SLOVAK UNIVERSITY OF TECHNOLOGY
IN BRATISLAVA

FACULTY OF CHEMICAL AND FOOD TECHNOLOGY

Reg. No.: FCHPT-16584-97449

Advanced Control of a Laboratory Distillation Column

MASTER THESIS

2023 **Bc. Tímea Mészárosová**

SLOVAK UNIVERSITY OF TECHNOLOGY IN BRATISLAVA

FACULTY OF CHEMICAL AND FOOD TECHNOLOGY

Reg. No.: FCHPT-16584-97449

Advanced Control of a Laboratory Distillation Column

MASTER THESIS

Study programme:	Automation and Information Engineering in Chemistry and Food Industry
Study field:	Cybernetics
Training workspace:	Department of Information Engineering and Process Control
Thesis supervisor:	doc. Ing. Radoslav Paulen, PhD.

Bratislava 2023 **Bc. Tímea Mészárosová**

Slovak University of Technology in Bratislava
Department of Information Engineering and Process Control

Faculty of Chemical and Food Technology
Academic year: 2022/2023
Reg. No.: FCHPT-16584-97449

STU FCHPT

MASTER THESIS TOPIC

Student:	**Bc. Tímea Mészárosová**
Student's ID:	97449
Study programme:	Automation and Information Engineering in Chemistry and Food Industry
Study field:	Cybernetics
Thesis supervisor:	doc. Ing. Radoslav Paulen, PhD.
Head of department:	doc. Ing. Martin Klaučo, PhD.
Consultant:	Ing. Martin Mojto
Workplace:	Department of Information Engineering and Process Control

Topic: **Advanced Control of a Laboratory Distillation Column**

Language of thesis: English

Specification of Assignment:

Distillation columns are among the most important equipment of the process industry. The quality of management of these facilities can have a significant impact on the overall profit. The complexity and significant economic impact of industrial distillation columns usually precludes effective analysis of various advanced control methods. The aim of this work is to design an advanced controller for a laboratory distillation column UOP3CC and to compare the quality of control using this controller with a basic control structure consisting of individual PI controllers.

Tasks:
- Selection and pairing of controlled and controlling quantities.
- Identification of partial transfer functions based on experimental measurements.
- Design of PI controllers for individual pairs of controlled and controlling quantities.
- Design of an advanced (LQ) controller.
- Assessment of control performance and comparison of studied control strategies.

Selected bibliography:

1. Ľ. Čirka, M. Fikar, J. Mikleš. Youla–Kučera Parameterisation Approach to LQ Tracking and Disturbance Rejection Problem, In Selected Topics in Modelling and Control, Editor(s): Mikleš, J., Veselý, V., Slovak University of Technology Press, vol. 5, pp. 8–12, 2007.
2. M. King, 2011. Process Control: A Practical Approach. John Wiley & Sons Ltd.

Deadline for submission of Master thesis:	14. 05. 2023
Approval of assignment of Master thesis:	07. 03. 2023
Assignment of Master thesis approved by:	prof. Ing. Miroslav Fikar, DrSc. – study programme supervisor

Honour Declaration

I declare that the submitted diploma thesis was completed on my own, in cooperation with my supervisor, with the help of professional literature and other information sources, which are cited in my thesis in the reference section. As the author of my diploma thesis, I declare that I didn't break any third-party copyrights.

.....
Signature

Acknowledgment

I would like to express my sincerest gratitude to my supervisor, doc. Ing. Radoslav Paulen, PhD., and my thesis consultant, Ing. Martin Mojto, for their invaluable insights, guidance, and feedback throughout my master thesis. Their unwavering support and patience have been instrumental in helping me reach this point.

I would also like to extend my heartfelt thanks to my family, whose unconditional love and encouragement have been a constant source of strength for me. To my partner, who has always stood by my side and believed in me, I am deeply grateful. Lastly, I would like to thank my dear friends for their unwavering support and motivation.

Thank you all for being a part of my journey and for helping me achieve my goals.

Abstract

Nowadays, the appropriate monitoring and control of industrial distillation columns play a crucial role in achieving the desired production. This thesis focuses on the design of the advanced control of Armfield's UOP3CC laboratory distillation column. Concentration is one of the difficult-to-measure variables in the industry, which is both time-consuming and costly to measure on a regular basis. As a solution to this issue, inferential sensors are used to indicate this difficult-to-measure quantity and then to use it in an advanced control structure. Inferential sensors enable fast and accurate measurement of key variables. Different approaches will be used in their design, including the data-driven approach, the model-based approach, and a combination of these, the hybrid approach. These sensors are capable to calculate difficult-to-measure or completely unmeasurable values that cannot be obtained directly from the sensors present on the laboratory distillation column. They are calculated based on the outputs from the physical sensors. The sensor inference methods studied in this work include ordinary least squares regression and the least absolute shrinkage and selection operator. Additionally, a custom procedure was developed based on data processing to identify one of the key properties of the process, steady state. Achievement of this state is essential for the identification of the process and for the subsequent design of the control system. To obtain the desired product concentrations, the appropriate variables must be selected, which can control the system correctly. This can be accomplished by using the variable pairing method. This approach determines which quantity will be controlled and which quantity should be manipulated to accomplish the goal. For a better understanding of the system, classical proportional integral derivative controllers are first proposed, which are designed by the trial-and-error method. Then a linear quadratic output feedback regulator is proposed for advanced control of the system. The goal is to minimize the optimisation function to ensure that the selected variable is controlled to the desired value. This work has several outputs, which also include a human-machine interface scheme in MATLAB Simulink or detailed documentation of how to proceed with the experiment on laboratory distillation column UOP3CC.

Keywords: binary mixture, distillation, inferential sensors, process control

Abstrakt

V súčasnosti zohráva pri dosahovaní požadovanej kvality výrobkov kľúčovú úlohu vhodné monitorovanie a riadenie priemyselných destilačných kolón. Táto práca sa zameriava na návrh pokročilého riadenia laboratórnej destilačnej kolóny Armfield UOP3CC. Koncentrácia je jednou z ťažko merateľných veličín v priemysle, ktorej pravidelné meranie je časovo aj finančne náročné. Ako riešenie tohto problému sa používajú inferenčné senzory na indikáciu tejto ťažko merateľnej veličiny a pre následný návrh pokročilej riadiacej štruktúry. Inferenčné senzory umožňujú rýchle a presné meranie kľúčových veličín. Pri ich návrhu sa použijú rôzne prístupy vrátane prístupu založeného na údajoch, prístupu založeného na modeli a ich kombinácie, hybridného prístupu. Tieto senzory sú schopné vypočítať ťažko merateľné alebo úplne nemerateľné hodnoty, ktoré nemožno získať priamo zo snímačov osadených na laboratórnej destilačnej kolóne. Vypočítavajú sa na základe výstupov z fyzikálnych snímačov. Metódy návrhu senzorov skúmané v tejto práci zahŕňajú obyčajnú regresiu pomocou najmenších štvorcov a operátor najmenšieho absolútneho zmenšenia a výberu. Okrem toho bol na základe spracovania údajov vyvinutý vlastný postup na identifikáciu jednej z kľúčových vlastností procesu, ustáleného stavu. Dosiahnutie tohto stavu je nevyhnutné pre identifikáciu procesu a pre následný návrh riadiaceho systému. Na dosiahnutie požadovaných koncentrácií produktu je potrebné vybrať vhodné premenné, ktoré dokážu systém správne riadiť. To sa dá dosiahnuť použitím metódy párovania premenných. Tento prístup určuje, ktorá veličina sa bude riadiť a s ktorou veličinou sa má manipulovať, aby sa dosiahol požadovaný cieľ. Pre lepšie pochopenie systému sa najprv navrhnú klasické proporcionálne integrálne derivačné regulátory, ktoré sa navrhujú metódou pokus-omyl. Potom sa na pokročilé riadenie systému navrhuje lineárny kvadratický regulátor so spätnou väzbou. Cieľom je minimalizovať optimalizačnú funkciu, aby sa zabezpečilo riadenie vybranej veličiny na požadovanú hodnotu. Táto práca má niekoľko výstupov, medzi ktoré patrí aj schéma rozhrania človek-stroj v prostredí MATLAB Simulink či podrobná dokumentácia postupu pri experimente na laboratórnej destilačnej kolóne UOP3CC.

Kľúčové slová: binárna zmes, destilácia, inferenčné senzory, riadenie procesu

List of symbols and abbreviations

Acronyms

β	Estimated value of inferential sensor
ℓ_i	i-norm penalty
$\hat{y}$	Estimated output
λ	Parameter of penalty
μ_x	Mean of measurement
σ	Standard devitation
σ_x^2	Variance
$\tilde{x}$	Matrix of deviating input variables
ε	Residual error
D^{Hist}	Historical domain
e	Control error
G	Transfer function
K_{i}	Control gain, where i = P,I,D
R	Residual
r	Pearson's correlation coefficient
$T_{\text{I, D}}$	Integral or derivative time constant
$u_{c,i}$	Input to the PID controller, where i = P,I,D
w	Desired value
X	Input data
x^s	Input in steady state
X^{Hist}	Historical input

x_{std}	Standardised variable
y	Measured output
y^{Hist}	Historical output
A,B,C,D	System matrices
E	Euclidean norm
HMI	Human machine interface
J	Objective (cost) function
LASSO	Least absolute shrinkage and selection operator
LQR	Linear quadratic regulator
MIMO	Multi-input multi-output system
MSE	Mean squared srror
MV	Manipulated variable
o	System order
OLSR	Ordinary least squares regression
P	Lyapunov equations matrix
P&ID	Piping and instrumentation diagram
PID	Proportional, integral and derivative controller
PV	Process variable
Q, R	Weighting matrices
QI	Quality indicator
RI	Refractive index
RMSE	Root-mean-square error
S	Lagrange multipliers matrix
SP	Set Point
VLE	Vapour-liquid equilibrium
Z	Symmetric matrix

Physical and chemical values

ρ	Fluid density [kg/m^3]
CH_4O or CH_3OH	Methanol, methyl alcohol, wood alcohol or carbinol
g	Gravitational acceleration [m/s^2]
k	Contant based on Markofsky-Harleman
p	Hydrostatic pressure [Pa]
$w_{\text{CH}_3\text{OH}}$	Mass fraction of methanol in liquid [kg/kg]

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Since the earliest civilizations, distillation has been significant to people. The first knowledge of the distillation process dates back to the first century AD. It is a separation process used to separate components based on different (relative) volativity [22]. Separation is achieved through the contact of the liquid and vapour phases. There is a direct transfer of heat from the vapour to the liquid. The feedstock that is processed is called the feed which is separated into the distillate and bottom product. It can be divided into two basic groups, batch distillation, and continuous distillation. Continuous distillation is a process of multiple distillations and condensation [16]. Currently, this process is already in common use, whether in the laboratory or the industrial field [39, 45].

An Armfield UOP3CC laboratory distillation column [27] is located at the Department of Information Engineering and Process Control, Faculty of Chemical and Food Technology, Slovak University of Technology in Bratislava. Despite distillation being a complex process, by measuring the appropriate quantities and implementing a control strategy, it is possible to ensure direct (online) measurement of the quality (concentration) of both products. The laboratory distillation column UOP3CC can be used as a prototype to test new techniques for industrial distillation columns, to train operators, or to test theoretical knowledge, among other things. However, to bring the column to such a level, it is necessary to equip it with many additional features.

Despite the frequent use of distillation, effective control of this process is difficult to achieve. It is caused by the complexity of the process, which expresses nonlinear behaviour and strong interactions between the internal variables. The optimal control of the distillation columns is achieved by combining input-output variables into effective configurations, such as LV, LB, DV, LD or Ryskamp [21, 45]. These basic control strategies outline the fundamental control of a distillation column with the integration of PID controllers [35, 36]. The PID control within the distillation columns has a rich history and has effectively adapted to changes in technology, from the analogue period

to current digital computer control systems [19]. Nevertheless, over time there was a need to increase the level of control, and advanced process controllers have been developed [26]. The implementation of an advanced control strategy can be performed in different ways. One way is that instead of multiple PID control loops, a controller with multiple inputs and outputs is implemented. Another possibility is that although PID controllers are still implemented, an advanced process controller is added. The advanced controller will act as a master and set the desired quantities of the slave, PID controllers [21, 26]. Their implementation is intended to optimize the process and deal with interactions between variables. Advanced process controllers, such as adaptive output-feedback linear quadratic regulator (LQR) [26], can improve the efficiency of the process. These controllers minimize the conventional quadratic criterion [44]. They use weight matrices to ensure the required quality of control. The tuning of these matrices is based on engineering decision-making resulting from the trade-off between speed and accuracy of control.

The performance of industrial control strategies is usually highly related to the quality of the monitoring of key process variables. In the process industry, devices are usually equipped with many sensors that provide important information about different desired quantities. The primary goal of the sensors is to provide data for process monitoring and control. Recently, the data from these sensors also find application in the prediction. Furthermore, this data can be used to design mathematical structures called soft (from software) or inferential sensors [20]. Such a sensor uses measurements from other easy-to-measure variables in the process (e.g., temperatures, pressures) to estimate difficult-to-measure or unmeasured variables. Inferential sensors also have the advantage of being relatively simple to design and transparent in their structure. The design of inferential sensors is based on the creation of mathematical models that determine the relationship between input and output quantities. Their design may be categorised into three main approaches based on the structure and modelling process: model-based inferential sensors, data-driven inferential sensors and hybrid inferential sensors [11, 36]. In the case of model-based sensors, it is assumed that some prior knowledge about the physical and chemical laws related to the studied process is available. Data-driven sensors are designed based on measured data from the process plant. Lastly, hybrid inferential sensors are a combination of model-based and data-driven approaches. The design of inferential sensors is usually (in the industrial sector) covered by simple regression approaches, such as ordinary least squares regression (OLSR) or least absolute shrinkage and selection operator (LASSO) [43, 46].

The aim of this work is to develop an advanced control structure for the studied Armfield UOP3CC laboratory distillation column. However, this cannot be achieved without increasing the level of automation. It is essential to measure critical input and output

variables, but this can also be time-consuming and expensive. A suitable alternative is to design inferential sensors that can measure some difficult-to-measure variables, such as the concentration of product components. These sensors can also facilitate the design of single-step control loops and are essential to identify the laboratory distillation column for advanced control design. After implementing different control strategies, we aim to compare the performance of the distillation column under different control methods and evaluate the advantages and disadvantages of each strategy. Ultimately, our goal is to develop a distillation column that can efficiently separate components and provide the desired product quality in a certain amount of time. We aim to make the distillation column available for experimentation with detailed documentation, during which all necessary variables can be measured and the process can be controlled using controllers via an improved human-machine interface.

The document consists of a theoretical and a practical part. The theoretical part, at first describes the distillation process, focusing on a particular type of distillation column, the continuous distillation column. This section involves the description of binary mixtures, with a focus on the mixture of methanol and water considered in the experimental section of the thesis. In this section of the work, readers can find an introduction to the basic principles of data processing and related basic steps required for further analysis, such as inferential sensor design or advanced process controller design. This part of the thesis also introduces various approaches to the inferential sensors and demonstrates the specific methods that have been used during this work, such as the OLSR and LASSO methods. In this part of the thesis, the accuracy performance indices are introduced. The next section describes the identification of the process, which is essential for designing advanced controllers. In the final section of this part, readers will gain a basic understanding of the control principles of distillation columns, including the state-of-the-art types of control schemes, how to choose and pair controlled and manipulated variables, and the functions of PID controllers and LQR.

The practical part of the thesis can be divided into three main sections. Firstly, it provides a detailed description of the Armfield UOP3CC laboratory distillation column. The next section focuses on the design and various applications of inferential sensors. The thesis explores four categories of inferential sensor application: (1) flow rate measurement for calibrating feed and bottom product pumps, (2) liquid level measurement in the reboiler, (3) composition analysis of the products (distillate and bottom product, feed), and (4) evaluating the steady state of the process. The third section of the practical part delves into advanced controller design. It begins with a description of the process for selecting and pairing input-output variables to create efficient control strategies. The subsequent section outlines the procedure for designing

PI controllers. Following that, process identification is introduced and described. Lastly, there is further information available regarding the design of advanced control structures.

Chapter 2

Theoretical Foundations

2.1 Distillation Process

The basic principle of the distillation process is the separation of two or more liquid mixtures based on the different volatilities. Distillation is characterised by the liquid-vapour phase change. This phenomenon ensures that with the different chemical potentials between the liquid and vapour components, the separation of the individual components occurs. This leads to the different compositions of the liquid and vapour phases. The distillation process is initialized with the first contact of the liquid and vapour phases. After a certain time interval of evaporation of the liquid phase, the vapour phase is enriched with a more volatile component. The more volatile component is considered to be a substance that has a lower boiling point. Logically, there will be a higher concentration of the less volatile substance in the liquid phase with a higher boiling point. These phases are in equilibrium and hence have the same temperature and pressure [16, 21, 45].

According to the feed inlet, distillation can be divided into two basic categories, batch distillation and continuous distillation. In batch distillation, it is necessary to wait until the process is finished, subsequently, the bottom product can be discharged. The distillate is withdrawn continuously. After the process is completed, another batch of feed can be added to the plant and the separation restarted. In the continuous distillation process, there is no need to wait for the total amount of feed to separate and it is possible to increase the feed flow rate during the distillation process. Moreover, the distillate and bottom products are continuously generated during the process. According to the number of distillation stages, it is possible to distinguish between single-stage and multi-stage (distillation column) distillation [16].

During the single- and multi-stage distillation the vapour passes through a condenser. This refers to the fact that the hot vapours are cooled by a condenser into which the coolant enters [16, 23]. The main difference between the types occurs after condensation.

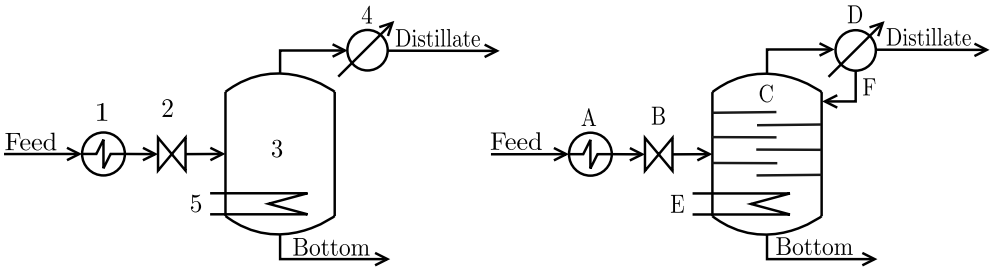


Figure 2.1: Scheme of single-stage (on the left-hand side) and multi-stage distillation column (on the right-hand side).

A comparison between the two types of distillation is provided in Figure 2.1. We consider that the same raw material, i.e., a feed of identical composition, enters both plants. This feed is often preheated by a heat exchanger (preheater) indicated by 1 and A in Figure 2.1 and then it is led to the column by a pump (2, B). Vapour flows upwards from the bottom of the column (3, C). In single-stage distillation, the cooled vapours, i.e. the liquefied vapours after condensation (4, D), are carried away and the final product of distillation, the distillate, is obtained. In the distillation column, only part of the condensed vapour is drawn off (as distillate product), but part of it is returned to the distillation column (as reflux flow). Because of this, there is contact between the vapour phase rising up and the liquid phase flowing down the column. The vapours are generated by the heat input within the reboiler (5, E). On the left-hand side, there is a simple distillation where all of the condensed vapour leaves the column as the distillate product. On the right-hand side, part of the condensed vapour is drawn off as the distillate product and the rest of the condensed vapour is returned to the distillation column as reflux (F). In both pictures, we can see another stream at the bottom of the column. This is the second product of the distillation, the bottom product [16, 45]. For additional information on various types and classifications of distillation, readers may refer to [16, 45].

2.1.1 Distillation Column

Continuous distillation provides better separation of the individual phases compared to batch distillation (see in Section 2.1) due to the intermittent contact between phases. This property is achieved by the trays (distillation stage) within the column section. The number of trays is directly proportional to the desired distillate quality. However, the number of trays also affects the overall cost of the distillation process. With a higher number of trays, it is possible to achieve separation even of difficult-to-separate products, but this has a significant impact on the dynamics of the process due to the

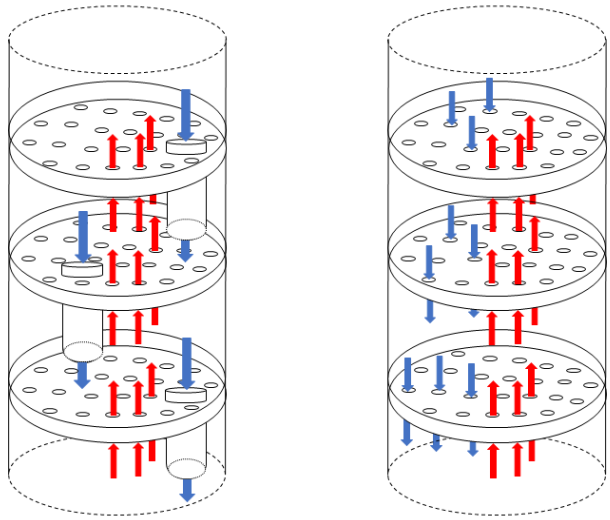


Figure 2.2: Scheme of the crosswise (on the left-hand side) and upstream (on the right-hand side) flow phases.

flow rate of the individual phases. The higher number of trays ensures a longer contact time between the phases at a larger interface, but it also increases the transport time for the phases to reach the top and bottom of the distillation column. In the case of trayed columns, it is crucial to ensure the appropriate insulation of the trays. This can have a significant negative impact on the temperatures within the column, which can lead to a decrease in the overall profit of the distillation column [16, 29, 21].

In general, the multi-stage distillation columns (or rectification columns) have cheap maintenance, a wide range of applications and are affordable. The contact of the vapour and liquid phases can occur in distillation columns in two ways, either crosswise or upstream [16, 29, 34, 38]. A comparison can be seen in Figure 2.2, where red arrows indicate the direction of the vapour phase coming from the lower section (reboiler) of the distillation column and blue arrows show the direction of the liquid phase coming from the upper section (condenser) of the distillation column. On the left-hand of the figure, the crosswise flow of the different phases can be seen. The liquid flows from the upper trays and it is partially retained on the individual trays within the column. The vapour flows from bottom to top through smaller holes located on the base plate of the trays. The vapour phase passes through the liquid layer and thus there is contact between the liquid and vapour phases. Once the maximum level on the tray is reached, further liquid flows through the overflow pipe to the lower trays.

On the right-hand side of the figure, the upstream flow of the phases is shown. As we can see, there are no overflow tubes, only small holes on the plate. These small holes ensure contact between the phases so that the individual phases are forced to pass through the same holes. Trays with the crosswise flow are more frequently used in the industry compared to upstream flow trays due to the more intense and longer contact of the liquid and vapour phases, ensuring higher profitability. Also, the trays are placed one above the other, horizontally, and there is often an equal distance between the trays [16, 29, 34, 38]. This arrangement can be seen at close range in Figure 2.2 too.

