Optimal Design of Flexible Heat-Integrated Crude Oil Distillation Systems

A thesis submitted to The University of Manchester for the degree of
Doctor of Philosophy
in the Faculty of Science and Engineering

2018

Dauda Ibrahim

School of Chemical Engineering and Analytical Science

Table of Content

List of Figures

List of Tables

Optimal Design of Flexible Heat-Integrated Crude Oil Distillation Systems

Dauda Ibrahim
The University of Manchester
2017

Abstract —PhD Thesis

The need for petroleum refineries to process different types of crude oil in order to maximise profit margin and to meet demand for products, calls for flexibility in the design and optimisation of crude oil distillation systems comprising distillation units and the heat recovery network. Crude oil distillation is a complex, capital- and energy-intensive process. The large number of degrees of freedom (column structure and operating conditions) and complex interactions within the system make the design and optimisation of crude oil distillation system a highly challenging task. This work develops new methodologies for the design of crude oil distillation systems that process a single crude oil feedstock and multiple crude oil feedstocks.

In this work, the crude oil distillation unit is modelled using a rigorous tray-by-tray model where the number of trays active in each section is also a design degree of freedom. The model is embedded in an optimisation framework, together with a heat recovery model (applying pinch analysis), for design of an energy-efficient and cost-effective distillation system. The optimisation framework addresses both structural and operational degrees of freedom of the system, capturing the trade-off between capital and energy costs, and accounting for heat integration. The distillation model is built in Aspen HYSYS, while the optimisation is carried out in MatLab using a genetic algorithm, where data is exchanged during process simulation and optimisation.

To overcome the shortcomings of the rigorous distillation model in the context of system optimisation, surrogate models based on artificial neural networks (ANN) and a support vector machine (SVM) are developed and applied in the optimisation framework. The ANN model simulates the crude oil distillation unit, while the SVM partitions the search space, increasing the likelihood that the optimised solution will converge when simulated using a rigorous model. The SVM helps to reduce computational effort by focusing the search on potentially feasible solutions. Both the ANN and SVM are fitted to results of multiple rigorous simulations of the distillation unit.

The proposed surrogate modelling approach is extended to take into account multiple crude oil feedstocks in the design of the distillation unit. The distillation column models for multiple crude oils and heat recovery model are embedded in a two-stage optimisation framework, in which a hybrid stochastic-deterministic approach is applied to optimise structural variables and distillation column operating conditions. The overall objective is to maximise net profit while meeting product quality (and flow rate) constraints.

The capabilities of the proposed methodologies are illustrated using industrially-relevant case studies. Results indicate that the used of surrogate model instead of rigorous models reduces computational time without compromising solution accuracy and optimality. The design approach to account for flexible operation is shown to identify effectively design alternatives that are economically viable and operable over the range of crude oil feedstocks.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Dauda Ibrahim

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademark sand other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University's policy on Presentation of Theses

Dedication

To my parents, Ibrahim D. Miringa and late Aishatu I. Dauda for their support, encouragement, and love throughout my life.

To my brothers and sisters, Maryam, Musa, Hauwa, Mohammed, Hauwa and Habiba for the prayers and support.

Acknowledgement

I would like to express my appreciation to my supervisors, Dr Megan Jobson, Dr Gonzalo Guillén-Gosálbez and Dr Jie Li for their assistance and guidance throughout this research work. Special thanks to Dr Megan for her support, advice and encouragement, constructive suggestions and feedbacks throughout this project. Also, special thanks to Dr Gonzalo for his support, guidance, advice, and for enrolling me in an optimisation course at Imperial College London. I would also like to thank Dr Jie for his advice and support.

I would like to express my appreciation to Prof. Robin Smith and Dr Nan Zhang for their valuable suggestions. Also worth acknowledging are all the member of staffs in Centre for Process Integration, especially Prof. Kostas (Director of Postgraduate Research) for warmly welcoming us into the PhD programme.

I would like to acknowledge the financial support from Petroleum Technology Development Fund (PTDF), Nigeria, for sponsoring this PhD research project. Without your support, this research wouldn't have been possible.

I would also like to thank all my friends and colleagues in Centre for Process Integration (here in Manchester) and Centre for Process Systems Engineering (Imperial College London). Especially Awwal for the brotherhood and wonderful time we spent together, and Minerva for the time we spent working and sharing ideas about crude oil distillation.

My profound gratitude goes to my parents, Ibrahim D. Miringa and Late Aishatu I. Dauda for their encouragement, patience, support and prayers, which are unexpressed by words, may the Almighty God reward them with Jannatul Firdausi.

Also, worth acknowledging is my beloved sisters and brothers, Maryam, Musa, Hauwa, Mohammed, Hauwa and Habiba for their encouragement and prayers.

Chapter 1 Introduction

1.1 Context and overview of research problem

Petroleum refineries are faced with uncertainties in terms of future quality and quantity of feedstocks and products (Castelo et al., 2010). To cope with these uncertainties, refinery processes must be flexible enough to adapt to various operating scenarios arising from, for example, changes in crude oil feedstock and products, and changes in throughput. Flexibility is defined as the inherent ability of refinery process to establish feasible operation (e.g., meeting product specifications) over a wide range of operating scenarios.

Process flexibility plays an important role in maximising refinery profit margins (i.e., the total amount by which revenue from the sales of refined petroleum products exceeds operating cost). Changes in market prices of crude oil and refined petroleum products can provide opportunities to improve profit margins if the refinery processes have the capability of processing various types of crude oil feedstocks to produce market-driven products. In general, flexible operation in a petroleum refinery starts in the crude oil distillation systems (Spangler et al., 2006).

Crude oil distillation is one of the primary processes in a petroleum refinery. The crude oil distillation system consists of a crude oil distillation unit (also known as column) and a heat recovery network (also known as preheat train) where the crude oil feedstock is heated and partially vaporised (see Figure 1.1). The crude oil distillation unit, as shown in Figure 1.1, has a complicated configuration, comprising a main column equipped with pump-arounds and side-strippers, and a condenser. The pump-arounds provide local reflux and create heat recovery opportunities, while the side-strippers remove light components from side draws.

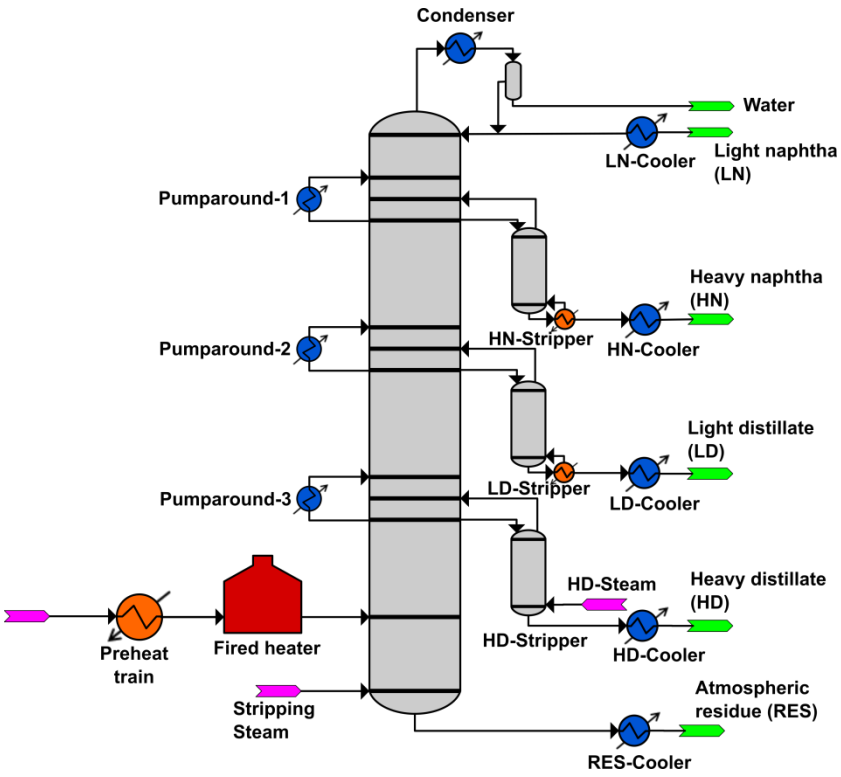


Figure 1.1 Typical refinery crude oil distillation system.

Crude oil distillation is a complex, capital- and energy-intensive operation. The overall system configuration typically consists of a furnace which consumes fuel equivalent to 1-2% of the entire crude oil being processed (Liebmann et al., 1998; Szklo and Schaeffer, 2007). This fuel combustion is associated with high CO_2 emissions and high operating costs. Due to the scale of the system, even a small energy savings can lead to significant economic and environmental benefits. Heat integration is implemented to enhance energy efficiency of the system by exchanging heat between hot streams that require cooling and cold streams that require heating.

The main purpose of the crude oil distillation system is to perform the initial separation of crude oil feedstocks into fractions or 'cuts,' which are either blended into marketable products or sold as feedstocks for the petrochemical industries. Crude oil is a complex mixture of different classes of hydrocarbons (see Chapter 2). Each type of crude oil has unique physical and chemical properties, which primarily depends on the source of the crude oil (Jones, 1995). Figure 1.2 illustrates product yield obtainable from three varieties of crude oil feedstocks, namely, Maya, Azeri light, and Brent, with

different properties, e.g., the density of the three crude oils are 924.8 kg m^{-3}, 850.9 kg m^{-3}, and 833.3 kg m^{-3} respectively.

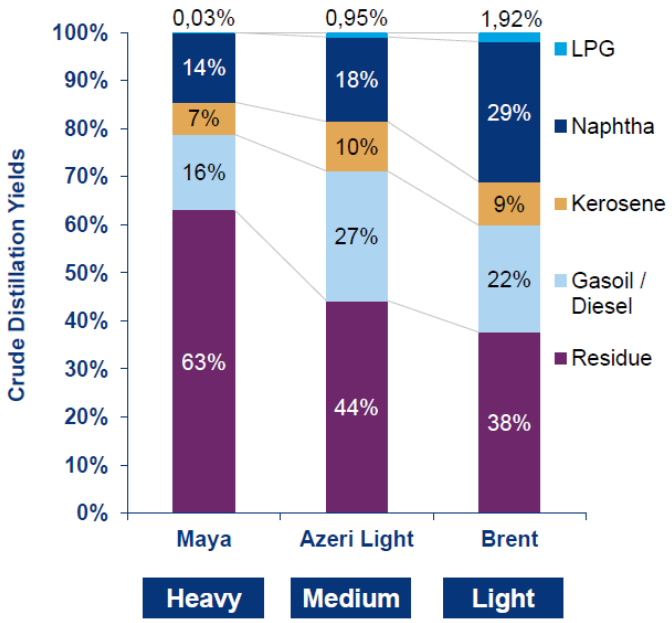


Figure 1.2 Crude oil yield for heavy, medium and light crude (Cooper and Mackenzie, 2013).

In general, light crude oils (density less than 859 kg m^{-3}) are less dense than medium (density between 859 kg m^{-3} and 921 kg m^{-3}) and heavy crude oils (density greater than 921 kg m^{-3}), and contain a significant amount of low-boiling hydrocarbon compounds (Favennec, 2001). Each type of crude oil is a unique mixture of various hydrocarbon compounds (paraffin, naphthenes, and aromatics), and therefore the crude oil distillation column should be design (fired heating, stripping steam flow rate, number of trays in column sections etc.) to accommodate changes in feedstock properties.

The variation in feedstock properties can have a significant impact on design and operation of crude oil distillation systems. For example, if a system is designed based on a specific type of crude oil, changes in feedstock properties can impact on the system performance, such as product qualities and flow rate, fired heating requirements, net profit and CO_2 emissions. Thus, to avoid economic penalties resulting from products not meeting market requirements and failing to capitalise on

the cheapest crude oil in the market, the crude oil distillation unit and the associated heat recovery network should be designed to operate satisfactorily over a wide range of operating scenarios, while ensuring product quality and flow rates are within their set values.

1.2 Grassroots design of heat-integrated crude oil distillation systems

In general, the process of design can be broadly classified into two groups based on design objectives, namely, grassroots (also known as 'greenfield design') and retrofit design. In grassroots design, the aim is to design a new process at minimum cost and/or maximum profit, while retrofit design involves making structural (e.g. installing new equipment, replacing an existing equipment, repiping etc.) and/or operational (e.g. changes in temperature, pressure, throughput etc.) modifications to an existing process in order to achieve a desired objective, for example, increased throughputs, increased capacity, and reduction in energy consumption and emissions (Smith, 2005).

Long and Lee (2017) estimated that 70-80% of capital investment project in the process industries is retrofit projects, while the remaining constitute grassroots projects. Retrofit design is more complicated than grassroots design as the retrofit design space is more restricted, and there are fewer degrees of freedom (Westerberg, 2004). Although many similarities exist between retrofit and grassroots design methodologies, there are still many fundamental differences between the two design approaches, for example, each approach requires a unique model (Grossmann et al., 1987; Westerberg, 2004). Moreover, some aspect of retrofit design requires the knowledge, understanding, skills and insights derived from grassroots design (Westerberg, 2004). This work focuses on grassroots design of a crude distillation system that can process multiple crude oil feedstocks. Heat integration is taken into account using pinch analysis (based on grand composite curve). The grand composite curve is an important tool for process heat integration. The curve is constructed using process stream data, consisting of stream supply and target temperatures and enthalpy

change. The grand composite curve is a useful tool for calculating energy targets (minimum hot and cold utilities) and for utility selection for lowest cost (Smith, 2005).

The crude oil distillation unit is strongly interconnected with the associated heat recovery network through pump-around duties, condenser duty, and product coolers. Changes in the design and operation of the atmospheric distillation unit affect the design and operation of the heat recovery network and vice versa.

Typically, grassroots design aims to select the structural (feed tray location, pump-around and side-stripper location, number of trays in each section of the column) and operational variables (feed temperature, pump-around duties and temperature drops, stripping steam flow rates and reflux ratio) of the crude oil distillation unit, while simultaneously selecting the configuration of the associated heat recovery network and the area requirements of the heat exchangers. Through effective design strategies, the synergy between the two subsystems can be exploited to achieve an overall good design performance (e.g. total annualised cost, energy consumption and profit).

The design of crude oil distillation systems involves selecting the structure and operating conditions of the distillation unit and the heat recovery network, and has the additional complexity that the overall system must be capable of processing multiple crude oil feedstocks and crude oil blends. That is, the crude oil distillation system should accommodate variations in crude oil feedstocks and operating conditions. A flexible crude oil distillation system must be capable of operating satisfactorily over a wide range of feedstocks.

The complex nature of the crude oil distillation system, including the strong interaction between the unit and the heat recovery network poses a very challenging design and optimisation problem. In addition, there are many degrees of freedom in the system, namely, operating conditions and structural variables of the crude oil distillation unit and the heat recovery network. Furthermore, these degrees of freedom need to be selected while taking into account all the crude oil feedstocks and/or crude oil blends that need to be processed.

In the past, design of crude oil distillation unit and the associated heat recovery network (preheat train) were carried out in separate steps, i.e., the distillation column was designed first, followed by the heat recovery network (Nelson, 1958; Watkins, 1979; Jones, 1995). These design approaches applied heuristic rules, empirical correlations, and experience to design the complex crude oil distillation unit. However, the design approaches proposed by Nelson (1958), Watkins (1979) and Jones (1995) require trial and error and do not account for interactions within the system.

To overcome the above limitations, several researchers focused on design of integrated crude oil distillation systems, taking into account the interactions between the distillation column and the heat recovery network. This design strategy leads to an energy-efficient design compared to earlier methods. In this regard, Liebmann and co-workers (Liebmann, 1996; Liebmann et al., 1998) integrate rigorous column simulation and pinch analysis to design the complex column. Their approach evolves the design in a stepwise manner, taking into account maximum heat recovery calculated using the grand composite curve. Based on these heat recovery targets/minimum utility targets, design modifications that improve separation and reduce energy consumption are proposed and adopted. A limitation of this work is that it requires iteration, and the distillation column is not optimised.

To avoid the use of rigorous model, Sharma et al. (1999) propose a stepwise design strategy that combines simple energy balance and column grand composite curve (Dhole and Buckingham, 1994). First, energy balance is carried out across several sections of the complex column to generate temperature-enthalpy data. The data are used to construct a column grand composite curve. Based on this curve, the maximum amount of energy that can be recovered without affecting separation is calculated. However, the number of trays in column section is fixed. Also, the design procedure is based on some rules of thumb and optimisation is not attempted.

To develop an optimisation-based design approach, Suphanit (1999) proposed simplified (based on Fenske–Underwood–Gilliland) models for crude oil distillation columns and apply pinch analysis in an optimisation framework to identify the crude

oil distillation column structure and operating conditions that minimise the total annualised cost. Rastogi (2006) improved the simplified model of Suphanit (1999) in order to account for pump-around locations and pressure drop. A non-linear optimisation technique (successive quadratic programming) is applied to optimise the column structure and operating conditions. Chen (2008) further modifies the model of Rastogi (2006) to allow for alternative pump-around locations (e.g., above the top side draw). The model is implemented in an optimisation framework, together with a heat exchanger network model, to simultaneously design the column and the heat exchanger network. In addition, Chen (2008) accounts for temperature-dependent properties of crude oil and product streams undergoing phase change. Despite being easier to handle numerically, simplified column models have the limitation that they tend to lead to less accurate results compared with rigorous models. Furthermore, they are not versatile with respect to the column configuration.

The methodologies presented so far focused on the design of crude oil distillation systems that separates a specific type of crude oil feedstock to produce intermediate products of specified quality. However, a significant change from the design conditions (e.g. change in crude oil feedstock) can impact on the overall system performance (energy consumption, profit, etc.) or even lead to infeasible operation (i.e. failure of the design to satisfy separation requirements, such as product quality). For example, ''light'' crude oil contains a large amount of low-boiling hydrocarbon compounds compared with "medium" and "heavy" crude oils; thus it is expected that the feed inlet temperature for the light crude oil will be lower than that of medium and heavy crude oils, which implies that the light crude oil will require less preheating than the medium and heavy crude oils.

To design a flexible crude oil distillation system that can process multiple crude oil feedstocks and/or blend of crude oils, it is crucial to determine the structural and operational variables of the system that can operate over different crude oil feedstock. For example, Bagajewicz and Ji (2001) propose a design approach that combines rigorous simulation and pinch anlysis based on heat demand-supply diagram. Their approach applies heuristic rules to_determine feasible operating conditions for light,

medium and heavy crude oils using a fixed crude oil distillation column structure. Nonetheless, the trade-offs between capital and energy cost are not considered. Furthermore, the resulting crude oil distillation system is not optimised.

To optimise the system, More et al. (2010) develop an optimisation-based approach for the design of crude oil distillation system that processes multiple crude oils and blend of crude oils. Their methodology consist of two stages. In stage 1, the crude oil distillation system is modelled rigorously in Aspen Plus. In stage 2, an optimisation framework is formulated, with a profit related objective function and constraints on product quality. The variables optimised include feed flow rate, steam flow rates, and product flow rates. The nonlinear optimisation tool embedded in the commercial process simulator is used to optimise the system. Although the standard optimisation tool used in this work can facilitate the search for better solutions compared to the approach presented by Bagajewicz and Ji (2001), the method of More et al. (2010) also has some drawbacks; for example, structural variables such as number of trays in each column section and the locations of feed tray, pump-arounds and side-strippers are not optimised. Furthermore, heat integration is not taken into account during the optimisation, although the heat duties are used to calculate operating cost.

Overall, early approaches (Nelson, 1958; Watkins, 1979; Jones, 1995) for the design of crude oil distillation system have not taken into account heat integration. Integrated design methods (Liebmann et al., 1998; Sharma et al., 1999) do consider heat integration, but the distillation column is not optimised. The optimisation-based approach integrates simplified column model and pinch analysis (Suphanit, 1999) / heat exchanger network (Rastogi, 2006; Chen, 2008) to design an optimised system. However, simplified models may lead to an unrealistic estimate of the distillation column performance. Again, the approaches above focused on the design of crude oil distillation system that processes one type of crude oil feedstock. While methodologies (Bagajewicz and Ji, 2001; More et al., 2010) for the design of flexible crude oil distillation system that process multiple crude oils and blends of crude oil are available, they are subject to many limitations such as lack of consideration of relevant design variables as well as trade-offs between capital and energy cost.

To develop a systematic approach for optimisation-based design of crude oil distillation system, there is a need for computationally efficient and accurate distillation column models that can be used to simulate alternative designs. The models should be capable of representing the complex behavior of the system and the relevant degrees of freedom (structural and operational variables). Rigorous and surrogate models are available for simulating crude oil distillation systems. If properly modelled, rigorous models can produce an accurate estimate of the distillation column performance as compared with surrogate models. However, rigorous models can be computationally demanding. On the other hand, surrogate models are relatively simple and are less computationally demanding. This feature makes surrogate models more suitable for implementation in an optimisation framework to design the crude oil distillation system.

Methodologies for modelling of crude oil distillation unit based on rigorous models (Bagajewicz and Ji, 2001; Basak et al., 2002; More et al., 2010) and surrogate models (López C. et al., 2013; Ochoa-Estopier and Jobson, 2015a; Osuolale and Zhang, 2017) have been developed and presented in literature. However, none of the methods have incorporate structural decisions (e.g. number of trays in column sections) as a design variable. Thus, the resulting column model cannot be applicable for optimisation-based design of crude oil distillation system. While structural variables have been implemented in modelling of simple column using rigorous models (Caballero et al., 2005), this approach can not be directly applied to design a heat-integrated crude oil distillation systems, due to the complicated nature of the unit configuration and large number of degrees of freedom.

1.3 Aims and objectives of this work

As discussed in Sections 1.1 and 1.2, there is a lack of systematic methodologies for the design of flexible heat-integrated crude oil distillation systems that process multiple crude oil feedstocks.

This research work aims to develop a new systematic approach for the design of flexible crude oil distillation systems, to address the limitations of existing design methods. The method applies an optimisation-based approach to design the crude oil distillation unit, while simultaneously considering heat recovery using pinch analysis, in a unified framework. The optimisation framework aims to select both column structural and operating degrees of freedom while taking into account multiple crude oil feedstocks, product quality constraints, capital investment and operating costs. The objectives to achieve these aims are to:

1. Develop an appropriate modelling approach for crude oil distillation units that takes into account both structural and operational degrees of freedom of the distillation column. The approach explores the use of both rigorous and surrogate models.
2. Propose a design methodology that incorporates rigorous simulation model and pinch analysis in a unified framework to facilitate the design of crude oil distillation systems.
3. Adapt the design methodology of Objective 2 to apply surrogate distillation column models, considering both column performance and definition of region of appropriate operating conditions.
4. Develop an optimisation framework that incorporates suitable distillation column models and pinch analysis to support the design of flexible crude oil distillation systems.
5. Propose an effective solution strategy to facilitate the search for flexible, cost-effective, and energy-efficient design option.
6. Demonstrate the capabilities of the proposed frameworks using industrially-relevant case studies.

1.4 Contributions of this work

The following outlines the contributions of the work presented in this thesis:

1. Design of crude oil distillation units using rigorous simulation model.

i. A new approach for representing crude oil distillation column (superstructure) using rigorous simulation model is introduced. The proposed distillation column superstructure takes into account both structural and operational degrees of freedom of the crude oil distillation unit. When the distillation column superstructure is implemented in an optimisation framework, it is possible to optimise both structural (number of trays in column sections) and operational (pump-around duty and temperature drops, feed inlet temperature, stripping steam flow rate, and reflux ratio) degrees of freedom of the system, thus allowing the inherent trade-offs between capital and energy cost to be fully captured, leading to an economically viable design alternative.

2. Design of crude oil distillation units using surrogate distillation column model.
   i. A surrogate model of the crude oil distillation unit is developed using artificial neural networks, taking into account both structural and operational degrees of freedom.
   ii. A feasibility constraint based on support vector machine is proposed in this work. The constraint is applied to rule out infeasible design alternatives from the solution space, thus improving computational efforts and increasing the likelihood that an optimal design would be feasible when simulated on a rigorous model.
   iii. Methodology that allows the optimisation of structural and operational degrees of freedom of crude oil distillation unit using surrogate model is proposed, considering feasibility constraint, heat integration and economic model.

3. Design of flexible crude oil distillation unit that process multiple feedstocks.
   i. Data sampling and surrogate modelling approaches for crude oil distillation unit that processes multiple crude oil feedstocks are proposed.

ii. A two-stage optimisation framework is proposed to facilitate the design of flexible crude oil distillation unit. The framework incorporates the surrogate column models, heat recovery models, feasibility constraints and economic model.

iii. A hybrid stochastic-deterministic strategy is introduced to aid the search for flexible and cost-effective design alternatives within the solution space.

4. Scenario-based design of flexible chemical processes.

   i. A scenario-based approach for the design of chemical processes in which some parameters and/or input variables are subject to variability is proposed. Compared to the strategy presented in Contribution 3, this approach is capable of handling, effectively, large number of operating scenarios (e.g. many crude oil feedstocks to be processed).

   ii. A decomposition strategy that breaks the complex multi-scenario design problem into sub-problems is introduced. The sub-problems include: (i) defining and characterizing process parameters that are subject to variability, (ii) design for each scenario, (iii) evaluating each design within the parameter space, and (iv) selecting the most economic and flexible design option.

   iii. Multi-criteria decision-making tool is introduced for selection of optimal design among many alternatives, taking into account both quantitative and qualitative judgement.

1.5 Overview of this Thesis

The remainder of this thesis is organised in six chapters, following the "Journal Format" style of the University of Manchester. Chapter 2 presents an overview of the crude oil distillation system, followed by a review of relevant work on modelling of these systems. Previous work on the design and optimisation of heat-integrated crude oil distillation systems, methodologies for process design for flexibility, and optimisation methods are critically discussed.

Chapter 3 presents Publication 1: Ibrahim, D., Jobson, M., Guillén-Gosálbez, G., 2017. Optimization-based Design of Crude Oil Distillation Units using Rigorous Simulation Models. *Ind. Eng. Chem. Res.*, **2017**, *56* (23), pp 6728–6740, **DOI:** 10.1021/acs.iecr.7b01014. In this work, a new approach for the design of crude oil distillation systems using rigorous simulation models is proposed. The strategy for modelling the crude oil distillation unit is presented, followed by the optimisation problem and its solution procedure. Two examples are presented to demonstrate the application of this approach.

Chapter 4 presents Publication 2: Ibrahim, D., Jobson, M., Li, J., Guillén-Gosálbez, G., 2018. Optimization-based Design of Crude Oil Distillation Units using Surrogate Models and a Support Vector Machine. *Chem. Eng. Res. Des*., **2018**, **DOI:** doi.org/10.1016/j.cherd.2018.03.006. In this work, a new approach for the design of crude oil distillation systems based on surrogate models is proposed. The distillation column modelling method is presented first, followed by a framework for optimisation of the column structure and operating conditions. The application of the approach is illustrated using an example.

Chapter 5 presents Publication 3 and 4: Ibrahim, D., Jobson, M., Lie J., Guillén-Gosálbez, G., 2017. Optimal Design of Flexible Heat-Integrated Crude Oil Distillation Units using Surrogate Models. *Chem. Eng. Res. Des.* [To be submitted] and Ibrahim, D., Jobson, M., Guillén-Gosálbez, G., 2017. Design of Chemical Processes under Uncertainty Combining the Sample Average Approximation and the Analytic Hierarchy Process. *Comput. Chem. Eng.* [Submitted], respectively. In Publication 3, a new approach for the design of flexible crude oil distillation systems that process multiple crude oil feedstocks is proposed. The capabilities of the proposed method are illustrated using a case study. Publication 4 extends the approach proposed in Publication 3 to address design problems with a large number of operating scenarios that may be encountered during plant operation.

Chapter 6 highlights the contribution of the research work, discusses the limitations of the research and recommends some future work.

Chapter 2 Literature review

In the petroleum refining industry, crude oil distillation plays a key role in the overall production process due to its economic and environmental importance. As discussed in Chapter 1, the need for refineries to process different types of crude oil feedstocks and/or blend of crude oils in order to meet market demand for products and to maximise their profit margins, calls for greater flexibility in the design and operation of crude oil distillation systems. To achieve this, there is a need for systematic tools that can facilitate the design and optimisation of the crude oil distillation systems.

To develop such systematic tools, three major challenges need to be addressed. Firstly, a simulation model that can represent the complex behaviour of the crude oil distillation system is required. Such a model should not only be computationally efficient but should also be accurate and robust enough to guarantee convergence. Secondly, to account for heat integration during design and optimisation, a heat recovery model is required. Heat recovery can be accounted for using either pinch analysis or heat exchanger networks (Smith, 2005). Finally, an efficient optimisation framework that incorporates the distillation column model and heat recovery model is required, in order to facilitate the search for energy-efficient and cost-effective design alternatives within the design space.

Several works have been carried out on crude oil distillation system, for example, to develop new modelling strategies and to improve established design and optimisation methodologies. This chapter reviews the relevant work on modelling, design, and optimisation of crude oil distillation systems. First, a technical overview of the crude oil distillation system is presented. Second, existing modelling strategies for crude oil distillation units are presented in Section 2.2. Third, previous work on the design and optimisation of heat-integrated crude oil distillation systems is presented in Section 2.3. Sections 2.2 and 2.3 complements the literature reviews of the papers presented in

Chapters 3, 4 and 5. Lastly, methodologies for process design for flexibility and optimisation methods are presented in Sections 2.4 and 2.5 respectively.

2.1 Technology background – crude oil distillation

2.1.1 Crude oil and its properties

Crude oil mixture

Crude oil or petroleum is a complex mixture of 100 000s of hydrocarbons ranging from compounds with one carbon atom (methane) to those with more than twenty. These compounds include paraffinic hydrocarbons, naphthenic hydrocarbons and aromatic hydrocarbons (Jones, 1995). Paraffinic hydrocarbons are saturated compounds, such as ethane, propane, butane and other members of the homologues series. Naphthenic hydrocarbons are saturated cyclic compounds, such as cyclo-pentane, cyclo-hexane and so on. Lastly, aromatic hydrocarbons are unsaturated cyclic compounds such as benzene. In addition, crude oil contains small number of inorganic compounds (impurities) such as sulphur, oxygen, nitrogen and metals.

Although all varieties of crude oil contain similar compositions, the proportion of individual components in the crude oil mixtures differs, depending on the origin of the crude oil (Jones, 1995). In its original state, crude oil is highly viscous and has low flash point and therefore has limited values and application; refining is usually required to transform the limited value crude oil into a more valuable high-quality product that meets the specifications of the energy market.

Properties of crude oil mixture

The value of crude oil in the market is determined by two relevant properties, namely, sulphur content and API (American Institute of Petroleum) gravity (density). The amount of sulphur in crude oil is of paramount importance to a refinery, as it determines the cost of treatment that will be required during refining, which has a considerable impact on the refinery economics. Crude oil with less than 0.5 wt% sulphur content is termed 'sweet,' while that with higher values are termed sour (Gary

et al., 2007). The higher the sulphur content, the lower the value of the crude oil and vice versa. Figure 2.1 illustrates the sulphur content and API gravity of selected crude oils around the world.

The API gravity expresses the density of the crude oil at 60 °F ($°API = [141.5/specific\ gravity] - 131.5$). The majority of crude oils have API gravity in the range of 20 to 45 (Gary et al., 2007). In general, heavy crude oils are less valuable than lighter ones, as they produce less high value products, and require considerable processing cost.

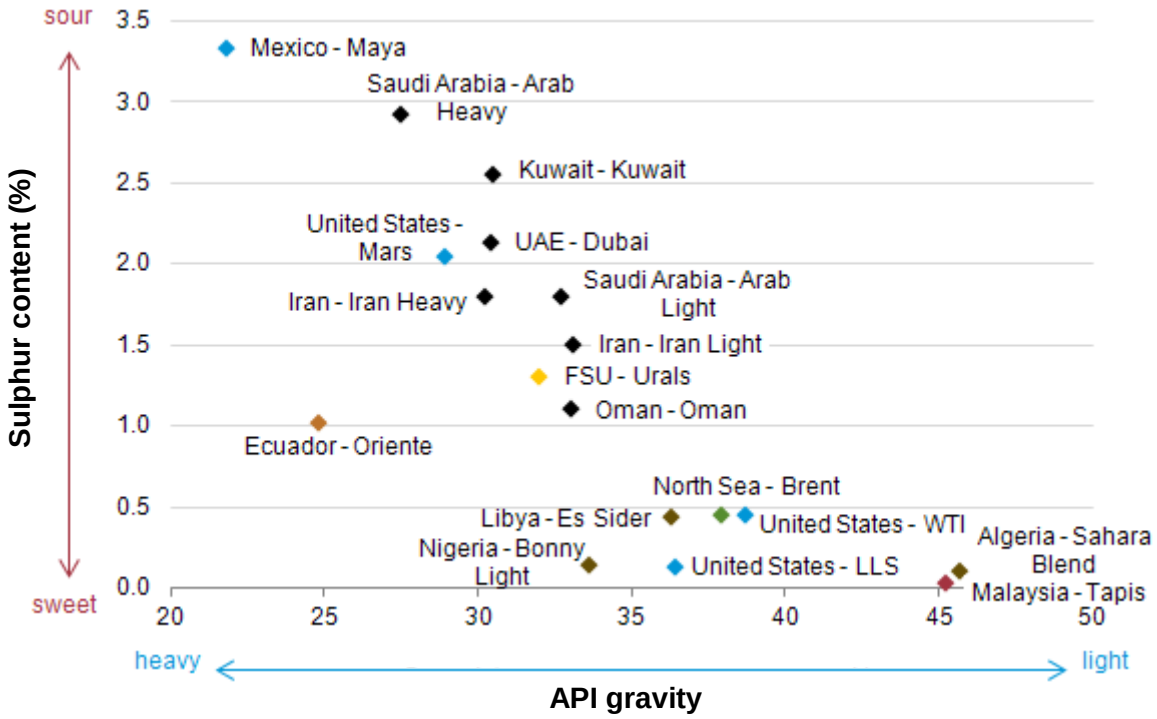


Figure 2.1 Density and sulphur content of selected crude oil (EIA, 2012)

Boiling range

Another important property of crude oil is the boiling range. Unlike sulphur content and API gravity, the boiling range of a crude oil reveals the amount of valuable products that can be recovered from the whole crude oil. The boiling range of a crude oil is typically determined experimentally using one of the following tests: true boiling point (TBP), ASTM D86 (American Society for Testing and Materials), and equilibrium flash vaporisation (EFV) (Watkins, 1979; Jones, 1995; Gary et al., 2007; Fahim et al.,

2009). Various correlations are available for inter-conversion between the distillation curves; the details of these correlations and their application are presented by Riazi (1989).

Other important properties of crude oil include salt content, pour point, carbon residue, nitrogen content, metal content, flash point, etc. This list is inexhaustible, as there are many properties used to characterise crude oil mixtures in the petroleum and refining industries.

Characterisation of crude oil mixture

The whole crude oil boiling point curve can be divided into several distinct segments. For example, during TBP distillation, the volume percent distilled can be collected over a narrow temperature range, as the temperature is increased gradually, several distinct fractions can be defined. Thus the whole crude oil is split into fractions. Each fraction still contains many unknown components. Nonetheless, the components have similar boiling point, which will be useful for calculating other properties and for the design of the separation system. For calculation convenience, each fraction can be treated as a pure component, which are commonly termed 'pseudo-components' (Fahim et al., 2009).

In engineering practice, it is not necessary to apply experimental procedures to generate the pseudo-components. Once the true boiling point curve for the whole crude oil is available, simple numerical calculations can be applied to divide the whole crude oil TBP curve into pseudo-components. There is no standard rule for selecting the number of pseudo-components to be used for a specific type of crude oil mixture, although a few sources suggested 30, 10 and 8 pseudo-components for boiling point ranges of 38–427 ℃, 427–649 ℃ and 649–899 ℃ respectively (Chang et al., 2012). In general, the higher the number of pseudo-components, the better the accuracy of the calculations, and the larger the computational effort required.

Over the years, industrial practitioners and petroleum engineers have developed several correlations that can be employed to divide the whole crude oil into pseudo-components, and subsequently determine the physical properties (density and volume), transport properties (viscosity, thermal conductivity, diffusivity, etc.) and thermodynamic properties (enthalpy, heat capacity, K-values, etc.) of each pseudo-component that can be used in design calculations (Fahim et al., 2009). Most of these correlations are embedded in commercial process simulation packages, e.g., Aspen HYSYS, Aspen Plus, UNISIM, and Pro II, and have been used to facilitate simulation, design, and optimisation of petroleum refining processes, including crude oil distillation systems.

2.1.2 Crude oil distillation products and separation specifications

As discussed in Chapter 1, crude oil distillation is required to perform the initial separation of crude oil into various intermediate products that are either blended into final products (e.g., gasoline, diesel, and kerosene) or sold as feedstocks to the chemical and petrochemical industries. Table 2.1 presents typical crude oil distillation intermediate products and their boiling range (specification).

Table 2.1 Recommended ASTM boiling ranges (in ℃) for products of atmospheric tower#

Product	A	B	C
Light naphtha	121 — 135*	121 — 135*	121 — 135*
Heavy naphtha	204*	163*	163*
Light distillate	191 — 316*	149 — 316*	149 — 288*
Heavy distillate	302 — 357*	302 — 357*	274 — 357*
Atmospheric gas oil	Determined by allowable oil temperature		

* End point i.e., the boiling temperature to vaporise the entire product

Watkins (1979)

A: Maximum naphtha operation

B: Maximum light distillate operation

C: Maximum heavy distillate operation

The amount of each distillation product can be estimated from the whole boiling point curve. To achieve this, the boiling point curve needs to be split into several segments that correspond to the boiling point range of the products. The temperature on the boiling point curve that represents the limit for each product is called the 'cut point,' and the volume percent corresponding to the cut point represents the amount of the product (see Figure 2.2).

For each cut (product), the temperature (T0%) at which the first component vaporises is called the 'initial boiling point,' while the vaporisation temperature (T100%) of the final component is called the 'endpoint.' These concepts are illustrated in Figure 2.2

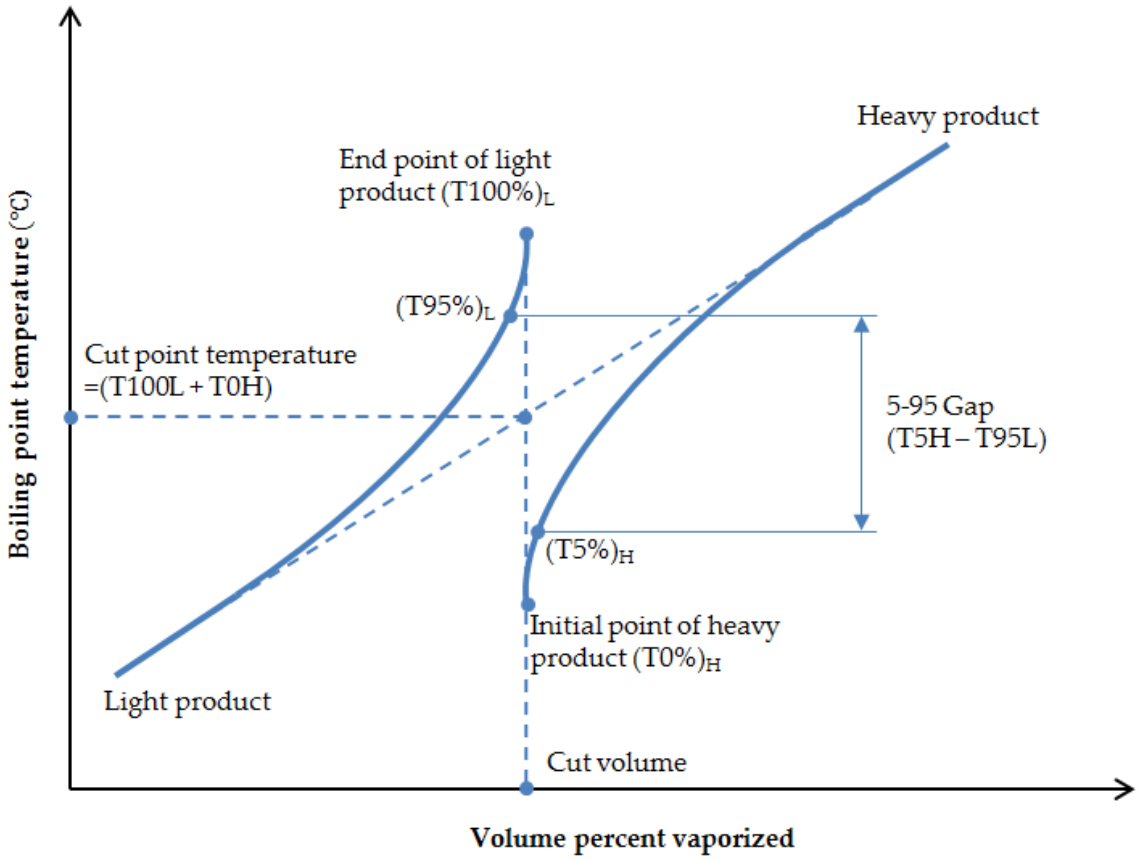


Figure 2.2 Cut point temperature between distillation products and 5-95 Gap (adapted from Watkins (1979))

Unlike conventional distillation in which almost pure products can be produced, in crude oil distillation, there is no sharp separation between adjacent products; thus,

components with boiling points lower than the cut point temperature, and components with boiling point higher than the cut point temperature are present in the product. This overlap has significant implications on the quality and boiling range of products. For example, the presence of components with boiling point below the cut point temperature, lowers the boiling point of the cut to below the required specification. To maintain the overlap within acceptable limits, product quality specifications are needed.

In crude oil distillation, two important terms are commonly used to specify the product quality and degree of separation, namely, ASTM boiling temperature and 5-95 gap (Watkins, 1979). ASTM boiling temperature is a key specification for most distillate products, and it defines the quality of the distillation product. The specification is commonly defined at T5% and T95% boiling points of the product, indicating the temperature at which 5% and 95% of the product will vaporise in the ASTM test. The 5-95 gap defines the degree of separation between two adjacent products. Quantitatively, it is the difference between T5% ASTM boiling temperature of a heavy product and T95% ASTM boiling temperature of an adjacent lighter product (see Figure 2.2). A positive difference indicates a gap, which is an indication of a good separation, while a negative difference indicates an overlap, which is an indication of a sloppy separation. Table 2.2 presents a typical ASTM 5-95 gaps between the indicated products.

Table 2.2 Separation criteria for atmospheric distillation products (Watkins, 1979)

Separation	5-95 Gap, ℃
Light naphtha — heavy naphtha	-6.67 to -1.11
Heavy naphtha — light distillate	-3.89 to 10
Light distillate — heavy distillate	-17.8 to -12.2
Heavy distillate — atmospheric gas oil	-17.8 to -12.2

2.1.3 Crude oil distillation system

Crude oil distillation is the first process in any petroleum refinery. Figure 2.3 shows a diagram of a typical crude oil distillation system. The system comprises a preheat train, a fired heater, and a crude oil distillation unit (consisting of a distillation column

equipped with side-strippers and pump-around loops). This section presents a detailed description of the individual components of the crude oil distillation unit.

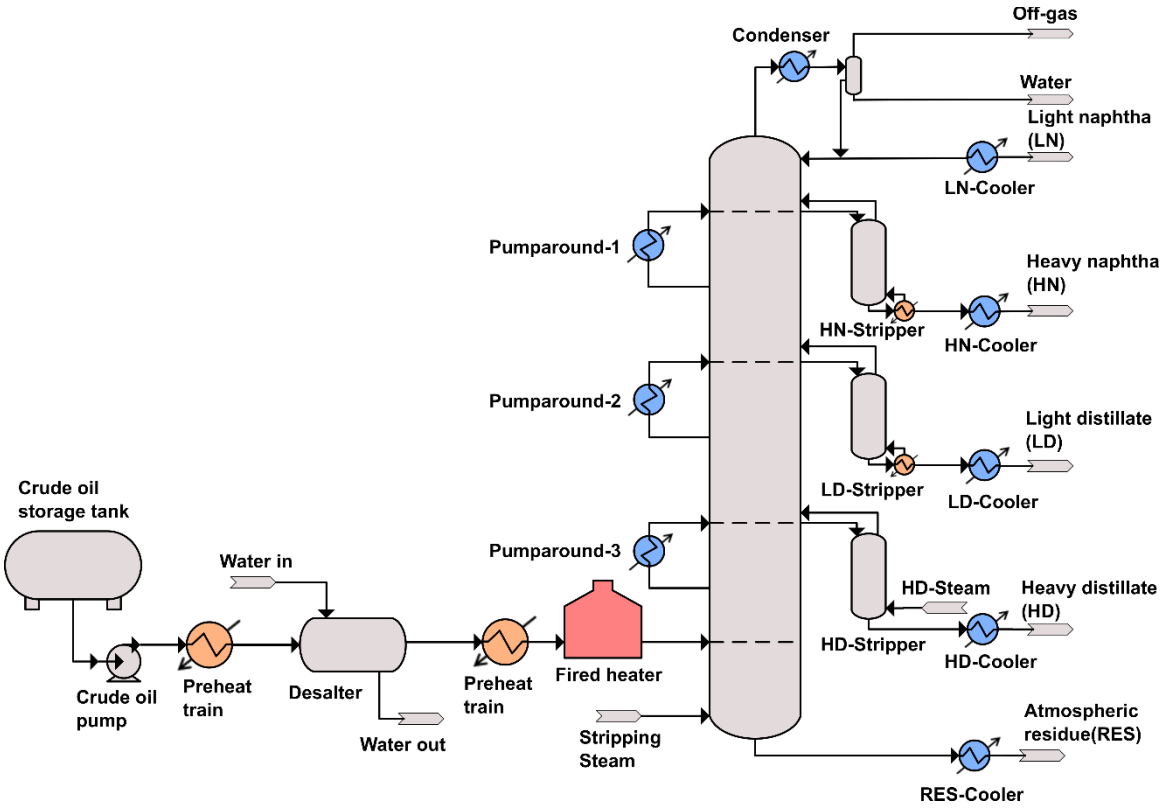


Figure 2.3 Conventional crude oil distillation system.

2.1.3.1 Pre-separation: preheat train, desalter and fired heater

Figure 2.3 shows the configuration of a typical refinery crude oil distillation system. Raw crude oil, usually at ambient temperature is pumped from storage tanks and preheated in two stages. Firstly, the crude oil is partially heated in the first part of the Preheat Train and fed to a desalter which removes dissolved or suspended salts from the crude oil feed (Gary et al., 2007). The inlet temperature of the desalter is approximately 160 ℃. The desalted crude oil is further heated in the second part of the Preheat Train. The heat used in the Preheat Train is mainly heat recovered from the product coolers, pump-around loops, and column condenser. The outlet temperature of the Preheat Train depends on the degree of heat recovery. The outlet temperature is typically between 270 ℃ and 290 ℃ (Gary et al., 2007).

Before the crude oil is fed to the crude oil distillation column, it undergoes further heating in a fired heater, also known as furnace. The outlet temperature of the furnace ranges between 350℃ and 370℃. The temperature is set by the maximum allowable temperature of the crude oil mixture; and the temperature should be sufficient to vaporise the entire distillate products to be recovered in the column, plus an extra 2 to 5 % vaporisation, called 'over flash' (Watkins, 1979); outlet temperatures that would cause thermal cracking of the crude oil mixture should be avoided.

2.1.3.2 The main distillation column

The heated, partially vaporised crude oil is fed to the flash zone of the atmospheric fractionation column a few trays above the bottom stage. Stripping steam is also supplied to the column at the bottom stage, which partly suppresses the boiling point of the crude oil and also, causes further vaporisation. The crude oil mixture is then recovered into various fractions, in this case, light naphtha (LN), heavy naphtha (HN), light distillate (LD) and heavy distillate (HD). The residue from the atmospheric column contains valuable hydrocarbons that are recovered under reduced pressure in a vacuum distillation column (not shown in Figure 2.3).

Typically, the atmospheric distillation column contains between 30 to 50 trays, depending on the type of product produced and the degree of separation required (Gary et al., 2007). Apart from the distillation top product, all products other than the residue are withdrawn from the column at intermediate trays (side-draws).

2.1.3.3 The side-strippers

To maintain satisfactory boiling range of products, side strippers are employed to recover light components from side-draws. Heat is supplied to the side-stripper in two ways: direct heating using live steam and indirect heating via a reboiler. In both cases, the stripped light material vapour is returned to the main column. In most column arrangements, side-strippers contain between 3 and 8 trays (Gary et al., 2007).

2.1.3.4 The pump-around loops

In a pump-around loop, hot liquid is drawn from the column, cooled in a heat exchanger by exchanging heat with a colder stream, and returned to the column two to three trays above the draw point. Pump-arounds are attached to the atmospheric distillation column for three main reasons: (i) to enhance separation efficiency by providing internal reflux; (ii) to increase the energy efficiency of the separation system by creating heat recovery opportunities; (iii) to reduce the diameter at the top of column by controlling the vapour and liquid traffic at various sections of the complex column. Heat recovered from the pump-around loops, together with other available heat sources (column condenser and product coolers) is used to preheat the raw crude oil, thus reducing fuel consumption in the furnace and most importantly the operating cost.

To summarise, crude oil, as well as distillation products, are complex mixtures of hydrocarbons, containing a small amount of non-hydrocarbon molecules. Both crude oil and distillation products are mainly described using boiling point curves. To facilitate modelling, design and optimisation of crude oil distillation systems, several useful correlations have been developed to aid characterisation of crude oil mixtures into pseudo-components.

Crude oil distillation system separates the complex crude oil mixture into valuable products. The system has a highly sophisticated configuration, consisting of side-strippers, pump-arounds, a fired heater and a column condenser. Accurate, robust and computationally efficient models are required for the design and optimisation of the system. These models are discussed in Section 2.2; Section 2.3 presents available methods for the design of crude oil distillation systems.

2.2 Crude oil distillation column models

In general, crude oil distillation column models can be broadly classified into three categories, namely, simplified, rigorous and surrogate models. In this thesis, simplified models are those models based on modified Fenske–Underwood–Gilliland (shortcut

models) design equations; rigorous models refer to models based on first principles (material and energy balance, equilibrium equations etc.), and surrogate models are data-driven models such as artificial neural network, polynomials etc. These models can be used in various applications including grassroots design, retrofit and operational optimisation (involve modifications to only operating conditions). The model used in each application requires a unique set of degrees of freedom. For example, models used for operational optimisation include only the continuous variables (i.e., operating conditions) of the crude oil distillation unit, such as feed inlet temperature, pump-around duties, and temperature drops, stripping steam flow rates and reflux ratio. Conversely, for a grassroots or retrofit design problem, the model should include discrete variables (representing column structure) in addition to continuous variables, since the model is required to simulate alternative distillation column structures and their operating conditions during optimisation. The discrete variables can include the locations of the feed tray and of pump-around and side-stripper draw streams, and the number of trays in each section of the column. This section discusses the three categories of models available for simulating crude oil distillation units, their limitations, and scope of application.

2.2.1 Simplified models

Simplified models constitute another alternative for the modelling of distillation units. In the context of crude oil distillation, these models are based on the column decomposition strategy of Liebmann and co-workers (Liebmann, 1996; Liebmann et al., 1998), in which the crude distillation unit is treated as a sequence of thermally coupled columns. Modified Fenske–Underwood–Gilliland design equations for simple columns are then applied for each column section, while the separation is specified in terms of purity or recovery of light key and heavy key components. These models can predict the minimum reflux ratios and number of trays in each column section corresponding to specified reflux ratios. Nevertheless, the predictions from simplified models are rather poor compared with rigorous model. Consequently, simplified models are not commonly applied in practice.

2.2.2 Rigorous models

Rigorous distillation models are based on the so-called MESH (Material–Equilibrium–Summation–Heat) equations which represent relevant phenomena for each stage within the column. These models can be reasonably accurate, as they are based on the first principles governing the separation process (i.e., stage-by-stage material and energy balance), although they assume phase equilibrium is achieved on each stage, no chemical reactions occur, and no entrainment of liquid drops in vapour and vapour bubbles in liquid (Seader et al., 2010). With these models, it is possible to estimate the temperature and pressure profiles within the column, in addition to stream flow rates and compositions. Figure 2.4 illustrate a simple equilibrium stage.

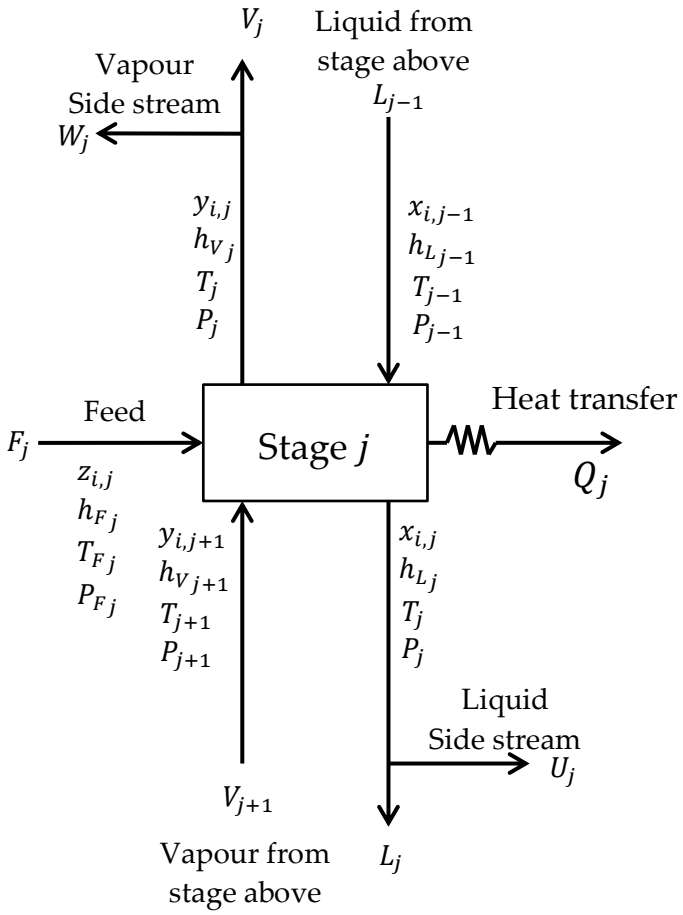


Figure 2.4 General equilibrium stage (adapted from Seader et al. (2010))

For the simple equilibrium stage in Figure 2.4, the MESH equations representing Material balance, Equilibrium equations, Summation equations and Heat balance can be written as follows (Seader et al., 2010):

Material balance for each component

$$L_{j-1}x_{i,j-1} + V_{j+1}y_{i,j+1} + F_j z_{i,j} - \left(L_j + U_j\right)x_{i,j} - \left(V_j + W_j\right)y_{i,j} = 0 \quad (2.1)$$

Phase-Equilibrium relation for each component

$$y_{i,j} - K_{i,j}x_{i,j} = 0 \quad (2.2)$$

Mole-fraction Summations (one per stage)

$$\sum_{i=1}^{C} y_{i,j} - 1.0 = 0 \quad (2.3)$$

$$\sum_{i=1}^{C} x_{i,j} - 1.0 = 0 \quad (2.4)$$

H equation, denoting energy balance (one per stage)

$$L_{j-1}h_{L_{j-1}} + V_{j+1}h_{V_{j+1}} + F_j h_{F_j} - \left(L_j + U_j\right)h_{L_j} - \left(V_j + W_j\right)h_{V_j} - Q_j = 0 \quad (2.5)$$

where indices i and j denote component and stage number respectively. F is the feed molar flow rate; V and L are vapour and liquid molar flow rates; U and W are liquid and vapour side streams; Q represent heat transfer to or from stage j; z denotes the component molar fraction in the feed stream; x and y are the liquid and vapour

component mole fractions; h and K denotes enthalpy and equilibrium constant respectively.

The outlined MESH equations are typically defined for each equilibrium stage. In a distillation column comprising C components and N number of stages, the number of MESH equations to be solved is equal to $N(2C + 3)$ (Seader et al., 2010). Due to the strong interactions and highly non-linearity of these equations, specially-tailored algorithms are required to generate feasible solutions. The solution algorithms include bubble point (BP) method for systems with narrow boiling point components, sum-rate (SR) method for systems with wide boiling point components and Newton-Raphson for systems with intermediate boiling case (Seader et al., 2010). Due to the iterative nature of the solution methods, the column simulation is computationally demanding. The MESH equations and their corresponding solution methods are implemented in several commercial process simulators (such as Aspen HYSYS, Aspen Plus, UNISIM and PRO/II), and have been used by several researchers and industrial practitioners to design (Liebmann, 1996; Liebmann et al., 1998; Bagajewicz and Ji, 2001), analysis (Errico et al., 2009; Benali et al., 2012; Waheed et al., 2014) and optimise (Basak et al., 2002; Al-Mayyahi et al., 2011; Gu et al., 2015) the crude oil distillation columns.

Liebmann (1996) presents a stepwise procedure for the design of crude oil distillation column that combines rigorous column simulation and pinch analysis. In each step, a grand composite curve is constructed using stream data generated from a rigorous simulation. The grand composite curve is used to facilitate the search for column structure and operating conditions that lead to significant savings in energy consumption.

Bagajewicz and Ji (2001) extends the work of Liebmann (1996) and also introduces the concept of heat demand-supply diagram. Firstly, a rigorous simulation of a column with no pump-arounds is setup. Then, a heat demand-supply diagram is applied to identify suitable location of the pump-arounds while taking into account the effect of stripping steam on the maximum heat recovery of a crude oil distillation system.

Rigorous model have been implemented in frameworks to facilitate optimisation of an existing crude oil distillation unit. For example, Bagajewicz (1998) combined pinch analysis and a rigorous column model into a framework to aid the search for column operating conditions that minimise utility requirements. The rigorous column model predicts the product quality and flow rate, column temperature profile and stream enthalpy change. Operating variables optimised include pump-around duties and return temperatures, over–flash ratio and stripping steam flow rates.

Basak et al. (2002) develop an approach for online optimisation of crude oil distillation units using a rigorous model. The model parameters, such as stage efficiencies are tuned to minimise the discrepancy between the measured plant data and model prediction. A gradient-based optimisation method is applied to search for the best combination of operating variables (steam and pump-around flow rates, reflux ratio and feed temperature) that maximise profit.

Inamdar et al. (2004) and Al-Mayyahi et al. (2011) develop frameworks for optimisation of crude oil distillation unit, taking into account multiple objectives. Both works applied the elitist non-dominated sorting genetic algorithm (NSGA-II). Inamdar et al. (2004) focused on maximising profit and minimising the cost of energy by varying product flow rates, pump-around flow rate, reflux ratio, and feed temperature, while Al-Mayyahi et al. (2011) focused on maximising profit and minimising CO_2 emissions. The optimisation variables considered in the work of Al-Mayyahi et al. (2011) include steam flow rate, feed temperature and flow rate, pump-around duty and reflux ratio.

Ali et al. (2013) applied rigorous models to optimise the net profit of an existing crude oil distillation unit. First, the crude oil distillation unit is modelled in Aspen HYSYS software. Then, the NLP solver embedded in the commercial software is used to select the operating variables (pump-around flow rates, feed temperature, bottom steam flow rate and product flow rates) that lead to maximum profit.

Rigorous models consist of complex nonlinear equations derived based on first principles. These models are reasonably accurate, although their solution methods are computationally demanding. Several works have applied rigorous models, for example, to design (Liebmann, 1996; Bagajewicz and Ji, 2001) and to optimise (Basak et al., 2002; Inamdar et al., 2004; Ali et al., 2013) the crude oil distillation unit. Only the work of Liebmann (1996) and Bagajewicz and Ji (2001) have accounted for heat recovery. None of the methodologies presented have incorporated column structural degrees of freedom (e.g., number of trays in column sections, feed tray location, pump-around, and side-stripper locations) as a design variable. Thus, these methodologies cannot be directly applied to perform optimisation-based design of the crude oil distillation unit.

2.2.3 Surrogate models

Surrogate models, also known as data-driven models, statistical models or meta-models are compact, scalable mathematical models that describe the relationship between specific inputs (e.g., feed temperature) and outputs (e.g., product quality) of complex systems. Surrogate models are less computationally demanding than rigorous simulation models. Thus they are suitable for implementation in an optimisation framework, and for sensitivity analysis. Various forms of surrogate models (e.g., linear, polynomial, artificial neural network, etc.) may require different sample size during training (fitting/ regression) in order to achieve a desired accuracy (Nuchitprasittichai and Cremaschi, 2012; Quirante et al., 2015). In general, a large sample can improve the accuracy of the model. However, the sampling and model fitting time may be increased significantly (Nuchitprasittichai and Cremaschi, 2012).

Surrogate modelling of chemical processes comprises three main steps, namely, sampling (also known as data generation), model selection, and model fitting (Biegler et al., 2014).

Prior to sampling, it is necessary to select the desired inputs and outputs of the system of interest. The inputs and outputs to be used depend on the scope of the model; for

example, a model used for design purposes should take into account structural and operational variables of the system. Typically, the input and output variables are selected based on experience and knowledge of the system. In most cases, it is desirable to select variables that have a significant impact on the system performance. Such variables can be identified via a sensitivity analysis.

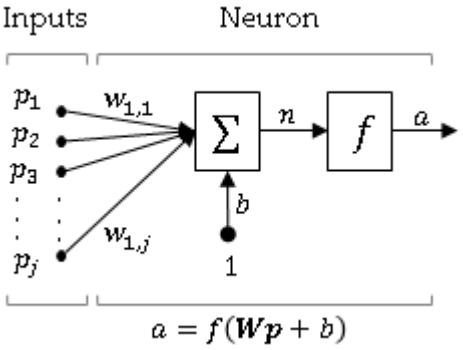
Once the variables are selected, a data set can be generated. The data set is typically created in two ways: real plant measurements and/or experimental data, and multiple rigorous simulations. In multiple rigorous simulations, statistical technique is applied to generate data points from models (Henao and Maravelias, 2010; Gueddar and Dua, 2011; Nuchitprasittichai and Cremaschi, 2013; Biegler et al., 2014; Boukouvala et al., 2015; Quirante and Caballero, 2016), while plant measurements are often used to build surrogate models of an existing process. These models are mostly used to facilitate operational optimisation (Ochoa-Estopier and Jobson, 2015a; Osuolale and Zhang, 2017), process troubleshooting (Mouli et al., 2016; Yang and Hou, 2016), and process control (Osuolale and Zhang, 2015; Xie et al., 2015)

In the next step, the form of the model to represent the data is selected. Different forms of model can be used for data fitting, for example, polynomial (Heiberger and Neuwirth, 2009), artificial neural networks (Beale et al., 2015), support vector regressions (Vapnik, 1995), etc. The form of model to be used for regression is crucial and should be carefully selected. For example, linear models are suitable for input–output data set with strong linear relationship. Lastly, an optimisation algorithm (Floudas, 1995; Biegler et al., 1997; Edgar et al., 2001) is applied to fit the model to the data set by minimising the error between the model predictions and the original data. After the model is built, several statistical tests need to be performed on the model to test its validity. Among the surrogate modelling techniques, artificial neural network leads to accurate and robust models that are easier to implement in an optimisation framework (Henao and Maravelias, 2010; Nuchitprasittichai and Cremaschi, 2012); thus this work makes use of artificial neural network to model the complex crude oil distillation unit.

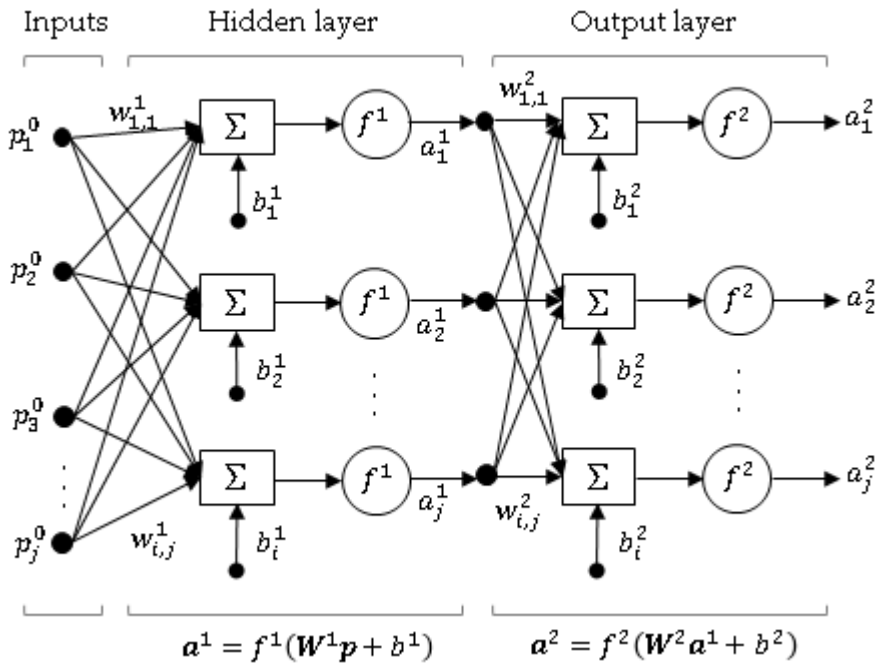
2.2.3.1 Artificial neural networks

Artificial neural networks are computational modelling tools applied to approximate complex non-linear systems and to classify data set. Artificial neural network have been successfully applied in various field of research, including system identification (Prasad and Bequette, 2003; Aguado et al., 2009), models reduction (Gueddar and Dua, 2011; Xie et al., 2015), fault detection (Kankar et al., 2011; Ben Ali et al., 2015; Yang and Hou, 2016), process troubleshooting (Mouli et al., 2016; Yang and Hou, 2016), operational optimisation (Liau et al., 2004; Ochoa-Estopier et al., 2012), system design (Henao and Maravelias, 2010; Fahmi and Cremaschi, 2012), and property prediction (Hussain, 1999; Gharagheizi et al., 2011; Afrand et al., 2016). The wide application of artificial neural network is attributed to their ability to capture complex non-linear relationships between input-output data, especially when the relation among the system variables is unknown (Dua, 2010).

