

IMPROVING THE RELIABILITY OF GEOSPATIAL MODELS FOR TIMBER HARVEST PREDICTION BY MEANS OF CONFORMAL PREDICTION

Jitse Deketelaere

Student ID: 02107184

Supervisors: Prof. Dr. Ir. Willem Waegeman, Dr. ir. Victor Deklerck

Tutor: Ir. Sebastian Gabriel Jimenez

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of master in Bioscience Engineering: Chemistry and Bioprocess Technology.

Academic year: 2025 - 2026

De auteur en de promotor geven de toelating deze masterproef voor consultatie beschikbaar te stellen en delen van de masterproef te kopiëren voor persoonlijk gebruik. Elk ander gebruik valt onder de beperkingen van het auteursrecht, in het bijzonder met betrekking tot de verplichting de bron uitdrukkelijk te vermelden bij het aanhalen van resultaten uit de masterproef. The author and the promotor give permission to use this thesis for consultation and to copy parts of it for personal use. Every other use is subject to the copyright laws, more specifically the source must be extensively specified when using results from this thesis.

Preface

This thesis is carried out as part of a collaboration between the Department of Data Analysis and Mathematical Modeling, within the Faculty of Bioscience Engineering, Botanic Garden Meise and World Forest ID. I would like to express my sincere appreciation and gratitude to both of my promoters; Dr. ir. Victor Deklerck and prof. Dr. ir. Willem Waegeman for always being approachable and enthusiastic. Every meeting, they provided helpful and meaningful feedback and consistently showed great interest in my work. They clearly communicated expectations and offered excellent guidance throughout the year.

Additionally, I also want to thank my tutor, Ir. Sebastian Gabriel Jimenez who always was ready to help. Especially at the beginning of the year, he was a big help, as I had no prior experience with Python before starting this dissertation. He was patient and allowed me to learn at my own pace, while remaining encouraging and positive. I also want to thank my fellow student, Wander Labaere, who attended monthly meetings. Working with the same data, he clearly explained his progress and showed interest in mine as well. Finally, I would like to thank my family for their genuine interest and for supporting me throughout the entire year.

Contents

Preface	iii
Contents	vi
List of symbols and abbreviations	ix
Samenvatting	xi
Summary	xiii
Introduction	xv
1 DEFORESTATION	1
1.1 Consequences of deforestation	1
1.2 Drivers of deforestation	2
1.2.1 Agricultural expansion	2
1.2.2 Illegal timber trade	3
1.2.3 The wrong policy as a deforestation driver	4
1.3 Legislation	5
1.3.1 A complex supply chain	5
1.3.2 The European Deforestation Regulation	5
1.4 The current problem with the regulations	7
2 THE CHEMICAL FINGERPRINT	9
2.1 Stable Isotope Ratios	9
2.1.1 Principle of isotopes	9
2.1.2 Origins of variations in isotopic abundances	11
2.1.2.1 Carbon	11
2.1.2.2 Hydrogen	11
2.1.2.3 Oxygen	12
2.1.2.4 Nitrogen	12
2.1.2.5 Sulfur	13
2.1.3 Real-life traceability applications of SIR	13
2.2 Multi-Element	13
3 ORIGIN ASSIGNMENT METHODS	15
3.1 Machine Learning: current state-of-the-art	15
3.1.1 Discriminative models	15
3.1.2 Generative models	16
3.2 Other gaps and limitations	17
4 OBJECTIVES AND OUTLINE OF THIS DISSERTATION	19

5	MATERIAL AND METHODS	21
5.1	Sample collection and measurements	21
5.1.1	Sample collection	21
5.1.2	Isotope analysis	22
5.1.3	X-ray fluorescence spectroscopy	23
5.2	Data pre-processing	23
5.2.1	ME pre-processing	23
5.2.1.1	Isometric log ratio transformation	23
5.2.1.2	Robust median standardization	24
5.2.1.3	Data whitening	25
5.2.2	SIRA pre-processing	25
5.3	Random Forest	25
5.3.1	Tree-based methods	26
5.3.2	Random Forest Algorithm	26
5.4	Conformal prediction	28
5.4.1	Full versus split	28
5.4.2	Conformalized quantile regression	29
5.4.3	Classification	31
5.4.3.1	Adaptive prediction sets	31
6	Results	33
6.1	Results Data exploration	33
6.2	Results Random Forest	36
6.2.1	Classification Random Forest Model	36
6.2.2	Regression Random Forest Model	37
6.3	Results Conformal Prediction	37
6.3.1	Results LAC Conformal Prediction	38
6.3.2	Results APS Conformal Prediction	38
6.3.3	Results Conformalized Quantile Regression	39
6.3.4	Comparison of the different methods	39
6.3.5	Results Per Genus	41
7	Discussion	44
7.1	Discussion Conformal Prediction performance	47
7.2	Discussion Conformal Prediction merits	49
7.3	Possible improvements	50
	Bibliography	54

Use of GenAI tools in the master's dissertation:

- ☐ No use of GenAI tools
- ☒ One or more GenAI tools used

Indicate and describe, in case of using one or more GenAI tools, how you have used them:

	GenAI tool used as	Description	Used tool(s)
☐	Idea generator		
☒	Search engine	Searching for fundamental academic literature of certain topics	Undermind
☒	Writing assistant	Making text more concise and fluent	Copilot
☐	Designer		
☐	Practice tool		
☐	Data analysis		
☐	Efficiency booster		
☒	Used in another way	Helping with complex coding tasks.	Copilot

List of symbols and abbreviations

Adaptive prediction set APS.

Conformal Prediction CP.

Conformalized Quantile Regression CQR.

Convention on the International Trade in Endangered Species CITES.

Ecosystem Services ES.

European Deforestation Regulation EUDR.

European Timber Regulation EUTR.

Food and Agriculture Organization FAO.

Forest Stewardship Council FSC.

Gaussian Process GP.

Isometric Log-ratio ILR.

Isotope Ratio Mass Spectrometer IRMS.

Least Ambiguous set-valued Classifier LAC.

Mass Spectrometry MS.

Median Absolute Deviation MAD.

Multi-element ME.

Multi-element analysis MEA.

Random Forest RF.

Stable Isotope Ratio SIR.

Stable Isotope Ratio analysis SIRA.

X-ray fluorescence XRF.

