
Otto-von-Guericke University Magdeburg



Faculty of Process Systems Engineering
Institute of Process Engineering [IVT]

Master Thesis

Surrogate-Based Optimization of Extractive Distillation Processes Using Active Learning

Author:
Saikiran Davangam Sreedhar

April 17, 2023

Reviewer
Prof. Kai Sundmacher
Supervisor
M.Sc. Zihao Wang

Faculty of Process Systems Engineering
Otto-von-Guericke University
Universitätsplatz 2
39106 Magdeburg, Germany

Reviewer
Ph.D. Xiang Zhang
Supervisor
M.Sc. Edgar Ivan Sanchez Medi

Faculty of Process Systems Engineering
Otto-von-Guericke University
Universitätsplatz 2
39106 Magdeburg, Germany



Antrag auf Zulassung zur Masterarbeit
 Application for Approval of the Master Thesis

Im Studiengang: Chemical and Energy Engineering
 In the course of study

Name, Vorname: Saikiran Davangam Sreedhar
 Surname, First Name

Geboren am: 03.09.1996 **in** India
 Birthday Place of Birth

Matrikelnummer: 229971 **Seminargruppe:** FVST
 Registration Number Seminar Group

Adresse: Walther-Rathenau-Straße 58
 Address

Tel. Nr.: 0176 5573 4467 **Student-Email:**
 Phone Number

Hiermit beantrage ich die Zulassung zur Masterarbeit. Meine Masterarbeit auf der Grundlage
 der für mich geltenden Prüfungsordnung hat folgendes Thema:

Herewith I apply for the approval of the master's thesis. My master's thesis is based on the applicable examination
 regulations and is on the following topic:

Surrogate-Based Optimization of Extractive Distillation Processes.....

Using Active Learning.....

.....
 Bitte ankreuzen / Please place checkmark:

Es ist eine Einzelarbeit:
 It is a single thesis

Gruppenarbeit:
 team thesis

Erstgutachter ist: Prof. Dr.-Ing. Kai Sundmacher
 The primary reviewer is FVST OVGU / PSE MPI
 Institut:
 Institute:

Ich bin bereit, die Arbeit als Erstgutachterin / Erstgutachter zu betreuen.
 She /He commits herself /himself to supervise the thesis as the primary supervisor.



(Unterschrift der Erstgutachterin / des Erstgutachters)
 (Signature of the primary supervisor)

Zweitgutachter ist: Ph.D. Xiang Zhang
 The secondary reviewer is FVST OVGU / PSE MPI
 Institut:
 Institute:

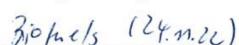
Ich bin bereit, die Arbeit als Zweitgutachterin / Zweitgutachter zu betreuen.
 She /He commits herself /himself to supervise the thesis as the secondary reviewer.



(Unterschrift der Zweitgutachterin / des Zweitgutachters)
 (Signature of the secondary supervisor)

Die folgenden gemäß Studien- und Prüfungsordnung geforderten Studien- und Leistungs nachweise sind noch offen:

According to the study and examination regulations the following records still have to be delivered:





Hiermit beantrage ich die Zulassung zur Masterarbeit
 Herewith I apply for the approval of the master thesis.

Magdeburg den 21.10.22
 Ort/Place Datum/Date
 Unterschrift der/ des Studierenden
 Signature of the student

Eine gültige Immatrikulationsbescheinigung ist dem Antrag beizufügen.
 A valid confirmation of registration must be attached to this application.

**Bestätigung des Prüfungsamtes der Fakultät für Verfahrens- und Systemtechnik der
Otto-von-Guericke-Universität Magdeburg**
Approval by the examination office of the Faculty of Process and Systems Engineering of the Otto von Guericke
University of Magdeburg

Alle gemäß Studien- und Prüfungsordnung geforderten Studien- und Leistungsnachweise,
ausgenommen der oben aufgeführten Leistungen, liegen vor.

According to the study and examination regulations, all records were delivered except for the ones listed above.

Datum, Unterschrift *14.11.22*
Date, Signature

Bestätigung der / des Prüfungsausschussvorsitzenden
Approval by the chairman of the board of examiners

Der Prüfungsausschuss ist mit der Ausgabe des o. g. Themas einverstanden.
The board of examiners agrees to the release of the topic mentioned above.

Der Ausgabetermin ist am: *1.11.22*
Date of release

Der Abgabetermin ist am: *11.3.23*
Date of submission

Datum, Unterschrift *15.11.2022* *Franz Scheffler*
Date, Signature
Prof. Dr. rer. nat. F. Scheffler
Prüfungsausschussvorsitzende



Otto-von-Guericke-Universität Magdeburg, Postfach 4120, 39016 Magdeburg

Student: Saikiran Davangam Sreedhar
Master course: Chemical and Energy Engineering
Immatriculation Number: 229971

INSTITUT VERFAHRENSTECHNIK
 Lehrstuhl Systemverfahrenstechnik

Prof. Kai Sundmacher

Otto-von-Guericke-Universität Magdeburg
 Universitätsplatz 2
 39106 Magdeburg
 Telefon: +49 391 67-58414
 Telefax: +49 391 67-11245
 kai.sundmacher@ovgu.de
 www.ovgu.de/ivt/vst

— Ihre Zeichen, Ihre Nachricht vom: Unsere Zeichen Durchwahl: Datum: 21.10.2022

Task of Master Thesis

Title: Surrogate-Based Optimization of Extractive Distillation Processes Using Active Learning

Background

As the trend in process systems engineering moves towards integrated optimization considering multiple hierarchical levels (i.e., the molecular, phase, unit, and process levels), we are benefiting from the ever-increasing model complexity at the expense of computational efficiency. Process optimization is usually formulated by complex mathematical models (e.g., thermodynamics, reaction, and transport), where many discrete degrees of freedom are involved [1]. This leads to mixed integer nonlinear programming (MINLP) problems that are hard to solve and often involve difficult initialization procedures [2], especially for distillation processes. In this context, process optimization presents an increasing demand for simple yet accurate process-related models as alternatives.

Surrogate modeling provides a practical approach to build simple yet accurate models to replace the original complex models involved in process optimization [3]. Different machine learning (ML) methods such as the artificial neuron network and Gaussian process regression are capable of generating powerful surrogate models to approximate key performance indicators of the chemical process using process variables as inputs. In general, a large amount of data is necessary to obtain an accurate surrogate. Therefore, an accurate global surrogate model requires high computational costs in data generation (i.e., data labeling).

Active learning (AL) offers an efficient way to build accurate models with less data. Specifically, an initial model is established with a very small dataset. Based on this model, a few unknown data to be labeled are selected from a large pool (i.e., unknown data selection). After labeling the selected unknown data (i.e., data labeling), the dataset is extended and correspondingly the model is updated to have higher accuracy (i.e., model update). With an iterative process consisting of unknown data selection, data labeling, and model update, a satisfactory model is obtained finally. In such an AL manner, minimal efforts are required in the data labeling to build accurate models.

Aim

The aim of this master thesis is to develop simple yet reliable surrogate models in an AL manner to support the efficient optimization of extractive distillation (ED) processes. ML is used to

approximate key performance indicators (i.e., product purity and energy consumption) of the ED process using process variables (i.e., solvent-to-feed ratio, the number of stages, reflux ratio, and operating pressure) as inputs. The data labeled by rigorous process simulations in Aspen Plus is used to train and test surrogate models. Based on the surrogate models, process optimization is performed to identify optimal process variables (i.e., unknown data selection) for subsequent verification (i.e., data labeling). With a series of surrogate update, surrogate-based optimization, and performance verification in an AL manner, sufficiently accurate surrogate models are obtained finally, with less computational efforts in the data labeling. Meanwhile, optimal process variables can be obtained to achieve desired process performance.

Task list

- Task 1: Create a ED process flowsheet via Aspen Plus
- Task 2: Perform the design of computer experiments to get key performance indicators of the ED process under varying process variables
- Task 3: Establish surrogate models for the ED process using ML
- Task 4: Optimize the ED process based on the established surrogate models
- Task 5: Perform surrogate update, surrogate-based optimization, and rigorous process simulation in an AL manner

Literature

- [1] Pistikopoulos, E. N., Barbosa-Povoa, A., Lee, J. H., Misener, R., Mitsos, A., Reklaitis, G. V., Venkatasubramanian, V., You, F., & Gani, R. Process systems engineering—the generation next?. *Computers & Chemical Engineering*, 2021, 147, 107252.
- [2] Quirante, N., Javaloyes, J., & Caballero, J. A. Rigorous design of distillation columns using surrogate models based on Kriging interpolation. *AIChE Journal*, 2015, 61(7), 2169-2187.
- [3] McBride, K., & Sundmacher, K. Overview of surrogate modeling in chemical process engineering. *Chemie Ingenieur Technik*, 2019, 91(3), 228-239.

Duration of thesis:	20 weeks
Start of thesis work:	01.11.2022
Reviewer I of thesis work:	Prof. Kai Sundmacher
Reviewer II of thesis work:	Ph.D. Xiang Zhang
Supervisor I of thesis work:	M.Sc. Zihao Wang
Supervisor II of thesis work:	M.Sc. Edgar Ivan Sanchez Medina

Magdeburg, den 1.11.2022


Prof. Kai Sundmacher
Chair of Process Systems Engineering

Davangam Sreedhar, Saikiran:

Surrogate-Based Optimization of Extractive Distillation Processes Using Active Learning

Master Thesis, Otto-von-Guericke University
Magdeburg, 2023.

Declaration by candidate

I hereby declare that this master's thesis is completely my own work and effort. Furthermore more, it has not been submitted anywhere for any award and other work used as a resource has been properly acknowledged by providing in the reference. My thesis work has not been presented in the same or a similar form to any other testing authority and has not been made public.

Magdeburg, 17-04-2023

(Saikiran Davangam Sreedhar)

Acknowledgement

I would like to express my sincere gratitude and appreciation to **M.sc. Zihao Wang** for providing me with invaluable guidance and support throughout my Master's thesis. His expertise, insights, and feedback have been invaluable in helping me navigate the research process. I would also like to thank **Prof. Kai Sundmacher, Ph.D. Xiang Zhang, M.Sc. Edgar Ivan Sanchez Medi** for providing an opportunity to do my thesis. Their encouragement and advice have been instrumental in shaping my work. I would also thank my fellow mates **Venkat Mannam Srinath, Muralidhar Reddy Kuluru, Kranti Kumar Talluri, and Mohammed Maqsood** for helping me throughout this work. Lastly, I would like to acknowledge the unwavering support and encouragement of my mother **Koti Manjulavani Garu** and my brother **Saiteja Devangam Sreedhar**. Their love and encouragement have kept me motivated throughout this journey.

Contents

1 Introduction

1.1 Extractive Distillation	13
1.2 Input variables	14
1.3 Single objective optimization	16

2 Background

2.1 Distillation	18
2.1.1 Mathematical model	18
2.1.2 Types of Distillation	20
2.2 Surrogate Modelling	22
2.2.1 Basics	22
2.2.2 Characteristics	22
2.2.3 Workflow of Surrogate modeling	23
2.3 Machine learning	25
2.3.1 Basics of Machine Learning	25
2.3.2 Machine Learning Systems	26
2.3.3 Gaussian Process	29
2.4 Active learning	33

3 Related Work

3.1 Surrogate based optimization	35
3.1.1 Batch distillation	35
3.1.2 Stirring tank	36
3.2 Sequential learning	37
3.2.1 Strategies and Utility	38

4 Evaluation and Experiments

4.1 Optimization	40
4.1.1 Basics of optimization	40
4.1.2 General problem statement	41
4.1.3 Global and Local minimum	42
4.2 Bayesian Optimization	43

4.2.1	Acquisition functions	44
4.3	Example	48
4.4	Problem statement	50
4.5	Results	53
4.5.1	Minimization of reboiler heat duty with low-quality data .	53
4.5.2	Minimization of reboiler heat duty with high-quality data .	57
4.5.3	Comparison	62
4.6	Discussion	63
4.6.1	Case-1(Low-quality data)	63
4.6.2	Case-2(High-quality data)	64
4.6.3	Comparison	64

5 Conclusions and Future Work

5.1	Conclusions	65
5.2	Future Work	66

A List of Figures**B List of Tables****C List of Libraries****D Bibliography**

Abstract

Azeotropic or close-boiling mixtures can be successfully separated using extractive distillation. However, because there are so many design factors and there are so many intricate interactions between the process variables, extractive distillation column design and optimization are difficult and costly in terms of computing. A possible method for improving the creation of the extractive distillation column is surrogate-based optimization. In this study, a surrogate model for the extractive distillation column is developed using an active learning-based methodology. The suggested method strikes a balance between accuracy and computing efficiency by utilizing an active learning algorithm and a Gaussian process (GP) surrogate model. Based on the Gaussian process model's uncertainty, the active learning algorithm chooses the next sample point for evaluation. Using a case study of an extractive distillation column used to separate a 1-butene/1,3-butadiene mixture using N-methyl-2-pyrrolidone(NMP) as the solvent, the performance of the suggested method is assessed. The findings demonstrate that the suggested strategy performs better in terms of accuracy and computing efficiency. As a result, the suggested method may be a helpful tool for improving the design of the extractive distillation column.

1

Introduction

Engineering process design heavily relies on a careful examination of the process's performance in relation to certain design aspects [2]. These evaluations are typically performed using time-consuming, high-resolution computer simulations. Accelerating these simulation-based analyses is critical for getting the product to market faster. Surrogate modeling[34], a data-driven approach, has recently gained popularity in various engineering domains. Computer simulations aim to forecast the conduct of a real-world system by solving the mathematical equations associated with its physical operations. Engineers, for instance, use aerodynamic experiments to compute an aircraft's drag and lift, and structural dynamic simulations to determine the strength of the aircraft's wings. These experiments are important because they offer developers valuable data on the product's performance without practically constructing the product , making them essential for virtual prototyping. Engineers are typically required to perform the following tasks to make the method of design more efficient and dependable:

- **Sensitivity analysis:** To investigate product behavior when design parameters vary;
- **Optimization:** To determine the best design parameters for maximum productivity and/or lowest possible cost;
- **Risk analysis/uncertainty quantification:** To calculate the risk of a product malfunction in instances where the design variables are uncertain.

All the analyses mentioned earlier share a common feature: they require a considerable number of simulation runs, with each run necessitating distinct design parameters as inputs. However, computer simulations are expensive and may take days to complete on an industry cluster. As a result, carrying out analyses that need a large number of simulations would be impractical due to high computational costs. Surrogate modeling can offer a potential solution to this issue.

Optimization can occur in several applications[11]. When it comes to the power plant, energy consumption, for instance, is one of the optimization challenges that can be taken into consideration. A process plant's energy use can be optimized to decrease costs and have a smaller negative impact on the environment. Designing a flow system is one of the optimization problems that can be taken into account while discussing fluid flow systems. Optimizing a fluid flow system can help improve efficiency, reduce operating costs, and increase equipment lifespan. When it comes to separation process optimization, both product purity and separation process optimization can be thought of as optimization problems. Optimizing separation processes can help improve product quality, increase throughput, and reduce operating costs. Finding the ideal set of operating parameters that maximize a particular objective function, such as reaction efficiency, product yield, or energy efficiency, is the goal of optimization problems in reactors.

1.1 Extractive Distillation

The best distillation process design can significantly impact the profitability of a plant because distillation is arguably the most significant separation technology in the chemical and petrochemical industries [3]. Distillation column work based on the different volatilities of the components that are being separated. But occasionally, when azeotropes are present or the relative volatility is almost unity, ordinary distillation fails to provide the appropriate separation. Heterogeneous azeotropic distillation and homogeneous extractive distillation, sometimes known as extractive distillation, are the two main commercial techniques used to separate close-boiling and azeotropic

mixtures. Extractive distillation often uses less energy than heterogeneous azeotropic distillation, especially when heat integration is taken into account. In extractive distillation, an appropriate separating solvent called an "entertainer" is introduced to the initial mixture to raise the relative volatility of the compounds and so facilitate separation.

