SLOVAK UNIVERSITY OF TECHNOLOGY
IN BRATISLAVA

FACULTY OF CHEMICAL AND FOOD TECHNOLOGY

Reg. No.: FCHPT-16584-105421

Modeling and Optimization of Chemical-Reaction Systems

MASTER THESIS

2026 **Bc. Adam Fedor**

SLOVAK UNIVERSITY OF TECHNOLOGY IN BRATISLAVA

FACULTY OF CHEMICAL AND FOOD TECHNOLOGY

Reg. No.: FCHPT-16584-105421

Modeling and Optimization of Chemical-Reaction Systems

MASTER THESIS

Study programme:	Process Control
Study field:	Cybernetics
Training workspace:	Institute of Information Engineering, Automation, and Mathematics
Thesis supervisor:	doc. Ing. Radoslav Paulen, PhD.

2026 **Bc. Adam Fedor**

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

Faculty of Chemical and Food Technology
Academic year: 2025/2026
Reg. No.: FCHPT-16584-105421

STU FCHPT

MASTER THESIS TOPIC

Student: **Bc. Adam Fedor**
Student's ID: 105421
Study programme: Automation and Information Engineering in Chemistry and Food Industry
Study field: Cybernetics
Thesis supervisor: doc. Ing. Radoslav Paulen, PhD.
Head of department: prof. Ing. Miroslav Fikar, DrSc.
Workplace: Department of Information Engineering and Process Control

Topic: **Modeling and Optimization of Chemical-Reaction Systems**

Language of thesis: English

Specification of Assignment:

The main objective of the project is to develop a mathematical model of a system with a chemical reaction (e.g., a batch reactor) and subsequently use it to optimize operating conditions. The model will be based on mass and energy balance laws and will describe the key dynamic properties of the process. Based on the model, an optimization problem will be formulated with the objective of, for example, maximizing the yield of the desired product or minimizing energy consumption while satisfying safety and quality constraints. The result will represent the optimal values of process variables (e.g., temperature and concentration).

Tasks:
Process analysis and formulation of mass and energy balances
Design of the mathematical model structure of the chemical reactor
Estimation of model parameters based on experimental or literature data
Formulation and solution of the optimization problem for optimal process operation

Selected bibliography:

1. D. Zhao, X. Wang, J. B. Miller, G. W. Huber, ChemSusChem, 2020, 13 (7), 1764-1774. doi: https://doi.org/10.1002/cssc.201903434

Acknowledgment

I would like to express my sincere gratitude to my supervisor, doc. Ing. Radoslav Paulen, PhD., for his guidance, valuable feedback, and continuous support throughout the preparation of this thesis. His expertise and constructive suggestions significantly contributed to the direction and quality of this work.

Abstract

This thesis addresses the general problem of process modelling and optimisation based on experimental data that may be physically inconsistent and therefore cannot be used directly for reliable model development. The main focus is on the use of data reconciliation to obtain a physically consistent dataset satisfying mass and elemental balance constraints, which can subsequently be used for the development of process models.

As a model case, the thesis considers the pyrolysis of polyethylene. Experimental yield data are first converted to a common mass basis and reconciled by weighted least squares. The reconciled data are then used to develop product-yield models based on linear regression and Gaussian Process Regression. Polyethylene conversion is described using an Arrhenius-type random-scission model, which is implemented together with the yield correlations in a steady-state CSTR model in gPROMS. The model is then used for simplified economic optimisation and global system analysis. The results show that data reconciliation improves the physical consistency of the experimental data and that the obtained optimum is strongly affected by the residence-time constraint and by the assumptions of the simplified economic objective.

Keywords: polyethylene pyrolysis; data reconciliation; yield modelling; random-scission kinetics; Gaussian Process Regression; CSTR reactor model; process optimisation

Abstrakt

Táto diplomová práca sa zaoberá všeobecným problémom modelovania a optimalizácie procesov na základe experimentálnych údajov, ktoré môžu byť fyzikálne nekonzistentné, a preto nemusia byť priamo použiteľné na spoľahlivý vývoj modelu. Hlavný dôraz je kladený na použitie rekonciliácie dát s cieľom získať fyzikálne konzistentný súbor údajov, ktorý spĺňa hmotnostné a elementárne bilančné obmedzenia a ktorý môže byť následne použitý na vývoj procesných modelov.

Ako modelový príklad je v práci uvažovaná pyrolýza polyetylénu. Experimentálne údaje o výťažkoch sú najprv prevedené na jednotnú hmotnostnú bázu a následne rekonciliované metódou vážených najmenších štvorcov. Rekonciliované dáta sú potom použité na vývoj modelov výťažkov produktov založených na lineárnej regresii a regresie Gaussoveho procesu. Konverzia polyetylénu je opísaná Arrheniovým modelom náhodného štiepenia reťazcov, ktorý je spolu s koreláciami výťažkov implementovaný do ustáleného stavu modelu prietokového reaktora v prostredí gPROMS. Model je následne použitý na zjednodušenú ekonomickú optimalizáciu a globálnu systémovú analýzu. Výsledky ukazujú, že rekonciliácia dát zlepšuje fyzikálnu konzistentnosť experimentálnych údajov a že získané optimum je výrazne ovplyvnené obmedzením zádržného času a predpokladmi zjednodušenej ekonomickej účelovej funkcie.

Kľúčové slová: pyrolýza polyetylénu; rekonciliácia dát; modelovanie výťažkov; kinetika náhodného štiepenia; regresia Gaussovho procesu; model CSTR reaktora; procesná optimalizácia

Contents

Declaration of Honour

I hereby declare that I have prepared this diploma thesis independently, based on the acquired theoretical knowledge and using the listed professional literature and other sources, which are properly cited in the thesis.

Artificial intelligence tools were also used during the preparation of the thesis, mainly for linguistic, stylistic, and formulation support. The professional content of the thesis, the methods used, the interpretation of the results, and the conclusions are the result of my own work.

List of symbols

Symbol	Description
Kinetic and degradation variables	
x	Bond-cleavage variable, fraction of broken polymer links.
x_{in}	Inlet value of the bond-cleavage variable.
α	Conversion / normalized mass-loss proxy obtained from the random-scission relation.
$A,\ A_1,\ A_2$	Arrhenius pre-exponential factors.
$E,\ E_1,\ E_2$	Activation energies.
R	Universal gas constant.
T	Absolute temperature.
T_C	Temperature in degrees Celsius.
T_{in}	Inlet temperature.
t	Time.
t_{op}	Selected operating time.
β_T	Heating rate, $\beta_T = \mathrm{d}T/\mathrm{d}t$.
N_0	Initial number of polymer linkages.
L	Minimum non-volatile fragment length.
$m_0,\ m,\ m_f$	Initial, current, and final sample mass in the degradation model.
Masses, phases, and elemental balances	
$m_{\text{in}},\ m_{\text{out}}$	Inlet and outlet mass.
Δm_{acc}	Accumulation or hold-up difference.
$m_p,\ m_{\text{gas}},\ m_{\text{liq}},\ m_{\text{char}}$	Phase mass, gas mass, liquid mass, and char mass.
$m_{p,\text{out}}$	Outlet mass of phase p.
$m_{e,\text{in}},\ m_{e,\text{out}}$	Inlet and outlet mass of element e.
$m_{e,p,\text{out}}$	Outlet mass of element e in phase p.
$m_{C,s},\ m_{H,s},\ m_{O,\text{CO}}$	Carbon, hydrogen, and oxygen masses assigned to species.
m_s	Mass of species s on the reconciliation basis.
$m_{\text{H}_2},\ m_{\text{CO}}$	Hydrogen and carbon monoxide masses.
ω_p	Phase mass fraction.

Symbol	Description
$\omega_{e,\mathrm{in}},\ \omega_{e,p}$	Elemental mass fractions in the inlet and in phase p.
$\omega_{C,\mathrm{PE}}$	Carbon mass fraction of polyethylene.
$\omega_{C,s},\ \omega_{H,s}$	Carbon and hydrogen mass fractions in species s.
$\omega_{O,\mathrm{CO}}$	Oxygen mass fraction in carbon monoxide.
Yields, amounts, and molar quantities	
Y_C	Carbon yield.
$Y_p^{\mathrm{w\%}}$	Phase yield reported as mass percentage.
$Y_s^{\mathrm{C\%}}$	Species yield reported as carbon yield.
Y_s^{emp}	Empirical mass yield of product s.
$\widetilde{Y}_{C,s}$	Carbon-based empirical yield of product s.
Y	Generic yield or output variable.
$n_C,\ n_{C,\mathrm{in}}$	Carbon amount and inlet carbon amount.
$n_{\mathrm{H_2}},\ n_{\mathrm{nonH_2}}$	Hydrogen amount and total amount of non-hydrogen gas species.
$\chi_{\mathrm{H_2}}$	Hydrogen mole fraction in the gas phase.
$M_C,\ M_H$	Atomic molar masses of carbon and hydrogen.
M_s	Molar mass of species s.
$M_{C,s}$	Mass of carbon contained in one mole of species s.
$M_{\mathrm{H_2}}$	Molar mass of hydrogen.
Reactor-model variables and mass flow rates	
$k_1,\ k_2$	Kinetic rate constants in the reactor model.
r_{kin}	Kinetic bond-cleavage rate.
τ	Residence time.
V	Reactor volume.
$\dot{V}_{\mathrm{in}}$	Inlet volumetric flow rate.
ρ_{in}	Inlet density.
$\dot{m}_{\mathrm{in}}$	Inlet mass flow rate.
$\dot{m}_{\mathrm{out}}$	Outlet mass flow rate.
$\dot{m}_{\mathrm{conv}}$	Converted polyethylene mass flow rate.
$\dot{m}_{\mathrm{PE,unconv}}$	Unreacted polyethylene mass flow rate.
$\dot{m}^*_{\mathrm{PE}}$	Polyethylene-like remaining stream.
$\dot{m}_{\mathrm{rem}}$	Unassigned converted material mass flow rate.

Symbol	Description
$\dot{m}_{C,\mathrm{in}},\ \dot{m}_{C,\mathrm{conv}}$	Inlet and converted carbon mass flow rates.
$\dot{m}_{C,g}$	Carbon mass flow assigned to fragment class g.
$\dot{m}_{C,s}$	Carbon mass flow in species s.
$\dot{m}_s$	Mass flow rate of species s.
$\dot{m}_{\mathrm{H_2}},\ \dot{m}_{\mathrm{char}}$	Hydrogen and char mass flow rates.
$\dot{m}_{\mathrm{prod,exp}}$	Explicitly calculated product-stream mass flow rate.
Random-scission and product-allocation quantities	
ϕ_g	Random-scission contribution of fragment class g.
$\phi_{\mathrm{sum}},\ \phi_{\mathrm{rest}}$	Sum of explicitly represented fragment contributions and remaining part of the distribution.
$\psi_{C,s}$	Carbon-based allocation to species s.
$\Sigma_{C2}^{\mathrm{emp}},\ \Sigma_{C4}^{\mathrm{emp}}$	Empirical carbon-yield sums for the C_2 and C_4 product groups.
Regression variables	
$u,\ \mathbf{u}$	Scalar regressor and input vector.
$u_r,\ u_{rj}$	Regressor value in observation r and value of regressor j in observation r.
$\mathbf{u}'$	Second input vector used in kernel definitions.
$y,\ y_r$	Response variable and response value in observation r.
$\hat{y},\ \hat{Y}$	Fitted response and fitted yield/output variable.
$\bar{u},\ \bar{y}$	Sample means of the regressor and response.
$\theta_0,\ \theta_1,\ \theta_j$	Regression coefficients.
$\theta_T,\ \theta_\tau$	Temperature and residence-time regression coefficients.
$\theta_{0,s},\ \theta_{T,s},\ \theta_{\tau,s}$	Linear empirical-yield coefficients for species s.
$\boldsymbol{\theta},\ \hat{\boldsymbol{\theta}}$	Regression coefficient vector and its least-squares estimate.
$\boldsymbol{\theta}_Y$	Coefficient vector for a generic output variable Y.
$\varepsilon,\ \varepsilon_r$	Regression error term and error in observation r.
$\boldsymbol{\varepsilon}$	Vector of regression errors.
σ_ε^2	Regression error variance.

Symbol	Description
δ_r	Residual at observation r.
$\mathbf{y},\ \mathbf{Y},\ \mathbf{U}$	Response vector, output vector, and regression matrix.
$\mathbf{1}$	Vector of ones used in the regression matrix.
S	Least-squares objective function.
$S_{uu},\ S_{uy}$	Scatter terms used in ordinary least squares.
N_{obs}	Number of observations.
n_u	Number of regressors or input dimensions.
n_z	Number of reconciled variables.
n_{exp}	Number of experiments.
Gaussian-process regression and model-quality metrics	
$f(\mathbf{u})$	Gaussian-process latent function.
$\mu(\mathbf{u})$	Gaussian-process mean function.
$k(\mathbf{u}, \mathbf{u}')$	Gaussian-process covariance kernel.
$\sigma_f^2,\ \sigma_n^2$	Gaussian-process signal and noise variances.
ℓ_d	Gaussian-process lengthscale in dimension d.
$\rho(\mathbf{u}, \mathbf{u}')$	Scaled distance in ARD kernels.
α_{RQ}	Rational-quadratic kernel shape parameter.
RMSE	Root mean square error.
R^2	Coefficient of determination.
$SS_{\text{res}},\ SS_{\text{tot}}$	Residual and total sums of squares.
σ_{noise}	Standard deviation of synthetic noise used in the robustness test.
Data reconciliation variables	
$\mathbf{z}^{\text{meas}},\ \mathbf{z}^{\text{rec}}$	Measured and reconciled variable vectors.
$z_j^{\text{meas}},\ z_j^{\text{rec}}$	Measured and reconciled value of variable j.
$\sigma_j,\ \sigma_s$	Variable-level and species-level uncertainty proxies.
$\Delta Y_{s,r}$	Reported absolute error of species yield s in experiment r.
λ_j	Reconciliation weight.
$\mathbf{\Lambda}$	Diagonal reconciliation weight matrix.

Symbol	Description
J	Reconciliation objective function.
Economic and energy variables	
$\Pi_{\mathrm{EUR}},\ \dot{\Pi}_{\mathrm{EUR}}$	Total profit and profit rate.
$\dot{\mathcal{R}}_{\mathrm{EUR}}$	Product revenue rate.
$\dot{C}_{\mathrm{heat,EUR}},\ \dot{C}_{\mathrm{waste,EUR}}$	Heating-cost rate and waste-energy credit rate.
$\dot{Q}_{\mathrm{in}}$	Heat-input rate.
$\dot{Q}_{\mathrm{waste,PE}}$	Recoverable heat rate from the polyethylene-like remaining stream.
$\dot{Q}_{\mathrm{waste,char}}$	Recoverable heat rate from char.
$\dot{Q}_{\mathrm{waste,total}}$	Total recoverable heat rate.
$c_{p,\mathrm{PE}}$	Specific heat capacity of polyethylene.
c_s^{prod}	Product price of species s.
$c_{\mathrm{el}},\ c_{\mathrm{ng}}$	Electricity and natural-gas prices on a kWh basis.
$c_{\mathrm{el,J}},\ c_{\mathrm{ng,J}}$	Electricity and natural-gas prices converted to a joule basis.
$c_{\mathrm{heat,J}}$	Effective heating-cost coefficient on a joule basis.
$\Delta h_{f,\mathrm{PE}},\ \Delta h_{R,\mathrm{PE}}$	Melting and reaction enthalpy terms of polyethylene.
$\Delta h_{\mathrm{comb,PE}},\ \Delta h_{\mathrm{comb,char}}$	Recoverable-energy coefficients for polyethylene-like material and char.
$\eta_{\mathrm{waste}},\ \eta_{\mathrm{char}}$	Waste-heat and char-energy recovery efficiencies.
Sets	
$\mathcal{P},\ \mathcal{E},\ \mathcal{S},\ \mathcal{H}$	Sets of phases, elements, gaseous species, and explicit hydrocarbon products.

Indices and labels

Index or label	Description
e	Element index, $e \in \mathcal{E} = \{C, H, O\}$.
p	Phase index, $p \in \mathcal{P}$.
s	Species index.

Index or label	Description
j	Generic variable or regressor index.
r	Observation or experiment index.
d	Input-dimension index.
g	Fragment-size index.
$\mathrm{in}, \mathrm{out}$	Inlet and outlet labels.
$\mathrm{meas}, \mathrm{rec}$	Measured and reconciled labels.
$\mathrm{conv}, \mathrm{unconv}, \mathrm{rem}$	Converted, unconverted, and remaining material labels.
$\mathrm{gas}, \mathrm{liq}, \mathrm{char}$	Gas, liquid, and char phase labels.
$\mathrm{sum}, \mathrm{rest}$	Summed explicit part and remaining part.
emp	Empirical-correlation label.
prod	Product-price or product-term label.
exp	Explicitly calculated product-stream label, or experiment-count label in n_{exp}.
$\mathrm{true}, \mathrm{fit}$	True and fitted parameter labels used in parameter-recovery tests.
noise	Synthetic-noise label.
PE	Polyethylene label.

Chapter 1

Introduction

In 2015, humans had produced 8.3 billion tons of virgin plastic, of which 6.3 billion tons became plastic waste [7]. In 2010, 275 million tons of plastic waste were estimated to be generated, with 4.8 to 12.7 million tons entering the ocean [8]. The global plastic production continued to increase over years due to the vast application of plastics in many sectors [15]. This leads to an urgent need for sustainable end-of-life management strategies beyond traditional landfilling.

Among the various recovery methods, thermochemical conversion offers a promising pathway to transform waste plastics into valuable resources. In this context, pyrolysis represents a relevant process because it converts polymeric material into gaseous, liquid, and solid products under oxygen-free conditions [2, 14]. However, the process is technologically complex and its sustainable operation requires mathematical models suitable for process analysis and optimization.

A key challenge in polyethylene (PE) pyrolysis modeling is to obtain a reliable and physically valid process model from experimental data that are imperfect, uncertain and often not fully consistent with mass and elemental balances.

Recent literature describes plastic pyrolysis not only as a waste-to-energy route, but also as a chemical recycling pathway capable of producing liquid hydrocarbons, gases, char, and petrochemical feedstocks from end-of-life plastics [21, 19]. The product distribution is strongly governed by the polymer structure and operating conditions. In the case of polyolefins such as polyethylene and polypropylene, thermal degradation proceeds mainly through free-radical chain scission reactions, producing a broad mixture of alkanes, alkenes, dienes, waxes, and light gaseous hydrocarbons [21, 11]. Temperature, residence time, heating rate, pressure, catalyst selection, and reactor configuration are repeatedly identified as key variables controlling the extent of cracking, secondary reactions, aromatization, coke formation, and the final quality of the pyrolysis products [21, 19].

From a process-design perspective, reactor selection and heat-transfer limitations are particularly important. Fluidized-bed reactors are often considered attractive for plastic pyrolysis because of their efficient heat and mass transfer, good temperature uniformity, residence-time control, and scale-up potential [11, 19]. However, even under similar feed and temperature conditions, different reactors can lead to different product distributions due to differences in heating rate, vapor residence time, contact with solids, and secondary vapor-phase reactions [6]. Recent modelling studies further emphasize that plastic pyrolysis is controlled by the coupled effect of reaction kinetics and heat transfer, especially when particle size, external heat supply, and endothermic reaction effects become significant [20].

These findings show that pyrolysis modelling cannot rely only on simplified stoichiometric or empirical descriptions. A useful model should account, at least in simplified form, for the dependence of conversion and product yields on operating conditions, while also respecting physical constraints such as mass and elemental balances. This is especially relevant when published experimental datasets are used for model development, because such data may contain measurement uncertainty, inconsistent reporting bases, or incomplete material closure.

1.1 Objectives and Scope

The main objective of this thesis is to develop and evaluate a modelling framework for polyethylene pyrolysis that combines kinetic conversion modelling, data reconciliation, yield modelling, reactor simulation, and simplified process optimization. The work focuses on the implementation and verification of an Arrhenius-type random-scission model, preprocessing and reconciliation of published LDPE pyrolysis data, development of empirical yield models, and implementation of a steady-state CSTR reactor model in gPROMS.

