CHEMICAL ENGINEERING TRANSACTIONS
VOL. 61, 2017
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet
Guest Editors: Petar S Varbanov, Rongxin Su, Hon Loong Lam, Xia Liu, Jiří J Klemeš

ISBN 978-88-95608-51-8; **ISSN** 2283-9216
DOI: 10.3303/CET1761306

Feasibility Bounds in Operational Optimization and Design of Crude Oil Distillation Systems Using Surrogate Methods

Megan Jobson[a,*], Lluvia M. Ochoa-Estopier[a], Dauda Ibrahim[a], Lu Chen[b], Gonzalo Guillén Gosálbez[c], Jie Li[a]

[a]Centre for Process Integration, School of Chemical Engineering and Analytical Science, The University of Manchester Sackville Street, Manchester M13 9PL, UK
[b]Process Integration Limited, Station House, Stamford New Road, Altrincham, Cheshire WA14 1EP, UK
cCentre for Process Systems Engineering, Imperial College of Science, Technology and Medicine, London SW7 2BY, UK
megan.jobson@manchester.ac.uk

Crude oil distillation systems, comprising distillation units and their associated heat recovery networks, are highly complex and integrated systems. Their function is to separate crude oil into several streams with different boiling ranges for downstream processing. In practice, these systems typically need to be operated efficiently, so that the value added by the separation units is maximized (e.g. by maximizing flows of the most valuable intermediate products while minimizing production costs). Process improvement projects typically seek to increase production and/or to reduce energy consumption in existing crude oil distillation systems.
Recent developments in design and operational optimization of crude oil distillation systems apply surrogate models, together with stochastic optimization techniques, for column design or operational optimization. Column operation is highly constrained by the product specifications and, in existing columns, by physical limitations related to column configuration and size. Column models must capture these constraints. The effectiveness of surrogate modelling of the columns is enhanced by this work that develops complementary screening and filtering correlations and surrogate models (using artificial neural networks and support vector machines) to define feasibility bounds. Applying these feasibility bounds enables more targeted searches, bringing robustness and efficiency to the optimization frameworks.
Examples and case studies illustrate the effectiveness of the correlations and surrogate models for defining constraints in design and operational optimization approaches.

1. Introduction

The importance of distillation in the chemical process industries is well known. Methods for column design are established, but challenges remain for distillation retrofit and operational optimization. Design, retrofit and operational optimization problems have different features related to objectives and constraints, which has implications for modelling and optimization. Models need to be sufficiently rigorous and versatile to describe real processes, whether for design or simulation. The effectiveness of optimizations applying these models depends on the complexity and accuracy of the model, on the optimization algorithm and on the definition of optimization bounds. A wide search space can facilitate global optimization, but may include many infeasible solutions, which can undermine the optimization search.
Crude oil distillation is an essential, capital- and energy-intensive step in petroleum refining that presents particular challenges for optimization-based process design. Challenges include the complexity of the multicomponent mixture, the complex column configuration, the interactions with the heat recovery system and the many degrees of freedom for design and operation. Constraints relate to product quality (defined e.g. in terms of boiling properties) and to equipment limitations (e.g. column hydraulics). These challenges and constraints motivate ongoing research to develop effective and robust new methodologies, considering feasibility bounds and practical constraints, with potential impact on engineering costs, product yield and production margins, energy demand and environmental impact, especially greenhouse gas emissions.