1.2 Input variables

Depending on the particular system and circumstances, several input variables may be used in extractive distillation. However, some of the most important input factors that could impact the effectiveness and efficiency of extractive distillation include:

Feed stream:

The combination of two or more components that need to be separated is known as this. The feed stream's composition and flow rate are crucial input factors.

Entrainor or solvent:

This is the third element that is added to the distillation column to aid in the feed stream separation. Critical input variables include the trainer's type, makeup, and flow rate.

Operating pressure:

Because the column's operating pressure has an impact on the components of the mixture's vapor-liquid equilibrium, it is a crucial input variable.

Operating temperature:

The temperature of the column controls how quickly the components vaporize and condense, making it a crucial input variable.

Reflux ratio:

The ratio of the vapor exiting the top of the column to the vapor that has condensed is known as the reflux ratio. It is a crucial input variable that has an impact on separation efficiency.

Column height:

The separation efficiency is influenced by the column's height, which is a crucial input factor.

Tray or packing design:

The layout of the trays or packing material in the column has an impact on the effectiveness of the separation and is a crucial input factor.

Heat integration:

In the distillation column, heat integration entails the transfer of heat across various streams. The process's energy efficiency is impacted by the heat integration technique, which is a crucial input factor.

Entrainment ratio:

The ratio of the solvent flow rate to the feed flow rate is known as the entrainment ratio. It is a crucial input variable that has an impact on separation efficiency.

Feed location:

An important input variable that influences separation efficiency is where the feed tray or column point is located.

1.3 Single objective optimization

Single objective optimization[51] is commonly used in the chemical field to find the optimal conditions for a chemical process that maximize a certain objective, such as yield, selectivity, or efficiency. For example, let's consider a chemical reaction that produces a certain compound from two reactants. The reaction is carried out in a batch reactor[38], and the objective is to maximize the target product's yield while minimizing the formation of unwanted byproducts. The reaction rate can be modeled by a kinetic equation [30] that depends on the concentrations of the reactants and the reaction conditions, such as temperature, pressure, and catalyst concentration. The yield of the desired product could be written using a function of the reaction rate and the stoichiometry of the reaction. The problem can be represented as a single objective optimization with constraints, and the objective is to maximize the yield of the desired product subject to the constraints on the reaction conditions and the formation of unwanted byproducts. The solution to this issue can be solved using optimization techniques such as gradient-based methods, genetic algorithms, or particle swarm optimization. The optimal reaction conditions can then be used to design and optimize the reactor, such as selecting the appropriate reactor size, type, and operating conditions.

Another example of single objective optimization in the chemical field is the optimization of a distillation column. The distillation column separates a mixture of two or more components into its components based on their boiling points. The objective is to maximize the purity or recovery of a certain component while minimizing the energy consumption of the distillation process. The problem can be represented as a single objective optimization with constraints, and the objective is to maximize the purity or recovery of the desired component subject to the constraints on the energy consumption and the separation efficiency of the distillation column. The solution to this problem can be found using optimization techniques such as dynamic programming, simulated annealing, or genetic algorithms. The optimal design and operating conditions of the distillation column can then be determined according to the solution of the optimization problem.

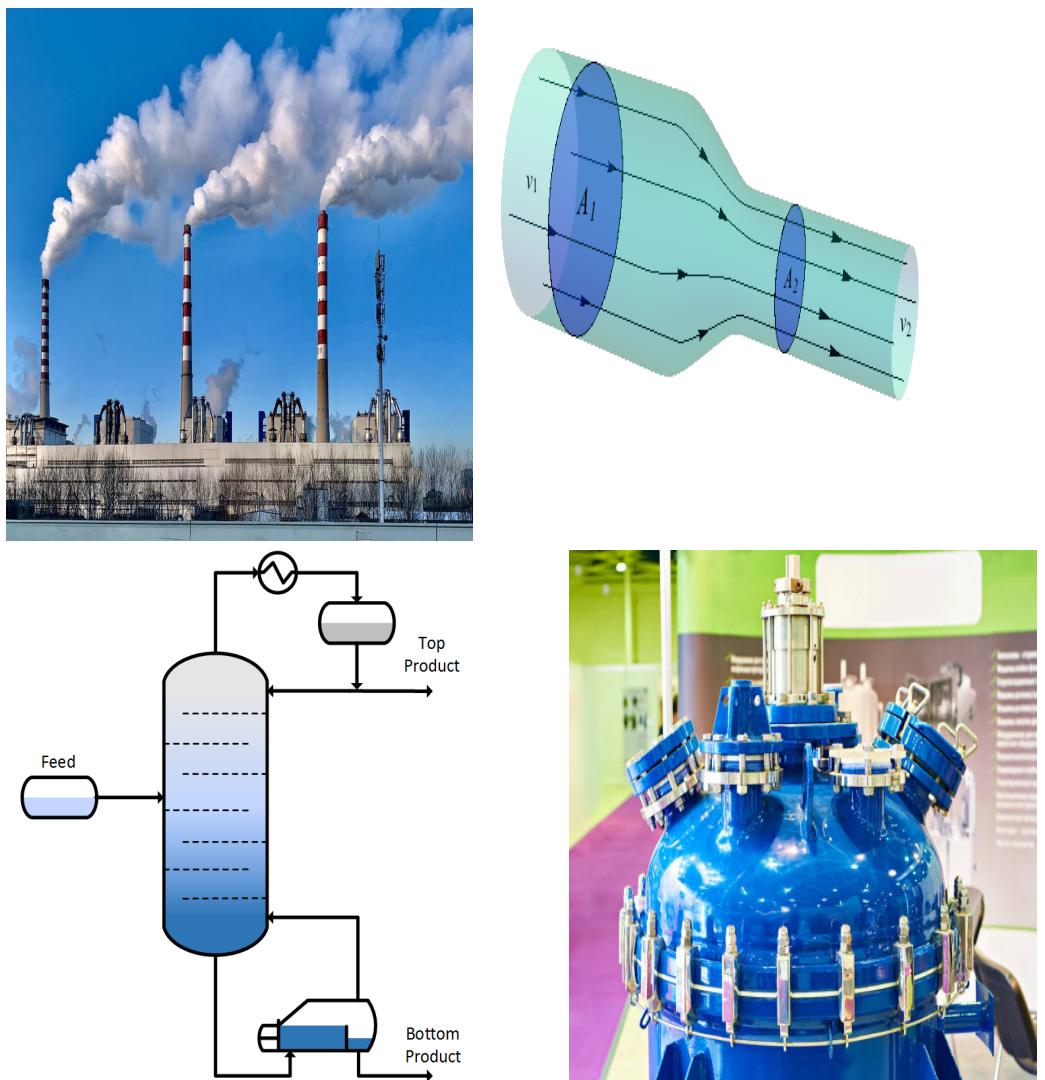


Figure 1.1: Different applications of Optimization

2

Background

This section presents a high-level overview of distillation, surrogate modeling, and machine learning concepts. More detailed handling of this topics can be found in [14], [36], [17] and [45] respectively

2.1 Distillation

Distillation is a significant separation task used to separate light substances from solutions or mixes[29]. It is established by Raoult's law[4], which declares that a solvent's partial vapor pressure is directly proportional to its mole fraction. Distillation is well recognized in the chemical process industries for liquid mixture separation; it accounts for approximately 95% of all liquid separations[23]. This function is energy-intensive, with significant needs for energy. The operation's effectiveness and productivity are dependent on its efficiency and economy. In several scenarios, it is feasible to integrate multiple operations to reduce drawbacks while maximizing benefits. Simulation and modeling techniques for distillation columns are also hot topics.

2.1.1 Mathematical model

The demand for condensate useful outputs is measured by the purification of the overhead distillate, X_D , which might be greater than/equal to 98%, as well as the bottom product's impurity X_B , might be less than or equal to 2%. A distillate column with 14 columns can be mathematically calculated to 16 non-

linear differential equations[26]. Some components of the feedstock have been grouped together for analysis. The feedstock might be a binary component, as in the mixture of liquefied petroleum gas (iso-butane, n-butane, and propane), and occurs in nature (iso-pentane, n-pentane, and heavier components). Depending on the distillate tower feeding composition, the binary component characteristics will vary within the attached range. Under many assumptions, rigorous governing equations for 16 columns, such as the section of reboiler and condenser, are developed:

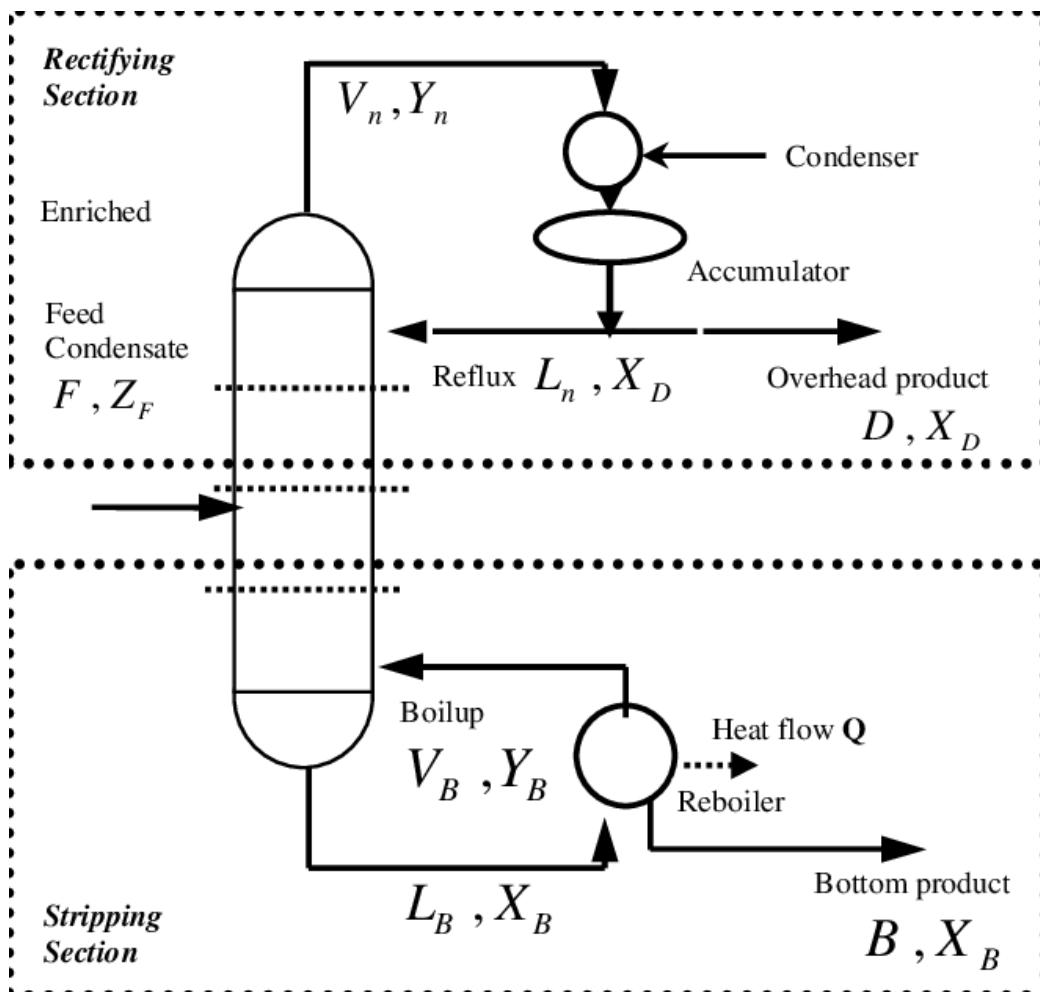


Figure 2.1: Mathematical model of distillation

(i) Condenser ($N + 2$) :

$$M_c X_N = (V_N + V_{N+1} + V_f) Y_{N+1} - L, X_D - D, X_D$$

(ii) Feed Tray :

$$M_N X_N = (V_N + V_{N+1} + V_f)(Y_{N+1} - Y_n) \\ - (L_N, X_N + L_{N+1}, X_{N+1})$$

(iii) Upper feed :

$$M_N X_N = (V_N + V_{N+1} + V_f)(Y_{N+1} - Y_n) \\ - (L_N, X_N + L_{N+1}, X_{N+1}) + V_f, Y_f$$

(iv) Bottom feed:

$$M_N X_N = (V_N + V_{N+1} + V_f)(Y_{N+1} - Y_n) \\ - (L_N, X_N + L_{N+1}, X_{N+1}) + L_f, Y_f$$

(v) Reboiler:

$$M_B X_N = (L_N + L_{N+1} + L_f) X_N \\ - B, X_B - V, Q_B$$

2.1.2 Types of Distillation

Simple distillation

Simple distillation[25] is among the most basic distillation processes, in which the primary component to be purified is evaporated along with impurities at a high temperature, and the generated vapor is transmitted to a container, where it solidifies at a lower temperature.

Fractional distillation

A mixture of fluids is heated in fractional distillation[5], and the resulting steam pass through a transparent tube called a "fractionating column" and separates.

Steam distillation

Steam distillation [7] is a technique used to cleanse, separate, or isolate thermally sensitive substances such as organic aromatic substances. The distillation unit must be supplied with steam or water, and the procedure can be

carried out at lower temperatures by leveraging the properties of insoluble, unmixable liquids.

Vacuum distillation

The vacuum distillation process is utilized to divide higher boiling components of petroleum products[27] which occurs at reduced pressure. The pressure maintained is between 50 to 100mmHg in vacuum distillation.

Reactive distillation

The connectivity of reaction and distillation within the same chamber of time and space of one section[43] is identified as reactive distillation. It is challenging to determine a general purpose operating frame between the reaction and separation that delivers satisfactory and malleable reaction rates and suitable volatilities to ensure high reactant intensities and low product concentration levels in the reactive zone. In such situations, reactive distillation comes handy to deal with such processes. Hydration of cyclohexene to obtain cyclohexanol[8], dehydration of glycerol to acetol[9], isoamyl acetate production via esterification of isoamyl alcohol and acetic acid[16] are some of the examples that come under reactive distillation process.

Extractive distillation

Extractive distillation is a popular distillation method for separating minimum- or maximum-boiling azeotropes and low-relative-volatility mixes. In contrast to azeotropic distillation, the entrainer is loaded in a different place than the main combination to be divided, resulting in an extractive section within the column. Process of separating the binary azeotropic mixture methanol-acetone[21], and separating ethyl acetate-ethanol azeotrope [44] are some of the examples that come under the extractive distillation process.

2.2 Surrogate Modelling

2.2.1 Basics

Surrogate modeling refers to a technique where a statistical model is built to mimic the output of a computer simulation. Data-driven models are the most commonly used among the different types of surrogates available. Surrogate modeling aims to replace the original simulation with a trained statistical model that can be used in various analyses, such as sensitivity analysis, optimization, or risk analysis. The advantage of surrogate models is that they can reduce the time and resources required to conduct expensive simulations. With surrogate models, conducting a large number of output evaluations is no longer a challenge because a single evaluation of simulation output is typically much slower than a single review of the trained statistical model. In short, surrogate modeling techniques can significantly lower the cost of conducting expensive analyses[19].

