of approximating the TBP curves so that the uncertainty could be quantifiable. The bulk property will have the same probability distribution as the parameters of the probability functions. In this paper, the uncertainty of the crude quality properties is assumed to have a normal distribution. After the crude oil TBP curve described as a beta function, the uncertainty in the crude oil TBP curve can be described by the uncertainty of the parameters. Since the parameters A and B determine the lower and upper bounds on the distribution function, these two parameters are fixed to the optimal value fitting the target crude oil TBP curve. The positive parameters α and β control the shape of the distribution function; they describe the uncertainty qualities of the crude oil. Since crude oil characterization using the pseudocomponent method is based on the TBP curve, changing the shape of the curve will change (i) the distribution of the pseudocomponents, (ii) the yield of the distillation products, and (iii) the bulk properties (e.g., specific gravity and sulfur content) of the crude oil and of the CDU products.

The uncertainty parameters α and β are assumed to have normal distribution $N(\mu, \sigma)$, and consequently, the bulk properties (e.g., specific gravity sg and sulfur content sul) also have an approximately normal distribution. In order to limit the variance of the uncertainty in the crude oil, it is assumed that the crude oil with uncertainty qualities have the probability $\Pr(\mu - 3\sigma \le sg \text{ and } sul \le \mu + 3\sigma) \approx 0.9973$ maintaining its type, which is the so-called three-sigma rule of thumb. The types of crude oil are defined by the specific gravity (light $sg < 0.8762$, medium $0.8762 \le sg < 0.9218$, heavy $sg \ge 0,9218$) and sulfur content (sweet $sul < 0.5\%$, sour $sul \ge 0.5\%$). Therefore, in this work, μ is set to be the optimal parameter fitting the target TBP curve for both uncertainty parameters α and β; $\sigma = 0.002$ is a propitiate value for α and β following the assumption, which will not change the type of the crude oil.

The crude oil in Fig. 1 is taken as an example using the Mote-Carlo sampling method. The uncertainty parameters $\alpha \sim N(0.4554, 0.002)$ and $\beta \sim N(0.3688, 0.002)$ are employed to generate 10000 samples which have been used as a basis for different scenarios for the crude oil uncertainty as shown in Figs. 2 and 3.

Two of the sampled random TBP curves and the deterministic fitted TBP curve are plotted in Fig. 4. It shows that the uncertainties in the parameters change the crude oil TBP curve. Consequently, at the same cut temperature, the yield of a product varies. This also leads to a different distribution of the pseudocomponents, which causes uncertainty in the bulk properties of the crude oil. The calculation of the bulk properties of the sampled crude oil is based on the pseudocomponents method assuming a constant property value of individual pseudocomponent.

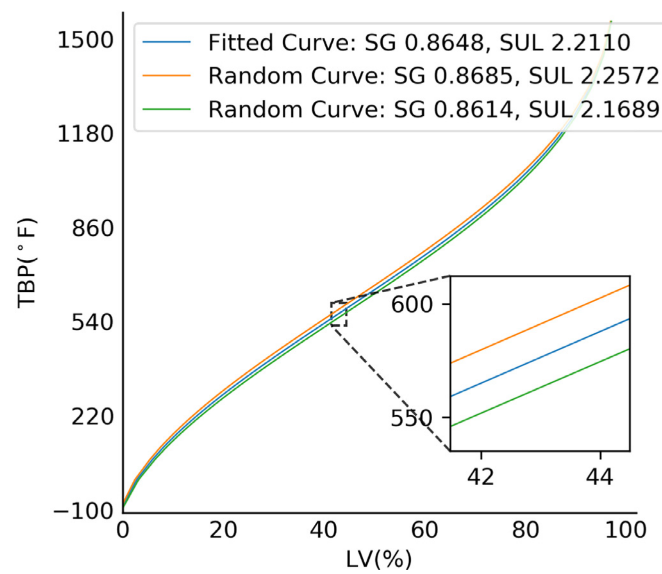

Fig. 4. Fitted and random curves of the crude oil 1.

The histograms of the calculated bulk property (specific gravity and sulfur content) of the samples are shown in Fig. 5 and Fig. 6. The mean of the specific gravity is 0.8648, and the standard deviation is 0.0010. The mean of the sulfur content is 2.2110, and the standard deviation is 0.0131. It is pointed out that the bulk properties have an approximatively normal distribution. Since the type of the crude oil is light and sour, the samples in $[\mu - 3\sigma, \mu + 3\sigma]$ belongs to the same type as the target crude oil.

In the remainder of this paper, the uncertainty in the parameters α and β is used as a surrogate to describe the uncertainty in the crude oil TBP curve, since each set of the uncertainty parameters corresponds to a different TBP curve. The same approach is used for crude oil 1 (CO1) and crude oil 2 (CO2); details are provided in the Supporting Information.

2.3. Scenario generation for the uncertain crude oil TBP curve

The uncertainty of a continuous variable can be described by its probability distribution. In the probabilistic scenarios based stochastic programming approach, the scenarios are discrete re-

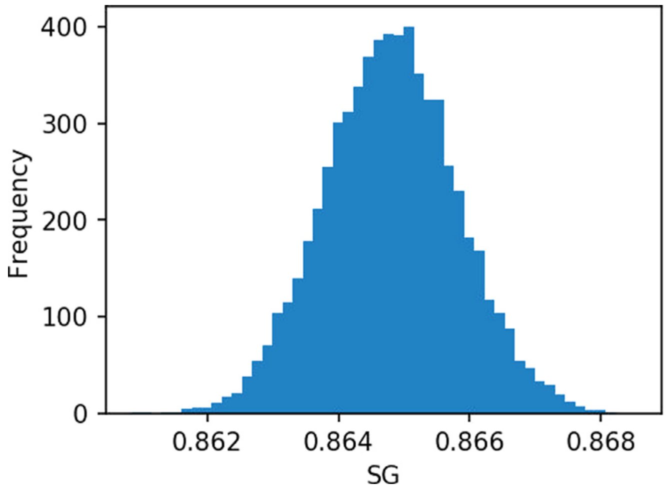

Fig. 5. Histogram of specific gravity.

