Beans to Bytes: Grey-Box Nonlinear System Identification Using Hybrid Physics-Neural Network Models

by

Morgen Pronk

B.S. Mechanical Engineering, University of Alberta, 2018

Submitted to the System Design and Management Program
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN ENGINEERING AND MANAGEMENT

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2024

Authored by: Morgen Pronk
System Design and Management Program
August 9, 2024

Certified by: Brian W. Anthony
Director, Master of Engineering in Advanced Manufacturing & Design
Associate Director, MIT Nano
Principle Research Scientist, MIT Mechanical Engineering
Thesis Supervisor

Accepted by: Joan Rubin
Executive Director, System Design & Management Program,

Beans to Bytes: Grey-Box Nonlinear System Identification Using Hybrid Physics-Neural Network Models

by

Morgen Pronk

Submitted to the System Design and Management Program
on August 9, 2024 in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN ENGINEERING AND MANAGEMENT

ABSTRACT

The advancement of neural networks in the last several years has yielded some astonishing results. However, the applicability to system identification and modelling dynamical systems still has a great amount of room for exploration. This thesis reviews different neural network architectures and their application to complex non-linear dynamic system identification. In particular, it uses the intricate process of coffee roasting as a case study to explore and demonstrate these techniques. Coffee roasting is a complex process that requires precise control to achieve the desired coffee quality. The ability to develop models that represent a system, i.e. system identification, is of great value to industry. Coffee roasting poses several challenges for system identification from complex chemical reactions occurring inside the bean, to temperature trajectories being dependent on several states that cannot be explicitly measured, such as moisture content, or reaction rate, making it an ideal candidate for exploring the application and limitations of neural networks. The primary contributions of this study are a proposed "grey-box" model that augments previously established physics based models, as well as illustrating the limits of LSTM, Deep NARX models using "one-step" forward prediction techniques. Although the study focuses explicitly on coffee roasting, the conclusions drawn are applicable to other similarly complex industrial and manufacturing processes.

Thesis supervisor: Brian W. Anthony
Title: Director of Master of Engineering in Advanced Manufacturing and Design Program, Associate Director of MIT Nano, and MIT Mechanical Engineering Principal Research Scientist

Acknowledgments

This thesis would not have been possible without the support of many, but first and foremost, my family.

To my wife and children, thank you for enduring my demanding schedule and the unconventional arrangements it required. Your understanding and flexibility, particularly during my time in Cambridge while you were in Japan, were essential to completing this work. To my parents, I deeply appreciate you travelling from Canada to Cambridge to provide much needed support during this challenging period.

I extend my sincere gratitude to my advisor, Professor Brian Anthony, for his guidance, patience, and mentorship throughout this research. His expertise in control theory was invaluable, and I am particularly thankful for the opportunity to teach in his Advance Analytics class. This experience broadened my understanding of advanced analytics and allowed me to share my knowledge with others.

I am deeply grateful to my friend Samuel Gomez, whose encouragement and introduction to Professor Anthony were pivotal in setting me on this path. His insights and support were instrumental throughout my studies.

I would also like to thank Bryan Moser and Joan Rubin, the SDM program advisor and executive director, for their guidance and support during my time in the program.

To everyone mentioned above, and to all those who have contributed to my journey, I offer my sincere thanks.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Coffee, a beverage enjoyed globally [1], undergoes a complex transformation from green bean to aromatic brew, intricately tied to the roasting process [2]. This procedure involves precise heat transfer and chemical reactions, where the quality and flavor of the final cup depend heavily on the precision and control during roasting. Inconsistent roasting can lead to underdeveloped flavors, burnt notes, or batch-to-batch variations, compromising both the sensory experience and economic value of the product [3].

To achieve the consistency and high quality demanded by discerning coffee drinkers, the coffee roasting industry increasingly relies on advanced control systems. Automatic control systems have been in use since antiquity, with examples like automatic fluid level control devices dating back to 300 to 1 B.C. The first industrial feedback controller, James Watt's flyball governor from 1769, marked the beginning of many advancements in control techniques [4]. Despite these advancements, modern tools and techniques still largely depend on linear, time-invariant models, even for systems that are nonlinear, time-varying, and uncertain [5]. Consequently, system identification for real systems often focuses on approximating them as linear to apply these control techniques.

One of the first steps in developing or optimizing control systems is understanding the system that is to be controlled. This often involves deriving differential equations from "first principles" or well-known physics and mathematical laws. Numerical techniques, such as computational fluid dynamics (CFD) and finite-element analysis (FEA), allow engineers to model and simulate systems when a high level of geometrical detail is known [6]. However, explicit equations or geometrical details are often unavailable or challenging to develop, necessitating system identification techniques based on input and output data.

Modern advancements in AI, e.g. Non-linear algorithms, and computing are quickly changing this landscape. Advanced algorithms, such as neural networks, and increasingly powerful, affordable computers, make these techniques more accessible. This shift is par-

ticularly relevant in industrial control settings involving complex dynamic systems that are difficult to model from first principles, have complex nonlinear inter-dependencies, or are hard to observe.

While system identification has a long history in control engineering, traditional methods often struggle with complex, nonlinear systems. Advanced machine learning (ML) techniques, particularly neural networks, offer new opportunities for system identification, providing greater accuracy and robustness in modeling and control. However, the intersection of these fields remains relatively unexplored. This study investigates the application of machine learning for system identification in the context of coffee roasting, a highly nonlinear process with stringent control requirements for achieving high-quality taste. Coffee roasting is an ideal test case due to its complexity and the potential for significant industry impact through improved process control.

This thesis delves into the intricacies of coffee roasting as a representative example of a complex nonlinear process. However, the principles and techniques developed in this work are broadly applicable to modeling and control challenges in various industrial settings. The objective is to develop a model that accurately captures the dynamic behavior of the coffee roasting process, facilitating the design and optimization of control systems for improved consistency and quality. This research bridges the gap between traditional modeling approaches and advanced data-driven methods by leveraging the power of neural networks.

Coffee roasting's complexity poses significant challenges for traditional modeling approaches. While physics-based and data-driven methods have been used, each has limitations in capturing the process's intricate dynamics. This thesis explores a hybrid, or grey-box, approach that combines the strengths of both methodologies. Physics-based models, grounded in thermodynamics and heat transfer, provide a structured framework and interpretable results but struggle with complex nonlinear relationships. Neural networks, on the other hand, excel at learning patterns from data but lack interpretability and robustness. By integrating these methodologies, this research aims to develop a model that accurately represents the coffee roasting process and possesses both approaches' advantages.

This research focuses on dynamically modeling the roasting chamber, a critical component of the coffee roasting system responsible for heat transfer influencing the coffee's flavor profile. Traditional models often rely on simplified assumptions and constant parameters, limiting their accuracy. This thesis proposes a novel approach, incorporating a neural network within the differential equations representing the physical processes. By training this network on real-world production data, the model learns the complex relationships between key variables, such as heat transfer coefficient, bean movement, and airflow rates, not easily captured by traditional methods.

The development of this model proceeds through several key stages, outlined in the following chapter structure:

- Background and Literary Review - This chapter provides a comprehensive overview of coffee bean characteristics, the stages of the roasting process, and the various roasting equipment designs. It then delves into a detailed review of existing coffee roasting models, encompassing white-box, black-box, and grey-box approaches. The limitations of existing models in capturing the dynamic nature of the roasting process are highlighted, setting the stage for the proposed grey-box approach.

- Methods and Experiments - - This chapter outlines the methodologies employed in this research. It begins by describing the implementation and evaluation of a black-box Non-linear Auto-regressive with exogenous input (NARX) model, demonstrating its limitations in accurately representing the roasting process. The chapter then details the development of the proposed grey-box model. This involves analyzing the performance of existing physics-based models, identifying key parameters influencing model accuracy, and explaining the integration of a neural network to enhance the model's predictive capabilities. The chapter concludes by outlining the experimental setup, training process, and datasets used for model development and validation.

- Results and Discussion - This chapter presents and analyzes the results obtained from the different modeling approaches. It compares the performance of the black-box, white-box, and grey-box models, highlighting the advantages of the proposed grey-box approach in capturing the dynamic behavior of the coffee roasting process. The chapter discusses the model's generalization capabilities, limitations, and potential avenues for future research.

- Conclusion and Future Work - This chapter summarizes the key findings of the research, emphasizing the contributions made towards developing a robust and accurate model for simulating and controlling the coffee roasting process. It reiterates the limitations and future research directions, highlighting the broader implications of this work for the coffee industry and the field of process control.

This research contributes to the growing body of knowledge in coffee roasting modeling and control, as well as providing a framework for developing robust and accurate models for any complex nonlinear process. By combining the strengths of physics-based and data-driven approaches, this thesis offers a promising avenue for developing advanced control systems, ultimately leading to more consistent, high-quality production across various industries.

Chapter 2

Background and Literary Review

2.1 Overview of Coffee and Coffee Roasting

This chapter provides a comprehensive review of the existing literature and background relevant to the study of coffee roasting and the development of models to simulate and control the roasting process. The objective is to establish a solid foundation of knowledge, identify gaps in the current research, and justify the methodologies adopted in this study. The chapter is divided into several sections, each focusing on different aspects of coffee roasting and system modeling.

The first section, "Coffee Beans", details the anatomy of coffee cherries, highlighting the various layers and components that affect the roasting process. The section discusses the primary species of coffee beans, Arabica and Robusta, and how factors like climate, processing methods, and bean species influence the taste and quality of coffee.

The "The Roasting Process and Roasting Profile" section delves deeper into the stages of roasting, describing how heat transfer induces chemical reactions in the beans, which ultimately shape the flavor and aroma of the coffee. The importance of controlling the roasting process to achieve consistent results is emphasized, along with a discussion on how taste is quantified and evaluated using standards like the Specialty Coffee Association's (SCA) cupping protocols.

The subsequent section, "Roasting Equipment", provides an overview of different coffee roaster designs, with a focus on rotating-drum roasters. It explains the mechanics of various roaster types and their impact on the roasting process. The section also includes a detailed description of the TMR-25 and TMR-250 [7] roasters used in this study, including the key measurements and parameters monitored during roasting.

The "Modelling" section reviews various approaches to modeling coffee roasting processes, categorized into white box, grey box, and black box models. White box models, like those by

Schwartzberg [8], rely on physical laws and heat transfer equations, while black box models are data-driven, such as Adiwijaya's [9] use of LSTM neural networks. Grey box models combine both methods. The review highlights the need for accurate models for controller design and simulation, differentiating system identification from time series forecasting. It discusses specific physics-based models, such as those by Schwartzberg and Vosloo [8, 10], and explores data-based models including FIR, ARX, RNNs, LSTMs, and GRUs, emphasizing recent advancements in machine learning techniques for system identification.

This chapter aims to provide a thorough understanding of the current state of coffee roasting research and modeling, setting the groundwork for the methodologies and experiments conducted in this study.

2.2 Coffee Beans

Coffee beans originate from cherries that grow on coffee trees. Each cherry usually contains two seeds, encased in several thin layers: first, a silverskin, followed by a yellowish layer known as the parchment. Between the parchment and the fruit flesh lies a viscous layer. The outer skin, which starts green and turns red as it ripens, covers the entire fruit. The anatomy of the coffee cherry is illustrated in Figure 2.1.

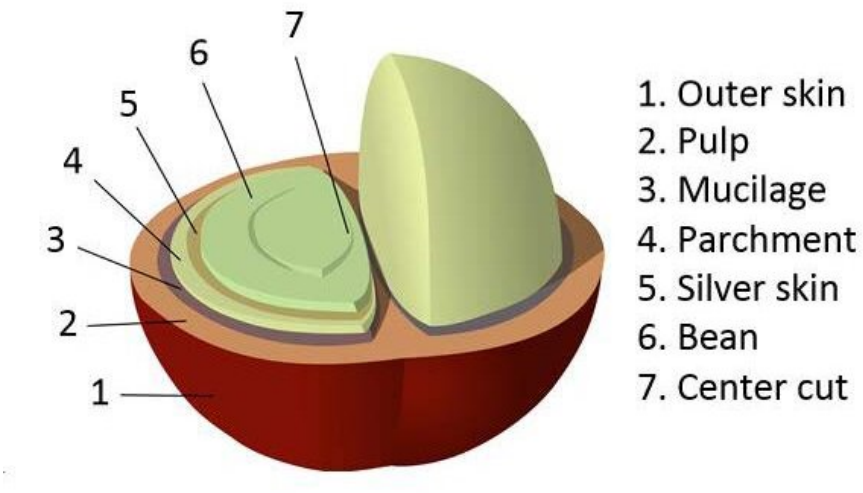


Figure 2.1: Anatomy of a coffee cherry. [11]

The composition of the beans is highly sensitive to various factors such as climate, processing method, bean species, and location. These differences effect the nature and ultimately the taste of the coffee bean. [12]. There are two primary species of coffee beans: Arbica and Robusta. However, there are several subspecies that are derived through selective breeding

to produce different varieties of traits. In general, there are a few species that are commercially important due to traits like disease resistance and yield [13]. Generally, Arabica Coffee is considered sweeter, with higher acidity and bright tastes, and is found typically at higher altitudes. Robusta is typically considered more bitter and less complex flavors. It is grown at lower altitudes and more resistance to temperature. It is used often for blending.

Like mentioned above, several factors influence the taste and quality of coffee. The harvesting process plays a crucial role; some producers hand-pick cherries to ensure only ripe ones are harvested, while others use mechanical methods like strip picking, where entire branches are stripped when most cherries are ready [12]. Once harvested, green coffee processing involves separating the fruit from the cherry and drying the beans to prevent rotting during storage or transportation [1]. There are three major methods for this: washed coffee processing, dry natural coffee processing, and pulped natural coffee processing. In the washed process, the outer fruit layer is separated by a de-pulper, and the mucilage layer is removed by fermentation and washed off with water [14]. In contrast, the dry natural method leaves the beans with the fruit to dry, potentially leading to a fermented taste [1, 2]. The pulped natural process also separates the outer fruit layer but retains the silverskin, parchment, and mucilage layer during drying, which increases the sweetness of the bean due to the extra sugar in the mucilage layer [14]. These processing methods result in variations in the taste and quality of the coffee, with the natural process being less controlled and cheaper, the pulped method producing sweeter coffee, and the washed method being the most controlled and sophisticated but more expensive to implement [14].

The roasting process significantly influences the final coffee quality. Heating green coffee beans triggers various chemical reactions, including hydrolysis, polymerization, reduction, oxidation, and decarboxylation, transforming them into coffee that can be ground, brewed, and consumed. The speed, timing, and location of these reactions greatly affect the flavor, aroma, and color of the final product [15, 16]. Achieving consistent and reproducible coffee roasts, and thus consistent flavor, aroma, and color, is a major challenge for designing roaster control systems.

2.3 The Roasting Process and Roasting Profile

Although many aspects of coffee flavor stem from the nature of the bean (species or subspecies) and environmental factors such as climate and soil content, as discussed in Section 2.2, several elements can be controlled, such as the harvesting and processing steps, which also influence the final taste. The roasting of the beans explicitly controls the development of coffee taste through heat transfer, inducing chemical reactions in the beans that

ultimately determine the final taste and aroma of the brewed coffee. Schwartzberg explains that various constituents in coffee, like sucrose, reducing sugars, proteins, chlorogenic acid, trigonelline, and amino acids, react extensively and nearly disappear during roasting. These reactions form the basis of coffee flavor [8]. He further illustrates that the timing and rate of these reactions are critical for taste development. Multiple reactions occur at different temperatures and rates, creating complex reaction networks. The temperature at various stages influences the availability of reactants and products, thereby affecting the final taste. Reaction rates depend on temperature and reactant concentrations, with later reactions depending on the rate and duration of earlier ones. Schwartzberg believes this sensitivity to temperature history is why reproducible coffee flavor is challenging to achieve [8].

In the following subsections, this paper will briefly review how taste is affected and evaluated by the roasting process, detailing the stages of coffee roasting and the prominent methods and equipment used for coffee roasting.

2.3.1 Stages of the Roasting Process

According to Vosloo, there are 5 general stages that should occur in the roasting process: drying, yellowing, first crack, roasting, and second crack. These stages are illustrated with an example of a temperature trend that is typically seen by a coffee roaster in Figure 2.2.

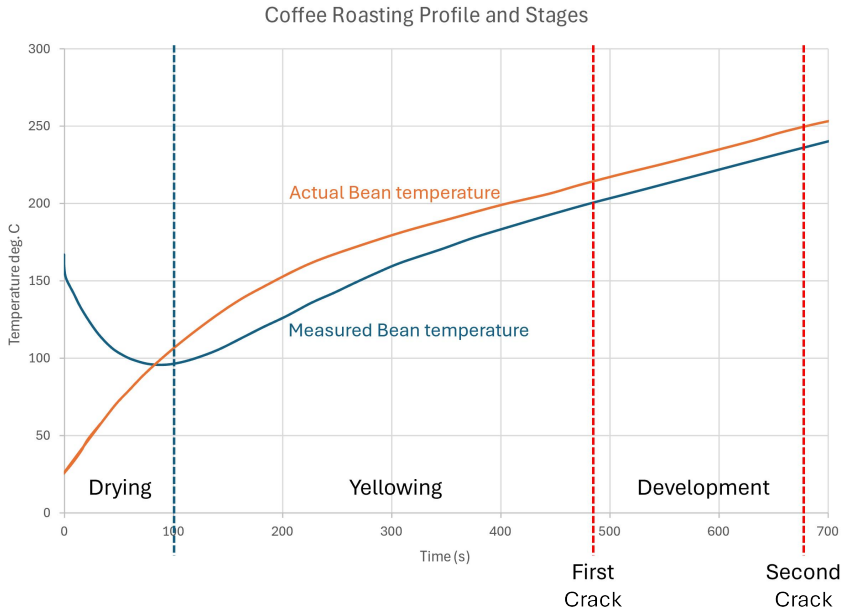


Figure 2.2: Roasting profile example with stages approximately marked. Adapted from [8]

It is important to note, that the actual bean temperature cannot be measured directly in most situations, and so a probe is placed in the drum where a majority of bean mass will

be. This is used as a proxy to the bean temperature. When the beans are initially placed in the roaster, they are typically at ambient temperature. The roaster typically is preheated, and so the probe starts at high temperature, but then falls rapidly as the cooler beans are introduced. At some point the probe temperature starts to increase instead of decrease, which is known as the "turning point" [2, 17]. The bigger the thermal mass of the sensor the longer this can take, and the larger the gap between the actual bean temperature and the measured sensor. The measured temperature is known as the "recipe" or the "profile" [17]. If a profile is found to produce good flavor, then a roaster professional will try to replicate the curve to consistently create that flavour. However, because different probes will have different dynamics and therefore profiles, it is difficult to transfer profiles from machine to machine, unless the differences in the probes, and other elements, have carefully been considered or calibrated [17].

Green coffee beans have moisture contents of around 8 - 12.5 percent, as per standards governing the storage and sale of green coffee [1, 18]. In the Drying Phase, the majority of the heat in the process is used converting this moisture to steam, and drying the coffee bean. In this period, there is very little taste development Next is the Yellowing Phase. Here the coffee begins to change color and more reactions occur, sugars begin to break down into acids as moisture level continues to decrease. This stage is characterized as having a "bread" like aroma [17]. chaff is produced here, as the beans begin to swell, and the silverskins break and come off. The chaffs pose a risk of catching fire, and so it is important that enough air flow is present to carry the chaff away. The chaff is collected typically by a cyclone [14]. The parts of a roaster will be discussed in further detail in section 2.3.3.

Bitter or sour coffee can result if the yellowing phase and drying phase are not completed properly. Increasing the bean temperature too high, before a sufficient amount of evaporation has taken place causes the outside of the bean to roast, but the inside to remain under-roasted [2]. This is relevant to controls as, once this occurs, the roast can not be fixed, by slowing the development. An inaccurate system dynamics model, or improper controller could overshoot temperatures if not properly designed, which would ruin the entire batch of coffee. This is part of the thermal inertia of the system and contributes to some unique control challenges.

The First crack is a milestone that indicates the start of the development phase. It is called this from the characteristic cracking sound that occurs as pressure from the products of the reactions, like carbon dioxide, in the beans build up and the bean cracks and lets the pressure out [19]. Several studies have been done, and show that first crack typically occurs around 200°C [3, 14, 19].

2.3.2 Taste

Developing and consistently reproducing coffee taste is one of the primary goals of coffee roasting. Thus, it is important to understand how coffee taste is quantified. According to a seasoned coffee roaster (personal interview, May 29, 2024), the Specialty Coffee Association (SCA) and their "Cupping Protocols" are among the most authoritative sources for quantifying coffee taste [17]. The SCA's cupping standards are designed to accurately assess coffee quality through rigorous procedures and qualified 'Q graders.' These graders evaluate several key flavor attributes, including Fragrance/Aroma, Flavor, Aftertaste, Acidity, Body, Balance, Uniformity, Clean Cup, Sweetness, and Defects. Each attribute is graded on a 16-point scale, ranging from 6 to 10 [20]. However, roasters often simplify these categories into three primary characteristics: bitterness, acidity, and "origin characteristics," which are considered to change during the roast, as illustrated in Figure 2.3 [17].

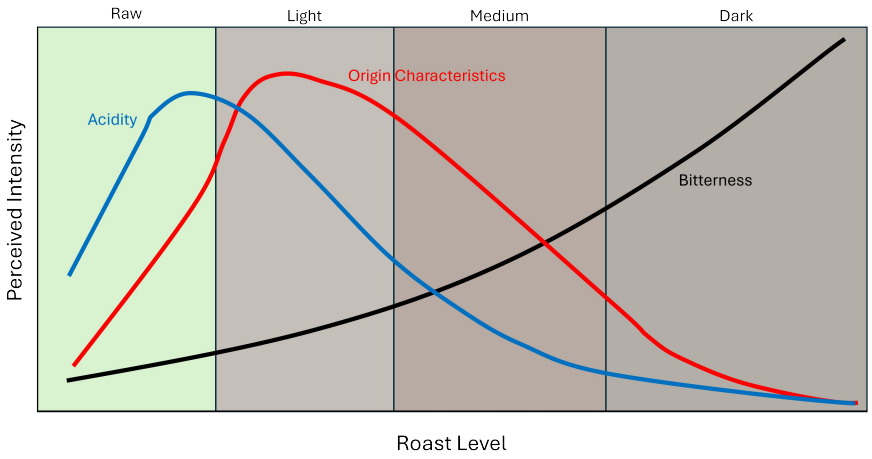


Figure 2.3: General trends of taste characteristics through the coffee roasting process.

The longer coffee is roasted, the more bitter it becomes. Acidity increases initially but decreases with prolonged heat exposure. Origin characteristics encompass all other flavor aspects and result from the chemical reactions within the bean. Green coffee initially has little taste, requiring heat to initiate the chemical reactions that develop interesting flavors. However, the longer these reactions continue, the more the chemistry of the beans converges, as reaction pathways near completion. This is why darker roasts tend to have more similar flavors, and why specialty coffees are typically lighter roasts. For specialty roasters, precision in controlling temperature and heat over time is particularly important [2, 14, 17].

Schwartzberg, and Clarke, clarified the relation between the roast color, reactions and taste. He explained that roast color can be an indicator of overall reaction extent, however, there are many ways for the reactions to advance to get to the same color. He further

explained that the color itself cannot tell us what pathways were favoured, and therefore, cannot tell us about the exact taste of the roast [8, 21]. The reaction pathways followed during the roast determine the taste and is sensitive to the temperature profile followed in the roasting process as previously discussed.

