A Framework for the Near-Real-Time Optimization of Integrated Oil & Gas Midstream Processing Networks

Abdullah Al Ghazal

Centre for Process Systems Engineering

Department of Chemical Engineering and Chemical Technology

Imperial College London

London, SW7 2BY

Statement of Originality

I hereby declare that this thesis and the work contained herein are my own. I spent a considerable amount of time collecting data, developing and validating models for this work. Some of the work is also drawn from my own experience, having worked in the oil & gas industry for 12 years. Wherever this is the case, I ensured that this is properly declared. All inputs from other sources have been appropriately referenced.

Copyright Declaration

Confidentiality Statement

The data used in this researched were masked and altered to allow publishing findings. While they were inspired from a real case, they were diligently altered and neutralized to preserve confidentiality. Moreover, most of the data used was acquired from the public domain. As such, data contained within this work cannot and should not be used as a reference to report on production data of any plants or fields.

Table of Contents

1. Abstract

The oil and gas industry plays a key role in the world's economy. Vast quantities of crude oil, their by-products and derivatives are produced, processed and distributed every day. Indeed, producing and processing significant volumes of crude oil requires connecting to wells in different fields that are usually spread across large geographical areas. This crude oil is then processed by Gas Oil Separation Plants (GOSPs). These facilities are often grouped into clusters that are within approximate distance from each other and then connected laterally via swing lines which allow shifting part or all of the production from one GOSP to another. Transfer lines also exist to allow processing intermediate products in neighbouring GOSPs, thereby increasing complexity and possible interactions. In return, this provides an opportunity to leverage mathematical optimization to improve network planning and load allocation.

Similarly, in major oil producing countries, vast gas processing networks exist to process associated and non-associated gases. These gas plants are often located near major feed sources. Similar to GOSPs, they are also often connected through swing lines, which allow shifting feedstock from some plants to others.

GOSPs and gas plants are often grouped as oil and gas midstream plants. These plants are operated on varied time horizons and plant boundaries. While plant operators are concerned with the day-to-day operation of their facility, network operators must ensure that the entire network is operated optimally and that product supply is balanced with demand. They are therefore in charge of allocating load to individual plants, while knowing each plants constraints and processing capabilities. Network planners are also in charge of producing production plans at varied time-scales, which vary from yearly to monthly and near-real time.

This work aims to establish a novel framework for optimizing Oil and Gas Midstream plants for near-real time network operation. This topic has not been specifically addressed in the existing literature. It examines problems which involve operating networks of GOSPs and gas plants towards an optimal solution. It examines various modelling approaches which are suited for this specific application. It then focuses at this stage of the research on the GOSP optimization problem where it addresses optimizing the operation of a complex network of GOSPs. The goal is to operate this network such that oil production targets are met at minimum energy consumption, and therefore minimizing OpEx and Greenhouse Gas Emissions. Similarly, it is often required to operate the network

such that production is maximized. This thesis proposes a novel methodology to formulate and solve this problem. It describes the level of fidelity used to represent physical process units. A Mixed Integer Non-Linear Programming (MINLP) problem is then formulated and solved to optimize load allocation, swing line flowrates and equipment utilization. The model demonstrates advanced capabilities to systematically prescribe optimal operating points. This was then applied to an existing integrated network of GOSPs and tested at varying crude oil demand levels. The results demonstrate the ability to minimize energy consumption by up to 51% in the 50% throughput case while meeting oil production targets without added capital investment.

2. Introduction

This section aims to introduce the subject of Oil & Gas Midstream plants optimization. It provides a background to the industry, its challenges and relates the work to Enterprise-Wide Optimization. It also aims to describe in brief the industrial problem which is to be studied and briefly highlights currently dominant approaches to solve similar problems. While both oil processing and gas facilities are introduced, the thesis will focus on applying the techniques on the former. Nevertheless, it will also explore techniques that allow applying it to gas plants (e.g. through use of surrogates). This section also highlights the expected benefits of formulating a framework for the robust optimization of this network. The introduction concludes with establishing the research objectives which shall set the stage for next chapters.

2.1 The Importance of the Oil & Gas Industry

The oil & gas upstream industry plays a very important role in today's global economy. The industry provides raw materials that are used in a vast array of other industries from refining, petrochemicals, automotive, pharmaceuticals and many others. Oil & Gas operators explore, produce, process, store and transport huge volumes of materials across the globe. These operations are often highly complex and are subject to a wide range of environmental, political, legal and economic pressures. Moreover, operators within this industry are often under significant pressure to optimize operations and cut cost to remain competitive throughout periods of highly volatile product prices (Clews, 2016).

The economic growth around the world is placing an increasingly high demand on oil, gas and their derivatives. Despite growing importance of alternative fuels, the U.S. Energy Information Administration (*EIA*) still projects that oil & gas will constitute the largest share of the US energy mix through 2050. In fact, it is projected that in 2050, oil & gas contribution to the US energy mix will only be slightly lower than it was in 2017 (U.S. EIA, 2018).

In 2015, the World's average consumption of oil was approximately 93 million barrels per day. One third of this consumption came from the U.S. and China (U.S. EIA, 2018). Additionally, the demand for oil has been steadily rising in many developing countries. In China, the demand for oil has been steadily increasing since 1965. In fact, it soared from 7.9 million b/d in 2008 to an impressive 12.8 million b/d in 2017 (CEIC, 2018). This exceeds the all-time highest oil production from Saudi Arabia (Bloomberg, 2018).

2.2 The Oil & Gas Industry Segments

Oil & Gas practitioners have traditionally segmented the industry in differing ways. In this thesis, the author shall adopt the convention of breaking the industry into 3 sectors, namely: Upstream, Midstream and Downstream.

The following is a high-level description of what each sector constitutes.

Upstream:

This sector typically includes exploration and drilling activities which precedes the development of a field. It also includes activities relating to the development of fields and the production from oil wells (Devold, 2013).

Midstream:

Midstream typically includes the separation of oil and gas and the stabilization of oil. It also includes gas treatment and Natural Gas Liquids (NGL) production activities. This sector also includes the oil and gas pipeline system (Devold, 2013).

Downstream:

Downstream includes activities in which oil is processed to a range of marketable products. This sector includes refineries, which produce final products such as diesel and gasoline. Downstream also includes petrochemical plants, which mainly use a range of hydrocarbon feedstock to produce chemical products such as plastics (Devold, 2013).

2.3 Midstream Network Operation Challenges

Gathering and processing crude oil in oil-rich regions, such as the Arabian Peninsula, involves connecting to wells in different fields that are usually within vastly remote distances from each other. This crude can then be separated by Gas Oil Separation Plants (GOSPs) which are also located remotely from each other (Abdel-Aal et al., 2003). These facilities are often clustered into groups based on their proximity from each other. They are then connected laterally via swing and transfer lines which allow shifting part or all of the production from one GOSP to another. The purpose of those pipelines is to provide an added flexibility to the operation of the overall network. For example, when a bottleneck exists in one GOSP relating to water processing capacity, the production from wells with higher water-cut may be diverted to a GOSP which is not similarly bottlenecked. This allows processing higher total crude volumes and leads to improving the utilization of these assets. Also, this allows distributing crude oil more optimally to minimize energy consumption for the overall network while meeting production quotas. Similarly, the availability of swing lines can allow shifting production from GOSPs experiencing planned or unplanned outages to operational ones (Liu et al., 2016).

Processing significant volumes of gas from vastly spread-out sources is similarly complex and involves routing feed to different plants. Gas plants receive feed either from associated or non-associated gas fields. The former produce both oil and gas while the latter only produce gas. They also receive condensate, which is also known as wet gas (Senthamaraikkannan et al., 2014).

Gas plants process the received feed in a number of stages based on its quality. Typically, the process starts with acid gas removal. This is followed by dehydration to remove water. The dehydrated gas is then sent to NGL recovery. This is often a cryogenic process that produces light gas (mostly methane) and a heavier stream (ethane+ or propane+) (Al-Sobhi & Elkamel, 2015)

Some oil majors operate vast networks of those Midstream plants. This study is inspired by a real case study of an oil company which operates networks of Upstream, Midstream and Downstream plants.

The operation of those networks is done at multiple hierarchal levels, which will be discussed in the subsequent section. However, this study is concerned with the near-real time optimization of these networks. This is done centrally by planners who decide the allocation of raw materials to different plants. The aim is to operate those assets to meet supply quotas in the most optimal manner. Indeed, given the vast complexity of the networks and the number of controlled variables, the problem lends itself very well to mathematical optimization.

2.4 Network Optimization Approaches

The area of 'near-real time optimization' has not been studied extensively in the literature. Researchers have mostly followed the industry's traditional approach of breaking the network's management process into a hierarchal structure. The structure consists of a planning, a scheduling and a control layer (Grossmann, 2012). Figure 1 illustrates this hierarchy.

The planning layer typically involves optimizing the process horizon for a period of few weeks to one year. The scheduling horizon, on the other hand, lasts for around a week. The control layer is used to optimize the process on real time basis. The last layer typically interfaces directly with the plant's (Model Based Controllers) MPCs. The size of the system being tackled traditionally decreases rapidly towards the bottom of the optimization hierarchy. For examples, planning models can cover full supply chains, which include multiple facilities and distribution networks. Scheduling models typically cover individual facilities or distribution schemes. The control layer typically covers individual units within plants (Grossmann & Biegler, 1995).

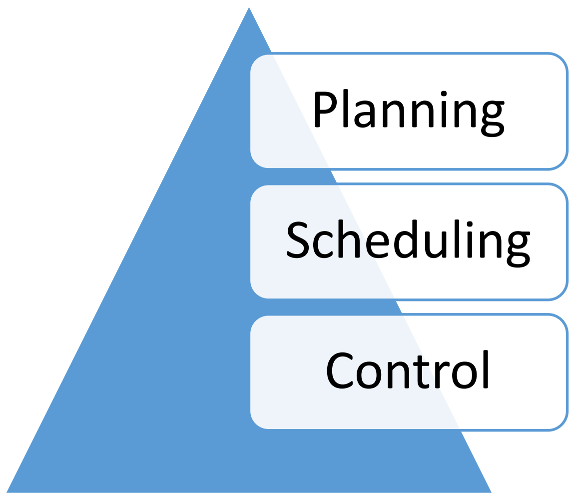


Figure 1 Common Enterprise-Wide Optimization Hierarchy (Grossmann, 2012)

Traditionally, tackling the optimization of large systems was mostly done using linear programming (LP) models. Those models are typically large and are used to answer high-level questions relating to the operation of those plants and networks. The use of rigorous or nonlinear models was mostly reserved for subsystems of the supply chain. This is mostly due to the significant added complexity of optimizing larger systems using rigorous approaches (Grossmann & Biegler, 1995).

More recently, nonlinear programming (NLP) models have started to gradually replace LP models. This is mainly attributed to the advancement of efficient optimization algorithms and computing power (Grossmann & Biegler, 1995).

NLP models can overcome several problems associated with LP models. NLPs allow delivering solutions that are more accurate and therefore implementable in the field. This is in contrast with the often-inaccurate solutions provided by LP models (Alhajri et al., 2008).

2.5 Case Study Brief

In this section, a brief is provided to describe the assets being optimized in this work.

Gas Oil Separation Plants

Unlike many privately-owned oil majors, most state-owned oil companies have the leverage and complexity of operating convoluted supply chains. This starts from exploring and developing oil and gas fields. Oil is sent from wells to Gas Oil Separation Plants (GOSPs). The purpose of those plants is to separate the three-phase feed into gas, water and oil. Gas is compressed and sent to gas plants for further processing while oil is sent either to refineries or export terminals. Water is re-injected to the reservoir to maintain reservoir pressure and improve oil recovery.

GOSPs are often constructed as networks that have swing lines interconnecting nearby plants. This is intended to allow for improved flexibility and better optimization capability. Such added flexibility makes the problem suitable for optimization.

Gas Plants

Gas is mainly received from two sources: either GOSPs or gas fields. Gas from the former source is termed 'associated gas' since it is produced as a by-product of oil production. Gas from the latter source is termed 'non-associated gas'. These gases, in addition to condensate (often termed 'wet gas') are sent to a network of gas plants. The system allows a certain degree of flexibility to swing gas amongst nearby gas plants.

A Gas plant (often termed 'LNG plant') consists of multiple units that aim to treat and maximize the value of the feed stream. Typically, gas is first received in slug catchers which mainly reduce slugging impact and separate heavy hydrocarbons using flash separators. Gas is then sent to acid gas recovery units (AGRs). These units consist of amine contactors which absorb acid gases. The rich amine is regenerated and the acid gas is sent for further processing. Based on its composition, the acid gas is either sent directly to Sulfur recovery units (SRUs), which are typically Claus based units, or sent to acid gas enrichment (AGE). Enrichment units are used for acid gases with high CO2 content. They utilize amine-based absorption to improve the quality of SRUs feed. This entails selectively absorbing H2S while slipping a CO2 rich stream. SRUs use a form of Claus based set-ups to recover sulphur. The aim here is to minimize SOx emissions, which have a high environmental impact. After sweetening in AGRs, gas is sent to dehydration units to strip out water. In gas plants where the feed is lean (containing a small C2+ fraction), the gas is sent directly after dehydration to the sales gas network. This is a vast network which includes many supply and demand points. Gas in this network is supplied to power plants and industrial complexes and is mostly used for power generation. In gas plants with rich feed (high C2+ fraction), gas from dehydration is sent for NGL recovery. Using a variety of recovery technologies, which are chosen based on the expected C2+ content, a C2+ stream is recovered and sent to NGL fractionation plants, where it is separated to C2, C3, C4 and C5+. Light gases from the NGL recovery are compressed and sent to the sales gas network.

The ability to swing feed to different gas plants, the differing feed gas composition based on its source, and the number of controlled variables within the gas plants lend this problem very well to optimization. Additionally, although there is no control on associated gas rates, it is possible to optimize the non-associated gas feed.

Refineries

Oil from GOSPs is sent for further stabilization then is either sent to export terminals or refineries for distillation and other operations that aim to maximize the value of the oil's fractions. This includes hydrotreating, catalytic reforming, hydrocracking, visbreaking and other operations. Refining will not be discussed in detail as it is outside the scope of this thesis.

As can be seen, there is a significant scope for optimizing Midstream networks owing to their flexibility, numerous controlled variables and operational constraints. The case study used for this thesis is inspired by a real-world network which provides an elaborate example of Midstream integrated plants, where the goal is to drive the entire network towards an optimal solution.

2.6 Benefits of Optimizing an Elaborate Midstream Network on a Near-Real-Time Basis

As previously stated, researchers have mostly adopted a hierarchy for Enterprise Wide Optimization that segments the optimization process based on its time horizon. This includes planning, scheduling and control (Shah, 1998), (Shah, 2004), (Grossmann, 2009), (Grossmann, 2012), (Mokhatab et al., 2019).

This thesis proposes to employ a strategy for near real time optimization (N-RTO). This layer is presented as having similarities of both scheduling and control. It is mostly applicable to the central real-time optimization of sizable networks.

In N-RTO, central operators are tasked with re-optimizing the network's operation to respond to plant disruptions, imbalances between supply and demand or to simply improve the network's economics. Like control, the purpose here is to optimize real-time operation without significant considerations of a future horizon. However, unlike control, there is no need for dynamic modelling or communication with MPCs. There is also no need to ensure process steady state ahead of running an optimization, which is the case for control when no dynamic modelling is used. Due to the significant network size and multiple plants, ensuring a steady state may not be achievable.

The nature of centralized N-RTO requires frameworks that strike a unique balance between solution accuracy and run-time. Operators expect to arrive at an answer in a short time to respond to operational requirements. Concurrently, they expect the answer to be sufficiently accurate and implementable. These requirements demand devising techniques and frameworks that allow meeting them. Such an approach may depend on the network's size and the complexity of the underlying processes. Ideally, the modelling approach will involve as much fidelity as possible to meet accuracy requirements. Fidelity will only be reduced by using other techniques once run time is too long. The goal of this thesis is to address Midstream processes and devise a framework that would be fit-for-purpose. The benefits are expected to be as follows:

- **Meeting Demand:** For state-owned companies, meeting local demand and export commitments is a top priority, which precedes profitability. The use of N-RTOs is expected to allow meeting the demand of the different hydrocarbons within the given constraints (assuming the problem is feasible). This is more complex than it first appears. For example, assume a case when the network is required to meet the sales gas demand. There is little flexibility in producing above or under the requirement since the piping network is the only place for inventory. Now assume that there is an increased demand for propane. It is now important to meet this demand keeping in mind the demand for other products and the constrained logistics.
- **Maximizing Profitability:** While meeting demand is essential, it is important to accomplish this while maximizing profitability. For example, it is easy to meet an increased demand for sales gas by relaxing NGL recovery and therefore slipping heavier fractions to the top of the demethanizer column at gas plants. However, this would lead to losing ethane, which is a key feed to petrochemical complexes.

In companies that own petrochemical plants, this is a detrimental operation since it would result in substantial loss of profitability.

- **Environmental Compliance:** Midstream N-RTOs can be essential for reducing environmental impact. Running the network in a manner which ensures that feed is allocated appropriately and operational variables are set optimally can result in reducing damaging emissions. For example, this can be accomplished by improving AGE's selectivity and therefore SRUs performance. A number of researchers have approached coupling environmental and economic mathematical modelling by designing elaborate schemes for multi-objective optimization (Martins & Costa, 2010), (Tautenhain et al., 2019). Others have also devised novel techniques to reduce the computational burden resulting from these formulations by systematically reducing the number of objectives (Guillén-Gosálbez, 2011), (Pozo et al., 2012). However, this is outside the scope of this thesis. In this work, a penalty associated with emissions shall be added to the objective function.

2.7 Research Objectives

To the best of the author's knowledge, there is no work in the open literature which proposes a framework for the N-RTO of Oil & Gas Midstream processing networks. The objectives of this research are to fill this research gap.

More specifically, this work shall address:

1. Using data inspired from the real world to develop a framework for the optimization of Midstream networks.
2. Exploring techniques and approaches that allow meeting the functional requirements of N-RTO. This includes exploring first principles modelling for simpler parts of the network. The use of surrogate-based modelling and semi-parametric modelling will also be explored to meet functional and run-time requirements for complex parts of the network.
3. Developing a mathematical formulation that enables maximizing profitability while meeting products demand.

The initial results presented in this thesis shall focus on developing a framework for the rigorous optimization of a network of GOSPs using mixed integer nonlinear programming (MINLP).

2.8 Thesis Outline

Having introduced the research topic at hand, Chapter 3 will provide a review of literature which is relevant to this work. This shall cover optimization techniques, modelling approaches and the state-of-the art around EWO.

Chapter 4 shall describe the Midstream network, its technical details and challenges.

Chapter 5 will provide a description of the proposed methodology.

Chapter 6 is to illustrate initial results that were achieved after tackling a part of the problem.