The scope of the thesis is limited to computational model development and analysis within the operating range supported by the available experimental data. The reactor model is formulated as a simplified steady-state CSTR, and the economic objective includes only product revenue, heating cost, and a recoverable-energy credit from unconverted polyethylene and char. Detailed reactor design, separation units, compression, purification requirements, capital costs, and a complete techno-economic analysis are outside the scope of this work.

1.2 Thesis Structure

The thesis is organized into five main chapters. Chapter 1 introduces the motivation, objectives, and scope of the work. Chapter 2 presents the theoretical background of polyethylene pyrolysis, random-scission kinetics, mass-balance and data-reconciliation principles, regression methods used for yield modelling, and the formulation of the steady-state CSTR reactor model, including the product-yield representation, simplified economic objective, optimization setup, and global system analysis methodology. Chapter 3 describes the experimental data processing, basis harmonisation, reconciliation procedure, calculation of mass yields, and yield-modelling workflow. Chapter 4 presents and discusses the obtained results, including kinetic model verification, reconciliation results, yield-modelling performance, reactor optimization, product-distribution analysis, and global system analysis interpretation. Chapter 5 summarizes the main conclusions, limitations of the developed modelling framework, and possible directions for future work.

Chapter 2

Theory

2.1 Pyrolysis Concept

Pyrolysis is the thermal degradation of organic materials in the absence of oxygen [2, 14], usually performed at 400–600°C. During pyrolysis, the polymer is heated and decomposed into smaller molecules, producing gaseous and liquid fractions and, in some cases, a solid residue [14]. The gaseous fraction can contain plastic-derived monomers that can be recovered and used to produce new polymers, as well as hydrogen that can be used as a reactant in further processing (e.g., hydrogenation) or as a low-carbon energy carrier. In contrast, heavier hydrocarbons are typically concentrated in the liquid fraction and can be utilized as fuels.

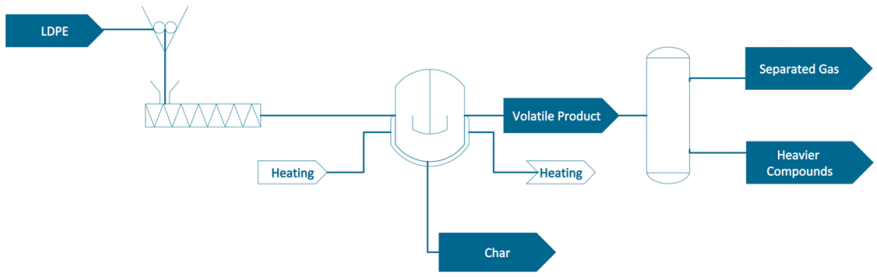

Figure 2.1: Schematic representation of the reactor model.

2.1.1 Thermal Degradation of LDPE

Polyethylene (PE) pyrolysis has been studied using kinetic models and more detailed mechanistic descriptions. [13] Experimental studies show that operating conditions,

especially temperature and residence time, strongly influence the distribution of gaseous, liquid, and solid products, which motivates predictive yield models for process analysis and optimization [22]. However, published yield datasets can exhibit imperfect balance closure and non-negligible experimental uncertainty, which can distort regression or data-driven modeling if the data are used without preprocessing [22].

2.1.2 Conversion and Interpretation

In this work, the bond-cleavage variable x is interpreted as an internal kinetic state that describes the fraction of broken polymer links. It is not identical to the overall mass conversion. The experimentally observable conversion or normalized weight loss is obtained from x through the random-scission relationship. This distinction is important because the reactor model uses x to describe the kinetic progress of degradation, while the mass flows of converted and unconverted polyethylene are calculated from the corresponding conversion expression.

2.2 Kinetic Description

2.2.1 Random-Scission Mechanism

This model links normalized weight loss α, which describes the normalized mass loss of volatile compounds, with the bond-cleavage variable x, which represents the fraction of broken linkages. Both variables take values in the interval $[0, 1]$: $\alpha = 0$ denotes no mass loss and $\alpha = 1$ denotes complete volatilization, while $x = 0$ denotes no scission and $x = 1$ denotes complete scission.

Bond-cleavage kinetics are given by

$$\frac{\mathrm{d}x}{\mathrm{d}T} = \frac{A}{\beta_T} \exp\left(-\frac{E}{RT}\right)(1 - x), \tag{2.1}$$

where A is the Arrhenius pre-exponential factor, E is the activation energy, R is the gas constant, and β_T is the heating rate, with $\beta_T = \frac{\mathrm{d}T}{\mathrm{d}t}$.

The normalized weight loss α is related to bond cleavage by [22]

$$\alpha = \frac{m_0 - m}{m_0 - m_f} = 1 - (1 - x)^{L-1}\left[1 + x\frac{(N_0 - L)(L - 1)}{N_0}\right], \tag{2.2}$$

where N_0 represents the initial number of linkages in the polymeric chain, L represents the minimal chain length that is not volatile, and m_0, m, and m_f denote the initial

sample mass, the sample mass at a given time, and the final residual mass, respectively. This model provides a reasonable compromise between predictive accuracy and mathematical simplicity, with only A and E as adjustable parameters. More complicated models such as parallel first-order models can be used for better results.

2.3 Mass Balances and Data Reconciliation

2.3.1 Total Mass Balance and Phase Yields

As mentioned earlier, in an ideal experiment the total mass of products leaving the reactor equals the mass of input entering the reactor. In practice, measurement errors lead to imperfect closure; therefore, mass-balance relations are used as physical constraints in the subsequent reconciliation step. Ensuring mass-balance closure is essential because it enforces physical consistency of the dataset, preventing biased parameter estimates and unreliable predictions in subsequent modeling and optimization. The mass balance is usually calculated in weight percents (w%) or, in the case of pyrolysis, it is often also disclosed in carbon yields.

The overall mass balance on a per-experiment mass basis is written as

$$m_{\text{in}} = m_{\text{out}} + \Delta m_{\text{acc}}, \tag{2.3}$$

where m_{in} is the total mass of polymer fed to the reactor, m_{out} is the total mass of all recovered products, and Δm_{acc} denotes the net mass accumulation or hold-up difference over the experiment. The balance relations are applied as overall closure constraints on a per-experiment basis. Any net accumulation or hold-up difference is neglected at this level, i.e.

$$\Delta m_{\text{acc}} \approx 0. \tag{2.4}$$

In this work, the outlet stream is partitioned into three phases,

$$\mathcal{P} = \{\text{gas, liq, char}\}, \tag{2.5}$$

so that the total outlet mass is the sum of phase masses,

$$m_{\text{out}} = \sum_{p \in \mathcal{P}} m_{p,\text{out}}. \tag{2.6}$$

The phase mass fractions ω_p (dimensionless) are defined by

$$m_{p,\text{out}} = \omega_p \, m_{\text{out}}, \qquad p \in \mathcal{P}, \tag{2.7}$$

and must satisfy the normalization constraint,

$$\sum_{p\in\mathcal{P}} \omega_p = 1. \tag{2.8}$$

For completeness, product yields are sometimes also reported on a carbon basis. A generic carbon yield is defined as

$$Y_C\ [\%] \;=\; 100\,\frac{n_C}{n_{C,\mathrm{in}}}, \tag{2.9}$$

where n_C is the amount of carbon (in moles) contained in the reported phase or compound, and $n_{C,\mathrm{in}}$ is the total amount of carbon in the input.

2.3.2 Elemental Balance

In addition to total mass closure, the dataset is constrained by elemental conservation. Let

$$\mathcal{E} = \{C, H, O\} \tag{2.10}$$

denote the set of considered elements. For each element $e \in \mathcal{E}$, the input mass of element e must be equal to the total outlet mass of element e:

$$m_{e,\mathrm{in}} = m_{e,\mathrm{out}}, \qquad e \in \mathcal{E}. \tag{2.11}$$

Using inlet elemental mass fractions $\omega_{e,\mathrm{in}}$ (dimensionless), the inlet elemental masses are

$$m_{e,\mathrm{in}} = \omega_{e,\mathrm{in}}\, m_{\mathrm{in}}, \qquad e \in \mathcal{E}. \tag{2.12}$$

On the outlet side, the elemental mass is the sum over phases:

$$m_{e,\mathrm{out}} = \sum_{p\in\mathcal{P}} m_{e,p,\mathrm{out}}, \qquad e \in \mathcal{E}. \tag{2.13}$$

Defining $\omega_{e,p}$ as the mass fraction of element e within phase p, we write

$$m_{e,p,\mathrm{out}} = \omega_{e,p}\, m_{p,\mathrm{out}}, \qquad e \in \mathcal{E},\ p \in \mathcal{P}. \tag{2.14}$$

By definition, the elemental fractions within each phase satisfy

$$\sum_{e\in\mathcal{E}} \omega_{e,p} = 1, \qquad p \in \mathcal{P}. \tag{2.15}$$

Combining the above relations yields the compact elemental-balance constraint used in reconciliation:

$$\omega_{e,\mathrm{in}}\, m_{\mathrm{in}} = \sum_{p\in\mathcal{P}} \omega_{e,p}\, m_{p,\mathrm{out}}, \qquad e \in \mathcal{E}. \tag{2.16}$$

2.3.3 Weighted Least-Squares Reconciliation Principle

Based on the total mass and elemental balance constraints introduced above, data reconciliation can be formulated as a constrained optimization problem. The measured variables are collected in the vector $\mathbf{z}^{\text{meas}}$, while the corresponding reconciled values are denoted by $\mathbf{z}^{\text{rec}}$. The weighted least-squares reconciliation problem is written as

$$\min_{\mathbf{z}^{\text{rec}}} J = \sum_{j=1}^{n_z} \frac{\left(z_j^{\text{rec}} - z_j^{\text{meas}}\right)^2}{\sigma_j^2}, \tag{2.17}$$

subject to the total mass balance, elemental balance, phase closure, species closure, and other physical constraints introduced in the previous sections. Where required, the reconciled variables are also constrained to be non-negative:

$$\mathbf{z}^{\text{rec}} \geq \mathbf{0}. \tag{2.18}$$

In Equation (2.17), z_j^{meas} denotes the measured value of the j-th variable, while z_j^{rec} denotes its reconciled value. The index $j = 1, \ldots, n_z$ runs over all variables included in the reconciliation, where n_z is the total number of reconciled quantities. The parameter σ_j represents the assumed standard deviation of the corresponding measurement. Therefore, the weighting coefficient assigned to variable j is defined as

$$\lambda_j = \frac{1}{\sigma_j^2}. \tag{2.19}$$

The weighting coefficient controls how strongly a deviation from the measured value is penalized. Measurements with lower uncertainty have smaller σ_j, and therefore receive larger weights. As a result, these values remain closer to the reported experimental data during reconciliation. In contrast, measurements with higher uncertainty are assigned smaller weights and can be adjusted more significantly in order to satisfy the imposed balance constraints.

In this work, measurement errors are assumed to be uncorrelated. Therefore, the weighting is applied component-wise, which is equivalent to using a diagonal weight matrix. In matrix form, the objective function can be written as

$$J = \left(\mathbf{z}^{\text{rec}} - \mathbf{z}^{\text{meas}}\right)^{\top} \mathbf{\Lambda} \left(\mathbf{z}^{\text{rec}} - \mathbf{z}^{\text{meas}}\right), \tag{2.20}$$

where

$$\mathbf{\Lambda} = \operatorname{diag}\left(\frac{1}{\sigma_1^2}, \frac{1}{\sigma_2^2}, \ldots, \frac{1}{\sigma_{n_z}^2}\right). \tag{2.21}$$

When complete uncertainty information is not available, approximate uncertainty values are assigned in order to preserve the relative reliability of the individual measurements. This approach allows the reconciled dataset to satisfy the physical constraints while remaining as close as possible to the measured data according to the assumed measurement reliability.

2.4 Regression Methods for Yield Modelling

2.4.1 Linear Regression

The simple linear regression model relates a single regressor u to a response y through a straight line,

$$y = \theta_0 + \theta_1 u + \varepsilon, \tag{2.22}$$

where the parameters θ_0 and θ_1 are regression coefficients. The intercept θ_0 and the slope θ_1 represent unknown constants, and ε stands for the random error component. The errors are assumed to have mean zero and unknown variance σ_ε^2. Additionally, we usually assume that the errors are uncorrelated, meaning that each error term is independent of the others.

In the following, the regressor u is treated as known with negligible measurement error, assuming a reliable control system that maintains the prescribed experimental conditions. The response y is therefore modelled as a random variable; for each fixed u, y has a conditional distribution with mean

$$\mathbb{E}(y \mid u) = \theta_0 + \theta_1 u. \tag{2.23}$$

If the regressors were measured with non-negligible error, an errors-in-variables regression approach would be more appropriate.

The conditional variance of y is given by

$$\mathrm{Var}(y \mid u) = \mathrm{Var}(\theta_0 + \theta_1 u + \varepsilon) = \sigma_\varepsilon^2. \tag{2.24}$$

Thus, the mean of y is a linear function of u, although the variance of y does not depend on the value of u. Furthermore, because the errors are uncorrelated, the responses are also uncorrelated.

These coefficients have a simple and often useful interpretation. The slope θ_1 is the change in the mean of the distribution of y produced by a unit change in u. If the range of data on u includes $u = 0$, then the intercept θ_0 is the mean of the distribution

of the response y when $u = 0$. If the range of u does not include zero, then θ_0 has no practical interpretation [9].

The parameters θ_0 and θ_1 are unknown and must be estimated by using the least-squares method, where N_{obs} pairs of data are obtained experimentally,

$$y_r = \theta_0 + \theta_1 u_r + \varepsilon_r, \quad r = 1, 2, \ldots, N_{\text{obs}}. \tag{2.25}$$

The objective function is formulated as a least-squares criterion

$$S(\theta_0, \theta_1) = \sum_{r=1}^{N_{\text{obs}}} \left(y_r - \theta_0 - \theta_1 u_r\right)^2, \tag{2.26}$$

and subsequently minimized. The minimization yields the first-order optimality conditions

$$\left.\frac{\partial S}{\partial \theta_0}\right|_{\hat{\theta}_0, \hat{\theta}_1} = -2 \sum_{r=1}^{N_{\text{obs}}} \left(y_r - \hat{\theta}_0 - \hat{\theta}_1 u_r\right) = 0, \tag{2.27}$$

and

$$\left.\frac{\partial S}{\partial \theta_1}\right|_{\hat{\theta}_0, \hat{\theta}_1} = -2 \sum_{r=1}^{N_{\text{obs}}} \left(y_r - \hat{\theta}_0 - \hat{\theta}_1 u_r\right) u_r = 0. \tag{2.28}$$

These forms of the equation can be simplified into least-squares normal equations,

$$N_{\text{obs}} \hat{\theta}_0 + \hat{\theta}_1 \sum_{r=1}^{N_{\text{obs}}} u_r = \sum_{r=1}^{N_{\text{obs}}} y_r, \tag{2.29}$$

and

$$\hat{\theta}_0 \sum_{r=1}^{N_{\text{obs}}} u_r + \hat{\theta}_1 \sum_{r=1}^{N_{\text{obs}}} u_r^2 = \sum_{r=1}^{N_{\text{obs}}} y_r u_r. \tag{2.30}$$

The solution to the normal equations is

$$\hat{\theta}_0 = \bar{y} - \hat{\theta}_1 \bar{u}, \tag{2.31}$$

and

$$\hat{\theta}_1 = \frac{\sum_{r=1}^{N_{\text{obs}}} y_r u_r - \frac{\left(\sum_{r=1}^{N_{\text{obs}}} y_r\right)\left(\sum_{r=1}^{N_{\text{obs}}} u_r\right)}{N_{\text{obs}}}}{\sum_{r=1}^{N_{\text{obs}}} u_r^2 - \frac{\left(\sum_{r=1}^{N_{\text{obs}}} u_r\right)^2}{N_{\text{obs}}}}. \tag{2.32}$$

The last two equations represent averages of the given variables,

$$\bar{y} = \frac{1}{N_{\text{obs}}} \sum_{r=1}^{N_{\text{obs}}} y_r, \qquad \bar{u} = \frac{1}{N_{\text{obs}}} \sum_{r=1}^{N_{\text{obs}}} u_r. \tag{2.33}$$

The estimators $\hat{\theta}_0$ and $\hat{\theta}_1$ are the least-squares estimators of the intercept and slope, respectively. The fitted simple linear regression model is then

$$\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 u. \tag{2.34}$$

This equation gives a point estimate of the mean of y for a particular u. It can be simplified and put into more compact form by defining

$$S_{uu} = \sum_{r=1}^{N_{\text{obs}}} u_r^2 - \frac{\left(\sum_{r=1}^{N_{\text{obs}}} u_r\right)^2}{N_{\text{obs}}} = \sum_{r=1}^{N_{\text{obs}}} (u_r - \bar{u})^2, \tag{2.35}$$

and

$$S_{uy} = \sum_{r=1}^{N_{\text{obs}}} y_r u_r - \frac{\left(\sum_{r=1}^{N_{\text{obs}}} y_r\right)\left(\sum_{r=1}^{N_{\text{obs}}} u_r\right)}{N_{\text{obs}}} = \sum_{r=1}^{N_{\text{obs}}} y_r (u_r - \bar{u}), \tag{2.36}$$

so that

$$\hat{\theta}_1 = \frac{S_{uy}}{S_{uu}}. \tag{2.37}$$

The difference between the observed value y_r and the corresponding fitted value $\hat{y}_r$ is a residual. Mathematically, the r-th residual is

$$\delta_r = y_r - \hat{y}_r = y_r - (\hat{\theta}_0 + \hat{\theta}_1 u_r), \qquad r = 1, 2, \ldots, N_{\text{obs}}. \tag{2.38}$$

The regression model used in this thesis is a multiple regression model that is an extension of the simple regression model. The multiple regression model contains more than one regressor variable and usually describes a plane (the two-regressor case) or a higher-dimensional object (the n_u-regressor case),

$$y = \theta_0 + \theta_1 u_1 + \theta_2 u_2 + \varepsilon, \tag{2.39}$$

$$y = \theta_0 + \theta_1 u_1 + \theta_2 u_2 + \cdots + \theta_{n_u} u_{n_u} + \varepsilon, \tag{2.40}$$

and, for N_{obs} observations,

$$y_r = \theta_0 + \theta_1 u_{r1} + \theta_2 u_{r2} + \cdots + \theta_{n_u} u_{rn_u} + \varepsilon_r = \theta_0 + \sum_{j=1}^{n_u} \theta_j u_{rj} + \varepsilon_r, \qquad r = 1, 2, \ldots, N_{\text{obs}}. \tag{2.41}$$

The parameters are once again estimated using the least-squares function

$$S(\theta_0, \theta_1, \ldots, \theta_{n_u}) = \sum_{r=1}^{N_{\text{obs}}} \delta_r^2 = \sum_{r=1}^{N_{\text{obs}}} \left(y_r - \theta_0 - \sum_{j=1}^{n_u} \theta_j u_{rj} \right)^2. \tag{2.42}$$

The function S must be minimized with respect to $\theta_0, \theta_1, \ldots, \theta_{n_u}$. The least-squares estimators $\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_{n_u}$ must satisfy

$$\left.\frac{\partial S}{\partial \theta_0}\right|_{\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_{n_u}} = -2 \sum_{r=1}^{N_{\text{obs}}} \left(y_r - \hat{\theta}_0 - \sum_{j=1}^{n_u} \hat{\theta}_j u_{rj} \right) = 0, \tag{2.43}$$

$$\left.\frac{\partial S}{\partial \theta_j}\right|_{\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_{n_u}} = -2 \sum_{r=1}^{N_{\text{obs}}} \left(y_r - \hat{\theta}_0 - \sum_{j=1}^{n_u} \hat{\theta}_j u_{rj} \right) u_{rj} = 0, \qquad j = 1, 2, \ldots, n_u. \tag{2.44}$$

The corresponding normal equations can be written in expanded scalar form,

$$N_{\text{obs}} \hat{\theta}_0 + \hat{\theta}_1 \sum_{r=1}^{N_{\text{obs}}} u_{r1} + \hat{\theta}_2 \sum_{r=1}^{N_{\text{obs}}} u_{r2} + \cdots + \hat{\theta}_{n_u} \sum_{r=1}^{N_{\text{obs}}} u_{rn_u} = \sum_{r=1}^{N_{\text{obs}}} y_r, \tag{2.45}$$

$$\hat{\theta}_0 \sum_{r=1}^{N_{\text{obs}}} u_{r1} + \hat{\theta}_1 \sum_{r=1}^{N_{\text{obs}}} u_{r1}^2 + \hat{\theta}_2 \sum_{r=1}^{N_{\text{obs}}} u_{r1} u_{r2} + \cdots + \hat{\theta}_{n_u} \sum_{r=1}^{N_{\text{obs}}} u_{r1} u_{rn_u} = \sum_{r=1}^{N_{\text{obs}}} u_{r1} y_r, \tag{2.46}$$

$$\vdots$$

$$\hat{\theta}_0 \sum_{r=1}^{N_{\text{obs}}} u_{rn_u} + \hat{\theta}_1 \sum_{r=1}^{N_{\text{obs}}} u_{rn_u} u_{r1} + \hat{\theta}_2 \sum_{r=1}^{N_{\text{obs}}} u_{rn_u} u_{r2} + \cdots + \hat{\theta}_{n_u} \sum_{r=1}^{N_{\text{obs}}} u_{rn_u}^2 = \sum_{r=1}^{N_{\text{obs}}} u_{rn_u} y_r. \tag{2.47}$$

This form of notation is complicated for further use, so it is more convenient to use matrix notation and solve it afterwards. The response vector, regressor matrix, parameter vector, and error vector are defined as

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N_{\text{obs}}} \end{bmatrix}, \qquad \boldsymbol{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_{n_u} \end{bmatrix}, \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{N_{\text{obs}}} \end{bmatrix},$$

$$\mathbf{U} = \begin{bmatrix} 1 & u_{11} & u_{12} & \cdots & u_{1n_u} \\ 1 & u_{21} & u_{22} & \cdots & u_{2n_u} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & u_{N_{\text{obs}}1} & u_{N_{\text{obs}}2} & \cdots & u_{N_{\text{obs}}n_u} \end{bmatrix}. \tag{2.48}$$

The multiple linear regression model is then expressed as

$$\mathbf{y} = \mathbf{U}\boldsymbol{\theta} + \boldsymbol{\varepsilon}. \tag{2.49}$$

In general, $\mathbf{y}$ is an $N_{\text{obs}} \times 1$ vector of the observations, $\mathbf{U}$ is an $N_{\text{obs}} \times (n_u + 1)$ matrix of the regressor values, $\boldsymbol{\theta}$ is an $(n_u + 1) \times 1$ vector of the regression coefficients, and $\boldsymbol{\varepsilon}$ is an $N_{\text{obs}} \times 1$ vector of random errors.