2.2.2 Characteristics

For developing a surrogate model, a data-driven strategy is used, which is also known as a replacer of the surrogate model or an emulator. It obtains training data by conducting simulation outputs at several intelligently chosen locations in the design parameter space. A full simulation is run at each of these selected locations to calculate the corresponding simulation output. We can create a statistical model by organizing pairs of input design parameters and their corresponding outputs into a training dataset obtained from the above step. Surrogate modeling is a subset of supervised machine learning that is used in engineering. Linear regression, Support vector machines, Gaussian processes, and neural networks are widely used as surrogate models to train the input-output dataset.

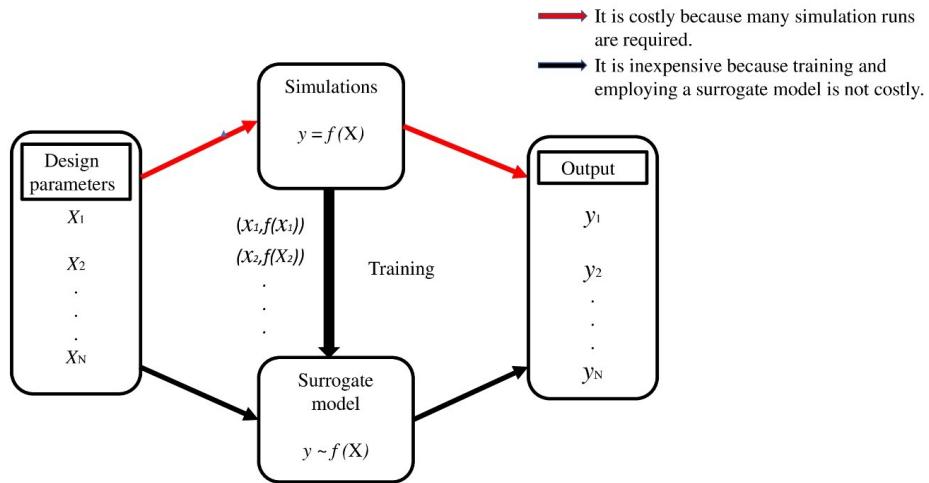


Figure 2.2: Surrogate model v/s Simulation model

2.2.3 Workflow of Surrogate modeling

Sampling:

We commence by gathering preliminary training data by selecting design parameter samples from their parameter space via a method called Experiment Design. It is ideal to have evenly distributed samples across the parameter space at this stage, as it ensures that we have representatives of the input-output relationship approximation from all areas of the parameter space we are considering. This approach offers several advantages.

Evaluations:

Once we have selected the preliminary training data, we need to execute them to obtain their corresponding output values. After combining the selected training samples and their corresponding output values, we have the initial training dataset.

Model building:

In this phase, we construct the surrogate model using the training data we obtained earlier. To ensure accurate model training, we should utilize well-established machine-learning approaches for model validation and selection. Additionally, advanced techniques such as bagging and boosting can enhance the performance of the surrogate model.

Active learning:

Generally, we cannot predict how many samples will be necessary to construct a precise surrogate model. The approximated input-output paired complexity decides this. As the process continues, it makes continuous contributions to enrich the training dataset in the previous step. This is known as active learning. In the end, specially designed learning functions are employed to assist us in order to locate the following sample with the greatest information value. These selected methods of learning are meant to assign samples to locations where the meta-model is expected to be erroneous or uncertain, or to regions that include particularly intriguing consolidation of design variables for instance a region with the potential to find global optimal design parameter values.

Adding new Samples:

After identifying the following sample, a full simulation experiment is carried out to map the input value with the corresponding simulation result as this makes input-output pair that should be added to the initial training data for model updation. We iterate this cycle until we are convinced of the surrogate/meta model's accuracy.

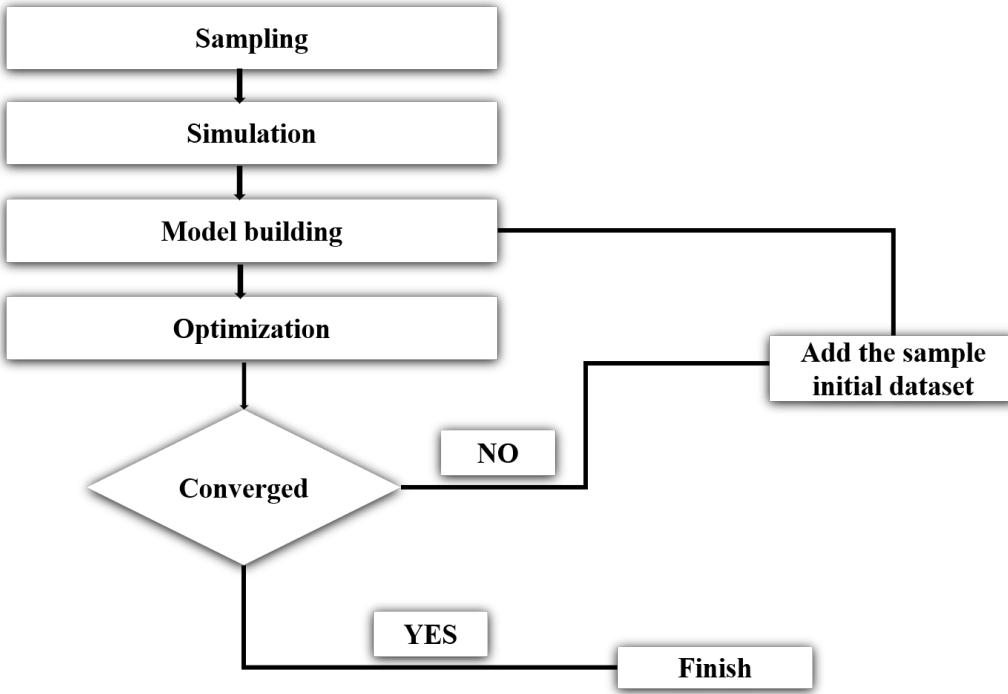


Figure 2.3: Process Workflow

2.3 Machine learning

2.3.1 Basics of Machine Learning

Machine learning (ML)[14] and Artificial intelligence (AI) have observed a surge in demand in the last few decades, fuelled by massive amounts of data and computational power, in addition to the development of enhanced learning techniques. The concept of a computer understanding technique that is abstracted from data and applying it to newly unexplored scenarios, on the other hand, is not new and has been around at least since the 1950s. The availability of data and the increase in computing power, in conjunction with the introduction of innovative training methods, have resulted in several scientific breakthroughs. In machine learning, we make use of data and datasets. A dataset comprises multiple data points (also known as samples),

each of which represents a sample to be analyzed, such as a chemical component or a characteristic of a chemical sample. Data-related challenges are common to all quantitative disciplines, including pharmacometrics but not only for Machine learning approaches. Each feature represents a dimension of the feature space. A feature vector is a compilation of all feature values for a data point. As the no. of features increases the dimensionality of the feature space increases which increases the complexity of the problem.

2.3.2 Machine Learning Systems

As there are several kinds of machine learning techniques they are being differentiated based[14] on the factors that follow:

- Whether systems take the help of humans for training .(supervised, unsupervised, semi-supervised, and Reinforcement learning).
- Whether systems are able to learn on the go in an incremental fashion or not (online versus batch learning).
- By simply comparing unknown points to known points or, like scientists, by predicting patterns in training data and building a predictive model.

Supervised/Unsupervised learning

Machine learning systems are differentiated depending on how much and what kind of supervision they receive and learn throughout training[6]. There are 4 main types of learning: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning.

Supervised learning: In supervised learning, the desired solutions, known as labels, are included in the training data that are given to the machine learning algorithm. Classification is the most well-known supervised learning problem. A spam classifier is considered a good example of a Classification task: it is trained with a large no. of example emails in which they are labeled whether ham or spam and the trained model is used to predict

unknown emails. Another well-known example of a classification task is to predict a target numeric value, such as the price of a car, given a set of car attributes like mileage, age, brand, and so on. This is considered a regression. To train the prediction system, you must provide it with large examples of cars, with their corresponding predictors and labels (i.e., their prices). Some of the common and important supervised learning algorithms are listed below:

- k-Nearest Neighbours
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVMs)
- Decision Trees and Random Forests
- Neural networks

Unsupervised learning: As you might expect, the training data in unsupervised learning is unlabeled. The system attempts to learn without the assistance of a teacher. Some of the most important unsupervised learning algorithms are listed below:

- Clustering
 - ⇒ *K-Means*
 - ⇒ *DBSCAN*
 - ⇒ *Hierarchical Cluster Analysis (HCA)*
 - Anomaly detection and novelty detection
 - ⇒ *One-class SVM*
 - ⇒ *Isolation Forest*
 - Visualization and dimensionality reduction
 - ⇒ *Principal Component Analysis (PCA)*
 - ⇒ *Kernel PCA*
 - ⇒ *Locally-Linear Embedding (LLE)*
 - ⇒ *t-distributed Stochastic Neighbor Embedding (t-SNE)*
-

- Association rule learning
 - ⇒ *Apriori*
 - ⇒ *Eclat*

Semi-supervised learning

Some algorithms can handle training data that has been partly tagged, which typically consists of a large amount of unlabelled data and a small amount of labeled data[1]. This is known as semi-supervised learning. Here are two examples of semi-supervised learning tasks:

1. **Image classification:** In this task, a model is trained to classify images into different categories, such as cats or dogs. A semi-supervised approach would involve training the model on a small set of labeled images, along with a larger set of unlabeled images. The model would learn from the labeled images to identify common features that correspond to different categories, and then use those features to classify the unlabeled images.
2. **Natural Language Processing:** Another example of a semi-supervised learning task is natural language processing (NLP), where the goal is to extract meaning from text data. For example, a model might be trained to identify the sentiment of a movie review as positive or negative. A semi-supervised approach would involve training the model on a small set of labeled reviews, along with a larger set of unlabeled reviews. The model would learn from the labeled reviews to identify common patterns of positive or negative sentiment, and then use those patterns to classify the unlabeled reviews.

Reinforcement Learning

Machine learning includes reinforcement learning as a subfield that deals with how an agent can learn to make decisions in an environment in order to maximize some notion of cumulative reward [18]. It is based on the concept of trial-and-error learning, where an agent interacts with an environment and learns from the feedback it receives in the form of rewards or penalties. The agent observes the state of the environment, takes an action and receives a reward or penalty based on the action taken. The agent's purpose is to discover a strategy that maximizes the predicted cumulative re-

ward over the course of time. Reinforcement learning algorithms use this feedback to learn the optimal policy for the agent to take in a given environment. One of the main challenges in reinforcement learning is the trade-off between exploration and exploitation. The agent needs to balance exploring new actions and exploiting its current knowledge to maximize the cumulative reward. Reinforcement learning has applications in many areas, including robotics, game playing, finance, and healthcare. Some popular algorithms in reinforcement learning include Q-learning, policy gradient methods, and actor-critic methods.

2.3.3 Gaussian Process

Basics

The Gaussian process is a powerful machine-learning technique that can be used for regression and classification tasks[40]. They are a type of probabilistic model that defines a probability distribution over functions, rather than just a point estimate like in many other machine learning models. In a Gaussian process, a function is represented as a distribution over its possible values at any set of input points. This distribution is defined by a mean function and a covariance function, both of which are assumed to be Gaussian. The mean function defines the expected value of the function at any input point, and the covariance function describes how the values of the function at different input points are related. Basic knowledge of multivariate normal distribution, kernels, non-parametric models, and joint and conditional probability is required to better understand Gaussian processes.

Mathematical basics

In regression, We would like to fit a function to some identified sample points followed by employing the function to predict additional points of data. For a given set of identified sample points shown in figure 2.4, there are 'N' different functions that can be used to represent the points shown.

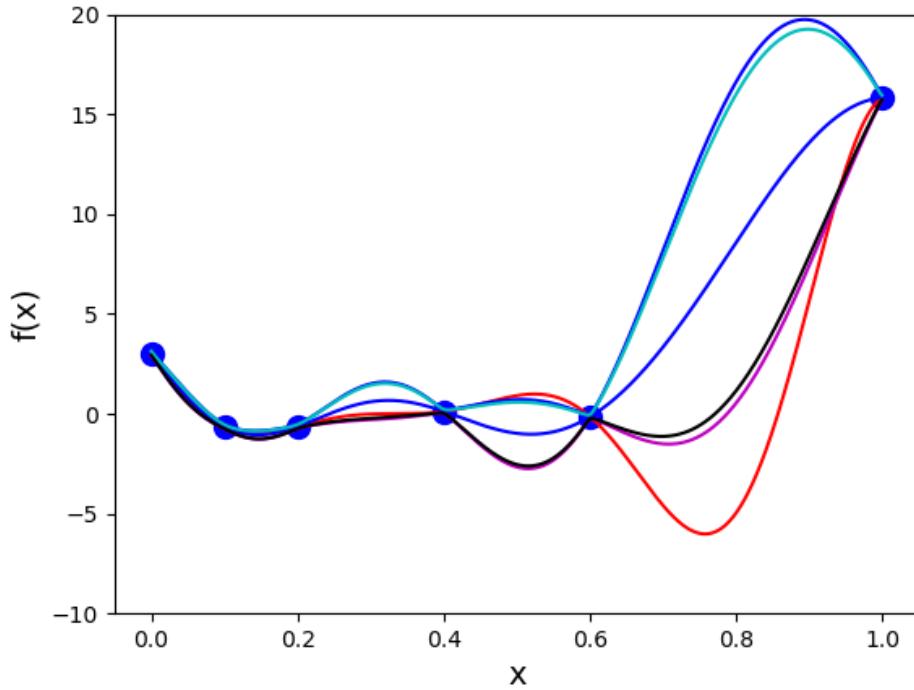


Figure 2.4: 'N' no. of functions

Gaussian Distribution

A random variable X with a mean μ and variance σ^2 can be described as Gaussian or normally distributed if and only if its probability density function (PDF) is represented by:

$$P_X(x) = \frac{1}{\sqrt{2 * \pi * \sigma}} * \exp\left(-\frac{(x - \mu)^2}{2 * \sigma^2}\right) \quad (2.1)$$

where X denotes random variables and x denotes the actual argument. X 's normal distribution is typically represented by $P_X(x) \sim N(\mu, \sigma^2)$. In figure 2.5, the probability density function of a univariate normal(or Gaussian) distribution is plotted. We randomly selected 1000 points from a univariate normal distribution and plotted them on the x-axis in figure 2.5.