Please cite this article as: Jobson M., Ochoa-Estopier L.M., Ibrahim D., Chen L., Gosálbez G.G., Li J., 2017, Feasibility bounds in operational optimization and design of crude oil distillation systems using surrogate methods, Chemical Engineering Transactions, 61, 1849-1854 DOI:10.3303/CET1761306

1.1 The State of the Art

Currently, industrial practice for crude oil distillation column design, retrofit and operational optimization is not systematic: it relies on trial and error, supported by rigorous distillation simulation models and user experience. Heat recovery is typically considered sequentially to column design, leading to iterative design processes that hinder generation of near-optimal designs (Kamel et al., 2013). Early research in this area used rigorous simulation models together with thermodynamic (pinch) analysis to analyze heat recovery implications of proposed design or operating conditions (Bagajewicz and Ji, 2001). Later, optimization-based approaches for column design and retrofit applied shortcut distillation models (Chen, 2008). However, these models present challenges for reflecting industrially-relevant specifications and configurations and for initialization and convergence. Nevertheless, these models can consider the column and heat recovery system in some detail.
More recent research exploits the versatility and accuracy of rigorous simulation models (Caballero et al. 2005) to overcome their inherent computational complexity by regressing significantly simpler surrogate models and later (Quirante et al., 2015)and the most recently (Osuolale and Zhang, 2017). External software (e.g. Visual Basic, MatLab) allows multiple simulations to be carried out, to generate samples for the regression. Simulation outputs include operational variables, as well as data classifying convergence (i.e. simulation specifications are met) and feasibility (i.e. additional specifications are also met). The data sets created allow training, testing and validation of a regressed model for application within an optimization framework to identify design and operational variables that perform best with respect to a given objective function.
Diverse approaches have been applied to regress surrogate distillation models to process simulation results for the purposes of operational optimization of crude oil distillation columns; these include support vector regression (Yao and Chu, 2012), polynomial surrogate models (López C. et al., 2013) and artificial neural networks (Smith et al., 2013 and later Ochoa-Estopier et al. (2015a). Surrogate models have also been developed by regression against plant data (Liau et al., 2004) and also later Motlaghi et al. (2007), but the data underpinning such models is inevitably available only within 'normal' operating ranges.
It is well known that random sampling techniques, such as the Latin hypercube sampling technique applied by López C. et al. (2013) and more recently Ochoa-Estopier et al. (2015a), ensure that samples are well distributed within a defined range. A drawback of these techniques is that many rigorous simulations need to be performed, which is time-consuming. Furthermore, many simulations may not converge; as a result only a small proportion of the samples may be useable in the regression.
Process optimization requires variable bounds to be specified; methods for identifying suitable bounds – leading to feasible solutions – have received relatively little attention in the research literature, but are known from crude oil distillation applications experience (Ochoa-Estopier et al., 2016) to impact significantly on the optimization solution. While identifying the available degrees of freedom for optimization is relatively easy, defining sensible bounds for these optimization variables requires an understanding of the trade-offs between these variables and process constraints. For example, bounds need to capture the extent to which a product flow rate could be increased before its quality specification is violated.
Data analysis is useful to understand these trade-offs as well as to identify the dominant degrees of freedom and constraints. Current methods for data analysis and modelling, such as principal component analysis and partial least squares modelling, use plant measurements to identify improvements to current process operation (Garcia-Munoz and MacGregor, 2016). However, plant measurements are not reliable when the purpose of optimization is to move away from 'normal' operation; in this case, data may be generated by simulation studies. Systematic approaches to effectively identify the optimization search space are currently lacking, but are needed where the surrogate model is regressed against simulation results. Bounds on the search space guide the optimizer to find solutions that meet all process constraints in relatively short computation times.
State-of-the-art methods for regressing surrogate models to crude oil distillation simulation results are gaining maturity. The potential – and therefore industrial relevance – of these methods is currently limited by inability to define systematically operating points for which simulation convergence and feasibility is likely. As a result, the quality of the surrogate model may be compromised by poor quality samples (many non-converged or infeasible samples) and the optimization may be compromised by a poorly defined search space, given the lack of suitable optimization bounds. This paper showcases recent work addressing this limitation.

2. Recent developments in operational optimization – statistical analysis of samples

Section 1.1 highlights recent developments in design and optimization of heat-integrated crude oil distillation systems where surrogate models regressed against process simulation results represent the distillation process. While it is straightforward to regress the surrogate model against only converged and feasible samples, too few useable samples will compromise the quality of the surrogate model. Also, the optimization search space is not easily partitioned or characterized into regions where feasibility and convergence is assured. If the surrogate models are effectively blind to the trade-offs between the degrees of freedom and process constraints, these important relationships cannot be incorporated into the surrogate models. These limitations motivate current development of methods to bound and partition the search space adequately.