2.1.2 Binary Separated Mixture

Compared to industrial separations, simplified settings are frequently employed performed in laboratory conditions, such as considering a mixture of only a few components [22]. The well-known and frequently used mixture for separation in laboratory distillation columns is a binary mixture of methanol and water. Both components within the separated mixture have one -OH functional group which is attached to the electropositive centre. Methanol has fewer tight hydrogen bonds compared to water. Due to its chemical properties, methanol has a lower boiling point than water [1, 38]. Methanol is more volatile compared to water and its boiling point is 64.7° C. Methanol is characterized by melting point −97.8° C and its critical temperature is 240° C. Its molecular formula is CH_4O or CH_3OH. Methanol or methyl alcohol has the simplest structure of primary alcohols. In practice, we may also hear the name wood alcohol or carbinol. It is a flammable alcohol, that is toxic and harmful to health. It is a clear, colourless and volatile liquid which has a slight alcoholic odour at laboratory temperature. It is highly soluble in water and in most organic solvents, such as ethanol, ether, acetone or chloroform. It is a hygroscopic product and reacts explosively when gets to the reaction with some metal powders or strong oxidizers [15, 38]. Risk and safety phrases for methanol can be found in [15].

During the distillation process, separation occurs when there is thermodynamic equilibrium between the vapour and liquid phases [16, 22]. The success of binary mixture separation can be determined from a vapour-liquid equilibrium (VLE) diagram shown in Figure 2.3. There are two types of VLE diagrams. The first type describes the temperature (top left chart) or pressure (bottom left chart) as a function of the liquid (blue trajectory) and vapour (red trajectory) phases. The second type explains the relationship between the composition of the liquid and vapour phase considering constant pressure (top right chart) or constant temperature (bottom right chart). In both cases, the composition of a more volatile component (methanol) is considered [6, 4].

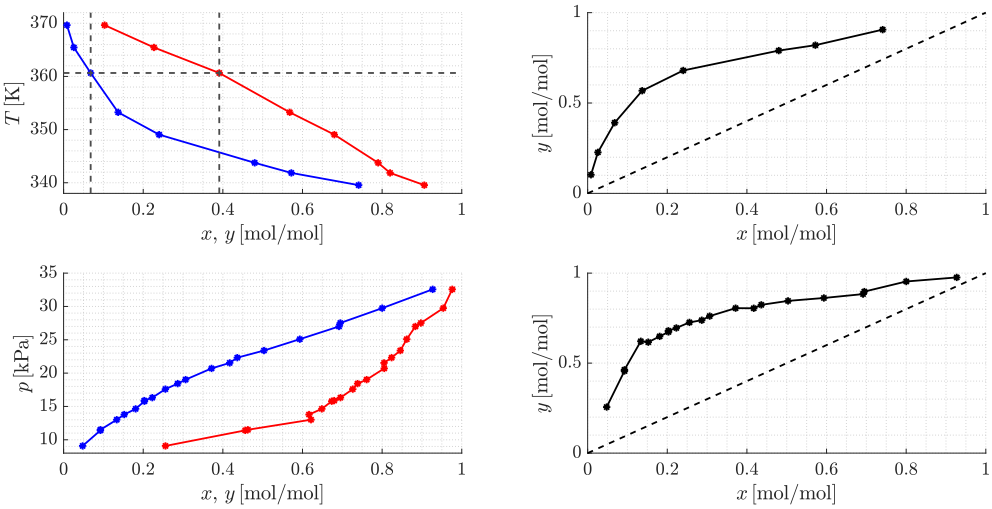


Figure 2.3: VLE diagram of binary methanol-water mixture, where on the upper left is the dependence of T on x (blue trajectory) and y (red trajectory) at p = 97.99 kPa with an auxiliary line (black dashed line), on the lower left is p on x (blue trajectory) and y (red trajectory) at T=312.91 K, the upper right is the dependence of y on x (black trajectory) at p = 97.99 kPa, and the lower right is the dependence of y on x (black trajectory) at T = 312.91 K [6, 4].

From these diagrams, we can obtain the composition of more volatile components in both the liquid and vapour phases. This can be useful to deduce the composition of the distillate and the bottoms during distillation separation based on pressure and temperature. Overall the VLE diagram helps to determine the operating conditions. For example, by choosing the constant pressure 97.99 kPa and temperature of T = 360.65 K, the composition of methanol in the liquid phase is $x = 0.068$ [mol/mol] and in the vapour phase, it is 0.391 [mol/mol]. Moreover, this chart shows that with increasing temperature, the value of y in the vapour phase decreases. This indicates that the vapour phase is enriched with more volatile component and vice versa (when the temperature decreases). The opposite effect occurs with decreasing temperature. It can be seen on the lower left graph that at a constant temperature, it is possible to expect an increase in the concentration of both phases with an increase in pressure [16, 22].

2.2 Distillation Column Control Principles

In the case of MIMO (multi-input multi-output) processes, the pairing of variables is problematic and it is often necessary to use predictive control to control the process in a robust way [19]. To ensure effective control of distillation columns, the first step is to gain an awareness about the limits of the system. Without appropriate limit checks and overrides designed to handle operating limits, frequent operator interventions are necessary during upsets. This can lead to switching the column from automatic control more frequently than required. Such behaviour harms safety and reduces the effectiveness of the system [29].

The primary application of actuators in the distillation column is to regulate product purity. Additionally, they are used to minimize unit disturbances caused by changes in inputs to the device. The design of the control strategy involves considering these factors and selecting suitable actuator combinations [29].

The primary application of actuators in the distillation column is to regulate product purity. Additionally, they are used to minimize unit disturbances caused by changes in inputs to the device. When designing the control strategy, these factors are taken into consideration, and appropriate couples are selected [29].

Controlled variables are typically obvious and are found when process objectives are stated and understood. However, deciding which variables to adjust by control might be more difficult. The general rule is to manipulate the stream that has the most impact on the controlled variable [29]. Such as manipulating with:

- the stream that is most highly correlated with the controlled variable,
- the stream that is least sensitive to surrounding conditions,
- the stream with a smaller flow if more streams have the same effect on the controlled variable,
- the stream least likely to produce interaction issues,
- the stream with the lowest probability of causing interaction problems.

However, these rules might conflict with each other, complicating the decision on how to pair variables. The relative gain calculation approach may be used to optimise the pairing of controlled and manipulated variables (MV) [29]. The control error can be calculated as:

$$e(t) = w(t) - y, \tag{2.1}$$

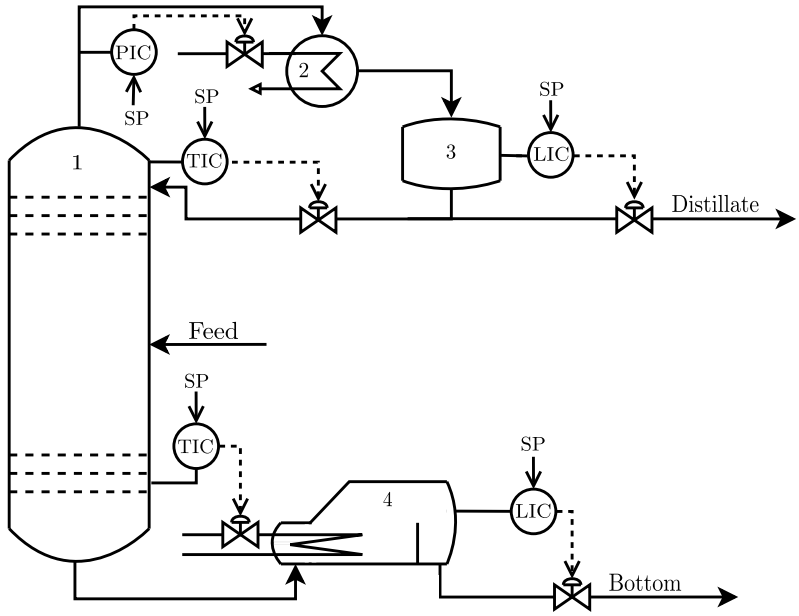


Figure 2.4: Scheme of distillation column with the LV control configuration [21], where (1) is the distillation column, (2) is the condenser, (3) is a reflux drum and (4) is a reboiler.

where $w(t)$ is the desired value or set point (SP) of the chosen variable and y is the measured output or process variable (PV) of the system [19].

In general, a distillation column has five basic process variables (PV). These are pressure, reflux drum level, column base level, and composition of distillate and bottom product. Correspondingly, there are typically five manipulated variables (MV): distillate flow, bottom flow, reflux flow, and the duty of the reboiler and condenser. The main problem is pairing, which determines the coupling between PVs and MVs. This results in a total of 5! (or 120) possible combinations. While many of these combinations are infeasible or illogical, there are still numerous viable pairings to consider. After pairing the variables, the control loops that are formed are integrated into the control scheme. Nevertheless, it is important to recognize that not all control schemes demonstrate economic advantages once the variables are paired. Distillation columns have individual characteristics, operating conditions, and objectives that require customized control schemes. Therefore, it is essential to acknowledge that control schemes with diverse combinations of process variables and manipulated variables may not be universally applicable to all distillation columns. Designing an appropriate control scheme requires

careful consideration of the distinct attributes of the distillation column, operational parameters, and objectives [21].

Another challenge is the interactions between variables, where one MV can affect multiple PVs (not only one desired). Furthermore, some parts of the plant can increase the transport delay, e.g. the reflux tank. The transport delay may increase if there is a large number of storage tanks present in a given device. Controllability may also be influenced by the fact that the distillation device is connected to another device and thus, e.g., there may be a fluctuation in the flow rate or enthalpy of feed. One possible solution is the variable pairing method, which can prevent controllers from fighting with each other [21].

Balance is crucial for process control, and pressure is a key variable that helps maintain equilibrium in a plant. In the case of a distillation column, the condenser consumes the most energy while the reboiler adds the most energy. If more energy is added to the reboiler than the condenser can consume, the pressure in the column will increase. Pressure has advantages as it responds quickly to changes, affects the behaviour of valves during faults and affects the dew point and bubble point of the mixture. For these reasons, it is obvious that one of the control loops will almost always be dedicated to pressure control [21].

As mentioned, there are many wiring strategies for input and output variables in distillation column control. Among the basic ones belong LV, LB, DV and LD configurations [21] where the two capital letters indicate which flows are used to control the quality of products. With five control loops available, one of which is dedicated to pressure control, the remaining two loops may be used to regulate product concentration and the remaining two to manage liquid levels in the distillation column. The LV configuration recommends to manipulating the flow rate of the reflux flow and the heating steam flow rate. The LB configuration suggests controlling the flow rate of the reflux flow and the flow rate of the bottom. In the DV configuration, the two main control loops are designed for the flow of distillate and the heating steam flow rate. The LD configuration controls the flow rate of the reflux flow and the flow of distillate [21].

One of the basic schemes can be seen in Figure 2.4. The figure shows the LV diagram. In this scheme can be seen, that the pressure at the head of the column is controlled by the flow of the cold medium to the heat exchanger. The liquid level in the reflux drum is controlled by the distillate flow rate and simultaneously the liquid level in the reboiler is controlled by the bottom product flow rate. The temperature on the head of the column is controlled by the reflux valve and the temperature at the bottom of

the column is controlled by the heating steam flow rate, which flows to the reboiler. In this scheme then the distillate concentration is controlled by the reflux and the bottom product concentration is controlled by the vapour flow in the reboiler. This is the most widely used configuration for controlling distillation columns [21].

The extensions of the basic control configurations (i.e., LV, LB, DV, LD) are hybrid schemes, such as the Ryskamp scheme. In this case, a reflux controller is added to make the composition of the bottom product more sensitive to changes in the reboiler [21].

2.2.1 Basic Control

The PID (proportional, integral, derivative) algorithm has been used in industry since the 1930s and has become the most commonly used controller type. This method proved appropriate for many fundamental processes found in the process industries. One of the key characteristics of PID is that there is a clear connection between plant responses and the manipulation and adjustments of the three controller terms when dealing with simple plants. For this reason, the trial-and-error method was often used. The tuning technique of the PID controller is divided into two parts, to determine the structure of the PID controller and to select the coefficients for the chosen controller or tune the controller [19]. Ziegler and Nichols [21] developed the first PID tuning rules. The Ziegler-Nichols rules were the top-ranked and most often applied rules for nearly sixty years, despite several additional empirical rules being later proposed [19, 21]. Equally well-known methods include the relay experiment of Åström and Hägglund [19], the principles of which have been used for twenty years. The most popular description of process dynamics is the first-order plus delay process model, which has parameters K, T and D. Although many distributed control system suppliers have made an effort to provide new, more effective algorithms, the basic control applications nearly universally still rely on this approach [21]. This type of controller is often used and developed in the academic field. According to industrial records, the PID controller is the primary control instrument applied for many control problems, especially those in the process industries [19].

The P part of PID is used when the controller action must be proportional to the size of the process error signal [21]. The control law for this member in the time domain representations is given as:

$$u_{\mathrm{c,\,P}}(t) = K_{\mathrm{P}} e(t), \tag{2.2}$$

where $u_{\mathrm{c,\,P}}(t)$ is the input to the proportional controller, K_{P} is the proportional gain and $e(t)$ is the control error [21].

The controller gain is always defined in absolute form. The control algorithm contains an extra engineer-defined parameter called action. When it is set to direct, the controller output grows as PV increases. When it is set to inverse, the output decreases as PV increases. The main drawback of proportional control is that in the steady state always will be a drift. Except in initial conditions, the PV will never approach the SP. By raising gain, the offset may be minimised while increasing oscillatory behaviour. In any process, due to this oscillations can be reached an unstable process, before the offset can be decreased to zero. There are cases in which the offset is useful. However, in most cases, the PV must reach the SP. Then the P controller is not enough in such situations [21].

When the controller has to compensate for any continuous offset from the reference signal value, the I part is applied. Integral control eliminates the control error from the P part of PID by removing offset without using an unnecessarily large controller gain [19]. It is also known as a reset action because its function is to continuously modify the controller output till the error persists. This can be achieved by ensuring that the rate of output change is proportional to the magnitude of error [21]. Its time domain representation is as follows:

$$u_{\mathrm{c,\ I}}(t) = K_{\mathrm{I}} \int_0^t e(\tau)\, d\tau, \tag{2.3}$$

where $u_{\mathrm{c,\ I}}(t)$ is the input to the integral controller, K_I represents the integral controller gain [19].

The offset can be removed by a small value of the integral action. However, rapid or aggressive removal can lead to oscillations. The solution may be to reduce the value of gain to remove the oscillations. In many cases, a PI controller is sufficient, and it is enough to find a compromise between the P and I components. At the same time, such a solution is also easier to implement as only two parameters are set, not three as in the case of a PID controller. Without considering the derivative component, it is easier to perform tuning by trial and error method. Although the PI controller can be further improved by adding the last component of the controller, the D part [21].

The last member is the D part, which uses the rate of change of the error signal to predict the action [19]. If it detects a sudden change in the error, it responds by taking action. Even if the error is currently zero, rapid changes in the process variables can cause it to increase with time. By modifying the output in proportion to the rate of error change, derivative action tries to stop this from occurring [21]. It can be described as:

$$u_{\mathrm{c,\ D}}(t) = K_{\mathrm{D}} \frac{\mathrm{d}e}{\mathrm{d}t}, \tag{2.4}$$

where $u_{\mathrm{c,\,D}}(t)$ is the input to the derivative controller, K_{D} is the derivative control gain [19].

This part of the controller can be useful when the process deadtime is large, to make up for the time that passed between the change in the PV and the disturbance that generated it [21]. Due to the potential increase of measurement noise pure derivative control can not often be implemented in real-life situations [19]. It is even frequently used with temperature controls which exhibit significant deadtime as at the output from the heat exchanger or on distillation column trays [21].

The key advantages of PID control are its past record of efficiency, wide availability, and simplicity of use. Despite the development of new regulators, PID control has not truly been able to be replaced. The control engineer actually uses other approaches when the process environment requires a more advanced control solution for managing a complex process [19].

2.2.2 Advanced Process Control

Classical control theory desired a one-loop-at-a-time strategy. The major advantage of advanced control is that it enables multi-input multi-output (MIMO) systems with multiple feedback loops. As a result, it is a simple layout design strategy. Another advantage is that this control also allows the designing of device controllers with an arbitrary structure [50, 26]. It is often impossible or not cost-effective to measure all the state variables. As a result, various new techniques have been constructed, leading to the generalisation of the state-feedback issue to output feedback [14]. One of the possible methods to use is called the Linear Quadratic Regulator (LQR) with output feedback [26]. The aim is to choose a performance criterion that stabilises the closed-loop, and then formulate matrix equations that can be solved for all control gains at the same time. Various computer programs can be used to solve these matrix problems. Once the performance requirement is chosen, matrix design equations explicitly compute the control gains and closed-loop stability will be ensured. In this way, engineering decision-making in the design of controllers is presented to select the required performance criterion [26].

Assume a system given by the following linear state space model:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t), \end{aligned} \tag{2.5}$$

where A, B, C are the system matrices [26].

The considered control law is as follows:

$$u(t) = -Ky(t), \tag{2.6}$$

where K is a searched gain matrix of coefficient for feedback controllers [26].

The dimensionality of K is mxn, where m is the number of inputs. This control rule reflects the fact that only measurable values may be used to calculate the input. Controlled systems depend only on the output not from entire states, thus this is a problem called output feedback [26].

State regulation in the system aims to stabilise the system by reducing the initial condition error to zero. This can be accomplished by choosing choosing the control input $u(t)$ that minimises a quadratic cost function using the formula below in the time-variant form:

$$J = E \int_0^\infty (x^{\mathrm{T}} Q x + u^{\mathrm{T}} R u) \mathrm{d}t, \tag{2.7}$$

where J is the objective (cost) function also called quality indicators (QI), Q and R are symmetric positive semidefinite weighting matrices [26].

Therefore, these matrices are limited as follows [26]:

$$Q \geq 0, \quad R \geq 0. \tag{2.8}$$

Different Q and R values can be used in various designs to fulfil different performance requirements that belong to certain functions of the state and input. The requirements can be balanced by minimising the state vector and control input by varying the relative magnitudes of Q and R. A greater R more severely penalises the control input, which causes a reduced norm in comparison to the state vector. In opposition, a higher Q speeds up the state vector reaches zero. In summary, for the linear system represented by (2.5), it is necessary to find the feedback coefficient matrix K at the input of the control that minimises the value of the quadratic QI [26].

To find the optimal gain matrix K, the following equations must be true:

$$\begin{aligned} 0 &= A_c^{\mathrm{T}} P + P A_c + C^{\mathrm{T}} K^{\mathrm{T}} R K C + Q, \\ 0 &= A_c S + S A_c^{\mathrm{T}} + X, \\ K &= R^{-1} B^{\mathrm{T}} P S C^{\mathrm{T}} (C S C^{\mathrm{T}})^{-1}, \end{aligned} \tag{2.9}$$

where the first two equations describe Lyapunov equations and the third one is used to compute K [26].

To solve these equations, it is necessary to specify two symmetric $n \times n$ matrices, P and S, to identify K. The matrix S is a matrix of Lagrange multipliers and P is a constant, positive semi-definite matrix [26]. For these equations A_c can be expressed as follows:

$$A_c = A - BKC, \tag{2.10}$$

and for initial conditions must pay, which is represented by:

$$Z = E\{x(0)x(0)^{\mathrm{T}}\}, \tag{2.11}$$

where Z is a symmetric matrix. Although the exact initial state is unknown, it is assumed that it is nonzero and has the suspected Euclidean norm. An exact derivation of this problem can be found in [26].

As previously mentioned, various computer programs are often used to solve this problem. One of the possible tools is a Toolbox from A. Lika [14], which can be used in programs Matlab or Octave. The key considerations during the development process of this Toolbox were simplicity, adjustability, and user-friendly interface. This toolbox contains numerous types of output-feedback controllers. The used toolbox function is called `oflqr` [14].

2.3 Identification

Identification is a process that declares the properties of a system. It can be used to construct process model structures and determine parameter values as well [21].

System identification involves establishing the relationships between input and output signals that describe a process, resulting in input-output models [17]. There are three basic categories of identification approaches:

1. Active or passive experiment-based identification: In active identification, input signals are intentionally generated for the purpose of identification, while in passive identification, the process is identified using data from normal operating conditions.
2. Deterministic and stochastic methods: Deterministic methods assume precise knowledge of inputs and outputs, while stochastic methods involve numerical solutions such as the least squares method or the Bayesian approach.
3. One-time or continuous processing: identification methods can be classified based on how the data are processed, either one-time or continuously over time [17].

For more detailed information on specific models, refer to the Literature [17].

Process identification can be performed in a steady state, where one of the input variables is changed while the others remain constant. By introducing a step change, we can obtain the step response of the process. The analysis of the step response is classified as a deterministic method, as it provides valuable information about the approximate gain, dominant time constant, and delay of the system [29, 17].

In the process of system identification, it is often beneficial to consider the deviation form of variables. This data processing technique involves deducing steady-state values from the measured variables and obtaining their deviations [17]. The deviation form of e.g. input data can be expressed as:

$$\tilde{x}_i = x_i - x_i^s, \tag{2.12}$$

where $\tilde{x}$ represents the matrix of deviating input variables, x_i denotes the matrix of measured input variables, and x_i^s represents the steady-state value of x_i [17].

Utilizing the deviation form can effectively highlight dynamic changes in the data, leading to improved sensitivity and more accurate identification. By transforming variables into a deviation form and comparing them, it ensures that the data are within a similar range, facilitating the identification process [17, 21].

During the system identification process of a distillation column, it is possible to make certain simplifications that greatly impact the analysis. The number of trays in the distillation column has a significant impact on the system order. The hold-up of liquid on trays causes an increase in the system order. By allowing the liquid to flow downward from tray to tray, changes made in the column can appear after a certain amount of time. However, this phenomenon is also influenced by the type and size of the distillation column, as well as the number of trays and the height between the trays. Therefore, for simplicity, such a system could be approximated by o times first-order exponential lags with approximately the same time constant. Then it can be described as [29].:

$$G(s) = \frac{K}{(Ts+1)^o}, \tag{2.13}$$

where K represents the gain, T denotes the time constant, and o represents the order of the system [29].

The order for the distillation column with 20 trays should be equal to 20. Therefore, such considerations will simplify the control design [29].

System identification can be carried out in either an open loop or closed loop configuration. In open-loop identification, the process is operated without a controller or with the controller in manual mode. By manipulating the input variable, a disturbance is introduced into the system. In the case of the closed loop, the controller is already connected and ensures a stable process. In this case, the change of MV occurs by changing the controller. First of all, the changes made must be agreed upon with the operator. It is advisable to avoid any changes in the operation mode. The size and length of the steps must be specified. The bigger the steps, the better, but care must be taken for safety. During the performance of step changes, no other changes should be made to the system as it may generate unwanted errors. During such a process it is advisable to collect all data, not only MV and PV, it is also advisable to take laboratory samples that can tell you about the scatter. It is important to be careful to avoid correlated steps. This problem should be solved by making changes of different sizes and durations. Identification will be most reliable if we are in a steady state after the steps have been performed. This is not a requirement and it is sufficient if the steady state is reached at least after some steps. The presence of hysteresis can also be a problem. Hysteresis is usually caused by wear and tear on the equipment. All regulators must be in manual mode. This can be problematic for unstable processes. It is necessary in these cases to make such a large step that it is visible on the output. If there is also noise, it will be difficult to determine. It is recommended that the change in MV should be at least 5 times the amplitude of the noise. Such changes can also cause complications in the process [21].