We aim to find the vector of least-squares estimators $\hat{\boldsymbol{\theta}}$ that minimizes

$$S(\boldsymbol{\theta}) = \sum_{r=1}^{N_{\text{obs}}} \varepsilon_r^2 = \boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{U}\boldsymbol{\theta})^\top (\mathbf{y} - \mathbf{U}\boldsymbol{\theta}). \tag{2.50}$$

Note that $S(\boldsymbol{\theta})$ may be expressed as

$$\begin{aligned} S(\boldsymbol{\theta}) &= \mathbf{y}^\top \mathbf{y} - \boldsymbol{\theta}^\top \mathbf{U}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{U}\boldsymbol{\theta} + \boldsymbol{\theta}^\top \mathbf{U}^\top \mathbf{U}\boldsymbol{\theta} \\ &= \mathbf{y}^\top \mathbf{y} - 2\boldsymbol{\theta}^\top \mathbf{U}^\top \mathbf{y} + \boldsymbol{\theta}^\top \mathbf{U}^\top \mathbf{U}\boldsymbol{\theta}, \end{aligned} \tag{2.51}$$

since $\boldsymbol{\theta}^\top \mathbf{U}^\top \mathbf{y}$ is a 1×1 matrix (a scalar) and its transpose $(\boldsymbol{\theta}^\top \mathbf{U}^\top \mathbf{y})^\top = \mathbf{y}^\top \mathbf{U}\boldsymbol{\theta}$ is the same scalar.

The least-squares estimators must satisfy

$$\left.\frac{\partial S}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} = -2\mathbf{U}^\top \mathbf{y} + 2\mathbf{U}^\top \mathbf{U}\hat{\boldsymbol{\theta}} = \mathbf{0}, \tag{2.52}$$

which simplifies to

$$\mathbf{U}^\top \mathbf{U}\,\hat{\boldsymbol{\theta}} = \mathbf{U}^\top \mathbf{y}. \tag{2.53}$$

The closed-form solution is

$$\hat{\boldsymbol{\theta}} = (\mathbf{U}^\top \mathbf{U})^{-1} \mathbf{U}^\top \mathbf{y}. \tag{2.54}$$

The parameter estimation of the multivariable model follows standard regression methodology as presented in Montgomery et al. [9]. A fundamentally similar approach is presented by Fikar and Mikleš [5] in the field of system identification, where the procedure is referred to as the identification of static models. Both methodologies rely on the minimization of a quadratic objective function, leading to the same matrix-based solution for the estimated coefficients.

2.4.2 Gaussian Process Regression (GPR)

Gaussian processes are a simple and general class of models of functions [3]. A Gaussian process is any distribution over functions such that any finite set of function values has a joint Gaussian distribution. A Gaussian process model, before conditioning on data, is completely specified by its mean function:

$$\mathbb{E}[f(\mathbf{u})] = \mu(\mathbf{u}), \tag{2.55}$$

and its covariance function, also called the kernel:

$$\mathrm{Cov}[f(\mathbf{u}),\ f(\mathbf{u}')] = k(\mathbf{u}, \mathbf{u}'). \tag{2.56}$$

There are many possible choices of covariance function, and we can specify a wide range of models just by specifying the kernel of a Gaussian process. For example, linear regression, splines, and Kalman filters are all examples of Gaussian processes with particular kernels. However, these are just a few familiar examples out of a wide range of possibilities. One of the main difficulties in using Gaussian processes is constructing a kernel which represents the particular structure present in the data being modeled.

The crucial property of Gaussian processes that allows us to automatically construct models is that we can compute the marginal likelihood of a dataset given a particular model. The marginal likelihood allows one to compare models, balancing between the capacity of a model and its fit to the data.

A flexible way to model functions having more than one input is to multiply together kernels defined on each individual input. For example, a product of Squared Exponential (SE) kernels over different dimensions, each having a different lengthscale parameter, is called the ARD Squared Exponential kernel:

$$k_{\mathrm{SE,ARD}}(\mathbf{u}, \mathbf{u}') = \sigma_f^2 \prod_{d=1}^{n_u} \exp\left(-\frac{1}{2}\frac{(u_d - u_d')^2}{\ell_d^2}\right) = \sigma_f^2 \exp\left(-\frac{1}{2}\sum_{d=1}^{n_u}\frac{(u_d - u_d')^2}{\ell_d^2}\right). \tag{2.57}$$

ARD stands for automatic relevance determination, so named because estimating the lengthscale parameters $\ell_1, \ell_2, \ldots, \ell_{n_u}$ implicitly determines the relevance of each dimension. Input dimensions with relatively large lengthscales imply relatively little variation along those dimensions in the function being modeled. [3]

2.4.2.1 Kernel Functions

The choice of the covariance function, or kernel, is a fundamental aspect of Gaussian Process Regression, as it defines the prior assumptions about the shape and smoothness

of the modeled function. In practice, ARD variants of common kernels can be used.

The scaled distance between two input vectors $\mathbf{u}$ and $\mathbf{u}'$ in an ARD context is defined as

$$\rho(\mathbf{u}, \mathbf{u}') = \sqrt{\sum_{d=1}^{n_u} \frac{(u_d - u'_d)^2}{\ell_d^2}}, \tag{2.58}$$

where ℓ_d represents the characteristic lengthscale for the d-th input dimension. The following kernels are available in the `fitrgp` function that is part of the MATLAB Statistics and Machine Learning Toolbox:

- **ARD Squared Exponential:** Assumes the process is infinitely differentiable, resulting in very smooth curves:

$$k_{\mathrm{SE}}(\mathbf{u}, \mathbf{u}') = \sigma_f^2 \exp\left(-\frac{1}{2}\rho^2\right). \tag{2.59}$$

- **ARD Exponential:** Also known as the Matern 1/2 kernel, it represents a less smooth, more stochastic process:

$$k_{\mathrm{Exp}}(\mathbf{u}, \mathbf{u}') = \sigma_f^2 \exp(-\rho). \tag{2.60}$$

- **ARD Matern 3/2 and 5/2:** These kernels offer a balance by being once or twice differentiable, respectively. They are often preferred for physical processes:

$$k_{\mathrm{M32}}(\mathbf{u}, \mathbf{u}') = \sigma_f^2(1 + \sqrt{3}\rho)\exp(-\sqrt{3}\rho), \tag{2.61}$$

$$k_{\mathrm{M52}}(\mathbf{u}, \mathbf{u}') = \sigma_f^2\left(1 + \sqrt{5}\rho + \frac{5}{3}\rho^2\right)\exp(-\sqrt{5}\rho). \tag{2.62}$$

- **ARD Rational Quadratic:** Acts as an infinite mixture of Squared Exponential kernels with different lengthscales, suitable for modeling multi-scale variations:

$$k_{\mathrm{RQ}}(\mathbf{u}, \mathbf{u}') = \sigma_f^2\left(1 + \frac{\rho^2}{2\alpha_{\mathrm{RQ}}}\right)^{-\alpha_{\mathrm{RQ}}}. \tag{2.63}$$

The performance of the GPR model is governed by its hyperparameters: the signal variance σ_f^2, the noise variance σ_n^2, and the individual lengthscales ℓ_d. These parameters were optimized during the training phase to minimize the negative log-marginal likelihood.

2.4.2.2 Evaluation Metrics for Model Comparison

To quantitatively assess the predictive performance of the developed models and to facilitate a rigorous comparison between linear regression (plane fit) and Gaussian Process Regression, two primary statistical metrics were employed: the Root Mean Square Error (RMSE) and the Coefficient of Determination (R^2).

The Root Mean Square Error is a measure of the average magnitude of the prediction errors. It is particularly useful in chemical process modeling as it maintains the same units as the response variable (yield %), providing a clear indication of the model's precision. According to Montgomery et al. [9], RMSE penalizes larger outliers more heavily, which is desirable when high deviations in yield prediction are critical. It is defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_{\mathrm{obs}}} \sum_{r=1}^{N_{\mathrm{obs}}} (y_r - \hat{y}_r)^2}, \tag{2.64}$$

where y_r represents the experimental response, $\hat{y}_r$ is the predicted value, and N_{obs} is the number of observations.

The Coefficient of Determination, denoted as R^2, represents the proportion of the variance in the dependent variable that is predictable from the independent variables. As discussed by Fikar and Mikleš [5] in the context of model identification, R^2 provides an intuitive measure of the goodness-of-fit:

$$R^2 = 1 - \frac{SS_{\mathrm{res}}}{SS_{\mathrm{tot}}} = 1 - \frac{\sum_{r=1}^{N_{\mathrm{obs}}} (y_r - \hat{y}_r)^2}{\sum_{r=1}^{N_{\mathrm{obs}}} (y_r - \bar{y})^2}, \tag{2.65}$$

where SS_{res} is the sum of squares of residuals and SS_{tot} is the total sum of squares proportional to the variance of the data.

While R^2 indicates the global strength of the linear or non-linear relationship, RMSE provides a direct measure of the absolute fit in terms of yield percentage. The combination of these metrics allows for a comprehensive evaluation: a high R^2 coupled with a low RMSE indicates a model capable of accurately capturing the observed yield trends within the investigated operating window.

2.5 Reactor Model Development

2.5.1 Application of Kinetics to Reactor Modelling

The reactor model developed in this work applies the kinetic and random-scission concepts described in the theoretical part to a continuous stirred-tank reactor (CSTR) formulation. In the reactor model, an enhanced two-term Arrhenius formulation based on Zhao et al. [22] is used, since the parallel kinetic model was shown to describe polyethylene pyrolysis more accurately than the single-step formulation.

The main variable of the kinetic state is the fraction of broken links, denoted as x. This variable describes the extent of bond cleavage in the polyethylene chain. At $x = 0$, no polymer bonds are considered to be broken and the polymer chain remains in its initial state. At $x = 1$, complete bond scission is assumed in the model, meaning that all relevant linkages have been broken. Intermediate values $0 < x < 1$ therefore represent partial degradation of the polymer chain.

In the reactor model, x is used as an internal kinetic variable that determines the progress of thermal degradation. The overall mass conversion of polyethylene is not taken directly as x, but is calculated from x using the random-scission conversion relationship introduced below.

The temperature dependence of the reaction rate is represented by two Arrhenius-type kinetic terms [22],

$$k_1 = A_1 \exp\left(-\frac{E_1}{RT}\right), \tag{2.66}$$

and

$$k_2 = A_2 \exp\left(-\frac{E_2}{RT}\right), \tag{2.67}$$

where A_1 and A_2 are pre-exponential factors, E_1 and E_2 are activation energies, R is the universal gas constant, and T is the reactor temperature. The total kinetic rate of bond cleavage is then written as

$$r_{\mathrm{kin}} = (k_1 + k_2)(1 - x). \tag{2.68}$$

Following the random-scission interpretation introduced in Section 2.2.1, the CSTR state balance for the bond-cleavage variable x is formulated as

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{1}{\tau}(x_{\mathrm{in}} - x) + r_{\mathrm{kin}}, \tag{2.69}$$

where x_{in} is the inlet value of the bond-cleavage variable and τ is the mean residence time. The term $(x_{\mathrm{in}} - x)/\tau$ represents the effect of continuous flow through the CSTR

on the intensive state variable x, while r_{kin} represents the kinetic increase of bond cleavage due to thermal degradation. In steady-state simulations, the time derivative is equal to zero, but the same state-balance structure is retained in the model.

The residence time is calculated from the reactor volume and inlet volumetric flow rate,

$$\tau = \frac{V}{\dot{V}_{\mathrm{in}}}, \tag{2.70}$$

where the inlet volumetric flow rate is obtained from the inlet mass flow rate and inlet density,

$$\dot{V}_{\mathrm{in}} = \frac{\dot{m}_{\mathrm{in}}}{\rho_{\mathrm{in}}}. \tag{2.71}$$

The bond-cleavage variable x is subsequently transformed into the overall conversion using the chain-depolymerization/random-scission relationship originally introduced by Simha and Wall [16] and later used in the polyethylene pyrolysis modelling framework of Zhao et al. [22]. The implemented conversion relation is

$$\alpha = 1 - (1 - x)^{L-1} \left[1 + x \frac{(N_0 - L)(L - 1)}{N_0} \right]. \tag{2.72}$$

Here, N_0 is the initial number of linkages in the polymer chain and L is the minimum length of a non-volatile fragment. The conversion is used to divide the inlet polyethylene mass flow into converted and unconverted parts,

$$\dot{m}_{\mathrm{conv}} = \dot{m}_{\mathrm{in}} \alpha, \tag{2.73}$$

and

$$\dot{m}_{\mathrm{PE,unconv}} = \dot{m}_{\mathrm{in}} (1 - \alpha). \tag{2.74}$$

The carbon flow in the inlet is calculated from the inlet mass flow rate and the carbon mass fraction of polyethylene,

$$\dot{m}_{C,\mathrm{in}} = \omega_{C,\mathrm{PE}}\, \dot{m}_{\mathrm{in}}. \tag{2.75}$$

The parameter $\omega_{C,\mathrm{PE}}$ denotes the carbon mass fraction of the polyethylene feed.

2.5.2 Steady-State Model Assumptions

The reactor model is based on several simplifying assumptions. The reactor is considered an ideal continuous stirred-tank reactor with perfect mixing. Therefore, the outlet composition is assumed to be equal to the composition inside the reactor. The reactor

is operating in steady state, which means that accumulation terms are neglected in the final optimization problem.

The reactor temperature is assumed to be uniform and is treated as a decision variable. Heat losses to the surroundings are not explicitly modelled. The energy demand is represented by a simplified heat-input term $\dot{Q}_{\mathrm{in}}$, which includes sensible heating of the inlet polyethylene stream, $\dot{m}_{\mathrm{in}} c_{p,\mathrm{PE}}(T - T_{\mathrm{in}})$, the melting enthalpy contribution, $\dot{m}_{\mathrm{in}} \Delta h_{f,\mathrm{PE}}$, and an approximate reaction enthalpy contribution, $\dot{m}_{\mathrm{conv}} \Delta h_{R,\mathrm{PE}}$.

A constant heat capacity of polyethylene was used in the heating calculation. The value used in the model,

$$c_{p,\mathrm{PE}} = 2092 \ \mathrm{J\,kg^{-1}\,K^{-1}}, \tag{2.76}$$

was selected as an approximate engineering value of the same order as reported polyethylene thermal-property data in the AFWL study of plastics and related materials [1]. The melting and reaction enthalpy contributions were represented by constant approximate values taken from Netsch et al. [10],

$$\Delta h_{f,\mathrm{PE}} = 142 \times 10^3 \ \mathrm{J\,kg^{-1}}, \qquad \Delta h_{R,\mathrm{PE}} = 473 \times 10^3 \ \mathrm{J\,kg^{-1}}. \tag{2.77}$$

With the sign convention used in the model, both enthalpy terms increase the required heat input. This approximation neglects the temperature dependence of heat capacity and represents melting and reaction effects only through constant lumped enthalpy terms.

The product distribution model also uses simplified assumptions. Only selected products were explicitly represented: CH_4, C_2H_4, C_2H_6, C_3H_6, C_3H_8, C_4H_8, C_4H_{10}, C_4H_6, C_5H_{12}, H_2, and char.

The economic evaluation is limited to operating terms included in the model. Product revenue, heating cost, and a credit for recoverable energy from the polyethylene-like outlet stream and char are considered. Capital costs, separation costs, compression costs, catalyst costs, labour costs, and product purification costs are not included. Therefore, the calculated profit should be interpreted as a simplified process-performance indicator rather than a complete techno-economic analysis.

2.5.3 CSTR Model Formulation in gPROMS

The reactor model was formulated in gPROMS ModelBuilder, a process-modelling environment used for the mathematical formulation, simulation, and optimization of process systems [12]. In this work, gPROMS was used to implement the steady-state

CSTR model, solve the model equations, and perform process optimization and global system analysis.

The random-scission relationship is used in the reactor model to describe how the converted part of the polyethylene feed is distributed into shorter chain fragments. Following Zhao et al. [22], the carbon-yield contribution of a chain fragment containing g monomer units is expressed as

$$\phi_g = g x^2 (1-x)^{g-1}, \qquad g = 1, 2, \ldots, 10. \tag{2.78}$$

The variable ϕ_g represents the relative carbon-yield contribution of fragments of size g. The index g denotes the number of monomer units in the chain fragment, and x is the fraction of broken linkages introduced in Section 2.2.1. Only the first ten fragment sizes are represented explicitly in the model, because the reactor model focuses on the light product range used later in the yield and economic calculations.

The explicitly represented fragment contributions are summed as

$$\phi_{\mathrm{sum}} = \sum_{g=1}^{10} \phi_g. \tag{2.79}$$

The remaining part of the distribution is calculated as

$$\phi_{\mathrm{rest}} = \max(0, 1 - \phi_{\mathrm{sum}}). \tag{2.80}$$

The term ϕ_{rest} represents the part of the converted material that is not assigned to the explicitly represented fragment range.

The random-scission distribution is applied only to the converted part of the feed. The converted carbon flow is calculated as

$$\dot{m}_{C,\mathrm{conv}} = \dot{m}_{C,\mathrm{in}} \alpha. \tag{2.81}$$

The variable α denotes the overall conversion obtained from the random-scission conversion relationship.