Artificial neural network models for operational optimization (Ochoa-Estopier and Jobson, 2015b) initially were applied in conjunction with an artificial neural network model (called the feasibility ANN) with hyperbolic tangent and identity functions to characterize simulation convergence of samples. This class of model helps to bias the optimizer towards operating points that satisfy the specifications of the full process simulation model, and will therefore converge. As the surrogate modelling approach gains maturity and is applied in various industrial projects of increasing complexity (Ochoa-Estopier et al., 2016), this approach for identifying likelihood of convergence representing feasibility has presented significant limitations.

Firstly, the original feasibility model only used information on convergence of the samples simulated. As the number of simulation specifications is limited, additional constraints, relating to feasibility of solutions (e.g. other product quality criteria and hardware limitations) also need to be considered when detecting the solution space. In addition, it is advantageous to appreciate interactions between process variables, and therefore to correlate convergence and feasibility with more than one variable; for example, product quality may depend on the sum of duties of two pump-arounds, each of which is an operational variable.

The industrial studies found that optimization capabilities are seriously compromised by the absence of well-defined bounds for sampling and optimization by lack of understanding about interacting variables. To address these issues, an approach to systematize the definition of bounds is proposed:

1) A set of samples is generated using initial values for variable bounds. For each sample, performance metrics, including those relating to constraints, are collated.
2) All pairs of variables are plotted to visualize interactions, highlighting convergence and feasibility status.
3) Scatter plots are generated to represent visually how each system performance metric (including compliance with a constraint) varies with each optimization variable. For a given optimization variable, a scatter plot highlights the range for which samples converged and the range for which samples meet a given constraint. These plots also reveal trends in performance with that optimization variable – e.g. whether the performance is insensitive, varies monotonically or presents a trade-off.
4) Simple linear and quadratic correlations are regressed for each optimization variable, to discern trends in the system performance with that variable.
5) Statistical metrics are applied to these scatter plots, aiming to define, for each optimization variable, bounds that include a specified proportion of converged samples.
6) If monotonic trends detected in step 3) are statistically significant, the variable bound may be defined to include points in the most attractive direction (even if relatively few converged points exist in that range) or to ensure that the base case is also included.

Section 4 illustrates these concepts for two case studies and demonstrates how refining the variable bounds can lead to generation of much more useful sample sets.

3. Recent developments in column design – surrogate models with a data classifier

Current research is extending the application of surrogate models to design of crude oil distillation columns (Ibrahim et al., 2017). Design optimization aims to select the column structure (number of stages in each column section), to determine the column diameter to avoid flooding, as well as to select optimal operating conditions. Design optimization objectives include net profit, total annualized cost and fired heating demand.

This design approach uses Pinch Analysis techniques to evaluate the minimum demand for hot and cold utilities, but does not consider the details or cost of the heat recovery system. To select the number of stages in each section, a maximum number of stages and stage efficiencies are defined as optimization variables; these efficiencies allow selection of a number of ‘active’ stages in each section (Caballero et al., 2005).

Surrogate models are showing excellent promise in allowing these complex design problems to be solved with moderate engineering and computation time. This work uses artificial neural network models to minimize total annualized cost, applying a genetic algorithm. Samples created using process simulation software that fail to converge are discarded from the set used to regress the surrogate model of the distillation column.

This work uses non-converged samples to define boundaries of the search space, in order to effectively exclude such non-converged points. As previous work (Ochoa-Estopier and Jobson, 2015b) found feasibility ANNs limited in their efficacy, this work uses non-linear (polynomial) regression to create support vector machines (Vapnik, 1995). Support vector machines are algorithms for classifying data into two categories; in this work, they demarcate regions with converged and non-converged samples. The support vector machine is regressed and validated using all simulation-generated samples. Defining the design optimization search space in this way significantly enhances the quality of the optimization in terms of computation time and optimal performance. The search space still includes infeasible designs (i.e. that do not meet all the problem constraints). When the column surrogate model generates these off-specification solutions, they are penalized in the objective function.