3. Literature Review

The near real time optimization (N-RTO) of Oil & Gas Midstream networks requires careful consideration of suitable modelling techniques and optimization approaches. These networks can vary in size and complexity and therefore demand varying mathematical structures.

There is extensive literature covering the areas of supply chain and Enterprise-Wide Optimization (EWO). However, there is no research covering the specific optimization of Midstream networks on N-RTO basis. This area presents its own set of unique challenges with benefits that make it sufficiently attractive to address.

This chapter shall tackle this problem by establishing three cornerstones in this review. Firstly, it will provide a background on optimization and the algorithms used for optimizing large systems. Then, it will address the possible modelling approaches, which shall be classified into parametric, non-parametric and semi-parametric. Finally, it will provide an overview of the major past efforts which tackled EWO problems.

These cornerstones shall pave the road to systematically explore this topic and decide on suitable methods to address the associated challenges.

3.1 Mathematical Optimization

3.1.1 Introduction to Optimization

An optimization problem can be generally formulated using mathematical relationships which consist of a set of equality and inequality constraints and an objective function. A set of variables are manipulated to serve the purpose of minimizing or maximizing the objective function. There are many variations to this formulation such as use of multiple objectives. The general form of the mathematical model is as below (Grossmann & Biegler, 1995):

$$\textit{Minimize } f(x)$$

Over $x \in \mathbb{R}^n$; subject to:

$$h(x) = 0$$

$$g(x) \leq 0$$

Where *f(x)* is the objective function. Additionally, *h(x)* and *g(x)* are the equality and inequality constraints, respectively.

A maximization problem can also be converted into the general form of a minimization problem by minimizing its negative value. Moreover, based on its structure, an optimization model can be classified into various categories. Common categorization can be based on the factors below:

- **Variable Type:** Variables can be either continuous or discrete. A problem belongs to discrete optimization even if some of the variables are in the form of integers (Biegler & Grossmann, 2004).
- **Variable Linearity:** A problem can be classified as linear if the objective function and constraints are all linear. Nonlinear problems can be further classified into convex and nonconvex. In the former a local optimum is always a global optimum. In the latter, there might be several local optima (Biegler & Grossmann, 2004).
- **Constraint availability:** A mathematical model can be either constrained or unconstrained. A varied design may be achieved by reformulating optimization problems to include some constraints as penalty functions in the objective function (Yeniay, 2005).
- **Number of objectives:** An optimization problem can have zero, one or multiple objectives. A zero objective problem is intended to determine the problem's feasibility given its constraints. Most optimization problems have a single objective. Problems with multiple objectives are often reformulated to have a single objective by weighing other objectives into the main one (Gennert & Yuille, 1988) or by formulating additional constraints. A property which makes multi-objective optimization unattractive to N-RTO is the Pareto-optimal solutions (Deb, 2005), which provide infinite optimal equilibria, making this approach impractical for this field.
- **Parameter Certainty:** An optimization problem can be either deterministic or stochastic. Stochastic optimization allows incorporating the uncertainty of some parameters into the optimization problems. This could represent the volatility in the price for products. This form allows finding solutions that are feasible for all or most of the expected variations (Diwekar, 2008).

The subsequent section will address variables type and linearity. Addressing unconstrained, multi-objective or stochastic optimization is outside the scope of this work.

3.1.2 Optimization Approaches

The size of the system being evaluated often dictates the type of appropriate formulation. System size can vary based on the number of variables and constraints. Variable type and linearity also have significant impact on formulation. For example, until recently, problems with 100 or more variables were considered large for NLP. Similarly, until recently MINLP problems used to be considered unsolvable (Grossmann & Biegler, 1995).

Direct Search Optimization

Optimization problems can be solved by employing direct search algorithms, without having to evaluate derivatives. These techniques are regarded as particularly effective in solving combinatorial optimization problems, such as the Traveling Salesman Problem (TSP) (Grefenstette et al., 1985). They typically involve performing many function evaluations to arrive at a solution. The overall strategy is based on exploitation, exploration or a combination of both. Popular direct search techniques include:

- Hill Climbing: This algorithm is designed such that a random solution is picked. This is followed by making a local move and re-evaluating the function. If it is better, the new solution is accepted and a new move is made. If it is worse, the original solution is kept and a new local move is made. This is repeated until no better move is available (Talbi & Muntean, 1993).
- Simulated Annealing: This algorithm is a modification on hill climbing. The modification is intended to improve the chance of arriving at or near a global optimum as opposed to being trapped in a local one. Initially, even if a move is found to be bad, it may be accepted based on an assigned probability that decreases exponentially with how bad this move is. As the 'temperature', of the system decreases, it becomes less likely to accept a bad move and therefore tries to focus the search (Talbi & Muntean, 1993).
- Genetic Algorithm: This technique is inspired by Natural Selection. The process starts by creating an initial population. This is followed by evaluating its members and selecting fitter individuals while discarding others. Crossover is achieved by combining traits of selected parents. Mutation is performed to make small changes to the genome in order to add a level of randomness (Talbi & Muntean, 1993).

These techniques are generally easy to apply. However, they are considered better suited for unconstrained optimization problems and are highly dependent on the algorithm parameters (Grossmann & Biegler, 1995). Their performance also degrades rapidly as the number of variables increase and are rarely applied with more than a few dozen controlled variables (Biegler, 2004).

Linear Optimization

There several advantages to LP formulations, which facilitated their wide adoption by the industry, particularly in evaluating sizable models. Firstly, in LP a locally optimal solution is guaranteed to be a global one. This optimal solution also lies at a vertex within the solution space. These attractive properties allowed linear models to dominate operations problems, such as those concerned with planning, scheduling and supply chain optimization (Biegler & Grossmann, 2004).

There are two main techniques to evaluate LP models, namely: the simplex and the interior point algorithms.

The Simplex method was developed by George Dantzig in the 1940s and is the most widely used LP algorithm until today. It efficiently goes through a sequence of testing adjacent vertices within the feasible space. The algorithm moves to a new vertex only if the objective function is either improved or remains the same (Nocedal & Wright, 2006). Other than specially formulated examples, the method usually solves in polynomial time (Spielman et al., 2004).

As opposed to the Simplex method, the interior point algorithms traverse the feasible region to find an optimal solution. John von Neumann first proposed this technique in the 1960s. However, it was not until 1984 that this method was popularized through the modifications proposed by Karmarkar's algorithm which made this technique efficient and solvable in polynomial time (Nocedal & Wright, 2006).

These algorithms can be highly effective. Biegler et al. reported that recent solvers can efficiently solve problems with millions of variables and constraints. Indeed, using decomposition techniques, it is possible to solve problems that are even 3 order of magnitudes larger (Biegler & Grossmann, 2004).

LP problems are mostly solved using the powerful CPLEX, XPRESS and GUROBI commercial codes (Meindl & Templ, 2012).

Discrete Linear Optimization

The inclusion of discrete or binary variables transforms LP problems into mixed integer linear programming (MILP) problems. This entails employing different techniques to evaluate those functions, which greatly complicates the solution process. Grossmann et al. reported that it can be demonstrated that these problems are in fact NP-complete (Grossmann & Biegler, 1995).

Branch and bound (Dakin, 1965) is the most common approach here. It successively breaks the optimization problem into a decision tree and solves a relaxed LP sub-problems at each node of the tree.

Initially, the original optimization problem is solved as a relaxed LP. If fathomed, the process ends. Otherwise, for the unfathomed sub-problem, an integer branching variable is chosen, which does not have an integer value in the relaxed solution. Two branches are then created by adding an upper and lower constraints. For example, if the chosen integer x has a value of 4.6 in the relaxed LP, the two branches shall have restrictions where $x \leq 4$ in one branch and $x \geq 5$ in the other. Then, for each branch, an LP relaxation is solved. The branch is considered fathomed if:

- Its objective function is more than an existing feasible solution for a minimization problem (the opposite for maximization)
- Its LP relaxation is infeasible
- The solution has integer values for all integer variables

The process is stopped when all sub-problems are fathomed and the lowest objective function (in a minimization problem) is then considered optimal.

Several techniques have emerged in modern solvers to reduce the computational burden of evaluating mixed integer models. Presolve techniques often greatly tighten these formulation and therefore improve numerical stability or detect infeasibilities. For example given that variables x1 and x2 are integers, the constraint ($x1 + x2 \leq 0$) can allow inferring by non-negativity that both x1 and x2 equal 0. On the other hand, the constraint ($x1 + 3 \geq 6$) makes the problem infeasible and it can be determined as such prior to evaluating the objective function (Mahajan, 2010).

The use of cutting planes has greatly reduce MILP solve time by iteratively adding constraints that move the algorithm towards the optimal solution.

MILP problems are mostly solved using CPLEX, XPRESS, OSL and GUROBI commercial codes (Meindl & Templ, 2012).

Nonlinear Optimization

In nonlinear programming problems (NLP), a locally optimal solution can be defined using the Karush-Kuhn-Tucker (KKT) conditions (Kuhn & Tucker, 1951).

In convex optimization where both the objective function and constraints are convex, a local optimum corresponds to a global one. However, this is not the case when the problem is non-convex (i.e. the Hessian of its Lagrangian is negative definite).

For the function described below:

$$Minimize\ \ f(x)$$

Subject to:

$$h_i(x) = 0$$

$$g_i(x) \leq 0$$

assuming that the objective function f(x) and the constraints $h_i(x)$ and $g_i(x)$ are continuously differentiable at point x*, there exists parameters μ_i *and* λ_i if a point x* is a local minimum such that:

A. Stationarity condition:

$$-\nabla f(x^*) = \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{j=1}^{l} \lambda_j \nabla h_j(x^*)$$

B. Primal feasibility

$$g_i(x^*) \leq 0\ ,\ i = 1, \ldots, m$$

$$h_i(x^*) = 0, j = 1, \ldots., l$$

C. Dual feasibility

$$\mu_i \geq 0, i = 1, \ldots, m$$

A. Complementary slackness

$$\mu_i g_i(x^*) = 0, i = 1, \ldots, m$$

The major algorithms for solving constrained nonlinear optimization problems have been the reduced gradient and sequential quadratic programming (SQP) techniques. It was also reported that while reduced gradient is often suited for NLP problems with mostly linear constraints, SQP is rather superior with highly nonlinear problems (Grossmann & Biegler, 1995).

SQP is based on iteratively modelling and solving NLPs using quadratic programming sub-problems to approximate the objective function and the constraints. This is in addition to applying Newton's method to the KKT conditions. It has been shown to be highly successful in solving NLPs as it requires the fewest function evaluations. This efficiently leads to fast convergence (Binder et al., 2001).

By applying the newton method to the KKT conditions at an iterate point x_k, the minimization objective function is replaced by the approximation:

$$f(x^k) + \nabla f(x^k)(x - x^k) + \frac{1}{2}(x - x^k)^T Hf(x^k)(x - x^k)$$

Similarly, the constraints are replaced by their approximation:

Inequality Constraint:

$$g(x^k) + \nabla g(x^k)(x - x^k)$$

Equality Constraint:

$$h(x^k) + \nabla h(x^k)(x - x^k)$$

by setting: $$dx = x - x^k$$

the below quadratic approximation of the original problem is deduced, noting that the term $f(x^k)$ may be removed since it is a constant.

$$\text{Min } \nabla f(x^k)dx + \frac{1}{2}dx^T Hf(x^k)dx$$

Subject to:

$$g(x^k) + \nabla g(x^k)dx \leq 0$$

$$h(x^k) + \nabla h(x^k)dx = 0$$

Much progress has been made in developing global optimization techniques for non-convex models with special structures. However, proving convexity for this type of problem is very challenging and often not possible. Practically, settling for local optimality is often considered a reasonable outcome (Grossmann & Biegler, 1995).

The use of stochastic methods that are mainly based on direct search techniques have also been applied to this domain. However, these are typically applied to models that are of inexpensive nature and are not highly constrained (Grossmann & Biegler, 1995).

As for deterministic global optimization approaches, Floudas (2013) reported that significant progress has been made in developing such techniques to solve important and challenging problems. This includes addressing general classes of NLP and MINLP problems. He listed the approaches below as the main ones for deterministic global optimization (Floudas, 2013):

- Branch and Bound methods
- Cutting Plane methods
- Primal-Dual Decomposition methods
- Outer Approximation methods
- Inner Approximation methods
- Difference of Convex and Reverse Convex methods
- Reformulation-Linearization methods
- Lipschitzian methods
- Trajectory and Homotopy methods
- Interval Analysis methods

Discrete Nonlinear Optimization

Biegler & Grossmann (2004) reported that MINLP combines the difficulties associated with MILPs and NLPs as described above. This includes the combinatorial features of the mixed integer portion and the nonconvexities often associated with nonlinearities. These problem classes are most often solved by the general branch and bound methods. Similar to MILPs an alternating sequence of solving the master mixed integer and the sub nonlinear problems is employed. Similar to NLPs, a globally optimal solution can only be guaranteed in convex problems (Grossmann & Biegler, 1995), (Biegler & Grossmann, 2004).

The Outer Approximation Equality Relaxation Augmented Penalty (OAERAP) method is also a popular technique for solving MINLP problems. This technique is also employed by gPROMS, which is the software used for performing the modelling and optimization for this problem. OAERAP is a modified version of the Outer Approximation/Equality Relaxation (OA/ER) method (Viswanathan & Grossmann, 1990). When the optimization problem is convex, this algorithm can guarantee global optimality.

The OAERAP algorithm applies decomposition techniques to break the problem into a master MILP problem and a series of NLP sub-problems. It starts by solving a relaxed NLP problem of the integer or binary variables. At this stage binary or mixed integer variables can be assigned continuous values. This allows obtaining an initial iteration to move to the next problem. The MILP master problem then aims to find integer values which feature an augmented penalty function to reach the minimum over the convex linearized function. An NLP is then solved to find the optimum value of the continuous variables while fixing the discrete variables to their pre-determined values. Then, the gradient is calculated based on the linearized functions. The solver then determines if optimality was satisfied or whether an additional iteration is needed.

A global optimum cannot be guaranteed because of the linearization being applied to non-convex functions. Due to the high nonlinearity of the GOSP optimization problem, arising from various physical separation and compression equations, the solution may find difficulty arriving at a global optimum. This was often addressed by changing initial guesses. Indeed, a good initial guess substantially improves the solver's chances of arriving at a global optimum.

3.2 Process Modelling

A process model can be classified in many different ways based on the purpose for which it was built. However, this section will focus on exploring the classification of models based on their generation methodology. This shall mainly address the extent to which a model is generated based on first principles versus being data-driven. This is a key component of this research since it will allow defining how a Midstream network model is to be built to meet the functional requirements of N-RTOs.

Classifying models based on the above criteria allows identifying three categories, namely: parametric, non-parametric and semi-parametric models.

Parametric models (PMs), otherwise known as first principles or white-box models, rely primarily on process mechanisms. They are based on the well-established principles of mass, energy and momentum conservation, in addition to thermodynamics, reaction kinetics, fluid flow, mass transfer and others. They can be mostly developed without relying on plant or experimental data. As such, they incorporate a priori knowledge of the process, its performance and the interactions amongst its components. They perform in accordance with their embedded assumptions, which may at times be invalid. It is indeed often difficult to match these models to actual data due to their high dependency on such assumptions. Indeed, many such models are also dependent on empirical correlations which were generated using experimental work. This includes Fick's law of chemical diffusion, Fourier's law of heat conduction and Darcy's law of flow in porous media (Zendehboudi et al., 2018). Indeed many laws that are perceived as first principles, rely significantly on experimentally deduced coefficients. However, these models still seek to establish a deep understanding of the process. It is also often possible to extrapolate well with these models and generalize them to different environments if their underlying assumptions remain valid.

Non-parametric models (NPMs), otherwise known as data-driven or black-box models, rely principally on available data. They can be generated using a variety of techniques that map input with output data without incorporating a priori knowledge of the process. Due to their generally simpler nature, these models often lend themselves well to optimization and allow construction of computationally inexpensive models. Since these models require availability of data, they are typically used for process optimization and control rather than design or debottlenecking. This also makes them restricted to the range of data over which they were generated. This makes them suited for data interpolation but significantly impaired in extrapolation. The structure of NPMs is typically based on neural, radial basis, wavelet, hinging hyperplanes and fuzzy models (Sjöberg et al., 1995).

Semi-parametric models (SPMs), otherwise known as grey-box or hybrid models, combine properties of both modelling approaches addressed above. They were first introduced in the 1990s by Psychogios et al., Thompson & Kramer and a few others (Psichogios & Ungar, 1992), (Thompson & Kramer, 1994). These models have been extensively used to model systems that exhibit parts which can be well defined by PMs and parts which are rather difficult to define using first principles. Accordingly, the combination of both modelling techniques often offers attractive properties. Indeed, some researchers recommended use of PMs as much as possible within SPMs structure to "impose rigor and discipline on purely data driven models" (Venkatasubramanian, 2019). This area has also grown in terms of maturity with 5-20 papers being published annually in peer-reviewed journals (Glassey & von Stosch, 2018).

Common NPM Models

This section seeks to briefly review common NPM models found in the chemical engineering literature. These are also often used as submodels within SPMs. Their properties shall enhance identifying proper modelling techniques for N-RTOs. Popular NPMs which will be explored are Artificial Neural Networks (ANNs), Support Vector Machines (SVMs) and Fuzzy Logic (FL).

Artificial Neural Networks (ANNs)

ANNs were first described and studied by McCulloch and Pitts in the 1940s.Their idea was to mathematically model a perceptron which mimics biological neurons. It receives inputs and calculates output based on weights which are associated to those inputs. After multiplying the individual inputs by their weights, their summation is computed and passed to an activation function. This allows determining if a neuron should produce an output based on passing the activation function's threshold (McCulloch & Pitts, 1943).

An activation function can be a simple step formula, but is often modelled as a nonlinear sigmoid, tangent hyperbolic, or other functions. ANNs popularization was in part a result of Werbos PhD thesis in the 1970s, in which he proposed the backpropagation algorithm (Werbos, 1974).

For a single neuron calculation, the input formula is shown below:

$$Z = \sum_{j=1}^{n} w_j x_j + b$$

where x_j is the input signal, w_j is the assigned weight and b is the bias.

The output signal is then:

$$y = f(z)$$

Nowadays, most ANNs utilize the multilayer perceptron (MLP), which allow adding multiple hidden layers between the input and output. Indeed, as the number of hidden layers increase, the network is generally able to learn more complicated patterns.

ANNs can be very effective for a variety of applications. After a network is trained, it can be used for function evaluation, prediction, clustering and classification. Moreover, their learning can be supervised, unsupervised or using a combination of both. (Jain &

Mohiuddin, 1996). Networks with recurrent features (RNNs) allow for adaptiveness and modelling sequential and time varying patterns, which lends them well to forecasting and dynamic systems.