Samenvatting

Wereldwijde ontbossing blijft een alarmerend groot probleem, met de uitbreiding van de landbouw en illegale houtkap als belangrijkste oorzaken. Illegaal verkregen hout maakt 10–30 % uit van de totale wereldwijde handel in houtproducten. Sinds juni 2023 heeft de Europese ontbossingsverordening (EUDR) de Europese houtverordening (EUTR) vervangen, waarbij nog steeds wordt vertrouwd op een *due diligence approach*. De EUDR erkent echter dat de volledige effectiviteit van deze regelgeving fundamenteel afhangt van de beschikbaarheid van robuuste, wetenschappelijk onderbouwde methoden waarmee de opgegeven geografische oogstlocatie van risicovolle bosproducten (hout, soja. . .) betrouwbaar kan worden geverifieerd. In deze masterscriptie maken we gebruik van georefereerde datasets met stabiele isotopen- en multi-elementgegevens van verschillende houtsoorten. We ontwikkelen een nieuwe voorspellingsmethode op basis van *conformal prediction* (CP), met onderliggende Random Forest-modellen. Deze methodologie wordt gebruikt om zowel landen als aaneengesloten gebieden te voorspellen. Voor de afzonderlijke landen maken we gebruik van zowel *Least Ambiguous set-valued Classifier* (LAC) CP als *adaptive prediction set* (APS) CP. Daarnaast vertalen we de voorspelde aaneengesloten gebieden naar landen, waardoor een directe vergelijking tussen de afzonderlijke en de aaneengesloten benadering mogelijk wordt. We constateren dat het gebrek aan referentiegegevens de grootste beperking blijft bij het toewijzen van de herkomst van hout. Onze resultaten tonen echter aan dat CP inderdaad in staat is om voorspellingen betrouwbaarder en robuuster te maken, en om interpreteerbare resultaten te leveren.

Summary

Global deforestation remains an alarmingly high problem, with agricultural expansion and illegal logging as its major drivers. Illegally obtained timber accounts for 10–30 % of the total global trade in timber products. Since June 2023, the European Deforestation Regulation (EUDR) has replaced the European Timber Regulation (EUTR), with a continued reliance on paperwork due diligence. However, the EUDR acknowledges that the full effectiveness of these regulations depends fundamentally on the availability of robust, science-based methods capable of reliably verifying the claimed geographic harvest location of forest risk commodities (timber, soy...). In this master's dissertation, we use a georeferenced datasets of stable isotope and multi element data of several timber species. We develop a novel prediction approach based on conformal prediction (CP), with underlying Random Forest models. This methodology is applied to both a discrete country prediction pipeline and a continuous area prediction pipeline. For the discrete setting, we employ both Least Ambiguous set-valued Classifier (LAC) CP and adaptive prediction sets (APS) CP. Additionally, we translate the predicted continuous areas into countries, allowing for a direct comparison between the discrete and continuous approaches. We find that the lack of reference data remains the biggest limitation in the setting of timber origin assignment. However, our results demonstrate that CP does indeed have the ability to make predictions more reliable, robust and to provide easily interpretable outputs.

Introduction

Despite multi-national, governmental and corporate pledges, global deforestation remains alarmingly high, driven primarily by agricultural expansion (Food and Agriculture Organization of the United Nations, 2024). Tropical forests are hit the hardest by agricultural encroachment, and have lost 3.7 million hectares to agricultural expansion in 2023 alone (Weisse et al., 2024). Alongside agricultural expansion, the continued occurrence of illegal logging is exerting increased pressure on forest ecosystems. Illegal logging is the most profitable natural resource crime, valued at US\$52 billion to US\$157 billion dollars per year (Mavrellis, 2017). Illegally obtained timber accounts for 10–30 % of the total global trade in timber products and, when focusing on Southeast Asia, Central Africa and South America, it is estimated that 50–90 % of timber is obtained illegally from those regions (Nellemann, 2012). After logging, the illegal timber often is traded via third countries, e.g. China, before it enters the consumer country. This hampers paper traceability efforts and makes it harder to identify the true timber harvest location. In order to obtain a long-term solution, policies that combine both socioeconomic benefits and ecological protection are needed. The European Deforestation Regulation (EUDR), which entered into force in June 2023 and replaces the European Timber Regulation (EUTR), expands the scope beyond timber to include a broader range of forest-risk commodities (soy, cacao. . .) and requires that these products originate from land that has not been subject to deforestation after the cut-off date. However, the EUDR still maintains a due diligence approach, which includes a risk assessment to analyze and evaluate the risk of placing illegally harvested forest risk commodity products on the market. Paper-based approaches are highly vulnerable to fraud. Evidence suggests that between 20 and 50% of traded timber and forest products are sold with fraudulent claims (World Forest ID, 2025). However, the full effectiveness of these regulations is fundamentally dependent on the availability of robust, science-based methods that can reliably verify the claimed geographic origin of Forest risk commodities.

Within this dissertation, conducted in collaboration with the Meise Botanic Garden and World Forest ID (www.worldforestid.org), we aim to deepen our understanding of timber traceability models and assess their capability to determine the harvest location of timber. In particular, this work explores a novel approach, conformal prediction, as a means of improving the reliability of origin assignments. Beyond introducing this methodology, the dissertation examines how model performance differs when geographic space is represented using a continuous versus a discrete formulation. The ultimate goal of this research is to develop a new state-of-the-

art methodology that delivers robust and trustworthy predictions of timber harvest origin. Such a framework is designed not only to achieve high accuracy in its current implementation, but also to serve as a foundation for further refinement and development by the research teams at Meise Botanic Garden and World Forest ID. In doing so, this master's dissertation seeks to make a meaningful contribution to improving monitoring and enforcement practices within the timber trade, thereby supporting broader efforts to combat illegal logging and promote sustainable forest management.

1. DEFORESTATION

Forests are fundamental to the health of our planet. They provide numerous ecosystem services (ES) that sustain both human societies and natural systems. These services include (1) provisioning functions, such as the supply of food, fuel, and timber; (2) regulating functions, such as carbon sequestration, biodiversity conservation, flood and erosion control, and air purification; and (3) cultural services, which provide opportunities for recreation, education, and broader societal well-being (Leal Filho et al., 2025; Maes et al., 2021). Together, these services illustrate how essential forests are for the resilience of natural systems and the communities that depend on them.

Deforestation remains a critical global concern. Forests currently cover roughly 35% of the total land surface, with more than half (54%) of these forests existing in only 5 countries: Russia, Brazil, Canada, the United States of America and China (in descending order, by area) (Food and Agriculture Organization of the United Nations, 2024). Unfortunately, deforestation is a major threat, with far-reaching and potentially severe global consequences, as illustrated in the following paragraphs.

1.1 Consequences of deforestation

Deforestation and forest degradation have far-reaching ecological impacts worldwide, including significant losses in biodiversity, accelerated climate change, soil degradation, and disruption of the water cycle. Over the long term, these environmental effects can also translate into socio-economic challenges, such as reduced long-term income opportunities for local communities and producer countries (Pendrill et al., 2019b). In what follows, three ecological consequences are examined in greater detail to illustrate the severity of these impacts.

First, the removal of forests contributes greatly to habitat loss and puts pressure on the fauna and flora. Together, climate change and deforestation could cause a decrease of up to 58% in the richness of Amazon tree species, while deforestation alone can cause 19–36% by 2050 (Gomes et al., 2019; Tilman et al., 2017; Pendrill et al., 2019b).

Secondly, deforestation and forest degradation release large amounts of stored carbon, in the form of CO₂, into the atmosphere and reduce long-term carbon se-

questration (Baccini et al., 2012). This weakens the global role of forests in the regulation of climate change. In the Amazon region, such losses critically diminish its role as a carbon sink, with significant repercussions for global climate stability (Leal Filho et al., 2025). Furthermore, forests play a crucial role in controlling the overall air quality, which directly impacts human health. (Reddington et al., 2015; Abraham, 2020).