The roaster interviewed emphasized that there are other aspects of controlling the roasting process that are critical to taste. For example, right after the coffee beans get to the "first crack" point (described in Section 2.3.2), if the heat is not controlled properly, a temperature "flick" or "crash" can occur. Both of which are undesirable and lead to what roasters refer to as "baked" tasting coffee [17]. This happens because after the first crack the beans undergo an exothermic reaction, so if heat is not controlled or compensated correctly, there can be a sharp rise in temperature, know as a "flick". When the flick is overcompensated for and the heat is lowered too much, and or too much ambient air is let in, the temperatures can drop drastically. Because this point is nearer to the end of the roast, the amount of time to get the temperatures back up might be too long and the required reactions for the desired coffee taste will be missed or have occurred at the wrong rates. [17].

In total, roasting is a vital process in developing the unique taste of different coffee flavours, and key in the development is controlling the heat transfer to the beans throughout time. Because of this, applications of controls, such as Proportional-Integral-Derivative (PID) systems, have been common for helping to get consistent desired tasting coffee, however, implementation is difficult due to nonlinear dynamics involved, which will be discussed further in later sections. Here, machine learning techniques and learned control systems hold great promise for improving the methods used to control the temperature of the coffee beans and therefore the taste of the coffee.

2.3.3 Roasting Equipment

This section provides an overview of the different technologies used for roasting coffee beans, particularly in commercial applications. Schwartzberg and Clarke have identified several of the most common roaster types used in the industry [8, 14]. These include Rotating-Drum Roasters, Spouted-Bed Roasters, Rotating-Bowl Roasters, Scoop-Wheel Roasters, and Swirling-Bed Roasters. Among these, Rotating-Drum Roasters, especially the "classic" design with directly heated drums, are considered the most prevalent [14].

In this section, each of these roaster types will be summarized and reviewed. Despite the differences in design, commercial roasters generally require several common systems: a heat source, a roasting chamber, and an airflow system. Additionally, there are several auxiliary systems often found in roasters, such as the bean loading system, cyclone, cooling tray, and

de-stoning system [14]. These auxiliary systems, although not specific to any particular roasting design, play crucial roles in the overall roasting process.

The heat source for most roasters is either electric or natural gas, but other new energy sources, such as hydrogen, are being researched [22]. While the heat source is primarily responsible for adding energy to the system, several other factors influence the rate of heat transfer to the coffee beans, as discussed in Section 3. The roasting chamber is where this heat transfer occurs, with the airflow system serving multiple purposes: removing the silver-skin and chaff that come off the beans during roasting (which are fire hazards) and influencing heat transfer to the beans. Some roasters control temperature profiles by adjusting airflow, while others maintain a consistent airflow range and control heat transfer through the heat source.

Industrial and commercial roasters often feature an air recycle system, which reuses hot air to heat the beans in the roasting chamber, improving energy efficiency and helping maintain high temperatures. Roasters that reuse air are known as "R roasters," while those that use air only once are referred to as "single-pass roasters" or "SP roasters" [23]. Another important component of the air system is the cyclone, which uses centrifugal force to remove silver-skin and chaff during roasting.

The bean loading system handles the transfer of beans into the roasting chamber. The cooling tray cools the beans after roasting to prevent further "cooking" and ensure consistent flavor. The de-stoning system removes any non-bean debris, such as small rocks, which are common contaminants during harvesting and transportation [14]. While these auxiliary processes are outside the scope of the modeling in this paper, they are important to mention. An overview of each roaster design will be provided below.

Rotating-drum roasters

These are characterized by horizontal rotating drums roasting chambers. The drums are often flighted to help with the mixing and heat transfer of the beans. For consistent roasts the beans should be roasted uniformly in the roaster, and the rotation of the drum helps ensure that each bean roughly gets heated similarly by mixing the beans around in the drum [8]. Since the air exiting the drum is cooler than the air entering, and heat transfer between the metal of the drum and the air is also different, the flights are designed to move the beans so that each beans roughly experiences heat transfer the same [8]. In modern drum roasters, the speed of the drum is often a variable that can be used to control the heat transfer to the beans, but in practice this is often set to a constant speed and other parameters such as the heat source is primarily used to adjust the bean temperatures and to meet the desired bean temperature profile [8]. This is especially the case in small to medium scale roasting

facilities, where roasting professionals are the primary source for operating and adjusting temperatures. Cristo [24], showed that the motion of the coffee beans vary depending on the amount of the drum filled, and the speed of the drum, and can be divided into several different regimes, shown in Figure 2.4 - Slumping, rolling, Cascading, Cataracting, and Centrifuging - which has an impact on the effective heat transfer experienced by the beans.

(a) Rolling (b) Cataracting (c) Centrifuging

Figure 2.4: Different Regimes of motion in a flighted rotating drum. The type of motion has an effect on the heat transfer experienced by the coffee beans in the roasting chamber.Figure adapted from [24]

Within the class of rotating-drum roasters, several differences in design influence how the beans are roasted. The oldest design is the direct heated drum roaster, where the drum is directly heated by the heat source, most commonly a natural gas flame. In this method, the metal of the drum heats both the roasting air and the beans [8]. This approach is more challenging to control due to the thermal mass of the drum, which creates a significant lag between heating inputs. Additionally, if the drum overheats, it is difficult to cool down.

Next are hot air roasters, which use a heat exchanger to heat the air that is then sent to the drum, where it roasts the coffee beans [14]. These roasters are more expensive but easier to control, as the temperature of the air, the primary heat source, can be directly and easily regulated. Some roasters use drums with perforated walls, and some do not. The major difference being in the perforated case, the air moves into the roaster drum axially, but leaves radially, whereas in the non-perforated case the air enters axially, and leaves axially [8].

Drum roasters typically operate with relatively high inlet gas temperatures, upwards to 550°C, and roasts typically last from 8 to 12 minutes, although this can depend on the size of the roaster [8].

The datasets for this study were collected from roasts performed on TMR-25 and TMR-250 roasters manufactured by IMA, which are categorized as "R"-roasters. A specific example is the TMR25 from IMA Petroncini, a solid-rotating drum R-roaster. Figure 2.5 illustrates

this system, and Table 2.1 details the key measurements shown in the figure, which are typical in commercial roasting settings.

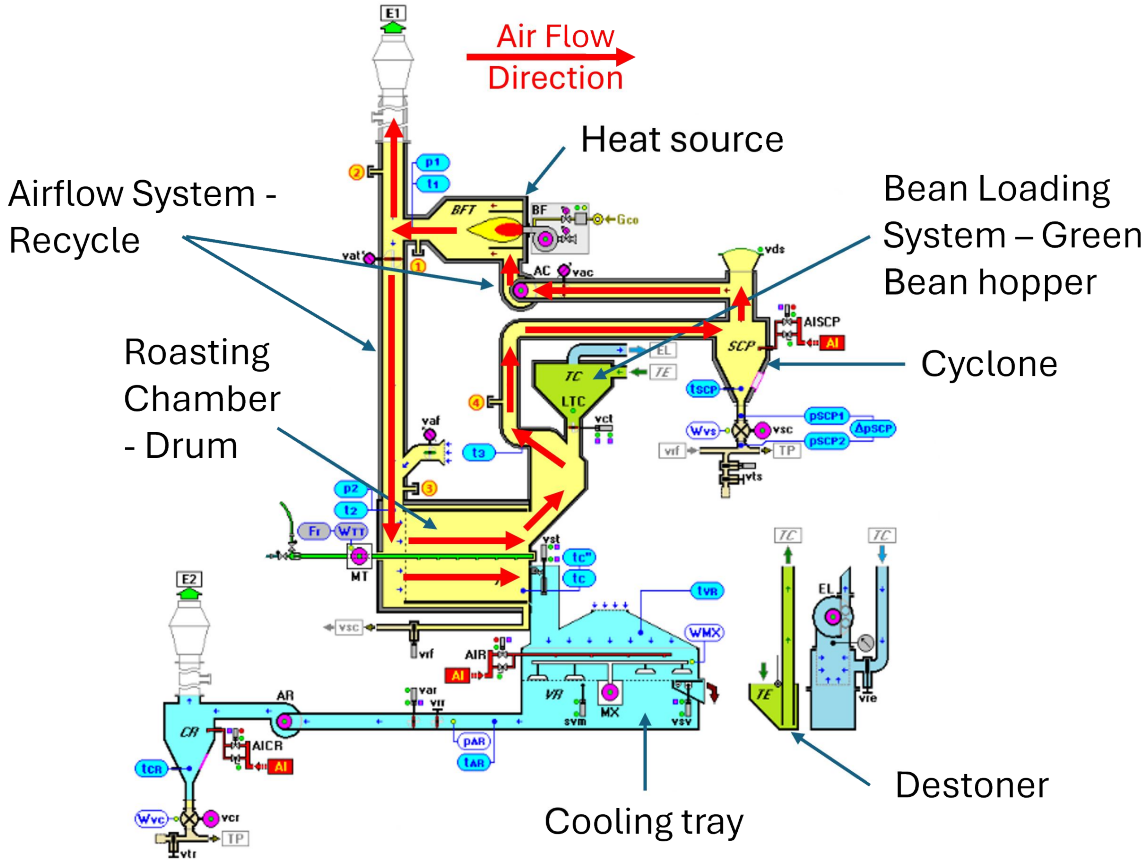


Figure 2.5: Representation of a hot air drum roaster with air recycle (R-roaster). Adapted from [7]

Spouted-Bed Roasters

Spouted-bed roasters and fluidized-bed roasters are terms that are sometimes used interchangeably. Although similar in that they both utilize high-velocity air to lift the beans, they are quite different. In fluidized-bed roasters, the beans are kept suspended in the air, creating a bed of constantly moving and circulating beans. However, most roasters referred to as fluidized-bed roasters are actually spouted-bed roasters. The difference is that while fluidized-bed roasters keep the beans fully suspended, spouted-bed roasters use air to lift a "spout" of beans into the air, which then fall back down around the edges, continuously mixing. Both methods primarily rely on convection for heating the beans.

True fluidized-bed roasters are uncommon in industrial or commercial settings [8]. Spouted-

Table 2.1: Typical Measured Roaster Parameters

Parameter	Description
T_1	The temperature of the air coming out of the burner.
T_2	The temperature of the air coming into the drum roasting chamber. This temperature is controlled by the burner fuel used, the amount of air recirculated, and the amount of fresh air drawn into the system
T_3	The temperature of the roasting air leaving the drum.
VAC	This is the valve that, along with VAT, controls the air flow rate through the burner. The valve position is typically measured as a percentage open or closed.
VAT	This valve mostly controls how much air is recirculated and how much is put through the exhaust. The valve position is typically measured as a percentage open or closed.
VAF	This valve controls how much ambient air is drawn into and mixed with the recirculated air. The valve position is typically measured as a percentage open or closed.
BF	This is a small control system that controls how much air and fuel the burner flame gets. It is typically measured as a percentage of maximum flame temperature.

bed roasters, however, are often used by hobbyists or for small batch roasting due to their smaller equipment footprint compared to drum roasters. These roasters allow for higher rates of heat transfer due to larger airflow rates and volumes of air. They are also mechanically simple and typically cheaper to produce. Spouted-bed roasters usually operate at lower inlet temperatures, typically less than 360°C, with roasting times ranging from 10 to 20 minutes [8].

Rotating-Bowl Roasters

In Rotating-bowl roasters, beans move across the surface of a heavy cast-iron bowl by centrifugal force. the beans move up the wall of the bow towards the rim and strike a stationary cover where they fall back inwards to the bowl. hot gas moves into the bowl through a central duct in the top cover and moves through the bowl into an annular duct [8]. Inlet gas temperatures up to 550°C are typically used, however, roasting times are typically quicker, being around 3 to 6 minutes.

Scoop-Wheel Roasters

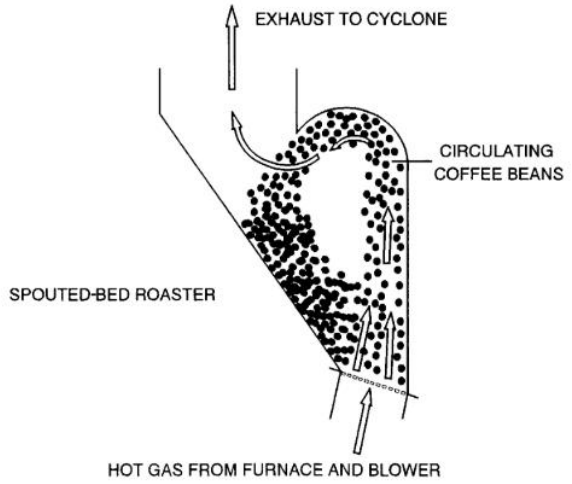
This design utilizes scoops mounted on a wheel in the roasting chamber, that rotates through the beans that collect in a trough at the base. The scoops move through the bed of beans, mixing them and throwing some beans upwards in the head space. Hot air moves tangentially to the scoops and moves through slits of perforated metal near the bottom of the trough. Inlet air temperatures of up to 420°C are typically used, and roasting times as short as 1.5 to 6 minutes are typical [8].

Swirling-Bed Roasters

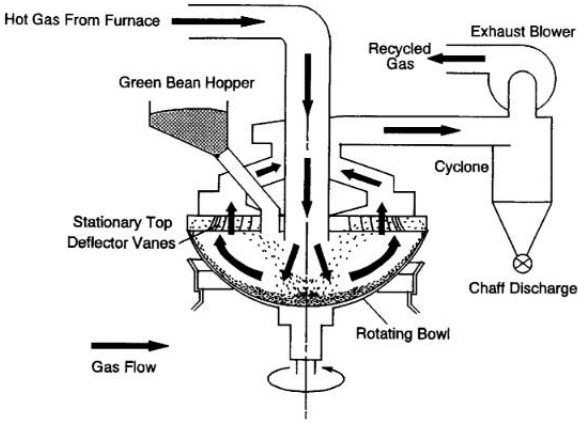
Swirling-bed roasters are similar to spouted-bed or fluidized-roasters in that it utilizes high air speeds to keep the beans in constant motion, increasing the effective surface area and rate of heat transfer. This results in high-efficiency roasts that utilize lower inlet gas temperatures, typically lower that 280°C and short roast times, 1.5 to 3.0 minutes. [8]. The roasting chamber is designed so that hot air swirls around and pushes beans against the slightly outward tapered walls, so that the move out and up, and results in a layer of beans to spiral around the chamber. Beans get pushed up inwards as the get to the top of the wall and fall back down onto an up-wards pointing cone, where the process repeats. Increasing the flow of the air increases the heat transfer, but also the drag and centrifugal forces proportionally [8].

Although there are many types and designs of roasters, as illustrated in Figure 2.6, this study focuses on the rotating-drum roaster. Specifically, rotating solid-drum R-roaster like illustrated in figure 2.5. This is because drum roasters are arguably the most common type of roaster used commercially [14], and because operational data for this type of roaster was available for this study. This data was used to examine the effectiveness of several physical based models, and develop a reflective physics based system for developing synthetic data that could be used to further explore machine-learning techniques for system identification and non-linear control.

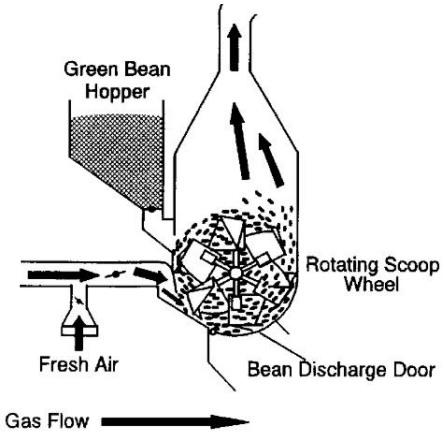
Although there are many types and designs of roasters, as illustrated in Figure 2.6, this study focuses on the rotating-drum roaster, specifically the rotating solid-drum R-roaster, as illustrated in Figure 2.5. Drum roasters are arguably the most common type of roaster used commercially [14], and operational data for this type of roaster was available for this study. This data was used to assess the effectiveness of several physics-based models and develop a grey-box model in Chapter 3.



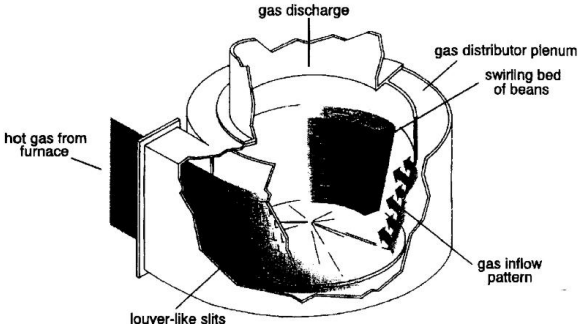
(a) Spouted-bed Roaster



(b) Rotating-Bowl Roaster



(c) Scoop-Wheel Roaster



(d) Swirling-Bed Roaster

Figure 2.6: Alternative Roaster Designs. Images taken from [8].

2.4 Modelling

In the previous sections, the background of coffee and the roasting process was summarized, highlighting the complexity and significance of this process in determining the taste and quality of the final product. Coffee roasting is a highly intricate procedure involving various chemical reactions and heat transfer mechanisms, which significantly impact the flavor, aroma, and overall quality of the coffee. Given this complexity, developing accurate models to understand and control the roasting process is of immense value to the coffee industry. These models enable the design and optimization of control systems, ensuring consistency and quality in coffee production.

Several models have been proposed for coffee roasting, ranging from those specific to certain sections of the roasting process and equipment to more general approaches. These models can generally be classified into three categories: white box, grey box, and black box models. White box models typically rely on physical, first principle relations to describe the system. Schwartzberg's model [8] is one of the most referenced in this category, providing a comprehensive framework for understanding the heat transfer and reaction kinetics in coffee roasting. Black box models, on the other hand, are characterized by their data-driven nature. They utilize machine learning and statistical techniques to model the system based on empirical data, making them particularly useful when the underlying physical processes are too complex to model accurately with first principles alone. Recent advancements in machine learning, especially neural networks, have opened new avenues for developing highly accurate data-driven models for coffee roasting. Grey box models combine elements of both methodologies, using known physical laws, empirical relations, and data techniques such as regression or machine learning to create robust and flexible models capable of capturing the intricacies of coffee roasting.

It is important to note that the intent of developing a model for this study is to design and simulate a controller. Therefore, the model must accurately predict how the system states develop with different inputs or control algorithms. This requirement highlights a key distinction between "time series forecasting" and "system identification." Forecasting aims to estimate future states without necessarily perturbing the system, whereas system identification seeks to understand the system dynamics to facilitate control and manipulation.

This section reviews various approaches to modeling the coffee roasting process. It begins with a review of physics-based models in literature. Next, it examines attempts at data-based modeling specifically for coffee roasters. Finally, it provides an overview of neural network techniques recently developed and used specifically for system identification.

2.4.1 Physics-based Roaster Models

The most referenced physics based model for coffee roasting comes from Schwartzberg [8], which encompasses the heat transfer between beans, air, metal, and accounts for moisture content of the beans as well as exothermic reactions that occur inside the beans. His model is not specific to any one roaster, although, in his paper he mentions how to adapt the equations and assumptions that can be made for different roaster types, such as rotating drum roasters, or spouted-bed roasters. The model primarily relies on heat balance equations between the roasting air, the bean mass, and the metal of the roaster, but it relies heavily on regression techniques and the assumption that several key parameters are constant, such as coefficients related to the rate of reactions inside the bean, and critically, the effective heat transfer coefficient. This assumption may be accurate when mass of beans in the drum, the movement of the beans, and the mass air flow rate are held constant, but a model on this assumption is shown in Section 3 to have significant errors when these variables vary. This is supported by Critso et al. [24] and Clarke [14], where observations that the effective heat transfer inside the drum of a roaster was affected by the motion of the bean, which in term was related to the amount the drum was filled, and the speed of the drum rotation.

Others have taken the Schwartzberg model, and modified it to try to account for the effective heat transfer, by modelling the flow rate around a coffee bean, such as Vosloo [10]. This method might capture the effect of that different air flow rate will have on the effective heat transfer but does not take into account the motion of the beans in the drum. Typically in roasters, the average air velocity is relatively low in relation to the motion on of beans, so the relative effect of the beans moving through the air will have a significant effect on the heat transfer rate [24]. Puranto and Chen [15] developed a model utilizing heat transfer equations similar to Schwartzberg, however, they proposed a moisture loss model which uses the lumped reaction engineering approach (L-REA) to estimate the moisture loss of the coffee beans. This process still relies on experimentally determining an "activation energy" parameter which is determined experimentally.

Other studies, such as Fabbri et al. [25] focus on modelling the the bean temperature by focusing on the specific geometry of a coffee bean, and utilized Fick's law to model moisture loss, and empirical relations to model water diffusivity within the coffee bean.

Vosloo [10] reviews several of these models and tries to validate their performance experimentally with a direct-flame heated drum roaster. Critically however, the inlet temperatures, drum speeds, and roasting air flow rates were held constant in experimentation. This allows for several critical assumptions, such as a constant effective heat transfer rate, required for many of the models above. In real roasting scenarios, the inlet gas temperature varies sig-

nificantly as part of the bean temperature control, as well as varying drum speeds, and air flow rates. Interestingly, between many of the models looked at, Schwartzberg based models performed, with the described caveats, very well with R2 values of around 0.98 for constant inlet temperatures [8]. For this reason, this is the model that this study utilizes for it's exploration into augmenting models to account for varying parameters, or for learning the system dynamics. Details of the models used can be found in the Appendix B.

2.4.2 Data-based Roaster Models

There have been some studies discovered that try black-box approaches for modelling coffee roasters. A relevant study utilizing Long Short-Term Memory (LSTM) based neural networks was conducted by Adiwijaya [9]. Adiwijaya's study focuses on fitting a "one-step-ahead" LSTM-based neural network to several roasts with different configurations to determine how well models generalize to different conditions. This study builds off some of this work done and explores "one-step-ahead" approaches performance when used in an auto-regressive fashion to simulate the system dynamics. This is done because the motivation for developing a model in this study is to facilitate the design and simulation of a controller. In order to do this, a model of the "plant" dynamics are important. In Adiwijaya's study, predictions were made only with actual measured values as input, and not auto-regressively with previous predictions as inputs to the LSTM.

Other studies look at black-box approaches for identifying system dynamics, but are not specific to coffee roasting. Cheng demonstrates a unique method for learning both the system dynamics and optimal controls of linear systems simultaneously. Cheng proposes a Neural Ordinary Differential Equation (ODE) based method which combines the the system identification and optimal control learning using a coupled neural structure. Cheng verifies the approach virtual simulations of deterministic systems, such as the classic mechanical "cart-pole" or "inverted-pole" problem [26]. Similarily, Bachhuber et al. also develop an approach for utilizing nerual ODEs for automatic system identification and test it on several virtual systems. Ramirez-Chavarria et al. presents a regularization scheme for identifying nonlinear finite response models (NFIR) using neural networks, and Hendriks et al. use neural networks to develop Deep Energy-Base models that can be used for system identification. There are several other studies focused at the indersection of machine learning, controls and system identification. Section 2.4.3, covers some of the common data-based system identification utilizing machine learning.