Artificial neural network architecture can be broadly classify into feedforward, recurrent or feedback, and mesh (Silva et al., 2017). Feedforward network is the most widely acceptable architecture due to its mathematical simplicity and ease of implementation within an optimisation algorithms (Nuchitprasittichai and Cremaschi, 2012). In a feedforward network, information is processed in the forward direction only. Figure 2.5b shows a typical multi-layer feedforward network with three layers: input, hidden and output layers, connected via neurons.



(a)

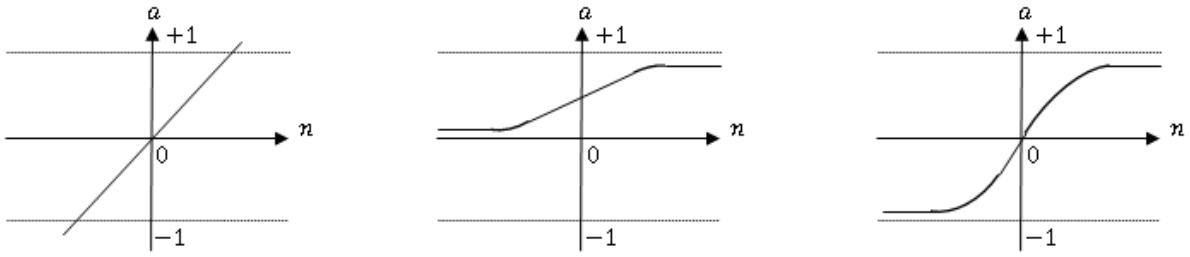


(b)

Figure 2.5 Artificial neural network (a) network neuron; (b) multi-layer feedforward neural network (Adapted from (Beale et al., 2015)).

In general, neurons are the fundamental building block of any artificial neural network. A neuron, as shown in Figure 2.5a, consists of two main components: the summation point, $\sum$, and the transfer function, f, (Basheer and Hajmeer, 2000; Himmelblau, 2008; Beale et al., 2015). The summation point adds the product of all inputs, $\boldsymbol{p}$, and their corresponding weights, $\boldsymbol{W}$, and bias, b, to produce a net scalar input, n, while the transfer function, f, takes the net input and produce a scalar output, a.

Various forms of transfer function used in building artificial neural network are available, such as linear, log-sigmoid, tan-sigmoid etc. Figure 2.6 shows the schematic of the most commonly used transfer function.



(a) (b) (c)

Figure 2.6 Transfer functions applied in artificial neural network architecture (a) linear, (b) log-sigmoid, and (c) tan-sigmoid (Adapted from (Beale et al., 2015)).

Linear transfer function takes a weighted input and transform, linearly, to an output between $-\infty$ and $+\infty$; log-sigmoid and tan-sigmoid transform weighted input to a range of 0 to 1, and -1 to $+1$, respectively. Log-sigmoid and tan-sigmoid are commonly used in the hidden layer, while linear transfer function is applied in output layer. In this work, log-sigmoid and linear functions are used in the hidden and output layers respectively.

Before applying the neural network, the weights and biases of the network (see Figure 2.5b) needs to be tuned such that the artificial neural network mimicked the behaviour of the input-out data. The process is called 'training' and it is carried out using optimisation method such as Levenberg-Marquardt, Bayesian regularization, BFGS Quasi-Newton, scaled conjugate gradient etc. (Beale et al., 2015). The fastest training algorithm is Levenberg-Marquardt, and it is the one used in this work. The objective of 'training' is to minimise a cost function, in this case the mean square error (see Eq. 2.6) between the network predictions and the input-output data by adjusting the weights and biases. To facilitate training and also enhance the performance of the built network, the input-output data set should be scaled between -1 and $+1$ (Beale et al., 2015).

$$mse = \frac{1}{N}\sum_{i=1}^{N}(t_i - a_i)^2 \tag{2.6}$$

where *mse* is the mean square error, *t* and *a* denote the target and predicted output respectively, *N* is the total number of sample. Section 2.2.3.2 presents the application of surrogate models in crude oil distillation, including the use of artificial neural networks.

2.2.3.2 Modelling of crude oil distillation unit based on surrogate models

Surrogate models have been used by several authors to simulate the crude oil distillation system. Liau et al. (2004) and Motlaghi et al. (2008) develop an artificial neural network (ANN) model of a crude oil distillation column using data from existing plants. In the work of Liau et al. (2004), the distillation model inputs include crude oil properties, feed temperature, product flow rates while the outputs include product quality. The model is optimised using successive quadratic programming to determine the operating conditions that improve product yield. Although the built model is accurate, several operating variables such as stripping steam and pump-around duties and temperature drops are not considered. Similarly, the artificial neural network model develop by Motlaghi et al. (2008) include crude oil properties and operating variables as inputs, while the outputs are product quality and their flow rates. A genetic algorithm is used to optimise the flow rate of products according to their market values.

Yao and Chu (2012) developed a surrogate model of the crude oil distillation using the concept of support vector regressions. The model is regressed using data generated via multiple rigorous simulations (in Aspen Plus). The surrogate model is implemented in a framework to optimise profit by varying operating variables. The variables optimised include feed temperature, reflux ratio, product flow rates, pump-around temperature drops and flow rates, and steam flow rates.

López C. et al. (2013) formulate a framework to optimise the operational variables of crude oil distillation system that processes crude oil blends. The distillation system comprises of three atmospheric columns and two vacuum columns and preheats trains. Meta-models based on second order polynomial functions are used to model the crude oil distillation units. The models are regressed using samples obtained from multiple rigorous simulations. The built models together with an energy balance representing the heat exchanger network are implemented in a framework to maximise net profit.

Ochoa-Estopier and Jobson (2015a) formulate an approach for operational optimisation of crude oil distillation systems using surrogate models. Several artificial neural networks are regressed against data generated via multiple rigorous simulations (in Aspen HYSYS). The heat exchanger network model of Rodriguez (2005) is adopted to represent the crude oil preheat train. The dependence of thermal properties on temperature in process streams is modelled using a combination of linear (for sensible heat) and third order polynomial (for phase change) correlations. A stochastic optimisation method based on simulated annealing is applied to optimise net profit. The variables considered for optimisation include operating conditions such as pump-around temperature drops and duties, feed inlet temperature, steam flow rate and product flow rates.

Recently, Osuolale and Zhang (2017) modelled a crude oil distillation system, comprising a prefractionator, atmospheric column, and vacuum column, using bootstrap aggregate artificial neural networks. In this strategy, several artificial neural networks are constructed for each distillation unit. The predictions (i.e., outputs) from all the artificial neural networks are aggregated and used as the network output. In this way, the accuracy and reliability of the artificial neural networks could be improved. The models are combined with an optimisation algorithm (successive quadratic programming) to optimise profit objective. Decision variables include flow rates of products and steam, pump-around temperature drop and duties, while constraints are imposed on product quality (in terms of T5% and T95% boiling temperatures).

Various regression techniques have been applied to construct surrogate models for crude oil distillation units. A few modelling approaches (Liau et al., 2004; Motlaghi et al., 2008) have applied real plant data to regress the parameters of the surrogate model. Therefore, the valid range of application for these models is restricted to previously known scenarios. Again, since sampling of real plant data is usually associated with measurement error, the accuracy of the built model could be compromised. Other approaches (Yao and Chu, 2012; López C. et al., 2013; Ochoa-Estopier and Jobson, 2015a; Osuolale and Zhang, 2017) applied data generated via multiple rigorous simulations to regress the parameters of the surrogate model. Thus the model can explore scenarios other than those encountered during previous operation. In this way, the chances of obtaining a better solution could be enhanced. Among the optimisation approaches, only the work of López C. et al. (2013) and Ochoa-Estopier and Jobson (2015a, 2015b) have adequately accounted for heat integration within the system, thus increasing the likelihood that the solution obtained can be valid in practice, and also the energy efficiency of the system is improved. Furthermore, Osuolale and Zhang (2017) have also improved the energy efficiency of the system by minimising exergy losses. None of the surrogate modelling approaches presented here have incorporated structural degrees of freedom (e.g., number of trays in column sections, feed tray location, pump-around, and side-stripper locations) as a design variable, hence the methodologies cannot be applied to perform optimisation-based design of the crude oil distillation unit.

2.3 Design of heat-integrated crude oil distillation systems

2.3.1 Design for single crude oil feedstock

Various methods are available for the design of conventional crude oil distillation systems. Some of these methods are carried out based on heuristic rules, experience, empirical correlations and simple calculations. For example, Nelson (1958) describes a design method for crude oil distillation columns. In this approach, the number of trays in each section of the column and stripping steam required are estimated based on empirical correlations that are constructed from previously established designs. Later,

Watkins (1979) describes a procedure for the design of atmospheric and vacuum distillation columns. The design procedure is guided by heuristic rules. In the approach of Watkins (1979), the required number of trays in each section of a column is selected from a predetermined range, and stripping steam required is estimated based on product flow rates.

The methods presented by Nelson (1958) and Watkins (1979) formed the basis for subsequent design methodologies of crude oil distillation columns. However, the design methodologies of Nelson (1958) and Watkins (1979) require trial and error in their calculations. Furthermore, the heat exchanger network is designed after the design of the crude oil distillation column is completed. Thus, the interactions between the distillation columns and the associated heat recovery systems are not taken into account.

To design an energy-efficient distillation column, the interactions between the column and the heat recovery network need to be taken into account. Several researchers have focused on developing an integrated design of the crude oil distillation column and the associated heat recovery network. For example, Liebmann (1996) develops a methodology for the design of crude oil distillation columns using rigorous column models and pinch analysis. In this design method, an indirect sequence of simple columns with no thermal coupling is first initialised by decomposing the conventional atmospheric distillation column. The advantage of the decomposed columns is that it allows the evaluation of feed tray location and number of trays in each section of the column to be carried out based on product specification and feed composition. The decomposed column is simulated in Aspen Plus, and a grand composite curve is constructed using the process stream data of the crude atmospheric distillation column. The grand composite curve is used to suggest column modifications that enhance separation and improve the potential for energy savings. The main strength of the method of Liebmann (1996) is that the interactions between the atmospheric distillation column and heat recovery network are taken into account. However, the heat exchanger network design is not considered by Liebmann (1996), and the distillation column is not optimised.

In a related work, Sharma et al. (1999) develop a design approach for crude oil distillation unit that uses the concept of a column grand composite curve (Dhole and Buckingham, 1994). First, the column grand composite curve is constructed using temperature–enthalpy data obtained from simple energy balance across several sections of the complex column. The column grand composite curve is then used to identify the maximum amount of energy that can be recovered without affecting the separation. This approach neglects the effect of stripping steam on separation and heat recovery network design has not been taken into account. Furthermore, the number of trays in each column section is fixed, and the distillation column is not optimised.

To optimise the system, simplified models have been implemented in an optimisation framework together with heat recovery models in order to design the crude oil distillation unit (see Chapter 1). The work of Suphanit (1999) accounts for heat recovery using pinch analysis, while Rastogi (2006) and Chen (2008) applied heat exchanger network. The simplified model allows optimisation of structural and operational degrees of freedom. Even though simplified models are relatively easy to handle numerically, their inability to produce an accurate estimate of the distillation column performance makes them less useful in real application as compared with other models (e.g., rigorous models and shortcut models).

Overall, the discussion presented above focused on the design of crude oil distillation systems that process one type of crude oil feedstock. In general, these design methodologies may lead to a crude oil distillation column that performs well for a particular crude oil feedstocks or operating scenario, and perform poorly for other crude oil feedstock and/or operating scenarios, particularly in future scenarios where the column is to be used to process crude oil types other than the one considered during the column design.

2.3.2 Design for multiple crude oil feedstocks

To accommodate changes of crude oil feedstock doing operation, the crude oil distillation system should be designed to work well over a range of crude oil

feedstocks. Along this line, Bagajewicz and Ji (2001) extended the design approach of Liebmann (1996) to propose a new method for the design of crude oil distillation units that process light, medium and heavy crude oil feedstocks. The heat demand-supply diagram is used instead of the grand composite curve. The design procedure begins with an initial column design with no pump-around loops constructed based on the design approach of Watkins (1979). The column is then simulated with the lightest crude oil to be processed and the process stream data of the column are used to construct the heat demand-supply diagram. Based on the heat demand-supply diagram constructed for the column, heat load is transferred from the column condenser to the top pump-around. A similar step is carried out to distribute condenser heat load to subsequent pump-arounds located between product draws. In each step, the product specifications are maintained by adjusting the flow rate of stripping steam in side-strippers. The stripping steam flowrate is increased if the boiling temperature gap between adjacent fractions becomes smaller than the appropriate value. The transfer of heat load continues as long as the cost of energy saved can offset the cost of steam added. The operating conditions of the medium and heavy crudes are also determined based on the outlined procedure. Although the methodology determines the operating conditions of the different types of crude oils (light, medium and heavy) to be processed, the trade-off between capital and energy cost is not taken into account; pump-around location is selected based on heuristic rules; number of trays in column sections are fixed and the distillation column is not optimised.

To optimise the distillation column, More et al. (2010) set up a framework to study the effect of binary feed selection on grassroots design of crude oil distillation system. First, the crude distillation systems, i.e., pre-flash unit, atmospheric column and vacuum column, are modelled in a process simulator (Aspen Plus). Then, light, medium and heavy crudes and their binary mixtures (ratio of 10:90) are used to set up different operating scenarios. The optimisation tool embedded in the commercial simulator is used to optimise the distillation column operating variables (feed flow rate stripping steam, and product flow rate) for each operating scenario. One major

drawback of this analysis is that heat integration is not considered. Also, the approach neglects essential design variables such as number of trays in each column section, and other operating variables such as pump-around temperature drops and duties, and feed inlet temperature.

A critical observation of the research literature presented in this section indicates that a systematic methodology for the design of flexible heat-integrated crude oil distillation system is lacking. A systematic design approach should simultaneously consider the selection of the distillation column structural and operating variables, heat integration and various crude oils to be processed in a unified framework; this can be carried out using sophisticated optimisation techniques for process design for flexibility.

2.4 Optimisation methods

Optimisation is a quantitative mathematical tool that facilitates best selection from a set of many alternatives. To identify the best alternative, a quantitative measure of goodness called the objective function is required (Biegler et al., 1997). In process synthesis and design, common objectives include maximising net profit, product yield, and net present value, or minimising total annualised cost, energy cost, utility consumption and CO_2 emissions.

The value of the objective function is calculated from the problem variables. The variables are classified into dependent and independent variables. Independent variables, also known as decision variables, manipulated variables or degrees of freedom, refer to those system variables that can be adjusted to improve the objective function value. In a chemical process, the independent variables can include temperatures, pressure, feed flow rate, etc. Dependent variables, also known as process outputs are variables that determine the system performance. Examples of dependent variables include column diameter, reactor volume, product flow rates, etc. To ensure the solution from an optimisation is valid, constraints are generally imposed to define the design space. Process constraints comprise inequality constraints (e.g., product

purity, column hydraulic limit) and equality constraints that define the physical system (e.g., material and energy balance, phase equilibrium)

The objective function, problem variables, and process constraints together form the optimisation problem. Optimisation problem encountered in process synthesis and design can be classified as linear programming (LP), mixed integer linear programming (MILP), nonlinear programming (NLP) and mixed integer nonlinear programming (MINLP) (Floudas, 1995; Biegler et al., 1997; Edgar et al., 2000). The primary methods for solving these problems can be group into deterministic and stochastic search methods.

2.5.1 Deterministic optimisation methods

Deterministic methods, also known as rigorous optimisation, are gradient-based approaches that rely on derivatives of the functions (objective function and constraints) to guide the search for the best solution. Several optimisation techniques (such as successive linear programming, successive quadratic programming, generalized reduced gradient, etc.) that apply deterministic method are implemented in commercial software such as GAMS, MatLab, and Excel , and can be employed to optimise different type systems.

One advantage of these methods lies in their ability to guarantee local optimality (Floudas, 1995; Edgar et al., 2001). These methods are versatile and may require few function evaluations to converge to the optimal solution. Deterministic methods that guarantee global optimality are available, e.g. BARON (Tawarmalani and Sahinidis, 2005) and ANTIGONE (Misener and Floudas, 2014), although they may require large storage capacity and considerable computational time.

Deterministic optimisation method have been applied extensively in many fields of research, including process system engineering. For example in biorefining (Corbetta et al., 2016; Pérez Rivero et al., 2016), thermally coupled distillation columns (Caballero and Grossmann, 2001; Caballero, 2015), heat exchanger networks (Papalexandri and

Pistikopoulos, 1994; Li et al., 2015; Isafiade and Short, 2016). A comprehensive review of various application of deterministic optimisation techniques in process system engineering can be found in the work by Grossmann et al. (2000). In the context of crude oil distillation, Bagajewicz (1998) formulated a nonlinear programming problem to improve the energy efficiency of crude oil distillation unit. Successive quadratic programming is applied to search for an optimal set of operating conditions that improves the objective function value. Basak et al. (2002) and More et al. (2010) applied successive quadratic programming to select the best combination of operating variables (such as steam and pump-around flow rates, reflux ratio and feed temperature), to improve net profit of the system. López C. et al. (2013) and Osuolale and Zhang (2017) set up frameworks to optimise the crude oil distillation system using surrogate models. López C. et al. (2013) applied a generalised reduced gradient to search for operating conditions that lead to maximum profit, while successive quadratic programming is applied in the work of Osuolale and Zhang (2017).

Deterministic methods are suitable for problems that are continuously differentiable (Floudas, 1995; Edgar et al., 2001). However, many real-life optimisation problems are highly nonlinear, non-convex and non-differentiable. Thus there is a need for alternative method, for example methods based on stochastic optimisation.

2.5.2 Stochastic optimisation methods

Unlike deterministic methods, stochastic optimisation methods do not rely on derivative information of the objective function and constraints while searching for the best solution; thus stochastic optimisation is more suitable for problems in which the calculation of the function derivatives are complex and for large-scale problems defined using black box models. Stochastic optimisation methods apply random choice to guide the search process.

Stochastic search methods can be used to solve various forms of optimisation problems, including LP, MILP, NLP, and MINLP. The use of random choice rather than numerical calculation to search for optimal solution help stochastic search

methods to be less susceptible to converge to locally optimal solutions. Stochastic approaches require many function evaluations before finding the best solution. Thus stochastic search methods are computationally demanding. Examples of stochastic methods include genetic or evolutionary algorithms (Mitchell, 1998), simulated annealing (Du and Swamy, 2016), pattern search (Wen et al., 2013), particle swarm (Marini and Walczak, 2015), and scatter search (Martí et al., 2006).

Despite the fact that stochastic methods are computationally demanding, many research work have applied these methods, for example in kinetic modelling (Pérez Rivero et al., 2016), regression analysis (Rogina et al., 2011), robust control of distillation column (Ghoreishi et al., 2011), synthesis of heat exchanger networks (Ravagnani et al., 2005; Ghanizadeh et al., 2013; Ochoa-Estopier et al., 2015), design of intensified distillation column (Vazquez-Castillo et al., 2009), process synthesis and design (Yuan et al., 2009; Odjo et al., 2011; Javaloyes-Antón et al., 2013; Skiborowski et al., 2015), scheduling of multiproduct batch chemical plant (Arbiza et al., 2008), process troubleshooting (Mouli et al., 2016), and supply chain management (Copado-Méndez et al., 2013).

In crude oil distillation, Motlaghi et al. (2008) develop a methodology to optimise product yields according to their market values, that is a genetic algorithm selects the optimal decision variables such as operating conditions that improve market-driven products. Similarly, Yao and Chu (2012) apply particle swarm optimisation to facilitate the search for operating conditions (feed temperature, reflux ratio, product flow rates, pump-around temperature drop and flow rates, and steam flow rates) that maximise net profit. Ochoa-Estopier and Jobson (2015a) develop an operational optimisation framework to improve the profitability of a crude oil distillation system. Decision variables include pump-around temperature drop and duties, steam flow rates and feed temperature. Simulated annealing is applied to optimise the system.

Deterministic and stochastic search methods have been applied by several researchers to optimise the crude oil distillation system using different types of objective function and constraints. The methodologies have reported an appreciable improvement in the

objective function of the optimal solution relative to the base case. However, simultaneous optimisation of structural and operational variables of the crude oil distillation unit have not been considered.

To take advantage of the benefits of the deterministic and stochastic search methods, the two optimisation methods can be integrated to form a hybrid approach (see Chapter 5) and can be used to facilitate the design and optimisation of the crude oil distillation systems.

2.5 Process design for flexibility

The traditional design approach for chemical processes considers one set of operating conditions (nominal conditions) at the design stage, thus ignoring deviations from the nominal conditions. This approach may result to a design that has good performance in one operating scenario but exhibits poor performance in other operating scenarios (Grossmann and Guillén-Gosálbez, 2010). Instead, ''flexible design'' of a chemical process considers deviations from the nominal conditions with the aim of identifying a flexible process. A flexible process is capable of establishing feasible steady-state operation for a wide range of variation in operating conditions that may be experienced during operation (Biegler et al. 1997). In process design, flexibility is defined as the inherent characteristic of a design to tolerate variations in process conditions (Biegler et al. 1997).

The design of a flexible chemical process that can handle variability in process conditions is a broad area of research within chemical engineering (Pistikopoulos and Ierapetritou, 1995; Sahinidis, 2004; Wang and Rong, 2010; Kostin et al., 2012; Rogers and Ierapetritou, 2015; Amaran et al., 2016; Wang et al., 2016). The main optimisation methods to address this category of problem are stochastic programming and so-called robust optimisation (Grossmann and Guillén-Gosálbez, 2010).

In stochastic programming, the process variability is described using random points generated from a probability distribution. Here it is assumed that the distribution of

the process variability is known or estimated (Gorissen et al., 2015). This type of problem is typically solved in two stages, i.e., the design and operating stage (Grossmann and Guillén-Gosálbez, 2010; Pistikopoulos and Ierapetritou, 1995). In the first (design) stage, the optimal vector of design variables that represents process structure and equipment size are selected and remains fixed during the operating stage. At the second (operating) stage, the adjustable operating variables are manipulated to determine the optimal vector of operating variables that satisfies the process constraints. The objective here is to minimise or maximise an expected value.

Unlike in stochastic programming in which the distribution of the process variability is assumed to be known, in the robust optimisation the process variability is assume to reside within a given set of scenarios, also known as "uncertain set" (Bertsimas et al., 2010; Gorissen et al., 2015). Robust optimisation aims to find decision variables (e.g. number of trays in a distillation column) that are optimal for the "worst–case" (Bertsimas et al., 2010; Gorissen et al., 2015). A unique feature of robust optimisation is that the optimal solution must satisfy all the problem constraints; no constraint violation is tolerated. The constraints are associated with each operating scenario. Robust optimisation is particularly important if the decision maker is very risk-averse (Gorissen et al., 2015).

In stochastic programming and robust optimisation, a discrete set of operating scenarios is pre-specified, then a process is designed to accommodate the entire scenarios. The flexibility level of the design is not optimised. To determine the optimum degree of flexibility of a design, a trade-off between an economic objective (e.g., total annualised cost and net profit) and design flexibility is required (Pistikopoulos and Ierapetritou, 1995; Biegler et al., 1997). The higher the degree of flexibility, the wider the range of operation and the less the chance of encountering infeasible operation (e.g., failure to meet product specifications). Therefore, the total annualised cost of the design increases with flexibility.

The design methodology for a process with an optimal degree of flexibility is formulated as a multi-objective optimisation problem, i.e., to maximise degree of

flexibility and minimise total annualised cost (Biegler et al., 1997; Pistikopoulos and Ierapetritou, 1995). The multi-objective problem explores the trade-offs between flexibility and total annualised cost within the feasible region of operation.

In general, multi-objective optimisation applies mathematical programming techniques to find optimal solution to a problem involving multiple assessment criteria that are often conflicting. A key characteristic of multi-objective optimisation methods is that no unique solutions exist; instead, a set of mathematically equally good solutions can be identified within the feasible design space (Miettinen, 2008). The set of solution is known as Pareto optimal solution (non-dominated or non-inferior solutions) (Miettinen, 2008). For a problem involving two conflicting performance criteria, the multi-objective optimisation problem can be represented as follows:

$$\text{P1} \qquad \begin{aligned} &\min \ f_1(x,y) \\ &\max \ f_2(x,y) \end{aligned} \tag{2.7}$$

$$\left.\begin{aligned} s.t. \quad h(x,y) &= 0 \\ g(x,y) &\leq 0 \\ x \in \mathbb{R}, y &\in \{0,1\} \end{aligned}\right\} \tag{2.8}$$

where f_1 and f_2 are the scalar objectives to be minimised and maximised respectively; h and g denote the equality and inequality constraints that the solution should satisfy, respectively; x and y are continuous and binary variables respectively; In this case, f_1 and f_2 are total annualised cost and flexibility level respectively.

To solve Problem P1, a quantitative measure of flexibility is required. Swaney and Grossmann (1985) developed a flexibility index that can be used to quantify the maximum deviation a design can accept without violating process constraints, e.g., product specification and distillation column hydraulic limits.

Problem P1 has been used by several authors to design flexible chemical processes other than the crude oil distillation system. For example, Problem P1 has been applied to design a process consisting of a reactor, a flash drum, a purge and two pumps, with

variability in component fraction and kinetic parameters (Pistikopoulos and Grossmann, 1988); a simple distillation column with variability in feed flow rate, condenser and reboiler heat transfer coefficients, and cooling water inlet temperature (Hoch et al., 1995); a simple chemical process comprising of a reactor and a heat exchanger, with variability in process parameters such as temperature, flow rate, reaction kinetics and heat transfer coefficient (Chacon-Mondragon and Himmelblau, 1996); and heat exchanger networks with variability in supply and target temperatures and heat capacity flow rates (Chen and Hung, 2004). For problems with a large number of constraints, the methods for evaluation of flexibility index (Swaney and Grossmann, 1985) can be computationally expensive to implement. Furthermore, methods based on flexibility analysis require the explicit form of equations representing the chemical process, in order to facilitate calculation of flexibility metric, and problem objective function and constraints. Thus, these methodologies cannot be directly applied to optimisation problems that are described using black box models, which is the case in this work.

2.6 Concluding remarks

This chapter presents an overview of the crude oil distillation system, methodologies for the design and optimisation of heat-integrated crude oil distillation systems, and techniques available for process design for flexibility as well as optimisation methods.

Various methodologies have been developed to facilitate the design and optimisation of heat-integrated crude oil distillation systems. However, there are still many important design issues that have not been addressed.

Simplified models are simple and relatively easy to handle numerically, although they produce less accurate estimate of the crude oil distillation unit performance (e.g. capital cost) compared to rigorous models. Therefore, simplified models are less frequently used in practice. Rigorous and surrogate models have been used to optimise the crude oil distillation system. Rigorous models are accurate and produce realistic results; however, they are very complicated and, computationally demanding. These features

make rigorous models less suitable for optimisation purposes, especially for large-scale industrial problems. While surrogate models are relatively simple, robust and less computationally demanding compared to their rigorous counterpart, their accuracy typically depends on the form of the models selected and the data used to fit the model.

The rigorous and surrogate modelling approaches presented in this chapter takes into account only the operating conditions (feed inlet temperature, pump-around duties, and temperature drops, stripping steam flow rates and reflux ratio) of the crude oil distillation system as variables. For grassroots design purposes, the distillation column model should take into account not only operating conditions but also the structural variables (e.g., number of trays in column sections) of the crude oil distillation system. Therefore the rigorous and surrogate models presented cannot be applied for optimisation-based design of crude oil distillation units.

Various methodologies have been presented in the literature for the design and optimisation of heat-integrated crude oil distillation systems. Early ('Traditional') design methodologies (Jones, 1995; Nelson, 1958; Watkins, 1979) apply heuristic rules, empirical correlations, and simple calculations to design the distillation column, without taking into account heat integration within the system. Integrated design methods (Liebmann, 1996; Sharma et al., 1999; Bagajewicz and Ji, 2001) account for heat integration between the distillation column and the heat recovery network, however, the distillation column is not optimised. While simplified models have been used to perform optimisation-based design of the system, their inability to produce accurate predictions may result to an unrealistic design. Most of these methodologies focused on the design of crude oil distillation system that processes only one type of crude oil feedstock. Therefore, change in crude oil feedstocks can affect the system performance, e.g., not meeting separation requirements (product quality specifications). Design methodologies (Bagajewicz and Ji, 2001; More et al., 2010) for crude oil distillation unit that processes multiple crude oil feedstocks are available. Based on a fixed column structure, these approaches identify suitable operating conditions for different varieties of crude oil to be processed. However, the approach propose by Bagajewicz and Ji

(2001) have not considered the trade-off between capital and energy cost, and the column structural and operational degrees of freedom are not optimised. Therefore, the approach may lead to a suboptimal design solution. Although More et al. (2010) presents an approach that optimises some operating variables of the system, many other degrees of freedom are not optimised, and heat integration is not considered. Therefore, the approach may lead to a design that is not energy efficient.

Mathematical programming techniques, such as stochastic programming, robust optimisation, and methods based on flexibility index analysis, have been employed to design chemical processes containing variables that are subject to variability. However, these methodologies cannot be directly applied to a very complicated heat-integrated chemical process such as the crude oil distillation system.

Various works have applied deterministic optimisation (Basak et al., 2002; López C. et al., 2013; Osuolale and Zhang, 2017) and stochastic optimisation (Motlaghi et al., 2008; Yao and Chu, 2012; Ochoa-Estopier and Jobson, 2015a) to search for a new set of decision variables that improve a specific objective relative to a base case. None of the approaches presented has integrated the deterministic and stochastic methods to rigorously explore the search space of the optimisation problem. Furthermore, simultaneous optimisation of structural and operational variables of the crude oil distillation system has not been considered.

This research work aims to develop a systematic methodology for the design of flexible heat-integrated crude oil distillation systems that processes multiple crude oil feedstocks. Chapter 3 presents a new approach for the design of crude oil distillation unit using rigorous model. The methodology incorporates both structural and operational degrees of freedom as design variables in order to facilitate the design of the complex system. Chapter 4 develops a new optimisation-based approach for the design of crude oil distillation units using surrogate models. Both structural and operational degrees of freedom are optimised. Chapter 5 presents a new approach for the design of flexible crude oil distillation unit that processes multiple crude oil feedstocks. The approach is optimisation-based; therefore, the final design is optimal.

The proposed method takes into account relevant operational and structural degrees of freedom and product quality constraints. A hybrid stochastic-deterministic approach is proposed to search for design alternative that is flexible and economically viable. Chapter 5 also presents a new scenario-based design method that handles a large number of operating scenarios. The approach takes into account multiple objectives and can be applied to design problems formulated using equation oriented models and black box models.

Chapter 3 Design of heat-integrated crude oil distillation systems using rigorous simulation models

As discussed in Chapter 2, there are two classes of distillation column models used in practice to represent crude oil distillation units, namely, rigorous models and surrogate models. These simulation models take into account only the operational variables (feed inlet temperature, stripping steam flow rate, pump-around temperature drop and duty, reflux ratio) of the crude oil distillation unit, thus limiting their scope of application. For design purposes, the distillation column model needs to incorporate both structural and operational degrees of freedom as variables in order to enable optimisation-based design of the unit.

This Chapter addresses the first and second objectives of this research work (see Section 1.3), that is, (i) develop an appropriate modelling approach for crude oil distillation units that take into account both structural and operational degrees of freedom of the distillation column. The approach explores the use of both rigorous and surrogate models; (ii) propose a design methodology that incorporates rigorous simulation model and pinch analysis in a unified framework to facilitate the design of crude oil distillation systems. Chapter 4 explores the use of surrogate models.

3.1 Introduction to Publication 1

This paper presents a new approach for the design of crude oil distillation unit using rigorous models. The modelling approach presented in this paper builds a superstructure of the distillation column, taking into account both structural (number of trays in column section) and operational (pump-around duty and temperature drops, feed inlet temperature, stripping steam flow rate, and reflux ratio) degrees of freedom. The superstructure embeds several alternative designs, and it is developed

using rigorous simulation model available in commercial process simulator (in this case Aspen HYSYS). The use of the rigorous tray-by-tray column model in Aspen HYSYS allows the proposed approach to take advantage of the physical, transport, and thermodynamic property models in the process simulator to generate accurate and reliable results.

The accuracy of rigorous distillation column model in Aspen HYSYS depends on the type of equation of state and/ or activity model (property models) applied. Examples of equation of state in Aspen HYSYS include Peng-Robinson (Peng and Robinson, 1980), Soave-Redlich-Kwong (Soave, 1980), Kabadi Danner (Kabadi and Danner, 1985), Lee Kesler Plocker (Plöcker et al., 1978), Zudkevitch-Joffee (Zudkevitch and Joffe, 1970), etc. On the other hand, activity models include Non-Random-Two-Liquids (Austgen et al., 1989), universal quasi-chemical (Maurer and Prausnitz, 1978), and Margules, van Laar, Wilson models (Perry and Green, 2008). These models have been previously tested and validated over a range of conditions (temperature and pressure), components, and component mixtures (AspenTech, 2011). The type of property model to be used depends on the components of the system under consideration and the operating conditions. For a crude oil distillation system containing complex hydrocarbon mixture and water, the recommended property model is Peng Robinson (Fahim et al., 2009; Chang et al., 2012), and it is the model used in this work.

In building the column superstructure, Murphree tray efficiency (Seader et al., 2010) related to each tray within the column is treated as a binary variable, i.e. an efficiency of one is specified if a tray is taking part in the separation ('active tray'), and zero otherwise ('inactive tray'). In this way, the total number of trays in column sections can be optimised. This is the first attempt to incorporate number of trays as a design variable in modelling complex crude oil distillation unit using rigorous simulation models.

To design the crude oil distillation system, an approach is proposed that incorporates the rigorous tray-by-tray distillation unit model (superstructure representation), pinch analysis, hydraulic model and cost model in an optimisation framework to facilitate

the design of the distillation system. A genetic algorithm is used to select the best column structure and operating conditions that minimises total annualised cost. As will be seen in the paper presented in this chapter, heat integration and hydraulic calculations are carried out in each iteration to enable the estimation of operating cost and capital cost respectively. This approach enables the inherent trade-offs between capital and energy cost to be exploited during optimisation, and thus guide the optimisation algorithm towards a cost-effective solution.

The proposed methodology is applied to a case study that concerns the design of a refinery crude oil distillation unit that separates Tia Juana light crude oil (Watkins, 1979) into intermediate products, such as light naphtha, heavy naphtha, light distillate, heavy distillate and residue. Numerical results show that energy efficient and cost-effective design alternative can be identified within the solution space. The supporting information for this paper is presented in Appendix A.1.

3.2 Publication 1

Ibrahim, D., Jobson, M., Guillén-Gosálbez, G., 2017. Optimization-based Design of Crude Oil Distillation Units using Rigorous Simulation Models. *Ind. Eng. Chem. Res.*, **2017**, *56* (23), pp 6728–6740, **DOI:** 10.1021/acs.iecr.7b01014

Optimization-based design of crude oil distillation units using rigorous simulation models

Dauda Ibrahim[1,*], Megan Jobson[1], Gonzalo Guillén-Gosálbez[2]

[1] Centre for Process Integration, School of Chemical Engineering and Analytical Science, University of Manchester, Manchester M13 9PL, UK

[2] Department of Chemical Engineering, Centre for Process Systems Engineering, Imperial College, South Kensington Campus, London SW7 2AZ, UK

Abstract

The complex nature of crude oil distillation units, including their interactions with the associated heat recovery network and the large number of degrees of freedom, makes their optimization a very challenging task. We address here the design of a complex crude oil distillation unit by integrating rigorous tray-by-tray column simulation using commercial process simulation software with an optimization algorithm. While several approaches were proposed to tackle this problem, most of them relied on simplified models that are unable to deal with the whole complexity of the problem. The design problem is herein formulated to consider both structural variables (the number of trays in each column section) and operational variables (feed inlet temperature, pump-around duties and temperature drops, stripping steam flow rates and reflux ratio). A simulation-optimization approach for designing such a complex system is applied, which searches for the best design while accounting for heat recovery opportunities using pinch analysis. The approach is illustrated by its application to a specific distillation unit, in which numerical results demonstrate that the new approach is capable of identifying appealing design options while accounting for industrially relevant constraints.

Keywords: Process design, heat integration, genetic algorithm, grand composite curve

1 Introduction

Crude oil distillation is one of the most complex and energy-intensive separation units in the petroleum refining industry. The crude oil distillation system comprises a complex distillation unit and a heat recovery network in which the crude oil feed is partially vaporised. Figure 1 illustrates a typical petroleum refinery crude oil distillation system. The system includes a fired heater that typically consumes fuel equivalent to 1 to 2% of the oil being processed[1,2]. This combustion of fuel is associated with high CO_2 emissions and high operating costs. Extensive heat recovery is routinely implemented in crude oil distillation systems to reduce energy costs.

In grassroots design, several degrees of freedom related to the column structure, its operating conditions and the associated heat recovery network are subject to optimization. The need to account for the complex interactions between these subsystems makes the design of crude oil distillation columns a highly challenging task. For a new ('grassroots') design, the column configuration (number of trays in each section of the column and location of the feed tray, pump-arounds and side-stripper draws) and the operating conditions (feed inlet temperature, pump-around duties and temperature drops, stripping steam flow rates and reflux ratios) need to be selected. In addition, the heat recovery network (known as the preheat train) needs to be designed simultaneously. In this way, the column can be designed to create heat recovery opportunities that can be further exploited by the heat exchanger network. The design of this heat recovery network aims to identify the network configuration and heat transfer area that minimise the total annualised cost while accounting for both capital and operation expenditures.

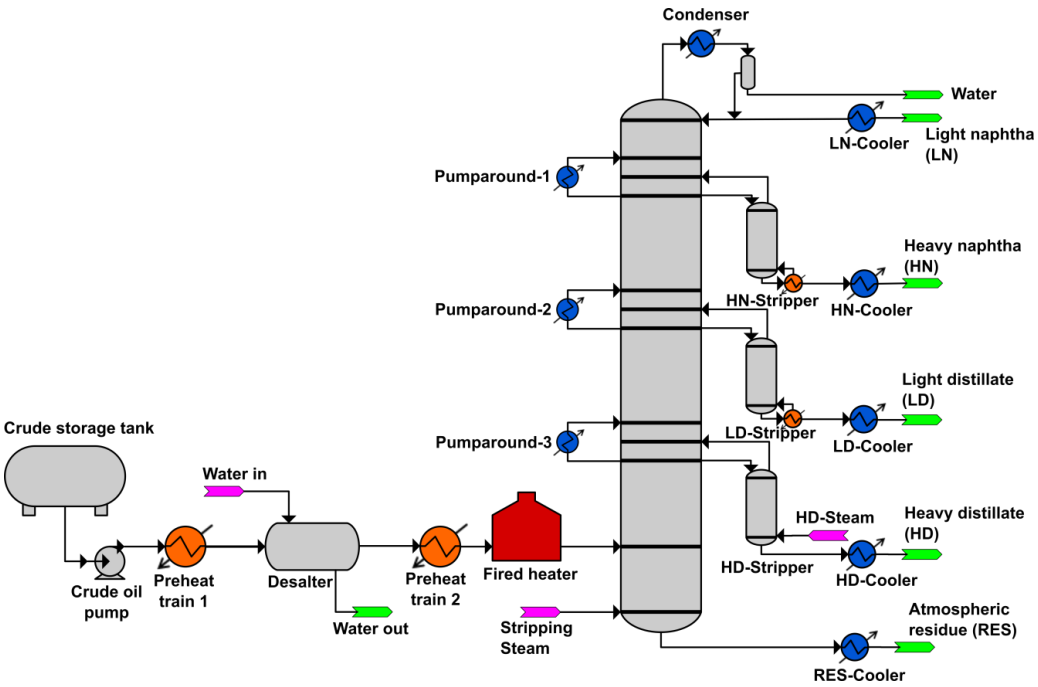


Figure 1 Typical crude oil distillation system

The operation of a standard crude oil distillation column is as follows (see Figure 1). Stored raw crude oil is partially heated in Preheat Train 1 and fed to a desalter, which removes dissolved or suspended salts from the crude oil feed[3]. The crude oil is further heated, in Preheat Train 2 and a fired heater, before being fed to the atmospheric distillation column. The preheat trains use heat recovered from the crude distillation unit, particularly the pump-arounds, condenser and product streams. The partially vaporised crude oil is fed to the atmospheric distillation column a few trays above the bottom stage. Stripping steam is supplied to the column at the bottom stage, which partly suppresses the boiling point of the crude mixture and further vaporises the crude oil mixture. The crude oil is separated into various fractions, such as light naphtha (LN), heavy naphtha (HN), light distillate (LD) and heavy distillate (HD). Side-strippers remove light components from side-draws using stripping steam or reboilers. Pump-arounds provide internal reflux and create heat recovery opportunities by cooling and returning liquid streams withdrawn from the column. The residue from the atmospheric column contains valuable hydrocarbons, which are typically further separated in a vacuum distillation column (not shown in Figure 1).

Conventional design methods[4–6] consider the complex column and the heat recovery network in separate steps, without taking into account interactions between the two subsystems. Various researchers[7–9] have applied optimization techniques to design the crude oil distillation column while simultaneously accounting for the heat recovery network. However, these approaches apply simple shortcut distillation models[7] to support the design task in an attempt to avoid the numerical problems encountered when optimising more rigorous simulation models. The use of these shortcut models can lead to large errors, as they often cannot accurately predict the behaviour of the complex crude oil distillation unit[8]. These shortcut models are also restricted to specific column configurations, which limits their applicability.

This work applies a simulation-optimization approach for the design of crude oil distillation units that integrates a rigorous tray-by-tray model of the distillation unit implemented in a commercial process simulator (Aspen HYSYS v8.6) with an optimization algorithm coded in Matlab R2015a. The optimization of process simulation models using external algorithms was addressed in other works[10–12], but to the best of our knowledge none of them applied this approach to the design of complex crude-oil distillation units. In essence, these approaches decouple the simulation from the optimization in order to simplify the modelling and subsequent optimization of the process model. The process model is thus implemented in a simulation package that solves a system of nonlinear equations, while the optimization is carried out by an external algorithm that seeks the best values of the independent values by iteratively interrogating the process model.

As will be later discussed in the article, these approaches differ in the optimization algorithm employed, which can be a deterministic method (e.g. gradient based) or based on stochastic optimization algorithms (e.g. genetic algorithms, simulated annealing). Hence, when applied to the design of complex distillation units, the simulation-optimization approach takes advantage of the physical property and thermodynamic models, as well as the crude oil characterization and column hydraulic models available in the process simulator. These tailored models ultimately lead to more accurate results compared with the use of shortcut methods. In addition, the

rigorous simulation environment is more versatile, allowing for a more flexible modelling of the column configuration. Furthermore, heat recovery opportunities for each proposed design were determined in this work using an open source algorithm implemented in Matlab[13]. Following this approach, pinch analysis was used iteratively to perform heat integration calculations for the designs proposed by an external optimizer.

The remainder of this article is organised as follows. Section 2 reviews existing methodologies for the design of crude oil distillation units. In Section 3, a superstructure representing the crude oil distillation unit is proposed and a detailed optimization formulation is presented together with a customized solution procedure. The solution procedure makes use of readily available commercial process simulator to simulate the crude oil distillation column, hence avoiding the need to formulate the complex column using explicit equations; moreover, the process simulator environment is versatile and user-friendly, thus making our approach easier to implement in practice and accessible to industrial practitioners. Section 4 introduces a case study that illustrates the capabilities of the proposed design methodology. The conclusions of the work are finally presented in Section 5.

2 Previous research on crude oil distillation unit design

In the past decades, various methods have been proposed and developed for the design of crude oil distillation units. Conventional methods apply heuristic rules, experience, empirical correlations and simple relationships. For instance, the number of trays in each section of the column and the stripping steam are often estimated based on empirical correlations obtained from previously established designs [4]. Similarly, in Watkins[5] method, the number of trays in each section of a column is selected from a predetermined range, while the stripping steam flow is estimated based on product flow rates. The approaches of Nelson[4] and Watkins[5] formed the basis for many subsequent design methodologies for crude oil distillation units that involve iterations and trial and error procedures. Furthermore, the heat recovery network is

omitted in these design approaches, which neglect the complex interactions between such a network and the distillation column.

Other research has focused on developing integrated design methods that address the design of the crude oil distillation unit and the associated heat recovery network simultaneously. Along these lines, Liebmann and co-workers[1,14] combined rigorous column models and pinch analysis to design a crude oil distillation unit. Their approach takes design decisions in a sequential manner considering heat recovery at each step using pinch analysis. To avoid the numerical difficulties associated with the rigorous simulation of the column, Sharma[15] proposed to use the concept of a column grand composite curve[16]. This strategy identifies the maximum amount of energy that can be recovered without affecting the separation. A limitation of this approach is that the role of the stripping steam is neglected. Bagajewicz and Ji[17] focused on overcoming the above limitation, incorporating the effect of the stripping steam on the maximum heat recovery of a crude oil distillation column and introducing the concept of a heat demand–supply diagram. This approach, however, does not account for the trade-off between capital and energy costs.

To design an integrated process system, it is necessary to design the complex column and the heat recovery network simultaneously. For example, Suphanit[7] applied the column decomposition strategy of Liebmann and co-workers[1,14] to develop a shortcut model for the crude oil distillation column. This model was then used within an optimization framework together with pinch analysis to simultaneously optimise distillation operating variables and the heat recovery network (utility demand and area) so as to minimize the total annualized cost. Rastogi[8] extended the shortcut model of Suphanit[7] to account for column pressure drop and pump-around location. A detailed model of both the heat exchanger network and the distillation column was incorporated into an optimization framework that optimized the column structure and operating conditions. Chen[9] modified the shortcut models of Rastogi[8] to allow for other pump-around locations and also modelled temperature-dependant properties of process streams undergoing phase change. In this work, the structure and operating

conditions of the column together with the heat exchanger network were optimized using simulated annealing.

A comprehensive overview of optimization methods applied to process synthesis and design can be found in the excellent work of Grossmann et al.[18] and Grossmann and Guillén-Gosálbez[19]. The optimization methods have been applied to design several chemical processess (other than crude oil distillation units)[20–24].

Several conclusions can be drawn from the literature review presented above. Traditional distillation design methods do not simultaneously consider heat recovery. Integrated design approaches do consider both, the column and the heat recovery network, but seldom analyse the trade-offs between capital and energy cost in a rigorous way. Optimization techniques have been used to design the column and heat recovery network[7–9]. However, to simplify the calculations, most of these approaches rely on shortcut distillation models[7] that provide less accuracy and versatility. No approaches have been identified that directly use rigorous distillation models for optimization-based design of heat-integrated crude oil distillation systems.

This research introduces a systematic framework for the design of heat-integrated crude oil distillation units that overcomes the limitations of established methods. Our approach applies rigorous tray-by-tray distillation column models to simulate alternative designs. These models are combined with a genetic algorithm that optimizes the column design. The number of trays in each column section together with the operating conditions (including the feed inlet temperature, pump-around duties and temperature drops, stripping steam flow rates and reflux ratios) are selected to minimize the total annualized cost. This cost accounts for the annualized capital cost and annual operating costs related to fuel consumption in the furnace.

3 Optimization-based design approach

This section presents a simulation-optimization based approach for the design of crude oil distillation units. First, the rigorous tray-by-tray model used to simulate the crude

oil distillation column is discussed. Then, the mathematical formulation of the optimization problem is presented. Next, an approach proposed to solve the optimization problem is described. We emphasize that we are dealing here with a very complex crude oil distillation system for which many decisions (including pump arounds, diameters and number of trays in different sections and operating conditions) must be optimized all together while considering the design of the HEN coupled with the unit. Developing a short-cut method for such a system is a very challenging task that would very likely result in larger approximation errors.

3.1 Crude oil distillation unit simulation model

In process design, it is crucial that models used to simulate design options are sufficiently realistic to deliver feasible solutions. Two main types of models are available for design of crude oil distillation units, namely, shortcut models[7–9] and rigorous models[14,17]. The shortcut models adapt the Fenske–Underwood–Gilliland design equations for simple columns. These models predict the number of trays in each column section and operating conditions, such as reflux and reboil ratios. When applied to crude oil distillation, these models are restricted in terms of allowable configurations and accuracy of the predictions.

On the other hand, the so-called rigorous models apply material and energy balances as well as equilibrium relations in every stage of the column[25]. These models provide more accurate predictions. However, they are more difficult to handle due to the need to start the calculations from a very good initial guess in order to avoid convergence problems. Procedures for solving rigorous models are well established, and have been implemented in commercial process simulation software such as Aspen HYSYS, Aspen Plus, UNISIM, and PRO II. Such software allows designers to simulate complex distillation column flowsheets using iterative and sequential modular algorithms. Here, there is no need to define in an explicit form the model equations, as they are already implemented in the process simulator. Simulation packages like ASPEN, HYSYS or gPROMS already contain specific routines to solve distillation columns (and other unit operations) that are highly efficient. In this work, without loss of generality,

the 'rigorous' distillation model available in Aspen HYSYS v.8.6 is used to simulate the crude oil distillation column under steady state. Section 3.3 presents a detailed description of how the rigorous column model is built.

3.2 Heat recovery consideration

In general, heat recovery is carried out using a heat exchanger network (HEN). After fixing the process configuration and operating conditions, information on process streams (i.e., inlet and outlet temperatures and duties of all streams requiring heating and cooling) becomes available. This information could be used to design the HEN, which determines the minimum utility requirements that will be used to evaluate the design options.

In an optimization-based design, many options need to be evaluated before selecting the best alternative. In this context, designing a full HEN for each potential design would require significant computational effort. This is because the HEN design can itself be posed as a nonconvex MINLP problem that is per se hard to solve, mainly due to the presence of bilinear terms in the constraints as well as concave ones in the objective function[26]. While there have been some recent attempts to solve the HEN design problem more efficiently[21], the methods proposed still scale poorly with the number of hot and cold streams. To overcome this limitation, pinch analysis is applied here. Hence, targets for minimum utility requirement are determined to screen the design options and propose improvements for existing designs[17]. In this work, the grand composite curve is coded in Matlab R2015a[10] and incorporated into the optimization procedure to calculate minimum utility requirements for the crude oil distillation unit. Detailed HEN design is not addressed. Nevertheless, pinch analysis is expected to minimize the dominant cost, i.e. fired heating, and it is well known that utility costs dominate distillation process economics. It is anticipated that the annualized HEN capital costs will be relatively similar for different column designs. Future work intends to account for HEN details.

3.3 Crude oil distillation column modelling – superstructure formulation

In this section, the column superstructure used to design the crude oil distillation column is developed. A process superstructure considers (ideally) all possible design alternatives simultaneously. The superstructure of the complex heat-integrated crude oil distillation column is built treating Murphree tray efficiencies[27] as binary variables that can activate or deactivate trays (following the approach developed by Yeomans and Grossmann[28] and Caballero et. al.[10]). In this approach, a column section containing equilibrium stages includes a set of 'temporary' trays (also known as inactive trays) and 'permanent' trays (also known as active trays). On a permanent tray, mass transfer takes place between the vapour and liquid phases; it is assumed that phase equilibrium is achieved. On a temporary tray, no mass transfer takes place; the temporary tray is modelled as a by-pass with inputs equal to the outputs in each phase. In a commercial process simulator, trays can be modelled by setting appropriately their Murphree tray efficiency[27]: zero (when the tray is inactive) or one (when it is active). On both types of trays, the material and energy balances and equilibrium relations are solved. However, on a temporary tray, no separation takes place. Figure 2 illustrates the superstructure for modelling the crude oil distillation column.

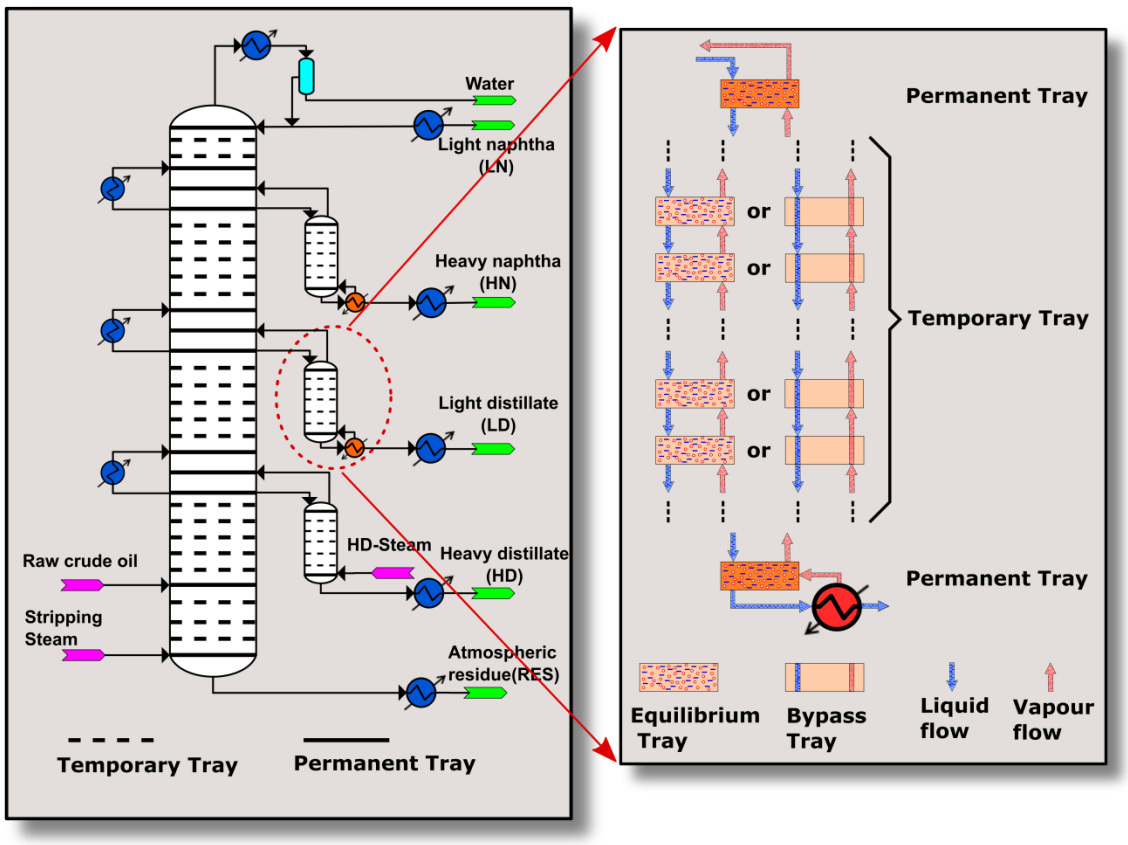


Figure 2 Superstructure representation of crude oil distillation column

As shown in Figure 2, the column superstructure consists of eight sections: five sections in the main column and three side strippers. The initial number of trays in each column section can be selected using traditional methods[4,5] or shortcut models[7–9]. In the superstructure, each section is set up to ensure that more trays are available than will be needed. During optimization, the optimal number of trays in each section will be determined. The minimum possible number of permanent (active) trays in the superstructure is 18; 12 in the main column and 2 in each side stripper. These trays are located at points where a stream enters or leaves the column.

3.4 Mathematical formulation

The crude oil distillation column design problem can be formulated as an MINLP model (M1) based on the superstructure proposed in Figure 2:

$$\text{(M1)} \qquad \min_{x_S, x_O} \quad \psi(x_D, x_S, x_O) \tag{1}$$

$$s.t. \quad h_I(x_D, x_S, x_O) = 0$$
$$h_E(x_D, x_S, x_O) = 0$$
$$g_E(x_D, x_S, x_O) \leq 0$$

$$x_D \in X_D, x_S \in X_S, x_O \in X_O$$

where ψ is the objective function; h_I denotes the set of implicit equality constraints representing material, energy and thermodynamic equations embedded in the process simulator; h_E is the set of explicit equality constraints while g_E is the set of inequality constraints. X_D, X_S and X_O are the feasible sets for the decision variables, namely x_D, x_S and x_O, which represent dependent, structural and operational variables, respectively. The dependent variables are calculated by the simulator for fixed values of X_S and X_O.

For the crude oil column design problem, the inequality constraints can be more specifically formulated as in Model M2:

$$\text{(M2)} \qquad \min_{x_S, x_O} \quad \psi(x_D, x_S, x_O) \tag{2}$$

$$s.t. \quad h_I(x_D, x_S, x_O) = 0$$
$$h_E(x_D, x_S, x_O) = 0$$
$$g_1: lb_i \leq N_i \leq ub_i \qquad i = 1,2,3,\dots,N_{section}$$
$$g_2: lb_j \leq Q_{PA,j} \leq ub_j \qquad j = 1,2,3$$
$$g_3: lb_j \leq \Delta T_{PA,j} \leq ub_j \qquad j = 1,2,3$$
$$g_4: lb_k \leq F_{s,k} \leq ub_k \qquad k = 1,2$$
$$g_5: lb \leq R \leq ub$$
$$g_6: lb \leq T_F \leq ub$$
$$g_7: lb_l \leq T5_l \leq ub_l \qquad l = 1,2,\dots,N_{product}$$
$$g_8: lb_l \leq T95_l \leq ub_l \qquad l = 1,2,\dots,N_{product}$$

where N_i is the number of active trays in column section i; $Q_{PA,j}$ and $\Delta T_{PA,j}$ are the duty and temperature drop of pump-around j; $F_{S,k}$ is the steam flow rate of stream k; R is the overhead reflux ratio; T_F is the feed inlet temperature; and $T5_l$ and $T95_l$ are the boiling temperatures of product l at 5% and 95% vaporization (according to ASTM standards; note that other specifications could be defined in a similar way); N_l is the number of product l.

In Model M2, g_1 to g_6 are bounds on structural and operational variables, while g_7 and g_8 represent constraints on product quality in terms of ASTM D86 boiling temperature: T5 and T95. To enhance the numerical robustness of the model, it is advantageous to include the latter constraints (product quality) in the objective function via penalty terms[21]. The resulting formulation, M3, is:

$$\text{(M3)} \quad \min_{x_S, x_O} \quad \psi(x_D, x_S, x_O) + \left[\Pi \sum_{i=1}^{n} [\max(0, (g_i))]^2\right] \tag{3}$$

$$s.t. \quad h_I(x_D, x_S, x_O) = 0$$
$$h_E(x_D, x_S, x_O) = 0$$
$$g_1, g_2, g_3, g_4, g_5, g_6$$

where g_i denotes the inequality constraints g_7 and g_8; Π is a scalar parameter that scales the magnitude of the violation of constraints, and hence ensures that the product quality specifications are maintained during the optimization. Note that this penalty term can be easily formulated using slack variables.

3.4.1 Objective function

The aim of the optimization-based design task is to search for those process structure and operating variables that best achieve a desired objective. Different types of objective functions are relevant, for example, net profit, energy cost, net present value and total annualized cost. The most appropriate objective function to be used depends on the aims of the design. For grassroots design, a suitable objective is to minimize the

total annualized cost (TAC), as it accounts for the trade-off between capital investment and operating expenses. The total annualized cost is the sum of the total operating cost (OC) and annualized capital cost (ACC) (Smith, 2005):

$$TAC = OC + ACC \tag{4}$$

For the particular case of crude oil distillation unit, the most significant operating costs are the cost of stripping steam and of hot and cold utilities, usually fuel for fired heating and cooling water. Pinch analysis allows minimum utility requirements to be calculated[29] for a given set of heating and cooling duties. In this way, opportunities for heat recovery are accounted for during the design optimization.

$$OC = \sum_{i=1}^{n} ST_i . C_{ST,i} + HU . C_{HU} + CU . C_{CU} \qquad n = 2 \tag{5}$$

In Eq. (5), C_{ST}, C_{HU} and C_{CU} are the unit costs of stripping steam, hot and cold utilities, respectively; HU and CU are minimum hot and cold utilities, respectively, while n represents the number of stripping steam streams associated with the column.

The annualized capital cost is the installed cost of the column shells (S_c) and the installed cost of trays within the column, (T_C), multiplied by an annualization factor (A_f)[29].

$$ACC = (S_c + T_C) * A_f \tag{6}$$

$$A_f = \frac{i(1+i)^t}{(1+i)^t - 1} \tag{7}$$

where i is the interest rate and t is the plant life.

3.4.2 Cost models

The column shell and tray costs are estimated using the correlations proposed by Guthrie[30].

$$S_C = \left(\frac{M\&S\ Index_{2011}}{280}\right) 101.9(D)^{1.066}(H)^{0.802}(2.18 + F_{c1}) \quad (8)$$

where M&S Index$_{2011}$ is the Marshall and Swift chemical equipment cost index for year 2011 (4th quarter)[31] allowing costs to be updated from 1969 (when the M&S Index was 280); the cost is updated to current equipment cost using Eq. (10); D is the sectional diameter of the column, H is the sectional height, which depends on tray spacing and $F_{c1}$is the column cost factor, which depends on the column material of construction and column operating pressure.

$$T_C = \left(\frac{M\&S\ Index_{2011}}{280}\right) 4.7(D)^{1.55} H F_{c2} \quad (9)$$

The tray cost factor F_{c2} depends on the type of tray, tray spacing and material of construction.

$$\left(\frac{CEPCI_{2014}}{CEPCI_{2011}}\right) \times Cost_{2011} \quad (10)$$

where CEPCI$_{2011}$ and CEPCI$_{2014}$ are the chemical engineering plant cost index for year 2011 (4th quarter)[31] and 2014 (4th quarter)[32] respectively; Cost$_{2011}$ is the equipment cost for year 2011, calculated using Eq. (8) and Eq. (9). The M&S Index$_{2011}$, CEPCI$_{2011}$ and CEPCI$_{2014}$ are 1536.5, 590.1 and 575.7 respectively.

The column diameter and height for a specific type of internal are determined using hydraulic models, as discussed in the Section 3.4.3.

3.4.3 Hydraulic models

In crude oil distillation column design, hydraulic analysis is required to identify an appropriate tray selection and to avoid entrainment (or jet flooding), weeping, coning and downcomer flooding[25]. Different design and types of trays and packings (e.g. sieve

tray, valve tray, high capacity tray) have a specific range of satisfactory vapour and liquid flow rates, defined by upper and lower bounds[25,33,34]. Therefore, the column diameter has to be sized appropriately to accommodate the flows of vapour and liquid in the column, and also provide enough active area for mass transfer[25], without incurring an excessively high pressure drop. The tray spacing needs to be chosen in order to avoid entrained liquid jetting onto the tray above[25]. Furthermore, the downcomers should have sufficient area to allow liquid flow[25]. Established methods are available for column sizing and selection of internals[25,34,35], many of which have been implemented in commercial process simulators. In this work, the column sizing capabilities implemented in the tray sizing utility of Aspen HYSYS v8.6 are employed to carry out the hydraulic calculations. The column diameter obtained from this calculation, together with the column height (determined based on permanent trays), are used to determine the purchase cost of the column.

3.5 Solution procedure

The optimization of the column naturally leads to an MINLP problem containing nonlinear equations and binary as well as continuous variables. Various approaches have been developed and proposed to solve this type of problem. These approaches can be broadly classified as deterministic (also known as gradient-based) methods[36–39] and stochastic (a class of derivative-free methods) methods[36,40]. A detailed discussion of MINLP algorithms can be found elsewhere[36–41]. Note that our MINLP is not defined in a fully explicit manner, but rather via both explicit and implicit equations implemented in the simulator and in an external modelling system (i.e., Matlab). MINLP problems encountered in the simulation-optimisation of chemical processes can be solved by several methods. Caballero et al.[10] applied gradient-based methods to solve one such MINLP, which was decomposed into two levels following an outer-approximation scheme. In this work, at the lower level continuous variables are optimized for a fixed design by solving an NLP problem in which a gradient-based NLP solver iterates with the simulation model. At the upper level, new designs encoded in the values of the binary variables are generated by solving an MILP. This MILP is constructed by linearizing the nonlinear equations at the optimal solution of the NLP. These two levels

are solved iteratively until they converge towards a final solution considering a given stopping criterion. Stochastic approaches, on the other hand, attempt to solve the MINLP in a simultaneous way by optimising the decisions variables all together and without relying on gradient-based methods. The standard approach here is to combine metaheuristics (e.g. simulated annealing, genetic algorithms, etc.) with the simulation model and let them iterate for a given time.