With the assistance of System Identification Toolbox [32] in the MATLAB environment even reasonably complicated processes, including those with higher orders, can be evaluated simply. This toolbox can analyse process data in a wide range of forms, including the transfer function and the state space model. It expands dynamic relationships between observed variables and builds models in either continuous or discrete time that might be based on time or frequency domain data [32].

2.4 Inferential Sensors

Physical sensing devices are often insufficient, inaccurate and unreliable. The alternative solution represents the inferential or software (soft) sensors. They provide measurements from the online sensors (online measurements) making it easier to work with the equipment. These sensors provide real-time estimates of key process variables based on their correlation with other process measurements, making them essential for industrial process control [8]. Inferential sensors are online devices built

on mathematical concepts. Based on modelling and identification methodologies, they provide estimations of the process state and product parameters [5]. The application of inferential sensors is to make the best forecast possible by employing a collection of process predictor variables and to predict important variables. These mathematical models support process behaviour diagnosis as well as monitoring, analysing, and prediction of product quality [41].

The advantage of inferential sensors is therefore that we can measure quantities that are hard-to-measure or completely unmeasurable in a given process. These quantities are calculated based on other measurable quantities that are available [37]. These methods are often applied as prognostic tools. For example, they have also found their use in macroeconomics. They also have the advantage of direct interpretation of variables that contribute to parameter estimation, they can be used without strict limits, and adjustment of penalty values is possible [7]. A further advantage of inferential sensors is that they also ensure higher safety in the industrial plant [8]. Although they also have their disadvantages. The accuracy and reliability of industrial measurements have a great impact on the resulting designed sensors [37]. Moreover, the measurements from all input (independent) variables are required for the estimation of the output variable from the inferential sensor. Furthermore, data-driven inferential sensors often assume that the process is in a steady state but this is not guaranteed in the industrial dataset [8].

Inferential sensor design procedure The other basic task of soft sensors is to choose the correct structure, which consists in selecting the variables. The sensor structure is characterized by the inputs to the inference sensor. These inputs are a combination of online measured variables [37, 41].

Based on the structure and modelling process, the design of inferential sensors can be divided into three main groups: (1) model-based inferential sensors [30], (2) data-driven inferential sensors [37] and (3) hybrid inferential sensors [47]. In most cases, the model-based inferential sensors use a first-principles model [17, 30]. The model-based inferential sensors require some prior knowledge about the process dynamics based on equations from physics or chemistry. Due to the complexity of industrial systems, it is often too difficult or impossible to design the inferential sensor based on the first-principles models with sufficient accuracy and reliability [30, 37]. Models based on data-driven approaches are cheap and less time-consuming to design. They do not require precise knowledge of the mathematical model of the process, so they are a suitable alternative to model-based inferential sensors [37]. Hybrid models offer a flexible approach to developing soft sensors by combining both the process knowledge

based on first-principle understanding and the data-driven techniques [47].

The aim is to find a nominally optimal model J, which has a general form (described by ℓ_2 norm) [9, 24]:

$$J = \left\| y - \hat{y} \right\|_2^2, \tag{2.14}$$

where y is the measured output and $\hat{y}$ is the estimated output. The mathematical representation of $\hat{y}$ is as follows:

$$\hat{y} = M(x_1, x_2, \ldots, x_\mathrm{n}), \tag{2.15}$$

where x refers to a set of input variables and n is a specific number of input variables in the equation. The result is a model (J) that predicts with minimum deviations from the process outputs [9].

A model defining the dependence between the input data (X) and output data (y) is able to approximate this relationship as

$$y = f(X) + \varepsilon, \tag{2.16}$$

where f is a function that represents the mathematical relationship between inputs and outputs and $\varepsilon \in N(0, \sigma)$ is a residual error of the estimation which is normally distributed with mean value zero and a standard deviation σ [5, 24].

We can also rewrite (2.16) into the form of a multivariate linear model [24]:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \ldots + \beta_\mathrm{n} x_\mathrm{n} + \varepsilon = \beta_0 + \sum_{i=1}^{\mathrm{n}} \beta_\mathrm{i} x_\mathrm{i} + \varepsilon, \quad \forall i = \{1, \ldots, \mathrm{n}\}, \tag{2.17}$$

where β is a vector of estimated model parameters within the inferential sensor.

The model is developed and trained using historical data [5], such as

$$D^{\mathrm{Hist}} = (X^{\mathrm{Hist}}, y^{\mathrm{Hist}}), \tag{2.18}$$

where D^{Hist} is the historical domain containing the input and output data set.

The formula for estimating output at each time (t) is as follows [5]:

$$\hat{y}_t = f(X_\mathrm{t}^\mathrm{f}), \tag{2.19}$$

where X_t^f is a matrix of the input (candidate) variables into the model structure [5].

Hybrid inferential sensors can be described as [24]:

$$\hat{y} = M(x_1, x_2, \ldots, x_\mathrm{n}) + \sum_{\mathrm{i}=1}^{\mathrm{n}} \beta_\mathrm{i} x_\mathrm{i}, \tag{2.20}$$

where the first part of the equation is from the model-based inferential sensor and the second part is from the data-driven inferential sensor [24].

2.4.1 Data processing

Preparation of the dataset is a critical and essential stage, with the primary purpose of obtaining final data sets that may be considered valid for further analysis [10]. The initial and crucial step of the mathematical model design is to obtain data. For analysis and design, it is necessary to perform an experiment or collect historical data from the plant. If the measurement was done manually and only analogue data is available, the data needs to be digitized [31]. The quality of the data used for the design of mathematical models has a major effect on the efficiency and reliability of the further analysis, such as the inferential sensor design [37]. Process data are usually correlated in industrial settings because of energy and material balances as well as control loops implemented on the plants [41].

Nearly all theoretical approaches presume that the data is adequately sampled and free of noise. However, this assumption does not hold for data collected through experimentation, as it is frequently correlated, confounded, and may contain missing values. These factors can affect the interaction between input and output [10]. The literature [10] suggests three fundamental strategies for reducing the effect of these issues. The first is the use of classifiers to generate robust learners. If this technique is not achievable, data polishing can be used instead. However, both of these strategies are time-consuming and are usually only applied to smaller data sets. The most extensive method is to use a noise filter. They can detect and eliminate noise to some degree, but it is possible that a filter learned on one data set will not be valid for another, leading to data corruption [10]. The filter produces lag and the increase in the order of the system leads to an increase in the perceived deadtime [21].

Industrial data is typically collinear, resulting in data with redundant information. The volume of data from many sources might be the cause of the collinearity. In models where a small change in the input creates large variations in the output, collinearity can induce instability. These phenomena may cause difficulties in prediction models, such as the design of inferential sensors [11]. To address this issue, simpler models are preferred and often the number of variables is reduced by removing irrelevant or

redundant variables. When two predictors exhibit high correlation, it means they are evaluating the same underlying information. Removing one should not negatively impact the efficiency of the model. This approach results in a smaller data set that still effectively describes the overall system [25, 10]. Dimension reduction is another technique in which data is reduced from a higher dimension to a lower dimension while the properties of the system are kept. Understanding the process can also help to solve the collinearity problem [11]. The advantages include reduced correlation, lower computational effort, decreased data storage requirements, and increased time and economic efficiency since fewer data are needed to be measured [25, 10].

The centering of variables used for prediction is one of the most often applied data processing techniques. The average value of the independent variable is subtracted from each data point in this stage. As a result, the predictor has a zero mean. Data scaling is a similar strategy in which the variable to be predicted is divided by the standard deviation, resulting in a common standard deviation of one. This method offers the benefit of improving data processing quality by centering the data at a specific distance from the median value, which facilitates faster and more accurate evaluation. The disadvantage is that the data may be interpreted incorrectly since they are no longer in the same units [25].

One feasible approach to data processing involves standardization. The standardization ensures that all variables have zero mean and unit standard deviation. It may be that data that have a higher variance affects the result more [33, 18]. The variance σ_x^2 of n measurements in the sample can be calculated as

$$\sigma_x^2 = \frac{\sum(x_i - \mu_x)^2}{n-1}, \tag{2.21}$$

where x is measurement and μ_x is mean of measurement [33].

For highly variable data, the variance will be large, whereas, for less variable data, the variance will be small [33]. The standardisation should ensure that everything is on the same scale [18, 33]. The standardised variable x_{std} of variable x can be described as:

$$x_{\text{std}} = \frac{x - \mu_x}{\sigma_x}, \tag{2.22}$$

where μ_x is an arithmetic mean and σ_x is standard deviation. For this case, the data means is equal to 0 and after dataset standardization the standard deviations are equal to one [33, 18].

In general, the industrial dataset is corrupted by systematic errors (physically infeasible measurements) and outliers, i.e., significantly deviated measurements from the typical

trend of the dataset. These outliers can be caused by instrument errors or various process or transmission failures. Such data can cause complications either in system identification or in the design of mathematical models. These data generally fall into two categories. The first is the obvious outliers, which represent physically inadmissible conditions and such values are easy to identify. Often a data limiter is used to explicitly tell which data are outliers. Another category is non-obvious outliers. They may be realisable but deviate from typical behaviour. Given the correlation of the data, it may be difficult to determine such values. Their recognition can be difficult and is very specific to each process [11].

The data-driven methods (see Section 2.4) usually consider the training and testing data. The training data is used to design the mathematical models and the performance of the designed models is analysed on the testing data set. There are several effective methods for splitting given data into training and testing sets. If a big amount of data is provided, it is suitable to integrate the collected data and apply a random distribution. In practice, typically 70-80,% of data is utilised for training and the remaining 20-30,% is used for testing. In cases where a large quantity of data is available, it is often possible to separate the data into three datasets, namely training, validation, and testing. The training set is used to learn parameters within the model structure, the validation set is used to produce the error estimate, and the test set discusses the model accuracy on previously unknown data. In this strategy, half of the data is typically utilised for training, a quarter for validation, and a quarter for testing. Therefore, the way of the distribution of the data into the training and testing datasets depends on the quantity and quality of the data, the complexity of the process and the model, and there is no general rule for the distribution. Adequate data is necessary to train mathematical models, but overfitting could lead to potential errors during the prediction [13]. It can result in poor performance when applied to new datasets, thus it is important to assess how well the model fits the training data [25].

2.4.2 Regression Methods

Regression methods are one of the most preferred due to their simplicity and ease of application [37]. Regression analysis is used in statistical modelling to explain correlations between an output variable (y) and a matrix of inputs (X). The most common type of regression analysis is linear regression [43]. The method works by minimising the sum of squared errors between measured data and the predicted data from inferential sensors [37] as follows:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + ... + \beta_P x_{P,i} + \varepsilon_i, \quad \forall i \in \{1, 2, \ldots, n\}, \tag{2.23}$$

where β_0 is the estimated intercept, β_j is the estimated coefficient for the j^{th} predictor, x_{ij} is the value of the j^{th} predictor for the i^{th} sample, P is the number of predictors or also known as the order of the model.

The linear model can be generalized into the form described as

$$y = X\beta + \varepsilon, \tag{2.24}$$

where

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{\text{n}} \end{pmatrix}, \tag{2.25}$$

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1\text{n}} \\ 1 & x_{21} & \dots & x_{2\text{n}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{\text{i}1} & \vdots & x_{\text{in}} \end{pmatrix}, \tag{2.26}$$

where β is an $1 \times (n+1)$ vector of unknown regression parameters and ε is an $(i \times 1)$ [43].

Additionally, these models allow for the computation of standard errors of the coefficients, provided that certain assumptions about the distribution of the model residuals are made [25]. This section is focused on the two most frequently used and popular methods for data-driven models, ordinary least squares regression (OLSR) and least absolute shrinkage and selection operator (LASSO) method.

Ordinary Least Squares Regression The ordinary least squares regression (OLSR) method is considering the linear model, for which are given input variables and p-predictors [43]. The OLSR approach solves the optimization problem where the aim is to derive the model parameters β as follows [43]:

$$\min_{\beta} \sum_{\text{i}=1}^{\text{N}} (y_{\text{i}} - \hat{y}_{\text{i}})^2. \tag{2.27}$$

The optimisation problem is defined as the minimisation of the sum of squares deviations between the measured (y_i) and predicted ($\hat{y_i}$) response [24, 25]. From (2.27), it is possible to derive the explicit solution for the optimal values of model parameters β^* as follows [49, 25, 24, 37]:

$$\beta^* = (X^{\text{T}}X)^{-1}X^{\text{T}}y. \tag{2.28}$$

As previously mentioned, linear regression is a popular modelling strategy due to the interpretability of its coefficients. However, these same characteristics that make the model interpretable can also lead to errors [25]. Multicollinearity, a circumstance where there is linear dependence among the predictors, causes a significant decrease in the OLSR performance [43]. Specifically, the $(X^{\mathrm{T}}X)^{-1}$ term in (2.28) is proportional to the covariance matrix of the predictors. If there are no predictors that can be inferred from a combination of other predictors and the number of samples is greater than the number of predictors, then a unique inverse of this matrix exists. However, if the data do not fulfil one of the criteria, there is no unique set of regression coefficients. If the data match the first requirement, a unique set of predicted values can be obtained by either replacing the conditional with the inverse or removing collinear predictors. In the case of data collinearity, linear regression can still be employed for prediction. However, because the regression coefficients used to determine these predictors are not unique, the ability to interpret the coefficients meaningfully is limited. If the second condition is met, it is possible to build a regression model after data pre-processing. In cases where the number of inputs still exceeds the number of observations even after data pre-processing (which may involve the exclusion of some variables due to collinearity), parameter estimate shrinking techniques may be employed. This includes methods such as ridge regression, LASSO or elastic net [25].

Least Absolute Shrinkage and Selection Operator Allowing the parameter estimations to be biased might result in a lower mean squared error (MSE). As bias increases, variance decreases significantly, resulting in a lower MSE than the coefficients of OLSR [25]. The least absolute shrinkage and selection Operator (LASSO) method is one of the shrinkage methods, which is searching for the best input variables to explain the predicted process value [49]. This type of biased model can be used to solve collinearity and offer competitive overall MSE [25]. Adding a penalty to the sum of the squared errors is one way to produce biased regression models. In case of overfitting or collinearity, calculation errors may occur when using OLSR. It is therefore advisable to control the magnitude of the estimation of these parameters. Such control can be achieved by adding a penalty. We can create a compromise between the variance of the model and bias by introducing a penalty. By accepting some bias, we can commonly minimize the variance to a point where the total MSE is lower than that of unbiased models [25].

The LASSO method is using ℓ_1 norm penalty. This is used to decide whether to put more weight on the exactness or the complexity of the model [37]. The mathematical

representation of LASSO is following:

$$\min_{\beta} \sum_{\mathrm{i}=1}^{\mathrm{N}} (y_{\mathrm{i}} - \hat{y}_{\mathrm{i}})^2 + \lambda \sum_{\mathrm{j}=1}^{\mathrm{p}} |\beta_{\mathrm{j}}|, \tag{2.29}$$

where λ is a penalization element [37].

If $\lambda = 0$, then the LASSO method estimates the same parameters β as the OLSR method. The amount of shrinkage depends on the parameter λ (tuning parameter), which expressed the penalty of the model[49]. This approach necessitates adjusting the parameter λ. This method mostly uses cross-validation. The cross-validation errors are computed for each value of λ, which was chosen. Then is chosen the tuning parameter value with the least error. The model is then retrained using the chosen λ on all training data. When there is collinearity in the data, the LASSO method frequently provides seemingly distinct sets of chosen variables with just modest alterations to the training set [41].

2.4.3 Performance Indicator

When designing inferential sensors, it is possible that to predict the same quantity, two different models will use different inputs. To compare the performance of considered models, the root-mean-square error (RMSE) can be evaluated [41]. This performance indicator or accuracy criterion indicates the deviation of the model estimates from the actual measured data. As mentioned in Section 2.4.1, the quality of the data can be affected by various aspects, which also impacts on the accuracy of the designed inferential sensors. The measured data (y_{m}) are typically affected by errors from experiments, noise, and other factors, as compared to theoretically perfect data. As a result, the true data is often unknown. Nonetheless, estimated data (y_{e}) can be calculated to fit the measured data, with the objective of minimizing the difference between y_{e} and y_{m} as much as possible. The residuals (or errors) R, are the differences between the measured y_{m} and estimated y_{e} data [31]. This can be mathematically written as follows:

$$R = y_{\mathrm{m}} - y_{\mathrm{e}}, \tag{2.30}$$

where y_{e} is a function of the model and parameters [31].

Through fitting, a set of parameters can be obtained that provides the most accurate estimate of the true data by minimizing the difference between the estimated data (y_{e}) and the measured data (y_{m}). As a consequence, it is possible to calculate the root

mean squared error (RMSE) of the elements of the matrix R [31]:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{\mathrm{i}=1}^{n} {R_{\mathrm{i}}}^2}{n}}. \tag{2.31}$$

In the process of model design, the aim is to identify the set of parameters for a given model that gives the lowest possible RMSE value. The process of fitting them becomes simpler the fewer parameters there are [31].

Chapter 3

Experimental Results

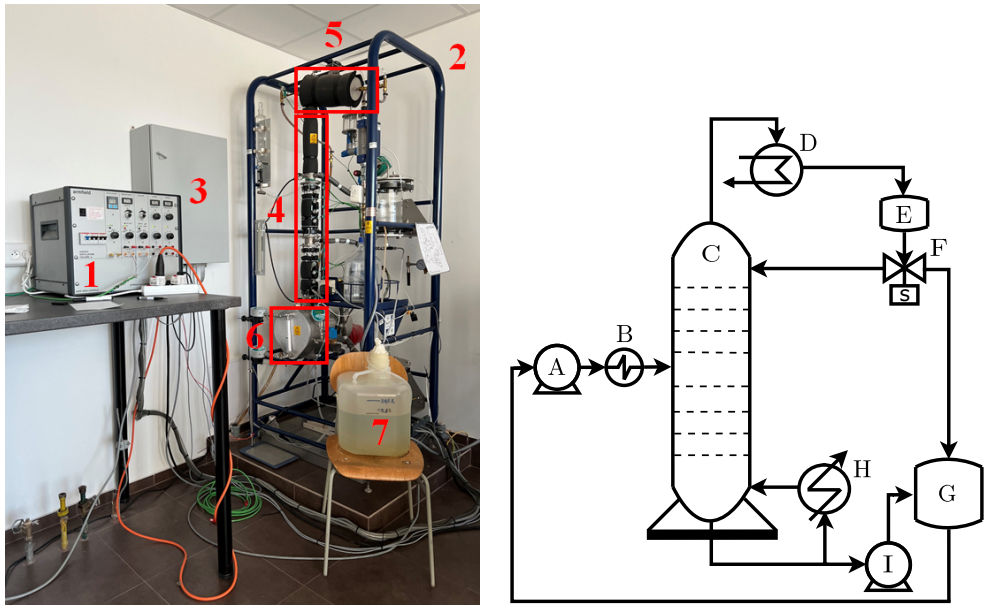
3.1 Laboratory Distillation Column Device

The thesis studies the laboratory distillation column Armfield UOP3CC. This device is situated at the Department of Information Engineering and Process Control, Faculty of Chemical and Food Technology, Slovak University of Technology in Bratislava. It can be operated as a continuous or batch distillation column. Operators can also choose between packed and plate versions, while the plate type is available at our institute [27]. It is a freestanding distillation plant, which has two interconnected parts. It is made up of a floor-standing process unit and a bench-mounted control console. This process is a computer-linked system, which means, that it can be externally controlled by a computer. Above all, it is designed for training to give the operators theoretical and practical knowledge of the distillation process. For example, students can get a better idea of how the separation process works and can directly see the impact of distillation column management. It allows researchers to try out scientific ideas and test theory but also train industry personnel on how to use real industrial equipment, such as plant operators [28].

The column is equipped with a preheater for the feed and the maximal power of the heat exchanger is 2 kW. The feed pump injects the feed into the distillation column. The pump dosing speed is 50-200 ml/min. The distillation column has eight trays with sieve plates. We number the trays within the distillation from the top, so on the head of the distillation column is the first stage and at the bottom is the eighth stage. Every tray has sample and feed ports. The column consists of two glass parts. Both of them have four trays. All trays are placed on one rod. The feed mixture is fed to the fifth tray from above. The individual trays are positioned one above the other, vertically, and aligned for the counter-current flow of the liquid and vapour. Each tray has a weir and downcomer, while plates have a 5 cm diameter and between successive trays is 5.5 cm space, except between trays 4 and 5. It is considered, that every tray has a weir and a downcomer and each plate have 122 holes with 1 mm in diameter. The height of

the extruding edge of the downcomer is 0.3 cm. This causes the liquid to be detained in each tray. The entire glass part is isolated to prevent heat loss. Each of the trays has a sensor for measuring the temperature. The approved maximum temperature on trays is 150° C. Each sheath of the plate temperature sensor has a maximum of 1.5 mm diameter to provide a quick dynamic response. The vacuum system allows for operation at pressures as low as 200 mbar. The column is equipped with a U-tube manometer, which serves to measure pressure drop over the distillation column. In the distillation column, the vapour phase flows upwards and the liquid phase flows downwards. From the column head, the vapour passes into the condenser, which is water-cooled and is a coil-in-shell type. When we performed step changes in the coolant flow rate, we found that this did not affect the condenser itself, because of its size. We have come to the observation that on the distillation column considered by us, there is an oversized condenser and we can not control this part of the plant. The coolant flow is controlled by a diaphragm valve. A parallel water reservoir that circulates cooling water cools the condenser. The column is equipped with a rotameter to measure and handle the cooling water flow rate. The mixture flows from the condenser into the decanter (phase separator), which is made of glass. It then flows on through a three-way solenoid valve. This valve divides how much liquid will be sent into the column, as reflux to the head of the distillation column, and how much will be removed. When it drains away, it is collected in a container for the distillate. The solenoid valve operates on a pulse-width modulation principle. The reflux control ratio can be controlled in the interval 0 % to 100 %. All of the condensates will flow to the column if the valve has the value 100 % or the timer is stopped. This occurrence is called total reflux. At the bottom of the distillation column, there is a reboiler, which is isolated. The reboiler has a flame-proof immersion-type heating element and is constructed of stainless steel. Inside the reboiler, a level sensor is located that monitors the level in the reboiler, preventing the heating spiral from overheating. There is electrical resistance in the reboiler which ensures the evaporation of the liquid mixture. The size of the reboiler is 21 l and the heating element provides heating up to 2.5 kW, while the ideal range for heating is between 1–2.5 kW. The outgoing liquid from the reboiler is cooled with the bottom product cooler and then introduced into the collection tank for the bottom product [27, 23, 38, 3].