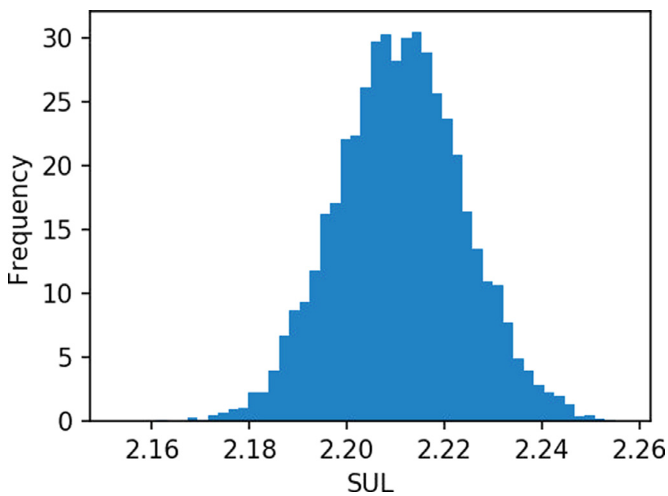

Fig. 6. Histogram of sulfur content.

alizations of the probability distribution. The simplest method to generate discrete scenarios is a sample average approximation method (S scenarios corresponding to the uniform probability $1/S$) via the Monte-Carlo sampling method. Although it is a simple way of generating the scenarios, care must be taken to ensure that a sufficient number of the scenarios is generated, which impacts the scale of the stochastic programming model.

Moment matching is another way of discretization of the continuous random variables. It enables an accurate representation of representing the probability distribution using a relatively small number of scenarios. Gaussian Quadrature based moment matching ensures the discrete approximation of a continuous distribution with N points exactly matching its first $(2N-1)$ statistical moments. $N=3$ is a common value that has been used in the study of the design of energy systems under uncertainty (Afzali et al., 2020; Mavromatidis et al., 2018) and is also employed in this work so that the important statistical moments (mean, variance, skewness, and kurtosis) can be matched accurately. The uncertainty of each parameter α and β is approximated via three valued points and their associated probability, as shown in Fig. 7.

The uncertain parameters can take on realizations corresponding to one of the three discrete points in an arbitrary scenario. In order to generate the probabilistic scenarios, Monte Carlo sampling randomly selects a point from the probabilistic distribution of the uncertain parameters according to the probability of each

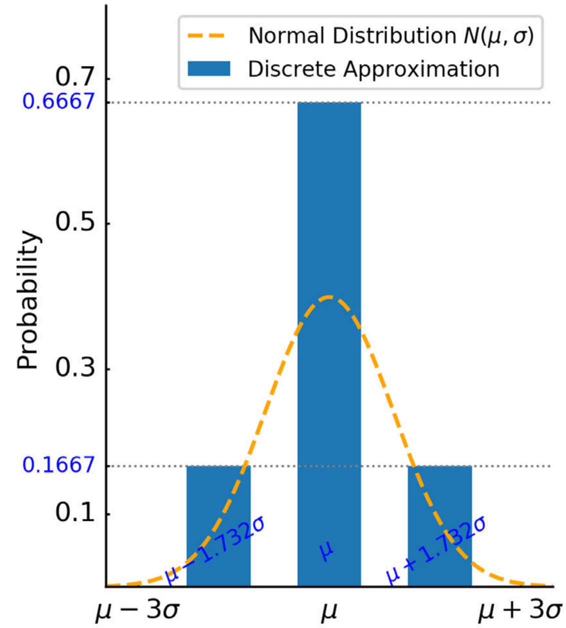

Fig. 7. Discrete approximation of the normal distribution.

point so that a large number of scenarios is needed to ensure that the points with lower probability are also selected. With the discretization of the uncertain parameters, there are only three candidate points in the probabilistic sets, and all of them are expected to be selected in the scenarios with a relatively small number of scenarios. The random vector sampling (RVS) guarantees that the uncertain parameter can take on the values of all discrete points in consecutive scenarios (Afzali et al., 2020). For the three discrete points, RVS approach builds three scenarios at each time. For each of the three scenarios, one of the three points is assigned to the uncertain parameter as a realization. All probable permutations (3-permutation of 3, $\mathrm{P}_3^3=6$) for assigning three discrete points $(\upsilon_1, \upsilon_2, \upsilon_3)$ to tree scenarios (s_1, s_2, s_3) are $\Omega_{\mathrm{P}_3^3}$:

$$\Omega(s_1, s_2, s_3) = \{(\upsilon_1, \upsilon_2, \upsilon_3), (\upsilon_1, \upsilon_3, \upsilon_2), (\upsilon_2, \upsilon_1, \upsilon_3), (\upsilon_2, \upsilon_3, \upsilon_1), (\upsilon_3, \upsilon_1, \upsilon_2), (\upsilon_3, \upsilon_2, \upsilon_1)\} \tag{4}$$

The number of scenarios S to be generated should be the multiple of three $S=3K$. For each K, one of the probable permutations Ω is randomly selected by the same probability $\pi = 1/\mathrm{P}_3^3$, which is applied for all uncertain parameters P.

After all scenarios S are generated, the probability for each scenario is computed by the following probability normalization. The probability for a parameter p realized at the value υ_i^p $(i=1,\ 2,\ 3)$ for a scenario s in the kth set of all the scenarios is denoted as $\omega_s^{p,k}$ and the associated normalized probability is denoted as $\omega_s^{p,k,Nr}$ calculated via

$$\omega_s^{p,k,Nr} = \frac{\omega_s^{p,k}}{\sum_s \omega_s^{p,k}} \ \forall k \in K,\ s = \{3k-2,\ 3k-1,\ 3k\} \tag{5}$$

The probability for a scenario s in the kth set of all the scenarios is denoted as ψ_s^k. The probability of the three scenarios in each K is calculated via

$$\psi_s^k = \prod_{p \in P} \omega_s^{p,k,Nr} \ \forall k \in K,\ \forall s \in S \tag{6}$$

Then, the three scenarios $s=k, k+1, k+2$ in the kth set are normalized by

$$\psi_s^{k,Nr} = \frac{\psi_s^k}{\sum_s \psi_s^k} \ \forall k \in K,\ s = \{3k-2,\ 3k-1,\ 3k\} \tag{7}$$

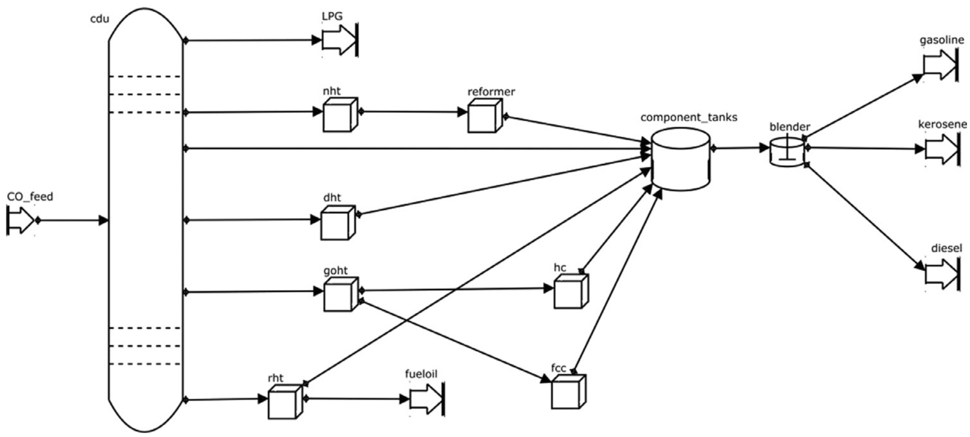

Fig. 8. Simplified flowsheet of the refinery model (Li et al., 2020).