Since Model M3 is nonlinear and non-convex, standard deterministic methods can only guarantee convergence to a local optimum. Furthermore, obtaining the derivatives of the NLP might be difficult, which may lead to convergence problems when applying gradient-based NLP algorithms. To overcome these limitations, this work applies a stochastic global search method to solve M3 based on genetic algorithms (GA). We note that, despite the various strategies implemented in the GA, this approach is unable to guarantee convergence to the global optimum. Global optimality can only be ensured using deterministic methods, but these require the explicit form of the equations. In our case, these equations are implemented by the simulator, which does not provide direct access to the equations.

The strategy proposed to optimize Model M3 is presented in Figure 3. The proposed approach combines a rigorous tray-by-tray crude oil distillation column model with cost models, a heat recovery model and a hydraulic model within a unified framework. This strategy searches for the best configuration and operating conditions that result in minimum total annualized cost. The crude oil distillation unit is simulated using a rigorous column model implemented in Aspen HYSYS v8.6, while the optimization algorithm is coded in Matlab R2015a. The exchange of information between Matlab R2015a and Aspen HYSYS v8.6 is established using the automation client-server application provided by Matlab R2015a.

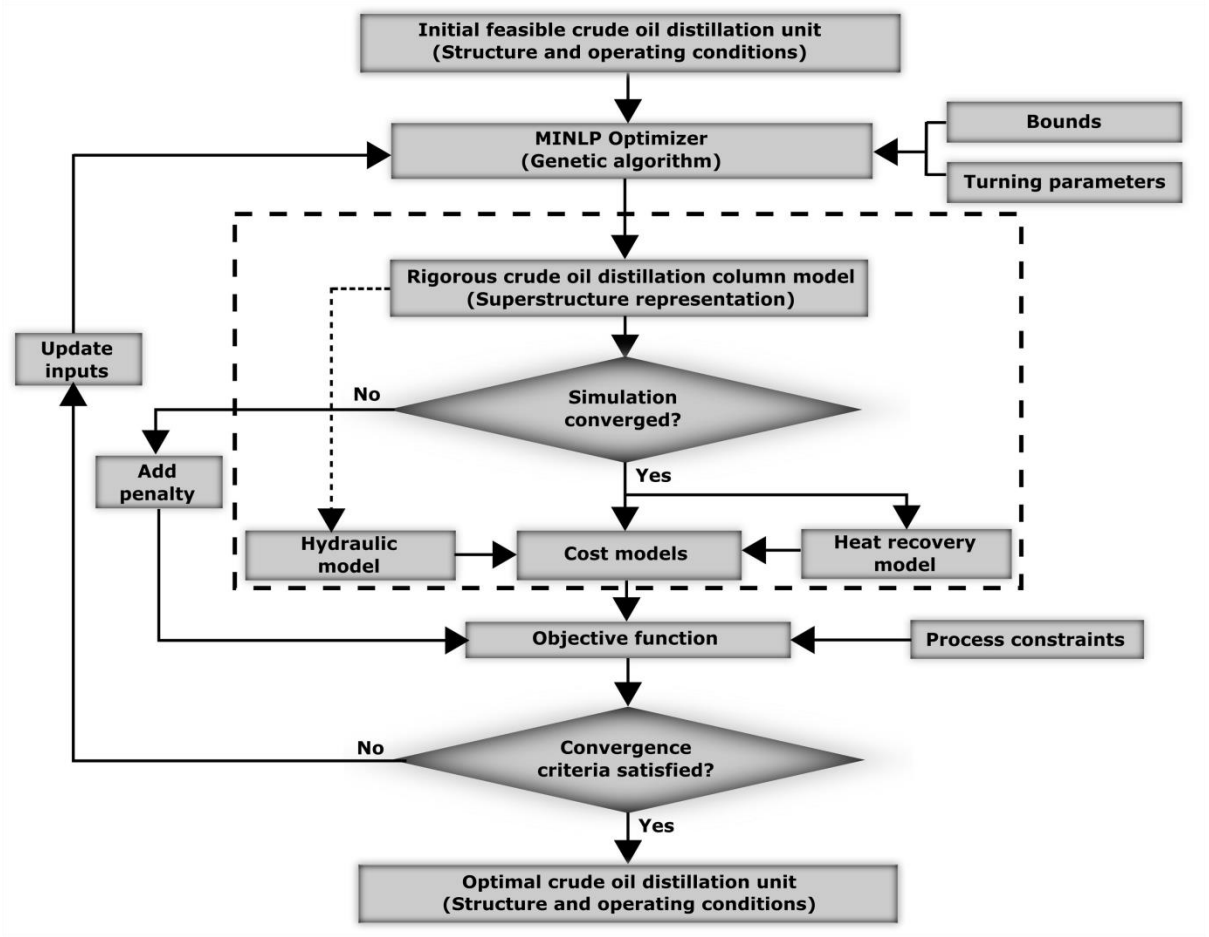


Figure 3 Framework for the optimization-based design of crude oil distillation unit

The implementation of the proposed framework is carried out in two steps. The distillation column superstructure is first defined in the Aspen HYSYS environment. This superstructure contains dependent, structural and operational variables that will be optimized by the GA. The mixed integer non-linear programing model (M3) is solved by a GA implemented in Matlab. The overall MINLP model includes equality constraints (explicit and implicit), inequality constraints, bounds on optimization variables and the objective function. The interface established between Matlab and Aspen HYSYS facilitates the transfer of the data required to calculate the objective function value and to assess whether the constraints are met. Once the model is defined, and the link between Matlab and Aspen HYSYS has been established, an optimization algorithm (e.g. genetic algorithm) is employed to search for the optimal column design.

Hence, in each iteration, the genetic algorithm proposes the column structure and operating conditions to be simulated in the rigorous column model. Process stream information from the converged column flowsheet is used by the heat recovery model to calculate the minimum hot and cold utility demand. For each permanent (active) tray inside the column, the vapour and liquid flow rates and fluid properties are calculated and used in the hydraulic model to determine the column diameter (using the tray sizing utility within the Aspen HYSYS v8.6 simulation environment). The column diameter and height, utility targets and steam flow rates are used to calculate the objective function. This objective value is used to guide the search for an optimal set of structural and operational variables that minimize the column total annualized cost. Certain combinations of inputs proposed by the optimization algorithm may lead to an infeasible simulation of the rigorous column, which may consequently halt the algorithm. This issue is overcome by adding a large penalty to the objective function whenever the column simulation fails. As discussed later in detail, numerical results show that this strategy leads to feasible designs that do not violate any inequality. The section that follows provides more detail about the genetic algorithm.

3.5.1 Optimization algorithm

The genetic algorithm falls into the class of stochastic optimization methods known as evolutionary programming. Genetic algorithms have been successfully applied to many complex chemical engineering problems[42–45]. The algorithm handles both integer decisions and continuous variables and does not require derivative information. Therefore, discontinuous functions can be handled with ease.

The implementation of the genetic algorithm involves four fundamental steps, namely, generation of a random population of individuals, evaluation of fitness of individuals, and selection of the best individuals and reproduction using genetic operators (cross-over and mutation) in order to create the population for the next generation[46]. Figure 4 provides an overview of how these steps evolve to an optimal solution. A detailed description of each step follows.

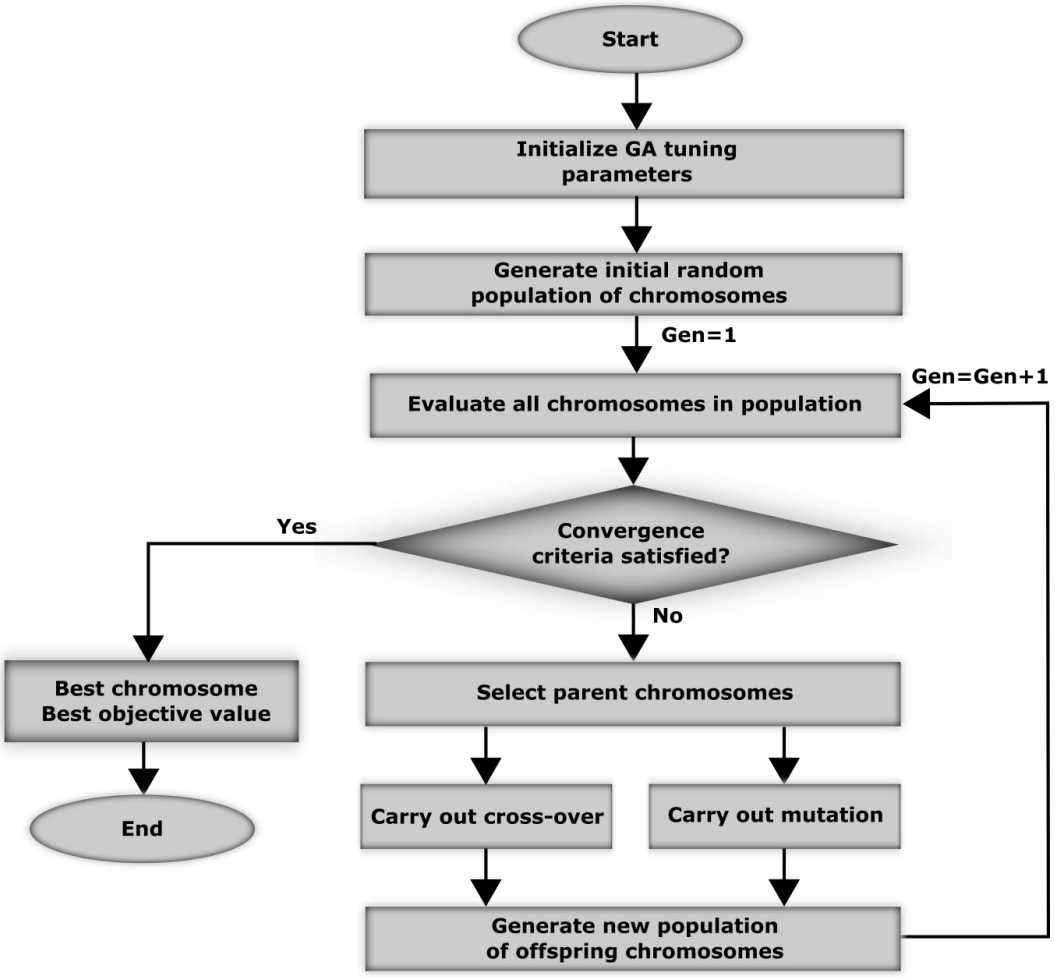


Figure 4 Solution strategy based on genetic algorithm (adapted from Odjo et al.[42])

Step 1: Generation of random population of chromosomes

The first step in the implementation of the genetic algorithm involves the generation of a random population of chromosomes (or individuals), which represents alternative structure and operating conditions of the crude oil distillation unit within the search space. The search space is defined by lower and upper bounds on the optimization variables and by a set of equality constraints that are defined implicitly in the simulation that enable the assessment of the units' performance. Each chromosome contains 18 genes: 8 integer variables representing the number of trays in each column section and 10 continuous variables representing feed inlet temperature, pump-around duties and temperature drops, stripping steam flow rates and reflux ratio. This number of genes comes from an analysis of degrees of freedom in our systems. Note that other

combinations of variables could be defined. However, numerical examples show that this particular choice leads to a system that is numerically more robust than others, meaning that there are fewer solutions proposed by the genetic algorithm that do not convergence in the simulator. The total number of chromosomes in a generation, the population size, is usually pre-specified by the designer.

Step 2: Evaluation of individual fitness

In this step, all the chromosomes generated in Step 1 representing alternative designs are simulated on the rigorous distillation column model in HYSYS in order to determine their relative fitness. The fitness function is usually the objective function of the optimization problem, which in this case corresponds to the total annualized cost. The specifications in the simulation are as follows: product quality, boil-up ratio(s), reflux ratio(s), pump-around duties and temperature drops. Chromosomes leading to simulations that do not convergence are penalised to prevent the genetic algorithm from proposing similar solutions again. Similarly, solutions that do converge but violate at least one constraint are also penalised (see Eq. (3)).

Step 3: Selection of best individuals

All the members of the population are evaluated in terms of their fitness. High scores are assigned to members with high fitness and low scores to those with low fitness. High performing chromosomes (also called parents), i.e. alternative designs with minimum total annualized cost are later chosen for reproduction (cross-over or mutation). Some of these high performing chromosomes (elites count) are retained and passed forward to the subsequent generation without changing their form[47]. In this work, cross-over fraction and mutation fraction are 0.8 and 0.2 respectively; and elites count is 5% of the total population.

Step 4: Reproduction

Reproduction consists of two operations, namely, cross-over and mutation. For the cross over operation, two parents are selected at random, and then some part of the genetic information (i.e. number of trays in column sections, feed inlet temperature, pump-around duties and temperature drops, stripping steam flow rates and reflux ratio) of one parent is swapped with the other to create two new chromosomes[47]. Mutation involves random alteration of the genetic information (either number of trays in column section or operating conditions) of one parent to produce a new chromosome[47]. This operation helps to ensure diversity of the population and to prevent the algorithm from been trapped in a locally optimal solution[47].

The steps listed above are repeated several times until one of the convergence criteria is satisfied. In this work, as convergence criteria we use the maximum number of generations and population convergence. The former establishes a maximum number of iterations after which the algorithm terminates while the latter stops when the difference in performance between two consecutive populations is less than a given tolerance. On termination, the best-performing solution (i.e. the best individual) in the latest population and the corresponding objective function value are reported as the optimal solution to the design problem (provided all the constraints are met, otherwise the calculations are repeated using larger penalties). Implementation of the methodology is illustrated in a case study in the next section.

4 Case study

This section demonstrates the capabilities of the novel design approach.

4.1 Problem description

The case study is based on that presented by Chen[9] and concerns the design of a crude oil distillation unit that separates 100,000 bbl/day (662.4 m^3 h^{-1}) of Venezuelan Tia Juana light crude oil[5] into five products, namely, light naphtha (LN), heavy naphtha

(HN), light distillate (LD), heavy distillate (HD) and residue (RES). The details of the crude oil assay are presented in Tables S1 and S2 in the supporting information.

The initial column design consists of a main column with three pump-arounds and three side-strippers, as shown in Figure 5. The main column has five sections (S-1 to S-5) with 5, 9, 10, 8 and 9 trays, respectively. The HD, LD and HN side-strippers have 5, 7 and 6 trays, respectively. The number of trays together with the utility demand is used to calculate the column annualized capital cost and operating cost, respectively, which are applied in Eq. (4) to estimate the total annualized cost. The operating conditions, product quality (in terms of ASTM 5% and 95% boiling temperature) and flow rates are presented in Tables S3 and S4 in the supporting information. The column operates at a uniform pressure of 2.5 bar.

Sieve trays are assumed in the column hydraulic calculations. In all sections of the main column, four passes per tray are used, while two pass trays are used in the side-strippers. The main column and all side-strippers are sized based on an 85% approach to jet flooding, 50% approach to downcomer and a tray spacing of 0.609 meters.

The economic evaluation assumes an interest rate of 5%, a plant life of 20 years and 8,700 operating hours per year in line with common practice (Maples, 2000). The initial column cost and costs of utilities are presented in Tables S5 and S6 in the supporting information. A minimum approach temperature of 30 ºC is used to calculate minimum utility requirements. The minimum approach temperature corresponds to typical values used in the design of industrial crude oil distillation unit (Smith et al., 2010).

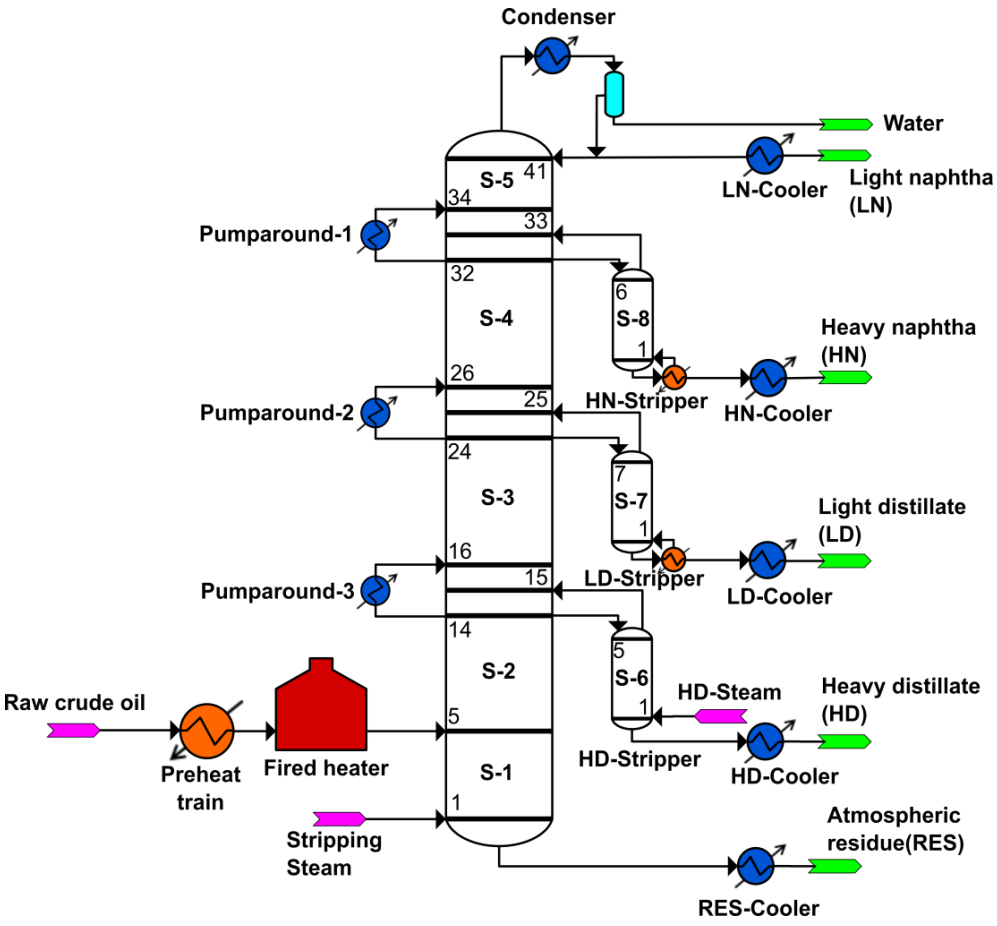


Figure 5 Case study: crude oil distillation unit design.

4.2 Initialization of the column optimization

The pre-optimization step starts by building the column superstructure that embeds all design alternatives using the approach presented in Section 3.3. The proposed superstructure for this case contains of 6, 10, 11, 9 and 10 trays in Sections S-1 to S-5 of the main column, and 6, 8 and 7 in the HD, LD and HN side-strippers respectively. Table 1a shows the lower and upper bounds on the number of trays in each column section.

Table 1a Bounds on numbers of trays per column section

Section	Lower bound	Upper bound	Initial value
1	3	6	5
2	7	10	9
3	8	11	10
4	6	9	8
5	7	10	9
6	3	6	5
7	5	8	7
8	4	7	6

The lower bounds indicate the minimum number of active trays in each column section, and the difference between the upper and lower bound indicates the maximum number of temporary (inactive) trays, the existence of which is determined by the optimization algorithm. To complete the definition of the search space, bounds are defined for the operating conditions in Table 1b.

Table 1b Bounds on operating conditions

Operating condition	Lower bound	Upper bound	Initial value
PA 1 duty (MW)	8.40	14.00	11.2
PA 2 duty (MW)	13.42	22.36	17.89
PA 3 duty (MW)	9.63	16.05	12.84
PA 1 DT (ºC)	10	30	20
PA 2 DT (ºC)	40	60	50
PA 3 DT (ºC)	20	40	30
Main stripping steam (kmol h^{-1})	900	1500	1200
HD stripping steam (kmol h^{-1})	188	313	250
Feed temperature (ºC)	340	375	365
Reflux ratio	3.17	6.17	4.17

The product quality specifications that must be satisfied are shown in Table S4 in the supporting information. An allowable range of 10 ºC is set for each product specification. Note that the initial design must be 'feasible'. This implies two things: the

solution must meet the product specifications and the associated simulation must in turn converge in Aspen HYSYS v8.6.

4.3 Optimization parameters

The optimization aims to identify a column with the best structure and operating conditions, corresponding to the minimum total annualized cost. Due to the stochastic nature of the genetic algorithm, each optimization run is likely to reach a different solution. In order to search for the best solution obtainable, the genetic algorithm is run ten times consecutively; each run is carried out using different initial population of chromosomes.

The MINLP problem formulated in Section 3.4 is coded and solved in Matlab R2015a using the genetic algorithm implemented in the Global Optimization Toolboox, on a HP desktop PC with Intel(R) Core i5 processor running at 3.2GHz, and 8 GB of RAM (see Section 3.5.1 for detailed discussion on genetic algorithm) . The initial population contains 100 chromosomes and the maximum number of generations is set to 200. Note that the initial population is randomly created, considering the lower and upper bounds specified (see Tables 1a and 1b). The initial population represents the initial guess used by the genetic algorithm. These parameter values were determined by running the genetic algorithm multiple times to finally select those parameter values that represent a good compromise between computational effort and quality of the final solution. Details of the computational results for the multiple runs (i.e. 10 consecutive runs) of the genetic algorithm are summarized in Tables S7 and S8 of the supporting information. The optimization time ranges between 3 to 6.5 hours.

4.4 Optimization results

4.4.1 Case 1: CDU design without constraint on product flow rate

The best results found for the crude oil distillation unit are summarized in Figure 6 and Tables 2a to 2c.

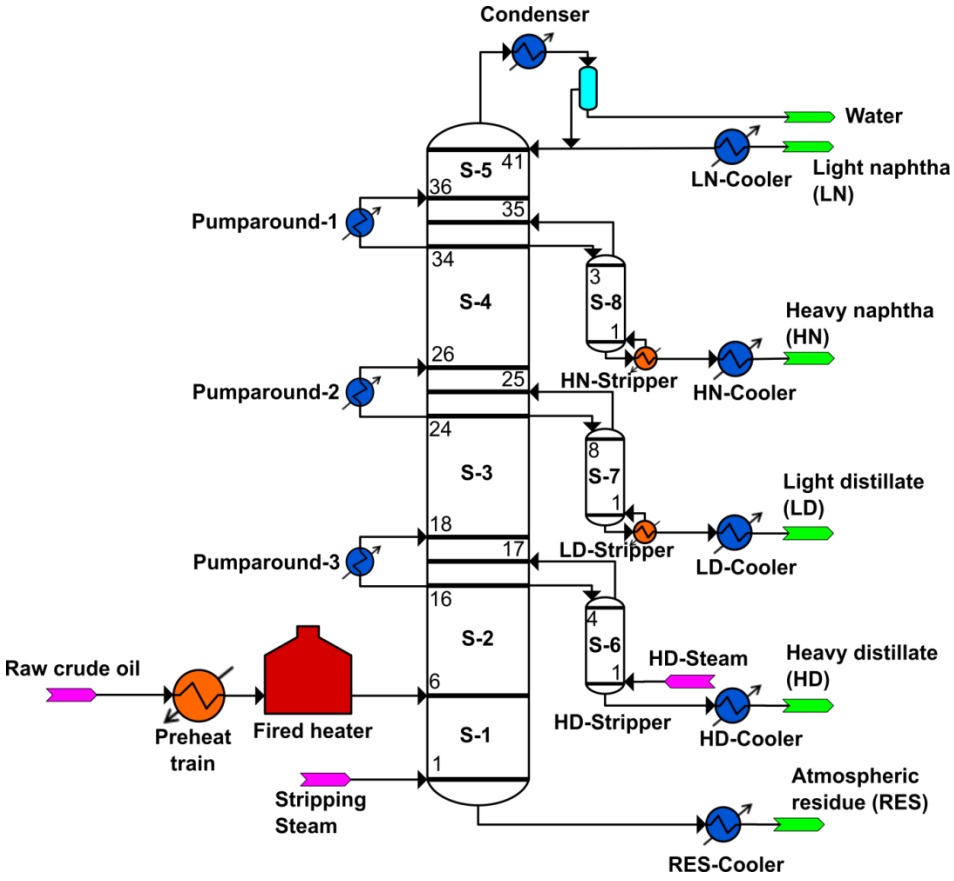


Figure 6 Optimal configuration of crude oil distillation unit (Case 1)

The column contains 6, 10, 8, customized 10 and 7 trays in the five sections of the main column, and 4, 8 and 3 trays in HD, LD and HN side-strippers respectively. The minimum utility targets for the optimal column, as calculated by the grand composite curve, are 40.35 MW of hot utility and 41.71 of cooling utility (respectively). This corresponds to a total utility cost of \$6.3 million per annum (\$MM y^{-1}). It should be noted that the utility targets reflect the minimum amount of utility required by the column from a thermodynamic viewpoint, without taking into account heat exchanger network details. For a more detailed analysis, a heat exchanger network model will be required to replace the grand compose curve in the optimization framework. Nevertheless, the stream data for the optimal column presented in Table S9 of the supporting information can be used to design the heat exchanger network for the column. The column structure together with the steam and utility requirement lead to a total annualised cost of 7.84 \$MM y^{-1}.

Table 2a CDU operating conditions (Case 1)

Variable	Initial value	Case 1
PA 1 duty (MW)	11.2	8.41
PA 2 duty (MW)	17.89	13.44
PA 3 duty (MW)	12.84	10.25
PA 1 DT (℃)	20	29
PA 2 DT (℃)	50	55
PA 3 DT (℃)	30	34
Main steam (kmol h^{-1})	1200	900
HD steam (kmol h^{-1})	250	188
Feed temperature (°C)	365	340
Reflux ratio	4.17	3.17

Table 2b Product quality and flow rate (Case 1)

Products	T5% (ºC, ASTM D86)		T95% (ºC, ASTM D86)		Flow rates ($m^3 h^{-1}$)	
	Initial value	Case 1	Initial value	Case 1	Initial value	Case 1
LN	26	24	111	111	103.5	98.8
HN	139	129	187	187	78.2	86.6
LD	216	217	301	301	140.3	130.5
HD	311	304	354	354	48.1	45.3
RES	361	352	754	752	292.5	301.4

Table 2c Utility demand and column cost (Case 1)

Variable	Initial value	Case 1	Units*
Utility requirements			
Hot utility	54.61	40.35	MW
Cold utility	61.18	41.71	MW
Cost			
Utility cost	8.51	6.27	\$MM y^{-1}
Steam cost	1.77	1.33	\$MM y^{-1}
Total operating cost	10.28	7.60	\$MM y^{-1}
Annualised capital cost	0.33	0.25	\$MM y^{-1}
Total annualised cost	**10.61**	**7.84**	**\$MM y^{-1}**

*\$MM y^{-1} denotes millions of dollars per annum

As shown in Table 2b, all the 5% and 95% ASTM boiling temperatures are within 10 ºC of the specified values; that is, no constraints on product quality are violated. On the

other hand, the flow rates of the five products have varied compared to the initial values. The most significant change is the increase of the atmospheric residue (RES) flow rate, representing a loss of valuable products from the atmospheric crude oil distillation unit. Such a large change could be avoided by adding constraints on the product flowrates and/or including the value of products in the objective function (i.e., maximizing profit instead of minimizing cost). Section 4.4.2 explores the use of constraints to restrict the flow rate of the atmospheric residue to within realistic limits.

4.4.2 Case 2: CDU design with constraint on product flow rate

Figure 7 and Tables 3a to 3c present the best results for the crude oil distillation unit design including constraint on product flow rates. A summary of the computational results (multiple runs) for this case is presented in Table S8 of the supporting information.

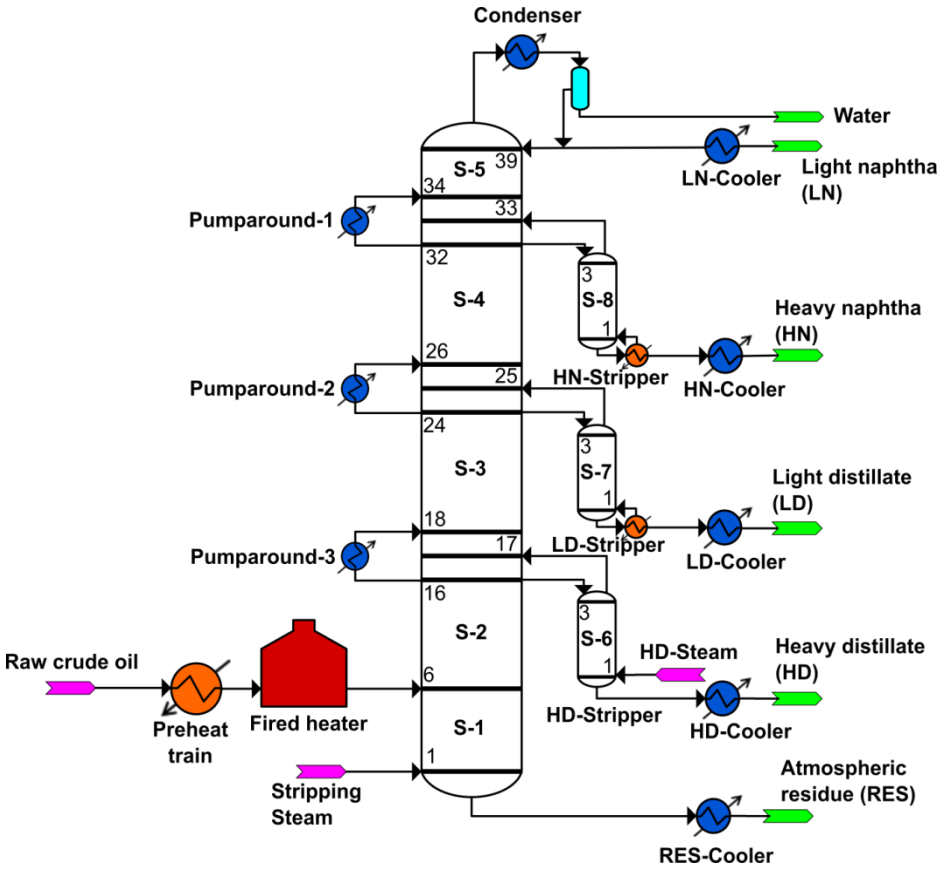


Figure 7 Optimal configuration of crude oil distillation unit (Case 2)

The total number of trays in the crude oil distillation unit is 48: 39 in the main column and 3 in each side-stripper. The stripping steam requirements for Case 2 are the same as in Case 1. However, the hot and cold utility requirements have increased by 8.9% and 9.1% respectively. This results from the fact that in Case 2, the crude oil feed needs heating at higher temperature (361°C) compare with Case 1 (340°C). The heating is required to vaporize more valuable products from the crude oil feed, which are recovered in the distillation unit. The additional heating is performed at the expense of increasing the fired heater duty, which eventually increases the total operating cost of the column. The total annualised cost of the new column structure is 8.47 \$MM y^{-1}.

Table 3a CDU operating conditions (Case 2)

Variable	Initial value	Case 2
PA 1 duty (MW)	11.2	9.58
PA 2 duty (MW)	17.89	16.19
PA 3 duty (MW)	12.84	15.20
PA 1 DT (ºC)	20	21
PA 2 DT (ºC)	50	43
PA 3 DT (ºC)	30	21
Main steam (kmol h^{-1})	1200	900
HD steam (kmol h^{-1})	250	188
Feed temperature (ºC)	365	361
Reflux ratio	4.17	3.39

Table 3b Product quality and flow rate (Case 2)

Products	T5% (ºC, ASTM D86)		T95% (ºC, ASTM D86)		Flow rates (m^3 h^{-1})	
	Initial value	Case 2	Initial value	Case 2	Initial value	Case 2
LN	25.9	24	110.6	111	103.5	99.9
HN	138.9	131	186.6	187	78.2	80.8
LD	215.9	213	301.5	301	140.3	135.1
HD	310.7	300	354.4	354	48.1	54.3
RES	361.4	361	754.3	754	292.5	292.5

Table 3c Utility demand and column cost (Case 2)

Variable	Initial value	Case 2	Units*
Utility requirements			
Hot utility	54.61	44.32	MW
Cold utility	61.18	45.90	MW
Cost analysis			
Utility cost	8.51	6.89	\$MM y^{-1}
Steam cost	1.77	1.33	\$MM y^{-1}
Total operating cost	10.28	8.22	\$MM y^{-1}
Annualised capital cost	0.33	0.26	\$MM y^{-1}
Total annualised cost	**10.61**	**8.47**	**\$MM y^{-1}**

*\$MM y^{-1} denotes millions of dollars

In Table 3b, the flow rate of the atmospheric residue (RES) remains unchanged, compared to the initial values. Therefore, no valuable product is lost from the atmospheric crude oil distillation unit. Furthermore, all of the product ASTM boiling temperatures are within 10 ºC of the specified values; again, no constraints on product quality are violated.

5 Conclusions

The design of a crude oil distillation unit is a nontrivial task due to the large number of degrees of freedom and complex interactions between individual units. Further challenges arise due to the need to design the complex column and heat recovery system simultaneously. Modified shortcut column models based on Fenske–Underwood–Gilliland have been applied for these purposes. However, the accuracy of these models can be low and they cannot be readily applied to any column configuration.

This article introduced a simulation-optimization framework for the design of crude oil distillation units that overcomes existing limitations, where maximum heat recovery is considered using pinch analysis. The approach presented combines a rigorous tray-by-tray model of a complex crude oil distillation unit implemented in

Aspen HYSYS v8.6 with a genetic algorithm coded in Matlab R2015a. The optimization of the superstructure of the crude oil distillation column naturally leads to an MINLP model. This MINLP, solved by a GA, optimizes the crude oil distillation unit configuration (i.e. feed tray location, pump-around and side-stripper locations, number of trays in each section of the column) along with its operating conditions (i.e. feed inlet temperature, pump-around duties and temperature drops, stripping steam flow rates and reflux ratio) simultaneously.

One important advantage of using stochastic optimizers is that they do not require derivatives and are designed to avoid getting trapped in locally optimal solutions (although they cannot guarantee global optimality within a given tolerance). Computational results for the case study demonstrate that the proposed framework is capable of identifying designs that significantly improve the starting solution. In addition, it was shown that product quality specifications can be met effectively using a suitable penalty function, although this may be at the expense of some valuable products slipping into the residue stream. This limitation can be overcome by constraining flow rates within allowable limits.

Future work will explore the use of surrogate models[48–51] of the distillation column in the optimization. Furthermore, to capture trade-offs between HEN capital cost and other costs, the framework needs to be extended to account for the synthesis of the heat exchanger network. Lastly, the proposed framework will be extended to retrofit of the complex crude oil distillation unit.

Associated content

*S Supporting Information

Data on crude oil properties and characterization; initial feasible design data; and results for optimization of the crude oil distillation unit presented in the case study. This information is available free of charge via the Internet at http:// pubs.acs.org/.

Author information

Corresponding Author

*E-mail: dauda.ibrahim@manchester.ac.uk

Notes

The authors declare no competing financial interest.

Acknowledgement

The authors would like to acknowledge the financial support from Petroleum Technology Development Fund (PTDF), Nigeria, for sponsoring this PhD research.

References

(1) Liebmann, K.; Dhole, V. R.; Jobson, M. Integrated Design of a Conventional Crude Oil Distillation Tower Using Pinch Analysis. *Chem. Eng. Res. Des.* **1998**, *76* (3), 335–347.

(2) Szklo, A.; Schaeffer, R. Fuel Specification, Energy Consumption and CO_2 Emission in Oil Refineries. *Energy* **2007**, *32* (7), 1075–1092.

(3) Gary, J. H.; Handwerk, G. E. *Petroleum Refining: Technology and Economics,* 4th ed.; Marcel Dekker: New York, 2001.

(4) Nelson, W. L. *Petroleum Refinery Engineering*; McGraw-Hill: New York, 1958.

(5) Watkins, R. N. *Petroleum Refinery Distillation*; Gulf Publishing Company, Book Division: TX, 1979.

(6) Jones, D. S. J. *Elements of Petroleum Processing*; John Wiley & Sons: West Sussex, UK, 1995.

(7) Suphanit, B. Design of Complex Distillation System, PhD Thesis, UMIST, Manchester, UK., 1999.

(8) Rastogi, V. Heat Integrated Crude Oil Distillation System Design., University of Manchester, UK., 2006.

(9) Chen, L. Heat-Integrated Crude Oil Distillation System Design. PhD Thesis, University of Manchester. Manchester, UK. **2008**.

(10) Caballero, J. A.; Milan-Yanez, D.; Grossmann, I. E. Rigorous Design of Distillation Columns: Integration of Disjunctive Programming and Process Simulators. *Ind. Eng. Chem. Res.* **2005**, *44* (17), 6760–6775.

(11) Brunet, R.; Cortés, D.; Guillén-Gosálbez, G.; Jiménez, L.; Boer, D. Minimization of the LCA Impact of Thermodynamic Cycles Using a Combined Simulation-Optimization Approach. *Appl. Therm. Eng.* **2012**, *48*, 367–377.

(12) Brunet, R.; Guillen-Gosalbez, G.; Perez-Correa, J. R.; Caballero, J. A.; Jimenez, L. Hybrid Simulation-Optimization Based Approach for the Optimal Design of Single-Product Biotechnological Processes. *Comput. Chem. Eng.* **2012**, *37*, 125–135.

(13) Morandin, M. Pinch Analysis cascade calculation, 2014. http://uk.mathworks.com/matlabcentral/fileexchange/47743-cascade-m (accessed Nov 20, 2014).

(14) Liebmann, K. Integrated Crude Oil Distillation Design. PhD Thesis, UMIST, Manchester, UK, 1996.

(15) Sharma, R.; Jindal, A.; Mandawala, D.; Jana, S. K. Design/Retrofit Targets of Pump-Around Refluxes for Better Energy Integration of a Crude Distillation Column. *Ind. Eng. Chem. Res.* **1999**, *38* (6), 2411–2417.

(16) Dhole, V. R.; Buckingham, P. R. Refinery Column Integration for Debottlenecking and Energy Saving. *Proc. ESCAPE IV Conf. Dublin* **1994**.

(17) Bagajewicz, M.; Ji, S. Rigorous Procedure for the Design of Conventional Atmospheric Crude Fractionation Units. Part I: Targeting. *Ind. Eng. Chem. Res.* **2001**, *40*, 617–626.

(18) Grossmann, I. E.; Caballero, J. A.; Yeomans, H. Advances in Mathematical Programming for Automated Design, Integration and Operation of Chemical Processes. *Korean J. Chem. Eng.* **1999**, *16*, 407–426.

(19) Grossmann, I. E.; Guillén-Gosálbez, G. Scope for the Application of Mathematical Programming Techniques in the Synthesis and Planning of Sustainable Processes. *Comput. Chem. Eng.* **2010**, *34* (9), 1365–1376.

(20) Dowling, A. W.; Biegler, L. T. A Framework for Efficient Large Scale Equation-Oriented Flowsheet Optimization. *Comput. Chem. Eng.* **2015**, *72*, 3–20.

(21) Skiborowski, M.; Rautenberg, M.; Marquardt, W. A Hybrid Evolutionary–Deterministic Optimization Approach for Conceptual Design. *Ind. Eng. Chem. Res.* **2015**, *54* (41), 10054–10072.

(22) Steimel, J.; Engell, S. Optimization-Based Support for Process Design under Uncertainty: A Case Study. *AIChE* **2016**, *62* (9), 3404–3419.

(23) Corbetta, M.; Grossmann, I. E.; Manenti, F. Process Simulator-Based Optimization of Biorefinery Downstream Processes under the Generalized Disjunctive Programming Framework. *Comput. Chem. Eng.* **2016**, *88*, 73–85.

(24) Javaloyes-Antón, J.; Ruiz-Femenia, R.; Caballero, J. a. Rigorous Design of Complex Distillation Columns Using Process Simulators and the Particle Swarm Optimization Algorithm. *Ind. Eng. Chem. Res.* **2013**, *52* (44), 15621–15634.

(25) Sinnott, R. K. Coulson & Richardson's Chemical Engineering Volume 6: Chemical Engineering Design. *Book* **2013**, *6*, 1038.

(26) Yee, T. F.; Gnossmann, I. E. Simultaneous Optimization Models for Heat Integration - II. Heat-Exchanger Network Synthesis. *Comput. Chem. Eng.* **1990**, *14* (10), 1165–1184.

(27) Seader, J. D.; Henley, E. J.; Roper, D. K. *Separation Process Principles, 3rd Edition*; John Wiley & Sons: New York, 2010.

(28) Yeomans, H.; Grossmann, I. E. Optimal Design of Complex Distillation Columns Using Rigorous Tray-by-Tray Disjunctive Programming Models. *Ind. Eng. Chem. Res.* **2000**, *39* (11), 4326–4335.

(29) Smith, R. *Chemical Process: Design and Integration*; Wiley: Chichester, UK, 2005.

(30) Guthrie, K. M. Data and Techniques for Preliminary Capital Cost Estimating. *Chem. Eng.* **1969**.

(31) Economic Indicator 2012. www.che.com (accessed Nov 20, 2014).

(32) Economic Indicator 2014. www.che.com (accessed Nov 20, 2014).

(33) Stichlmair, J.; Fair, J. R. *Distillation: Principles and Practices*; A Wiley-Liss publication; Wiley-VCH: New York, 1998.

(34) Kister, H. *Distillation Design*; McGraw-Hill chemical engineering series; McGraw-Hill Education: New York, 1992.

(35) Branan, C. R. *Rules of Thumb for Chemical Engineers*; Gulf Professional Publishing: TX, 2005.

(36) Edgar, T. F.; Himmelblau, D. M.; Lasdon, L. S. *Optimization of Chemical Processes*; McGraw-Hill: New York, 2001.

(37) Floudas, C. A. *Deterministic Global Optimization: Theory, Methods and Applications*; Nonconvex Optimization and Its Applications; Springer: US, 2013.

(38) Biegler, L. T.; Grossmann, I. E.; Westerberg, A. W. *Systematic Methods of Chemical Process Design*; Prentice Hall PTR: NJ, 1997.

(39) Duran, M. A.; Grossmann, I. E. An Outer-Approximation Algorithm for a Class of Mixed-Integer Nonlinear Programs. *Math. Program.* **1986**, *36* (3), 307–339.

(40) Heyman, D. P.; Sobel, M. J. *Stochastic Models in Operations Research: Stochastic Optimization*; Dover Publications: New York, 2003.

(41) Floudas, C. A. *Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications*; Topics in Chemical Engineering; Oxford University Press, 1995.

(42) Odjo, A. O.; Jr., N. E. S.; Yuan, W.; Marcilla, A.; Eden, M. R.; Caballero, J. A. Disjunctive-Genetic Programming Approach to Synthesis of Process Networks. *Ind. Eng. Chem. Res.* **2011**, *50* (10), 6213–6228.

(43) Fiandaca, G.; Fraga, E. S. A Multi-Objective Genetic Algorithm for the Design of Pressure Swing Adsorption. *Eng. Optim.* **2009**, *10*, 1–24.

(44) Vazquez-Castillo, J. A.; Venegas-Sánchez, J. A.; Segovia-Hernández, J. G.; Hernández-Escoto, H.; Hernández, S.; Gutiérrez-Antonio, C.; Briones-Ramírez, A. Design and Optimization, Using Genetic Algorithms, of Intensified Distillation Systems for a Class of Quaternary Mixtures. *Comput. Chem. Eng.* **2009**, *33* (11), 1841–1850.

(45) Lee, E. S. Q.; Rangaiah, G. P. Optimization of Recovery Processes for Multiple Economic and Environmental Objectives. *Ind. Eng. Chem. Res.* **2009**, *48* (16), 7662–7681.

(46) Deep, K.; Singh, K. P.; Kansal, M. L.; Mohan, C. A Real Coded Genetic Algorithm for Solving Integer and Mixed Integer Optimization Problems. *Appl. Math. Comput.* **2009**, *212* (2), 505–518.

(47) Mitchell, M. An Introduction to Genetic Algorithms: MIT Press: Cambridge, US, **1998**, 1–40.

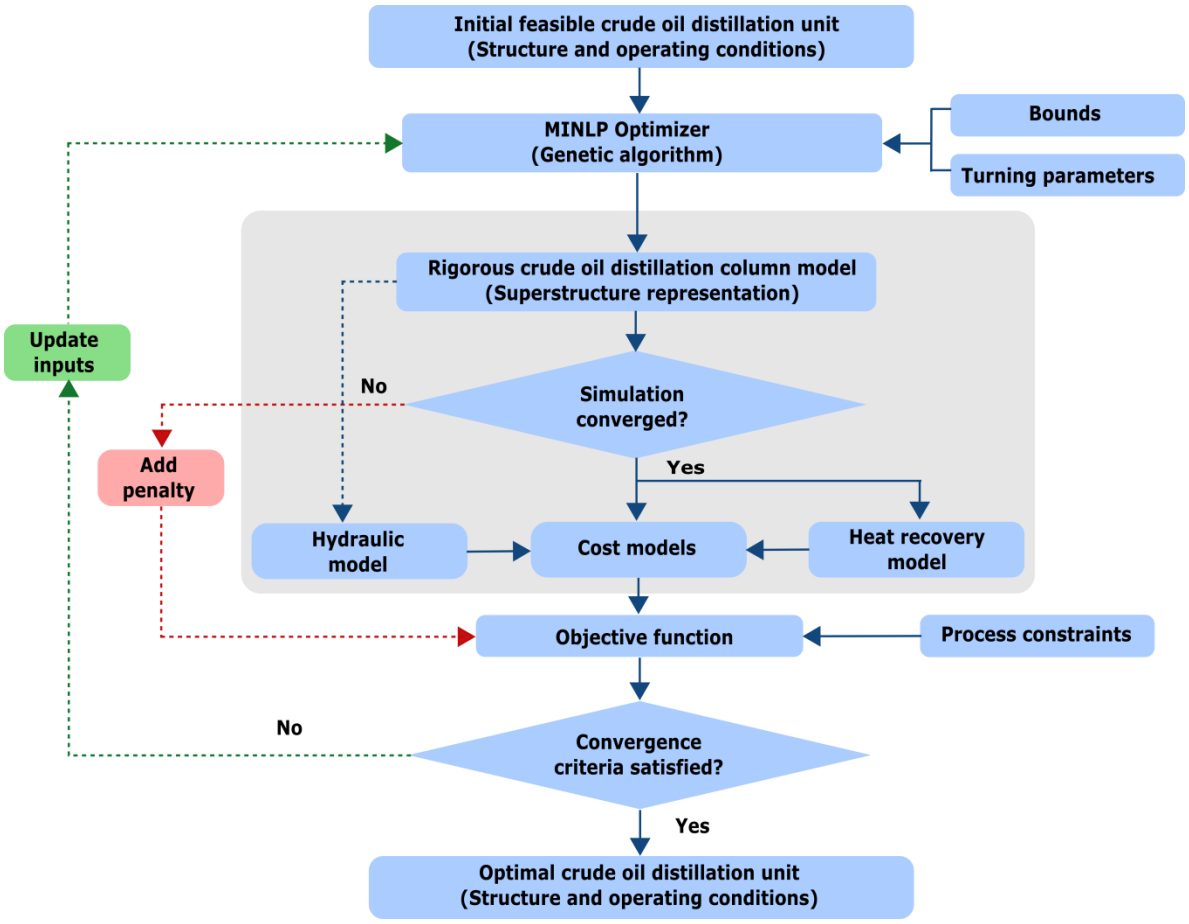
(48) Ochoa-Estopier, L. M.; Jobson, M.; Smith, R. Operational Optimization of Crude Oil Distillation Systems Using Artificial Neural Networks. *Comput. Chem. Eng.* **2013**, *59*, 178–185.

(49) Ochoa-Estopier, L. M.; Jobson, M. Optimization of Heat-Integrated Crude Oil Distillation Systems. Part I: The Distillation Model. *Ind. Eng. Chem. Res.* **2015**, *54* (18), 4988–5000.

(50) Quirante, N.; Javaloyes, J.; Caballero, J. A. Rigorous Design of Distillation Columns Using Surrogate Models Based on Kriging Interpolation. *AIChE* **2015**, *61* (7), 2169–2187.

(51) López, D. C.; Hoyos, L. J.; Mahecha, C. A.; Arellano-Garcia, H.; Wozny, G. Optimization Model of Crude Oil Distillation Units for Optimal Crude Oil Blending and Operating Conditions. *Ind. Eng. Chem. Res.* **2013**, *52* (36), 12993–13005.

Table of contents (TOC) graphic



Chapter 4 Design of heat-integrated crude oil distillation systems using surrogate models

Chapter 3 presents a new approach for modelling crude oil distillation units using rigorous simulation models, taking into account both structural variables and operating conditions of the unit. The rigorous model proposed in Chapter 3 can give accurate predictions of the column performance; however, large computational times are required to find the optimal column structure and its operating conditions. This large computational time makes rigorous column models unsuitable for design problems that involve a large number of operating scenarios, for example, flexible crude oil distillation system that processes multiple crudes. To improve computational time, without compromising solution accuracy greatly, surrogate models can be used, where there are derived from the results of rigorous simulation models.

This chapter addresses the third objective of this research, i.e., to adapt the design methodology of Chapter 3 to apply surrogate distillation column models, considering both column performance and definition of region of appropriate operating conditions.

4.1 Introduction to Publication 2

This paper presents a new modelling approach for crude oil distillation systems using surrogate models. The main advantage of the modelling approach presented here compared with those in literature (see Chapter 2) is that both structural and operational variables of the crude oil distillation unit are taken into account (and heat recovery). The surrogate modelling technique used is artificial neural networks due to their ability to capture complex non-linear relationships between input-output data (See Chapter 2). A comprehensive discussion on artificial neural network fundamentals and application can be found in Chapter 2, Sections 2.2.3.1 and 2.2.3.2.

In this work, the parameters of the artificial neural network (weight and bias) are regressed using data generated via multiple rigorous simulations, set up in a commercial process simulator (Aspen HYSYS). Firstly, Latin hypercube (Wang et al., 2004) samples are generated for the relevant input variables (number of trays in column section, pump-around temperature drop and duty, stripping steam flow rate, feed inlet temperature and reflux ratio) using the function "*ihsdesign*" in MatLab (MATLAB, 2014). Through an interface (AspenTech, 2011) established between Aspen HYSYS and MatLab, all the samples are simulated on the rigorous model and their corresponding outputs (product quality [in terms of T5% and T95%], product flow rate, stream supply, target and enthalpy change, and column diameter) are collected and recorded. Due to the combinatorial nature of data sampling, not all the samples could lead to a converged simulation on the rigorous model. In this work, the maximum number of iteration is set to 20 (in the process simulator), therefore any sample that fails to converge after 20 iterations is deemed infeasible or unconverged sample. It should be noted that convergence can be improved by allowing for large number of iteration, better starting point, and repeating simulation multiple times with the same sample. However, the overall sampling time could increase significantly.

In this work, only the converged samples were used to train the artificial neural network, as they represent feasible points within the design space. To prevent the optimisation algorithm from converging to an infeasible solution, it is necessary to remove the unconverged samples from the design space. This work developed a feasibility constraint using a support vector machine. Here, the feasibility constraints are equality constraints that are applied to remove infeasible solutions from the design space, thus increasing the likelihood that the optimal solution would be feasible when simulated on a rigorous model, and also computational effort can be reduced significantly. The support vector machine is trained using the entire samples (i.e. both converged and unconverged samples) generated from the multiple rigorous simulation.

The surrogate models of the crude oil distillation unit are implemented in an optimisation framework, together with pinch analysis, in order to search for cost-

effective solutions. The approach is applied to a relevant case study. Results indicate that cost-effective crude oil distillation unit can be identified using the proposed method at a significantly reduce computational time compared to when rigorous simulation model is used (see Chapter 3). The supporting information for this paper is presented in Appendix A.2. Chapter 5 extends the surrogate model to design a flexible heat-integrated crude oil distillation system.

4.2 Publication 2

Ibrahim, D., Jobson, M., Li, J., Guillén-Gosálbez, G., 2018. Optimization-based Design of Crude Oil Distillation Units using Surrogate Models and a Support Vector Machine. *Chem. Eng. Res. Des.*, **2018**, **DOI:** doi.org/10.1016/j.cherd.2018.03.006.

Optimization-based Design of Crude Oil Distillation Units using Surrogate Column Models and a Support Vector Machine

Dauda Ibrahim[1,*], Megan Jobson[1], Jie Li[1], Gonzalo Guillén-Gosálbez[2]

[1]**Centre for Process Integration, School of Chemical Engineering and Analytical Science, University of Manchester, Manchester M13 9PL, UK**

[2]**Department of Chemical Engineering, Centre for Process Systems Engineering, Imperial College, South Kensington Campus, London SW7 2AZ, UK**

Abstract

This paper presents a novel optimization-based approach for the design of heat-integrated crude oil distillation units, which are widely used in refineries. The methodology presented combines, within a unified framework, surrogate distillation column models based on artificial neural networks, feasibility constraints constructed using a support vector machine, and pinch analysis to maximize heat recovery, in order to optimize the distillation column configuration and its operating conditions. The inputs to the surrogate column model are given by the column structure and operating conditions, while the outputs are related to the column performance. The support vector machine classifier filters infeasible design alternatives from the search space, thus reducing computational time, and ultimately improves the quality of the final solution. The overall optimization problem takes the form of a mixed-integer nonlinear program, which is solved by a genetic algorithm that seeks the design and operating variables values that minimize the total annualized cost. The capabilities of the proposed approach are illustrated using an industrially–relevant case study. Numerical results show that promising design alternatives can be obtained using the proposed method. The approach can help engineers to design and operate petroleum

[*] Corresponding author
E-mail address: dauda.ibrahim@manchester.ac.uk; didauda@yahoo.com

refineries optimally, where these are expected to continue to play a major role in the energy mix for some years.

Keywords: Process design, atmospheric distillation unit, heat integration, genetic algorithm, artificial neural network

1 Introduction

1.1 Crude oil distillation: technical background

Distillation is the most widely used separation technique in the chemical and petroleum industries. Petroleum refining starts with crude oil distillation, in which the entire crude oil feedstock undergoes initial separation to produce intermediate products, such as light naphtha, heavy naphtha, light distillate, heavy distillate, and atmospheric residue (see Figure 1). These cuts are further enhanced and blended into marketable products (e.g. gasoline, kerosene, jet fuel, diesel, bunker fuel and fuel oil etc.) that are supplied to the global energy market.

Crude oil distillation is a capital- and energy-intensive process and is the largest consumer of energy in petroleum refineries (Gu et al., 2014). The energy consumed in crude oil distillation is equivalent to 1 to 2% of the total crude oil being processed (Szklo and Schaeffer, 2007). Consequently, it contributes significantly to the overall refinery CO_2 emissions. Heat integration is typically applied to enhance energy efficiency in crude oil distillation by recovering heat from 'hot' streams requiring cooling to 'cold' streams requiring heating, thus reducing demand for fired heating, and thus both greenhouse gas emissions and operating cost.

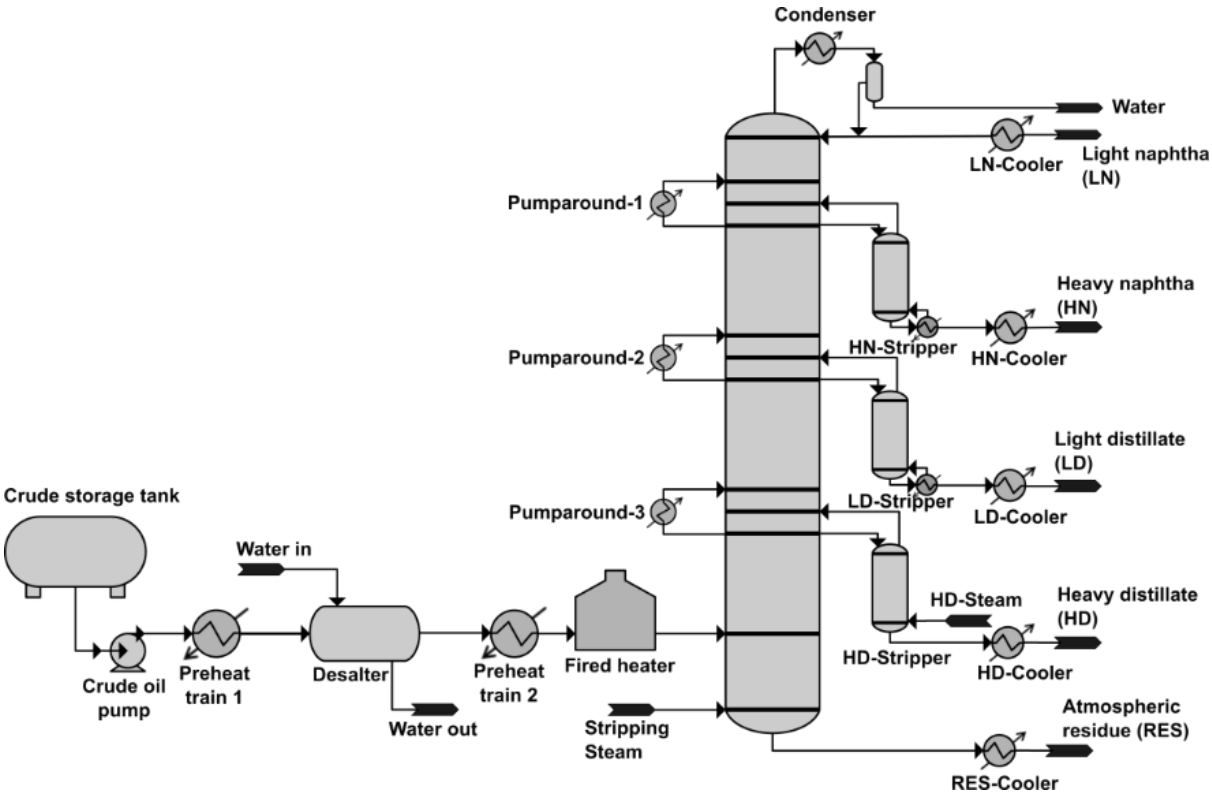


Figure 1 – Typical crude oil distillation system

Figure 1 illustrates a typical crude oil distillation system, which comprises a crude oil distillation unit, a heat recovery network (preheat train) and a fired heater (furnace) where the crude oil feedstock is heated and partially vaporized. The crude oil distillation unit is equipped with side-strippers and pump-arounds. Side-strippers are utilized to remove light component from side draws using stripping steam or reboilers, while pump-around loops provide internal reflux and create heat recovery opportunities by cooling and returning liquid streams withdrawn from the column. The crude oil distillation unit is strongly interlinked with the associated heat recovery network through pump-arounds, the condenser and product coolers. Changes in the design and operation of the crude oil distillation unit, therefore, affect the design and operation of the heat exchanger network and fired heater.

A new ('grassroots') design of a crude oil distillation unit aims to determine the optimal values of the design degrees of freedom, which include structural variables and operating conditions while taking into account the complex interactions between the column and the heat recovery network. The structural variables include the locations of the feed tray and of pump-around and side-stripper draw streams, and the number of trays in each section of the column. Operating conditions include the feed inlet temperature, pump-around duties and temperature drop, stripping steam flow rates and reflux ratio. The large numbers of degrees of freedom, complex interactions between individual units, and the need to design the column while accounting for heat recovery makes the design of crude oil distillation units highly challenging.

1.2 Crude oil distillation column modelling

Distillation column models may be employed to optimize both the column structure and its operating conditions with respect to various performance criteria. These models should be sufficiently realistic to provide meaningful solutions and robust enough to converge over a wide optimization search space.

Existing design methods (Bagajewicz and Ji, 2001; Liebmann et al., 1998) applied a sequential approach that combines rigorous simulation models embedded in commercial process simulator and pinch analysis to support the search for energy efficient column structure and its operating conditions. However, these approaches require trial and error before arriving at the final design, and trade-off between capital and energy cost has not been considered. To overcome these limitations, an approach the combine rigorous tray-by-tray model of the distillation column and pinch analysis in an optimization framework has been developed (Ibrahim et al., 2017a). In this approach, a superstructure of the crude oil distillation column (comprising many design alternatives) is built in Aspen HYSYS, while the optimization is carried out using an external solver in MatLab. The approach takes advantage of the physical property and thermodynamic model as well as the crude oil characterization and column hydraulic models available in the process simulator to generate an accurate

estimate of the crude oil distillation unit performance. Optimization variables include number of trays in column sections and operating conditions.

Despite the fact that rigorous models are versatile and produce accurate estimates of the distillation column performance, like most models, they require good initial points to converge to a feasible solution. Furthermore, from the optimization point of view, incorporating a rigorous tray-by-tray column model in an optimization framework so as to design the complex crude oil distillation unit is computationally intensive. To overcome these limitations, this work proposes a new strategy for the design and optimization of crude oil distillation units using surrogate column models. As will be seen in Section 4 of this article, the surrogate column model developed herein leads to significant reduction in computational time without sacrificing the model accuracy.

1.3 Surrogate modelling of crude oil distillation columns

In recent years, various regression and fitting techniques have been used to create surrogate models using different mathematical techniques, including polynomial regressions, artificial neural networks, and support vector regressions. Liau et al. (2004) and Motlaghi et al. (2008) applied artificial neural networks to build surrogate models of the crude oil distillation column using data collected from an existing plant. The models were used to perform operational optimization. Liau et al. (2004) focused on improving the yield of kerosene, diesel and atmospheric gas oil, while Motlaghi et al. (2008) optimized the flow rate of products according to their market values.

Yao and Chu (2012) developed a surrogate model of the crude oil distillation using the concept of support vector regressions. The model was implemented in a framework to optimize profit by varying operating variables. Gueddar and Dua (2012) applied artificial neural network to construct a reduced model of the crude oil distillation unit that is suitable for refinery-wide optimization. The model inputs include crude oil properties (e.g. true boiling point) and flow rate, while the outputs consist of refine product yield and their specifications. A deterministic optimization is used to search for the best inputs that improves energy efficiency. López C. et al. (2013) applied a

polynomial function to build models of a crude oil distillation system. The models, together with energy balances representing the heat exchanger network, were implemented in a framework to maximize net profit.

Ochoa-Estopier and Jobson (2015) applied data from multiple rigorous simulations to build a surrogate model of the crude distillation system using artificial neural networks. The column models, together with a heat exchanger network model, were implemented in an operational optimization framework to improve net profit while fulfilling practical constraints. More recently, Osuolale and Zhang (2017) applied bootstrap artificial neural networks to build a model of a crude oil distillation system, consisting of pre-fractionator, atmospheric column, and vacuum column. Sequential quadratic programming is used to optimize exergy efficiency by varying relevant operating conditions, where heat recovery opportunities are not explicitly considered.

1.4 Scope and objectives

The surrogate models for crude oil distillation columns presented so far assume that the design variables are fixed. Furthermore, these methodologies typically deal with continuous rather than discrete variables. This work extends the use of surrogate models to account for discrete design variables representing the number of trays in each column section (which cannot be handled via the approaches mentioned above). This work uses artificial neural networks, as these have been shown to provide simple, accurate and robust simulation models.

This work proposes a new methodology for the design of heat-integrated crude oil distillation units that implement a surrogate model based on artificial neural networks. Unlike previous work applying surrogate models to represent a crude oil distillation unit (Ochoa-Estopier and Jobson, 2015), the proposed approach takes into account both discrete and continuous variables. The artificial neural network is constructed using 'samples' that are results of multiple rigorous simulations.

Due to the highly combinatorial nature of the data sampling, not all samples within the search space are likely to lead to a converged simulation (i.e. a feasible design) using the rigorous model. In addition, numerical issues, e.g. poor initialization, can lead to non-convergence of rigorous simulations. In this work, a feasibility constraint is constructed using a support vector machine (Vapnik, 1995): the resulting classification model helps to restrict the search space to the region of feasible designs, thus avoiding the need to search in the region of infeasible designs during optimization. Restricting the search in this way increases the likelihood that the optimal design will feasible when modelled rigorously. Jobson et al. (2017) and Ibrahim et al. (2017b) provide fuller discussions of feasibility constraints in the context of surrogate modelling of crude oil distillation systems.

In this work, the artificial neural network column model and feasibility constraint constructed using a support vector machine are implemented in an optimization framework, in which a genetic algorithm is used to search for column structural variables and operating conditions that minimizes total annualized cost. Heat integration is taken into account using pinch analysis, without considering the detailed heat exchanger network structure. Figure 2 summarizes the overall design framework.