The carbon flow assigned to fragment size g is then calculated as

$$\dot{m}_{C,g} = \phi_g \dot{m}_{C,\mathrm{conv}}. \tag{2.82}$$

This formulation ensures that the product distribution is generated only from the converted carbon flow, while the unconverted part of the polyethylene feed remains outside the product-distribution calculation.

In the steady-state model, reactor temperature and inlet mass flow rate are treated as time-invariant variables. These variables affect the kinetic rate, residence time, conversion, product distribution, and therefore also the economic objective used later in the optimization.

2.5.4 Representation of Product Yields in the Model

The product-yield representation builds on the random-scission formulation introduced in the previous section. The random-scission model provides carbon-based fragment contributions, whereas the empirical yield correlations obtained from the processed and reconciled dataset are used to distribute these contributions among selected chemical species. The experimental basis for the product-yield modelling comes from the polyethylene pyrolysis data of Zhao et al. [22].

The empirical yields were implemented as linear functions of temperature and residence time. The temperature was converted from Kelvin to degrees Celsius,

$$T_C = T - 273.15. \tag{2.83}$$

For each product s, the empirical mass yield was written in the general form

$$Y_s^{\mathrm{emp}} = \theta_{0,s} + \theta_{T,s} T_C + \theta_{\tau,s} \tau. \tag{2.84}$$

To avoid negative yields outside the local validity of the linear correlations, all empirical yields were clipped to non-negative values:

$$Y_s^{\mathrm{emp}} = \max(0, \theta_{0,s} + \theta_{T,s} T_C + \theta_{\tau,s} \tau). \tag{2.85}$$

Here, Y_s^{emp} denotes the empirical mass yield of product s, $\theta_{0,s}$, $\theta_{T,s}$, and $\theta_{\tau,s}$ are the fitted coefficients, T_C represents the reactor temperature in degrees Celsius, and τ denotes the residence time.

Since these empirical yields describe the total mass of the corresponding compounds, they were first converted to a carbon basis. For a carbon-containing product s, the carbon-based empirical yield was calculated as

$$\widetilde{Y}_{C,s} = Y_s^{\mathrm{emp}} \frac{M_{C,s}}{M_s}, \tag{2.86}$$

where M_s is the molar mass of product s, and $M_{C,s}$ represents the mass of carbon contained in one mole of that product. For example,

$$\widetilde{Y}_{C,\mathrm{C_2H_4}} = Y_{\mathrm{C_2H_4}}^{\mathrm{emp}} \frac{24}{28}, \tag{2.87}$$

and

$$\widetilde{Y}_{C,\mathrm{CH_4}} = Y^{\mathrm{emp}}_{\mathrm{CH_4}} \frac{12}{16}. \tag{2.88}$$

The quantities $\widetilde{Y}_{C,s}$ therefore serve as carbon-based weighting factors derived from the empirical yield correlations.

In the random-scission formulation, the index g denotes the number of polyethylene monomer units in a fragment. Since one polyethylene monomer unit contains two carbon atoms, the contribution ϕ_1 corresponds to the C_2 group, while ϕ_2 corresponds to the C_4 group. These group-level contributions were then distributed among the selected compounds using the carbon-based empirical yields.

The C_2 group was divided between ethylene and ethane. The sum of the carbon-based empirical yields in this group was defined as

$$\Sigma^{\mathrm{emp}}_{C2} = \widetilde{Y}_{C,\mathrm{C_2H_4}} + \widetilde{Y}_{C,\mathrm{C_2H_6}}. \tag{2.89}$$

The carbon-based yield assigned to ethylene was calculated as

$$\psi_{C,\mathrm{C_2H_4}} = \phi_1 \frac{\widetilde{Y}_{C,\mathrm{C_2H_4}}}{\Sigma^{\mathrm{emp}}_{C2}}, \tag{2.90}$$

and the corresponding contribution of ethane was obtained from

$$\psi_{C,\mathrm{C_2H_6}} = \phi_1 \frac{\widetilde{Y}_{C,\mathrm{C_2H_6}}}{\Sigma^{\mathrm{emp}}_{C2}}. \tag{2.91}$$

The variable $\psi_{C,s}$ represents the final carbon-based yield assigned to product s after combining the random-scission group contribution with the empirical yield ratio.

Analogously, the C_4 group was divided between butene, butane, and 1,3-butadiene. The carbon-based empirical yield sum for this group was written as

$$\Sigma^{\mathrm{emp}}_{C4} = \widetilde{Y}_{C,\mathrm{C_4H_8}} + \widetilde{Y}_{C,\mathrm{C_4H_{10}}} + \widetilde{Y}_{C,\mathrm{C_4H_6}}. \tag{2.92}$$

The individual carbon-based yields were then evaluated as

$$\psi_{C,\mathrm{C_4H_8}} = \phi_2 \frac{\widetilde{Y}_{C,\mathrm{C_4H_8}}}{\Sigma^{\mathrm{emp}}_{C4}}, \tag{2.93}$$

$$\psi_{C,\mathrm{C_4H_{10}}} = \phi_2 \frac{\widetilde{Y}_{C,\mathrm{C_4H_{10}}}}{\Sigma^{\mathrm{emp}}_{C4}}, \tag{2.94}$$

and

$$\psi_{C,\mathrm{C_4H_6}} = \phi_2 \frac{\widetilde{Y}_{C,\mathrm{C_4H_6}}}{\Sigma^{\mathrm{emp}}_{C4}}. \tag{2.95}$$

The remaining selected hydrocarbons were calculated by empirical ratios relative to chemically related products that had already been obtained from the C_2 or C_4 group allocation. Methane was calculated using ethane as the reference compound:

$$\psi_{C,\mathrm{CH_4}} = \psi_{C,\mathrm{C_2H_6}} \frac{\widetilde{Y}_{C,\mathrm{CH_4}}}{\widetilde{Y}_{C,\mathrm{C_2H_6}}}. \tag{2.96}$$

Propene was calculated using butene as the reference compound,

$$\psi_{C,\mathrm{C_3H_6}} = \psi_{C,\mathrm{C_4H_8}} \frac{\widetilde{Y}_{C,\mathrm{C_3H_6}}}{\widetilde{Y}_{C,\mathrm{C_4H_8}}}, \tag{2.97}$$

and propane was calculated using butane as the reference compound,

$$\psi_{C,\mathrm{C_3H_8}} = \psi_{C,\mathrm{C_4H_{10}}} \frac{\widetilde{Y}_{C,\mathrm{C_3H_8}}}{\widetilde{Y}_{C,\mathrm{C_4H_{10}}}}. \tag{2.98}$$

This step preserves the relative empirical trends between selected compounds while keeping the main carbon allocation tied to the random-scission distribution.

After the final carbon-based product yields had been obtained, the converted carbon flow defined in the previous section was used to calculate the amount of carbon assigned to each product:

$$\dot{m}_{C,s} = \psi_{C,s} \dot{m}_{C,\mathrm{conv}}. \tag{2.99}$$

Here, $\dot{m}_{C,s}$ denotes the carbon mass flow contained in product s, and $\dot{m}_{C,\mathrm{conv}}$ is the converted carbon mass flow.

The carbon mass flows were subsequently converted back to total product mass flows using the corresponding molar-mass ratios:

$$\dot{m}_s = \dot{m}_{C,s} \frac{M_s}{M_{C,s}}. \tag{2.100}$$

For example,

$$\dot{m}_{\mathrm{CH_4}} = \dot{m}_{C,\mathrm{CH_4}} \frac{16}{12}, \tag{2.101}$$

and

$$\dot{m}_{\mathrm{C_2H_4}} = \dot{m}_{C,\mathrm{C_2H_4}} \frac{28}{24}. \tag{2.102}$$

Thus, the product calculation first assigns carbon to individual compounds and then converts the carbon flow into the corresponding total compound mass flow.

Hydrogen was treated separately because it does not contain carbon and therefore cannot be obtained from a carbon-based random-scission allocation. Its mass flow

was calculated using ethane as the reference compound. Since the ethane mass flow had already been obtained from the carbon-based calculation, hydrogen was evaluated from the empirical mass-yield ratio:

$$\dot{m}_{\mathrm{H_2}} = \dot{m}_{\mathrm{C_2H_6}} \frac{Y^{\mathrm{emp}}_{\mathrm{H_2}}}{Y^{\mathrm{emp}}_{\mathrm{C_2H_6}}}. \tag{2.103}$$

In this expression, $Y^{\mathrm{emp}}_{\mathrm{H_2}}$ and $Y^{\mathrm{emp}}_{\mathrm{C_2H_6}}$ represent the empirical mass yields of hydrogen and ethane, respectively.

Char was calculated by a ratio-based empirical relation using ethane as the reference compound:

$$\dot{m}_{\mathrm{char}} = \dot{m}_{\mathrm{C_2H_6}} \frac{Y^{\mathrm{emp}}_{\mathrm{char}}}{Y^{\mathrm{emp}}_{\mathrm{C_2H_6}}}. \tag{2.104}$$

In this expression, $Y^{\mathrm{emp}}_{\mathrm{char}}$ and $Y^{\mathrm{emp}}_{\mathrm{C_2H_6}}$ represent the empirical mass yields of char and ethane, respectively.

The set of explicitly calculated hydrocarbon products is defined as

$$\mathcal{H} = \{\mathrm{CH_4}, \mathrm{C_2H_4}, \mathrm{C_2H_6}, \mathrm{C_3H_6}, \mathrm{C_3H_8}, \mathrm{C_4H_8}, \mathrm{C_4H_{10}}, \mathrm{C_4H_6}, \mathrm{C_5H_{12}}\}. \tag{2.105}$$

The total explicitly calculated product stream was then obtained by summing the mass flows of all calculated products:

$$\dot{m}_{\mathrm{prod,exp}} = \sum_{s \in \mathcal{H}} \dot{m}_s + \dot{m}_{\mathrm{H_2}} + \dot{m}_{\mathrm{char}}. \tag{2.106}$$

Here, $\dot{m}_{\mathrm{prod,exp}}$ denotes the total mass flow of explicitly calculated products, and $\mathcal{H}$ represents the set of explicitly calculated hydrocarbon products.

The total converted mass flow was calculated from the overall conversion as

$$\dot{m}_{\mathrm{conv}} = \dot{m}_{\mathrm{in}} \alpha. \tag{2.107}$$

The genuinely unreacted part of the polyethylene feed was therefore expressed as

$$\dot{m}_{\mathrm{PE,unconv}} = \dot{m}_{\mathrm{in}} - \dot{m}_{\mathrm{conv}}. \tag{2.108}$$

The variable $\dot{m}_{\mathrm{PE,unconv}}$ represents the unreacted polyethylene stream.

The difference between the converted mass flow and the explicitly calculated product stream represents converted material that was not assigned to the selected products:

$$\dot{m}_{\mathrm{rem}} = \dot{m}_{\mathrm{conv}} - \dot{m}_{\mathrm{prod,exp}}. \tag{2.109}$$

The variable $\dot{m}_{\mathrm{rem}}$ denotes the unassigned converted material. This material is attributed mainly to higher hydrocarbons that were not described individually in the simplified product representation.

For the following model calculations, this unassigned converted material was treated using polyethylene properties. The stream carried forward with polyethylene properties was therefore written as

$$\dot{m}^{*}_{\mathrm{PE}} = \dot{m}_{\mathrm{PE,unconv}} + \dot{m}_{\mathrm{rem}}. \tag{2.110}$$

The superscript $(*)$ indicates that this stream is not pure unreacted polyethylene, but a modelling approximation consisting of unreacted polyethylene and the unassigned converted material.

2.5.5 Economic Evaluation and Process Optimization

The economic part of the model was added to evaluate how operating conditions influence the simplified profitability of the reactor. The objective function combines three main contributions: product value, heating cost, and a credit for recoverable energy from the polyethylene-like outlet stream and char.

The economic model was evaluated at steady state. Therefore, the main economic quantity used in optimization was the profit rate, expressed in $\mathrm{EUR\,s^{-1}}$. The total profit over the selected operating time t_{op} was also calculated as

$$\Pi_{\mathrm{EUR}} = \dot{\Pi}_{\mathrm{EUR}} t_{\mathrm{op}}. \tag{2.111}$$

2.5.5.1 Wholesale Price Data for Products and Energy

The product price data were collected as approximate wholesale price indicators from the Echemi database [4]. Where available, values corresponding to Northwest Europe were preferred; otherwise, available China-market price data were used. Since these prices are market-dependent and time-dependent, they are used only as approximate values for the simplified economic objective. The originally collected product prices were reported in USD/kg and were converted to EUR/kg using a fixed conversion factor of 1 USD = 0.85 EUR.

The price ranges used in the model are summarized in Table 2.1. For compounds where a minimum and maximum value were available, the average value was used in the economic calculation.

Table 2.1: Product price values used in the simplified economic model.

Product	Price used in model
CH_4	0.85–1.02 EUR/kg
C_2H_4	3.40–5.10 EUR/kg
C_2H_6	0.85–0.98 EUR/kg
C_3H_6	2.55–4.25 EUR/kg
C_3H_8	2.55–3.40 EUR/kg
C_4H_8	1.24 EUR/kg
C_4H_{10}	2.98–3.40 EUR/kg
C_4H_6	1.33 EUR/kg
H_2	0.85–0.89 EUR/kg

The electricity and natural gas prices were taken from the official SPP price documents for firms and organizations [18, 17]. The values implemented in the model were

$$c_{\mathrm{el}} = 0.10896 \text{ EUR/kWh}, \tag{2.112}$$

and

$$c_{\mathrm{ng}} = 0.04921 \text{ EUR/kWh}. \tag{2.113}$$

Representing EUR/J:

$$c_{\mathrm{el,J}} = 3.0267 \times 10^{-8} \text{ EUR/J}, \tag{2.114}$$

and

$$c_{\mathrm{ng,J}} = 1.3669 \times 10^{-8} \text{ EUR/J}. \tag{2.115}$$

The model uses the average of the electricity and natural gas energy costs as a simplified heating-cost coefficient,

$$c_{\mathrm{heat,J}} = \frac{1}{2}\left(c_{\mathrm{el,J}} + c_{\mathrm{ng,J}}\right). \tag{2.116}$$

2.5.5.2 Profit Function Formulation

The simplified profit rate is defined as

$$\dot{\Pi}_{\mathrm{EUR}} = \dot{\mathcal{R}}_{\mathrm{EUR}} - \dot{C}_{\mathrm{heat,EUR}} + \dot{C}_{\mathrm{waste,EUR}}, \tag{2.117}$$

where $\dot{\mathcal{R}}_{\mathrm{EUR}}$ is the product revenue rate, $\dot{C}_{\mathrm{heat,EUR}}$ is the heating cost rate, and $\dot{C}_{\mathrm{waste,EUR}}$ is the waste-heat credit rate. The profit function is intentionally simplified.

It is used to compare operating points and identify economically favourable regions of the operating space. It is not intended to represent a complete industrial economic analysis.

Heating Cost Term

The heat required to raise the inlet polyethylene stream from the inlet temperature to the reactor temperature is calculated as

$$\dot{Q}_{\mathrm{in}} = \dot{m}_{\mathrm{in}} \left[c_{p,\mathrm{PE}} (T - T_{\mathrm{in}}) + \Delta h_{f,\mathrm{PE}} \right] + \dot{m}_{\mathrm{conv}} \Delta h_{R,\mathrm{PE}}. \tag{2.118}$$

The parameter $c_{p,\mathrm{PE}}$ [1] represents the heat capacity of polyethylene, $\Delta h_{f,\mathrm{PE}}$ [10] denotes the enthalpy of melting, and $\Delta h_{R,\mathrm{PE}}$ [10] represents the reaction enthalpy of polyethylene pyrolysis. The converted polyethylene mass flow is denoted by $\dot{m}_{\mathrm{conv}}$. Since $c_{p,\mathrm{PE}}(T - T_{\mathrm{in}})$, $\Delta h_{f,\mathrm{PE}}$, and $\Delta h_{R,\mathrm{PE}}$ are expressed on a mass basis, the resulting heat input rate is obtained in $\mathrm{J\,s^{-1}}$.

The corresponding heating cost rate is

$$\dot{C}_{\mathrm{heat,EUR}} = c_{\mathrm{heat,J}} \dot{Q}_{\mathrm{in}}. \tag{2.119}$$

As discussed above, this term represents a simplified heat-input estimate including sensible heating, melting enthalpy, and an approximate reaction enthalpy contribution. The formulation should therefore be interpreted as an engineering approximation rather than a detailed heat-balance model.

Waste Heat Term

The model includes a simplified credit for recoverable energy from the polyethylene-like outlet stream and char. The polyethylene-like stream, denoted as $\dot{m}^{*}_{\mathrm{PE}}$, consists of unreacted polyethylene and the unassigned converted material introduced in the product-yield formulation.

The recoverable heat rate from this polyethylene-like stream is calculated as

$$\dot{Q}_{\mathrm{waste,PE}} = \eta_{\mathrm{waste}} \dot{m}^{*}_{\mathrm{PE}} \Delta h_{\mathrm{comb,PE}}, \tag{2.120}$$

where η_{waste} is the assumed energy recovery efficiency and $\Delta h_{\mathrm{comb,PE}}$ denotes the positive heating-value-type energy content assigned to the polyethylene-like material.

The recoverable heat rate from char is calculated as

$$\dot{Q}_{\mathrm{waste,char}} = \eta_{\mathrm{char}} \dot{m}_{\mathrm{char}} \Delta h_{\mathrm{comb,char}}, \tag{2.121}$$

where η_{char} is the assumed char energy recovery efficiency and $\Delta h_{\mathrm{comb,char}}$ denotes the positive heating-value-type energy content of char.

In this formulation, $\Delta h_{\mathrm{comb,PE}}$ and $\Delta h_{\mathrm{comb,char}}$ are expressed in $\mathrm{J\,kg^{-1}}$. They are used as recoverable-energy coefficients in the economic objective, not as thermodynamic reaction enthalpies with a negative combustion sign convention.

The total recoverable heat rate is therefore

$$\dot{Q}_{\mathrm{waste,total}} = \dot{Q}_{\mathrm{waste,PE}} + \dot{Q}_{\mathrm{waste,char}}. \tag{2.122}$$

The corresponding energy credit is calculated as

$$\dot{C}_{\mathrm{waste,EUR}} = c_{\mathrm{heat,J}} \dot{Q}_{\mathrm{waste,total}}. \tag{2.123}$$

This term should be interpreted as a simplified utility credit. In a real process, the practical recovery of this energy would depend on combustion efficiency, heat integration, equipment limitations, product composition, and environmental constraints.

Product Value Term

The product revenue is calculated from the mass flow rates of selected valuable products and their assumed wholesale prices. In EUR, the revenue rate is

$$\begin{aligned}\dot{\mathcal{R}}_{\mathrm{EUR}} = {} & \dot{m}_{\mathrm{CH_4}} c^{\mathrm{prod}}_{\mathrm{CH_4}} + \dot{m}_{\mathrm{C_2H_4}} c^{\mathrm{prod}}_{\mathrm{C_2H_4}} + \dot{m}_{\mathrm{C_2H_6}} c^{\mathrm{prod}}_{\mathrm{C_2H_6}} + \dot{m}_{\mathrm{C_3H_6}} c^{\mathrm{prod}}_{\mathrm{C_3H_6}} \\ & + \dot{m}_{\mathrm{C_3H_8}} c^{\mathrm{prod}}_{\mathrm{C_3H_8}} + \dot{m}_{\mathrm{C_4H_8}} c^{\mathrm{prod}}_{\mathrm{C_4H_8}} + \dot{m}_{\mathrm{C_4H_{10}}} c^{\mathrm{prod}}_{\mathrm{C_4H_{10}}} + \dot{m}_{\mathrm{C_4H_6}} c^{\mathrm{prod}}_{\mathrm{C_4H_6}} \\ & + \dot{m}_{\mathrm{H_2}} c^{\mathrm{prod}}_{\mathrm{H_2}}.\end{aligned} \tag{2.124}$$

Hydrogen was formally included in the product-value term through its market price. However, because hydrogen is not allocated through the carbon-based random-scission distribution, its mass flow is determined separately from the remaining converted mass after hydrocarbon allocation.