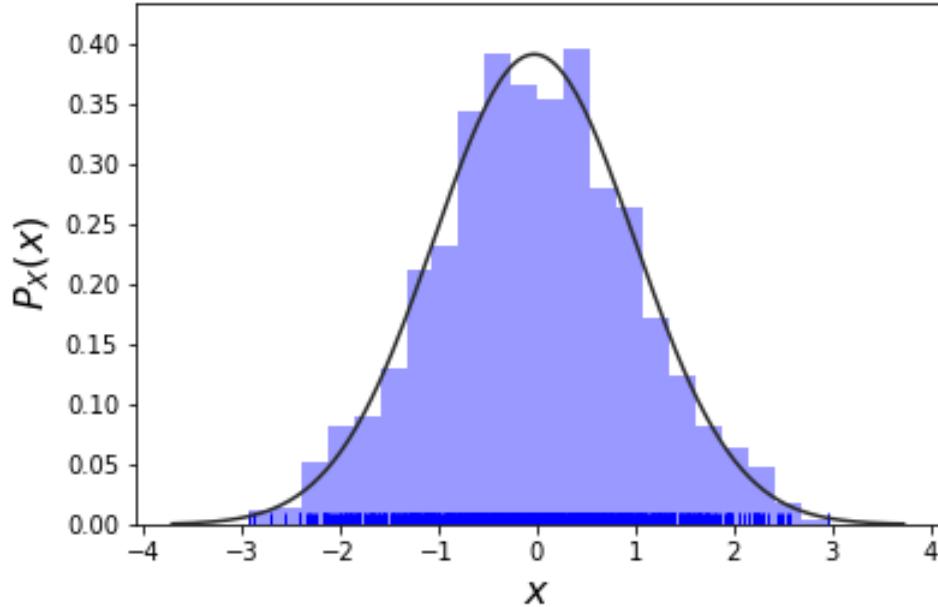
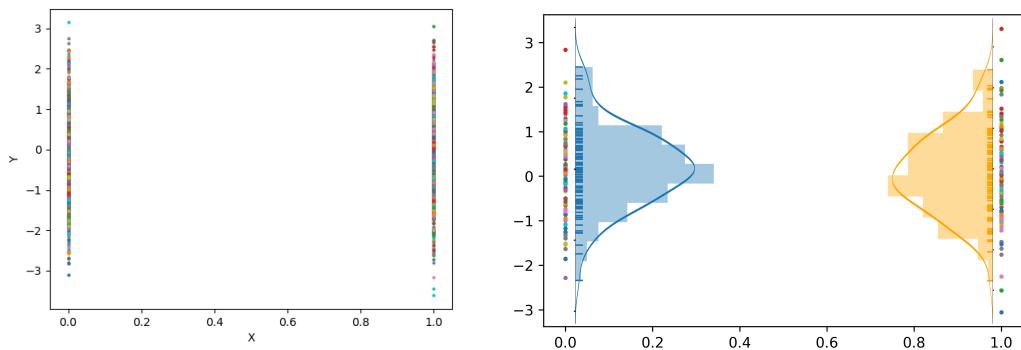


Figure 2.5: Gaussian Distribution

These generated random data points can be represented as a vector $x_1 = [x_1^1, x_1^2, \dots, x_1^n]$. By plotting the vector x_1 on a new Y axis at Y=0, we can see that $[x_1^1, x_1^2, \dots, x_1^n]$ is projected onto another space as shown in left figure 2.6. Another Gaussian vector $x_2 = [x_2^1, x_2^2, \dots, x_2^n]$ is plotted at Y=1 as shown in left figure 2.6. Either x_1 or x_2 is a univariate normal distribution shown in right figure 2.6

Figure 2.6: x_1 and x_2 vectors

Next, we choose 10 points at random from the vectors x_1 and x_2 shown in figure 2.7 and connect them in the order shown by lines. These connected lines appear to be linear functions spanning the $[0,1]$ domain. These lines are still too noisy for us to use in regression tasks to generate predictions. These functions have to be smoother, which necessitates similar output values for input locations that are close together. Because the "functions" produced by joining independent Gaussian vector points are not smooth enough for regression tasks, we require these independent Gaussian correlated to each other as a joint Gaussian distribution. The joint Gaussian distribution is described by the multivariate normal distribution theory.

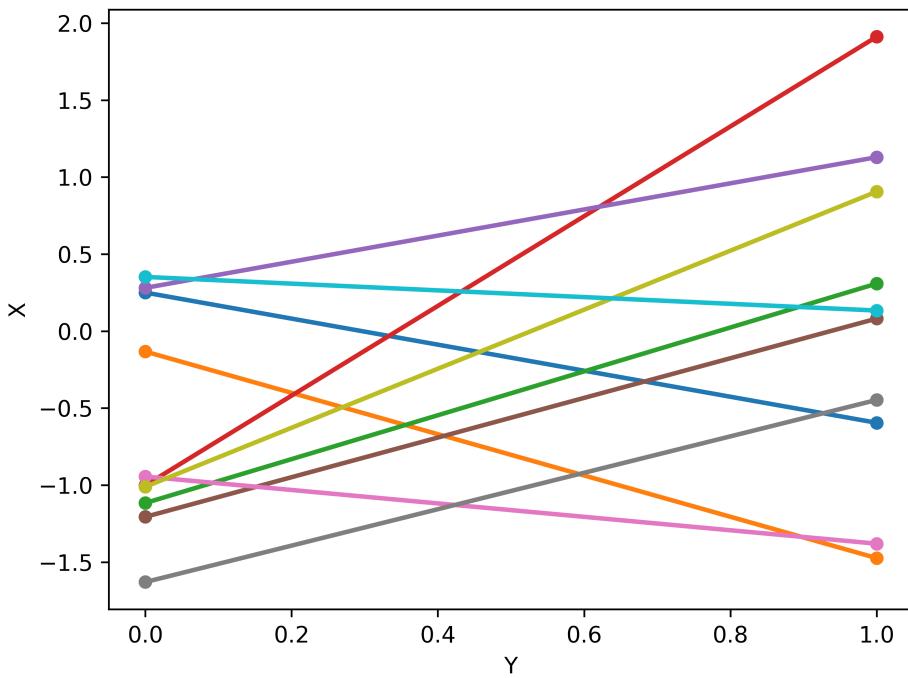


Figure 2.7: lines connecting 10 random points between x_1 and x_2

Multivariate Normal Distribution

It is common for a system to be described by multiple feature variables $[x_1, x_2, \dots, x_D]$ that are correlated between themselves. If we would like to create a model, with all of these variables as a single Gaussian model, we must

use a multivariate Gaussian/normal distribution[22] model. The Probability density function of a multivariate normal distribution with a D dimension can be represented by

$$\mathcal{N}(x | \mu, \Sigma) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right] \quad (2.2)$$

Where D represents no. of dimensions, x represents the variable, μ is the mean vector, and Σ is the $D \times D$ covariance matrix. Σ is a symmetric matrix containing all collectively modeled random variables' nearest neighbor covariance.

Gaussian processes kernels

In order to determine the form of the prior and posterior, kernels are an essential component of Gaussian processes [41]. They capture the assumptions about the function being learned by defining the resemblance over two data points and assuming that comparable data points should have similar target values. Fallsows stationary kernels, which depend solely on the space between two points and not on their absolute values, and non-stationary kernels, which depend on both absolute values and the space between two points, are two different types of kernels. A kernel's primary function is to calculate the covariance between data points[49].

2.4 Active learning

An effective strategy to create accurate models with fewer data is through the use of active learning[48]. To create good models, traditional machine-learning techniques need a lot of labeled data. But acquiring a lot of labeled data can be difficult, expensive, or even time-consuming in many real-world situations. This problem is solved by active learning, which iteratively chooses a small subset of data points for expert or crowd-sourced labeling. The first model is trained on a limited quantity of labeled data, which can be collected either by picking a representative subset or randomly sampling from the available data. A sizable collection of unlabeled data points are then utilized to predict labels using the model. Based on the prediction uncertainty or diversity of the model, the active learning algorithm

chooses a small fraction of the unlabeled data points that are the most uncertain or informative. After being labeled by a professional or a member of the crowd, the chosen data points are subsequently included in the labeled data set. The model is then updated using the newly labeled data points to provide a model that is more precise. Up till a good model is obtained, the active learning process is performed iteratively. The intricacy of the problem and the amount of data available determine how many iterations are necessary. In general, compared to conventional machine learning techniques, active learning takes fewer labeled data points to obtain the same degree of accuracy. This makes it a desirable alternative in situations where labeling data is time-consuming, costly, or challenging.

3

Related Work

In this section, we represent the most recent works in the optimization of chemical systems using surrogate modeling. A more detailed description can be found in [20],[50],[47] and [48]

3.1 Surrogate based optimization

3.1.1 Batch distillation

Because of the differences in their amount and composition, disposal solvent mixes are commonly treated using batch distillation (BD)[20]. An approach based on surrogate models for optimizing batch distillation processes[20] is suggested and utilized for the restoration of methanol from a five-component azeotropic disposal solvent mixture with contaminants removed in 2 fore-cuts and one after-cut. The profit growth of batch distillation processes can be enhanced through optimization, thereby reducing their electrical energy usage and environmental effect . In order to optimize BD processes, there are three different optimization problems: [35] maximum distillate, minimum time, and maximum profit. 0.07 mass% acetone, 37.4% methanol, 4.89% tetrahydrofuran, 56.34% water, and 1.56% toluene are present in the disposal solvent mixture to be considered. methanol must be recovered with 99.5% purity. The profit of a single batch was the objective function (OF), with constraints on the purity of the main cut and the composition of the recycled, 2nd fore-cut. With the help of this work, we have extended our work to extractive distillation processes.

3.1.2 Stirring tank

Stirring tanks seem to be critical components of machinery in commercial processes like adsorption and crystallization for solid-liquid suspension operations, and their performance varies depending on the pulp process, framework, and operating parameters.[50]. However, the conventional modeling approach for stirred tank framework is complex and time-consuming, limiting design efficiency greatly. A surrogate model is employed to deal with the cost and ineffectiveness of the optimal design of the stirring tank. To begin, computational fluid dynamics(CFD) was used to simulate the state of the fluid flow[46] in stirring tanks with varying structural dimensions and operational parameters and validated with Particle Image Velocimetry. The Latin Hypercube technique was then used to create a Kriging model[19] depending on the simulated data for the important parameters such as stirring speed, stirrer diameter, stirrer height from the bottom, and paddle pitch.

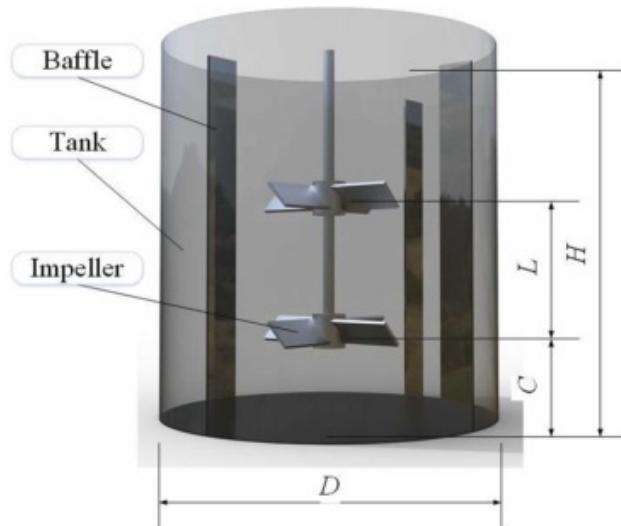


Figure 3.1: Layout of stirring tank[50]

Parameters/Unit	Values
Mineral particle density kg/m ³	1900
Solution density kg/m ³	1200
Mineral volume fraction	15%
Solid particle diameter mm	0.1
Viscosity mPa.s	2

Table 3.1: Process conditions of stirring tank[50]

Parameters/Unit	Values
Diameter of stirring tank D/m	3
Height from the bottom C/m	0.6
Impeller pitch L/m	1.2
Height of stirring tank H/m	3.3
Baffle width W/m	0.4
Diameter of impeller d/m	1

Table 3.2: Key structural parameters of the stirring tank[50]

3.2 Sequential learning

In terms of CO₂ emissions, alkali-activated binders (AAB) can be an eco-friendly alternative to traditional cement. The Sequential learning[47] approach blends machine-learning algorithms and real-world experiment feedback. The underlying concept of Sequential learning is that not every one of the experiments is considered equal. At the start of the process, the first set of training data having known target properties is fed into the prediction model. The selection of candidates decides the performance of sequential learning. The workflow of sequential learning is shown in the figure below.

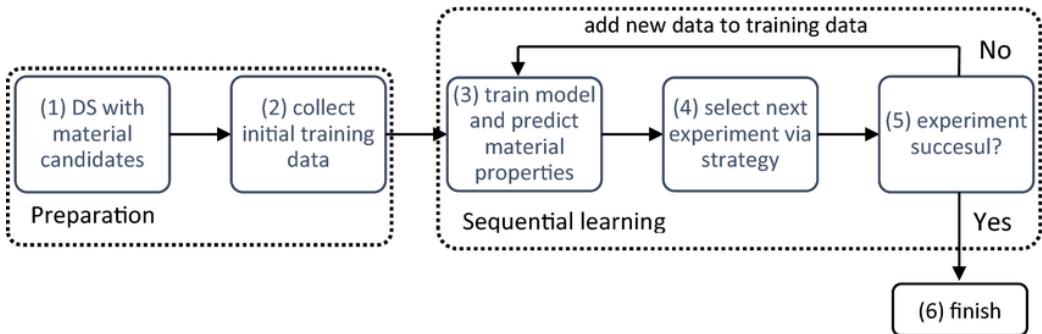


Figure 3.2: Workflow of sequential learning[47]

3.2.1 Strategies and Utility

Maximum Expected Improvement(MEI)

The MEI strategy [32] is purely exploitative, simply choosing the next candidate on the basis of the highest prediction value. The $i - th$ prediction utility is:

$$u_{MEI,i} = \mu_i \quad (3.1)$$

μ_i is the mean prediction

Maximum Likelihood of Improvement(MLI)

The MLI strategy [32] is a strategy of exploration and exploitation. It chooses the candidate who is most likely to display the particular target property. In the case of standard normal distribution prediction, equation 3.2 can be used to evaluate the candidate with the greatest 95 percent likelihood.

$$u_{MLI,i} = Q(y_i, 0.95) = \mu_i + 1.93 * \sigma_i \quad (3.2)$$

$Q(95\%) = 95\%$ quantile, μ_i = meanprediction, σ_i = Standarddeviation

MEI and MLI with maximum Euclidean distance

$$u_{MEI+D,i} = \text{mean dist } (x, u_{MEI,Q(90)}) = \frac{1}{n} * \sum_{j=1}^n |x_j - x_{Q(u_{MEI},0.9),i}|$$

$$u_{MLI+D,i} = \text{mean dist } (x, u_{MLI,Q(90)}) = \frac{1}{n} * \sum_{j=1}^n |x_j - x_{Q(u_{MLI},0.9),i}| \quad [47] \quad (3.3)$$

mean dist = mean Euclidian distance

x_j = jth coordinates of known learned data,

$x_{Q(u_{MEI},0.9),i}, x_{Q(u_{MLI},0.9),i}$

are the coordinates of the i th candidate > 90% of MEI or MLI utility

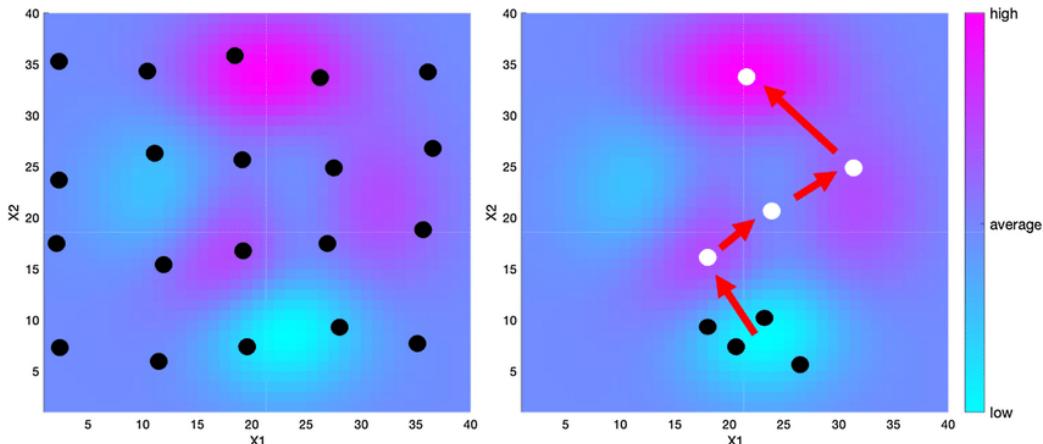


Figure 3.3: Left: As in classic Machine Learning, identically distributed selection (black dots) is desired; Right: Initial data (black dots) sampling with sequential learning and sequential sampling (red arrows) of new samples (white dots). [47]

4

Evaluation and Experiments

4.1 Optimization

4.1.1 Basics of optimization

Finding the optimal solution to a problem within a set of limitations is a key component of optimization [24] in chemical engineering. Profit maximization, cost reduction, increased product quality, and waste reduction are the four main objectives of optimization. Chemical synthesis, separation, purification, and reaction engineering are just a few of the processes that are subject to optimization in this field. Many factors, including reaction rates, temperature, pressure, flow rates, and concentrations, can be optimized using various strategies. Chemical engineering employs a number of optimization methods, including simulated annealing, neural networks, evolutionary algorithms, genetic algorithms [15], and mathematical programming. The most widely used optimization method in chemical engineering is known as mathematical programming, which makes use of linear programming, non-linear programming, mixed-integer programming, and dynamic programming. While solving an optimization problem in chemical engineering, the right optimization technique must be used after modeling the system or process, identifying the decision factors, setting up the objective function and constraints, and resolving any conflicts. The process can then be designed or improved, its efficiency increased, and its costs decreased using the resulting optimized solution.