4. Examples and case studies

Operational optimization was carried out for a highly constrained crude oil distillation system of a large Chinese refinery with two prefractionators, two atmospheric units and one vacuum unit. The optimization accounted for details of its considerably complex heat recovery system (50 process-to-process heat exchangers, 20 utility exchangers and 11 stream splitters). In this study, the vast majority of samples did not meet all constraints, and a strategic approach was needed to systematically identify and correlate problem bounds. Figure 1 shows how the algorithm identified a correlation between convergence and the kerosene product flow rate, and also identified ranges in which feasibility constraints are met; based on this information, the algorithm proposed new upper and lower bounds. Similarly, Figure 2 shows how data analysis identified trends in product quality and hydraulic performance and therefore adjusted sampling bounds. The probability of finding feasible solutions was initially 1/83,000; refining the sampling bounds and using neural network models increased this value to 1/5.
In a second industrial case, artificial neural network models were developed for a medium-scale European crude oil distillation system with one prefractionator, one atmospheric unit, 16 process heat exchangers and 7 coolers. Product quality was very tightly specified and operation was very close to the hydraulic limitations of the column. Sensitivity studies identified that the total duty of pump-arounds was the dominant variable affecting the furnace duty, flooding and product quality (as shown in Figure 3). Given that decreasing the upper bound could negatively impact on product revenue, new bounds aimed to prioritize feasibility. After analysis, the fraction of converged samples increased from 36 % to 75 % and the fraction of feasible samples increased from 1 % to 10 %.

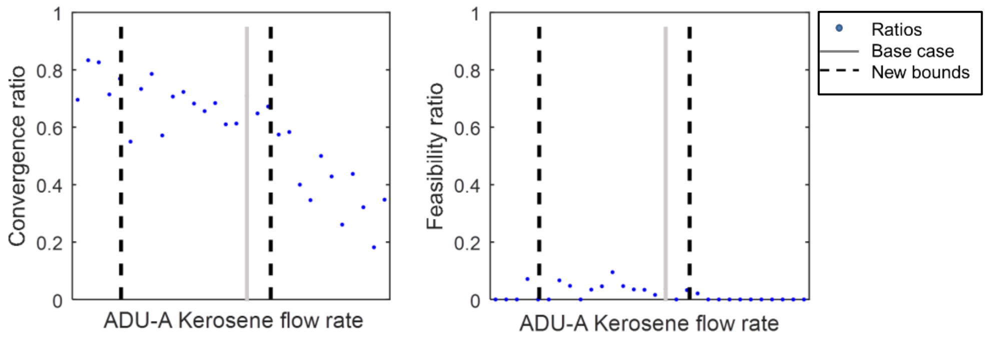

Figure 1: Scatter plots used to select bounds based on convergence and feasibility.

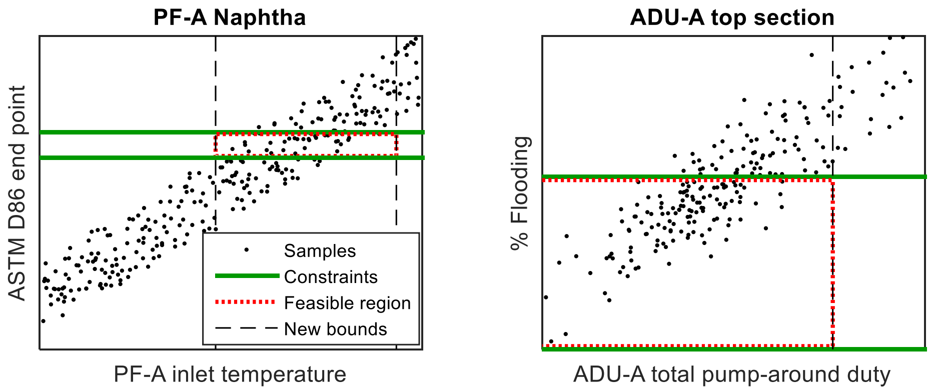

Figure 2: Scatter plots show interdependence of independent variables on product quality and flooding.

