DISSERTATION

Data-Driven Evolutionary Optimisation for the design parameters of a Chemical Process: A Case Study

Author:
Liezl STANDER (715347)

Supervisor:
Prof. Terence VAN ZYL
Dr. Matthew WOOLWAY

UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

submitted to
the Faculty of Science, in fulfilment of the requirements for the degree of
MSc in Computer Science

in the

Wits Institute of Data Science (WIDS)
School of Computer Science and Applied Mathematics

June 21, 2021

Declaration of Authorship

I, Liezl STANDER (715347), declare that this Dissertation titled, "Data-Driven Evolutionary Optimisation for the design parameters of a Chemical Process: A Case Study" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this Dissertation has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this Dissertation is entirely my own work.
- I have acknowledged all main sources of help.
- Where the Dissertation is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: LStander

Date: June 21, 2021

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Abstract

Faculty of Science
School of Computer Science and Applied Mathematics

MSc in Computer Science

Data-Driven Evolutionary Optimisation for the design parameters of a Chemical Process: A Case Study

by Liezl STANDER (715347)

The optimisation of Chemical plant design and operation has proven to be challenging due to the complexity of real-world systems. The complexity translates into high computational costs from the mathematical representations and simulation models for these systems. Research has illustrated the benefits of using surrogate models as substitutes for computationally expensive models. This research investigates two main concepts. The first one being the resource cost reduction when implementing surrogate assisted genetic algorithms for the optimisation of computationally expensive simulation models representing chemical systems. The second component focuses on determining the robustness of these algorithms towards stochastic and multi-objective systems. Two main algorithms were developed to optimise four different chemical plant systems. The Chemical Plant System - Basic (CPS-B) is a stochastic chemical process including buffer tanks, processing units, and a tank with multiple feed streams. The Chemical Plant System - Parallel (CPS-P) and Chemical Plant System - Feedback (CPS-F) are more complex variants of the CPS-B introducing additional complexities in the form of parallel and feedback loop systems respectively. The Surrogate Assisted Genetic Algorithm (SA-GA) was used to optimise these three systems. The SA-GA algorithm was adapted for multi-objective optimisation. The new adapted algorithm called the Surrogate Assisted NSGA-II (SA-NSGA) algorithm was tested on a popular literature case, the Pressure Swing Adsorption (PSA) system. The optimisation results for all the chemical systems illustrated that using Genetic Algorithms as an optimisation framework for complex stochastic, single and multi-objective chemical plant systems results in significant computational benefits. Introducing Machine Learning Surrogate models as substitutes for computationally expensive simulation models into a Genetic Algorithm framework yielded significant computational efficiency improvements. The optimisation of CPS-B, CPS-P, CPS-F, and PSA achieved 1.8, 1.74, 1.95, and 5 times speedup of the total elapsed run time, despite the increased complexity in the systems. It is worth noting that the SA-GA and SA-NSGA algorithms implemented in this research yielded results confirming both their flexibility and robustness towards more complex stochastic, single and multi-objective systems.

Acknowledgements

I would like to thank Professor Terence Van Zyl and Dr. Matthew Woolway for their continued support, guidance, teaching, and sharing their wealth of knowledge with me throughout this research project. I would also like to thank my Mentor, Ebert Caewood for the initial simulation model which he had developed and shared with me which formed the basis of this research.

Contents

List of Figures

List of Tables

List of Abbreviations

CPS	**C**hemical **P**lant **S**ystem
PSA	**P**ressure **S**wing **A**dsorption
ML	**M**achine **L**earning
KPI	**K**ey **P**erformance **I**ndicator
ANN	**A**rtificial **N**eural **N**etworks
m.u	**m**onetary **u**nits
MTBF	**M**ean **T**ime **B**etween **F**ailures
MTTR	**M**ean **T**ime **T**o **R**epair
GA	**G**enetic **A**lgorithm
NSGA	**N**on-dominated **S**orting **G**enetic **A**lgorithm
CPS - B	**C**hemical **P**lant **S**ystem **B**asic
CPS - P	**C**hemical **P**lant **S**ystem **P**arallel
CPS - F	**C**hemical **P**lant **S**ystem **F**eedback
SA - GA	**S**urrogate **A**ssisted **G**enetic **A**lgorithm
SA - NSGA	**S**urrogate **A**ssisted **N**on-dominated **S**orting **G**enetic **A**lgorithm
EA	**E**volutionary **A**lgorithm
HV	**H**yper**V**olume
GD	**G**enerational **D**istance
IGD	**I**nverted **G**enerational **D**istance
AES	**A**verage number of **E**valuations to **S**ucces

List of Publications

[1] L. Stander, M. Woolway, and T. L. van Zyl, "Data-driven evolutionary optimisation for the design parameters of a chemical process: A case study", in *2020 23rd International Conference on Information Fusion (FUSION 2020)*, 2020.

[2] L. Stander, M. Woolway, and T. L. van Zyl, "Extended surrogate assisted continuous process optimisation", in *2020 7th International Conference on Soft Computing Machine Intelligence (ISCMI)*, 2020, pp. 275–279. DOI: 10.1109/ISCMI51676.2020.9311560.

[3] L Stander, M Woolway, and T. van Zyl, "Surrogate assisted optimisation for a pressure swing adsorption system", *Neural Computing and Applications*, In Review.

Chapter 1

Introduction

1.1 Background

Optimisation of chemical processes plays a crucial role in decision-making related to process improvement and enhancing performance while reducing costs [4]. The Industrial Chemicals Industry contends with a significant amount of complexity, usually involving a large number of control variables and multiple (often) conflicting objectives [5]. Optimisation of systems like these requires the system's behaviour to be represented in a model [6]. A popular approach to these models within chemical plant design is computer simulation as system representations [7]–[9]. The simulation models play a central role in enabling the optimisation of these systems as they are often used as the primary evaluation platform [8], [10].

Evolutionary Algorithms (EA) are common in multi-objective optimisation domains [6], [11]–[13]. EA techniques are flexible across a broad range of problems as they do not require information specific to the problem such as derivatives [6]. The framework for EA's start with an initial population (set of solutions) and converges these towards improved solution spaces using selection, crossover, and mutation operators [14]; with the literature showing promising results when applied to chemical process problems [15]–[18].

Combining simulation and EAs can result in higher computational costs as multiple evaluations of the simulation model may be required to achieve an optimal solution. *Surrogate Models* are computationally cheaper models acting as proxies for the original computationally costly simulation models [7], [19]. The existing literature on the topic of surrogate assisted optimisation has demonstrated that optimal solutions can be achieved with fewer function evaluations [11], [20]–[22].

For various complex chemical systems, the use of a Surrogate Assisted Genetic Algorithm speeds up the optimisation of complex chemical systems while maintaining the quality of solutions [1], [2]. This research focuses on testing Surrogate Assisted Genetic Algorithms on four different complex chemical systems to determine the computational efficiency gains. The order of the systems represents the order of complexity (i.e. lowest complexity systems first and highest complexity systems last). The first system, representing the motivating example of this research, is a basic chemical process including buffer tanks, processing units, and a tank with multiple feed streams. This system will be referred to as the Chemical Plant System - Basic (CBP-B) and is detailed in Section 2.3. The second

and third systems, namely the Chemical Plant System - Parallel (CPS-P) and Chemical Plant System - Feedback (CPS-F) are two more complex variants of CPS-B. The additional complexities introduced include a parallel system in CPS-P and a feedback loop in CPS-F. The details of CPS-P and CPS-F are detailed in Sections 2.4 and 2.5 respectively. The fourth chemical process is the Pressure Swing Adsorption System (PSA). The optimisation of the PSA system involves maximising two objectives, namely the CO_2 *purity* and *recovery*. A specific variant of the PSA system studied by Yancy-Cabellero *et al.* (2020) [23] has been used in this research. The PSA introduces a spectrum of complexities, including multiple feedback loops. The details of the PSA system are included in Section 2.6.

The optimisation of PSA systems receive a significant amount of interest [24]–[30]. The PSA system usually involves optimisation across multiple objectives. NSGA-II is a popular multi-objective optimisation approach used for multiple-objective type systems [23], [26], [28]. Additionally, NSGA-II is robust in optimising a wide variety of problems [30]–[32]. Problems having multiple objectives, usually have a set of "Pareto-optimal" solutions, none of which is better than the others [33].

The first three CPS systems involve multiple objectives that can be combined into one for optimisation. The optimisation performance metric for the first three systems is the annual revenue generated from the final product produced. The multi-objective nature of the PSA system necessitates additional performance metrics. Riquelme *et al.* (2015) [34] investigated and reviewed 54 performance metrics for multi-objective optimisation and concluded that the most popular metrics include the generational distance (GD),hypervolume (HV), epsilon indicator and the inverted generational distance (IGD) [34]. The HV, GD, IGD, and visual inspection of the Pareto frontiers have been adopted for this research.

1.2 Problem Statement

Having an accurate surrogate model to reproduce the results from a complex simulation model has computational cost benefits. Implementing these surrogate models in combination with evolutionary algorithms will reduce the computational resource cost associated with the optimisation of Chemical Plant design and operation.

Although these simulation models provide an improved representation of the system, most of the current optimisation in literature is implemented with a non-stochastic target function. The stochasticity of the simulation introduces challenges for the surrogate model to learn and the Genetic Algorithm (GA) to search the solution space as each simulation run yields different results. The additional complexity of multiple objectives in a chemical system also poses a challenge to the optimisation algorithm as multiple optimal solutions exist.

In response to this, we propose to derive, implement and evaluate the performance of an algorithm that makes use of both the simulation and the surrogate model at certain criteria to evaluate the fitness function for the optimisation.

1.3 Significance and Motivation

The decision-making process within chemical process performance enhancements, process improvement, and cost reduction relies heavily on chemical process optimisation [4]. This dependence highlights the importance of having true optimal design and operation of your chemical systems. The chemical systems that are categorised within the Industrial Chemicals Industry involve a significant amount of complexity. The chemical systems usually involve multiple (often) conflicting objectives and are made up of large numbers of control variables [5].

Stochastic simulation models are important to enable the accurate representation of the systems. Additional complexities in the form of multiple objectives also need to be taken into account for accurate system representation. Both of these components result in higher computational costs for chemical plant design and operation optimisation. To ensure that costs are minimised, an accurate surrogate model to reproduce the results from computationally expensive simulation models can be implemented. The optimisation process can be done with reduced computational resource cost which results in reduced overall costs.

1.4 Research Aims and Objectives

The research aims to investigate two primary concepts. The first being the potential computational cost reduction in making use of surrogate assisted genetic algorithms for chemical plant design and operation optimisation. The second concept is to investigate the robustness of the surrogate assisted genetic algorithms towards stochastic and multi-objective systems.

The aims of this research project will be achieved through the following objectives (goals):

- Investigate the current algorithms and approaches to optimising chemical plant design and operation in the literature;
- Develop and source various stochastic simulation models representing chemical plan systems;
- Source a simulation representing a real-world complex chemical system that includes multiple objectives;
- Develop the surrogate assisted algorithms to optimise the various simulation models.
- Collect data on the performance of the algorithms on the various test systems.
- Analyse and discuss the results and summarise the findings.

1.5 Research Questions

1. What are the impacts of the stochasticity of the simulation model on the optimisation algorithms?
2. What are the impacts of introducing multiple objectives into the optimisation algorithms?
3. By how much, can the addition of a Machine Learning surrogate model in the optimisation of complex chemical systems represented by simulation models, improve the speed of convergence towards the optimal solution?

1.6 Delineations, Limitations and Assumptions

This research will only focus on four specific chemical plant system test cases detailed in Chapter 2. As a result, the outcomes of the research may not be generalisable to case studies outside the realm of chemical system design and operation.

1.7 Outline

The remainder of the dissertation is structured as follows. Chapter 2 explores the available literature on the research topics. Chapter 3 outlines the research methodology and algorithms executed in this research. The research results have been summarised and discussed in Chapter 4 and the conclusions are included in Chapter 5.

Chapter 2

Literature Review

This chapter details the topic of genetic algorithms and the existing literature related to surrogate assisted optimisation within the field of chemical engineering, including the unpacking of various papers, their techniques, cases, and results. The four chemical plant system use cases are detailed in this chapter.

2.1 Genetic Algorithms

Genetic algorithms (GA) is an extremely popular and effective optimisation technique for solving complex, often multi-objective, optimisation problems [6], [11], [12]. The algorithm has been developed in such a way that it can simultaneously process multiple solutions. This capability makes the genetic algorithms robust towards multi-objective problems [13]. Genetic algorithms are also able to be applied to a broad range of different types of problems because they do not require any information specific to the system, like derivatives [6]. The general framework of the genetic algorithm originated from organic evolution [14]. The general process of the GA starts with an initial population set and progressively moves towards improved solution spaces [14]. This progression is achieved by implementing selection, crossover and mutation techniques [14], [15], [17].

2.2 Surrogate Assisted Optimisation

Multi-objective optimisation in the field of chemical engineering is an extremely popular topic illustrated by the existing substantive literature. Reduced function evaluation requirements and improved computation costs for optimisation are some of the benefits shown by from the topic of surrogate assisted optimisation [11], [20], [31], [35]–[39].

Wahid et al. [36] investigated minimising the compression energy in a single mixed refrigerant process of natural gas liquefaction. The paper illustrated significant time reductions by using surrogate assisted optimisation, specifically using Radial basis functions as their surrogate model. This approach allowed their optimisation algorithms to be reduced from 5.87 hours to only 6.75 minutes (for the Genetic Algorithm implementation) while still improving their objective value from the base case. The solutions achieved are either extremely close or at the optimal solutions.

Beck et al. [31] achieved a five times reduction in computational effort by implementing a Kriging surrogate model. The system they were optimising was the design of a Vacuum/pressure swing adsorption system for the separation of CO2 and N2 mixture.

The optimisation of a Pressure Swing Adsorption unit for methane nitrogen separation was investigated by Anna et al. [35]. The objectives were to maximise Purity and Recovery of Nitrogen. The optimisation time decreased from 15.7 hours to only 50 seconds when they implemented an Artificial Neural Network as their surrogate model while simultaneously improving their recovery and purity objectives.