Pentane, char, carbon monoxide, and the polyethylene-like outlet stream were not included as direct product-revenue terms in the implemented objective function. The

polyethylene-like outlet stream and char influence the objective only through the waste-energy credit.

2.5.5.3 Decision Variables

The steady-state optimization used two time-invariant continuous decision variables: the reactor temperature T and the inlet mass flow rate $\dot{m}_{\mathrm{in}}$. The residence time was not optimized directly, but was calculated from the inlet mass flow rate, reactor volume, and inlet density.

Temperature

The reactor temperature T was selected as a decision variable because it directly affects the Arrhenius kinetic terms, the bond-cleavage variable x, the conversion, and the empirical yield correlations. In the optimization setup, the temperature was bounded as

$$673.15\ \mathrm{K} \leq T \leq 873.15\ \mathrm{K}. \tag{2.125}$$

Residence Time

The inlet mass flow rate was treated as a decision variable with bounds

$$0.05\ \mathrm{kg\,s^{-1}} \leq \dot{m}_{\mathrm{in}} \leq 0.20\ \mathrm{kg\,s^{-1}}. \tag{2.126}$$

The residence time was calculated as

$$\tau = \frac{V \rho_{\mathrm{in}}}{\dot{m}_{\mathrm{in}}}. \tag{2.127}$$

In addition to the bounds on $\dot{m}_{\mathrm{in}}$, the residence time was constrained directly by endpoint inequality constraints:

$$10\ \mathrm{s} \leq \tau \leq 20\ \mathrm{s}. \tag{2.128}$$

2.5.5.4 Steady-State Optimization Problem

The steady-state optimization problem can be summarized as

$$\max_{T,\dot{m}_{\mathrm{in}}} \dot{\Pi}_{\mathrm{EUR}}, \tag{2.129}$$

subject to the CSTR model equations, product-yield equations, economic equations, decision-variable bounds, residence-time bounds, and non-negativity constraints on the relevant product mass flow rates.

The main constraints imposed in the optimization were

$$673.15\ \mathrm{K} \leq T \leq 873.15\ \mathrm{K}, \tag{2.130}$$

$$0.05\ \mathrm{kg\,s^{-1}} \leq \dot{m}_{\mathrm{in}} \leq 0.20\ \mathrm{kg\,s^{-1}}, \tag{2.131}$$

$$10\ \mathrm{s} \leq \tau \leq 20\ \mathrm{s}, \tag{2.132}$$

and

$$\dot{m}_s \geq 0 \tag{2.133}$$

for all selected product mass flows.

The optimization was solved in gPROMS as a steady-state optimization problem. The numerical results of the optimization are presented and discussed in Chapter 4.

2.5.6 Global System Analysis

A global system analysis was performed to evaluate the behaviour of the reactor model over a wider range of operating conditions. The analysis varied the inlet mass flow rate and reactor temperature as sampled factors. Both variables were treated as time-invariant continuous quantities. The main responses included product mass flow rates, residence time, the polyethylene-like remaining mass flow rate, and the profit function.

The sampled factors were the reactor temperature T and the inlet mass flow rate $\dot{m}_{\mathrm{in}}$. The residence time was calculated from the sampled inlet mass flow rate according to

$$\tau = \frac{V\rho_{\mathrm{in}}}{\dot{m}_{\mathrm{in}}}. \tag{2.134}$$

The GSA results were exported and post-processed in MATLAB. Since the model output was sampled in terms of T and $\dot{m}_{\mathrm{in}}$, the calculated residence time τ was used to construct two-dimensional response surfaces in the T-τ plane.

Chapter 3

Data and Methods

This chapter describes the datasets used in this work and the practical implementation steps for (i) implementing and validating the conversion-level model for PE thermal degradation, (ii) formulating and solving the yield data reconciliation problem using weighted least squares under balance constraints, and (iii) modeling the reconciled yields as functions of temperature and residence time using a baseline linear regression model and Gaussian Process Regression (GPR), including a comparison across kernel choices.

3.1 Experimental Data Processing in MATLAB

Two types of information are used: (i) conversion-related data used for verification of the primary-pyrolysis kinetic model, and (ii) a published yield dataset containing phase fractions and gas-phase composition, used for reconciliation and subsequent yield modelling. The used dataset was published by Zhao et al. [22].

For the yield dataset, the reported quantities are expressed using different definitions and units:

- **Phase fractions** (gas/liquid/char) are reported as **mass fractions** (w%).
- **Gas composition** (individual gas species) is reported as **carbon yield** (C-yield%), i.e., the distribution of inlet carbon among individual gaseous products.
- **Hydrogen** is reported separately as **vol%** in the gas phase.

The dataset contains $n_{\text{exp}} = 6$ operating points. The operating window covers $T_C = 500$–600 °C and residence times τ of approximately 12–21 s. The available measured gas

species are: CH_4, C_2H_4, C_2H_6, C_3H_6, C_3H_8, C_4H_8, C_4H_{10}, C_4H_6 (1,3-butadiene), C_5H_{12}, CO, and H_2.

In addition to the yields, the source provides absolute error bars (reported as $\pm$ values) for most gas species. These are used to construct weighting coefficients for reconciliation (Section 3.2.2). For pentane, no uncertainty is available and an educated guess is used to avoid a singular weighting definition.

3.2 Data Reconciliation Procedure

For each experiment, reconciliation is formulated as a constrained weighted least-squares problem on the mass basis established in Section 3.3.1. The decision vector $\mathbf{z}^{\mathrm{rec}}$ includes:

- inlet mass m_{in} and inlet elemental masses $(m_{C,\mathrm{in}}, m_{H,\mathrm{in}})$,
- phase masses $(m_{\mathrm{gas}}, m_{\mathrm{liq}}, m_{\mathrm{char}})$ and corresponding elemental partitions,
- individual gas-species masses m_s (including H_2) and their elemental contributions.

The objective penalises deviations from measured values:

$$\min_{\mathbf{z}^{\mathrm{rec}}} (\mathbf{z}^{\mathrm{rec}} - \mathbf{z}^{\mathrm{meas}})^{\top} \mathbf{\Lambda} (\mathbf{z}^{\mathrm{rec}} - \mathbf{z}^{\mathrm{meas}}), \tag{3.1}$$

where $\mathbf{z}^{\mathrm{meas}}$ is obtained by the conversions in Section 3.3.1. The weight matrix $\mathbf{\Lambda}$ is diagonal.

In the implementation, the diagonal weights are constructed as $\Lambda_{jj} = 1/\sigma_j^2$, where σ_j is an error (uncertainty) proxy assigned to each variable.

3.2.1 Balance Constraints

Reconciliation is constrained by overall and elemental balances:

$$m_{\mathrm{in}} = m_{\mathrm{gas}} + m_{\mathrm{liq}} + m_{\mathrm{char}}, \tag{3.2}$$

$$m_{H,\mathrm{in}} = m_{H,\mathrm{gas}} + m_{H,\mathrm{liq}} + m_{H,\mathrm{char}}, \tag{3.3}$$

$$m_{C,\mathrm{in}} = m_{C,\mathrm{gas}} + m_{C,\mathrm{liq}} + m_{C,\mathrm{char}}, \tag{3.4}$$

$$m_{O,\mathrm{in}} = m_{O,\mathrm{gas}} \quad (= 0). \tag{3.5}$$

Phase decomposition constraints enforce consistency between total and elemental phase masses:

$$m_{\text{char}} = m_{C,\text{char}} + m_{H,\text{char}}, \qquad m_{\text{liq}} = m_{C,\text{liq}} + m_{H,\text{liq}}. \tag{3.6}$$

Let

$$\mathcal{S} = \{\text{CH}_4, \text{C}_2\text{H}_4, \text{C}_2\text{H}_6, \text{C}_3\text{H}_6, \text{C}_3\text{H}_8, \text{C}_4\text{H}_8, \text{C}_4\text{H}_{10}, \text{C}_4\text{H}_6, \text{C}_5\text{H}_{12}, \text{CO}, \text{H}_2\} \tag{3.7}$$

denote the set of gaseous products. For the gas phase, closure is enforced on (i) total gas mass and (ii) elemental gas masses:

$$m_{\text{gas}} = \sum_{s\in\mathcal{S}} m_s, \tag{3.8}$$

$$m_{H,\text{gas}} = \sum_{s\in\mathcal{S}} m_{H,s}, \tag{3.9}$$

$$m_{C,\text{gas}} = \sum_{s\in\mathcal{S}} m_{C,s}, \tag{3.10}$$

$$m_{O,\text{gas}} = m_{O,\text{CO}}. \tag{3.11}$$

Chemical consistency is imposed by linking species masses to their elemental partitions:

$$m_{C,s} = \omega_{C,s}\, m_s, \qquad m_{H,s} = \omega_{H,s}\, m_s, \tag{3.12}$$

and for CO:

$$m_{O,\text{CO}} = \omega_{O,\text{CO}}\, m_{\text{CO}}. \tag{3.13}$$

Finally, all decision variables are constrained to be non-negative:

$$\mathbf{z}^{\text{rec}} \geq \mathbf{0}. \tag{3.14}$$

3.2.2 Weight Construction

The yield dataset does not provide a full statistical characterization (replicate counts and pointwise standard deviations for all quantities). Therefore, weighting is constructed from the reported absolute errors (published as ± intervals) using a pragmatic aggregation approach.

For each gas species s, a vector of absolute errors across all operating points is available, with the exception of pentane, e.g. $\Delta Y_{s,1}, \ldots, \Delta Y_{s,n_{\text{exp}}}$. The mean absolute error

$$\sigma_s \equiv \frac{1}{n_{\text{exp}}} \sum_{r=1}^{n_{\text{exp}}} \left|\Delta Y_{s,r}\right|, \tag{3.15}$$

is treated as the uncertainty proxy for that species. For decision variables associated with species s, the corresponding σ_j value is assigned from σ_s. For variables where no uncertainty is reported (e.g., pentane), a small positive number is assigned to avoid a zero uncertainty, which would lead to an undefined weight. In addition, the inlet mass and the phase totals are assigned small uncertainties (large weights), which keeps them close to the reported values and forces the reconciliation to primarily adjust the gas-species yields while satisfying the global balance constraints.

3.2.3 Solver Setup and Implementation Details

The optimization problem is implemented in MATLAB using YALMIP. Constraints are collected into a single constraint set, and each experiment is solved independently. Because the objective is quadratic and all constraints are linear equalities and inequalities, the resulting problem is solved using quadratic programming.

The reconciliation produces, for each experiment: (i) reconciled phase masses (gas/liquid/char), (ii) reconciled masses of individual gas species (including H_2), and (iii) reconciled elemental partitions consistent with the imposed balance constraints.

3.2.4 Consistency of Reconciled Data

The consistency of the reconciled dataset was evaluated by checking the residuals of the total mass balance and elemental balances after optimization. For each operating point, the total mass, carbon, hydrogen, and oxygen balance residuals were inspected together with the non-negativity of all reconciled variables. The reconciled dataset was considered physically consistent only if all imposed constraints were satisfied within numerical tolerance.

3.3 Calculation of Mass Yields

The independent variables for yield modeling are:

- temperature T_C (in °C),
- gas residence time τ (in s).

The dependent variables are product yields expressed on a single consistent basis. Because the raw dataset is reported on mixed bases (wt%, C-yield%, and vol%), all

reported quantities are first converted to **absolute masses** by normalising the polymer inlet to a unit mass:

$$m_{\text{in}} = 1, \tag{3.16}$$

which can be interpreted as 1 g (or any other convenient unit). This scaling is without loss of generality: the reconciliation formulation and the relative corrections are invariant to the selected mass unit.

3.3.1 Preprocessing and Basis Harmonisation

The reconciliation is performed on a mass basis using the following conversion steps.

Polymer elemental composition. The polymer inlet is treated as polyethylene with repeating unit CH_2. Using $M_C = 12.011$ and $M_H = 1.00784$, the mass fraction of carbon in polyethylene is:

$$\omega_{C,\text{PE}} = \frac{M_C}{M_C + 2M_H} = \frac{12.011}{12.011 + 2.01568}. \tag{3.17}$$

Thus, for $m_{\text{in}} = 1$, the inlet elemental masses are:

$$m_{C,\text{in}} = \omega_{C,\text{PE}}\, m_{\text{in}}, \qquad m_{H,\text{in}} = (1 - \omega_{C,\text{PE}})\, m_{\text{in}}, \qquad m_{O,\text{in}} = 0. \tag{3.18}$$

Conversion of phase yields (w%) to masses. If the gas, liquid, and char phase fractions are reported as mass percentages $Y_p^{\text{w\%}}$ of the inlet, their measured masses are obtained as:

$$m_p = \frac{Y_p^{\text{w\%}}}{100}\, m_{\text{in}}, \qquad p \in \mathcal{P}. \tag{3.19}$$

Assumed elemental composition of liquid and char. For the liquid phase, the elemental composition is assumed to follow the inlet polymer:

$$m_{C,\text{liq}} = \omega_{C,\text{PE}}\, m_{\text{liq}}, \qquad m_{H,\text{liq}} = (1 - \omega_{C,\text{PE}})\, m_{\text{liq}}. \tag{3.20}$$

For the char phase, a fixed elemental split is assumed:

$$m_{C,\text{char}} = 0.95\, m_{\text{char}}, \qquad m_{H,\text{char}} = 0.05\, m_{\text{char}}. \tag{3.21}$$

Gas species (C-yield%) to species masses. For each gaseous product s reported as carbon yield ($Y_s^{\mathrm{C\%}}$), the carbon mass assigned to s is:

$$m_{C,s} = \frac{Y_s^{\mathrm{C\%}}}{100}\, m_{C,\mathrm{in}}. \tag{3.22}$$

The total mass of species s is then reconstructed using its carbon mass fraction $\omega_{C,s}$:

$$m_s = \frac{m_{C,s}}{\omega_{C,s}}, \tag{3.23}$$

where $\omega_{C,s}$ is computed from the molecular formula (e.g., $\omega_{C,\mathrm{CH_4}} = 12.011/(12.011 + 4.03136)$, $\omega_{C,\mathrm{C_2H_4}} = 24.022/(24.022+4.03136)$, and $\omega_{C,\mathrm{CO}} = 12.011/(12.011+15.999)$). Hydrogen and oxygen masses within each gaseous product are obtained similarly via $\omega_{H,s} = 1 - \omega_{C,s}$ for hydrocarbons and $\omega_{O,\mathrm{CO}} = 1 - \omega_{C,\mathrm{CO}}$ for carbon monoxide.

Conversion of hydrogen (vol%) to mass. Hydrogen is reported as vol% in the gas phase and is treated as a mole fraction (ideal-gas approximation). Let $\chi_{\mathrm{H_2}}$ denote the reported hydrogen mole fraction. The total amount of non-hydrogen gaseous products is first converted to moles and summed to obtain $n_{\mathrm{nonH_2}}$. The hydrogen amount is then reconstructed by:

$$n_{\mathrm{H_2}} = \frac{\chi_{\mathrm{H_2}}}{1 - \chi_{\mathrm{H_2}}}\, n_{\mathrm{nonH_2}}, \qquad m_{\mathrm{H_2}} = n_{\mathrm{H_2}}\, M_{\mathrm{H_2}}. \tag{3.24}$$

3.4 Yield Modeling

3.4.1 Temperature and Residence Time

The reconciled dataset is used to model yields as functions of the input vector $\mathbf{u} = [T_C, \tau]^\top$. Models are fitted for both the **measured** values and the **reconciled** values in order to quantify how reconciliation affects the identified trends and smooth response surfaces.

3.4.2 Baseline Linear Model

As a baseline, each output variable Y is fitted by a linear function in (T_C, τ):

$$\hat{Y}(T_C, \tau) = \theta_0 + \theta_T T_C + \theta_\tau \tau. \tag{3.25}$$

The coefficients are estimated by ordinary least squares (OLS) using the matrix formulation:

$$\boldsymbol{\theta}_Y = \begin{bmatrix} \theta_0 \\ \theta_T \\ \theta_\tau \end{bmatrix} = (\mathbf{U}^\top \mathbf{U})^{-1} \mathbf{U}^\top \mathbf{Y}, \qquad \mathbf{U} = [\mathbf{1},\, T_C,\, \tau]. \tag{3.26}$$

Model quality is quantified by the coefficient of determination R^2 computed on the available operating points. The plane fit is applied uniformly to all variables of interest, enabling a simple interpretable comparison between measured and reconciled datasets.

3.4.3 Gaussian Process Regression and Kernel Comparison

To capture nonlinear dependencies under a small number of datapoints, Gaussian Process Regression models are trained for each yield variable using the two-dimensional input space (T_C, τ). Several ARD kernel families available in MATLAB are evaluated as introduced in Chapter 2.

In all cases, inputs are standardized (`Standardize=true`). For each species and kernel, the model is trained using `fitrgp` and assessed on the available datapoints by the re-substitution RMSE and R^2.

For each yield variable, the **best kernel** is selected as the one maximizing R^2 on the available datapoints. The selected models are then used to generate smooth response surfaces on a dense grid in (T_C, τ) for visualization and comparison between measured and reconciled fits.

Chapter 4

Results and Discussion

4.1 Kinetic Model Verification

4.1.1 Conversion–Temperature Profiles

The Arrhenius-type random-scission model produced a characteristic sigmoidal conversion curve $\alpha(T)$ under linear heating. Conversion remained close to zero at low temperatures, followed by a rapid rise in the main degradation interval and a gradual approach to $\alpha \approx 1$ at higher temperatures. The predicted shape reflects the typical onset–acceleration–completion pattern observed for LDPE primary thermal degradation in an inert environment.

As an initial consistency check, the model was simulated using the parameter values reported by Zhao et al. [22] and compared to manually scraped experimental data of $\alpha(T)$ from the reference figure.

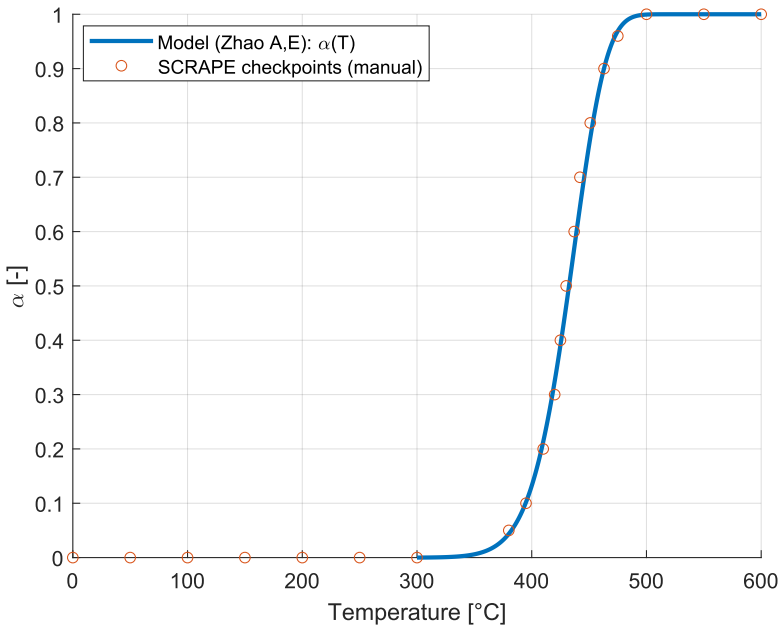

Figure 4.1: Simulated conversion curve $\alpha(T)$ using the parameter values reported by Zhao et al. compared with manually scraped experimental datapoints from the reference figure.

Figure 4.1 shows that the simulated curve closely follows the scraped points throughout the transition region, confirming that both the numerical integration of the kinetic ODE and the random-scission mapping $\alpha(x)$ were implemented consistently with the reference trend.

4.1.2 Self-Consistency and Robustness Tests

To verify robustness to data imperfections, two parameter estimations were performed.