2.4.3 Data-based models and System Identification

System identification is a crucial discipline in control theory, focused on building mathematical models of dynamic systems from measured data. Traditionally, the classical approach to system identification relies on techniques from mathematical statistics, primarily using prediction error methods (PEM) and discrete model orders [27]. This method involves selecting model structures like FIR, ARX, NFIR, and NARX models, which depend on unknown parameter vectors [27]. The complexity of these models is managed by adjusting the dimension of these vectors, striking a balance between bias and variance. Complexity measures such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), as well as cross-validation methods, are commonly employed to select the most suitable model structure [28].

In classical approaches, the process follows a loop, where data is collected, a family of models are chosen, and finally the best model in this family is determined, by using model validation tests [27]. However, it is not guaranteed that any of the models explored produce satisfactory results. This can happen because of several reasons, such as the the model set or family was not appropriate and doesn't contain a satisfactory description of the system, or the data was not informative enough to provide guidance in selecting an appropriate model [27].

According to Pillonetto, there has been significant interest and research in using machine learning, in particular deep neural networks, as a tool for system identification, offering even more complex model spaces [28]. These networks, composed of multiple layers of linear transformations and nonlinearities, can approximate a wide range of dynamic systems. Advances in deep learning, including improved training algorithms, larger datasets, and enhanced computational resources, have propelled their application in system identification. Hierarchical models such as deep NFIR and NARX, temporal convolutional networks, and deep state-space models illustrate the versatility and power of deep learning in capturing complex system dynamics [28].

The integration of deep learning with system identification promises to bridge the gap between these fields, offering new methodologies and insights. Techniques like energy-based models [29], concatenated representer theorems [30], and Koopman operator theory combined with deep autoencoders [28, 31] are at the forefront of this interdisciplinary research. By leveraging the strengths of deep learning, system identification can achieve more accurate and robust models, paving the way for innovative applications across various domains. Here we will go into some detail about a few of the more common neural network architectures

used for system identification [28].

FIR and ARX Neural Networks

Finite Impulse Response (FIR) neural networks are model architectures used for system identification where the system output depends on a finite sequence of past inputs [32]. This concept originates from FIR filters in signal processing, which respond to an impulse input over a limited duration.

In FIR neural networks, the input layer captures both the current and a fixed number of previous input values. This structure allows the network to inherently manage temporal dependencies within the data. The hidden layers then process this sequence through typical neural network operations, including weighted summations followed by activation functions, to extract relevant features and patterns [32].

The output layer of an FIR neural network predicts the current system output based on the processed input vector, effectively modeling the relationship between the inputs and the output over the specified window of past inputs [32]. This makes FIR neural networks particularly suited for time-series prediction and dynamic system modeling tasks.

For example, an FIR neural network designed to predict the output $y(t)$ based on the past N inputs $x(t), x(t-1), \ldots, x(t-N+1)$ would operate as follows:

- The input layer takes the current input $x(t)$ along with the past $N-1$ inputs $x(t-1), \ldots, x(t-N+1)$.
- The hidden layers process this sequence, extracting temporal features and dependencies.
- The output layer provides the predicted output $\hat{y}(t)$.

The mathematical representation of an FIR neural network is shown in Equation 2.1

$$\hat{y}(t) = f(x(t), x(t-1), \ldots, x(t-N+1); \theta) \tag{2.1}$$

where f denotes the function implemented by the neural network, and θ represents the network's parameters, including weights and biases.

FIR neural networks have been recognized for their capability to handle dynamic systems with finite memory effectively and their structure allows them to capture the essence of time-dependent data [32]. Finite Impulse Response (FIR) neural networks, while advantageous in many applications, come with several notable disadvantages, such as requiring a large number or parameters due to the large inputs and increases training complexity. They are

also sensitive to the input window size as well as to noise making them difficult to train and tune.

ARX (AutoRegressive with eXogenous inputs) neural networks are another popular approach in system identification, extending the traditional ARX model by incorporating neural network components, thus enabling the modeling of nonlinear relationships [28]. These are similar to FIR architectures, but differ primarily in ARX architectures include exogenous inputs, such as past measured outputs. This is represented in Equation 2.2

$$y(t) = f(x(t), x(t-1), \ldots, x(t-N+1), y(t-1), y(t-2), \ldots, y(t-M); \theta) \tag{2.2}$$

Here M is the number of past outputs considered.

One important consideration with ARX models is that they require past outputs to make future predictions. This dependency poses a challenge for long-term predictions since future outputs are not available until they are predicted. Using these future values prematurely can lead to temporal leaking, which complicates the training process [28]. Consequently, ARX models are typically constrained to "one-step-ahead" prediction techniques, where the model predicts the output of the dynamic system one time-step into the future. These one-step-ahead predictions are then used in an autoregressive manner to model the system's dynamics.

RNNs, LSTM, and GRU Networks

Recurrent Neural Networks (RNNs) are a class of neural network architectures designed to handle sequential data by modeling dependencies from past inputs and states. RNNs maintain a hidden state that is updated at each time step, allowing them to capture temporal dependencies. The general structure of an RNN includes a state-propagation equation and an output equation. For instance, the Elman RNN updates its hidden state and computes the output as follows [28]:

$$h(t+1) = \sigma(W_{hh}h(t) + W_{hx}x(t) + b_h), \tag{2.3}$$

$$\hat{y}(t) = W_{yh}h(t) + W_{yx}x(t) + b_y, \tag{2.4}$$

Here $h(t)$ is the hidden state, $x(t)$ is the input, $\hat{y}(t)$ is the predicted output, and σ is a nonlinear activation function. The parameters $W_{hh}, W_{hx}, W_{yh}, W_{yx}, b_h$, and b_y are learned during training.

One common issue with RNNs is the exploding and vanishing gradient problem, which arises during backpropagation through time. This problem can make training difficult and

slow convergence. Solutions include gradient clipping, non-saturating activation functions, and orthogonal RNNs.

To address the limitations of standard RNNs, particularly the vanishing gradient problem, gated mechanisms such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) have been developed.

Long Short-Term Memory (LSTM) models introduce a set of gates (input, output, and forget gates) that regulate the flow of information, allowing the network to maintain long-term dependencies. The equations for an LSTM are as shown in Equations 2.5 through 2.10.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{2.5}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{2.6}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{2.7}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \tag{2.8}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \tag{2.9}$$

$$h_t = o_t * \tanh(C_t), \tag{2.10}$$

Here f_t is the forget gate, i_t is the input gate, o_t is the output gate, C_t is the cell state, and h_t is the hidden state.

Gated Recurrent Units (GRUs) simplify the LSTM architecture by combining the forget and input gates into a single update gate and merging the cell state and hidden state. The GRU equations shown in Equations 2.11 to Equations 2.14.

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z), \tag{2.11}$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r), \tag{2.12}$$

$$\tilde{h}_t = \tanh(W_h \cdot [r_t * h_{t-1}, x_t] + b_h), \tag{2.13}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t, \tag{2.14}$$

Here z_t is the update gate, r_t is the reset gate, and h_t is the hidden state.

The benefit of being able to learn long range dependencies is is obvious when it comes to modeling language, where the context for from words far removed from the present word is important, however, this is not always the case in dynamical systems, where only nearby temporal properties are relevant in understanding the current trajectory [28]. This can be important and beneficial for systems that are cyclic in nature, but one must be careful as it

might find patterns or dependencies in the data that are not really there.

Probabilistic Recurrent Neural Networks (Probabilistic RNNs) extend the deterministic nature of traditional RNNs by incorporating probabilistic elements [33]. Instead of predicting a single output value, Probabilistic RNNs predict a distribution over possible outputs. This approach provides a measure of uncertainty in the predictions and can be particularly useful in applications where it is important to quantify the confidence in the model's predictions. An example of how this might be expressed mathematically is shown below in Figure 2.15 [28].

$$\begin{bmatrix} \mu(t) \\ \log \sigma(t)^2 \end{bmatrix} = W_{yh}h(t) + W_{yx}x(t) + b_y. \tag{2.15}$$

The logarithm ensures that the variance remains non-negative. The model is trained by maximizing the likelihood of the observed data under the predicted distribution. The benefit of using a probabilistic RNN is that it can provide a measure of uncertainty, which is important in some applications, as well as it can be more robust to noise and outliers in the data.

Overall, the advancements in RNN architectures, particularly the introduction of LSTM and GRU, have significantly enhanced the capability of neural networks to handle complex temporal dependencies in system identification.

Chapter 3

Methods and Experiments

3.1 Method Overview

This chapter presents the methodology and experiments conducted to model a coffee roaster using both black-box and grey-box approaches. The focus is on developing a robust model that sufficiently represents the coffee roasting process with a level of accuracy appropriate for use in a control system, leveraging the strengths of neural networks and physics-based equations.

The chapter begins with an exploration of black-box modelling using a Neural ARX (NARX) approach. Various neural network architectures, including NARX and RNN, are discussed, with an emphasis on the NARX model trained on real production data from approximately 4000 coffee roasts. Key variables used in training the model are listed and described. The NARX model's structure, implementation, and training process are thoroughly explained. Despite achieving high R-squared values and low RMSE, the model's performance is critically evaluated against a naive model, which predicts no change in bean probe temperature. The comparison reveals significant limitations in the NARX model, particularly when used for autoregressive predictions, highlighting the need for a more accurate and reliable modelling approach.

To address the limitations of the black-box approach, the chapter transitions to a grey-box modelling strategy that integrates physics-based equations with neural networks. The theoretical foundation for the grey-box approach is laid out, drawing from the work of Schwartzberg [8], Putranto et al. [34], and Vosloo [10]. Schwartzberg's model, comprising a system of coupled differential equations, serves as the basis for the grey-box model. Key parameters and their relations are described, with a focus on the effective heat transfer coefficient (h_e), a critical parameter influencing the model's accuracy.

A detailed analysis of the heat transfer coefficient (h_e) is conducted. Initially, Vosloo's

method [10] for determining h_e is replicated and evaluated using real production data. The regression technique used to optimize h_e and other parameters is explained, revealing the challenges and limitations of assuming a constant h_e. The impact of drum speed and air velocity on h_e is discussed, highlighting the importance of capturing these dynamics accurately.

The chapter then explores the integration of neural networks into the differential equations to model h_e as a function of drum rotation speed, air velocity, and other variables. A neural network is trained within the system of differential equations to minimize the prediction error of the bean probe temperature. The training algorithm, incorporating backpropagation and optimization techniques, is detailed. This hybrid approach aims to leverage the flexibility and learning capabilities of neural networks while maintaining the interpretability and reliability of physics-based models.

The chapter concludes by summarizing the methods and experiments conducted to develop a robust coffee roaster model. The limitations of the black-box approach are addressed through the proposed grey-box model, which combines empirical data with first-principles equations. This hybrid approach is anticipated to provide a more accurate and reliable representation of the coffee roasting process, suitable for controller design and simulation.

Overall, this chapter lays the groundwork for the results and analysis presented in the following chapters, demonstrating the importance of combining neural networks with physics-based models to overcome the challenges of accurately modelling complex systems like coffee roasting.

3.2 Roaster Black-box Modelling

The initial method explored in this study was a "one-step-ahead" prediction approach using neural networks to model the coffee roaster as a black-box system. As discussed by Pillonetto, in his survey on machine learning techniques for system identification, various architectures can be employed for this purpose, with NARX and RNN architectures being common [28]. In this study, a NARX model was trained to evaluate the effectiveness of this approach for modeling a coffee roaster. Although this section focuses on detailing the methodology, it also highlights the challenges these models face when used for controller simulation and design, particularly when the variable to be controlled changes slowly compared to the rate at which the controller inputs change.

In this section, the NARX model developed for this study is first described. Next, a "naive" model is introduced, which assumes no change between the current bean probe temperature and the next time step. Finally, the implications of this comparison are discussed,

motivating the exploration of a grey-box model to address the challenges encountered with the black-box approach.

The theory behind Neural ARX architectures is covered in Chapter 2. This section provides details on the implementation of the this technique on the application of coffee roasting. The data comprised approximately 4000 production roasts collected from a real commercial coffee production. Details about this dataset are found in Appendix C. Table 3.1 lists the variables used in training the neural network, including both inputs to the model and the target outputs it aims to predict.

Variable	**Description**
Set % command BF [%]	Command percentage for Burner Fuel (BF) control. This is a percentage of the maximum fuel rate.
Setpoint VAC [% * 0.1]	Setpoint for open position of valve VAC (valve controlling airflow into furnace), scaled by 0.1.
Setpoint drum speed TT [% * 0.1]	Setpoint for drum speed, scaled by 0.1. The setpoint is a percentage of the maximum speed
Setpoint Closing VAF [%]	Setpoint for closed position of VAF (valve controlling ambient air intake) percentage.
Setpoint Opening VAT [%]	Setpoint for open position of VAT (valve controlling roaster air recycle) percentage.
Actual value Tc [°C]	Actual temperature Tc (bean probe) in degrees Celsius.
Actual value T1 [°C]	Actual temperature T1 (roaster air temperature out of the furnace) in degrees Celsius.
Actual value T2 [°C]	Actual temperature T2 (roaster air temperature into the roaster drum) in degrees Celsius.
Actual value T3 [°C]	Actual temperature T3 (roaster air temperature leaving the roaster) in degrees Celsius.
Actual value T4 [°C]	Actual temperature T4 (roaster air temperature entering the furnace) in degrees Celsius.
Flow gas BF [%]	Actual BF (Burner Fuel) flow rate as a percentage of total rate.
Present Value VAC [% * 0.1]	Actual open percentage of the VAC.
Actual value TT [rpm]	Actual drum speed TT in rpm.
Actual value closing VAF [%]	Actual value of the VAF closed percentage.
Actual value opening VAT [%]	Actual value of the VAT open percentage.
Air speed meter fSCP [m/sec]	Air speed measured by a Pitot tube in meters per second in the roaster outlet ducting. Sometimes denoted as v_g.

Table 3.1: Variables used in training the neural network.

These variables were chosen based on their relevance to the system dynamics, their potential impact on the model's predictive capabilities, and their availability. It is essential to distinguish between control "inputs" and model "inputs." Control inputs are signals controlled by a controller to achieve desired outputs, while model inputs are features used to predict the system's output response. For the coffee roaster system, the control inputs are the setpoints for BF, VAF, VAT, and VAC. Model inputs, however, are the lagged values of the control inputs and all the other variables listed in Table 3.1. The model input variables provide the necessary information about the system's state and control commands in the past, while the model outputs are the states of the system the model aims to predict for the next time step. Since this is a auto-regressive model, and the intent is to simulate a the "plant" dynamics for controller design, the outputs have to include all of the necessary information, that will be inputs for the next prediction. The importance of this will be discussed later.

An NARX model uses lagged values of both model inputs and outputs for prediction. Therefore, our neural network inputs consisted of lagged values for each variable in Table 3.1, with the output being the non-controller inputs (actual temperatures, flow rates, and pressures) for the current time step. The measurements were synchronized and taken 1 second apart. A dataloader was constructed to create "datapoints" containing lagged features for each model input variable as the x vector and the current values of those vectors as the model output y vector, as shown in Figure 3.1.

Various lagged values were experimented with (testing up to 30 seconds), but values beyond 5 seconds did not significantly improve results and increased training time. This is supported by Pillonetto's comments, where in dynamics systems, long term dependencies may not necessarily be beneficial [28]. The final model used 80 features (5 control inputs with 5 lagged values, and 11 model inputs with 5 lagged values) and produced 11 outputs (system states at the next time step) for each data point. The neural network architecture, shown in Figure 3.1, consisted of two dense hidden layers with 128 and 64 neurons, respectively.

The model was implemented using TensorFlow and was kept relatively simple due to resource constraints. It is important to note that all non-controller inputs must be predicted to make recursive predictions when simulating a roast. The data was split into training, validation, and test sets in a 70-15-15 ratio, ensuring that individual roasts were not divided between sets to avoid bias.

The neural networks were trained for approximately 50 epochs, sufficient to observe an "elbow" in the loss trend, reducing the risk of over-fitting. The obtained R-squared value and RMSE for the test set were 0.9999 and 4.812, respectively, calculated using all 11 output variables. Figure 3.2 shows an example of the model's fit for the bean probe temperature

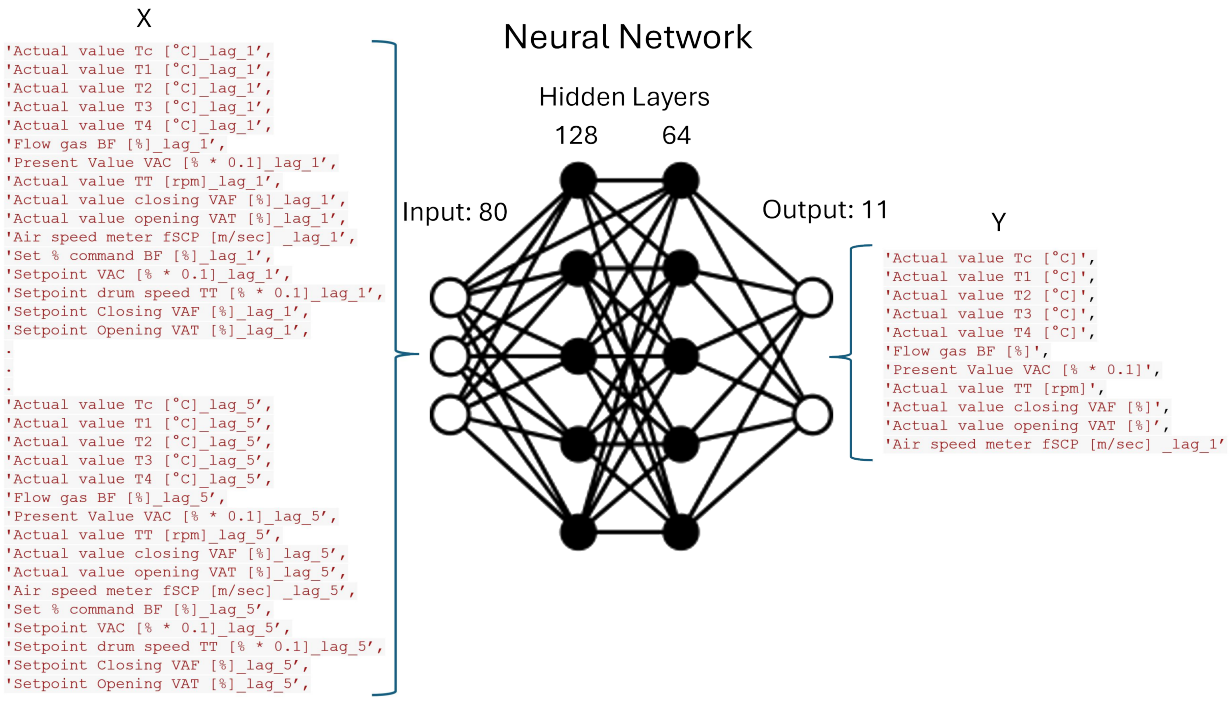


Figure 3.1: Structure of the NARX Neural Network developed for modeling the Coffee Roaster as a Black-box system.

using one-step-ahead predictions with actual data available.

Despite the high R-squared values and low RMSE, caution is necessary. When compared to a "naive" model, which predicts no change in the bean probe temperature at the next time step using actual measurements from the previous time steps, the performance is indistinguishable.

In machine learning, particularly in classification tasks, model performance is often compared to a naive model. For example, the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve is used to compare a model against a naive model that makes random guesses. An AUC value of 0.5 indicates a model that performs no better than random guessing, while an AUC less than 0.5 indicates a model performing worse than random guessing. In this study, a similar concept is employed by comparing the NARX model's performance to a naive model that assumes no dynamics, predicting that any variable's value remains constant from one time step to the next. This comparison provides a baseline to evaluate the usefulness of the NARX model, as expressed mathematically in Equation 3.1.

$$\hat{y}(t) = y(t-1) \tag{3.1}$$

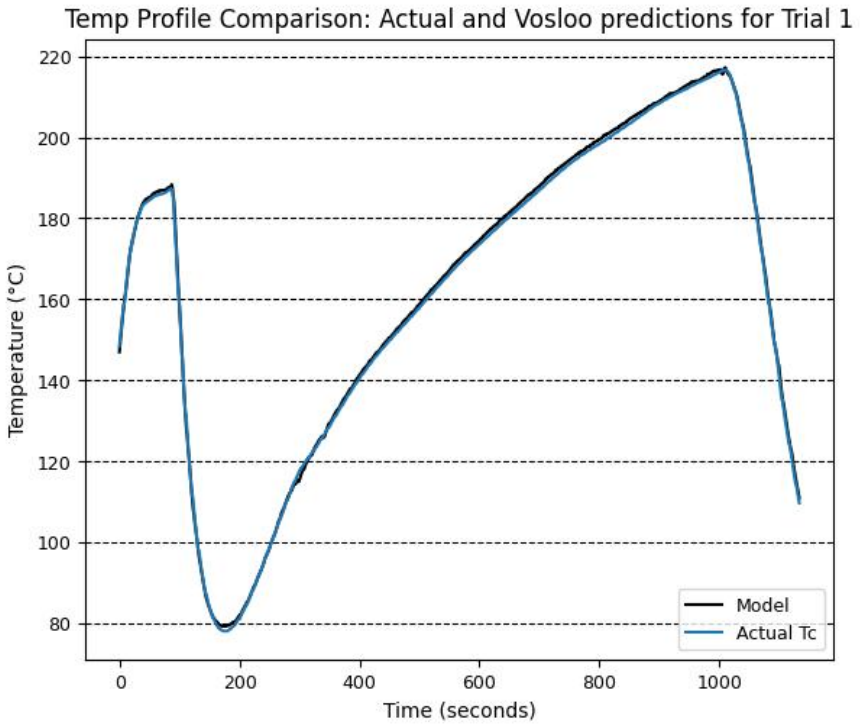


Figure 3.2: Neural ARX model predictions when prior actual data is available for each prediction. The model fit appears nearly perfect.

Equation 3.1 indicates that the predicted value at time t, $\hat{y}(t)$, is assumed to be the same as the observed value at the previous time step, $y(t-1)$. This assumption of no change provides a simple baseline against which the performance of more complex models can be compared. The naive model's performance, with an average R-squared value of 0.999 and a total RMSE of 1.446 for the three output variables (T_c, T_1, and T_2) is shown in Figure 3.3.

Performance metrics such as R-squared can sometimes be misleading when assessing the practical utility of a model. For instance, in the classic example of cancer prediction, a model that predicts no one has cancer might achieve a high accuracy because the majority of the population does not have cancer [35]. However, such a model would be practically useless if the intent was to identify actual cancer cases. This underscores the importance of comparing models against naive baselines to assess their true effectiveness. By benchmarking the NARX model against a naive model, the relative performance and practical value can be understood more clearly.