Ibrahim et al. [20] investigated the optimisation of the design of crude oil distillation units using Artificial Neural Networks as the surrogate model. The results of the paper presented that the surrogate assisted optimisation yielded a 44% time reduction while maintaining the accuracy and optimality of the solution.

Another paper investigating the optimisation of crude oil distillation units by Shi et al. [39] used adaptive radial basis functions for their surrogate modelling strategy and achieved a reduction in function calls from 1236 to 562.

A review paper detailing the advances in surrogate based modeling, feasibility analysis and optimisation done by Bhoesekar et al. [40] highlights the popularity of surrogate modeling in a range of applications within multiple engineering disciplines. Each surrogate modeling approach yields its own set of trade-offs making the selection process of the correct model for the problem quite challenging. Categorising the problem into either feasibility, prediction or optimisation focuses the selection of surrogate models as certain surrogate models are more suited towards specific problem categories.

Carpio et al. [38] made use of surrogate modeling and achieved better numerical results that were more stable, requiring less rigorous simulation model evaluations in comparison to the classical optimisation and achieving near-optimal solutions. The case that was investigated was the optimisation of a Biorefinery aimed at profit maximisation.

Complex problems such as the one detailed by Chugh et al. [11] where an Industrial blast furnace with eight competing objectives was successfully simultaneously optimised using surrogate models for the various objectives.

Various optimisation techniques are used in case studies focusing on the optimisation of chemical plant design [20], [24], [25], [31], [39], [41]. Evolutionary Algorithms (EA) is a popular technique within literature which is illustrated by the amount of existing research [20], [31], [39], [41]. This includes Genetic Algorithms (GA) [20], [31] and Differential Evolution [41].

The optimisation of a Pressure Swing Adsorption (PSA) unit for methane nitrogen separation was investigated by Anna et al. [35]. The objectives were to maximise Purity and Recovery of Nitrogen. The optimisation time decreased from 15.7 hours to only 50 seconds when they implemented an Artificial Neural Network as their surrogate model while simultaneously improving their recovery and purity objectives.

The PSA system is an extremely popular chemical process within literature [24], [25], [31], [37], [42]. This system achieves the separation of certain required gases and plays pivotal roles in applications such as CO2 capture [24], [25], the Separation and purification of

Hydrogen [24], and other Industrial systems requiring gas separation [26]. The optimisation of the PSA system has received a significant amount of interest, based on the number of available literature articles [24], [25], [31], [37]. As detailed in the paper by Anna et al. [35], surrogate modelling has been a solution implemented across various studies for the optimisation of the PSA system [24], [25], [31], [37].

The remainder of this chapter includes the description and details of the four systems that were used to test the algorithms described in Chapter 3. The first system, referred to as the Chemical Plant System Basic (CPS-B), represents the motivating example for the research and is detailed in Section 2.3. The next two systems are more complex variants of CPS-B which include additional complexities such as an additional parallel system and a feedback loop. The system including the parallel stream is referred to as the Chemical Plant System Parallel (CPS-P) and is detailed in Section 2.4. The feedback loop is included in Section 2.5 and is referred to as the Chemical Plant System Feedback (CPS-F). Due to the similarities in the CPS-B,CPS-P and CPS-F, only the additional complexities have been detailed in Sections 2.4 and 2.5. The simulation modelling approach remained the same with differences including the decision variables and their bounds and the objective functions. The final system is a Pressure Swing Adsorption (PSA) system for post-combustion capture of carbon and is detailed in Section 2.6.

2.3 Chemical Plant System - Basic

2.3.1 Design Parameters

CPS-B is illustrated in Figure 2.1. This system is a general representation of a chemical plant system and has been used as the motivating example for this research.

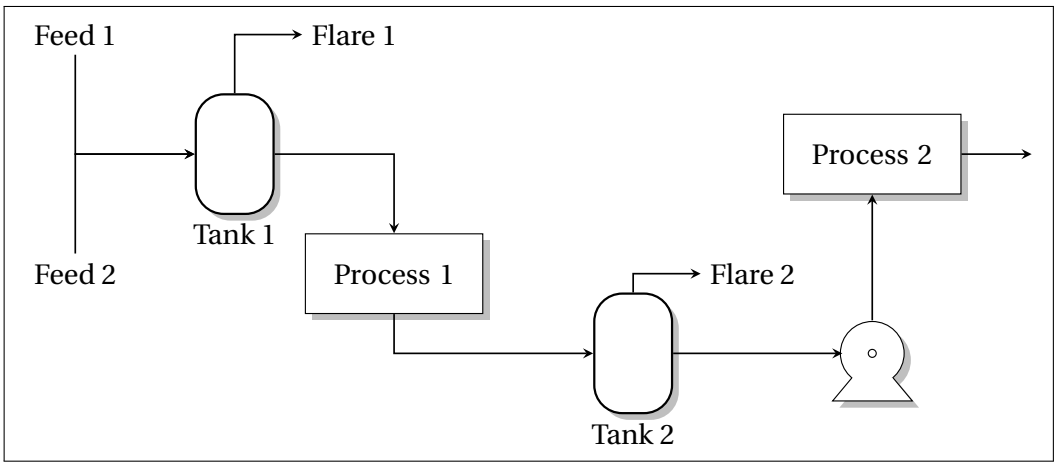

FIGURE 2.1: Chemical Plant System - Basic

The plant system in Figure 2.1 starts with two feed streams flowing into Tank 1, representing the first buffer tank. The feed from Tank 1 is processed by Process 1 and gets stored in the second buffer tank, Tank 2. Flare 1 is activated if tank 1's capacity is exceeded due to unplanned events such as process 1 not being able to absorb the feed streams. Flare 1 then

flares (release into the atmosphere through a process of burning) the unprocessed product. The product in Tank 2 is then transported via Pump 1 to Process 2, which produced the final product. Flare 2 is activated if the tank 2's capacity is exceeded due to unplanned events such as Process 2 not being able to absorb the feed stream or pump 1 being out of service. The product produced by process 2 is the system's major revenue generator. The cost of flaring product and failures on the various system components (Process 1, Process 2 and Pump 1) negatively impact the final revenue of the system. The minimum number of spares to keep on-site, tank capacities,minimum number of pumps to procure and the pump capacity were the decision variables identified as the primary features to optimise the system. Table 2.1 includes the objective functions for CPS-B.

TABLE 2.1: Objective Functions for Chemical Plant System - Basic

No.	Max/Min	Variable
1	Maximise	Final Product Stream
2	Minimise	Flare 1
3	Minimise	Flare 2
4	Minimise	Tank Sizes
5	Minimise	Pump Sizes
6	Minimise	Holding stock level
7	Minimise	Amount of stock purchased

2.3.2 Simulation Model

The operation of the CPS-B system was replicated in a continuous-time stochastic event simulation model. All seven objective functions described in Table 2.1 are able to be expressed into one single objective. At every run of the simulation model, the seven objectives were evaluated resulting in a single Revenue (monetary units/year (m.u)) value.

The overall duration of the simulation run is 365 days to allow the model to iterate all the real-life events at a frequency experienced in practice. An hourly time increment was chosen for the simulation model to account for the lowest true frequency variables.

Table 2.2 illustrates the design specifications for the system. These specifications include the maximum and minimum capacities of the various components in CPS-B.

TABLE 2.2: Processing Capacities of Chemical Plant System - Basic

Equipment/Process	Minimum	Maximum
Process 1	100 m^3/hour	100 m^3/hour
Process 2	100 m^3/hour	100 m^3/hour
Pump 1	60 m^3/hour	120 m^3/hour
Tank 1 & 2	50 m^3	1000 m^3/hour

The mean time between failures (MTBF) summarised in Table 2.3 represent the first component of modelling the failures. These times were implemented as sampled values

from the exponential distribution due to its applicability in the reliability environment for equipment failure. The assumptions on failures include:

- Repairs done on equipment achieve an almost new state;
- Failures are not as a result of deterioration; and
- Failures are entirely random.

The mean time to repair (MTTR) or failure times included in Table 2.4 represents the second component of modelling the system failures. The log-normal (LOGN), uniform (UNIF), and normal (NORM) distributions were used and determined from industry norms. The “Spares Available" column illustrates the relationship between the failure times and the available spares on-site (i.e. if spares are available, the failure time is quicker, and where there is a N/A in the column it represents that there is no relationship between the spares and the failure distribution).

TABLE 2.3: Mean Time Between Failures (MTBF) Distributions

Equipment/Process	**MTBF (Hours)**
Process 1	EXPO(480)
Process 2	EXPO(720)
Pump 1	EXPO(600)

TABLE 2.4: Mean Time To Repair (MTTR) Distributions

Equipment/Process	**MTTR (Hours)**	**Spares Available**
Process 1	LOGN(2.5, 0.5) +2	Yes
Process 1	UNIF(72, 168)	No
Process 2	NORM(36,5)	N/A
Pump 1	LOGN(1.5, 0.46) + 4	N/A

The empirical distributions illustrated in Figure 2.2 were randomly sampled for the flow rates for feed 1 and 2. Two separate peaks at either end of the figure are evident in the distributions of the data. The peak on the left represents times when the feed producing units are offline due to planned or unplanned maintenance. The peak on the right represents stable operations for the units. These peaks were included in the data to incorporate the impacts of any challenges upstream resulting in low to no feed to the system.

2.4 Chemical Plant System - Parallel

2.4.1 Design Parameters

CPS-P is represented in Figure 2.3. This system is different to CPS-B, due to a parallel system being introduced. The additional complexity from the parallel system results in two stochastic systems relying on a centralised spares pool. The shared spares pool needs to

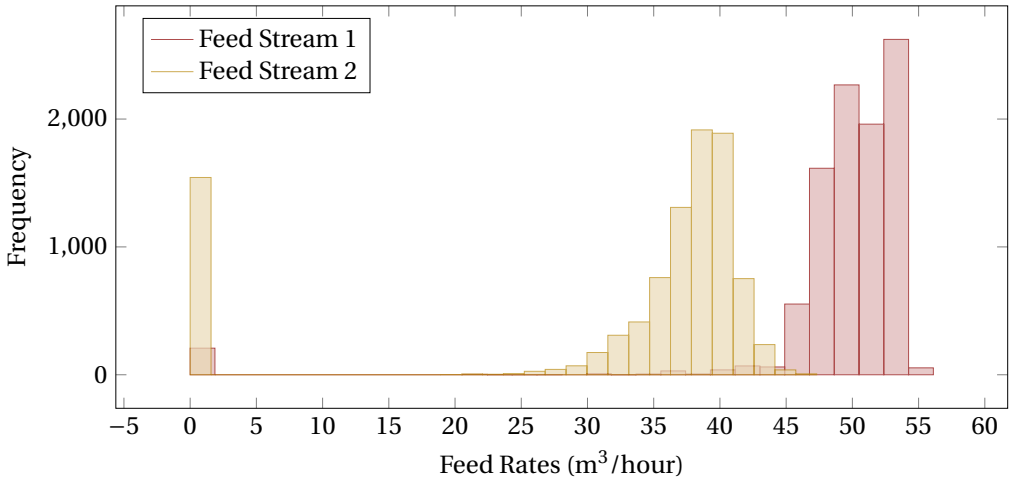

FIGURE 2.2: Empirical Distributions of Feed Streams

TABLE 2.5: Fitness Function Values for the Chemical Plant System - Basic

Equipment/Process	**Value (Monetary Units)**	**Cost/Revenue**
Product	$6042/m^3$	Revenue
Flare 1	$2848/m^3$	Cost
Flare 2	$3907/m^3$	Cost
Per Tank	$9.94 \times 10^7 + 1.52 \times 10^6/m^3$	Cost
Per Pump	$4.44 \times 10^6 + 2.96 \times 10^5/m^3$	Cost

be optimised to reduce the failure times for both systems while balancing the cost associated with the number of spares bought. The minimum number of spares to keep on-site, tank capacities,minimum number of pumps to procure and the pump capacities represent the decision variables for the system. The maximum bounds of the number of spares to keep and purchase were doubled as well as introducing three additional decision variables, specifically two tank capacities and a pump capacity.

2.5 Chemical Plant System - Feedback

2.5.1 Design Parameters

CPS-F is illustrated in Figure 2.4. This system includes the complexity of product that needs to be reprocessed due to quality specification failures (recycle stream). To quantify this recycle stream, α represents the ratio of recycled product to the final product. To reduce this ratio, more man-hours can be spent maintaining the system. These maintenance hours spent on the system (online basis) represents the critical lever to influencing this ratio. The α ratio was incorporated into the model by sampling from a normal distribution with a mean that is a function of the number of maintenance hours spent and has a maximum variance of 10% of the mean of the maintenance hours. The maintenance hours decision variable range is 0 and 1314 hours where the maximum bound represents 15% of the entire cycle time of the simulation model. The remaining decision variables

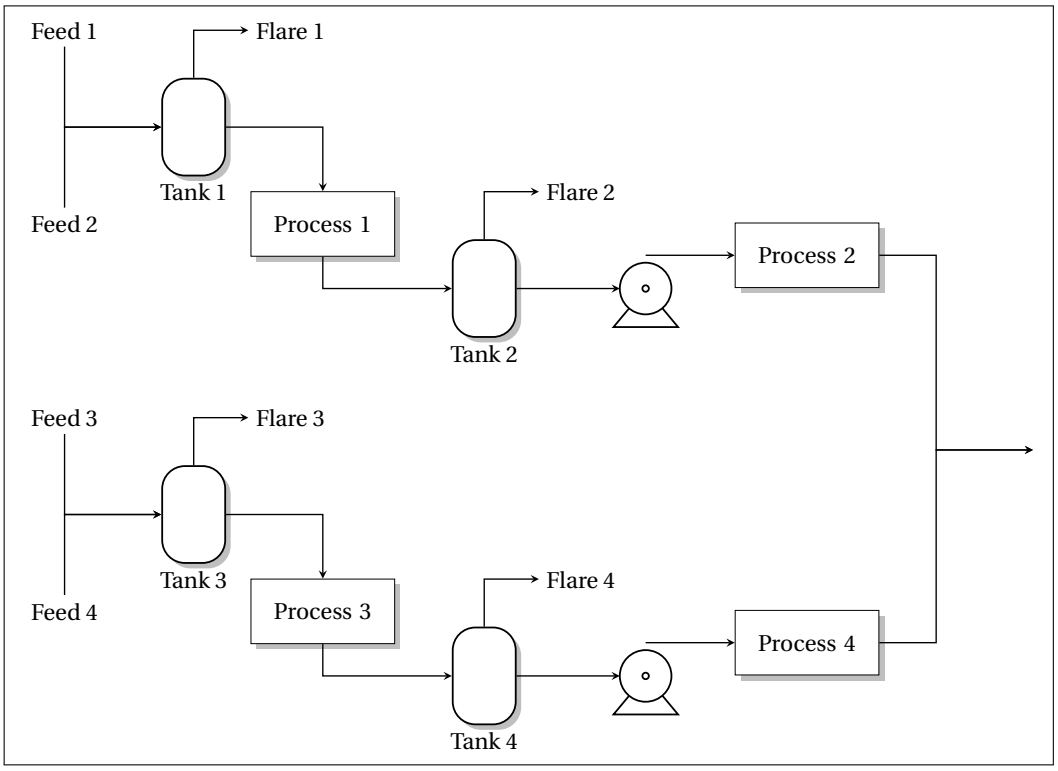

FIGURE 2.3: Chemical Plant System - Parallel

include the minimum number of spares to keep on-site,tank capacities, the minimum number of pumps to procure and the pump capacity.