Finally, the normalized probability of a scenario s is

$$\psi_s^{Nr} = \frac{3}{S}\psi_s^{k,Nr}\ \forall s \in S,\ \ k = INT(3/s) \tag{8}$$

3. Refinery planning model

The refinery model in this work is based on the deterministic production planning optimization studied by our previous work of the tri-section CDU model (Li et al., 2020). The configuration of the refinery flowsheet is the same as our previous work. The simplified flowsheet of the refinery model is shown in Fig. 8.

The refinery planning problem under crude quality uncertainty is described as follows:

Given

A planning horizon of a specific length.

A set of crude oils, the probabilistic scenarios of the crude oil TBP curves the pseudocomponent profiles, and supply profiles (e.g., costs and available amounts).

A set of final products (e.g., LPG, gasoline, kerosene, diesel, fuel oil, coke), their quality specifications, prices and demand in the planning horizon.

A set of storage tanks and their initial, minimum, and maximum initial inventories.

A set of blenders and their minimum and maximum blending rates.

Processing units (i.e., CDU, hydrotreaters, catalytic reformer, fluid catalytic cracker, etc.) with minimum and maximum throughput rates and their operating modes.

The topology of all the processing units and storage tanks.

Determine

First-stage

The amount of each crude oil feedstock to be used.

Second-stage

The amount of each final product to be produced.

The production volumes in each processing unit, each operating mode.

The final inventories of the storage tanks.

The components blended in each blender.

While

Maximizing the expected profit.

Objective Function

$$\begin{aligned} max\ \text{profit} = & \sum\nolimits_{s\in\mathbf{S}} \psi_s^{Nr} \sum\nolimits_{p\in\mathbf{P}} \text{Price}_p \times V_{p,s} \\ & - \sum\nolimits_{m\in\mathbf{M}} \text{Cost}_m \times V_m \\ & - \sum\nolimits_{s\in\mathbf{S}} \psi_s^{Nr} \sum\nolimits_{u\in\mathbf{PU},\mathbf{CDU},\mathbf{HTU}} \text{Cost}_u \times V_{u,s} \\ & - \sum\nolimits_{s\in\mathbf{S}} \psi_s^{Nr} \sum\nolimits_{t\in\mathbf{T}} \text{Cost}_t \times \left(V_{t,s} - \text{V}_t^{\text{i}}\right) \end{aligned} \tag{9}$$

The objective function of the refinery planning under uncertainty is maximizing the expected profit as shown in Eq. (9). The total cost of the materials constitutes $f_1(\cdot)$ in Eq. (1); the revenue of the final products, the total cost of the units, and the inventory constitute $f_2(\cdot)$ in Eq. (1) and Eq. (2). First-stage variables are the amount of materials (**M**) including the crude oils and blend components (e.g., alkylate and n-butane). Second-stage variable are the amount of final products (**P**), the other operational variables, the capacity of the main units (e.g., processing units **PU**, crude oil distillation units **CDU**, and hydrotreating units **HTU**), the final inventory of the tanks (**T**). Price_p, Cost_m, Cost_u, and Cost_t are the deterministic parameters of the price and cost values. It should be noted that the following constraints Eqs. (11)-((46)) on the second-stage variables are indexed with scenario subscript s.

Supply constraints

The supply materials are constrained between the minimum and maximum available amounts from the market, as shown in Eq. (10). TBP temperatures $TBP_{m,s}$ of the crude oil ($\mathbf{CO} \subseteq \mathbf{M}$) are uncertain variables described before. The probabilistic scenarios **S** of the TBP temperatures $\text{TBP}_{m,s}$ are generated via the RVS method, and the probability ψ_s^{Nr} of each scenario have been calculated via the normalization approach. The other qualities **Q** are fixed to deterministic values $\text{Q}_{m,q}^{\text{fix}}$.

$$\text{A}_m^{\min} \le V_m \le \text{A}_m^{\max}\ \ m \in \mathbf{M} \tag{10}$$

$$TBP_{m,s} = (\text{TBP}_{m,s}, \psi_s^{Nr})\ \ m \in \mathbf{CO}, s \in \mathbf{S} \tag{11}$$

$$Q_{m,q,s} = \text{Q}_{m,q}^{\text{fix}}\ \ m \in \mathbf{M}, q \in \mathbf{Q}\ \ , s \in \mathbf{S} \tag{12}$$

Demand constraints

The product demands are constrained between minimum and maximum required amounts in the planning horizon, as shown in Eq. (13).

$$\text{D}_p^{\min} \le V_{p,s} \le \text{D}_p^{\max}\ \ p \in \mathbf{P}, s \in \mathbf{S} \tag{13}$$

Blender costraints

The material balances of the blenders are described by Eqs. (15) and (16). The blending components (**BC**) are blended into the products. The recipes of the blend products are defined by the tuples (c, p) which state that a component c is in the recipe for the product p. The blenders (**B**) have multiple blend products defined by the tuples (p, b) which means a product p is blended in the blender b. Binary variable b_p is equal to 1 or 0 representing a product is blended or not in Eq.(17). The minimum and maximum blend rates of a blender during the planning horizon L are denoted as R_b^{min} and R_b^{max} in Eq. (16) and (17). The specifications of the blend product are constrained by Eqs. (18)-(22), where $R_{c,p}^{min}$ and $R_{c,p}^{max}$ are the minimum and maximum fractions of the blend recipe from a component c to a product p. The quality properties **Q** blended on a volumetric basis ($\mathbf{VQ} \subset \mathbf{Q}$) and on a weight basis ($\mathbf{WQ} \subset \mathbf{Q}$) are assumed to be linear in Eqs. (19)-(22), where $Q_{p,q}^{min}$ and $Q_{p,q}^{max}$ are the bounds of the quality q specification on a product p, the correlation of the component, the product and the quality are defined by tuples (c, p, q).