The studied laboratory distillation column is illustrated in Figure 3.1. Part (a) displays a real plant with the following important components: 1. bench-mounted control console, 2. freestanding distillation machine, 3. wall-mounted switchboard, 4. column section, 5. condenser, 6. reboiler, 7. product collection tank. Part (b) shows the corresponding piping and instrumentation diagram (P&ID), where (A) is the feed pump, (B) is the feed heat-exchanger, (C) is the distillation column, (D) is the condenser, (E) is a tank for the condensed liquid, (F) is the reflux solenoid valve, (G) is the product



(a) Real plant. **(b)** P&ID flowchart.

Figure 3.1: Laboratory distillation column Armfield UOP3CC with important parts.

collection tank, (H) is the reboiler and (I) is the pump for the bottom products.

During the separation process, the selected cold feed is dispensed using a feed pump (A), which pumps it further to the preheater (B) to be reheated to the desired temperature. Then, the feed flows into the distillation column (C) for the feed tray. The vapours in the column flow upwards and pass into the condenser (D). Here, the vapours are condensed using a heat exchanger and then pass into a collecting tank (E). A part of this mixture is gradually fed back into the column, and a part is removed as the final product, using a solenoid valve (F) for separation. The withdrawn product is collected in a product collection tank (G), which also collects the bottoms. Simultaneously, the fresh mixture is used as the feed into the distillation column. If the process is operating correctly, equilibrium occurs, and the container will not be emptied. In this way, it is not possible for the feed pump to pump air when turned on. The liquid that flows downward in the distillation column passes into the reboiler (H), where the vapours are generated. A part of the liquid is discharged from the reboiler using a waste pump (I) as the bottom product. Finally, the feed from the collector tank (G) is recycled back to the column by the feed pump (A). After the experiments are completed, the collection tank (G) is closed to prevent the mixture from evaporating.

The same mixture can be used for further experiments.

A precise description of the working procedure with the Armfield UOP3CC laboratory distillation column can be found in Appendix B in Section B.1.

As previously mentioned, the distillation column can be controlled through MATLAB. Within this environment, the basic Human Machine Interface (HMI) scheme is created to ensure the interface between the operator and laboratory distillation column UOP3CC. This interface allows for observation and control of the various parts of the distillation column. Initially, 19 variables could be monitored on the device, including temperature sensors (14 pieces), pressure sensors (3 pieces), as well as data about the heat provided to the reboiler and the feed flow rate. At present, the device provides 39 variables, nearly double the original amount. A table of these variables can be found in the Appendix, in Table B.1. The basic HMI has been expanded to include several new elements that help to evaluate the current situation during experiments. First and foremost, graphs are provided to help in immediate analysis. Individual control loops are designed and implemented in a tunable fashion to make the HMI scheme more approachable to other users or operators. Data is now saved throughout the entire experiment, not just after the shutdown of the column. Moreover, the HMI scheme is enhanced by the models of designed inferential sensors (e.g., steady state indication, product concentrations and others) significantly increasing the overall quality of the process monitoring during the experiments.

The distillation unit has a data acquisition that makes it possible to do the online monitoring of the temperatures, control of the feed pump rotation, reflux flow rate and reboiler heat duty [3]. The feed can be injected at the base, centre or top of the column. There are 14 thermocouples at the plant, 8 of which are located on individual trays and the others in other strategic locations or can be positioned as required. However, it is necessary to mention that the inferential sensor on the 3rd tray is damaged, and therefore we do not have information about the temperature from this stage. Additionally, this distillation column is unique in that it allows the use of combustible solvents like acetone, benzene, or p-xylene [27, 38]. However, during the experiments performed for this work, the binary methanol-water mixture is only separated.

3.2 Design of Inferential Sensors

In this part of the thesis, the design of nine inferential sensors for different parts of the laboratory distillation column UOP3CC is presented. These mathematical models

have been developed to indicate quantities that we are not able to measure directly. The set of indicated variables involves the flow rate of feed and bottom product, the liquid level in the reboiler based on model-based and data-driven approaches, the concentration of products with a refractometer, the composition of feed, distillate and bottom streams, and steady-state indication. The designed inferential sensors are divided according to their use and characteristics and presented in further sections of the practical part of the thesis. The design methods for each sensor are described in the specific sections. The advantage of these designs is that the techniques used are not limited to this particular distillation column, but can be used for other devices as well. After the design, each inferential sensor is implemented into the HMI scheme, communicating with the distillation column. Subsequently, the performance of each designed inferential sensor is evaluated based on the experiment data from the last experiment and also from the previous experiments (if available).

More than 10 experiments were performed on this distillation column and during these experiments, data were collected. The focus of these experiments varies, therefore the length of the experiments varies accordingly. In Section 2.4.1 was mentioned, that the data source has a large impact on the quality of the inferential sensors. Considering this fact, the experiments with the longest duration are selected from all those performed. In this way, we also tried to improve the design of the sensors and increase their prediction accuracy. When designing models for product concentration, we use well-known and popular regression methods such as ordinary least squares regression (OLSR) or least absolute shrinkage and selection operator (LASSO). These methods allow us to predict the variables due to the partitioning of the data into training and testing sets. Once the output variables are estimated, the RMSE of the testing set is determined and then the accuracy of the model can be evaluated. In order to monitor key process variables, the designed inferential sensors are implemented into the HMI scheme communicating with the UOP3CC distillation column from Armfield.

3.2.1 Design of an Inferential Sensor for Flow rate Indication

We design the inferential sensor (calibration curve) to indicate the flow rate of two pumps. The first one introduces the feed into the distillation column and the second one removes the bottom product (waste) from the reboiler. The first step to obtaining inferential sensors is to measure the efficiency of the pumps. As previously mentioned in Section 3.1, one of the measured variables within the original dataset from the UOP3CC distillation column is the feed flow rate. Nevertheless, the feed pump (which monitors the feed flow rate) must be recalibrated due to biased indications. The pump for the bottom product has been mounted additionally on the device.

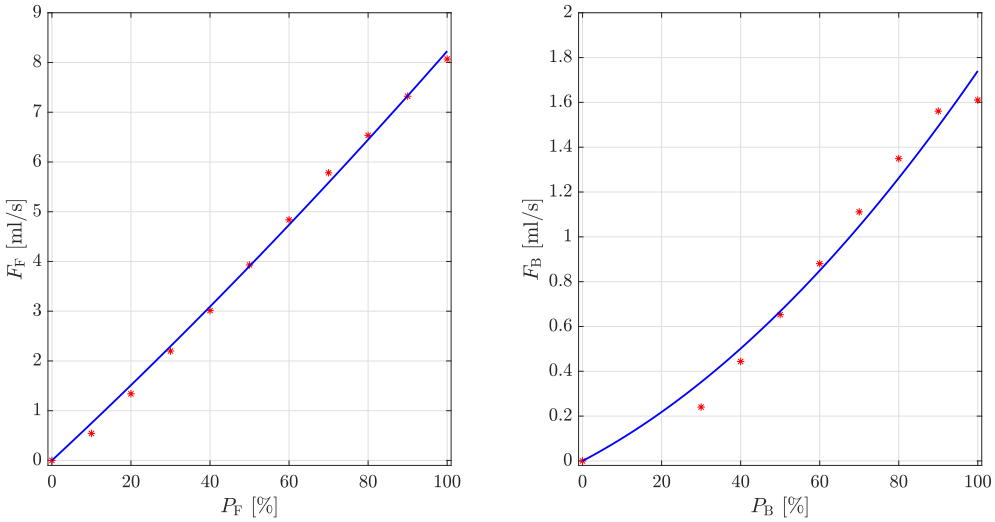


Figure 3.2: Comparison of measured data (red points) with the resulting model (blue trajectory). On the right is the model for the feed pump, and on the left, is the pump for the bottom product.

To measure the flow rates, it is necessary to measure the time during which the pump withdraws a certain amount of liquid at different pump capacities. Due to the fact that the mixture of methanol-water and pure water have relatively equal densities, the experiment is performed with pure water as follows:

1. In the first place it is important to select the amount of water we want to pump. Due to the fact that the pumps have different capacities, we use 100 ml in the case of the feed pump and 20 ml in the case of the bottom product pump.
2. In the next step, we measure with a timer how long it takes for the pumps to pump the selected quantities of water at different capacities (powers %). Due to the size of the pump for the bottom product, we start at the 30 % capacity of this pump as it is not able to work efficiently at lower speeds.
3. We repeat it several times and the averaged value from the corresponding measurements is calculated.
4. We use the regression method to fit the data, which falls under the OLSR method.

The measured data with the resulting fitted model (calibration curve) is illustrated in Figure 3.2, where on the left-hand side is the design for the feed pump and on the

right-hand side is the design for the bottom product pump. The power of pumps (P_F and P_B) is shown on the x-axis and the flow rate of liquids (F_F and F_B) is shown on the y-axis. We can directly see that the flow rate of the bottom product is smaller compared to the flow rate of the feed. This is a reasonable observation considering the overall mass balance of the distillation column. On the left-hand side, we can see that during experiments, we can pump a maximum of 8.0 ml/s with the feed pump but only a maximum of 1.6 ml/s with the bottom product pump.

Feed Pump The resulting mathematical model for the flow rate of feed (F_F) has the following form:

$$F_\mathrm{F} = 0.0001P_\mathrm{F}^2 + 0.0740P_\mathrm{F}, \tag{3.1}$$

where P_F is the feed pump power.

Bottom Pump The final sensor for the bottom product flow rate (F_B) reads as follows:

$$F_B = 0.0001P_\mathrm{B}^2 + 0.0093P_\mathrm{B}, \tag{3.2}$$

where P_B is the bottom (waste) pump power.

It can be seen that the two inferential sensors differ only in the second parameter. The small coefficients of the quadratic terms arise from the fact that the data used for the design are almost linear. In both cases, the parameter values are rounded to four decimal places.

3.2.2 Design of an Inferential Sensor for the Liquid Level Indication

One of the desired quantities in the monitoring of the distillation column is the liquid level in the reboiler. We use two different approaches to achieve the monitoring of this quantity. The main difference between the studied approaches is that the first one considers a model-based inferential sensor, while the second one considers a hybrid inferential sensor.

Model-based Approach To measure the liquid level in the reboiler, we first developed a model which is designed based on hydrostatic pressure. In this case, we know the model and can directly access the resulting variable using the equation:

$$p = h\rho g, \tag{3.3}$$

where h is the liquid level in the reboiler, ρ is the density of the fluid and g gravitational acceleration equal to $9.81\,\mathrm{m/s^2}$ [40].

From this, we can express the liquid level height in the reboiler as:

$$h_{\mathrm{MB}} = \frac{\Delta p}{\rho g} = \frac{p_{\mathrm{Reb,T}} - p_{\mathrm{Reb,B}}}{100\rho(T_{\mathrm{reb}})g}, \tag{3.4}$$

when h_{MB} is the liquid level calculated based on the inferential sensor designed by the model-based approach, Δp is the difference between the pressure on the top of the reboiler $p_{\mathrm{Reb,T}}$ and on the bottom of the reboiler $T_{\mathrm{Reb,B}}$, function $\rho(T_{\mathrm{reb}})$ expresses the density of the liquid in the reboiler as a function of the temperature as follows:

$$\rho(T_{\mathrm{reb}}) = \rho_{H_2O}(1 - k(T_{\mathrm{Reb}} - 4)^2) = 999.973(1 - 6.63 \cdot 10^{-6}(T_{\mathrm{Reb}} - 4)^2), \tag{3.5}$$

where $\rho_{\mathrm{H_2O}}$ is the density of the water at 4° C ($999.973\ \mathrm{kg/m^3}$) and k is a constant based on Markofsky-Harleman.

The (3.5) is obtained from the literature [48]. After that, this model has been fitted using linear regression.

Hybrid Approach In this case, we use OLSR and LASSO methods (described in Section 2.4.2). In Figure 3.3 can be seen a comparison of the training and testing of IS with two different methods. On the x-axis can be seen the measurement index and on the y-axis the liquid level in reboiler (h_{Reb}). Manually measured data is recorded by taking readings of the liquid level in the reboiler on the ruler that is placed on the side of the reboiler. The measurements are taken randomly during the experiment.

On the upper graph can be seen the training of the inferential sensor. It is interesting to note that while OLSR used all the variables provided for the method, LASSO only selected just some of them for the model. Both models use information from the previous model-based approach, which makes them hybrid. We can see, that during the training the results are quite similar. However, based on the RMSE criterion the OLSR is better. This is because the OLSR method attempts to use all available variables to provide a perfect description of the data. The RMSE value for the OLSR is 0.5966 and for the LASSO 0.8680. During the training of the IS, we used 58 measurements.

The lower graph shows the testing of the inferential sensor. During the testing 28 measurements are used. The graph shows that during the testing of IS OLSR tries to fit more points, and the models result in greater peaks. The RMSE of OLSR is 4.5648 and for the LASSO is 2.6256.

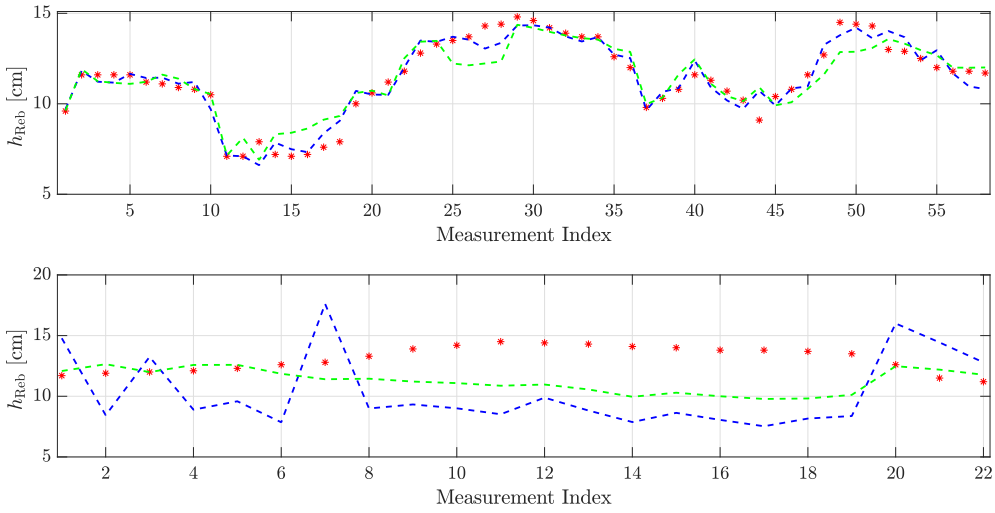


Figure 3.3: Training and testing of IS for liquid level in reboiler with methods OLSR (blue dashed trajectory) and LASSO (green dashed trajectory) on measured data (red points).

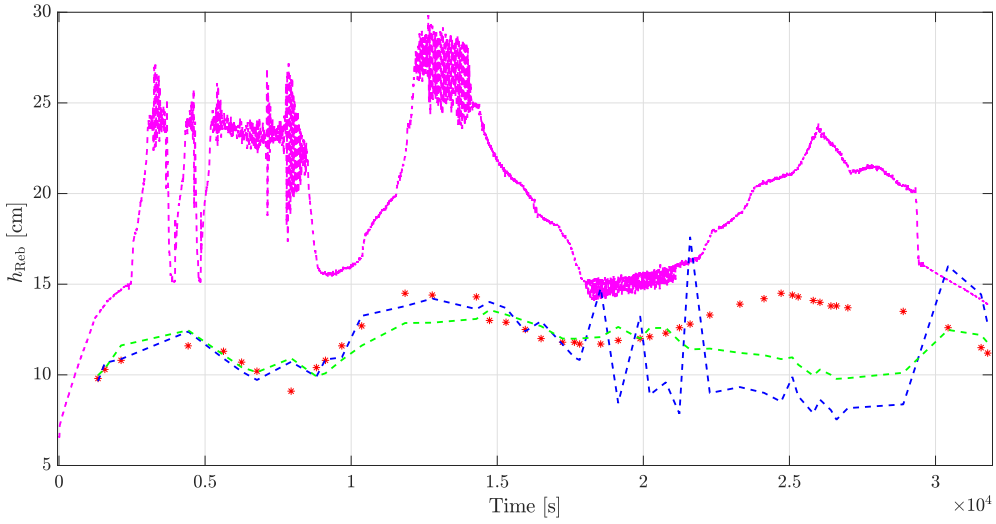


Figure 3.4: Testing of IS for liquid level in reboiler with methods OLSR (blue dashed trajectory), LASSO (green dashed trajectory) and hydrostatic pressure model (magenta dashed trajectory) on measured data (red points).

In Figure 3.4 we can see the comparison of the derived inferential sensors with manually measured data from one of the performed experiments with 44 measurements. The graph shows the time dependence of the liquid level in the reboiler. When comparing the models with manually measured level data, it can be seen that the inferential sensor designed with a hybrid approach more accurately describes the experimental data. This is confirmed by the root mean square error values. In the case of a model-based approach sensor, the RMSE value is approximately two times greater. The value of RMSE for inferential sensor based on a model-based approach is 7.6445 and in the case of hybrid inferential sensors, OLSR equals 3.2632, for LASSO it is 1.9478.

The phenomenon of the increase in RMSE can also be observed on the graph (3.4) where the model-based IS suggests that the liquid level in the reboiler should reach up to 30 cm. However, this is not physically possible as the maximum recommended liquid height in the reboiler is approximately 23 cm (this is the maximum that can be measured by the ruler). The model error could be caused by the fact that in the hydrostatic pressure model (Equation 3.5), we assumed a constant density of water (ρ_{H_2O}), while the temperature in the reboiler varied up to 100°C. Another factor that influenced this model is the neglect of methanol in the calculations. As the feed contained only 20 % methanol in the initial phase, this did not have a large effect on the measurements. However, this factor should not be ignored during continuous distillation of the same mixture. Finally, based on the RMSE criterion during testing the IS we can conclude, that the LASSO model is the better one and should be used during further experiments. The estimated height of the liquid in the reboiler is calculated as:

$$\begin{aligned}\hat{h}_{\mathrm{H}} = &-1.0292Q_{\mathrm{Reb}} - 0.0112\mathrm{val}_{\mathrm{R}} - \\ &- 1.2692val_B + 0.1606\mathrm{sf}_{\mathrm{R}} - 0.0761P_{\mathrm{F,out}} - 0.3283Q_{\mathrm{Reb,out}} - \\ &- 0.0090T_5 - 1.6812T_{\mathrm{C,in}} - 0.0023T_{\mathrm{C,out}} - 0.0708T_{\mathrm{R}} + \\ &+ 0.0769T_{\mathrm{F,in}} + 0.0720T_{\mathrm{F,out}} + 0.2402\hat{h}_{\mathrm{MB}} - 1.7200,\end{aligned} \tag{3.6}$$

where the individual coefficients can be found in the Appendix B, Table B.1.

3.2.3 Design of an Inferential Sensor for the Product Composition

The product composition is the most important variable in separation as it directly indicates the performance of the distillation column. However, we do not have physical sensing devices on the distillation column providing this information online. These sensing devices are usually very expensive and often inaccurate. In order to design the inferential sensors measuring the composition of the products, we need to provide

Figure 3.5: Digital refractometer DR201-95 from A.KRÜSS Optronic.

the measurements of the corresponding composition. For this purpose, we use a refractometer. Before we start to use this device, it is necessary to create the calibration curve, which determines the concentration of a solution based on its refractive index. By measuring the refractive index of a sample and referencing it to the calibration curve, the corresponding concentration of the solute can be determined. The refractometer is a very precise instrument and it allows us to measure each composition during the experiment.

Refractometer We have at our disposal a refractometer DR201-95 from company A.KRÜSS Optronic. This equipment can be seen in Figure 3.5. It is a digital mobile device that can automatically measure the Refractive Index [1] (RI) of substances. It has a high measurement accuracy, $\pm$ 0.0003 RI, and can operate in the temperature range of 10–40° C. With this device, we can check the purity of products during chemical processes and subsequently determine their concentrations [12].

We start with the design of the inferential sensor (calibration curve) for the refrac-

[1] Refractive index defines the ratio of the speed of light in a vacuum to the speed of light in the material [2].

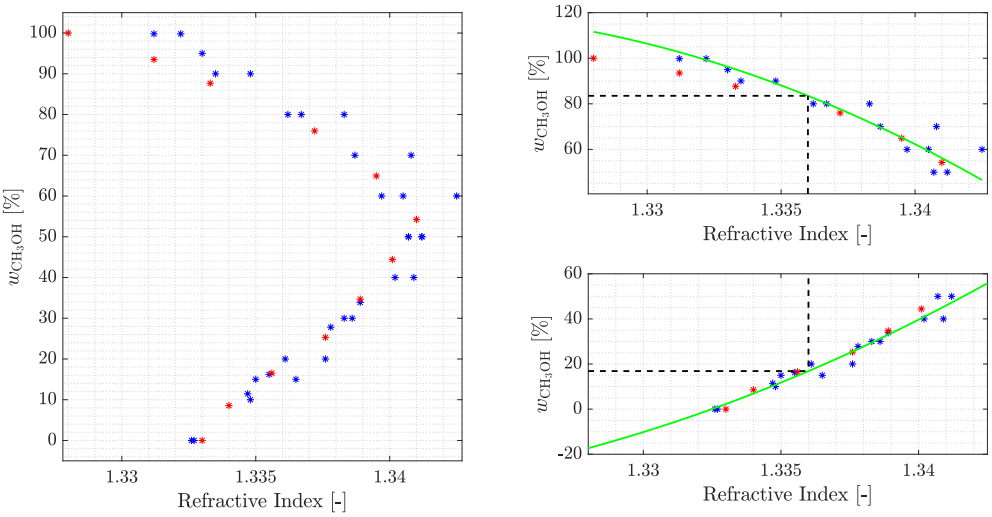


Figure 3.6: Comparison of the measurements (on the left-hand side) from the literature (red points) with our experimental data (blue points), designed IS (green trajectory) in interval 50–100 % (on the right-hand side, top) and designed IS (green trajectory) in interval 0–50 % (on the right-hand side, bottom).

tometer by manually measuring the composition of the products (feed, distillate and bottom product) several times during the experiment using a refractometer. During the calibration of this device, we measured the composition of certain samples. Using a refractometer, we get the mass fraction of the highly volatile substance w_{CH_3OH} in liquid mixtures (in our case for methanol). By comparing the data from the literature [42], we found that the experimental data were correct because they have the same development. This comparison of data can be seen in Figure 3.6, where the blue dataset indicates our manually measured data and the red dataset is the data from the literature. On this graph on the x-axis can be found the refractive index of measurements (RI) and on the y-axis the mass fraction of methanol (w_{CH_3OH}). Readers can observe that two w_{CH_3OH} compositions can correspond to one measured RI value. Therefore, due to the shape of the measured data, we divided these measurements into two graphs and created two models. That is shown in the right part of Figure 3.6.