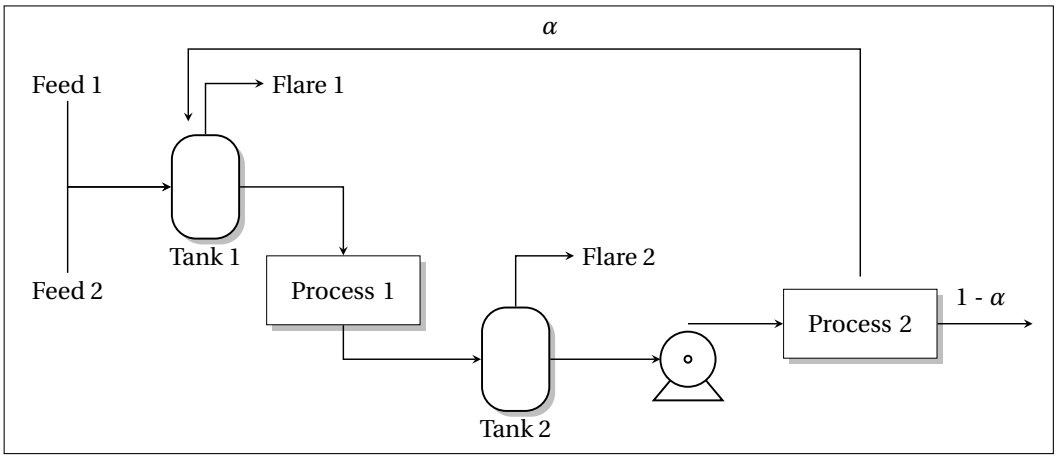

FIGURE 2.4: Chemical Plant System - Feedback

The key performance indicator for CPS-B, CPS-P, and CPS-F is to maximise the overall revenue of the system. The final product produced is the main revenue generator. The cost of failures on the various system components (processing units and pumps) and flaring product negatively impact all three systems' revenue. The objective functions for all three systems focus on simultaneously maximising the final product stream while minimising all the other variables. Table 2.6 includes all the revenues and costs for the fitness

function values for all three systems. These are defined as:

$$\underset{x_1}{\text{argmax}} \quad c_1 x_1 - \sum_{i=2}^{n} c_i x_i, \tag{2.1}$$

where n is the number of costs to include for the respective problems.

TABLE 2.6: Fitness Function Values for all the Chemical Plant Systems

(i)	Equipment/Process (x_i)	Revenue/Costs (c_i)	Problem Set	Bounds
1	Final Product	$6042/m^3$	All	-
2	Flare 1	$2848/m^3$	All	-
3	Flare 2	$3907/m^3$	All	-
4	Per Tank	$9.94 \times 10^7 + 1.52 \times 10^6/m^3$	All	[500, 1000]
5	Per Pump	$4.44 \times 10^6 + 2.96 \times 10^5/m^3$	All	[60, 120]
6	Flare 3	$2848/m^3$	CPS-P	-
7	Flare 4	$3907/m^3$	CPS-P	-
8	Number of Man Hours	474036/hr	CPS-F	[0, 1314]

2.6 Pressure Swing Adsorption

2.6.1 Design Parameters

A post-combustion CO_2 capture PSA process has been used in this research. The PSA system was sourced from work completed by Yancy-Cabellero *et al.* (2020) [23]. The carbon capture process plays a significant role in reducing the CO_2 emissions from gas-fired and coal power plants [23]. The PSA system forms part of the first step of the three-step carbon capture process [23].

The PSA cycle is illustrated in Figure 2.5. The PSA cycle configuration that has been implemented in this research is called the modified Skarstrom cycle. The system consists of five steps within a cycle, namely: A. pressurisation, B. adsorption, C. heavy reflux, D. counter-current depressurisation, and E. light reflux [23]. The cycle starts at low pressure and is then pressurised (step A) by the flue gas up to the adsorption pressure. Once the adsorption pressure is reached, the adsorption step (step B) takes place by having the top end of the column open, allowing the feed gas to be fed through and the CO_2 to be concentrated at the opposite end of the column. The heavy reflux step involves substituting the flue gas flowing into the column by the heavy product collected during the light reflux step (step E) after a set amount of time. This substitution results in a higher concentration of CO_2 at the bottom of the column due to the heavy product's high CO_2 concentration. The bottom of the column is then closed, and the pressure in that section is dropped to the starting pressure during the counter-current depressurisation step (step D). This step also includes collecting emissions at the bottom of the column as the CO_2 product. The final step, the light reflux step (step E), involves feeding the light product produced in the adsorption phase into the top of the column once the initial pressure is achieved, and the cycle repeats [23].

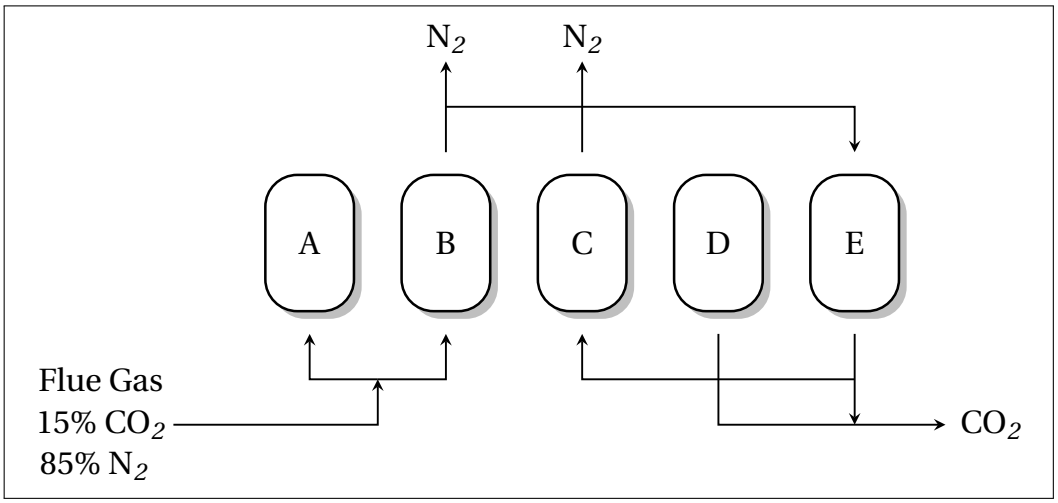

FIGURE 2.5: Pressure Swing Adsorption System - 5 step modified Skarstrom cycle (A - Pressurisation, B - Adsorption, C - Heavy Reflux, D - Counter-Current Depressurisation and E - Light Reflux) [23]

The PSA system's simulation model was developed in MATLAB from a set of partial differential equations describing the system. To optimise the PSA system, the two operating parameters that are maximised at each cycle are CO_2:

$$Purity = \frac{\text{Moles of } CO_2 \text{ in product}}{\text{Total moles in the Product}} \times 100\% \tag{2.2}$$

and:

$$Recovery = \frac{\text{Moles of } CO_2 \text{ in to product}}{\text{Moles of } CO_2 \text{ fed into cycle}} \times 100\%. \tag{2.3}$$

2.7 Conclusion

This chapter has illustrated the vast amount of available literature on the topic of surrogate assisted optimisation. Combining genetic algorithms with surrogate assisted optimisation is a common approach for multi-objective optimisation domains. The use of surrogate assisted genetic algorithms for the optimisation of stochastic systems is a topic that still requires further investigation. The existing literature has illustrated the significant benefits in implementing surrogate assisted optimisation. This research will further investigate this by implementing surrogate assisted genetic algorithms to optimise both stochastic and multi-objective systems.

Chapter 3

Research Methodology

3.1 Research Design

The research design that has been followed is an Empirical Experimental approach [43]. The Empirical Experimental research design methodology focuses on testing and confirming certain hypotheses [43]. Experimental research design uses a methodology where certain variables are kept constant and others are actively changed and manipulated to determine the change in outcome [44].

3.2 Methodology

The research methodology architecture is illustrated in Figure 3.1. The various research instruments, data sets, and analyses have stemmed from this framework and are detailed in the sections that follow.

The first research instrument that is detailed in Figure 3.1 is the simulation models. These models represent the primary evaluation platforms for the optimisation framework. In this research, four main systems were modelled to test the surrogate assisted evolutionary algorithms. The first system, Chemical Plant System - Basic (CPS-B) is outlined in Section 2.3. The details of how the simulation model was developed and all the parameters are included in Section 2.3.2. Chemical Plant System - Parallel (CPS-P) is detailed in section 2.4. This system is a more complex variant of CPS-B by introducing a parallel system. The Chemical Plant System - Feedback (CPS-F) system is detailed in Section 2.5 and is also a more complex variant of CPS-B. The added complexity is introduced with a feedback loop which represents the off-spec product that needs to be reworked. The final system is the Pressure Swing Adsorption (PSA) chemical process detailed in section 2.6. This simulation model has been sourced from the paper by Yancy-Cabellero *et al.* (2020)[23] and is intended to test the generalisability of the algorithm for optimising complex real-world chemical systems.

Each of the systems described above has its own simulation model including its own set of decision variables. The structure of the data generated from the simulation models includes the decision variables and their respective the performance indicators for the various models. This structure enables the algorithm to optimise each system effectively. The CPS systems described above only have a single performance indicator which is the

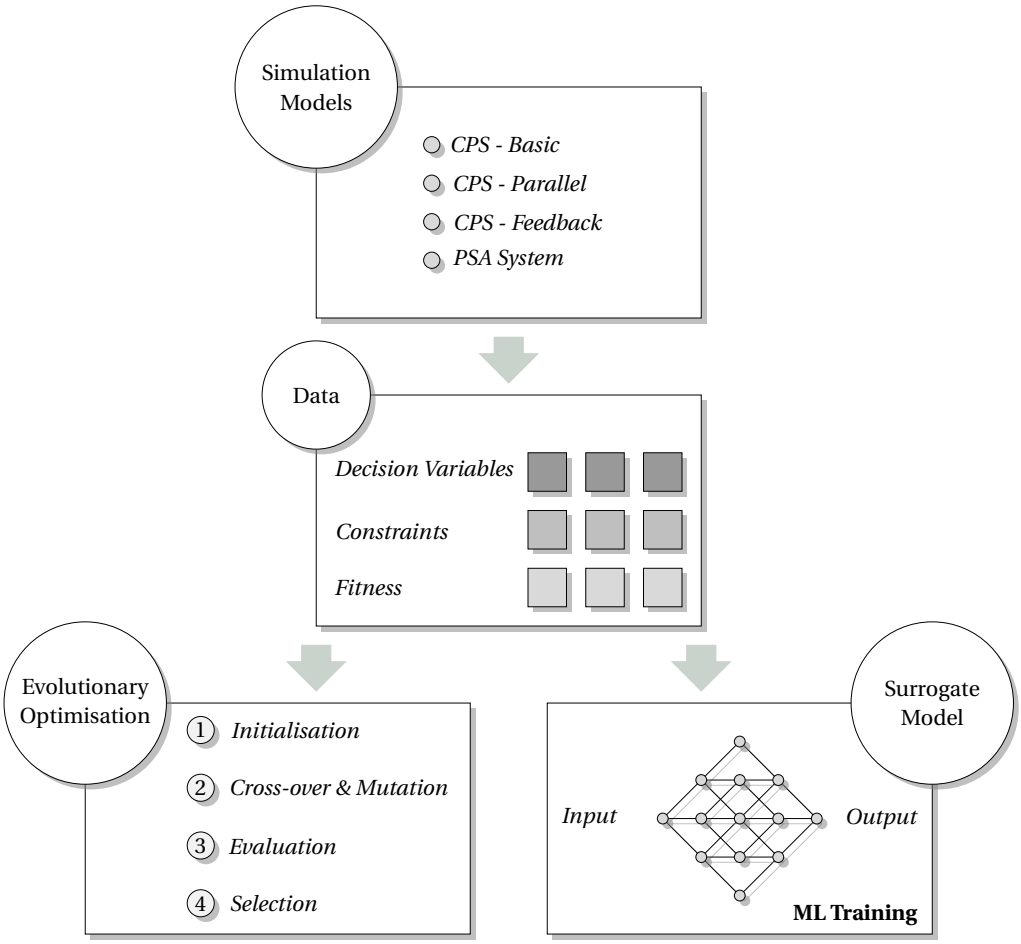

FIGURE 3.1: Research Methodology Architecture

overall system revenue. The PSA system has two objectives, namely the CO_2 *purity* and *recovery* ratios.