$$V_{p,s} = \sum\nolimits_{c\in \mathbf{BC}|(c,p)} V_{c,p,s} \; p \in \mathbf{P}, s \in \mathbf{S} \tag{14}$$

$$\sum_{p\in \mathbf{P}|(c,p)} V_{c,p,s} = V_{c,s} \; c \in \mathbf{BC}, s \in \mathbf{S} \tag{15}$$

$$R_b^{min} \times b_{p,s} \le V_{p,s} \le R_b^{max} \times \mathrm{L} \times b_{p,s} \; (p \in \mathbf{P}, b \in \mathbf{B}), s \in \mathbf{S} \tag{16}$$

$$\sum\nolimits_{p\in \mathbf{P}|(p,b)} V_{p,s} \le R_b^{max} \times \mathrm{L} \; b \in \mathbf{B} \tag{17}$$

$$R_{c,p}^{min} \times V_{p,s} \le V_{c,p,s} \le R_{c,p}^{max} \times V_{p,s} \; (c \in \mathbf{BC}, p \in \mathbf{P}), s \in \mathbf{S} \tag{18}$$

$$QV_{c,p,q,s} = V_{c,p,s} \times Q_{c,q,s} \; (c \in \mathbf{BC}, p \in \mathbf{P}, q \in \mathbf{VQ}), s \in \mathbf{S} \tag{19}$$

$$QM_{c,p,q,s} = QV_{c,p,\text{'sg'},s} \times Q_{c,q,s} \; (c \in \mathbf{BC}, p \in \mathbf{P}, q \in \mathbf{WQ}), s \in \mathbf{S} \tag{20}$$

$$Q_{p,q}^{min} \times V_{p,s} \le \sum_{c\in \mathbf{BC}|(c,p)} QV_{c,p,q,s} \le Q_{p,q}^{max} \times V_{p,s} \quad (c \in \mathbf{BC}, p \in \mathbf{P}, q \in \mathbf{VQ}), s \in \mathbf{S} \tag{21}$$

$$Q_{p,q}^{min} \times \sum_{c\in \mathbf{BC}|(c,p)} QV_{c,p,\text{'sg'},s} \le \sum_{c\in \mathbf{BC}|(c,p)} QM_{c,p,q,s} \le Q_{p,q}^{max} \times \sum_{c\in \mathbf{BC}|(c,p)} QV_{c,p,\text{'sg'},s} \; (c \in \mathbf{BC}, p \in \mathbf{P}, q \in \mathbf{WQ}), s \in \mathbf{S} \tag{22}$$

Storage tank constraints

The blend products and blend components are stored in individual tanks. Each tank has its initial inventory V_t^i and initial quality $Q_{t,q}^i$ in Eq. (23), and is constrained between the minimum V_t^{min} and maximum V_t^{max} inventory levels in Eq.(24). The minimum and maximum total withdrawal rates are R_t^{min} and R_t^{max} in Eq.(25). The blending rules are the same as blenders assuming to be linear in Eqs. (26) and (27).

$$\sum_{ti\in \mathbf{TI}|(ti,t)} V_{ti,s} + V_t^i = V_{t,s} + \sum_{to\in \mathbf{TO}|(t,to)} V_{to,s} \; t \in \mathbf{T}, s \in \mathbf{S} \tag{23}$$

$$V_t^{min} \le V_{t,s} \le V_t^{max} \; t \in \mathbf{T}, s \in \mathbf{S} \tag{24}$$

$$R_t^{min} \times \mathrm{L} \le \sum_{to\in \mathbf{TO}|(t,to)} V_{to,s} \le R_t^{max} \times \mathrm{L} \; t \in \mathbf{T}, s \in \mathbf{S} \tag{25}$$

$$\sum_{ti\in \mathbf{TI}|(ti,t)} V_{ti,s} \times Q_{ti,q,s} + V_t^i \times Q_{t,q}^i = V_{t,s} \times Q_{t,q,s} + \sum_{to\in \mathbf{TO}|(t,to)} V_{to,s} \times Q_{to,q,s} \; (ti \in \mathbf{TI}, t \in \mathbf{T}, to \in \mathbf{TO}, q \in \mathbf{VQ}), s \in \mathbf{S} \tag{26}$$

$$\sum_{ti\in \mathbf{TI}|(ti,t)} V_{ti,s} \times Q_{ti,q,s} \times Q_{ti,\text{'sg'},s} + V_t^i \times Q_{t,q}^i \times Q_{t,\text{'sg'}}^i = V_{t,s} \times Q_{t,q,s} \times Q_{t,\text{'sg'},s} + \sum_{to\in \mathbf{TO}|(t,to)} V_{to,s} \times Q_{to,q,s} \times Q_{to,\text{'sg'},s} \; (ti \in \mathbf{TI}, t \in \mathbf{T}, to \in \mathbf{TO}, q \in \mathbf{WQ}), s \in \mathbf{S} \tag{27}$$

Unit constraints with a single operating mode

The mixers (**MIX**), splitters (**SPL**), and hydrotreating units (**HTU**) in the refinery model are the units with a single operating model. The constraints consist of the material balance in Eq. (28)-(30), the capacity in Eq. (31), and quality balance in Eq. (32) and (33). Each hydrotreating unit has a fixed yield $Y_{uo,u}^{fix}$ operating mode in Eq. (30) and fixed quality $Q_{uo,q}^{fix}$ of the outlet streams in Eq. (34).

$$V_{u,s} = \sum_{ui\in \mathbf{UI}|(ui,u)} V_{ui,s} \; u \in \mathbf{MIX}, \mathbf{SPL}, \mathbf{HTU}, s \in \mathbf{S} \tag{28}$$

$$\sum_{uo\in \mathbf{UO}|(uo,u)} V_{uo,s} = V_{u,s} \; u \in \mathbf{MIX}, \mathbf{SPL}, s \in \mathbf{S} \tag{29}$$

$$V_{uo,s} = Y_{uo,u}^{fix} \times V_{u,s} \; (uo \in \mathbf{UO}, u \in \mathbf{HTU}), s \in \mathbf{S} \tag{30}$$

$$R_u^{min} \times \mathrm{L} \le V_{u,s} \le R_u^{max} \times \mathrm{L} \; u \in \mathbf{MIX}, \mathbf{SPL}, \mathbf{HTU}, s \in \mathbf{S} \tag{31}$$

$$\sum_{ui\in \mathbf{UI}|(ui,u)} V_{ui,s} \times Q_{ui,q,s} = \sum_{uo\in \mathbf{UO}|(u,uo)} V_{uo,s} \times Q_{uo,q,s} \; (ui \in \mathbf{UI}, u \in \mathbf{MIX}, \mathbf{SPL}, uo \in \mathbf{UO}, q \in \mathbf{VQ}), s \in \mathbf{S} \tag{32}$$

$$\sum_{ui\in \mathbf{UI}|(ui,u)} V_{ui,s} \times Q_{ui,q,s} \times Q_{ui,\text{'sg'},s} = \sum_{uo\in \mathbf{UO}|(u,uo)} V_{uo,s} \times Q_{uo,q,s} \times Q_{uo,\text{'sg'},s} \; (ui \in \mathbf{UI}, u \in \mathbf{MIX}, \mathbf{SPL}, uo \in \mathbf{UO}, q \in \mathbf{WQ}), s \in \mathbf{S} \tag{33}$$

$$Q_{uo,q,s} = Q_{uo,q}^{fix} \; (u \in \mathbf{HTU}, uo \in \mathbf{UO}), \; q \in \mathbf{Q}, s \in \mathbf{S} \tag{34}$$