Third, the water vapor produced by the transpiration of tropical forests in South America, forms 'flying rivers'. These are transported by wind to drier regions, particularly those with a strong agroindustry. Therefore, the hydroclimate in the south-central region partially regulates the agroindustry by moisture fluxes originating in the Amazon rainforest, and is responsible for approximately 70% of the country's GDP (Getirana et al., 2021). Consequently, forest loss contributes not only to global climate change but also to regional climatic imbalances, intensifying droughts and altering rainfall regimes (Leal Filho et al., 2025).

The consequences described above illustrate the crucial role forests play. Despite that, deforestation remains a major global issue. A clear illustration can be found in the Amazon rainforest, which covers an area of more than 6 million km² across South-America. Since the 1970s, it has lost approximately 15% of its surface area, pushing the region dangerously close to a tipping point (Flores et al., 2024). A tipping point for this gigantic biome refers to an inability to continue sustaining itself once the limit of destructive anthropic interference has been exceeded. A deeply degraded forest or even non-forest vegetation will then replace the rainforest (Marques, 2024).

Addressing deforestation requires a holistic approach. To achieve this, we must first identify the underlying drivers of deforestation and determine how best to address them.

1.2 Drivers of deforestation

1.2.1 Agricultural expansion

DeFries et al. (2010) showed that deforestation is primarily driven by urban population growth and agricultural expansion. Tropical forests are particularly vulnerable to agricultural encroachment, losing 3.7 million hectares to agricultural expansion in 2023 alone (Weisse et al., 2024). For instance, agricultural expansion accounted for 62% of deforestation in the tropics and subtropics between 2005 and 2014, with Brazil and Indonesia together responsible for nearly half of this total (Pendrill et al., 2019b).

1. DEFORESTATION

Soy cultivation is the third largest driver of tropical deforestation, with particularly strong impacts in Brazil (Maor et al., 2026; da Silva Araújo et al., 2025). Between 2000 and 2019, soybean cultivation more than doubled from 26.4 Mha to 55.1 Mha, with the most rapid expansion occurring in the Brazilian Amazon, where soybean area increased more than tenfold (Song et al., 2021).

Another example involves the Ghanaian tropical forest, which continues to face severe threats of deforestation and degradation. The country is facing an alarming annual deforestation and forest degradation rate of 2% (about 135,000 hectares each year). Agricultural expansion is again the main driver, in particular the cocoa production, as cocoa crops require high light levels, encouraging farmers to remove shade trees (Kyere-Boateng and Marek, 2021).

Overall, deforestation driven by agricultural expansion is closely linked to global market demand for soy, cocoa, palm oil, and other non-timber forest commodities. The transition from the European Timber Regulation (EUTR) to the European Deforestation Regulation (EUDR) reflects growing recognition of the need to address agricultural drivers of deforestation. The EUDR expands its scope beyond timber to include key forest-risk commodities, such as cattle, oil palm, soy, cocoa, coffee, and rubber, highlighting their increasing importance in anti-deforestation legislation (Köthke et al., 2023). The EUDR is discussed in more detail in Section 1.3.2.

1.2.2 Illegal timber trade

Alongside agricultural expansion, the continued occurrence of illegal logging and illicit timber trade is exerting increased pressure on forest ecosystems. The illegal timber trade is the most profitable natural resource crime, valued at US\$ 52 billion to US\$ 157 billion dollars per year (Mavrellis, 2017; Nellemann, 2012). Globally, illegally sourced timber constitutes approximately 10–30% of all traded timber products. In regions such as Southeast Asia, Central Africa, and South America, this proportion is even more pronounced, with estimates suggesting that 50–90% of timber is obtained through illegal means (Nellemann, 2012). After logging, the illegal timber often is traded via third countries, such as China, before it enters the consumer country. These complex supply chains hampers paper traceability efforts and makes it harder to identify the true timber harvest location. Fraud was the second most common form of corrupt practice used to introduce illegal wood into the market, following bribery (INTERPOL, 2016). Evidence suggests that between 20 and 50% of traded timber and forest products are sold with fraudulent claims (World Forest ID, 2025).

A recent example illustrates these challenges: following Russia's invasion of Ukraine, the USA and the EU imposed economic sanctions on Russia including restrictions or bans on wood imports from Russia and Belarus (EU, 2023). However, so called

'conflict timber' continues to enter European markets through secondary markets. EU officials report that it is nearly impossible to determine the true origin of such wood and, consequently, to identify who profits from it (EarthSight, 2023). The loss of the EU market forced a decrease in Russian domestic processing and an expansion of exports to secondary markets in countries without sanctions, such as Turkey and Kazakhstan. At the same time, increased volumes of plywood from third countries buying from Russia, appear in EU plywood imports, indicating the continued leakage of 'conflict timber' into European supply chains (Mortier et al., 2024).

1.2.3 The wrong policy as a deforestation driver

Lastly, deforestation is also driven by other factors, such as poor governing. Sassen et al. (2013) observed that around Mount Elgon, Uganda, deforestation increased when management institutions broke down, while forest recovery took place when boundaries were re-established. Although policy frameworks that address deforestation have multiplied rapidly in recent decades, their effectiveness varies considerably across regions. In some critical areas, strengthened regulations and enhanced monitoring systems have contributed to notable reductions in forest loss. For example, deforestation has decreased by 8.4% in Indonesia and by 50% in Brazil's Legal Amazon (Food and Agriculture Organization of the United Nations, 2024). These improvements were crucial, as Indonesia and Brazil together accounted for nearly half (44%) of annual global deforestation between 2005 and 2013, representing an average of 5.5 Mha yr⁻¹ of forest loss across tropical and subtropical regions (62% of the global total) (Pendrill et al., 2019a).

According to Sassen et al. (2013), the context (e.g. law enforcement, collaborative management, political interference) under which drivers operate, rather than the drivers per se, determines the impact on forest cover. Clear and well-designed policies are crucial for effective forest governance. However, contemporary forest management strategies often prioritize short-term economic gains, frequently intervening in natural processes to maximize timber production or land conversion (Duncker et al., 2012).

In addition, effective communication and transparency across the entire supply chain are essential. The global wood sector faces continuous scrutiny, meaning that value-added wood products must consistently meet stringent quality requirements. Yet, research shows that many wood-exporting companies in the West Africa sub-region are unaware of most of these standards (Mensah et al., 2025).

1.3 Legislation

1.3.1 A complex supply chain

As discussed in the previous sections, timber and wood-based products are valuable commodities traded around the world. Like many other high-value natural resources, these products face challenges of corruption, fraud, and illegal harvesting. These gray and black market activities in the wood and forest products sector are not limited to the countries where the wood was harvested, but they also occur throughout the global supply chain and have been linked to illicit financial flows, including trade-based money laundering, document fraud, species mislabeling, and other suspicious or potentially illegal practices (Wiedenhoeft et al., 2019). At every stage of the forest-product supply chain, law enforcement agencies struggle to determine whether specific shipments may be connected to illegal activity. Global supply chains are often notoriously complex and involve multiple countries. The places where raw materials are harvested, grown, or extracted are frequently different from the countries where the final product is manufactured, processed, and eventually shipped for sale. To achieve long-term solutions, policies must be holistic, addressing both the socioeconomic needs of communities and the protection of forest ecosystems.