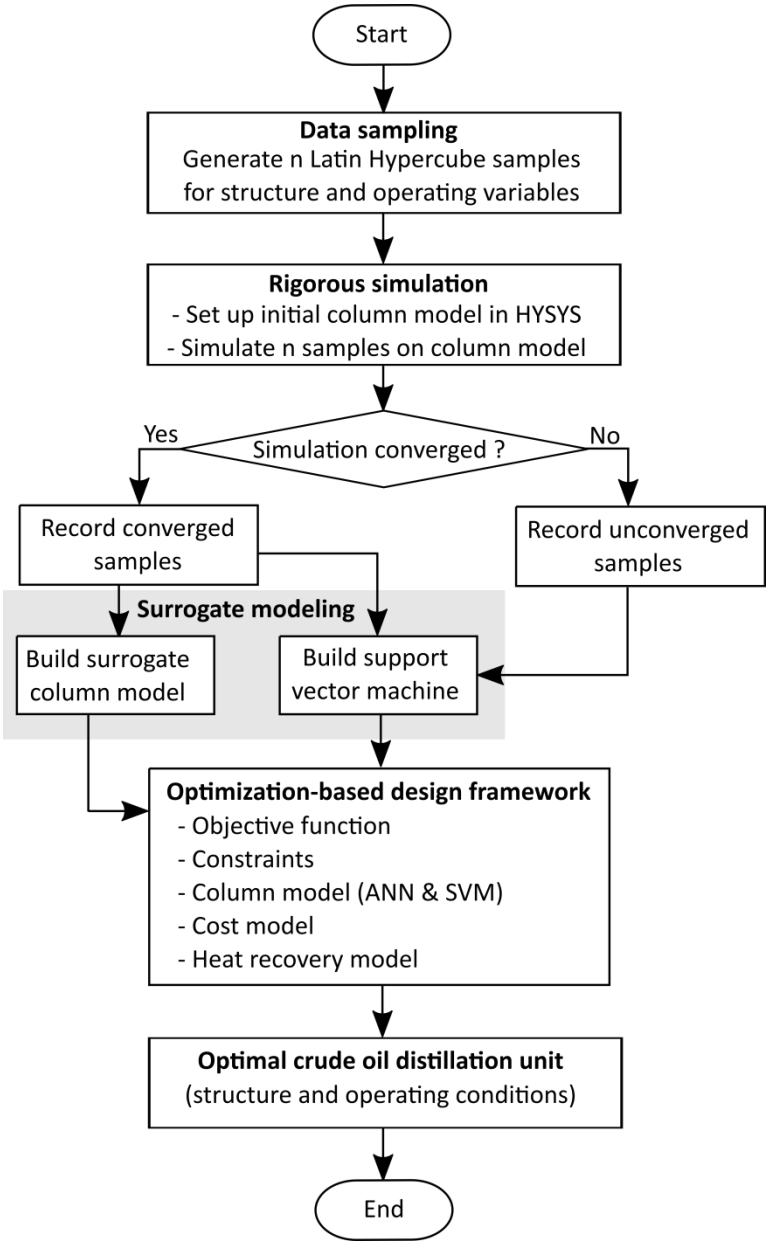


Figure 2 – Overview of design framework

The remainder of this paper is organized as follows. Section 2 presents an overview of the modeling tools adopted in this work, including the artificial neural network and the support vector machine. In Section 3, a detailed stepwise approach for modeling the complex crude oil distillation unit is developed. Section 4 presents the proposed mathematical formulation of the design problem together with a solution procedure. Section 5 illustrates and reviews the capabilities of the novel design methodology using a case study. Lastly, conclusions are presented in Section 6.

2 General background on surrogate modeling and support vector machine

The modeling tools applied to model the crude oil distillation column include an artificial neural network and a support vector machine. As mentioned in Section 1, the artificial neural network is used to build a surrogate model of the crude oil distillation column that is subsequently implemented in an optimization framework, while the support vector machine filters out infeasible design alternatives during optimization, thus increasing the likelihood that the design alternatives proposed by the genetic algorithm will be feasible in the process simulator. A detailed description of the two modeling tools follows.

2.1 Artificial neural networks

Artificial neural networks (ANNs) are computational data-driven modeling systems that can represent complex nonlinear relationships between process input-output data points. The network consists of highly interconnected simple elements called neurons. The neurons are composed into at least two or more layers as shown in Figure 2.

Various types of artificial neural network architectures (Basheer and Hajmeer, 2000) have been applied to model process systems of different levels of complexity (Henao and Maravelias, 2010; Gueddar and Dua, 2011; Nuchitprasittichai and Cremaschi, 2013; Ochoa-Estopier and Jobson, 2015). A multi-layer feedforward network is the most widely acceptable architecture due to its mathematical simplicity. This feature makes it suitable for implementation in an optimization framework, where the reduction of the mathematical complexity is an important factor (Nuchitprasittichai and Cremaschi, 2012). Figure 3 shows a typical multi-layer feed forward network comprising two layers, namely, a hidden layer and an output layer. The hidden layer may consist of one or more sub-layers, although one sub-layer is a commonplace. Established literature (Henao and Maravelias, 2010) have shown that multi-layer feedforward networks with one hidden layer are capable of approximating arbitrary multivariate functions with a finite number of discontinuities; these are features of the problem of interest.

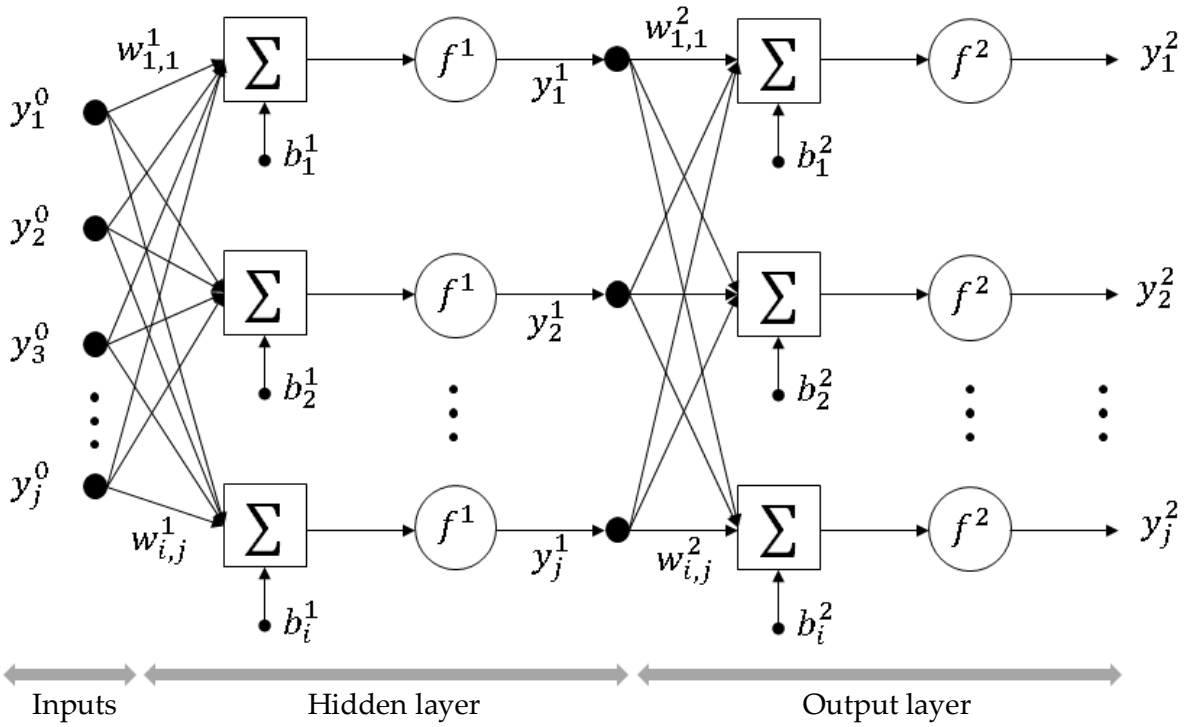


Figure 3 – Schematic representation of a multi-layer feed forward neural network (adapted from Beale et al. (2015))

In Figure 3, the superscript and subscript denote the indices of the network layer and neuron respectively. Mathematically, the multi-layer feedforward network in Figure 3 can be formulated as (Beale et al., 2015)

$$\boldsymbol{y} = f^2(\boldsymbol{W}^2 f^1(\boldsymbol{W}^1 \boldsymbol{y}^0 + \boldsymbol{b}^1) + \boldsymbol{b}^2) \tag{1}$$

where $\boldsymbol{y}^0$ and $\boldsymbol{y}$ represent the vectors of inputs (independent) and outputs (dependent) variables; $\boldsymbol{b}$ denotes the vector of biases and $\boldsymbol{W}$ denotes the matrix of weights. f^1 and f^2 are transfer functions of the hidden and output layers, respectively. In this case, the hidden and output layers are represented by a sigmoid function and a linear function, respectively (Beale et al., 2015).

The multi-layer feedforward network is trained to input-output data points using training algorithms. Different training algorithms are available (Beale et al., 2015). Backpropagation is the most commonly used method, which is the one used in this work. Backpropagation consists of two steps. First the prediction error, e.g., mean

square error, is computed using fixed value of weights and biases. Second, the weights and biases are adjusted to minimize the prediction error. The objective function used in most training algorithm can be defined as (Beale et al., 2015):

$$MSE = \sum_{i=1}^{N} \frac{(t_i - y_i)^2}{N} \tag{2}$$

where MSE denotes the mean square error, N is the total number of sample points, t and y denote the target output and predicted output, respectively. Section 3.2 presents further detail on the development of artificial neural network model for a crude oil distillation column.

2.2 Support vector machine

A support vector machine (SVM) is a widely used statistical technique for regression analysis (Ławryńczuk, 2016; Zaidi, 2012) and data classification (De Boves Harrington, 2015; Oliynyk et al., 2016). Vapnik (1995) provides a detailed description of the fundamentals of support vector machines. Here, the focus is on application of support vector machines for binary classification of samples (corresponding to column designs) as either feasible or infeasible. Feasible samples are those that lead to converged simulation, while infeasible samples are otherwise. The authors acknowledged that feasibility could be improved by additional iterations and/ or better initial guess.

The main idea of support vector machines is to define an optimal hyperplane that separates two or more classes of data points. For a given data set consisting of two separable classes, the optimal hyperplane is the one with the largest distance to the nearest data point, where distance is the Euclidian distance in the n-dimensional space.

Figure 4 illustrates a hyperplane separating two classes of data: feasible samples denoted by squares and infeasible samples denoted by circles. In this work, samples leading to a converged simulation on the rigorous crude distillation column model built in Aspen HYSYS are termed feasible, while those leading to unconverged simulations are termed infeasible.

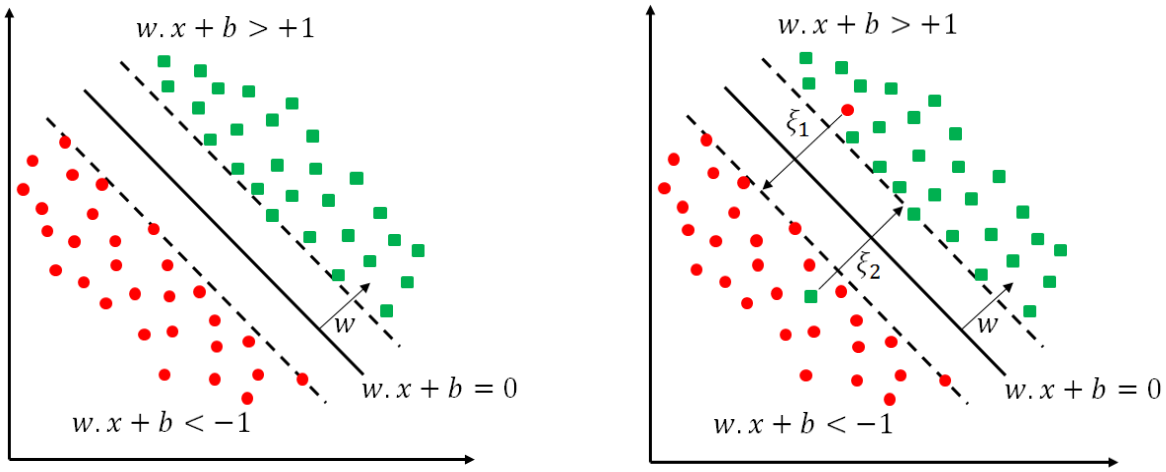


Figure 4 – Schematic representation of a support vector machine showing an optimal hyperplane for: (a) perfectly separable data sets, and (b) non-separable data sets in a two-dimensional space (adapted from Mahe et al. (2005)). Eq. (4) shows the detailed formulation for finding the optimal hyperplane.

In Figure 4, the optimal hyperplane can be determined as follows: Let $(\boldsymbol{x}_i, \boldsymbol{y}_i)$, $where\ i = 1, \ldots, n$, represent the set of training sample points, where $\boldsymbol{x} \in \mathbb{R}^n$ are decision variables (e.g. feed inlet temperature, pump-around duties and temperature drops, stripping steam flow rates and reflux ratio) and the corresponding sample class $\boldsymbol{y} \in \{+1, -1\}$ represents whether the decision variables values lead to a feasible or infeasible solution, where +1 denotes a feasible sample and -1 an infeasible sample. The support vector machine classifier can be defined by (Vapnik, 1995)

$$y(\boldsymbol{x}) = sign(\boldsymbol{w} \cdot \boldsymbol{x} + b) \tag{3}$$

where $w \in \mathbb{R}^n$ is the normal vector of the classification hyper plane, while $b \in \mathbb{R}$ is the bias.

Given a new instance $\boldsymbol{x}$, Eq. (3) allows classification of the sample as either feasible (positive sign) or infeasible (negative sign). Before Eq. (3) is applied to classify new samples, the weight vector ($\boldsymbol{w}$) and bias (b) need to be fitted to the training sample

points. This process is called 'learning,' and is usually carried out by solving Problem P1 using optimization (Vapnik, 1995). The procedure seeks the hyperplane with the largest separation margin between the two classes of data.

$$\text{[P1]} \qquad \min_{w,b,\xi} \ \frac{1}{2} w^T w + C \sum_{i=1}^{n} \xi_i \qquad (4)$$
$$\text{s.t} \quad y_i(w^T \cdot x_i + b) \geq 1 - \xi_i$$
$$\xi_i \geq 0, i = 1, \dots, n$$

where n represents the number of training data points; ξ_i are slack variables to account for misclassification of data points (see Figure 3b); C is a scalar penalty constant to scale constraint violations.

Eq. (3) represents a linear formulation of the support vector machine, which are mostly applied to linearly separable data points (Vapnik, 1995). Other formulations for non-linearly separable data points are also available such as polynomial functions, radial base functions and multi-layer perceptron (Vapnik, 1995).

Section 3.3 presents the detailed application of the support vector machine, Eq. (3), to construct a feasibility constraint to be implemented in an optimization framework with the aim of filtering out infeasible designs during the optimization task.

3 Modeling and solution procedure applied to a crude oil distillation column

The proposed surrogate modeling method comprises three main steps, namely, data generation (known as sampling), regression of a surrogate column modeling using the concept of artificial neural networks, and construction of feasibility constraints using a support vector machine.

3.1 Data generation

Data generation is an essential step in the construction of surrogate models, since the performance of the surrogate model strongly depends on the quality of the data used in the training stage. As discussed in the introduction, this work makes use of multiple rigorous simulations to generate sample points used for training the surrogate model.

Data generation using multiple rigorous simulation consist of three (3) major steps: selection of relevant inputs (or independent variables) and outputs (or dependent variables), generation of random sample points for each input variable, and lastly, for each set of input data, a rigorous simulation of the crude oil distillation is carried out to obtain the corresponding set of output variables.

In this work, the input variables selected include variables that can be adjusted to improve the crude distillation column performance. These variables include both structural (number of trays in each column section) and operational (feed inlet temperature, pump-around duties, and temperature drops, stripping steam flow rates and reflux ratio) degrees of freedom. The output variables selected include those that represent product quality (e.g., ASTM T5 and T95 boiling temperatures of each product), product flow rates, supply and target temperatures of streams requiring heating and cooling and corresponding enthalpy changes, and diameters of each column section. These variables allow calculation of the objective function and checking whether the product constraints are satisfied.

Next, a sampling technique is applied to generate random samples for each input variable. Several sampling techniques are available, for example, Monte Carlo sampling, Hammersley sequence and Latin hypercube sampling (Subramanyan et al., 2011). Without loss of generality, Latin hypercube sampling is used in this work, since it has shown to produce accurate statistical estimates of a probability distribution (Subramanyan et al., 2011). Latin hypercube sampling divides the input variable space into intervals, where samples are created randomly from each interval. The sampling method thus guarantees well-distributed samples.

After generating the samples, multiple rigorous simulations are carried out by simulating each such sample of the independent variables using a rigorous distillation column model implemented in a process simulation package. In this work, the multiple rigorous simulations are facilitated via an interface between Aspen HYSYS v8.6 and MatLab R2015a (AspenTech, 2011). Results of the simulations of the crude oil distillation in Aspen HYSYS v8.6 are recorded for each sample set. Samples are labeled as feasible if the simulation converges and infeasible if it does not. It is desirable to remove infeasible samples from the solution space to prevent the optimization algorithm from searching for and converging to infeasible solutions. Section 3.2 describes how an artificial neural network model representing the distillation column is constructed using the converged samples and Sections 3.3 explains how the whole sample set is used to construct a support vector machine.

3.2 Creating a surrogate distillation column model

This work uses an artificial neural network to create a surrogate model of the crude oil distillation column. The network correlates input and output data representing independent and dependent variables of the crude oil distillation column. A major advantage of artificial neural networks over other statistical techniques is the ability to correlate multiple inputs to multiple outputs, leading to compact models that can be implemented in an optimization environment with ease. To take advantage of this important feature, the artificial neural network representing the crude distillation column is built by correlating all the independent variables to a group of specific dependent variables. In this work, the crude oil distillation unit consists of 18 independent variables and 46 dependent variables. The dependent variables are divided into seven groups: ASTM T5 and T95 boiling temperatures, product flow rates, supply and target temperatures, enthalpy changes and diameters of each column section. Then seven ANN models are constructed to correlate the independent variables to each group of dependent variables.

This work uses the Artificial Neural Network Tool Box in MatLab R2015a to construct, validate and test the surrogate crude oil distillation column model. A multi-layer feed-

forward network with one input layer, one output layer, and one hidden layer is constructed. The hidden layer contains 10 neurons; the size of the output layer depends on the number of dependent variables. The former uses a sigmoid function, and the latter uses a linear transfer function, in line with common practice (Beale et al., 2015). The feasible sample points generated from the multiple rigorous simulations are randomly divided into a training set (70%), a validation set (15%) and a testing set (15%). The ANN model is first constructed using the points in the training set; note that the training data is normalized between -1 and 1 in order to improve training (Beale et al., 2015). The points in the validation set are employed during training to avoid model overfitting. The testing set is used for checking the performance of the model. The coefficient of determination is used to assess the performance of the seven networks. The coefficient of determination is a dimensionless quantity, typical range is 0 to 1, and it indicates the fraction of the variability of the dependent variable that is explained by the surrogate model (Diamond and Jefferies, 2001). The larger the coefficient of determination (close to one), the better the fitting and vice versa. In building the artificial neural network, the size of the hidden layer and number of neurons can be varied until the desired performance is achieved. While a methodology to optimize the number of layers and neurons has been developed by Dua (2010), implementing this methodology can be time consuming and challenging. In this work, several neural networks were constructed; trial and error was used to identify how many layers and neurons are needed for a good performance for optimization-based design of the crude oil distillation unit.

3.3 Feasibility constraint

This work uses feasibility constraints formulated as a support vector machine to enhance the optimization task. The support vector machine constructed here has a third-order polynomial function with an output of +1 and -1. Positive one represents converged samples (feasible designs), and negative one denotes unconverged samples (infeasible designs). The inputs (18 variables) of the support vector machine are similar to that of the artificial neural network discussed in Section 3.2. All the samples generated via multiple rigorous simulations, consisting of converged and unconverged

samples, are used to build the support vector machine. The data set is split into training set (75%) and validation set (25%). The function *fitcsvm*, which is implemented in MatLab 2015a, is applied to train and validate the support vector machine.

4 Framework for the design of crude oil distillation units

In this section, the artificial neural networks representing the crude oil distillation unit and the feasibility constraint developed using support vector machine are implemented a framework to design the unit. First, the mathematical formulation of the optimization problem is presented, followed by an approach proposed to solve the optimization problem.

4.1 Mathematical formulation

The design of crude oil distillation column based on surrogate models (artificial neural network and support vector machine) can be formulated as a mixed integer nonlinear programming problem (P1) as follows:

$$\text{(P2)} \qquad \min_{x_S, x_O} \quad f(x_D, x_S, x_O) \tag{4}$$

$$s.t. \quad h\,(x_D, x_S, x_O) = 0$$
$$g\,(x_D, x_S, x_O) \leq 0$$

$$x_D \in X_D, x_S \in X_S, x_O \in X_O$$

where f is the objective function, h is the set of equality constraints represented by a surrogate model, g is the set of inequality constraints; while X_D, X_S and X_O are the feasible sets of the variables, namely: x_D, x_S and x_O, which represent dependent, structural and operational variables, respectively. As will be later discussed in more detail, this model is solved using a genetic algorithm.

In this work, the surrogate model comprises seven artificial neural networks. Each neural network predicts a specific dependent variable of the crude oil distillation column. In addition to the artificial neural network, a support vector machine is also

trained to predict the feasibility of the crude oil distillation column. Hence, the equality constraints in Problem P2 can be more specifically represented as follows

$$
\begin{aligned}
h_1&: \quad [T5_i] = ANN_1(x_s, x_o) && i = 1,2,3,\dots,N_{product}\\
h_2&: \quad [T95_i] = ANN_2(x_s, x_o) && i = 1,2,3,\dots,N_{product}\\
h_3&: \quad [F_i] = ANN_3(x_s, x_o) && i = 1,2,3,\dots,N_{product}\\
h_4&: \quad [D_j] = ANN_4(x_s, x_o) && j = 1,2,3,\dots,N_{section}\\
h_5&: \quad [E_k] = ANN_5(x_s, x_o) && k = 1,2,3,\dots,N_{stream}\\
h_6&: \quad [TS_l] = ANN_6(x_s, x_o) && l = 1,2,3,\dots,N_{stream}\\
h_7&: \quad [TT_l] = ANN_7(x_s, x_o) && l = 1,2,3,\dots,N_{stream}\\
h_8&: \quad [CC] = SVM(x_s, x_o)\\
h_9&: \quad CC = 1
\end{aligned}
\qquad (5)
$$

where *T5* and *T95* represents the ASTM boiling temperatures of product i at 5% and 95% vaporization, *F* is the flow rate of product i, *D* is the diameter of section j, *E* is the enthalpy change of stream k, *TS* and *TT* are the supply and target temperatures of stream l, *CC* represents the predicted convergence criterion (+1 for converged column and -1 otherwise), $\mathbf{x_s}$ and $\mathbf{x_o}$ are structural and operating variables (independent variables) and lastly, *ANN* and *SVM* represent artificial neural network and support vector machine functions respectively

Inequality constraints in Problem P2 define bounds imposed on independent variables and product quality specifications. These can be represented as follows

$$
\begin{aligned}
g_1&: \quad lb_i \le N_i \le ub_i && i = 1,2,3,\dots,N_{section}\\
g_2&: \quad lb_j \le Q_{PA,j} \le ub_j && j = 1,2,3\\
g_3&: \quad lb_j \le \Delta T_{PA,j} \le ub_j && j = 1,2,3\\
g_4&: \quad lb_k \le F_{s,k} \le ub_k && k = 1,2\\
g_5&: \quad lb \le R \le ub\\
g_6&: \quad lb \le T_F \le ub\\
g_7&: \quad lb_l \le T5_l \le ub_l && l = 1,2,\dots,N_{product}\\
g_8&: \quad lb_l \le T95_l \le ub_l && l = 1,2,\dots,N_{product}
\end{aligned}
\qquad (6)
$$

where N_i is the number of active trays in column section i; $Q_{PA,j}$ and $\Delta T_{PA,j}$ are the duty and temperature drop of pump-around j; $F_{S,k}$ is the flow rate of stream k; R is the overhead reflux ratio; T_F is the feed inlet temperature; $T5_l$ and $T95_l$ are the boiling temperatures of product l at 5% and 95% vaporization, for example, according to ASTM standards.

The objective function [$f(x_D, x_S, x_O)$] employed in this work is the total annualized cost, since the aim is to identify the design alternative that minimizes both capital expenditures and operating costs of the crude oil distillation unit. Although other types of objective function such as net profit, net present value, energy cost, and CO_2 emission could be used depending on the purpose of the design.

The total annualized cost is the sum of the annualized capital cost (ACC) and the total operating costs (OC) of the crude oil distillation unit. The annualized capital cost is the sum of the installed cost of the column shells (S_c) and of the trays within the column, (T_C). The column shell and tray costs are estimated using the correlations proposed by Guthrie (1969) (see Support Information). The annualization factor described by Smith (2005) is applied to split the total column cost across the entire plant life at a specific interest rate.

$$ACC = (S_c + T_C) \cdot A_f \tag{7}$$

$$A_f = \frac{i(1+i)^t}{(1+i)^t - 1} \tag{8}$$

where i is the interest rate and t is the plant life.

The operating cost of the crude oil distillation unit is dominated by the cost of the fired heating, typically using fuel oil or natural gas; the cost of steam for stripping indirect heating and the cost of cold utilities also. In this work, the utility demand is estimated using pinch analysis, i.e., using composite curves to determine the minimum utility demand (Smith, 2005). The pinch calculation is carried out using an open source MatLab code (Morandin, 2014). In this way, heat recovery is incorporated during the

column optimization without the need to designing explicitly the associated heat exchanger network (HEN). Hence, the operating cost (*OC*) is evaluated using Eq. (9):

$$OC = \sum_{i=1}^{n} ST_i * C_{ST,i} + HU * C_{HU} + CU * C_{CU} \qquad n = 2 \qquad (9)$$

where C_{ST}, C_{HU} and C_{CU} are the unit costs of stripping steam and of the hot and cold utilities respectively; HU and CU are the minimum hot and cold utilities, respectively, while n represents the number of stripping steam streams associated with the column.

4.2 Optimization framework

In Problem P2, the inequality constraints, g_1 to g_6 are bounds on structural and operational variables, while g_7 and g_8 represent constraints on product quality in terms of ASTM D86 boiling temperature (T5 and T95). For the sake of simplicity and to facilitate the search for the optimal solution, the latter constraints are included into the objective function as a penalty function (Edgar et al., 2001); the new formulation, P3, is:

$$\text{(P3)} \qquad \min_{x_S, x_O} \quad f(x_D, x_S, x_O) + \left[\Pi \sum_{i=1}^{n} [\max(0, (g_i))]^2\right] \qquad (10)$$

$$s.t. \quad h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8$$
$$g_1, g_2, g_3, g_4, g_5, g_6$$

where g_i denotes the inequality constraints g_7 and g_8; Π is a scalar parameter that scales the magnitude of the violation of constraints, and hence ensures that the product quality specifications are maintained during the optimization. The magnitude of the scalar parameter may impact on the final optimal solution. A small value may allow some constraints to be violated, leading to infeasible solutions, while large values impose constraints very stringently. Computational experience suggests that a value of a similar magnitude as the objective function yields good results. In this work, Π is 10^6.

Figure 5 presents the strategy proposed to optimize Problem P3. The proposed approach integrates the artificial neural networks, the support vector machine, the heat

recovery model (pinch analysis) and the cost models into a unified framework. The framework applies a stochastic optimization algorithm (i.e., a genetic algorithm) to search for the best column structure and operating conditions that minimize capital expenditure and operating expenses.

The surrogate models employed in this work are easier to optimize than rigorous models (e.g. as used by Ibrahim et al., 2017a). Nevertheless, the sigmoid function used in this work to construct the surrogate model and the objective function (including non-linear capital cost correlations) together give rise to a nonlinear, non-convex optimization problem. Gradient-based searches are unlikely to locate the global optimum. Therefore, a stochastic optimization method, namely a genetic algorithm (GA), is used. The GA can search the solution space more effectively, even though it still cannot guarantee convergence to the global optimum

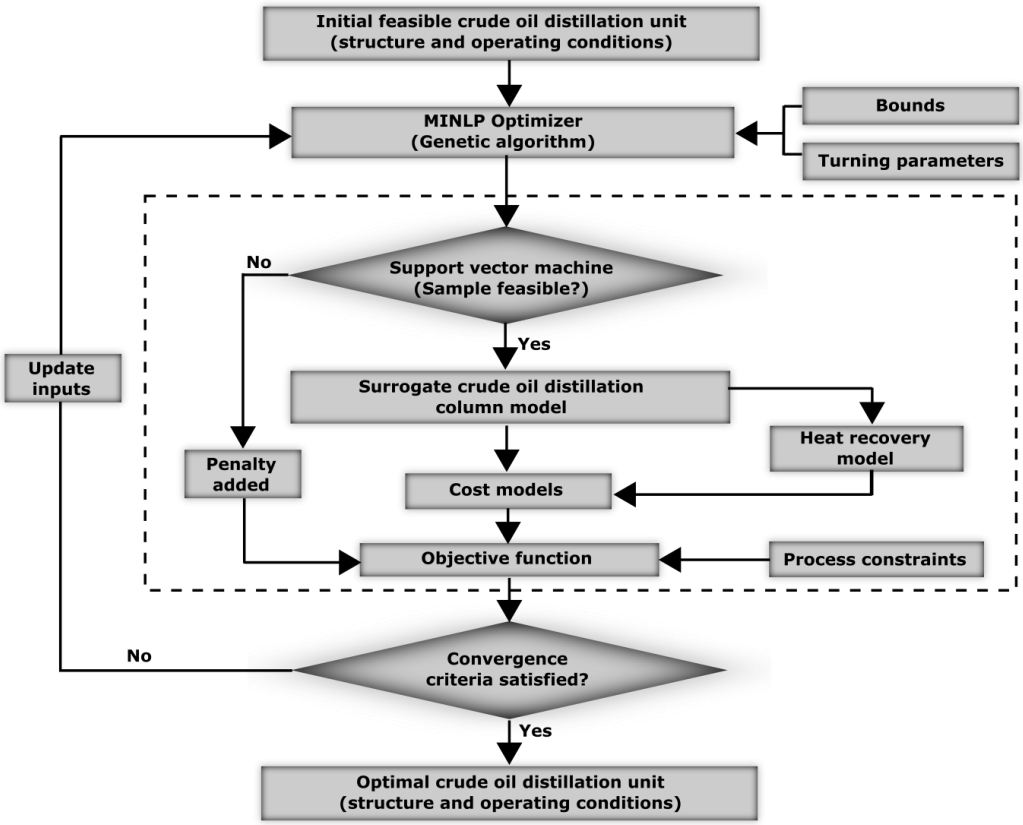


Figure 5 – Optimization framework for the design of crude oil distillation unit

The framework shown in Figure 5 is implemented in MatLab. A genetic algorithm available in the global optimization toolbox in MatLab ('*ga*') is employed to carry out the search for cost-effective design alternative. In each iteration, the genetic algorithm proposes a column structure and its operating conditions; the proposed inputs are then checked by the support vector machine to confirm whether the design is likely to be feasible. Inputs identified as feasible are then used by the artificial neural network model to simulate the proposed design solution and predict its performance. Among the model outputs are process stream data used in the heat recovery model to determine utility targets and other information required to calculate the value of the objective function. Inequality constraints are applied to check whether product qualities and flow rates specifications are met. In cases that the support vector machine identifies solutions generated by the optimisation algorithm as likely to be infeasible, the corresponding designs are not simulated by the artificial neural network model. Instead, a penalty is added to the objective function. Attributing a poor performance to potentially infeasible solutions helps to train the genetic algorithm not to propose similar design alternatives again. In this way, the search focuses on the feasible region, increasing the likelihood that the optimization will converge to a feasible design option. Furthermore, the computational time is reduced, since potentially infeasible solutions are removed from the search space. The next section illustrates how the methodology is applied to a relevant case study.

5 Case study

5.1 Problem description

This section demonstrates the capabilities of the novel design approach via the design of a crude oil distillation system. The crude oil distillation unit that separates 100,000 bbl day^{-1} (662.4 m^3 h^{-1}) of Venezuelan Tia Juana light crude oil (Watkins, 1979) into five products, namely, light naphtha (LN), heavy naphtha (HN), light distillate (LD), heavy distillate (HD) and residue (RES). The initial design of the feasible column is obtained from Chen (2008); the unit comprises a main column with three pump-arounds and three side-strippers. The main column has five sections with 5, 9, 10, 8 and 9 sieve trays

respectively. The HD, LD and HN side-strippers have 5, 7 and 6 sieve trays, respectively (Chen, 2008), presented also in Ibrahim et al. (2017).

Given data include operating conditions, product quality metrics (in terms of ASTM 5 % and 95 % boiling temperatures) and product flow rates. The column operating pressure is taken to be uniform and equal to 2.5 bar. The economic calculation assumes an interest rate of 5%, a plant life of 20 years and 8700 operating hours per year. The costs of utilities, which include stripping steam (260 ^{o}C, 4.5 bar), fired heating (1500–800 ^{o}C) and cooling water (supplied at 10 and returned at 40 ^{o}C) are \$0.14 kmol^{-1}, \$150 kW^{-1} a^{-1} and 5.25 \$ kW^{-1} a^{-1} respectively. A minimum approach temperature of 30 ^{o}C in all streams is used to calculate the minimum utility requirements. Details of the crude oil assay, initial operating conditions, product quality specifications and flow rates (initial values) of the base case are presented in Tables S1 to S5 of the supplementary material.

5.2 Surrogate crude oil distillation column model

The surrogate modeling approach proposed in Section 3 is applied to model the crude oil distillation unit. First, a rigorous simulation model of the crude oil distillation unit is implemented in Aspen HYSYS using Peng-Robinson as thermodynamic property package. Sensitivity analysis is carried out to identify suitable bounds for each independent variable as shown in Table 5.1. Initially, bounds are defined for each independent variable, i.e., $\pm$ 10 °C for temperature related variables, $\pm$ 25 % for stripping steam and duty, $\pm$ 2 for reflux ratio and -2/+1 for number of trays in column sections; then multiple simulations were carried out, and the bounds are adjusted accordingly, to facilitate convergence. The range of each bound is defined with reference to the initial/ nominal value.

Next, the Latin Hypercube sampling method is applied to generate 7000 samples, each consisting of different combinations of the independent variables within the bounded region. Through an interface established between HYSYS and MatLab, all the samples are simulated on the rigorous column model developed in HYSYS. Out of these samples, 59% (4130) simulations converged; for the remaining 41% (2870), the

simulations did not converge. The sampling is carried out on a HP desktop PC with Intel(R) Core i5 processor running at 3.2 GHz, and 8 GB of RAM. It took around 1.5 h to generate the set of 7000 samples.

Table 5.1 – Bounds on trays in column sections and operating conditions

Variables	Lower bound	Upper bound	Initial value
Number of tray			
Section 1	3	6	5
Section 2	7	10	9
Section 3	8	11	10
Section 4	6	9	8
Section 5	7	10	9
Section 6	3	6	5
Section 7	5	8	7
Section 8	4	7	6
Operating condition			
PA 1 duty (MW)	8.40	14.00	11.2
PA 2 duty (MW)	13.42	22.36	17.89
PA 3 duty (MW)	9.63	16.05	12.84
PA 1 DT (°C)	10	30	20
PA 2 DT (°C)	40	60	50
PA 3 DT (°C)	20	40	30
Main stripping steam (kmol/h)	900	1500	1200
HD stripping steam (kmol/h)	188	313	250
Feed temperature (°C)	340	375	365
Reflux ratio	3.17	6.17	4.17

Figure 6 shows the performance results of the seven artificial neural networks representing the crude oil distillation units. The performance of the artificial neural network is calculated using the data in the testing set (see Section 3.2). The coefficient of determination for all artificial neural network models is approximately 0.9999. The values of the coefficient of determination indicate that the models built are able to explain 99.99% of the variance of the outputs, which indicates the goodness of fit.

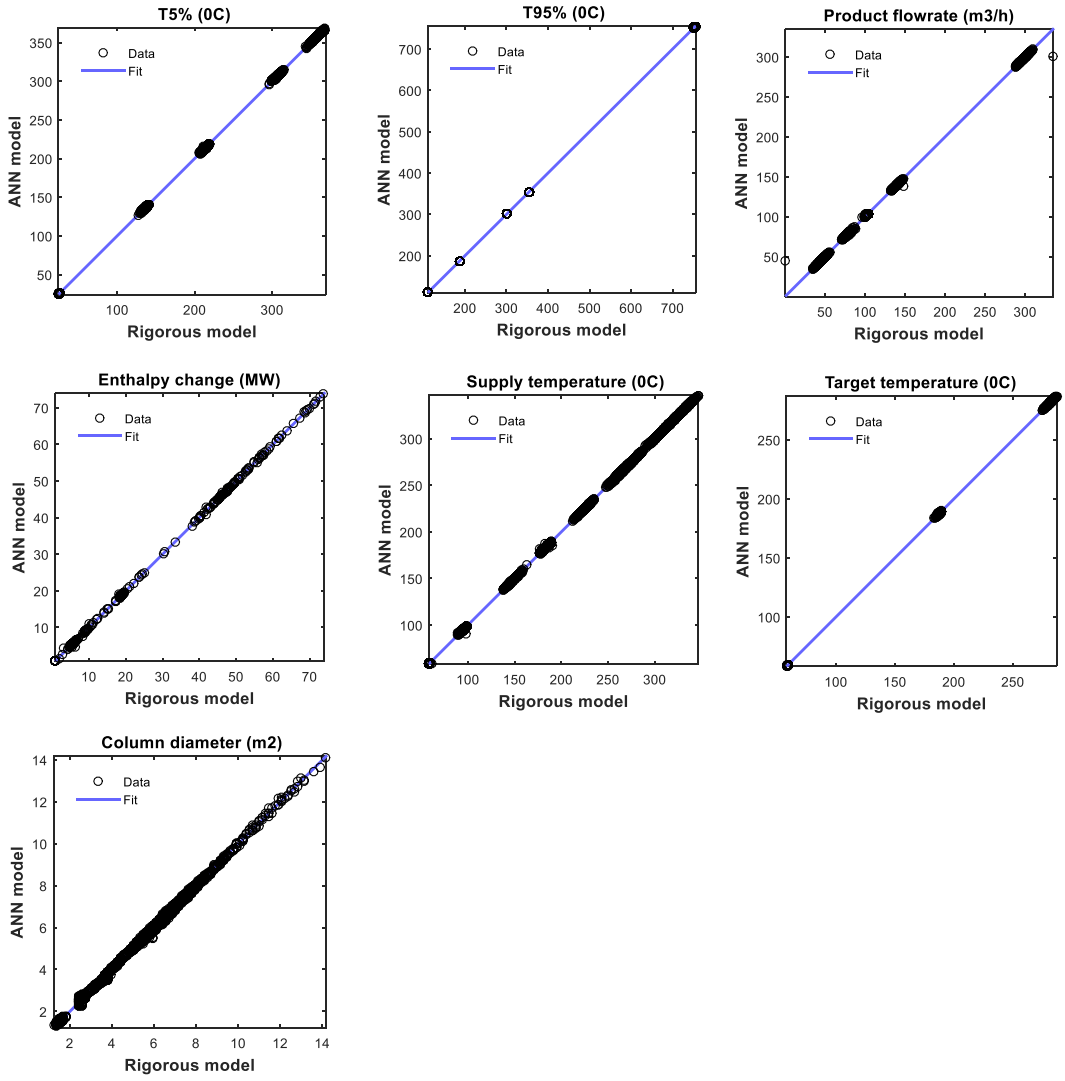


Table 6 Parity plots showing predictions of artificial neural network versus rigorous model

To gain further confidence on the ANN models, additional statistical tests are carried out to further assess the model accuracy. The performance criteria are average absolute error and average relative error, calculated for temperature related variables and other variables (product flow rate, enthalpy change and diameter in column sections) respectively. These tests are carried out using 100 converge samples generated using the approach presented in Section 3.1. Table 5.2 shows the average error of the predictions of the artificial neural network column model compared to the rigorous simulation model results. The seven artificial neural network models have 46 outputs in total: five each for T5 and T95 boiling temperatures; five product flow rates; 11 supply temperatures; three target temperatures; eight duties for the condenser, three

pump-arounds, two reboilers and five product coolers; the eight diameters, one for each column section.

Table 5.2 – Validation results for the ANN column model

Variables	Average absolute error (°C)	Variables	Average relative error (%)
Product quality		**Product flow rates**	
LN T5%	0.098	LN	0.169
HN T5%	0.250	HN	0.368
LD T5%	0.454	LD	0.168
HD T5%	0.358	HD	0.736
RES T5%	0.269	RES	0.084
LN T95%	0.009	**Exchanger duties**	
HN T95%	0.028	ADU condenser	0.295
LD T95%	0.013	LN cooler	2.227
HD T95%	0.002	HN cooler	1.035
RES T95%	0.027	LD cooler	0.301
Supply temperature		HD cooler	1.825
ADU condenser	0.173	RES cooler	0.248
LN cooler	0.054	HN reboiler	2.122
HN cooler	0.196	LD reboiler	0.683
LD cooler	0.199	Fired heater	0.101
HD cooler	0.732	**Column diameter**	
RES cooler	0.423	Main column	
PA1	0.322	Section 1	0.898
PA2	0.208	Section 2	0.903
PA3	0.277	Section 3	0.915
HN reboiler	0.220	Section 4	0.942
LD reboiler	0.235	Section 5	0.767
Target temperature		HN side-stripper	2.409
ADU condenser	0.057	LD side-stripper	0.345
HN reboiler	0.178	HD side-stripper	3.012
LD reboiler	0.165		

As shown in Table 5.2, the largest average absolute error in temperature-related variables is less than 0.5°C and average relative error for all other variables is less than 3%. The largest deviation is observed for the section diameter in the HN and HD side strippers, yet the associated error is small and may have a marginal impact on the

column capital cost calculations. These results confirm the effectiveness and accuracy of the artificial neural network column model.

5.3 Feasibility constraint based on support vector machine

Table 5.3 presents the validation results of the support vector machine classifier. Note for positive class: true prediction means the SVM classifies the sample as feasible, and it later converges in the simulation model; false means that it is labeled as feasible, but it does not converge in the rigorous model. For negative class: true prediction means the SVM labels the sample as infeasible, and it indeed does not converge in the rigorous model; false prediction means the SVM labels the point as infeasible, yet it converges in the rigorous model.

Table 5.3 – Validation results for support vector machine.

Prediction class	True prediction	False prediction
Positive class (Converged samples, +1)	966 [91.1%]	95 [8.9%]
Negative class (Unconverged samples, −1)	396 [58%]	293 [42%]
Overall: Correct prediction	78%	
Wrong prediction	22%	

From the optimization point of view, false positives cut-off the feasible region, therefore rejecting design alternatives that may be feasible (and potentially optimal). False negatives wrongly enlarge the search space by adding infeasible points. False positives need to be minimized, as otherwise there is the danger that the optimization algorithm will converge towards an infeasible solution. On the other hand, low false negatives will lead to larger CPU times, but do not compromise the quality of the final solution.

The support vector machine results are compared with previous work (Ochoa-Estopier and Jobson, 2015), in which an artificial neural network was used to classify sample points. The same data set used in training the support vector machine is used to train the artificial neural network. The procedure for setting up the artificial neural network

can be found elsewhere (Beale et al., 2015). Table S10 in the supplementary material shows the performance results obtained from this analysis. From the results, artificial neural network retains slightly (3.4%) more feasible designs than the support vector machine. On the other hand, the support vector machine removes high proportion (4.1%) of the infeasible designs than the artificial neural network. This analysis indicates that the support vector machine is advantageous when the emphasis is to increase the likelihood for optimization solutions to be feasible when simulated on a rigorous model.

5.4 Crude oil distillation unit optimization results

The MINLP problem [PD2] is solved using MatLab R2015a, employing the genetic algorithm, '*ga*' implemented in the Global Optimization Toolbox. The initial population contains 100 chromosomes (representing alternative designs), and the maximum number of generations is set to 300. These parameters values are tuned via some preliminary tests.

The genetic algorithm is run ten times in order to confirm that the solutions obtained are of a reasonable quality, as indicated by standard deviation.

To check for the effectiveness of the support vector machine introduced in this work, the optimization is run (using the artificial neural network column model) with and without the support vector machine. As shown in Table S11 (see supplementary material), the order of magnitude of the objective function for the two cases are within similar range, although when the support vector machine is not used, only 70% of the optimization results lead to a feasible simulation on the rigorous model compared with 100% when the support vector machine is used, confirming the need to include support vector machine within an optimization framework for the design of crude oil distillation unit

Details of the computational results for the multiple runs of the genetic algorithm (including support vector machine) are summarized in Tables S6 (see supplementary

material). The solution with the lowest objective function value and minimum computational time is reported Figure 7 and Tables 5.5 to 5.7. The five sections of the main column consist of 6, 10, 11, 8 and 7 trays respectively (counting from the bottom), while HD, LD, and HN side strippers have three trays each. The minimum hot and cold utility demand has decreased by 20% and 28%, respectively, compared with the initial design. These decrease are due to: (i) reduction in the temperature of the feed to the column, which reduces the fired heating duty for a given flow rate and furnace inlet temperature; (ii) redistribution of product flow rates between adjacent streams, i.e., transfer of product from streams with low temperature to those with high temperature, without compromising product quality specifications (see Table 5.6). The latter strategy improves the total amount of recoverable heat at high temperature (high-quality heat) within the system, thereby decreasing hot utility demand.

The total operating cost, consisting of utility cost and steam cost, amounts to 6.81 \$MM a^{-1} (where MM is million), while the annualized capital cost of the column is 0.21 \$MM a^{-1}. Therefore, the column total annualized cost amounts to 8.45 \$MM a^{-1}, which is 20% lower than the initial design.

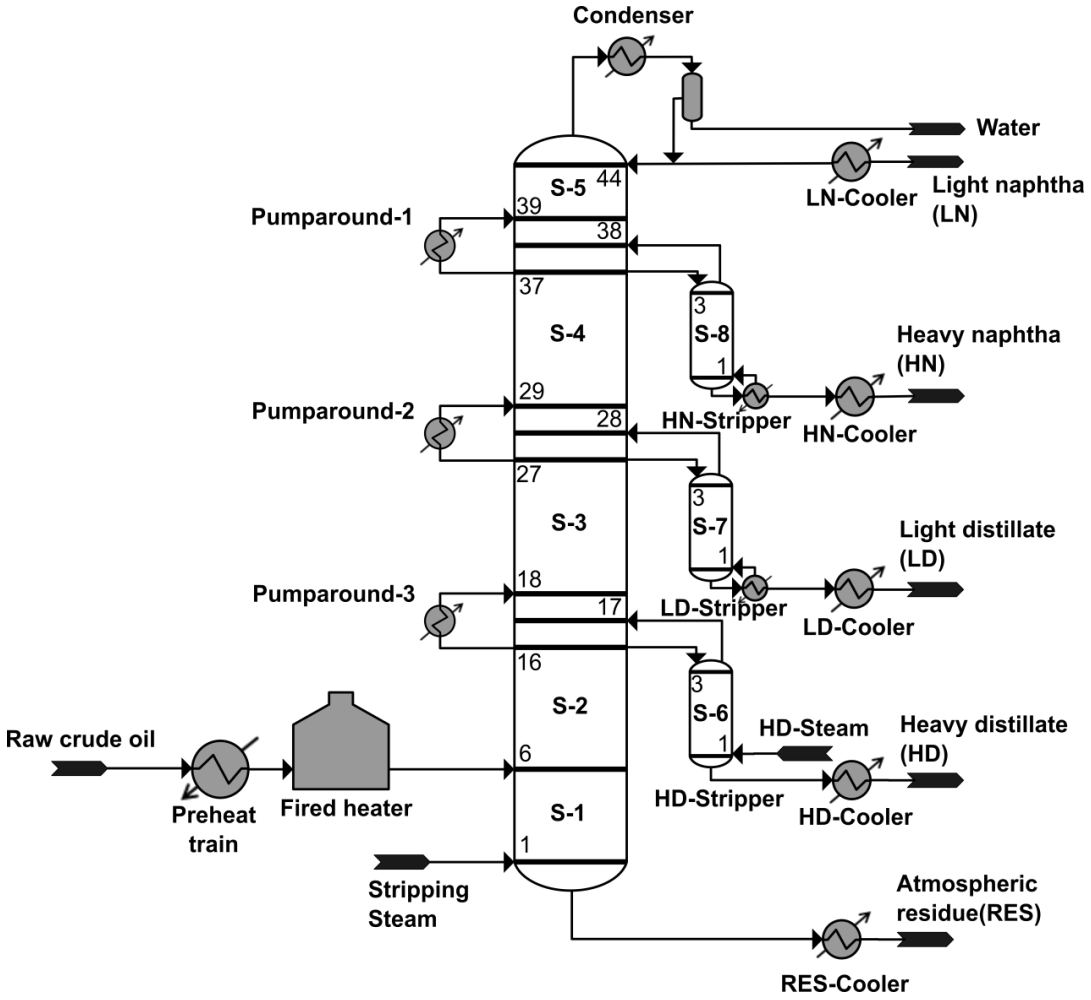


Figure 7 – Optimal configuration of crude oil distillation unit

Table 5.5 – Crude oil distillation unit operating conditions

Variable	Initial design	Optimized design
PA 1 duty (MW)	11.2	11.05
PA 2 duty (MW)	17.89	14.01
PA 3 duty (MW)	12.84	14.35
PA 1 DT (°C)	20	20
PA 2 DT (°C)	50	56
PA 3 DT (°C)	30	28
Main steam (kmol h^{-1})	1200	900
HD steam (kmol h^{-1})	250	189
Feed temperature (°C)	365	361
Reflux ratio	4.17	3.33

Table 5.6 – Product quality and flow rate

Products	T5% (°C, ASTM D86)		T95% (°C, ASTM D86)		Flow rates ($m^3\ h^{-1}$)	
	Initial design	Optimized design	Initial design	Optimized design	Initial design	Optimized design
LN	25.9	24.5	110.6	110.6	103.5	99.4
HN	138.9	131.1	186.6	186.8	78.2	85.5
LD	215.9	215.7	301.5	301.5	140.3	134.7
HD	310.7	306.9	354.4	354.4	48.1	49.8
RES	361.4	360.4	754.3	754.1	292.5	293.3

Table 5.7 – Utility demand and column cost

Variable	Initial design	Optimized design	Unit*
Utility requirements			
Hot utility	54.61	44.46	MW
Cold utility	61.18	46.04	MW
Cost			
Utility cost	8.51	6.91	\$MM a^{-1}
Steam cost	1.77	1.33	\$MM a^{-1}
Total operating cost	10.28	8.24	\$MM a^{-1}
Annualized capital cost	0.33	0.21	\$MM a^{-1}
Total annualized cost	10.61	8.45	\$MM a^{-1}

*\$MM denotes millions of dollars

The minimum energy demand is calculated using pinch analysis, without taking into account details of the heat exchanger network. Nevertheless, the stream information in Table S8 of the supplementary material can be used to design the heat exchanger network for the column. For a more detailed analysis, a heat exchanger network model will be required to replace the grand composite curve. In this way, the column and the heat exchanger network can design simultaneously.

The optimization results i.e., the optimal column configuration and operating conditions obtained using the surrogate model were then used as inputs to a rigorous simulation in Aspen HYSYS, in order to check whether: (i) the simulation will converge; (ii) the ANN results are in good agreement with the converged simulation results; (iii) the performance in the two cases is similar, i.e., the ANN gave good guidance about the economic performance. The column validation, shown in Tables S6

(see supplementary material), indicates that the optimal results obtained using the surrogate model are in excellent agreement with those generated by the rigorous simulation, confirming the accuracy and effectiveness of the proposed approach.

The optimization results were also compared to results obtained via a direct simulation–optimization approach using a rigorous distillation column model, the results of which are shown in Table 5.8. The same input data (as presented in Tables S1 to S4) were used in both approaches. Details of the procedure for setting up the problem and generating the results for the simulation-optimization approach can be found elsewhere (Ibrahim et al., 2017). The simulation-optimization is facilitated via an interface established between Aspen HYSYS v8.6 and MatLab, where the rigorous column modeling is carried out in Aspen HYSYS v8.6, and the optimization is performed in MatLab using a genetic algorithm.

Table 5.8 – Comparison of surrogate and rigorous model performance

Variable	Optimization with the rigorous model	Optimization with the# ANN+SVM	Unit*
Utility requirements			
Hot utility	44.32	44.46	MW
Cold utility	45.9	46.04	MW
Cost analysis			
Utility cost	6.89	6.91	\$MM a^{-1}
Steam cost	1.33	1.33	\$MM a^{-1}
Total operating cost	8.22	8.24	\$MM a^{-1}
Annualized capital cost	0.26	0.21	\$MM a^{-1}
Total annualized cost	8.47	8.45	\$MM a^{-1}

*\$MM denotes millions of dollars

#Validated results

As can be seen in Table 5.8, the heating and cooling duties for two different methods are more or less the same, leading to similar utility cost and the same stripping steam cost. A total annualized cost of 8.45 \$MM a^{-1} is obtained using the surrogate model-based approach, which is slightly lower than the one from the simulation-optimization approach, although the two results have the same order of magnitude.

The CPU time for the simulation-optimization approach was 180 CPU minutes compared with about 100 CPU minutes obtained using the surrogate column model. It should be noted that the CPU time for the surrogate model includes sampling (94 minutes), model construction and validation (5.2 minutes), and column optimization and pinch analysis (1.25 minutes). Sampling required most of the computation time. Overall, significantly less time was required (44%) without affecting the optimality and accuracy of the solution.

It is also worth mentioning that model development needs to be carried out only once, and then enables the optimization of the system considering different objectives and constraints. For example, in a design problem where many scenarios might be explored, the proposed approach may significantly overcome the direct optimization of the rigorous model. As an example, consider the application of a standard sensitivity analysis that aims to understand how the design changes according to variations in a cost parameter. In the proposed approach, the artificial neural network and support vector machine are built only once, and then the optimization is run (which is quite quick) as many times as desired. On the contrary, carrying out the same calculations with the rigorous model would entail much longer times as the time-consuming optimization of approximately three hours would need to be repeated many times.

6 Conclusions

Optimization-based design of crude oil distillation unit using surrogate models naturally leads to a mixed integer nonlinear programming problem that is very difficult to solve. Heat recovery needs to be considered during column design in order to account effectively for the trade-off between capital and operating cost.

This work proposes a novel systematic framework for the design of heat-integrated crude oil distillation unit that combines a surrogate column model (in particular an artificial neural network), the support vector machine as a feasibility classifier and pinch analysis within an optimization framework that applies a genetic algorithm.

Suitable dependent and independent variables identified via sensitivity analysis were applied to generate sample points. Those samples that resulted in converged rigorous simulations were employed to build a surrogate model of the column using an artificial neural network, while the entire set of samples (whether or not the rigorous simulation converged) was used to train the support vector machine which aimed to classify the solution space into feasible (i.e. likely to converge) and infeasible designs. The artificial neural network predictions are shown to be in good agreement with the results from the rigorous column model, and the support vector machine is shown to have some ability to remove a significant proportion of the infeasible design options from the search space. The column design was optimized, considering both structure and operational variables, using a genetic algorithm. The proposed approach was demonstrated to be capable of identifying cost-effective designs in significantly reduced times, compared to direct simulation-optimization using rigorous column models, rather than surrogate column models.

Future work will extend the proposed approach to design a flexible heat-integrated crude oil distillation unit that optimally processes multiple crude oil feedstocks.

Appendix A: Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.cherd.2018.03.006

Acknowledgement

The authors would like to acknowledge the financial support from Petroleum Technology Development Fund (PTDF), Nigeria, for sponsoring this PhD research.

References

Bagajewicz, M., Ji, S., 2001. Rigorous Procedure for the Design of Conventional Atmospheric Crude Fractionation Units. Part I: Targeting. Ind. Eng. Chem. Res. 40, 617–626.

Basheer, I.A., Hajmeer, M., 2000. Artificial neural networks: Fundamentals, computing, design, and application. J. Microbiol. Methods 43, 3–31.

Beale, M.H., Hagan, M.T., Demuth, H.B., 2015. Neural Network Toolbox™ User's Guide.

Chen, L., 2008. Heat-Integrated Crude Oil Distillation System Design. PhD Thesis, Univ. Manchester, UK.

De Boves Harrington, P., 2015. Support Vector Machine Classification Trees. Anal. Chem. 87, 11065–11071.

Diamond, I., Jefferies, J., 2001. Beginning Statistics: An Introduction for Social Scientists. SAGE Publications. California, US.

Dua, V., 2010. A mixed-integer programming approach for optimal configuration of artificial neural networks. Chem. Eng. Res. Des. 88, 55–60.

Edgar, T.F., Himmelblau, D.M., Lasdon, L.S., 2001. Optimization of chemical processes, McGraw-Hill. New York.

Goda, T., Sato, K., 2014. History matching with iterative Latin hypercube samplings and parameterization of reservoir heterogeneity. J. Pet. Sci. Eng. 114, 61–73.

Gu, W., Huang, Y., Wang, K., Zhang, B., Chen, Q., Hui, C.W., 2014. Comparative analysis and evaluation of three crude oil vacuum distillation processes for process selection. Energy 76, 559–571.

Gueddar, T., Dua, V., 2012. Novel model reduction techniques for refinery-wide energy optimisation. Appl. Energy 89, 117–126.

Gueddar, T., Dua, V., 2011. Disaggregation-aggregation based model reduction for refinery-wide optimization. Comput. Chem. Eng. 35, 1838–1856.

Guthrie, K.M., 1969. Data and Techniques for Preliminary Capital Cost Estimating. Chem. Eng. 76, 114.

Henao, C.A., Maravelias, C.T., 2010. Surrogate-Based Superstructure Optimization Framework. AIChE J. 57, 1216–1232.

Himmelblau, D.M., 2008. Accounts of Experiences in the Application of Artificial Neural Networks in Chemical Engineering. Ind. Eng. Chem. Res. 47, 5782–5796.

Ibrahim, D., Jobson, M., Guillén-Gosálbez, G., 2017a. Optimization-Based Design of Crude Oil Distillation Units Using Rigorous Simulation Models. Ind. Eng. Chem. Res. 56, 6728–6740.

Ibrahim, D., Jobson, M., Li, J., Guillén-Gosálbez, G., 2017b. Surrogate Models Combined with a Support Vector Machine for the Optimized Design of Crude Oil Distillation Units Using Genetic Algorithms. Proc. ESCAPE 27.

Jobson, M., Ochoa-Estopier, L.M., Ibrahim, D., Chen, L., Guillén Gosalbez, G., Li, J., 2017. Feasibility Bounds in Operational Optimization and Design of Crude Oil Distillation Systems Using Surrogate Methods. Chem. Eng. Trans. 61.

Ławryńczuk, M., 2016. Modelling and predictive control of a neutralisation reactor using sparse support vector machine Wiener models. Neurocomputing 205, 311–328.

Liau, L.C.-K., Yang, T.C.-K., Tsai, M.-T., 2004. Expert system of a crude oil distillation unit for process optimization using neural networks. Expert Syst. Appl. 26, 247–255.

Liebmann, K., 1996. Integrated Crude Oil Distillation Design. PhD Thesis, UMIST, Manchester, UK.

Liebmann, K., Dhole, V.R., Jobson, M., 1998. Integrated Design of a Conventional Crude Oil Distillation Tower Using Pinch Analysis. Chem. Eng. Res. Des. 76, 335–347.

López C., D.C., Hoyos, L.J., Mahecha, C.A., Arellano-Garcia, H., Wozny, G., 2013. Optimization model of crude oil distillation units for optimal crude oil blending and operating conditions. Ind. Eng. Chem. Res. 52, 12993–13005.

Mahe, P., Ueda, N., Akutsu, T., Perret, J.-L., Vert, J.-P., 2005. Graph kernels for molecular structure-activity relationship analysis with support vector machines. J. Chem. Inf. Model. 45, 939–951.

Motlaghi, S., Jalali, F., Ahmadabadi, M.N., 2008. An expert system design for a crude oil distillation column with the neural networks model and the process optimization using genetic algorithm framework. Expert Syst. Appl. 35, 1540–1545.

Nuchitprasittichai, A., Cremaschi, S., 2013. Optimization of CO2 capture process with aqueous amines - A comparison of two simulation-optimization approaches. Ind. Eng. Chem. Res. 52, 10236–10243.

Nuchitprasittichai, A., Cremaschi, S., 2012. An Algorithm to Determine Sample Sizes for Optimization with Artificial Neural Networks. AIChE J. 59, 805–812.

Ochoa-Estopier, L.M., Jobson, M., 2015. Optimization of Heat-Integrated Crude Oil Distillation Systems. Part I: The Distillation Model. Ind. Eng. Chem. Res. 54, 4988–5000.

Ochoa-Estopier, L.M., Jobson, M., 2015b. Optimization of Heat-Integrated Crude Oil Distillation Systems. Part III: Optimization Framework. Ind. Eng. Chem. Res. 54, 5018–5036.

Oliynyk, A.O., Adutwum, L.A., Harynuk, J.J., Mar, A., 2016. Classifying crystal structures of binary compounds AB through cluster resolution feature selection and support vector machine analysis. Chem. Mater. 28, 6672–6681.

Osuolale, F.N., Zhang, J., 2017. Thermodynamic optimization of atmospheric distillation unit. Comput. Chem. Eng. 103, 201–209.

Smith, R., 2005. Chemical Process: Design and Integration. Wiley. Chichester, UK.

Szklo, A., Schaeffer, R., 2007. Fuel specification, energy consumption and CO_2 emission in oil refineries. Energy 32, 1075–1092.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer. US.

Watkins, R.N., 1979. Petroleum Refinery Distillation. Gulf Publishing Company, Book Division. Texas, US.

Yao, H., Chu, J., 2012. Operational optimization of a simulated atmospheric distillation column using support vector regression models and information analysis. Chem. Eng. Res. Des. 90, 2247–2261.

Zaidi, S., 2012. Development of support vector regression (SVR)-based model for prediction of circulation rate in a vertical tube thermosiphon reboiler. Chem. Eng. Sci. 69, 514–521.

Chapter 5 Design of flexible heat-integrated crude oil distillation systems

As discussed in Chapters 1 and 2, crude oil feedstocks processed in a petroleum refinery are inevitably subject to variability during operation. Refinery processes designed based on one type of crude oil feedstock may fail to accommodate variability of crude oil feedstocks. To enable the crude oil distillation system to operate satisfactorily over a range of feedstocks, it is necessary to incorporate some degree of flexibility at the design stage.

Chapters 3 and 4 present new methodologies for the design of crude oil distillation system that processes single crude oil feedstock. Numerical results show that the use of surrogate distillation column model is computationally efficient compared to rigorous simulation models. Thus, surrogate model is used in this chapter for the design of flexible crude oil distillation system.

This chapter addresses objectives four, five and six of this research, that is, (i) develop an optimisation framework that incorporates suitable distillation column models and pinch analysis to support the design of flexible crude oil distillation systems; (ii) propose an effective solution strategy to facilitate the search for flexible, cost-effective, and energy-efficient design option; (iii) demonstrate the capabilities of the proposed framework using industrially-relevant case studies. This chapter includes two papers, i.e., Publication 3 and Publication 4.

5.1 Introduction to Publications 3

Publication 3 presents a new approach for the design of flexible crude oil distillation systems that can process multiple feedstocks. Unlike the methodologies presented in literature (see Chapter 2), the developed approach takes into account important design

issues such as trade-offs between capital and energy cost, heat integration, and simultaneous optimisation of structural and operational degrees of freedom of the system.

The surrogate modelling approach for crude oil distillation units presented in Chapter 4 is modified and extended to take into account multiple crude oil feedstocks. Here, the artificial neural network and support vector machine are constructed for the individual crude oil to be processed. The inputs of the artificial neural networks include pump-around temperature drop and duty, stripping steam flow rate, feed inlet temperature, reflux ratio, and number of trays in column sections, while the outputs are product quality (ASTM T5% and T95% boiling temperature) and flow rate, stream enthalpy change, stream supply and target temperatures, and column diameter. As stated in Chapter 4, the input and output variables selected are those that have significant influence on the distillation column performance (e.g. energy demand, profit margin etc.)

To design the flexible crude oil distillation system, an optimisation framework that integrates the surrogate column models representing the individual crude oil feedstocks and pinch analysis is developed. The optimisation problem is posed as a mixed integer nonlinear programming problem with many scenarios. The size of the problem increases with increase in number of crude oil to be processed. This type of problem (non-convex MINLP) is very difficult to solve (Floudas, 1995; Biegler et al., 1997; Edgar et al., 2001). In an attempt to ease the numerical difficulty during optimisation, the problem is solved in two stages, i.e., the design stage (Stage 1) and the operating stage (Stage 2). Stage 1 aims to select the optimal structural design variables (number of trays in each column section) of the distillation column. In Stage 2, the operating variables (for the individual crude oils) that maximise profit margin are selected, where the column structure is fixed in the second stage. It is acknowledged that solving the problem in a single stage could lead to a better optimal solution, however, as the size of the problem becomes larger, single stage solution can be difficult and computationally demanding.