The simulation model data also plays a pivotal role in the performance of the surrogate model as it is used to train the model. The algorithm that is used for the surrogate models is the Random Forest Machine Learning algorithm. The surrogate Machine Learning algorithm was chosen by evaluating various algorithms, namely the K-Nearest Neighbours, Lasso, Ridge, Random Forest, Extra Tree Regressor,Support Vector Machine, and Gradient Boosting Regressor for the data generated by CPS-B detailed in section 2.3. Ten-fold cross-validation was used to compare these algorithms. The data set used for validation was created using the simulation model of the CPS-B. The best performance was achieved by the Random Forest algorithm yielding an out-of-sample accuracy between 0.77 and 0.93. The high accuracy achieved was sufficient for its application in the optimisation framework. The out-of-the-box (default) parameters were used and no further hyper-parameter tuning was done. Table 3.1 includes the Random Forest hyper-parameter set. For each system, an additional simulated data set was kept aside as a test during the optimisation of the system. At the start and every retraining instance of the algorithm, the accuracy of the surrogate model was monitored using the test set. When the divergence criteria was met during optimisation, the additional data generated by the simulation model would be

used to retrain the Random Forest model.

TABLE 3.1: Random Forest Surrogate Model Hyper Parameter Set

Number	Hyper Parameter	Value
1	Number of Trees	100
2	Criterion	Mean Squared Error
3	Splitting Minimum Samples	2
4	Leaf Node Minimum Samples	1
5	Splitting Maximum Features	5
6	Bootstrapping	True

The final component in Figure 3.1 is the evolutionary optimisation component. This incorporates the Genetic Algorithm with the function evaluation platform which comprises a combination of the simulation and surrogate model to manage the divergence of the Machine Learning model predictions.

The optimisation of the PSA system required a slightly different optimisation algorithm to the one implemented for the CPS systems due to the difference in performance metrics. For this reason, the final component is included in two separate algorithms. Each of these algorithms is detailed in their sections, namely Sections 3.2.1 and 3.2.2.

3.2.1 Surrogate Assisted Genetic Algorithm

The algorithm that was used to optimise the CPS systems is the Surrogate Assisted Genetic Algorithm (SA-GA) and is further detailed in the following sections.

Algorithm

The SA-GA algorithm represented in Figure 3.2 starts with an initial random population of size 800. The best 75 candidates of the initial population are selected as a warm start population. This warm start population is then used to intialise the surrogate machine learning model (Random Forest).

The SA-GA continues from here following the flowsheet steps in Figure 3.2. The top 15% of the initial ranked (highest to lowest revenue) population (elite) is selected. Using the BLX-α crossover technique, offspring are generated from the elite population set [45]. Mutation is then applied at a rate of 30% using the random substitution technique.

To manage deviations between their predictions: at each generation, the elite population is evaluated by both the surrogate and the simulation model. The deviation is assessed by calculating the difference between the surrogate predictions and the simulation model's output. If the difference is greater than one sigma, the simulation model is used as the evaluation platform, and the surrogate model is retrained with the new data generated by the simulation.

The terminating criteria for the algorithm is fifty generations.

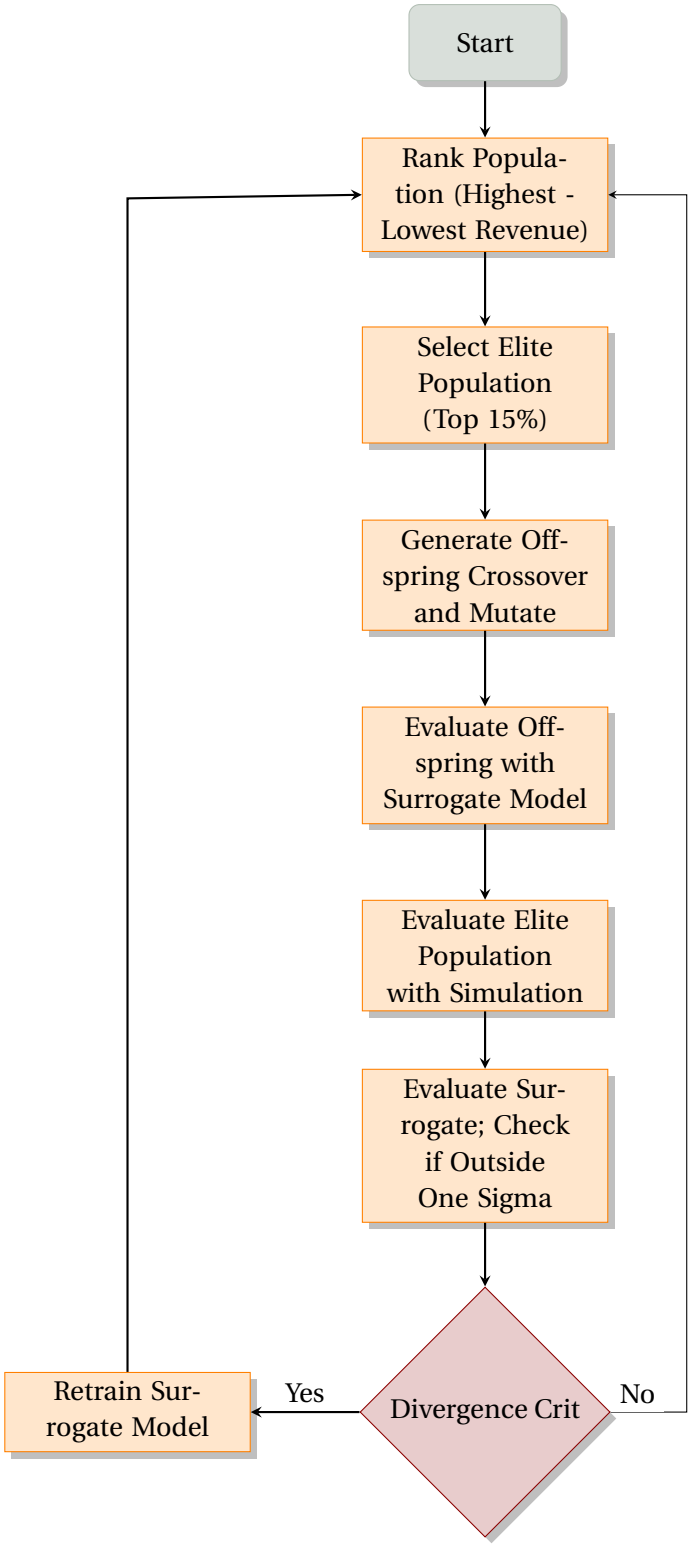

FIGURE 3.2: Flow Chart for Surrogate Assisted Genetic Algorithm

Crossover and Mutation

The BLX-α crossover technique involves combining a pair of chromosomes [45]. The BLX-α algorithm is implemented using two steps for each gene in a chromosome:

1. The range is calculated range = max − min by finding the minimum (min) and the maximum (max) values of the parent genes,
2. The child gene will be a random number in the interval [min − (range × α), max + (range × α)],

where α controls how much outside the [min,max] interval you would like to consider. A value of $\alpha = 0$ gives Uniform Crossover. A value of $\alpha = 0.15$ is employed for this research previously shown to give good results [1]. At every iteration of the SA-GA algorithm, each new child's fitness value is compared to the fitness values of the elite population. If it is discovered to have a better fitness value, the chromosome is substituted into the elite population, and the cycle continues [46]. Figure 3.3 illustrates an example of the crossover procedure for CPS-F.

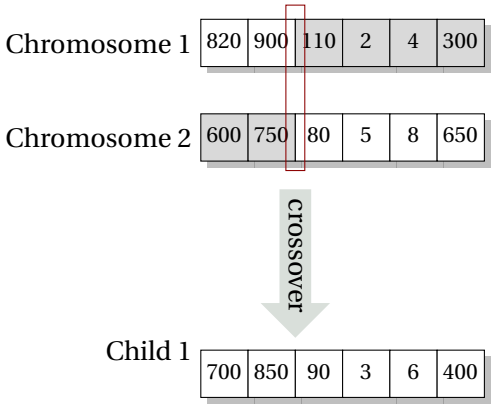

FIGURE 3.3: An Example of the crossover operation for two of the potential candidates used by the Genetic Algorithm

A 100 % crossover rate was used for the optimisation which means that the entire offspring was created from the elite population. The final step was the mutation of the offspring. The mutation was executed using the random substitution technique which involves substituting the values in the gene with a random number within its respective maximum and minimum bounds. Figure 3.4 is an example of the mutation procedure.

Genetic Structure

The decision variables for the CPS systems formed part of the chromosomes that made up the genetic algorithm population. The chromosomes' genetic structures used in the CPS systems are summarised in Table 3.2.

3.2.2 Surrogate Assisted NSGA-2

The PSA system was optimised using the Surrogate Assisted NSGA-II (SA-NSGA) algorithm. This algorithm is detailed in the sections to follow.

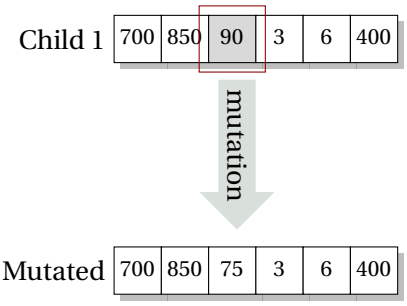

FIGURE 3.4: An Example of the mutation operation for the potential candidate used by the Genetic Algorithm

TABLE 3.2: Genetic Structure for the Chemical Plant Systems

#	Description	CPS Set	Bounds
1	Tank 1 Size	All	[500,1000]
2	Tank 2 Size	All	[500,1000]
3	Tank 3 Size	CPS-P	[500,1000]
4	Tank 4 Size	CPS-P	[500,1000]
5	Pump 1 Size	All	[60,120]
6	Pump 2 Size	CPS-P	[60,120]
7	Minimum Spares Level	All	[0,20]
8	Minimum Spares to Purchase	All	[1,20]
9	Number of Maintenance Man Hours	CPS-F	[0, 1314]

Algorithm

The SA-NSGA algorithm is represented in Figure 3.5. The initial population is selected as the best 60 candidates from the warm start random population of size 800. The surrogate machine learning models (Random Forest) for the *purity* and *recovery* objectives are both initially trained and parameterised using the random warm start population. To handle the PSA system's multi-objective nature, for the surrogate modelling component of the optimisation, two surrogate models for each of the *purity* and *recovery* fitness values were used [31].

The initial population is sorted and ranked according to the non-dominated sorting technique using Pareto fronts and the crowding distance metrics (estimate of the surrounding solution density for each gene [47]). The top sixty genes are selected for Tournament Selection. Intermediate crossover and Gaussian mutation are applied with the rate of $\frac{2}{6}$. The offspring are either evaluated using the simulation or the surrogate model based on the divergence criteria. The terminating criterion for the GA is sixty generations.

The deviation between the surrogate models and the simulation model is managed by evaluating the top 15% of the ranked population using both the simulation and surrogate models and retraining the surrogate model if the difference is greater than one sigma. The function evaluation is done in MATLAB and the results are sent back to python using the Python `matlab.engine` library to integrate the two platforms. It is important to note that the integration of these platforms does introduce an additional computational cost.

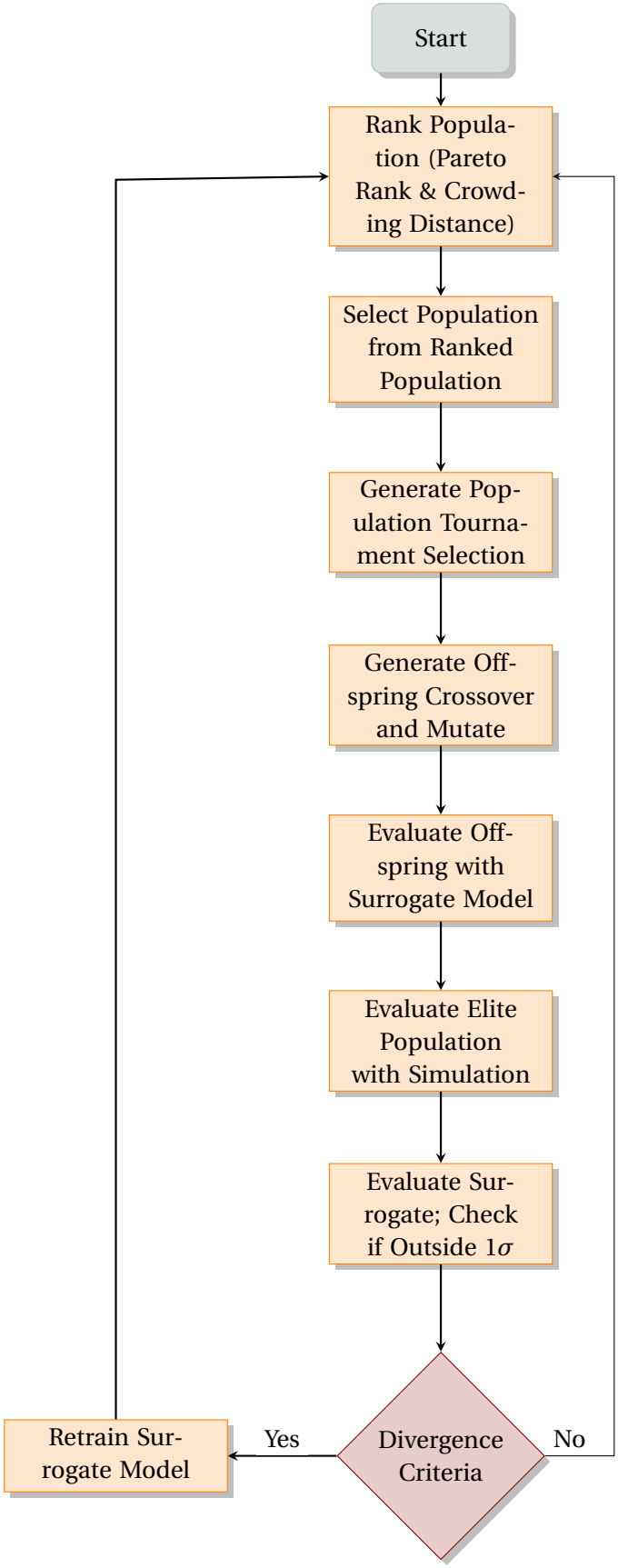

FIGURE 3.5: Flow Chart for Surrogate Assisted NSGA-II

Selection

The SA-GA [1], [2] has been adjusted to substitute the existing sorting technique with the NSGA-II non-dominated sorting methodology. The sorting algorithm is detailed in Algorithm 1[33].