The high performance of the naive model likely stems from the fact that in any 1-second interval the temperature changes are very small due to the large thermal masses involved in coffee roasting. In other words, since the temperatures do not move much in comparison to their magnitudes over any 1 second interval, a model that assumes that there is no movement

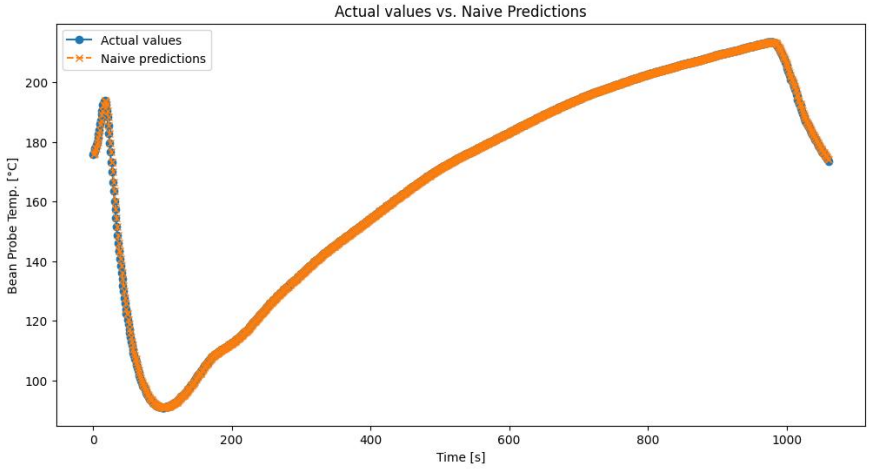


Figure 3.3: Predicted vs actual temperatures for a naive model, which predicts no change for the next step. Despite this, the R-squared and RMSE are deceivingly good, 0.999 and 1.446, respectively.

results in little error. This issue is further highlighted when the NARX model is tested using auto-regressive predictions, where prior predictions are used in place of actual data, resulting in significantly worse performance (Figure 3.4).

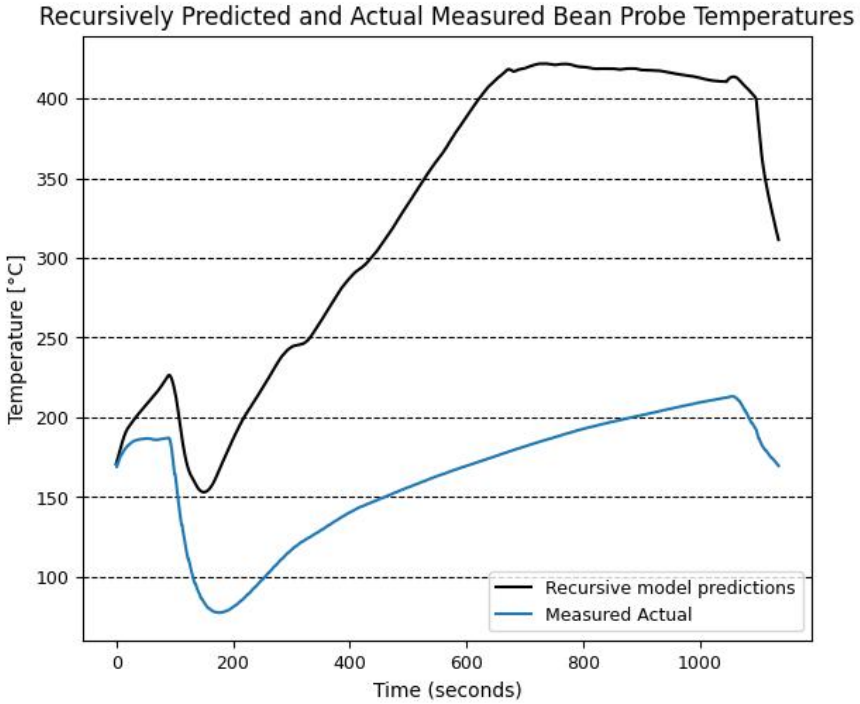


Figure 3.4: Auto-regressive predictions using the NARX model show significantly decreased performance.

This issue also applies to feed-forward neural networks and RNN architectures used for

"one-step-ahead" predictions. When the changes are small relative to the magnitude of what is being predicted, performance metrics may appear impressive when models are tested with measured data to predict the next step. However, this may not be sufficient for the model's intended purpose. These models may fail when used autoregressively, where predictions are fed back into the inputs to simulate system dynamics for controller design.

For the trained NARX model, performance visibly decreases when used autoregressively. However, it may appears the model has learned something about the system dynamics, as the trend shape is representative and not random. To further test this, the model was used to simulate a scenario the same as that shown in Figure 3.4, only the fuel control input was changed to a constant 1%. Figure 3.5 displays the results.

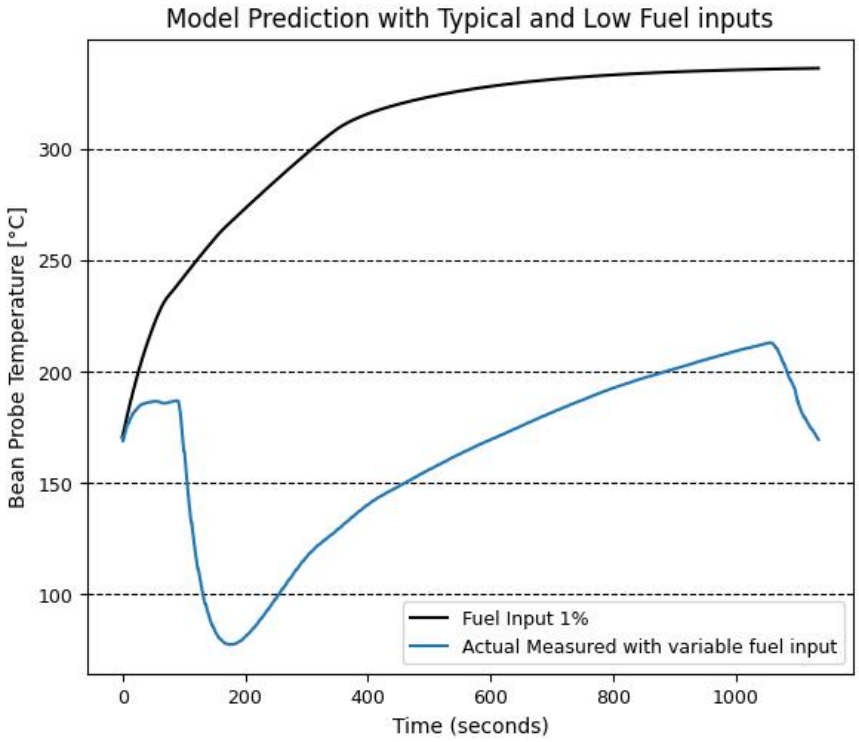


Figure 3.5: Simulation of black-box roaster model with a constant 1% fuel input results in unrealistic model behavior, missing the drying phase dip and showing higher-than-expected temperatures.

Although production data from a real system with a fuel rate kept at 1% is not available in the dataset, the performance can still be assessed by an intuitive understanding and domain knowledge of what the system should do. As discussed in the Section 2.3.1, the cold beans are dropped into the drum, which should lower the temperature of the probe during the drying stage. If the input heat from the furnace is lowered, we should expect the temperatures to drop potentially faster, and deeper. The model completely misses this and

conversely predicts that the probe temperature will continue to increase instead of decrease. This indicates that the model is not truly representative of the system. Additionally, when compared to the actual data (with all controls except for the fuel used in the simulation), the simulated bean temperature is higher. This is unrealistic, as the actual fuel rate was much higher than 1% throughout the roast. Since the fuel is the primary heat source, the bean temperature could not be higher in the simulated 1% fuel rate scenario.

This error in the model could be due to insufficient data for learning dynamics in regions where the fuel rate is low, such as the 1% scenario, or it could result from the model architecture or hyperparameters. While there may be many other reasons for the error, this example highlights a critical flaw in this method: verifying the correctness of the model without additional experimentation and data collection is challenging. Moreover, model performance metrics such as R-squared and RMSE can be misleading, depending on the method used for model testing. Overall, it is difficult to determine how accurately the model reflects the system dynamics when control settings differ from those in the dataset. Unfortunately, additional experimentation and extra data collection are not available for this study. This limitation is common in the industry, as running equipment under different regimes, apart from the current operational conditions, to verify models can be expensive or prohibitive.

In the examples above, it is likely that the NARX neural network has difficulties in accurately capturing the system dynamics due to the small changes in the variables compared to their magnitudes and the precision of measurements taken at 1-second intervals. This issue is highlighted by the naive model's ability to achieve high R-squared values despite predicting no change. Additionally, there are several states of the system, identified by physics-based models in other research [8, 10, 36], such as the rate of chemical reactions in the beans and moisture content, that are not directly measurable. This lack of complete data makes it challenging for any black-box approach to learn the correct dynamics accurately.

In contrast, traditional physics-based, or white-box, modeling offers advantages that address the challenges faced by the black-box approach. First, it ensures that the model behaves predictably within the constraints of well-understood physical laws, providing confidence in the model's consistency across different inputs. Additionally, these models are interpretable, allowing for the inspection and understanding of the results. Although there are many studies and methods for interpreting neural networks, it remains a challenging task [37]. Despite these advantages, explicitly modeling all important dynamics in a system, such as determining heat transfer coefficients, can be complex and difficult. Therefore, this study proposes combining physics-based models with neural networks to leverage the strengths of both approaches, resulting in a more robust grey-box model.

The following section explores this hybrid approach in detail, addressing the limitations

encountered with the black-box models and demonstrating how a grey-box model can provide a more accurate and reliable representation of the coffee roasting process.

3.3 Roaster Grey-box Modelling with Neural Networks

As discussed in Section 3.2, one-step-ahead black-box neural network models struggled to learning the correct dynamics for the coffee roaster due to the small changes in data over the time-steps between controller inputs. Conversely, white-box models that rely purely on first-principle physics laws and equations that accurately describe the roaster dynamics are difficult to write explicitly. This section details the exploration of a grey-box approach that combines first-principles, empirical relations, and neural networks to model the coffee roaster. First, the physics-based approaches developed primarily by Schwartzberg, and Vosloo (introduced in Chapter 2) and their performance using real commercial production data are reviewed. Next, development of a neural network integrated into the differential equations established in the physics-based approaches is discussed.

3.3.1 Physics-based Models

This section explores various physics-based approaches for modeling a coffee roaster system, laying the groundwork for developing a grey-box model. Among the different methods considered, the equations established by Schwartzberg [8] serve as the primary basis for exploration. Two main approaches are examined: Schwartzberg's original method, as detailed in his seminal paper [8], and a modified approach by Vosloo [10]. Vosloo's method integrates findings from additional studies, including those by Putranto and Chen [34], to enhance the model's accuracy and applicability. By comparing these methods and their performance with real production data, we aim to identify the most effective strategy for incorporating physics-based principles into a grey-box model.

As discussed in Section 2.3.3, several different roaster designs exist. The datasets used in this study are from various roasters in the TMR series developed by IMA. These are "R" roasters, meaning they recirculate heated air for re-use, unlike "SP" roasters where the air passes through only once. Although system dynamics can vary depending on the roaster design, the dynamics of the roasting drum are generally similar. Because of this, physical models developed tend to focus on the roasting chamber. For simplicity, and to align with other studies ([8, 10, 34, 36, 38]), the grey-box model focuses on modeling the roasting chamber.

In particular, Volsoo reviewed several roasting models including Schwartzberg model,

Heyd et al. model, fabbri et al. model, L-REA model proposed by Puratranto and Chen. He found that all proposed models showed good correlation between simulated and experimental results when the inlet temperature, T_2, and mass air flow rate are held constant. Vosloo determined that the models proposed by Schwartzberg, and Putranto and Chen were the most effective, as they required less computational effort while maintaining relatively accurate results [10]. Consequently, this study's model heavily relies on Vosloo's findings and uses Schwartzberg's model as a foundation due to its simplicity, widespread use, and Vosloo's endorsement.

Schwartzberg's model is described by a system of coupled differential equations with the states being bean temperature (T_b), moisture content (X_b), exothermic heat produced (H_e) and bean probe temperature (T_{rp}). The controlled inputs are considered as the inlet temperature (T_{gi}) and mass flow rate (G_g). The system of equations are summarized by equations 3.2 to 3.6:

$$\frac{dT_b}{dt} = \frac{G_g C_{pg}(T_g)(T_{gi} - T_{go}) + m_{db}(\frac{dH_e}{dt} + \Delta H_v \frac{dX_b}{dt})}{m_{db}(1 + X_b)C_{pb}(T_g)} \tag{3.2}$$

$$\frac{dX_b}{dt} = -\frac{4.32 \times 10^9 X_b^2}{(d_b \times 10^3)^2} \exp\left(-\frac{9889}{T_b}\right) \tag{3.3}$$

$$\frac{dH_e}{dt} = A_r \exp\left(-\frac{\Delta E_R}{T_b}\right)\left(\frac{H_{et} - H_e}{H_{et}}\right) \tag{3.4}$$

$$\frac{dT_{rp}}{dt} = K(T_b - T_{rp}) \tag{3.5}$$

Critically, the outlet air temperature, T_{go} is considered to be a algebraic constraint and is described by the following Equation 3.6.

$$T_{go} = T_{gi} - (T_{gi} - T_b)\left[1 - \exp\left(-\frac{h_e A_b}{G_g C_{pg}}\right)\right] \tag{3.6}$$

The thermal properties of the roasting air, such as heat capacity (C_{pg}), are evaluated at the average temperature between the inlet and the outlet for simplicity. This temperature is denoted as T_g. Detailed derivations and descriptions of the variables can be found in Appendix B. Relations determined experimentally by Vosloo [10] are used for the thermal properties of the roasting air, while the relation for the heat capacity of the coffee beans is taken from the study by Putranto and Chen [15]. Vosloo, and Putranto and Chen, found that the thermal properties of the roasting air could be sufficiently described as a function of air temperature, while the coffee bean properties could be described by a function of both moisture content and temperature [10, 15]. Details of these relations are detailed in

Appendix A.

Although the rotational speed of the drum does not appear explicitly in these equations, it significantly affects the effective heat transfer rate (h_e) according to Cristo et al [24]. The heat transfer term, h_e, is explicitly represented in equation 3.6, and thus implicitly shows up in the equation for the bean temperature, equation 3.2, and by extension of the bean temperature in Equations 3.3, 3.4, and 3.5. Because of this, it is a very influential parameter that must be carefully evaluated to ensure the model's accuracy.

Schwartzberg and Vosloo take different approaches in addressing the effective heat transfer coefficient. Schwartzberg assumes a constant heat transfer coefficient (h_e) and uses regression techniques to determine a constant that minimizes the error between actual measurements and model predictions, whereas Vosloo's study tries to explicitly model the h_e relation. Vosloo does this by approximating the heat transfer between the bean and the air using a model of a sphere in a laminar air stream [10]. The exact method is shown in Equation 3.7 and derivation is explained in Appendix B. This approach is challenging because it requires knowledge of the relative speed of the bean to the air flow, not just the absolute average velocity of the air. Given that the absolute air movement in a roaster is relatively slow, heat transfer is more influenced by the movement of the beans within the drum. This movement is primarily determined by the drum's rotational speed, the amount the drum is filled, and the drum geometry according to Cristo et al. [24].

Schwartzberg proposes that h_e can be considered time invariant if the inlet temperature and the mass flow rate are held constant, and determines this value, and other parameters, by minimizing the error between actual measurements and model predictions [8]. Vosloo suggests that this value might be approximated by modeling the airflow around a bean, using the Nusselt, Biot, Prandtl, and Reynolds numbers, making it a function of the air temperature, the general geometry of the bean, and the properties of roaster air, which depend on the roasting air temperature [10]. The values for the air velocity come from the mass flow rate and geometry of the drum. Details of this approach are in Appendix B, and the final representation of the effective heat constant, using Vosloo's method, is found in Equation 3.7.

$$h_e = \frac{\lambda_b \left(0.6\lambda_g \left(\frac{c_{pg} u_g}{\lambda_g} \right)^{1/3} (d_b \rho_g v_g)^{1/2} + 2 \right)}{d_b \left(0.6 d_b + \lambda_b + 0.18\lambda_g \left(\frac{c_{pg} \mu_g}{\lambda_g} \right)^{1/3} \left(\frac{d_b \rho_g v_g}{\mu_g} \right)^{1/2} \right)} \tag{3.7}$$

Critically, both approaches neglect the effect of the beans' movement in the drum, which significantly affects the heat transfer between the beans and the air, metal, and other beans, as pointed out by Cristo et al. [24]. By modeling the effective heat transfer coefficient, h_e,

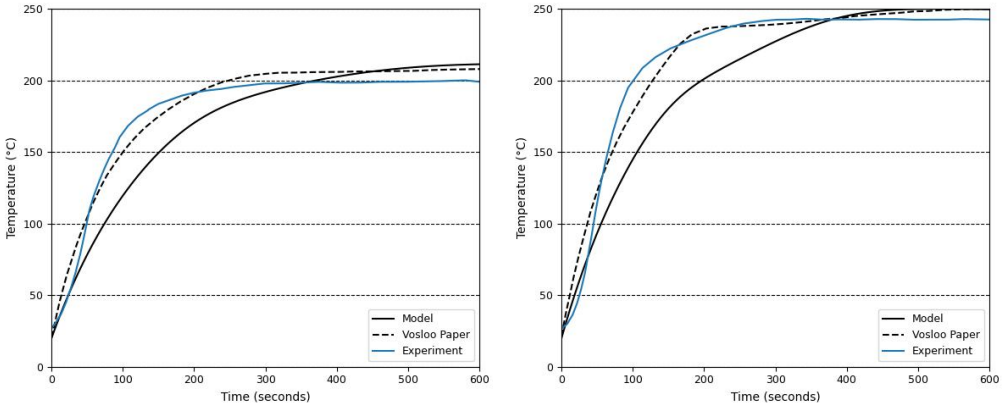
as a neural network these effects can potentially be included. Details such an approach are described in 3.3.2.

Initially, the results from Vosloo's model were replicated to ensure consistency with his paper [10]. Vosloo's parameters, shown in Table 3.2, were used to simulate the roaster drum dynamics. The results from Vosloo [10] and from this studies reproduced model are shown in Figure 3.6.

Table 3.2: Simulation Parameters used by Vosloo [10]

Parameter	**Value**
A_b	0.08 m^2
d_b	6.6×10^{-3} m
G_g	0.02 kg/s
$m_b(d.b)$	91.8×10^{-3} kg d.b
$T_{b,i}$	20 °C
$T_{g,i}$	210, 250 °C
t	600 s
v_g	0.2 m/s
$X_{b,i}$	0.082 kg/kg d.b.
ΔH_v*	2790×10^3 J/(kg K)
A_r*	1.162×10^8 W/ kg
ΔE_R*	5500 K
H_{et}*	232×10^3 $J/\ kg$

Parameters with "*" are values Vosloo takes from literature [10]

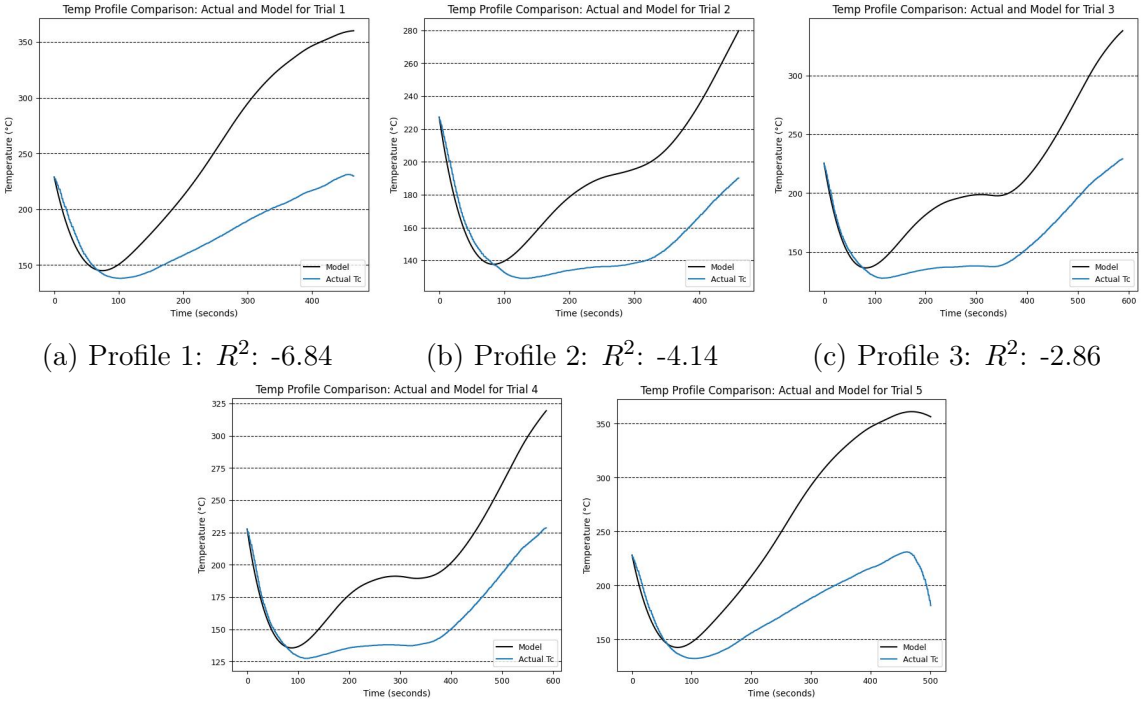


(a) 200°C Constant inlet temperature (b) 210°C Constant inlet temperature

Figure 3.6: Comparison of recreated model, and results from Vosloo [10]. "Experiment" refers to the physical experiment performed by Vosloo, and "Vosloo Paper" is the physics based model that was developed by Vosloo.

Although the solutions are similar and have good general agreement, there are some discrepancies in the bean temperature. Vosloo's model was reviewed, and the discrepancies were considered small enough for the methods to be verified.

Next a baseline for Vosloo's model was constructed, using constants in Table 3.2 and Equation 3.7, with control inputs from real production data. The roast profile constant, K, was picked to be 0.0012. This value was chosen to be close to values seen in other papers [8, 10, 38]. A few of the results are shown in Figure 3.7.



(a) Profile 1: R^2: -6.84 (b) Profile 2: R^2: -4.14 (c) Profile 3: R^2: -2.86

(d) Profile 4: R^2: -1.77 (e) Profile 5: R^2: -6.86

Figure 3.7: Comparison of Vosloo's method with the actual measured roaster data of variable roasting air temperature, velocity, and drum rotation speed. Although R^2 values are poor, these physics based models are able to get the correct general shapes of the profiles.

For this model, the R^2 values are poor, as can be seen visually in Figure 3.7. Despite this, the model does seem to be somewhat representative of the dynamics as the shape of the profile is largely correct and is not random. This could be evidence that the underlying physics are largely correct, but values such as K, or other parameters, are not accurate.

As suggested in Figure 3.7, and will also be emphasized later, the performance of physics-based models significantly depends on accurately fitting unknown parameters, such as K, to the production data. The more unknown parameters there are, the greater the uncertainty introduced into the model. To minimize this uncertainty, only trials from a TMR25 roasting machine were selected from the dataset. These trials included detailed information about

the piping where the average air velocity was measured, the mass of green beans introduced into the roast, and the approximate geometry of the drum. These parameters are crucial for the model and are directly measurable, making them readily available if the equipment is physically accessible. This detailed information allows for a precise assessment of how well Vosloo's physics-based model performs with real production data, reducing the added uncertainty from unknown parameters.