To effectively solve the non-convex MINLP problem, a hybrid-optimisation approach is proposed, combining stochastic and deterministic optimisation methods. Hence a genetic algorithm is applied to optimise the integer variables (number of trays in column sections) in Stage 1, while in Stage 2, successive quadratic programming is applied to select the optimal set of operating conditions for the individual crude oil to be processed. The proposed methodology is applied to an industrially-relevant case study that involves the design of a flexible crude distillation unit for the separation of three crude oil feedstocks (Tia Juana light, Bonny light, and Brent) into five refined petroleum products (light naphtha, heavy naphtha, light distillate, heavy distillate and residue). Numerical results indicate that a flexible crude oil distillation unit that can operate optimally across the three crude oil feed stock can be identified within the solution space. Furthermore, all product quality specifications are within their constraint limits, and the validation results indicate that the optimal distillation column obtained using the surrogate model is in good agreement with the rigorous simulation, confirming the effectiveness and accuracy of the proposed approach. The supporting information for this paper is presented in Appendix A.3.

5.2 Introduction to Publications 4

This paper presents a new methodology for the design of chemical processes under uncertainty. The approach presented in Section 5.1 is capable of dealing with problems involving a few operating scenarios. For a large number of operating scenarios (e.g. several crude oil feedstocks to be processed at different times of the year), the two-stage optimisation approach leads to a large-scale multi-scenario optimisation problem with discrete and continuous variables. The solution of such problem can be computationally intensive, leading to large CPU times. Publication 4 develops an alternative scenario-based flexible design approach that effectively deals with a large number of operating scenarios.

Overall, the proposed methodology comprises four main steps. Firstly, the system parameters that are subject to variability are identified and characterised using a probability distribution function (e.g. normal, triangle, uniform, Gaussian etc.). The

distribution is then discretised to create several points that represent distinct operating scenarios within the design space. This step can be skipped when the scenarios are defined *a priori*, for example different crude oil to be processed.

Secondly, process synthesis is carried out to generate alternative design for each representative scenario. This step is facilitated using a process superstructure (e.g. HEN model proposed by Yee and Grossmann (1990)) and optimisation algorithm such as outer approximation (Duran and Grossmann, 1986), genetic algorithm (Mitchell, 1998) etc. This step requires solving a complex mixed integer nonlinear programming problem.

Thirdly, each design alternative created in the previous step is assessed across the entire operating scenarios, taking into account different performance metric, e.g. economic, feasibility and risk metrics. This step is carried out by fixing the process configuration and unit size, then the system is optimised by solving a nonlinear programming problem, considering the entire operating scenarios.

Finally, a multi-criteria decision-making tool, Analytic Hierarchy Process (Saaty, 2008), is applied to select the most flexible and economically viable design among several alternative. Two case studies are used to demonstrate the capabilities of the proposed method. Numerical results indicate that the proposed methodology is capable of handling multi-scenario multi-criteria problem effectively. The supporting information for this paper is presented in Appendix A.4.

5.2 Publication 3

Ibrahim, D., Jobson, M., Lie J., Guillén-Gosálbez, G., 2017. Optimal Design of Flexible Heat-Integrated Crude Oil Distillation Units Chem. Eng. Res. Des. [To be submitted]

Optimal design of flexible heat-integrated crude oil distillation units using surrogate models

Dauda Ibrahim[1,*], Megan Jobson[1], Jie Li[1], Gonzalo Guillén-Gosálbez[2]

[1]Centre for Process Integration, School of Chemical Engineering and Analytical Science, University of Manchester, Manchester M13 9PL, UK

[2]Department of Chemical Engineering, Centre for Process Systems Engineering, Imperial College, South Kensington Campus, London SW7 2AZ, UK

Abstract

This paper presents a new optimization-based approach for the design of flexible heat-integrated crude oil distillation units that can process multiple crude oil feedstocks. In this work, the crude oil distillation unit is modeled based on artificial neural network and a support vector machine. The artificial neural network model predicts the performance of the distillation unit for a given crude oil feedstock. The inputs to the artificial neural network include the column structural variables and operating conditions, while the outputs are variables required to evaluate the column performance. The support vector machine classifier filters out infeasible design alternatives (i.e. designs that are unlikely to converge when simulated using a rigorous model) from the solution space. The artificial neural network models and support vector machines constructed for different crude oil feedstocks are integrated into a two-stage optimization framework in order to optimize the column structural variables and operating conditions. Pinch analysis is used to estimate minimum utility demand. An effective solution strategy that combines stochastic and deterministic optimization algorithms is applied to search for economically viable and flexible design alternatives that can operate over a given range of crude oil feedstocks while satisfying product

* Corresponding author
E-mail address: dauda.ibrahim@manchester.ac.uk

quality specifications. The capability of the proposed approach is illustrated using an industrially-relevant case study.

Keywords: Flexible process design, heat integration, genetic algorithm, artificial neural network, support vector machine

1 Introduction

Petroleum refineries typically process various crude oil feedstocks in order to accommodate changes in feedstock supply and product demand and to maximize profit margins. The need to process different crude oil feedstocks motivates the development of design methodologies that facilitate flexible operation of a refinery. Here, flexibility refers to the capability of a refinery process to accommodate various operating scenarios related to changes in crude oil feedstock, in market price of refined petroleum products, and in product specifications to meet the market requirements.

Crude oil distillation is the first major processing step in any petroleum refinery; therefore it is imperative that the distillation unit can process different crude oil feedstocks and can accommodate various operating scenarios. The crude oil distillation process is complex, capital- and energy-intensive, consuming fuel equivalent to 1 to 2% of the entire crude oil feedstock being processed (Szklo and Schaeffer, 2007). This tremendous amount of energy consumption is associated with significant CO_2 emissions and operating costs. Heat integration is usually implemented to improve the energy efficiency of the distillation process, where heat is recovered from 'hot' streams that require cooling and used to heat 'cold' streams, thus reducing both CO_2 emissions and operating cost.

A typical crude oil distillation system comprises crude oil distillation units and a heat recovery network in which the raw crude oil feedstock is preheated and partially vaporized. An atmospheric distillation unit, equipped with side-strippers and pump-around loops, is typically complemented by pre-separation units and/or a vacuum distillation unit. This work focuses on the atmospheric distillation unit. The two sub-systems exhibit strong interactions through cooling in pump-arounds, the column

condenser and product coolers. The design and operation of the crude oil distillation unit affect the design and operation of the heat recovery network and vice-versa. The interactions between the two sub-systems are of paramount importance in the design of the overall system: the distillation unit can be designed to create opportunities for heat recovery, while the heat recovery network can be designed to exploit those opportunities, leading to good design solutions, considering capital investment, product yield and operating costs..

In general, grassroots design of a crude oil distillation unit aims to select the distillation column structural variables (namely, feed tray location, pump-around and side-stripper location, number of trays in each section of the column) and its operating conditions (i.e. feed inlet temperature, duties and temperature drops of pump-arounds, stripping steam flow rates and reflux ratio). The requirement to process multiple crude oil feedstocks and/ or blends of crude oils introduces further challenges and complexity for distillation system design.

An industrially relevant design methodology needs to account for challenges related to the large number of design degrees of freedom, complex interactions within the system, heat recovery opportunities, and multiple feedstocks, as well as design optimality. Most existing design methodologies (e.g. Liebmann et al., 1998; Sharma et al., 1999; Ibrahim et al., 2017a, 2017b) focus on distillation of a single, specific crude oil. However, operating a distillation unit of a fixed design for other crude oil feedstocks can impact on the overall system performance (considering, for example, energy consumption, CO_2 emissions or profit) or can even lead to infeasible operation (i.e. failure of the design to satisfy product quality specifications or to respect hydraulic limits, e.g. related to flooding). Few published methodologies are available for the design of crude oil distillation units that process multiple crude oil feedstocks, and that of Bagajewicz and Ji (2001) does not adequately address trade-offs between capital investment, operating cost and product yield, and does not consider structural design variables along with operating conditions.

This work proposes a new optimization-based methodology for the design of flexible crude oil distillation units that can process multiple feedstocks. Extending previous work (Ochoa-Estopier and Jobson, 2015; Ibrahim et al., 2017b), the complex crude oil distillation unit is modeled using surrogate models regressed against data generated via multiple rigorous simulations. The surrogate model for the crude oil distillation unit, together with pinch analysis (to calculate minimum utility requirements), is applied within a two-stage optimization framework to facilitate the search for design alternatives that are economically viable and also able to process a range of crude oil feedstocks. The new approach addresses shortcomings of existing methodologies to consider trade-offs between capital investment and energy cost, and to account for relevant practical constraints (e.g. product yield and hydraulic limits of the column).

Section 2 of this manuscript provides a review of relevant research literature on design methodologies for crude oil distillation; Section 3 presents the new design approach for flexible crude oil distillation units. This two-stage approach applies surrogate models and both stochastic and deterministic optimization algorithms. In Section 4, an industrially-relevant case study demonstrates the capabilities and benefits of the proposed methodology, conclusions and recommendations for future work are presented in Section 5.

2 Previous work: design of crude oil distillation units

Methodologies for design (and optimization) of crude oil distillation units have been presented and developed over many decades. Conventional design methods (Nelson, 1958; Watkins, 1979; Jones, 1995) rely on heuristic rules, empirical correlations, and simple mass and energy balance calculations. The crude oil distillation unit and heat recovery network are considered separately, without accounting for interactions between the two sub-systems. These methods require trial and error and require significant engineering effort.

Several researchers have accounted for interactions between the distillation unit and the heat recovery network by considering heat recovery together with design of the

distillation unit. Liebmann et al. (1998) pioneered an approach that applied rigorous column simulation and pinch analysis sequentially to facilitate the search for an energy-efficient column configuration and its operating conditions. Sharma et al. (1999) proposed a two-step design approach: first material and energy balance are applied to each section of a proposed crude oil distillation column and then the corresponding temperature–enthalpy data are used to construct the column grand composite curve (Dhole and Buckingham, 1994) to maximize heat recovery without adversely affecting separation. Bagajewicz and Ji (2001) applied the related concept of a heat demand–supply diagram to identify the best location for pump-arounds, while taking into account the effect of stripping steam on heat recovery, as well as to identify suitable operating conditions for a column processing a range of crude oil feedstocks. The methods proposed by Liebmann et al. (1998), Sharma et al. (1999) and Bagajewicz and Ji (2001) assume a fixed number of trays in each column section, and thus do not consider the trade-off between capital investment and energy cost. Furthermore, the design is not optimized.

Ibrahim et al. (Ibrahim et al., 2017a) proposed a superstructure for optimizing the structure of a crude oil distillation unit, extending the approach developed by Caballero et al. (2005) for simple columns, where pinch analysis is used to account for heat recovery. The crude oil distillation unit is modeled in Aspen HYSYS, while the optimization framework is developed in MatLab. An interface (AspenTech, 2011) is established between the two software packages to exchange information during optimization. In this approach, both the column structural variables (number of trays in each column section) and operating conditions are optimized. Using the rigorous column model for system optimization was shown to be computationally expensive; subsequent work (Ibrahim et al., 2017b) improved the computational performance of the methodology through the use of surrogate models based on artificial neural networks in place of the rigorous column model, extending the approach of Ochoa-Estopier and Jobson (2015). Support vector machines were applied to filter the search space to improve the quality of solutions and to expedite the optimization. The

resulting optimization framework was shown to be effective for generating cost-effective design solutions.

The design methodologies presented above focus on crude oil distillation units that process a specific feedstock. While the resulting design may perform well for a given feedstock, the design is likely to perform poorly for alternative crude oil feedstocks with properties that are significantly different from those considered during the design. Although Bagajewicz and Ji (2001) presented an approach for selecting the operating conditions of a crude oil distillation unit that can process light, medium and heavy crude oils, the distillation column configuration is not optimized. To the best of the authors' knowledge, no approaches have been identified that apply distillation models, together with optimization algorithms, to design flexible heat-integrated crude oil distillation units that can perform well for a range of crude oil feedstocks.

This manuscript proposes a new methodology for the design of flexible crude oil distillation units that can process multiple crude oil feedstocks. The method extends surrogate modeling techniques (Ibrahim et al., 2017b) together with stochastic and deterministic optimization techniques In this work, a separate artificial neural network model (Beale et al., 2015) and support vector machine (Vapnik, 1995) representing the distillation unit is constructed for each crude oil feedstock to be processed. These models and associated feasibility constraints (Jobson et al., 2017) are applied in a two-stage optimization procedure (Grossmann and Guillén-Gosálbez, 2010); a hybrid stochastic-deterministic approach is proposed to optimize the structural variables of a column that can operate for all proposed scenarios, and the operating conditions of the column are optimized for each feedstock under consideration. The approach considers performance in terms of product yield, product quality specifications, column hydraulic constraints, capital investment and operating costs.

3 Proposed methodology

This section presents the new framework for the design of flexible crude oil distillation units that applies surrogate models. Figure 3.1 provides an overview of the approach,

showing its four main steps: problem initialization, data sampling, surrogate column modeling and two-stage optimization-based design. First, the information required to simulate the distillation of the crude oil feedstocks is collated. These data include the crude oil assay and product quality specifications. An initial feasible design, i.e. the column structure and operating conditions, is established. This information, together with the knowledge of the crude oil distillation unit, is applied to identify appropriate dependent and independent variables. Samples can then be generated: these are the results of multiple rigorous simulations. The samples for each crude oil are used to construct surrogate models of the crude oil distillation column. The surrogate models for all the crude oils and a pinch analysis algorithm are applied in an optimization framework, where a hybrid stochastic-deterministic algorithm is used to search for column configuration that can operate optimally across all the crude oil to be processed. The details of each step are covered in subsequent sections.

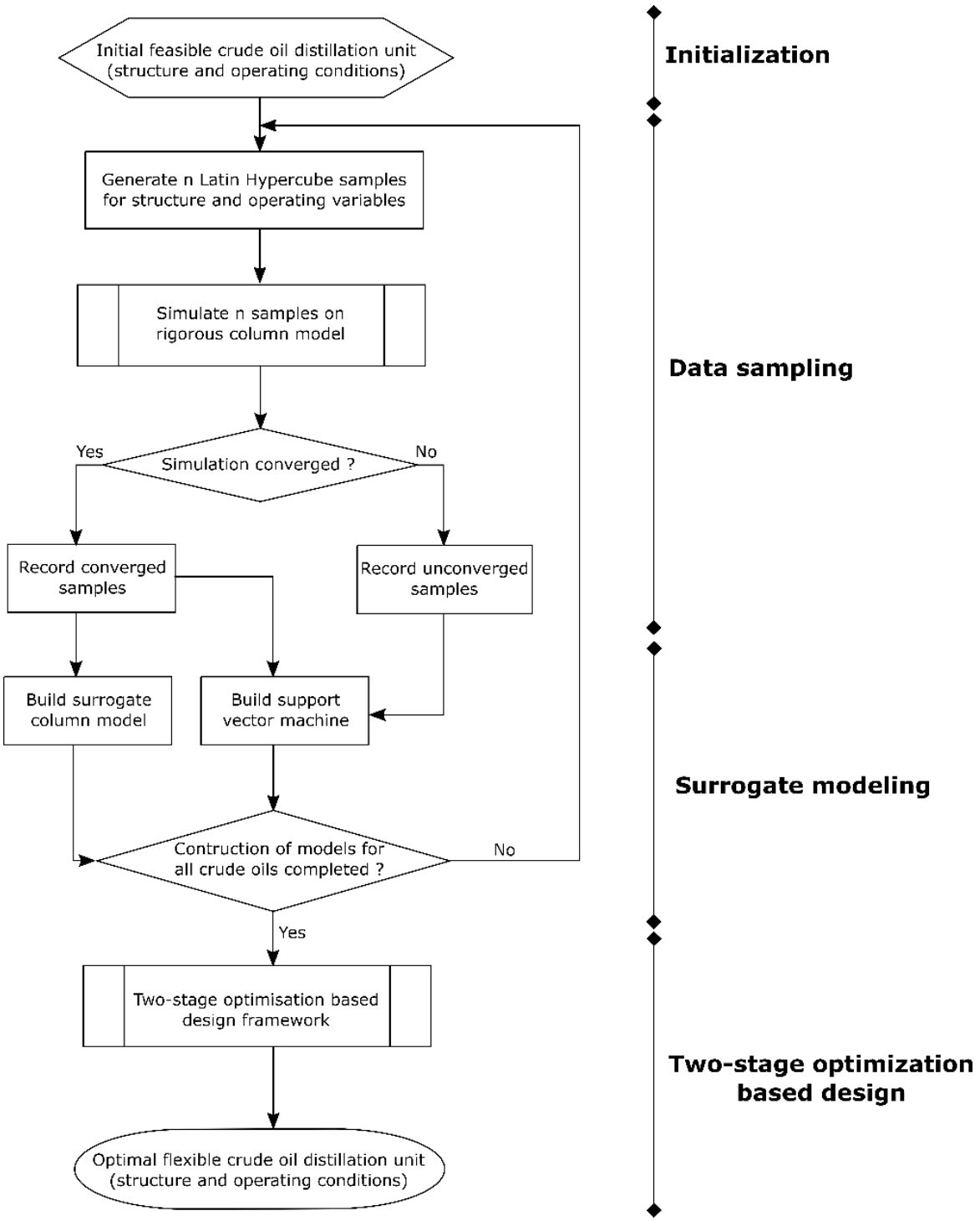


Figure 3.1 Framework for the design of flexible crude oil distillation units

3.1 Initialization

Surrogate models are used to simplify calculations and to improve computational performance, compared to more rigorous models, while also providing a relatively accurate representation of the process being modeled. A key challenge is to select an

appropriate set of dependent and independent variables and a form of the surrogate model that balances accuracy and ease of computation.

For the case of design of a crude oil distillation column, the independent variables include the structural and operational design degrees of freedom; these are the inputs of the surrogate model. The dependent variables – the outputs of the surrogate model – allow calculation of the value of the objective function, and also allow relevant inequality constraints to be checked, for example those related to product quality. Ibrahim et al. (2017b) propose a suitable set of dependent and independent variables for modeling a crude oil distillation unit for design purposes. In this work, the independent variables are: the feed inlet temperature, stripping steam flow rates, pump-around temperature drops and duties, reflux ratio and number of trays in each column section. The dependent variables are: supply and target temperatures of heaters and coolers, variables characterizing product quality (for example, ASTM T5% and T95% boiling temperatures), product flow rates and diameter required in each column section.

Prior to optimization, it is necessary to define the limits of the solution space to ensure that the optimized solutions are meaningful and potentially realizable. Therefore, before sampling, sensitivity analyses are carried out. Simulations for which the specified independent variables do not converge are assumed not to lead to realistic solutions; the corresponding inputs are identified as 'infeasible'. Initially, bounds are defined for each independent variable, i.e., $\pm$ 10 °C for temperature related variables, $\pm$ 25 % for stripping steam and duty, $\pm$ 2 for reflux ratio and -2/+1 for number of trays in column sections. Single-variable sensitivity studies (via multiple rigorous simulations) for all independent variables are therefore carried out to determine the range for which convergence is obtained and thus to define their upper and lower bounds for sampling.

An important premise of this work is that the vector of inputs corresponding to an unconverged simulation signifies a set of independent variables that will not lead to a feasible solution. A 'feasible' solution is one that meets all problem constraints, including material and energy balances, phase equilibrium relations and problem-specific constraints, e.g. those relating to product quality or yield. It is acknowledged

that non-convergence is not absolutely correlated with feasibility – in some cases, convergence could be achieved with better initialization or additional iterations. Furthermore, it is recognised that not all converged solutions will meet all problem-specific constraints, since the rigorous simulation model only imposes a limited set of constraints, according to the number of degrees of freedom in the model; a penalty function applied during optimization may be needed to impose additional constraints. Nevertheless, the intention of reducing the sampling space is to eliminate samples that do not converge, so as to avoid generating spurious optimal solutions that simply cannot satisfy all constraints.

3.2 Data sampling

Once the sets of dependent and independent variables are identified, and bounds are defined for each independent variable, samples can be generated within the design space. The sampling approach applied in this work comprises two main steps:

First, random samples are generated for each independent variable. Various types of sampling techniques such as Monte Carlo sampling, Latin hypercube sampling, and Hammersley are widely used to generate random samples (Diwekar and Kalagnanam, 1997). Without loss of generality, the Latin hypercube sampling is used in this work, since it has been shown to facilitate statistically useful distributions of samples (Subramanyan et al., 2011).

Second, all the Latin hypercube samples generated are sent as inputs to a rigorous simulation of the crude oil distillation column; the corresponding outputs are recorded. In this work, the rigorous distillation model is built in Aspen HYSYS, while the Latin Hypercube sampling is implemented in MatLab. An interface is established to link the two software packages (AspenTech, 2011) for automated exchange of information during sampling.

Not all the samples generated using the Latin hypercube sampling algorithm lead to converged simulations. To facilitate building an accurate and robust surrogate column model, the samples classified as ‘converged’ and ‘unconverged’, according to whether

the simulation using rigorous models converged. Only the converged samples are used to build the surrogate models, as for these samples there is a meaningful relationship between the inputs and outputs. The entire sample set is classified (as converged and unconverged samples); this classification underpins the construction of feasibility constraints using a support vector machine.

3.3 Surrogate modeling of the distillation unit

In this work, the concept of an artificial neural network is applied to build surrogate models of the crude oil distillation unit. Comprehensive descriptions of artificial neural networks can be found elsewhere (Basheer and Hajmeer, 2000; Himmelblau, 2008). Here, the artificial neural network toolbox embedded in MatLab 2015a is used to facilitate the construction, validation, and testing of surrogate models. The artificial neural network correlates system inputs and outputs, either by correlating multiple inputs with multiple outputs, or by correlating multiple inputs to a single output. The former requires a relatively small number of networks for a given problem and leads to a relatively compact model, where compactness is invaluable for systems with a large number of outputs, for example the crude oil ditillation system.

The structure of the artificial neural network consists of a multilayer perceptron with one input layer, a variable number of hidden layers and one output layer (Beale et al., 2015). The hidden and output layers comprise a specified number of neurons. Given that algorithms for selecting the number of layers and numbers of neurons are challenging or time consuming to implement, in this work, the number of layers and numbers of neurons are determined by trial and error: alternative networks are generated for different numbers of layers and numbers of neurons and the accuracy of the resulting network is assessed for the 'testing' data in terms of the 'prediction error.' Following the findings of published studies (Basheer and Hajmeer, 2000; Himmelblau, 2008), the sigmoid function was chosen to connect hidden layers and the linear function was selected to connect to the output layer.

The data set generated in the sampling step is used to train the networks. The converged data set is randomly divided into three subsets, for training (70%),

validation (15%) and testing (15%), where these splits follow available guidance (Beale et al., 2015). First, the artificial neural networks are regressed against the training set, where the aim is to minimize the difference between predicted outputs and the corresponding values in the sample set (note that the data set is scaled between -1 and 1 in order to improve training (Beale et al., 2015)). Then the validation set is utilized during the training process to avoid over-fitting (Beale et al., 2015). Finally, the testing set is used to check the accuracy of the artificial neural network model, where the performance criterion used in this work is the coefficient of determination (Allen, 1997). The artificial neural network toolbox in MatLab enables all these steps and permits the user to choose the size and structure of the model, the performance criterion and to review the performance for each of the data subsets. In addition to the artificial neural network model for the column, a model is developed to partition the solution space, in order to promote the search for feasible solutions (i.e. those that are likely to converge). The aim is to restrict the optimization search space to the region comprising feasible solutions, to enhance computational efficiency and increase the likelihood of identifying optimal solutions that are also feasible.

A support vector machine is constructed to partition the solution space. The entire data set (both converged and unconverged samples) is randomly split into a training set and a validation set. In this work, a 75%–25% split is used, based on previous work (Ibrahim et al., 2017b). The MatLab function *fitcsvm* is used to train and validate the support vector machine (MATLAB, 2014). In this work, the support vector machine is chosen to be a third-order polynomial that generates outputs equal to +1 and − 1, denoting feasibility and infeasibility, respectively.

Sampling and surrogate modeling, resulting in artificial neural network models and support vector machine, is carried out separately for each type of crude oil under consideration. These models are then applied within the optimization framework discussed below.

3.4 Optimization-based design

The two-stage optimization framework is presented below in terms of its mathematical formulation, solution strategy for solving the optimization problem and implementation.

In developing the mathematical formulation, the following assumptions are made:

i. Each crude oil feedstock represents a discrete operating scenario;
ii. Only one crude oil is processed per operating scenario and blending is not considered;
iii. Weighting factors are used to reflect the relative importance of a scenario.

The first stage considers the column structure, where a structure is proposed and evaluated in the second stage. Structural decisions comprise the number of trays in each column section; the number and locations of feeds, pump arounds, draw streams, vapour return streams, reboilers and stripping steam feeds are fixed (i.e. are located at the top of bottom of a given section, where the number of trays in that section is a variable). In the second stage, the fixed structure is applied to all the operating scenarios (i.e. to all the crude oil feedstocks to be processed) and the operating conditions for each scenario are optimized to achieve the best overall performance. The overall performance is the weighted average of the performance of the individual scenarios. Model M1 (eq. 1) summarizes the problem formulation, which takes the form of a mixed integer nonlinear programming (MINLP) problem:

$$[\text{M1}] \qquad \max_{x_s, y} \quad U(\boldsymbol{y}) + \sum_{s=1}^{S} w_s \cdot V_s(\boldsymbol{x}_s, \boldsymbol{y}) \tag{1}$$

$$s.t. \quad g_s(\boldsymbol{x}_s, \boldsymbol{y}) \leq 0$$

$$h_s(\boldsymbol{x}_s, \boldsymbol{y}) = 0$$

$$\boldsymbol{x}_{L,s} \leq \boldsymbol{x}_s \leq \boldsymbol{x}_{U,s}$$

$$\boldsymbol{y}_L \leq \boldsymbol{y} \leq \boldsymbol{y}_U$$

$$\boldsymbol{x} \in X, \boldsymbol{y} \in Y$$

where $U(\boldsymbol{y})$ denotes the objective function of the first stage, i.e. the annualized capital cost of the distillation unit for the set of discrete variables $\boldsymbol{y}$, where $\boldsymbol{y}$ is the vector representing the number of trays in each column sections; $V_s(\boldsymbol{x}_s, \boldsymbol{y})$ denotes the second stage objective function and $\boldsymbol{x}_s$ is the vector representing the operating variables for a given scenario, s. The weighting factor w is pre-specified, e.g. according to the probability or importance of each scenario. X, Y, S represent the sets of operating conditions, $\boldsymbol{x}$, discrete variables, $\boldsymbol{y}$, and operating scenarios, s; g and h represent inequality constraints and equality constraints, respectively. Subscripts L and U denote lower and upper bounds.

3.4.2 Optimization algorithm

Optimization methods are broadly categorized as deterministic and stochastic, where deterministic approaches require information about gradients of the objective function and constraints in the search for optimal solutions. Unlike deterministic methods, stochastic methods apply random numbers in the search for optimal solutions. An advantage of stochastic methods is that, by introducing randomness and ignoring gradients, they are less likely to become trapped near locally optimal solutions. Stochastic methods can suffer from computational inefficiency as a result of the randomness and it is difficult to guarantee that an optimum solution, whether local or global, has been obtained. Stochastic search methods such as genetic algorithms, simulated annealing and particle swarm optimization have been applied extensively in process design optimization (Edgar et al., 2001).

To mitigate the disadvantages of both approaches, while also taking advantage of their strengths, this work combines stochastic and deterministic methods in the search for optimal solutions. In the first stage, which is dominated by discrete design alternatives, the column structure is optimized using stochastic methods. The second stage applies deterministic search methods, as only continuous variables are involved.

In particular, a genetic algorithm is applied in the first stage, since it has been shown to be well suited for solving large-scale highly combinatorial problems (Steimel et al., 2013). Successive quadratic programming (Edgar et al., 2001) is used to optimize

continuous variables in the second stage. This deterministic algorithm is selected because of its ability to guarantees optimality (local) of a solution.

3.4.3 Implementation of the optimization framework

The artificial neural network model of the column and the support vector machine describing the solution space can be created in MatLab, where these models are fitted to samples generated and recorded using an interface between MatLab and Aspen HYSYS. The models are conveniently formulated as functions in MatLab and can readily be called to evaluate the performance of specific proposed solutions and to optimize the system performance. On the other hand, the mathematical functions expressing these models can be coded in other equation-based software environments, such as GAMS and Fortran, and their optimization capabilities can be exploited.

In this work, the surrogate models provided as MatLab functions are applied for optimization. In the first stage, the genetic algorithm *'ga'* in the global optimization toolbox of MatLab 2015a is employed; in the second stage, the successive quadratic programming algorithm *fmincon* is applied.

4 Case study

4.1 Problem description

The case study, which aims to demonstrate the capabilities of the proposed design methodology, considers the design of a crude oil distillation unit that is flexible enough to process three specified crude oil feedstocks and that is optimized in terms of profit margin.

The three feedstocks of interest are Tia Juana, Bonny light and Brent crude oil (Watkins, 1979; TOTAL, 2015) into five products: light naphtha, heavy naphtha, light distillate, heavy distillate, and residue. Each feedstock represents an operating scenario, and it is assumed that each is processed for 4 months each year, so the three scenarios are equally weighted. The flow rate of each crude oil is 100,000 bbl day^{-1}

(662.4 m^3 h^{-1}). The crude oil assay information for the three crude oils is presented in Table S1 of the supplementary material.

Figure 4.1 illustrates the crude oil distillation column configuration: the main column has three pump-arounds and is connected to three side-strippers. The main column has five sections (S-1 to S-5), and the three side-strippers represent three sections (S-6 to S-8).

As the methodology requires an initial feasible design, the initial stage distribution and operating conditions for Tia Juana light crude oil are adapted from the design of Chen (2008), presented also in Ibrahim et al. (2017a). Trial and error is applied to adapt the stage distribution to the two other crude oils, i.e., Bonny light and Brent. The details of the initial stage distribution and initial operating conditions for the three crude oils are presented in Tables S2 and S3 of the supplementary material.

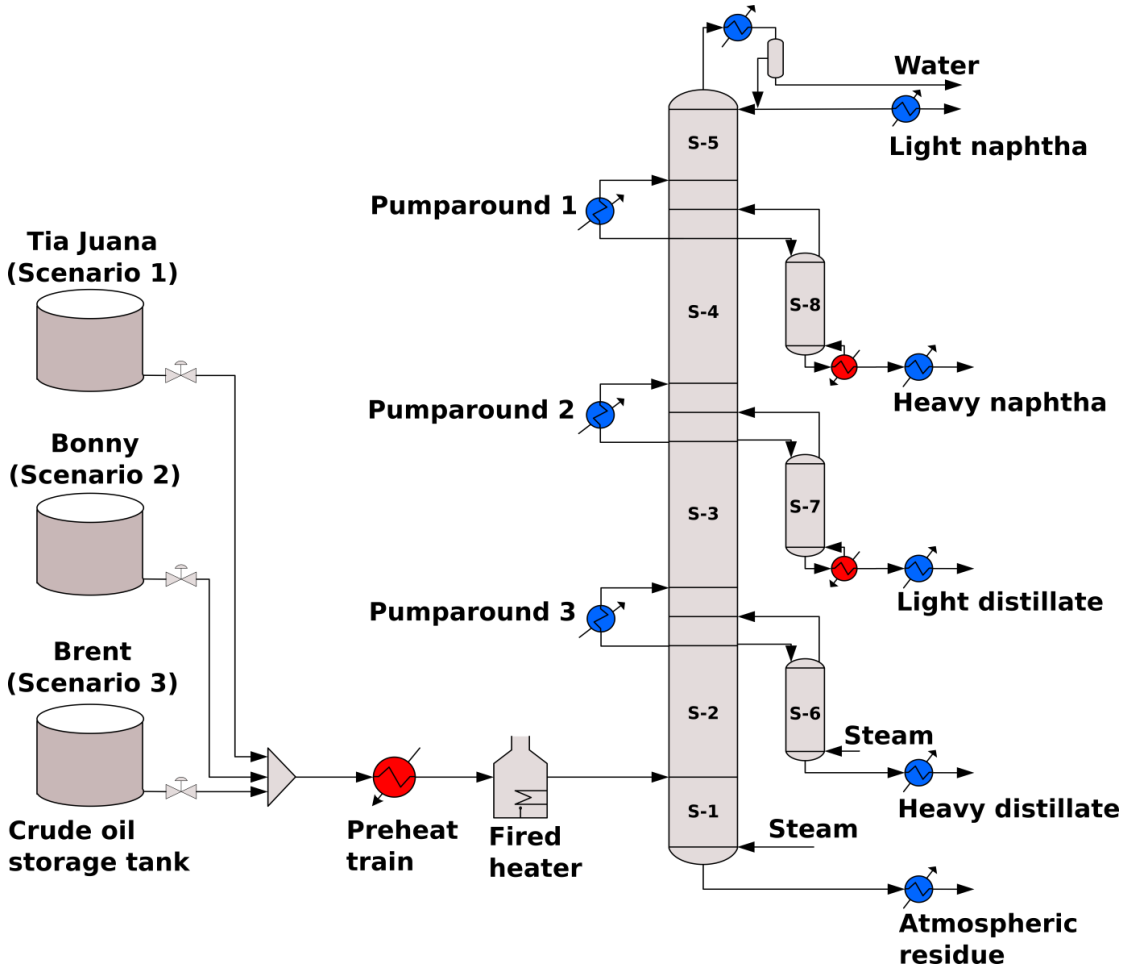


Figure 4.1 Configuration of crude oil distillation unit to be optimized

The initial flow rates of products and their respective specifications (in terms of ASTM T5% and T95% boiling temperatures) are presented in Table S4 of the supplementary material. The operating pressure is uniformly 2.5 bar. The main column and side-strippers are equipped with sieve trays spaced 0.61 m apart. The required diameter for each section is estimated using the tray sizing utility in Aspen HYSYS, assuming that the approach to jet flooding should be at most 80% and that downcomer backup should not exceed 50%.

The economic evaluation of the design assumes an interest rate of 5%, a plant life of 20 years and 8700 operating hours per year. Table S5a presents the transfer value of intermediate products and costs of the crude oils; Table S5b presents the cost of utilities. A minimum approach temperature of 30℃ is used to calculate the minimum utility requirement.

4.2 Initialization and data sampling

The data in Section 4.1 are used to set up a rigorous simulation model of the crude oil distillation unit in Aspen HYSYS (v8.6). For each independent variable, a sensitivity analysis is carried out. Based on the range of the variable for which simulation convergence was obtained, bounds were identified for the three crude oils; these are presented in Tables S6 to S9 of the supplementary material.

Next, 7000 Latin Hypercube samples are generated for each crude oil. Sampling took about 2 seconds per simulation on a HP desktop PC with an Intel® Core i5 processor running at 3.2 GHz and 8 GB of RAM.

4.3 Surrogate model of crude oil distillation column

In this case study, the crude oil distillation unit outputs are divided into seven groups as shown in Table 4.1, following previous work (Ibrahim et al., 2017b). Each group of outputs is correlated with the inputs using a dedicated artificial neural network. The 18 inputs are the number of trays in each column section (8), pump-around temperature drops (3) and duties (3), stripping steam flow rates (2), feed inlet temperature (1), and reflux ratio (1). Trial and error was used to select the structure of the surrogate model,

as discussed in Section 3.3. Each of the 7 artificial neural networks has one input layer, one hidden layer and one output layer, where the hidden layers each have ten neurons. Table 4.1 summarizes the purpose of the artificial neural network models and the associated number of outputs. The system comprises of 11 target temperatures, three of which are predicted using ANN number seven (see Table 4.1), five of which are exit temperature of product coolers (values fixed at 40℃ for LN, HN and LD cooler; 50℃ and 100℃ for HD and residue coolers respectively), and three of which are return temperatures of pump-arounds (values equal to pump-around supply temperature less temperature drop—input variable)

Table 4.1 Description of artificial neural network models

ANN number	Description	Outputs
1	Product quality (T5 %)	5
2	Product quality (T95 %)	5
3	Product flow rate	5
4	Column diameter	8
5	Enthalpy change	9
6	Supply temperature	11
7	Target temperature*	3

Of the 7000 samples, only the converged samples were used to fit the artificial neural network model, i.e. 4130 samples for Tia Juana light; 6782 for Bonny light and 4419 for Brent. The performance of the artificial neural network models representing the distillation column of Tia Juana, Bonny light, and Brent is presented in Table 4.2 in terms of the coefficient of determination. As can be seen in Table 4.2, the coefficient of determination for all the artificial neural networks ranges from 0.992 to 0.999. These results reveal that all the networks can explain at least 99% of the variance of the outputs (from the rigorous model). The performance of each artificial neural network is calculated using data in the testing set (see Section 3.3)

Figures S1 to S3 of the supplementary material provide parity plots which illustrate how almost all of the models are in very good agreement with the results of the rigorous simulation models, apart from models 4 and 5 for Brent crude oil, where there

are significant, but randomly distributed, deviations for the models predicting column diameter and enthalpy change.

Table 4.2 Performance of artificial neural networks (coefficient of determination)

Serial no.	Network ID	Description	Outputs	Tia Juana	Bonny light	Brent
1	ANN_1	T5 %	5	0.999	0.999	0.999
2	ANN_2	T95 %	5	0.999	0.999	0.999
3	ANN_3	Product flow	5	0.999	0.999	0.999
4	ANN_4	Column diameter	8	0.999	0.999	0.992
5	ANN_5	Enthalpy change	9	0.999	0.999	0.998
6	ANN_6	Supply temp.	11	0.999	0.999	0.999
7	ANN_7	Target temp.	3	0.999	0.999	0.999

To increase confidence in the artificial neural network models, an error analysis is carried out, where the error is the difference between the rigorous simulation results and the predictions of the artificial neural network. The average absolute error is calculated for temperature-related variables and is presented in Table 4.3; Table 4.4 presents the average relative error is calculated for other variables. These calculations are carried out using new data set (100 converged simulation) generated via Latin Hypercube sampling technique presented in Section 3.1.

As can be seen in Table 4.3, the highest average absolute error for stream supply and target temperatures is 1.5℃. All other temperature predictions are within 1°C of the more rigorous prediction. Table 4.4 shows that prediction of product flow rates is good (less than 1.4% error). The maximum average error in exchanger duties is case-specific, ranging from 2.2% for Tia Juana to 16.4% for Brent. The errors in the prediction of required diameter is around 8%; the largest errors relate to side-stripper diameters. The difference of around 0.3 m in the stripper diameter gives rises to a difference of around 575.9 $/y in the capital cost of the side-stripper, amounting to a difference of less than 0.0006% in the net profit for the process. Overall, the statistical test reveals that predictions of the artificial neural network models representing the distillation columns processing Tia Juana, Bonny light and Brent crude oils are in good agreement with those of the more rigorous Aspen HYSYS simulations.

Table 4.3 Average absolute error (°C) calculated for the ANN representing each crude oil

Variables	Tia Juana	Bonny light	Brent
Product quality (°C)			
LN T5%	0.098	0.099	0.194
HN T5%	0.250	0.701	0.934
LD T5%	0.454	0.059	0.258
HD T5%	0.358	0.839	0.438
RES T5%	0.269	0.666	0.434
LN T95%	0.009	0.103	0.725
HN T95%	0.028	0.004	0.496
LD T95%	0.013	0.527	0.002
HD T95%	0.002	0.003	0.004
RES T95%	0.027	0.215	0.344
Supply temperature (°C)			
ADU condenser	0.173	0.263	0.351
LN cooler	0.054	0.101	0.116
HN cooler	0.196	0.361	0.518
LD cooler	0.199	0.269	0.332
HD cooler	0.732	0.778	0.609
RES cooler	0.423	0.420	0.403
PA1	0.322	0.535	1.519
PA2	0.208	0.453	0.620
PA3	0.277	0.360	0.244
HN reboiler	0.220	0.341	0.452
LD reboiler	0.235	0.395	0.491
Target temperature (°C)			
ADU condenser	0.057	0.066	0.117
HN reboiler	0.178	0.266	0.501
LD reboiler	0.165	0.096	0.313

Table 4.4 Average relative error calculated for the ANN representing each crude oil

Variables	Tia Juana	Bonny light	Brent
Product flow rate (m h^{-1})			
LN	0.169	0.482	0.266
HN	0.368	0.566	0.588
LD	0.168	0.276	0.334
HD	0.736	1.388	0.641
RES	0.084	0.504	0.159
Exchanger duty (MW)			
ADU condenser	0.295	0.250	0.342
LN cooler	2.227	1.232	4.073
HN cooler	1.035	0.572	0.590
LD cooler	0.301	0.398	0.739
HD cooler	1.825	1.279	3.798
RES cooler	0.248	0.414	1.073
HN reboiler	2.122	5.684	16.349
LD reboiler	0.683	2.829	8.979
Fired heater	0.101	0.035	0.213
Column diameter (m)			
Main column			
Section 1	0.898	1.105	1.262
Section 2	0.903	1.029	2.492
Section 3	0.915	0.970	0.828
Section 4	0.942	0.856	0.746
Section 5	0.767	0.795	1.570
HN side-stripper	2.409	4.850	8.257
LD side-stripper	0.345	2.165	7.382
HD side-stripper	3.012	3.313	2.034

Table 4.5 shows the validation results of the support vector machine for Tia Juana, Bonny light and Brent. For each crude oil, 75% of the total samples are used to construct the support vector machine, while the remaining 25% is used to check the effectiveness of the SVMs.

The output of the support vector machine can be either 1 (converged samples) or –1 (unconverged samples). In Table 4.5, a true prediction is when the support vector machine predicts the output of the rigorous model correctly, for example, if the rigorous model converged and the support vector machine predicts 1. On the other hand, false prediction is when the support vector machine wrongly predicts the output of the rigorous model. From the optimization perspectives, true positive prediction retains the feasible region containing converged samples (feasible designs), while true

negative prediction removes infeasible region (containing infeasible designs) from the design space. As can be seen in Table 4.5, a considerable portion of the feasible region of the three crude oils is retained: 91.1%, 97.3% and 93.4, while a moderate portion of the infeasible region is removed from the design space: 58%, 7%, and 70%. In the other case, false positive prediction means feasible designs have undesirably removed from the design space, while false negative prediction means infeasible designs are retained within the design space, which is undesirable; leading to large CPU time, but the final solution is not affected.

Table 4.5 Performance of support vector machines

Prediction class	**Tia Juana**		**Bonny light**		**Brent**	
	True prediction	False prediction	True prediction	False prediction	True prediction	False prediction
Positive class (Converged samples, 1)	966 [91.1%]	95 [8.9%]	1646 [97.3%]	47 [2.7%]	998 [93.4%]	71 [6.6%]
Negative class (Unconverged samples, 0)	396 [58%]	293 [42%]	4 [7%]	53 [93%]	477 [70%]	204 [30%]
Overall: Correct prediction	78%		94.3%		84.3%	
Wrong prediction	22%		5.7%		15.7%	

4.4 Flexible crude oil distillation unit – optimization results

The artificial neural network column models and support vector machine classifiers for the three crude oils are implemented into Model M1 (see Section 3.4) in order to search for the best column structure as well as the operating conditions for each of the three crude oils. More specifically, the equality constraints in Model M1 are represented as;

$$
\begin{aligned}
h_{1,s}:&\quad [T5_i] = ANN_{1,s}(x_s, y) && i = 1,2,3,\dots,N_{product}\\
h_{2,s}:&\quad [T95_i] = ANN_{2,s}(x_s, x_o) && i = 1,2,3,\dots,N_{product}\\
h_{3,s}:&\quad [F_i] = ANN_{3,s}(x_s, x_o) && i = 1,2,3,\dots,N_{product}\\
h_{4,s}:&\quad [D_j] = ANN_{4,s}(x_s, x_o) && j = 1,2,3,\dots,N_{section}\\
h_{5,s}:&\quad [E_k] = ANN_{5,s}(x_s, x_o) && k = 1,2,3,\dots,N_{stream}\\
h_{6,s}:&\quad [TS_l] = ANN_{6,s}(x_s, x_o) && l = 1,2,3,\dots,N_{stream}\\
h_{7,s}:&\quad [TT_l] = ANN_{7,s}(x_s, x_o) && l = 1,2,3,\dots,N_{stream}\\
h_{8,s}:&\quad [CC] = SVM_s\,(x_s, x_o)\\
h_{9,s}:&\quad CC = 1
\end{aligned}
\qquad (2)
$$

where $T5_i$ and $T95_i$ represent the ASTM boiling temperatures of product i at 5 vol% and 95 vol% vaporization; F_i is the volumetric flow rate of product i; D_j is the diameter of section j; E_k is the enthalpy change of stream k; TS_l and TT_l are the supply and target temperatures of stream l; CC represents the convergence criterion predicted by the support vector machine (+1 or -1); ANN and SVM represent the artificial neural network model and support vector machine function, respectively.

The inequality constraints in Model M1 can be expressed as

$$\left.\begin{aligned}
g_{1,s}&: \quad lb_j \le N_j \le ub_j \qquad j = 1,2,3,\dots,N_{section}\\
g_{2,s}&: \quad lb_l \le Q_{PA,l} \le ub_l \qquad l = 1,2,3\\
g_{3,s}&: \quad lb_l \le \Delta T_{PA,l} \le ub_l \qquad l = 1,2,3\\
g_{4,s}&: \quad lb_m \le F_{s,m} \le ub_m \qquad m = 1,2\\
g_{5,s}&: \quad lb \le R \le ub\\
g_{6,s}&: \quad lb \le T_F \le ub\\
g_{7,s}&: \quad lb_i \le T5_i \le ub_i \qquad i = 1,2,\dots,N_{product}\\
g_{8,s}&: \quad lb_i \le T95_i \le ub_i \qquad i = 1,2,\dots,N_{product}
\end{aligned}\right\} \quad (3)$$

where N_j is the number of active trays in column section j; $Q_{PA,l}$ and $\Delta T_{PA,l}$ are the duty and temperature drop of pump-around l; $F_{S,m}$ is the steam flow rate of stream m; R is the overhead reflux ratio; T_F is the feed inlet temperature; $T5_i$ and $T95_i$ are the boiling temperature of product i at 5% and 95% vaporization, for example, according to ASTM standards.

In Model M1, g_1 to g_6 are bounds on independent structural and operational variables, while g_7 and g_8 represent constraints on dependent variables, related to product quality. The inequality constraints pertaining to all these variables can be included in the objective function as a penalty function (Edgar et al., 2001). In this way, Model M1 is converted to an unconstrained optimization problem, which is considerably easier to solve than the original constrained problem. The resulting reformulation, M2, is:

$$[\text{M2}] \qquad \max_{x_s, y} \; U(\boldsymbol{y}) + \sum_{s=1}^{S} w_s \cdot V_s(\boldsymbol{x}_s, \boldsymbol{y}) + \left[\Pi \sum_{i=7}^{8} \left[\max(0, (g_{i,s})) \right]^2 \right] \qquad (4)$$

$$s.t. \quad h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8, h_9$$

$$g_1, g_2, g_3, g_4, g_5, g_6$$

where g_i denotes the inequality constraints g_7 and g_8; Π is a large scalar value that amplifies the consequences of violating constraints g_7 and g_8, which helps to ensure that the optimal solution meets all the constraints.

4.4.1 Objective function

Optimization-based design aims to select the best design option among several alternatives according to a given performance criterion. In process synthesis and design, many different performance criteria can be used, including economic indicators such as profit margin, total annualized cost, net present value, and other indicators, such as CO_2 emissions and safety. The choice of performance criterion depends significantly on the design objectives. In this work, the aim is to design a flexible crude oil distillation system that maximizes the economic performance, in particular, profit margin.

The first stage of the optimization procedure aims to fix the structure of the crude oil distillation column, without considering the associated operating conditions for any of the scenarios. Therefore the objective function at this stage can only consider capital expenditure. In this work, the objective to be minimized is the annualized capital cost (ACC) of the distillation unit, which is the sum of installed column shell cost (S_C) and total tray cost (T_C). Without loss of generality, the capital cost correlations of Guthrie (1969) are used, given the number of trays and diameter in each section of the column. These predicted costs are multiplied by a suitable installation factor and cost index to update the costs. Further details are provided in Section S3 of the supplementary material. The annualization factor (A_f) described by Smith (2005) is applied to apportion the total capital cost over the plant life (t), for a specified interest rate (i).

$$ACC = (S_c + T_C) \cdot A_f \tag{5}$$

$$A_f = \frac{i(1+i)^t}{(1+i)^t - 1} \tag{6}$$

The objective to be maximized at the second stage, *V*, is the profit margin (*PM*), i.e. transfer value (*or "transfer price"*) (*R*) of intermediate products less the cost of the crude oil feedstock and the operating cost (*OC*), specifically, the cost of stripping steam and of heating and cooling utilities. Note that Eq. (5) to (7) relate to a given operating scenario and the price of crude oil is different in each scenario (depends on crude oil type been processed).

$$PM = R - C_{CO} \cdot F_{CO} - OC \tag{7}$$

$$R = \sum_{i=1}^{N} F_i \cdot P_i \qquad i = 1,2,3,\dots,N_{Product} \tag{8}$$

$$OC = \sum_{m=1}^{M} ST_m \cdot C_{ST} + Q_{HU} \cdot C_{HU} + Q_{CU} \cdot C_{CU} \qquad m = 1,2 \tag{9}$$

In Eq. (7) to (9), C_{CO} is the unit price of a given crude oil feedstock; F_{CO} is the flow rate of crude oil feedstock; F_i and P_i are the flowrate and transfer value of intermediate petroleum products respectively; C_{ST}, C_{HU} and C_{CU} are the unit costs of the stripping steam and the hot and cold utilities respectively; Q_{HU} *and* Q_{CU} are the minimum demand for hot and cold utilities, respectively, while *k* represents the number of stripping steam streams associated with the column.

In this work, the minimum demand for hot and cold utilities is estimated using pinch analysis (Smith, 2005). The algorithm of Morandin (2014) is applied to generate the grand composite curve and determine Q_{HU} *and* Q_{CU}. This approach has the advantages of enabling operating costs to be evaluated for every proposed solution, without needing to consider the details of the associated heat recovery system. However, the approach does not allow meaningful estimates of the cost of the heat exchanger network to be made; nor does it account for challenges relating to obtaining a design

for the heat exchanger network that can accommodate all the operating scenarios of interest.

The optimization variables in Model M2 include the number of trays in each column section and the column operating conditions for each crude oil. Tables S6 to S9 presents the bounds on the optimization variables. The product quality specifications that must be satisfied are presented in Table S4, where the product quality indicators have a tolerance of ±10°C (Chen, 2008).

As explained in Section 3.4, the design variables are optimized in the first stage using a genetic algorithm, while the operating conditions of the individual crude oils are optimized in the second stage using successive quadratic programming. The tuning parameters of the genetic algorithm include the population size and number of generations. In this work, an initial population size of 30 alternative column configurations and 20 generations are selected by trial and error: after applying the algorithm multiple times, these parameters were selected as they represented a good compromise between the performance of the optimum solution and the computation time.

Stochastic optimization methods do not guarantee optimality of a solution. To gain confidence in the optimization results, the optimization framework is applied ten times consecutively and the best result is selected. Each run is carried out using a different set of initial population that are randomly generated. For the results presented in Table S10, optimization took 3426 to 4480 seconds. Figure 4.2 illustrates the optimal column design for processing the three crude oil feedstocks, while Tables 4.6 and 4.7 presents the optimal operating conditions and details of the product quality and flow rates. As shown in Figure 4.2, the distillation unit has 5, 10, 8, 10 and 10 trays in the five sections of the main column, and 5, 7 and 4 trays in the HN, LD and HD side-strippers, respectively. The total annualized cost of the distillation unit is $\$0.41 \cdot 10^6\ a^{-1}$.

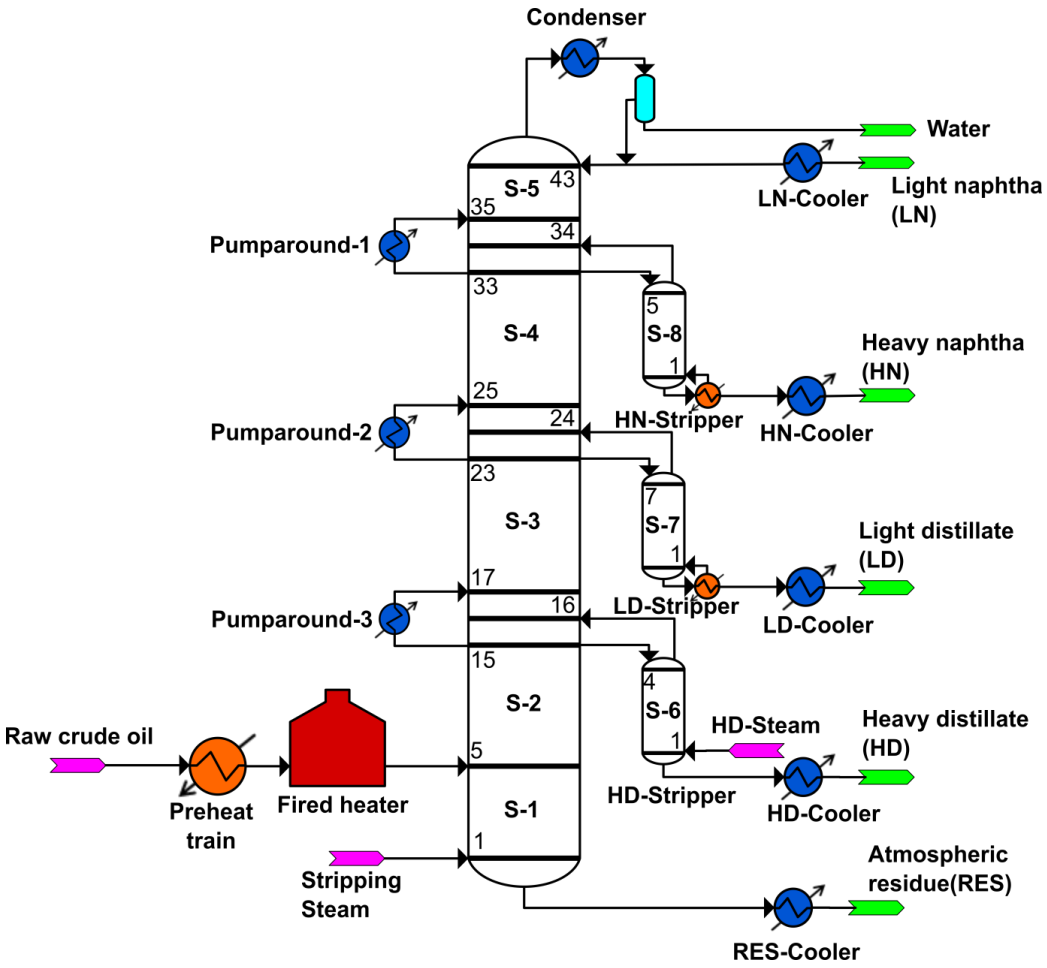


Figure 4.2 Optimized design of flexible crude oil distillation unit

Table 4.6 Optimal operating conditions of crude oil distillation unit

Operating condition	Tia Juana	Bonny light	Brent
PA 1 duty (MW)	14.00	15.00	7.00
PA 2 duty (MW)	14.89	17.58	23.31
PA 3 duty (MW)	16.05	13.35	22.00
PA 1 DT (°C)	20	70	50
PA 2 DT (°C)	40	15	55
PA 3 DT (°C)	20	19	25
Main steam (kmol h^{-1})	1500	1500	1300
HD steam (kmol h^{-1})	188	118	232
Feed temperature (°C)	375	375	375
Reflux ratio	3.78	4.35	2.97

Table 4.7 Optimal product quality and flow rates for three crudes

Variables	Tia Juana	Bonny light	Brent
Product quality (℃)			
LN T5%	25.5	25.1	3.9
HN T5%	135.9	132.8	136.5
LD T5%	217.9	215.9	215.9
HD T5%	308.2	299.4	308.2
RES T5%	367.8	377.4	373.9
LN T95%	110.6	110.6	110.9
HN T95%	186.5	186.6	186.2
LD T95%	301.5	301.4	301.5
HD T95%	354.4	354.4	354.4
RES T95%	755.6	620.2	657.8
Product flow rates ($m^3\ h^{-1}$)			
LN	103.3	115.5	160.4
HN	82.2	104.9	92.0
LD	131.7	184.5	133.1
HD	60.3	86.8	64.8
RES	285.1	171.0	212.3

Table 4.8 Optimal product quality and flow rates for three crudes

Variables	Tia Juana	Bonny light	Brent
Utility requirements			
Hot utility (MW)	49.79	52.75	69.94
Cold utility (MW)	58.71	57.32	70.29
Cost analysis			
Utility cost (MM$ a^{-1})	7.80	9.16	9.35
Steam cost (MM$ a^{-1})	2.06	1.97	1.87
Total operating cost (MM$ a^{-1})	9.85	11.14	11.22
Profit margin (MM$/y)	140.59	97.42	101.39
Annualized capital cost (MM$ a^{-1})		0.41	
Net profit (MM$ a^{-1})	130.74	86.29	90.17
Expected profit (MM$ a^{-1})		102.40	

MM$ a^{-1} denotes millions of dollars per annum

The minimum hot utility requirement for Tia Juana, Bonny light and Brent, as calculated by the grand composite curve (Morandin, 2014), is 49.8 MW, 52.8 MW, and 69.9 MW, respectively. Evidently, lightest crude oil has the highest hot utility demand.

This trend is consistent with the findings of Al-Mayyahi et al. (2011), who reported a significant hot utility demand in processing a crude oil blend comprising a large fraction of a light crude oil and vice versa. This behavior may be explained by considering that, when distilling light crude oils, a large quantity of recoverable heat is available at relatively low temperatures (in the top sections of the column), while for heavy crude oils, recoverable heat is predominantly available at higher temperatures (in the bottom sections of the column). The lower-temperature heat is less useful in the heat recovery system than the higher-temperature heat; consequently, more of the heat needs to be supplied by a fired heater, increasing fuel costs and CO_2 emissions, as can be seen in Table 4.8.

In Table 4.6, the flow rates of stripping steam (main steam and HD steam) for Tia Juana, Bonny light and Brent are 1688 kmol h^{-1}, 1618 kmol h^{-1}, and 1532 kmol h^{-1}, respectively. The main purpose of stripping steam is to suppress the boiling point of the hydrocarbons and to supply heat for vaporizing the hydrocarbon mixture in the flash zone (feed inlet location) of the crude oil distillation unit. The increased demand for stripping steam as the density of the crudes oil increases results from the greater need to suppress the boiling point of heavier hydrocarbons.

The optimum expected profit is $\$102.4 \cdot 10^6$ a^{-1}, with Tia Juana light, Bonny light and Brent contributing 43%, 28% and 29%, respectively. The optimum product flow rate and product quality for the three crude oils are presented in Table 4.7. It may be seen that, in all cases, the maximum profit is attained by increasing the flow rate of the most valuable product at the expense of less valuable products. Table 4.7 shows that all product quality specifications are met, within ±10°C.

The predicted performance for the optimal flexible design of the crude oil distillation unit is validated using rigorous simulation in Aspen HYSYS. Tables S11 and S12 of the supplementary material show that the results obtained with the surrogate model are in good agreement with the rigorous simulation results, with errors in temperatures of less than 2°C and other errors within 3%.

4 Conclusions

The complex nature of crude oil distillation coupled with the large number of degrees of freedom poses a highly challenging design and optimization problem. Additional challenges arise due to the need to simultaneously design the complex column and address heat recovery and to take into account multiple feedstocks. Existing methods address column design only for single crude oil feedstocks, or do not adequately account for capital–energy trade-offs as well as design optimization.

This work proposes a systematic framework for the design of flexible crude oil distillation units that process multiple crude oil feedstocks. The optimization framework integrates surrogate column models based on ANNs and support vector machine and a hybrid stochastic–deterministic optimization algorithm. The framework comprises two stages: structural design and optimization of operating conditions. In the first stage, a genetic algorithm is employed to search for the best structure, accounting for all crude oil feedstocks of interest; in the second stage, successive quadratic programming searches for the optimal operating conditions for each feedstock. A case study indicates that the methodology enables a flexible design to be identified in reasonable computational times (around 1 h).

Future work aims to extend the methodology to consider a larger number of crude oil feedstocks and crude oil blends. An important limitation of the work is that no details of the heat recovery system are considered: an approach to design a single cost-effective heat exchanger network that can service all envisaged processing conditions is also required. Future work could replace the pinch-based calculation with a detailed heat exchanger network model for flexible operation, for example, the HEN models of de Oliveira Filho et al. (2007) or Yee and Grossmann (1990).

Appendix A: Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://

Acknowledgement

The authors would like to acknowledge the financial support from Petroleum Technology Development Fund (PTDF), Nigeria, for sponsoring this PhD research project.

References

Allen, M.P., 1997. The coefficient of determination in multiple regression, in: Understanding Regression Analysis. Springer. Boston, US. pp. 91–95.

Al-Mayyahi, M.A., Hoadley, A.F.A., Smith, N.E., Rangaiah, G.P., 2011. Investigating the trade-off between operating revenue and CO_2 emissions from crude oil distillation using a blend of two crudes. Fuel 90, 3577–3585.

AspenTech, 2011. Aspen HYSYS: Customization Guide. Aspen Technol. Inc.

Bagajewicz, M., Ji, S., 2001. Rigorous Procedure for the Design of Conventional Atmospheric Crude Fractionation Units. Part I: Targeting. Ind. Eng. Chem. Res. 40, 617–626.

Basheer, I.A., Hajmeer, M., 2000. Artificial neural networks: Fundamentals, computing, design, and application. J. Microbiol. Methods 43, 3–31.

Beale, M.H., Hagan, M.T., Demuth, H.B., 2015. Neural Network Toolbox™ User's Guide.

Caballero, J.A., Milan-Yanez, D., Grossmann, I.E., 2005. Rigorous design of distillation columns: Integration of disjunctive programming and process simulators. Ind. Eng. Chem. Res. 44, 6760–6775.

Chen, L., 2008. Heat-Integrated Crude Oil Distillation System Design. PhD Thesis, University of Manchester, UK.

de Oliveira Filho, L. O.; Queiroz, E. M.; Costa, A. L. 2007, A matrix approach for steady-state simulation of heat exchanger networks. Appl. Therm. Eng. 27, 2385–2393

Dhole, V.R., Buckingham, P.R., 1994. Refinery column integration for debottlenecking

and energy saving. Proc. ESCAPE IV Conf. Dublin.

Diwekar, U.M., Kalagnanam, J.R., 1997. Efficient sampling technique for optimization under uncertainty. AIChE J. 43, 440–447.

Edgar, T.F., Himmelblau, D.M., Lasdon, L.S., 2001. Optimization of chemical processes, McGraw-Hill chemical engineering series. McGraw-Hill.

Grossmann, I.E., Guillén-Gosálbez, G., 2010. Scope for the application of mathematical programming techniques in the synthesis and planning of sustainable processes. Comput. Chem. Eng. 34, 1365–1376.

Guthrie, K.M., 1969. Data and Techniques for Preliminary Capital Cost Estimating. Chem. Eng. 76, 114.

Himmelblau, D.M., 2008. Accounts of Experiences in the Application of Artificial Neural Networks in Chemical Engineering. Ind. Eng. Chem. Res. 47, 5782–5796.

Ibrahim, D., Jobson, M., Guillén-Gosálbez, G., 2017a. Optimization-based Design of Crude Oil Distillation Units using Rigorous Simulation Models. Ind. Eng. Chem. Res.

Ibrahim, D., Jobson, M., Li, J., Guillén-Gosálbez, G., 2017b. Surrogate Models Combined with a Support Vector Machine for the Optimized Design of Crude Oil Distillation Units Using Genetic Algorithms. Proc. ESCAPE 27.

Jobson, M., Ochoa-Estopier, L.M., Ibrahim, D., Chen, L., Guillén Gosalbez, G., Li, J., 2017. Feasibility Bounds in Operational Optimization and Design of Crude Oil Distillation Systems Using Surrogate Methods. Chem. Eng. Trans. 61.

Liebmann, K., 1996. Integrated Crude Oil Distillation Design. PhD Thesis, UMIST, Manchester, UK.

Liebmann, K., Dhole, V.R., Jobson, M., 1998. Integrated Design of a Conventional Crude Oil Distillation Tower Using Pinch Analysis. Chem. Eng. Res. Des. 76, 335–347.

MATLAB, 2014. MathWorks. https://uk.mathworks.com/help/stats/fitcsvm.html (accessed 2.1.15).

Morandin, M., 2014. Pinch Analysis cascade calculation.

http://uk.mathworks.com/matlabcentral/fileexchange/47743-cascade-m (accessed 11.20.14).

Nelson, W.L., 1958. Petroleum refinery engineering, McGraw-Hill series in chemical engineering. McGraw-Hill.

Ochoa-Estopier, L.M., Jobson, M., 2015. Optimization of Heat-Integrated Crude Oil Distillation Systems. Part I: The Distillation Model. Ind. Eng. Chem. Res. 54, 4988–5000.

Sharma, R., Jindal, A., Mandawala, D., Jana, S.K., 1999. Design/Retrofit Targets of Pump-Around Refluxes for Better Energy Integration of a Crude Distillation Column. Ind. Eng. Chem. Res. 38, 2411–2417.

Smith, R., 2005. Chemical Process: Design and Integration. Wiley. Chichester, UK.