Algorithm 1 Fast non-dominated sort (P) [33]

```
for each p ∈ P do
    S_p = ∅
    n_p = 0
    for each q ∈ P do
        if p < q then
            S_p = S_p ∪ q
        else if q < p then
            n_p = n_p + 1
        end if
    end for
    if n_p = 0 then
        p_rank = 1
        F_1 = F_1 ∪ q
    end if
end for
i = 1
while F_i ≠ ∅ do
    Q = ∅
    for each p ∈ F_i do
        for each q ∈ S_p do
            n_q = n_q − 1
            if n_q = 0 then
                q_rank = i + 1
                Q = Q ∪ q
            end if
        end for
        i = i + 1
        F_i = Q
    end for
end while
```

The sorting algorithm executes by iterating over all the solutions in the population. It uses two parameters, namely the domination count (n_p), which represents the number of solutions that dominate the current solution, and the solution set that the current solution dominates (S_p). The first nondominated front includes solutions with a domination count of zero. For each solution p with a zero domination count, the algorithm iterates through its solutions q from the set S_p and reduces its domination count by one. When the domination count of any solution q gets to zero, the solution is assigned to a separate list Q. These solutions represent the second nondominated front [33].

Once the non-dominated sorting is applied, the sorting algorithm's nondomination ranking output is used in combination with each solution's crowding distance metric to

sort the population. The next phase of the SA-NSGA is the tournament selection process which is executed to generate a population for the offspring generation process. The tournament selection process involves the random sampling of a set number of individuals, with replacement from the existing population with the intent of comparing their fitness and selecting the best individual for the new population [48]. The number of individuals sampled at every iteration is two, and the population size is sixty, which results in 120 tournaments taking place to achieve a population size of sixty.

Crossover and Mutation

The population generated by the tournament selection process is then used for crossover using the intermediate crossover technique. This technique creates offspring by taking a weighted average of the parents using a ratio value to specify the weights. The offspring is a function of two parents, parent1 and parent2. The first child is given by:

$$\text{Child 1} = \text{Parent 1} + \text{Rand} \times \text{Ratio} \times (\text{Parent 2 - Parent 1}) \tag{3.1}$$

and the second by:

$$\text{Child 2} = \text{Parent 2} - \text{Rand} \times \text{Ratio} \times (\text{Parent 2 - Parent 1}) \tag{3.2}$$

where the Ratio is a (uniform) randomly generated weight used to determine the change in the genetic structure of the child from the parent's. The probability of a child changing its genetic structure from its parent is based on the Rand vector, a set of binary values, the same length as the number of decision variables. The Rand vector is determined by initially having a generated vector set of random numbers and checking if they are less than the specified fraction (crossover rate) of $\frac{2}{6}$.

Figure 3.6 is an example of the Intermediate crossover technique applied for the PSA optimisation. The grey blocks represent the parent elements that will be changed, and the white blocks represent the elements that will remain from chromosome 1 in child 1.

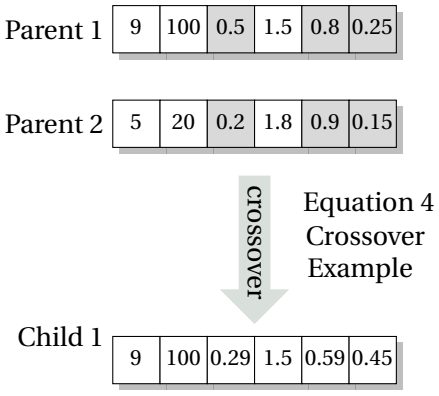

FIGURE 3.6: An example of the Intermediate crossover operation for two potential parent candidates and a resulting new chromosome

The Gaussian mutation technique adds a random number from the Gaussian distribution, with a mean of 0 and standard deviation set by two parameters: the shrink and

scale parameters, to each child's entry. The mutation is applied at a rate of $\frac{2}{6}$. Figure 3.7 is an example of the mutation procedure.

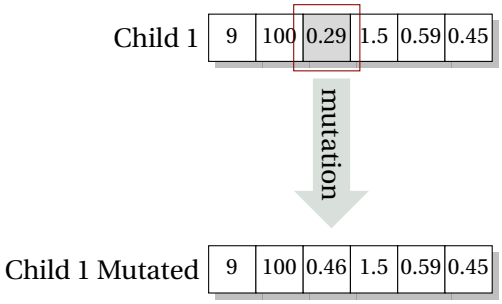

FIGURE 3.7: An example of the Gaussian mutation operation for the potential candidate used by the Genetic Algorithm

At the end of both of these procedures, the decision variables are checked against their bounds. If the decision variables lie above or below their respective bounds, they are substituted with either the maximum or the minimum values, respectively.

Genetic Structure

The genetic algorithm population was made up of chromosomes that included the decision variables for the PSA system. The genetic structure of the chromosomes for the PSA system optimisation is summarised in Table 3.3.

TABLE 3.3: Genetic Structure for Gene 1 to 6 for the Pressure Swing Adsorption System

#	Description	Units	Bounds
1	Adsorption Pressure	Bar	[1,10]
2	Time of Adsorption	Seconds	[10, 1000]
3	Light Product Reflux Ratio	-	[0.01, 0.99]
4	Feed Velocity	m/s	[0.1, 2]
5	Heavy Product Reflux Ratio	-	[0,1]
6	Purge Pressure	Bar	[0.1, 0.5]

Performance Metrics

The metrics used to analyse the performance of the different experiments implemented for the PSA optimisation are the generational distance (GD), hypervolume (HV) and inverted generational distance (IGD). The HV indicator provides the area (two-dimensional problem) in which the solution set dominates for a specific selected reference point [49].

The GD metric is represented by the following formula:

$$GD(X) = \frac{1}{|X|}\left(\sum_{i=1}^{|X|} d_i^p\right)^{\frac{1}{p}}, \tag{3.3}$$

where X represents the solution vector achieved by the algorithm. The Euclidean distance from x_i to the nearest reference point in the Pareto front or reference set of points is represented by d_i. The formula determines the average distance from any point on X to the closest point on the (ideal) Pareto front.

The IGD formula is represented as:

$$IGD(X) = \frac{1}{|Z|}\left(\sum_{i=1}^{|Z|} \hat{d}_i{}^{p}\right)^{\frac{1}{p}}, \tag{3.4}$$

where Z represents the Pareto front or reference set of points. $\hat{d}_i{}^{p}$ represents the euclidean distance from Z to the nearest reference point in X, which is the solution vector achieved by the algorithm. The inverted generation distance metric provides the average distance from Z to X's closest points, where the generational distance metric measures the average distance from X to the nearest point in Z.

3.2.3 Software Details

The simulation models and the SA-GA Algorithm for the optimisation of the CPS systems were all implemented in Python. Three different implementations were executed for the PSA optimisation case. The first implementation is the direct implementation which was executed in MATLAB using the source code provided by Yancy-Cabellero *et al.* (2020) [23]. The second is the reference implementation involving the replication of the algorithm used in the direct implementation in python. The final is the surrogate assisted implementation, which introduced surrogate models into the reference approach as substitutes for the simulation model. The surrogate assisted implementation is focused on the potential computational efficiency improvements that can be achieved by using surrogate models. The performance metrics detailed in Section 3.2.2 were implemented in Python using *Pymoo*. *Pymoo* is the multi-objective optimisation library used to calculate the HV, GD, and IGD. The hypervolume metric within the *Pymoo* library was implemented using the algorithm developed by Distributed Evolutionary Algorithms in Python (DEAP). All the implementations for the optimisation of both the primary and motivating examples are available on Github [50].

3.3 Limitations

The limitations of the research methodology include the number of use cases for the testing of the algorithm. Due to time and resource constraints, only four chemical plant systems were used to test the optimisation algorithm. These cases have proven the generalisability and robustness of the algorithms but the more cases that are tested, the better the validity of the results. This limitation provides an opportunity for further work to be completed to test more complex and constrained chemical plant design and operation Optimisation cases. The second limitation of the methodology includes the number of experiments that were run. This is also a limitation due to time and resource constraints.

The selected number of experiments have yielded representative results by firstly accounting for the stochasticity of the CPS simulation models and secondly accounting for the variability in the results.

3.4 Conclusion

The research methodology has been structured such that it can enable the investigation of the potential computational resource cost reduction in making use of surrogate assisted evolutionary algorithms for four chemical plant systems. The robustness and generalisability of the algorithms have also been validated by optimising four different chemical plant systems. The primary components of the methodology include the simulation models as were used to test and validate the results of the algorithms. The limitations of the research methodology have highlighted an opportunity for further investigation and testing to be done.

Chapter 4

Results

This chapter includes the results for the optimisation of all four chemical plant systems detailed in Chapter 2.

For the CPS systems, the results were acquired through an average of 30 experiments, for both the simulation-only and surrogate assisted modes. The primary performance metric for the chemical plant systems (CPS) is the final revenue. For each CPS, the revenue metric has been summarised using different techniques to highlight the comparative performance of the simulation-only and surrogate assisted modes. These techniques are detailed in the Sections below.

A set of twenty experiments were run for the direct and surrogate assisted implementations for optimisation of the Pressure Swing Adsorption (PSA) system. The reference implementation was executed for five runs due to this implementation focusing on replicating the direct implementation and having significantly higher computational cost. Three primary quantitative metrics were used to illustrate the performance of the various implementations. These metrics included the generational distance (GD), the hypervolume (HV), and the inverted generational distance (IGD). Each of these metrics was calculated per generation for all experiments. The HV metric is an area type calculation that is independent of any ideal solution. The GD and IGD metrics are distance measures to some reference vector. For the twenty experiments, the average of the final generation's population generated from the direct implementation was used as the reference vector. The qualitative metric that was used across all the experiment sets was the Pareto frontier for the last generation. This metric was intended to provide visual comparisons of the performances across the different implementations.

The remainder of this chapter has been divided into four subsections each representing the optimisation results per system.

4.1 Chemical Plant System - Basic

4.1.1 Simulation Evaluation Results

The results where the simulation model was used as the only evaluation platform is summarised in figures 4.1, 4.3, and 4.2. The data summarised in Figure 4.1 represents the average of all 30 experiments of the average revenue value per population per generation

smoothed using a rolling mean with window size two. The results illustrate that the genetic algorithm (GA) constantly improves the fitness of the CPS-B up to generation 20. After generation 20 the GA converges.

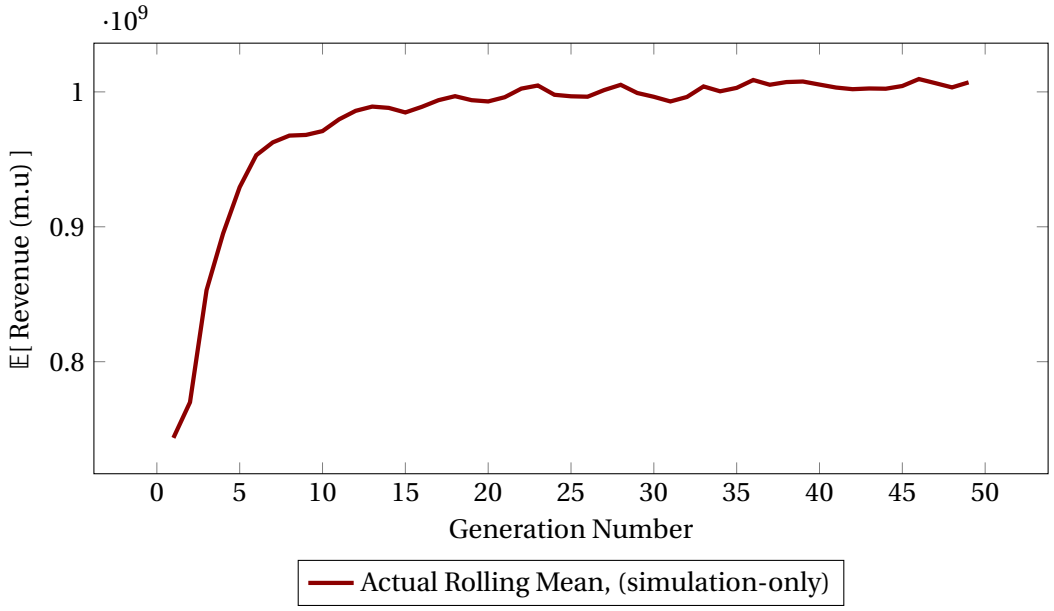

FIGURE 4.1: Rolling (window=2) Mean Revenue per generation

It is illustrated in Figure 4.2 that the GA can consistently improve the elite population. This figure is the only non-smoothed figure and illustrates that GA can improve the solution of the elite population in a relatively stable manner.

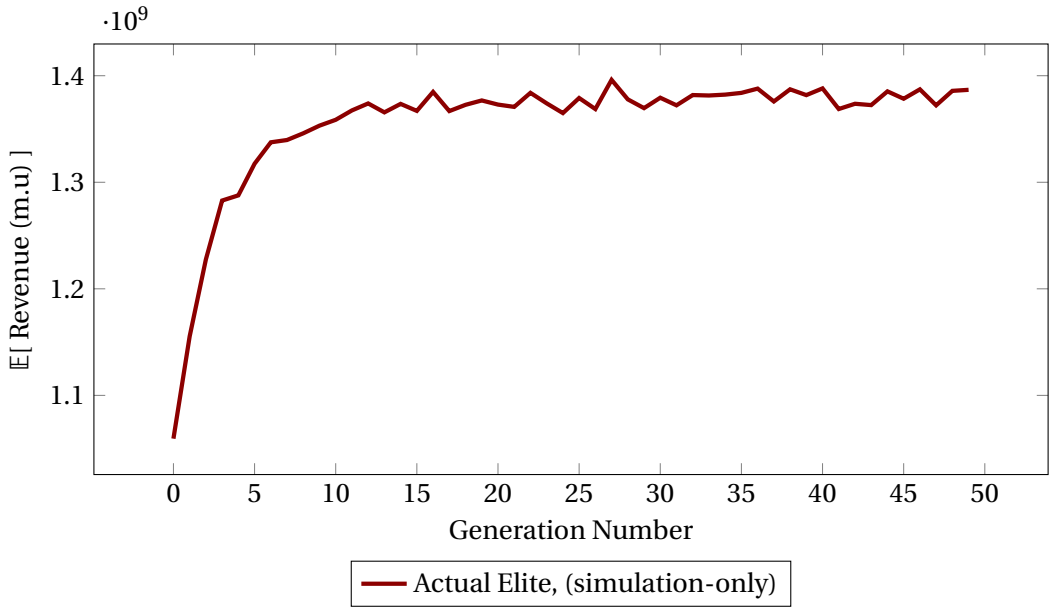

FIGURE 4.2: Mean Revenue of Elite per generation

Figure 4.3 illustrates a less smoothed increasing trend in the mean of the maximum revenue values per population per generation. Both the average and the maximum revenue for the elite population is increasing. This means that new elite parents are being

identified as the GA progresses. The results also indicate that the GA is successful in its search for optimal solutions.