Challenges associated with ANNs include data bias, over-fitting to a particular dataset without being able to generalize to new situations, and hyperparameter optimization. The latter is related to choosing an optimal set of parameters to design the network such as the number of layers and type of activation function. Additionally, using such a black-box method makes it often difficult to interpret their behaviour and verify their performance (Gupta, 2018).

Support Vector Machines (SVMs)

SVMs were first introduced in the 1960s (Vapnik, 1963) although not popularized until the 1990s. They are based on supervised learning and have been shown to be often very effective for certain classes of classification, clustering and regression problems. SVMs are also reported to perform better in terms of not over-generalizing classification problems in comparison with ANNs (Mitchell, 1997)

In its basic form of classifying two categories of data that are linearly separable, the concept is based on placing a hyperplane that divides the data with a maximum margin from both categories. Additionally, SVMs are highly impacted by the choice of support vectors which are lines that intersect chosen points within both categories. Support vectors are chosen, such that when the hyperplane is placed, it has the maximum margin from both categories. SVM classifiers can be more complex and can involve the use of soft margins and non-linear classifiers. (Jakkula, 2006).

Considering the simple binary classification set with sample points (x_i, y_i), where i = 1, ..., n, where x_i represent the training input and y_i represent the output. yi corresponds to +1 when the result belongs to the first class and -1 when it belongs to the second. The SVM classifier is therefore defined as (Cortes & Vapnik, 1995):

$$f(x) = sign(w.x + b)$$

Where w and b are the weight vector and bias, respectively. These values are determined by fitting the above equation to the training data.

In cases where data is not linearly separable, SVMs can be used to classify data using functions in kernels to classify data in a higher dimensional space. Popular transformation kernels use polynomial, sigmoid or Gaussian functions. The approach aims to reach a decision space with the least possible Vapnik-Chervonenkis (VC) dimension and training error such that the generalization performance is improved (Agrawal et al., 2003).

Fuzzy Logic (FL)

Fuzzy Logic theory enables taking account of inaccuracies and uncertainties by considering 'degrees of truth' to allow a condition to be specified in a 'grey' state other than the Boolean true or false. Introduced by Zadeh in 1965, it provides attractive properties to non-crisp classification. The idea behind it stems from human reasoning, which may be too difficult to model through binary algorithms.

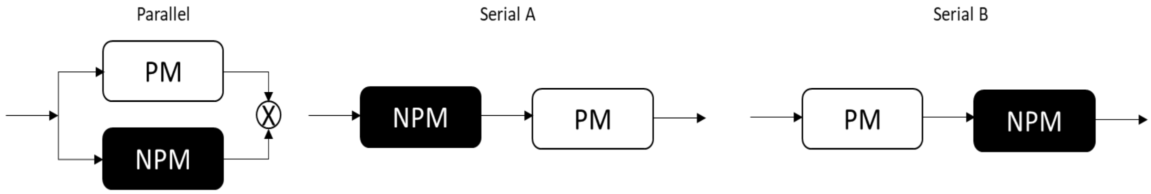
In Fuzzy Logic, crisp input data is first 'fuzzified' using membership functions which maps input to membership values that ranges from 0-1 based on its degree of truth. Fuzzy reasoning is then applied which uses a variety of fuzzy rules, operators and conditions. A process of defuzzification is then applied to output crisp values (Hellmann, 2001).

Semi-Parametric Models

The use of semi-parametric models is considered attractive since it balances the advantages and disadvantages of using full parametric or non-parametric models. In their pioneering work in this field, Psichogios and Ungar applied the first reported SPM to a fed-batch fermentation reactor. They reported that a semi-parametric structure of ANNs and first principles models resulted in a model which was more capable of predicting the process state. It was additionally capable of interpolating well and most importantly extrapolating. It also overcame the difficulty of interpreting models which are exclusively non-parametric. In addition, the model appeared to overcome the phenomena of overfitting, which is commonly associated with ANNs (Psichogios & Ungar, 1992).

Engineering research, and in particular process systems engineering, benefited greatly from the application of SPMs, which was pioneered in the 1990s also by Su et al. (Su et al., 1996), Kramer and Thompson (Thompson & Kramer, 1994) and Johansen and Foss (Johansen & Foss, 1992).

There is little guidance in the open literature to establish techniques for modelling SPMs to serve the variety of possible purposes. Available work has been mostly based on the researchers' expertise to design architectures that are fit for purpose. There are several techniques that were reported for the construction and arrangement of SPMs. These structures can broadly be classified into parallel and serial as shown below.



Parallel SPMs

Parallel SPMs have received less attention in the literature than their counterparts. They have mostly been used in situations where the performance of the parametric model is not satisfactory for a subset of the outputs. Such errors can result from invalid assumptions, linearization or not representing certain impactful parameters. The PM's output errors is then estimated using the NPM portion, which then acts to adjust the PMs estimations (von Stosch et al., 2014).

This structure is mostly suited when it is possible to uncouple certain effects by a separate model. Those effects can then be adjusted in this architecture (Su et al., 1996).

There are several techniques to capture the error compensation. A popular method is the pure superposition technique which adds a residual term to the PM's output. This residual term represents the mismatch between the PM calculations and real data (von Stosch et al., 2014).

Serial SPMs

A more popular arrangement for SPMs is using a serial structure. These structures are typically used when a part of the first principles model is not available or is too complex to describe. Complexity can arise either from experimental or computational burden.

In serial SPMs, certain parts which are well established or easily derived are modelled as parametric. This may include mass, momentum and energy conservation laws.

On the other hand, the NPM portion includes parts that are not easy to capture accurately such as reaction kinetics, friction factor, thermodynamic and heat transfer parameters.

The arrangement which is mostly popular in chemical engineering literature is the one tagged as Serial A above. The availability of sufficient process data allows the establishment of a non-parametric model, which estimates certain parameters and provide an input to the parametric model. A popular use is in employing NPMs to estimate reaction parameters then feeding those to PMs for further processing (Glassey & von Stosch, 2018).

The arrangement tagged as Serial B is less popular in the process systems engineering literature. It is often used where data is not sufficiently available to train NPMs (Zendehboudi et al., 2018). The author believes the structure may be well suited for situations where critical parameters are not directly measured in the field, such as oxygen concentration in amine regeneration reboilers. This parameter is impactful to the unit's control and optimization since it is a corrosion precursor.

Important to note is the fact that the efficacy of such modelling arrangements is dictated by the modelling approach and accuracy of the underlying models. In some cases, use of an NPM may be more effective than SPM if the parametric portion lacks reasonable performance (Zendehboudi et al., 2018).

There are many other structures for SPMs reported in the literature. This includes building multiple NPMs, PMs or having a mixture of parallel and serial structures. An interesting use of SPMs is to employ multiple NPMs, where one NPM performs pretreatment and conditioning of data ahead of feeding them to subsequent models. Researchers identified significantly improved performance by using clustering, classifying and filtering techniques to address data noise and uncertainty (von Stosch et al., 2014).

Surrogate Modelling

There are often many benefits associated with substituting a simulation model with its simplified surrogate. Oftentimes, the modeller aims to optimize a derivative free system with minimum black-box calls. In N-RTOs, this system can be a rigorous simulation model describing multiple plants. For this system, derivatives are often not available or are very expensive to estimate. Accordingly, it may be prohibitive to call such model iteratively during the optimization process. On the other hand, a surrogate model can provide a satisfactory accuracy with a significantly shorter run time.

Surrogate modelling is a mature research area with a significant amount of literature. The aim in this review is to briefly describe popular techniques, which are enablers to this work.

Assuming the original model is represented by:

$$y = f(x)$$

The surrogate model can be written as:

$$y^* = g(x)$$

such that

$$y = y^* + \varepsilon$$

where ε represents the error resulting from approximating the original function.

Given the computational burden of running a vast number of simulations to generate surrogates, it is often necessary to optimize the generation process. This process starts with defining an efficient design of experiments (DOE). The design aims to identify a variable design space and a sampling plan to efficiently obtain sample points (Forrester et al., 2008).

Firstly, the input and output variables are determined through multiple steps that involve variable analysis, classification, reduction and identification (Simpson et al., 2001). Optimizing the number of variables is essential due to the exponential increase of required sampling points when the problem increases in terms of dimensionality. A sampling method is then selected. A number of methods are available such as simple random sampling, Latin hypercube and orthogonal Latin hypercube (Forrester et al., 2008).

A variety of techniques can then be utilized to construct the surrogate model. This includes polynomial regression, decision trees and random forests, symbolic regression, kriging and ANN (Bartz-Beielstein et al., 2016). It is then often necessary to improve weak areas of the surrogate design by methodically adding infill points (Forrester et al., 2008).

Ibrahim, Jobson, Li and Gosalbez proposed the use of support vector machine (SVM) classifiers to optimize the design of heat-integrated crude oil distillation units. Their approach resulted in significantly reducing the computational expense by filtering infeasible designs from the search space. Their MINLP formulation was solved using a genetic algorithm to minimize annualized cost (Ibrahim et al., 2018).

NPMs and SPMs in the Process Industry

The process industry provides a rich set of challenges and opportunities for optimization. The literature is therefore mirroring these aspects by providing models of a wide variety of classes and categories.

There are indeed many examples in the literature that used NPMs to model and optimize chemical processes.

(Al-Enzi and Elkamel, 2000) presented a feed-forward ANN framework to model the operation of a refinery's Fluid Catalytic Cracking (FCC) Unit. The objective was to be able to predict product rates and qualities given limited feed properties, which included API, sulphur and Watson characterization factor. They reported better predictions than commercial rigorous simulators.

(Zahedi et al., 2006) used an ANN model to simulate a refinery hydrotreater plant. They proposed a radial basis function (RBF) to predict hydrogen demand and product stream's properties using a radial basis function. They tested their method with 7 different feed stocks and reported the superiority of their model in predicting product qualities against a conventional rigorous simulator.

(Zahedi et al., 2008) utilized an ANN framework to predict the quality of the gasoline production from a Catalytic Reformer Unit. Their model was able to accurately capture the unit's performance and led to a 2.38% increase in gasoline production.

(Aminian & Shahhosseini, 2008) used an ANN to predict the fouling in the crude units preheat exchangers. Their technique sought to determine the effect of various parameters on fouling.

(Alhajree et al., 2011) used an ANN to model and study the sensitivity of a refinery's hydrocracker unit. Their sensitivity analysis sought to determine the effect of various parameters on the unit's performance. Finally, they used MATLAB to maximize the yield of gas oil, kerosene, heavy and light naphtha.

There are also numerous examples of SPM models focusing on the process industry.

In reaction engineering, SPMs are usually arranged such that the reaction kinetics are described by the NPMs submodel while conservation laws are described by PMs.

(Bollas et al., 2003) compared the performance of pure ANN and hybrid models to scale up a pilot FCC plant into an industrial size plant. The ANN model performed well in predicting weight percent conversion and coke yield. Nevertheless, the hybrid model provided better extrapolation performance.

(Bollas et al., 2004) studied the coupling of a rigorous and an ANN models to predict the performance of a refinery's hydrodesulpharization reactor. The ANN model was used to describe kinetic parameters. The rigorous model, on the other hand, was used to study the reactor performance and hydrogen consumption.

(Zendehboudi et al., 2014) sought to compare the performance of a FPM against a SPM in modelling and optimizing a urea production plant. Using a rigorous FPM model, satisfactory matching with plant data was achieved, except with regard to CO2 conversion and outlet temperature. Both are important parameters in industrial urea production. They then constructed an SPM with an ANN as the NPM submodel. The ANN was used to estimate the conversion of CO2 as a function of temperature and the feed composition. They reported significant improvement in the SPMs capability of predicting the plant's performance. They also reported a significant improvement in computational time.

There are also various applications of SPMs to separation processes.

(Safavi et al., 1999) developed a SPM with a series structure to model a distillation column. They sought to examine the prediction accuracy of the SPM against the rigorous model. Their results show an excellent agreement with the rigorous model while significantly reducing computational time.

There are also many other applications of other domains within the chemical industry. This included predicting the friction factor of pipelines in turbulent flow regimes (Shayya & Sablani, 1998) and determining mass and heat transfer coefficients for a catalytic fixed bed reactor (Mjalli & Al-Mfargi, 2009).

3.3 Supply Chains & Enterprise-Wide Optimization

In his paper which addressed advances and challenges in the process industry's supply chains, (Shah, 2005) anticipated that the industry will need to further evaluate, report and improve on its sustainability metrics. To achieve this, the industry must, in part, improve on its supply chain planning activities. He categorized the representation of the production process within the industry based on gross margin into 'recipe-based' and 'property-based'. The former, which typically operates on higher margins, is mostly related to the pharmaceutical and food processing industries. The latter generally includes businesses with slimmer margins such as refining and petrochemicals.

(Shapiro, 2006) defines Enterprise-Wide Optimization (EWO) as coordinating the optimization of the various operations within a defined supply chain. This includes activities, such as R&D, receipt of raw materials, processing and distribution of products. The goal is often to minimize costs, inventories and environmental impact, while improving profitability, assets utilization and agility. Researchers have addressed the optimization of a variety of complex supply chains, ranging from pharmaceuticals (Shah, 2004; Papageorgiou, 2001) to refineries (Menezes et al., 2017) and integrated chemical sites (Wassick, 2009).

Many EWO problems can be formulated as MILP models. Those models are often very large. Their size can also be several times larger when considering multiple periods. As most real-world problems involve non-linearities, these were often addressed by introducing new variables and equations to perform piecewise linear approximations or exact linearization. However, this can only be done in limited cases (Grossmann, 2012).

On the other hand, there remains a class of problems which necessitates handling nonlinearities. This leads to MINLP formulations. As stated previously, these are typically solved using various techniques, such as branch and bound methods with NLP solvers at each node. As EWO problems are usually nonconvex, a local optimal solution is often accepted as sufficient outcome. Applying rigorous global optimization techniques, such as the one used in the Baron solver (Sahinidis, 1996) is often computationally expensive and is not practiced for sizable EWO problems (Grossmann, 2012).

(Papageorgiou, 2009) presented an overview of the mathematical models used for the optimization of the process industries' supply chain. This focused on the strategic and tactical level. He addressed the issue of modelling uncertainty using multi-stage stochastic models. He also highlighted the use of multi-objective formulations to address environmental impact. These aspects were also highlighted by (Sahebi et al., 2004) who presented a review of the existing work addressing crude oil supply chains. He also noted the importance of developing efficient algorithms and techniques to address the complexity of these problems.

Researchers have employed various frameworks to optimize oil and gas planning and scheduling models.

(Wassick, 2009) employed a Resource Task Network (RTN), initially described by (Pantelides, 1994) to optimize the scheduling of an integrated chemical site. He showed how an integrated site can be composed of various sub-systems, which perform different tasks. He then used the RTN to optimize the scheduling of the site's waste water treatment.

A framework for the optimization of petroleum supply chains was proposed by (Neiro & Pinto, 2004). It was mostly focused on oil & gas downstream assets including refineries, storage tanks and pipelines. They described a large scale multiperiod MINLP model to optimize the complex topology which consisted of connecting multiple nodes representing each element of the network. Their manipulated variables included flow rates, operational parameters and inventory and facilities assignments.

Several researchers have addressed the optimization of oil & gas midstream operations to various extents.

In his PhD thesis, (Wang, 2003) reviewed oil & gas upstream optimization problems and categorized them into: lift gas and production allocation; processing plants design and operation optimization and reservoir development and planning optimization. He further divided optimization problems based on timescale into operational, tactical and strategic problems.

(Al-Sobhi & Elkamel, 2015) simulated and optimized a natural gas processing network consisting of LNG, GTL and methanol plants. They used the commercial package Aspen Plus to perform the simulation then used an LP formulation to maximize the network profitability.

(Li et al., 2017) proposed a stochastic model for the design and operation of natural gas networks under uncertainty. They applied a modified nonconvex generalized Benders decomposition (NGBD) method to solve it. They started from a network superstructure, with the objective of determining an optimal structure to maximize the system's NPV over a period of 25 years. Using a multi-loop NGBD, they solved each primal subproblem using global optimization. This resulted in reducing the solution time by more than an order of magnitude.

(Li et al., 2007) examined building a network planning tool. The presented network consisted of wellhead platforms to oil and gas export facilities. They used Aspen Hysys to perform process simulation and then inputted the results to a network optimization module. This allowed bypassing nonlinearities.

Liu, Alhasan and Papageorgiou proposed a MILP model to optimize a network of gas-oil separation plants (GOSPs) in the Saudi Arabian Ghawar field. The objective was to ultimately minimize the network's operating expenditure. Their model made use of transfer lines to swing production fully or partially from some plants to others. They represented the network using a state task network (STN), which was described by Kondili, Pantelides and Sargent in 1993, and employed piecewise linearization to handle nonlinear power consumption curves. They reported an average of 12.8% OPEX savings, resulting mainly from reduced power and chemicals consumption. They also noted that due to the unconventional nature of GOSP networks, this problem class was given little attention in the literature (Liu et al., 2016).

3.4 Research Gaps

To the best of the author's knowledge and as demonstrated in the literature review:

- There is no proposed framework in the literature which addresses optimizing oil & gas midstream networks per the requirements of N-RTOs.
- There is no modelling or optimization framework that address introducing NPMs or SPMs for general EWOs or specifically oil & gas midstream EWO. These are expected to be beneficial in improving convergence and minimizing computational cost.
- There is no reported framework in the open literature which utilizes MINLP for optimizing the operation of integrated GOSPs or gas plants.

The objective of this work is to address these gaps and present initial results, to quantify and demonstrate the benefits of the proposed methodology.

4. The Midstream Network

This research examines two types of midstream networks, namely: GOSP and Gas Processing networks. The two networks will not be combined into one holistic network due to their segregated operating philosophies. GOSPs are generally operated to maximize oil production or optimally meet set targets. Gas produced from GOSPs is a by-product of oil production. On the other hand, in oil & gas midstream networks, gas plants are operated to meet set targets for sales gas, which is mostly composed of methane, while maximizing profitability by maximizing NGLs production and minimizing energy consumption. This work will primarily focus on GOSP networks.

4.1 Gas Oil Separation Plants (GOSPs)

As previously stated, GOSPs are often connected laterally via swing and transfer lines which would allow shifting part or all of the production from one GOSP to another. This provides for significant flexibility and room for optimization.

GOSPs often contain equipment items which are highly energy-intensive. This mainly includes low- and high-pressure gas compressors and water injection pumps. Accordingly, optimizing the operation of such equipment can have a significant impact on improving energy efficiency, and consequently further minimizing processing cost and greenhouse gas emissions. It is therefore prudent to target minimizing energy consumption while meeting oil production targets and reservoir management strategies.

Figure 2 shows a simplified block flow diagram representing a typical GOSP. Crude is received at one or multiple high-pressure production trap(s) (HPPT). This is a three-phase separator, which separates high pressure gas, crude and oily water. High pressure gas either free-flows to a nearby gas plant or is sufficiently compressed through a high-pressure compressor to allow overcoming pipeline pressure drop and reaching gas plants that are remotely located from the GOSP at the desired pressure.