Vosloo approximates the properties of the beans, such as the heat generated by chemical reactions and the heat of vaporization, based on values found in literature, as shown in Table 3.2 [10]. Other parameters, specific to the equipment but not directly measurable, such as K, were determined using optimization techniques. The state variables of the system and the control inputs are expressed as Equation 3.8 and Equation 3.9.

$$x = [T_b, T_{go}, X_b, H_e, T_{rp}] \tag{3.8}$$

$$u = [T_{gi}, v_g, T_T] \tag{3.9}$$

The system of equations describing the system (Equations 3.2 to 3.6) can be rewritten and summarized as Equations 3.10 to 3.14.

$$\frac{dT_b}{dt} = f_b(T_b, T_{go}, X_b, H_e, T_{rp}, T_{gi}, v_g, T_T, K) \tag{3.10}$$

$$\frac{dX_b}{dt} = f_{X_b}(T_b, X_b) \tag{3.11}$$

$$\frac{dH_e}{dt} = f_{H_e}(T_b, H_e) \tag{3.12}$$

$$\frac{dT_{rp}}{dt} = K(T_b - T_{rp}) \tag{3.13}$$

$$\frac{dT_{go}}{dt} = f_{T_{go}}(T_{gi}, T_b, h_e, A_b, G_g, c_{pg}) \tag{3.14}$$

The optimization process involves finding the parameter K that minimizes the objective function $J(K)$. The objective function to be minimized is the mean squared error (MSE) between the predicted and actual temperature profiles. Equation 3.15 can be written for each roast profile i.

$$\text{MSE}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \left(T_{rp}^{(i)}(t_j) - T_c^{(i)}(t_j) \right)^2 \tag{3.15}$$

Here, $T_{rp}^{(i)}(t_j)$ is the predicted bean probe temperature at time t_j, $T_c^{(i)}(t_j)$ is the actual bean temperature at time t_j, and N_i is the number of data points in each roast profile i. The

overall objective function to be minimized across all roast profile is written as Equation 3.16.

$$J(K) = \frac{1}{M} \sum_{i=1}^{M} \text{MSE}_i \tag{3.16}$$

Where M is the total number of datasets. The optimization problem can be expressed as Equation 3.17 and subject to condition 3.18.

$$\min_{K} J(K) \tag{3.17}$$

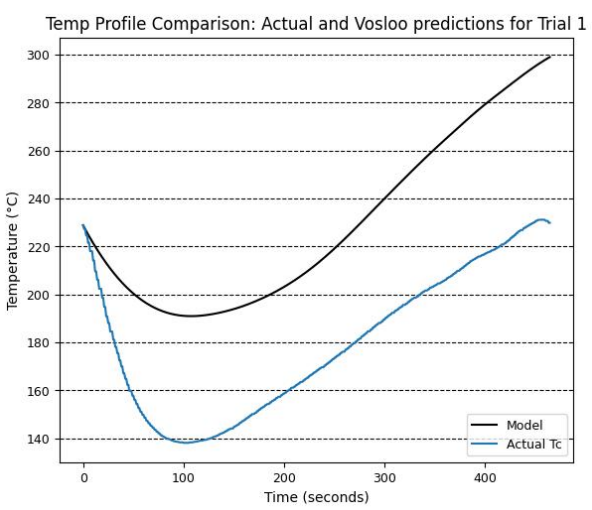
$$K > 0 \tag{3.18}$$

The entire process can be summarized mathematically as in Equation 3.19.

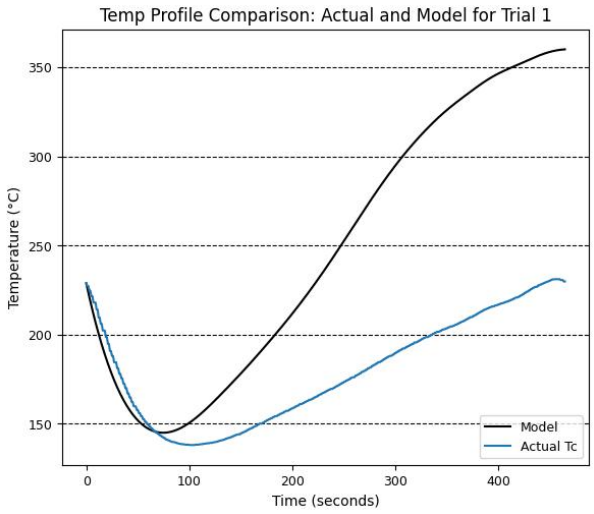
$$K^* = \arg\min_{K} \left\{ \frac{1}{M} \sum_{i=1}^{M} \frac{1}{N_i} \sum_{j=1}^{N_i} \left(T_{rp}^{(i)}(t_j) - T_c^{(i)}(t_j) \right)^2 \right\} \tag{3.19}$$

Here K^* is the optimized parameter that minimizes the overall mean squared error between the predicted and actual temperature profiles across all datasets. This was done using the L-BFGS-B optimization algorithm.

Using this method, the optimized value of K was found to be approximately 0.00376, resulting in an average R^2 value of -1.384. An example of the model's prediction for Trial 1 is shown in Figure 3.8a. Figure 3.8b compares the difference between the baseline case, using K of 0.0012, and this optimized case. Although the R^2 value improved compared to the method used in Figure 3.7, the overall chart with the optimized value of K is less representative, particularly in the drying section of the roast process, which is no longer captured as accurately. This discrepancy suggests that the model might be missing some critical dynamics or that assumptions about other parameters, such as the constants adapted from literature, might be inaccurate. Either way, this shows that the method proposed by Vosloo is insufficient to effectively model the production data.



(a) Trial 1, K=0.00376, R^2: -1.84



(b) Trial 1, K=0.0012, R^2: -6.48

Figure 3.8: Comparison of optimized K and Benchmark K

Although this approach uses the same base set of equations as Schwartzberg's method, it differs by using literature values and explicitly attempting to model h_e. Schwartzberg's method involved optimizing several parameters simultaneously, including values for $h_e A_b$ and factors related to the exothermic reaction, such as A_r, H_{et}, $\Delta E/R$, and H_v. For the next experiment, Vosloo's method for estimating h_e was maintained, however, instead of using values from literature, a methodology was adapted to determine the unknown parameters using a similar optimization technique as the prior experiment. We will denote this as the "Schwartzberg Method" as it seeks to optimize several parameters all at once like Schwartzberg does in his original paper [8]. The optimization problem is now expressed as a function of A_r, H_{et}, $\Delta E/R$, and H_v. To address collinearity issues that arise when using regression on these parameters individually, several parameters are combined: $A = A_r$, $B = \frac{A_r}{H_{et}}$, $C = \frac{\Delta E}{R}$, $D = H_v$.

The optimization problem is then expressed as equation 3.20 and results are summarized as Figure 3.9.

$$\min_K J(A, B, C, D, K) \tag{3.20}$$

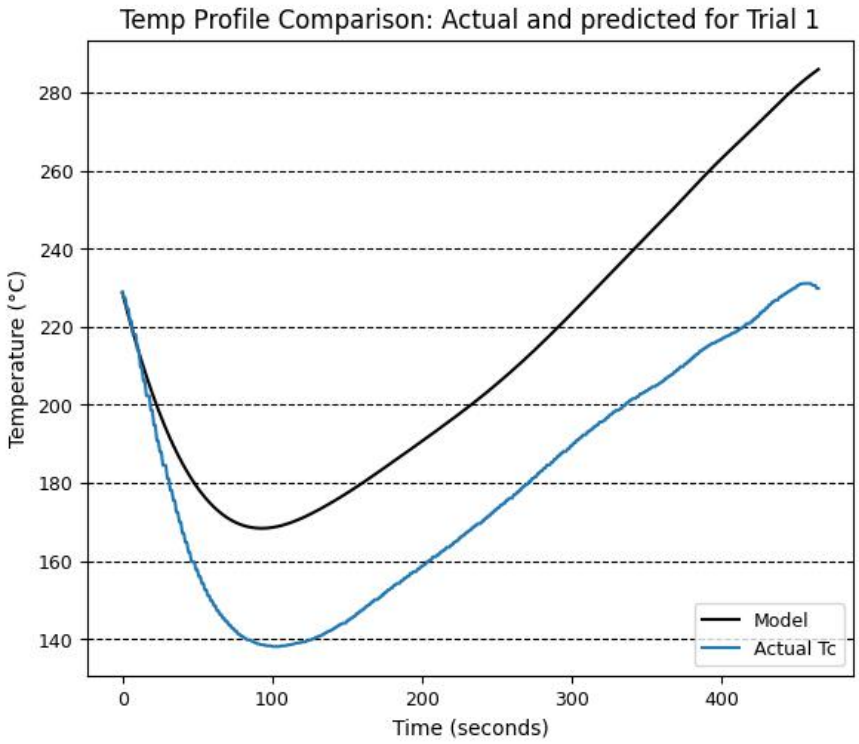


Figure 3.9: Trial 1, Schwartzberg method, R^2: -0.350. The fit improves from the Vosloo method and the benchmark K method

As shown in Figure 3.9, the fit improves marginally compared to the results in Figure 3.8a and 3.8b when using the Schwartzberg's method. Interestingly, the values of A (A_r) and D (H_v) did not change significantly from the values used in Vosloo's study, whereas the values of K, B ($\frac{A_r}{H_{et}}$), and C ($\frac{\Delta E}{R}$) did. Vosloo's study uses values sourced from literature, but there is uncertainty about the consistency of these properties across different coffee bean types. The results suggest that some parameters may remain consistent across various coffee beans. In contrast, other values are likely more specific to the physical properties of the individual coffee beans, which might be expected given their relation to the chemical reactions occurring within the beans.

Notably, the value of K differs from the value determined by regression when optimizing for that parameter alone. When K was optimized independently, the value was approximately 0.00376, whereas now it is nearly twice as much, at 0.00721. This highlights that for these techniques, the estimates of the parameter values can vary depending on the accuracy of the other parameter values. Therefore, it is crucial to optimize parameters simultaneously rather than individually, as optimizing one parameter at a time can lead to inaccuracies.

Finally, a regression was conducted assuming that the effective heat transfer coefficient (h_e) could be considered constant. This approach, along with parameterizing unknown

factors, aligns with Schwartzberg's original proposal [8]. Schwartzberg's model implicitly assumes that h_e is a constant value determined by minimizing the loss between model predictions and actual data. However, Cristo et al. [24] argue that h_e likely varies with bean movement. The assumption of a constant h_e may be valid if the mass flow rate of the air and the drum's rotational speed are held constant, as in Schwartzberg's study. This assumption makes the equations specific to those conditions and potentially introduces errors when conditions vary.

For the dataset used in this study, these conditions are approximately met, as shown in Figure 3.10. The average air velocity is around 7.5 to 8.0 m/s, and the drum rotation speed remains close to 70 RPM for most of the roast. Although the RPM varies slightly, it decreases primarily at the end, where the temperature difference between the beans and the roasting air is small, minimizing the impact on h_e, since h_e is a factor of the temperature difference which is smallest at the end of the roast.

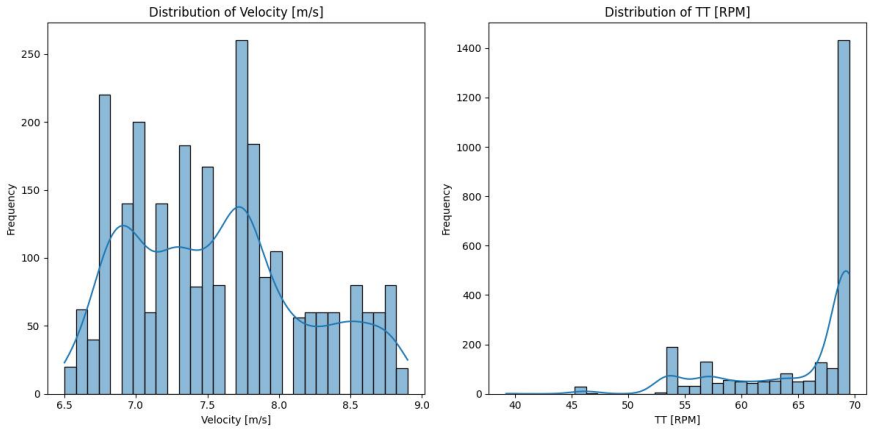


Figure 3.10: Distributions of the average air velocity and drum speed, TT. This shows that the values are relatively constant

An example of the results of the optimization under the assumption of a constant effective heat transfer coefficient can be seen in Figure 3.11 and the overall R^2 for the values is about 0.946.

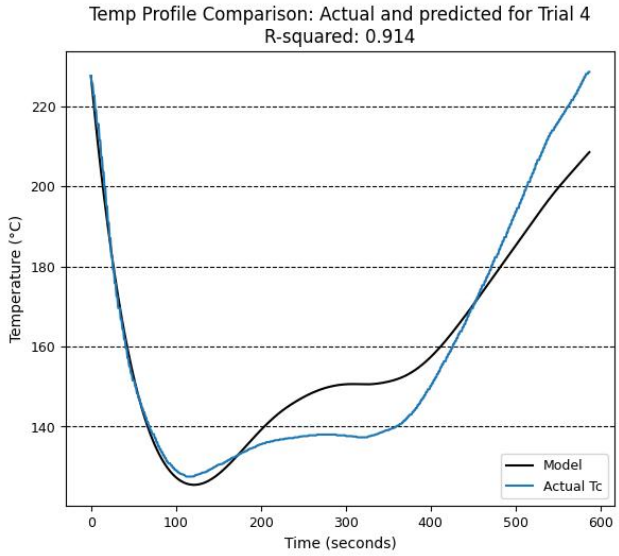


Figure 3.11: Model predictions compared to the actual measured data assuming that H_e can be considered roughly constant

While the improvement is significant, there are some critical caveats. It is unclear how well this method generalizes to other flow rates and drum speeds, as the available data is fairly uniform. Cristo et al. demonstrated that bean motion in the drum changes significantly with varying speeds, transitioning through slumping, rolling, cascading, cataracting, and centrifuging regimes [24]. Consequently, as conditions (particularly drum speed and mass flow rate) deviate from the averages used for optimization, the model's accuracy is likely to decrease. To use this model for a control system, it would be necessary to maintain constant parameters such as drum speed and airflow, using only the inlet temperature to control the roast profile. Alternatively, a schedule of h_e values could be created for different ranges of tumbling speeds and airflows, tailored to a specific drum geometry and fill level. Using a neural network to learn the relation of h_e instead of relying on constant assumptions may allow a more generalizable model to be developed without such caveats.

In summary, the exploration of physics-based models for coffee roasting underscores the strengths and limitations of both Schwartzberg's and Vosloo's approaches. Vosloo's method, while providing representative results, lacks the accuracy needed for simulating controllers and dynamics with real production data. On the other hand, Schwartzberg's method, despite offering a good fit for specific drum speeds and airflow rates, proves impractical for generalizing the roasting dynamics and controlling simulations that may involve varying these inputs. These findings highlight the necessity of a more flexible modeling approach that can accurately represent the complex interactions within the roasting process, paving the way for the development of a grey-box model that integrates neural networks with physics-based

principles for enhanced control and simulation accuracy.

3.3.2 Grey-box Model

This section delves into the development of a grey-box model for the coffee roasting process, combining the strengths of physics-based modeling and data-driven neural networks. While physics-based models, such as those proposed by Schwartzberg and Vosloo, provide a foundational understanding of the roasting dynamics, they struggle with accuracy when applied to real-world production data due to their reliance on certain assumptions, such as constant inlet temperatures, and limited parameterization. Conversely, purely data-driven approaches like neural networks can capture complex relationships within the data but can struggle depending on the structure of the data and lack the interpretability and grounded nature of physical principles.

To bridge this gap, a grey-box approach is adopted. This method leverages the underlying physics described by differential equations while incorporating neural networks to model complex, non-linear relationships that traditional methods may fail to capture accurately. The key challenge addressed in this subsection is the modeling of the effective heat transfer coefficient (h_e), which significantly impacts the accuracy of the roasting process simulation.

The development of the grey-box model involves several steps: exploring production and synthetic datasets, establishing the model architecture, and training the neural network to predict h_e within the differential equations. This integration aims to create a more robust and flexible model capable of accurately simulating the coffee roasting dynamics under varying conditions. The process, implementation details, and the results of this approach are thoroughly examined, highlighting the potential of grey-box modeling to enhance control and prediction in coffee roasting systems.

Data

Unfortunately, the data for roasts that includes the mass of the coffee beans, the geometry where the average speed of the air is measured, and the size of the drum is limited and falls within a narrow distribution for values such as the drum rotation and air mass flow rates as described in Section 3.3.1. This limitation makes it challenging to confidently apply the aforementioned techniques in a way that accurately trains a model that generalizes to other conditions.

To address this, two approaches were take simultaneously to train and verify the model. First a relationship for the effective heat transfer was assumed, specific parameters were chosen, and synthetic data was generated. The details of which can be found in Appendix B.

This approach allowed for the theoretical exploration of a technique to identify h_e as a function of drum rotation speed, air velocity, and mass. The synthetic data was then used to test a grey-box approach that utilizes neural networks to effectively evaluate the h_e value within the system of differential equations.

The second approached used production data. However, since mass of the coffee beans, the geometry where the average speed of the air is measured, and the size of the drum were not recorded explicitly, some knowledge about the business practices and general size of equipment was used to estimate these values in hopes the model could be trained and tested to verify the approach. This consisted of roughly 300 roasts with the temperature profiles summarized in Figure 3.12. Summary statistics for the dataset can be found in Appendix C.

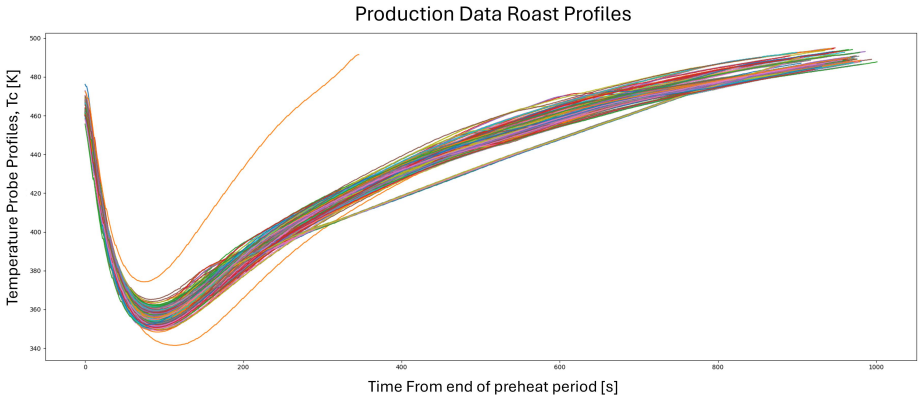
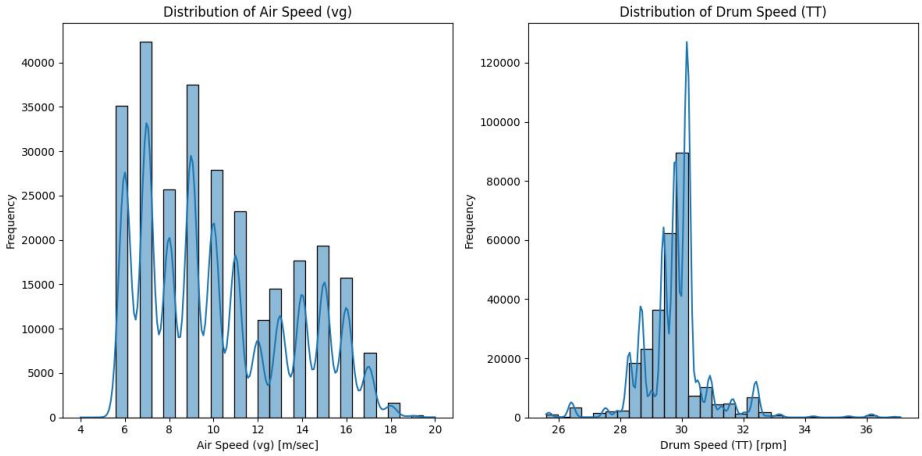


Figure 3.12: Roasting profiles for the training grey-box model with production data.

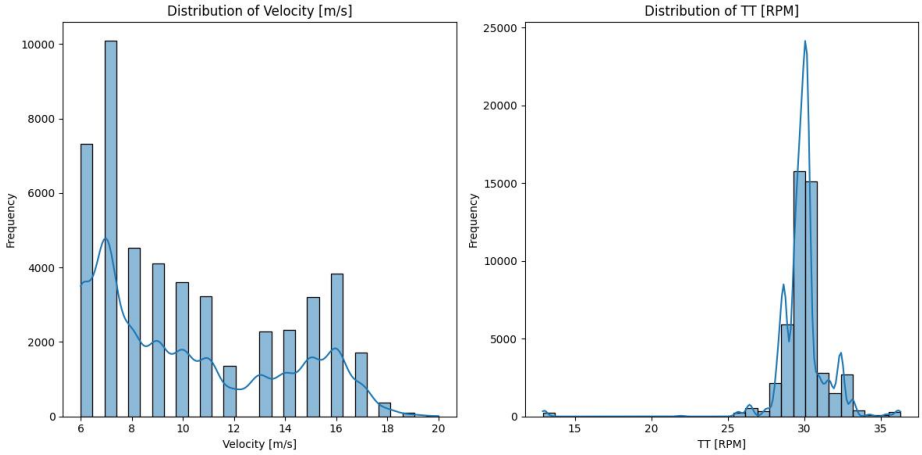
The generation of the synthetic data involved taking the available production data, regardless of whether bean mass and other details were known, extracting the initial conditions and control inputs, and simulating the roasts with the assumed relations for h_e, h_c, and h_k found in Appendix B. This resulted in simulated roast data with distributions similar to the production data but on hypothetical roaster equipment. By using synthetic data, we could incorporate all available data, even those without known parameters, thus achieving a much wider distribution. This wider distribution is still realistically obtainable on a hypothetical machine as defined by the model generating the synthetic data. The data was collected into a new dataset.

Figure 3.13a and Figure 3.13b summarize the distributions of inputs (air velocity and drum rotation speed) for the production training and test data. While the distributions between the test and training sets are similar, they are distinct enough to demonstrate that the model generalizes well and avoids overfitting. As the roaster operates within defined operational bounds, the model does not need to generalize to all possible values but only

those within the operating range. Since the control inputs are identical for both the synthetic data and the production data the distributions are nearly identical.



(a) Distribution of velocity and drum rotation speed for production training data.



(b) Distribution of velocity and drum rotation speed for production test data.

Figure 3.13: The distributions for the inputs velocity and drum rotation speeds are similar, but slightly different between the test and training sets. This helps with avoiding overfitting during training.

Model Theory

For this Grey-box approach, the system of equations described in equations 3.2 to 3.6 with "Schwartzenberg method" described in Section 3.3.1 will be used and augmented with a neural network. To do this the "unknown" parameters such as A, B, C, D and K from Equation 3.20 were parameterized and the relation for the effective heat transfer coefficient,

h_e, from Vosloo or Schwartzberg is replaced with a neural network, NN_{h_e}, as described in Equation 3.21.

$$h_e = \text{NN}_{h_e}(T_g, v_g, T_T) \tag{3.21}$$

Where T_g is the average gas temperature, ω is the drum rotation speed, v_g is the air velocity, T_T is the drum temperature. The neural network is trained to minimize the prediction error of the bean probe temperature T_{rp} compared to actual measurements T_c.

Here the parameters A, B, C, D and K, as well as the weights in the h_e neural network will be determined simultaneously using backpropagation with the loss function being minimized described by Equation 3.22.

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(T_{rp}^{(i)} - T_c^{(i)})^2 \tag{3.22}$$

Where $T_{rp}^{(i)}$ is the predicted bean probe temperature at time step i, $T_c^{(i)}$ is the actual measured temperature at time step i, N is the total number of time steps. The neural network is trained using Algorithm 1.

Data: List of training trials $T = \{D_1, D_2, \ldots, D_M\}$, where each trial $D_j = \{(x_i, y_i)\}_{i=1}^{N_j}$ consists of N_j training samples.