Self-consistency test. Synthetic conversion data $\alpha(T)$ were generated from the reported parameters in Table 4.1 and subsequently used as input for estimation of the Arrhenius parameters (A, E) via least-squares fitting (reconstructing $x(T)$ from $\alpha(T)$ and fitting the integrated model). The fitted curve nearly overlapped the synthetic truth (Figure 4.2), achieving $\mathrm{RMSE}_\alpha = 1.55 \times 10^{-2}$ and $R^2_\alpha = 0.99878$.

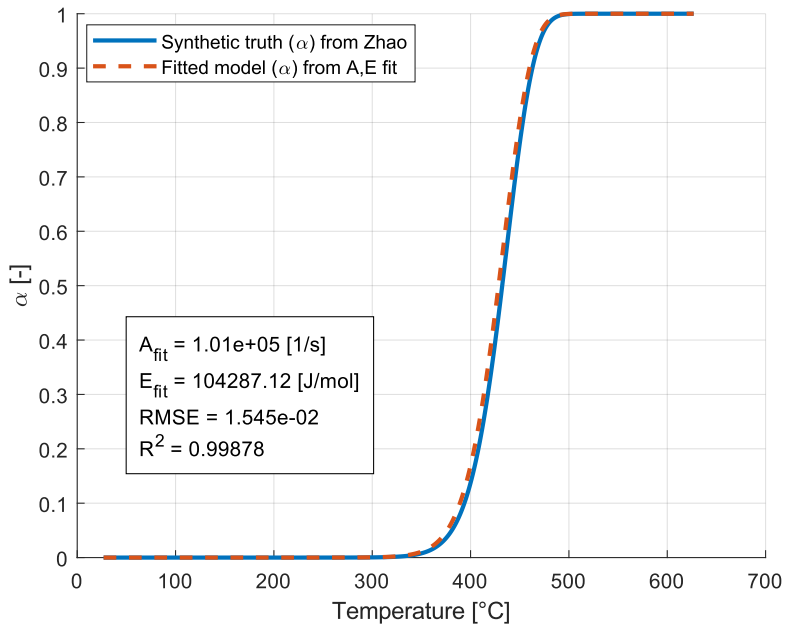

Figure 4.2: Self-consistency test: recovery of (A, E) from synthetic conversion data generated by the reference parameter set. The fitted curve nearly overlaps the synthetic truth.

Robustness test To emulate measurement noise, Gaussian noise with standard deviation $\sigma_{\text{noise}} = 0.01$ was added to a sparse set of synthetic conversion points (clipped to $[0, 1]$). The parameters (A, E) were again re-estimated from the noisy data. The fitted model retained excellent predictive agreement on the noisy datapoints (Figure 4.3), with $\text{RMSE}_\alpha = 1.61 \times 10^{-2}$ and $R^2_\alpha = 0.99841$. Across both recovery tests, the fitted activation energy remained close to the true value (within ≈ 3.4–3.5%), while the fitted pre-exponential factor exhibited a larger deviation. This behavior is consistent with practical parameter correlation in Arrhenius-type models, where multiple (A, E) combinations can reproduce a very similar $\alpha(T)$ shape. For the purposes of implementation verification and robustness demonstration, the key result is the stable reproduction of the conversion curve with very high goodness-of-fit.

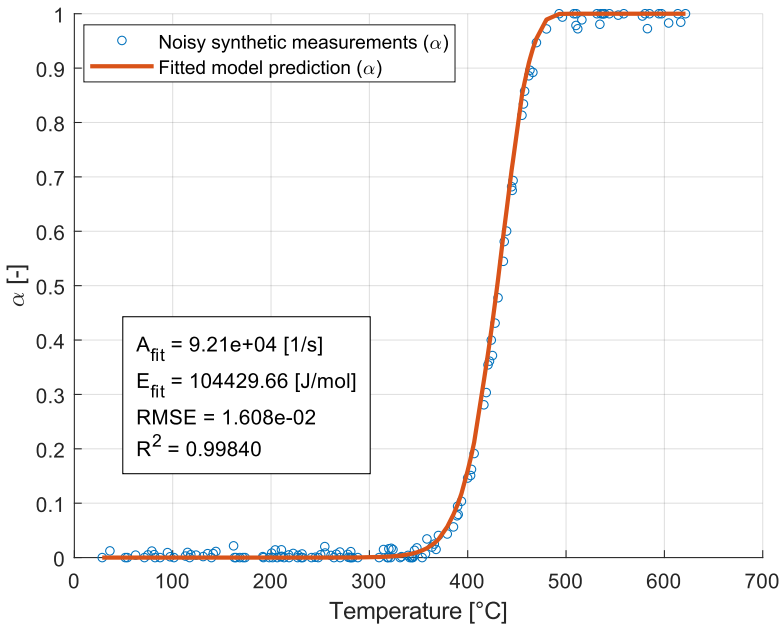

Figure 4.3: Robustness test: parameter estimation from noisy synthetic conversion data ($\sigma_{\text{noise}} = 0.01$). The model provides a good fit to the noisy data points.

Table 4.1: Parameter-recovery tests for the Arrhenius-based random-scission conversion model.

Test	A_{true} [s^{-1}]	E_{true} [kJ mol^{-1}]	A_{fit} [s^{-1}]	E_{fit} [kJ mol^{-1}]	RMSE_α	R^2_α
Self-consistency	1.80×10^5	108.1	1.01×10^5	104.3	1.55×10^{-2}	0.99878
Noisy ($\sigma_{\text{noise}} = 0.01$)	1.80×10^5	108.1	9.21×10^4	104.4	1.61×10^{-2}	0.99841

4.2 Data Reconciliation Results

This section summarises the key outcomes of the weighted least-squares reconciliation. All reconciled values satisfy the imposed conservation constraints, while staying as close as possible to the reported measurements under the specified weighting strategy.

4.2.1 Measured and Reconciled Phase Yields

Figure 4.4 compares measured and reconciled phase yields over all experiments.

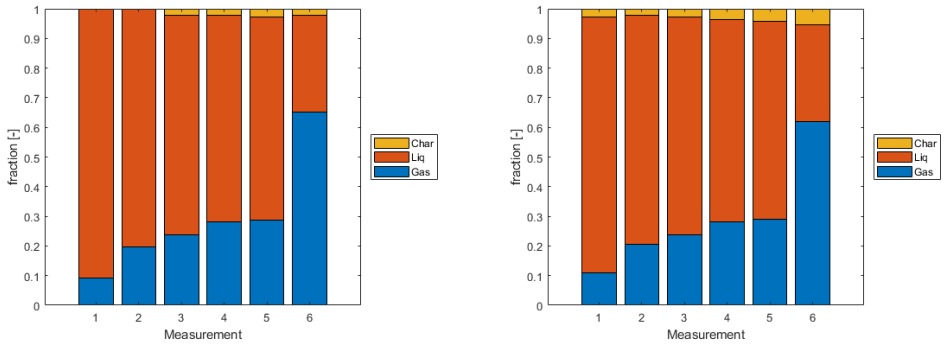

Figure 4.4: Phase yields (gas/liquid/char) before and after reconciliation. Left: measured phase distribution. Right: reconciled phase distribution satisfying exact closure.

As shown in Figure 4.4, the reconciliation redistributes the mass-balance discrepancy in a way that is consistent with the uncertainty weighting. In particular, an increase in the char fraction and a corresponding decrease in the liquid fraction can be observed after reconciliation.

Table 4.2: Measured and reconciled phase yields.

Exp.	T_C [°C]	τ [s]	$Y_{\text{gas}}^{\text{meas}}$ [wt%]	$Y_{\text{gas}}^{\text{rec}}$ [wt%]	$Y_{\text{liq}}^{\text{rec}}$ [wt%]	$Y_{\text{char}}^{\text{rec}}$ [wt%]
1	500	13.4	8.20	10.77	83.84	2.74
2	550	12.4	18.00	20.00	75.90	2.10
3	550	13.9	23.40	23.71	73.19	2.62
4	550	17.8	26.30	27.75	66.92	3.72
5	550	20.4	27.30	28.47	66.08	4.08
6	600	13.7	56.80	59.93	31.67	5.23

4.2.2 Mass and Elemental Balance Closure

After reconciliation, the total mass balance and the elemental balances (C, H, O) are satisfied exactly. The unreconciled measurements do not close perfectly; hence the reconciliation redistributes the discrepancy among the variables according to the

assigned uncertainties. Figure 4.5 shows the fixed inlet C/H split implied by the CH_2 composition; this quantity is constant across all datasets.

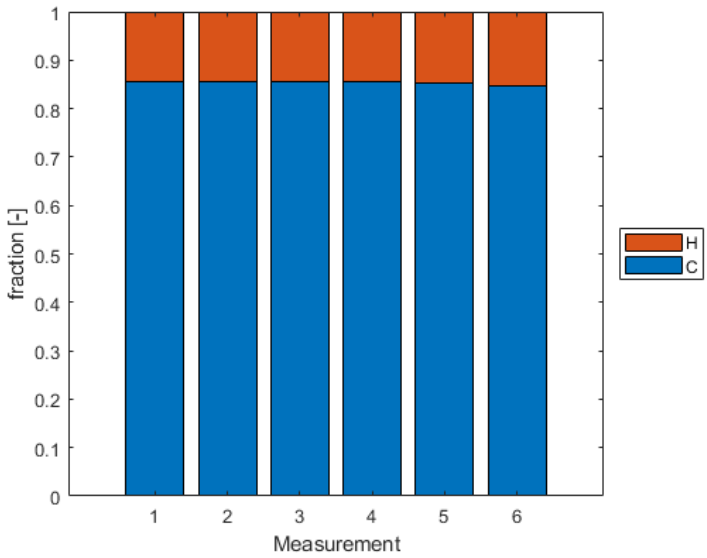

Figure 4.5: Normalized inlet elemental masses (C and H) used to define elemental-balance constraints for the reconciliation.

This confirms that the reconciled dataset is physically consistent and suitable as an input for regression and GPR yield modelling without propagating systematic imbalance errors into fitted correlations.

4.2.3 Effect of Reconciliation on Gas-Phase Composition

Reconciliation affected individual gaseous components primarily through redistribution within the gas-phase species set while respecting: (i) total gas mass consistency, (ii) elemental consistency, and (iii) the reported uncertainty-driven weights. Figure 4.6 compares the measured and reconciled gas composition (normalized) across operating points.

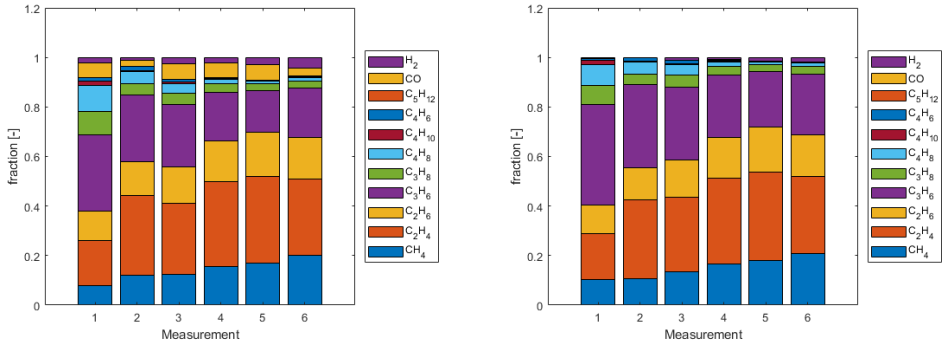

Figure 4.6: Gas composition before and after reconciliation. Left: measured gas composition (normalised). Right: reconciled gas composition (normalised) consistent with phase totals and elemental balances.

In Figure 4.6 it can be seen that, after reconciliation, the CO and C_5H_{12} contributions are driven to (near-)zero, while the relative share of lighter hydrocarbons increases. This shift is consistent with the imposed balance structure and modelling assumptions: the feed is treated as oxygen-free and CO is the only oxygen-carrying species, so enforcing the elemental balances implies a vanishing oxygen inventory and therefore suppresses CO. Similarly, species with weak or missing uncertainty information can be adjusted more strongly by the reconciliation to satisfy global closure. Overall, the reconciled composition is chemically and elementally consistent with the constraints, indicating that the reconciliation achieved the intended physical closure.

4.3 Yield Modelling Results

This section compares a baseline linear plane fit against Gaussian Process Regression models trained on the same small dataset ($n_{\text{exp}} = 6$). The comparison is performed for both measured and reconciled datasets to assess the benefit of reconciliation for subsequent modelling.

4.3.1 Linear Regression Results

For each yield variable Y, the linear regression model

$$\hat{Y}(T_C, \tau) = \theta_0 + \theta_T T_C + \theta_\tau \tau$$

was fitted by OLS. Despite its simplicity, the plane fit provides a transparent baseline and helps identify dominant directional trends over the operating window. In most cases, the fitted surfaces suggest that temperature drives the primary trend, while residence time provides only a secondary correction. This is also consistent with the fact that the available range of τ is relatively narrow compared to the temperature variation, which limits the identifiability of residence-time effects from the dataset.

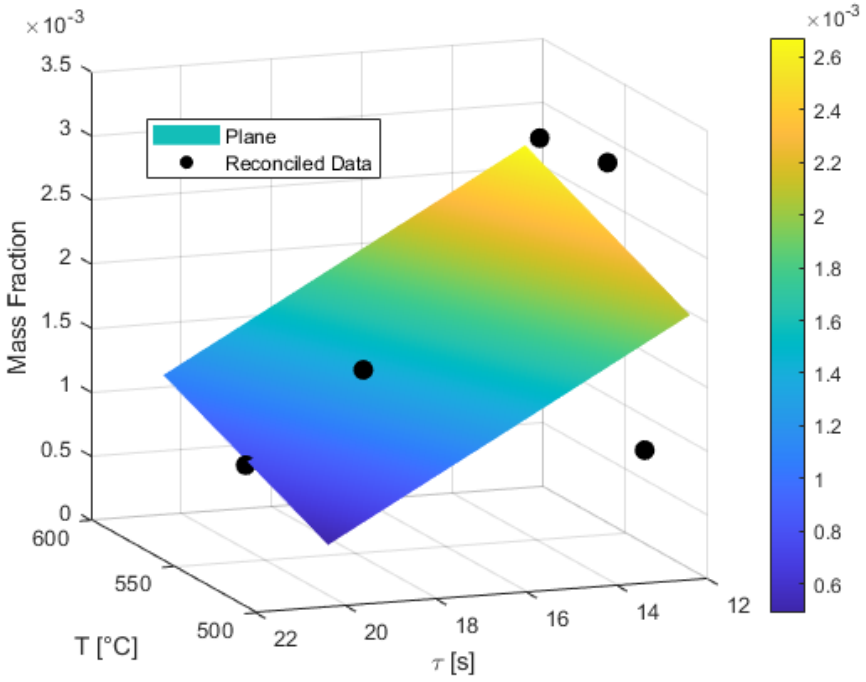

Figure 4.7: Representative plane-fit response surface (linear regression, OLS; reconciled data) for C_4H_6.

As can be seen in Table 4.3, the coefficient of determination R^2 decreases after reconciliation for most variables. This is likely because the balance-constrained adjustments can modify the apparent trends and may reveal nonlinear dependence of yields on temperature and residence time, which a simple linear plane model cannot capture adequately.

Table 4.3: Plane-fit metrics for measured vs reconciled data (R^2 only).

Variable Y	R^2_{meas}	R^2_{rec}
Gas (wt%)	0.8865	0.8648
Liquid (wt%)	0.8022	0.8270
Char (wt%)	0.7759	0.6196
CH_4	0.8223	0.7866
C_2H_4	0.9303	0.9212
C_2H_6	0.8715	0.8552
C_3H_6	0.9171	0.8736
C_3H_8	0.7718	0.7547
C_4H_8	0.9645	0.9755
C_4H_{10}	0.8225	0.7994
C_4H_6	0.4006	0.4248
C_5H_{12}	0.5546	0.5546
CO	0.7921	NaN
H_2	0.7181	0.7180

4.3.2 Gaussian Process Regression Results

GPR models were trained for each yield variable using multiple ARD kernels. Because the dataset is very small, the reported RMSE and R^2 values (computed on the training points) should be interpreted primarily as an indicator of interpolation capability rather than true generalisation. Nevertheless, kernel comparison remains informative: different kernels encode different smoothness assumptions and can produce meaningfully different surfaces between measured datapoints.

As shown in Tables 4.4–4.5, the predictive accuracy changes noticeably after reconciliation. In particular, the fits improve for C_3H_6, C_3H_8, and C_4H_8, whereas H_2 and C_4H_6 exhibit a clear degradation in accuracy. Across most variables, the most frequently selected kernel is the Exponential ARD kernel (Exp-ARD).

Table 4.4: GPR kernel comparison — measured data (resubstitution metrics). Best kernel per variable is the one with highest R^2.

Variable Y	**SE-ARD** R^2 RMSE	**Exp-ARD** R^2 RMSE	**M3/2-ARD** R^2 RMSE	**M5/2-ARD** R^2 RMSE	**RQ-ARD** R^2 RMSE	**Best kernel**
H_2	0.9860 0.000928	**0.9927** 0.000670	0.9858 0.000933	0.9859 0.000932	0.9860 0.000928	**Exp-ARD**
CH_4	0.99992 0.000316	**0.99999** 7.85e-05	0.99995 0.000256	0.99992 0.000309	0.99992 0.000316	**Exp-ARD**
C_2H_4	0.99997 0.000249	**1.0000** 4.72e-05	1.0000 0.000107	0.99999 0.000153	0.99999 0.000175	**Exp-ARD**
C_2H_6	0.9968 0.00154	**0.99999** 0.000102	0.9976 0.00133	0.9972 0.00144	0.9970 0.00150	**Exp-ARD**
C_3H_6	**0.9793** 0.00396	0.9793 0.00396	0.9793 0.00396	0.9793 0.00396	0.9793 0.00396	**SE-ARD**
C_3H_8	0.84392 0.00126	0.84392 0.00126	0.84392 0.00126	**0.84392** 0.00126	0.84392 0.00126	**M5/2-ARD**
C_4H_6	**0.2968** 0.000758	7.85e-06 0.000904	6.45e-06 0.000904	4.09e-06 0.000904	1.86e-06 0.000904	**SE-ARD**
C_4H_8	0.9737 0.000396	**0.9745** 0.000390	0.9723 0.000407	0.9729 0.000402	0.9737 0.000396	**Exp-ARD**
C_4H_{10}	2.16e-06 0.000421	2.08e-06 0.000421	2.15e-06 0.000421	**2.31e-06** 0.000421	1.60e-06 0.000421	**M5/2-ARD**
C_5H_{12}	1.41e-06 4.85e-05	**1.70e-06** 4.85e-05	1.58e-06 4.85e-05	1.54e-06 4.85e-05	1.04e-06 4.85e-05	**Exp-ARD**
CO	0.99898 0.000180	**0.99901** 0.000177	0.99901 0.000178	0.99900 0.000178	0.99898 0.000180	**Exp-ARD**
Gas (wt%)	0.9927 0.0128	**0.9993** 0.0038	0.9942 0.0113	0.9935 0.0121	0.9935 0.0120	**Exp-ARD**
Liquid (wt%)	0.9976 0.0083	**1.0000** 6.11e-05	0.9996 0.0033	0.9983 0.0069	1.0000 0.000307	**Exp-ARD**
Char (wt%)	**0.9999** 0.000103	0.9999 0.000104	0.9999 0.000108	0.9999 0.000109	0.9999 0.000107	**SE-ARD**

Table 4.5: GPR kernel comparison — reconciled data (resubstitution metrics). Best kernel per variable is the one with highest R^2.