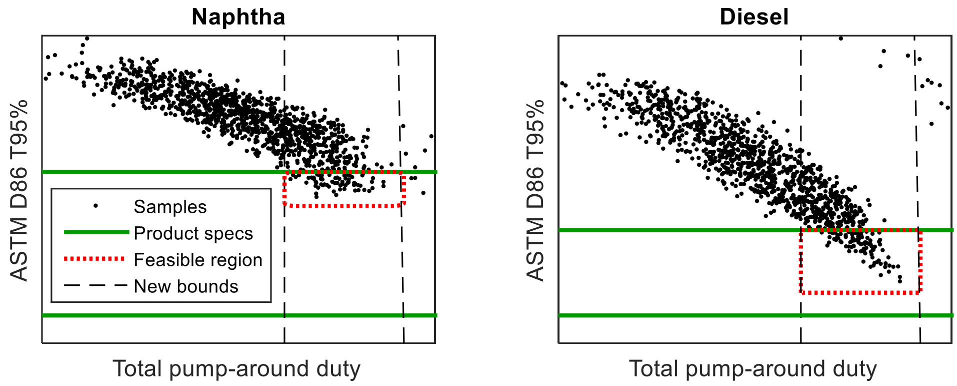

Figure 3: Scatter plots show that total pump-around duty is correlated with product quality.

Table 2: Validation of support vector machine – classification results

Convergence?	Sample size	Yes (predicted to converge)	No (predicted not to converge)
Yes (converged)	1061	966	95
No (not converged)	689	293	396
Total	1,750		

These scatter plots show that accounting for trends in and relationships between optimization variables can help to improve sampling and therefore the surrogate model, and can refine the optimization bounds. Understanding of interdependence between variables can thus inform the optimizer. Statistical analysis of the distribution of points and correlation of linear and quadratic trends helps to define bounds.

Surrogate models are applied for the design of an atmospheric unit separating 100,000 bbl/d (662 m^3/h) of Venezuelan Tia Juana light crude oil (Watkins, 1979) into five products (Ibrahim et al., 2017). The initial, non-optimized design is drawn from Chen (2008); there are five sections and three side strippers. 7,000 samples were generated using Aspen HYSYS v8.6 and MatLab R2015a, of which 59 % converged. The surrogate model is created using the Artificial Neural Network Tool Box in MatLab. 5,250 (75 %) of the samples are used to train the non-linear support vector machine, applying the function fitcsvm in MatLab, and the remainder are used to validate its classification predictions. The surrogate model shows excellent agreement with the original simulation results for all the converged samples in the validation set: temperature predictions are all within 0.5 °C and predictions of all other variables are within 3 % (Ibrahim et al., 2017). Table 2 summarizes the predictions made by the support vector machine.

Table 2 shows that 9 % of samples which converged were classified incorrectly (95/1,061), implying that only 5 % of all potential samples would be neglected because of a 'false negative' prediction. Around 43 % of samples that did not converge (293/689) were incorrectly classified as converging, which implies that around 17 % of all possible samples would be unnecessarily included because of a 'false positive' prediction. These results provide strong evidence that the classification provided by a support vector machine can help to reduce the number of samples generated (and the time spent generating these) by 28 % in this case, with only around 5 % of all samples being erroneously rejected. The best 10 optimized solutions obtained using the surrogate model all converged when used as Aspen HYSYS simulation inputs. In contrast, when the support vector machine is not used to guide the optimization, only 6 of the best 10 solutions converged. The optimization time was similar in the two cases: 60s – 166 s with the support vector machine and 77s – 80 s without.