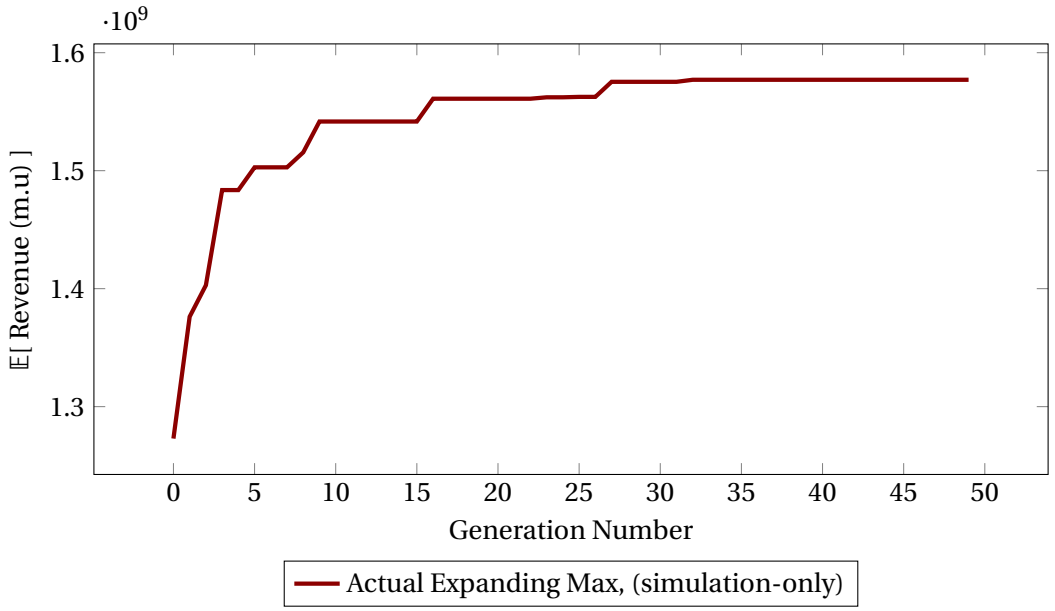

FIGURE 4.3: Expanding Max Revenue per generation

The optimal target value for the stochastic simulation is changing from generation to generation and the GA is expected to acquire an optimal value. The GA's ability to discover an optimal parameterisation in this stochastic environment proves the robustness of the algorithm.

4.1.2 Surrogate Assisted Results

The positive simulation only results leads to the next consideration of the potential speedup of the overall elapsed run time that may be be attained when implementing the surrogate model. Figures 4.4, 4.5, and 4.6 include results for the 30 consolidated experiments in which a surrogate model was introduced. Despite the stochastic nature of the system, the Random Forest machine learning model achieved an accuracy of 77%. In addition to this, 1.82 times speedup of the elapsed run time was achieved using the surrogate model (average across five runs of the experiment).

In Figure 4.6 it is evident that the surrogate assisted model is able to achieve a max value faster in its overall elapsed run time. The results are also illustrating that the model is able to achieve a better max revenue than the simulation-only model. The stability introduced by the surrogate model predictions may have had an impact on these better max values achieved because it allows the GA to optimise the target function better.

Figure 4.4 shows the average predicted values produced by the surrogate model per population per generation alongside the average actual values produced by the simulation. The upward trend in Figure 4.4 illustrates positive improvements for the predicted average for the surrogate assisted experiment set. The values in figure 4.4 are significantly lower than the ones in figure 4.1 but are exhibiting a general upward trend. The use of

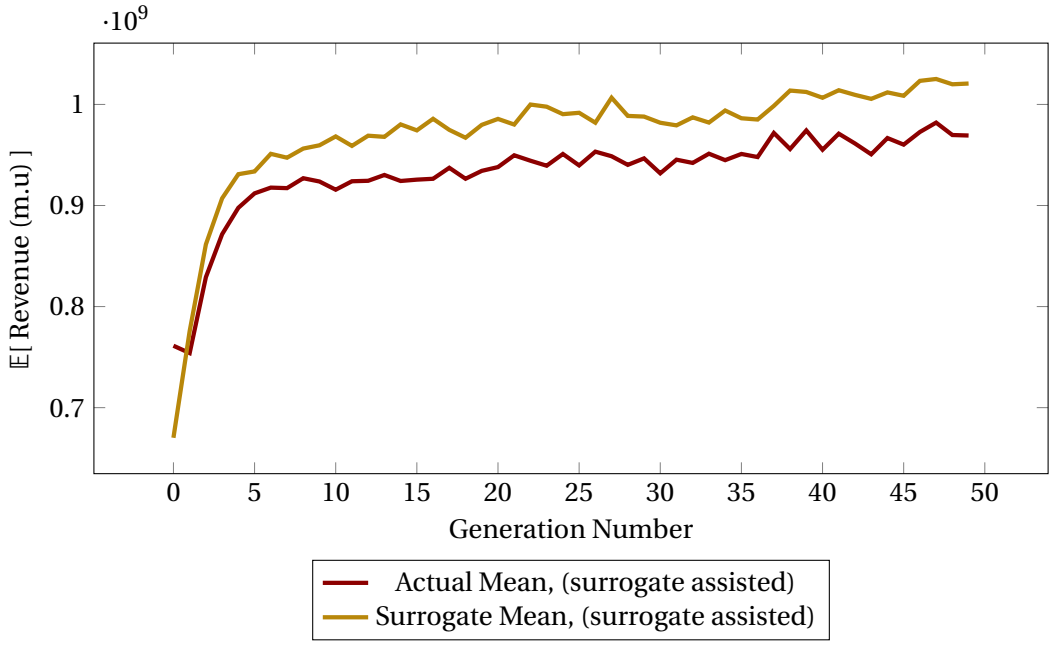

FIGURE 4.4: Comparison of the Surrogate Mean Revenue And Actual Mean Revenue of the surrogate assisted model

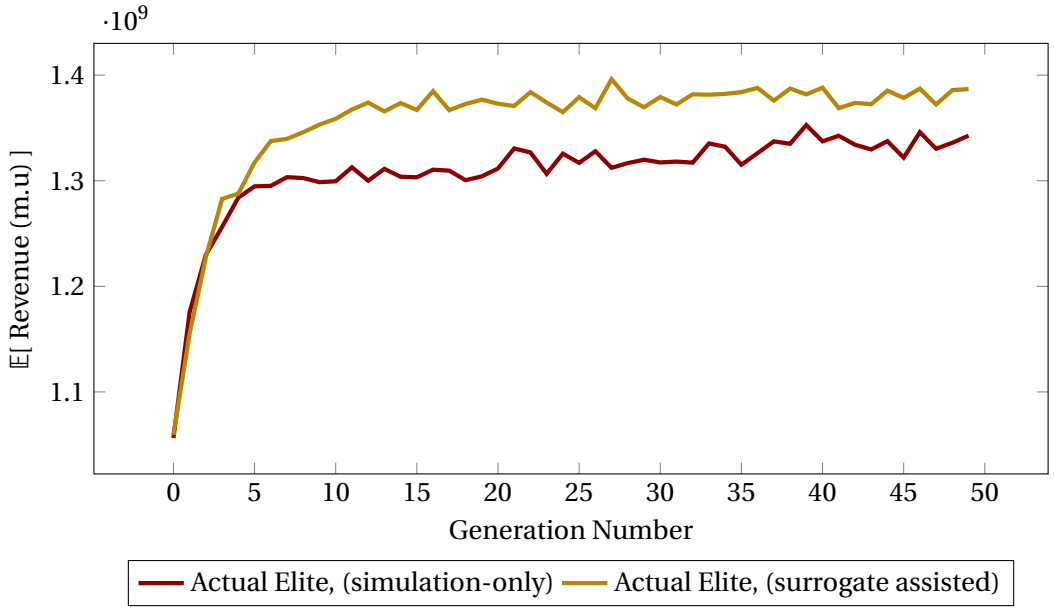

FIGURE 4.5: Comparison of the Actual Mean Elite revenue of the surrogate assisted and simulation-only models

a surrogate model has achieved the speedup of the optimisation process as well as improved optimal solutions.

The summary of the results as detailed above has shown that the use of a GA can facilitate the optimisation of chemical plant design despite the stochastic nature of these systems. The accuracy of 77% in conjunction with a 1.8 times speedup (average across 5 experiments) of the overall elapsed run time when implementing the surrogate model is a significant improvement.

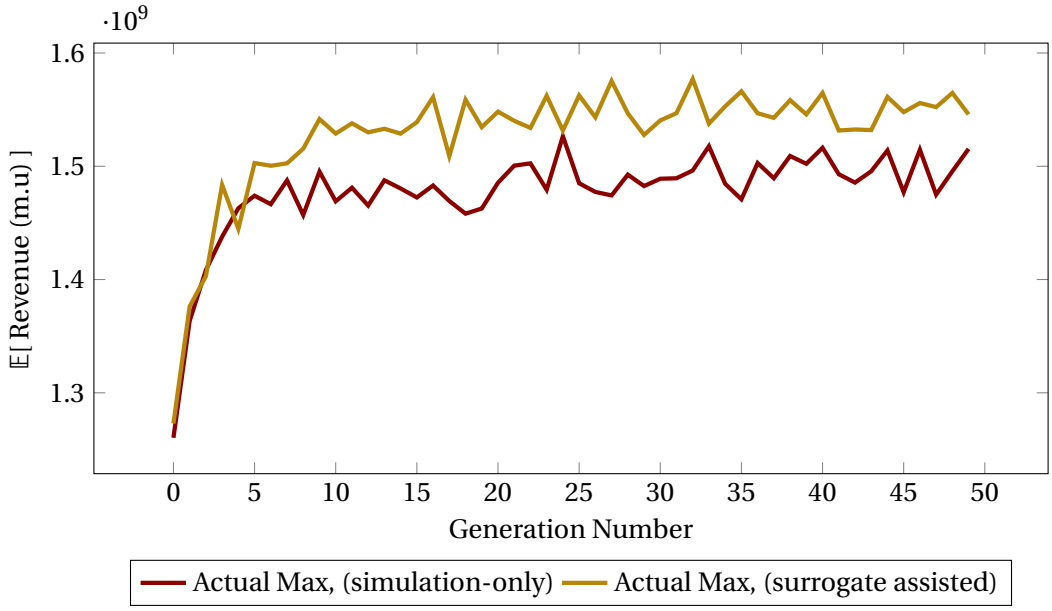

FIGURE 4.6: Comparison of the Actual Max Revenue of the surrogate assisted and simulation-only model

4.2 Chemical Plant System - Parallel

4.2.1 Simulation Evaluation Results

The results from the simulation-only mode for CPS-P showed that the GA was able to optimise across a more complex stochastic system with increased decision variables.

4.2.2 Surrogate Assisted Results

Figure 4.7 illustrates the SA-GA algorithm results for the optimisation of the CPS-P system. The actual lines (solid dark red, solid gold, and dotted gold) represent the results where the simulation model was used to evaluate the population. The surrogate line (dotted gold) represents where the surrogate model was used to evaluate the population. The solid dark red and dotted red lines represent the mean actual revenues for the elite population in the simulation-only and surrogate assisted optimisation modes, respectively. The solid gold and dotted gold lines represent the mean actual and surrogate revenues for the entire population in the surrogate assisted optimisation modes. The performance of the SA-GA algorithm is evident from Figure 4.7. Despite the added complexity and stochastic nature of the system, the GA was still able to optimise across the target. The GA is also improving the elite population across generations for both the simulation-only and surrogate assisted implementations. The surrogate assisted algorithm can predict the revenue of the CPS-P with an accuracy of 90%. The algorithm also achieved a significant speedup of 1.95 times over the simulation-only mode.