Steimel, J., Harrmann, M., Schembecker, G., Engell, S., 2013. Model-based conceptual design and optimization tool support for the early stage development of chemical processes under uncertainty. Comput. Chem. Eng. 59, 63–73.

Subramanyan, K., Diwekar, U., Zitney, S.E., 2011. Stochastic modeling and multi-objective optimization for the APECS system. Comput. Chem. Eng. 35, 2667–2679.

Szklo, A., Schaeffer, R., 2007. Fuel specification, energy consumption and CO_2 emission in oil refineries. Energy 32, 1075–1092.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer.

Watkins, R.N., 1979. Petroleum Refinery Distillation. Gulf Publishing Company, Book Division. Texas, US.

Yee, T.F., Grossmann, I.E., 1990. Simultaneous optimization models for heat integration - II. Heat-exchanger network synthesis. Comput. Chem. Eng. 14, 1165–1184.

5.3 Publication 4

Ibrahim, D., Jobson, M., Guillén-Gosálbez, G., 2017. Design of Chemical Processes under Uncertainty Combining the Sample Average Approximation and the Analytic Hierarchy Process. Comput. Chem. Eng. [Submitted]

Design of chemical processes under uncertainty combining the sample average approximation and the analytic hierarchy process

Dauda Ibrahim[1,*], Megan Jobson[1], Gonzalo Guillén-Gosálbez[2]

[1] **Centre for Process Integration, School of Chemical Engineering and Analytical Science, University of Manchester, Manchester M13 9PL, UK**

[2] **Department of Chemical Engineering, Centre for Process Systems Engineering, Imperial College, South Kensington Campus, London SW7 2AZ, UK**

Abstract

This paper introduces a novel methodology for the synthesis of chemical processes under uncertain operating conditions. The approach comprises four main steps: (i) uncertainty characterization and generation of scenarios using sampling methods; (ii) process synthesis based on the 'sample average approximation algorithm' to generate feasible designs; (iii) evaluation of each such design in the space of uncertain parameters using both feasibility and probabilistic metrics computed over the scenarios generated in Step 2; and (iv) application of the analytic hierarchy process to identify the designs that best reflect decision-makers' preferences. We illustrate the capabilities of our methodology through its application to the design of heat exchanger networks under uncertain inlet conditions, and distillation column for multicomponent separation with uncertain feed conditions. Numerical results obtained demonstrate that the proposed methodology is capable of synthesizing a network and a distillation column that are flexible, operable under a wide range of operating conditions, and with better overall performance than designs that do not explicitly account for flexible operation.

[*] Corresponding author
E-mail address: dauda.ibrahim@manchester.ac.uk

Keywords: Process design, Uncertainty, Flexibility, Optimization, Analytical hierarchy process

1. Introduction

Chemical process design is typically performed under the assumption of fixed operating parameters (temperature, pressure, composition, flow rate, etc.) at so-called nominal conditions. During plant operation, however, it is likely that the operating parameters will change, affecting the operability, controllability and feasibility of the design. To address these variations in processing conditions, it is necessary to introduce some degrees of flexibility at the design stage, to ensure that the process will be able to accommodate deviations from the nominal state during its operation (Grossmann et al., 1983). In the past, flexibility considerations were incorporated at the design stage based on engineers' experience and judgment (Wagler and Douglas, 1988), usually relying on empirical overdesign factors (Swaney and Grossmann, 1985). These approaches depend little on understanding of or insight into the required degrees of flexibility of a given design and may lead to significant overdesign, typically associated with unnecessarily high capital investment (Grossmann et al., 1983).

Chemical process design under uncertainty is an active area of research (Halemane and Grossmann, 1983; Pistikopoulos and Ierapetritou, 1995; Sahinidis, 2004; Wang and Rong, 2010; Kostin et al., 2012; Rogers and Ierapetritou, 2015; Amaran et al., 2016; Wang et al., 2016). Two main methods are applied to address optimization problems containing uncertain parameters, namely, robust optimization and stochastic programming (Georgiadis and Pistikopoulos, 1999; Bertsimas et al., 2010; Grossmann and Guillén-Gosálbez, 2010; Zhang et al., 2016). In robust optimization approaches, the solution to the uncertain problem is determined for every possible realization of the uncertain parameter within the feasible region for a certain probability of satisfying the problem constraints. On the other hand, in stochastic programming, uncertainties are described by discrete scenarios. Each scenario corresponds to a particular realization of the uncertain parameter within the uncertain space. This type of problem is typically solved in two stages (Pistikopoulos and Ierapetritou, 1995; Grossmann and Guillén-

Gosálbez, 2010). In the first stage, design variables and the process configuration are selected, while in the second stage the adjustable variables are optimized according to the realization of the scenarios.

The design of flexible processes can be seen as a special case of design under uncertainty, where the engineer seeks a design capable of remaining feasible – i.e. meeting the design objective – within a given operating interval. Instead of using any of the methods for optimization under uncertainty mentioned above, many researchers have focused on developing customized approaches for this problem based on the definition of a flexibility index. This flexibility index, which is calculated considering the variability of the uncertain parameters (typically the operating conditions in a process design problem), is then optimized as an additional criterion (i.e. flexibility objective in addition to economic objective). In a pioneer work, Swaney and Grossmann (1985) developed a flexibility index, F, as a quantitative measure of the maximum tolerable deviation of an uncertain parameter within the feasible operation region. This original index has been the basis for many methodologies for the design of flexible chemical processes.

Pistikopoulos and Grossmann (1988a) proposed a linear model for retrofitting an existing flowsheet in order to improve flexibility at minimum total annualized cost. By exploring the linear model, they explicitly included flexibility constraints into the formulation. In a later work, Pistikopoulos and Grossmann (1988b) extended the linear formulation to design problems with infeasible nominal points and nonlinear behavior. Chacon-Mondragon and Himmelblau (1996) proposed an integrated approach for process design addressing flexibility and total annualized cost. To incorporate the two potentially conflicting objectives (flexibility index and total annualized cost) at the design stage, a multi-objective optimization (MOO) problem was formulated to simultaneously maximize the flexibility index and minimize the total annualized cost considering the potential realizations of the uncertain parameter within the feasible region of operation.

Chen and Hung (2004) proposed an iterative three-step approach that integrates flexibility analysis and process synthesis. First, an MINLP formulation derived from a process superstructure is used to identify a candidate design. In the second step, the flexibility of the design is quantified to assess whether the design is feasible over the full disturbance range. Finally, a constraint is applied to exclude designs that fail to satisfy the flexibility criterion from the search space used for the next iteration. Recently, Wang et al. (2016) explored the application of flexibility analysis in the context of chemical process supply chain design in the face of uncertainties in product demand, raw material supply, production yield and price of final product. Zhang et al. (2016) compared flexibility analysis and robust optimisation from the historical context.

Most of the research highlighted above relies on the definition and optimization of flexibility indices. However, these approaches lead to nonlinear problems with auxiliary constraints required to calculate the flexibility index. As a result, such approaches can be computationally expensive and also difficult to implement for problems described using black box models.

This work presents a systematic framework for the optimal design of chemical processes under uncertainty that overcomes the above limitations. First, the proposed method does not rely on auxiliary constraints, thereby simplifying the calculations from the viewpoint of implementation (i.e. it can easily be implemented in process simulators). Second, the new approach considers in an explicit manner the inherent trade-off between level of flexibility and economic performance through the application of a multi-criteria decision-support methodology. Third, the approach can handle different types of distribution of uncertainties, where established methods for flexible design do not incorporate probabilistic information in the analysis. The capabilities of the methodology, which integrates the Sample Average Approximation (SAA) algorithm (Kleywegt et al., 2001) with the Analytic Hierarchy Process (Saaty, 1990), are illustrated through its application to the design of a heat exchanger network (HEN) and a distillation column.

The remainder of this article is organized as follows: Section 2 defines the problem to be addressed. Section 3 presents the detailed mathematical formulation of process design under uncertainty together with an efficient solution method. Section 4 applies the proposed approach to a case study on heat exchanger network design with uncertain inlet conditions and distillation column design with uncertain feed conditions. Finally, comprehensive conclusions on the current study are drawn.

2. Problem statement

The chemical process design problem to be addressed can be formally stated as follows:

The design methodology aims to identify the configuration and operating conditions of a chemical process that will allow the process to operate feasibly (i.e. to meet production goals, within a specified tolerance) and optimally (i.e. with good performance, defined in terms of appropriate performance metrics), over the whole range of the space of uncertain parameters. The approach is based on the assumption that these uncertain parameters can be represented using a set of pre-defined scenarios, each of which has a known probability of occurring.

Given are sufficient process data to allow design and evaluation of proposed designs. Some of these design variables may be uncertain (i.e. their values may vary according to a given probability function within a given range). For example, data may include process stream data (i.e. supply and target temperatures, heat capacity flow rates and heat transfer coefficients, temperatures of available utilities), process specifications (e.g. flow rate and composition of product streams), constraints (e.g. minimum approach temperature, range of operating temperatures or pressures) and economic data (e.g. unit costs of utilities, parameters of capital cost models, operating hours, project life).

Two design problems are used to illustrate the capabilities of the proposed approach: heat exchanger network design and distillation column design.

3. Methodology

To solve the problem defined in Section 2, a general mathematical formulation for design under uncertainty is proposed, together with an effective solution procedure. The application of this approach is presented in Section 4.

3.1 Mathematical formulation

Problem (1) presents a general mathematical formulation for the problem of process design under uncertainty:

$$\min_{d,z,x} \quad f(d,z,x,\theta) \tag{1}$$

$$s.t \quad h(d,z,x,\theta) = 0$$
$$g(d,z,x,\theta) \leq 0$$
$$d \in D, z \in Z, x \in X, \theta \in \Theta$$

where d, z, x, θ represent the vectors of design, control, state and uncertain variables respectively. $f(d,z,x,\theta)$ is the objective function; $h(d,z,x,\theta)$ and $g(d,z,x,\theta)$ denote the equality and inequality constraints, respectively; X, Z and D represent sets of state, control and design variables, respectively, while Θ denotes the set of all possible values that the process parameters can take.

For simplicity, the state variable, x, is eliminated from Problem (1), leading to the alternative representation in Problem (2):

$$\min_{d,z} \quad f(d,z,\theta) \tag{2}$$

$$s.t \quad h(d,z,\theta) = 0$$
$$g(d,z,\theta) \leq 0$$
$$d \in D, z \in Z, \theta \in \Theta$$

This work assumes that the uncertain parameters can be described via a set of pre-defined scenarios. A scenario is taken to be a sample of the uncertain parameter space, in which each parameter takes a specific value (i.e. corresponding to a particular realization in the space of uncertain parameters). This leads to the formulation presented as Problem (3):

$$\min_{d,z} \quad f(d, z_s, \theta_s) \tag{3}$$

$$\begin{aligned} s.t \quad & h(d, z_s, \theta_s) = 0 \qquad \forall s \in S \\ & g(d, z_s, \theta_s) \leq 0 \qquad \forall s \in S \\ & d \in D, z_s \in Z \end{aligned}$$

where θ_s is the vector of parameters values in scenario s belonging to the set of scenarios S. Note that structural design decisions are not scenario-dependent, as they must be taken here and now, before (and whether or not) any particular uncertain scenario occurs. On the other hand, operating decisions are scenario-dependent, as it is assumed that the corresponding value of a manipulated variable can be set in response to the specific materialization of the uncertain parameter.

The objective function $f(d, z_s, \theta_s)$ of the model might include more than a single criterion, as in general we will be interested in maximizing the economic performance (optimality), while at the same time ensuring adequate operation under all possible conditions (feasibility), that is, while maximizing the flexibility level. Hence, different metrics might be used to represent the design objectives (further details on this topic is presented in Section 3.2.2), leading to a multi-objective model. In general, these objectives will tend to conflict; for example, ensuring that the design will work well over a wide range of operating conditions will tend to increase capital costs. Hence, the multi-objective and multi-scenario problem, M1, will take the form:

$$\text{(M1)} \qquad \min_{d,z_s} \quad \{f_1(d, z_s, \theta_s), \dots, f_k(d, z_s, \theta_s)\} \tag{4}$$

$$
\begin{aligned}
s.t \quad & h(d, z_s, \theta_s) = 0 \qquad \forall s \in S \\
& g(d, z_s, \theta_s) \leq 0 \qquad \forall s \in S \\
& d \in D, z_s \in Z
\end{aligned}
$$

where f_1 to f_k are the scalar objectives to be minimized. A key point in this formulation is the approach in which flexibility can be quantified using metrics described by algebraic equations.

The solution of model M1 is not unique, but rather given by a set of Pareto points, each achieving a unique combination of objective function values. Hence, assuming that the model could be efficiently solved, one should still address the challenge of selecting a specific Pareto solution to be implemented in practice. To this end, this work applies the Analytic Hierarchy Process (AHP), a multi-criteria decision-support tool that translates decision-makers' preferences into weights. The weights provided by the AHP allow Model M1 to be reformulated into the following single-objective multi-scenario problem, M2:

$$
\text{(M2)} \qquad \min_{d, z_s} \quad \sum_{i=1}^{k} w_i f_i(d, z_s, \theta_s) \qquad (5)
$$

$$
\begin{aligned}
s.t \quad & h(d, z_s, \theta_s) = 0 \qquad \forall s \in S \\
& g(d, z_s, \theta_s) \leq 0 \qquad \forall s \in S \\
& d \in D, z_s \in Z
\end{aligned}
$$

where w_i represents the weight assigned to objective i (calculated via the AHP). Model M2 is easier to solve than M1, as it has a single objective. However, it might still be difficult to calculate the global optimum due to its size and nonlinearities, as its number of equations depends on the number of scenarios (e.g. mass and energy balances are defined for every scenario). To simplify its calculation, we use a variant of the sample average approximation (SAA) algorithm (Kleywegt et al., 2001) that is explained in more detail in Section 3.2.2.

3.2 Solution procedure

The solution strategy applied to solve M2 efficiently is illustrated in Figure 1 and involves four main steps:

- Step 1: Stochastic modeling: characterization and sampling of the uncertain operating conditions.
- Step 2: Generation of design solutions using the SAA algorithm (see Section 3.2.2): process synthesis for every sample generated from the uncertain distributions.
- Step 3: Assessment of the designs in the space of uncertain parameters: process optimization of the designs generated in Step 2 considering all the scenarios that represent the uncertain parameters.
- Step 4: Identification of the best trade-off solution: Application of the analytic hierarchy process to identify the best process design among the alternatives identified in the previous steps.

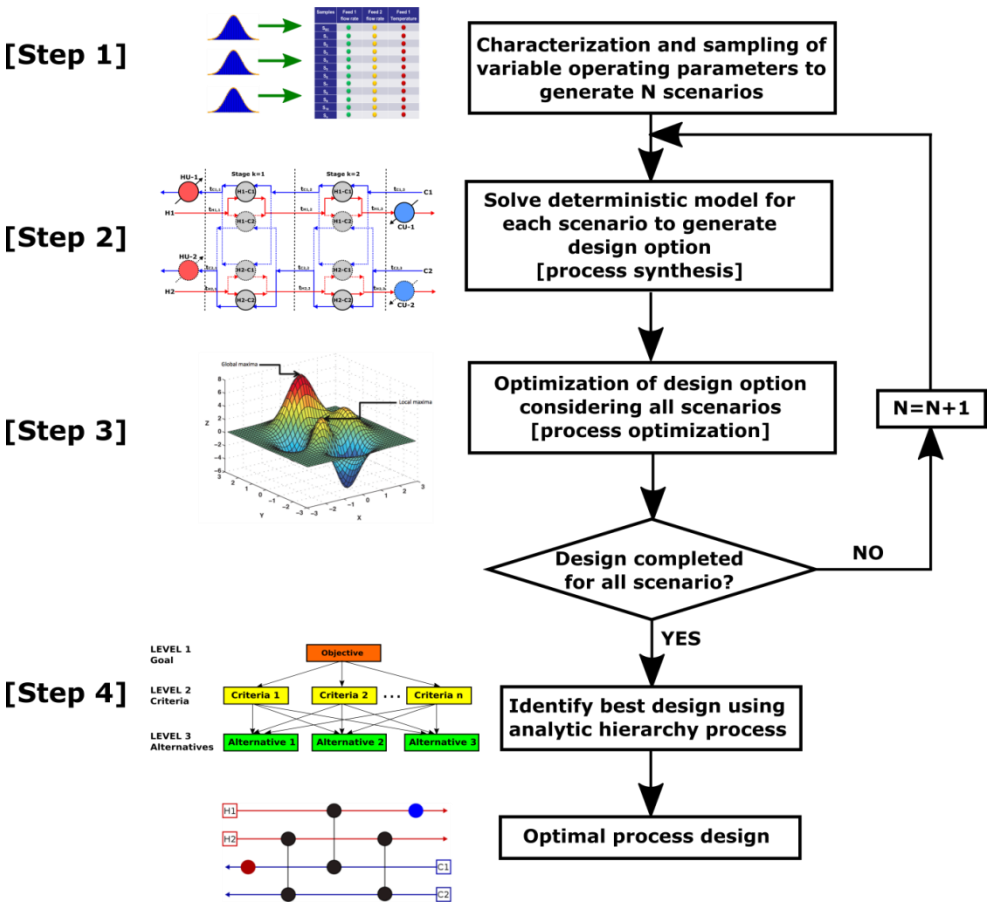


Figure 1 Strategy for optimal process design with uncertain operating conditions.

In the new design method proposed in this work, the uncertain parameters are described via probability distributions, as opposed to other published research on flexibility that defined minimum and maximum levels within which uncertain parameters fall. Monte Carlo sampling of the uncertain parameters is then applied to generate representative scenarios (i.e. samples). For each such scenario, a deterministic model is solved to generate potentially attractive design alternatives. The performance of each of these designs is evaluated by fixing the design variables in a deterministic model that considers all the scenarios simultaneously and optimizes only the operating variables (for all the scenarios that represent the uncertain parameters). This optimization evaluates the performance of each design in the space of uncertain parameters; these outputs are the inputs to the analytic hierarchy process that identifies the best process design that best matches the preferences of the decision-makers, as represented using weights in equation (15). Details of each step of the proposed strategy are presented in Sections 3.2.1 to 3.2.4.

3.2.1 Step 1: Stochastic modeling

The first step involves describing and quantifying the uncertain parameters via probability functions, such as uniform, normal and triangular distributions, among others (Diwekar, 2008). After the probability distributions are constructed, an appropriate sampling technique is applied to sample and propagate the effect of the uncertainty into the model. Different sampling techniques are presented in the literature (Diwekar, 2003). Without loss of generality, the Monte Carlo sampling method is adopted in this work, as it shows equidistribution properties of the sets of points in the space of uncertain parameters (Diwekar and Rubin, 1994). Monte Carlo sampling is based on a pseudo-random number generator that approximates a given distribution, where the specific values of each input are selected by inverse transformation over the cumulative probability distribution (Diwekar and Rubin, 1994). The sample from the Monte Carlo simulation is defined by a combination of different random values for each input (Diwekar and Rubin, 1994); this information is used in Step 2 to synthesize a different design for each scenario.

3.2.2 Step 2: Generation of design solutions using the SAA algorithm

The sampled scenarios are used as inputs in the deterministic model M2, which is solved using an algorithm based on the SAA algorithm. The SAA algorithm solves recursively an approximate model containing fewer scenarios than the original model and stores the values of the first-stage decision-variables for each such run. These first-stage variables are then fixed in the original problem containing all the scenarios and the second-stage decisions are optimized. The first-stage solution that performs best in the original problem approximates the global optimum of the original model.

Here we apply the SAA algorithm; in the first step of the algorithm we use deterministic models constructed for each scenario. Hence, following this approach, model M2 is solved at the process synthesis step to generate designs for each sampled scenario (i.e. we solve M2 for every scenario separately). Note that the original problem has several objective functions (e.g. expected performance, risk metric, etc.), some of which need to be calculated over the whole range of scenarios. Nevertheless, in this stage of the algorithm, one scenario is considered at a time, which prevents the calculation of the objectives required to assess the level of flexibility of every design alternative. To overcome this limitation, the alternatives are optimized considering a single criterion (i.e. the economic performance). Hence, for every scenario, s^*, the value of the first-stage variables $\bar{d}_{s*}$ that define the optimal design for that outcome of the uncertain parameters is obtained by solving the following problem:

$$\text{(M3)} \qquad \bar{d}_{s*} = \underset{d, z_{s*}}{\operatorname{argmin}} f_{eco}(d, z_{s*}, \theta_{s*}) \tag{6}$$

$$s.t \qquad h(d, z_{s*}, \theta_{s*}) = 0$$
$$g(d, z_{s*}, \theta_{s*}) \leq 0$$
$$d \in D, z_{s*} \in Z$$

where z_{s*} and θ_{s*} are the optimal control variables and parameter values, respectively, associated with design $\bar{d}_{s*}$ and scenario $s^* \in S$ and f_{eco} is the economic objective

function. It should be noted that *argmin* is a function that returns the values of d and z_{s*} that minimizes $f_{eco.}$

3.2.3 Step 3: Assessment of the designs generated in the space of uncertain parameters

The designs generated in step 2 are optimized in the parameter space in order to assess their performance. Note that each design is optimized for a single scenario (i.e. it is optimal for that combination of uncertain parameter values, but might be suboptimal for others); therefore it might be unfeasible when it is evaluated under conditions that differ from those for which it was optimized. To avoid infeasibilities during the assessment stage, the equality and inequality constraints must be relaxed, and a penalty added to the objective function. The model is therefore modified as follows:

$$\text{(M4)} \quad \min_{z_s} \ \sum_{s\in S} prob_s\left(f_{eco}\left(\bar{d}_{s*}, z_s, \theta_s\right) + \pi\left(slack_s^{h+} + slack_s^{h-} + slack_s^{g}\right)\right) \qquad (7)$$

$$s.t \qquad h\left(\bar{d}_{s*}, z_s, \theta_s\right) + slack_s^{h+} + slack_s^{h-} = 0 \ \ \forall s \in S$$
$$g\left(\bar{d}_{s*}, z_s, \theta_s\right) - slack_s^{g} \leq 0 \ \ \forall s \in S$$

$$z_s \in Z, slack_s^{h+}, slack_s^{h-}, slack_s^{g} \in \mathbb{R}^+$$

where $prob_s$ is the probability of scenario, s, occurring, π is a penalty weighting that penalizes the violation of the equality and inequality constraints, while $slack_s^{h+}, slack_s^{h-}$ and $slack_s^{g}$ (slack variables) represent the extent of deviation from feasible operation. Note that here we optimize a single objective, the economic performance, calculated as $f_{eco}\left(\bar{d}_{s*}, z_s, \theta_s\right)$ plus the penalty related to deviation from feasible operation, represented by a weighted sum of the slack variables.

From the set of operating conditions, $\overline{z_{sd}}$, that is optimal for each scenario, we can finally calculate the different objectives. In this work, we focus on slack variables (TS), expected total cost (ETC) and worst case (WC). These objectives were selected since we are interested in finding an optimal design that represent the best trade-off between

extent of flexibility, process economics and risk associated to process operation under different scenarios. The details of the individual objectives follow:

- The sum of the so-called 1-norms of the slack variable (TS), which indicates the deviation from the feasible region of operation for a specific design over all expected scenarios. This objective is mathematically defined as follows:

  $$TS_d = \sum_{s \epsilon S} prob_s \left(\left\| slack_s^{h+} \right\|_1 + \left\| slack_s^{h-} \right\|_1 + \left\| slack_s^{g} \right\|_1 \right) \quad (8)$$

  1-norm is the sum of the absolute values of the slack variables

- The expected total cost (ETC), which represents the weighted sum of the total annualized cost (TAC) over all the scenarios:

  $$ETC_d = \sum_{s \epsilon S} prob_s TC_s \quad (9)$$

  where TC_s is the cost in scenario, *s*, of the design being assessed and $prob_s$ is the probability of occurrence of scenario, *s*.

- The worst case (WC), which is the maximum value of the total cost over all of the scenarios.

  $$WC_d = \max_s \{TC_s\} \quad (10)$$

The best process design is identified after aggregating the values of each performance indicator for each potential design alternative. Section 3.3 introduces the AHP and shows how it can be applied to carry out such aggregation.

3.2.4 Step 4: Application of Analytic Hierarchy Process (AHP) to identify optimal design

The analytic hierarchy process (AHP) (Saaty, 1990) is an important tool used in the field of multi-criteria decision-making to select alternatives according to multiple objectives. One important feature of AHP is the ability to handle both qualitative (subjective opinion) and quantitative inputs. AHP has been widely used to solve problems in different areas including manufacturing (Yurdakul, 2004), logistics (Wang et al., 2005), management (Saaty et al., 2003) and engineering (Su et al., 2003).

The application of AHP comprises three main steps: (i) decomposition, (ii) comparative judgment and (iii) synthesis of priorities (Korpela and Lehmusvaara, 1999). In the decomposition step, the complex multi-criteria problem is decomposed into a hierarchy of decisions. The hierarchy places the overall objective (goal) as the highest element, followed by criteria, sub-criteria and alternatives (which are the lowest level elements in the hierarchy), as illustrated in Figure 2. The second step, comparative judgment, involves the construction of a matrix of pairwise comparisons of the elements at each level (criteria, sub-criteria and alternatives) in the hierarchy using the standard Saaty scale presented in Table 1. In the third step, weights are assigned according to the comparison matrix and based on the experience and judgment of the designers. An aggregated score is then calculated for each design to allow the best designs to be ranked.

Table 1 Pairwise comparison scale of preference (Adapted from Saaty et al., 2003)

Magnitude of Importance	Definition	Description
1	Equally important	Two criteria contribute equally to the overall objective
3	Moderately more important	Experience and judgment slightly favor one criterion over the other
5	Strongly more important	Experience and judgment strongly favor one criterion over the other
7	Very strongly more important	Experience and judgment very strongly favor one criterion over the other
9	Extremely more important	Experience favoring one criterion is of the highest possible or of magnitude

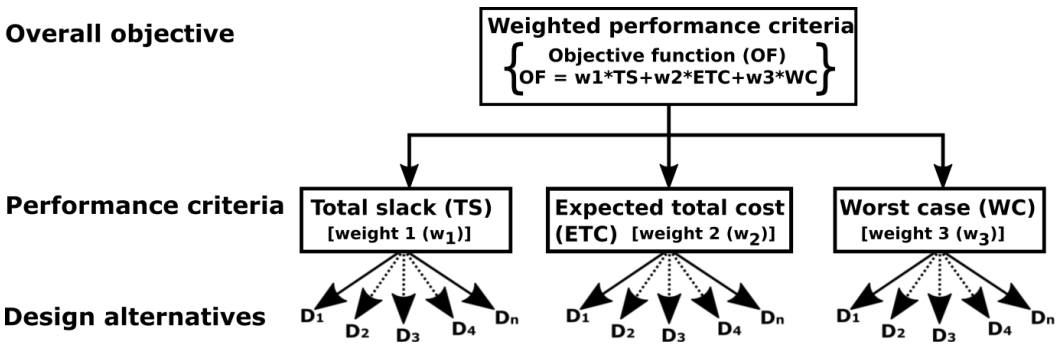


Figure 2 Hierarchy for AHP – based analysis for the proposed methodology

In the context of our methodology, we start by defining a hierarchy as shown in Figure 2. The objective is to identify a flexible process design considering three performance criteria namely: extent of violation of constraints (as indicated by 'total slack' metric), expected total cost and worst case performance. The next step is to generate a matrix of pairwise comparisons based on the scale of preferences described in Table 1. The pairwise comparison matrix is used to generate the weights (w_1, w_2 and w_3) assigned to each criterion (performance criterion) according to the decision-makers' judgment.

Let us consider a coefficient matrix A showing the relative importance of k different objectives.

$$\begin{pmatrix} 1 & a_{1i} & a_{1k} \\ \vdots & \ddots & \vdots \\ \frac{1}{a_{k1}} & \frac{1}{a_{ki}} & 1 \end{pmatrix}$$

Before calculating the weights associated with such matrix, we first check its consistency by using a consistency index (CI), the value of which should fall below a given threshold (Saaty, 1990):

$$CI = \frac{\lambda_{max} - k}{k-1} \tag{11}$$

where λ_{max} is the maximum eigenvalue of the matrix A and k is the number of objectives. (In our case k is 3, but more objectives could be defined). The consistency index is then compared with a random index (RI) (Saaty, 1990) to determine the consistency ratio (CR):

$$CR = CI/RI \tag{12}$$

The random index is the consistency index of a matrix with random inputs, which has the same order with matrix A (Saaty, 1990). If the consistency ratio is above a threshold value ($CR \leq 0.10$) then the pairwise comparisons must be revised (Saaty, 1990). Otherwise, we can proceed to calculate the weights for each objective by solving the following system of linear equations:

$$\sum_{i'=1}^{k} a_{ii'} w_{i'} - \lambda_{max} w_{i'} = 0 \qquad \forall i \tag{13}$$

where i and i' are row and column index of the elements in matrix A respectively.

Once the weights are obtained, we can sort the designs according to an aggregated score. Let $\bar{d}_{s*}$ be the optimal design for scenario s*, its aggregated score (AG_{s*}) is calculated as follows:

$$AG_{s*} = \sum_{i=1}^{k} w_i \hat{f}_i\left(\bar{d}_{s*}, \bar{z}_s, \theta_s\right) \tag{14}$$

where $\bar{z}_s$ are the optimal values of the second stage decision-variables associated with that design; $\hat{f}_i$ is the normalized value of objective function i. Different normalization methods can be applied to normalize $\hat{f}_i$; without loss of generality, we have used the following:

$$\hat{f}_i = \frac{f_i - \underline{f_i}}{\overline{f_i} - \underline{f_i}} \tag{15}$$

where $\hat{f}_i$ is the normalised value of sample f_i, $\underline{f_i}$ and $\overline{f_i}$ are the corresponding minimum and maximum objective values across all the alternative designs (i.e. the minimum and maximum value taken by objective i for all the design alternatives generated by the SAA algorithm).

Finally, the different designs can be ranked in terms of the aggregated score. A detailed stepwise application of the AHP is illustrated using a case study in Section 4.

4. Case Study

The capabilities of the proposed methodology are demonstrated on two case studies: heat exchanger network with uncertain inlet temperatures and heat capacity flow rates demonstrates, and design of a distillation column with uncertain feed condition (temperature, flow rate and composition).

4.1 Case 1: Heat exchanger network design

Design problem data include stream data for one hot stream (H1), two cold streams (C1 and C2), steam (S1), cooling water (W1), costs of utilities and heat exchanger capital cost. The problem data are presented in Table 2 and equation (16) (Biegler et al., 1997).

Table 2 Stream data for HEN design

Streams	T_{in} (K)	T_{out} (K)	FCp (kW K^{-1})	h (kW m^{-2} K^{-1})	Cost (\$ kW^{-1} yr^{-1})
H1	440	350	22	2	-
C1	349	430	20	2	-
C2	320	368	7.5	0.67	-
S1	500	500	-	1	120
W1	300	320	-	1	20

Minimum approach temperature (ΔT_{min}) = 1 K

Exchanger cost = $6{,}600 + 670(Area)^{0.83}$ (16)

The overall heat transfer coefficient for each heat exchanger match is calculated from the heat transfer coefficient, h considering the individual streams in Table 2. A minimum approach temperature of 1 K is defined.

4.1.1 Step 1: Description and sampling of uncertain operating conditions

The uncertain parameters (i.e. HEN inlet temperature and heat capacity flow rate) are described using normal distributions. The mean value of the distribution is specified to be the nominal conditions, while standard deviations of 7% and 2% are defined for the distribution of inlet temperatures (in K) and heat capacity flow rates, respectively. Ten sets of operating conditions, i.e. supply temperatures and heat capacity flow rates of hot stream H1 and cold stream C1, are generated using Monte Carlo sampling; these are presented in Table 3. In this work, it is assumed that all scenarios are equally probable, i.e. the probability of each occurring is 0.091.

Table 3 Sample data for uncertain stream conditions.

Scenario	$T_{H1(in)}$ (K)	$T_{C1(in)}$ (K)	$FCp_{(TH1)}$ (kW K^{-1})	$FCp_{(TC1)}$ (kW K^{-1})
S_{NC}	440	349	22.0	20.0
S_1	442	346	21.9	20.4
S_2	437	349	22.1	19.7
S_3	438	348	22.2	19.6
S_4	443	347	21.2	20.0
S_5	437	345	21.7	19.6
S_6	434	356	21.6	19.6
S_7	434	352	21.8	20.3
S_8	441	353	22.4	20.6
S_9	445	346	21.7	20.9
S_{10}	435	352	22.0	19.8

*NC = Nominal case

4.1.2 Step 2: Synthesis of flexible HENs for each scenario

The simultaneous MINLP model proposed by Yee and Grossmann (1990) and the corresponding stagewise superstructure are adopted for the synthesis of a flexible HEN. Figure 3 illustrates the stagewise superstructure for the HEN design problem.

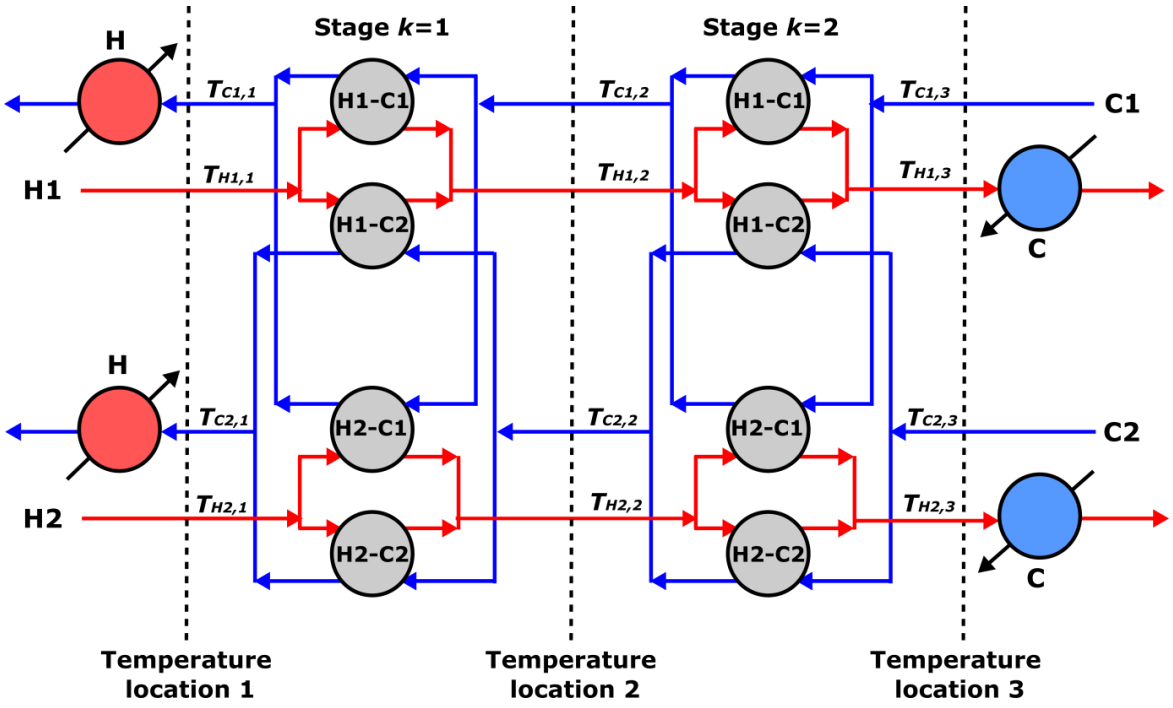


Figure 3 Heat exchanger network superstructure

The number of stages in the superstructure can be determined by the maximum number of hot and cold streams (Biegler et al., 1997). In each stage, various heat recovery 'matches' between hot and cold streams are possible. The network with the minimum total annualized cost can be determined via optimization of the MINLP model introduced by Yee and Grossmann (1990); this model assumes isothermal mixing at the exchanger outlets and elimination of bypass and split streams with more than two matches per stage. The objective function and constraints of the original MINLP were modified to allow them to be used in the context of our approach; in particular, the single scenario model needed to be reformulated as a multi-scenario model, the equality and inequality constraints were formulated using slack variables and penalties related to slack variables were added to the objective function. The detailed mathematical formulation is presented in Section S1 of the Supplementary material.

To generate the potential HEN design options, the modified MINLP model was implemented in the General Algebraic Modeling System (GAMS version 24.0.2, 2012) and solved with the BARON solver on a HP desktop PC with Intel(R) Core i5 processor running at 3.2 GHz, and 8 GB of RAM. The SAA algorithm was implemented in GAMS, while the AHP approach was coded in MATLAB R2014a.

The time taken to solve each iteration of the deterministic MINLP model ranged from 1.0 to 3.14 seconds. The CPU time to solve the stochastic (multi-scenario) MINLP models where the design variables were fixed took 1.0 to 1440 seconds. Note that the relative gap of the global solver is set to 10 %, in order to ensure an acceptable margin between primal (upper bound) and dual (lower bound) problem.

Table 4 shows the details of the optimized HEN structure, required heat transfer area, relative gap and total annualized cost (TAC) for each scenario, as evaluated using the deterministic HEN model (see appendix). The design for the nominal case, D_{NC}, is that obtained given the nominal input values. It can be observed that three heat exchangers are needed (1,1,1; 1,2,1; 1,2,2) in the nominal case design and the heat transfer area

needed is 183 m^2; no heaters or coolers are needed, i.e. the design is fully heat integrated.

Other scenarios have different HEN structures. Designs D_1, D_4 and D_9 each have two heat exchangers and a heater, while design D_8 has two heat exchangers and a cooler, and design D_6 has two heat exchangers, one heater and one cooler. The remaining designs have three heat exchangers and one heater. With the exception of designs D_1, D_4, D_6, D_8 and D_9, all other designs incorporate a splitter.

Design D_5 has the lowest total HEN area (142 m^2) with total annualized cost of \$89,951/y, while Design D_{10} has the largest requirement for heat transfer area (220 m^2), corresponding to TAC of \$97,288/y. Among the potential alternative designs, D_6 is moderate in terms of required heat transfer area, and thus capital investment. However, to select the HEN design that is flexible and operable over a wider range of operating conditions, further evaluation of the alternative designs and of the nominal case design is required.

Table 4 Optimal HEN design for each scenario

Optimal HEN design	**Individual unit area (m^2) for match (i,j,k), qhu,i and qcu,j**						**Relative gap (%)**	**Total annualized cost (\$ y^{-1})**
	1,1,1	1,2,1	1,2,2	qhu,1	qhu,2	qcu,1		
D_{NC}	144.7	7.6	30.6				0.09	76,502
D_1	121.8		38.8		1.5		0.09	80,434
D_2	146.5	7.8	30.6	0.8			0.09	88,410
D_3	148.2	7.9	30.2	0.3			0.09	85,920
D_4	108.9		55.3		0.9		0.09	77,502
D_5	108.8	9.5	21.2	2.8			0.09	89,951
D_6	94.7		45.8	2.8		4.8	0.09	94,641
D_7	150.3	7.5	33.6	2.3			0.09	99,909
D_8	112.1		66.2			3.3	0.03	78,775
D_9	104.7		60.0		0.8		0.04	77,421
D_{10}	174.5	7.5	37.0	0.8			1.00	97,288

i,j,k denotes the match between cold stream i and hot stream j in stage k

qhu,i denotes a heater on cold stream i

qcu,j denotes a cooler on hot stream j

4.1.3 Step 3: Assessment of flexible HENs

For each alternative design, the decision variables that define the equipment size and configuration are fixed, and the design is optimized within the uncertain space (i.e. for all the scenarios defined in Table 3). The value of the expected total cost, worst case performance and penalty related to slack variables (total slack) for each design alternative are shown in Figure 4. Since the performance indicators have different units and orders of magnitude, normalized values are used to compare the design options. The normalization is performed using Eq. 15 as presented in Section 3, in a way such that each performance indicator falls in the range 0 to 1, as can be seen in Figure 5.

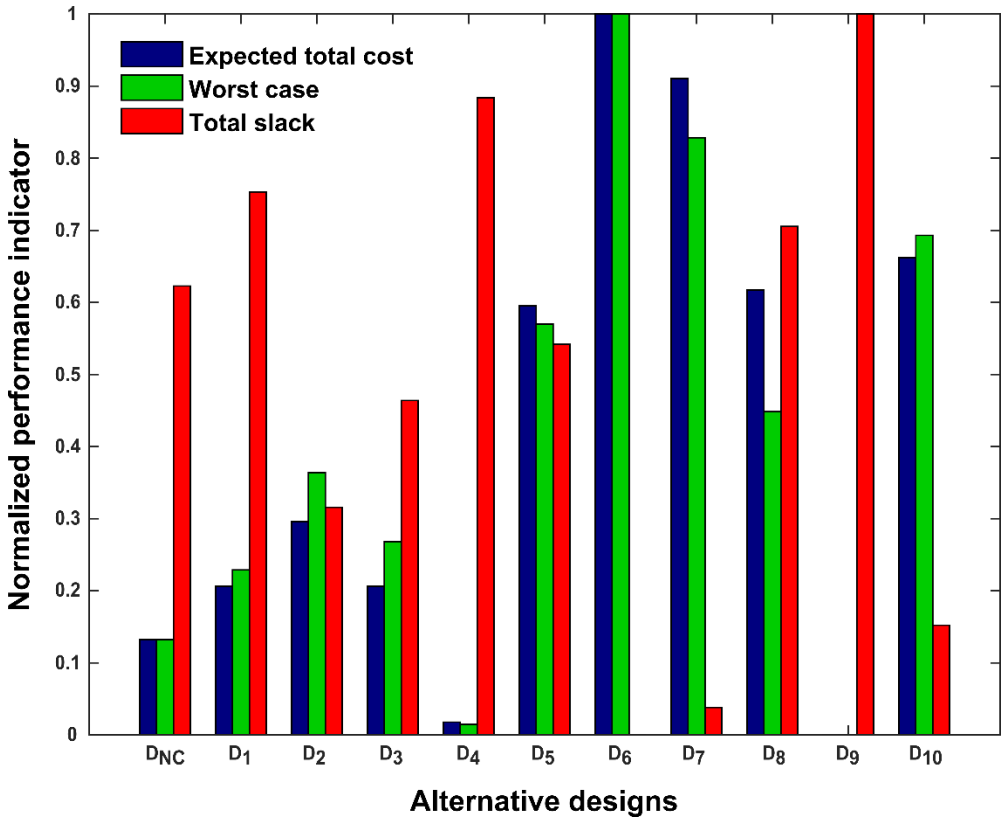


Figure 4 Normalized performances for the potential HEN design options, where lower values indicate better performance with respect to all the criteria

It can be observed from Figure 4 that the nominal case design (D_{NC}) is not the design with the lowest expected total cost, ranking third of 11 designs. Furthermore, this design does not perform particularly well with respect to the other two performance

indicators – in the worst case, its performance ranks third, and with respect to feasibility, as reflected by the 'total slacks' indicator, it ranks sixth. This observation supports a core premise of this work, i.e. a design that is based only on the nominal conditions may not perform well over the range of scenarios that can reasonably be expected to occur.

Figure 4 also shows that three designs (D_4, D_6 and D_9) are best in at least one or more objectives. Among the best solutions, D_9 has the lowest expected total cost and lowest worst case performance, while D_6 has the lowest penalties relating to slack variables. D_4 has a moderate expected total cost and slack-related penalty: these both are higher than in D_6.

It is valuable to gain a physical understanding of why the three best-performing designs did perform well. Firstly, it may be observed that D_6, with four rather than three heat transfer units, has the third highest total annualized cost. The presence of two heat exchangers, one heater and one cooler in D_6 apparently allows the network to cope well with variations in demand for heating and cooling and in stream supply temperatures. In particular, the heater and cooler provide flexibility to satisfy changes in the heating and cooling demand by increasing or decreasing the flow of steam and cooling water entering the heater and cooler respectively. In addition, D_6 includes a utility path in the network structure, as illustrated in Figure 7. A utility path in the heat exchanger network establishes a link between two or more utility exchangers. The utility path within the network allows both heat exchanger and utility exchanger duties to be adjusted, as heat demand and stream temperatures vary from one scenario to another.

While design D_4 is similar to D_9, it includes slightly less heat exchanger area (165.1 m^2), although the area of the heater is slightly higher than that in D_9. On the other hand, there is no utility path in design D_4. However, the presents of large heat transfer area that is 10 % greater than that in D_6 enables D_4 to cope with variation in stream temperatures

Given the performance information presented in Figure 4, it is not straightforward to select which of the three (identified) designs is the best. These results support that premise that a multi-criteria decision-making approach is required to identify the most flexible and operable HEN design.

4.1.4 Step 4: Selection of flexible and operable HEN

The AHP was applied using the pairwise comparison presented in Table 5. The Saaty scale of preferences introduced in Table 1 was applied to determine the relative importance of each performance indicator relative to the others. For example, if we assume the total slack indicator is five times more important than total expected cost, then the corresponding entry in the matrix takes the value of 5/1. Note that in the pairwise comparison of criteria in Table 5, only the upper part of the matrix is required, as the lower part is obtained from the inverse of the upper part.

Table 5 Pairwise comparison matrix for criteria employed for selecting best HEN design

	Expected total cost	Worst case	Total slack	Weights
Expected total cost	1	1/1	1/9	0.091
Worst case	1	1	1/7	0.112
Total slack	9	7	1	0.797

The weights associated with the matrix are shown in Table 5, which are calculated using equation (13). The performance in terms of penalties related to slack variables is the most important criterion, followed by the worst case and then the total cost. Note that these weights depend on decision-makers' preferences.

In the final step, the design alternative with the lowest overall performance indicator is determined. To carry out this task, the weights presented in Table 5 are used in equation (14) to calculate the aggregate performance for each design option shown in Figure 4. The weighted performance of each alternative design is then obtained, as

shown in Figure 5, where a lower value indicates a better performance in terms of both cost and violation of constraints.

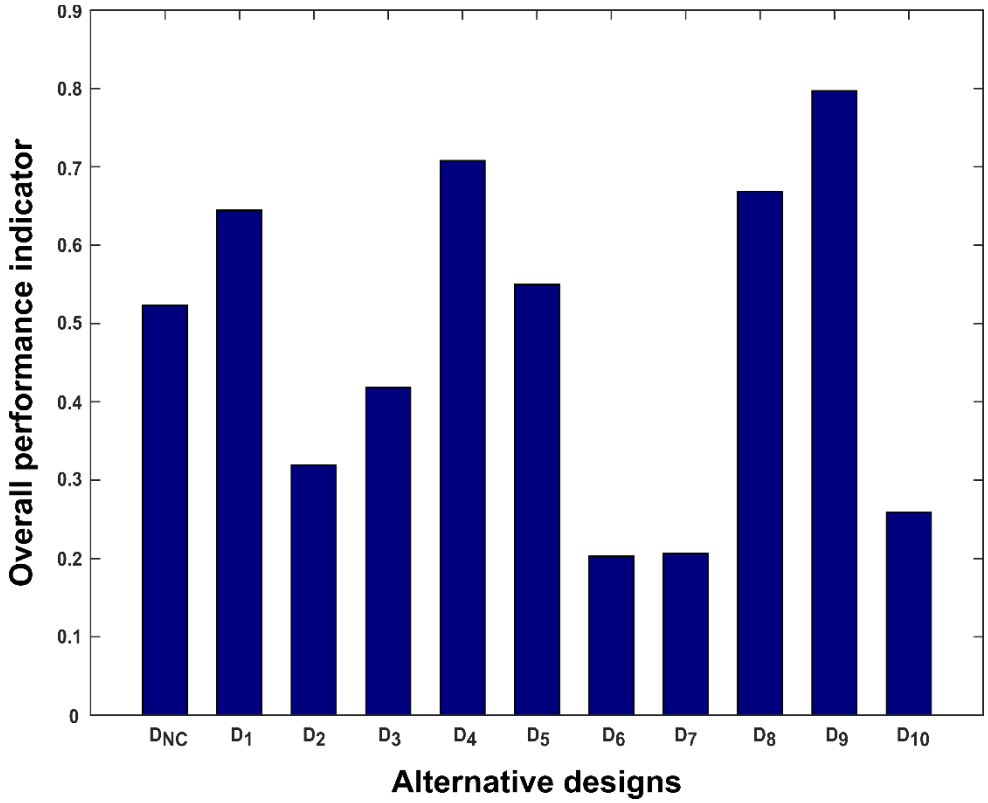


Figure 5 Overall ranking of alternative HEN design options

As can be observed in Figure 5, design D_8, with its exceptionally poor performance with respect to feasibility metrics, has the worst overall performance (0.82). Designs D_2, D_7 and D_{10} have a moderate performance of 0.32, 0.21 and 0.26, respectively, while D_6 has the best performance of 0.20, which is 0.32 lower than that of the nominal case design. Therefore, it can be concluded that design D_6 is the best one, according to the decision maker's preferences presented in Table 5, followed by design D_7.

While it is clear that design D_6 has the best overall performance using the values in Table 5, it is not necessarily the best for other weightings. We can assess the impact of the designer's preferences, expressed using different combinations of weighting factors, by constructing the ternary diagram presented in Figure 6. In the ternary diagram, each axis reflects the weight of that performance indicator, and the colors of the points reveal whether designs D_2 (red diamond), D_3 (gray square), D_4 (black

hexagram), D_6 (blue circle), D_7 (green pentagram), D_9 (red cross), and D_{NC} (green asterisk) is the best design overall.

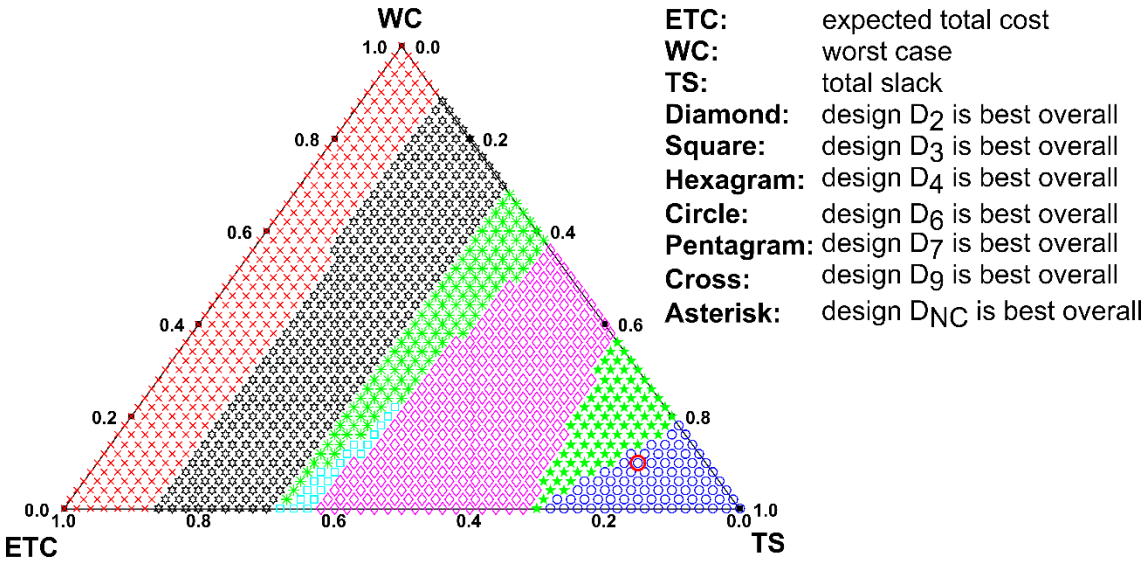
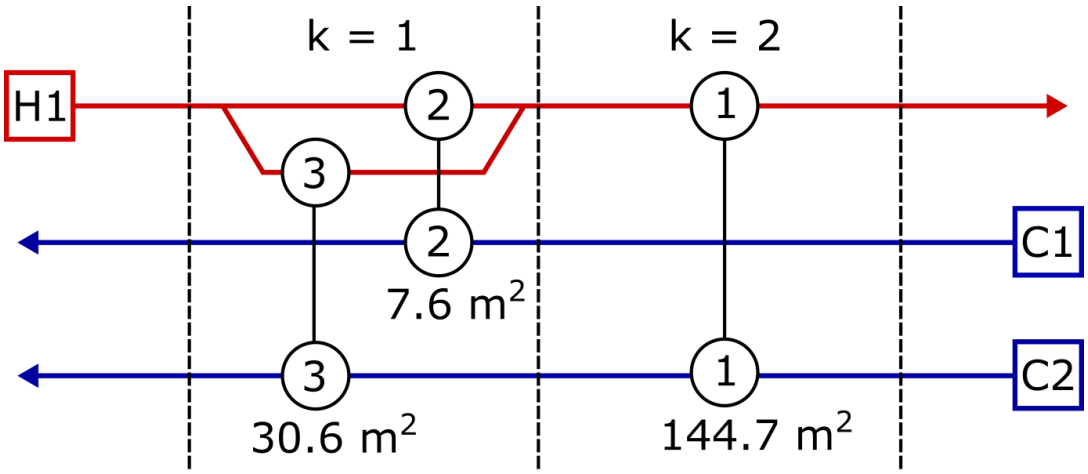


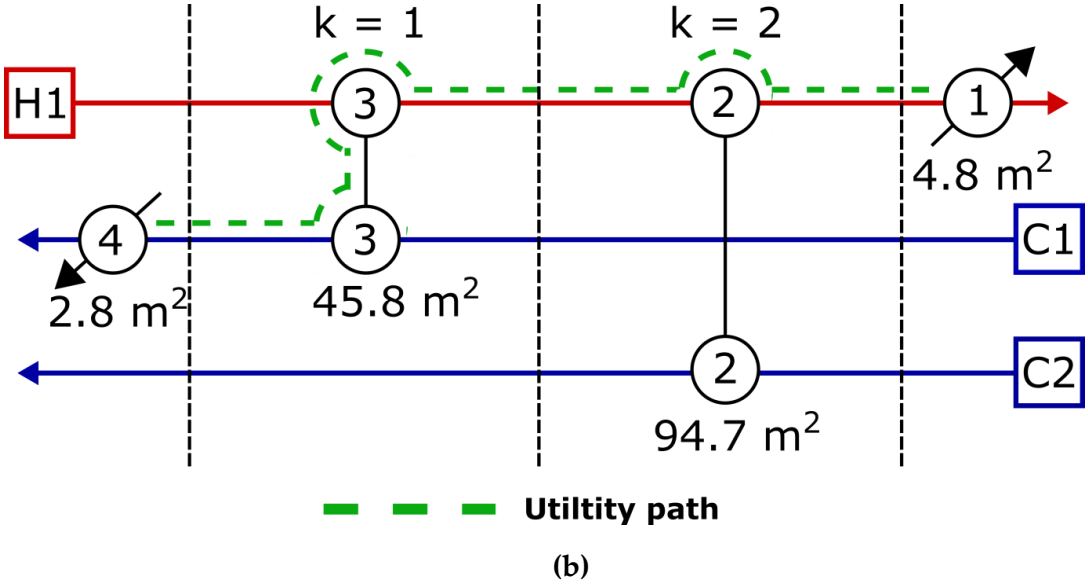
Figure 6 Relative ranking of alternative HEN designs

In Figure 6, the best design option identified in Figure 5 is indicated by a red circle. Clearly design D_9 dominates when decision-makers prefer low expected total costs and priorities lean towards minimizing costs for the worst-case scenario, while D_6 emerges as the optimal solution when decision-makers prefer constraints not to be violated (the total slack metric).

Next, we compare the best design option (D_6) with the nominal design; both are as presented in Figure 7 and Table 6.



(a)



(b)

Figure 7 Heat exchanger network (a) nominal case design D_{NC}, (b) optimal design D_6

The nominal case design has three exchangers with a total area of 183 m^2. Therefore, D_{NC} operates at maximum energy recovery with zero heater and cooler required. On the other hand, D_6 has two exchangers, one cooler and one heater, with a total heat transfer area of 148 m^2, respectively. As discussed in Section 4.4, the heater in design D_9 enhances flexibility of the heat exchanger network, as the flow rate of steam can be adjusted according to the heating demand in each scenario. Inevitably, a price must be paid for flexibility: the additional unit (heater and cooler) in design D_6 increases its operating cost and thus total annualized cost of the design, which is 19% higher than that of the nominal case design.

Table 6 Comparison of nominal case design (D_{NC}) and design D_6

	Nominal case design (D_{NC}) TAC = \$76,502/yr	Design D_6 TAC = \$94,641/yr
Exchanger	Area [m^2]	Area [m^2]
1	144.7	94.7
2	7.6	45.8
3	30.6	
4 (heater)	-	2.8
5 (cooler)		4.8
Total	183	148

4.2 Case 2: Design of flexible distillation column

In this case study, the design of a flexible distillation column for the separation of benzene from toluene and diphenyl, in a typical hydrodealkylation of toluene plant is used as a test bed example. Details of the whole flowsheet are presented in the supplementary material.

The column feed consists of benzene, toluene and diphenyl at a flow rate of 344.8 kmol/h; molar fractions of the three components are 0.11, 0.88 and 0.001 respectively. At least 95% of benzene in the feed stream is to be recovered, with purity of at least 0.95. The column consists of 40 initial trays (feed entering at the 20^{th} tray – counting from the bottom) and operates at a uniform pressure of 200 kPa. Hydraulic calculations are performed based on sieve tray, 85% approach to jet flooding, 50% down comer backup and tray spacing of 0.609 m.

4.2.1 Step 1: Description and sampling of uncertain operating conditions

Following a similar approach to the one demonstrated in Case study 1, the variability in the feed condition of the benzene distillation column is generated, as shown in Table 7. The feed condition considered includes temperature and component molar flow. Further details of the steps followed to generate the data are presented in Section S2 of the supplementary material. Note that for this case study, all the scenarios are assumed to be equally probable.

Table 7 Sampled data for uncertain feed conditions

Scenario	Temperature (^{0}C)	Benzene flow (kmol/h)	Toluene flow (kmol/h)	Biphenyl flow (kmol/h)
S_{NC}	132	37.6	306.0	1.2
S_1	131	38.5	279.2	1.5
S_2	132	36.5	322.5	1.0
S_3	130	40.5	215.5	2.3
S_4	132	35.6	344.0	0.9
S_5	130	40.0	232.0	2.0
S_6	133	34.3	378.6	0.7
S_7	131	38.4	299.6	1.4
S_8	131	39.3	246.1	1.8
S_9	129	40.4	196.6	2.4
S_{10}	131	38.7	269.6	1.5

*NC = Nominal case

4.2.2 Step 2: Design of distillation column for each scenario

The column superstructure originally proposed by Caballero et al. (2005) is adopted and used to design the flexible distillation column. To apply this approach, Murphree tray efficiency related to each tray in column sections (stripping and rectifying sections) is define as a binary variable (1 for active trays and 0 for inactive trays); then an upper and lower bound for trays in column sections is defined; and lastly, an optimization method is applied to select the number of active trays and column operating conditions required to achieve the desired separation at minimum total annualized cost. In this work, the column superstructure is set up in Aspen HYSYS simulation environment and the column optimization is carried out in MatLab. The optimization problem consisting of an objective function and constraints is presented in Section S2 of the supplementary material.

The MINLP problem is solved in MATLAB R2014a using genetic algorithm in the *Global Optimization Toolboox*, on a HP desktop PC with Intel(R) Core i5 processor running at 3.2 GHz, and 8 GB of installed RAM. An initial population of 50 chromosomes is used and the maximum number of generation is set to 100. An automation tool developed by Microsoft is used to facilitate the exchange of data between Aspen HYSYS and MatLab during optimization. The data exchanged includes

independent variables required to simulate the distillation column and dependent variables required to calculate the objective function and constraints. The independent variables consist of number of trays in column sections, and reflux and boil-up ratios, while the dependent variables consists of product specification and recovery, stream information, column diameter, jet flooding and down comer backup. It takes between 247 to 492 seconds to solve the deterministic MINLP model. Table 8 shows the details of the optimized designs, i.e., number of trays in column sections, required column diameter and total annualized cost (TAC) for each scenario.

Table 8 Optimal column design for each scenario

Optimal column design	Number of trays in rectifying section	Number of trays in stripping section	Column diameter (m^2)	Total annualized cost ($ y^{-1})
D_{NC}	9	10	1.98	103,877
D_1	10	11	1.83	100,687
D_2	10	11	1.98	109,824
D_3	9	7	1.83	85,172
D_4	10	13	1.98	116,004
D_5	9	8	1.83	87,966
D_6	11	13	2.13	128,928
D_7	9	10	1.98	103,934
D_8	9	8	1.83	89,308
D_9	8	7	1.83	81,350
D_{10}	9	8	1.98	96,315

*NC = Nominal case

As can be seen in Table 8, the nominal case design, D_{NC}, consists of 19 trays: 9 in the rectifying section and 10 in the stripping section, and a column diameter of 1.98 m – resulting to a total annualized cost of 103,877 $/y. Design D_9 has the least number of trays standing at 15, and a diameter of 1.83 m – resulting to the least total annualized cost (81,350 $/y) among the alternative designs. On the other hand, design D_6 has a total of 24 trays, which is the highest of all the scenario designs, and also the design has the largest diameter of about 2.13 m; consequently resulting to the greatest total annualized cost. Apparently, the design with best total annualized cost can be easily be identify from Table 8, however, it is not possible to identify the flexible design that can feasibly operate over the whole scenario. To achieve this task, further analysis is

necessary, which will involve evaluation of each alternative design over the entire scenarios.

The next task is to identify the column design that can feasibly operate over the entire scenario, that is, the column design that can meet product quality specifications for all scenarios without exceeding jet flooding and down comer backup limits.

4.2.3 Step 3: Assessment of flexible columns

In this step, the performance of the individual scenario designs together with the nominal case design is assessed over the entire uncertain space, as represented using scenarios. The assessment is performed by fixing the number of trays in the rectifying and stripping sections, and the column diameter in each design, then all the operating scenarios (in Table 7) are optimized on the designs. Similar to Case study 1, the performance indicators used to assess each design, including expected total cost, worst case performance and total slack (constraint violation) are recorded and presented in Figure 8. In this case, the total slack relates to violation in product quality specifications and recovery (benzene), and column jet flooding and down comer backup. Note that computational results indicate that all the designs were able to meet product specifications for all scenarios, although some design alternatives have failed to satisfy the hydraulic demand for some scenarios. Therefore, the magnitude of the penalty related slack is predominately jet flooding and down comer backup.

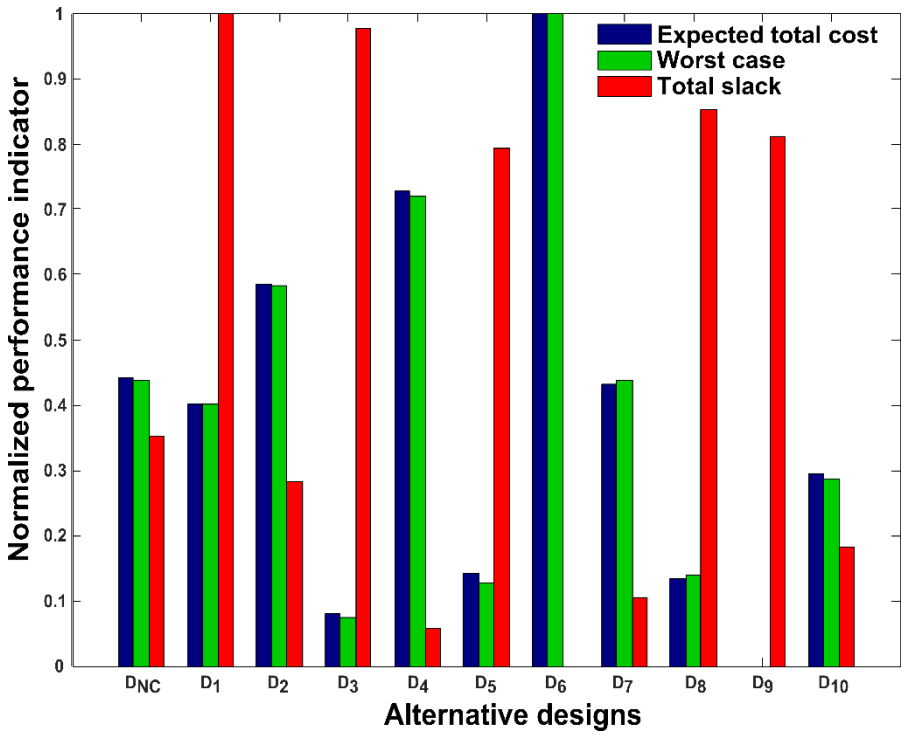


Figure 8 Normalized performances for the potential column design options, where lower values indicate better performance with respect to all the criteria

The assessment of the alternative designs against the expected variation in feed condition provides more insight on the individual design performance. As can be seen in Figure 8, the nominal case design (D_{NC}) is unable to cope with the changing feed conditions, as indicated by the performance indicator with respect to total slack (constraint violation). Again, this observation further emphasis the core premise of this work; that is, a design carried out based on the nominal conditions may failed to perform well over an expected variability in operating condition.