4.1.2 General problem statement

Standard Optimization general form:

Minimize objective function $f(X)$

subjected to

n number of inequality constraints $g_j(X) \leq 0, j = 1, 2, 3, \dots, n$

m number of equality constraints $h_j(X) = 0, j = 1, 2, 3, \dots, m$

where the no of design variables is given by $x_j, j = 1, 2, 3, \dots, s$

or by design vector

$$X = \begin{Bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{Bmatrix}$$

- This standard problem statement may also be used to find a solution where the objective function is to be maximized (rather than minimized) because maximization of a function $f(X)$ is the same as minimizing the negative of $f(X)$
- Similarly, we can handle the \geq type of inequality constraints by reversing the sign of the constraint function, which transforms them into the \leq type of inequality constraints.
- There may be simple limits on the allowable range of values for a design variable, which are known as side constraints: $x_i' \leq x_i \leq x_i''$
- Where x_i' and x_i'' denote the lower and upper bounds of x_i . However, these side constraints can be simply turned into regular inequality constraints.

4.1.3 Global and Local minimum

Let X^* represent the collection of design parameters that yield a minimum [42] of the objective function $f(X)$ (the * represents quantities and terms referring to an optimum point). At point X , an objective $f(X)$ is at its global (or absolute) minimum[28]if and only if:

$$f(X^*) \leq f(X) \text{ for all } X \text{ in the feasible region}$$

The objective has a local(or relative) minimum [28] at point X^* if and only if:

$$f(X^*) \leq f(X) \text{ for all } X \text{ within a small neighbourhood of } X^*$$

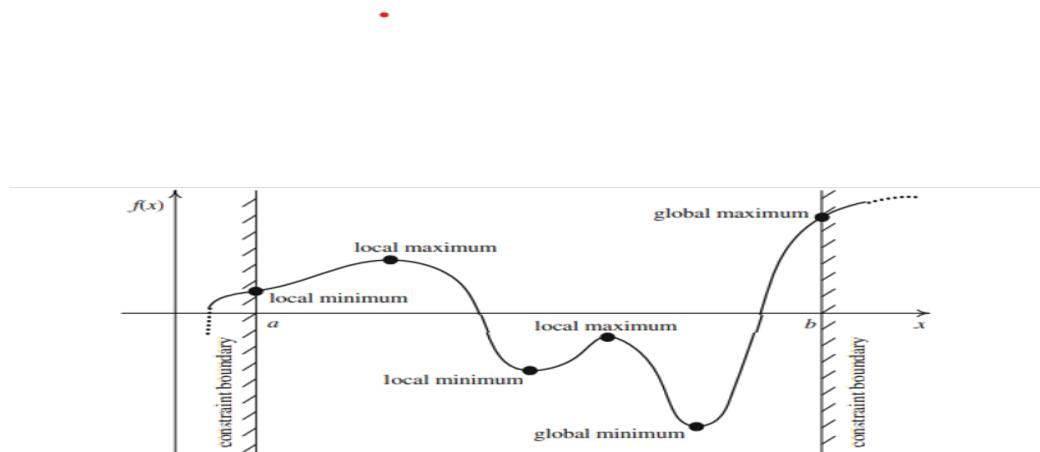


Figure 4.1: Local and Global minima

The above figure depicts a graphical representation of these concepts for the case of a single variable x over a closed feasible region $a \leq x \leq b$.

4.2 Bayesian Optimization

Consider a relation $f : X \rightarrow R$ that we wish to reduce on a field X [33] $\subseteq X$. That implies, we want to find out:

$$x^* = \arg\min f(x), x \in X$$

The above function is known as global optimization in numerical analysis and this research is under study for decades and continues. We differentiate by considering overall optimization, which involves discovering the overall optimum in X , and which involves finding a local optimum near a given starting location x_0 is known as local optimization. Making assumptions about f is a common approach to optimization problems, if the optimization function f is known to have a convex form and the domain X on which we optimize is also convex, then the optimization is referred to as convex optimization and has attracted a great deal of attention. Convex optimization is a machine learning tool that is commonly utilized. If a suitable form of math for f is not predictable, (f is considered a "black box") bayesian optimization works by holding a probabilistic belief about f and in addition to the acquisition function determining the evaluation of the function next. Bayesian optimization [37] is especially the best match for global optimization tasks in which f is a costly black-box relation. How do we decide where to evaluate the function next given a set of initial data? The meta-strategy in bayesian optimization is to establish an acquisition function $a(x)$ that is generally inexpensive to evaluate at a particular point. The value of this function is directly proportional to how beneficial it is expected to be to evaluate the objective function, f , at that point x in order to solve the minimization problem. In other words, the acquisition function helps to determine which point to choose next to evaluate the objective function in order to find the minimum value more efficiently. The low-cost function is then optimized to determine the next point to evaluate. we have just mimicked our actual optimization task with another optimization task, but this time on a considerably less expensive function $a(x)$.

4.2.1 Acquisition functions

Acquisition functions [39][47] can be explained as calculating an estimated loss related to testing f at a location x in the context of Bayesian decision theory. As usual, we select the spot that has the smallest projected loss.

Probability of improvement

The Probability of improvement acquisition function measures the probability that a new sample point will have a better function value than the current best sample point found so far. In other words, it quantifies the probability of improving upon the best solution found so far. Suppose

$$f' = \min f$$

is the smallest measure of f discovered until now. The Probability of improvement assesses f at a location where it is most widely believed to get better on current value. This is equivalent to the utility function related to calculating f at a specific point x :

$$u(x) = \begin{cases} 0, & f(x) > f' \\ 1, & f(x) \leq f' \end{cases}$$

i.e., if $f(x)$ is less than f' , then a unit is rewarded, and no reward otherwise. The utility expected as an outcome of x is considered as the probability of improvement acquisition function:

$$\begin{aligned} a_{PI}(x) &= \mathbb{E}[u(x) | x, \mathcal{D}] = \int_{-\infty}^{f'} \mathcal{N}'(f; \mu(x), K(x, x)) df \\ &= \Phi(f'; \mu(x), K(x, x)). \end{aligned} \tag{4.1}$$

The location with the greatest chance of enhancement (the maximum expected utility) is chosen.

Expected Improvement

The acquisition function related to the probability of improvement is a little strange. Regardless of how much the improvement takes place based on the present minimum we are being rewarded. This can at times result in strange behavior, and in the real case, there is a chance of getting stuck in local optima and not exploring globally. Expected improvement is another acquisition function that considers how much the improvement is. Assume that f' is the smallest value of f found out thus far. Expected improvement calculates f at the location where it is expected to improve f' the largest. The below function replicates that:

$$u(x) = \max(0, f' - f(x))$$

if $f(x)$ comes out to be smaller than f' then "Improvement" is equal $f' - f(x)$ is awarded and no award if it turns to another case. The expected utility as a function of x is then considered as an expected improvement acquisition function:

$$\begin{aligned} a_{EI}(x) &= \mathbb{E}[u(x) | x, \mathcal{D}] = \int_{-\infty}^{f'} (f' - f) \mathcal{N}(f; \mu(x), K(x, x)) df \\ &= (f' - \mu(x)) \Phi(f'; \mu(x), K(x, x)) + K(x, x) \mathcal{N}(f'; \mu(x), K(x, x)) \end{aligned} \quad (4.2)$$

The area with the greatest expected improvement/the maximal expected utility is chosen. There are two components of expected improvement. Decreasing the mean $\mu(x)$ can be used to increase the first component. Increasing the variance function $K(x, x)$ can be used to increase the second component. The above two factors can be read as clearly containing a trade-off between exploitation and exploration (evaluating at low-mean points) (evaluating at locations having high uncertainty). The exploitation-exploration [10] trade-off is common attention in such scenarios, and as a result of the bayesian decision-theoretic treatment, the expected improvement function automatically captures both.

Entropy

The idea behind the entropy acquisition function is to select a sample point that maximizes the amount of information gained about the objective func-

tion. This is achieved by selecting a point that has high uncertainty (i.e., high entropy) in its prediction of the objective function. In other words, the entropy acquisition function aims to balance exploration (trying new points to gain information) and exploitation (focusing on points that are likely to be optimal).

Mathematically, the entropy acquisition function can be expressed as follows:

$$x^* = \arg \min_{x \in X} f(x)$$

The purpose of entropy search is to estimate locations in order to reduce the entropy associated with the resulting probability $p(x^* | D)$. In this case, the utility function is the decrease in entropy as a consequence of a novel evaluation at $x, (x, f(x))$:

$$u(x) = H[x^* | \mathcal{D}] - H[x^* | \mathcal{D}, x, f(x)] \quad (4.3)$$

An acquisition function is created by calculating the expected utility provided by f at a given location x , exactly as we did with the likelihood of improvement and expected improvement. This is somewhat difficult because of the complications in terms of a probability distribution of the $p(x^* | D)$, as well as several instances of estimates that must be calculated.

Upper Confidence bound

An additional acquisition function is commonly referred to as UCB, with 'UCB' standing for the upper confidence bound. UCB is largely characterized by maximizing f rather than minimizing f . However, for a minimizing problem, the acquisition function can be written as:

$$a_{\text{UCB}}(x; \beta) = \mu(x) - \beta\sigma(x) \quad (4.4)$$

where $\beta > 0$ is trade-off parameter and $\sigma(x) = \sqrt{k(x, x)}$ is the marginal deviation of $f(x)$. Again, the Upper confidence bound includes explicit terms

for exploitation $\mu(x)$ and exploration $\sigma(x)$. Despite this, powerful theoretical results for UCB are known, specifically, the repetitive utilization of this acquisition function will result in the true global minimum of f within certain conditions.

4.3 Example

let's consider the problem of optimizing the following function:

$$Y = (X * 6 - 2)^2 . np.sin(X * 12 - 4)$$

where X is a scalar variable in the range $[0, 1]$

This function is computationally expensive to evaluate, as it involves evaluating a sine function. Therefore, we will use surrogate-based optimization with active learning to find the minimum of this function.

- First, we start by generating a small set of initial samples, say $X = [0, 0.1, 0.2, 0.4, 0.6, 1]$.
- We evaluate the function at these points and use them to train a Gaussian process, a surrogate model.
- Next, we use the surrogate model to select the next point to evaluate using the acquisition function.

One commonly used acquisition function is the expected improvement (EI)[12] function:

$$EI(x) = \max(0, \mu(x) - f_{best} - xi)$$

where $\mu(x)$ is the mean prediction of the surrogate model at x , f_{best} is the best function value observed so far, and xi is a parameter that controls the balance between exploration and exploitation. In this example, we set $xi = 0.01$.

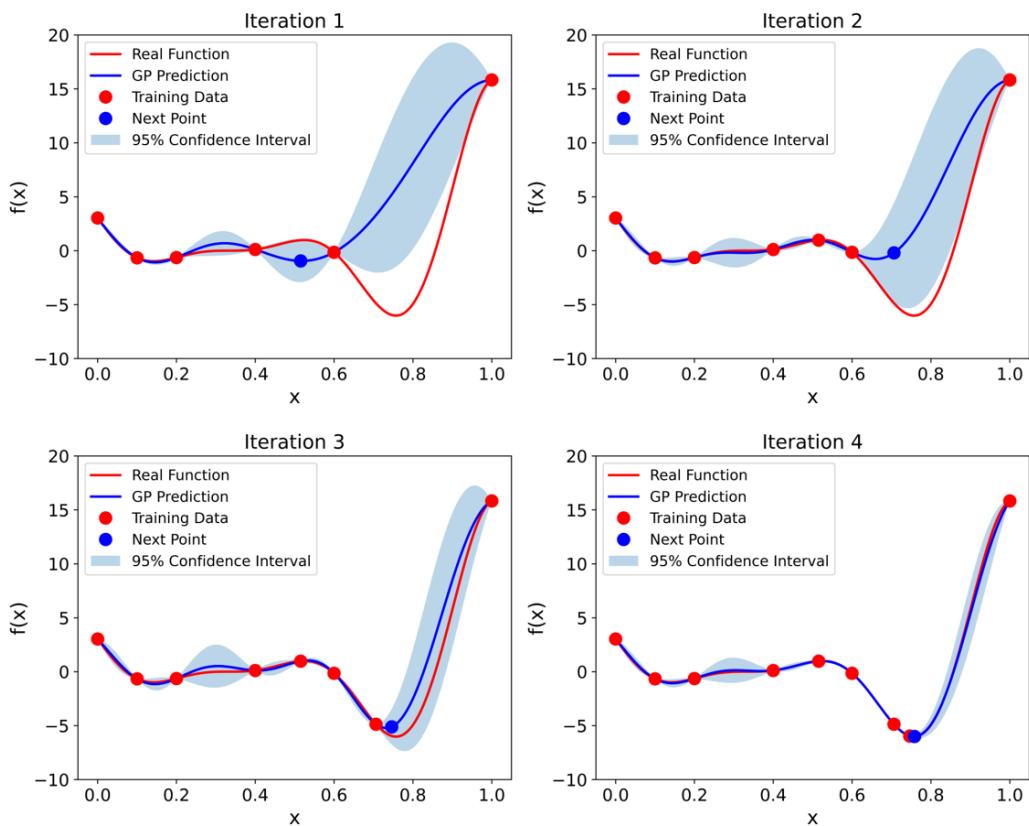


Figure 4.2: Optimization of $Y = (X*6-2)^2 \cdot np.sin(X*12-4)$ using active learning

4.4 Problem statement

Styrene butadiene rubber, polybutadiene rubber, and styrene butadiene latex are only a few examples of rubber and plastics made from 1,3-butadiene, a significant petrochemical product. It can be found in C_4 mixes, which are leftovers from the production of naphtha crackers. 1,3-butadiene must be isolated from C_4 mixtures before use in polymerization operations[3]. C_4 mixes typically consist of butanes, butenes, 1,2-butadiene, 1,3-butadiene and acetylenes among other substances. Because several of the C_4 components have relative volatilities that are very comparable to those of 1,3-butadiene, the C_4 mixtures cannot be segregated into each component using traditional distillation. The number of trays or the reflux ratio must be drastically raised in a typical distillation operation to separate hydrocarbons with identical boiling points. Because the other C_4 components are less soluble than 1,3-butadiene in some solvents, such as acetonitrile (ACN), dimethylformamide (DMF), and N-methyl-2-pyrrolidone, 1,3-butadiene can be separated by extractive distillation (NMP). For this study, only a part of the separation is considered in which C_4H_8 (1 -Butene) and C_4H_6 (1,3-Butadiene) as a separation mixture and NMP (N-methyl-2-pyrrolidone) as a solvent is considered. The process involves the following steps:

- **Getting the mixture ready:** By cracking hydrocarbons, such as naphtha, in the presence of a catalyst, 1-butene, and 1,3-butadiene are created.
 - **Adding NMP:** As a solvent, NMP is added to the mixture. The ratio of butene to butadiene in the mixture determines how much NMP should be added.
 - **Distillation:** In order to separate the components based on their respective boiling points, the mixture is then heated in a distillation column. The butene and butadiene components evaporate as the temperature rises, rising up the column. By shifting the vapor-liquid balance with the addition of NMP, butadiene is more preferentially concentrated in the liquid phase, making it easier to separate from 1-butene.
-

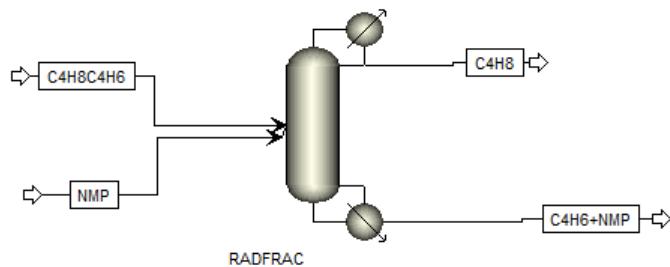


Figure 4.3: Flow sheet Extractive Distillation

- **Separation:** The butadiene and NMP-containing liquid phase is then taken out of the column and transferred to a second distillation column, where the butadiene and NMP are separated. While the NMP can be recovered and recycled for additional use, the butadiene can be separated out as a pure product.