Water from HPPTs flows to a water-oil separator (WOSEP). The water is separated and re-injected to the reservoir. This serves two purposes, namely: finding a suitable disposition for contaminated water and maintaining reservoir pressure. Separated oil flows to the LPPT.

The Low-Pressure Production Trap (LPPT) receives feed from the HPPT and the WOSEP. A further pressure drop allows releasing more gases. Those gases are compressed at the low-pressure compressor and then further compressed at the high pressure compressor which also receives feed from the HPPT.

Oil from the LPPT is pumped to the de-salter and then the de-hydrator to remove salt and further remove water. Dry crude is then pumped out of the GOSP.

The configuration of GOSPs can vary significantly. For example, some GOSPs may lack any compressors, while some may lack LP compressors. Similarly, some GOSPs may divert their oily water to other nearby GOSPs for separation and underground injection. GOSPs also vary significantly in terms of the capacity and efficiency of their equipment and the HPPT's operating pressures.

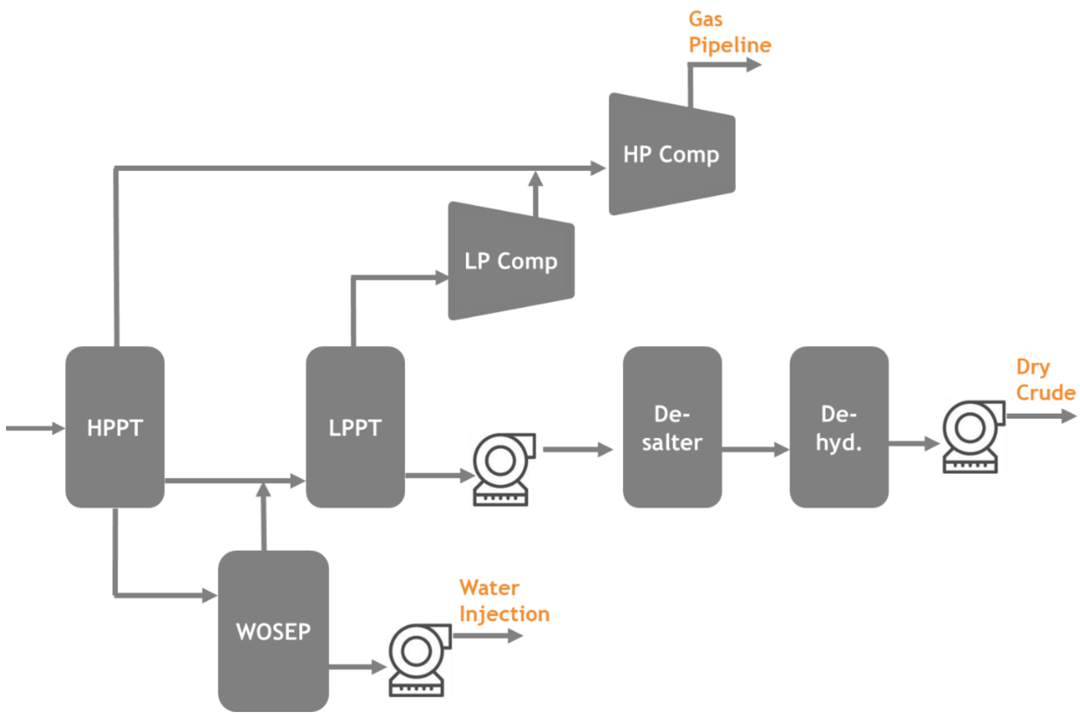


Figure 2 Generic GOSP Representation

The presence of these numerous variables and differing constraints makes operating the network most profitably very difficult due to the thousands of options which are possible. This lends the problem very well to mathematical optimization. A mathematical solver is often able to highlight solutions which are not readily obvious to the network's planner.

Network optimization of GOSPs can also lead to opportunities whereby some production from GOSPs which are shutdown can be recouped by diverting their feed to nearby GOSPs while considering the various limitations such as gas compression and water processing capacities.

Within this work, 2 GOSP networks will be considered, namely Area B and C. Both networks are located within the same reservoir and produce Arabian Light crude. As such, they can be lumped in the same problem formulation towards a single objective function.

Figure 3 shows Area B GOSP network which is connected by swing lines. Some groups of wells are considered swingable while others are not. Feed from a group of swingable wells is typically sent to the primary GOSP, but can be diverted to the secondary one. Feed from non-swingable wells can only be diverted to their primary GOSP. GOSP B-2 is a compression station. Its purpose is to compress the gas from GOSPs B-3 and B-6, both of which contain no compressors.

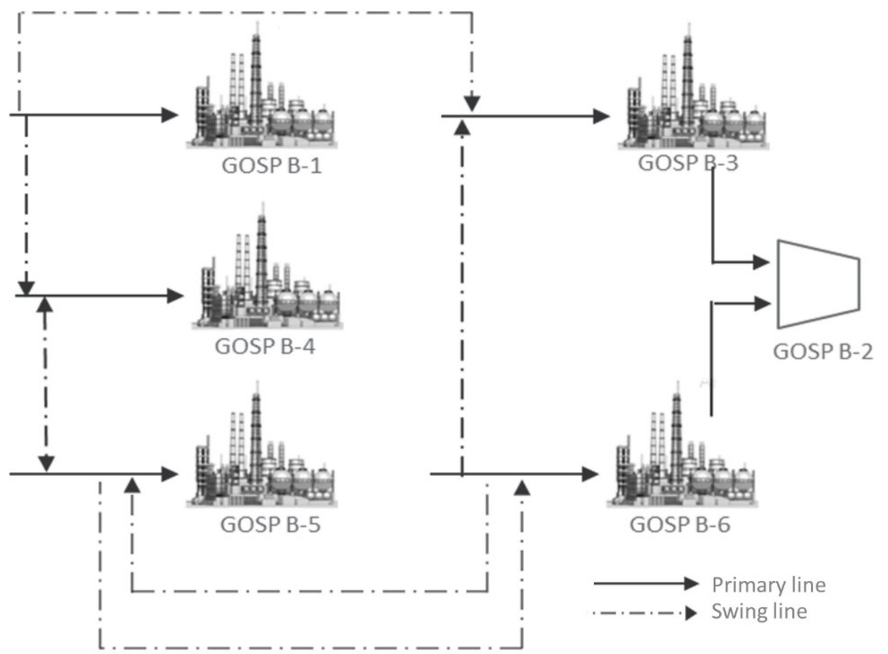


Figure 3 Area B GOSP Network

Figure 4 shows Area C GOSP network, where only a single swing line exists from GOSP C-4 to C-3. This provides a more limited swinging capability.

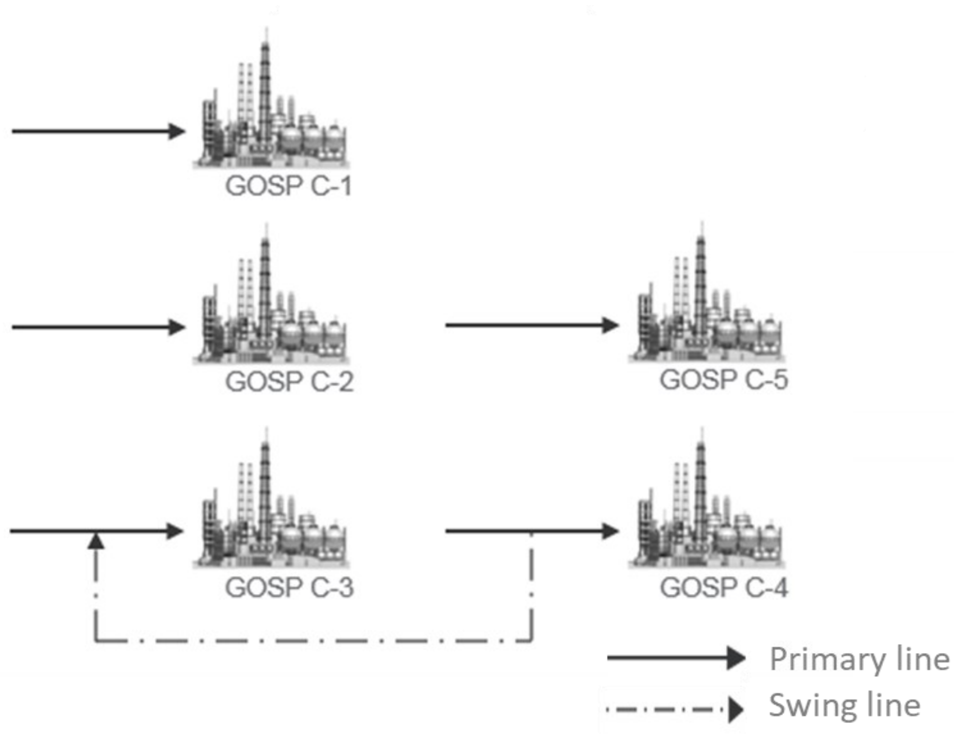


Figure 4 Area C GOSPs Network

5. Methodology

5.1 gPROMS ProcessBuilder

The software which was used to build both the simulation models and the optimization formulations is gPROMS ProcessBuilder v1.3 provided by Process Systems Enterprise ("PSE"). This platform provides capabilities to build high-fidelity models for various applications within the process industry. ProcessBuilder is an Equation Oriented (EO) modelling platform which lends itself to be used for process optimization. Unlike sequential-modular (SM) simulators, EO does not require directionality of computation. Moreover, it allows for efficient handling of multiple recycles, which is essential for optimizing GOSPs, as they contain a variety of recycle streams, such as the water being recycled from dehydrators to WOSEPs. A disadvantage of EO simulators is that their numerical solvers require good initial guesses. Failing to provide those may lead to failures. gPROMS provides capabilities to initialize without the need to preset variables using homotopy-continuation techniques. Nevertheless, EO simulators remain less popular in the process industry due to the general difficulty of providing meaningful diagnostics on failure. This makes debugging difficult and therefore limit the use of those systems to experienced users.

The equation system within a ProcessBuilder model involves parameters and variables. The former can hold an integer or real value. It can also be a foreign object, which allows capturing an external entity such as performance curves of pumps and compressors. On the other hand, for each variable, a variable type has to be created and assigned. Both bounds and default values have to be defined, as well.

To allow running simulations, a *Process* is created and the remaining degrees of freedom are specified. In addition, a solver and solution parameters are selected.

Similarly, to allow running optimization, an optimization entity is created. Within an optimization entity, the objective function, constraints and controlled variables are selected. For controlled variables, bounds and initial guesses are specified. Constraints can include both equality and inequality types.

This section describes the methodology adopted for developing the multi-GOSP optimization model.

5.2 Model Development Workflow

The model development workflow to set up a GOSP optimization application included the below key steps:

1. Technical Specification Document

In the first step, a thorough technical specification document was written and reviewed with the project's proponent. This described the models' functionality, level of fidelity, main inputs, main outputs, optimization problem formulation, assumptions and required data.

2. Collecting & aggregating well data

Each GOSP is connected to one or more groups of oil producing wells, through complicated pipeline network, trunkline and headers. A non-swingable group of wells can only feed a single GOSP. Swingable wells can feed either a primary or a secondary GOSP. In the model, each group is represented by a single feed stream with an averaged gas to oil ratio (GOR) and water cut (WC). At this stage, the models did not include a representation of individual wells as this was deemed outside the scope of this work. A future improvement can add and connect all individual wells to flowlines for added fidelity and higher accuracy in estimating flow conditions and capturing hydraulics.

3. Constructing updated equipment curves

One of the key objectives of this study is to minimize the power consumption. The major power consumers are the compressors and salt water injection pumps. It is therefore essential to construct and embed updated performance curves in the models to accurately estimate power consumption. The performance for the HP, LP compressors and multi-stage salt water injections pumps was simulated and tuned to match current plant performance.

Figure 5 demonstrates the adjustment of the flowrate-head curve for a SWI pump at GOSP B-3. Initially, a simulation model was constructed to match the performance of each equipment at design conditions. The rotating equipment inlet conditions were then adjusted per actual plant performance. This resulted in predicting an equipment performance which sometimes mismatches with expected design performance. A bias was then added to the full equipment curve. At this point, design conditions were used to simulate performance. Equipment performance at the design inlet conditions was then recorded. Once all points were processed, the bias between actual and expected performance was calculated. This was then correlated to equipment flowrate. This residual term represents the mismatch between the PM calculations and real data as discussed in section 3.2. As described in parallel semi-parametric modelling, a pure superposition technique was used to add a residual term to the PM's output. This allowed generating performance curves for the whole range of design data, although available plant data did not sufficiently cover this range. This is because those equipment are usually operated in narrower regions than originally designed.

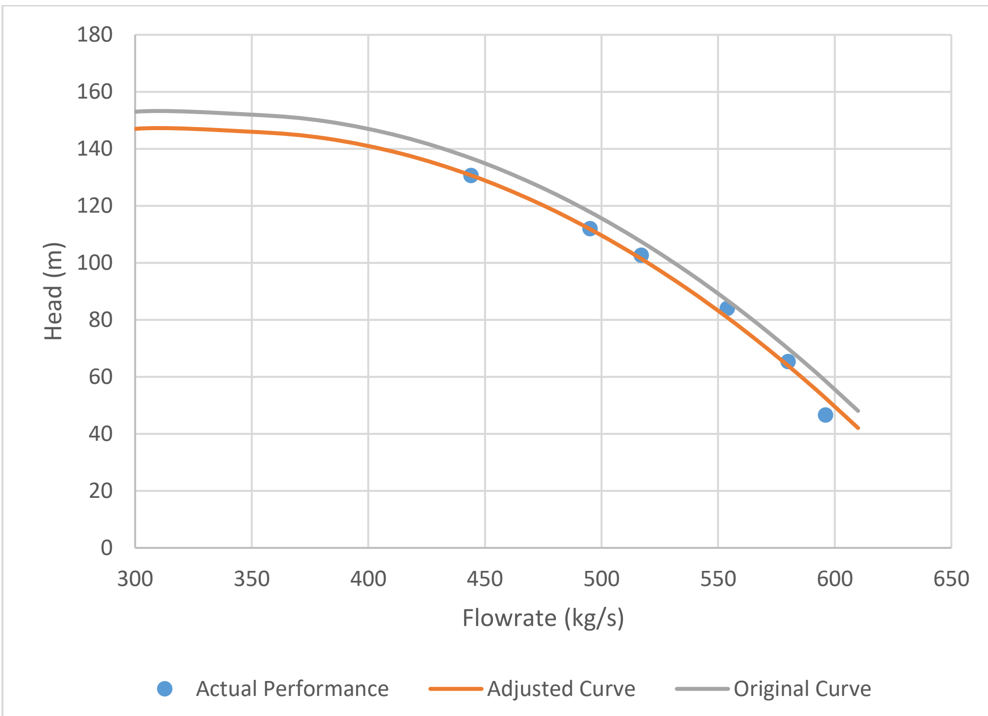


Figure 5 Adjusting flowrate-head curve for a SWI pump at GOSP B-3

To take into consideration the anti-surge controls for the compressors, the compressors were modelled with a minimum flowrate at 80% of the compressor design flowrate. This was done by utilizing the recycle flows. Since the reservoir GOR and WC changes over time, the compressor's suction conditions, such as the temperature and pressure, were adjusted to maintain the minimum flowrate while minimizing the recycle flowrate. The suction historical conditions were collected and utilized to accurately model power consumption.

4. Building physics-based models

gPROMS ProcessBuilder was used to build unit operation models and gRPOMS Multiflash Properties were used to represent the process thermodynamics. Subsequent sections will provide more description on the models' development process and its components.

5. Validating simulation against plant data

Model validation was carried out to validate the accuracy of the simulation against production data. The operating conditions in the model were tuned to match the actual operations. It was found that the model is able to accurately predict phase separation as well as power consumption as demonstrated in Table 1 and Table 2.

Table 1 Comparing measured with modelled energy consumption at base case (75% Loading)

GOSP	Measured Energy Consumption	Modelled Energy Consumption	Relative Error
ANDR-1	12.82	13.08	1.99%
ANDR-2	11.97	11.75	-1.87%
ANDR-3	8.24	8.57	3.85%
ANDR-4	6.41	6.47	0.93%
ANDR-6	0.17	0.18	5.56%
SDGM-1	4.74	4.82	1.66%
SDGM-2	9.21	8.91	-3.37%
SDGM-3	4.61	4.39	-5.01%
SDGM-4	5.72	5.36	-6.72%
SDGM-5	3.5	3.44	-1.74%
SDGM-6	2.78	2.67	-4.12%

Table 2 Relative errors in gas, water and oil predictions at base case (75% Loading)

GOSP	Gas Flow Relative Error	Water Flow Relative Error	Oil Flow Relative Error
ANDR-1	3.17%	2.47%	-4.13%
ANDR-2	-2.96%	0.46%	1.90%
ANDR-3	1.07%	3.34%	-2.82%
ANDR-4	-2.67%	-1.32%	3.80%
ANDR-6	6.45%	-4.46%	-1.28%
SDGM-1	0.56%	4.90%	-2.66%
SDGM-2	1.10%	5.70%	-3.71%
SDGM-3	1.58%	4.26%	-4.88%
SDGM-4	-1.90%	2.10%	-0.87%
SDGM-5	2.91%	-2.65%	-0.39%
SDGM-6	-1.70%	-2.98%	5.16%

6. **Formulating the optimization problem**

Once the simulation model was validated, the optimization model was developed. The optimization model drives the simulation model towards minimizing or maximizing a defined objective function based on a set of constraints and controlled variables. Section 5.6 will provide more details on the optimization formulation.

7. **Running case studies**

A variety of case studies were conducted. This mainly aimed to benchmark the model's performance against past data and operating conditions. It was essential to highlight the model's ability to produce reasonable solutions. It was also key to highlight variations between the base and optimized cases and to quantify the perceived benefits.

5.3 Physical Properties

The standard physical property package used in gPROMS ProcessBuilder is Infochem Multiflash, which is provided by KBC Advanced Technologies. This package is well suited for hydrocarbons modelling and the given application because of its ability to generate tight convergence of iterations and of partial derivatives with respect to composition, pressure and temperature. In addition, Phase equilibria is determined for a variety of PVT, enthalpy, entropy and internal energy combinations. The package also provides the composition of a given phase at a given pressure or temperature.

Multiflash avails a variety of commonly-used models for equations of state, such as Soave-Redlich-Kwong (SRK) and Peng-Robinson (PR), which are classified as cubic equations of state. Moreover, a variety of non-cubic equations of state exist, such as Lee-Kesler and the Benedict-Wee-Rubin-Starling.