Another important observation is that design D_9 has the best performance with respect to both expected total cost and worst case, however, the design will perform poorly over the expected variations in feed condition, as indicated by the slack variable (rank 7th). Although D_6 has the overall highest performance with respect to expected total cost and worst case, the design has the best performance with respect to slack variable followed by designs D_4, D_7 and D_{10} – ranking 2nd, 3rd and 4th respectively. The low value of slack variable indicates that these designs have the ability to cope with the expected variability in the feed condition.

A critical look at design D_6 reveals two important factors that contribute to the flexibility and operability of the column. First, Design D_6 has the highest number of trays (24) among all the alternative designs, which apparently increases the installed capital cost of the column. The large number of trays enables the column to operate over different operating scenarios, providing flexibility for the design to meet the required separation whenever the feed composition, flow rate and temperature changes from one scenario to another. Second, design D_6 has the largest column diameter of about 2.13 m. Although the large column diameter incurs high installed capital cost, it allows efficient handling of substantial amount of vapour and liquid traffic in the column, thereby avoiding the column from being bottlenecked, as throughput changes from one scenario to another.

Design D_4 has the same diameter as D_7 and D_{10}, although the number of trays is slightly higher. Other designs such as D_1, D_3, D_5, D_8 and D_9 have the least diameter of 1.83 m, consequently resulting to a very poor performance with respect to total slack variable; this implies that the aforementioned designs cannot efficiently handle the hydraulic demand of the entire operating scenarios.

Considering the number of competing alternative designs in Figure 8, it is difficult to select the best design. Hence, the results further support the idea that a multi-criteria-decision-making tool will be required to identify the least expensive distillation column that is flexible and operable.

4.2.4 Step 4: Selection of flexible and operable column

In this step, the pairwise comparison matrix described in Section 3 is applied to generate weights corresponding to each performance criteria (expected total cost, worst case performance and total slack), as shown in Table 9. The magnitude of the weight indicates the importance of one performance criterion relative to the other ones. Table 9 is generated in same way as Table 5.

Table 9 Pairwise comparison matrix for criteria employed for selecting the best column design

	Expected total cost	Worst case	Total slack	Weights
Expected total cost	1	1/3	1/7	0.096
Worst case	5	1	1/5	0.162
Total slack	7	3	1	0.742

The calculated weights in Table 9 indicates that total slack relating constraint violation is the most important criterion, followed by worst case and expected total cost (an indication that the design with the lowest constraint violation is the most preferred chioce). Note that the weight are not unique, but depends on the designer's preference.

Next, the weights are applied in equation (14) together with the performance indicators in Figure 8 to determine the aggregate perfomance indicator for each alternative design, as shown in Figure 9

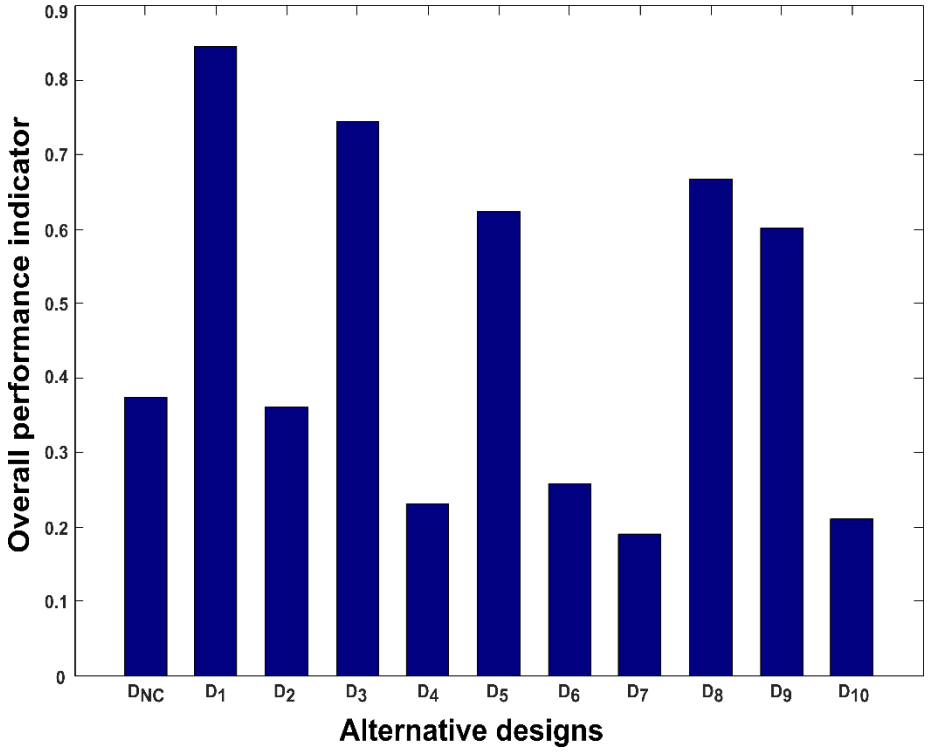


Figure 9 Overall ranking of alternative column design options

As can be seen in Figure 9, designs D_6, D_4, D_7 and D_{10} have better overall performance compared with the remaining designs. Design D_1 with an overall poor performance with respect to total slack criterion, has the overall worst performance (0.86). The nominal case design, D_{NC}, and design, D_2, have a moderate performance of 0.48 and 0.47 respectively. Although, it is apparent that design D_7 is the best design (0.2) identified thus far, it is limited to the information the decision maker enters into the pairwise comparison matrix (i.e. it may not necessarily be the overall best design in all cases). To assess the impact of decision maker's preference (i.e. by applying different combination of weights representing expected total cost, worst case performance and total slack), a ternary diagram is constructed, as presented in Figure 10.

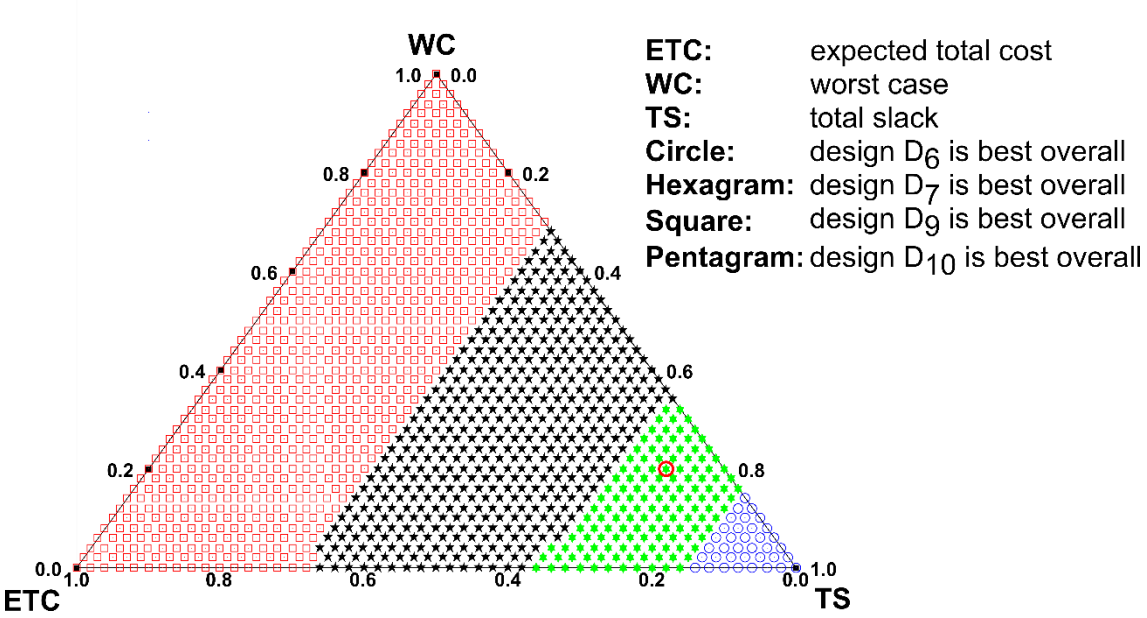


Figure 10 Relative ranking of alternative column designs

In the ternary diagram, each axis represents a performance criterion, and the points in red (square), black (pentagram), green (hexagram) and blue (circle) reveals when designs D_9, D_{10}, D_7 and D_6 are the best design overall. Apparently, design D_9 dominates when the decision maker prefers minimum expected total cost and minimum worst case performance. Conversely, D_6 is the best choice when the decision maker prefers smooth operation over the entire scenario (i.e. minimum constraint violation). Other

designs such as D_7 and D_9 dominate when the minimum of all three objectives is preferred.

5. Conclusions

This paper proposes a new methodology for design of chemical processes under uncertainty. The methodology characterizes uncertainty using probability distributions and Monte Carlo sampling. Process synthesis and evaluation are carried out using mathematical modeling and optimization techniques. The best design alternative(s) are selected using the analytic hierarchy process.

The methodology is illustrated through its application to the design of two relevant chemical processes: flexible HENs, applying a modified version of the MINLP model of Yee and Grossmann (1990) to generate the flexible HEN alternatives; and flexible distillation column.

The numerical results for both case studies demonstrate that the proposed methodology is capable of synthesizing a chemical process that is flexible and feasible (operable) under variable operating conditions. In addition, it was found that subjective preferences represented using a pairwise comparison matrix, significantly influence which design alternative is best.

Future work will expand the scope and complexity of the problems addressed towards developing an approach that is capable of providing practical and economic solutions of industrial relevance, for example flexible crude oil distillation systems.

Acknowledgement

The authors would like to acknowledge the financial support from Petroleum Technology Development Fund (PTDF), Nigeria, for sponsoring this PhD research.

Nomenclature

Abbreviations

HEN heat exchanger network

TAC total annualized total cost

AHP analytic hierarchy process

WC worst case

WTC weighted total cost

TS total slack

CR consistency ratio

LP linear programming

NLP non-linear programming

MINLP mixed integer non-linear programming

PFD probability distribution function

Indices:

i hot process stream

j cold process stream

k superstructure stage

s scenario

Sets:

HP set of hot process streams

CP set of cold process streams

ST set of superstructure stages

S set of scenarios

Parameters:

TIN inlet temperature of process streams

$TOUT$ outlet temperature of process streams

F heat capacity flow rate of process streams

U overall heat transfer coefficient

CCU unit cost for cold utility

CHU unit cost for hot utility

CF fixed charge for exchangers (heaters and coolers included)

C area cost coefficient

β index for area cost

NOK total number of stages

Ω upper bound for heat exchange duty

Γ upper bound for temperature difference

ΔT_{min} minimum approach temperature

$Aex_{i,j,k}$ heat exchanger area for match i and j in stage k

Acu_i cooler area for match i

Ahu_j heater area for match j

π penalty (scalar multiplier)

$prob_s$ probability of scenario, s, occurring

Variables:

$t_{i,k,s}$ temperature of hot stream i at hot end of stage k in scenario s

$t_{j,k,s}$ temperature of cold stream j at hot end of stage k in scenario s

$q_{i,j,k,s}$ heat exchanged between hot stream i and cold stream j in stage k for scenario s

$qcu_{i,s}$ heat exchanged between hot stream i and cold utility

$qhu_{j,s}$ heat exchanged between cold stream j and hot utility

$dt_{i,j,k,s}$ temperature approach for match i and j at temperature location k

$dtcu_{i,s}$ temperature approach for match between hot stream i and cold utility

$dthu_{i,s}$ temperature approach for match between cold stream j and hot utility

$z_{i,j,k}$ binary variable to signify that match i and j exist in stage k

zcu_i binary variable to signify that cold utility exchange heat with hot stream i

zhu_j binary variable to signify that hot utility exchange heat with cold stream j

$Sh1_{i,s}$ positive slack variable for hot stream i in scenario s

$Sh2_{i,s}$ negative slack variable for hot stream i in scenario s

$SC1_{i,s}$ positive slack variable for cold stream j in scenario s

$SC2_{i,s}$ negative slack variable for cold stream j in scenario s

Appendix A: Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://

References

Amaran, S., Zhang, T., Sahinidis, N. V., Sharda, B., Bury, S.J., 2016. Medium-term maintenance turnaround planning under uncertainty for integrated chemical sites. Comput. Chem. Eng. 84, 422–433.

Bertsimas, D., Brown, D.B.D., Caramanis, C., 2010. Theory and Applications of Robust Optimization. Oper. Res. 50.

Biegler, L. T., Grossmann, I. E. and Westerberg, A. W. (1997). *Systematic Methods of Chemical Process Design*. Upper Saddle River, N.J.: Prentice Hall PTR.

Caballero, J.A., Milan-Yanez, D and Grossmann I.E. (2005). Rigorous design of distillation columns: Integration of disjunctive programming and process simulators. *Ind Eng Chem Res*, 44(17), 6760-6775.

Chacon-Mondragon, O. L. and Himmelblau, D. M. (1996). Integration of flexibility and control in process design. *Computers & Chemical Engineering,* 20(4)**,** 447-452.

Chen, C. L. and Hung, P. S. (2004). Simultaneous synthesis of flexible heat-exchange networks with uncertain source-stream temperatures and flow rates. *Industrial & Engineering Chemistry Research,* 43(18)**,** 5916-5928.

Diwekar, A. M. (2008). *Introduction to Applied Optimization, Second Edition.* New York: Springer.

Diwekar, U. M. (2003). A novel sampling approach to combinatorial optimization under uncertainty. *Computational Optimization and Applications,* 24(2-3)**,** 335-371.

Diwekar, U. M. and Rubin, E. S. (1994). Parameter design methodology for chemical processes using a simulator. *Industrial & Engineering Chemistry Research,* 33(2)**,** 292-298.

Floudas, C. A. and Grossmann, I. E. (1986). Synthesis of flexible heat-exchanger networks for multiperiod operation. *Computers & Chemical Engineering,* 10(2)**,** 153-168.

GAMS, (2012). Release 24.0.2. GAMS Development Corporation. Washington, DC, USA.

Georgiadis, M.C., Pistikopoulos, E.N., 1999. An Integrated Framework for Robust and Flexible Process Systems An Integrated Framework for Robust and Flexible Process Systems 38, 133–143.

Grossmann, I. E. and Guillen-Gosalbez, G. (2010). Scope for the application of mathematical programming techniques in the synthesis and planning of sustainable processes. *Computers & Chemical Engineering,* 34(9)**,** 1365-1376.

Grossmann, I. E., Halemane, K. P. and Swaney, R. E. (1983). Optimization strategies for flexible chemical processes. *Computers & Chemical Engineering,* 7(4)**,** 439-462.

Guthrie, K.M. (1969). Capital Cost Estimating. *Chem Eng*, 114–142.

Halemane, K.P., Grossmann, I.E., 1983. Optimal process design under uncertainty. AIChE J. 29, 425–433.

Ho, W. (2008). Integrated analytic hierarchy process and its applications - A literature review. *European Journal of Operational Research,* 186(1)**,** 211-228.

Ho, W., Dey, P. K. and Higson, H. E. (2006). Multiple criteria decision-making techniques in higher education. *International Journal of Educational Management,* 20(5)**,** 319-337.

Kleywegt, A. J., Shapiro, A. and Homem-De-Mello, T. (2001). The sample average approximation method for stochastic discrete optimization. *Siam Journal on Optimization,* 12(2)**,** 479-502.

Korpela, J. and Lehmusvaara, A. (1999). A customer oriented approach to warehouse network evaluation and design. *International Journal of Production Economics,* 59(1-3)**,** 135-146.

Kostin, A.M., Guillén-Gosálbez, G., Mele, F.D., Bagajewicz, M.J., Jimenez, L., 2012. Design and planning of infrastructures for bioethanol and sugar production under demand uncertainty. Chem. Eng. Res. Des. 90, 359–376

MATLAB, (2014). version R2014a. The MathWorks, Inc.: Natick, Massachussetts, United States.

Pistikopoulos, E.N., Ierapetritou, M.G., 1995. Novel approach for optimal process design under uncertainty. Comput. Chem. Eng. 19, 1089–1110.

Pistikopoulos, E. N. and Grossmann, I. E. (1988a). Optimal retrofit design for improving process flexibility in linear-systems. *Computers & Chemical Engineering,* 12(7), 719-731.

Pistikopoulos, E. N. and Grossmann, I. E. (1988b). Evaluation and redesign for improving flexibility in linear-systems with infeasible nominal conditions. *Computers & Chemical Engineering,* 12(8), 841-843.

Rogers, A., Ierapetritou, M., 2015. Feasibility and flexibility analysis of black-box processes part 2: Surrogate-based flexibility analysis. Chem. Eng. Sci. 137, 1005–1013.

Saaty, R. W. (1987). The analytic hierarchy process - what it is and how it is used. *Mathematical Modelling,* 9(3-5), 161-176.

Saaty, T. L. (1990). How to make a decision - the analytic hierarchy process. *European Journal of Operational Research,* 48(1), 9-26.

Saaty, T. L., Vargas, L. G. & Dellmann, K. (2003). The allocation of intangible resources: the analytic hierarchy process and linear programming. *Socio-Economic Planning Sciences,* 37(3), 169-184.

Sahinidis, N. V. (2004). Optimization under uncertainty: state-of-the-art and opportunities. *Computers & Chemical Engineering,* 28(6-7), 971-983.

Shang, J. and Sueyoshi, T. (1995). A unified framework for the selection of a flexible manufacturing system. *European Journal of Operational Research,* 85(2), 297-315.

Smith, R. (2005). *Chemical Process Design And Integration*. Chichester, West Sussex, England ; Hoboken, NJ: Wiley.

Su, J. C. Y., Chen, S. J. and Lin, L. (2003). A structured approach to measuring functional dependency and sequencing of coupled tasks in engineering design. *Computers & Industrial Engineering,* 45(1), 195-214.

Swaney, R. E. and Grossmann, I. E. (1985). An index for operational flexibility in chemical process design .2. Computational algorithms. *AIChE Journal,* 31(4), 631-641.

Wagler, R. M. and Douglas, P. L. (1988). A method for the design of flexible distillation sequences. *Canadian Journal of Chemical Engineering,* 66(4)**,** 579-590.

Wang, G., Huang, S. H. and Dismukes, J. P. (2005). Manufacturing supply chain design and evaluation. *International Journal of Advanced Manufacturing Technology,* 25(1-2)**,** 93-100.

Wang, H., Mastragostino, R. and Swartz, C. L. E. (2016). Flexibility analysis of process supply chain networks. *Computers & Chemical Engineering,* 84**,** 409-421.

Wang, J., Rong, G., 2010. Robust optimization model for crude oil scheduling under uncertainty. Ind. Eng. Chem. Res. 49, 1737–1748.

Yee, T. F. and Grossmann, I. E. (1990). Simultaneous-optimization models for heat integration .2. Heat-exchanger network synthesis. *Computers & Chemical Engineering,* 14(10)**,** 1165-1184.

Yee, T. F., Grossmann, I. E. and Kravanja, Z. (1990). Simultaneous-optimization models for heat integration .3. Process and heat-exchanger network optimization. *Computers & Chemical Engineering,* 14(11)**,** 1185-1200.

Yeomans, H. and Grossmann I.E. (2000). Optimal design of complex distillation columns using rigorous tray-by-tray disjunctive programming models. *Ind Eng Chem Res*, 39(11), 4326-4335.

Yurdakul, M. (2004). Selection of computer-integrated manufacturing technologies using a combined analytic hierarchy process and goal programming model. *Robotics and Computer-Integrated Manufacturing,* 20(4)**,** 329-340.

Zhang, Q., Lima, R. M. and Grossmann, I. E. (2016). On the relation between flexibility analysis and robust optimization for linear systems. *AIChE Journal***,** doi: 10.1002/aic.15221.

Chapter 6 Conclusions and future work

6.1 Conclusions

Today, the petroleum refining industry is faced with several challenges, including high operating cost; stringent regulations on product specifications and CO_2 emissions (greenhouse gas emissions leading to global warming); and uncertainties in terms of the quality and quantity of crude oil feedstocks that need to be processed in order to maximise the production of market-driven petroleum refined products. To address these challenges, systematic methodologies for the design and optimisation of refinery heat-integrated crude oil distillation systems are needed.

This thesis presents new methodologies for the design of heat-integrated crude oil distillation systems. First, two new methods that facilitate design of a crude oil distillation system that process a specific type of crude oil feedstock are proposed (see Chapters 3 and 4). Second, a methodology that takes into account multiple crude oil feedstocks to design a flexible heat-integrated crude oil distillation systems is developed (see Chapter 5). The three proposed methodologies are optimisation-based, in which the distillation column models, pinch analysis for minimum utility calculations, and process constraints are incorporated into a unified framework to facilitate the search for a cost-effective design configuration that maximises net profit and/or minimises total annualised cost. Third, a scenario-based approach is developed for design of chemical processes in which some parameters are subject to variability (see Chapter 5).

The main contributions of the work presented in this thesis are summarised below:

6.1.1 Design of heat-integrated crude oil distillation systems using rigorous simulation models

Existing methodologies for modelling of crude oil distillation units using rigorous models takes into account only the continuous variable (operating conditions) of the system. Design variables such as number of trays in column sections and locations of feed, pump-around, and side-strippers are fixed. Thus, the distillation column models are not applicable for optimisation-based design of crude oil distillation systems.

In this thesis, a new approach for representing crude oil distillation units using rigorous models is developed (see Chapter 3). The distillation column model presented here takes into account both structural (number of trays in column section) and operational degrees of freedom of the system. When this model is implemented in an optimisation framework, it is possible to vary both column structure and operating conditions in order to select the best design option among several alternatives.

In the new modelling approach, the column superstructure of Caballero et al. (2005) for simple columns is adapted to build the column superstructure of the crude oil distillation unit using a rigorous tray-by-tray model in Aspen HYSYS. In the column superstructure, Murphree tray efficiencies are treated as binary variables. Thus it is possible to vary the total trays in the column by specifying a tray efficiency of either one or zero. It is important to emphasise here that this is the first attempt to include number of trays as a design variable in modelling of crude oil distillation units using rigorous simulation models.

In this work, the superstructure of the crude oil distillation unit is implemented in an optimisation framework together with pinch analysis (for calculating minimum utility requirements) in order to search for column design that minimises consumption of cooling water and fuel in the furnace, while ensuring product quality specifications are satisfied. The use of pinch analysis here enables the framework to account for interactions between the distillation column and the heat recovery network. Both the structure and operating conditions of the column are optimised, thus exploiting trade-offs between capital and energy costs.

The proposed methodology takes advantage of many useful models (computational routines) available in the commercial process simulation package to produce an accurate and realistic design solution that can be implemented in practice. Examples of models available in the process simulator include crude oil characterisation models, physical and thermodynamic property models and column hydraulic models. Moreover, the commercial process simulator environment is versatile and user-friendly, and Aspen HYSYS software is widely applied, making the proposed approach easy to implement in practice and accessible to both researchers and industrial practitioners.

The capabilities of the proposed design methodology are illustrated using industrially-relevant examples. In the first example, the total annualised cost of the crude oil distillation unit is optimised by varying column number of trays and operating conditions such as pump-around temperature drops and duties, stripping steam flow rate, feed temperature and reflux ratio. Numerical results show that a crude oil distillation unit with an improved total annualised cost can be obtained without affecting product quality constraints.

However, these results also reveal a slight increase in residue flow rate which is undesirable (some valuable products are lose to the residue stream). To avoid losing valuable product to the residue, the second example includes constraints on both product quality and flow rate (more specifically the residue stream flow rate). In the two examples, the optimal distillation column structure differs from the initial starting design (base case). This result highlights the need to optimise the column structure in addition to operating conditions, which cannot be achieved using existing methods. Although the proposed method has been successfully applied to design cost-effective crude oil distillation units, significant computational time is required. This makes the approach unsuitable for process design with a large number of operating scenarios, for example, the design of flexible crude oil distillation system.

6.1.2 Design of heat-integrated crude oil distillation systems using surrogate models

Chapter 4 explores the use of surrogate models to replace the rigorous simulation models. These surrogate models are fitted to a set of samples generated using the rigorous simulation model.

Established approach using surrogate models for crude oil distillation units do not consider both structural and operational degrees of freedom of the complex system. Thus, these surrogate models are not applicable for the design of crude oil distillation systems.

This thesis proposes a new methodology for modelling crude oil distillation unit using surrogate models. Unlike previous methods, this work takes into account the structural and operational degrees of freedom, thus expanding the scope of application to grassroots design of crude oil distillation systems.

An artificial neural network is applied to construct the surrogate model of the complex distillation column. The parameters of the artificial neural network are regressed using data set generated from using Latin Hypercube sampling. Independent variables that have a significant effect on capital investment and energy requirements of the system are used. Dependent variables of the system that enable evaluation of the system performance and allow checking of constraints on product quality and flow rate are applied.

Data sampling considering both structural and operating degrees of freedom for design is highly combinatorial: hence not all samples will satisfy material and energy balance, phase equilibrium and product quality constraints, i.e. will correspond to 'infeasible' design. In this work, a support vector machine is used to remove infeasible points from the design space, thus increasing the likelihood that a design selected during optimisation will be feasible and helping to reduce computational effort by restricting the design space.

In this thesis, the artificial neural network column model and a support vector machine are incorporated in an optimisation framework to aid the design of the crude oil distillation system. Pinch analysis is used within the framework to predict minimum utility requirements. Thus, the interactions between the column and heat recovery network are captured.

In Chapter 4, a case study is presented to illustrate the capabilities of the proposed methodology for identifying cost-effective design solutions that meets both product quality and flow rate constraints. When the optimised solution is simulated using the rigorous model, good agreement is seen between the performance predicted by the surrogate and rigorous models. Thus the approach is reliable and can be applied in the refining industries for design, analysis, and optimisation of crude oil distillation systems. As expected, the computational time has been significantly reduced as compared with the approach that implements rigorous models (see Chapter 3).

6.1.3 Design of flexible heat-integrated crude oil distillation systems

In practice, petroleum refineries process different types of crude oil feedstocks and blends. A few works (Bagajewicz and Ji, 2001; More et al., 2010) focus on the design of crude oil distillation systems that can process multiple crude oil feedstocks. Nevertheless, important practical issues, such trade-offs between capital and energy costs and simultaneous optimisation of column structural and operational degrees of freedom have not been addressed. That is, there are no systematic approaches for design and optimisation of flexible heat-integrated crude oil distillation systems

This thesis presents a systematic methodology for the design of flexible crude oil distillation systems that addresses the limitations of existing methods (see Chapter 5). The methodology models the distillation column, heat recovery system and all crude oil feedstocks to be processed in a unified framework. Optimisation is applied to select a column configuration that is not only operable across the range of feedstocks, but also economical and energy-efficient.

The surrogate modelling approach for a single crude oil proposed in Chapter 4 is modified and extended to address multiple crude oils. A surrogate model is developed for each crude oil to be processed. A two-stage optimisation framework applies the surrogate model for each crude oil, together with pinch analysis. One advantage of this approach is that the best column configuration is selected while accounting for minimum utility requirements, thus capturing the complex interactions between the two sub-systems, and the corresponding trade-offs between capital and energy costs.

An effective solution strategy has been proposed that decouples the problem into two stages. In Stage 1, the optimal flexible column configuration that can operate over the entire set of crude oils to be processed is selected, while Stage 2 selects the optimal operating conditions for each crude oil to be processed. The approach has the advantage of allowing the optimisation algorithms to thoroughly explore the problem design space (in Stage 1) and operating space (in Stage 2).

Chapter 5 presents a case study that demonstrates the capabilities of the proposed methodology. The case study is concerned with the design of a flexible crude oil distillation system that separates three types of crude oil. The objective is to maximum profit, taking into account column capital cost, maximum energy recovery and product quality constraints. The results provide evidence that the proposed approach can be applied in the refining industries to design flexible crude oil distillation systems. One advantage of this design approach, compared with established methods, is that the final design is optimal and satisfies relevant practical constraints (e.g., product quality), helping the proposed approach to generate realistic solutions.

The design of a flexible crude oil distillation unit is a highly combinatorial problem. Decoupling the problem into two levels reduces the level of complexity, helping the problem to be solved effectively. If a large number of crude oil is to be processed, the solution method becomes computationally demanding and very difficult to solve.

To overcome the limitation of the two-stage optimisation approach, a new scenario-based design approach is also proposed in this thesis. The main advantage of this

method is that the complex and highly combinatorial problem is fully decoupled into two sub-problems that are solved in separate stages. In the first stage, an optimal design is generated for each operating scenario. In the second stage, each design is optimised considering all the operating scenarios, taking into account important criteria, such as expected total cost, total slack variables (related to violation of constraints) and worst case performance. In the final step, a multi-criteria decision-making tool (in this case, the Analytic Hierarchy Process) is used to select the best design alternative among several others.

The capabilities of the scenario-based design approach is demonstrated using two industrially-relevant examples: design of a heat exchange network with uncertain inlet stream conditions (inlet temperature and heat capacity flow rate) and design of a distillation column with variable inlet conditions (feed flow rate, temperature and pressure).

In this thesis, the proposed methodologies for the design of crude oil distillation systems accounts for heat integration using pinch analysis, without considering the design of the heat exchanger network. Therefore the capital cost of the heat exchanger network is not considered in the analysis. Nevertheless, the optimisation algorithm minimises the cost of hot and cold utility demand (calculated using pinch analysis), which represents the dominant cost of the heat exchanger network. Additionally, the stream information of the optimised crude oil distillation system can be used to design the heat exchanger network.

This research work focuses on the design of flexible crude oil distillation system that can process multiple crude oil feedstocks. In practice, for a refinery to effectively process multiple feedstocks, other downstream units such as vacuum distillation unit, fluid catalytic cracking unit, catalytic reformer unit, hydrocracker unit etc. need to be flexible as well. Although with little engineering effort, the methodology proposed in Chapter 5 can be extended to design the aforementioned downstream units.

6.2 Future work

Future work in this area can extend the methodology presented in this thesis to address the following issues:

1. The new modelling approach presented in this work takes into account both structural and operating conditions of the crude oil distillation unit. The structural variables considered in this work are limited to the number of trays in each column section. The approach can be extended to incorporate additional structural variables, such as pump-around and side-stripper locations, and the feed inlet location.

2. The new modelling approach presented in this work focuses on the atmospheric distillation column. Modelling of other processing units such as a flash unit, a pre-fractionator and the vacuum distillation unit can be addressed, modelled and incorporated into the proposed methodology.

3. In this work, heat integration is incorporated into the design framework using pinch analysis. The proposed methodology can be extended to include details of the heat exchanger network configuration and of the individual heat exchangers.

4. The methodology proposed in this work focuses on flexible design of new crude oil distillation units. The approach can be extended to retrofit design, taking into account various modifications to allow an existing distillation column to process wide range of feedstock.

5. The flexible design approach presented in this thesis is quite general, thus there is potential for it to be extended to other chemical and petrochemical processes, for example fluid catalytic cracking unit, catalytic reformer unit, hydrocracker unit, hydrotreater unit etc.

Appendix A Data for Publications 1, 2, 3 and 4

This appendix consists of four sections, in which the Supporting Information/ supplementary material for Publications 1, 2, 3 and 4 are presented respectively. Each Supporting Information is submitted together with the corresponding paper for publication in reputable journals in the field of Chemical Engineering.

A.1 Supporting Information for Publication 1

Ibrahim, D., Jobson, M., Guillén-Gosálbez, G., 2017. Optimization-based Design of Crude Oil Distillation Units using Rigorous Simulation Models. *Ind. Eng. Chem. Res*., 2017, 56 (23), pp 6728–6740, **DOI:** 10.1021/acs.iecr.7b01014

Supporting Information

Table S1 shows the crude oil assay that was used in the case study in Section 4.

Table S1 Physical properties of crude oil (crude oil assay)

Name (country of origin)	Tia Juana light (Venezuela)	
Type	Light	
Bulk properties		
i. Density (kg m^{-3})	867.6	
ii. API gravity (ºAPI)	31.6	
Distillation properties		
i. Light end analysis	Component name	Volume %
	Ethane	0.04
	Propane	0.37
	i-Butane	0.27
	n-Butane	0.89
	i-Pentane	0.77
	n-Pentane	1.13
ii. TBP curve	Temperature (°C)	Volume %
	36.1	0
	64.4	5
	100.6	10
	163.9	20
	221.1	30
	278.9	40
	337.2	50
	397.2	60
	463.9	70
	545.0	80

The crude oil assay in Table S1 is cut into 25 pseudo-components using the standard oil characterization procedure in Aspen HYSYS v8.6. Table S2 presents the normal boiling temperature, compositions and volumetric flow rates of both the pure components and the pseudo-components.

Table S2 Crude oil feedstock characterization

Component name	NBP (°C)	Volume fraction	Volumetric flow ($m^3\ h^{-1}$)
Ethane	-89	0.040	0.265
Propane	-42	0.370	2.451
i-Butane	-12	0.270	1.789
n-Butane	-1	0.890	5.896
i-Pentane	28	0.770	5.101
n-Pentane	36	1.130	7.486
NBP_47	47	4.250	28.157
NBP_72	72	3.371	22.331
NBP_97	97	3.263	21.616
NBP_122	122	3.654	24.207
NBP_146	146	3.848	25.491
NBP_171	171	4.028	26.683
NBP_195	195	4.148	27.479
NBP_219	219	4.147	27.473
NBP_244	244	4.083	27.048
NBP_268	268	4.071	26.970
NBP_293	293	4.063	26.916
NBP_317	317	4.031	26.705
NBP_341	341	4.018	26.615
NBP_366	366	3.947	26.147
NBP_390	390	3.822	25.319
NBP_415	415	3.673	24.332
NBP_449	449	6.153	40.762
NBP_493	493	5.466	36.208
NBP_538	538	4.875	32.294
NBP_581	581	4.264	28.246
NBP_625	625	3.214	21.288
NBP_685	685	4.781	31.674
NBP_771	771	2.809	18.609
NBP_858	858	1.408	9.328
NBP_950	950	1.142	7.563

Tables S3 to S6 present the initial design data for the base case.

Table S3 Initial operating conditions

Operating variable	Initial value
PA 1 duty (MW)	11.2
PA 2 duty (MW)	17.89
PA 3 duty (MW)	12.84
PA 1 DT (ºC)	20
PA 2 DT (ºC)	50
PA 3 DT (ºC)	30
Main stripping steam (kmol h^{-1})	1200
HD stripping steam (kmol h^{-1})	250
Feed temperature (ºC)	365
Reflux ratio	4.17

Table S4 Product quality and flow rate

Products	Quality (°C, ASTM D86)		Flow rate (m^3 h^{-1})
	T5%	T95%	
LN	25.9	110.6	103.5
HN	138.9	186.6	78.2
LD	215.9	301.5	140.3
HD	310.7	354.4	48.1
RES	361.4	754.3	292.5

Table S5 Utility demand and column cost

Variable	Values	Units
Utility requirements		
Hot utility	54.61	MW
Cold utility	61.18	MW
Cost analysis		
Utility cost	8.51	\$MM y^{-1}
Steam cost	1.77	\$MM y^{-1}
Total operating cost	10.28	\$MM y^{-1}
Annualised capital cost	0.33	\$MM y^{-1}
Total annualised cost	10.61	\$MM y^{-1}

\$MM y^{-1} denotes millions of US dollars per annum

Table S6 Utility costs (Chen, 2008)

Utility	Price	Unit
Stripping steam (260 °C, 4.5 bar)	0.14	\$ $kmol^{-1}$
Fired heating (1500-800 °C)	150	\$ $(kWy)^{-1}$
Cooling water (10-40 °C)	5.25	\$ $(kWy)^{-1}$

The solution time for the multiple run genetic algorithm ranges between 4 and 6 hours for Case 1; and between 3 and 6 hours for Case 2. The computation results are summarised in Tables S7 and S8.

Table S7 Results for CDU design based on genetic algorithm (Case 1)

Run	Objective function ($MM y^{-1})*	CPU time (hours)	No. of Generations	Convergence criterion#	Number of simulations
1	7.84	6.49	199	1	20001
2	8.10	6.57	200	0	20101
3	7.85	5.33	173	1	17401
4	7.84	4.58	200	0	20101
5	7.84	5.89	200	0	20101
6	7.84	5.32	200	0	20101
7	8.15	5.76	175	1	17601
8	7.84	5.51	188	1	18901
9	**7.84**	**4.06**	**171**	**1**	**17201**
10	7.85	4.94	162	1	16301

* $MM y^{-1} denotes millions of US dollars per annum

1 denotes population convergence; 0 denotes maximum no of generation reached

Table S8 Results for CDU design based on genetic algorithm

Run	Objective function ($MM y^{-1})*	CPU time (hours)	No. of Generations	Convergence criterion#	Number of simulations
1	8.49	4.57	200	0	20101
2	8.53	2.91	200	0	20101
3	8.52	5.64	200	0	20101
4	8.51	3.46	188	1	19001
5	8.52	3.96	200	0	20101
6	**8.47**	**3.00**	**200**	**0**	**20101**
7	8.50	3.93	153	1	15401
8	8.49	5.46	200	0	20101
9	8.52	4.76	200	0	20101
10	8.47	3.12	200	0	20101

* $MM y^{-1} denotes millions of US dollars per annum

1 denotes population convergence; 0 denotes maximum no of generation reached

Figures S1 and S2 shows how the population evolves using genetic algorithm to the best solution; Run 9 and Run 6 representing the best solutions for Case 1 and Case 2 respectively. As can be seen, at the start of the algorithm, the mean fitness of the individuals in the population is greater than the fitness of the best individual. This results from the fact that the initial randomly generated population of individuals is diverse. As the algorithm progresses, the fitness of all members of the subsequent generations tends to converge to the best solution. As a result, the mean fitness of the population and the fitness of the best individual become similar. The algorithm terminates on generation 171 in Case 1, when the population converge; i.e. there is no significant improvement of the objective value for a given number of generations. In Case 2, the algorithm terminates at the maximum generation.

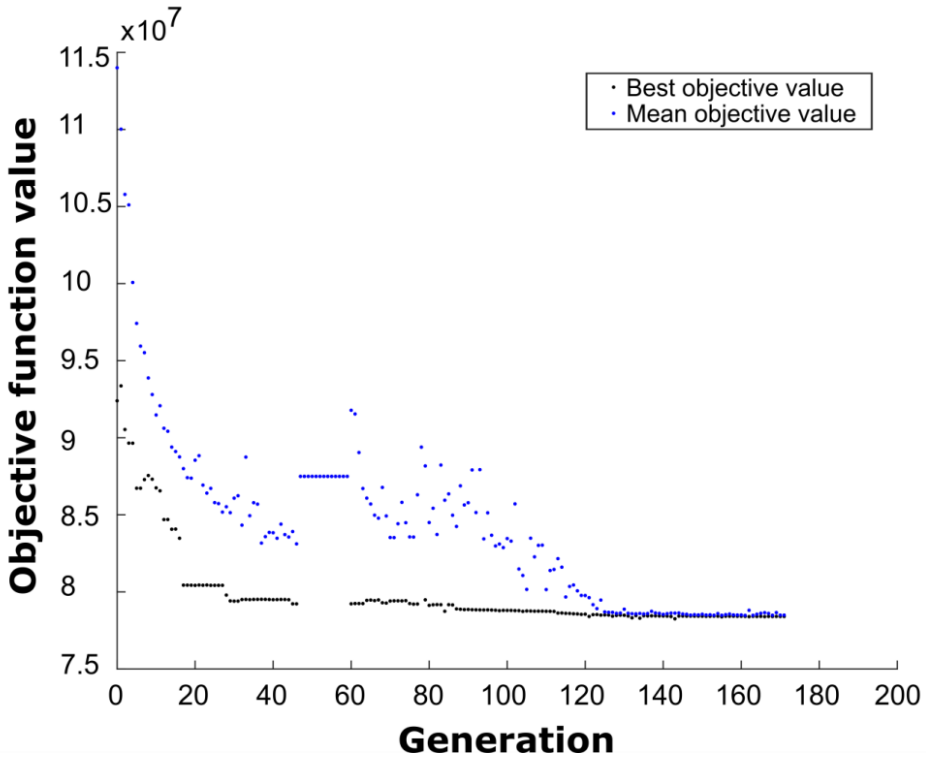


Figure S1 Optimisation progress (Case 1)

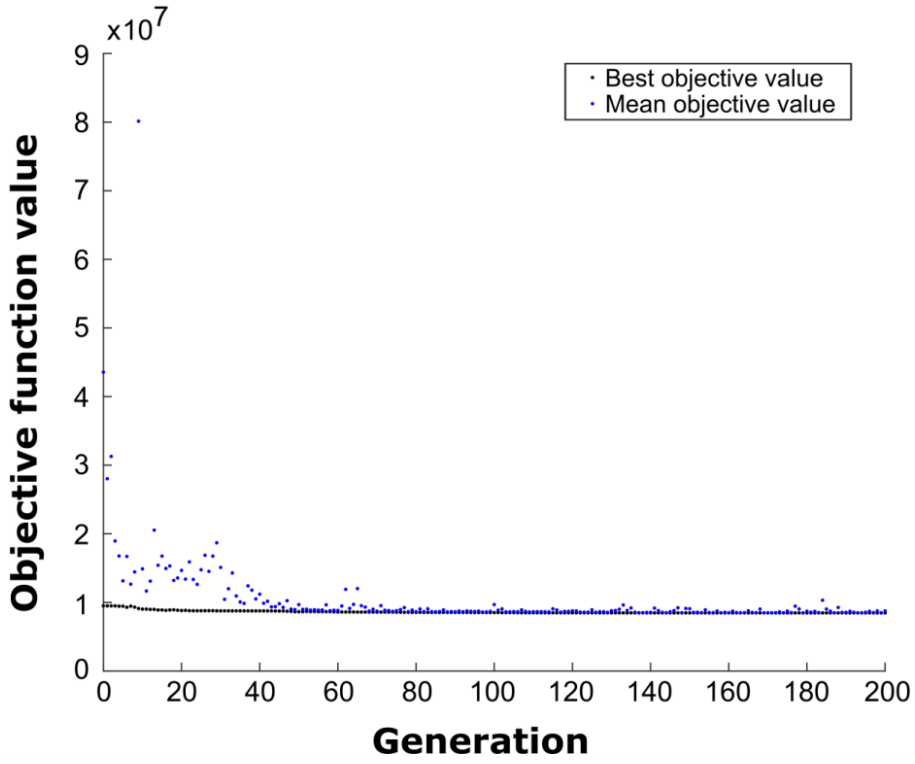


Figure S2 Optimisation progress (Case 2)

Tables S9 and S10 present the process stream information for the optimal crude oil distillation unit. This information was used in the heat recovery model to calculate the minimum energy targets for the design.

Table S9 Process stream information (Case 1)

Stream Name	T_{supply} (°C)	T_{target} (°C)	ΔH (MW)
ADU condenser	93	58	43.15
LN cooler	58	40	0.74
HN cooler	184	40	6.14
LD cooler	285	40	17.81
HD cooler	266	50	5.54
RES cooler	322	100	44.80
PA1	157	128	8.41
PA2	229	174	13.44
PA3	300	266	10.25
HN reboiler	178	184	2.70
LD reboiler	274	285	8.16
Raw crude oil	25	340	138.11

Table S10 Process stream information (Case 2)

Stream Name	T_{supply} (°C)	T_{target} (°C)	ΔH (MW)
ADU condenser	93	58	45.16
LN cooler	58	40	0.76
HN cooler	185	40	5.76
LD cooler	281	40	18.09
HD cooler	275	50	7.04
RES cooler	338	100	47.08
PA1	152	131	9.58
PA2	228	185	16.19
PA3	305	284	15.20
HN reboiler	178	185	5.15
LD reboiler	268	281	8.81
Raw crude oil	25	361	149.35

References

Chen, L., 2008. Heat-Integrated Crude Oil Distillation System Design. PhD Thesis, University of Manchester, UK.

A.2 Supporting Information for Publication 2

Ibrahim, D., Jobson, M., Lie J., Guillén-Gosálbez, G., 2018. Optimization-based Design of Crude Oil Distillation Units using Surrogate Column Models and a Support Vector Machine Chem. Eng. Res. Des., **2018**, **DOI:** doi.org/10.1016/j.cherd.2018.03.006.

Supplementary material

S1 Initial feasible design information

Table S1 shows the crude oil assay (Watkins, 1979) that was used in the case study in Section 4.

Table S1 – Physical properties of crude oil (crude oil assay)

Name (Country of origin)	Tia Juana light (Venezuela)	
Type	Light	
Bulk properties		
i. Density (kg/m^3)	867.6	
ii. API gravity (^{o}API)	31.6	
Distillation properties		
i. Light end analysis	Comp. name	Vol. %
	Ethane	0.04
	Propane	0.37
	i-Butane	0.27
	n-Butane	0.89
	i-Pentane	0.77
	n-Pentane	1.13
ii. TBP curve	Temp. (^{0}C)	Vol. %
	36.1	0
	64.4	5
	100.6	10
	163.9	20
	221.1	30
	278.9	40
	337.2	50
	397.2	60
	463.9	70
	545.0	80

The crude oil assay in Table S1 is cut into 25 pseudo-components using the standard oil characterization procedure in Aspen HYSYS v8.6. Table S2 presents the normal boiling temperature, compositions and volumetric flow rates of both the pure components and the pseudo-components.

Table S2 – Crude oil feedstock characterization

Comp. name	NBP (°C)	Volume fraction	Volumetric flow ($m^3\ h^{-1}$)
Ethane	-89	0.040	0.265
Propane	-42	0.370	2.451
i-Butane	-12	0.270	1.789
n-Butane	-1	0.890	5.896
i-Pentane	28	0.770	5.101
n-Pentane	36	1.130	7.486
NBP_47	47	4.250	28.157
NBP_72	72	3.371	22.331
NBP_97	97	3.263	21.616
NBP_122	122	3.654	24.207
NBP_146	146	3.848	25.491
NBP_171	171	4.028	26.683
NBP_195	195	4.148	27.479
NBP_219	219	4.147	27.473
NBP_244	244	4.083	27.048
NBP_268	268	4.071	26.970
NBP_293	293	4.063	26.916
NBP_317	317	4.031	26.705
NBP_341	341	4.018	26.615
NBP_366	366	3.947	26.147
NBP_390	390	3.822	25.319
NBP_415	415	3.673	24.332
NBP_449	449	6.153	40.762
NBP_493	493	5.466	36.208
NBP_538	538	4.875	32.294
NBP_581	581	4.264	28.246
NBP_625	625	3.214	21.288
NBP_685	685	4.781	31.674
NBP_771	771	2.809	18.609
NBP_858	858	1.408	9.328
NBP_950	950	1.142	7.563

Tables S3 to S5 present the initial design data for the base case.

Table S3 – Initial operating conditions

Operating variable	Initial value
PA 1 duty (MW)	11.2
PA 2 duty (MW)	17.89
PA 3 duty (MW)	12.84
PA 1 DT (ºC)	20
PA 2 DT (ºC)	50
PA 3 DT (ºC)	30
Main stripping steam (kmol h^{-1})	1200
HD stripping steam (kmol h^{-1})	250
Feed temperature (ºC)	365
Reflux ratio	4.17

Table S4 – Product quality and flow rate

Products	Quality (°C, ASTM D86)		Flow rate (m^3 h^{-1})
	T5%	T95%	
LN	25.9	110.6	103.5
HN	138.9	186.6	78.2
LD	215.9	301.5	140.3
HD	310.7	354.4	48.1
RES	361.4	754.3	292.5

Table S5 – Utility demand and column cost

Variable	Values	Units
Utility requirements		
Hot utility	54.61	MW
Cold utility	61.18	MW
Cost analysis		
Utility cost	8.51	\$MM a^{-1}
Steam cost	1.77	\$MM a^{-1}
Total operating cost	10.28	\$MM a^{-1}
Annualised capital cost	0.33	\$MM a^{-1}
Total annualised cost	10.61	\$MM a^{-1}

\$MM a^{-1} denotes millions of US dollars per annum

S2 Capital cost models

The column shell costs (S_C) and tray costs (T_C) are estimated using the correlations proposed by Guthrie (Guthrie, 1969).

$$S_C = \left(\frac{M\&S\ Index_{2011}}{280}\right) 101.9(D)^{1.066}(H)^{0.802}(2.18 + F_{c1}) \quad \text{(S1)}$$

where M&S Index$_{2011}$ is the Marshall and Swift chemical equipment cost index for year 2011 (4th quarter)("Chemical Engineering magazine," 2012) allowing costs to be updated from 1969 (when the M&S Index was 280); the cost is updated to current equipment cost using Eq. (S3); D is the sectional diameter of the column, H is the sectional height, which depends on tray spacing and F_{c1} is the column cost factor, which depends on the column material of construction and column operating pressure.

$$T_C = \left(\frac{M\&S\ Index_{2011}}{280}\right) 4.7(D)^{1.55} HF_{c2} \quad \text{(S2)}$$

The tray cost factor F_{c2} depends on the type of tray, tray spacing and material of construction.

$$\left(\frac{CEPCI_{2014}}{CEPCI_{2011}}\right) \times Cost_{2011} \quad \text{(S3)}$$

where CEPCI$_{2011}$ and CEPCI$_{2014}$ are the chemical engineering plant cost index for year 2011 (4th quarter) ("Chemical Engineering magazine," 2012) and 2014 (4th quarter) ("Chemical Engineering magazine," 2014) respectively; Cost$_{2011}$ is the equipment cost for year 2011, calculated using Eq. (S1) and Eq. (S2). The M&S Index$_{2011}$, CEPCI$_{2011}$ and CEPCI$_{2014}$ are 1536.5, 590.1 and 575.7 respectively.

S3 Optimization results

Table S6 Results for multiple optimization runs

Run	Objective function ($MM a^{-1})*	CPU time (s)	No. of Generations
1	8.54	85	300
2	8.49	90	300
3	**8.44**	**83**	**300**
4	8.48	68	240
5	8.46	84	300
6	8.47	82	300
7	8.44	75	270
8	8.46	87	300
9	8.49	80	300
10	8.48	79	300

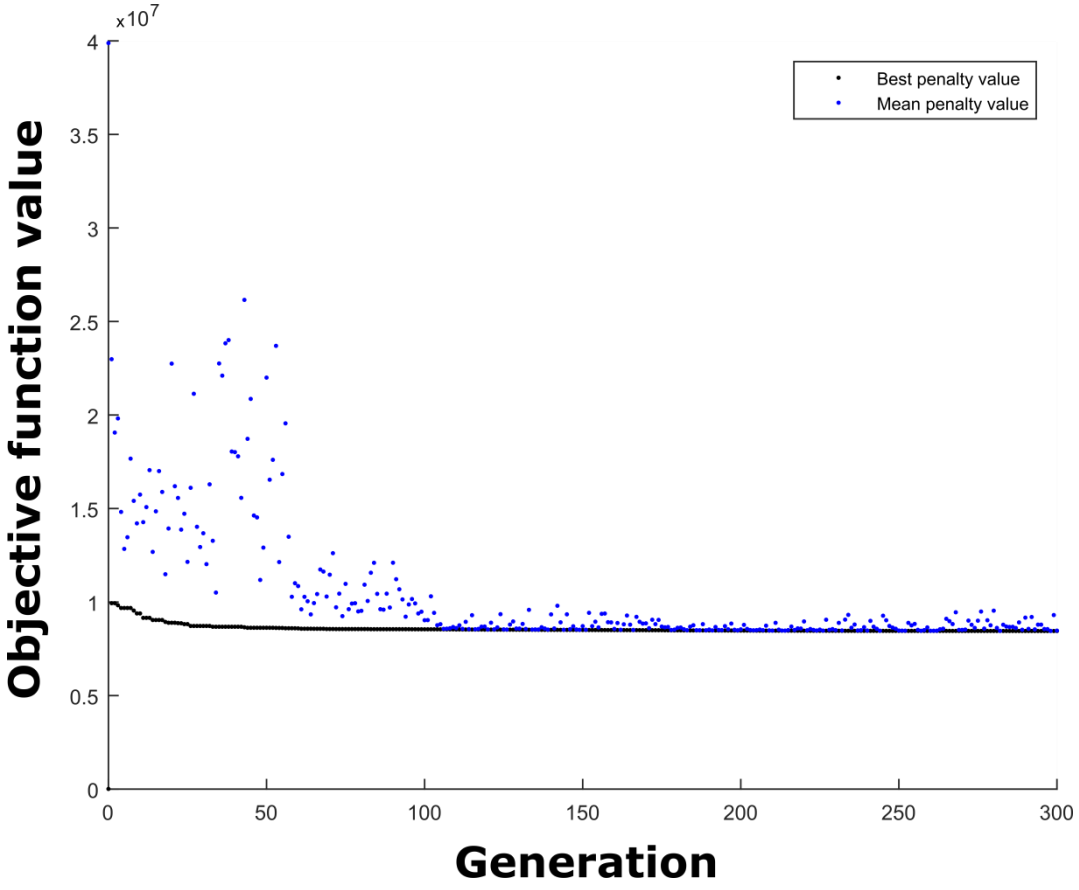


Figure S1 Optimisation progress for Run 3

S4 Simulation of ANN results on rigorous model

Table S7 Product quality, supply and target temperature

Variables	Tia Juana light			Variables	Tia Juana light		
Product quality (℃)	Rigorous model	ANN model	Diff.	**Product flow rates ($m^3\ h^{-1}$)**	Rigorous model	ANN model	Diff.
LN T5%	24.51	24.68	-0.17	LN	99.39	98.87	0.52
HN T5%	131.09	128.58	2.51	HN	85.47	86.25	-0.78
LD T5%	215.71	215.63	0.07	LD	134.71	134.28	0.43
HD T5%	306.87	307.47	-0.60	HD	49.76	50.71	-0.95
RES T5%	360.38	360.61	-0.24	RES	293.27	292.52	0.76
LN T95%	110.56	110.57	-0.01	**Exchanger duties (MW)**			
HN T95%	186.79	186.56	0.23	ADU condenser	44.58	44.57	0.00
LD T95%	301.47	301.48	0.00	LN cooler	0.75	0.76	-0.01
HD T95%	354.38	354.38	0.00	HN cooler	6.11	6.12	-0.01
RES T95%	754.14	754.22	-0.07	LD cooler	18.43	18.50	-0.07
Supply temperature (℃)				HD cooler	6.52	6.48	0.04
ADU condenser	93.13	91.64	1.49	RES cooler	47.12	47.10	0.02
LN cooler	58.21	57.81	0.41	HN reboiler	3.41	3.37	0.05
HN cooler	185.15	183.87	1.29	LD reboiler	8.71	8.68	0.02
LD cooler	284.65	284.65	0.01	Fired heater	49.62	49.62	0.00
HD cooler	276.92	277.16	-0.25	**Column diameter (m^2)**			
RES cooler	337.46	336.83	0.62	Main column			
PA1	156.49	156.40	0.09	Section 1	6.10	6.30	-0.21
PA2	232.32	232.94	-0.62	Section 2	5.94	6.06	-0.11
PA3	309.32	309.56	-0.24	Section 3	6.40	6.43	-0.03
HN reboiler	178.77	177.56	1.21	Section 4	5.49	5.56	-0.07
LD reboiler	272.38	272.98	-0.59	Section 5	5.03	5.04	-0.01
Target temperature (℃)				HN side-stripper	2.59	2.36	0.23
ADU condenser	58.21	58.04	0.17	LD side-stripper	2.59	2.59	0.00
HN reboiler	185.15	184.11	1.05	HD side-stripper	1.37	1.48	-0.11
LD reboiler	284.65	285.16	-0.50				

Table S8 Validation of artificial neural network results on rigorous model in Aspen HYSYS

Variable	Rigorous model	ANN model	Diff.	Units
Utility requirements				
Hot utility	44.46	44.53	-0.07	MW
Cold utility	46.04	44.87	1.17	MW
Cost				
Utility cost	6.91	6.92	-0.01	\$MM a^{-1}
Steam cost	1.33	1.33	0.00	\$MM a^{-1}
Total operating cost	8.24	8.24	-0.01	\$MM a^{-1}
Annualized capital cost	0.21	0.21	0.00	\$MM a^{-1}
Total annualized cost	8.44	8.45	-0.01	\$MM a^{-1}

Table S9 presents the process stream information for the optimal crude oil distillation unit. This information was used in the heat recovery model to calculate the minimum energy targets for the design.

Table S9 Process stream information

Stream Name	**T_{supply} (°C)**	**T_{target} (°C)**	**ΔH (MW)**
ADU condenser	93	58	44.58
LN cooler	58	40	0.75
HN cooler	185	40	6.11
LD cooler	285	40	18.43
HD cooler	277	50	6.52
RES cooler	337	100	47.12
PA1	156	136	11.05
PA2	232	176	14.01
PA3	309	281	14.35
HN reboiler	179	185	3.41
LD reboiler	272	285	8.71
Raw crude oil	25	361	149.24

Table S10 Validation results for artificial neural network classifier.

Prediction class	True prediction	False prediction
Positive class	952	56
(Converged samples, +1)	[94.4%]	[5.6%]
Negative class	400	342
(Unconverged samples, −1)	[53.9%]	[446.1%]
Overall: Correct prediction	77%	
Wrong prediction	23%	

Table S11 Optimisation with and without support vector machine

	With SVM				Without SVM		
Run	Objective function ($MM y–1)*	CPU time (s)	No. of Generations		Objective function ($MM y–1)*	CPU time (s)	No. of Generations
1	8.54	85	300		fail to converge on rigorous model		
2	8.49	90	300		8.64	55	300
3	8.44	83	300		8.59	55	300
4	8.48	68	240		fail to converge on rigorous model		
5	8.46	84	300		8.60	54	300
6	8.47	82	300		8.64	56	300
7	8.44	75	270		8.63	55	300
8	8.46	87	300		8.61	55	300
9	8.49	80	300		8.58	55	300
10	8.48	79	300		fail to converge on rigorous model		

All results presented in Table 11 have been validated on a rigorous simulation model.

S5 References

Chemical Engineering magazine, 2014. www.che.com (accessed 11.20.14).

Chemical Engineering magazine, 2012. www.che.com (accessed 11.20.14).

Guthrie, K.M., 1969. Data and Techniques for Preliminary Capital Cost Estimating. Chem. Eng. 76, 114.

Watkins, R.N., 1979. Petroleum Refinery Distillation. Gulf Publishing Company, Book Division.

A.3 Supporting Information for Publication 3

Ibrahim, D., Jobson, M., Lie J., Guillén-Gosálbez, G., 2017. Optimal Design of Flexible Heat-Integrated Crude Oil Distillation Units Chem. Eng. Res. Des. [To be submitted]

Supplementary material

S1 Crude oil assay

Table S1 Physical properties of three variety of crude oil used in this work (Watkins, 1979; TOTAL, 2015)

Name	Tia Juana light*		Bonny light**		Brent**			
(Country of origin)	(Venezuela)		(Nigeria)		(United Kingdom)			
Type	Light		Light		Light			
Bulk properties								
Density (kg m^{-3})	867.6		849		833.7			
API gravity (^{0}API)	31.6		35.1		38.1			
Distillation properties								
Light end analysis	Comp. name	Vol. (%)	Comp. name	Vol. (%)	Comp. name	Vol. (%)		
	Ethane	0.04	Ethane	0.05	Ethane	0.23		
	Propane	0.37	Propane	0.35	Propane	1.48		
	i-Butane	0.27	i-Butane	0.42	i-Butane	0.44		
	n-Butane	0.89	n-Butane	0.89	n-Butane	2.00		
	i-Pentane	0.77	i-Pentane	-	i-Pentane	-		
	n-Pentane	1.13	n-Pentane	-	n-Pentane	-		
TBP curve	Temp. (^{0}C)	Vol. %	Temp. (^{0}C)	Vol. %	Temp (^{0}C)	Vol. %	Temp (^{0}C)	Vol. %
	36.1	0	-0.50	0	-0.50	0	330	59.76
	64.4	5	015	1.37	080	11.95	340	61.53
	100.6	10	080	9.26	140	25.65	350	63.26
	163.9	20	150	23.80	150	27.80	360	64.94
	221.1	30	175	28.67	160	29.79	370	66.56
	278.9	40	230	39.65	180	33.43	380	68.12
	337.2	50	325	64.37	200	36.77	390	69.61
	397.2	60	350	69.63	220	40.03	400	71.03
	463.9	70	375	74.30	240	43.39	450	77.39
	545.0	80	550	93.68	250	45.14	475	80.26
			565	95.39	260	46.93	500	83.02
					300	54.28	525	85.78
					310	56.13	550	88.61
					320	57.96	565	90.36

The assay for each crude oil in Table S1 is cut into 25 pseudo-components using the standard oil characterization procedure in Aspen HYSYS v8.6. In addition to the

pseudo-components, there are six light ends in the case of Tia Juana and four light ends in the case of Bonny light and Brent (see Table S1).

S2 Initial feasible design information

Table S2 Initial crude oil distillation unit design parameters (Ibrahim et al., 2017a)

Parameter	Tia Juana light*	Bonny light	Brent
Number of trays in main column	41	41	42
Number of trays in side strippers – HN, LD, HD	6, 7, 5	7, 7, 6	5, 5, 6
HN side-stripper draw and return trays	32, 33	31, 32	32, 33
LD side-stripper draw and return trays	24, 25	22, 23	23, 24
HD side-stripper draw and return trays	14, 15	12, 13	13, 14
Pump-around 1 (PA 1) draw and return trays	32, 34	31, 33	32, 34
Pump-around 2 (PA 2) draw and return trays	24, 26	22, 24	23, 25
Pump-around 3 (PA 3) draw and return trays	14, 16	12, 14	13, 15
Stripping steam feed tray	1	1	1
Crude oil feed tray	5	3	4

Table S3 Initial crude oil distillation unit operating conditions

Operating condition	Tia Juana light	Bonny light	Brent
PA 1 duty (MW)	11.20	12.00	10.00
PA 2 duty (MW)	17.89	14.58	26.31
PA 3 duty (MW)	12.84	14.00	25.00
PA 1 DT (°C)	20	60	60
PA 2 DT (°C)	50	25	65
PA 3 DT (°C)	30	20	35
Main steam (kmol h^{-1})	1200	1200	1000
HD steam (kmol h^{-1})	250	250	250
Feed temperature (°C)	365	365	365
Reflux ratio	4.17	4.17	3

Table S4 Product type, quality and flow rate

Products	Quality (^{0}C) [ASTM D86]		Flow rates (m^3 h^{-1})		
	T5%	T95%	Tia Juana light	Bonny light	Brent
LN	25.99*	110.56	103.5	116.3	161.1
HN	138.89	186.57	78.2	103.5	91.1
LD	215.96	301.47	140.3	190.5	135.5
HD	310.66	354.38	48.1	70.3	51.6
RES	361.40	754.31*	292.5	181.9	223.5

* The T5% boiling temperature for Bonny light and Brent are 25.36℃ and 4.07℃, while the T95% boiling temperature are 617.90℃ and 652.50℃ respectively.

S3 Capital cost models and price of feed, products and utilities

S3.1 Capital cost correlations

The column shell costs (S_C) and tray costs (T_C) are estimated using the correlations proposed by Guthrie (1969):

$$S_C = \left(\frac{M\&S\ Index_{2011}}{280}\right) 101.9(D)^{1.066}(H)^{0.802}(2.18 + F_{c1}) \quad \text{(S1)}$$

where M&S Index$_{2011}$ is the Marshall and Swift chemical equipment cost index for year 2011 (4th quarter)(Chemical Engineering, 2012) allowing costs to be updated from 1969 (when the M&S Index was 280); the cost is updated to current equipment cost using Eq. (S3); D is the sectional diameter of the column, H is the sectional height, which depends on tray spacing and F_{c1} is the column cost factor, which depends on the column material of construction and column operating pressure. In this case study, the column is design using sieve trays, and the diameter of each column section is estimated using the tray sizing utility in Aspen HYSYS. The eight sections of the column are cost according to their respective diameter and number of trays.

The column factor F_{c1} is the sum of material factor F_{m1} (value is 1 for sieve trays) and pressure factor F_p (value is 1 for sieve trays). Similarly, F_{c2} is the correction factor for sieve trays, and is the sum of tray spacing factor, type factor and material factor, with values of 1, 0 and 0 respectively.

$$T_C = \left(\frac{M\&S\ Index_{2011}}{280}\right) 4.7(D)^{1.55} HF_{c2} \quad \text{(S2)}$$

The tray cost factor F_{c2} depends on the type of tray, tray spacing and material of construction.

The cost is further updated using CEPCI values for 2011 and 2014:

$$\left(\frac{CEPCI_{2014}}{CEPCI_{2011}}\right) \times Cost_{2011} \quad \text{(S3)}$$

where $CEPCI_{2011}$ and $CEPCI_{2014}$ are the Chemical Engineering Plant Cost Indices for 2011 (4th quarter) (Chemical Engineering, 2012) and 2014 (4th quarter) (Chemical Engineering, 2014) respectively; $Cost_{2011}$ is the equipment cost for year 2011, calculated using Eq. (S1) and Eq. (S2). The values used for $M\&S\ Index_{2011}$, $CEPCI_{2011}$ and $CEPCI_{2014}$ are 1536.5, 590.1 and 575.7, respectively.