It was necessary to develop two inferential sensors to determine the composition above and below 50 % of this more volatile substance. The measurements are divided into two parts and in the upper graph, we can see the data from 50 % to 100 % and in the lower graph from 0 % to 50 %. We apply a polynomial fitting to the resulting

refractometer calibration curve to fit the data. We have thus used the OLSR method. The model is then fit for use (green). The upper graph is used to indicate the composition of the distillate and the lower graph is in the case of the feed and bottom product. In Figure 3.6, you can see the auxiliary lines (black dashed lines) indicating the concentration corresponding to a given RI. For example, if we measure with a refractometer the RI value of 1.336, we get concentrations of 83.51 % and 16.87 %. If the sample came from a distillate we will consider the higher value, but if the sample was from a feed or bottom product then the lower value should be the correct one.

It is specified by the producer that the refractometer can operate in the interval 1.333–1.532 RI. However, during our experimental measurements, values in the range of 1.328–1.343 RI were found. These values are from multiple experimental measurements. Due to this lower limit therefore it is assumed that there was a problem with the refractometer equipment and thus may have corrupted the data and caused a bias for the inferential sensors that were designed further. Despite these facts, it is desirable to design an IS. Measuring concentration with a refractometer is time-consuming and the operator can only obtain information about the concentration at certain times. Taking a product sample for analysis with a refractometer can take up to 15 minutes, which is inefficient. Therefore, the implementation of an IS for concentration control is necessary.

Inferential Sensor Design Procedure The design of IS for the bottom product composition is based on the OLSR and LASSO methods (see in Section 2.4). We use the same procedure for the IS design to indicate the concentration of other streams within the distillation column. The procedure consists of the following steps:

1. It is important to exclude experiments where the desired input variables (measured by online sensors) are not available. As it is mentioned in Section 3.1, only quantities directly measured by certain instruments, such as temperature and pressure sensors, were initially available. The number of available input variables increased over time. We consider only experiments in which variables 1–30 from Table B.1 are available to measure.
2. It is important to combine data from several experiments. As there are relatively small datasets with lab measurements (manual measurements with the refractometer) from each experiment, all remaining data (after selection from the first step) were combined.
3. In the next step, the lab measurements (from the refractometer) are paired with the input variables (from online sensors) according to the measurement time.

4. Subsequently, the constant input variables throughout the experiments (e.g., safety valves) are discarded. The number of variables was also reduced by removing other inferential sensors. These input variables do not provide further useful information as they have zero variance. Furthermore, the measurements from the pressure sensors were discarded because they were recalibrated during the work with the laboratory distillation column.

5. The linear correlations between the input variables are indicated by `corrplot` function in Matlab, which compares the variance between the input variables. According to this plot, it is possible to exclude the input variables with an obvious linear correlation. For decision-making, data that had a higher Pearson's correlation coefficient[2] r than 0.9 were excluded. Such variables include temperatures measured at two subsequent trays, where the correlation is expected. For instance, when comparing the temperatures at the first and second trays, they are found to be similar. This is leading to the subsequent discarding of one of the variables. A similar situation arises with the reboiler heater energy input and feed pump power variables. Both variables involve measurements of user input and device output, yet the differences between these measurements are negligible. Consequently, one of the variables was neglected based on the correlation coefficient, simplifying our model in this way.

6. The dataset consisting of various experiments is subsequently mixed and divided into training (80 %) and testing (20 %) sets.

7. The standardization is performed to ensure that the input and output variables are on the same scale.

8. The next step involved training the inferential sensors using both the OLSR and LASSO methods. For all the inferential sensors, a LASSO penalty element of 0.1 is considered.

9. The designed inferential sensor is validated on the testing dataset (unseen during the training step). The accuracy of the inferential sensor is determined by using the RMSE criterion.

10. The penultimate step is de-standardization transforming the variables into the original space (i.e., variables have original units).

[2]Pearson's correlation coefficient describes the variance and covariance between two random variables. It indicates the degree of linearity, while the value r must be in interval $-1 \leq r \leq 1$. The lower bound denotes perfect negative correlation, 0 no correlation and the upper bound is perfect positive correlation [33].

11. The last step is the implementation of the developed inferential sensor in the HMI scheme and subsequently, it is ready for the estimation of the desired quantity during the next experiment.

This procedure is recommended if the design is limited by the size of the available dataset. Furthermore, continuous visualization of the data is recommended which can help to improve the model during the design.

Bottom Product Composition After completing these steps, we obtained the resulting IS models for the bottom product. The performance of these IS on the training and testing dataset is shown in Figure 3.7. The inferential sensors are trained on the lab measurements, which are manually indicated by using the refractometer. This graph shows the time dependency of methanol composition in bottom product x_{B}.

On the upper graph can be seen the training of the IS. After the analysis, we found that the RMSE value for the inferential sensor for the bottom composition based on the OLSR method is 1.9075 and for the LASSO method is equal to 2.6842. This means that OLSR describes better the manually measured data and should have been elected. From this, we can deduce that we should choose the OLSR method for determining the composition.

On the lower graph in Figure 3.7, one can see the tested inference sensors that we performed on data from other experiments. The description of the axis of the graphs is the same as in the sensor training. In this case, the RMSE of OLSR is 9.3273 and for LASSO this value equals 1.7245. It can be seen that we get the opposite results when testing and the IS designed by the LASSO method are closer to reality. For our purposes this value is acceptable. As can be seen, only 5 measurements were used for testing, which is a very small sample for evaluation. We experimented with different combinations of available datasets, in total 29 measurements. The sensor with the lowest RMSE value during testing was considered the best option.

Despite having limited data available, we found that using 80 % measurements to train the sensor is the best option. This allows us to use the largest possible number of measurements for training (desired due to many input variables), as we needed to have some data left for testing. As mentioned previously, obtaining concentration measurements can be very time demanding. To obtain a total of 29 measurements of distillate, bottom and feed concentrations, it took approximately 8 hours. Moreover, we only have information about these values at certain irregular times. Therefore, any

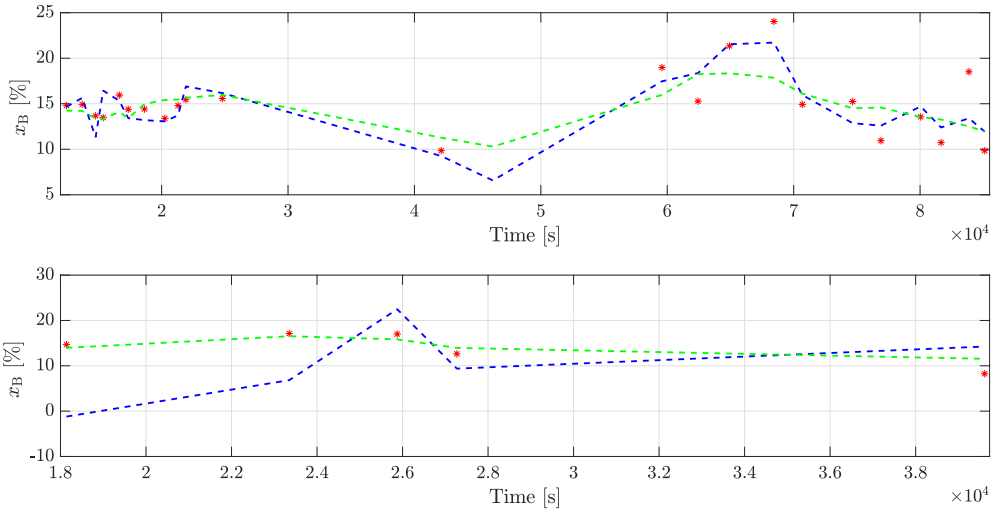


Figure 3.7: Training (upper graph) and testing (lower graph) of IS for bottom product composition designed by OLSR (blue dashed trajectory) and LASSO (green dashed trajectory) method on manually measured data (red points).

assistance in obtaining concentration information will be helpful and even inferential sensors designed on such simple methods as linear regression can predict the required quantities. In this case, OLSR used all 15 variables available after reduction, while LASSO used only 4. The deviation between the models was quite small, but the simpler LASSO had a smaller RMSE value during testing and should therefore be chosen for implementation in the control scheme.

The LASSO-based inferential sensor for the bottom product composition in its final form is as follows:

$$\hat{x}_{\mathrm{B}} = 1.038Q_{\mathrm{Reb}} - 0.044\mathrm{val}_{\mathrm{c}} - 0.6782T_{\mathrm{Reb}} - 4.462T_{\mathrm{c,in}} + 136.4, \tag{3.7}$$

where Q_{Reb} is a heat input to the reboiler, the opening of the valve for the cooling medium $\mathrm{val}_{\mathrm{c}}$, the temperature in the reboiler T_{Reb}, temperature of coolant entering to the condenser $T_{\mathrm{c,in}}$, and the last value, which is the bias.

Verification of the resulting model in (3.7) is possible by checking the values of the β coefficients and their signs. For example, increasing the heating in the reboiler leads to the generation of more vapour. As a consequence, the concentration of methanol increases in the distillate product and decreases in the bottom product. Therefore,

in the case of Q_{Reb} the sign should be the opposite than in (3.7). Similarly for the temperature in the reboiler, the temperature rises by increasing the heating. Therefore, we can conclude that the sign for T_{Reb} is correct because with rising temperature the composition of methanol in the bottoms is decreasing. When we reduce the temperature of the cooling medium entering the condenser, the temperature within the condenser itself decreases. This decrease in $T_{\mathrm{c,in}}$ creates a cooler environment for the vapour phase, allowing the more volatile substances to condense more efficiently and at a faster rate. This improved condensation process plays a crucial role in enhancing the quality of the distillate. This increased cooling causes a higher degree of separation between the volatile components and the desired distillate product. This leads to a higher concentration of the more volatile components in the distillate. In summary, decreasing the temperature of the inlet medium into the condenser in a distillation column improves the separation between different components, favouring the condensation and collection of the more volatile components in the distillate. The effect on the composition of methanol in the bottom product is the opposite. When the temperature of the cooling medium is decreased, the composition of methanol in the bottom product decreases. Therefore, the sign for $T_{\mathrm{c,in}}$ should be reversed than in (3.7). Closing the valve for the cooling medium ($\mathrm{val_c}$) has the reverse effect. By reducing the valve opening, less cooling medium is introduced, which does not contribute to separation. Consequently, a larger amount of methanol remains in the bottom product. In this case, the sign is correct.

Distillate Product Composition We followed the same steps in the case of the distillate product composition. Again, it is necessary to appropriately distribute the entire available dataset into the training and testing datasets.

Figure 3.8 illustrates the training and testing of the inferential sensor. On the upper graph can be seen the training and on the lower graph is the testing of this inferential sensor. The graph represents the time dependence (x-axis) of the methanol concentration in the distillate (y-axis). The training dataset consists of 23 measurements while the testing dataset involves only 5 measurements. The RMSE value during training using the OLSR method is 3.1278, whereas for the LASSO method achieves 5.7348. Consistent with the previous case, the OLSR model outperforms the LASSO model on the training dataset. However, notable differences between the two methods arise during testing. In this case, the OLSR method yields a higher RMSE (18.5101) compared to LASSO (4.5879). The OLSR method exhibits larger deviations than the LASSO method and provides unrealistic values since the methanol composition should fall within the range of 0 to 100 %. This demonstrates that LASSO is more effective compared to OLSR, as we could see in the case of inferential sensor design for bottom

product composition. Similar to the previous case, the designed inferential sensors can be enhanced by collecting additional data.

The fact that OLSR considers all variables could lead to multicollinearity, a situation in which some of the independent variables are highly correlated with each other. For this reason, it may not be meaningful to evaluate the model designed with OLSR in this way and instead, we should rely on additional quality criteria like RMSE. In practice, the effects of modifying variables can be complex and may depend on many factors, such as the specific equipment and conditions employed. As a result, experimental data can help to understand it.

The final model of IS designed with the LASSO method has a description:

$$\begin{aligned}\hat{x}_{\mathrm{D}} = {} & 0.1035P_{\mathrm{F}} + 0.167\mathrm{val}_{\mathrm{c}} + 0.1159\mathrm{val}_{\mathrm{R}} - 9.2601\mathrm{val}_{\mathrm{B}} - \\ & - 0.2299T_{\mathrm{top}} - 2.0523T_{\mathrm{c,in}} + 0.1739T_{\mathrm{R}} + 0.3503T_{\mathrm{F,in}} + 96.7150,\end{aligned} \tag{3.8}$$

where the inputs to the inferential sensor, in this case, are the following: the power of the feedpump P_{F}, the opening of the valve for the cooling medium $\mathrm{val}_{\mathrm{c}}$, reflux valve opening $\mathrm{val}_{\mathrm{R}}$, valve for draining the bottom product $\mathrm{val}_{\mathrm{B}}$, temperature on the top of the distillation column T_{top}, temperature of coolant entering to the condenser $T_{\mathrm{c,in}}$, temperature in the reboiler T_{Reb}, temperature of the feed entering to the preheater $T_{\mathrm{F,in}}$ and bias.

Similarly, the composition of the distillate product can be analyzed to assess the impact of variable changes on the final model of the inferential sensor. Increasing the flow rate of the feed into the distillation column can disrupt the equilibrium between the liquid and vapour phases, resulting in reduced separation efficiency. The faster flow of the feed limits the interaction time between phases and prolongs the attainment of equilibrium. This leads to decreased purity in the distillate and therefore, the sign for P_{F} should be negative.

As described for the inferential sensor for bottom composition, as $\mathrm{val}_{\mathrm{c}}$ increases, the concentration of more volatile components in the distillate increases. Therefore, the sign is correct. Similarly, the sign for $T_{\mathrm{c,in}}$ is correct because a decrease in temperature results in an increased composition of methanol in the distillate. When more bottom product is drained from the reboiler, the concentration of volatile matter in the distillate decreases, thus the sign of $\mathrm{val}_{\mathrm{B}}$ is correct.

The increased temperature at the top of the column promotes a higher degree of separation between the different components. Components with lower boiling points reach their vaporization temperature more quickly and are preferentially vaporized.

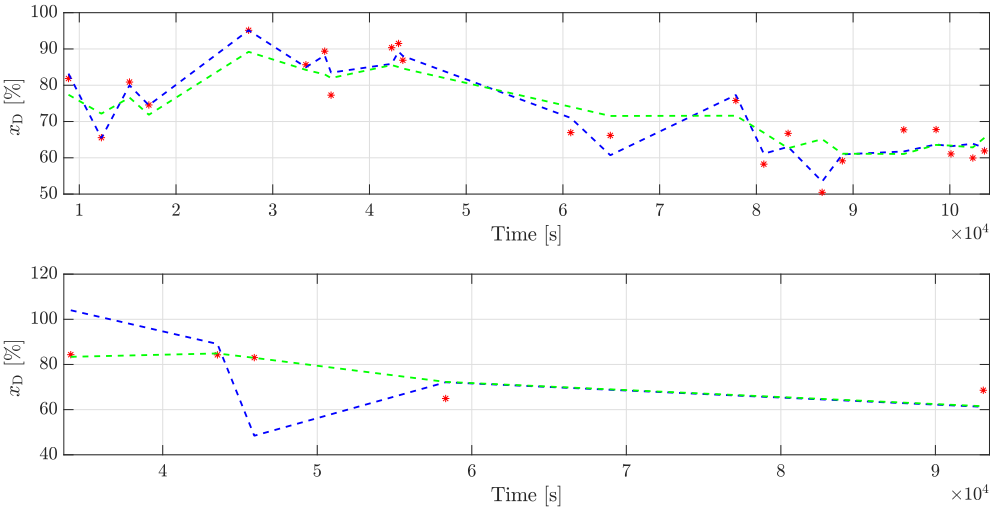


Figure 3.8: Training (upper graph) and testing (lower graph) of IS for distillate composition with OLSR (blue dashed trajectory) and LASSO (green dashed trajectory) method on manually measured data (red points).

Consequently, the distillate collected from the top of the column will have a higher concentration of more volatile components. Therefore, increasing the temperature at the top of the distillation column results in a distillate with a higher concentration of more volatile components. The impact of the variable T_{top} on the composition of methanol in the distillate is linear, and the sign is correct.

Increasing the reflux temperature (T_{R}) generally leads to a decrease in the concentration of more volatile components in the distillate. The reflux temperature influences the reflux ratio, where a higher reflux ratio means more condensed liquid. This condensed liquid is richer in less volatile components and it is returned to the column. However, increasing the reflux ratio (increasing val_{R}) improves product quality. It ensures better contact between the liquid and vapour phases, resulting in increased mass transfer and improved product quality. Therefore, in the case of T_{R}, the sign is incorrect, but the sign for val_{R} is correct.

The model also considers the feed temperature before preheating ($T_{\text{F,in}}$). If the feed temperature increases before the heat exchanger, it will also increase in the distillation column. This promotes vaporization, particularly of methanol. By increasing the feed temperature and enhancing the vaporization of methanol, a larger proportion of methanol will be present in the vapour phase and, consequently, in the distillate.

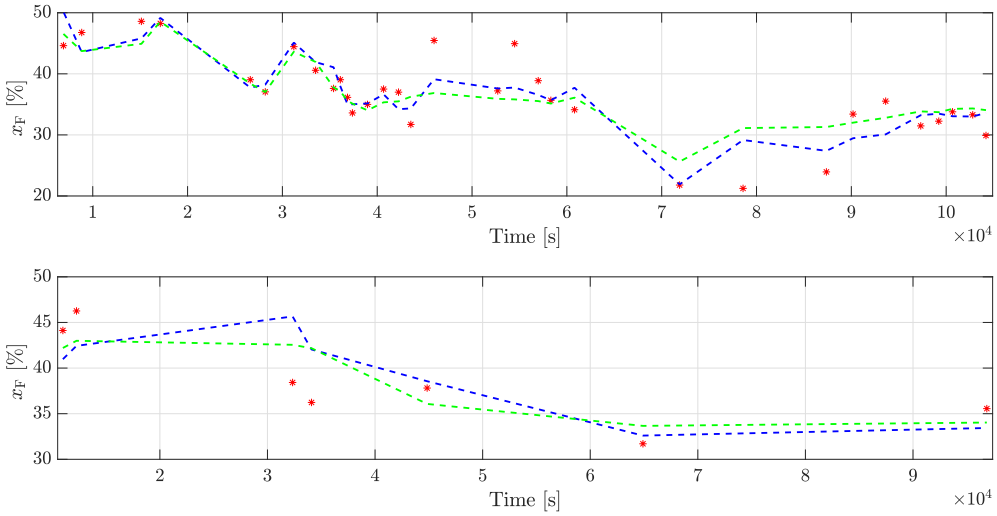


Figure 3.9: Training (upper graph) and testing (lower graph) of IS for feed composition with OLSR (blue dashed trajectory) and LASSO (green dashed trajectory) method on manually measured data (red points).

Therefore, the sign of $T_{\mathrm{F,in}}$ is correct.

Feed Composition Although the feed is almost always considered a disturbance quantity, the information obtained about its concentration can contribute to a better understanding of the process. In Section 3.1, it was described that in the case of the UOP3CC distillation column, both the distillate and the bottom product are collected into a common container. Subsequently, the feed is drawn from this container. Therefore, the quality of the products directly affects the concentration of the feed. The feed has undergone multiple recycling processes, resulting in considerable changes in its concentration throughout the experiments.

Similarly to previous IS designs for compositing, it is also true that the OLSR method was better during training and LASSO during testing. This can be observed in Figure 3.9. We chose datasets where 32 measurements were used for training and 7 for testing. During training, the RMSE value for the IS aggregated by the OLSR method is equal to 3.2395 and 3.7113 for LASSO. During testing the RMSE value for OLSR is 4.40793 and for LASSO 3.2936. In this case, LASSO considers 5 variables while OLSR considers almost 3 times as much, i.e. 14. This phenomenon was due to the fact that OLSR tries to reach the best results during data training, but as a consequence,

redundancy will occur and all variables will be considered unnecessarily.

The final inferential sensor model for feed composition based on the LASSO method has the following form:

$$\begin{aligned}\hat{x}_{\mathrm{F}} = {} & 0.0547P_{\mathrm{F}} + 0.0199\mathrm{val}_{\mathrm{c}} + 0.3741Q_{\mathrm{Preh}} - \\ & - 0.5928T_{\mathrm{c,in}} + 1.0535T_{\mathrm{F,in}} + 0.003T_{\mathrm{F,out}} + 13.92\end{aligned} \tag{3.9}$$

where the inferential sensor consists of the following variables: feed pump power P_{F}, the cooling medium valve opening $\mathrm{val}_{\mathrm{c}}$, feed pre-heater energy input Q_{Preh}, temperature of coolant entering $T_{\mathrm{c,in}}$ to the condenser, temperature of the feed entering to the preheater $T_{\mathrm{F,in}}$ and leaving the preheater $T_{\mathrm{F,out}}$ plus bias.

The power of the feed pump in a distillation column can indirectly affect the feed concentration. As described earlier, increasing P_{F} leads to a decrease in the methanol concentration in the distillate. This results in less effective separation, and therefore, the concentration of methanol in the feed does not increase. Consequently, the sign of the parameter P_{F} should be negative. Increasing $\mathrm{val}_{\mathrm{c}}$ results in an increase in the concentration of more volatile components in the distillate. However, the regression parameter for this variable has a relatively low value, indicating a small impact. The same applies to $T_{\mathrm{F,out}}$. Increasing the heat input to the preheater enhances the separation of the mixture, which should lead to an increased concentration in the feed. Thus, we can confirm that the sign of Q_{Preh} is correct. Decreasing the temperature of $T_{\mathrm{c,in}}$ results in an increased composition of the distillate, indicating more effective separation. Therefore, the sign of $T_{\mathrm{c,in}}$ is correct. Increasing $T_{\mathrm{F,in}}$ raises the temperature in the column, promoting effective separation. Hence, the sign of $T_{\mathrm{F,in}}$ is correct, indicating that the methanol composition in the feed rises with increasing $T_{\mathrm{F,in}}$.

Errors may have occurred also since the manually measured data were evaluated based on another inferential sensor. Thus, the errors from the refractometer mathematical model were also transferred to these inferential sensors for compositions. Improving the sensors would be possible with more experimental data.

3.2.4 Design of an Inferential Sensor for Steady State Indication

The last problem is the indication of a steady state. Our aim is to design controllers for the distillation column and therefore we need to perform the system identification. Process identification requires step changes and these cannot be generated without a steady state. The correctness of the steady state indication in the experimental data represents a key process variable. It allows a better understanding of the process and

subsequent design of controls to ensure the desired product composition. However, the indication of the steady state within the distillation column is problematic, whether in laboratory or industrial conditions. Due to the large number of variables involved, it is usually difficult to estimate this state with sufficient accuracy. The available measurements are mostly affected by noise or by correlation, resulting in the operator being unable to determine whether the desired steady state has already occurred or whether some variables are still in a state of transition.