Current timber regulations and international frameworks include the US Lacey Act, the EUDR, the UK Timber Regulation, the Convention on the International Trade in Endangered Species (CITES), and as well as various national laws in countries such as Australia, South Korea, and Japan. In the following section, we focus specifically on the European context and examine the EUDR in more detail.

1.3.2 The European Deforestation Regulation

It is evident that a clear authority framework is needed to ensure that legal aspects are being obeyed. Not only timber, but also other agricultural commodities linked to deforestation, such as soy and cacao, are increasingly significant. This is exemplified by the implementation of the EUDR since December 2024, which replaces the EUTR. Compared to the EUTR, the EUDR expands its scope to cover six additional agricultural commodities (soy, cattle, rubber, coffee, cocoa, and oil palm) and includes most products made from these commodities (European Commission, 2026; Köthke et al., 2023). It also demands faster, interoperable, and scientifically validated tools to detect and deter illegal trade (European Commission, 2026).

The EUDR serves as the EU's primary anti-deforestation law, designed to ensure that key products traded and consumed within the EU no longer contribute to global deforestation or forest degradation. It is a critical element of the European Green

Deal intended to mitigate deforestation associated with the EU’s market activities (European Commission, 2026). The EUDR requires that these products originate from land that has not been subject to deforestation after the cut-off date (31/12/2020). This means that the regulation prohibits placing on the EU market any commodity or product linked to deforestation, regardless of the country of origin (Nadras et al., 2024; Köthke et al., 2023).

The EUDR continues to rely on the due diligence approach, including a risk assessment to analyze and evaluate the risk of placing illegally harvested timber products on the market. Regulations such as the EUDR carry significant implications for commodity producers worldwide. In Ghana, for instance, cocoa producers often lack access to the technical solutions and data required to comply with EUDR traceability and due-diligence standards (Nadras et al., 2024). The full effectiveness of these regulations is fundamentally dependent on the availability of robust, science-based methods that can reliably verify the claimed geographic origin of Forest risk commodities.

The EUDR identifies three key entities in the supply chain of relevant products: operators, traders, and small and medium-sized enterprises (SMEs). Each has distinct responsibilities as described in Table 1.1.

Table 1.1: Entities and their roles in the EUDR supply chain

Entity	Description
Operators	Entities that place relevant products on the EU market or export them and are responsible for conducting due diligence under the EUDR. Operators submit due diligence statements confirming compliance and negligible risk.
Traders	Entities that buy and sell relevant products within the EU market, with obligations depending on their SME status.
SMEs	Small and medium-sized enterprises subject to reduced EUDR obligations, mainly related to information retention and provision.

Source: Adapted from Nadras et al. (2024)

In addition to regulations, certification schemes are becoming increasingly important. This trend is driven by growing consumer awareness, regulatory pressures such as the EUDR and the U.S. Lacey Act, and expanding corporate sustainability commitments (Mensah et al., 2025). For certification systems such as the Forest Stewardship Council (FSC), compliance with all applicable laws is a fundamental requirement. Consequently, certificate holders must be able to demonstrate that all timber has been sourced and traded legally.

1.4 The current problem with the regulations

At present, authorities mainly verify the declared geographic origin of timber by checking trade documentation, leaving considerable opportunity for tampering. Although the EUDR introduces a requirement to provide precise geolocation, this information remains vulnerable to fraud because reported origins can be falsified and are used primarily to assess potential deforestation rather than to confirm the actual site of harvest.

A fully independent approach that can verify the claimed origin of timber is crucial to enforce regulations such as the EUDR. For example, by using chemical and genetic wood characteristics, we can determine ("where was it harvested") or verify ("is the claimed harvest location plausible") the geographic origin of timber. Because it relies solely on wood properties, this method is fully independent of trade documentation.

Three methods to verify harvest claims have taken the forefront: (1) Stable Isotope Ratio Analysis (SIRA) (da Silva Araújo et al., 2025; Maor et al., 2026; Gori et al., 2025), (2) Multi-Element Analysis (MEA) (Aparicio-Muriana et al., 2025; Boeschoten et al., 2023) and (3) Genetic Analysis (Boeschoten et al., 2025; Degen et al., 2025). Mortier et al. (2024) combined both SIRA and MEA, while Boeschoten et al. (2025) integrated all three, namely the chemical and genetic data.

In this master's dissertation, we will focus on a combination of the first two (SIRA and MEA) and we will use the datasets described in Mortier et al. (2024), increased with some additional datapoints since publishing.

2. THE CHEMICAL FINGERPRINT

The composition of stable isotopes ratio's (SIR) and multi-elements (ME) in plant tissue is determined by the soil, climate, metabolic fractionation and other factors characteristic of the species and its habitat (Camin et al., 2017; Sharp, 2017). Consequently, the natural variation of these factors across space allows us to discriminate between geographic areas.

2.1 Stable Isotope Ratios

Most of the abundant elements in organic compounds (hydrogen, oxygen, carbon, sulfur, nitrogen) have stable isotopes and can be analyzed relatively routinely with mass spectrometry (MS) (Boner et al., 2007). In the following section, we examine the chemical principles underlying isotopic methods in greater detail.

2.1.1 Principle of isotopes

An atom is composed of protons, electrons, and neutrons. Protons and electrons carry positive and negative charges respectively, whereas neutrons are electrically neutral. Both protons and neutrons have comparable masses, while the mass of an electron is negligible. Elements are defined by their number of protons, known as the atomic number Z . For a given element, all atoms have the same Z , but the number of neutrons N may vary. The sum $A = Z + N$ defines the mass number.

Variations in neutron number do not alter the fundamental chemical behavior of an element or its compounds, but they do change atomic mass and give rise to isotopes. Thus, isotopes of the same element differ only in the number of neutrons in the nucleus. For example, oxygen has three naturally occurring isotopes: ^{16}O with 8 protons and 8 neutrons, ^{17}O with 8 protons and 9 neutrons, and ^{18}O with 8 protons and 10 neutrons. Figure 2.1 shows a schematic representation of the stable oxygen isotopes. These mass differences result in subtle chemical and physical variations that lead to isotopic fractionation (Sharp, 2017).