Variable Y	**SE-ARD** R^2 RMSE	**Exp-ARD** R^2 RMSE	**M3/2-ARD** R^2 RMSE	**M5/2-ARD** R^2 RMSE	**RQ-ARD** R^2 RMSE	**Best kernel**
H_2	0.9369 0.000850	**0.9403** 0.000827	0.9326 0.000879	0.9345 0.000866	0.9369 0.000850	**Exp-ARD**
CH_4	0.9995 0.000863	**1.0000** 7.39e-05	0.9997 0.000679	0.9995 0.000844	0.9995 0.000870	**Exp-ARD**
C_2H_4	0.9996 0.00106	**1.0000** 4.33e-05	0.9999 0.000138	0.9999 0.000192	0.9999 0.000155	**Exp-ARD**
C_2H_6	0.9969 0.00155	**0.99999** 9.53e-05	0.9978 0.00131	0.9974 0.00144	0.9971 0.00152	**Exp-ARD**
C_3H_6	0.9963 0.00198	**0.9963** 0.00198	0.9963 0.00198	0.9963 0.00198	0.9963 0.00198	**Exp-ARD**
C_3H_8	0.8753 0.00118	**0.8753** 0.00118	0.8753 0.00118	0.8753 0.00118	0.8753 0.00118	**Exp-ARD**
C_4H_6	**0.2841** 0.000760	7.49e-06 0.000898	5.71e-06 0.000898	5.49e-06 0.000898	1.86e-06 0.000898	**SE-ARD**
C_4H_8	**0.9834** 0.000317	0.9812 0.000337	0.9819 0.000331	0.9826 0.000325	0.9834 0.000317	**SE-ARD**
C_4H_{10}	2.14e-06 0.000424	2.07e-06 0.000424	2.14e-06 0.000424	**2.30e-06** 0.000424	1.58e-06 0.000424	**M5/2-ARD**
C_5H_{12}	1.41e-06 4.84e-05	**1.70e-06** 4.84e-05	1.58e-06 4.84e-05	1.54e-06 4.84e-05	1.04e-06 4.84e-05	**Exp-ARD**
CO	– 0.000000	– 0.000000	– 0.000000	– 0.000000	– 0.000000	–
Gas (wt%)	0.9996 0.00284	**1.0000** 7.55e-05	0.9996 0.00316	0.9996 0.00293	0.9996 0.00325	**Exp-ARD**
Liquid (wt%)	1.0000 0.000333	**1.0000** 6.07e-05	1.0000 0.000141	1.0000 0.000214	1.0000 0.000190	**Exp-ARD**
Char (wt%)	0.9995 0.000225	**0.9999** 0.000111	0.9996 0.000196	0.9996 0.000221	0.9995 0.000225	**Exp-ARD**

Data reconciliation can change the apparent shape of yield trends across the operating points. For some species (e.g., C_3H_6, C_3H_8, and C_4H_8), enforcing overall and elemental-balance closure tends to make the data more mutually consistent across experiments, so the subsequent regression and GPR models can represent these yields more accurately. In contrast, for H_2 and C_4H_6 the balance-driven corrections may alter the local dependence on (T_C, τ) (e.g., increase apparent nonlinearity or change the direction of the trend), which can reduce the achievable fit quality for the same model class.

For each yield variable, the best kernel in Tables 4.4–4.5 was used to train a GPR model.

Figure 4.8 shows representative examples, illustrating that GPR can capture non-linear trends between sparse datapoints, especially for yields that exhibit curvature with temperature or residence time.

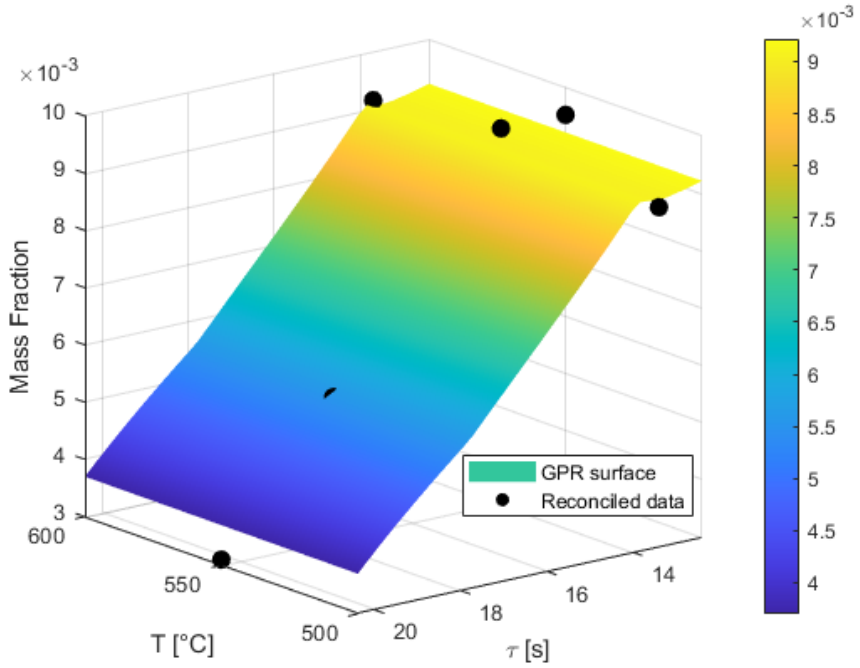

Figure 4.8: Representative GPR response surface (GPR, reconciled data) for C_4H_8 using the ARD exponential kernel.

4.3.3 Summary of Yields Modelling Results

Overall, the plane fit provides an interpretable baseline and reveals dominant linear severity trends, whereas GPR offers flexible non-linear interpolation over the small dataset. Moreover, with only $n_{\text{exp}} = 6$ datapoints, the GPR hyperparameter optimization can easily overfit the training points; therefore, an improved resubstitution fit does not necessarily imply reliable predictions away from the observed operating conditions.

4.4 CSTR Reactor Model Results

The reactor model was evaluated through the optimized steady-state solution. The following results focus on the optimized operating point, calculated product distribution, objective-function value, and material-balance consistency of the final solution.

4.4.1 Optimal Operating Point

The steady-state optimization converged successfully and returned a feasible optimum. The objective function value at the optimum was

$$\dot{\Pi}_{\mathrm{EUR}} = 0.278283 \ \mathrm{EUR\,s^{-1}}. \tag{4.1}$$

This value represents the simplified profit rate calculated from the economic objective used in the reactor model.

The optimized operating point is summarized in Table 4.6. The optimized inlet mass flow rate was $0.136896\ \mathrm{kg\,s^{-1}}$, the reactor temperature was 796.101 K, and the residence time reached its imposed lower bound of 10.000 s.

Table 4.6: Optimized operating point obtained from the steady-state optimization.

Variable	Initial guess	Bounds	Optimized value
$\dot{m}_{\mathrm{in}}$ [$\mathrm{kg\,s^{-1}}$]	0.200000	0.050000–0.200000	0.136896
T [K]	804.000	673.150–873.150	796.101
τ [s]	–	10.000–20.000	10.000

The optimized temperature lies inside the allowed temperature interval, while the residence time is located exactly at the lower constraint. This indicates that, under the selected economic objective and imposed bounds, the optimum favoured the shortest admissible residence time. The resulting operating point therefore corresponds to a relatively high-throughput regime within the constrained operating window.

4.4.2 Product Distribution at the Optimized Point

The calculated outlet mass flows at the optimized operating point are summarized in Table 4.7. The largest calculated product stream was methane, followed by ethylene and propene. Hydrogen and char were practically zero at this operating point.

Table 4.7: Calculated outlet mass flows at the optimized operating point.

Stream	Mass flow [$\mathrm{kg\,s^{-1}}$]
CH_4	0.0525747
C_2H_4	0.0246637
C_2H_6	0.0119131
C_3H_6	0.0179849
C_3H_8	0.00302732
C_4H_8	0.00786607
C_4H_{10}	0.00117037
C_4H_6	0.00176326
C_5H_{12}	0.00541271
H_2	0.000000
Char	6.93889×10^{-18}
Polyethylene-like remaining stream	0.0105198
Total outlet flow	0.13689593

The sum of the explicitly calculated product mass flows, excluding the polyethylene-like remaining stream, was

$$\dot{m}_{\mathrm{prod,exp}} = 0.12637613\ \mathrm{kg\,s^{-1}}. \tag{4.2}$$

The total outlet mass flow rate was defined as the sum of the explicitly calculated product stream and the polyethylene-like outlet stream:

$$\dot{m}_{\mathrm{out}} = \dot{m}_{\mathrm{prod,exp}} + \dot{m}^{*}_{\mathrm{PE}}. \tag{4.3}$$

At the optimized operating point, this gives

$$\dot{m}_{\mathrm{out}} = 0.12637613 + 0.0105198 = 0.13689593\ \mathrm{kg\,s^{-1}}. \tag{4.4}$$

This value agrees with the optimized inlet mass flow rate within numerical precision. The optimized solution is therefore materially consistent, since the calculated outlet stream closes the inlet mass flow.

The product distribution shows that the optimized point is dominated by light hydrocarbon formation. Methane reached 0.0525747 $\mathrm{kg\,s^{-1}}$, which was the highest individual product mass flow. Ethylene and propene also formed significant product streams, with 0.0246637 $\mathrm{kg\,s^{-1}}$ and 0.0179849 $\mathrm{kg\,s^{-1}}$, respectively. The polyethylene-like remaining stream was 0.0105198 $\mathrm{kg\,s^{-1}}$, which is small compared with the total inlet flow.

4.4.3 Interpretation of the Objective Function

The objective value of 0.278283 EUR s^{-1} represents the net value of the simplified economic function. It includes product revenue, heating cost, and the recoverable-energy credit. The result therefore indicates how favourable the optimized operating point is within the assumptions of the model, but it does not represent a complete industrial profit.

The location of the optimum at the lower residence-time bound is an important result. It shows that the simplified objective benefits from shorter residence time and higher throughput within the investigated operating window. This behaviour is influenced not only by product formation, but also by the energy credit assigned to the polyethylene-like remaining stream.

The result should therefore be interpreted with caution. If the lower residence-time bound were removed, the objective could favour even shorter residence times, where the material would have insufficient time for effective pyrolysis. In such a case, the apparent economic performance could be dominated by throughput and recoverable energy rather than by the formation of valuable pyrolysis products.

The optimized point is valid only for the simplified economic objective and the imposed operating constraints. Capital costs, separation and purification costs, compression costs, labour, maintenance, environmental costs, and product-quality requirements were not included. Consequently, the obtained optimum should be interpreted as a constrained model optimum and not as a final industrial design optimum.

4.5 Global System Analysis Results

4.5.1 Response Surface in the Temperature-Residence-Time Plane

The global system analysis was used to evaluate the behaviour of the reactor model over a wider operating region than the single optimum obtained from the steady-state optimization. In the GSA study, the reactor temperature T and inlet mass flow rate $\dot{m}_{\mathrm{in}}$ were varied as sampled factors. Since the residence time is calculated from the inlet mass flow rate, the results were subsequently transformed and visualized in the temperature–residence-time plane.

This transformation is useful because temperature and residence time are physically interpretable operating variables of the pyrolysis process. Temperature directly affects

the Arrhenius kinetic terms, while residence time determines how long the reacting material remains inside the reactor.

The main response evaluated in the GSA was the simplified profit rate,

$$\dot{\Pi}_{\mathrm{EUR}}, \tag{4.5}$$

which combines product revenue, heating cost, and the recoverable-energy credit. The resulting response surface, including the original optimization boundaries and the optimized operating point, is shown in Figure 4.9.

The dashed boundaries in Figure 4.9 represent the constrained operating region used in the optimization problem. These limits are important because the yield correlations used in the reactor model were developed from experimental data available only within a limited temperature and residence-time range. Therefore, points outside this region should be interpreted with caution, since they correspond to extrapolation of the fitted yield models.

The response surface shows that, within the investigated region, the simplified profit function is strongly influenced by residence time. The objective changes more significantly in the direction of decreasing τ than in the temperature direction, which indicates that residence time has the dominant influence on the economic response of the current model. Temperature still affects the response surface, but mainly by shaping the local position of the more profitable region within the constrained operating window.

The tendency of the objective function to increase at shorter residence times is a consequence of the simplified economic formulation. A lower residence time corresponds to a higher inlet throughput and, in the present objective function, also increases the contribution associated with the polyethylene-like remaining stream through the recoverable-energy term. If the residence-time bound were removed, the objective would tend to move toward very short residence times. Such a result would not represent a physically meaningful pyrolysis optimum, because the reactor would approach a situation where only a small fraction of polyethylene reacts and the polyethylene-like remaining stream dominates the outlet.

For this reason, the lower residence-time bound is an important modelling and interpretation constraint. It prevents the optimizer from selecting an artificial solution dominated by throughput and energy recovery from the polyethylene-like remaining stream rather than by the actual formation of valuable pyrolysis products. In practical terms, an unconstrained movement toward $\tau \to 0$ would correspond to a process in

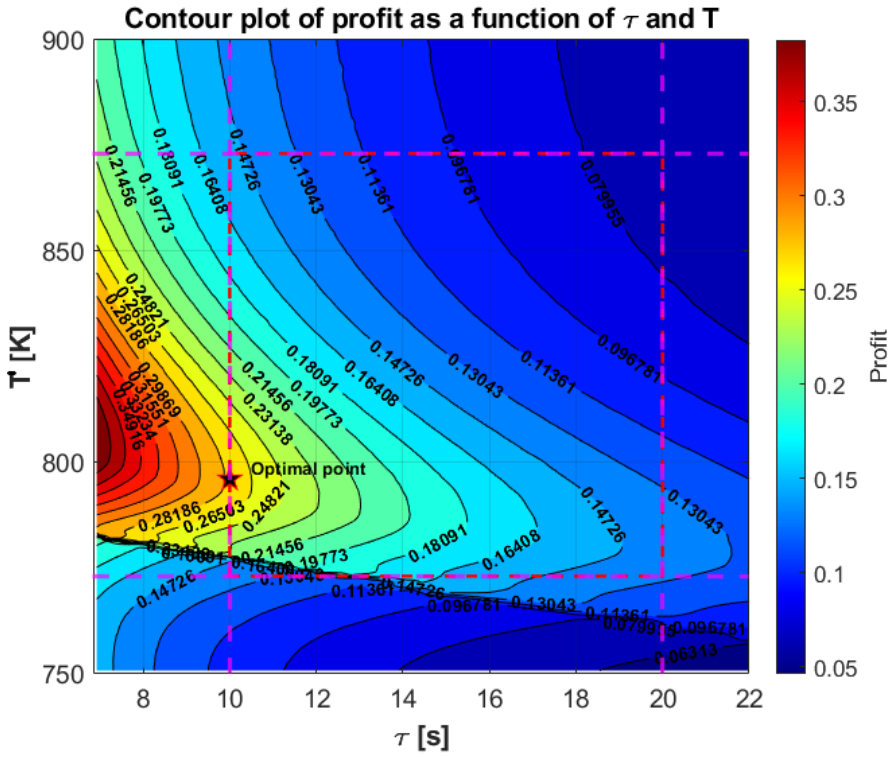

Figure 4.9: Global system analysis response surface of the simplified profit rate in the temperature–residence-time plane. The dashed lines indicate the boundaries of the optimization problem, and the optimized operating point is marked inside this constrained region.

which the material has insufficient time to react, and the economic result would be driven mainly by treating the remaining polyethylene-like stream as an energy source.

The temperature dependence is also visible in the response surface. At lower temperatures, the kinetic rate and product formation are limited. At higher temperatures, the heating cost increases. The resulting objective therefore reflects a compromise between product formation, energy demand, throughput, and the recoverable-energy credit included in the simplified profit function.

4.5.2 Comparison Between GSA and Optimization

The optimized point obtained from the steady-state gPROMS optimization is located on the lower residence-time constraint and inside the imposed temperature bounds. This is consistent with the trend observed in the GSA response surface, where the simplified objective favours shorter residence times within the allowed operating region.

However, this result should not be interpreted as a general recommendation to minimize the residence time without limitation. The GSA shows that, if the optimization boundaries were relaxed, the simplified objective would continue to favour even shorter residence times. This behaviour is mainly a consequence of the selected economic formulation and the inclusion of the recoverable-energy credit from the polyethylene-like remaining stream and char.

The comparison between the GSA response surface and the optimized point therefore provides an important interpretation of the optimization result. The optimizer did not identify a universal industrial process optimum, but rather the best operating point within the constrained model region and under the selected simplified objective function. The imposed boundaries are therefore essential, because they keep the optimization within a range where the empirical yield representation is still reasonably connected to the available experimental data.

Overall, the GSA confirms that the reactor model responds consistently to changes in temperature and residence time. At the same time, it highlights an important limitation of the simplified economic objective: without physically motivated bounds, the optimizer may favour operating points with very short residence time, where the process would no longer represent effective pyrolysis but rather a high-throughput case with a large amount of polyethylene-like remaining material.

Chapter 5

Conclusions

This thesis developed a modelling framework for polyethylene pyrolysis that combines literature-based kinetic modelling, data reconciliation, empirical yield modelling, steady-state reactor simulation, and simplified economic optimisation. The work was motivated by the need to build physically meaningful process models from experimental data that are reported on different bases and may not satisfy mass and elemental balances exactly.

A small literature dataset of LDPE pyrolysis experiments was first converted to a common mass basis and reconciled by weighted least squares subject to total-mass and elemental-balance constraints. This step improved the physical consistency of the data and provided a reconciled dataset suitable for further model development. The reconciled data were then used to construct empirical models for individual product yields as functions of temperature and residence time. A linear regression model was used as a transparent baseline, whereas Gaussian Process Regression provided a more flexible representation of nonlinear trends within the studied operating window.

The reactor model was formulated as a steady-state CSTR model and implemented in gPROMS. The model combined an Arrhenius-type random-scission description of polyethylene conversion with empirical yield correlations for selected products. This integrated model was subsequently used for steady-state optimisation and for global system analysis in order to evaluate the influence of operating conditions on conversion, product distribution, and the simplified economic objective.

The optimisation results showed that the obtained optimum is strongly influenced by the residence-time constraint and by the structure of the simplified objective function. In the present formulation, the optimum was located at the lower residence-time bound, and the global system analysis confirmed a general tendency of the objective to increase as residence time decreases. This result should not be interpreted as a general industrial optimum for chemical recycling. Rather, it shows that the reduced

objective function rewards short residence time and assigns value to the remaining material through the recoverable-energy term. The current optimisation therefore reflects the behaviour of the reaction-and-heating model under simplified assumptions, not a complete process-level economic optimum.

Several important costs and process features were intentionally excluded from the optimisation problem. These include separation and purification of products, compression, storage, labour, maintenance, shutdowns, equipment service, electricity consumption of auxiliary devices such as motors and compressors, and heat losses to the surroundings. For this reason, the calculated profit should be interpreted only as a simplified process-performance indicator. Under different market conditions, alternative product-price assumptions, or a more complete techno-economic formulation, the preferred operating region could shift toward conditions favouring a different balance between conversion, throughput, and production of valuable compounds.

The present work nevertheless shows that data reconciliation is a useful preprocessing step when small experimental pyrolysis datasets are used for model development. The thesis also demonstrates how reconciled literature data can be connected to reactor-level modelling and optimisation in a single workflow. This creates a basis for future work on larger and more diverse datasets, where more advanced data-driven models such as Gaussian Process Regression, or richer hybrid mechanistic-data-driven models, may become more beneficial.

Future work should focus on three main directions. First, the yield models should be trained and validated on a larger dataset covering a wider operating region. Second, the reactor model should be extended to represent the full product slate more rigorously, including heavier liquid products and downstream process requirements. Third, the economic model should be expanded toward a more realistic techno-economic assessment that includes utility consumption, heat losses, downstream separation, and additional operating and maintenance costs. These extensions would improve both the predictive value of the model and the practical relevance of the optimisation results.

Appendix A

Resumé

Táto diplomová práca sa zaoberá modelovaním pyrolýzy polyetylénu so zameraním na prepojenie kinetického modelovania, spracovania experimentálnych dát, rekonciliácie dát, empirického modelovania výťažkov, reaktorovej simulácie a zjednodušenej procesnej optimalizácie. Východiskom práce je skutočnosť, že produkcia plastov dlhodobo rastie, s čím súvisí aj zvyšujúce sa množstvo plastového odpadu, ktorého udržateľné spracovanie predstavuje významný technický a environmentálny problém. Popri tradičných spôsoboch nakladania s odpadmi sa preto zvyšuje význam technológií umožňujúcich jeho materiálové alebo energetické zhodnotenie. Pyrolýza patrí medzi relevantné termochemické prístupy, pretože umožňuje spracovať polymérny materiál v prostredí bez prístupu kyslíka na plynné, kvapalné a tuhé produkty. Výsledná produktová distribúcia je však silne závislá od typu polyméru, teploty, zdržného času a ďalších prevádzkových podmienok, čo vyžaduje použitie vhodných matematických modelov na analýzu a optimalizáciu procesu.