5. Conclusions

Surrogate models allow the results of multiple process simulation runs to be regressed as multiple input–multiple output models for the purposes of design and operational optimization. It is necessary to filter the simulation results to meaningfully define a suitable search space for process optimization. New techniques are enhancing current capabilities for operational optimization and design of crude oil distillation systems. Surrogate modelling methods for operational optimization are sufficiently powerful and robust to be used in commercial studies and have been implemented in prototype software tools. Graphical and statistical methods to bias sampling and optimization to points that also meet other process specifications have been found useful, but these are not yet fully systematized.

Ongoing work in design of crude oil distillation systems is exploring other surrogate modelling techniques and other classification algorithms in order to develop robust and computationally efficient approaches to design. Support vector machines are helpful in partitioning the search space based on non-convergence of samples and a relatively small proportion of potentially promising solutions is likely to be discarded.

Acknowledgments

Work by Ochoa-Estopier was supported by Innovate UK via a Knowledge Transfer Partnership [No. KTP009567, 2014] between the University of Manchester and Process Integration Limited. Work by Ibrahim is financially supported by the Petroleum Technology Development Fund (PTDF), Nigeria.

References

Bagajewicz M., Ji S., 2001, Rigorous Procedure for the Design of Conventional Atmospheric Crude Fractionation Units. Part I: Targeting. Industrial & Engineering Chemistry Research, 40, 617–626.

Caballero J.A., Milan-Yanez D., Grossmann I.E., 2005, Rigorous Design of Distillation Columns: Integration of Disjunctive Programming and Process Simulators, Industrial and Engineering Chemistry Research, 44, 6760–6775.

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

Garcia-Munoz S., MacGregor J.F., 2016, Success Stories in the Process Industries, Chemical Engineering Progress, 112, 36–40.

Ibrahim D., Jobson M., Lie J., Guillén Gosálbez G., 2017, Surrogate Models Combined with a Support Vector Machine for the Optimized Design of Crude Oil Distillation Units Using Genetic Algorithms, Proceedings of ESCAPE 27, Barcelona, Spain (accepted).

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 Systems with Applications, 26, 247–255.

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, Industrial & Engineering Chemistry Research, 52, 12993–13005.

Kamel D., Gadalla M., Ashour F., 2013, New Retrofit Approach for Optimisation and Modification for a Crude Oil Distillation System, Chemical Engineering Transactions, 35, 1363-1368.

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 Systems with Applications, 35, 1540–1545.

Ochoa-Estopier L.M., Jobson M., 2015a. Optimization of Heat-Integrated Crude Oil Distillation Systems. Part I: The Distillation Model. Industrial and Engineering Chemistry Research, 54, 4988–5000.

Ochoa-Estopier L.M., Jobson M., 2015b. Optimization of Heat-Integrated Crude Oil Distillation Systems. Part III: Optimization Framework. Industrial and Engineering Chemistry Research, 54, 5018−5036.

Ochoa-Estopier L.M., Jobson M., Chen L., Smith R., 2016, Retrofit and Off-Line Operational Optimisation of Heat-Integrated Crude Oil Distillation Systems. In Proceedings of the AIChE Spring Meeting 2016.

Osuolale F.N., Zhang J., 2017, Thermodynamic Optimization of Atmospheric Distillation Unit, Computers and Chemical Engineering, 103, 201–209.

Osuolale F.N., Zhang J., 2014 Energy Efficient Control and Optimisation of Distillation Column Using Artificial Neural Network, Chemical Engineering Transactions, 39, 37-42.

Quirante N., Javaloyes J., Caballero J.A., 2015, Rigorous Design of Distillation Columns Using Surrogate Models Based on Kriging Interpolation, AIChE J., 61, 2169-2186.

Smith, R., Ochoa-Estopier, L.M., Jobson, M., 2013, The Use of Reduced Models in the Optimisation of Energy Integrated Processes, Chemical Engineering Transactions, 35, 139-144.

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