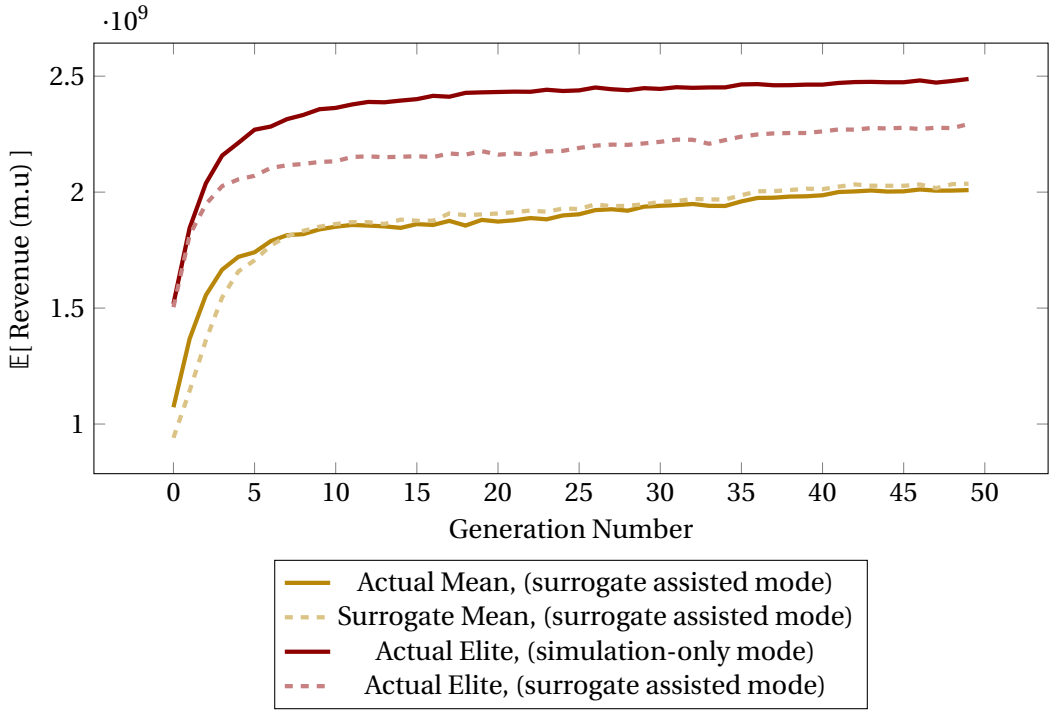

FIGURE 4.7: Actual mean revenue, surrogate mean revenue, actual elite mean revenue for the surrogate assisted mode and the actual elite mean revenue for the simulation-only mode

4.3 Chemical Plant System - Feedback

4.3.1 Simulation Evaluation Results

The optimisation of CPS-F for the simulation-only implementation yielded positive results. The GA was able to achieve optimal solutions despite the added complexity and stochastic nature of the system.

4.3.2 Surrogate Assisted Results

Figure 4.8 includes the results for three different experiments completed on the deterministic version of CPS-F. The dark and light red lines represent the revenues when the maintenance hours are fixed at the maximum and minimum bound of 1314 and 0 hours. The dotted red lines represent the revenue of the experiments where the GA could select and optimise the maintenance hours. These three different experiments have also been completed for the stochastic version of CPS-F and are illustrated in Figure 4.9. These results indicate that the surrogate assisted implementation improved the revenue of CPS-F despite the additional complexity added. This implementation also yielded a overall elapsed run time speedup of 1.74 times the simulation-only mode. The surrogate model was able to predict the revenue from the CPS-F system at an overall accuracy of 69%. The GA was able to handle the additional complexity added by the system's feedback component, indicating its robustness towards this type of component.

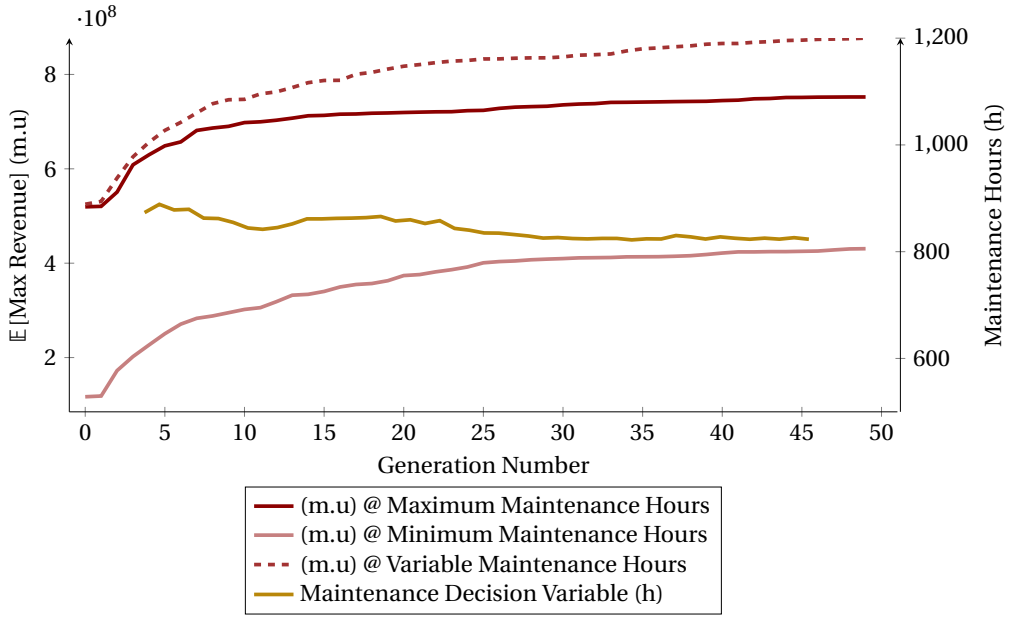

FIGURE 4.8: Mean of the max revenue of the deterministic system with constant and variable maintenance hours

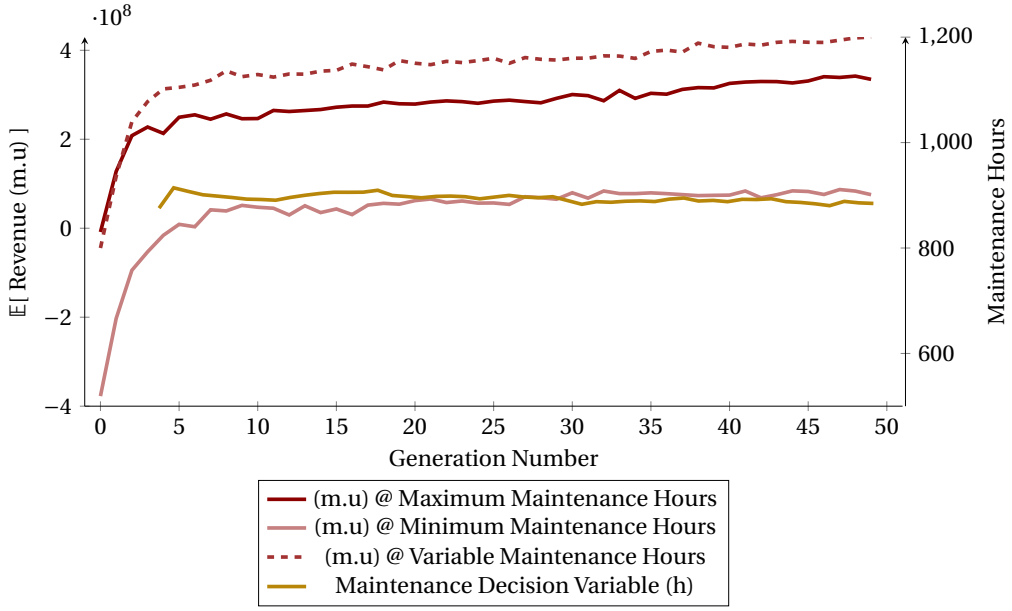

FIGURE 4.9: Mean Revenue of the stochastic system with constant and variable Maintenance Hours

4.4 Summary of Chemical Plant System Results

The results for all three CPS systems have illustrated the benefits of implementing a surrogate machine learning model in combination with a genetic algorithm framework. The overall elapsed time speedups achieved were 1.8, 1.74 and 1.95 times the direct (simulation only) optimisation for CPS-B, CPS-P and CPS-F respectively. The accuracy of each system's surrogate model was 77%, 90% and 69% for CPS-B, CPS-P and CPS-F respectively. These results have been achieved despite the complexity and stochasticity of the systems.

4.5 Pressure Swing Adsorption

The final generation frontiers for all three implementations are represented in Figure 4.10.

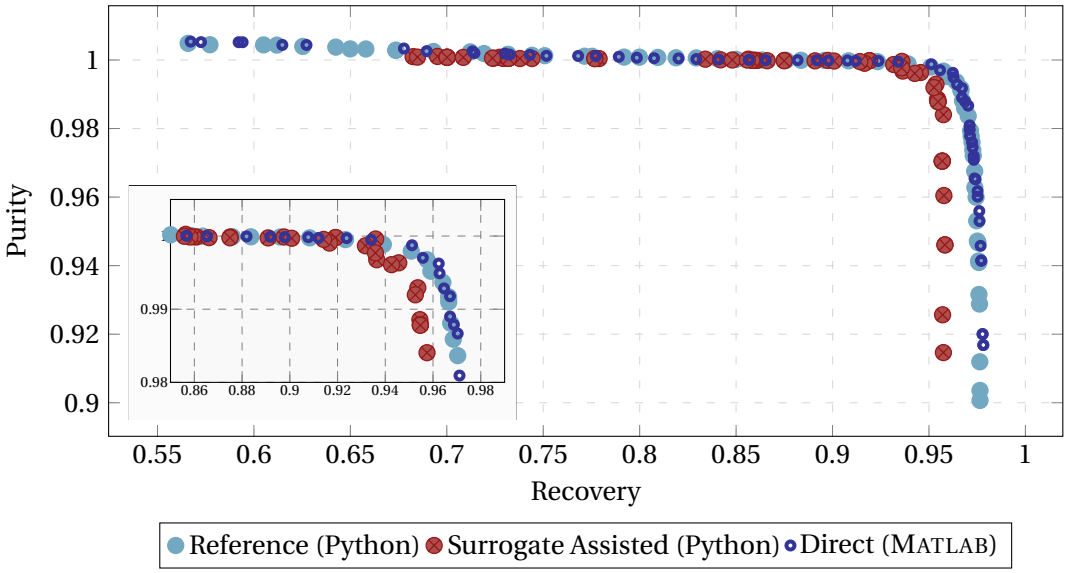

FIGURE 4.10: Frontier Comparison between the direct, surrogate assisted and reference implementations

The reference implementation achieved a similar shape and range to the direct implementation from visual inspection of these frontiers. The frontier of the surrogate assisted implementation also follows the same shape as the direct implementation. To further validate the results illustrated by the Pareto frontiers, quantitative measures have been used.

The average performance metrics for the three different PSA optimisation implementations are summarised in Table 4.1. These metrics represent the average performances achieved across all experiments for the final generation (60). The standard deviations have also been included to illustrate the variation across the experiments. It is important to note that for all the metrics in Table 4.1, the smaller the value, the better the performance.

The direct and reference implementations achieved the same HV values indicating that the reference implementation successfully achieved its intent of replicating the direct implementation. The surrogate assisted implementation has a larger HV than the direct and reference implementations, illustrating that its performance is inferior but can still be highly competitive. The GD and IGD metrics illustrate that the reference implementation results are closer to the direct than the surrogate assisted implementation.

To further investigate each implementation's convergence, the average of the HV per generation is illustrated in Figure 4.11 for all implementations. We note no material difference between progressions other than additional noise in the surrogate's solution space in early generations. *Python direct optimisation provided for reference.*

Each implementation has a specific convergence rate, and to illustrate this in a comparative nature, the average number of evaluations (generations) to success (AES) for 99.5%, 99%, 98.5%, and 98% HV success rates has been summarised in Table 4.2. The

TABLE 4.1: Average Performance Metrics for Final Generation

Metric	**Implementation**	**Value**
Hypervolume	Direct	0.007±0.001
	Reference	**0.007±0.000**
	Surrogate Assisted	0.009±0.003
Generational Distance	Direct	-
	Reference	**0.012±0.000**
	Surrogate Assisted	0.014±0.003
Inverted Generational Distance	Direct	-
	Reference	0.065±0.001
	Surrogate Assisted	**0.034±0.021**

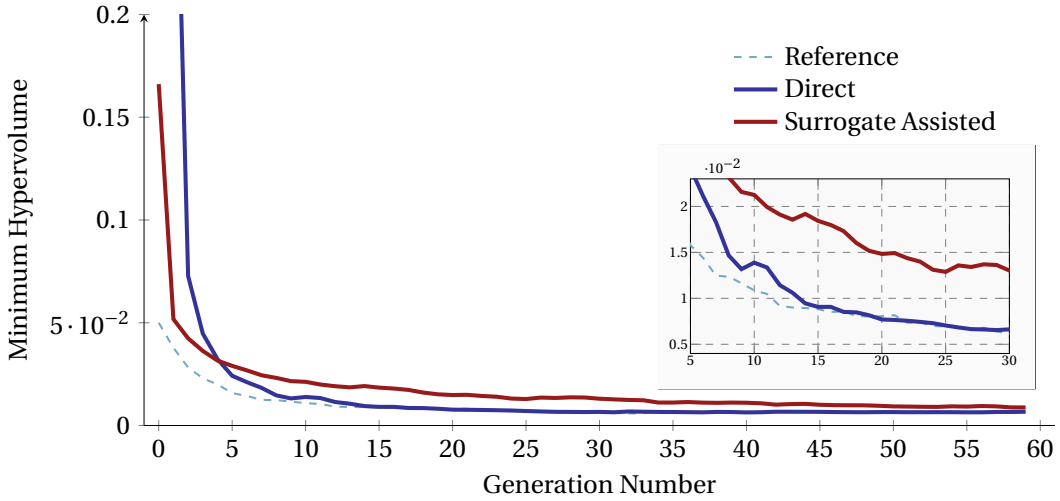

FIGURE 4.11: Mean hypervolume progression at each generation across repeated experiments

surrogate assisted implementation yields the slowest convergence rate across all the implementations. It is, however, able to achieve the 99.5% HV success rate within the maximum number of generations (60). The Surrogate Assisted model requires significantly more generations to achieve the same success rate as the other approaches. However, these additional generations are evaluated using the surrogate model and as such still result in a significant speedup.

TABLE 4.2: Average Evaluations to Success (AES) for various Success Rates (Number of Generations)

	Success Rate			
Implementation	**98%**	**98.5%**	**99%**	**99.5%**
Direct	*6*	*7*	*10*	*16*
Reference	**3**	**5**	**6**	**12**
Surrogate Assisted	*7*	*12*	*22*	*49*

From the results above, we have evidence demonstrating that the reference implementation achieved similar results to the direct implementation. The surrogate assisted technique achieved slightly inferior results to the direct optimisation cases but has significantly reduced the computation cost of the optimisation. The surrogate assisted implementation achieved a five times speedup of the overall elapsed run time compared to the reference implementation with an accuracy of 91.8% for the *purity* and 99.1% for the *recovery*.

Chapter 5

Conclusion

To conclude, the aims of the research were focused on two main components. Firstly the investigation into the potential resource cost reduction in making use of a surrogate assisted evolutionary algorithms for the optimisation of computationally expensive simulation models. Secondly, the research aimed to determine the robustness of these algorithms towards stochastic and multi-objective systems.