For this work, the SRK equation of state was chosen since it is able to sufficiently account for fugacity calculations. It was also compared against other equations of state and demonstrated better capabilities in representing multiphase crude oil separation processes. This model is described by the below set of equations.

$$P = \frac{RT}{V_m - b_{mixture}} - \frac{a_{mixture}(T)}{V_m(V_m + b_{mixture})}$$

$$\sqrt{a_{mixture}(T)} = \sum_{i}^{NC} y_i \sqrt{a_i(T_i)}$$

$$b_{mixture} = \sum_{i}^{NC} y_i b_i$$

$$a_i(T) = \frac{0.4275 R^2 T_{ci}^2}{P_{ci}} a_i(T), \qquad i = 1, \ldots, NC$$

$$a_i(T) = \left[1 + m_i(1 - \sqrt{\frac{T}{T_{ci}}})\right]^2, \qquad i = 1, \ldots, NC$$

$$m_i = 0.48 + 1.574 w_i - 0.176 w_i^2, \quad i = 1, \ldots, NC$$

$$b_i = 0.08644 \frac{RT_{ci}}{P_{ci}}, \qquad i = 1, \ldots, NC$$

where R is the ideal gas constant and V_m is the molar volume. The parameters $a_{mixture}$ and $b_{mixture}$ are calculated using ai and bi which are the pure components parameters that are determined from the critical pressure P_{ci}, the critical temperature T_{ci} and the acentric factor w_i.

Various activity coefficient models can be used, such as NRTL, UNIFAC and UNIQUAC. These are all available within the Multiflash package. For this work, the NRTL model was selected because it can be efficiently used for vapor-liquid equilibrium calculations.

5.4 gPROMS gML Library

The purpose of this section is to describe the physical meaning and functionality of the main model components used to build the Multi-GOSP high-fidelity models.

The gML library contains a variety of models for steady state and dynamic processes. The gML library contains all the base components required to set-up the required models.

5.4.1 Centrifugal Compressor

The centrifugal compressor model in gPROMS can be used to represent both single and multiple compression stages. However, in this study, only single stage compressors were used.

Isentropic and polytropic efficiencies

In a single stage compressor, the isentropic and polytropic efficiencies are described by the below relationships:

$$\eta_P = \eta_{is} \frac{\left\{\left(\frac{P_{out}}{P_{in}}\right)^{\frac{n-1}{n}} - 1\right\} \frac{n}{n-1} \frac{k-1}{k}}{\left(\frac{P_{out}}{P_{in}}\right)^{\frac{k-1}{k}} - 1}$$

$$n = \frac{ln\left(\frac{P_{out}}{P_{in}}\right)}{ln\left(\frac{\rho_{out}}{\rho_{in}}\right)}$$

$$k = \frac{ln\left(\frac{P_{out}}{P_{in}}\right)}{ln\left(\frac{\rho_{is}}{\rho}\right)}$$

Where η_P and η_{is} are the polytropic and isentropic efficiencies, respectively. P_{out} and P_{in} are the fluid outlet and inlet pressure, respectively. ρ_{out} and ρ_{in} are the fluid outlet and inlet density, respectively. P_{is} is the outlet isentropic fluid density. n is the polytropic index and k is the isentropic index.

The compressor's polytropic or isentropic efficiencies can then be used to determine the power demand using the following equations:

$$W_{is} = F(h_{is} - \ h_{in})$$

$$W_P = W \frac{\eta_P}{100}$$

$$W_{is} = W \frac{\eta_{is}}{100}$$

where F is the mass flow rate. h_{in} and h_{is} are the inlet and outlet mass specific enthalpy for isentropic compression. W_P and W_{is} are the power demand for polytropic and isentropic compression, respectively. W is then the compressor thermodynamic power demand.

Similarly, the fluid's enthalpy can be determined after calculating the power supply to the fluid and the fluid's flow rate using the below equation.

$$W = F(h_{out} - h_{in})$$

where h_{out} is the outlet mass specific enthalpy.

It is then possible to account for the mechanical losses which lead to the mechanical power demand being larger than the power required to compress the fluid. Assuming steady state conditions, the mechanical losses can be determined as follows:

$$W_{mech} \frac{\eta_{mech}}{100} = W$$

where W_{mech} is the compressor mechanical power demand and η_{mech} is the mechanical efficiency.

For the multi-GOSP optimization problem, performance maps were used to predict compressors' outlet pressure based on the flow rate. These maps provide relationships between the volumetric flowrate, the compressor's efficiency and the polytropic head. In gPROMS, performance maps can be either 1- or 2- dimensional. One-dimensional maps do not include a dimension for varying compressors speeds, while 2-dimensional maps allow varying performance at feasible compressor speeds.

The compressor's polytropic pressure head is given by:

$$H_P = 1000 \frac{W_P}{F}$$

where H_P is the polytropic head.

It is then possible to use the below equations to determine the head and efficiency from the volumetric flow rate:

$$H_{P,0} = M_{head}(F_{v,0})$$

$$\eta_p = M_{efficiency}(F_{v,0})$$

where $F_{v,0}$ is the inlet design volumetric flow rate. $H_{P,0}$ is the design polytropic head. M_{head} is the map head function and $M_{efficiency}$ is the map efficiency function.

5.4.2 Centrifugal Pumps

The centrifugal pump model in gPROMS can be used to represent both single and multiple stages. The power demand W_{is} for an ideal isentropic process is given by:

$$W_{is} = F(h_{is} - h_{in}) \approx F\frac{10^2\left(P^{out} - P^{in}\right)}{\rho}$$

Then, the actual power W supplied to the fluid can be determined by:

$$W\frac{\eta_{is}}{100} = W_{is}$$

The fluid's outlet enthalpy and torque requirement can then be determined in a similar fashion to the centrifugal compressors as described in section 5.4.1. Similarly, the equations in section 5.4.1 concerning calculating head and efficiency from volumetric flowrate can be used here.

5.4.3 Separators

In separators, the mass balance for each component i is provided by the following equation:

$$\hat{V}\frac{d\tilde{m}_i}{dt} = F^{in}w_i^{in} - F_L x_i - F_V y_i \quad, i = 1, \ldots, NC$$

where $\hat{V}$ is the characteristic volume, $\tilde{m}_i$is the volumetric mass holdup, F_{in}, F_L and F_V are the inlet , outlet liquid and outlet vapour mass flowrates, respectively. Finally, x_i and y_i are the mass fractions of component i in the liquid and vapour phases, respectively.

Since the models have an underlying assumption of a steady state, all terms involving the holdup are zero.

On the other hand, the energy balance is expressed by the following equations:

$$\hat{V}\frac{d\tilde{u}}{dt} = F^{in}h^{in} + Q^{in} - F_L h_l - F_V h_v$$

$$\tilde{u} = \tilde{m}_T h - 10^2 P$$

where ũ is the volumetric energy holdup, h_{in} is the inlet mass specific enthalpy and Q_{in} is the energy rate for the heat supplied to the vessel. h_L and h_V are the outlet liquid and vapour mass specific enthalpy, respectively.

Similarly, since the models have an underlying assumption of a steady state, all terms involving the holdup are zero.

The physical property package is then used to calculate the fugacity coefficients. The phase equilibrium is then obtained as follows:

$$\frac{y_i \emptyset_{i,v}}{\sum_{j \in NC} \frac{y_i}{MW_j}} = \frac{x_i \emptyset_{i,L}}{\sum_{j \in NC} \frac{x_i}{MW_j}}, \; i = 1, \ldots, NC$$

$$\sum_{i \in NC} x_i = 1$$

$$\sum_{i \in NC} y_i = 1$$

where MW_i is the molecular weight of component i. $\emptyset_{i,v}$ and $\emptyset_{i,L}$ are the fugacity coefficients of component i in the vapour and liquid phases.

5.5 Flowsheet Implementation

As was described in chapter 4, the multi-GOSP optimization process includes optimizing the operation of network-level and plant-level variables. Accordingly, a representation of the network took into account the multi-level nature of these plants.

Initially, individual GOSP models were built using the gML Library components as described in section 5.4. This is in addition to other auxiliary components which are necessary for flowsheet construction.

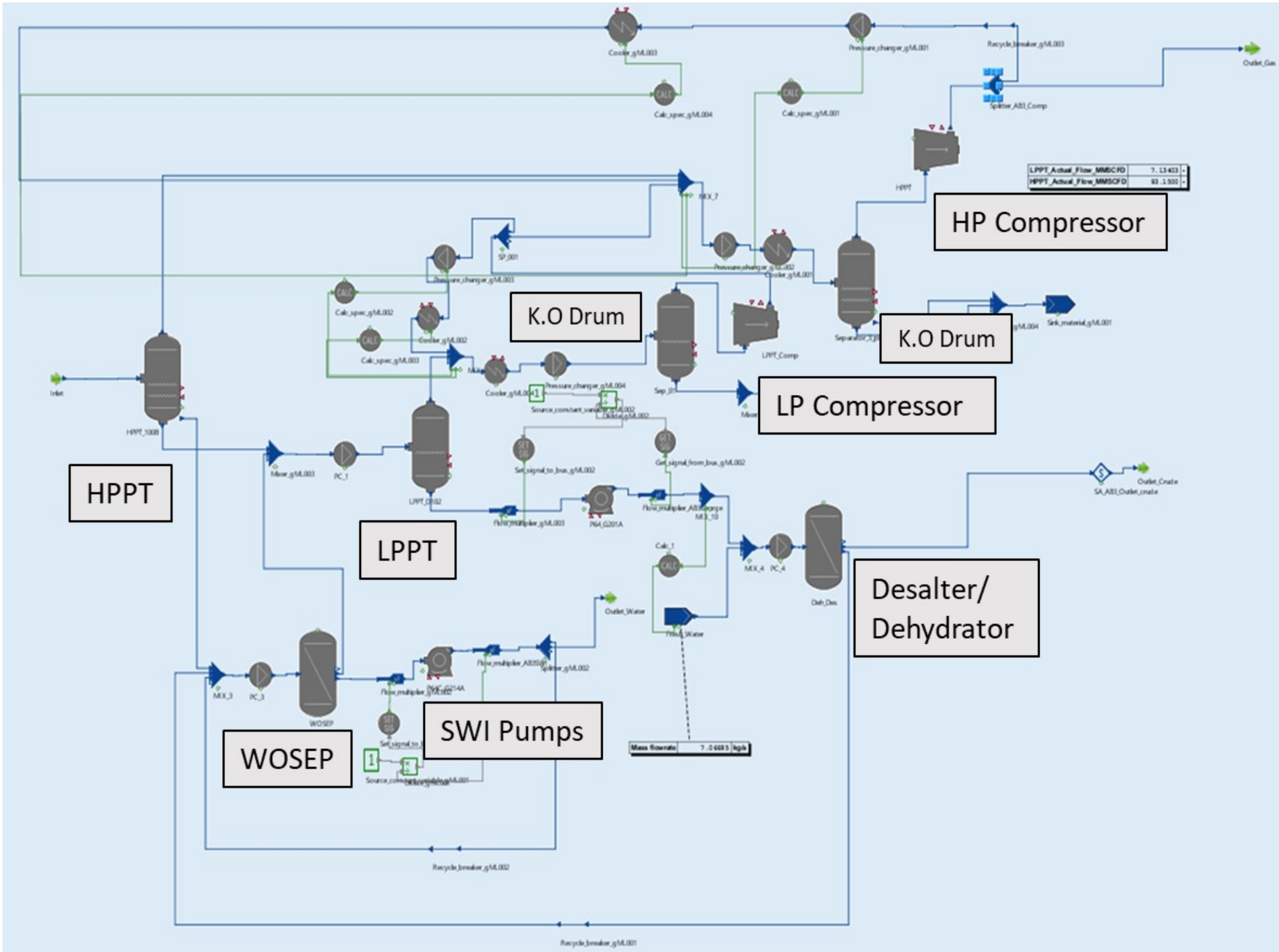


Figure 6 Flowsheet implementation of GOSP C-1 in gPROMS ProcessBuilder

A feed stream was added as a generic stream representing typical plant feed. This took no account of the possibility to swing feed streams as this was left to be a decision variable on the network level.

Feed composition reflected the same on plant Process Flow Diagrams (PFDs). Additionally, two pseudo-components were generated in this model to represent actual properties. These are listed below:

Table 3 Pseudo-components Properties

Pseudo-component	Molecular Weight	SPG	Tc, °F	Pc, psig	Acentric Factor (ω)
C7+	221	0.8463	862	259	0.485
C12+	312	0.8824	988	193	0.778

Table 4 summarizes crudes composition used in the model.

Table 4 Crude Composition (Mass Fraction)

	Composition
N2	0.0000
H2S	0.0003
CO2	0.0003
METHANE	0.0001
ETHANE	0.0006
PROPANE	0.0021
ISOBUTANE	0.0008
N-BUTANE	0.0077
PENTANE	0.0042
C6	0.0036
C7+	0.1868
C12+	0.7747

Coupled with the gas phase composition, GOR and WC, the feed to the GOSP is hence fully defined.

PFDs of each GOSP were followed to model the process rigorously.

The feed stream was connected to a separator model, which separates gas, water and liquid components. This separator represents the HPPT. The outlet gas stream was connected to a pressure changer, which is followed by a cooler and a separator. This separator represents the knock-out drum. The water stream was connected to a pressure changer then a component splitter, which represents the WOSEP. The goal of the pressure changer is to reflect the operating pressure of the WOSEP. The liquid hydrocarbon stream from the HPPT was connected to another pressure changer and then another separator. This separator represents the LPPT. The aim of the pressure changer is to represent the drop of pressure and allow flashing gases in the LPPT.

From the knock-out drum, which is connected to the gas stream from the HPPT, the overhead stream was connected to a high-pressure compressor. The compressor was configured to operate based on performance maps as described in section 5.4.1. Additionally, the compressor was equipped with a recycle. This was configured as a split stream from the compressor's outlet, which goes through a pressure changer then a cooler before going to the inlet of the knock-out drum. The recycle is key since several compressors were operated below design limits and were operating on a recycle mode. The bottom of the knock-out drum wasn't considered since it is a very small stream.

From the LPPT, the gas stream was connected to a cooler, a pressure changer then a knock-out drum. From the knock-out drum the overhead is sent to the low-pressure compressor. This is followed by a cooler, a knock-out drum and then the outlet is connected to the high-pressure compressor.

The liquid leaving the LPPT was connected to a charge pump with a multiplier. The goal of the multiplier is to allow for ease of controlling the number of parallel equipment. It is assumed that all parallel pumps operate similarly and use the same performance maps as described in section 5.4.2. The pump was connected to a recycle stream that can allow recycling product around it similar to the arrangement with the compressors.

The WOSEP was configured such that it perfectly splits hydrocarbons from water. The former is sent to the inlet of the LPPT. The latter is sent to the Salt Water Injection Pumps (SWIs). Those pumps were equipped with performance curves similar to previously described rotating equipment. Water leaving SWIs is sent through a sink and leaves the model.

The crude stream leaving LPPTs was connected to a dehydrator/desalter through the charge pump. Dehydrator/Desalters were modelled as a component splitter, which receives a wash water stream in addition to LPPT's crude. The component splitter then perfectly split hydrocarbons and send this stream as the crude oil output. Water, on the other hand, is recycled to the WOSEP.

As explained in chapter 4, GOSPs often differ in configuration. The above implementation process is therefore a generic one, which was adapted per actual plant set-up.

After building individual plant models, those were integrated into a network model. gPROMS allows setting up an interface for each model, such that it is transferred into a sub-model with ports to connect inlet and outlet streams.

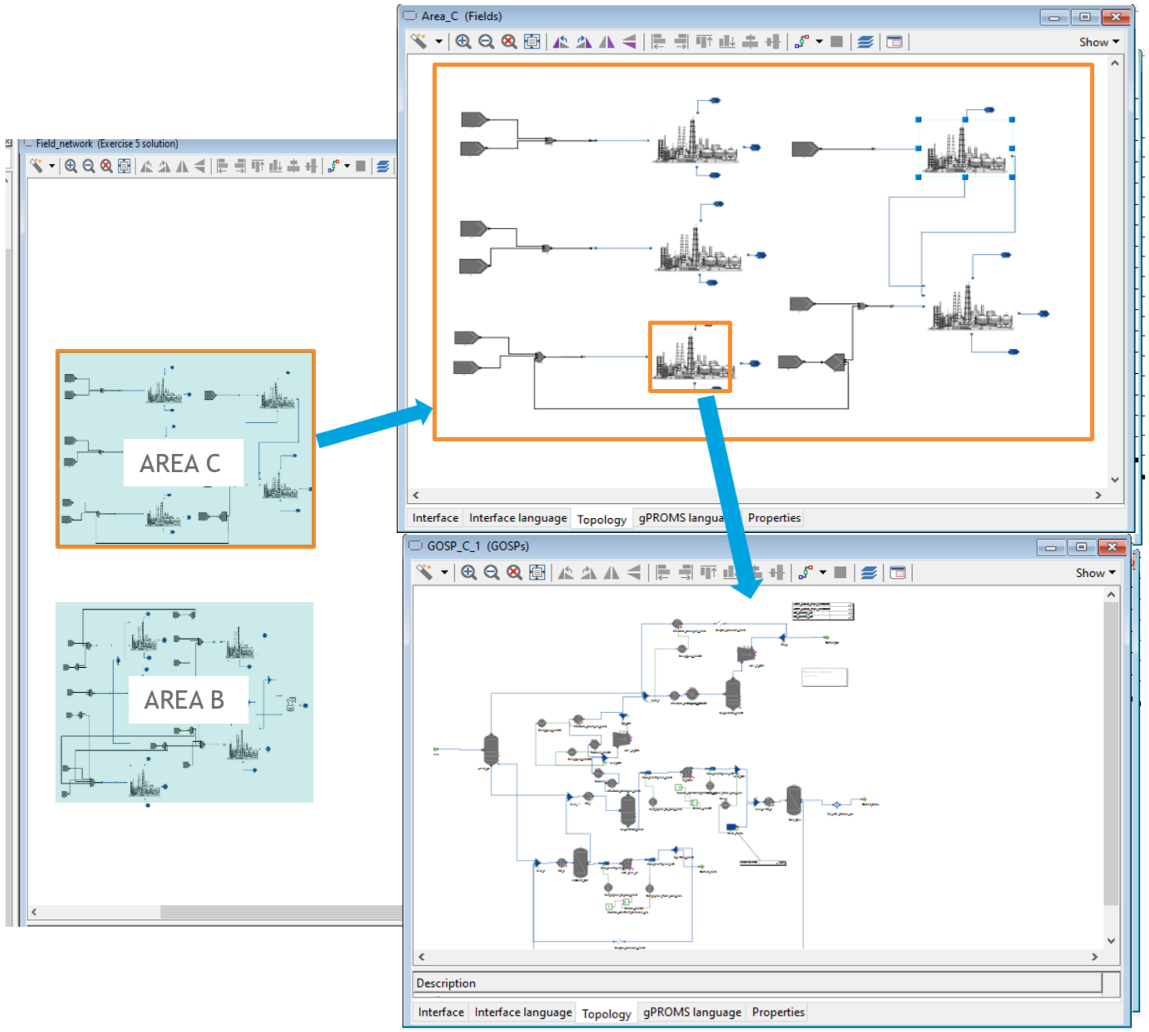


Figure 7 Hierarchal topology implementation for the network's overall topology in gPROMS ProcessBuilder

The network model was built to match the description in chapter 3. Feed streams were connected with the plants through a series of aggregators and routers. Aggregators allow combining multiple streams into a single stream. Routers allow specifying one or more possible outlets. Those are specified discretely, such that only one outlet stream can be selected at any time.