Isotopes are classified as either radioactive or stable. Radioactive isotopes possess unstable nuclei that decay over time into other isotopes, whereas stable isotopes remain unchanged on geological timescales. Most applications of stable isotope

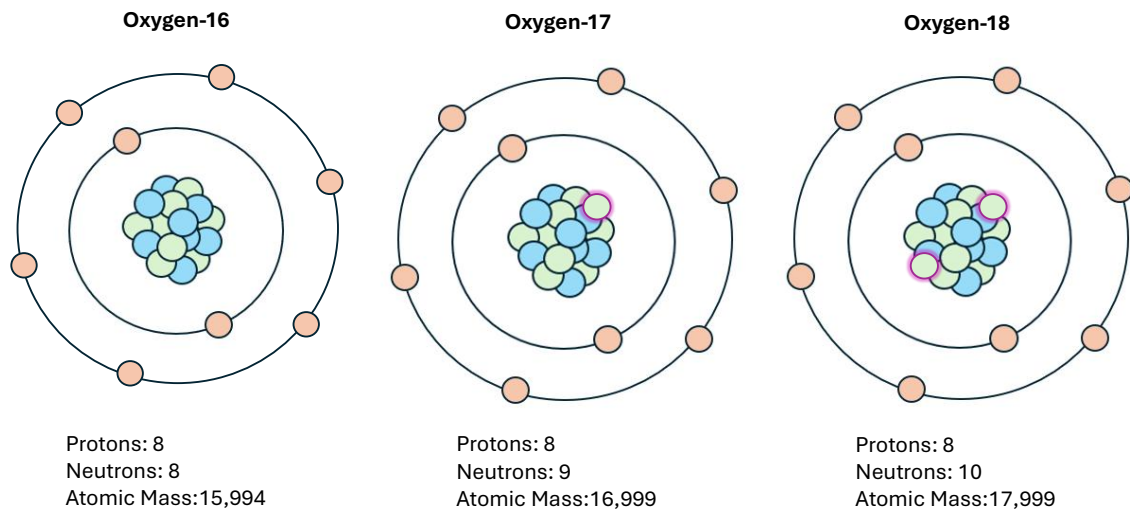


Figure 2.1: Schematic representation of the stable oxygen isotopes ^{16}O (left), ^{17}O (middle), and ^{18}O (right). Protons are blue. Neutrons are green, with a purple outline indicating that the neutron is an extra one compared to ^{16}O . Electrons are orange.

analysis focus on five elements: H, C, N, O, and S. This is because these elements constitute the majority of biological tissues.

The isotopic ratio refers to the ratio between two isotopes of an element as it occurs in nature. Conventionally, these ratios are expressed as the dimensionless ratio of the heavy (rarer) isotope to the light (more abundant) isotope, e.g., $R = \frac{^{18}\text{O}}{^{16}\text{O}}$

In general:

$$R = \frac{\text{Abundance heavy isotope}}{\text{Abundance of light isotope}}$$

Because absolute isotope ratios are difficult to measure with high precision, isotopic ratios are always reported relative to internationally recognized standards, such as Vienna Pee Dee Belemnite (VPDB) for carbon, Vienna Canyon Diablo Troilite (V-CDT) for sulfur, Vienna Standard Mean Ocean Water (VSMOW) for oxygen and hydrogen, and atmospheric nitrogen for nitrogen (Muccio and Jackson, 2009). To express small deviations from these standards, McKinney et al. (1950) introduced the δ -notation:

$$\delta = \left(\frac{R_x - R_{std}}{R_{std}} \right) \times 1000$$

2. THE CHEMICAL FINGERPRINT

Differences in isotope ratios arise from fractionation processes, which are natural mechanisms that partition lighter and heavier isotopes during chemical, physical, or biological reactions.

2.1.2 Origins of variations in isotopic abundances

Although the average isotope ratio of each terrestrial element was fixed around the time of the earth's formation, localized variations occur based on selective enrichment or depletion of the heavier isotopes relative to the average values. The signal of stable isotopes within an organism is mainly determined by soil, climate, and other biotic and abiotic components (Siegwolf et al., 2022). Fractionation occurs for all common elements such as carbon, hydrogen, oxygen, nitrogen and sulfur.

2.1.2.1 Carbon

Although all plants rely on atmospheric or dissolved CO₂ as their carbon source, various factors can influence a plant's ability to enrich or deplete ¹³C from common sources (fractionation). One major source of variation is genetic. For example, C₃ and C₄ plants differ fundamentally in their carbon fixation pathways, resulting in distinct fractionation signatures (Muccio and Jackson, 2009). Additionally, Rubisco, the primary enzyme involved in photosynthesis, strongly discriminates against ¹³C, resulting in the fractionation of the proportion of ¹³C incorporated into plant organic matter (Roden et al., 2022).

Next to genetic factors, environmental factors such as temperature, rainfall, and hours of sunlight also influence fractionation. These factors can influence kinetic processes such as the diffusion of CO₂ through the stomata in plant leaves (Muccio and Jackson, 2009). Clear evidence for environmental sensitivity to fractionation was presented by Ehleringer et al. (2000), who demonstrated that carbon isotope ratios could be used to determine the geographic regions of South America in which different cocaine plants were cultivated.

2.1.2.2 Hydrogen

During photosynthesis, plants remove hydrogen from water and incorporate it into organic compounds. The component most commonly analyzed is cellulose, the major structural carbohydrate in plants, although lipids can also provide valuable insights into isotopic fractionation. Compound-specific analyses of individual lipids have revealed a wide range of δ -values, from approximately -400 to $+200\text{‰}$ (Sachse et al., 2012). This variability reflects isotope fractionations associated with different biosynthetic pathways, as well as dehydrogenation exchange reactions (Hoefs,

2009). As summarized by Osburn et al. (2016), four major factors control the hydrogen isotope composition of hydrocarbons: (1) the deuterium/hydrogen (D/H) isotope composition of environmental water, (2) metabolic processes that incorporate water-derived hydrogen into organic molecules, (3) hydrogen exchange processes that alter D/H ratios over geologic time, and (4) kinetic fractionations occurring during the thermal cracking of sedimentary organic matter.

2.1.2.3 Oxygen

Oxygen can enter organic matter from three primary sources: CO₂, H₂O, and O₂. DeNiro and Epstein (1979) demonstrated that the ¹⁸O content of cellulose from two sets of plants grown in waters with similar oxygen isotope ratios, but exposed to CO₂ with initially different oxygen isotope compositions, did not differ significantly. This indicates that CO₂ rapidly equilibrates in oxygen isotopes with water. Consequently, the isotopic composition of environmental water governs the oxygen isotope composition of organically bound oxygen. As a result, when analyzing plants, the same reference standard as hydrogen is used for oxygen, namely VSMOW. Oxygen isotope fractionation does not occur during uptake of soil water through the root, but rather in the leaf, where evapotranspiration enriches the remaining water in ¹⁸O, due to evapotranspiration, causing isotope enrichment. The magnitude of this enrichment depends on the ratio of external to internal water vapor pressures (Hoefs, 2009). External vapor pressures can include a rise in temperature, or a decrease in relative humidity (Jahren et al., 2002).

2.1.2.4 Nitrogen

A variety of pathways allow inorganic nitrogen to be fixed into organic matter during photosynthesis. Nitrogen-autotrophs can utilize multiple nitrogen sources, resulting in a broad range of $\delta^{15}\text{N}$ values that reflect environmental conditions. Most plants, however, typically exhibit $\delta^{15}\text{N}$ values between approximately -5 and $+2\text{‰}$ (Hoefs, 2009).