Result: Trained neural network parameters θ

Initialization: Initialize the neural network parameters θ and the Adam optimizer.

for *epoch = 1 to max_ epochs* **do**

  Shuffle the order of trials in T.

  for *each trial D_j in the shuffled trial list T* **do**

    Reset the accumulated loss: $\mathcal{L}_{total} \leftarrow 0$.

    Extract the state data and control data from trial D_j.

    Initialize the initial state s_0 based on the first data point in D_j.

    for *each time step $t = 1$ to $N_j - 1$* **do**

      Perform forward propagation to predict the next state derivative: $\hat{s}_{t+1} = s_t + f(s_t, u_t; \theta)$, where u_t is the control input at time step t.

      Compute the loss for the state variable 'Tc': $\mathcal{L}_t = \text{MSE}(\hat{s}_{t+1}[3], y_{t+1}[3])$, where the index 3 corresponds to 'Tc'.

      Accumulate the loss: $\mathcal{L}_{total} \leftarrow \mathcal{L}_{total} + \mathcal{L}_t$.

      Update the state: $s_t \leftarrow \hat{s}_{t+1}$ (detach to avoid backpropagation through time).

    end

    Compute the gradients of the total trial loss: $\nabla_\theta \mathcal{L}_{total}$.

    Update the parameters using the Adam optimizer: $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}_{total}$.

  end

end

Algorithm 1: Pseudo code Training Algorithm with Embedded Neural Network in ODE

The overall process involves simulating the coffee roasting process using the neural network to predict the heat transfer coefficient, computing the temperature profiles, and updating the neural network and model parameters to minimize the prediction error. The optimization problem can be summarized as:

$$\arg\min_{\theta} \left\{ \frac{1}{N} \sum_{i=1}^{N} (T_{rp}^{(i)} - T_c^{(i)})^2 \right\} \tag{3.23}$$

Where θ represents the parameters of the neural network for determining the effective heat transfer coefficient and other parameterized values like A, B, C, D, and K in Equation **??**.

Neural Network Architecture

The neural network used to model the effective heat transfer coefficient (h_e) is a fully connected feedforward network with three hidden layers. The final architecture, after iterative tuning, consists of 128, 64, and 32 units in the hidden layers, respectively, with ReLU activation functions. Table 3.3 provides a detailed breakdown of the network's architecture.

Table 3.3: Neural Network Architecture

Layer	**Units**	**Activation**
Input Layer	3 (features: T_g, v_g, TT)	-
First Hidden Layer	128	ReLU
Second Hidden Layer	64	ReLU
Third Hidden Layer	32	ReLU
Output Layer	1	Linear

Due to the extensive time and resources required for model training, hyperparameter tuning efforts were limited. The investigation began with a simple single hidden layer architecture of 128 units and iteratively added layers until satisfactory results were achieved. The neural network itself is a fully connected feedforward network, utilizing the air temperatures (T_g), measured air velocity (v_g), and drum rotation speed (TT) as input features. The output is the effective heat transfer coefficient (h_e).

Implementation Details

The implementation was done in Python, utilizing the PyTorch library for building and training the neural network. The training process involved using the Adam optimizer and the mean squared error (MSE) as the loss function. Due to the complexity of the model and the need to solve the ODEs for the entire roast duration, the training was computationally expensive, taking approximately 48 hours on a Nvidia RTX 4050 GPU.

This approach allows for a more flexible and accurate representation of the heat transfer dynamics in the coffee roasting process, leveraging the power of neural networks to learn complex relationships from data. The results are discussed in Chapter 4.

3.4 Residual Modelling

A second exploration into improving the model involved modeling the residuals of the initial grey-box model. While the physics-based approach was augmented by adding a neural network to capture the dynamics of the effective heat transfer coefficient (h_e), higher-order

terms might still not be captured. To address this, a neural network was added to estimate these residuals.

The approach involved using the grey-box model descirbed in Section 3.3.2 and enhancing it further with an additional neural network to capture the residuals between the predictions and the actual data. The process to do this is illustrated in Figure 3.14.

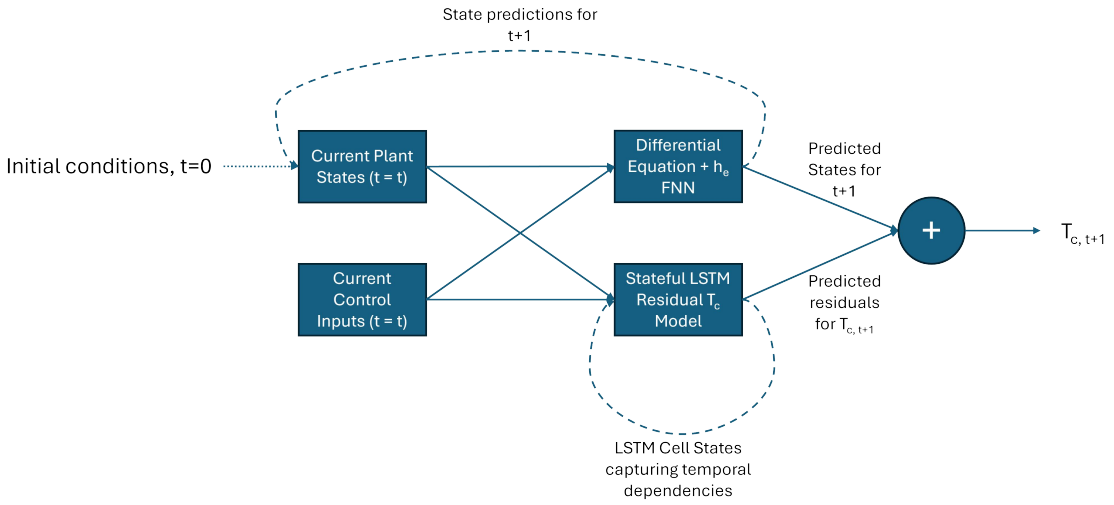


Figure 3.14: Process showing the flow of inputs to output of the model incorporating an LSTM network to capture residuals

In this process, the new LSTM network takes in the current states and control inputs, and outputs estimates of the residual values, or difference between the actual bean probe temperatures and the predicted values. This approach tries to avoid the problem of small differences between temperatures resulting in small relative errors that was described in section 3.2 by isolating only differences from the rest of the model and predicting these values. In order to train the LSTM network, the residuals between between the actual and predicted Tc values for each time step needed to be calculated. This was done in as part of the data processing stage and these values are only used in training, process. They are not needed for predictions. This is critical, as actual values would not be available in the scenario where a controller or plant is being simulated.

An LSTM based architecture was chosen, since these are well known at capturing temporal dependencies, which are likely present for determining higher order dynamics. Also, since the inputs of the model are predicted states of the initial grey-box model, because of error propagation, the error is likely related to the time in the roast and therefor have some temporal dependency. The final architecture used for the LSTM is summarized in Table 3.4.

The input layer consists of seven features: three control variables (Tgi, TT, vg) and the four state variables predicted by the initial grey-box model (Tb, Xb, He, Trp). A fully

Table 3.4: Architecture of the Residual LSTM Network

Layer	**Description**	**Parameters**
Input Layer	Input features	7 (T_{gi}, TT, vg, T_b, X_b, H_e, T_rp)
LSTM Layer 1	LSTM layer with hidden states	Hidden size: 127
LSTM Layer 2	LSTM layer with hidden states	Hidden size: 127
LSTM Layer 3	LSTM layer with hidden states	Hidden size: 127
Fully Connected Layer	Fully connected layer for residual prediction	Output size: 1 ($T_{c,t+1}$ residual)

connected (dense) layer predicts the residuals for Tc from the LSTM's output. The LSTM network was implemented in PyTorch, with an architecture optimized through hyperparameter tuning.

Hyperparameter tuning was conducted using Optuna, a hyperparameter optimization framework [39]. The goal was to identify the optimal configuration of the LSTM network that minimizes the validation loss. The hyperparameters tuned included hidden size, number of layers, and learning rate. Optuna's Tree-structured Parzen Estimator (TPE) [39] was used to explore the hyperparameter space efficiently. Fifty trials were conducted with a subset of the training data (30 roasts) to identify the best combination of hyperparameters. A smaller subset of roasts were used because of resources and time constraints, and optimization using all 300 roasts per epoch was not feasible. With the 30 roasts the hyperparameter tuning took more than 48 hours. Early stopping was implemented with a patience of 10 epochs to prevent overfitting, and a learning rate scheduler was used to adjust the learning rate based on the validation loss. The best parameters identified were a hidden size of 127, three layers, and a learning rate of about 0.000134.

The Tree-structured Parzen Estimator (TPE) is a sequential model-based optimization (SMBO) algorithm. Unlike traditional random or grid search methods, TPE constructs two probabilistic models: one for the set of good hyperparameter configurations and one for the set of all observed configurations. The goal is to maximize the expected improvement (EI) of the objective function. Mathematically, the TPE approach is grounded in the concept of modeling the density functions $l(x)$ and $g(x)$:

$$l(x) = p(x|y < y^*) \tag{3.24}$$

$$g(x) = p(x|y \geq y^*) \tag{3.25}$$

Here, y^* is a quantile (e.g., the median) of the observed values of the objective function, $l(x)$ represents the density of good configurations, and $g(x)$ represents the density of all configurations. The TPE algorithm selects the next hyperparameters to evaluate by maximizing the ratio $\frac{l(x)}{g(x)}$, which effectively prioritizes configurations that are likely to improve

the objective function.

After tuning the hyperparameters on the subset, the model was trained on the full set containing 300 roasts. Mean Squared Error (MSE) between the residuals of the initial grey-box and the LSTM networks estimated residuals were used for the loss function. The loss here as accumulated across the entire roast before gradients were updated. The Adam optimizer [40] was used to update the network parameters, and the loss was backpropagated through the network. This involved backpropegation through time (BPTT), because of the temporal nature of the LSTM networks. After each epoch, the network was validated on a subset of the data to monitor its performance. The validation loss was used to update the learning rate scheduler and to implement early stopping if no improvement was observed.

A key aspect of the training process was the stateful nature of the LSTM. Instead of feeding a sequence of states and inputs to the model, the current states and control inputs were fed in sequentially. This approach leverages the LSTM's ability to maintain cell states over long sequences, capturing dependencies that are important for modeling the dynamics of the system. The stateful LSTM maintains the state across different roasts, and were manually reset between each roast profile, allowing the model to remember the previous information throughout the entire roast profile but ensuring specific cell states were not carried over across roasts.

The pseudo-code for the training process is as follows:

The performance of the LSTM network was evaluated by comparing the corrected predictions (initial grey-box model predictions plus predicted residuals) with the actual Tc values. The evaluation metrics included R-squared (R^2) values and visual inspection of the predicted vs. actual temperature profiles. R^2 values were calculated for individual trajectories and overall to assess the model's accuracy. The corrected predictions showed significant improvements in R^2 values compared to the initial grey-box model. The corrected predictions and actual Tc values were plotted for visual inspection, demonstrating that the LSTM network effectively captured the residuals, resulting in predictions that closely matched the actual data.

Overall, the LSTM-based residual modeling approach successfully improved the accuracy of the coffee roasting model. By capturing the higher-order dynamics and dependencies not accounted for by the initial grey-box model, the LSTM network provided a more reliable and accurate representation of the roasting process. The results validated the effectiveness of combining physics-based models with data-driven approaches to enhance the modeling of complex systems.

Data: List of training trials $T = \{D_1, D_2, \ldots, D_M\}$, where each trial $D_j = \{(x_i, y_i)\}_{i=1}^{N_j}$ consists of N_j training samples.
Result: Trained LSTM parameters θ
Initialization: Initialize the LSTM parameters θ and the Adam optimizer.
for *epoch = 1 to max_ epochs* **do**
  Shuffle the order of trials in T.
  for *each trial D_j in the shuffled trial list T* **do**
    Reset the accumulated loss: $\mathcal{L}_{total} \leftarrow 0$.
    Extract the state data and control data from trial D_j.
    Initialize the initial state s_0 based on the first data point in D_j.
    for *each time step $t = 1$ to $N_j - 1$* **do**
      Perform forward propagation to predict the next state derivative: $\hat{s}_{t+1} = s_t + f(s_t, u_t; \theta)$, where u_t is the control input at time step t.
      Compute the residuals between predicted and actual Tc: $r_t = y_{t+1}[Tc] - \hat{s}_{t+1}[Tc]$.
      Feed current state and control inputs to LSTM to predict residual: $\hat{r}_t = \text{LSTM}(s_t, u_t; \theta)$.
      Compute the loss for the residual: $\mathcal{L}_t = \text{MSE}(r_t, \hat{r}_t)$.
      Accumulate the loss: $\mathcal{L}_{total} \leftarrow \mathcal{L}_{total} + \mathcal{L}_t$.
      Update the state: $s_t \leftarrow \hat{s}_{t+1}$.
    end
    Compute the gradients of the total trial loss: $\nabla_\theta \mathcal{L}_{total}$.
    Update the parameters using the Adam optimizer: $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}_{total}$.
  end
end

Algorithm 2: Pseudo Code Training Algorithm for Residual LSTM

Chapter 4

Results and Discussion

4.1 Overview

This chapter presents the results of the coffee roaster modeling using black-box, white-box and grey-box approaches. It evaluates the performance of the Neural ARX (NARX) model, highlights the challenges encountered, and discusses the enhanced results achieved with the proposed grey-box model. This chapter also includes a comparative analysis of the different models and provides insights into their generalization capabilities.

4.2 Black-box Model Performance

The first model to be explored was a black-box approach using a NARX model was employed to predict the coffee roasting process and methods used are detailed in Section 3.2. These experiments showed that while the NARX model achieved high R-squared values and low RMSE during one-step-ahead predictions, its performance degraded significantly when used for autoregressive predictions. This was emphasized by showing that a naive model, which assumed no change in bean probe temperature, performed comparably. Moreover, when tested for generalization to other control regimes, such as fixing the control inputs to low temperatures, the predictions became very unrepresentative, highlighting the limitations of the one-step-ahead approaches.

4.3 White-box Approach

The experiments detailed in Section 3.3.1 evaluated popular white-box, physics-based approaches for modeling roasting chambers developed by Schwartzberg [8] and Vosloo [10].

Schwartzberg's model assumes constant air flow and rotational speed, optimizing parameters to fit experimental data. Vosloo's model, on the other hand, uses dimensionless parameters like Reynolds, Nusselt, Prandtl, and Biot numbers to estimate the effective heat transfer coefficient (h_e). Both studies assumed a time-invariant inlet temperature and constant mass airflow rates. Neither explicitly addressed the effects of bean movement within the drum.

Section 3.3.1 specifically examined the accuracy of these models when input parameters such as inlet temperature, tumbling speed, and mass flow rates varied throughout the roast. This variability was crucial because the goal was to develop a model capable of predicting changes caused by adjustments in these control inputs, facilitating controller simulation and design. While both Schwartzberg's and Vosloo's methods produced temperature trajectories that qualitatively matched the general phases of a coffee roast, their quantitative fits were poor, with R-squared values of -0.35 for Schwartzberg's model and -1.84 for Vosloo's model.

These results suggest that the assumptions inherent in these models were not sufficiently accurate for the dynamic conditions of real coffee roasting processes and that the models may be missing some critical dynamics, such as the variable effects of drum rotation speed and bean movement on the heat transfer. Trying to incorporate and identify these dynamics provided the motivation for developing more sophisticated grey-box models.

4.4 Grey-box Model Approach

To address the limitations of the black-box and white-box approaches, a grey-box model was developed. This model integrates the empirical and physical relations established by Schwartzberg and Vosloo with a neural network. The neural network was embedded within the system of differential equations to model the effective heat transfer coefficient (h_e), which varies with air flow rate, bean movement, and properties of the roasting air.

In Section 3.3.2, the methodology for integrating a neural network into the system of differential equations established by Schwartzberg was explained. In summary, the heat transfer coefficient, h_e, could not be directly measured from operational data and was a complex function dependent on many factors. This complexity made it challenging to model using purely white-box or black-box approaches. Therefore, a neural network was created to represent h_e, and both the parameters of this neural network and certain parameters in the physics-based model were simultaneously learned by minimizing the accumulated loss between the predicted and actual values of the bean probe temperature. This approach was applied to two datasets: one synthetically generated and one based on actual production data.

The model trained on the synthetic dataset resulted in a R^2 of 0.945 when calculated

using a test set of about 50 roasts. Figure 4.1 illustrates an example of the model predictions.

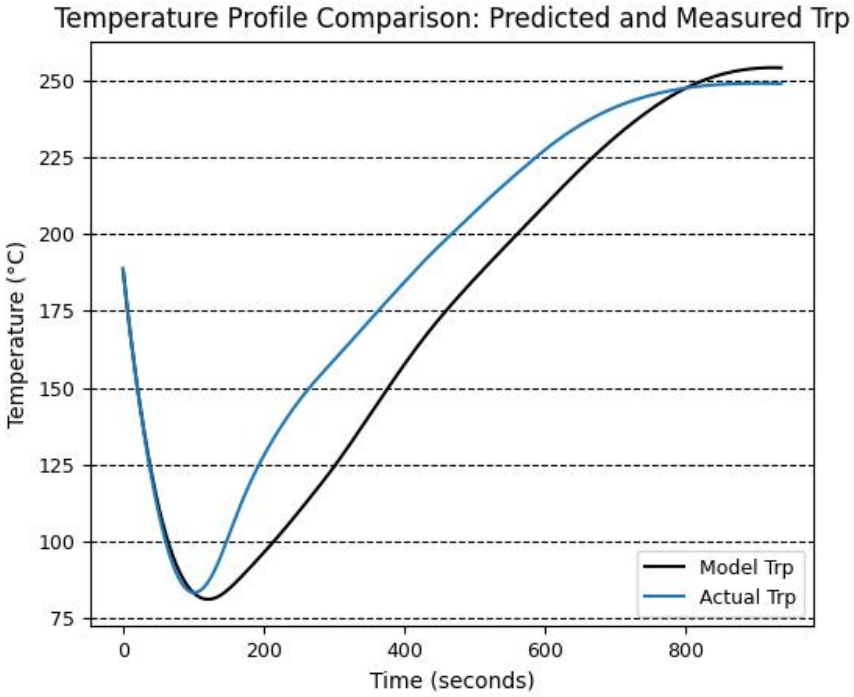


Figure 4.1: Predicted and actual bean probe temperature using synthetic data

When the model is tested in a similar manner to the black-box model in Section 3.2, using low-temperature inputs, the predictions remain accurate, as shown in Figure 4.2. This contrasts with the black-box model, where predictions were unrepresentative in a similar scenario that significantly deviated from the production data used for training. This provides evidence that the model can generalize well and still accurately represent the real dynamics.

The model trained on the production data resulted in an R^2 of 0.887 when the test set was used. Figure 4.3 illustrates an example of the models prediction for a roast.

Comparing the results from the two different data sources reveals that the synthetic data yields a better R^2 value. This outcome is expected, as the model used to generate the synthetic data and the model used to predict the data share the same underlying physics equations. This remains a worthwhile test, however, as it isolates the model's ability to learn the relation for h_e, given that other elements of the data generation and prediction processes are identical. This demonstrates the model's capabilities of modelling h_e without the influence of higher-order dynamics or noise.

Although the predictions made with the model trained on production data were lower than those from the model trained on the synthetic dataset, they were still quite good.

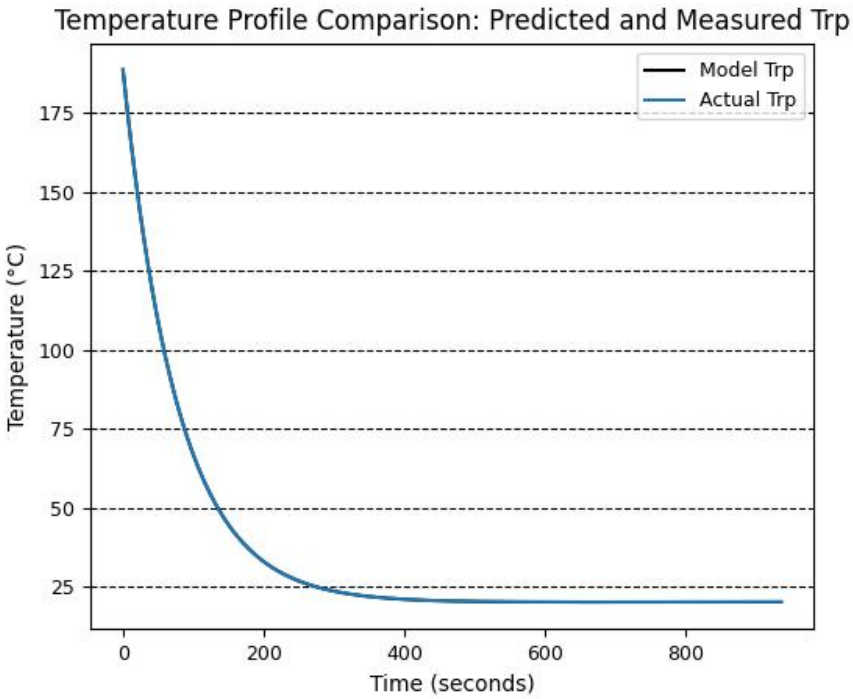


Figure 4.2: Model prediction for 20°C inlet air

However, it was found that adding an LSTM-based neural network to estimate the residuals significantly increased model performance. This enhancement potentially compensated for inaccuracies in parameter estimation and captured other higher-order dynamics. When trained and tested on the same production dataset, the overall R^2 value increased from 0.887 to 0.987. This improvement is illustrated by an example in Figure 4.4.

Interestingly, there are some fluctuations in the model. This might represent uncertainty due to the noise in the measurements. However, the general trend is that the new predictions are pulled closer to the actual values. Here, it might be helpful to use some smoothing function to make things more representative. Another approach that could be further considered is making a probabilistic RNN and using this as a measure of uncertainty with robust control techniques.

Table 4.1 summarizes the models and their performance.

The grey-box models developed in this study shows promise but also highlights several areas for further research. One key limitation is that the model is trained specifically for a single mass of beans and a particular drum size. The movement of the beans, which significantly influences the roasting process, depends on the drum fill level, as documented by Cristo et al. [24]. Consequently, this approach may not generalize well to other masses or

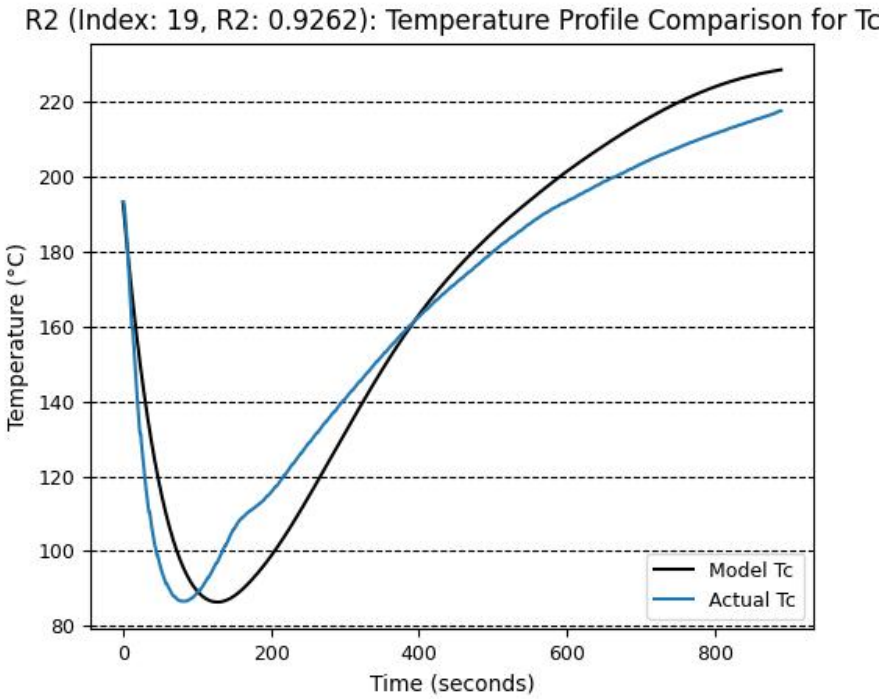


Figure 4.3: Predicted and actual bean probe temperature using production data

Table 4.1: Performance Summary of Different Models

Model	**Dataset**	**R-squared Value (R^2)**
NARX Black-box Model (one-step-ahead/autoregressive)	Production Data	0.99/-11.92
Schwartzberg White-box Model	Experimental Data	-0.35
Vosloo White-box Model	Experimental Data	-1.84
Grey-box Model (Synthetic)	Synthetic Data	0.945
Grey-box Model (Production)	Production Data	0.887
Grey-box Model with Residuals	Production Data	0.987

different drum geometries. Future research should focus on developing models that can adapt to various drum sizes and fill amounts, enhancing the model's versatility and applicability across different roasting setups.