Pressure changers were also used to allow for matching the pressure at each GOSP's HPPT. This allows flashing feed streams at the right conditions.

5.6 Optimization Problem Formulation

This section describes how the optimization problem was posed and presents details on how constraints and the objective function were formulated.

The total power consumption (TPC) of $GOSP_g$ is defined as:

$$TPC_g = \sum_{hpc \in HPC} PC_{hpc} + \sum_{lpc \in LPC} PC_{lpc} + \sum_{wip \in WIP} PC_{wip} + \sum_{sp \in SP} PC_{sp}$$

Where PC is the power consumption, and hpc and lpc are the high and low pressure compressors, respectively, while

wip and sp are the water injection and crude shipping pumps respectively.

The total feed (TF) to a GOSP is described by the following equation:

$$CAP_g^{min}.Y_g \leq TF \leq CAP_g^{max}.Y_g$$

Where CAP_g^{min} and CAP_g^{max} are the minimum and maximum capacities of GOSP g, respectively.

Y_g is a binary variable. It equals 0 if the GOSP is off and 1 if the GOSP is on.

The water handling capacity (WHC) of GOSP g is define as:

$$WHC_g \leq WHC_g^{max}.Y_g$$

The capacity of the HP and LP compressors is defined as:

$$0.8 \,.\, CompCAP_{hpc,g}^{max}.Y_g \ \leq CompCap_{hpc,g} \leq CompCAP_{hpc,g}^{max}.Y_g \quad \forall\ hpc\ \in HPC$$

$$0.8 \,.\, CompCAP_{lpc,g}^{max}.Y_g \ \leq CompCap_{lpc,g} \leq CompCAP_{lpc,g}^{max}.Y_g \quad \forall\ lpc\ \in LPC$$

Due to operational requirement, compressors were restricted to run at a minimum of 80% of their maximum capacity.

Both LP and HP compressors are operated to deliver gas at a minimum head pressure. This is defined as:

$$CompHead_{hpc,g} \leq CompHead_{hpc,g}^{max}.Y_g \qquad \forall\ hpc\ \in HPC$$

$$CompHead_{lpc,g} \leq CompHead_{lpc,g}^{max}.Y_g \qquad \forall\ lpc\ \in LPC$$

The total oil production, which is one of the main constraints is described as:

$$TotalOil\ \geq \sum SPR_g \qquad \forall\ g\ \in G$$

Where SPR_g is the total rate leaving the shipping pump at each GOSP g.

The minimum pressure of the water injection pumps was described by polynomials that define flow/pressure relationships. The polynomial was regressed as a best fit of historical data. The model is constrained to deliver a pressure which is at least equivalent to the minimum pressure at the water flowrate. This often mandates adding additional parallel pump(s) to meet the required pressure.

Moreover, it was often necessary to add some mathematical equations that describe the relationship between various GOSPs. For example, if the gas leaving GOSP g is processed at GOSP g', it is essential to maintain that the first GOSP cannot be operated while the second is shutdown. On the other hand, the second GOSP can still operate if the first is shutdown as long as it is able to meet the compressors' flowrates.

This is described through the following key equation.

$$Y_{g'}\ \leq\ Y_g$$

The objective function is simply defined as the minimization of the total power consumption of all the GOSPs. This is described below:

$$Min \quad TPC_G = \sum TPC_g\ \forall\ g\ \in G$$

An alternate objective function was similarly formulated to solve the problem of maximizing oil production during the planned/unplanned shutdown of one or more GOSPs.

The alternate objective function is described as:

$$Max \quad TotalOil = \sum SPR_g \qquad \forall\ g\ \in G$$

After formulating the problem, gPROMS was then used to perform steady-state optimization, where both continuous and discrete variables were optimized to minimize energy consumption while satisfying the given constraints.

Due to the nature of the described equation system, a nonlinear behaviour can be observed in addition to several binary variables. This makes this an MINLP problem, which follows the particularities described in section 3.1.2.

6. Case Studies & Results

In this section, we apply the proposed model to solve an optimization problem for the 2 integrated GOSP areas proposed in chapter 4. Both areas produce the same type of crude oil, Arab Light. It was therefore possible to combine both areas in one model, so that both are operated towards one objective function while increasing the flexibility by combining the controlled variables of both.

The modelling and optimization runs were performed using a Microsoft Windows 10 platform (64-bit operating system) with a 2.90 Ghz dual-core processor (AMD A6-5350M APU) and 8 GB RAM.

Table 5 provides a summary of the main capacities for GOSPs in both areas under consideration.

Table 5 Overall Crude, Water and Gas Handling Capabilities

GOSP	Maximum Crude Capacity (% of Total Area B & C)	Design WOSEP Capacity (% of Crude Capacity)	Number of Salt Water Injection Pumps	Number of Low-Pressure Compressors	Number of High-Pressure Compressors
B-1	5.52%	86.8%	2	1	0
B-2	N/A	N/A	N/A	1	2
B-3	12.57%	49.5%	4	0	0
B-4	12.57%	49.5%	4	1	2
B-5	12.57%	49.5%	3	0	0
B-6	6.48%	96.1%	2	0	0
C-1	12.57%	75.8%	4	1	2
C-2	12.57%	49.5%	4	1	2
C-3	12.57%	49.5%	3	1	1
C-4	6.29%	99.0%	4	1	1
C-6	6.29%	N/A	N/A	0	0

For the purpose of this case study, the feed rates to each GOSP were altered by the optimizer while being maintained within the provided avails. Avails present production capabilities as determined by Production Engineering and Reservoir Management. Typically, rates are controlled by adjusting the choke valves of the producing wells at the well pads. Operators will aim to ensure that each GOSP receives the required production rates.

It was assumed that the flowrates can be perfectly controlled to produce the required flowrates and averaged GORs and water cuts. Well flowrates were aggregated and the model was provided with an initial guess, a minimum and a maximum rate for each feed stream. As previously described, feed streams comprise aggregates of both swingable and non-swingable wells. The minimum for each feed stream was set to zero. The initial guess was provided as the starting production target, which is between the minimum and the maximum. The maximum is the avails flow rate.

6.1 Energy Optimization

In this section, the proposed model is applied with the objective to minimize energy consumption as set forth in the methodology section. The total oil production target was set as a percentage of the total avails. The goal is to allow studying the model's response as oil production targets as altered.

The base case in all scenarios assumed that all GOSPs are maintained operational as is the common practice. Moreover, in generating the base case, the optimizer was allowed to optimize the operation within individual GOSPs. However, swinglines were not utilized, since they're not commonly used in current operation for energy optimization. On the other hand, the optimizer was provided the flexibility to recommend shutting down GOSPs and using swing lines if it is deemed optimal.

6.1.1 Energy Optimization at 50% Throughput

The case where the throughput is 50% of total availability is initially tackled. The model statistics and computational results for this case are presented in Table 6. As demonstrated in the below table, the CPU time of 248 seconds is sufficiently short to serve the requirements of N-RTOs. Accordingly, the models' structure and level of fidelity are adequate and use of lower fidelity techniques are not needed. Indeed, it is expected that lower fidelity models would be required if the approach is to be used for gas plants.

Table 6 Model Statistics for 50% Throughput Level Scenario

Case	50% Throughput

Number of variables (discrete and continuous)	2344
Base Case Objective Function (MW)	66.06
Optimized Case Objective Function (MW)	32.52
CPU (s)	248

Figure 8 presents the power consumption comparison for each GOSP between the optimal and base solutions for the 50% production level. It shows that the optimal solution does not guarantee that all GOSPs in the optimal solution consume lower power than in the base solution. Indeed, it can be observed that 4 GOSPs consume more power in the optimal case. On the other hand, 7 GOSPs consume less power in the optimal case. This is because the optimization model considers the whole network at once and manipulates controlled variables amongst all GOSPs to achieve a better overall saving rather than optimizing each GOSP individually. This indeed demonstrates the value of optimizing the network as a single entity.

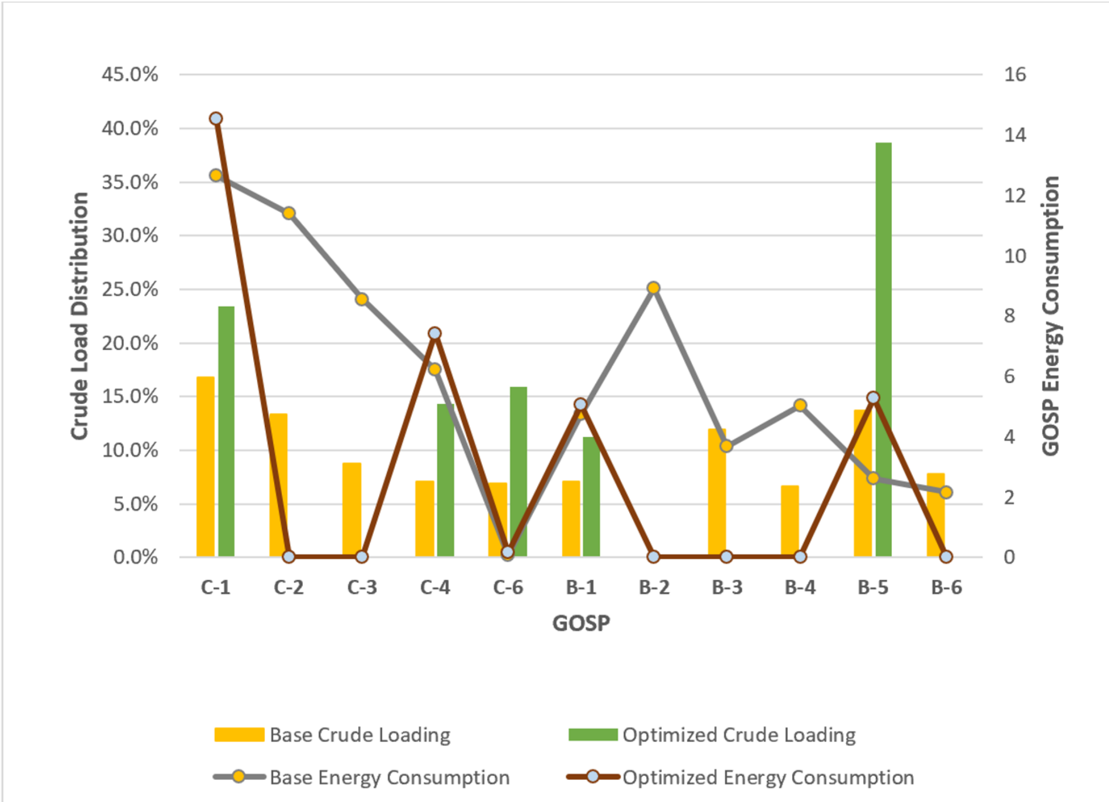


Figure 8 The power consumption comparison for each GOSP between the optimal and base solutions for the 50% production level

Figure 9 shows a schematic of optimal swing lines utilization between Area B and Area C GOSPs. It shows that only 2 swing lines were used. It also shows that the optimizer chooses to shut down a total of 6 GOSPs. This is an exceptional case due to the very low throughput level being tested. The case was devised purposely to test the algorithm's performance and robustness. The model's results are largely sensible since it elects to shutdown GOSPs with the higher power consumption. The optimizer recommends shutting down GOSPs B-3 and B-6 although their power consumption is relatively low, as both have no compressors. This is because they are linked with GOSP B-2, which receives their gases and consumes significant energy to compress it.

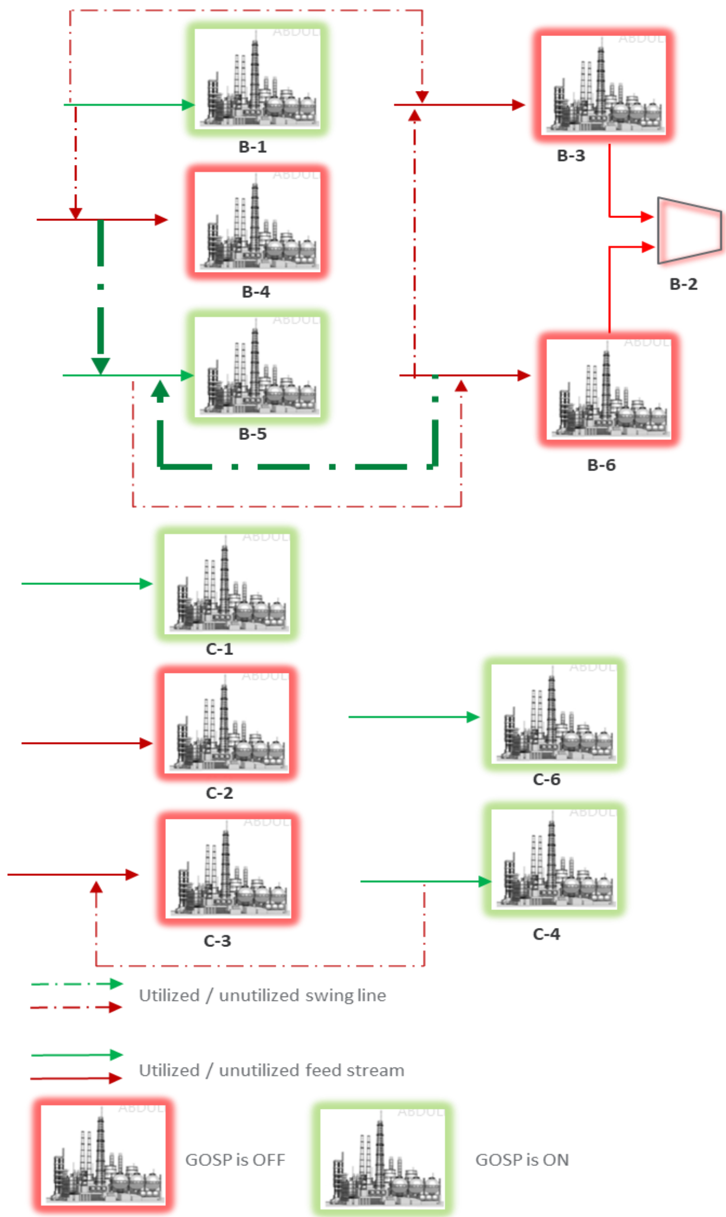


Figure 9 A schematic of optimal swing lines utilization between Area B and Area C GOSPs at 50% throughput

The optimum objective function in this case is 32.52 MW, compared to a base energy consumption of 66.06 MW.

6.1.2 Energy Optimization at 75% Throughput

In this case, throughput level is increased to 75% of maximum availability, thus only changing the constraint relating to the minimum oil throughput. The objective function, other constraints and all variables are maintained to be the same. The model statistics and computational results for this case are presented in Table 7.

Table 7 Model Statistics for 75% Throughput Level Scenario

Case	75% Throughput
Number of variables (discrete and continuous)	2344
Base Case Objective Function (MW)	69.64
Optimized Case Objective Function (MW)	56.46
CPU (s)	232

Figure 10 presents the power consumption comparison for each GOSP between the optimal and base solutions for the 75% production level. Similar to the previous case, it shows that the optimal solution does not guarantee that all GOSPs in the optimal solution consume lower power than in the base solution. In this case, it can be observed that 4 GOSPs consume more power in the optimal case. On the other hand, 7 GOSPs consume less power in the optimal case. This is because the optimization model considers the whole network at once and manipulates controlled variables amongst all GOSPs to achieve a better overall saving rather than optimizing each GOSP individually. This indeed demonstrates the value of optimizing the network as a single entity.

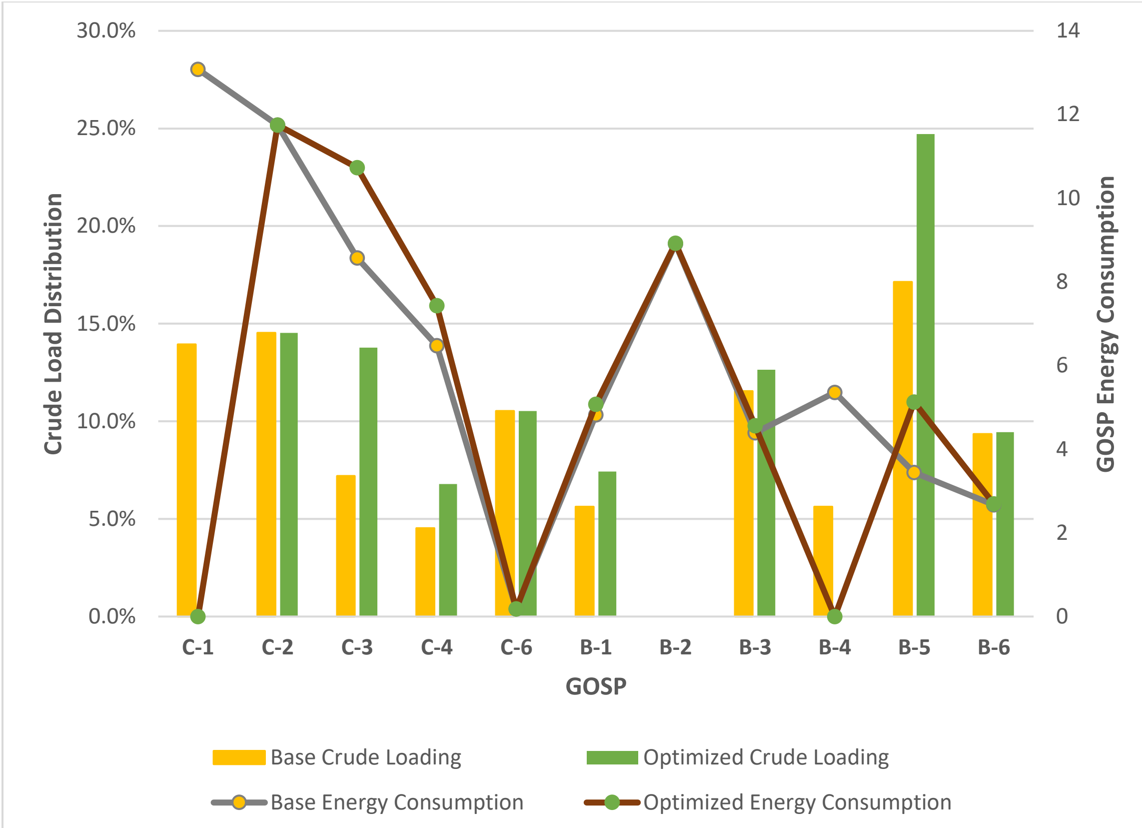


Figure 10 The power consumption comparison for each GOSP between the optimal and base solutions for the 75% production level

Figure 11 shows a schematic of optimal swing lines utilization between Area B and Area C GOSPs. It shows that only 2 swing lines were used. It also shows that the optimizer is able to shut down 2 GOSPs only. This is a more realistic case and provides a more reasonable throughput level. The model's results are largely sensible since it elects to shut down 2 GOSPs (B4 and C1) with relatively high power consumption. Both GOSPs are equipped with low- and high-pressure compressors and sets of sour water injection pumps, making them sensible choices for shutdown. As opposed to the 50% production level case, the model is not able to shut down more GOSPs as it is constrained to meet oil production target.