Variation in $\delta^{15}\text{N}$ within and among areas typically arises from numerous proximal factors, including the isotopic composition of deposited nitrogen, the extent to which plants acquire nitrogen from geologic sources, the amount of nitrogen obtained through symbiotic fixation, the specific chemical form of nitrogen taken up, mycorrhizal associations, and the isotopic signature of nitrogen lost from ecosystems (Craine et al., 2015).

2.1.2.5 Sulfur

The primary processes responsible for producing organically bound sulfur include the direct assimilation of sulfate by plants and various microbial assimilatory pathways that synthesize organic sulfur compounds. In most cases, inorganic sulfate and atmospheric SO₂ serve as the dominant sulfur sources for plants (Hoefs, 2009). Because the abundance of ³⁴S in total plant biomass generally reflects the isotopic composition of the source sulfur, the $\delta^{34}\text{S}$ values are commonly used to detect pollution originating from acidic rain (Tcherkez and Tea, 2013).

2.1.3 Real-life traceability applications of SIR

SIR data is well studied and has been used numerous times to determine the harvest location or origin of natural products. Examples include natural resources-such as forest products (Boner et al., 2007; Watkinson et al., 2022; Mortier et al., 2024; Gori et al., 2025; Maor et al., 2026), agricultural products (Camin et al., 2017; Maor et al., 2026), wildlife (Bowen et al., 2005), fish/seafood (Pereira et al., 2021), illicit drug trafficking (Ehleringer et al., 2000), food and drinks (Tipple et al., 2013), pollutants (Sturchio et al., 2014) and many others.

2.2 Multi-Element

Although less explored, MEA can be used for origin assignment as well. It is mainly based on the uptake of chemical elements into an organism given the amount of that element present in the soil (Boeschoten et al., 2022, 2023), but also on other factors such as the root system or nutrient assimilation strategies (Boeschoten et al., 2022). Elements that can be analyzed include Al, Si, P, S, Cl, K, Ca, Mn, Fe, Ni, Cu, Zn, Br, Rb, Sr, Ba, Pb and others.

Just as with SIR data, the ME composition of plant tissues is strongly influenced by external factors, most notably the elemental composition of the soil (Camin et al., 2017; Boeschoten et al., 2023). Consequently, the natural variation of these factors across space allows us to discriminate between geographic areas, in a manner comparable to SIR datasets (Gomes et al., 2025).

The multi-element fingerprint of vegetal samples reflects the composition of the soils in which they grow and can therefore provide information on their geographic origin. However, a direct, one-to-one comparison between the elemental composition of wood and its corresponding soil is not always feasible. Characteristic trends are nonetheless preserved in wood tissues and can assist in determining provenance (Boeschoten et al., 2023). Certain inorganic elements present in the soil

are incorporated into plant tissues through their physiological processes. Thus, MEA for origin determination relies on the fact that elemental absorption by the plant is directly linked to the availability and quantity of those elements in the soil (Boeschoten et al., 2023; Maor et al., 2026). In this way, the ME profile of wood, together with SIR data, contributes to a chemical fingerprint that can be used to determine or verify the harvest location of timber (Mortier et al., 2024). Such chemical profiles provide site-specific signatures that support geographic origin verification (Boeschoten et al., 2022).

Elemental composition-based approaches have already shown strong potential. For example, Boeschoten et al. (2023) demonstrated that ME signatures could distinguish the country of origin (Democratic Republic of Congo versus Côte d'Ivoire) of *Terminalia superba*. Similarly, Maor et al. (2026) found that the chemical profiles (SIR and ME) of soybean can be used to verify or even determine origin. Another example is provided by Moreira et al. (2024), who found that the chemical profiles of the *Ipê* species varied significantly between species, indicating that multi-elemental profiling could have potential for species identification. However, this could also mean that perhaps species specific models are needed for timber tracing, and not genus specific. They also showed that, while chemical profiles differ between wood tissues, the radial distribution patterns of the elements remain consistent.

3. ORIGIN ASSIGNMENT METHODS

The chemical fingerprint allows us to discriminate between geographical areas and makes timber harvest origin assignment possible. However, current practices could be improved. To date, the use of SIRA and MEA to identify origin is limited by the relatively simple models that are used, as well as by the limited number of reference samples used as input data for such models.

In this section, methods, currently applied for timber origin prediction, are explored. We compare them and summarize advantages and limitations.

3.1 Machine Learning: current state-of-the-art

Recent studies have approached the timber tracing problem using a variety of modeling techniques, including discriminative model approaches (Finch et al., 2017; Boeschoten et al., 2022, 2023; Wang et al., 2023; Watkinson et al., 2022) as well as probabilistic approaches based on Gaussian Process (GP) models (Mortier et al., 2024; Truszkowski et al., 2025).

3.1.1 Discriminative models

In many cases, the timber tracing problem is addressed as a supervised classification problem. Supervised learning involves training a model on labeled data, while a classification problem setting implies that the model output is categorical (Alnuaimi and Albaldawi, 2024). An alternative approach is a regression problem, where the output is continuous (Alnuaimi and Albaldawi, 2024).

Discriminative models focus on modeling the relationship between the observed features X and the class labels C directly. They estimate the posterior probability $P(C | x)$ (Nallapati, 2004). In other words, rather than modeling how the data are generated, these models aim to distinguish between classes based on the input features.

In literature, the timber tracing problem has been addressed using a range of discriminative models, with Random Forest (RF) among the most commonly applied

methods. Many studies use RF and often report good, and in some cases near-perfect, results. For instance, Boeschoten et al. (2022) performed a binary classification task between two forest sites in Cameroon, located approximately 100 km apart, achieving an accuracy of 97.3%. However, this result was obtained using a relatively small dataset, with only 37 trees sampled, of which only one was misclassified. In a following study situated in Central-Africa with more samples (991) and more classes (31), the performance decreased, while still scoring a high accuracy of 86% (Boeschoten et al., 2023). Both studies relied on multi-element (ME) data as input features.

In addition to ME data, (Watkinson et al., 2022) demonstrated that SIR data can also be effectively used to distinguish timber origin at relatively small spatial scales. They found that the SIR data of C, O, H, S exhibited significant differences between islands in the Solomon Islands. Wang et al. (2023) combined MEA and SIRA, and reported an overall accuracy of 96.2% when combining ten ME features with the stable isotope ratios $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$. However, these near-perfect result should again be interpreted with caution, as the study was based on a limited amount of samples, correctly assigning 25 out of 26 samples.

Beyond ME and SIR data, other data types can be used as well, such as metabolites (Finch et al., 2017), and genetic data (Degen et al., 2025). These approaches, however, are not the focus of this master's dissertation. Finally, although RF is commonly used for this timber tracing problem, other discriminative methods have also been applied. For instance, (Mensah et al., 2025) achieved an accuracy of 96.4% on a dataset of 56 samples based on ME analysis, using linear discriminant analysis (LDA) as a machine learning method.