Hlavným cieľom práce bolo vytvoriť a vyhodnotiť modelovací rámec pre pyrolýzu polyetylénu, ktorý spája čiastkové kroky do uceleného pracovného postupu. Konkrétne išlo o implementáciu a overenie konverzného modelu založeného na Arrheniovej kinetike a mechanizme náhodného štiepenia reťazca, spracovanie publikovaných experimentálnych dát pre pyrolýzu LDPE, ich prevod na jednotnú bázu, rekonciliáciu pomocou metódy vážených najmenších štvorcov, vytvorenie empirických modelov výťažkov a následné využitie týchto modelov v zjednodušenom ustálenom modeli CSTR reaktora. Súčasťou práce bolo aj zjednodušené ekonomické hodnotenie, optimalizácia ustáleného stavu a globálna systémová analýza. Rozsah práce bol vedome obmedzený na výpočtové modelovanie a analýzu v rámci podmienok podporených dostupnými experimentálnymi údajmi. Práca preto nezahŕňa detailný návrh reaktora, separačných a čistiacich uzlov, kapitálové náklady ani kompletnú techno-ekonomickú analýzu.

Teoretická časť práce opisuje základný princíp pyrolýzy a tepelnú degradáciu polyetylénu. Polyetylén pri zvýšenej teplote degraduje najmä mechanizmom štiepenia polymérnych

reťazcov, pri ktorom vznikajú ľahšie uhľovodíkové fragmenty. Z hľadiska matematického opisu je v práci dôležité rozlíšenie medzi premennou x, reprezentujúcou frakciu rozštiepených väzieb v polymérnom reťazci, a celkovou konverziou α, ktorá vyjadruje experimentálne pozorovateľný hmotnostný úbytok, respektíve premenu vstupného materiálu na prchavé produkty. Premenná x sa používa ako vnútorná kinetická veličina, zatiaľ čo celková konverzia α sa z nej počíta pomocou vzťahu odvodeného z modelu náhodného štiepenia reťazca. Toto rozlíšenie je kľúčové, pretože umožňuje oddeliť vnútorný kinetický opis degradácie polyméru od výpočtu skutočných hmotnostných tokov konvertovaného a nekonvertovaného polyetylénu v modeli reaktora.

Súčasťou teoretického rámca sú aj celkové hmotnostné a elementárne bilancie, ktoré tvoria základ rekonciliácie dát. Pri ideálnom experimente by mali publikované údaje spĺňať celkovú hmotnostnú bilanciu a zároveň aj bilancie jednotlivých prvkov, najmä uhlíka, vodíka a kyslíka. V praxi však experimentálne dáta často obsahujú neistoty, čiastkové bilančné neuzávery alebo sú vykazované na odlišných bázach, čo komplikuje ich priame použitie na identifikáciu modelov. Z tohto dôvodu boli v práci bilančné vzťahy použité ako fyzikálne obmedzenia pri rekonciliácii. Rekonciliácia bola formulovaná ako úloha vážených najmenších štvorcov, v ktorej sa hľadajú upravené hodnoty meraných veličín tak, aby boli čo najbližšie k publikovaným údajom, ale zároveň presne spĺňali stanovené bilančné a fyzikálne obmedzenia. Váhy boli určené na základe dostupných informácií o neistotách, pričom veličiny s menšou neistotou boli penalizované prísnejšie.

Experimentálny súbor dát použitý v práci pochádza z literatúry a obsahuje šesť prevádzkových bodov pre pyrolýzu LDPE. Teplotný rozsah dát je približne 500 až 600 °C a zdržné časy sa pohybujú v intervale 12 až 21 s. Dataset je špecifický tým, že jednotlivé veličiny sú reportované na rôznych bázach. Fázové výťažky plynu, kvapaliny a tuhého zvyšku sú uvedené ako hmotnostné percentá, jednotlivé plynné uhľovodíky ako uhlíkové výťažky a vodík ako objemové percento v plynnej fáze. Pred samotnou rekonciliáciou preto bolo potrebné previesť všetky údaje na jednotnú hmotnostnú bázu. Vstupná hmotnosť polyméru bola normalizovaná na jednotkovú hodnotu, polyetylén bol modelovaný ako opakujúca sa jednotka CH_2, z čoho boli vypočítané vstupné hmotnosti uhlíka a vodíka. Na rovnakej báze boli následne vyjadrené aj hmotnosti jednotlivých fáz a plynných zložiek. Pri spracovaní dát boli použité aj doplnkové predpoklady, napríklad idealizovaný prepočet vodíka z objemového podielu na hmotnosť alebo zjednodušené predpoklady o elementárnom zložení kvapalnej a tuhej fázy (charu).

Rekonciliácia bola riešená samostatne pre každý experimentálny bod. Rozhodovacie premenné zahŕňali celkovú a elementárnu vstupnú hmotnosť, hmotnosti fáz (plyn, kvapalina, char), hmotnosti jednotlivých plynných produktov a ich elementárne príspevky. Na tieto veličiny boli aplikované obmedzenia celkovej hmotnostnej bilancie, bilancie

uhlíka, vodíka a kyslíka, uzatvorenia plynnej fázy a podmienky nezápornosti. Váhová matica bola zostavená ako diagonálna, pričom jednotlivé váhy vychádzali z odhadov neistôt. Pri plynných zložkách sa na určenie neistoty využili publikované absolútne chyby; pre veličiny bez udanej neistoty bola použitá aproximácia, aby sa predišlo singulárnej definícii váh. Vstupná hmotnosť a celkové fázové podiely boli viazané vyššími váhami, aby rekonciliácia zachovávala globálne veličiny blízko publikovaných hodnôt a redistribuovala neuzávery prednostne na úrovni jednotlivých plynných zložiek.

Výsledky rekonciliácie potvrdili, že pôvodné merané dáta nie sú úplne konzistentné z pohľadu celkovej hmotnostnej ani elementárnej bilancie. Po vyriešení úlohy vážených najmenších štvorcov však všetky rekonciliované údaje spĺňali bilančné obmedzenia v rámci numerickej tolerancie, čím vznikol fyzikálne konzistentný dataset vhodný na ďalšie modelovanie. Z hľadiska fázových výťažkov došlo k systematickej redistribúcii: podiel plynnej fázy sa oproti meraným údajom zvýšil, pri kvapalnej fáze sa prejavil pokles a pri tuhom zvyšku mierny nárast. Táto zmena naznačuje, že pôvodné literatúrne údaje obsahovali neuzáveru, ktorá sa po uložení bilančných podmienok rozdelila podľa nastavenej váhovej stratégie medzi menej spoľahlivé premenné.

Podobne sa zmenilo aj zloženie plynnej fázy. Rekonciliácia upravila relatívne zastúpenie zložiek tak, aby bola dodržaná celková plynná bilancia aj elementárna konzistencia. Keďže vstupný polyetylén bol uvažovaný ako bezkyslíkatý a oxid uhoľnatý bol jedinou zložkou v plynnej fáze obsahujúcou kyslík, uplatnenie kyslíkovej bilancie viedlo k potlačeniu CO na nulové hodnoty. Výraznejšie boli upravené aj tie zložky, pri ktorých chýbala spoľahlivá informácia o neistote. Výsledkom rekonciliácie teda nebolo len číselné spracovanie dát, ale najmä vytvorenie datasetu, ktorý rešpektuje fyzikálny význam veličín a môže byť bezpečnejšie použitý pri regresnom modelovaní.

Na základe meraných aj rekonciliovaných údajov boli následne zostavené modely výťažkov ako funkcie teploty a zdržného času. Ako základný model bola použitá lineárna regresia vo forme roviny. Tento prístup bol zvolený ako jednoduchá a interpretovateľná referencia na identifikáciu základných trendov. Výsledky ukázali, že v dostupnom rozsahu dát má výraznejší vplyv teplota než zdržný čas, čo súvisí aj s úzkym intervalom experimentálnych zdržných časov. Pri viacerých výťažkoch sa po rekonciliácii znížila hodnota koeficientu determinácie R^2, čo naznačuje, že lineárny model už nedokáže dostatočne zachytiť zmeny trendov po vynútení bilančnej konzistencie.

Popri lineárnej regresii bola použitá aj metóda Gaussian Process Regression (GPR). Tento prístup umožňuje flexibilnejší opis nelineárnych závislostí a je vhodný pri malom počte dát, ak sa predpokladá hladký priebeh odozvy. Pre jednotlivé produkty boli v prostredí MATLAB porovnané viaceré ARD jadrá (squared exponential, exponential,

Matérn 3/2, Matérn 5/2 a rational quadratic). Výkonnosť modelov bola hodnotená pomocou metrík R^2 a RMSE. Keďže dataset obsahoval iba šesť experimentov, tieto metriky slúžia najmä ako ukazovateľ schopnosti interpolácie. Ukázalo sa, že GPR poskytuje flexibilnejší opis než lineárna rovina a vo väčšine prípadov bolo ako najvhodnejšie vybrané exponential ARD jadro. Rekonciliácia prispela k lepšej modelovateľnosti niektorých zložiek (napr. C_3H_6, C_3H_8), zatiaľ čo pri iných (H_2) sa kvalita modelu zhoršila, čo naznačuje, že bilančné obmedzenia menia lokálny charakter závislostí.

Ďalšia časť práce bola venovaná formulácii zjednodušeného ustáleného CSTR modelu v prostredí gPROMS. Reaktorový model vychádza z kinetického opisu degradácie a prevodu vnútorného stavu štiepenia väzieb na celkovú konverziu. Rýchlosť štiepenia väzieb bola reprezentovaná dvojčlenným Arrheniovým vyjadrením prevzatým z literatúry. Premenná x bola modelovaná bilančne ako intenzívny stav ideálne miešaného prietokového reaktora, pričom v ustálenom stave výsledná hodnota vyplývala z rovnováhy medzi prietokovým transportom a kinetickým príspevkom. Z nej sa následne vypočítala celková konverzia α, pomocou ktorej sa vstupný tok polyetylénu rozdelil na konvertovanú a nekonvertovanú časť.

Produktová reprezentácia v modeli kombinovala random-scission rozdelenie fragmentov s empirickými koreláciami výťažkov. Random-scission opis poskytol uhlíkovo založené príspevky fragmentov rôznej veľkosti, ktoré boli následne rozdelené medzi explicitne reprezentované chemické zložky (metán, etylén, etán, propén, propán, bután, 1,3-butadién, pentán, vodík a char). Zvyšná časť konvertovaného materiálu bola interpretovaná ako nepriradený konvertovaný zvyšok (ťažšie uhľovodíkové frakcie). Týmto spôsobom vznikol konzistentný, hoci zjednodušený model produktového spektra vhodný na ekonomické vyhodnotenie.

Ekonomická účelová funkcia bola formulovaná ako rozdiel medzi hodnotou produktov, nákladmi na ohrev a kreditom za energiu z nekonvertovaného polyetylénu a charu. Produktové ceny boli zavedené ako orientačné veľkoobchodné hodnoty a náklady na energiu boli reprezentované koeficientom odvodeným z cien elektriny a zemného plynu. Táto funkcia neslúži ako kompletná techno-ekonomická analýza, ale ako indikátor procesnej výkonnosti na porovnávanie prevádzkových bodov. Do analýzy neboli zahrnuté investičné náklady, náklady na separáciu, prácu ani environmentálne aspekty.

Optimalizácia ustáleného stavu bola vykonaná s dvoma rozhodovacími premennými: teplotou reaktora a vstupným hmotnostným tokom. Zdržný čas bol vypočitaný z objemu reaktora a prevádzkových podmienok. Teplota bola obmedzená na interval $673,15$ až $873,15$ K, hmotnostný tok na $0,05$ až $0,20$ $\mathrm{kg\,s^{-1}}$ a zdržný čas na 10 až 20 s. Riešenie optimalizačnej úlohy viedlo k prípustnému optimu s hodnotou účelovej

funkcie $0,278283\ \mathrm{EUR\,s^{-1}}$. Optimálny hmotnostný tok bol $0,136896\ \mathrm{kg\,s^{-1}}$, teplota $796,101\ \mathrm{K}$ a vypočítaný zdržný čas dosiahol dolnú hranicu $10,000\ \mathrm{s}$. Výsledok ukázal, že pri zvolenom modeli a účelovej funkcii je preferovaná prevádzka s kratším zdržným časom a vyšším prietokom.

V optimalizovanom bode dominovali ľahké uhľovodíky (metán, etylén, propén). Hmotnostná konzistencia modelu bola potvrdená numerickou presnosťou bilancie tokov. Výsledok však treba interpretovať v kontexte zjednodušeného ekonomického opisu. Keďže účelová funkcia oceňuje aj energiu z nekonvertovaného materiálu, môže dochádzať k zvýhodňovaniu kratších zdržných časov aj v oblastiach s nižšou chemickou konverziou.

Na hlbšie pochopenie tohto správania bola vykonaná globálna systémová analýza. Odozvová plocha ziskovej rýchlosti potvrdila dominantný vplyv zdržného času na účelovú funkciu. Najvýraznejší rast zisku sa prejavoval smerom ku kratším zdržným časom, čo je v súlade s výsledkom optimalizácie. Analýza ukázala, že bez dolnej hranice zdržného času by optimum smerovalo k nereálne nízkym hodnotám τ, čo by predstavovalo skôr vysokoprietokový transport než zmysluplnú pyrolýzu. Časť analyzovanej oblasti zasahuje na hranicu experimentálnych dát, preto je potrebné body mimo tohto intervalu interpretovať ako lokálnu extrapoláciu.

Z vedeckého hľadiska práca potvrdzuje, že rekonciliácia dát je kľúčovým krokom pri práci s malými literatúrnymi datasetmi. Umožnila vytvoriť fyzikálne konzistentný súbor údajov vhodnejší pre regresiu. Výsledky ukázali, že zatiaľ čo lineárne modely sú dobrou referenciou, GPR dokáže pružnejšie zachytiť nelineárne trendy. Reaktorový model demonštroval možnosť prepojenia kinetiky, rekonciliovaných dát a ekonomickej analýzy do jedného celku.

Práca má však aj svoje obmedzenia. Experimentálny súbor je malý (6 bodov) a pokrýva úzky rozsah podmienok. Model CSTR je zjednodušený a nereprezentuje detailnú konštrukciu reálneho zariadenia. Ekonomická časť nezahŕňa všetky priemyselné náklady. Výsledky optimalizácie preto slúžia skôr ako modelová analýza správania systému za definovaných predpokladov než ako definitívny technologický návrh.

Na záver možno konštatovať, že diplomová práca vytvorila ucelený modelovací rámec pre pyrolýzu polyetylénu. Budúci výskum by sa mal zamerať na rozšírenie datasetu, validáciu modelov na nezávislých údajoch, detailnejší opis ťažkých frakcií a komplexnejšiu techno-ekonomickú analýzu zahŕňajúcu downstream operácie. Takéto rozšírenie by zvýšilo praktickú použiteľnosť modelu pri hľadaní optimálnych podmienok priemyselnej pyrolýzy.

Bibliography

[1] James R. Asay, S. R. Urzendowski, and Arthur H. Guenther. Ultrasonic and thermal studies of selected plastics, laminated materials, and metals. Technical Report AFWL-TR-67-91, Air Force Weapons Laboratory, Air Force Systems Command, Kirtland Air Force Base, New Mexico, January 1968. AD827596.

[2] Ayhan Demirbas. Pyrolysis of municipal plastic wastes for recovery of gasoline-range hydrocarbons. *Journal of Analytical and Applied Pyrolysis*, 72(1):97–102, 2004.

[3] David Kristjanson Duvenaud. *Automatic Model Construction with Gaussian Processes*. PhD thesis, University of Cambridge, Cambridge, 2014. PhD thesis.

[4] Echemi. Chemical product price data. online, 2026. Wholesale chemical price data; Northwest Europe values were preferred where available, otherwise available China-market values were used.

[5] Miroslav Fikar and Ján Mikleš. *Identifikácia systémov (System Identification)*. STU Bratislava, Bratislava, 1998.

[6] Ribhu Gautam, Shekhar R. Kulkarni, Ilman N. Zaini, Hengda Han, S. Mani Sarathy, Bassam B. Dally, and William L. Roberts. A review on interactions and reactor implications during pyrolysis of biomass- and plastic-based waste. *Applications in Energy and Combustion Science*, 26:100507, 2026.

[7] Roland Geyer, Jenna R. Jambeck, and Kara Lavender Law. Production, use, and fate of all plastics ever made. *Science Advances*, 3(7):e1700782, 2017.

[8] Jenna R. Jambeck et al. Plastic waste inputs from land into the ocean. *Science*, 347(6223):768–771, 2015.

[9] Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining. *Introduction to Linear Regression Analysis*. John Wiley & Sons, Hoboken, 5 edition, 2012.

[10] Niklas Netsch, Michael Zeller, Frank Richter, Britta Bergfeldt, Salar Tavakkol, and Dieter Stapf. Energy demand for pyrolysis of mixed thermoplastics and waste plastics in chemical recycling: Model prediction and pilot-scale validation. *Sustainable Resource Management*, 1(7):1485–1492, 2024.

[11] Elmeri Pienihäkkinen, Christian Lindfors, Roman Tschentscher, Stephan Kubowicz, Syed Farhan Hashmi, Muhammad Hassam Khan, and Pouya Sirous-Rezaei. Catalytic and thermal fast pyrolysis of agricultural plastic waste: Comparison of btex and steam cracker feed production in a bubbling fluidized bed reactor. *Fuel*, 425:139426, 2026.

[12] Process Systems Enterprise Ltd. *gPROMS ModelBuilder Guide*, n.d. Introductory/user guide for gPROMS ModelBuilder.

[13] Pedro E. Sánchez-Jiménez et al. A new model for the kinetic analysis of thermal degradation of polymers driven by random scission. *Polymer Degradation and Stability*, 95(5):733–739, 2010.

[14] John Scheirs and Walter Kaminsky, editors. *Feedstock Recycling and Pyrolysis of Waste Plastics: Converting Waste Plastics into Diesel and Other Fuels*. John Wiley & Sons, Chichester, 2006.

[15] Shafferina Dayana Anuar Sharuddin et al. A review on pyrolysis of plastic wastes. *Energy Conversion and Management*, 115:308–326, 2016.

[16] Robert Simha and L. A. Wall. Kinetics of chain depolymerization. *The Journal of Physical Chemistry*, 56(6):707–715, 1952.

[17] Slovenský plynárenský priemysel, a.s. Cenníky plynu. online, 2026.

[18] Slovenský plynárenský priemysel, a.s. Ceny elektriny. online, 2026.

[19] Yanshan Yin, Shuo Wang, Liang Zhao, Wei Zhang, Wenran Gao, Bo Gu, Rui Liu, Shitong Liu, Ruiqi Wang, Hong Tian, Nai Shi, and Zhiliang Wu. A review of high-grade pyrolysis oils from plastic waste: Mechanisms, process strategies, and sustainability perspectives. *Renewable and Sustainable Energy Reviews*, 238:117035, 2026.

[20] Feichi Zhang, Muhao Li, Xiaoyue Ma, Niklas Netsch, Salar Tavakkol, Thorsten Zirwes, Rui Zhang, and Dieter Stapf. Assessment of the interplay between convective heat transfer and reaction kinetics during plastic pyrolysis. *Particuology*, 113:180–188, 2026.

[21] Mingyuan Zhang, Lei Ren, Yong Wang, Jingwei Geng, Zifeng Li, Haiping Shen, Ming Dong, and Charles Chunbao Xu. A short review of waste plastic chemical recycling by pyrolysis: From process parameters and catalytic mechanisms to product upgrading. *Journal of Analytical and Applied Pyrolysis*, 196:107710, 2026.

[22] Dongting Zhao et al. The chemistry and kinetics of polyethylene pyrolysis: A process to produce fuels and chemicals. *ChemSusChem*, 13(7):1764–1774, 2020.