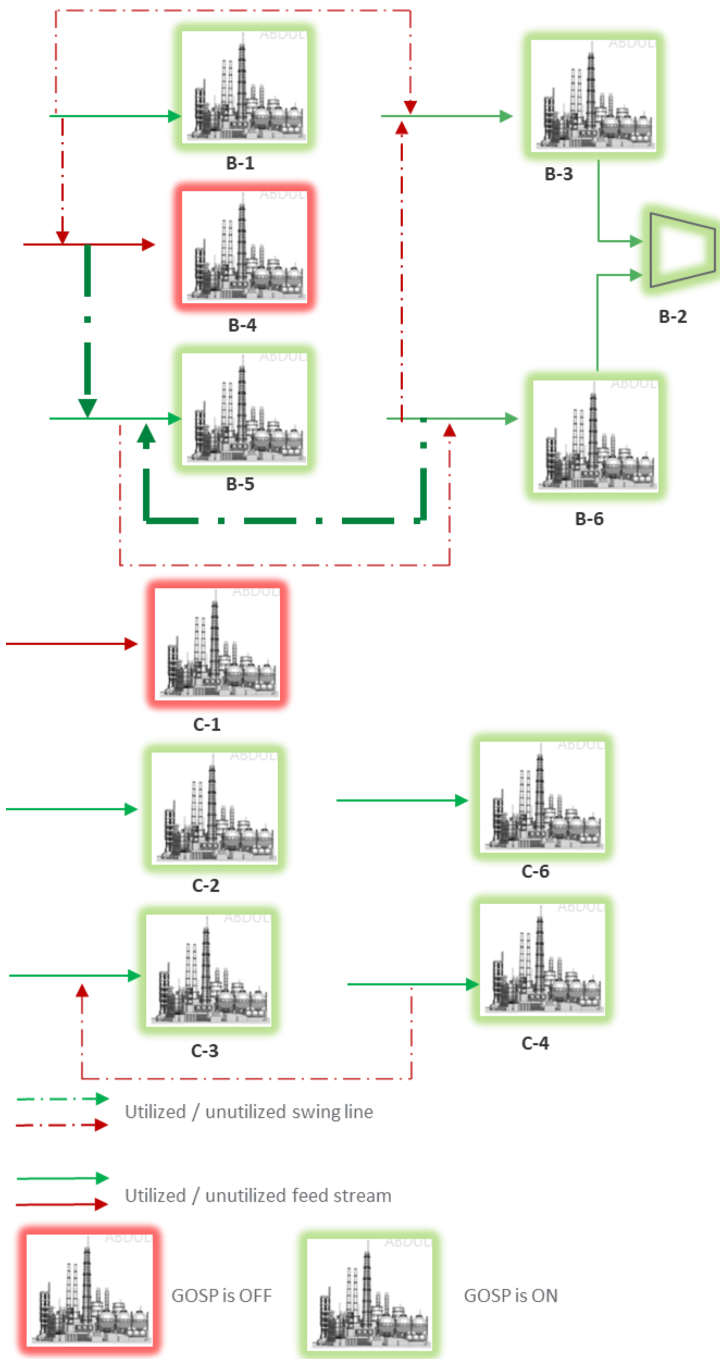


Figure 11 A schematic of optimal swing lines utilization between Area B and Area C GOSPs at 75% throughput

The optimum objective function in this case is 56.46 MW, compared to a base energy consumption of 69.64 MW.

6.1.3 Energy Optimization at Other Throughput Levels

To evaluate the optimal solutions achieved for all cases under consideration, the results of the optimal cases are compared with those of the base case where GOSPs are not shut down and swing lines are not utilized.

The power consumption comparison between the two solutions for all cases is demonstrated in Figure 12. It can be seen that the savings progressively decline as the throughput level is increased, which is sensible. As throughput levels increase, the optimization potential decreases and there is less room for substantial savings due to reduced optimization potential. The only thing that can be optimized in this case is the swing line utilization. This would also be limited to optimally distributing the load across rotating equipment while largely using the same number of them. The figure shows savings start at 51% corresponding to a throughput of 50% and reach 1% at a throughput of 100% of avails.

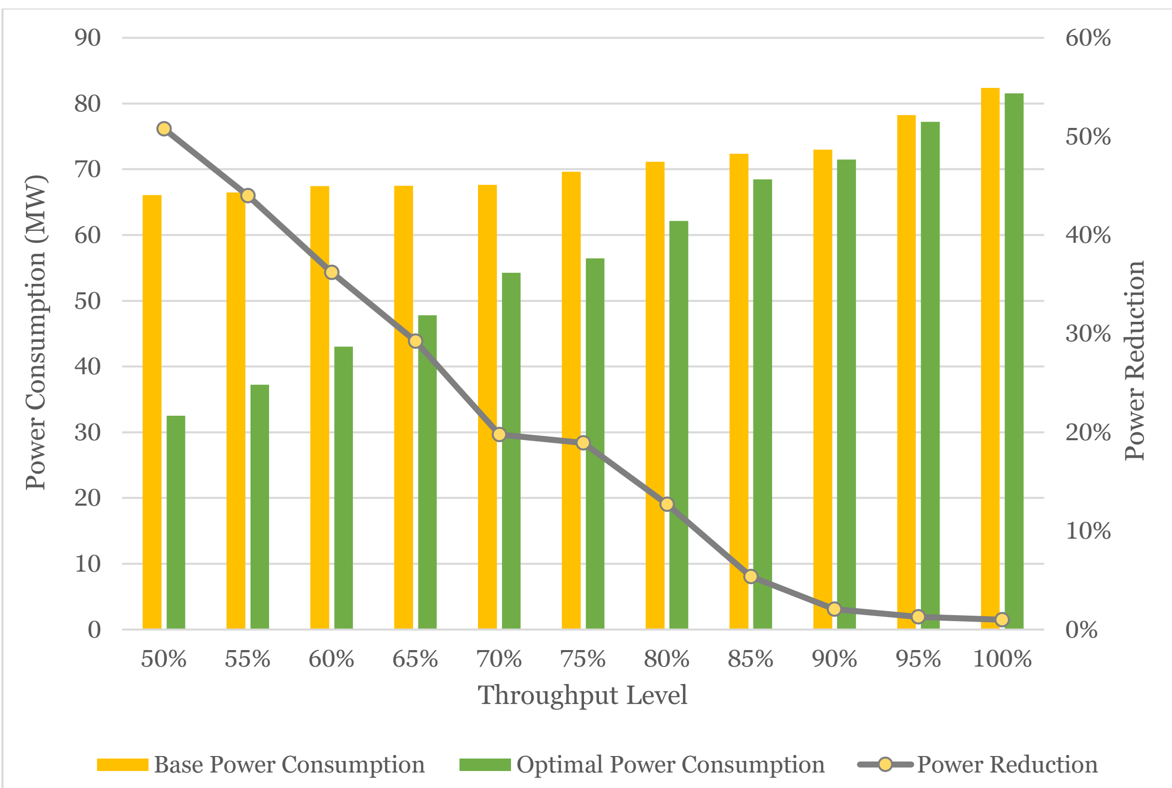


Figure 12 The power consumption comparison between the base and optimal solutions for all cases

Figure 13 shows the number of rotating equipment being utilized in the base versus optimal solutions. It can be observed that as throughput levels increase, the optimal number of rotating equipment starts approaching the base solution's scenario. Indeed, at the 100% case, both numbers are equal. This is sensible as increasing the throughput mandates using more rotating equipment, leaving little room for optimization.

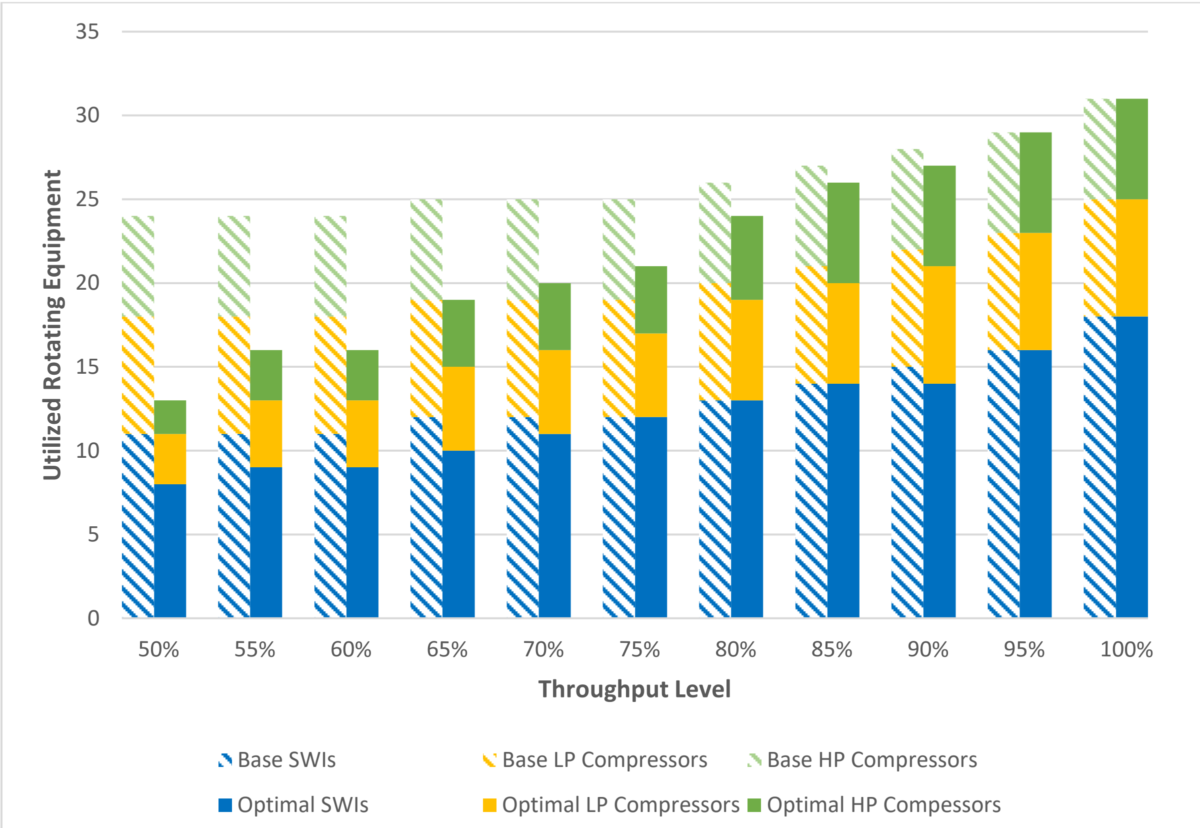


Figure 13 The number of rotating equipment being utilized in the base versus optimal case

6.2 Maximizing Oil Production with GOSPs Under Shut Down

Every year, each GOSP undergoes planned maintenance for eight days. Furthermore, these GOSPs undergo a major turnaround every 7 years, which last for 30 days. During this period, oil production is often reduced. Swing lines are often used but this is not done systematically. On the other hand, using the models developed in this work, it is possible to identify optimal solutions, such that oil production is maximized, while ensuring other constraints are met. On the other hand, while using heuristics it may be difficult to ensure that an optimal solution was achieved without violating equipment constraints.

To perform this analysis, the objective function is therefore changed to maximize crude production rather than minimizing energy consumption.

Additionally, we assign a fixed value of zero for the binary variable relating to GOSP status. As described in the methodology section, the total feed (TF) to a GOSP is described as:

$$CAP_g^{min}.Y_g \leq TF \leq CAP_g^{max}.Y_g$$

Where CAP_g^{min} and CAP_g^{max} are the minimum and maximum capacities of GOSP g, respectively.

Accordingly, we assign the binary variable Y_g a fixed value of zero.

6.2.1 Maximizing Oil Production with GOSP B-1 and B-5 Under Shut Down

In this case, we examine the scenario of shutting down both GOSP B-1 and B-5. Both GOSPs account for a total of 18.1% of the total production for Areas B and C as demonstrated in Table 5. Accordingly, we assign a fixed value to the binary variable relating to the status of both GOSPs.

In the base case, all production assigned to these GOSPs is lost. On the other hand, the result of the optimized solution is demonstrated in Figure 14.

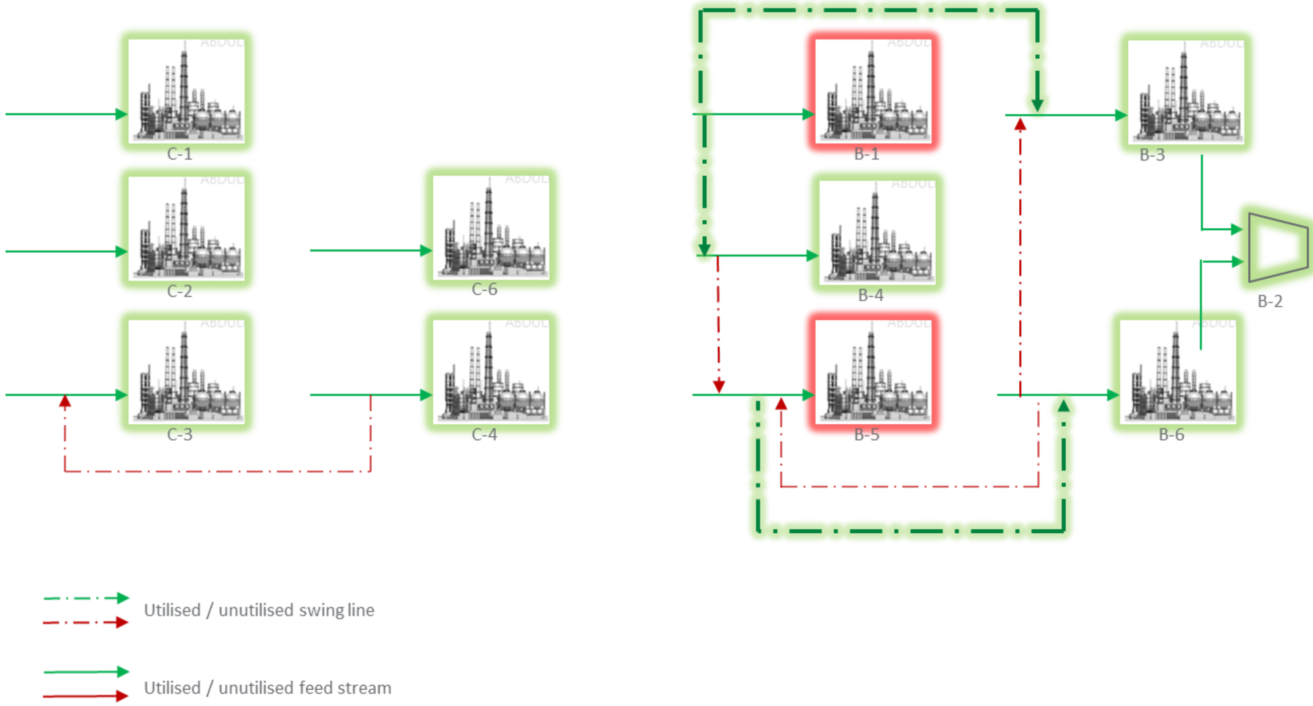


Figure 14 A schematic of optimal swing lines utilization between Area B and Area C GOSPs during shutdown of GOSPs B-1 and B-5

As can be observed, the model sensibly swings production away from the shut down GOSPs. Production is transferred from GOSP B-1 to GOSPs B-3 and B-4. Similarly, production is shifted from GOSP B-5 to GOSP B-6.

It is interesting to note that this is the only case where production is shifted away from GOSP B-5. This GOSP is typically used due to its low energy consumption. Therefore, it can be seen that it is usually fully loaded when the objective function aims to minimize energy consumption.

In this case production loss is reduced by 35.7% from the base case. In fact, the optimizer is able to load nearby GOSPs until constraints are made active in both GOSPs. Interestingly, both GOSPs are limited by SWI pumps capacities. Indeed, it may be interesting to study the benefit of adding additional SWI pumps to both GOSPs and studying how this contributes towards further maximizing production in this case.

6.2.2 Maximizing Oil Production with GOSP C-4 Under Shut Down

In this case, we examine the scenario of shutting down GOSP C-4 which accounts for 6.29% of the total production for Areas B and C as demonstrated in in Table 5.

Accordingly, we assign a fixed value to the binary variable relating to the status of this GOSP.

In the base case, all production assigned to this GOSP is lost. On the other hand, the result of the optimized solution is demonstrated in Figure 15.

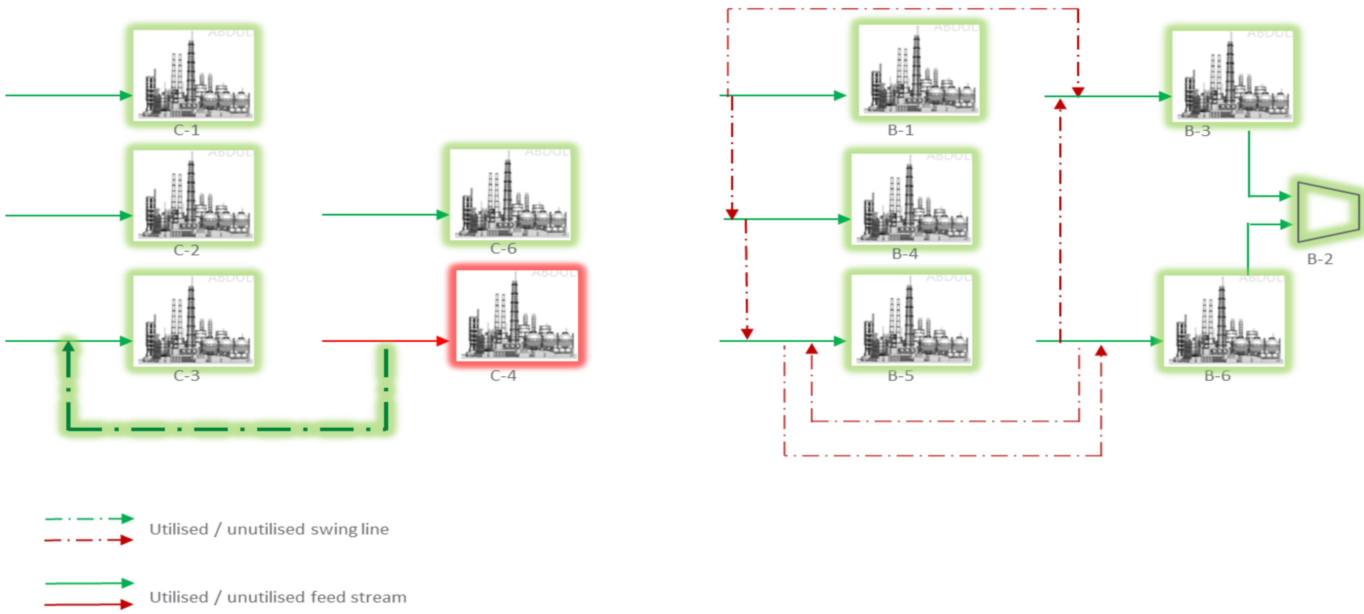


Figure 15 A schematic of optimal swing lines utilization between Area B and Area C GOSPs during shutdown of GOSP C-4

As can be observed, the model sensibly swings production away from the shutdown GOSP. Production is transferred from GOSP C-4 to GOSP C-3.