Units constraints with multiple operating modes

The processing units (**PU**) have more than one operating mode. The operating modes of processing units are denoted as $op \in \mathbf{OP}$. The associations between processing units and operating modes are described by tuples (op, u). The constraints of material balance in Eqs. (35) and (36), capacity in Eq. (37), and quality balance in Eq. (38) for the operating modes are almost the same as those for the hydrotreating units, except for extra binary variables b_{op} which equal to 1 or 0 specifying whether a mode is operated or not in Eq. (37). The total number of the operating modes and the total throughput of processing units are constrained by Eqs. (39) and (40).

$$V_{op,s} = \sum_{ui\in \mathbf{UI}|(ui,op)} V_{ui,s} \; op \in \mathbf{OP}, s \in \mathbf{S} \tag{35}$$

$$V_{uo,s} = Y_{uo,op}^{fix} \times V_{op,s} \; (op \in \mathbf{OP}, uo \in \mathbf{UO}), s \in \mathbf{S} \tag{36}$$

$$R_{op}^{min} \times \mathrm{L} \times b_{op,s} \le V_{op,s} \le R_{op}^{max} \times \mathrm{L} \times b_{op,s} \; op \in \mathbf{OP}, s \in \mathbf{S} \tag{37}$$

$$Q_{uo,q,s} = Q^{\text{fix}}_{uo,q}\ \left(op \in \mathbf{OP}, uo \in \mathbf{UO}\right), q \in \mathbf{Q}, s \in \mathbf{S} \tag{38}$$

$$\sum_{op \in \mathbf{OP}|(op,u)} b_{op,s} \le \mathrm{N}_u\ u \in \mathbf{PU}, s \in \mathbf{S} \tag{39}$$

$$\mathrm{R}^{\min}_u \times \mathrm{L} \le \sum_{op \in \mathbf{OP}|(op,u)} V_{op,s} \le \mathrm{R}^{\max}_u \times \mathrm{L}\ u \in \mathbf{PU}, s \in \mathbf{S} \tag{40}$$

Crude oil distillation unit constraints

Product tri-section CDU model developed by Li et al. (Li et al., 2020) is employed in this work, since it has accuracy comparable to a rigorous tray to tray CDU model, and its computational efficiency has been shown in the studies of the deterministic refinery planning problem. Since the CDU model determines how crude oil is processed via downstream units, we will describe it briefly. Tri-section CDU model describes the TBP curves of the distillation products (e.g., outlet streams of the unit uo) via three sections, namely the front-section ($1\% - 30\%$LV), the middle-section ($30\% - 70\%$LV), and the back-section ($70\% - 99\%$LV). In the middle-section, the TBP curve is approximated by a straight line, and the TBP temperatures at ($30\%, 50\%, 70\%$LV). are calculated from that line function, which is computed via partial least squares (PLS) models. In the front-section and in the back-section, the TBP temperatures at ($1\%, 5\%, 10\%; 90\%, 95\%, 99\%$LV) are calculated by adding deviations to the extrapolation of the straight line through the middle-section. The deviations, which have also been computed via PLS models, correspond to the cumulative cut width (ccw_{uo}), cut point temperatures (ct_{uo}), the TBP temperatures of the mixture inlet stream (TBP_{ui}), and operating variables (ov_u). The general formulation of the tri-section CDU model is shown in Eq. (41), where $\boldsymbol{F}_u$ is a series of linear functions corresponding to different TBP points and different distillation products.

$$TBP_{uo,s} = \boldsymbol{F}_u\left(ccw_{uo,s}, ct_{uo,s}, TBP_{ui,s}, ov_{u,s}\right)$$
$$\left(ui \in \mathbf{UI}, u \in \mathbf{CDU}, uo \in \mathbf{UO}\right), s \in \mathbf{S} \tag{41}$$

The TBP temperatures of the distillation products and quality profiles Q^{profile}_q are used to compute the quality of the products via Eq. (42).

$$Q_{uo,q,s} = g_q\left(TBP_{uo,s}, Q^{\text{profile}}_q\right)\ q \in \mathbf{Q}, s \in \mathbf{S} \tag{42}$$

The constraints of material balance and capacity for the CDU are shown in Eq.(43)-(46):

$$V_{u,s} = \sum_{ui \in \mathbf{UI}|(ui,u)} V_{ui,s}\ u \in \mathbf{CDU}, s \in \mathbf{S} \tag{43}$$

$$V_{uo,s} = Y_{uo,u} \times V_{u,s}\ \left(uo \in \mathbf{UO}, u \in \mathbf{CDU}\right), s \in \mathbf{S} \tag{44}$$

$$Y_{uo,u} = h\left(ccw_{uo,s}\right)\ \left(uo \in \mathbf{UO}, u \in \mathbf{CDU}\right), s \in \mathbf{S} \tag{45}$$

$$\mathrm{R}^{\min}_u \times \mathrm{L} \le V_{u,s} \le \mathrm{R}^{\max}_u \times \mathrm{L}\ u \in \mathbf{CDU}, s \in \mathbf{S} \tag{46}$$

The entire refinery planning model is a mixed-integer nonlinear programming (MINLP) problem formulated as a two-stage stochastic programming model.