3.1.2 Generative models

In contrast to discriminative algorithms, generative methods model the class-conditional probability $P(x | C)$ together with the prior probability $P(C)$ (Nallapati, 2004). The posterior probability $P(C | x)$ is then obtained by applying Bayes' rule:

$$P(C | x) = \frac{P(x | C)P(C)}{P(x)}$$

As illustrated by the examples in the previous section 3.1.1, existing SIRA and MEA studies commonly treat origin determination as a classification task, predicting a sample's origin from a set of predefined locations. Although this approach can be effective at local or regional scales, this approach struggles at larger scales (Mortier et al., 2024). GP methods allow for the direct prediction of continuous geographic locations instead of discrete classes, meaning they are not limited to a predefined set of outputs, which can be advantageous compared to classifiers.

Truszkowski et al. (2025) employed GP-based probabilistic models to predict timber harvest location. In this study, predictions were made by inferring timber isoscapes (isotopic landscapes). These isoscapes were constructed through the spatial interpolation of SIR data, resulting in a map of isotopic distributions. This method resulted in the origin estimation within 520 km.

Similarly, Mortier et al. (2024) addressed the timber tracing problem using GP regression. In a study with 905 chemically analyzed samples, they were able to estimate the geographic origin within 180 to 230 km of the true location. Their model incorporated both SIR data of C, H, N, O, and S, as well as 17 elements (Al, Si, P, S, Cl, K, Ca, Mn, Fe, Ni, Cu, Zn, Br, Rb, Sr, Ba, and Pb).

3.2 Other gaps and limitations

Besides the differences in machine learning techniques applied, previous mentioned examples illustrate that there are many other inconsistencies when addressing the timber tracing problem. There are notable differences between studies in terms of geographical scale (global, continental, national, regional), and spatial representation (discrete vs continuous). Additionally, some research consider species distinctions, while others don't. In other words, there is no consensus in the field regarding the optimal approach, with no single method clearly emerging as superior.

Disregarding the lack of unified framework, some other gaps and limitations still remain, such as the reliability and robustness of the predictions. Model-based predictions remain at risk of biased predictions when model assumptions are violated (Kumar et al., 2025). Robust and safe predictions are seen as increasingly important in machine learning, particularly in contexts where predictions have legal implications (Cresswell et al., 2024). Where simple models predict the class with the highest score (i.e., the most likely class), there can be cases where two or more classes have similar scores. In such cases, the model may select one class arbitrarily based on marginal differences in values, making the prediction brittle and potentially unstable. This lack of reliable confidence estimates can be problematic in high-stakes applications. To improve the robustness of model inferences, we will investigate strategies based on conformal prediction (CP), which produces prediction sets (Angelopoulos and Bates, 2023). CP is explained in more detail in Section 5.4. A key advantage of the CP approach is the fact that coverage is guaranteed, regardless of underlying model assumptions. Rather than returning a single prediction that may be wrong, the method provides a small set of possibilities for each test sample (Angelopoulos and Bates, 2023). CP is an upcoming technique. It can be used for applications like the classification of chemicals like pharmaceuticals (Booksh et al., 2025), or in medical image classification (Mehrtens et al., 2023). Furthermore, CP

is a highly useful tool in scenarios where human decision-making is required (Wang and Dobriban, 2026; Cresswell et al., 2024).

4. OBJECTIVES AND OUTLINE OF THIS DISSERTATION

This dissertation, carried out in collaboration with the Meise Botanic Garden and World Forest ID (www.worldforestid.org), seeks to deepen our understanding of timber traceability models and to evaluate whether they are capable of determining the harvest location of timber. In particular, it explores a novel and innovative statistical approach, namely conformal prediction, as a means of improving the reliability of origin assignments. In addition to introducing this method, the work investigates how model performance differs when using a continuous versus a discrete formulation of geographic space. Ultimately, the end-goal of this dissertation is to develop a new, state-of-the-art methodology that provides robust and trustworthy predictions of timber harvest origin. Such a framework is intended not only to deliver accurate results in its current form, but also to serve as a foundation for continued refinement and enhancement by the research team at Meise Botanic Garden and World Forest ID. In doing so, this master dissertation aims to contribute meaningfully to the improvement of monitoring and enforcement practices within the timber trade, supporting broader efforts to combat illegal logging and promote sustainable forest management.

5. MATERIAL AND METHODS

5.1 Sample collection and measurements

The data were provided by Meise Botanic Garden and World Forest ID. Chemical measurements (SIRA and MEA) are available for 1411 distinct timber samples. Each sample is characterized by a unique WFID identifier, an assigned genus and species, geographic coordinates (latitude and longitude), a country of origin, and associated chemical data. The raw dataset contains samples originating from 21 countries, including Belarus, Estonia, Latvia, Poland, Romania, Slovakia, Croatia, Finland, Hungary, Lithuania, Moldova, Ukraine, Czech Republic, France, Spain, Bosnia and Herzegovina, Solomon Islands, Russia, Kazakhstan, Austria, and Turkey. Furthermore, seven different genera are present in the raw dataset: *Betula* (386 samples), *Quercus* (298 samples), *Fagus* (366 samples), *Pinus* (207 samples), *Fraxinus* (83 samples), *Populus* (22 samples), and *Larix* (38 samples). Of the seven genera included in the raw dataset, four account for the majority of the samples, while *Fraxinus*, *Larix*, and *Populus* are poorly represented, with only 83, 38, and 22 samples, respectively. Additionally, all *Populus* samples originate exclusively from Ukraine. Due to the limited sample size and lack of distinct countries, these three genera were excluded from the dataset, resulting in a final total of 1257 samples from 19 remaining countries.

For a more detailed exploration of the dataset, we refer to the results presented in Section 6.1.

5.1.1 Sample collection

Samples were collected in the same manner as described in Mortier et al. (2024). For each species, within each country, the aim was to sample three trees within an area of 50 km. The nearest set of trees was then 100 to 250 km away, taking the species distribution into account. Based on priority genera for timber traceability in Eastern Europe, a subset was chosen to analyze. This included genera such as *Betula*, *Fagus*, *Pinus*, and *Quercus*.

All samples were analyzed using both SIRA and MEA. In total, five different isotope ratios were measured, including $\delta^{18}\text{O}_{\text{lipidextracted}}$, $\delta^2\text{H}_{\text{lipidextracted}}$, $\delta^{13}\text{C}_{\text{lipidextracted}}$,

$\delta^{15}\text{N}_{\text{lipidextracted}}$, $\delta^{34}\text{S}_{\text{lipidextracted}}$. For more a detailed explanation about isotopes and their notation, we refer to Section 2.1.2. The 17 multi-elements used in this master dissertation are: Al, Si, P, S, Cl, K, Ca, Mn, Fe, Ni, Cu, Zn, Br, Rb, Sr, Ba and Pb.

5.1.2 Isotope analysis

Stable isotopes of H, O, C, N, and S are analyzed using mass spectrometry. To achieve this, a dedicated and precise multi-collector magnetic sector instrument, known as an isotope ratio mass spectrometer (IRMS) is required (Muccio and Jackson, 2009).