In this case production loss is reduced by 40.2% from the base case. In fact, the optimizer is able to load GOSP C-3 with the entire quantity that can be shifted from GOSP C-4 using the swingable production line.

As can be observed in this chapter, the proposed model offers good capabilities to consistently provide implementable and sensible solutions. It is also able to supersede the base-case solution in most scenarios. This is particularly the case when oil demand is lower than the maximum avails. It can also be noted that run time, which is a key consideration for N-RTOs, was acceptable in the presented cases. Moreover, it is noted that throughout the process of running all cases used within this work, run time ranged between 32 and 3,788 seconds. As such, even considering the most extreme run time, the approach is considered suitable.

7. Conclusions and Future Work

In this work, we examined elements which allow developing a novel framework for the near-real time optimization of oil and gas midstream networks. Modelling and optimization approaches were examined and their features as pertains to this problem were evaluated. We then applied the proposed techniques to optimize the operation of a complex GOSPs network.

We built a rigorous model that aimed to optimize the operation of an integrated network of GOSPs. The rigorous physics-based simulation model was augmented with an MINLP model with an objective function that aimed to minimize power consumption at times and to maximize production at other situations. This is key to maintaining both profitability and sustainability through further reducing greenhouse gas emissions. We applied the model to data inspired from the real world and demonstrated the capability of this novel approach to propose optimized solutions which can lead to reducing energy consumption by up to 51% in the 50% throughput scenario.

Future work includes expanding the scope to apply the examined techniques for the optimization of integrated gas plants networks. To meet the requirements, we set forth for N-RTO, we must apply surrogate-based modelling techniques as examined in chapter 3. This is due to the added complexity of optimizing these complexes. Indeed, the use of high-fidelity optimization is expected to fall short with regard to meeting various requirements relating to run-time and convergence. Nevertheless, we also expect to apply MINLP techniques as the use of linear models is expected to fall short in terms of meeting the required level of accuracy, yielding unimplementable solutions.

This novel work is expected to yield significant value and contribute towards Process Systems Engineering research and fill multiple gaps in the literature.

8. References

Abdel-Aal, H.K., Aggour, M., and Fahim, M.A., 2003. Petroleum and Gas Field Processing. Marcel Dekker, Inc. New York.

Agrawal, M., Jade, A.M., Jayaraman, V.K. and Kulkarni, B.D., 2003. Support vector machines: A useful tool for process engineering applications. *Chemical engineering progress*, *99*(1), pp.57-62.

Al-Enzi, G. and Elkamel, A. (2000)., Predicting the effect of feedstock on product yields and properties of the FCC process. Petrol. Sci. Technol., Vol. 18, No. 3-4, pp. 407–428.

Alhajree, I., Zahedi, G., Manan, Z.A. and Zadeh, S.M., 2011. Modeling and optimization of an industrial hydrocracker plant. *Journal of Petroleum Science and Engineering*, *78*(3-4), pp.627-636.

Alhajri, I., Elkamel, A., Albahri, T., 2008. A nonlinear programming model for refinery planning and optimisation with rigorous process models and product quality specifications. International Journal of Oil Gas and Coal Technology.

Al-Sobhi, S. A., Elkamel, A., 2015, Simulation and Optimization of Natural Gas Processing and Production Network consisting of LNG, GTI, and methanol facilities. Journal of Natural Gas Science and Engineering, Volume 23, pp. 500-508.

Aminian, J. and Shahhosseini, S., 2008. Evaluation of ANN modeling for prediction of crude oil fouling behavior. *Applied thermal engineering*, *28*(7), pp.668-674.

Bartz-Beielstein, T., Naujoks, B., Stork, J. and Zaefferer, M., 2016. Tutorial on surrogate-assisted modelling.

Biegler, L.T. and Grossmann, I.E., 2004. Retrospective on optimization. Computers & Chemical Engineering, 28(8), pp.1169-1192.

Binder, T., Blank, L., Bock, H.G., Bulirsch, R., Dahmen, W., Diehl, M., Kronseder, T., Marquardt, W., Schlöder, J.P. and von Stryk, O., 2001. Introduction to model based optimization of chemical processes on moving horizons. In *Online optimization of large scale systems* (pp. 295-339). Springer, Berlin, Heidelberg.

Bloomberg, 2018. Bloomberg [Online], Available at: https://www.bloomberg.com/news/articles/2018-09-28/saudi-arabia-faces-the-ultimate-oil-test-producing-at-the-limit [Accessed 10 February 2019]

Bollas, G.M., Papadokonstadakis, S., Michalopoulos, J., Arampatzis, G., Lappas, A.A., Vasalos, I.A. and Lygeros, A., 2003. Using hybrid neural networks in scaling up an FCC model from a pilot plant to an industrial unit. *Chemical Engineering and Processing: Process Intensification*, *42*(8-9), pp.697-713.

Bollas, G.M., Papadokonstantakis, S., Michalopoulos, J., Arampatzis, G., Lappas, A.A., Vasalos, I.A. and Lygeros, A., 2004. A Computer-Aided Tool for the Simulation and Optimization of the Combined HDS–FCC Processes. *Chemical Engineering Research and Design*, *82*(7), pp.881-894.

CEIC, 2018. ISI Emerging Market Group Limited [Online], Available at: https://www.ceicdata.com/en/indicator/china/oil-consumption [Accessed 03 February 2019]

Clews R. J., 2016. Project Finance for the International Petroleum Industry, pp. 83-99.

Cortes, C. and Vapnik, V., 1995. Support-vector networks. *Machine learning*, *20*(3), pp.273-297.

Dakin, R.J., 1965. A tree-search algorithm for mixed integer programming problems. *The computer journal*, *8*(3), pp.250-255.

Devold, H., 2013. Oil and gas production handbook. An introduction to oil and gas production, transport, refining and petrochemical industry, pp. 4-10.

Deb, K., 2005. Multi-objective optimization. In Search methodologies, pp. 273-316. Springer, Boston, MA.

Diwekar, U., 2008. *Introduction to applied optimization* (Vol. 22). Springer Science & Business Media.

Floudas, C.A., 2013. *Deterministic global optimization: theory, methods and applications* (Vol. 37). Springer Science & Business Media.

Forrester, A., Sobester, A. and Keane, A., 2008. *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons.

Gennert, M.A., Yuille, A.L., 1988, December. Determining the optimal weights in multiple objective function optimization. In *ICCV*, pp. 87-89.

Glassey, J. and von Stosch, M., 2018. Benefits and Challenges of Hybrid Modeling in the Process Industries: An Introduction. In *Hybrid Modeling in Process Industries* (pp. 1-12). CRC Press.

Grefenstette, J., Gopal, R., Rosmaita, B. and Van Gucht, D., 1985. Genetic algorithms for the traveling salesman problem. In *Proceedings of the first International Conference on Genetic Algorithms and their Applications* (Vol. 160, No. 168, pp. 160-168). Lawrence Erlbaum.

Grossmann, I.E., Biegler, L.T., 1995. Optimizing Chemical Processes. CHEMTECH, Dec. 27-35.

Grossmann, I.E., 2009. Research Challenges in Planning and Scheduling for Enterprise-Wide Optimization of Process Industries. Computer Aided Chemical Engineering, pp. 15-21.

Grossmann, I.E., 2012. Advances in mathematical programming models for enterprise-wide optimization. Computers and Chemical Engineering, pp. 2-18.

Guillén-Gosálbez, G., 2011. A novel MILP-based objective reduction method for multi-objective optimization. Application to environmental problems Comput. Chem. Eng., pp. 1469-1477.

Gupta, A., 2018. Introduction to Deep Learning: Part 1.

Hellmann, M., 2001. Fuzzy logic introduction. *Université de Rennes*, *1*.

Ibrahim, D., Jobson, M., Li, J. and Guillén-Gosálbez, G., 2018. Optimization-based design of crude oil distillation units using surrogate column models and a support vector machine. *Chemical Engineering Research and Design*, *134*, pp.212-225.

Jakkula, V., 2006. Tutorial on support vector machine (svm). *School of EECS, Washington State University*, *37*.

Jain, A.K., Mao, J. and Mohiuddin, K.M., 1996. Artificial neural networks: A tutorial. *Computer*, (3), pp.31-44.

Johansen, T.A. and Foss, B.A., 1992, June. Representing and learning unmodeled dynamics with neural network memories. In *1992 American Control Conference* (pp. 3037-3043). IEEE.

Kondili, E., Pantelides, C.C. and Sargent, R.W.H., 1993. A general algorithm for short-term scheduling of batch operations—I. MILP formulation. *Computers & Chemical Engineering*, *17*(2), pp.211-227.

Kuhn, H. W.; Tucker, A. W., 1951. Nonlinear Programming. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 481--492, University of California Press, Berkeley, Calif.

Li, D., and Li, X. 2015. Global Optimization of an Industrial Natural Gas Production Network. IFAC-PapersOnLine, 9th IFAC Symposium on Advanced Control of Chemical Processes ADCHEM 2015, Whistler

Liu S, Alhasan I, Papageorgiou LG., 2016. A mixed integer linear programming model for the optimal operation of a network of gas oil separation plants. Chem Eng Res Des 111:147–160.

Mahajan, A., 2010. Presolving mixed–integer linear programs. *Wiley Encyclopedia of Operations Research and Management Science*.

Martins, F., Costa C.A., 2010. Multiobjective optimization with economic and environmental objective functions using Modified Simulated Annealing, 28. pp. 919-924.

McCulloch, W.S. and Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5*(4), pp.115-133.

Meindl, B. and Templ, M., 2012. Analysis of commercial and free and open source solvers for linear optimization problems. *Eurostat and Statistics Netherlands within the project ESSnet on common tools and harmonised methodology for SDC in the ESS*, *20*.

Menezes, B.C., Grossmann, I.E. and Kelly, J.D., 2017. Enterprise-wide optimization for operations of crude-oil refineries: Closing the procurement and scheduling gap. In *Computer Aided Chemical Engineering* (Vol. 40, pp. 1249-1254). Elsevier.

Mitchell, T., 1997. Machine Learning, McGraw-Hill Computer science series.

Mjalli, F.S. and Al-Mfargi, A., 2009. Neural network–based heat and mass transfer coefficients for the hybrid modeling of fluidized reactors. *Chemical Engineering Communications*, *197*(3), pp.318-342.

Mokhatab, S., Poe, W., Mak, J., 2019. Handbook of Natural Gas Transmission and Processing, pp. 643-668.

Neiro, S.M. and Pinto, J.M., 2004. A general modeling framework for the operational planning of petroleum supply chains. *Computers & Chemical Engineering*, *28*(6-7), pp.871-896.

Nocedal, J. and Wright, S., 2006. *Numerical optimization*. Springer Science & Business Media.

OPEC, 2019 [Online], Available at: https://www.opec.org/opec_web/en/about_us/169.htm [Accessed 11 February 2019]

Papageorgiou, L.G., Rotstein, G.E. and Shah, N., 2001. Strategic supply chain optimization for the pharmaceutical industries. *Industrial & Engineering Chemistry Research*, *40*(1), pp.275-286.

Papageorgiou, L.G., 2009. Supply chain optimisation for the process industries: Advances and opportunities. *Computers & Chemical Engineering*, *33*(12), pp.1931-1938.

Pantelides, C.C., 1994. Unified frameworks for optimal process planning and scheduling. In *Proceedings on the second conference on foundations of computer aided operations* (pp. 253-274). Cache Publications New York.

Pozo, C., Ruíz-Femenia R., Caballero, J., Guillén-Gosálbez, G., 2012. On the use of Principal Component Analysis for reducing the number of environmental objectives in multi-objective optimization: application to the design of chemical supply chains. Chem. Eng. Sci., pp. 146-158.

Psichogios, D.C. and Ungar, L.H., 1992. A hybrid neural network-first principles approach to process modeling. *AIChE Journal*, *38*(10), pp.1499-1511.

Safavi, A.A., Nooraii, A. and Romagnoli, J.A., 1999. A hybrid model formulation for a distillation column and the on-line optimisation study. *Journal of Process Control*, *9*(2), pp.125-134.

Sahebi, H., Nickel, S. and Ashayeri, J., 2014. Strategic and tactical mathematical programming models within the crude oil supply chain context—A review. *Computers & Chemical Engineering*, *68*, pp.56-77.

Sahinidis, N.V., 1996. BARON: A general purpose global optimization software package. *Journal of global optimization*, *8*(2), pp.201-205.

Saudi Aramco, 2017. Saudi Aramco Annual Review, Available at: https://www.saudiaramco.com/-/media/images/annual-review-2017/pdfs/en/2017-annualreview-full-en.pdf [Accessed 11 February 2019]

Senthamaraikkannan G., Chakrabarti, D., Prasad V., 2014. Future Energy, pp. 271-288.

Shah, N., 1998. Single and multisite planning and scheduling: current status and future challenges. Proceedings of the third international conference on foundations of computer-aided process operations, pp. 75–90.

Shah, N., 2004. Pharmaceutical supply chains: key issues and strategies for optimisation. *Computers & chemical engineering*, *28*(6-7), pp.929-941.

Shah, N., 2005. Process industry supply chains: advances and challenges. Computer Chem. Eng., pp. 1225-1235.

Shapiro, J., 2006. *Modeling the supply chain*. Nelson Education.

Shayya, W.H. and Sablani, S.S., 1998. An artificial neural network for non-iterative calculation of the friction factor in pipeline flow. *Computers and Electronics in Agriculture*, *21*(3), pp.219-228.

Simpson, T.W., Poplinski, J.D., Koch, P.N. and Allen, J.K., 2001. Metamodels for computer-based engineering design: survey and recommendations. *Engineering with computers*, *17*(2), pp.129-150.

Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P.Y., Hjalmarsson, H. and Juditsky, A., 1995. Nonlinear black-box modeling in system identification: a unified overview. *Automatica*, *31*(12), pp.1691-1724.

Spielman, D.A. and Teng, S.H., 2004. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM)*, *51*(3), pp.385-463.

Su, H.T. and McAvoy, T.J., 1993. Integration of multilayer perceptron networks and linear dynamic models: a Hammerstein modeling approach. *Industrial & engineering chemistry research*, *32*(9), pp.1927-1936.

Talbi, E.G. and Muntean, T., 1993. Hill-climbing, simulated annealing and genetic algorithms: a comparative study and application to the mapping problem. *Proceedings of the Twenty-sixth Hawaii International Conference on System Sciences* (Vol. 2, pp. 565-573). IEEE.

Tautenhain, C.P.S., Barbosa-Póvoa, A.P., Nascimento M.C.V., 2019. A multi-objective matheuristic for designing and planning sustainable supply chains. Computer Ind. Eng.

Thompson, M.L. and Kramer, M.A., 1994. Modeling chemical processes using prior knowledge and neural networks. *AIChE Journal*, *40*(8), pp.1328-1340.

U.S. EIA, 2017. U.S. EIA [Online], Available at: https://www.eia.gov/beta/international/analysis_includes/countries_long/Saudi_Arabia/saudi_arabia.pdf [Accessed 04 April 2019]

U.S. EIA, 2018. U.S. EIA [Online], Available at: https://www.eia.gov/energyexplained/index.php?page=oil_use [Accessed 03 February 2019]

Vapnik, V., 1963. Pattern recognition using generalized portrait method. *Automation and remote control*, *24*, pp.774-780.

Venkatasubramanian, V., 2019. The promise of artificial intelligence in chemical engineering: Is it here, finally?. *AIChE Journal*, *65*(2), pp.466-478.

Viswanathan, J. and Grossmann, I.E., 1990. A combined penalty function and outer-approximation method for MINLP optimization. Computers & Chemical Engineering, 14(7), pp.769-782.

Von Stosch, M., Oliveira, R., Peres, J. and de Azevedo, S.F., 2014. Hybrid semi-parametric modeling in process systems engineering: Past, present and future. *Computers & Chemical Engineering*, *60*, pp.86-101.

Wang, P., 2003. *Development and applications of production optimization techniques for petroleum fields* (Doctoral dissertation, Stanford University).

Wassick, J.M., 2009. Enterprise-wide optimization in an integrated chemical complex. *Computers & Chemical Engineering*, *33*(12), pp.1950-1963.

Werbos, P.J., 1974. Beyond regression: New tools for prediction and analysis in the behavioral sciences. Ph. D. thesis, Harvard University, Cambridge, MA, 1974.

Yeniay, Ö., 2005. Penalty function methods for constrained optimization with genetic algorithms. Mathematical and computational Applications, 10(1), 45-56.

Zahedi, G., Fgaier, H., Jahanmiri, A. and Al-Enezi, G., 2006. Artificial neural network identification and evaluation of hydrotreater plant. *Petroleum science and technology*, *24*(12), pp.1447-1456.

Zahedi, G., Mohammadzadeh, S. and Moradi, G., 2008. Enhancing gasoline production in an industrial catalytic-reforming unit using artificial neural networks. *Energy & Fuels*, *22*(4), pp.2671-2677.

Zendehboudi, S., Zahedi, G., Bahadori, A., Lohi, A., Elkamel, A. and Chatzis, I., 2014. A dual approach for modelling and optimisation of industrial urea reactor: Smart technique and grey box model. *The Canadian Journal of Chemical Engineering*, *92*(3), pp.469-485.

Zendehboudi, S., Rezaei, N. and Lohi, A., 2018. Applications of hybrid models in chemical, petroleum, and energy systems: A systematic review. *Applied energy*, *228*, pp.2539-2566.

Li, K. J., Thijssen, N., & Mittendorff, I. (2007, January 1). Enhancement Of Oil & Gas Production Using Integrated Network Analysis. International Petroleum Technology Conference. doi:10.2523/IPTC-11638-MS