4. Case studies

The refinery planning model used in the case studies is based on the flowsheet shown in Fig. 8. The detailed flowsheets of the refinery are shown in the Supporting Information. In this refinery model, there is one crude distillation unit (cdu) dealing with the mixture of two crude oils (CO1 and CO2); five hydrotreaters with a single operating mode for naphtha, diesel, gasoil to FCC, gasoil to hydrocracker, and residue (nht, dht, goth_fcc, goth_hc, and rht); one catalytic reforming unit with two operating modes for maximum n-butane and maximum reformate (reformerA and reformerB); one hydrocracking unit with three operating modes for gasoline, kerosene, and diesel (hc_gm, hc_km, and hc_dm); one fluid catalytic cracking unit with two operating modes for regular gasoline and premium gasoline (fccA and fccB); three blenders for gasoline, kerosene and diesel products. The products have two grades of gasoline (PG and RG), one grade of kerosene (K1), and two grades of diesel (D1 and D2). The total crude oil throughput is 100000 bbl during the planning horizon. It is assumed that the crude oil assay data is known when the decisions of purchase are made. But the quality is still uncertain until the crude arrives at the plant. The refinery planning problem under crude oil quality uncertainty is to find the optimum crude oil purchase that has the maximum expected profit.

The two-stage stochastic refinery planning model has been first studied by using a different number of scenarios. Between 3 to 30 scenarios, in increments of 3 (10 cases in total), have been used to test the computational time of different numbers of scenarios. Crude oil TBP curves vary $\pm 15°$F to $\pm 20°$F from the deterministic TBP curves. The statistics of the models are listed in Table 1. All models have been built in GAMS 26.1.0 and have been solved by ANTIGONE 1.1. The MILP solver and NLP solver used by ANTIGONE are CPLEX and CONOPT. The computing platform was Windows 10 PC with i7-8700 3.2GHz CPU and 16 GB memory. An optimality gap of 0.01% and a maximum solving time of 36000s have been set to be the termination criteria for the solver.

The size of the model increases linearly with the number of scenarios. Since the models are fairly large, the solver was not able to solve most of them from a cold start. We have adopted a two-step approach to be able to solve these models. The first step is to solve the deterministic models corresponding to each scenario (crude TBP curve) simultaneously, since the deterministic refinery model needs relatively short computational times (about 200s as reported by Li et al.) (Li et al., 2020). The crude oil recipes of all deterministic models are fixed at CO1:CO2=0.5:0.5, and all results of the deterministic models are collected and saved via GAMS grid feature. Then, the second step is to solve the stochastic model using the results from the first step as a starting point, where the crude oil recipe and operational conditions are optimized. This approach makes the large-scale stochastic models solvable in modest times. The times required to find the start points and to optimize the stochastic models are listed in Table 1. The total times of the two-step approach are also shown in Table 1. The profit, expected revenue, expected total cost, and crude oil recipes of all models are listed in Table 2.

The difference between the deterministic optimization and stochastic optimization of the crude oil selection is illustrated for the 30-scenario model. The optimum recipes of the deterministic models are plotted in Fig. 9 as two color bars, while the blue dashed line at 0.7012 fraction of CO2 shows the optimum recipe for the 30-scenarios stochastic model. Crude oil 2 is more profitable than crude oil 1. In deterministic cases, the fraction of CO2 varies between 0.8162 and 1.000. When the uncertain quality of crude oils is considered, the fraction of crude oil 2 decreases to 0.7012 to attain the optimum expected profit of the entire refinery.

Additional crude oils can be included, as long as their distillation curves are between the lightest and the heaviest crude oils which have been used to develop the product tri-section CDU model. Product tri-section CDU model development needs to be based on a set of distillation curves which cover the span between the lightest and the heaviest crude oil. Increased number of crude

Table 1
Model statistics for different number of scenarios.

No. of scenarios	No. of variables	No. of integer variables	No. of equations	No. of nonlinear terms	Solution status	Optimization time (s)	Find start point (s)	Total solving time (s)
3	6110	1272	6289	2743	Optimum	5267.68	234.32	5502.00
6	12197	2544	12547	5485	Optimum	2988.10	297.41	3285.51
9	18284	3816	18805	8227	Optimum	3689.23	363.54	4052.77
12	24371	5088	25063	10969	Optimum	154.51	405.01	559.52
15	30458	6360	31321	13711	Optimum	5434.67	493.61	5928.28
18	36545	7632	37579	16453	Optimum	14106.01	768.15	14874.16
21	42632	8904	43837	19195	Optimum	964.94	691.93	1656.87
24	48719	10176	50095	21937	Optimum	4633.02	1039.34	5672.36
27	54806	11448	56353	24679	Optimum	382.80	1112.44	1495.24
30	60893	12720	62611	27421	Optimum	1704.55	1104.26	2808.81

Table 2
Objective values and crude oil recipes results.

No. of scenarios	Profit (M$)	Expected revenue (M$)	Expected total cost (M$)	Fraction of CO1	Fraction of CO2
3	2.9231	9.0930	6.1698	0.2942	0.7058
6	2.9000	9.0734	6.1734	0.2985	0.7015
9	2.8990	9.0526	6.1536	0.2826	0.7174
12	2.9615	9.0354	6.0739	0.2060	0.7940
15	2.6650	8.8495	6.1845	0.3325	0.6675
18	2.6969	8.8457	6.1487	0.2988	0.7012
21	2.6680	8.8529	6.1850	0.3326	0.6674
24	2.7230	8.8481	6.1251	0.2717	0.7283
27	2.6832	8.8335	6.1503	0.2987	0.7013
30	2.6912	8.8395	6.1483	0.2988	0.7012

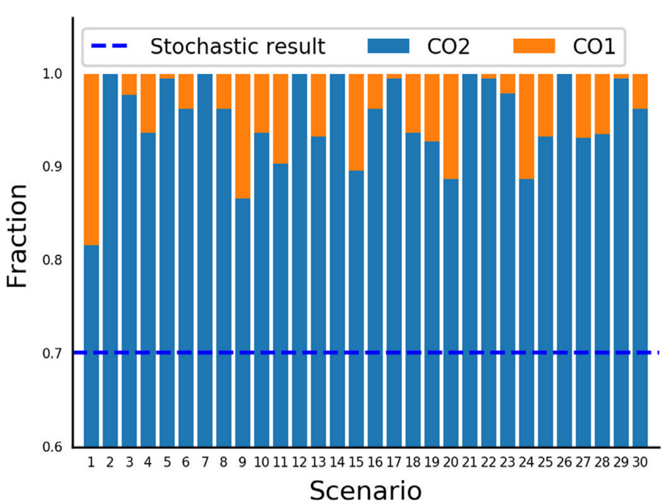

Fig. 9. Crude oil recipes of the deterministic models.