Collecting more production data with known parameters for the initial bean mass and the general geometries of the roaster is another important step. This data would help verify the method used in this study, which was tested primarily on synthetic data. Such empirical validation is crucial for confirming the model's accuracy and robustness in real-world conditions.

Additionally, the ultimate goal of this research is to develop a model that can be used for simulating and designing controllers for the entire roasting process. While this study

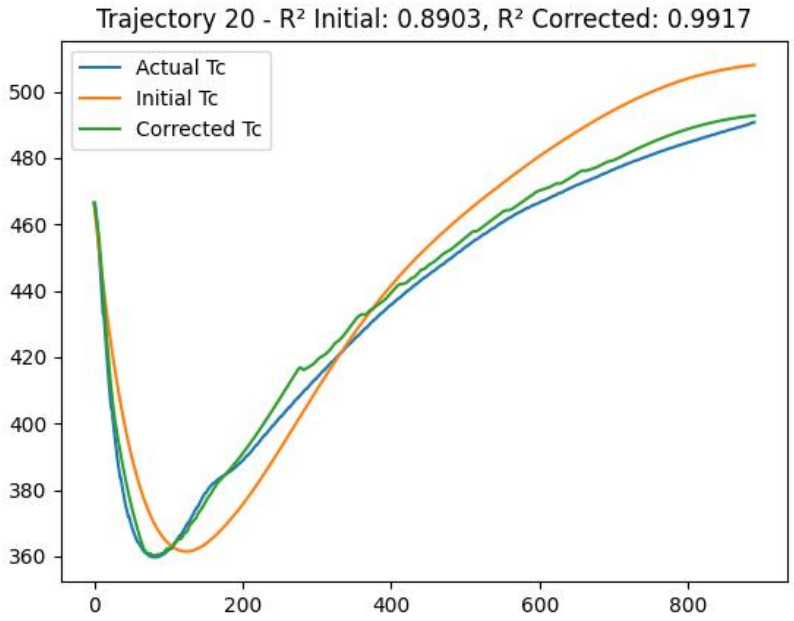


Figure 4.4: Example trajectory of predictions of the grey-box model without and with residual neural network.

focused on modeling the roasting chamber with controlled inputs such as inlet temperature, drum speed, and mass flow rates, practical applications require considering the entire system dynamics. For example, lower-level controllers, such as those managing the furnace and the valves that control recirculated and ambient air in an R-type roaster, play a significant role, and the dynamics involved needs to be considered. Future models should incorporate these elements to create a comprehensive system capable of designing and optimizing controllers for the entire roasting process.

Moreover, exploring more efficient training methods and architectures could help overcome the difficulties encountered with the black-box models and improve the resource requirements for the grey-box models. Techniques such as the adjoint method used in neural ordinary differential equations (NODEs) [41] might offer a way to reduce computational costs and enhance the efficiency of the training process.

While this study has made significant strides in modeling the coffee roasting process, addressing these areas in future research would build on the current research and lead to a robust, adaptable, and practical model that can be applied in the coffee industry.

Overall these results show that the model is able to learn a function for h_e that can consider variable drum rotation speeds, air velocities, and inlet temperatures which builds off of and improves the methods proposed by Schwartzberg [8] and Vosloo [10]. Instead of

assuming constant rotational speeds and air flow rates, the model is able to account for these within the physical equations and provide acceptable accuracy. However, There are some caveats to this process. The first is that the training process is very slow, as it can only take one optimizer step per iteration of solving the entire system of differential equations for an entire roast. This was a limitation on hyper-parameter tuning, and complexity of the neural network. A larger more complex neural network would add time for the forward and backward passes, and would likely require more epochs for training, which would add significantly more time. Possibly using the adjoint method to calculate gradients, as done in neural ordinary differential equations (NODEs) [41], could add efficiency to this process. However, this and other areas still have potential for future research.

Chapter 5

Conclusion and Future Work

This thesis explored the application of machine learning techniques, specifically neural networks, to model the complex and nonlinear dynamics of coffee roasting. The overarching goal of this research was to develop a model capable of accurately simulating the roasting process under varying control inputs, thus laying the groundwork for improved control system design and optimization.

Initial explorations utilizing black-box approaches, specifically a Neural NARX model, aimed to capture the system dynamics solely from data. While initial results appeared promising, with the model achieving high performance metrics during one-step-ahead predictions, further investigation revealed critical shortcomings. A simple naive model, which assumed no change in the bean probe temperature from one time step to the next, achieved surprisingly similar performance metrics. This suggested that the NARX model might not be effectively learning the subtle dynamics of the roasting process. This suspicion was confirmed when the model was used for autoregressive predictions, where its performance degraded significantly, and when tested for generalization to different control regimes, where it produced unrealistic results. These findings underscored the some challenges of purely data-driven approaches for this specific problem and highlighted the importance of incorporating domain knowledge into the modeling process. Faced with these challenges, two paths forward emerged: either develop significantly more complex neural network architectures, which might prove infeasible given resource constraints, or leverage the established understanding of the underlying physical principles governing the roasting process. This latter path, incorporating physical principles into the model, formed the basis for the subsequent exploration of grey-box modeling.

Traditional white-box models, as proposed by Schwartzberg [8] and Vosloo [10], were also evaluated. These models, grounded in fundamental heat transfer principles, provided a valuable framework for understanding the heat transfer mechanisms within the roasting chamber.

However, their reliance on simplified assumptions and constant parameters ultimately limited their accuracy when applied to real-world production data with varying control inputs. The models struggled to adequately capture the dynamic interactions between parameters like inlet temperature, tumbling speed, and mass flow rates, leading to noticeable deviations from actual temperature profiles.

To overcome the shortcomings of both black-box and white-box approaches, this thesis proposed a novel grey-box modeling approach. This hybrid methodology seeks to leverage the strengths of both physics-based and data-driven methods. At its core, the model utilizes the differential equations derived from heat transfer principles to represent the core dynamics of the roasting chamber. Recognizing the complexity and dynamic nature of the effective heat transfer coefficient (h_e), the model incorporates a neural network to learn this crucial parameter as a function of air velocity, bean movement, and bean mass. This integration allows for a more nuanced and accurate representation of the heat transfer processes occurring within the roaster. The results presented in this thesis clearly demonstrate that the grey-box model significantly outperforms both the black-box and white-box approaches in accurately predicting the temperature profiles of the coffee beans during roasting. This enhanced performance, however, comes at the cost of increased computational demands associated with training a neural network within a system of differential equations.

This research makes several key contributions to the field of coffee roasting modeling and control. Firstly, it highlights the limitations of using purely data-driven methods for modeling systems with subtle dynamics, emphasizing the importance of incorporating domain knowledge and physical principles. Secondly, it introduces a novel grey-box model that effectively combines physics-based equations with a neural network to accurately capture the dynamic heat transfer characteristics of the coffee roasting process. This model demonstrably outperforms existing models in terms of prediction accuracy. Lastly, the methodology developed in this thesis provides a general framework for modeling complex, nonlinear processes in various industrial settings, especially where traditional modeling approaches struggle to achieve sufficient accuracy.

Despite these advancements, there remain several limitations and potential avenues for future research. One key challenge is the model's dependence on high-quality training data. The accuracy of the model relies on having data that adequately represents the operating range of the roaster. Future work should explore methods to improve the model's robustness and generalization capabilities, particularly when data is limited or incomplete. Additionally, the computational cost of training the grey-box model is a significant consideration. Investigating more efficient training algorithms and architectures, such as those based on neural ODEs [41], could significantly reduce the computational burden. Furthermore, the current

model is trained for a specific roaster type and bean mass. Future studies should investigate methods to adapt the model to different roaster designs, bean varieties, and roasting conditions, enhancing its versatility and practical applicability. Finally, expanding the model to encompass the entire roasting system, including the burner, airflow control mechanisms, and other components, would facilitate the design and optimization of comprehensive control strategies for the entire roasting process.

In conclusion, this research demonstrates the potential of integrating machine learning techniques with physics-based models to address the challenges of modeling and controlling complex nonlinear processes. The developed grey-box model for coffee roasting provides a promising avenue for enhancing the consistency, quality, and efficiency of coffee production. Future research building on this foundation has the potential to revolutionize the coffee industry by enabling the development of intelligent and adaptive control systems capable of consistently producing high-quality coffee while optimizing resource utilization. Moreover, the principles and techniques presented in this work can be applied to a wide range of industrial applications, advancing the field of process control and paving the way for more efficient and sustainable manufacturing processes.

Appendix A

Thermal Properties for Roaster Air and Coffee Beans

This appendix details the thermal properties used in the model that was created to generate synthetic data, or called the "Representative Roaster Model". These relations were experimentally derived and fitted by Vosloo [10].

A.1 Heat Capacity of Roaster Air

The following relation was determined by Vosloo as accurate for fitting the heat capacity of air over the temperature of the air that may be experienced in a coffee roaster.

$$\begin{aligned} c_{pg}(T_g) = {} & 5.3091 \times 10^{-17} T_g^6 - 4.1550 \times 10^{-13} T_g^5 \\ & + 1.3621 \times 10^{-9} T_g^4 - 2.3267 \times 10^{-6} T_g^3 \\ & + 2.1034 \times 10^{-3} T_g^2 - 7.2075 \times 10^{-1} T_g \\ & + 1.0839 \times 10^{3} \end{aligned} \tag{A.1}$$

Where T_g is in Kelvin (K), and C_{pg} is in J/(kg K). The fit of the equation and the experimental data is shown by Vosloo in A.1

A.2 Density of Roaster Air

The following relationship for roaster air density was determined experimentally by Vosloo [10]:

$$\rho_g = 353.34 \times T_g^{-1.002} \tag{A.2}$$

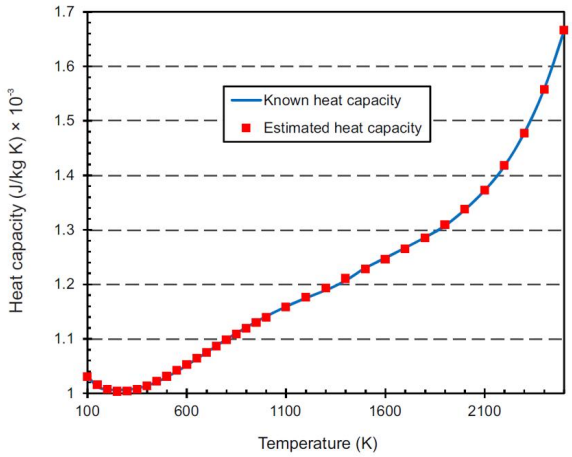


Figure A.1: Experimental data collected by Volsoo for roaster air heat capacity and fitted function outputs. Figure taken from [10]

Where T_g is in degrees Kelvin, and the density is in kg per meters cubed. The fit of the function to the experimental data is shown in figure A.2

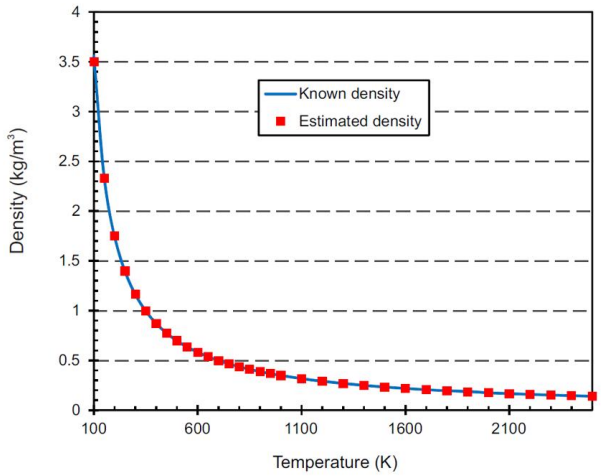


Figure A.2: Experimental data collected by Volsoo for Roaster air density and fitted function outputs. Figure taken from [10]

A.3 Viscosity of Roaster Air

Equation A.3 is the relation used by Vosloo [10] for the viscosity of the roasting air. This relation was developed by empirical observation and a polynomial fit to the data, as can be

seen in Figure A.3.

$$\begin{aligned}\mu_g =& 1.2184 \times 10^{-24}[T_g^6] - 8.1123 \times 10^{-21}[T_g^5] + 1.6089 \times 10^{-17}[T_g^4] \\ &+ 1.1460 \times 10^{-15}[T_g^3] - 3.9733 \times 10^{-11}[T_g^2] \\ &+ 7.1226 \times 10^{-8}[T_g] + 4.8855 \times 10^{-7}\end{aligned} \tag{A.3}$$

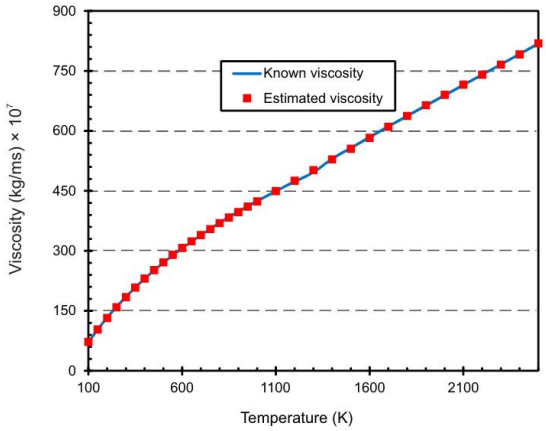


Figure A.3: Experimental data collected by Volsoo for Roaster air viscosity and fitted function outputs. Figure taken from [10]

A.4 Thermal Conductivity of Roaster Air

Equation A.4 is the relation used by Vosloo [10] for the viscosity of the roasting air. This relation was developed by empirical observation and a polynomial fit to the data, as can be seen in Figure A.4.

$$\begin{aligned}\lambda_g =& 1.3819 \times 10^{-20}[T_g^6] - 9.1506 \times 10^{-17}[T_g^5] + 2.2342 \times 10^{-13}[T_g^4] \\ &- 2.2872 \times 10^{-10}[T_g^3] + 6.8867 \times 10^{-8}[T_g^2] \\ &+ 8.0128 \times 10^{-5}[T_g] + 7.6694 \times 10^{-4}\end{aligned} \tag{A.4}$$

A.5 Heat Capacity of Coffee Beans

Equation A.5 is the relation developed by Putranto et. al. for the heat capacity of a coffee bean [34].

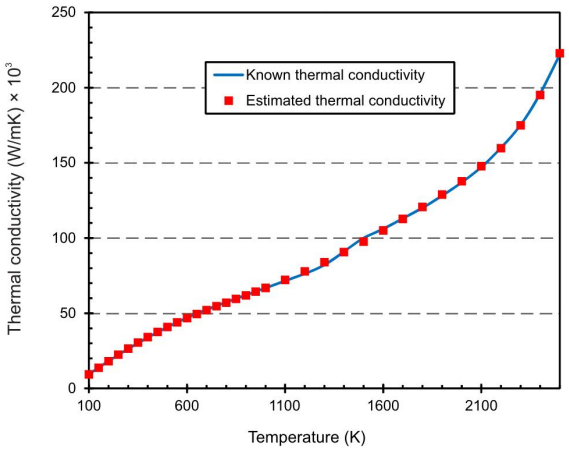


Figure A.4: Experimental data collected by Volsoo for Roaster air thermal conductivity and fitted function outputs. Figure taken from [10]

$$cp_b(T_b, X_b) = 10^3(1.674 + 2.51\left(\frac{X_b}{1 + X_b}\right)) \tag{A.5}$$

A.6 Thermal Conductivity of Coffee Beans

Equation A.6 is the relation proposed by Hernandez and used by Vosloo for modelling the heat transfer in a drum roaster [10]. It is a function of both the bean temperature and the moisture content of the bean.

$$cp_b(T_b, X_b) = 10^3(1.674 + 2.51\left(\frac{X_b}{1 + X_b}\right)) \tag{A.6}$$

Appendix B

Model Derivations

This appendix details the derivations of the physics based-models investigated in this paper and used by Vosloo and Schwartzberg as well as the relation for h_e used in generating a synthetic dataset to test the proposed grey-box's neural network.

B.1 Drum Model Derivation

The complete system of equations for the drum roasting chamber is described in Equations 3.2 through 3.6 in chapter 3. This section goes through the derivations and assumptions associated with these equations. In general, this Derivation primarily consists of a heat balance over the coffee bean mass, a heat balance looking at the volume of air in the roaster drum and equations governing the moisture content and exothermic reactions taking place inside the coffee beans. Starting with the heat balance of the coffee bean mass, according to Vosloo [10], general heat balance over the coffee beans can be written as follows:

$$m_b C_{p_b} \frac{dT_b}{dt} = \Phi_{gb} - \Phi_{gm} + \Phi_{mb} + \Phi_r - \Phi_{ev} \tag{B.1}$$

where m_b is the mass of the coffee beans, C_{p_g} is the heat capacity of the coffee beans, $\frac{dT_b}{dt}$ is the rate of temperature change of the coffee beans, Phi_{gm} is the heat transfer from the hot gas in the roaster to the coffee beans, Phi_{gm} is the heat transfer from the metal to the beans, Phi_r is the heat transfer due to exothermic reactions inside the beans, and Phi_{ev} is the heat loss due to evaporation of the moisture in the coffee beans. Vosloo further simplified this with the following assumptions:

- Negligible heat transfer from roasting air to metal. This assumption is supported by Schwartzberg for roasters where heat is not applied directly to the outside of the drum

as is the case in our roaster [8]. Hernandez et. Al also used this assumption in their analysis and justified it with the fact that roasters are preheated, and so heat transfer to the metal will be very small [36].

- Negligible heat transfer from beans to the metal. This assumption is made by Bottazzi et al. and Hernandez et al., justifying it by the fact that the contact area and time of a bean is very small as they are tumbling in the roaster [36, 38]

After these terms are dropped, the equation becomes:

$$m_b C_{p_b} \frac{dT_b}{dt} = \Phi_{gb} + \Phi_r - \Phi_{ev} \tag{B.2}$$

Terms are then rearranged to give the following in terms of the change in bean temperature:

$$\frac{dT_b}{dt} = \frac{\Phi_{gb} + \Phi_r - \Phi_{ev}}{m_b C_{p_b}} \tag{B.3}$$

The expressions for the heat transfer between the gas and the bean can be described by change in heat of the air between the inlet and the outlet of the drum:

$$\Phi_{gb} = G_g C_{pg} [T_{gi} - T_{go}] \tag{B.4}$$

Where G_g is the mass flow rate of the hot air in the roaster, C_{pg} is the heat capacity of the air, ant T_{gi} and T_{go} are the temperature of the air at the inlet and the outlet respectively. Here, it is important to note that the heat capacity of air changes drastically depending on the moisture content of the air, and the temperature, which makes this term very complex. Thankfully, however, Vosloo has experimentally determined the a heat capacity function for roaster air, as well as other properties such as density, thermal conductance, viscosity. Thermal properties for the beans were experimentally determined as well and are used this paper's model. These relations can be found in Appendix A for details about the calculations. Importantly for the current derivation, is that C_{pg} is only a function of temperature. Our approach is a fixed volume analysis to simplify the model, and the properties of the air are all evaluated at the average value between the inlet temperature, T2, and the outlet temperature, T3. This we call Tg, or gas temperature. The expression for the heat generated by the exothermic reaction is described by Vosloo and Schwartzberg to be the following [8, 10]:

$$\Phi_r = Q_r m_{b,d.b.} \tag{B.5}$$

Where Q_r is the rate of heat generated by the reaction for a given mass of dry coffee beans, $m_{b,d.b.}$. The expression of Q_r will be detailed later using an Arrhenius-based differential equa-

tion to model how quickly the reactions carry out as a function of the bean temperature [8]. This is expressed mathematically as Equation B.6.

$$Q_r = A_r \frac{H_{et} - H_e}{H_{et}} \exp \frac{-\Delta E}{RT_b} \tag{B.6}$$

Where A_r is the Arrhenius equation pre-factor combined with a coefficient representing the amount of heat generated per unit amount of reaction so that the units are kJ/ kg s on a dry coffee basis. ΔE is the reaction activation energy, R is the ideal gas constant, and T_b is the real temperature of the beans in Kelvin, H_{et} is the total amount of reaction heat produced per kilogram of dry coffee beans, and H_e is the amount of reaction heat that has been generated up to the current time. The rate of the reaction, $\frac{dH_e}{dt}$ is the same as the heat transfer rate, Q_r. This makes B.6 the same as 3.4. There are several assumptions made in the formulation of this relation. First, it assumes that the rate of heat generated is proportional to the rate of reactions inside the coffee beans, reaction rates are proportional to reactant concentrations, reactants are consumed in the reaction, and that the coefficients governing the relation, A_r, H_{et}, and ΔE are constants that are time-invariant, and not functions of temperature, pressure, or any other external factors. The last term in the Equation B.3, is the heat consumed by the latent heat of vaporization, caused by evaporating the moisture contained in the coffee bean into steam. It is expressed as Equation B.7.

$$\Phi_{ev} = \Delta H_v \left(-\frac{dX_b}{dt} \right) m_{b,d.b.} \tag{B.7}$$

Where Φ_{ev} is the heat transfer consumed by evaporating the moisture from the coffee beans, ΔH_v is the enthalpy of vaporization, $\frac{dX_b}{dt}$ is the change in moisture content in the coffee beans, and $m_{b,d.b.}$ is the mass of the coffee beans on a dry basis. Combining and rearranging equations B.3, B.4, B.5, B.6, and B.7 gives equation 3.2.

The differential equation for the change in the moisture content in the bean, $\frac{dX_b}{dt}$, is taken from [8] and given by the Equation B.8.

$$\frac{dX_b}{dt} = -\frac{4.32 \times 10^9 X_b^2}{(d_p \times 10^3)^2} \exp \frac{-9889}{T_b} \tag{B.8}$$

This is an empirical relation developed by looking at roasting bean data, and assuming that moisture loss was diffusely regulated and governed by an Arrhenius-type equation, with the diffusion rate and mass transfer both proportional to the moisture content, X_b. The data was only for beans with an effective average diameter, d_p of 6 mm, which may be as source of error, if the diameter of the coffee beans vary significantly from this. This is the same as

equation 3.3.