S3.2 Price of commodities

Table S5a presents the prices of intermediate products and the spot price of crude oil (EIA, 2016). The price of the intermediate products are calculated using the method presented in (Chen, 2008)

Table S5a Price of products, and cost of crude oils and utilities

Commodity	**Price** ($ m^{-3})	**Price** ($ bbl^{-1})
Tia Juana light	288.32	45.84
Bonny light	290.27	46.15
Brent	263.79	41.94
Light naphtha (LN)	386.34	61.42
Heavy naphtha (HN)	358.95	57.07
Light distillate (LD)	280.91	44.66
Heavy distillate (HD)	369.84	58.80
Residue (RES)	218.06	34.67

Table S5b presents the cost of utilities (Chen, 2008).

Table S5b Cost of utilities

Utility	Price	Unit
Stripping steam (260 ºC, 4.5 bar)	0.14	\$ $kmol^{-1}$
Fired heating (1500-800 ºC)	150	\$ kW^{-1} a^{-1}
Cooling water (10-40℃)	5.25	\$ kW^{-1} a^{-1}

S4 Bounds for data sampling and optimization

Table S6 Initial distribution of trays in column sections and their bounds

Section	Tia Juana	Bonny light	Brent	Lower bound	Upper bound
1	5	3	4	3	6
2	9	9	9	7	10
3	10	10	10	8	11
4	8	9	9	6	10
5	9	10	10	7	11
6	5	6	6	3	7
7	7	7	5	3	8
8	6	7	5	3	8

Table S7 Initial distillation operating conditions and their bounds (Tia Juana)

Operating condition	Lower bound	Initial feasible design	Upper bound
PA 1 duty (MW)	8.4	11.2	14
PA 2 duty (MW)	13.42	17.89	22.36
PA 3 duty (MW)	9.63	12.84	16.05
PA 1 DT (℃)	20	20	30
PA 2 DT (℃)	40	50	60
PA 3 DT (℃)	20	30	40
Main steam (kmol h^{-1})	900	1200	1500
HD steam (kmol h^{-1})	188	250	313
Feed temperature (℃)	340	365	375
Reflux ratio	3.17	4.17	6.17

Table S8 Initial distillation operating conditions and their bounds (Bonny light)

Operating condition	Lower bound	Initial feasible design	Upper bound
PA 1 duty (MW)	9	12	15
PA 2 duty (MW)	11.58	14.58	17.58
PA 3 duty (MW)	11	14	17
PA 1 DT (℃)	50	60	70
PA 2 DT (℃)	15	25	35
PA 3 DT (℃)	15	20	30
Main steam (kmol h^{-1})	900	1200	1500
HD steam (kmol h^{-1})	118	250	313
Feed temperature (℃)	345	365	375
Reflux ratio	4.17	4.18	4.35

Table S9 Initial distillation operating conditions and their bounds (Brent)

Operating condition	Lower bound	Initial feasible design	Upper bound
PA 1 duty (MW)	7	10	13
PA 2 duty (MW)	23.31	26.31	29.31
PA 3 duty (MW)	22	25	28
PA 1 DT (℃)	50	60	70
PA 2 DT (℃)	55	65	75
PA 3 DT (℃)	25	35	45
Main steam (kmol h^{-1})	700	1000	1300
HD steam (kmol h^{-1})	118	250	313
Feed temperature (℃)	345	365	375
Reflux ratio	2.97	3	4

S5 Surrogate modelling results: performance of artificial neural network model

S5.1 Tia Juana

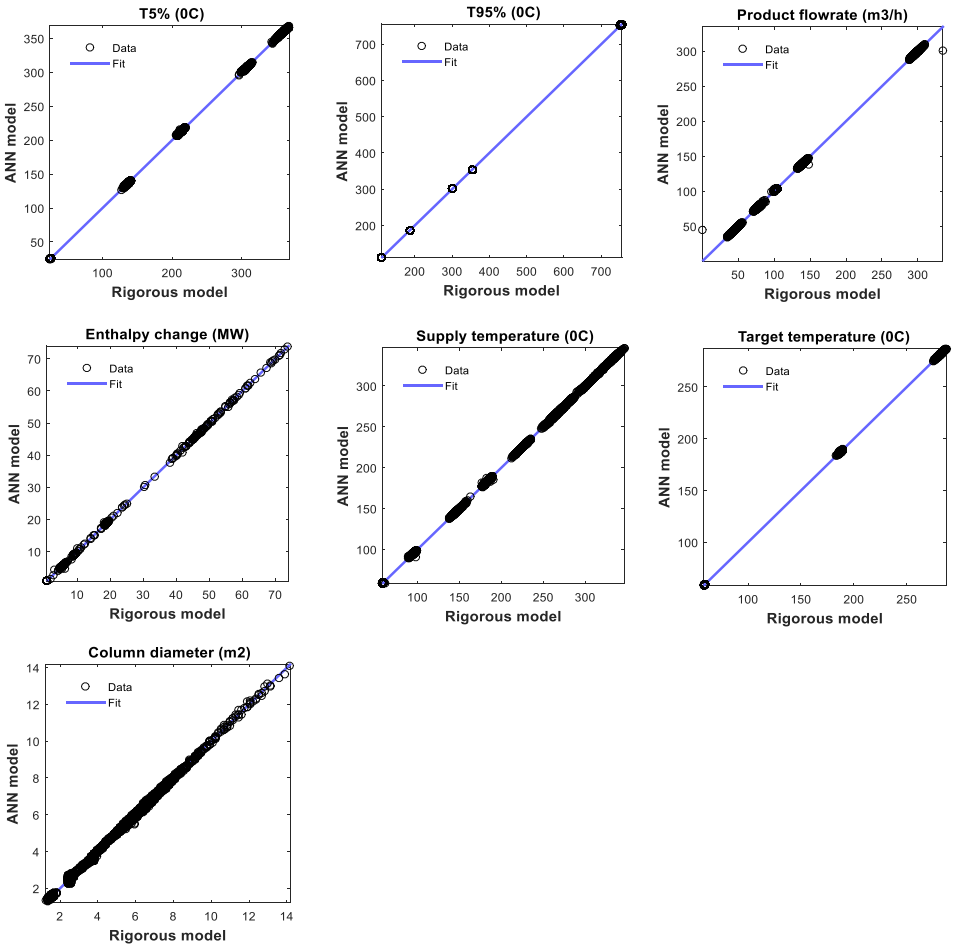


Table S1 Parity plots comparing predictions of ANN and rigorous model: Tia Juana

Note that the figures are presented in a different order to the models, as described in Table S5.1:

Table S5.1 Description of artificial neural network models

ANN number	Description	Outputs	R^2
1	Product quality (T5 %)	5	0.999
2	Product quality (T95 %)	5	0.999
3	Product flow rate	5	0.999
4	Column diameter	8	0.999
5	Enthalpy change	9	0.999
6	Supply temperature	11	0.999
7	Target temperature	3	0.999

S5.2 Bonny light

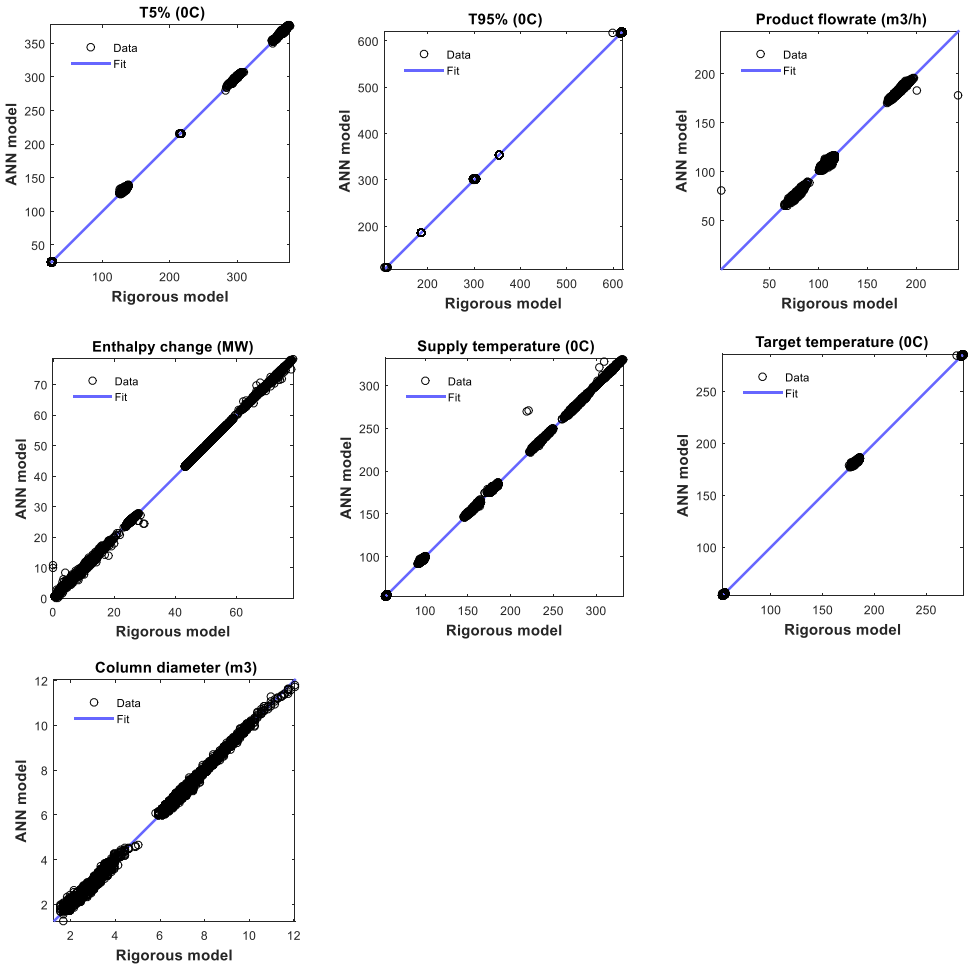


Table S2 Parity plots comparing predictions of ANN and rigorous model: Bonny light

Table S5.2 Description of artificial neural network models

ANN number	Description	Outputs	R^2
1	Product quality (T5 %)	5	0.999
2	Product quality (T95 %)	5	0.999
3	Product flow rate	5	0.999
4	Column diameter	8	0.999
5	Enthalpy change	9	0.999
6	Supply temperature	11	0.999
7	Target temperature	3	0.999

S5.3 Brent

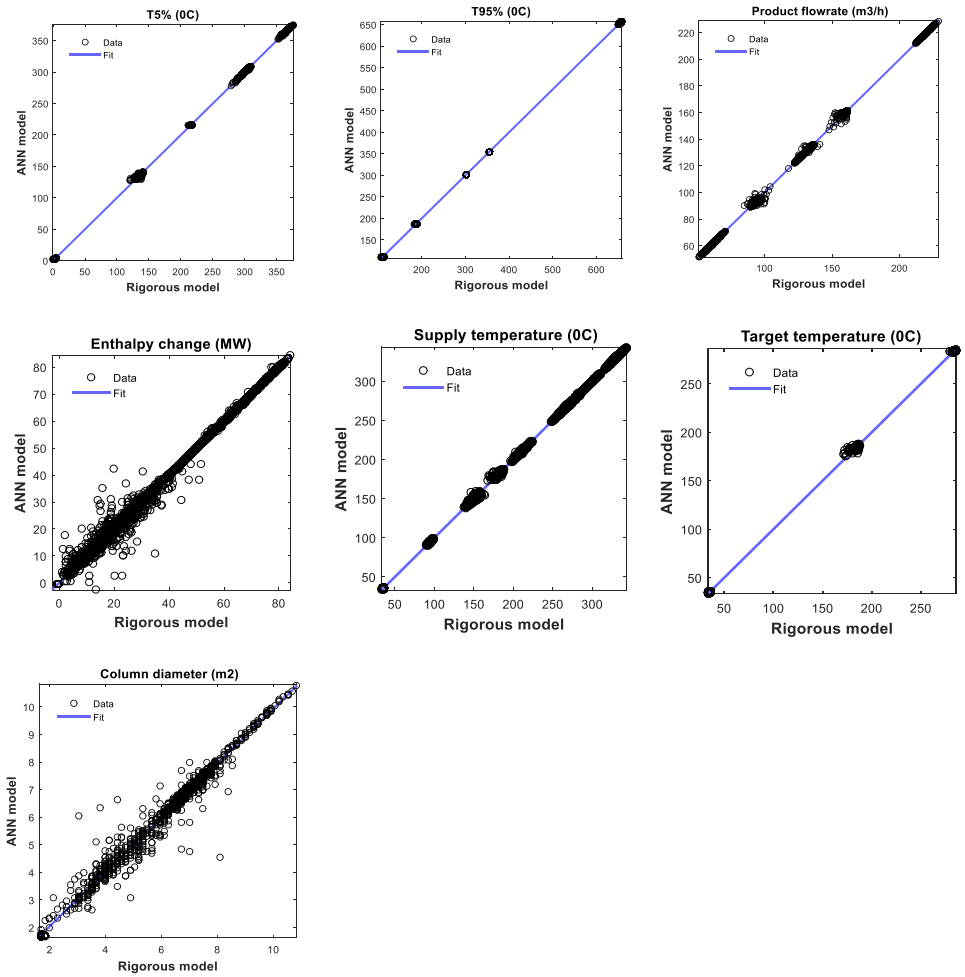


Table S3 Parity plots comparing predictions of ANN and rigorous model: Brent

Table S5.3 Description of artificial neural network models

ANN number	Description	Outputs	R^2
1	Product quality (T5 %)	5	0.999
2	Product quality (T95 %)	5	0.999
3	Product flow rate	5	0.999
4	Column diameter	8	0.992
5	Enthalpy change	9	0.998
6	Supply temperature	11	0.999
7	Target temperature	3	0.999

S6 Optimization results

S6.1 Optimization results for multiple runs

Table S10 Optimization results for flexible crude oil distillation unit

Run	Objective function ($MM y^{-1})*	CPU time (seconds)
1	100.91	3426
2	101.16	3940
3	101.27	3976
4	101.27	4081
5	101.27	3998
6	100.91	3817
7	**101.27**	**3686**
8	101.27	3754
9	101.27	3566
10	101.27	4480

*$MM denotes millions of dollars

Figures S4 shows how the initial population evolves to the best solution using genetic algorithm: Run 7.

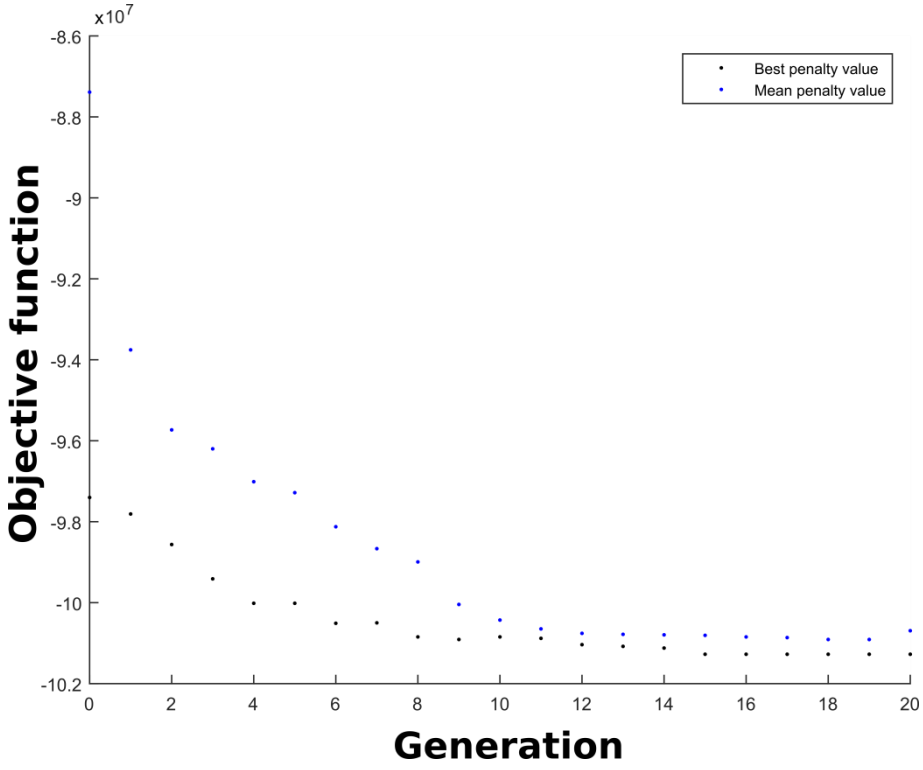


Figure S4 Optimization progress

S6.2 Validation: Rigorous simulation of optimized designs

Table S11 Product quality, supply, and target temperature

Variables	Tia Juana light			Bonny light			Brent		
Product quality (℃)	Rigorous model	ANN model	Diff.	Rigorous model	ANN model	Diff.	Rigorous model	ANN model	Diff.
LN T5%	25.92	25.49	0.43	25.00	25.08	-0.07	3.90	3.90	0.01
HN T5%	137.82	135.94	1.88	133.07	132.75	0.32	136.07	136.47	-0.40
LD T5%	215.97	217.87	-1.90	215.93	215.91	0.02	214.48	215.87	-1.40
HD T5%	308.28	308.18	0.10	299.93	299.43	0.51	308.14	308.17	-0.03
RES T5%	369.28	367.80	1.49	377.91	377.37	0.54	375.75	373.85	1.89
LN T95%	110.56	110.56	0.00	110.56	110.58	-0.02	110.02	110.91	-0.89
HN T95%	186.72	186.55	0.18	186.57	186.57	0.00	186.57	186.16	0.41
LD T95%	301.47	301.47	0.00	301.47	301.44	0.03	301.47	301.47	0.00
HD T95%	354.38	354.39	0.00	354.38	354.39	0.00	354.38	354.38	0.00
RES T95%	755.68	755.61	0.06	620.64	620.22	0.43	658.26	657.82	0.44
Supply temperature (℃)									
ADU condenser	92.43	91.92	0.51	96.09	96.34	-0.25	93.74	93.71	0.03
LN cooler	59.27	59.14	0.13	55.47	55.46	0.02	35.68	35.64	0.04
HN cooler	188.38	188.13	0.25	184.28	183.88	0.40	186.40	186.10	0.30
LD cooler	283.58	285.20	-1.62	284.70	284.71	-0.01	285.06	285.16	-0.10
HD cooler	280.97	282.84	-1.87	295.48	293.61	1.87	275.83	276.36	-0.53
RES cooler	336.25	338.31	-2.06	318.04	319.00	-0.96	330.66	331.49	-0.83
PA1	151.82	154.25	-2.43	158.73	158.56	0.17	148.29	146.70	1.59
PA2	232.96	231.42	1.55	244.47	244.62	-0.16	225.14	225.40	-0.25
PA3	311.88	312.41	-0.53	315.38	314.85	0.54	307.54	308.00	-0.47
HN reboiler	181.79	182.17	-0.38	179.27	178.90	0.37	180.68	180.18	0.50
LD reboiler	273.41	274.37	-0.96	276.04	275.96	0.08	273.93	274.17	-0.24
Target temperature (℃)									
ADU condenser	59.27	59.08	0.19	55.47	55.41	0.06	35.68	35.54	0.14
HN reboiler	188.38	188.35	0.03	184.28	183.69	0.59	186.40	185.87	0.53
LD reboiler	283.58	285.50	-1.91	284.70	284.75	-0.05	285.06	285.03	0.03

Diff: difference between temperature predicted by rigorous model and that of optimum solution obtained using surrogate models

Table S12 Product flow rate, enthalpy change and column diameter

Variables	Tia Juana light			Bonny light			Brent		
Product flow rates ($m^3\ h^{-1}$)	Rigorous model	ANN model	Diff.	Rigorous model	ANN model	Diff.	Rigorous model	ANN model	Diff.
LN	103.32	103.33	-0.01	115.00	115.46	-0.46	160.52	160.40	0.12
HN	80.10	82.17	-2.06	105.46	104.86	0.60	92.17	91.95	0.22
LD	134.53	131.71	2.82	183.60	184.49	-0.88	133.21	133.12	0.10
HD	58.60	60.33	-1.73	87.84	86.82	1.01	64.68	64.85	-0.17
RES	286.00	285.10	0.91	170.64	171.01	-0.37	211.95	212.26	-0.31
Exchanger duties (MW)									
ADU condenser	56.21	55.08	1.13	69.04	69.05	-0.01	71.28	70.78	0.50
LN cooler	0.83	0.76	0.07	0.74	0.74	0.00	0.43	0.42	0.01
HN cooler	5.88	6.13	-0.25	7.54	7.52	0.02	6.66	6.61	0.05
LD cooler	18.31	18.31	0.00	25.23	25.31	-0.08	18.25	18.81	-0.55
HD cooler	7.85	7.77	0.08	12.67	12.55	0.12	8.43	8.01	0.42
RES cooler	45.76	45.57	0.19	24.47	24.56	-0.09	32.59	32.26	0.34
HN reboiler	6.17	3.60	2.57	5.04	5.13	-0.09	12.58	14.45	-1.87
LD reboiler	6.74	8.23	-1.49	6.26	6.17	0.09	8.99	7.66	1.33
Fired heater	57.28	56.86	0.43	58.91	58.90	0.01	54.97	55.04	-0.07
Column diameter (m^2)									
Main column									
Section 1	7.77	7.61	0.16	7.01	6.98	0.03	6.86	6.62	0.24
Section 2	6.71	6.69	0.02	10.67	10.51	0.16	6.86	6.56	0.30
Section 3	7.01	9.19	-2.18	7.62	7.62	0.00	7.62	7.62	0.00
Section 4	8.84	6.33	2.51	10.06	9.99	0.07	9.75	9.59	0.16
Section 5	8.53	5.51	3.02	6.55	6.50	0.05	7.16	7.10	0.07
HN side-stripper	2.44	2.36	0.08	2.29	2.19	0.09	3.51	3.84	-0.33
LD side-stripper	2.59	2.59	0.00	2.59	2.60	-0.01	2.90	2.37	0.52
HD side-stripper	1.68	1.55	0.13	1.83	1.83	0.00	1.68	1.73	-0.06

Diff: difference between variable predicted by rigorous model and that of optimum solution obtained using surrogate models

References

Chemical Engineering magazine, 2014. www.che.com (accessed 11.20.14).

Chemical Engineering magazine, 2012. www.che.com (accessed 11.20.14).

Chen, L., 2008. Heat-Integrated Crude Oil Distillation System Design. PhD Thesis, University of Manchester, UK.

EIA, 2016. PETROLEUM & OTHER LIQUIDS. https://www.eia.gov/dnav/pet/pet_pri_imc3_k_m.htm (accessed 12.1.16).

Guthrie, K.M., 1969. Data and Techniques for Preliminary Capital Cost Estimating. Chem. Eng. 76, 114.

TOTAL, 2015. Crude assay. https://www.totsa.com/pub/crude/crude_assays.php?rub=1 (accessed 11.30.16).

Watkins, R.N., 1979. Petroleum Refinery Distillation. Gulf Publishing Company, Book Division. Texas, US.

A.4 Supporting Information for Publication 4

Ibrahim, D., Jobson, M., Guillén-Gosálbez, G., 2017. Design of Chemical Processes under Uncertainty Combining the Sample Average Approximation and the Analytic Hierarchy Process: Application to the Synthesis of Heat Exchanger Networks. Comput. Chem. Eng. [Submitted]

Supplementary material

S1 Information for HEN design

The modified version of the MINLP heat exchanger network model proposed by Yee and Grossmann (1990) is presented below. The modifications implemented include adding slacks to both stage and overall heat balance equations; slack related penalties are also added to the objective function.

Objective function

The objective of the heat exchanger network design problem is to identify a network with the smallest total annualized cost. Hence, the objective function is formulated as the sum of annual utility costs, annualized cost of exchangers (including fixed charge and area-dependent cost), and penalty terms expressed in terms of costs. The model presented next is the stochastic (multi-scenario) model used to assess the design alternatives. A deterministic model can easily be obtained from the stochastic (multi-scenario) model by considering a single scenario (and removing the penalty terms).

$$\begin{aligned}\min \quad \sum_{s\in S} prob_s \Bigg(& \sum_{s\in S}\sum_{i\in HP} CCU\ qcu_{i,s} + \sum_{s\in S}\sum_{j\in CP} CHU\ qhu_{j,s} + \sum_{i\in HP}\sum_{j\in CP}\sum_{k\in ST} CF_{i,j}\ z_{i,j,k} \\ & + \sum_{i\in HP} CF_{i,CU}\ zcu_i + \sum_{j\in CP} CF_{j,HU}\ zhu_j + \sum_{i\in HP}\sum_{j\in CP}\sum_{k\in ST} C_{ij}\ Aex_{i,j,k} \\ & + \sum_{i\in HP} C_{i,CU}\ Acu_i + \sum_{j\in CP} C_{j,HU}\ Ahu_j + \pi \sum_{s\in S}\sum_{i\in HP} Sh1_{i,s} + Sh2_{i,s} \\ & + \pi \sum_{s\in S}\sum_{j\in CP} SC1_{i,s} + SC2_{i,s} \Bigg)\end{aligned}$$

where the areas ($Aex_{i,j,k}$, Acu_i,Ahu_j) are given by the following equations (Yee and Grossmann, 1990).

$$Aex_{ijk} = \left[q_{i,j,k,s} / \left(U_{i,j,s} \left[(dt_{i,j,k,s} dt_{i,j,k+1,s}) ((dt_{i,j,k,s} + (dt_{i,j,k+1,s})/2 \right]^{1/3} \right) \right]^{\beta_{ij}}$$

$$Acu_i = \left[qcu_{i,s} / \left(U_{i,CU,s}\left[(dtcu_{i,s})(TOUT_{i,s} - TIN_{CU})(dtcu_{i,s} + (TOUT_{i,s} - TIN_{CU}))/2\right]^{1/3}\right)\right]^{\beta_{i,CU}}$$

$$Ahu_j = \left[qhu_{j,s} / \left(U_{j,HU,s}\left[(dthu_j)(TIN_{HU} - TOUT_j)(dthu_{j,s} + (TIN_{HU} - TOUT_j))/2\right]^{1/3}\right)\right]^{\beta_{j,HU}}$$

The objective function presented above includes both continuous and discrete variables. The discrete variables ($z_{i,j,k}$, zcu_i and zhu_j) model the existence (or not) of an exchanger match in each stage of the superstructure, while the continuous variables determine the area and duty of each exchanger match.

Constraints

The objective function presented above is evaluated within a search space that is defined by the process constraints. The constraints include overall heat balance, stage heat balance, initialization of inlet temperature for exchangers the superstructure, temperature feasibility, utility loads, existence of matches and minimum approach temperature. Each constraint is presented below.

(i) Overall heat balance

$$\left(TIN_{i,s} - TOUT_{i,s}\right)F_{i,s} + Sh1_{i,s} - Sh2_{i,s} = \sum_{k \in ST} \sum_{j \in CP} q_{i,j,k,s} + qcu_{i,s} \qquad i \in HP, s \in S$$

$$\left(TOUT_{j,s} - TIN_{j,s}\right)F_{j,s} + SC1_{i,s} - SC2_{i,s} = \sum_{k \in ST} \sum_{i \in HP} q_{i,j,k,s} + qhu_{j,s} \qquad i \in CP, s \in S$$

(ii) Stage heat balance

$$\left(t_{i,k,s} - t_{i,k+1,s}\right)F_{i,s} + Sh1_{i,s} - Sh2_{i,s} = \sum_{j \in CP} q_{i,j,k,s} \qquad k \in ST, i \in HP, s \in S$$

$$(t_{j,k,s} - t_{j,k+1,s})F_{j,s} + SC1_{i,s} - SC2_{i,s} = \sum_{j \in HP} q_{i,j,k,s} \qquad k \in ST, j \in CP, s \in S$$

(iii) Inlet temperature assignment for superstructure

$TIN_{i,s} = t_{i,1,s} \quad TIN_{j,s} = t_{j,NOK+1,s}$

(iv) Temperature feasibility

$t_{i,k,s} \geq t_{i,k+1,s} \qquad k \in ST, i \in HP, s \in S$

$t_{j,k,s} \geq t_{j,k+1,s} \qquad k \in ST, j \in CP, s \in S$

$TOUT_{i,s} \leq t_{i,NOK+1,s} \qquad i \in HP, s \in S$

$TOUT_{j,s} \geq t_{i,1,s} \qquad j \in CP, s \in S$

(v) Utility load (hot and cold)

$(t_{i,NOK+1,s} - TOUT_{i,s})\, F_{i,s} = qcu_{i,s} \qquad i \in HP, s \in S$

$(TOUT_{j,s} - t_{i,1,s})F_{j,s} = qhu_{j,s} \qquad j \in CP, s \in S$

(vi) Logical constraints

$q_{i,j,k,s} - \Omega\, z_{i,j,k} \leq 0 \quad i \in HP, j \in CP, k \in ST, s \in S$

$qcu_{i,s} - \Omega\, zcu_i \leq 0 \qquad i \in HP, s \in S$

$qhu_{j,s} - \Omega\, zhu_j \leq 0 \qquad j \in CP, s \in S$

(vii) Minimum approach temperature constraints

$dt_{i,j,k,s} \leq t_{i,k,s} - t_{j,k,s} + \Gamma(1 - z_{i,j,k}) \qquad i \in HP, j \in CP, k \in ST, s \in S$

$dt_{i,j,k+1,s} \leq t_{i,k+1,s} - t_{j,k+1,s} + \Gamma(1 - z_{i,j,k}) \quad i \in HP, j \in CP, k \in ST, s \in S$

$dtcu_{i,s} \leq t_{i,NOK+1,s} - TOUT_{CU} + \Gamma(1 - zcu_i) \qquad i \in HP, s \in S$

$dthu_{i,s} \leq TOUT_{CU,s} - t_{j,1,s} + \Gamma(1 - zhu_j) \qquad i \in CP, s \in S$

(viii) Minimum approach temperature bounds

$dt_{i,j,k,s}, dt_{i,j,k+1,s}, dtcu_{i,s}, dthu_{i,s} \geq \Delta T_{min}$

(ix) Positivity constraints

$t_{i,k,s}, t_{i,k+1,s}, t_{j,k,s}, t_{j,k+1,s} \geq 0$

$q_{i,j,k,s}, qcu_{i,s}, qhu_{j,s} \geq 0$

$Sh1_{i,s}, Sh2_{i,s}, SC1_{i,s}, SC2_{i,s} \geq 0$

(x) Binary variables

$z_{i,j,k}$, zcu_i, $zhu_j \in \{0,1\}$

$\forall\, i \in HP, j \in CP, k \in ST, s \in S$

S2 Information for distillation column design

Description and sampling of uncertain operating conditions (benzene column)

The overall flowsheet (shown in Figure S1) for the production of benzene via hydrodealkylation of toluene (Douglas, 1985) consists of feed mixer, fired heater, reactor, cooler, flash unit, purge stream, stabilizer, benzene column and toluene column.

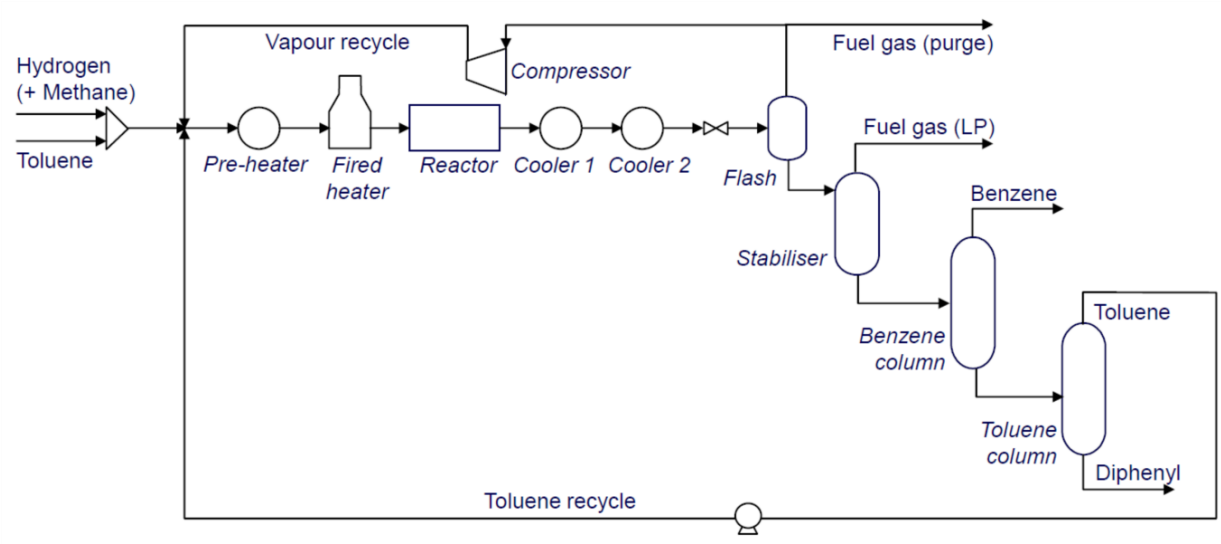


Figure S1 hydrodealkylation of toluene

Two homogenous reactions (Eq. 1 and Eq. 2) occur in the reactor. The first reaction produces benzene (desired product), while the second reaction produce diphenyl (undesired product).

Primary reaction (desirable)

$$C_6H_5CH_3 + H_2 \rightarrow C_6H_6 + CH_4 \quad (1)$$

Side reaction (undesirable)

$$2C_6H_6 \leftrightarrow C_6H_5–C_6H_5 + H_2 \quad (2)$$

The reactor exit gas, which consists of benzene, methane, hydrogen, diphenyl and unreacted toluene, is quenched in a partial condenser and subsequently in a flash separator to condense and separate the aromatics from the non-condensable methane and hydrogen. The remaining traces of hydrogen and methane in the aromatics are removed in a stabilizer column; benzene is recovered in the benzene column, and lastly, toluene is separated from diphenyl and is recycled back into the process. The flashed vapour is rich in hydrogen (with traces of methane) and is recycled back into the reactor. The design specification required to simulate the overall flowsheet are presented in Tables S1 and S2.

Table S1 Key data and specifications for HDA processing

Item	Specifications
Pre-heater	Pressure drop: 20 kPa outlet temperature: 230°C
Fired heater (furnace)	Pressure drop: 20 kPa outlet temperature: 725°C
Reactor	Void fraction: 1.0 Geometry: 1 tube Pressure drop: 20 kPa Volume: 300 m3 Diameter: 4.0 m
Cooler 1	Pressure drop: 20 kPa outlet temperature: 200°C
Cooler 2	Outlet temperature: 40°C Pressure drop: 20 kPa
Flash unit	operating pressure: 1000 kPa
Stabiliser column	99% recovery of benzene to bottom product the feed enters on the top tray. Design: 4 trays Operating pressure: 200 kPa
Benzene column	95 mol% purity of benzene product Design: 20 trays; feed stage: 10 Operating pressure: 200 kPa Specification: 95% recovery of benzene
Toluene column	>95% recovery of diphenyl Design: 20 trays; feed stage: 10 Operating pressure: 200 kPa Value for recovery of toluene: 95% Value for reflux ratio: 0.05
Purge splitter	Value for purge (fraction of flash vapour product): 0.20
Hydrogen feed	160 kmol/h 40°C, 3500 kPa Composition: 99 mol% H2, 1 mol% CH4
Toluene	58 kmol/h 40°C, 3500 kPa Composition: 100% pure
Steam	5 bar, 160°C (7.78 $/GJ)
Cooling water	30 – 40°C (0.354 $/GJ)
Heat transfer coefficient–column condenser	800 W/(m^2 K)
Heat transfer coefficient column reboiler	820 W/(m^2 K)
Operating hours per year	8000
Interest rate	10%
Plant life	5 years

Table S2 Key data for HDA reactions (T: Toluene; H: Hydrogen; B: Benzene; D: Diphenyl)

Reaction rate expression	Parameters
$r_T = -Ae^{\frac{E_A}{RT}} p_T p_H{}^{0.5}$ Units of r: kmol m^{-3} s^{-1}	A = 737 kmol m^{-3} s^{-1} kPa$^{-1.5}$ E_A = 2.3·10^5 kJ kmol^{-1}
$r_B = -2A_1 e^{\frac{E_{A1}}{RT}} p_B{}^2 + A_2 e^{\frac{E_{A2}}{RT}} p_D p_H$ Units of r: kmol m^{-3} s^{-1}	A_1 = 336 kmol m^{-3} s^{-1} kPa^{-2} E_{A1} = 2.13·10^5 kJ kmol^{-1} A_2 = 1434 kmol m^{-3} s^{-1} kPa^{-2} E_{A2} = 2.13·10^5 kJ kmol^{-1}

The main purpose of the purge stream in the process flowsheet is to avoid build-up of methane in the vapour recycle loop (Douglas, 1985). Mainly for the purpose of illustrating the proposed methodology, here the purged gas fraction is treated as an uncertain parameter that is represented using scenarios. In the way, stream conditions (calculated based on material and energy balances) at different purged gas fraction can be generated and recorded. Therefore, the unit operations on the process flowsheet are to be designed to accommodate these variations. Of particular interest in this work is the application of our proposed methodology to design the benzene distillation column with variations in feed condition (such as temperature, flow rate and compositions). Nonetheless, the proposed methodology can also be applied to design the remaining unit operations in the process flowsheet.

To this end, the uncertain parameter (i.e. purged gas fraction) is described using uniform distribution. Table S3 shows the nominal condition and the ten set of operating scenarios generated from the distribution using sampling method. Note that for this case study, all the scenarios are assumed to be equally probable.

Table S3 Sampled data for purged gas fraction

	S_{NC}	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
Split ratio	0.200	0.227	0.181	0.313	0.164	0.289	0.142	0.212	0.262	0.330	0.105

Next, a simulation model of the overall flowsheet for the production of benzene via hydrodealkylation of toluene (Douglas, 1985) is built in Aspen HYSYS using the data

presented in Tables S1 and S2. Each split ratio presented in Table S3 is used as input to the flowsheet and the corresponding inlet conditions to the benzene column are recorded and presented in Table 7.

Objective function and constraints

The objective of distillation column design is to identify the best column structure and operation conditions that can perform a desired separation at minimum total annualized cost. Therefore, the objective function and constraints can be defined as follows

$$\min \quad ACC + \sum_{s \in S} prob_s \times (OC_s + \pi(s_s^{h+} + s_s^{h-} + s_s^{g}))$$

$$\begin{aligned} s.t \quad & h_I(d, z_s, \theta_s) + s_s^{h+} + s_s^{h-} = 0 \quad \forall s \in S \\ & h_E(d, z_s, \theta_s) - s_s^{g} \leq 0 \quad \forall s \in S \\ & g_E(d, z_s, \theta_s) - s_s^{g} \leq 0 \quad \forall s \in S \end{aligned}$$

$$z_s \in Z;\ s_s^{h+}, s_s^{h-}, s_s^{g} \in \mathbb{R}^+$$

where d, z_s, θ_s represent design variable, operating condition, and uncertain parameter respectively. Subscript s denotes scenario and S is the total number of scenarios. The objective function is the sum of operating cost (*OC*), annualized capital cost (*ACC*) and a penalty term, which consists of a penalty weighting (π) that penalizes the violation of the equality and inequality constraints and slack variables (s_s^{h+}, s_s^{h-} and s_s^{g}) representing the extent of deviation from feasible operation. The annualized capital cost including cost of column shell, trays, condenser and reboiler is calculated using Guthrie cost correlations (Guthrie, 1969), and updated using 2014 Chemical Engineering Plant cost Index (4th quarter) ("Chemical Engineering magazine," 2014). The capital cost is annualized using the annualization factor described by Smith (2005). The operating cost is the sum of hot utility (steam) and cold utility (cooling water).

For the constraints; h_I denotes the set of implicit equality constraints representing material, energy, thermodynamic equations and hydraulic constraints embedded in the process simulator; h_E is the set of explicit equality constraints while g_E is the set of inequality constraints.

S3 References

Chemical Engineering magazine, 2014. www.che.com (accessed 11.20.14).

Douglas, J.M., 1988. Conceptual Design of Chemical Processes, McGraw-Hill chemical engineering series. McGraw-Hill. New York

Guthrie, K.M., 1969. Data and Techniques for Preliminary Capital Cost Estimating. Chem. Eng.

Smith, R., 2005. Chemical Process: Design and Integration. Wiley. Chichester, UK.

Yee, T.F., Grossmann, I.E., 1990. Simultaneous optimization models for heat integration - II. Heat-exchanger network synthesis. Comput. Chem. Eng. 14, 1165–1184.

References

Afrand, M., Ahmadi Nadooshan, A., Hassani, M., Yarmand, H., Dahari, M., 2016. Predicting the viscosity of multi-walled carbon nanotubes/water nanofluid by developing an optimal artificial neural network based on experimental data. Int. Commun. Heat Mass Transf. 77, 49–53.

Aguado, D., Ribes, J., Montoya, T., Ferrer, J., Seco, a., 2009. A methodology for sequencing batch reactor identification with artificial neural networks: A case study. Comput. Chem. Eng. 33, 465–472.

Al-Mayyahi, M.A., Hoadley, A.F.A., Smith, N.E., Rangaiah, G.P., 2011. Investigating the trade-off between operating revenue and CO2 emissions from crude oil distillation using a blend of two crudes. Fuel 90, 3577–3585.

Ali, S.F., Yusoff, N., Ganguly, S., Abidin, M.Z., 2013. Profit Maximization of a Crude Distillation Unit, 25–27.

Amaran, S., Zhang, T., Sahinidis, N. V., Sharda, B., Bury, S.J., 2016. Medium-term maintenance turnaround planning under uncertainty for integrated chemical sites. Comput. Chem. Eng. 84, 422–433.

Arbiza, M.J., Bonfill, A., Guillén, G., Mele, F.D., Espuña, A., Puigjaner, L., 2008. Metaheuristic multiobjective optimisation approach for the scheduling of multiproduct batch chemical plants. J. Clean. Prod. 16, 233–244.

AspenTech, 2011. Aspen HYSYS: Customization Guide. Aspen Technol. Inc.

Austgen, D.M., Rochelle, G.T., Peng, X., Chen, C.C., 1989. Model of Vapor—Liquid Equilibria for Aqueous Acid Gas—Alkanolamine Systems Using the Electrolyte—NRTL Equation. Ind. Eng. Chem. Res. 28, 1060–1073.

Bagajewicz, M., Ji, S., 2001. Rigorous Procedure for the Design of Conventional Atmospheric Crude Fractionation Units. Part I: Targeting. Ind. Eng. Chem. Res. 40, 617–626.

Bagajewicz, M.J., 1998. Energy savings horizons for the retrofit of chemical processes. Application to crude fractionation units. Comput. Chem. Eng. doi:10.1016/S0098-1354(98)00269-5

Basak, K., Abhilash, K.S., Ganguly, S., Saraf, D.N., 2002. On-Line Optimization of a Crude Distillation Unit with Constraints on Product Properties. Ind. Eng. Chem. Res. 41, 1557–1568.

Basheer, I.A., Hajmeer, M., 2000. Artificial neural networks: Fundamentals, computing, design, and application. J. Microbiol. Methods 43, 3–31.

Beale, M.H., Hagan, M.T., Demuth, H.B., 2015. Neural Network Toolbox™ User's Guide.

Ben Ali, J., Fnaiech, N., Saidi, L., Chebel-Morello, B., Fnaiech, F., 2015. Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. Appl. Acoust. 89, 16–27.

Benali, T., Tondeur, D., Jaubert, J.N., 2012. An improved crude oil atmospheric distillation process for energy integration: Part I: Energy and exergy analyses of the process when a flash is installed in the preheating train. Appl. Therm. Eng. 32, 125–131.

Bertsimas, D., Brown, D.B.D., Caramanis, C., 2010. Theory and Applications of Robust Optimization. Oper. Res. 50.

Biegler, L.T., Grossmann, I.E., Westerberg, A.W., 1997. Systematic methods of chemical process design, Prentice-Hall international series in the physical and chemical engineering sciences. Prentice Hall PTR.

Biegler, L.T., Lang, Y., Lin, W., 2014. Multi-scale optimization for process systems engineering. Comput. Chem. Eng. 60, 17–30. doi:10.1016/j.compchemeng.2013.07.009

Boukouvala, F., Hasan, M.M.F., Floudas, C.A., 2015. Global optimization of general constrained grey-box models: new method and its application to constrained PDEs for pressure swing adsorption. J. Glob. Optim.

Caballero, a, Grossmann, I.E., 2001. Generalized Disjunctive Programming Model for the Optimal 2260–2274.

Caballero, J.A., 2015. Logic hybrid simulation-optimization algorithm for distillation design. Comput. Chem. Eng. 72, 284–299.

Caballero, J.A., Milan-Yanez, D., Grossmann, I.E., 2005. Rigorous design of distillation columns: Integration of disjunctive programming and process simulators. Ind. Eng. Chem. Res. 44, 6760–6775.

Castelo, B.D., Gomes, G.L., Szklo, A.S., 2010. Challenges and technological opportunities for the oil refining industry: A Brazilian refinery case. Energy Policy 38, 3098–3105.

Chacon-Mondragon, O.L., Himmelblau, D.M., 1996. Integration of flexibility and control in process design. Comput. Chem. Eng. 20, 447–452.

Chang, A.F., Pashikanti, K., Liu, Y.A., 2012. Refinery Engineering: Integrated Process Modeling and Optimization. Wiley.

Chen, C.-L., Hung, P.-S., 2004. Simultaneous Synthesis of Flexible Heat-Exchange Networks with Uncertain Source-Stream Temperatures and Flow Rates. Ind.Eng.Chem.RES 43, 5916–5928. doi:10.1021/ie030701f

Chen, L., 2008. Heat-Integrated Crude Oil Distillation System Design. PhD Thesis, Univ. Manchester, UK.

Cooper, S., Mackenzie, W., 2013. Crude Oil in Europe : Production , Trade and Refining Outlook. http://www.easyfairs.com/fileadmin/groups/8/Shop_2012/Day_2__09.05__Steve_Cooper_pdf.pdf (accessed 1.2.15).

Copado-Méndez, P., Blum, C., Guillén-Gosálbez, G., Jiménez, L., 2013. Application of Large Neighborhood Search to Strategic Supply Chain Management in the Chemical Industry. Hybrid Metaheuristics SE - 12 434, 335–352.

Corbetta, M., Grossmann, I.E., Manenti, F., 2016. Process simulator-based optimization of biorefinery downstream processes under the Generalized Disjunctive Programming framework. Comput. Chem. Eng. 88, 73–85.

Dhole, V.R., Buckingham, P.R., 1994. Refinery column integration for debottlenecking and energy saving. Proc. ESCAPE IV Conf. Dublin.

Diwekar, U.M., Rubin, E.S., 1994. Parameter design methodology for chemical processes using a simulator. Ind. Eng. Chem. Res. 33, 292–298.

Du, K.L., Swamy, M.N.S., 2016. Search and optimization by metaheuristics: Techniques and algorithms inspired by nature. Search Optim. by Metaheuristics Tech. Algorithms Inspired by Nat. 1–434.

Dua, V., 2010. A mixed-integer programming approach for optimal configuration of artificial neural networks. Chem. Eng. Res. Des. 88, 55–60.

Duran, M.A., Grossmann, I.E., 1986. An outer-approximation algorithm for a class of mixed-integer nonlinear programs. Math. Program. 36, 307–339. doi:10.1007/BF02592064

Edgar, T.F., Himmelblau, D.M., Lasdon, L.S., 2001. Optimization of chemical processes, McGraw-Hill chemical engineering series. McGraw-Hill.

EIA, 2016. PETROLEUM & OTHER LIQUIDS. https://www.eia.gov/dnav/pet/pet_pri_imc3_k_m.htm (accessed 12.1.16).

EIA, 2012. TODAY IN ENERGY. https://www.eia.gov/todayinenergy/detail.php?id=7110 (accessed 1.2.15).

Errico, M., Tola, G., Mascia, M., 2009. Energy saving in a crude distillation unit by a preflash implementation. Appl. Therm. Eng. 29, 1642–1647.

Fahim, M.A., Al-Sahhaf, T.A., Elkilani, A., 2009. Fundamentals of Petroleum Refining. Elsevier Science. Amsterdam.

Fahmi, I., Cremaschi, S., 2012. Process synthesis of biodiesel production plant using artificial neural networks as the surrogate models. Comput. Chem. Eng. 46, 105–123.

Favennec, J., 2001. Refinery Operation and Management. Technip. Paris.

Floudas, C.A., 2013. Deterministic Global Optimization: Theory, Methods and Applications, Nonconvex Optimization and Its Applications. Springer US.

Floudas, C.A., 1995. Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications, Topics in Chemical Engineering. Oxford University Press.

Gary, J.H., Handwerk, G.E., Kaiser, M.J., 2007. Petroleum Refining: Technology and Economics, Fifth Edition. CRC Press.

Ghanizadeh, S., Fazli, M.S., Branch, D., 2013. Application of Genetic Algorithm on Heat Exchanger Network Optimization 6, 3378–3383.

Gharagheizi, F., Eslamimanesh, A., Mohammadi, A.H., Richon, D., 2011. Use of artificial neural network-group contribution method to determine surface tension of pure compounds. J. Chem. Eng. Data 56, 2587–2601.

Ghoreishi, S. a, Nekoui, M. a, Partovi, S., Basiri, S.O., 2011. Application of genetic algorithm for solving multi-objective optimization problems in robust control of distillation column. Int. J. 3, 32–43. doi:10.4156/ijact.vol3.

Gorissen, B., Yanikoglu, I., den Hertog, D., 2015. A practical guide to robust optimization. Omega 53, 124–137.

Grossmann, I.E., Caballero, J.A., Yeomans, H., 2000. Advances in Mathematical Programming for the Synthesis of Process Systems.pdf. Lat. Am. Appl. Res.

Grossmann, I.E., Guillén-Gosálbez, G., 2010. Scope for the application of mathematical programming techniques in the synthesis and planning of sustainable processes. Comput. Chem. Eng. 34, 1365–1376.

Grossmann, I.E., Westerberg, A.W., Biegler, L.T., 1987. Retrofit design of processes. FOCAPO Proc. 405–442.

Gu, W., Wang, K., Huang, Y., Zhang, B., Chen, Q., Hui, C.W., 2015. Energy Optimization for a Multistage Crude Oil Distillation Process. Chem. Eng. Technol. 38, 1243–1253.

Gueddar, T., Dua, V., 2011. Disaggregation-aggregation based model reduction for refinery-wide optimization. Comput. Chem. Eng. 35, 1838–1856.

Heiberger, R.M., Neuwirth, E., 2009. Polynomial Regression, in: R Through Excel: A Spreadsheet Interface for Statistics, Data Analysis, and Graphics. Springer New York, New York, NY, pp. 269–284.

Henao, C.A., Maravelias, C.T., 2010. Surrogate-Based Superstructure Optimization Framework. AIChE J. 57, 1216–1232.

Himmelblau, D.M., 2008. Accounts of Experiences in the Application of Artificial Neural Networks in Chemical Engineering. Ind. Eng. Chem. Res. 47, 5782–5796.

Hoch, P.M., Eliceche, A.M., Grossmann, I.E., 1995. Evaluation of design flexibility in distillation columns using rigorous models. Comput. Chem. Eng. 19, 669–674. doi:10.1016/0098-1354(95)87112-8

Hussain, M.A., 1999. Review of the applications of neural networks in chemical process control - simulation and online implementation. Artif. Intell. Eng. 13, 55–68.

Inamdar, S. V, Gupta, K.S., Saraf, D.N., 2004. Crude Distillation Unit Using the Elitist Non-. Chem. Eng. Res. Des. 82, 611–623.

Isafiade, A.J., Short, M., 2016. Simultaneous synthesis of flexible heat exchanger networks for unequal multi-period operations. Process Saf. Environ. Prot.

Javaloyes-Antón, J., Ruiz-Femenia, R., Caballero, J. a., 2013. Rigorous Design of Complex Distillation Columns Using Process Simulators and the Particle Swarm Optimization Algorithm. Ind. Eng. Chem. Res. 52, 15621–15634.

Jones, D.S.J., 1995. Elements of petroleum processing. John Wiley & Sons, West Sussex, UK.

Kabadi, V.N., Danner, R.P., 1985. A Modified Soave-Redlich-Kwong Equation of State for Water-Hydrocarbon Phase Equilibria. Ind. Eng. Chem. Process Des. Dev. 24, 537–541.

Kankar, P.K., Sharma, S.C., Harsha, S.P., 2011. Fault diagnosis of ball bearings using machine learning methods. Expert Syst. Appl. 38, 1876–1886.

Kostin, A.M., Guillen-Goselbez, G., Mele, F.D., Bagajewicz, M.J., Jimenez, L., 2012. Design and planning of infrastructures for bioethanol and sugar production under demand uncertainty. Chem. Eng. Res. Des. 90, 359–376.

Li, J., Du, J., Zhao, Z., Yao, P., 2015. Efficient Method for Flexibility Analysis of Large-Scale Nonconvex Heat Exchanger Networks. Ind. Eng. Chem. Res. 54, 10757–10767.

Liau, L.C.-K., Yang, T.C.-K., Tsai, M.-T., 2004. Expert system of a crude oil distillation unit for process optimization using neural networks. Expert Syst. Appl. 26, 247–255.

Liebmann, K., 1996. Integrated Crude Oil Distillation Design. PhD Thesis, UMIST, Manchester, UK.

Liebmann, K., Dhole, V.R., Jobson, M., 1998. Integrated Design of a Conventional Crude Oil Distillation Tower Using Pinch Analysis. Chem. Eng. Res. Des. 76, 335–347.

Long, N.V.D., Lee, M., 2017. Advances in Distillation Retrofit. Springer Nature, Singapore.

López C., D., Hoyos, L.J., Mahecha, C.A., Arellano-Garcia, H., Wozny, G., 2013. Optimization model of crude oil distillation units for optimal crude oil blending and operating conditions. Ind. Eng. Chem. Res. 52, 12993–13005.

Maples, R.E., 2000. Petroleum Refinery Process Economics. PennWell Corporation.

Marini, F., Walczak, B., 2015. Particle swarm optimization (PSO). A tutorial. Chemom. Intell. Lab. Syst. 149, 153–165.

Martí, R., Laguna, M., Glover, F., 2006. Principles of scatter search. Eur. J. Oper. Res. 169, 359–372.

MATLAB, 2014. MathWorks.
https://uk.mathworks.com/help/stats/fitcsvm.html (accessed 2.1.15).

Maurer, G., Prausnitz, J.M., 1978. On the derivation and extension of the uniquac equation. Fluid Phase Equilib. 2, 91–99.

Miettinen, K., 2008. Introduction to Multiobjective Optimization : Noninteractive Approaches. Springer. New York, US.

Misener, R., Floudas, C. a., 2014. ANTIGONE: Algorithms for coNTinuous / Integer Global Optimization of Nonlinear Equations. J. Glob. Optim. 59, 503–526.

Mitchell, M., 1998. An introduction to genetic algorithms 1–40.

More, R.K., Bulasara, V.K., Uppaluri, R., Banjara, V.R., 2010. Optimization of crude distillation system using aspen plus: Effect of binary feed selection on grass-root design. Chem. Eng. Res. Des. 88, 121–134.

Motlaghi, S., Jalali, F., Ahmadabadi, M.N., 2008. An expert system design for a crude oil distillation column with the neural networks model and the process optimization using genetic algorithm framework. Expert Syst. Appl. 35, 1540–1545.

Mouli, C., Madhuranthakam, R., Penlidis, A., 2016. Surrogate Models for Online Monitoring and Process Troubleshooting of NBR Emulsion Copolymerization.

Nelson, W.L., 1958. Petroleum refinery engineering, McGraw-Hill series in chemical engineering. McGraw-Hill. New York, US.

Nuchitprasittichai, A., Cremaschi, S., 2013. Optimization of CO2 capture process with aqueous amines - A comparison of two simulation-optimization approaches. Ind. Eng. Chem. Res. 52, 10236–10243.

Nuchitprasittichai, A., Cremaschi, S., 2012. An Algorithm to Determine Sample Sizes for Optimization with Artificial Neural Networks. AIChE J. 59, 805–812.

Ochoa-Estopier, L.M., Jobson, M., 2015a. Optimization of Heat-Integrated Crude Oil Distillation Systems. Part I: The Distillation Model. Ind. Eng. Chem. Res. 54, 4988–5000.

Ochoa-Estopier, L.M., Jobson, M., 2015b. Optimization of Heat-Integrated Crude Oil Distillation Systems. Part III: Optimization Framework. Ind. Eng. Chem. Res. 54, 5018–5036.

Ochoa-Estopier, L.M., Jobson, M., Chen, L., Rodríguez-Forero, C.A., Smith, R., 2015. Optimization of Heat-Integrated Crude Oil Distillation Systems. Part II: Heat Exchanger Network Retrofit Model. Ind. Eng. Chem. Res. 54, 5001–5017.

Ochoa-Estopier, L.M., Jobson, M., Smith, R., 2012. Operational optimization of crude oil distillation systems using artificial neural networks. Comput. Aided Chem. Eng. 30, 982–986.

Odjo, A.O., Jr., N.E.S., Yuan, W., Marcilla, A., Eden, M.R., Caballero, J.A., 2011. Disjunctive-Genetic Programming Approach to Synthesis of Process Networks. Ind. Eng. Chem. Res. 50, 6213–6228.

Osuolale, F., Zhang, J., 2015. Distillation control structure selection for energy-efficient operations. Chem. Eng. Technol. 38, 907–916.

Osuolale, F.N., Zhang, J., 2017. Thermodynamic optimization of atmospheric distillation unit. Comput. Chem. Eng. 103, 201–209.

Papalexandri, K.P., Pistikopoulos, E.N., 1994. Synthesis and Retrofit Design of Operable Heat Exchanger Networks. 1. {F}lexibility and Structural Controllability Aspects. Ind. Eng. Chem. Res. 33, 1718–1737.

Peng, D.-Y., Robinson, D.B., 1980. Two- and Three-Phase Equilibrium Calculations for Coal Gasification and Related Processes, in: Thermodynamics of Aqueous Systems with Industrial Applications, ACS Symposium Series. American Chemical Society, pp. 20–393.

Pérez Rivero, C., Sun, C., Theodoropoulos, C., Webb, C., 2016. Building a predictive model for PHB production from glycerol. Biochem. Eng. J. 116, 113–121.

Perry, R., Green, D., 2008. Perry's Chemical Engineers' Handbook, Eighth Edition, McGraw Hill professional. McGraw-Hill Education.

Pistikopoulos, E.N., Grossmann, I.E., 1988. Optimal design for improving process flexibility in linear systems. Comput. Chem. Eng. 12, 719–731.

Pistikopoulos, E.N., Ierapetritou, M.G., 1995. Novel approach for optimal process design under uncertainty. Comput. Chem. Eng. 19, 1089–1110.

Plöcker, U., Knapp, H., Prausnitz, J., 1978. Calculation of High-Pressure Vapor-Liquid Equilibria from a Corresponding-States Correlation with Emphasis on Asymmetric Mixtures. Ind. Eng. Chem. Process Des. Dev. 17, 324–332.

Prasad, V., Bequette, B.W., 2003. Nonlinear system identification and model reduction using artificial neural networks. Comput. Chem. Eng. 27, 1741–1754.

Quirante, N., Caballero, J.A., 2016. Large scale optimization of a sour water stripping plant using surrogate models. Comput. Chem. Eng. 92, 143–162.

Quirante, N., Javaloyes, J., Caballero, J.A., 2015. Rigorous Design of Distillation Columns Using Surrogate Models Based on Kriging Interpolation. AIChE 61, 2169–2187.

Rastogi, V., 2006. Heat Integrated Crude Oil Distillation System Design. University of Manchester, UK.

Ravagnani, M.A.S.S., Silva, A.P., Arroyo, P.A., Constantino, A.A., 2005. Heat exchanger network synthesis and optimisation using genetic algorithm. Appl. Therm. Eng. 25, 1003–1017.

Riazi, M.R., 1989. Characterization and Properties of Petroleum Fractions. ASTM Intenational. Pennsylvania, US.

Rogers, A., Ierapetritou, M., 2015. Feasibility and flexibility analysis of black-box processes part 2: Surrogate-based flexibility analysis. Chem. Eng. Sci. 137, 1005–1013.

Rogina, A., Šiško, I., Mohler, I., Ujević, Ž., Bolf, N., 2011. Soft sensor for continuous product quality estimation (in crude distillation unit). Chem. Eng. Res. Des. 89, 2070–2077.

Saaty, T.L., 2008. Decision making with the analytic hierarchy process. Int. J. Serv. Sci. 1, 83. doi:10.1504/IJSSCI.2008.017590

Sahinidis, N. V., 2004. Optimization under uncertainty: State-of-the-art and opportunities. Comput. Chem. Eng. 28, 971–983.

Seader, J.D., Henley, E.J., Roper, D.K., 2010. Separation Process Principles, 3rd Edition. John Wiley & Sons, New Jersey, US.

Sharma, R., Jindal, A., Mandawala, D., Jana, S.K., 1999. Design/Retrofit Targets of Pump-Around Refluxes for Better Energy Integration of a Crude Distillation Column. Ind. Eng. Chem. Res. 38, 2411–2417.

Silva, I.N. da, Spatti, D.H., Flauzino, R.A., Luisa Helena Bartocci Liboni, S.F. dos R.A., 2017. Artificial Neural Networks.

Skiborowski, M., Rautenberg, M., Marquardt, W., 2015. A Hybrid Evolutionary–Deterministic Optimization Approach for Conceptual Design. Ind. Eng. Chem. Res. 54, 10054–10072.

Smith, R., 2005. Chemical Process: Design and Integration. Wiley.

Smith, R., Jobson, M., Chen, L., 2010. Recent development in the retrofit of heat exchanger networks. Appl. Therm. Eng. 30, 2281–2289.

Soave, G., 1980. Rigorous and simplified procedures for determining the pure-component parameters in the Redlich-Kwong-soave equation of state. Chem. Eng. Sci. 35, 1725–1730.

Spangler, R., Varraveto, D., Schoonover, R., Hanke, T., 2006. REFINERY REVAMP-1: ConocoPhillips revamps crude unit to increase flexibility, profitability. Oil Gas J.

Suphanit, B., 1999. Design of Complex Distillation System. PhD Thesis, UMIST, Manchester, UK.

Swaney, R.E., Grossmann, I.E., 1985. An Index for Operational Flexibility in Chemical Process Design - Part I: Formulation and Theory. AIChE J. 31, 621–630.

Szklo, A., Schaeffer, R., 2007. Fuel specification, energy consumption and CO_2 emission in oil refineries. Energy 32, 1075–1092.

Tawarmalani, M., Sahinidis, N. V, 2005. Digital Object Identifier (A polyhedral branch-and-cut approach to global optimization. Math. Program., Ser. B 103, 225–249.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer. US

Vazquez-Castillo, J.A., Venegas-Sánchez, J.A., Segovia-Hernández, J.G., Hernández-Escoto, H., Hernández, S., Gutiérrez-Antonio, C., Briones-Ramírez, A., 2009. Design and optimization, using genetic algorithms, of intensified distillation systems for a class of quaternary mixtures. Comput. Chem. Eng. 33, 1841–1850.

Waheed, M.A., Oni, A.O., Adejuyigbe, S.B., Adewumi, B.A., 2014. Thermoeconomic and environmental assessment of a crude oil distillation unit of a Nigerian refinery. Appl. Therm. Eng. 66, 191–205.

Wang, H., Mastragostino, R., Swartz, C.L.E., 2016. Flexibility analysis of process supply chain networks. Comput. Chem. Eng. 84, 409–421.

Wang, J., Rong, G., 2010. Robust optimization model for crude oil scheduling under uncertainty. Ind. Eng. Chem. Res. 49, 1737–1748.

Wang, R., Diwekar, U., Padró, C.E.G., 2004. Efficient sampling techniques for uncertainties in risk analysis. Environ. Prog. 23, 141–157.

Watkins, R.N., 1979. Petroleum Refinery Distillation. Gulf Publishing Company, Book Division.

Wen, L., Gao, L., Li, X., Zhang, L., 2013. Free Pattern Search for global optimization. Appl. Soft Comput. J. 13, 3853–3863.

Westerberg, A.W., 2004. A retrospective on design and process synthesis. Comput. Chem. Eng. 28, 447–458.

Xie, W., Bonis, I., Theodoropoulos, C., 2015. Data-driven model reduction-based nonlinear MPC for large-scale distributed parameter systems. J. Process Control 35, 50–58.

Yang, C., Hou, J., 2016. Fed-batch fermentation penicillin process fault diagnosis and detection based on support vector machine. Neurocomputing 190, 117–123.

Yao, H., Chu, J., 2012. Operational optimization of a simulated atmospheric distillation column using support vector regression models and information analysis. Chem. Eng. Res. Des. 90, 2247–2261.

Yee, T.F., Grossmann, I.E., 1990. Simultaneous optimization models for heat integration - II. Heat-exchanger network synthesis. Comput. Chem. Eng. 14, 1165–1184.

Yuan, W., Odjo, A., Sammons, Jr., N.E., Caballero, J., Eden, M.R., 2009. Process Structure Optimization Using a Hybrid Disjunctive-Genetic Programming Approach. 10th Int. Symp. Process Syst. Eng. Part A 27, 669–674.

Zudkevitch, D., Joffe, J., 1970. Correlation and prediction of vapor-liquid equilibria with the redlich-kwong equation of state. AIChE J. 16, 112–119.