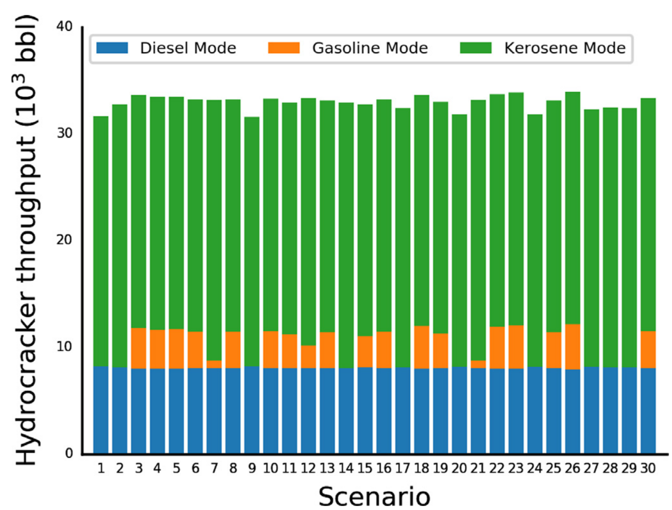

Fig. 10. Hydrocracker throughput of the deterministic models.

oils will require a larger number of scenarios and longer execution times.

If the probability scenarios of the crude oil quality (TBP curve) are taken into account in individual deterministic planning models, the resulting crude oil selection and mixed recipe fluctuate significantly, as shown in Fig. 9. Meanwhile, the variability of the crude oil quality impacts the throughput and operating modes of the downstream secondary process units. For instance, the hydrocracking unit, as shown in Fig. 10, has different throughput (total height of a bar) and different throughput distributions of diesel mode, gasoline mode, and kerosene mode for each scenario, which means the operation conditions need to change according to the uncertainty of the crude oil quality.

The comparison between the proposed two-stage stochastic programming method with the uncertainty of crude oil TBP curves and the deterministic approach with crude oil assay data are carried by two scenes to address the crude oil selection and production planning problem. Both scenes consider the planning (decision making) before and after the realization of the crude quality uncertainty. (i) Before the realization, select the crude oil recipe considering the uncertainty by stochastic programming. (ii) Before the realization, select the crude oil recipe using the qualities in crude assay data, which means the uncertainty is not considered. After the realization, since the crude recipe has been decided, the operation planning is optimized. The profits of the decisions made before the realization are treated as the benchmark profits to cal-

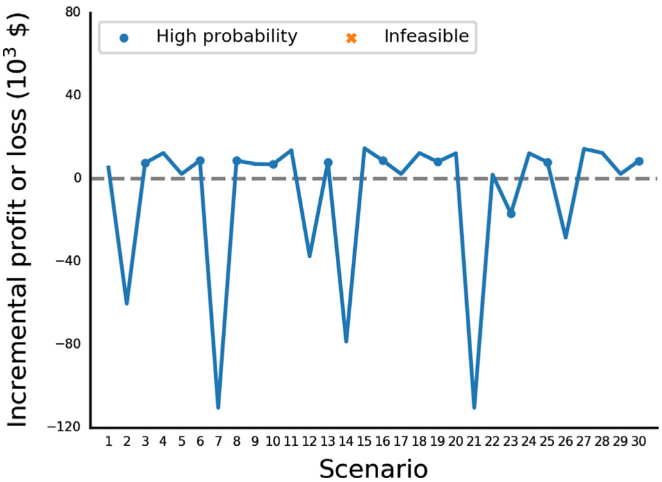

Fig. 11. Results from stochastic programming method crude selection.

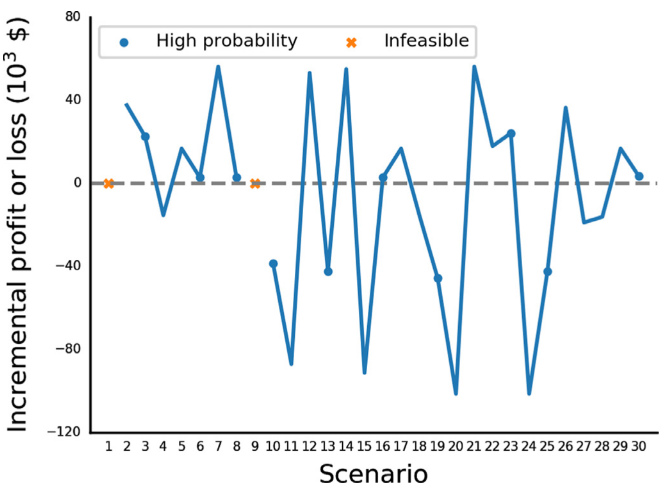

Fig. 12. Results from deterministic approach to crude selection

culate the profit increment or loss after the realization, as shown in Fig. 11 for the stochastic programming method and Fig. 12 for the deterministic approach. In Fig. 11 and Fig. 12, the scenarios with high probabilities are marked as blue points, which totally contain 99% probability. The decision made by the stochastic programming method has an expected incremental profit of 5463.8$ after the realization of the uncertainty. On the other hand, the deterministic approach has expected incremental 11033.1$ loss and even has two infeasible scenarios.

5. Conclusions

We have proposed a novel approach to address the refinery production planning problem under uncertainty of crude oil qualities in this work. Unlike Yang and Barton (Yang and Barton, 2015), who described crude oil quality uncertainty indirectly via an uncertainty in the yield of one of the products (which implies that the yields of other products are scaled proportionally), we employ variations in crude oil TBP curves to describe crude oil quality uncertainty. This approach enables us to consider the actual crude oil quality variations, which in turn enables us to predict realistic CDU yields via the product tri-section CDU model. The selection and purchase of crude oils with uncertain qualities are solved via a two-stage stochastic model. The amount of crude oils are determined as the first-stage decision while maximizing the expected profit. Other operating conditions and the amount of products are determined as the second-stage decisions after the realization of the uncertainty of crude oil qualities while minimizing the expected cost.

The uncertainty in the crude oil quality is described by the uncertain parameters that define the beta function approximating the crude oil TBP curve. The uncertain TBP curve simultaneously describes the uncertainty of the product yields and of the crude oil bulk properties. Amount of each pseudo-component is calculated to represent a given instantiation of the TBP crude oil curve. Each pseudo-component has associated values of other qualities (sulfur, gravity, etc.). By using the product-tri section CDU model, we predict the TBP curves of each product and calculate associated yields and qualities.

The two-stage stochastic refinery planning model stems from a deterministic model with the product tri-section CDU model. In order to consider a wide range of probabilistic scenarios for the two-stage stochastic programming, the discretization of the continuous random variables and the random vector sampling method have been used, which improves the approximation of the population using a relatively small number of scenarios. The scale of the stochastic linearly with the number of scenarios resulting in a large-scale mixed-integer nonlinear programming model.

The solution times of the models show that the two-step methodology (calculating initial starting point from deterministic cases, followed by the optimization of the stochastic model) combined with the product tri-section CDU model is an effective way to solve crude selection under uncertainty, while including a realistic, detailed refinery representation in the model.