The optimisation of the three CPS systems achieved results demonstrating the SA-GA algorithm's robustness towards more complex stochastic systems. The SA-GA algorithm was able to achieve significant computational efficiency gains of 1.8 times speedup for the CPS-B system, 1.95 times speedup for CPS-P and 1.74 times speedup for CPS-F. The accuracies achieved for the surrogate models were 77% for the CPS-B system, 90% for the CPS-P system and 69% for the CPS-F system. The PSA system optimisation yielded results supporting that the SA-NSGA algorithm is robust towards more complex multi-objective systems, including real-world multi-objective chemical plant systems. The results illustrated the achievement of replicating the direct implementation and further achieving a five times speedup with the surrogate assisted technique with an accuracy of 91.8% and 99.1% for the purity and recovery models, respectively. The research aims have been met resulting in positive outcomes of significant speedups, accuracies and algorithm robustness towards stochastic and mutli-objective systems.

The complete result set for the surrogate assisted optimisation of all four chemical plant systems illustrated the computational improvements obtained using a GA combined with a Machine Learning Surrogate model as a substitute for computationally expensive simulation models. It is worth highlighting the SA-GA and SA-NSGA algorithm's flexibility and robustness in adapting to more complex multi-objective systems. The surrogate assisted optimisation approach used in this research has proven successful across complex chemical systems and should be further verified and tested within Industry. This approach can enable faster decision making for the optimisation of chemical plant design and operations.

The research questions were aimed at determining three main concepts. Firstly to investigate the impacts of stochastic systems on the optimisation algorithm. The results from all three (stochastic) CPS systems illustrated that despite the complexity of the stochasticity of the systems, the SA-GA algorithm was able to converge to optimal solutions faster.

This was done while maintaining accuracies of between 69% - 90% for the machine learning surrogate models. The second question was aimed at understanding the impacts of multi-objective systems on the optimisation algorithm. The optimisation of the PSA system also achieved a significant speedup in convergence to optimal solutions while maintaining accuracies of 91.8% and 99.1% for the *purity* and *recovery* surrogate models. The final research question was aimed at determining the speed improvement for the optimisation of the chemical systems when implementing surrogate assisted optimisation. This question was answered through the 1.8, 1.95, 1.74 and five times overall elapsed run time speedups achieved for the surrogate assisted optimisation of the CPS-B, CPS-P, CPS-F and PSA respectively.

The current state of this research poses an opportunity for a wide variety of future studies. Topics include the further investigation and testing of the existing algorithm using different complex systems and introducing more objectives. There is also an opportunity to branch this research into various different industries that are challenged with high dimensional optimisation problems to improve the generalisability of the results.

Bibliography

[1] L. Stander, M. Woolway, and T. L. van Zyl, "Data-driven evolutionary optimisation for the design parameters of a chemical process: A case study", in *2020 23rd International Conference on Information Fusion (FUSION 2020)*, 2020.

[2] L. Stander, M. Woolway, and T. L. van Zyl, "Extended surrogate assisted continuous process optimisation", in *2020 7th International Conference on Soft Computing Machine Intelligence (ISCMI)*, 2020, pp. 275–279. DOI: 10.1109/ISCMI51676.2020.9311560.

[3] L Stander, M Woolway, and T. van Zyl, "Surrogate assisted optimisation for a pressure swing adsorption system", *Neural Computing and Applications*, In Review.

[4] R. Morales-Rodriguez, A. S. Meyer, K. V. Gernaey, and G. Sin, "A framework for model-based optimization of bioprocesses under uncertainty: Identifying critical parameters and operating variables.", in *Computer Aided Chemical Engineering*, vol. 29, Elsevier, 2011, pp. 1455–1459.

[5] F. P. Bernardo, E. N. Pistikopoulos, and P. M. Saraiva, "Quality costs and robustness criteria in chemical process design optimization", *Computers & Chemical Engineering*, vol. 25, no. 1, pp. 27–40, 2001.

[6] G. P. Rangaiah, *Multi-objective optimization: techniques and applications in chemical engineering*. world scientific, 2009, vol. 1.

[7] S. Razavi, B. A. Tolson, and D. H. Burn, "Review of surrogate modeling in water resources", *Water Resources Research*, vol. 48, no. 7, 2012.

[8] M. Emmerich, M. Grötzner, and M. Schütz, "Design of graph-based evolutionary algorithms: A case study for chemical process networks", *Evolutionary Computation*, vol. 9, no. 3, pp. 329–354, 2001.

[9] M. Meyer, H. Robinson, M. Fisher, A. Van der Merwe, G. Streicher, J. J. Van Rensburg, H. Van den Berg, E. Dreyer, J. Joubert, G. Bonthuys, *et al.*, "Innovative decision support in a petrochemical production environment", *Interfaces*, vol. 41, no. 1, pp. 79–92, 2011.

[10] D. Whitley, "A genetic algorithm tutorial", *Statistics and computing*, vol. 4, no. 2, pp. 65–85, 1994.

[11] T. Chugh, N. Chakraborti, K. Sindhya, and Y. Jin, "A data-driven surrogate-assisted evolutionary algorithm applied to a many-objective blast furnace optimization problem", *Materials and Manufacturing Processes*, vol. 32, no. 10, pp. 1172–1178, 2017.

[12] Y. Jin, H. Wang, T. Chugh, D. Guo, and K. Miettinen, "Data-driven evolutionary optimization: An overview and case studies", *IEEE Transactions on Evolutionary Computation*, vol. 23, no. 3, pp. 442–458, 2018.

[13] L Jourdan, D. Corne, D. Savic, and G Walters, "Lemmo: Hybridising rule induction and nsga ii for multi-objective water systems design", in *Procceedings of the Eighth International Conference on Computing and Control for the Water Industry*, vol. 2, 2006, pp. 45–50.

[14] T. Back, *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*. Oxford university press, 1996.

[15] M. Woolway and T. Majozi, "On the application of a metaheuristic suite with parallel implementations for the scheduling of multipurpose batch plants", *Computers & Chemical Engineering*, vol. 126, pp. 371–390, 2019.

[16] ——, "A novel metaheuristic framework for the scheduling of multipurpose batch plants", *Chemical Engineering Science*, vol. 192, pp. 678–687, 2018.

[17] Z. Bowditch, M Woolway, and T. van Zyl, "Comparative metaheuristic performance for the scheduling of multipurpose batch plants", in *2019 6th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, IEEE, 2019, pp. 121–125.

[18] T. van Zyl and M Woolway, "Makespan minimisation for multipurpose batch plants using metaheuristic approaches", in *2020 7th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, IEEE, 2020, pp. 56–60.

[19] T. Simpson, V. Toropov, V. Balabanov, and F. Viana, "Design and analysis of computer experiments in multidisciplinary design optimization: A review of how far we have come-or not", in *12th AIAA/ISSMO multidisciplinary analysis and optimization conference*, 2008, p. 5802.

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

[21] B. Beykal, F. Boukouvala, C. A. Floudas, and E. N. Pistikopoulos, "Optimal design of energy systems using constrained grey-box multi-objective optimization", *Computers & chemical engineering*, vol. 116, pp. 488–502, 2018.

[22] S. Park, J. Na, M. Kim, and J. M. Lee, "Multi-objective bayesian optimization of chemical reactor design using computational fluid dynamics", *Computers & Chemical Engineering*, vol. 119, pp. 25–37, 2018.

[23] D. Yancy-Caballero, K. T. Leperi, B. J. Bucior, R. K. Richardson, T. Islamoglu, O. K. Farha, F. You, and R. Q. Snurr, "Process-level modelling and optimization to evaluate metal–organic frameworks for post-combustion capture of co 2", *Molecular Systems Design & Engineering*, vol. 5, no. 7, pp. 1205–1218, 2020.

[24] F. Ye, S. Ma, L. Tong, J. Xiao, P. Bénard, and R. Chahine, "Artificial neural network based optimization for hydrogen purification performance of pressure swing adsorption", *International Journal of Hydrogen Energy*, vol. 44, no. 11, pp. 5334–5344, 2019.

[25] K. T. Leperi, D. Yancy-Caballero, R. Q. Snurr, and F. You, "110th anniversary: Surrogate models based on artificial neural networks to simulate and optimize pressure swing adsorption cycles for co2 capture", *Industrial & Engineering Chemistry Research*, vol. 58, no. 39, pp. 18 241–18 252, 2019.

[26] S. G. Subraveti, Z. Li, V. Prasad, and A. Rajendran, "Machine learning-based multiobjective optimization of pressure swing adsorption", *Industrial & Engineering Chemistry Research*, vol. 58, no. 44, pp. 20 412–20 422, 2019.

[27] J. Xiao, C. Li, L. Fang, P. Böwer, M. Wark, P. Bénard, and R. Chahine, "Machine learning–based optimization for hydrogen purification performance of layered bed pressure swing adsorption", *International Journal of Energy Research*, vol. 44, no. 6, pp. 4475–4492, 2020.

[28] K. N. Pai, V. Prasad, and A. Rajendran, "Experimentally validated machine learning frameworks for accelerated prediction of cyclic steady state and optimization of pressure swing adsorption processes", *Separation and Purification Technology*, vol. 241, p. 116 651, 2020.

[29] L. E. Perez, P. Sarkar, and A. Rajendran, "Experimental validation of multi-objective optimization techniques for design of vacuum swing adsorption processes", *Separation and Purification Technology*, vol. 224, pp. 553–563, 2019.

[30] R. Haghpanah, A. Majumder, R. Nilam, A. Rajendran, S. Farooq, I. A. Karimi, and M. Amanullah, "Multiobjective optimization of a four-step adsorption process for postcombustion co2 capture via finite volume simulation", *Industrial & Engineering Chemistry Research*, vol. 52, no. 11, pp. 4249–4265, 2013.

[31] J. Beck, D. Friedrich, S. Brandani, and E. S. Fraga, "Multi-objective optimisation using surrogate models for the design of vpsa systems", *Computers & Chemical Engineering*, vol. 82, pp. 318–329, 2015.

[32] R. Haghpanah, R. Nilam, A. Rajendran, S. Farooq, and I. A. Karimi, "Cycle synthesis and optimization of a vsa process for postcombustion co2 capture", *AIChE Journal*, vol. 59, no. 12, pp. 4735–4748, 2013.

[33] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii", *IEEE transactions on evolutionary computation*, vol. 6, no. 2, pp. 182–197, 2002.

[34] N. Riquelme, C. Von Lücken, and B. Baran, "Performance metrics in multi-objective optimization", in *2015 Latin American Computing Conference (CLEI)*, IEEE, 2015, pp. 1–11.

[35] H. R. S. Anna, A. G. Barreto Jr, F. W. Tavares, and M. B. de Souza Jr, "Machine learning model and optimization of a psa unit for methane-nitrogen separation", *Computers & Chemical Engineering*, vol. 104, pp. 377–391, 2017.

[36] W. Ali, M. S. Khan, M. A. Qyyum, and M. Lee, "Surrogate-assisted modeling and optimization of a natural-gas liquefaction plant", *Computers & Chemical Engineering*, vol. 118, pp. 132–142, 2018.

[37] F. Boukouvala, M. F. Hasan, and C. A. Floudas, "Global optimization of general constrained grey-box models: New method and its application to constrained pdes for pressure swing adsorption", *Journal of Global Optimization*, vol. 67, no. 1-2, pp. 3–42, 2017.

[38] R. R. Carpio, R. C. Giordano, and A. R. Secchi, "Enhanced surrogate assisted global optimization algorithm based on maximizing probability of improvement", in *Computer Aided Chemical Engineering*, vol. 40, Elsevier, 2017, pp. 2065–2070.

[39] X. Shi, C. Tong, and L. Wang, "Evolutionary optimization with adaptive surrogates and its application in crude oil distillation", in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, 2016, pp. 1–8.

[40] A. Bhosekar and M. Ierapetritou, "Advances in surrogate based modeling, feasibility analysis, and optimization: A review", *Computers & Chemical Engineering*, vol. 108, pp. 250–267, 2018.

[41] R. Angira and B Babu, "Evolutionary computation for global optimization of nonlinear chemical engineering processes", in *Proceedings of International Symposium on Process Systems Engineering and Control, Mumbai*, 2003, pp. 87–91.

[42] L. M. C. Oliveira, H. Koivisto, I. G. Wakiri, J. M. Loureiro, A. M. Ribeiro, and I. B. Nogueira, "Modelling of a pressure swing adsorption unit by deep learning and artificial intelligence tools", *Chemical Engineering Science*, p. 115 801, 2020.

[43] R. C. Calfee and M. Chambliss, "The design of empirical research", in *Methods of Research on Teaching the English Language Arts*, Routledge, 2005, pp. 53–88.

[44] D. M. I. Akhtar, "Research design", *Available at SSRN 2862445*, 2016.

[45] L. J. Eshelman and J. D. Schaffer, "Real-coded genetic algorithms and interval-schemata", in *Foundations of genetic algorithms*, vol. 2, Elsevier, 1993, pp. 187–202.

[46] C. Chudasama, S. Shah, and M. Panchal, "Comparison of parents selection methods of genetic algorithm for tsp", in *International Conference on Computer Communication and Networks CSI-COMNET-2011, Proceedings*, 2011, pp. 85–87.

[47] C. R. Raquel and P. C. Naval Jr, "An effective use of crowding distance in multiobjective particle swarm optimization", in *Proceedings of the 7th Annual conference on Genetic and Evolutionary Computation*, 2005, pp. 257–264.

[48] Y. Fang and J. Li, "A review of tournament selection in genetic programming", in *International Symposium on Intelligence Computation and Applications*, Springer, 2010, pp. 181–192.

[49] C. M. Fonseca, L. Paquete, and M. López-Ibánez, "An improved dimension-sweep algorithm for the hypervolume indicator", in *2006 IEEE international conference on evolutionary computation*, IEEE, 2006, pp. 1157–1163.

[50] L. Stander, *Surrogate assisted optimisation*, https://github.com/liezlstander/Surrogate_Assisted_Optimisation, 2021.