To find the steady state of the system during the experiment, we compiled our own technique for creating an inferential sensor. This proceeds in the following steps:

1. selection of the important input variables,
2. sum all input variables,
3. filtration of the input,
4. calculation of the time differences of the filtered input,
5. calculation of a moving average.

In the first step, we selected a subset of inputs to the inferential sensor design to simplify the mathematical model. After attempting to minimize the number of variables, we decided to retain all variables from the measurement element and exclude only the variables from the inferential sensors. Our analysis during the experiments showed that every single variable affected the accuracy of the sensor. Inferential sensors were excluded due to the fact that they are calculated from other measured quantities and thus do not provide any new information.

The next step is to make a sum of the input signals. We further work with the sum of signals to simplify our model.

After the selection of input variables, we can filter the input data to smooth it. A low-pass Finite Impulse Response (FIR) type filter is used to filter the experimental data. The fundamental form of the discrete FIR filter is following

$$F(z) = f_0 + \frac{f_1}{z} + \frac{f_2}{z^2} + \ldots + \frac{f_n}{z^n}, \tag{3.10}$$

where coefficients of the delayed input f_n can vary. In our case, we consider a discrete filter of the fifteenth order ($n = 15$) with sampling period 10 s.

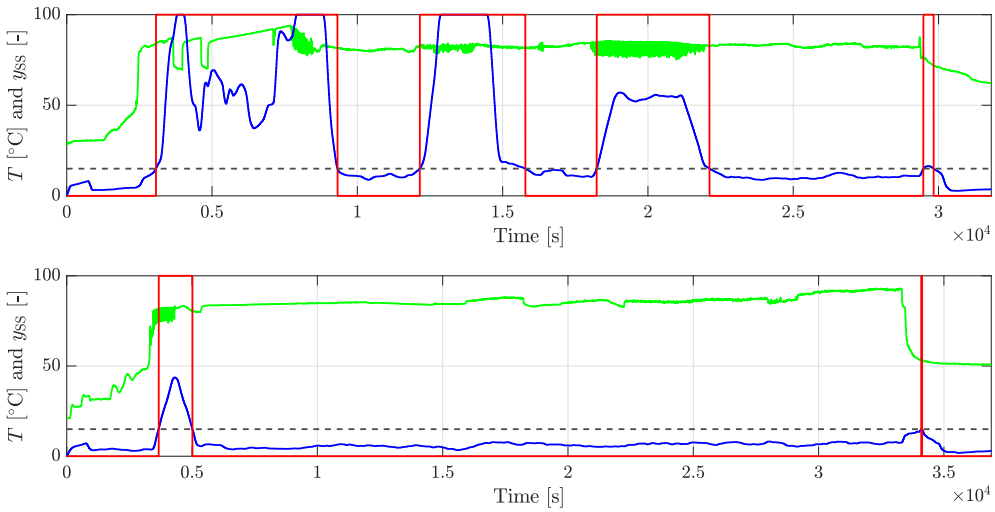


Figure 3.10: Comparison of output from steady state IS (blue trajectory), binary output from steady state IS (red trajectory) with the temperature at the eight tray (green trajectory) on training (upper graph) and testing (lower graph)of IS.

The next step is to calculate the time differences from the filtered input variables. These time differences are necessary to determine the steady state of the process. At this stage, we are looking for derivatives at each point.

The last step is to compute the moving average. Using a discrete integrator we are looking 25 minutes back in time to find a steady state.

Optionally, we can design a binary sensor for a more explicit display during the experiment to make the steady state indication more convenient for the operator.

The visual performance analysis of the designed inferential sensor is illustrated in Figure 3.10. This analysis involves the comparison of the steady state IS with a randomly chosen variable. In this case, we choose the temperature on the eight tray. The temperature is indicated with green colour, the output from the inferential sensor is blue and the binary output of the IS is red. On the x-axis can be seen the time (t) and on the y-axis the temperature on the eight tray (T) and the outputs from designed inferential sensors (y). The upper graph shows the training of the IS. We tried, based on our knowledge of steady states during the experiment, to tune the sensor to be as close as possible to the real observations. In the graph below you can see the testing

Table 3.1: Comparison of steady states from the inferential sensor with manually determined steady states.

	Inferential Sensor		**Manually determined**	
Num.	From	To	From	To
1.	0	3070	The start of the column	
2.	9319	12150	11100	11500
3.	15783	18236	Not determined	
4.	22119	29476	23500	25000
5.	29832	31815	The shutdown of the column	

of the sensor during a random experiment.

The output from inferential sensor for the steady state (y_{SS}) can obtain values $0\,\% \leq y_{\mathrm{SS}} \leq 100\,\%$. The binary inferential sensor $y_{\mathrm{SS,b}}$ can assume a value $0\,\%$, if indicates the steady state or $100\,\%$ in other cases. This value was determined by assuming that $y_{\mathrm{SS}} \leq 15\,\%$, the binary inferential sensor is $y_{\mathrm{SS,b}} = 0\,\%$. If $y_{\mathrm{SS}} > 15\,\%$, the binary inferential sensor shows $y_{\mathrm{SS,b}} = 100\,\%$. We can see in the upper figure, that based on the IS there are 5 steady states during this experiment. We manually checked each one of the variables to find out when approximately the steady state was reached and then compared them. This comparison is summarised in Table 3.1 in seconds.

The sensor shows five steady states, but we neglect the first and last ones since they occur during the start-up and shutdown of the column. For the second and fourth ones, the sensor indicated steady states correctly, even though they are shifted and last longer. The shift may be due to the high order of the filter but also to the dynamics of the process. The third steady state is incorrectly indicated and an error has occurred on the side of the IS. Despite this, we confirm that the IS correctly indicated two out of three steady states and can also detect the startup and shutdown interval of the distillation process. Furthermore, the graph clearly illustrates the improper functioning of the controllers. It is evident from the graph that there are intervals where the temperature exhibits oscillations, indicating a lack of stability in the control system. In these intervals, the sensors correctly do not indicate a steady state.

The graph presented below illustrates the results of the IS testing conducted during one of the experiments. By analyzing the IS, we can conclude that the process remained in a steady state from the initial start-up of the column until its shutdown. However, upon closer examination of the individual variables, it becomes apparent that temperature exhibited noticeable oscillations within the time interval ranging from 1.6 $times10^4$ s

seconds until the plant shutdown. Specifically for this temperature, the oscillations exceeded the acceptable threshold of $\pm$ 3°C, rendering it unsuitable for identification purposes. It is worth noting that even during the start-up and shutdown phases of the plant, a slight shift in the IS response can be observed. From approximately 3 $times10^4$ s onward, the process deviated from the steady state, as evidenced by a gradual increase in temperature. This indicates the necessity for more sensitive settings to ensure accurate measurements for this particular sensor. A suitable solution could be a direct adjustment of the sensor during further experiments. Despite these observations, we can still consider the sensor to be successful in indicating crucial operating points during the operation of the distillation column.

3.3 Control of Laboratory Distillation Column

Our objective is to design a control system for the distillation column that can achieve the desired concentration of products within a specified time frame. At the same time, our aim is also to increase the degree of automation of the laboratory distillation column UOP3CC. To fulfil these objectives, it is necessary to fix the recycling of the raw material in the distillation column, which is described in more detail in Section 3.1. Not only does this enhance the convenience of working with the process, but it also makes the distillation column more self-operating.

3.3.1 Variable Pairing

As mentioned in Section 2.2, there are many control strategies while the basic ones include the LV, LB, DV, and LD control configurations. Based on our experience with the Armfield UOP3CC laboratory distillation column, we chose the LV configuration, which is also the most widely used in practice. This control strategy was also described in the previously mentioned section, where it is shown in Figure 2.4. Such a control strategy is intended to help to make the process easier to control, to get to a steady state faster after the column start-up, to stay in the steady state even after step changes have been made, and last but not least to increase the level of separation. Due to limitations and deficiencies of our distillation column, we have modified it and adapted the aforementioned configuration to our conditions based on our experience. This denoted structure is shown in Figure 3.11. While the original LV configuration includes five control loops, we decided to control our distillation column with four.

In Section 2.2, it is described that one of the most important control loops is the one that maintains the overall balance of the system and the quality of the separation by

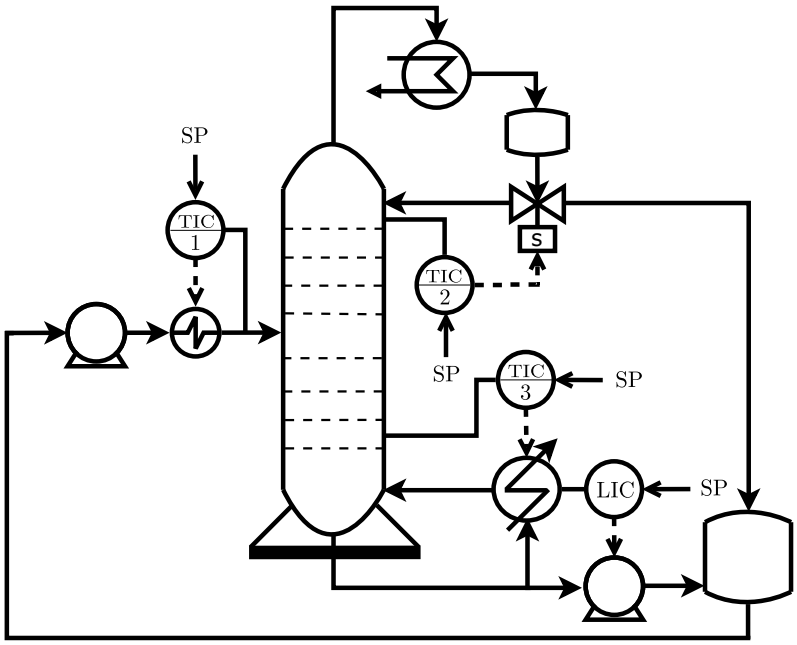


Figure 3.11: P&ID of laboratory distillation column UOP3CC with simplified LV configuration.

regulating the pressure on the head of the distillation column. However, due to the oversized condenser on our distillation column (as described in 3.1), we have not had the opportunity to consider this controller loop.

The LV configuration includes two temperature control loops, which are identical to our case. These loops regulate the temperature at the head of the distillation column with reflux flow (2. TIC) and the temperature at the bottom of the distillation column with heating in the reboiler (3. TIC). In our case, the relationship is created between the reflux valve and the temperature at the first tray. On the distillation column, there is a solenoid valve which splits the condensed vapour into distillate flow and reflux flow. The solenoid valve is therefore the manipulated variable and the temperature at the first stage remains the controlling variable, as is also the case of LV configuration.

Regarding the temperature control in the lower part of the distillation column, we decided to use the average of the temperatures at the seventh and eighth trays as the control variable. This decision results from the fact that during the experiments we noticed some shortcomings and inaccuracies of the temperature sensors. Typically, the

highest temperature is in the reboiler, and then the temperature drops as you move upwards through the distillation column at each tray. However, in our case, it often happens that the temperature on the seventh tray is higher than on the eighth tray, and therefore, we decided to average the outputs from these two temperature sensors. As mentioned above (Section 3.1), an oversized condenser can cause many problems in the distillation process. One of these issues is the rapid cooling of vapour, which can cause flooding in the distillation column. This flooding directly impacts the measured temperatures on the trays, resulting in similar temperatures across multiple trays or even causing higher temperatures on upper trays compared to lower trays. Through the heat exchanger hot steam flows to ensure the transformation of the flowing liquid phase into the vapour phase. In such cases, as indicated in Figure 2.4, the flow of the hot medium, i.e. the opening of the valve at the inlet to the reboiler, is used to ensure the desired temperature at the bottom of the distillation column. In our case, it is an electric reboiler and therefore it is only possible to control its power input. Therefore, as can be seen in Figure 3.11, the manipulated variable is the heating input in the reboiler.

According to the LV configuration, it would be necessary to implement two additional control loops for regulating the level in the reboiler and condenser. On the Armfield UOP3CC distillation column, we do not have the possibility to measure the level in the condenser and therefore this control loop can not be employed. Nevertheless, the second control loop which ensures the desired liquid level in the reboiler can be installed. The proposed inferential sensor also helped to achieve this. As it is mentioned in Section 3.2.2, we design the inferential sensors based on the hybrid approach to measure the liquid level in the reboiler. This variable is subsequently utilised as the control variable. Similar to the LV configuration, we selected the bottom flow rate as the pairing variable. In our case, a pump is installed at the bottom of the distillation column to ensure the discharge of the bottom product from the reboiler. The calibration with the design of the inferential sensor for the pump of the bottom product is presented in Section 3.2.1. This provides us with information about the bottom product flow rate in real-time.

The last pair consists of the feed temperature at the inlet to the distillation column and the preheater performance. Although the feed is typically considered as a disturbance variable, we found it necessary to control it. We observed that by changing the feed temperature we can reach a steady state in a shorter time. Another factor is that if the column is dried or flooded, this variable can significantly affect the condition in the column. The preheater for the feed inlet that is mounted on the plant is very sensitive and even small changes can have a significant effect on the distillation column. As a result, we create a third temperature control loop (1. TIC). Such a procedure

of variable pairing should ensure more efficient control of the laboratory distillation column UOP3CC.

3.3.2 PI Control

We have decided to use the trial and error method, which is often applied in practice. Our goal was to implement a PI controller. As it is described in Section 2.2.1, the P component gives a permanent control error. While it aids in achieving faster process control, the control deviation remains constant. However, using too large a P component in the controller will cause oscillations and consequently instability of the system. An improvement can be the application of the I component as well. The I component is responsible for precisely controlling the desired value by incorporating time information. Although the D component is used to measure the rate of deviation from the desired value (derivative part) and can help in fast response to changes in the input variable, we decided not to use it. This component does not work well with a noisy signal and can quickly cause system instability. This component can also increase the signal overshoot and worsen the effect of the controller. Therefore, we decided to adopt PI controllers and proceed with the implementation of a classical control structure. We added the controllers one by one, each of them having a parallel setting.

We begin with the implementation of the controller, which is responsible for ensuring the required temperature at the bottom of the distillation column. In the simulation environment, we consider a discrete controller with a sampling period of 10 s for all control loops. After implementation we activate first only the P part, starting with the value 1 and we watched how the system reacts to an increase or decrease of this value. When we were satisfied with the results, we add the I part and set it in a similar way. We start with the smaller value of the I component compared to the P component, as it is multiplied with the sampling period, causing its contribution to the controller to be ten times increased. The amount of vapour in the distillation column and the overall conditions within it are influenced by the heating of the reboiler. During the start-up of the distillation column or when making large step changes to quickly reach the desired value, the controller can immediately set its maximum, which can cause problems for the device. This can negatively impact the stability of the process and the lifespan of the distillation column components. To mitigate these issues, a constant value is added to the control action. This constant provides additional support and helps the controller start at a higher value, thereby preventing significant steps in the process.

Similarly, we continue with the other three controllers, i.e. temperature control at the first stage, control of the feed temperature at the inlet to the distillation column

and control of the level in the reboiler. During the setup of the other controllers, the previously implemented and tuned controllers are active. In this way, we want to incorporate the potential interactions between the action of the controllers into the design of the controllers to effectively expand the control configuration by other controllers. With the gradual addition of controllers, the control software responsiveness started to slow down and the program became increasingly overloaded. At the beginning of the work with the distillation column, only a 2-second sampling period is used and later on, we increase this value to 5 seconds. During this modification, we observed a noticeable reduction in the load experienced by the control circuit. Therefore, we doubled the sampling period to 10 s which still guaranteed a consistent and plentiful supply of sampled data throughout the process. After the implementation of the fourth controller, strong interactions between controllers appear. These interactions caused the process appears to be unstable under some circumstances.

Therefore we begin testing the individual control loops to identify the faults. We test each controller separately, so only the currently tested controller is on and all others are in manual mode, i.e. turned off. Although the controllers performed smoothly and were controlled well individually, they caused complications when operated together. Therefore, we decided to gradually deactivate the I components. The reason for this is that by reducing the complexity of the controller structure, we can avoid interactions that may have caused the aforementioned complications. This approach has helped and although only the P regulators are currently installed on the distillation column, the control is acceptable. This is also evidenced by the fact that while without controllers the distillation process took up to 4 hours after start-up to become steady, we can now achieve the steady state operation mode within 3180–5400 seconds. As a result, the controllers have the following P-values: 0.1 for temperature control at the bottom of the column, -6 for the temperature control at the column head, 1 for the control of the feed temperature at the inlet of the distillation column, -25 for the control of the level in the reboiler.

The gain values in a control system can be positive or negative depending on whether the quantity being manipulated has a direct or indirect effect on the controlled variable. In this case, increasing heating at the bottom of the distillation column and introducing hotter feed can improve product quality, so the corresponding gain values would be positive. On the other hand, to improve product quality, it is necessary to decrease the temperature at the head of the distillation column and the liquid level in the reboiler, so the corresponding gain values would be negative.

A necessary step is also the activation of saturation for control action. The controller responsible for the temperature at the column head can have an action range of 0 to

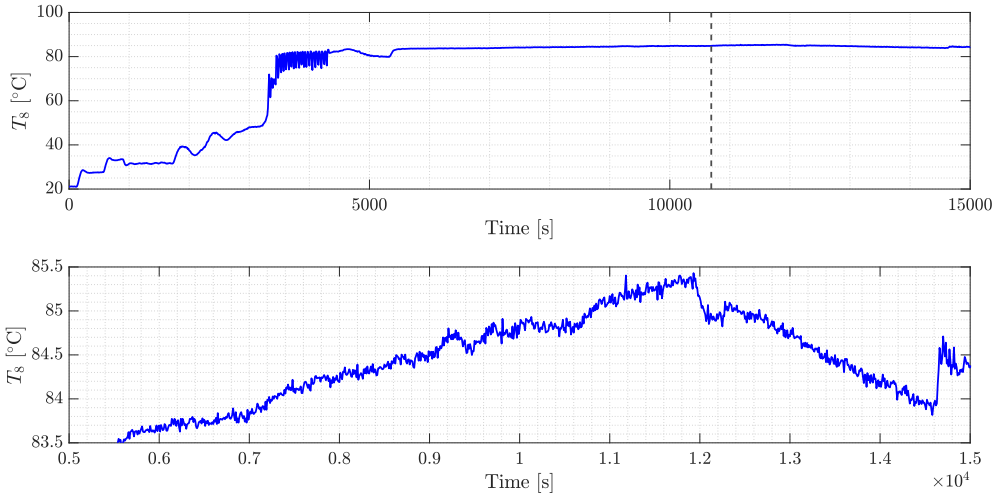


Figure 3.12: Behaviour of the temperature on the 8th tray (upper graph, blue trajectory), auxiliary line (black dashed line) and the settling area of the temperature on the 8th tray (lower graph, blue trajectory).

100 %, the same interval also applies for the control of the level in the reboiler. Due to the sensitivity of the preheater, its action is limited to 0 to 20 %. We decided to limit the control action of this controller, mainly from a safety point of view. The power of the reboiler can be normally changed manually in the range of 0 to 1.75 kW. As it is mentioned above, for the reboiler it is necessary to add a constant value of 1 and therefore, we adjusted the limits for the control action accordingly.

Currently, regulators serve us primarily for the start-up of the distillation column, while our main goal is to get to a steady state. In this state, we should be able to make changes. Figure 3.12 shows the time series chart with the temperature on the eight tray (blue trajectory). According to this variable, we can observe the progress of the column start-up. The controller has been switched on for 10690 s, which is also indicated by the auxiliary line (black dashed line). On the upper graph, it is possible to see some jumps up to about 3000 s. This was caused by the fact that right when the distillation column was switched on, we did not set the regulator to the desired steady state value, but gradually increased the set point to get to the desired temperature. From a time of about 3300 s it is possible to see a significant oscillation of the regulator, which can be caused by too strong P component. Then the temperature, representing the process behaviour, slowly settles down. From 5400 s onwards, a steady state is reached, which is also indicated by the inferential sensor as shown at the bottom graph in Figure 3.10.

For the temperature at the bottom of the distillation column, the steady state is considered to be 84.92° C. Although the oscillation still occurs. We allow ±2° C temperature oscillation for the eighth stage. Since a completely constant temperature cannot be achieved, some temperature oscillation is expected. So in this case it took about 5400 s to reach the steady state. Nevertheless, we let the regulator continue to operate and switched it off only at 10690 s. As you can see in the graph below, which illustrates the settling area, from the steady state to the controller shutdown the temperature oscillated a little, ±1° C around 84.50° C.

3.3.3 Identification of the Laboratory Distillation Column

Although the classical control structure is efficient, implementing it required adding many new blocks to the control scheme. Also, when changing the desired quantity, everything requires separate adjustments. To automate the process and make the work of the laboratory distillation column operator easier, we decided to implement an advanced control. For this, the identification of the process is essential.

Implementing an advanced control strategy offers various approaches. One approach involves enhancing the existing PID controllers by incorporating an advanced process controller. In this setup, the advanced controller assumes the role of a master, determining the desired set points for the slave P controllers. Another option is to replace multiple P control loops with a single controller capable of handling multiple inputs and outputs. In our particular case, we chose to pursue the second option.

As we already know, the basic principle of distillation is the separation of components based on their different boiling points. It is therefore necessary to ensure the correct conditions for the entire distillation column for the separated mixture. After the implementation of the classical control structure, we observe that temperature changes are directly related to the changes in separation. Therefore, we decided to use a combination of two control loops for the LQ controller, which were already present in Section 3.3.2. These are the temperature change at the head of the column with the reflux valve opening and the temperature change at the bottom of the column by adjusting the heating in the reboiler. For simplicity, we will use only the temperature from the last (8th) tray for advanced control.

To design an effective control system, it is necessary to understand the impact of input variables on the studied process. To gain insight into their influence on the temperature profile of the column, particularly the temperatures at the first and last tray, step changes were performed on the input variables. The input variables that were subjected to step changes were the heating in the reboiler and the opening of

the reflux valve. Through process identification, we can accurately characterize the process based on the variables of interest.

To ensure a sufficient description of the process, a series of experiments was conducted to reach a steady state. These experiments enabled us to determine the steady states associated with specific values of individual variables, which are required for the design of the advanced process controller. The indicated steady state values are 20 % open solenoid valve to ensure reflux flow, 83.83° C temperature at the first stage of the distillation column, output in the reboiler is lowered to 0.95 kW, average temperature 88.09° C at 7th and 8th trays.

Subsequently, we perform experiments to achieve the desired steady state of the process to be able to perform the step changes to identify the system. To reach the steady state, we used all P regulators. Once we reached a steady state, we gradually turned off the controllers and put them into manual mode. We then made step changes with the opening of the reflux valve and the heating in the reboiler. During the experiment, we change only these two input variables, and all others are constant. We made both upward and downward step changes. The steps were large enough to notice a change in the output, in this case of the temperatures. At the same time, the steps had to be different lengths and different sizes, even enough to avoid any correlation.