Overall, the extractive distillation process with NMP as a solvent provides a selective and efficient method for separating 1-butene and 1,3-butadiene. The addition of NMP as a solvent can improve the separation efficiency and yield a higher-purity product.

No. of stages	[40,80]
Reflux ratio	[1,10]
Operating pressure	[3.5,6]
Solvent-to-feed ratio	[1,8]

- **Input variables-** [No. of stages, Operating pressure, reflux ratio, Solvent to feed ratio]
- **Output variables-** [reboiler heat duty , product purity]
- 16 Initial data points from design space.
- Gaussian process regression(surrogate model)

- **Acquisition function:** Maximum probability of improvement and Upper Confidence bound.
- **Low-quality data:** Sampling of initial training data without considering purity constrain.
- **High-quality data:** Sampling of initial training data which has product purity ≥ 0.995 as a constrain.

Importance of Considering minimization of reboiler heat duty:

Minimizing the reboiler heat duty in extractive distillation is important because it can lead to several benefits, including:

- **Improved solvent recovery:** The solvent used in extractive distillation is often expensive and needs to be recovered and reused to minimize the operating costs [31] of the process. By minimizing the reboiler heat duty, the amount of solvent that is lost to vaporization can be reduced, leading to improved solvent recovery and lower operating costs.
- **Enhanced separation efficiency:** The amount of heat supplied to the reboiler affects the rate of vaporization and the separation efficiency of the process. By minimizing the reboiler heat duty, the rate of vaporization can be reduced, allowing for more time for the separation process to occur, which can improve the overall separation efficiency.
- **Lower operating costs:** The amount of energy required to supply the reboiler heat duty is a significant operating cost for extractive distillation processes[13]. By minimizing the reboiler heat duty, the energy consumption and operating costs of the process can be reduced.
- **Improved product quality:** The reboiler heat duty can affect the composition and purity of the product streams produced by extractive distillation. By minimizing the reboiler heat duty, the quality and purity of the product streams can be improved, leading to higher yields and better product quality.

4.5 Results

Figures 4.4, 4.5 represents the variation of Heat duty and Product purity with respect to iterations where heat duty and product purity are plotted on the y-axis and Iteration is plotted on the x-axis respectively for low-quality data.

4.5.1 Minimization of reboiler heat duty with low-quality data

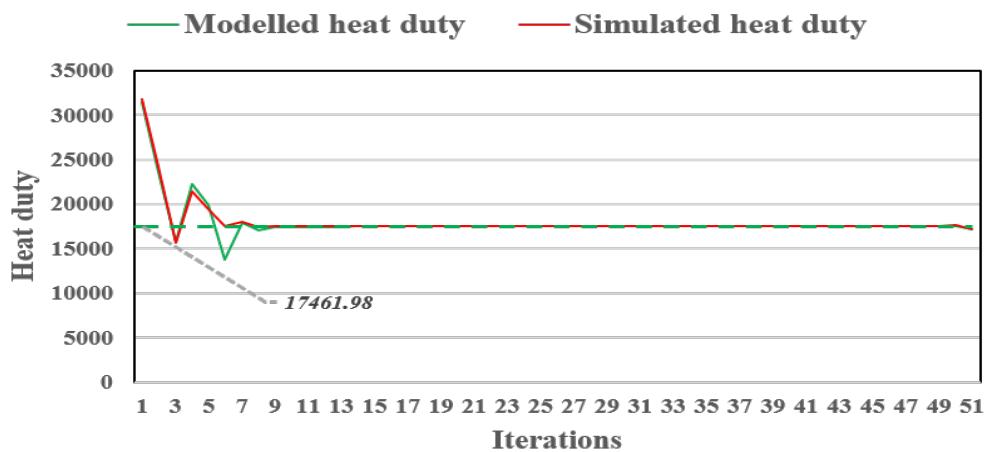


Figure 4.4: low quality - Heat duty v/s Iterations

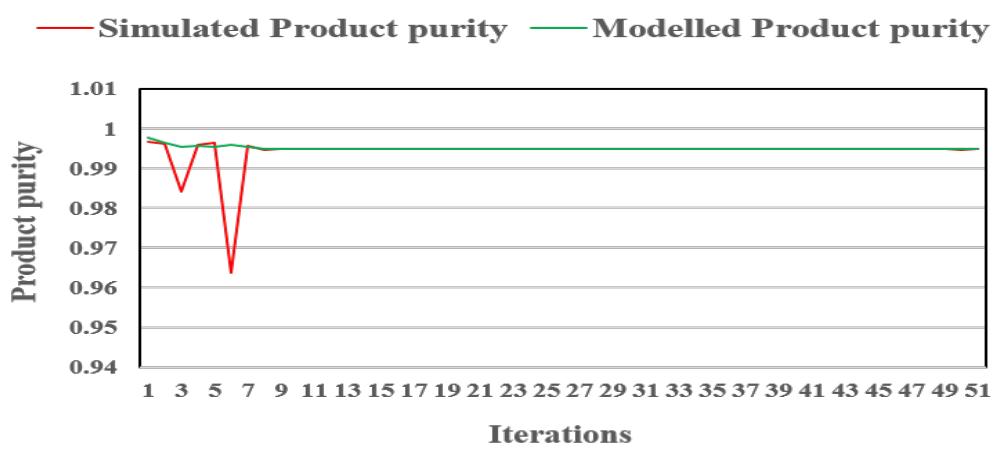


Figure 4.5: low quality - Product purity v/s Iterations

Figure 4.6 represents the variation of the input variable No. of stages on the y-axis and Iterations on the x-axis and figure 4.7 represents the variation of the input variable Operating pressure on the y-axis and Iterations on the x-axis.

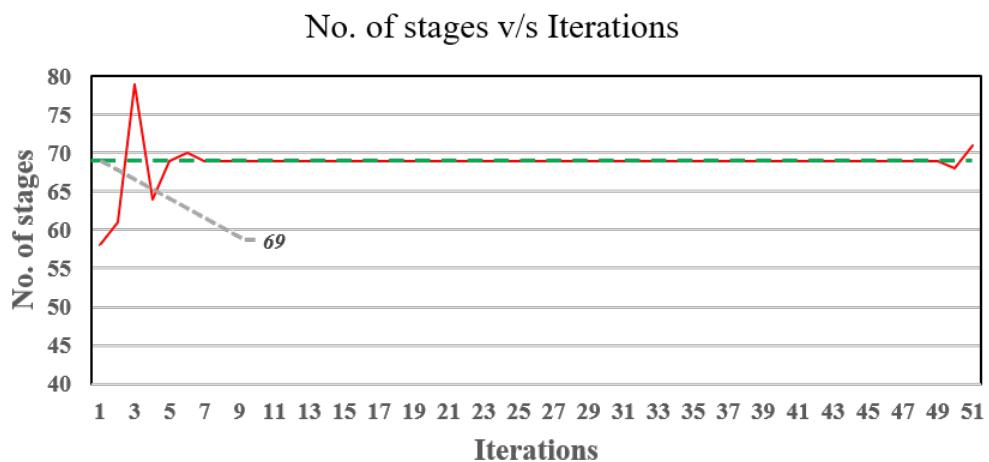


Figure 4.6: low quality - No. of stages v/s Iterations

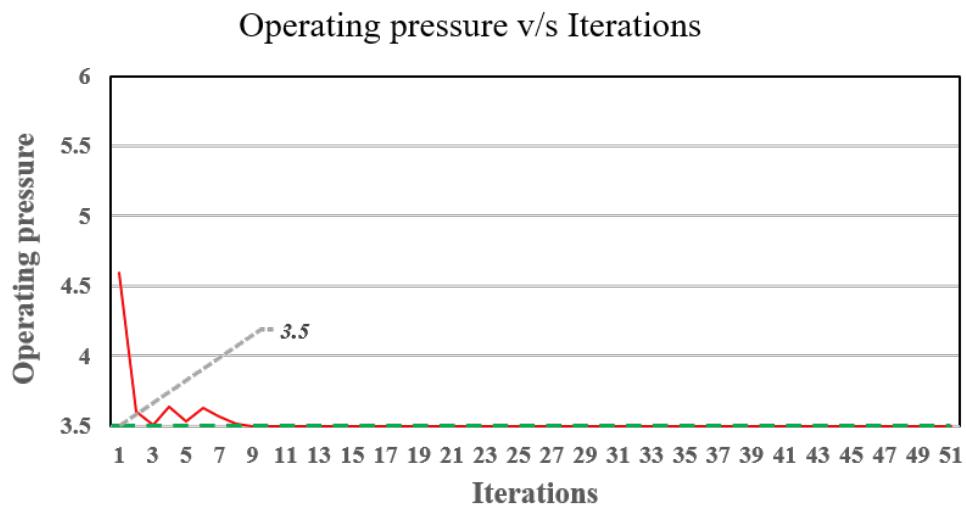


Figure 4.7: low quality - Operating pressure v/s Iterations

Figure 4.8 represents the variation of the input variable reflux ratio on the y-axis and Iterations on the x-axis and figure 4.9 represents the variation of the input variable solvent to feed ratio on y-axis and Iterations on the x-axis.

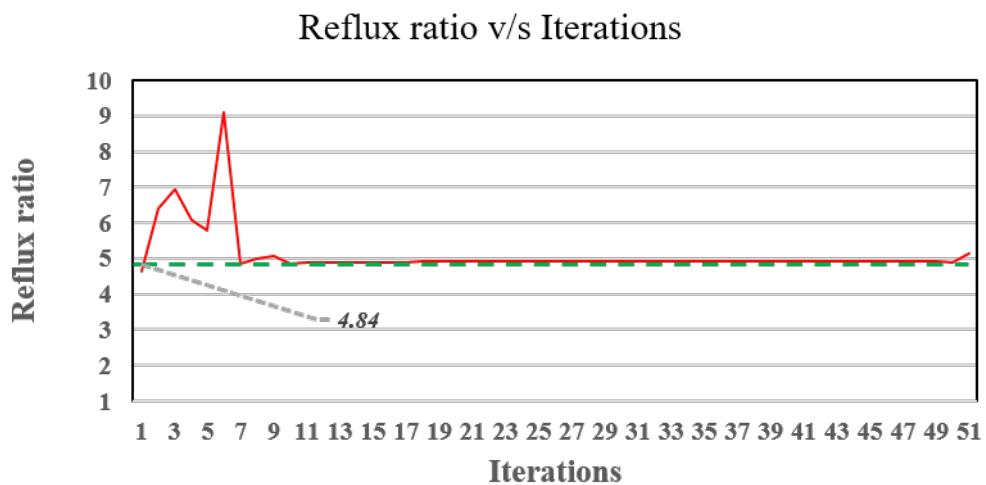


Figure 4.8: low quality - Reflux ratio v/s Iterations

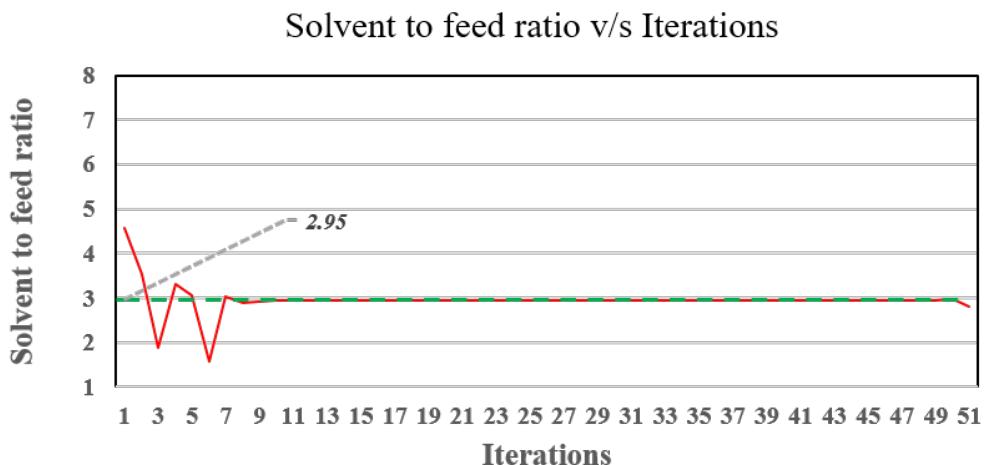


Figure 4.9: low quality - Solvent to feed ratio v/s Iterations

Convergence for Low-quality data occurs at the input variables:	No. of stages: 69 Reflux ratio: 4.84 Operating pressure: 3.50 Solvent-to-feed ratio: 2.95
Iterations required for high-quality data:	10
Heat duty for high-quality data:	17461.98 (KW)
The sampling time for high-quality data:	17.32 seconds

Table 4.1: Low-quality data convergence results

After analyzing the graphs 4.4, 4.5, 4.6, 4.7, 4.8, 4.9 we can see that after certain Iterations there is no change or minor change in the target variable that is considered. The input variables at which there is convergence or minor changes occur are tabulated in 4.1.

4.5.2 Minimization of reboiler heat duty with high-quality data

Figures 4.10,4.11 represent the variation of Heat duty and Product purity with respect to iterations where heat duty and product purity are plotted on the y-axis and Iterations is plotted on the x-axis respectively for high-quality data.