The case studies on the production planning under uncertainty of crude oil qualities show that the proposed approach of using the uncertain TBP curve is appropriate to address the uncertainty in crude oil purchasing and production planning. The methodology provides an optimal crude oil selection and recipe supporting the decision-making on the purchases of crude oils. This is the first study that considers uncertainties in crude oils TBP curves and successfully solves the associated entire refinery and crude oil selection stochastic optimization problem.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Fupei Li: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Feng Qian:** Supervision, Funding acquisition. **Wenli Du:** Conceptualization, Resources, Project administration. **Minglei Yang:** Resources, Investigation. **Jian Long:** Resources, Investigation. **Vladimir Mahalec:** Conceptualization, Writing – review & editing, Resources, Supervision.

Acknowledgements

This work has been supported by National Natural Science Foundation of China (Basic Science Center Program: (61988101), International (Regional) Cooperation and Exchange Project (61720106008) and National Natural Science Fund for Distinguished Young Scholars (61925305). and by McMaster University WBooth School of Engineering Practice and Technology. We thank Sayyed Faridoddin Afzali for proposing the random vector sampling method to minimize the number of scenarios used in the two-stage stochastic programming problem.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.compchemeng.2021.107361.

References

Afzali, S.F., Cotton, J.S., Mahalec, V., 2020. Urban community energy systems design under uncertainty for specified levels of carbon dioxide emissions. Appl. Energy 259, 114084. doi:10.1016/j.apenergy.2019.114084.

Chen, C., Zhou, L., Ji, X., He, G., Dai, Y., Dang, Y., 2020a. Adaptive modeling strategy integrating feature selection and random forest for fluid catalytic cracking processes. Ind. Eng. Chem. Res. 59, 11265–11274. doi:10.1021/acs.iecr.0c01409.

Chen, Y., Lin, M., Jiang, H., Yuan, Z., Chen, B., 2020b. Optimal design and operation of refinery hydrogen systems under multi-scale uncertainties. Comput. Chem. Eng. 138, 106822. doi:10.1016/j.compchemeng.2020.106822.

Dai, X., Wang, X., He, R., Du, W., Zhong, W., Zhao, L., Qian, F., 2019. Data-driven robust optimization for crude oil blending under uncertainty. Comput. Chem. Eng. 106595. doi:10.1016/j.compchemeng.2019.106595.

Durrani, M., Ahmad, I., Kano, M., Hasebe, S., 2018. An artificial intelligence method for energy efficient operation of crude distillation units under uncertain feed composition. Energies 11. doi:10.3390/en11112993, 2993–2912.

Evazabadian, F., Arvan, M., Ghodsi, R., 2019. Short-term crude oil scheduling with preventive maintenance operations: a fuzzy stochastic programming approach. Int. T Oper. Res. 26, 2450–2475. doi:10.1111/itor.12408.

Franzoi, R.E., Menezes, B.C., Kelly, J.D., Gut, J.A.W., Grossmann, I.E., 2020. Cutpoint temperature surrogate modeling for distillation yields and properties. Ind. Eng. Chem. Res. 59, 18616–18628. doi:10.1021/acs.iecr.0c02868.

Fu, G., Sanchez, Y., Mahalec, V., 2015. Hybrid model for optimization of crude oil distillation units. AIChE J. 62, 1065–1078. doi:10.1002/aic.15086.

Gao, X., Shang, C., Jiang, Y., Huang, D., Chen, T., 2014. Refinery scheduling with varying crude: a deep belief network classification and multimodel approach. AIChE J. 60, 2525–2532. doi:10.1002/aic.14455.

Gutierrez, G., Galan, A., Sarabia, D., Prada, C.de, 2018. Two-stage stochastic optimization of a hydrogen network. Ifac-Papersonline 51, 263–268. doi:10.1016/j.ifacol.2018.09.310.

Leiras, A., Hamacher, S., Elkamel, A., 2010. Petroleum refinery operational planning using robust optimization. Eng. Optimiz. 42, 1119–1131. doi:10.1080/03052151003686724.

Li, F., Qian, F., Yang, M., Du, W., Mahalec, V., 2020. Product tri-section based crude distillation unit model for refinery production planning and refinery optimization. Aiche J. 67. doi:10.1002/aic.17115.

Lima, C., Relvas, S., Barbosa-Póvoa, A., Morales, J.M., 2019. Adjustable robust optimization for planning logistics operations in downstream oil networks. Process 7, 507. doi:10.3390/pr7080507.

Mavromatidis, G., Orehounig, K., Carmeliet, J., 2018. Design of distributed energy systems under uncertainty: a two-stage stochastic programming approach. Appl. Energ. 222, 932–950. doi:10.1016/j.apenergy.2018.04.019.

Minh, L.Q., Duong, P.L.T., Lee, M., 2018. Global sensitivity analysis and uncertainty quantification of crude distillation unit using surrogate model based on gaussian process regression. Ind. Eng. Chem. Res. 57, 5035–5044. doi:10.1021/acs.iecr.7b05173.

Qian, F., Zhong, W., Du, W., 2017. Fundamental theories and key technologies for smart and optimal manufacturing in the process industry. Engineering 3, 154–160. doi:10.1016/j.eng.2017.02.011.

Sánchez, S., Ancheyta, J., McCaffrey, W.C., 2007. Comparison of probability distribution functions for fitting distillation curves of petroleum. Energy Fuels 21, 2955–2963. doi:10.1021/ef070003y.

Shin, J., Lee, J.H., 2018. Multi-timescale, multi-period decision-making model development by combining reinforcement learning and mathematical programming. Comput. Chem. Eng. 121, 556–573. doi:10.1016/j.compchemeng.2018.11.020.

Wang, J., Feng, Y., Rong, G., 2009. Optimizing crude oil operations under uncertainty. Ifac Proc. Volumes 42, 1020–1025. doi:10.3182/20090603-3-ru-2001.0061.

Wang, S., Wallace, S.W., Lu, J., Gu, Y., 2020. Handling financial risks in crude oil imports: Taking into account crude oil prices as well as country and transportation risks. Transp. Res. Part E Logist. Transp. Rev. 133, 101824. doi:10.1016/j.tre.2019.101824.

Yang, Y., Barton, P.I., 2015. Integrated crude selection and refinery optimization under uncertainty. AIChE J. 62, 1038–1053. doi:10.1002/aic.15075.

Zhang, J., Zhu, X.X., Towler, G.P., 2001. A level-by-level debottlenecking approach in refinery operation. Ind. Eng. Chem. Res. 40, 1528–1540. doi:10.1021/ie990854w.