After collecting the data, we proceeded with the identification of the process. To simplify the system, we focused on four variables, and these are the inputs of heating in the reboiler and opening of the reflux valve and the outputs of the temperature at 1. and 8. trays. Firstly, we created a deviation form of the variables by subtracting the steady state values from each measurement. This step results in obtaining the outputs from the system, starting from point zero. We also excluded the start-up and shut-down intervals of the distillation column as they could introduce non-linearity to the system. Our main objective is to focus on the operational working area, and therefore, for the sake of simplicity, we chose not to include these intervals. By doing so, we ensure that we only consider the data by which the step changes were made for identification. We then utilised the System Identification Toolbox in Matlab [32] to identify the Armfield UOP3CC laboratory distillation column. We have opted for an identification model in the form of first-order transfer functions. We have defined the desired accuracy of the model and the maximum number of iterations. These parameter values have a direct impact on the computational effort involved, so it is advisable to select them thoughtfully.

The identified first-order transfer functions have the following form:

$$G_{u_1,y_1} = \frac{-0.1142}{112.1093s + 1}, \tag{3.11}$$

$$G_{u_2,y_1} = \frac{7.9651}{0.5176s + 1}, \tag{3.12}$$

$$G_{u_1,y_2} = \frac{-0.0158}{29.3395 + 1} e^{-400s}, \tag{3.13}$$

$$G_{u_2,y_2} = \frac{7.5033}{8.2996s + 1}, \tag{3.14}$$

where u_1 is the reflux valve opening, u_2 is the heating in the reboiler, y_1 is the temperature at the first tray, y_2 is the temperature at the 8. tray.

Based on these transfer functions, it is evident that a transfer delay occurs only in the case of the input-output pair u_1 and y_2. This finding is intriguing because, considering the slow dynamics of the distillation process, one would expect a transport delay throughout the entire distillation column. While it is true that among the inputs and outputs under consideration, the largest transport delays are expected for $G_{u_1y_2}$ due to the time required for a cold liquid to flow through the entire column to the last tray. The transfer delay of 400 seconds is still relatively small given the dynamics of the distillation column. Thus, it is advisable to wait for a longer duration, such as at least 15 minutes, when making step changes to ensure that the impact of the input is adequately reflected in the output.

We also employed a manual approach to estimate the parameters K, T, and D for each subsystem individually. Our calculations yielded larger time constants, deviating from the values identified by the toolbox. These differences might arise from the configuration settings of the identification toolbox itself. Based on our observations while working on the distillation column, we anticipate significantly higher time constants. Consequently, refining the parameters in the toolbox (e.g., adjusting the number of iterations, employing different methods to identify the process) and conducting tests of this toolbox on a distinct dataset could lead to further improvements.

Another interesting observation is that the gain is several times smaller when the reflux valve is opened compared to when heating in the reboiler. In the case of u_2, there is no transport delay, which was also confirmed during experiments as the temperature immediately started to rise on the last trays upon changing the heating. Additionally, for u_1, it is worth noting the negative signs in the transfer functions $G_{u_1y_1}$ and $G_{u_1y_2}$. This property arises because opening the reflux valve allows more cold liquid to flow back into the column, causing a cooling effect on all stages of the distillation column. Conversely, when heating in the reboiler, the phenomenon is the opposite. The addition

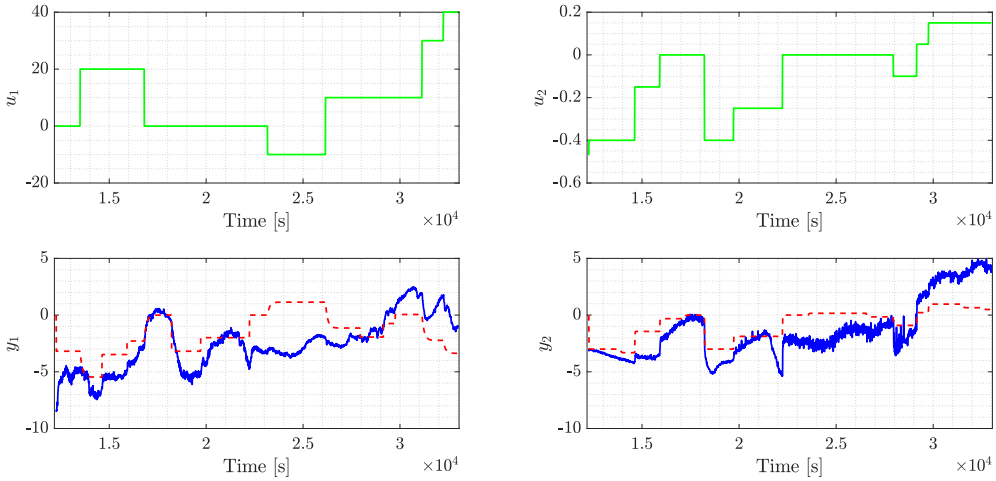


Figure 3.13: Step change of input variables (upper graphs, green trajectory) and the comparison (lower graphs) of measured (blue trajectory) and identified (red dashed trajectory) output variables responses in deviation form.

of heat increases the temperature on the trays and therefore the signs of $G_{u_2y_1}$ and $G_{u_2y_2}$ are positive.

The comparison of the outputs y_1 and y_2 for step changes of inputs u_1 and u_2, based on the data and time-dependent identification in deviation form, can be observed in Figure 3.13. Within the considered interval, step changes were only performed with u_1 and u_2. The graph of $y_1 = f(t)$ demonstrates that although u_1 remained constant in the interval from $1.7 \cdot 10^4$ to $2.3 \cdot 10^4$, there were changes in y_1. These changes can be attributed to the step changes in u_2. Similarly, the reverse is also true. The graphs indicate that the identified models calculated using the toolbox accurately capture the real system behaviour. The trajectories of the two outputs exhibit a similar trend, confirming the success of the identification process and supporting the development of an advanced control strategy. This conclusion is further confirmed by the individual RMSE values. The quality criterion value for y_1 is 2.2836, while the RMSE for y_2 is 1.8638. Although these results could be enhanced through additional experiments, data normalization, or other data treatment methods, they provide a solid foundation for further analysis and control design.

3.3.4 Design of a Linear Quadratic Regulator for the Laboratory Distillation Column

Creating a control strategy based on system states is often challenging due to the difficulty in obtaining and measuring states, particularly in the case of MIMO systems such as the distillation process. Therefore, it is generally easier to monitor the system outputs instead. We decided to use the output feedback Linear Quadratic Regular method, which is an advanced control technique based on the state-space description of the process. This version of LQR is specific in using the outputs from the system instead of states for calculations. Therefore, it is essential to identify the system for the design of an advanced control structure. After obtaining the transfer functions that describe the behaviour of the distillation column, we transformed them into a state-space representation to design the control structure.

As already described in Section 2.2.2, solving the optimization problem using LQR requires knowledge of the state description of the problem. Therefore, in the first step, we transformed the obtained transfer functions described by Equation 3.11 and transformed them into the state-space model such as:

$$\begin{aligned} \dot{x} &= Ax + Bu, \\ y &= Cx + Du. \end{aligned} \tag{3.15}$$

For the transformation, we used the System Identification Toolbox [32], which directly recalculated the values of the matrices A, B, C and D from the mentioned transfer functions. The continuous-time identified state-space model for the studied distillation column are

$$A = \begin{pmatrix} -0.0127 & 0.0030 & -0.0139 & 0.0052 \\ 0.0144 & -0.0261 & 0.0309 & 0.0136 \\ -0.0100 & -0.0094 & 0.0159 & -0.0002 \\ -0.0079 & -0.0121 & 0.0153 & 0.0066 \end{pmatrix}, \tag{3.16}$$

$$B = \begin{pmatrix} 0.0003 & 0.0761 \\ -0.0003 & -0.3502 \\ -0.0003 & -0.1729 \\ -0.0002 & -0.1488 \end{pmatrix}, \tag{3.17}$$

$$C = \begin{pmatrix} 14.2271 & -14.5138 & 39.3121 & -5.2594 \\ 10.5513 & -15.9015 & 17.6404 & 21.6983 \end{pmatrix}, \tag{3.18}$$

$$D = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \tag{3.19}$$

We checked the stability of the system by calculating the eigenvalues of matrix A. Their values are:

$$\mathrm{eig}(A) = \begin{pmatrix} -0.0001 \\ -0.0009 \\ -0.0065 \\ -0.0088 \end{pmatrix}. \tag{3.20}$$

As can be seen, the real part of the numbers is negative, and therefore we can evaluate that the system is stable and it is possible to proceed with the design of the controller.

We decided to use the toolbox provided by A. Ilka [14] called `oflqr` for the design of the output feedback LQR control which is built for the design of such a control structure. In order to verify the correctness of our implementation of this toolbox, we back-validated the results presented in literature [26].

Then we continue with the design of the controller for the studied distillation column. The shape of the gain matrix is in our case as follows:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} \begin{pmatrix} u_1 & u_2 \end{pmatrix}, \tag{3.21}$$

where u_1 is the reflux valve opening, u_2 is the input power of the reboiler, y_1 is the temperature on the first tray and y_2 is the temperature on the eighth tray.

We proceeded to design the LQR controller for the distillation column with the calculated gain matrix as follows:

$$K = \begin{pmatrix} -0.0004 & 0.0005 \\ -0.0017 & -0.0019 \end{pmatrix}, \tag{3.22}$$

while we were considering the weighting matrix are following:

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad R = \begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix}, \qquad N = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}. \tag{3.23}$$

Matrix Q has the form $n \times n$ based on the number of states (based on the matrix A), i.e. $n = 4$, matrix R has the form $m \times m$ based on the number of inputs (based on the matrix B), i.e. $m = 2$, and finally, the matrix N has the shape $n \times m$.

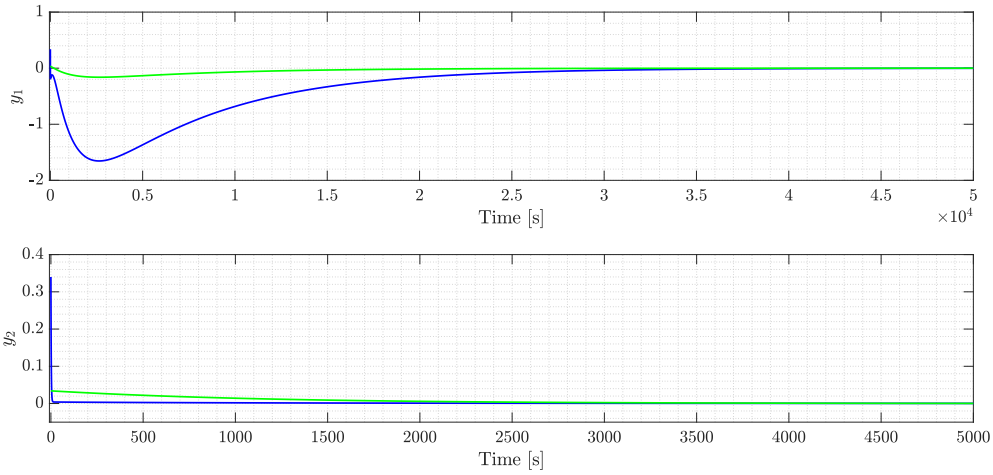


Figure 3.14: Comparison of control y_1 (upper graph) and y_2 (lower graph) with P controllers (blue trajectory) and LQR (green trajectory).

The stability of the system with the controller can be checked as follows:

$$\text{eig}(A - BKC) = \begin{pmatrix} -0.0001 & & \\ -0.0009 & & \\ -0.0074 & + & 0.0005\text{i} \\ -0.0074 & - & 0.0005\text{i} \end{pmatrix}, \tag{3.24}$$

where the real parts of the eigenvalues are negative, indicating that the system is stable.

Figure 3.14 illustrates a comparison between classical (blue trajectory) and advanced (green trajectory) control strategies, considering non-zero initial conditions over time based on the state space description of the UOP3CC distillation column. The purpose is to remove the disturbance (disturbance rejection). The upper graph shows the change in y_1, while the lower graph shows the change in y_2. Comparing the results reveals that the temperature at the distillation column head takes approximately ten times longer to stabilize. In the case of y_1, the settling time is significantly longer with the P controller, displaying a noticeable undershoot. The LQR achieves steady-state in about half the time. Regarding y_2, it takes approximately ten times less time to settle the variable compared to y_1. The LQR exhibits a longer settling time for y_2, while the P controller achieves it much faster. The P controller exhibits a notable overshoot at the beginning of the time interval, which can be detrimental to the lifespan of the distillation column actuators. The graph indicates that although the

LQR settles more slowly, it provides less aggressive control compared to the classical P controller. However, the speed and quality of the LQR can still be adjusted by tuning the weighting matrices.

Insufficient process identification or oversimplification can cause problems in control. Another problem can be caused by the setup of the toolbox, which was used for the identification. With LQR control, we can identify the most influential parts of the distillation column, which will help us better understand problems with the column or the control itself.

As mentioned earlier, one disadvantage of LQR control is its requirement for a more detailed system description. Inaccurate or incomplete system identification can result in suboptimal control performance. Moreover, LQR control often requires a larger number of sensors to track multiple outputs, which can be problematic and expensive.

In comparison, implementing classical PID controllers typically requires less time, but operating and controlling MIMO systems with multiple PID controllers can be more complex. Despite the increased complexity involved in preparing the control structure for LQR, which includes system identification, it offers greater efficiency compared to the classical control structure with PID controllers.

Chapter 4

Discussion

In the practical part of this paper, we are dealing with various methods of designing inferential sensors. We performed several experiments on the distillation column to obtain data from the measuring components, and we also took samples manually. These data were then utilised during the design procedure using different methods. To design the inferential sensors and obtain the composition of the products we applied data-based models, namely ordinary least squares regression and least absolute shrinkage and selection methods. An interesting observation and possible improvement could be the prediction of the desired quantities by other advanced approaches, such as partial least squares (PLS), artificial neural networks (ANN) or the implementation of an observer such as an (extended) Kalman filter.

The main problem in designing inferential sensors is the amount of available data. After data processing, the number of measurements from lab analysis is reduced to only 29 measurements for the distillate product, 34 for the bottom product, and 41 for the feed. Although we are able to design mathematical models, further experimental data could significantly improve the accuracy and reliability of the designed inferential sensors.

To check the performance of the designed sensors we use the RMSE criterion, but there are many methods for evaluating the quality criterion of inferential sensors, such as prediction error variance (PEV), relative average deviation (RAD), Bayesian information criterion and many more.

In designing the inferential sensor, we also indicate certain process characteristics based on pressure. We observed that if the pressure measured in the centre of the reboiler starts to rise significantly, the distillation column has a tendency to become flooded. However, these observations need further confirmation by additional experiments. It would be useful to add an inferential sensor to indicate this problem which may be of assistance to the operator.

In the next part of the thesis, we are dealing with the control itself. Firstly, we focused on the pairing of variables. Although we assumed the LV control configuration and knowledge from the literature or practice, it would be useful to also use the Relative Gain Array (RGA) method, which in case of multiple input/multiple output systems can help to tell exactly which variable should be paired with which one. In this section, we also mentioned that there were complications with measuring temperatures at the bottom of the distillation column. As a result, we used the average from the last two trays (within the distillation column) for control. We are not able to identify any reason for this phenomenon during our experiments. Therefore, we have deduced from this, that the poor calibration of these temperature sensors caused their unreasonable behaviour. We recommend checking the accuracy of these sensors and recalibrating them if necessary in the future.

After the pairing of the variables, we come to the design of classical P controllers using the trial and error method, which is commonly used in industry. However, it is time-consuming, and we recommend trying other tuning methods, such as the Ziegler-Nichols method, the Cohen-Coon method, robust control, or fuzzy control which can improve the controllers.

We also performed process identification, where we used the toolbox from Matlab after performing the step tests. It would be interesting to compare the results from this toolbox with the manual identification of the process to confirm the accuracy of the information obtained.

As mentioned earlier, we had a relatively small amount of data available, and therefore further experiments could help improve the LQR controller, which we addressed in the last section of the paper. We had limited space to tune the Q and R weight matrices, therefore, additional experiments would also contribute in this area.

In conclusion, it should be added that when deciding on the implementation of a control structure, several factors such as process knowledge, system complexity or economic consequence have to be taken into account. Therefore, there is no clear recommendation for which control structure to choose, as both methods have their advantages and disadvantages.

Throughout the thesis, we emphasised the importance of the quantity and quality of data in designing sensors and controllers. Working with a distillation column can be time-consuming, and collecting data takes time. However, the results are worth the effort as they can simplify the work even on a complex device such as a distillation column.

Chapter 5

Conclusions

This thesis focused on the design of an advanced control system for the Armfield UOP3CC laboratory distillation column. Initially, the measurement of several key process variables was required. Therefore, it was divided into two main parts: preparing the column to measure all the essential variables, and developing and testing various control strategies.

In the practical part of the work, we were introduced to the distillation column on which we performed several experiments. For this work, an Armfield UOP3CC laboratory distillation column was used on which all the aforementioned experiments were performed. During the experiments performed, a binary methanol-water mixture was separated, which was also recycled for further experiments.

As we have found out, for the design of the advanced control strategy, it is required to measure some variables as they are indirectly used for the design. We solved this problem by designing and then implementing inferential sensors. In this work, we have focused on creating inferential sensors based on different approaches and methods. They provide us with information about hard-to-measure quantities thanks to measurable quantities. Nine inferential sensors were designed for different parts of this distillation column. These are the calibration of the feed and bottom product removal pumps, an indicator for the liquid level in the reboiler, calibration of the refractometer and subsequent sensors to obtain the composition of the feed, distillate and bottom product and an inferential sensor to indicate the steady state of the process.

The inferential sensors were trained and tested using historical data. We verified their reliability and then, based on the results obtained, we selected the method that showed the best results in testing the given sensors. According to the results we can conclude, that the OLSR method was better for training and the LASSO method was better for testing. Consequently, we decided to implement the LASSO method in the control

scheme for all sensors. Then based on obtained experimental measurements we design an indicator of a steady state. This indicator is contingent upon individual parameters and their respective tuning. In this particular case, the sensor adeptly identified two steady states, although it also indicated an additional, extraneous one.

As we equipped the distillation column with sensors, we could move to the design of simpler controllers first. Using the method of variable pairing, we chose the pairs that would help us to reach the steady state sooner and ensure the controlled variables reach the desired values. We chose the LV configuration control scheme, which is one of the most applicable in practice. Although this scheme has five regulated loops, we chose only four to avoid overloading the control. We found that although our goal was to implement PI loops, using only P loops to control our process was also sufficient. Implementing the I element caused interactions that made the column unstable. Thus, we implemented only P loops in the resulting scheme, which were simpler and sufficient to achieve the desired values, even if there was a certain deviation. To design the controllers a trial and error method was employed. This control scheme unnecessarily imposes many single-controller loops, so we decided to apply advanced control.

We performed an experiment where we focused on step changes and identified the process. We used two inputs and two outputs and obtained four transfer functions that were transformed into state space descriptions. Based on the matrices, we designed a feedback LQR. In conclusion, we can summarize that although simple controllers were sufficient, implementing LQR control improved the process control. Overall, all steps depended on the quantity and quality of data. Both sensors and controllers could be improved by further experiments.

Compared to its previous state, the Armfield UOP3CC laboratory distillation column has undergone substantial improvements, rendering it more automated and user-friendly. These enhancements are crucial not only for our research but also for other users who may employ the device in future research endeavours, such as exploring alternative control strategies or advancing the current state-of-the-art.

Appendix A

Resumé

Prvé zmienky o destilačnom procese sú z rímskej ríše, avšak dodnes je to jeden z najpoužívanejších chemických procesov v priemysle. Pomocou tohto chemickotechnologického procesu je možné separovať kvapalné zmesi na základe ich (relatívnych) prchavostí. Destilačné kolóny je možné rozdeliť na vsádzkové a kontinuálne typy, pričom počas vsádzky je možné separovať zmes v určitých intervaloch a počas kontinuálnej destilácie neustále. Do destilačného zariadenia vstupuje nástrek ako kvapalina, ktorá môže byť predhriata. Zo spodku kolóny prúdia pary, ktoré sa dostanú do kontaktu s kvapalinou a takýmto spôsobom dôjde k obohateniu pár o viac prchavú zložku. Pary vystupujú cez hlavu kolóny do kondenzátora, kde sa skondenzujú. V tomto bode je možné hovoriť o ďalšom type destilácie. Jedná sa o jednoetážovú a viacetážovú kolónu. V prípade jednoetážovej kolóny sa skondenzovaná kvapalina v celom množstve odvádza. V prípade druhého typu, tzv. rektifikácie, je odvádzaná iba časť kvapaliny a časť sa vracia do destilačnej kolóny ako reflux. Odvodená kvapalina je hlavným produktom destilačného procesu a nazýva sa destilátom. Kvapalina, ktorá sa počas rektifikácie vedie naspäť do kolóny, tečie smerom ku spodku kolóny. Táto kvapalina prechádza cez etáže, ktoré slúžia na zabezpečenie väčšieho kontaktu parnej a kvapalnej fázy. Následne vstupuje do varáka, kde sa časť kvapaliny odparí a časť sa odvedie ako zvyšok, druhý produkt destilačného procesu.

Táto práca sa zameriava na návrh pokročilého procesného regulátora pre laboratórnu destilačnú kolónu Armfield UOP3CC, ktorá sa nachádza na Ústave automatizácie, informatizácie a matematiky. Jedná sa o voľne stojacu rektifikačnú kolónu s ôsmymi etážami, ktorá je predovšetkým určená pre akademické účely, na otestovanie teoretických znalostí, avšak svoje uplatnenie nájde aj v priemysle ako tréningový nástroj pre operátorov priemyselných destilačných kolón. Počas vykonaných experimentov sme separovali binárnu zmes metanol-voda. Destilačná kolóna je považovaná za komplikovaný proces, ktorý je silno nelineárny a označuje sa ako systém s viacerými vstupmi a výstupmi. Riadenie takéhoto zariadenia na zabezpečenie požadovanej kvality produktov môže byť náročné, preto sme sa rozhodli vytvoriť riadiacu stratégiu, ktorá uľahčí a

urýchli separačný proces.