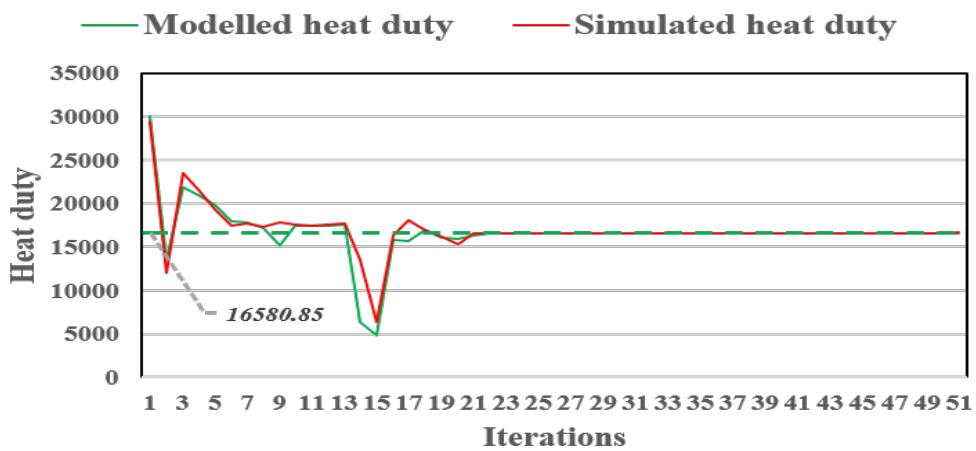


Figure 4.10: high quality - Heat duty v/s Iterations

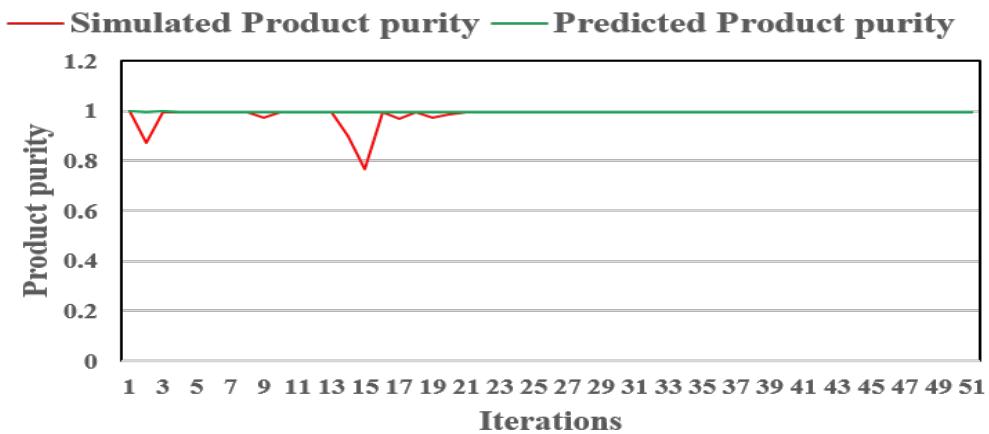


Figure 4.11: high quality - Product purity v/s Iterations

Figure 4.12 represents the variation of the input variable No. of stages on the y-axis and Iterations on the x-axis and figure 4.13 represents the variation of the input variable Operating pressure on the y-axis and Iterations on the x-axis.

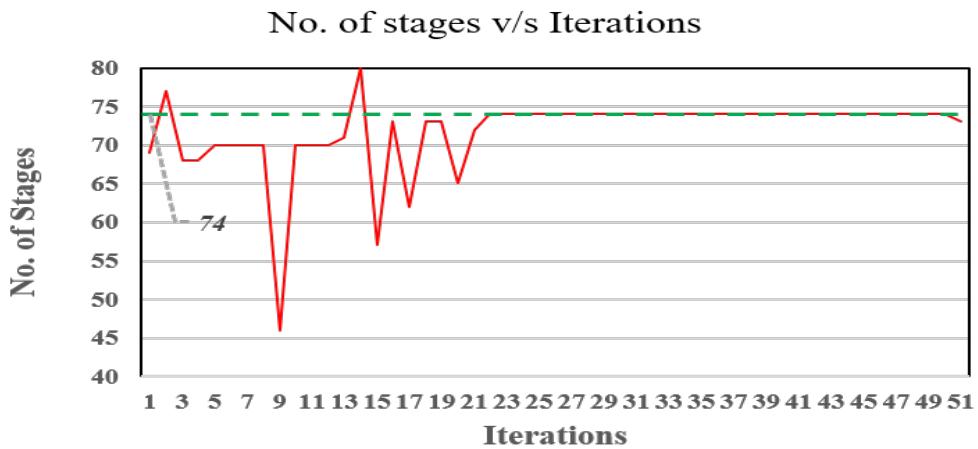


Figure 4.12: high quality - No. of stages v/s Iterations

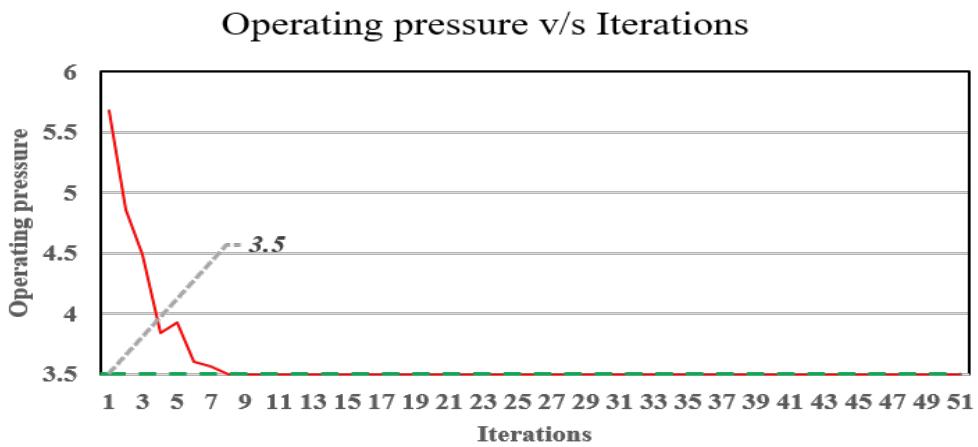


Figure 4.13: high quality - Operating pressure v/s Iterations

Figure 4.14 represents the variation of the input variable reflux ratio on the y-axis and Iterations on the x-axis and figure 4.15 represents the variation of the input variable Solvent to feed ratio on the y-axis and Iterations on the x-axis.

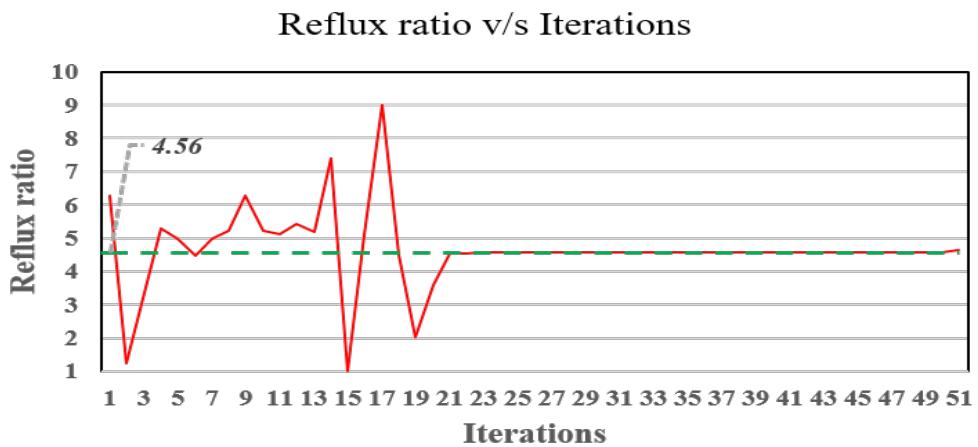


Figure 4.14: high quality - Reflux ratio v/s Iterations

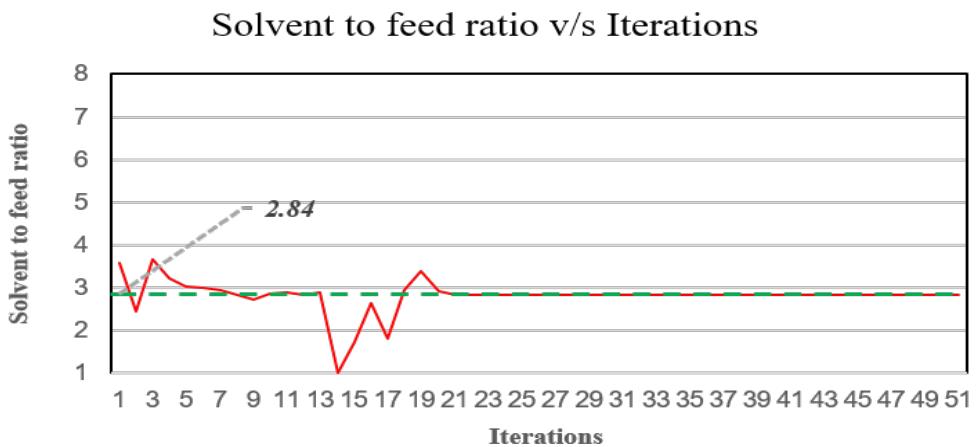


Figure 4.15: high quality - Solvent to feed ratio v/s Iterations

Convergence for high-quality data occurs at the input variables:	No. of stages: 74 Reflux ratio: 4.56 Operating pressure: 3.50 Solvent-to-feed ratio: 2.84
Iterations required for high-quality data:	25
Heat duty for high-quality data:	16580.85 (KW)
The sampling time for high-quality data:	43.40 seconds

Table 4.2: High-quality data convergence results

After analyzing 4.10, 4.11, 4.12, 4.13, 4.14, 4.15 we can see that after certain Iterations there are no change or minor change in the target variable that is considered. The input variables at which there is convergence or minor changes occur are tabulated in 4.2.

Figure 4.16, 4.17 represents the error between simulation result and model prediction result under different iterations for low-quality and high-quality data.

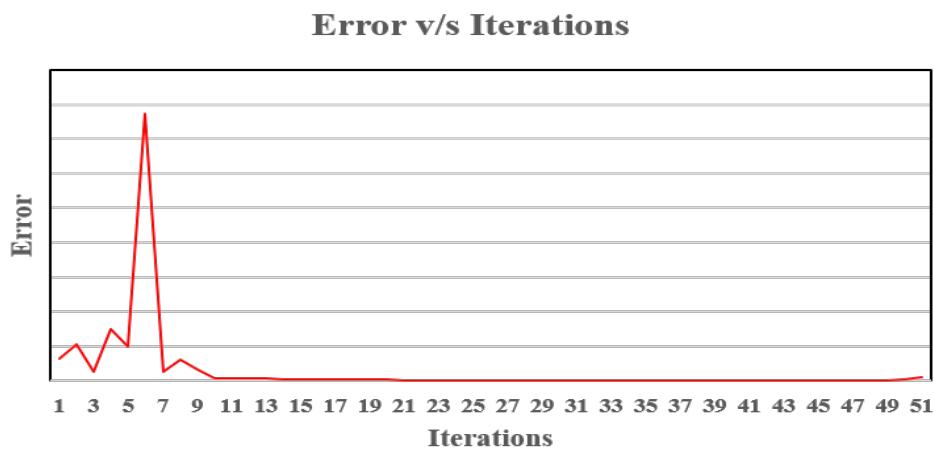


Figure 4.16: Low-quality v/s Iterations

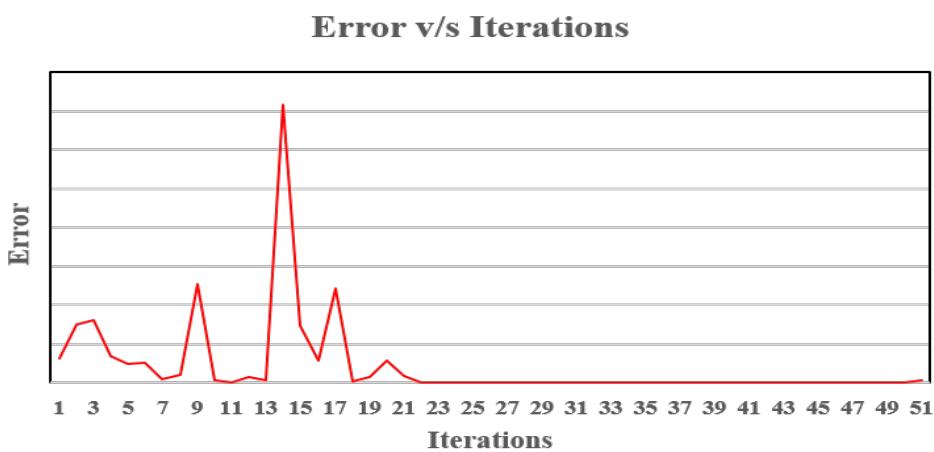


Figure 4.17: High-quality v/s Iterations

4.5.3 Comparison

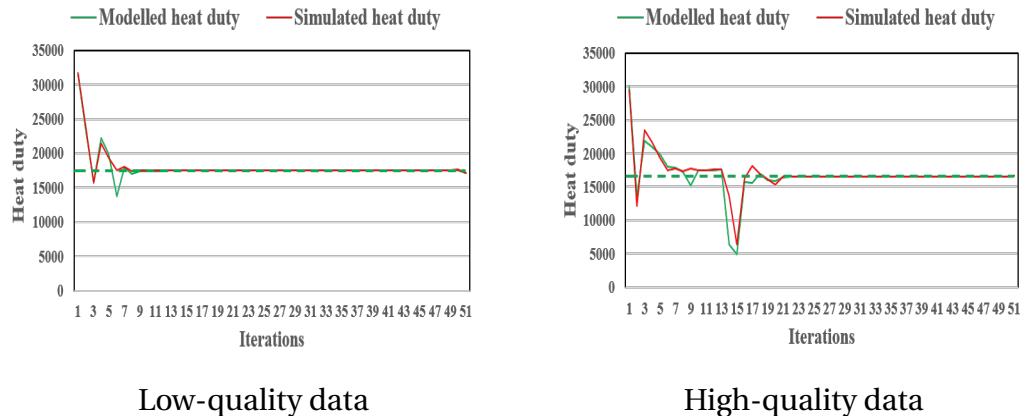


Figure 4.18: Comparison of Low-quality data v/s High- quality data

Case	Iteration	Heat duty(KW)	Product purity
Low-quality data	10	17461.98	0.95
High-quality data	25	16580.85	0.95

Table 4.3: Comparison results

With high-quality data, active learning needs more iterations but achieves better process performance

4.6 Discussion

4.6.1 Case-1(Low-quality data)

Iterations	No of stages	Reflux ratio	Operating pressure	Solvent to feed ration	Modelled heat duty	Simulated heat duty	Simulated Product purity	Modelled Product purity
1	58	4.632	4.595	4.558	31422.66	31738.44	0.996756	0.997701
2	61	6.402	3.598	3.543	23677.51	24201.13	0.99623	0.99646
3	79	6.938	3.506	1.878	15797.94	15659.57	0.984222	0.995346
4	64	6.084	3.634	3.294	22183.56	21431.48	0.995954	0.99566
5	69	5.806	3.531	3.045	19844.52	19349.06	0.996447	0.995339
6	70	9.104	3.627	1.577	13692.6	17561.73	0.963681	0.995888
7	69	4.838	3.564	3.036	17905.42	18035.52	0.995526	0.995288
8	69	5.000	3.514	2.889	17032.3	17344.52	0.994638	0.99501
9	69	5.079	3.502	2.911	17382.95	17542.17	0.99496	0.995001
10	69	4.848	3.500	2.957	17416.93	17461.98	0.994979	0.995

Table 4.4: Minimization of reboiler heat duty with low-quality - results

In the above-mentioned table, there are the first 10 iterations with different configurations of input variables and corresponding output variables with low-quality data.

The main aim of using the Maximum probability of improvement and Upper confidence bound is to balance between exploration and exploitation [10] where in Maximum probability of improvement is biased towards exploration and is good for non-linear or noisy functions while UCB is biased towards exploitation and is good for smooth or well-defined search spaces. In this case, I used low-quality data where there is no constraint on product purity in the initial training data. This process is continued until the required convergence is found in our target variable i.e., reboiler heat duty. Both model prediction and simulation output are being considered and tabulated.