Next to consider is a heat transfer with the volume of air in the roasting chamber. Here several assumptions are made. First it is assumed that the airflow through roaster is uniform, and the the flow rate is low enough that we can approximate it as being incompressible. It is assumed as well that all thermophysical properties of air are functions of the air temperature [36] and all other heat transfer other than that between the gas and the beans are neglected. This assumption assumes that the metal either mostly has contact with the beans, and that the transfer of heat to the air is small, since the metal is preheated, the temperature differences between the air and the metal should be small as well. using these assumptions, a heat balance equation over the air can be written as Equation B.9 [8].

$$-G_g C_{pg} \left(\frac{dT_g}{dz} \right) = h_e \left(\frac{dA_b}{dz} \right) (T_g - T_b) \tag{B.9}$$

Where G_g is the mass flow rate of the air, C_{pg} is the heat capacity of the air, $\frac{dT_g}{dz}$ is the change of the roaster air temperature with respect to the axial direction of the drum and the average velocity of the air, h_e is the effective heat transfer coefficient, $\frac{dA_b}{dz}$ is the change in the bean surface area with respect to the drum axial direction, T_g is the temperature of the roaster air, and T_b is the temperature of the beans. In the roaster, the temperature of the beans and the air is a function of z, where the beans get cooler towards the outlet of the drum as the heat is transferred as it moves through the drum.

Equation B.9 can be rearranged to give Equation B.10 and then integrated over the inlet and outlet temperatures to get Equation B.11 [10].

$$\int_{T_{g,i}}^{T_{g,o}} \frac{1}{(T_g - T_b)} dT_g = -\frac{h_e A_b}{G_g C_{p_g}} \tag{B.10}$$

$$(T_{g,o} - T_b = (T_{g,i} - T_b) \left(\exp \left[-\frac{h_e A_b}{G_g C_{p_g}} \right] \right) \tag{B.11}$$

Equation B.11 can be rearranged to give the outlet roaster air temperature in Equation B.12.

$$T_{go} = T_{gi} - (T_{gi} - T_b)(1 - \exp - \frac{h_e A_b}{G_g C_p g}) \tag{B.12}$$

This is the same equation as equation 3.6. h_e is the heat transfer coefficient, as discussed previously. Schwartzberg proposes that this be considered constant, which can be a good assumption if the air flow rate, and the tumbling speed of the drum does not change drastically [8]. Volsoo proposes that this value be estimated by using the Biot number and the

Ranz-Marshall correlation, as shown in equations B.13 through B.18 [10].

$$h_e = \frac{h}{1 + 0.3Bi} \tag{B.13}$$

$$B_i = \frac{hd_b}{\lambda_b} \tag{B.14}$$

$$Nu = 2 + 0.6Re^{1/2}Pr^{1/3} \tag{B.15}$$

$$Nu = \frac{hd_b}{\lambda_g} \tag{B.16}$$

$$Re = \frac{v_g d_b \rho_g}{\mu_g} \tag{B.17}$$

$$Pr = \frac{C_{pg}\mu_g d_b}{\lambda_g} \tag{B.18}$$

Here h is the heat transfer coefficient, Bi is the Biot number, Nu is the Nusselt number, Re is Reynolds number, and Pr is the Prandtl number. The dimensionless numbers - Biot, Nusselt, Reynold, and Prandtl - rely on several properties of the roasting air: λ_g - thermal conductivity, rho_g - density, and μ_g - viscosity. Relations for the properties of roaster air were determined experimentally by Vosloo and are summarized in Appendix A [10]. The equation that expresses the effective heat transfer coefficient is obtained by combining Equation B.13 through Equation B.18 and rearranging to obtain Equation B.19.

$$h_e = \frac{\lambda_b \left(0.6\lambda_g \left(\frac{c_{pg}u_g}{\lambda_g} \right)^{1/3} (d_b \rho_g v_g)^{1/2} + 2 \right)}{d_b \left(0.6d_b + \lambda_b + 0.18\lambda_g \left(\frac{c_{pg}\mu_g}{\lambda_g} \right)^{1/3} \left(\frac{d_b \rho_g v_g}{\mu_g} \right)^{1/2} \right)} \tag{B.19}$$

Finally, there are dynamics associated with the roaster profiles. The beans are dropped into the roaster at roughly ambient temperatures, and the roaster will be preheated to roasting temperatures, which means the temperature of the profile bean probe sensor will not be the same as the bean temperature, especially at the beginning of the roast. Schwartzberg, proposed the linear differential equation in Equation B.20.

$$\frac{T_{rp}}{dt} = K(T_b - T_{rp}) \tag{B.20}$$

Here T_{rp} is the temperature read by the bean probe, T_b is the temperature of the beans, and K is the roast profile constant. This is an empirical constant that Vosloo and Schwartzberg proposed determining it by data fitting and regression [8, 10]. This is the same as Equa-

tion 3.5.

B.2 Synthetic Data for Effective Heat Transfer Coefficient

The data generated to train the neural network for the proposed grey-box was generated synthetically. To do this, a function for h_e was developed and defined as in Equations B.21, B.22, and B.23.

$$h_c = C_1 Re(\omega)^m Pr^n \left(\frac{\lambda_g}{d_b}\right) \tag{B.21}$$

$$h_k = k_{contact}(AA + BB\omega + CC\omega^2 + DD\omega^3 + EE\omega^4) \tag{B.22}$$

$$h_e = \frac{1}{\frac{1}{h_c} + \frac{1}{h_k}} \tag{B.23}$$

The effective heat transfer coefficient (h_e) between the coffee beans and their surroundings within the rotating drum roaster is modeled as a combination of two primary heat transfer mechanisms: convection (h_c) and conduction (h_k).

Convective Heat Transfer Coefficient (h_c) is estimated using a common empirical correlation (Equation B.21), which relates h_c to the Reynolds number (Re), Prandtl number (Pr), thermal conductivity of the gas (λ_g), and the bean diameter (d_b). The parameters C_1, m, and n. Importantly, the Reynolds number has been changed from what it was before, to be a function of the rotational speed of the drum.

Conductive Heat Transfer Coefficient (h_k) is modeled as a polynomial function of the drum's rotational speed (ω) (Equation B.22). This function, with coefficients AA, BB, CC, DD, and EE, accounts for the varying contact area and pressure between the beans and the drum wall as the drum rotates. The parameter $k_{contact}$ represents the contact conductance between the beans and the drum surface.

Effective Heat Transfer Coefficient (h_e): The overall effective heat transfer coefficient (h_e) is calculated using a thermal resistance analogy (Equation B.23). This equation assumes that the convective and conductive heat transfer pathways act in parallel, with their thermal resistances being additive.

The values of AA, BB, CC, DD, EE, $k_{contact}$, C_1, m, n, K, A, B, C, D were all determined by regression. The initial mass of the coffee beans, m_{bd} was guessed, and the surface area, A_b, was determined using the values measured by Schwartzberg [8]. The data

used for regression comprised 50 roasting trials. Although the details of the mass of the beans and duct geometry required to convert average air velocity to mass flow were consistent across trials, these values are not precisely known and were arbitrarily chosen.

This approach does not guarantee that the model accurately represents the data or generalizes well to other conditions. However, this is not the primary intent. The goal is to create a hypothetical system that is somewhat representative of a coffee roaster, not necessarily the specific roaster from which the data was obtained. This allows for the generation of synthetic data to evaluate how well a neural network can learn the effective heat transfer coefficient as described earlier. The results of this approach are discussed in Chapter 4.

Appendix C

Production Coffee Roasting Dataset

This appendix provides a detailed overview of the production coffee roasting dataset used in this study. Understanding the characteristics and potential limitations of this dataset is crucial for interpreting the performance and generalizability of the black-box models developed and described in Chapter 3.

Table C.1 presents basic descriptive statistics summarizing the key variables in the dataset. Each variable represents a specific measurement or control input related to the coffee roasting process. A comprehensive description of each variable and its role in the roasting process can be found in [cross-reference to relevant chapter/section]. Notably, the dataset encompasses a substantial 3898522 data points from 4285 roasts, providing a robust foundation for training and evaluating our models.

Table C.1: Statitics of Production Roast Data and Variables

Index	Actual Value Tc [°C]	Actual Value T1 [°C]	Actual Value T2 [°C]	Flow Gas BF [%]	Present Value VAC [% * 0.1]	Actual Value TT [rpm]	Actual Value Closing VAF [%]	Actual Value Opening VAT [%]	Gas Pressure BF [mbar]	Air Speed Meter fSCP [m/sec]	Set % Command BF [%]	Setpoint VAC [% * 0.1]	Setpoint Drum Speed TT [% * 0.1]	Setpoint Closing VAF [%]	Setpoint Opening VAT [%]
mean	163.0	574.3	357.0	47.48	534.1	28.66	56.01	69.87	138.7	10.25	31.81	53.27	67.76	55.96	70.33
std	40.42	99.00	100.0	14.44	134.2	4.225	34.99	31.56	7.558	3.259	17.69	13.49	11.65	35.94	33.05
min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25%	131.2	511.0	284.0	40.00	452.0	28.70	23.00	50.00	136.0	8.000	22.00	45.00	68.00	22.00	51.00
50%	174.3	574.0	358.0	47.00	531.0	29.80	59.00	83.00	140.0	9.000	31.00	53.00	71.00	58.00	84.00
75%	197.0	658.0	445.0	58.00	603.0	30.20	95.00	97.00	142.0	13.00	45.00	60.00	72.00	94.00	98.00
max	229.8	796.0	553.0	100.0	932.0	40.90	100.0	100.0	162.0	21.00	100.0	92.00	180.0	100.0	100.0

Figure C.1 illustrates the bean probe temperature profiles for all roasts in the dataset. As evident from the figure, the data includes the preheating phase of the coffee roasters, explaining why numerous variables have minimum values of zero. The non-uniform start of the drying phases further highlights this inclusion.

A key consideration when interpreting the results of this study is the inherent limitations of the dataset, stemming from its origin in a production environment. Under such conditions,

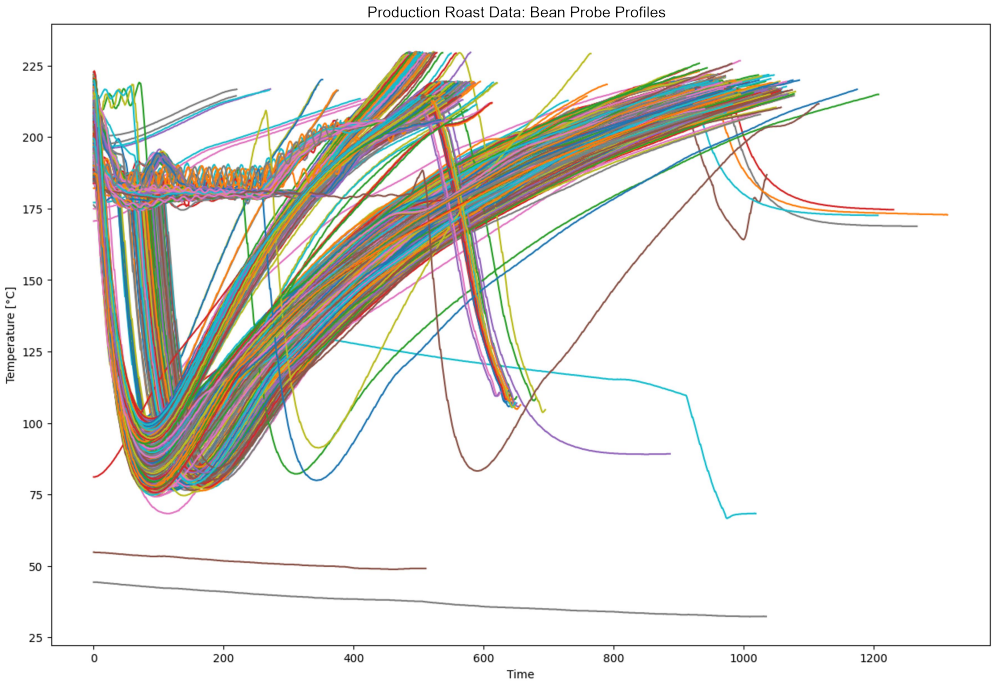


Figure C.1: Bean probe profiles for 4285 production roasts

the coffee roasting process is continually governed by a series of PID controllers. While these controllers play a crucial role in maintaining consistent product quality, their influence introduces a potential bias into the data.

PID controllers function by maintaining process variables within predefined desirable ranges, actively minimizing deviations. This control mechanism, however, may inadvertently mask the true dynamic behavior of the system. The controllers' corrective actions can dampen or obscure natural responses, making it challenging to extract information about the underlying physics or develop models accurately reflecting those dynamics. Moreover, the dataset likely exhibits a scarcity of data points in operational regions far from the desired setpoints. PID controllers inherently prevent the system from entering these less desirable states, hindering the models' ability to learn the system's behavior in such conditions.

Building comprehensive models requires a thorough understanding of the system across a wide range of conditions. This limitation of the dataset emphasizes the importance of integrating domain knowledge and exploring physics-based modeling approaches to complement the data-driven black-box models. By incorporating such knowledge, more robust and generalizable models can be developed, potentially capable of handling a broader range of operating conditions and extrapolating to scenarios not explicitly captured in the production dataset.

References

[1] J. Hoffmann. *The world atlas of coffee: From beans to brewing - coffees explored, explained and enjoyed*. World atlas of. Mitchell Beazley, 2014. ISBN: 978-1-84533-787-2. URL: https://books.google.com/books?id=02JOngEACAAJ.

[2] S. Rao. *The coffee roaster's companion*. Scott Rao, 2014. ISBN: 978-1-4951-1819-7. URL: https://books.google.com/books?id=cvP8oAEACAAJ.

[3] S. Schenker and T. Rothgeb. "The roast—Creating the Beans' signature". In: *The craft and science of coffee*. Elsevier, 2017, pp. 245–271.

[4] O. Mayr. *The Origins Of Feedback Control*. The MIT Press, July 15, 1975. ISBN: 978-0-262-63056-6. (Visited on 04/02/2024).

[5] R. C. Dorf and R. H. Bishop. *Modern control systems*. Fourteenth edition, global edition. Harlow: Pearson, 2022. 1022 pp. ISBN: 978-1-292-42237-4 978-0-13-730727-2 978-0-13-730725-8.

[6] B. Alonso-Torres, J. A. Perez, F. Sierra, S. Schenker, and C. Yeretzian. "Modeling and Validation of Heat and Mass Transfer in Individual Coffee Beans during the Coffee Roasting Process Using Computational Fluid Dynamics (CFD)". In: *Chimia* 67 (Apr. 2013), pp. 291–4. DOI: 10.2533/chimia.2013.291.

[7] IMA. *Use and Maintenance Instruction Manual: Roaster TMR 25-125-250-400-660 CA - GAS*. (Visited on 06/12/2024).

[8] H. Schwartzberg. "Modeling bean heating during batch roasting of coffee beans". In: *Engineering and Food for the 21st Century*. Mar. 2002. ISBN: 978-1-56676-963-1. DOI: 10.1201/9781420010169.ch52.

[9] Z. Adiwijaya. "Revamping Manufacturing Systems: Utilization of Data Driven Models, Interpretable Machine Learning, and Data-Product Stakeholder Flow Analysis". S.M. Thesis. System Design and Management Program: Massachusetts Institute of Technology, June 2023. URL: https://dspace.mit.edu/handle/1721.1/151252.

[10] J. Vosloo. "Heat and mass transfer model for a coffee roasting process". In: 2017. URL: https://api.semanticscholar.org/CorpusID:102743055.

[11] A. Brunerová, A. Haryanto, U. Hasanudin, D. A. Iryani, M. Telaumbanua, and D. Herák. "Sustainable management of coffee fruit waste biomass in ecological farming systems at West Lampung, Indonesia". In: *IOP Conference Series: Earth and Environmental Science* 345.1 (Oct. 2019). Publisher: IOP Publishing, p. 012007. DOI: 10.1088/1755-1315/345/1/012007. URL: https://dx.doi.org/10.1088/1755-1315/345/1/012007.

[12] H.-D. Belitz. *Food Chemistry*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. ISBN: 978-3-540-69933-0 978-3-540-69934-7. DOI: 10.1007/978-3-540-69934-7. URL: http://link.springer.com/10.1007/978-3-540-69934-7 (visited on 05/15/2024).

[13] T. Oberthür, P. Läderach, H. Pohlan, and H. Cock. "Specialty coffee: Managing quality". In: *International Plant Nutrition Institute (IPNI). Penang Malaysia* (2012).

[14] J.-C. Vincent. "Chapter 1: Green Coffee Processing". In: *Coffee: Volume 2: Technology*. Ed. by R. Clarke and R. Macrae. Number: v. 2. Springer Netherlands, 2012. ISBN: 978-94-009-3417-7. URL: https://books.google.com/books?id=NlcGCAAAQBAJ.

[15] A. Putranto and X. D. Chen. "Roasting of barley and coffee modeled using the lumped-reaction engineering approach (l-REA)". In: *Drying Technology* 30.5 (2012), pp. 475–483. DOI: 10.1080/07373937.2011.647185. URL: https://doi.org/10.1080/07373937.2011.647185.

[16] A. S. Franca, J. C. Mendonça, and S. D. Oliveira. "Composition of green and roasted coffees of different cup qualities". In: *LWT-Food Science and Technology* 38.7 (2005). Publisher: Elsevier, pp. 709–715.

[17] Anonymous Coffee Roaster Professional. *Personal interview*. May 2024.

[18] ICC. *Resolution 420: Coffee Quality-Improvement Programme - Modifications*. London, England. URL: https://www.ico.org/documents/iccres420e.pdf (visited on 05/29/2024).

[19] X. Wang and L.-T. Lim. "A kinetics and modeling study of coffee roasting under isothermal conditions". In: *Food and bioprocess technology* 7 (2014). Publisher: Springer, pp. 621–632.

[20] Specialty Coffee Association. *Cupping protocols*. URL: https://sca.coffee/research/protocols-best-practices/cupping-protocols.

[21] R. Clarke, ed. *Coffee: Volume 2: Technology*. Number: v. 2. Springer Netherlands, 2012. ISBN: 978-94-009-3417-7. URL: https://books.google.com/books?id=NlcGCAAAQBAJ.

[22] Energy Efficiency and Conservation Authority. *New technology opportunities for coffee manufacturers: Technology scan*. Technical Report. Wellington, New Zealand: Energy Efficiency and Conservation Authority, May 2024. URL: https://www.eeca.govt.nz/assets/EECA-Resources/Co-funding/Sector-Decarb-Files/Technology-Scan-Coffee.pdf.

[23] H. Schwartzberg. "Batch Coffee Roasting; Roasting Energy Use; Reducing That Use". In: *Food Engineering Series*. Sept. 2013, pp. 173–195. ISBN: 978-1-4614-7905-5. DOI: 10.1007/978-1-4614-7906-2_10.

[24] H. Cristo, M. Martins, L. Oliveira, and A. Franca. "Transverse flow of coffee beans in rotating roasters". In: *Journal of Food Engineering* 75 (July 2006), pp. 145–148. DOI: 10.1016/j.jfoodeng.2005.04.010.

[25] A. Fabbri, C. Cevoli, L. Alessandrini, and S. Romani. "Numerical modeling of heat and mass transfer during coffee roasting process". In: *Journal of Food Engineering* 105 (July 2011), pp. 264–269. DOI: 10.1016/j.jfoodeng.2011.02.030.

[26] C. Chi. *Neural Control: Concurrent System Identification and Control Learning with Neural ODE*. _eprint: 2401.01836. 2024. URL: https://arxiv.org/abs/2401.01836.

[27] C. Simpkins. "System Identification: Theory for the User, 2nd Edition (Ljung, L.; 1999) [On the Shelf]". In: *Robotics & Automation Magazine, IEEE* 19 (June 2012), pp. 95–96. DOI: 10.1109/MRA.2012.2192817.

[28] G. Pillonetto, A. Y. Aravkin, D. Gedon, L. Ljung, A. H. Ribeiro, and T. Schon. "Deep networks for system identification: a Survey". In: *ArXiv* abs/2301.12832 (2023). URL: https://api.semanticscholar.org/CorpusID:256389947.

[29] J. N. Hendriks, F. K. Gustafsson, A. H. Ribeiro, A. G. Wills, and T. B. Schön. "Deep energy-based NARX models". In: *IFAC-PapersOnLine* 54.7 (2021). Publisher: Elsevier, pp. 505–510.

[30] B. Bohn, M. Griebel, and C. Rieger. "A Representer Theorem for Deep Kernel Learning". In: *Journal of Machine Learning Research* 20.64 (2019), pp. 1–32. URL: http://jmlr.org/papers/v20/17-621.html.

[31] N. Takeishi, Y. Kawahara, and T. Yairi. "Learning Koopman invariant subspaces for dynamic mode decomposition". In: *Advances in neural information processing systems* 30 (2017).

[32] R. G. Ramírez-Chavarría and M. Schoukens. "Nonlinear Finite Impulse Response Estimation using Regularized Neural Networks". In: *IFAC-PapersOnLine* 54.7 (2021), pp. 174–179. ISSN: 2405-8963. DOI: https://doi.org/10.1016/j.ifacol.2021.08.354. URL: https://www.sciencedirect.com/science/article/pii/S2405896321011289.

[33] M. Fortunato, C. Blundell, and O. Vinyals. "Bayesian Recurrent Neural Networks". In: *CoRR* abs/1704.02798 (2017). arXiv: 1704.02798. URL: http://arxiv.org/abs/1704.02798.

[34] A. Putranto and X. Chen. "Roasting of Barley and Coffee Modeled Using the Lumped-Reaction Engineering Approach (L-REA)". In: *Drying Technology - DRY TECHNOL* 30 (Apr. 2012), pp. 475–483. DOI: 10.1080/07373937.2011.647185.

[35] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis. "Machine learning applications in cancer prognosis and prediction". In: *Computational and Structural Biotechnology Journal* 13 (2015), pp. 8–17. ISSN: 2001-0370. DOI: https://doi.org/10.1016/j.csbj.2014.11.005. URL: https://www.sciencedirect.com/science/article/pii/S2001037014000464.

[36] J. Hernández, B. Heyd, C. Irles, B. Valdovinos, and G. Trystram. "Analysis of the heat and mass transfer during coffee batch roasting". In: *Journal of Food Engineering* 78.4 (2007), pp. 1141–1148. ISSN: 0260-8774. DOI: https://doi.org/10.1016/j.jfoodeng.2005.12.041. URL: https://www.sciencedirect.com/science/article/pii/S0260877406000239.

[37] S. A and S. R. "A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends". In: *Decision Analytics Journal* 7 (2023), p. 100230. ISSN: 2772-6622. DOI: https://doi.org/10.1016/j.dajour.2023.100230. URL: https://www.sciencedirect.com/science/article/pii/S277266222300070X.

[38] D. Bottazzi, S. G. Farina, M. Milani, and L. Montorsi. "A numerical approach for the analysis of the coffee roasting process". In: *Journal of Food Engineering* 112 (2012), pp. 243–252. URL: https://api.semanticscholar.org/CorpusID:98249885.

[39] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. "Optuna: A Next-generation Hyperparameter Optimization Framework". In: *CoRR* abs/1907.10902 (2019). arXiv: 1907.10902. URL: http://arxiv.org/abs/1907.10902.

[40] D. P. Kingma and J. Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[41] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud. "Neural ordinary differential equations". In: *Advances in neural information processing systems* 31 (2018).