Pre návrh riadenia destilačnej kolóny je potrebné merať niekoľko premenných. Na začiatku práce s danou destilačnou kolónou bolo možné merať len 19 základných veličín, vrátane teplôt (14 meracích členov), tlakov (3 meracie členy), ohrevu vo varáku a prietoku nástreku. Avšak jeden z meracích prístrojov, konkrétne prietok nástreku, bol zle nakalibrovaný. Inštalácia dodatočných senzorov a analyzátorov by bola ekonomicky náročná bez garancie dostatočnej presnosti. Tento problém sme vyriešili návrhom a následnou implementáciou tzv. inferenčných senzorov, ktoré slúžia na monitorovanie ťažko merateľných veličín. Jedná sa o matematické modely, ktoré na základe dostupných premenných dokážu odhadnúť iné, ťažko merateľné premenné. Sú charakteristické nízkymi nákladmi na údržbu, keďže využívajú iné merané veličiny na výpočet v reálnom čase. Ich veľkou výhodou je kontinuálne meranie, nízka náročnosť na úložisko a predovšetkým, že sú nenáročné na údržbu. V práci sa zaoberám návrhmi inferenčných senzorov s využitým troch princípov, a tým sú dátovo orientovaný návrh, modelovo orientovaný návrh a ich kombináciou vzniká hybridný návrh. Celkovo bolo navrhnutých 9 inferenčných senzorov na indikovanie ľahko aj ťažko merateľných veličín, ktoré zahŕňajú prietok nástreku a zvyšku, výšku hladiny vo varáku, indikovanie zloženia refraktometrom, koncentrácia hlavných prúdov (nástrek, destilát a zvyšok), a posledným je návrh inferenčného senzora na indikovanie ustáleného stavu. Pri návrhu dátovo a hybridne orientovaných modelov sme využívali metódy, ktoré sú založené na lineárnej regresii. Týmito metódami sú Ordinary Least Squares Regression (OLSR) a Least Absolute Shrinkage and Selection Operator (LASSO). Vykonali sme niekoľko experimentov, počas ktorých sme nazbierali aj manuálne dáta o koncentrácii produktov. Tieto dátové sady sme následne spracovali a rozdelili na trénovacie a testovacie sady. Úspešnosť predikcie jednotlivých modelov bola vyhodnotená na základe kritéria kvality Root-Mean-Squared Error (RMSE). Zistili sme, že počas trénovania senzorov bola vždy úspešnejšia metóda OLSR na základe tohto kritéria, avšak počas testovania nižšie hodnoty RMSE dosiahol senzor navrhnutý pomocou metódy LASSO. Vyskúšali sme rôzne kombinácie dátových sád, aby sme mohli vybrať najlepšie navrhnutý senzor. Za finálny model na implementáciu do riadiacej schémy sme si vybrali senzor s najnižšou hodnotou RMSE. Pri návrhu inferenčného senzora pre ustálený stav sme vychádzali z úpravy dát. Vytvorili sme model, ktorý na základe historických dát najlepšie indikuje ustálené stavy. Úspešnosť senzora sme v tomto prípade vyhodnotili na základe počtu správne identifikovaných ustálených stavov. Po návrhu a implementácii inferenčných senzorov sme mohli prejsť k samotnému návrhu riadiacej stratégie.

V prvom kroku bolo potrebné určiť premenné, ktoré sa použijú na riadenie destilačnej kolóny. Na tento účel sme využili metódu párovania veličín, pričom sme vychádzali z riadiacej konfigurácie LV. Existuje mnoho konfigurácií, ale LV patrí medzi najpouží-

vanejšie v praxi. Táto konfigurácia obsahuje 5 regulačných slučiek, ale po niekoľkých experimentoch sme sa rozhodli, že pre náš prípad bude vhodnejšie uvažovať iba štyri a tým sú:

1. riadenie teploty na prvej etáži pootvorením refluxného ventilu,
2. riadenie teploty vypočítanej na základe priemernej hodnoty teploty zo siedmej a z ôsmej etáže ohrevom vo varáku,
3. riadenie teploty nástreku na vstupe do destilačnej kolóny výkonom predohrievača,
4. riadenie výšky hladiny vo varáku s čerpadlom na odvádzanie zvyšku.

Cieľom nášho návrhu bolo implementovať PI regulátory, ale na základe našich skúseností sme sa rozhodli neuvažovať integračnú zložku. Tento člen spôsoboval nestabilitu procesu a pomalšie sa ustaľovala destilačná kolóna po skokových zmenách. Výsledkom sú regulačné slučky iba s P zložkou, pričom každá z nich bola navrhnutá na základe metódy pokus-omyl. Hoci sme nedosiahli žiadané veličiny, kolóna nebola zaťažená a odchýlky boli pre nás akceptovateľné. Nevýhodou takejto riadiacej schémy je množstvo jednoslučkových regulátorov. Preto by vhodnejšou alternatívou mohol byť výstupný spätnoväzbový lineárny kvadratický regulátor (LQR), ktorý uvažuje viac vstupov a viac výstupov. Základom tejto metódy je minimalizácia účelovej funkcie. V účelovej funkcii je možné nastaviť váhové matice Q a R, ktoré určujú, či žiadame rýchlejšie alebo presnejšie riadenie. Na návrh tohto regulátora sme použili toolbox `oflqr` od [14]. Pre návrh takého regulátora je však potrebné spoznať proces. Preto bolo nutné vykonať si identifikáciu systému. Vykonali sme niekoľko skokových zmien, pričom sme sa zamerali na dve veličiny, a to reflux a ohrev vo varáku. Počas návrhu P-regulátorov sme zistili, že práve tieto dve veličiny majú najväčší vplyv na separáciu zložiek, a preto sme sa rozhodli, že vstupom do LQR bude ohrev vo varáku a pootvorenie refluxného ventilu a výstupom bude teplota na prvej a ôsmej etáži. Pri porovnaní takéhoto typu regulátora s klasickou riadiacou stratégiou je LQR jednoduchšie na implementáciu a taktiež na ovládanie počas experimentov, hoci treba dodať, že bola potrebná identifikácia procesu, čo pre P-regulátory nebolo nutné.

Na záver môžeme konštatovať, že obidve stratégie majú svoje výhody, avšak na určenie vhodnejšej riadiacej stratégie je nutné vykonať ďalšie experimenty. Počas tejto práce sme využívali jednoduché modely inferenčných senzorov, aby sme dokázali zabezpečiť meranie potrebných veličín. Výhodou týchto senzorov sú nízke náklady, pričom fyzické senzory môžu byť ekonomické náročné na údržbu. Kvalita inferenčných senzorov a samotných regulátorov závisí od množstva a kvality dát, preto v budúcnosti by bolo

možné ich vylepšiť ďalšími experimentmi. Všeobecne platí, že sa podarilo zvýšiť stupeň automatizácie destilačnej kolóny oproti pôvodnému stavu a momentálne dokážeme získať informácie o 39 premenných. Všetky navrhnuté inferenčné senzory a regulátory boli implementované do riadiacej schémy a laboratórna destilačná kolóna Armfield UOP3CC je vhodná na ďalšie použitie a výskumné práce.

Appendix B

Appendices

B.1 Working procedure on the Armfield UOP3CC Laboratory Distillation Column

The LDC shown in the Figure 3.1 is operated with a control console. On the back of the control console, there are two data ports, a USB connector for connecting to a computer, and a twenty-way signal output port that displays information from all of the sensors [23, 28]. The computer which is connected to the distillation column features the recent version (R2022a) of MATLAB. Several variables can be set or displayed using the control console. The console enables control of various parts of the distillation column, including:

1. activates the reflux valve and sets up the reflux ratio timer in seconds (Figure B.1, item 1),
2. set the power input to the reboiler heater and observe the indication of the low level in the reboiler (the mixture is possible to start heating in the reboiler only in case, the desired height of the liquid in the reboiler has been reached, app. 10−12l, Figure B.1, item 2),
3. manual adjustment of the feed pump power and an indication of the motor fault (Figure B.1, item 3),
4. indication of the distillation column temperature selected by the operator (one of the trays, Figure B.1, item 4),
5. indication of the process temperature selected by the operator (usually the temperature in the reboiler, Figure B.1, item 5) [27].

The start-up of the UOP3CC Laboratory Distillation Column from Armfield can be divided into the following steps:

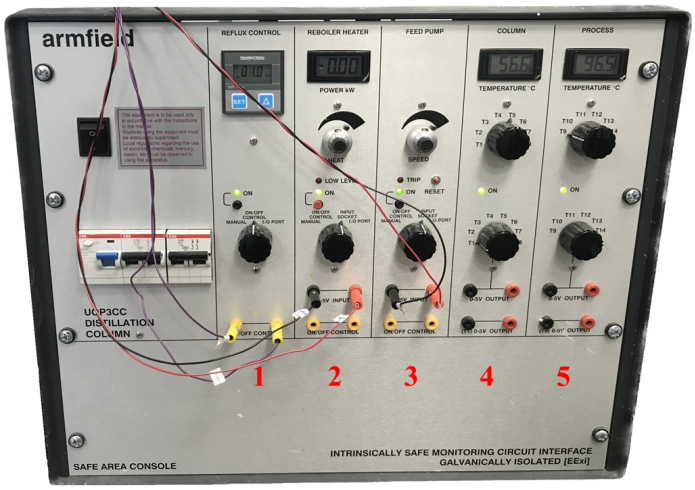


Figure B.1: Bench-mounted control console of the laboratory distillation column Armfield UOP3CC.

1. In the first step one needs to ensure proper venting of the laboratory, as we are going to work with hazardous chemicals.

2. Next, we turn on the electricity for the distillation column, i.e. all the extension cables that are responsible for starting the control console, and at the same time, we turn on the desktop computer.

3. Activate the distillation column by turning on all the necessary buttons on the front of the control console. These buttons are in charge of the different functions of the distillation column and also indicate some variables. Without activating these buttons it would not be possible to control the distillation column via the computer.

4. Open the water valve (fully), which is located on the wall (second from the left). The water will be used to cool the distillation column and will flow through the condenser. Set the flow rate of the cooling medium manually.

5. Then the containers are checked, in which the feed is contained. Ideally, both containers are full (max 4.5 l) and we put them in the designated place on the standing equipment.

6. The tubing hose must be inserted into the collection tank through the hole in the lid and positioned all the way to the bottom (one) of the container. The distillation column will draw the feed from here. If feed recycling is considered,

the tubing hoses from the feed, distillate and bottom product must be dipped into the collection tank.

7. Check the other tubing hose and valves. It is important that the tubing hoses are tight and also that they are not broken. For the pump for feed, it is important that we put the tubing hose directly into the box for the pump, then close the box and check again for any broken tubing hose. Open the distillate valve and therefore the distillate will flow directly from the distillate container into the collection tank. The valve at the bottom of the reboiler must be closed.

8. On the computer, turn on the MathWorks Matlab/Simulink program and find the necessary files to run the distillation column (*C:/UOP3CCMatlab/library/...*).

The distillation column is then prepared for experimentation [23]. When leaving, we check that everything is turned off on the distillation column in the program used. We turn off all programs on the computer, then the computer. One needs to close all valves on the distillation column and the water valve on the wall. One should switch off all buttons on the control box and then the electricity. When one leaves the laboratory, it is necessary to pull the extension cables from the power socket and close the windows.

B.1.1 List of measured variables

A list of variables currently can be measured on the Armfield UOP3CC laboratory distillation column.

Table B.1: List of measured variables.

Order	Label	Description	Unit
1.	$P_{\mathrm{F,in}}$	Feed pump power (User Input)	0−100 %
2.	$Q_{\mathrm{Reb,in}}$	Reboiler heater energy input (User Input)	0−1.75kW
3.	$\mathrm{val_c}$	Cooling medium valve opening	0−1.75kW
4.	$\mathrm{val_R}$	Reflux valve opening	0−100 %
5.	Q_{Preh}	Feed preheater energy input	0−100 %
6.	P_{B}	Bottom (waste) pump power	0−100 %
7.	$\mathrm{val_B}$	Bottom (waste) valve	on/off
8.	$\mathrm{sf_R}$	Safety variable for reflux valve	on/off
9.	$\mathrm{sf_{Preh}}$	Safety variable for feed preheater	on/off
10.	$P_{\mathrm{F,out}}$	Feed pump power (Device Output)	0−100 %
11.	$p_{\mathrm{Reb,B}}$	Pressure 1. (bottom) in reboiler	Pa
12.	$p_{\mathrm{Reb,M}}$	Pressure 2. (middle) in reboiler	Pa
13.	$p_{\mathrm{Reb,T}}$	Pressure 3. (top) in reboiler	Pa
14.	$Q_{\mathrm{Reb,out}}$	Reboiler heater energy input (Device Output)	0−100 %
15 - 22.	T_{1-8}	Temperature on the 1. - 8. tray	$^{\circ}C$
23.	T_{Reb}	Temperature in the reboiler	$^{\circ}C$
24.	T_{top}	Temperature on the top of the column	$^{\circ}C$
25.	$T_{\mathrm{c,in}}$	Temperature for the coolant entering the condenser	$^{\circ}C$
26.	$T_{\mathrm{c,out}}$	Temperature for the coolant leaving the condenser	$^{\circ}C$
27.	T_{R}	Temperature for the reflux flow	$^{\circ}C$
28.	$T_{\mathrm{F,in}}$	Temperature for the feed entering the preheater	$^{\circ}C$
29.	$T_{\mathrm{F,out}}$	Temperature for the feed leaving the preheater	$^{\circ}C$
30.	a_{Reb}	Alarm for low liquid level in reboiler	-
31.	$\hat{h}_{\mathrm{H}}$	Liquid level computed by hybrid approach	cm
32.	$\hat{h}_{\mathrm{MB}}$	Liquid level computed by model-based approach	cm
33.	$\hat{x}_{\mathrm{D}}$	Distillate product composition computed by IS	0−100 %
34.	$\hat{x}_{\mathrm{B}}$	Bottom product composition computed by IS	0−100 %
35.	$\hat{x}_{\mathrm{F}}$	Feed composition computed by IS	0−100 %
36.	$\hat{\dot{V}}_{\mathrm{B}}$	Bottom product flow rate computed by IS	$\mathrm{m^3/s}$
37.	$\hat{\dot{V}}_{\mathrm{F}}$	Feed flow rate computed by IS	$\mathrm{m^3/s}$
38.	y_{SS}	Steady state computed by IS	0−100
39.	$y_{\mathrm{SS,b}}$	Steady state computed by IS, binary signal	0/100

Bibliography

[1] F. Devínsky a kolektív. *Organická chémia*, chapter 5, pages 348–353. Osveta, Bratislava, 2. edition edition, 2013. ISBN 978-80-8063-388-2.

[2] Hossein Aghamollaei, Shiva Pirhadi, Soodabeh Shafiee, Mohammad Sehri, Vahabodin Goodarzi, and Khosrow Jadidi. Application of polymethylmethacrylate, acrylic, and silicone in ophthalmology. In Valentina Grumezescu and Alexandru Mihai Grumezescu, editors, *Materials for Biomedical Engineering*, chapter Chapter 15, pages 507–554. Elsevier, 2019.

[3] Willy Araújo, Marcílio Máximo, Jailson Nicácio, and Heleno Bispo. Data-driven digitalization of an armfield uop3cc distillation unit. *Digital Chemical Engineering*, 6, 2023.

[4] G W Bennett. A laboratory experiment on the boiling-point curves of non-azeotropic binary mixtures. *J.Chem.Educ.*, 6(12):1544–1549, 1929.

[5] Salah Bouhouche, Laib dit Leksir Yazid, Hazem Tarek, and Bast Jurgen. Inferential sensor – based adaptive principal components analysis for mechanical properties prediction and evaluation. *Measurement*, 46(9):3683–3690, 2013.

[6] G Bredig and R Bayer. Die dampfdrücke des binären systems methylalkohol-wasser. *Zeitschrift für Physikalische Chemie*, 130:1–14, 1927.

[7] Annika Camehl. Penalized estimation of panel vector autoregressive models: A panel lasso approach. *International Journal of Forecasting*, 2022.

[8] Liang Cao, Feng Yu, Fan Yang, Yankai Cao, and R. Bhushan Gopaluni. Data-driven dynamic inferential sensors based on causality analysis. *Control Engineering Practice*, 104:104626, 2020.

[9] Rajamani Doraiswami and Lahouari Cheded. Robust model-based soft sensor: Design and application. *IFAC Proceedings Volumes*, 47(3):5491–5496, 2014. 19th IFAC World Congress.

[10] Salvador García, Julián Luengo, and Francisco Herrera. Tutorial on practical tips of the most influential data preprocessing algorithms in data mining. *Knowledge-Based Systems*, 98:1–29, 2016.

[11] Sambit Ghosh, Shu Yang, and B. Wayne Bequette. Inferential modeling and soft sensors. In Masoud Soroush, Michael Baldea, and Thomas F. Edgar, editors, *Smart Manufacturing*, chapter 12, pages 323–351. Elsevier, 2020.

[12] A.KRÜSS Optronic GmbH. *Techniacal Data – Digital Handheld Refractometers: DR201-95*. Germany, 9. 2023.

[13] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009.

[14] Adrian Ilka. Matlab/octave toolbox for structurable and robust output-feedback lqr design. *IFAC*, 51(4):598–603, 2018. 3rd IFAC Conference on Advances in Proportional-Integral-Derivative Control PID 2018.

[15] Val Tech Diagnostics Inc. Methanol safety data sheet., 2020. [online]. Available at: https://www.labchem.com/tools/msds/msds/VT430.pdf.

[16] J. Longauer J. Dojčanský. *Chemické inžinierstvo II*, pages 149–214. Malé Centrum, Bratislava, 1. edition edition, 2000. ISBN 80-967064-8-9.

[17] M. Fikar J. Mikleš. *Process Modelling, Identification, and Control*. Berlin Heidelberg: Springer Berlin Heidelberg, New York, 2007. ISBN 978-3-540-71969-4.

[18] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning with Applications in R*. Springer Texts in Statistics. Springer, 2. edition edition, 2014.

[19] Michael A. Johnson and Mohammad H. Moradi. *PID Control: New Identification and Design Methods*. Springer, 1. edition edition, 2005.

[20] Petr Kadlec, Bogdan Gabrys, and Sibylle Strandt. Data-driven soft sensors in the process industry. *Computers & Chemical Engineering*, 33(4):795–814, 2009.

[21] M. King. *Process Control: A Practical Approach*. John Wiley & Sons Ltd., 2011.

[22] Norbert Kockmann. History of distillation. In Andrzej Górak and Eva Sorensen, editors, *Distillation: Fundamentals and Principles*, chapter 1, pages 1–43. Academic Press, Boston, 2014.

[23] Imrich Koncz. Creating video instructions and video presentations of our laboratory experiments, 2017. Bachelor thesis, Slovak University of Technology in Bratislava.

[24] Darko Križan. Software sensors for industry. Master's thesis, Slovak University of Technology in Bratislava, Bratislava, Slovakia, 2021.

[25] Max Kuhn and Kjell Johnson. *Applied Predictive Modeling*. Springer New York, 2013.

[26] Frank L. Lewis, Draguna L. Vrabie, and Vassilis L. Syrmos. *Optimal Control*. John Wiley & Sons, Hoboken, 3. edition edition, 2012.

[27] Armfield Limited. *Continuous Distillation Column UOP3CC, Instruction Manual*. England, 9. 2007.

[28] Armfield Limited. Uop3 distillation columns product description., 2023. [online]. Available on: https://armfield.co.uk/product/uop3-distillation-columns.

[29] B.G. Liptak. *Instrument Engineers' Handbook, Volume II: Process Control and Optimization*. CRC Press, 4. edition edition, 2006.

[30] Velislava Lyubenova, Georgi Kostov, and Rositsa Denkova-Kostova. Model-based monitoring of biotechnological processes—a review. *Processes*, 9(6), 2021.

[31] M. Maeder, N. McCann, S. Clifford, and G. Puxty. 3.21 - model-based data fitting. In Steven Brown, Romà Tauler, and Beata Walczak, editors, *Comprehensive Chemometrics*, pages 477–495. Elsevier, Oxford, 2. edition edition, 2020.

[32] MathWorks. *System Identification Toolbox*. MathWorks, Natick, MA, 2021.

[33] W. Mendenhall, R.J. Beaver, and B.M. Beaver. *Introduction to Probability and Statistics*. Brooks/Cole, 2009.

[34] Vu Trieu Minh and Ahmad Majdi Abdul Rani. Modeling and control of distillation column in a petroleum process. *Mathematical Problems in Engineering*, 2009, 6. 2009. ISSN 404702.

[35] M. Mojto, R. Paulen, K. Ľubušký, and M. Fikar. Modelling and analysis of control pairings of an industrial depropanizer column. In *Advanced Process Modelling Forum 26-27 March 2019*, pages 5–6. Process Systems Enterprise Ltd., 2019.

[36] Martin Mojto. Advanced process control of a depropanizer column. Master's thesis, Slovak University of Technology in Bratislava, Bratislava, Slovakia, 2019.

[37] Martin Mojto, Karol Ľubušký, Miroslav Fikar, and Radoslav Paulen. Data-based design of inferential sensors for petrochemical industry. *Computers & Chemical Engineering*, 153:107437, 2021.

[38] Tímea Mészárosová. Modelovanie laboratórnej destilačnej kolóny v gproms modelbuilder, 2021. Bachelor thesis, Slovak University of Technology in Bratislava.

[39] Anuj Narang, Amos Ben-Zvi, Artin Afacan, David Sharp, Sirish L. Shah, and Biao Huang. Undergraduate design of experiment laboratory on analysis and optimization of distillation column. *Education for Chemical Engineers*, 7(4):e187–e195, 2012.

[40] Gérard C. Nihous. Notes on hydrostatic pressure. *Journal of Ocean Engineering and Marine Energy*, 2, 2016.

[41] S. Joe Qin and Yiren Liu. Stable lasso for model structure learning of inferential sensor modeling. *IFAC-PapersOnLine*, 54(7):228–233, 2021. 19th IFAC Symposium on System Identification SYSID 2021.

[42] Adnan Ripin, Siti Kholijah Abdul Mudalip, and Rosli bin Mohd Yunus. Effects of ultrasonic waves on enhancement of relative volatilities in methanol–water mixtures. *Jurnal Teknologi*, 48, 2008.

[43] G.A. Shewa and F.I. Ugwuowo. A new hybrid estimator for linear regression model analysis: Computations and simulations. *Scientific African*, 19:e01441, 2023.

[44] Sigurd Skogestad and Manfred Morari. Control configuration selection for distillation columns. *AIChE Journal*, 33(10):1620–1635, 1987.

[45] J.G. Stichlmair, H. Klein, and S. Rehfeldt. *Distillation: Principles and Practice*. Wiley, 2021.

[46] Robert Tibshirani. Regression shrinkage and selection via the lasso: A retrospective. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(3):273–282, 04 2011.

[47] Zilong Wang, Reza Kamyar, Hamidreza Mehdizadeh, and Pushkar Yashvant Pendse. Moisture soft sensor for agitated pan dryers using a hybrid modeling approach. *International Journal of Pharmaceutics*, 586:119518, 2020.

[48] W.O. Wunderlich and Tennessee Valley Authority. Water Systems Development Branch. *Heat and Mass Transfer Between a Water Surface and the Atmosphere*. Water resources research laboratory report. Tennessee Valley Authority, Office of Natural Resources and Economic Development, Division of Air and Water Resources, Water Systems Development Branch, 1972.

[49] Xin Yan and Xiao Gang Su. *Linear Regression Analysis: Theory and Computing*. World Scientific Publishing Company, 1. edition edition, 6 2009.

[50] Ľuboš Čirka, Miroslav Fikar, and J. Mikleš. Youla-kučera parameterisation approach to lq tracking and disturbance rejection problem. *IFAC Proceedings Volumes (IFAC-PapersOnline)*, 16:1005–1010, 01 2005.