4.6.2 Case-2(High-quality data)

Iterations	No of stages	Reflux ratio	Operating pressure	Solvent to feed ration	Modelled heat duty	Simulated heat duty	Simulated Product purity	Modelled Product purity
1	69	6.268	5.679	3.58	30076.26	29441.98	0.996705	0.997267
2	77	1.234	4.857	2.438	13598.01	12090.21	0.873879	0.995542
3	68	3.338	4.491	3.665	21844.16	23452.97	0.996079	0.998026
4	68	5.306	3.843	3.230	20920.82	21614.29	0.996363	0.995905
5	70	4.983	3.924	3.037	19855.37	19361.91	0.995536	0.995571
6	70	4.463	3.607	3.002	17957.4	17436.12	0.994705	0.995145
7	70	4.983	3.561	2.950	17848.08	17765.67	0.995244	0.995247
8	70	5.211	3.503	2.842	17187.11	17387.4	0.99473	0.995004
9	46	6.285	3.599	2.726	15231.35	17784.73	0.972467	0.995
10	70	5.232	3.500	2.866	17460.22	17536.04	0.994974	0.995003

Table 4.5: Minimization of reboiler heat duty with High-quality - results

In the above-mentioned table, there are the first 10 iterations with different configurations of input variables and corresponding output variables with High-quality data. In this case, I used High-quality data where there is a constraint on product purity in the initial training data. This process is continued until the required convergence is found in our target variable i.e., reboiler heat duty. Both model prediction and simulation output are being considered and tabulated.

4.6.3 Comparison

From the above Table 4.3 on page 62 we can see that the no. of iterations required for convergence in reboiler heat duty in high-quality data is more when compared to low-quality data. When we compare reboiler heat duty the value is minimum in high-quality data than in low-quality data. The computational efficiency required in high-quality data is more than in low-quality data but reboiler heat duty is low which is the fetching point for the analysis.

5

Conclusions and Future Work

5.1 Conclusions

This study explored the use of surrogate-based optimization and active learning to optimize the design and operation of an extractive distillation process. By using a surrogate model to approximate the computationally expensive process model, we were able to significantly reduce the computational cost of the optimization process, while still achieving high-quality solutions.

Furthermore, by incorporating active learning into the optimization process, we were able to further improve the performance of the surrogate model, and obtain better solutions with fewer function evaluations.

The results of this study demonstrate the effectiveness of surrogate-based optimization and active learning for extractive distillation process optimization. The approach presented in this work can be extended to other chemical processes and can be used to solve other optimization problems where the underlying process models are computationally expensive.

Overall, this study provides a framework for efficient optimization of extractive distillation processes, which can lead to significant improvements in process performance, energy efficiency, and cost savings. The application of surrogate-based optimization and active learning to process optimization is a promising area of research, and further studies in this area can lead to even more efficient and effective optimization algorithms for a wide range of chemical processes.

5.2 Future Work

The future of surrogate-based optimization in extractive distillation using active learning is promising. As computational resources become more powerful, it will be possible to optimize more complex systems with greater accuracy. Additionally, advances in machine learning and artificial intelligence will allow for the development of more sophisticated surrogate models, capable of capturing complex interactions between variables.

However, there are also challenges to be addressed. One key challenge is the development of efficient active learning(AL) algorithms that can handle the high-dimensional parameter spaces typically encountered in extractive distillation. Another challenge is the need for robust optimization methods that can handle noisy or incomplete data.

Overall, the future of Surrogate based optimization in extractive distillation using active learning(AL) looks bright, with the potential to significantly reduce the computational cost of process optimization and improve process efficiency.

A

List of Figures

1.1	Different applications of Optimization	17
2.1	Mathematical model of distillation	19
2.2	Surrogate model v/s Simulation model	23
2.3	Process Workflow	25
2.4	'N' no. of functions	30
2.5	Gaussian Distribution	31
2.6	x_1 and x_2 vectors	31
2.7	lines connecting 10 random points between x_1 and x_2	32
3.1	Layout of stirring tank[50]	36
3.2	Workflow of sequential learning[47]	38
3.3	Left: As in classic Machine Learning, identically distributed selection (black dots) is desired; Right: Initial data (black dots) sampling with sequential learning and sequential sampling (red arrows) of new samples(white dots).[47]	39
4.1	Local and Global minima	42
4.2	Optimization of $Y = (X * 6 - 2)^2 \cdot np.sin(X * 12 - 4)$ using active learning	49
4.3	Flow sheet Extractive Distillation	51
4.4	low quality - Heat duty v/s Iterations	53

4.5 low quality - Product purity v/s Iterations	53
4.6 low quality - No. of stages v/s Iterations	54
4.7 low quality - Operating pressure v/s Iterations	54
4.8 low quality - Reflux ratio v/s Iterations	55
4.9 low quality - Solvent to feed ratio v/s Iterations	55
4.10 high quality - Heat duty v/s Iterations	57
4.11 high quality - Product purity v/s Iterations	57
4.12 high quality - No. of stages v/s Iterations	58
4.13 high quality - Operating pressure v/s Iterations	58
4.14 high quality - Reflux ratio v/s Iterations	59
4.15 high quality - Solvent to feed ratio v/s Iterations	59
4.16 Low-quality v/s Iterations	61
4.17 High-quality v/s Iterations	61
4.18 Comparison of Low-quality data v/s High- quality data	62

B

List of Tables

3.1	Process conditions of stirring tank[50]	37
3.2	Key structural parameters of the stirring tank[50]	37
4.1	Low-quality data convergence results	56
4.2	High-quality data convergence results	60
4.3	Comparison results	62
4.4	Minimization of reboiler heat duty with low-quality - results . .	63
4.5	Minimization of reboiler heat duty with High-quality - results . .	64

C

List of Libraries

Python programming language

Python is a high-level, interpreted programming language that was first released in 1991. It was created by Guido van Rossum and is currently maintained by the Python Software Foundation. Python is known for its simplicity, readability, and ease of use, making it a popular language for beginners and experienced developers alike.

Sklearn

Scikit-learn (also known as sklearn) is a popular machine-learning library in Python, developed by a group of contributors under the guidance of the European Organization for Nuclear Research (CERN). It is an open-source library that provides a range of tools for machine learning, including data preprocessing, feature selection, classification, regression, clustering, and model selection.

Surrogate Modeling Toolbox (SMT)

Surrogate Modeling Toolbox (SMT) is a Python library for building surrogate models, also known as metamodels or response surface models. Surrogate models are used to approximate complex mathematical functions, simulations, or data-driven models, and they can be much faster to evaluate than

the original models, making them useful in optimization, uncertainty quantification, and sensitivity analysis.

SciPy

Scipy is a Python library for scientific computing and technical computing, built on top of the NumPy library. Scipy provides a range of functions and algorithms for numerical integration, optimization, interpolation, signal processing, linear algebra, and statistics.

D

Bibliography

- [1] M. Alloghani, D. Al-Jumeily, J. Mustafina, A. J. Hussain, and A. J. Aljaaf. A systematic review on supervised and unsupervised machine learning algorithms for data science. *Unsupervised and Semi-Supervised Learning*, 2019.
- [2] C. Atman, R. Adams, M. Cardella, J. Turns, S. Mosborg, and J. Saleem. Engineering design processes: A comparison of students and expert practitioners. *Journal of Engineering Education*, 96:359–379, 10 2007.
- [3] G. Bohner, K. Kindler, M. Pahl, and G. Kaibel. Method for recovering crude 1, 3-butadiene by extractive distillation from a c4 cut, June 5 2007. US Patent 7,226,527.
- [4] M. Bowler, D. Bowler, and M. Bowler. Raoult’s law revisited: Accurately predicting equilibrium relative humidity points for humidity control experiments. *Journal of Applied Crystallography*, 50, 11 2016.
- [5] R. D. Campbell. Fractional distillation: A laboratory experiment. *Journal of Chemical Education*, 39(7):348, 1962.
- [6] J. G. Carbonell, R. S. Michalski, and T. M. Mitchell. An overview of machine learning. *Machine learning*, pages 3–23, 1983.
- [7] F. Chemat and C. Boutekedjiret. *Extraction // Steam Distillation*. 12 2015.

- [8] B.-C. Chen, B.-Y. Yu, Y.-L. Lin, H.-P. Huang, and I.-L. Chien. Reactive-distillation process for direct hydration of cyclohexene to produce cyclohexanol. *Industrial Engineering Chemistry Research*, 53:7079–7086, 04 2014.
 - [9] C.-W. Chiu, M. Dasari, G. Suppes, and W. Sutterlin. Dehydration of glycerol to acetol via catalytic reactive distillation. *AICHE Journal*, 52:3543 – 3548, 10 2006.
 - [10] G. De Ath, R. M. Everson, A. A. Rahat, and J. E. Fieldsend. Greed is good: Exploration and exploitation trade-offs in bayesian optimisation. *ACM Transactions on Evolutionary Learning and Optimization*, 1(1):1–22, 2021.
 - [11] S. Dutta. *Optimization in chemical engineering*. Cambridge University Press, 2016.
 - [12] P. I. Frazier. A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*, 2018.
 - [13] V. Gerbaud, I. Rodriguez-Donis, L. Hegely, P. Lang, F. Denes, and X. You. Review of extractive distillation. process design, operation, optimization and control. *Chemical Engineering Research and Design*, 141:229–271, 2019.
 - [14] A. Géron. Hands-on machine learning with scikit-learn and tensorflow: Concepts. *Tools, and Techniques to build intelligent systems*, pages 73–274, 2017.
 - [15] D. E. Goldberg. Genetic algorithms in search, optimization, and machine learning. *Addion wesley*, 1989(102):36, 1989.
 - [16] D. González, P. Bastidas, G. Rodríguez, and I. Gil Chaves. Design alternatives and control performance in the pilot scale production of isoamyl acetate via reactive distillation. *Chemical Engineering Research and Design*, 123:347–359, 07 2017.
 - [17] R. B. Gramacy. *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences*. CRC press, 2020.
-

- [18] A. Hammoudeh. A concise introduction to reinforcement learning. 02 2018.
 - [19] Z.-H. Han and k.-s. Zhang. *Surrogate-Based Optimization*. 03 2012.
 - [20] L. Hegely, Ö. F. Karaman, M. T. Szucs, and P. Lang. Surrogate model-based optimisation of a batch distillation process. *Chemical Engineering Research and Design*, 2023.
 - [21] Y. Hu, F. Li, S. Wei, S. Jin, and W. Shen. Design and optimization of the efficient extractive distillation process for separating the binary azeotropic mixture methanol-acetone based on the quantum chemistry and conceptual design. *Separation and Purification Technology*, 242:116829, 2020.
 - [22] D. Kalpic and N. Hlupić. *Multivariate Normal Distributions*, pages 907–910. 01 2011.
 - [23] G. Kanat, B. Gezer, Z. Yilmaz, and Akbal. Distillation: Types and applications (kinetics and reactor design project). 06 2020.
 - [24] T. Kanika and T. Kirti. A comparative analysis of optimization techniques. *Int. J. Comput. Appl.*, 131(10):6–12, 2015.
 - [25] S. Khaleefa Ali and M. Qitan. simple distillation. 11 2021.
 - [26] B. Khoo and V. Minh. Mathematical modeling and simulation of a distillation column. 10 2010.
 - [27] P. A. King, J. C. Stalley, and W. L. Thomas. Paper 1: Vacuum distillation in petroleum technology. *Proceedings of the Institution of Mechanical Engineers, Conference Proceedings*, 179(7):80–92, 1964.
 - [28] A. Kulkarni, G. Krishnasamy, and A. Abraham. *Introduction to Optimization*, volume 114, pages 1–7. 09 2017.
 - [29] S. J. Kulkarni. Distillation-research, studies and reviews on modeling, simulation and combined mode separations.
 - [30] M. Le Lann, M. Cabassud, and G. Casamatta. Modeling, optimization and control of batch chemical reactors in fine chemical production. *Annual Reviews in Control*, 23:25–34, 1999.
-

- [31] J. Lee, Y. Son, K. S. Lee, and W. Won. Economic analysis and environmental impact assessment of heat pump-assisted distillation in a gas fractionation unit. *Energies*, 12(5):852, 2019.
 - [32] J. Ling, M. Hutchinson, E. Antono, S. Paradiso, and B. Meredig. High-dimensional materials and process optimization using data-driven experimental design with well-calibrated uncertainty estimates. *Integrating Materials and Manufacturing Innovation*, 6:207–217, 2017.
 - [33] T. Lookman, P. V. Balachandran, D. Xue, and R. Yuan. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *npj Computational Materials*, 5(1):21, 2019.
 - [34] K. McBride and K. Sundmacher. Overview of surrogate modeling in chemical process engineering. *Chemie Ingenieur Technik*, 91(3):228–239, 2019.
 - [35] I. M. Mujtaba. *Batch distillation: design and operation*, volume 3. World Scientific Publishing Company, 2004.
 - [36] A. C. Müller and S. Guido. *Introduction to machine learning with Python: a guide for data scientists*. " O'Reilly Media, Inc.", 2016.
 - [37] M. S. of Engineering. Lecture notes on bayesian optimization, 2015.
 - [38] E. Pahija, F. Manenti, and I. M. Mujtaba. Optimization of batch and semi-batch reactors. In *Computer Aided Chemical Engineering*, volume 32, pages 739–744. Elsevier, 2013.
 - [39] A. A. Pawar and U. Warbhe. Optimizing bayesian acquisition functions in gaussian processes. *arXiv preprint arXiv:2111.04930*, 2021.
 - [40] C. E. Rasmussen, C. K. Williams, et al. *Gaussian processes for machine learning*, volume 1. Springer, 2006.
 - [41] I. Roman, R. Santana, A. Mendiburu, and J. A. Lozano. Evolution of gaussian process kernels for machine translation post-editing effort estimation. *Annals of Mathematics and Artificial Intelligence*, 89(8-9):835–856, 2021.
-

- [42] S. Salhi and N. M. Queen. A hybrid algorithm for identifying global and local minima when optimizing functions with many minima. *European Journal of Operational Research*, 155(1):51–67, 2004.
 - [43] B. Shah and P. Shah. Reactive distillation in process industries. 05 2015.
 - [44] T. Shi, W. Chun, A. Yang, Y. Su, S. Jin, J. Ren, and W. Shen. Optimization and control of energy saving side-stream extractive distillation with heat integration for separating ethyl acetate-ethanol azeotrope. *Chemical Engineering Science*, 215:115373, 2020.
 - [45] A. Sobester, A. Forrester, and A. Keane. *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons, 2008.
 - [46] B. Van Tran, S. I. Ngo, Y.-i. Lim, W. Kim, K.-S. Go, and N.-S. Nho. Computational fluid dynamics of gas-liquid bubble column with hydrocracking reactions. In *Computer Aided Chemical Engineering*, volume 44, pages 313–318. Elsevier, 2018.
 - [47] C. Völker, R. Firdous, D. Stephan, and S. Kruschwitz. Sequential learning to accelerate discovery of alkali-activated binders. *Journal of materials science*, 56:15859–15881, 2021.
 - [48] C. Völker, S. Kruschwitz, B. Moreno Torres, R. Firdous, G. Zia, and D. Stephan. Accelerating the search for alkali-activated cements with sequential learning. In *FIB Conference 2022*, pages 1–9, 2022.
 - [49] A. Wilson and R. Adams. Gaussian process covariance kernels for pattern discovery and extrapolation. *30th International Conference on Machine Learning, ICML 2013*, 02 2013.
 - [50] Z. Yao, H. Xu, J. Li, and T. Xu. Multi-objective optimization of stirring tank based on multiphase flow simulation. *Chemical Engineering Research and Design*, 189:680–693, 2023.
 - [51] M. Z. Zakaria, H. Jamaluddin, R. Ahmad, and S. M. Loghmanian. Comparison between multi-objective and single-objective optimization for the modeling of dynamic systems. *Proceedings of the institution of mechanical engineers, part i: journal of systems and control engineering*, 226(7):994–1005, 2012.
-