First samples are prepared for the IRMS as described in Mortier et al. (2024), who followed the method described by Watkinson et al. (2020). Samples were oven-dried at 103 °C prior to preparation and the increment cores were coarsely ground with a cutting mill. All material was subsequently milled to a fine powder using a ball mill. The powder was Soxhlet-extracted for 6 h with dichloromethane (non-polar) and methanol (polar), then dried in a laboratory drying cabinet for at least 1 h, stored in airtight vials, and weighed for analysis. To prevent equilibration and humidity effects during oxygen and hydrogen isotope analyses, samples were equilibrated overnight in desiccators at 10.6% relative humidity and subsequently vacuum-dried for at least 2 h.

Then, after pyrolysis of the samples, an IRMS is used to measure the analyte relative to a reference material or reference gas. A typical IRMS consists of five major components (Muccio and Jackson, 2009):

1. A sample introduction system.
2. An electron ionization source to ionize the analyte.
3. A magnetic sector analyzer, which separates ions according to their mass-to-charge ratio (m/z) based on their trajectories in a magnetic field.
4. A Faraday-collector detector array, which detects and sends a signal.
5. A computer-controlled data acquisition system.

Within the magnetic field, ions with different m/z values follow distinct trajectories and are directed onto separate Faraday collectors positioned at fixed locations. The magnitude of the ion currents measured by the Faraday collectors is proportional to the abundance of each isotope (Muccio and Jackson, 2009).

5.1.3 X-ray fluorescence spectroscopy

The ME concentrations were measured as described by Mortier et al. (2024). Ball-milled and oven-dried wood powder samples were analyzed using X-ray fluorescence (XRF) spectroscopy to determine the relative abundance of 17 trace elements (Al, Si, P, S, Cl, K, Ca, Mn, Fe, Ni, Cu, Zn, Br, Rb, Sr, Ba and Pb). The XRF instrument consist of an X-ray tube with a thick binary Pd/Co alloy anode and adaptive excitation, resulting in low background signals and low detection limits. For sample comparisons, normalized net intensities were used. Net intensities were normalized using a Compton-scatter region as an internal standard, which results in relative ME concentrations. Two scatter regions at different energies were applied for correction: scatter region 1 (6.75–6.83 keV) for Al, Si, P, S, Cl, K, and Ca, and scatter region 2 (19.71–20.55 keV) for Mn, Fe, Ni, Cu, Zn, Br, Rb, Sr, Ba, and Pb.

5.2 Data pre-processing

Before the data can be used to train machine-learning models and perform origin predictions, it must first undergo pre-processing. Data pre-processing is required for several reasons. For example, compositional data are subject to specific constraints that should be removed before applying machine learning (Van Nguyen et al., 2022; Leung, 2021). SIRA data and ME data are pre-processed using different approaches.

5.2.1 ME pre-processing

A key aspect of multi-element concentrations, is the fact that this data is compositional. Compositional data are defined as vectors of proportions whose components sum to unity (Aitchison, 1983; Egozcue et al., 2003; Greenacre, 2021). As a result, the individual components are constrained to non-negative values and are subject to a constant-sum constraint. Three different steps can be distinguished during the pre-processing of the ME data.

5.2.1.1 Isometric log ratio transformation

Aitchison (1983) first proposed the usage of a logarithmic ratio (log-ratio) transformation. Building on this work, Egozcue et al. (2003) developed the isometric log-ratio (ILR) transformation. This transformation is used as the pre-processing technique for the ME data in this master dissertation. Leung (2021) demonstrated that in almost all cases, ILR improves results of ML classifiers, including RF models.

When working with concentrations, the values of all the features sum up to a certain unit, often equal to 100% or 1. This is called the unit-sum constraint. The composition of J components is then $\sum_{j=1}^J x_j = 1$ (Greenacre, 2021).

The most simple logratio transformation is the pairwise logratio transformation. In this case, all possible ratios between pairs of components are considered. This results in $\frac{1}{2}J(J-1)$ different logratios of the form

$$LR(j, j_0) = \log\left(\frac{x_j}{x_{j_0}}\right), \quad j, j_0 = 1, \dots, J, \quad j < j_0.$$

However, a smaller, linearly independent subset is sufficient to generate all other logratios through linear combinations.

To transform compositional data from the Aitchison simplex to a Euclidean space while preserving the relative information, a more complex transformation is required. This is achieved by constructing logratios between groups of components instead of individual pairs. ILR transformation considers the ratio of the geometric means of two subsets of components.

The ILR transformation is defined as:

$$\text{ilr}(J_1, J_2) = \sqrt{\frac{|J_1||J_2|}{|J_1| + |J_2|}} \log\left(\frac{g(\{x_j : j \in J_1\})}{g(\{x_j : j \in J_2\})}\right)$$

with J_1 and J_2 being subsets of the J components and $g(\cdot)$ being the geometric mean, defined as

$$g(\{x_j : j \in J\}) = \left(\prod_{j \in J} x_j\right)^{1/|J|}.$$

(Egozcue et al., 2003; Greenacre, 2021).

The logarithm in the ILR transformation compares the geometric average of the components in J_1 with those in J_2 . The additional scaling factor

$$\sqrt{\frac{|J_1||J_2|}{|J_1| + |J_2|}}$$

standardizes the resulting ILR variables such that they all have equal length and are mutually orthonormal (Greenacre, 2021).

5.2.1.2 Robust median standardization

The next step involves the standardization of the data. This is done with a robust median standardization as described by Rousseeuw and Croux (1993). Here, the

5. MATERIAL AND METHODS

median from each value is subtracted, and this is divided by the median absolute deviation (MAD).

$$z_j = \frac{x_j - \text{median}(\mathbf{x})}{1.4826 * \text{MAD}(\mathbf{x})}$$

with the MAD being defined as $\text{MAD}(\mathbf{x}) = \text{median}(|x_j - \text{median}(\mathbf{x})|)$.

The constant 1.4826 scales the MAD to obtain a consistent estimator of the standard deviation under the assumption of normally distributed data (Rousseeuw and Croux, 1993).

5.2.1.3 Data whitening

Lastly, in order to uncorrelate the data's features, we apply a Ledoit-Wolf whitening transformation following Ledoit and Wolf (2004). First, we center the data by subtracting the mean vector μ , yielding $X_c = X - \mu$. The covariance structure of X_c is then estimated using the Ledoit-Wolf shrinkage estimator $\hat{\Sigma}_{\text{LW}}$. The whitening transformation is defined as:

$$Y = X_c \hat{\Sigma}_{\text{LW}}^{-1/2}$$

where $\hat{\Sigma}_{\text{LW}}^{-1/2}$ is the inverse square root of the covariance estimator.

5.2.2 SIRA pre-processing

Because the SIR data is not compositional, their pre-processing is relatively straightforward. The only transformation applied is the z-score normalization, which standardizes the data to a mean of 0 and a standard deviation of 1 (Cheadle et al., 2003). Z-scores are calculated by subtracting the overall mean value from each measured value and dividing the result by the standard deviation of all measurements, according to (Cheadle et al., 2003):

$$z_j = \frac{x_j - \mu}{\sigma}.$$

5.3 Random Forest

After pre-processing the data, a machine learning model was trained to make harvest origin predictions for the timber samples. As illustrated by the examples in Section 3.1.1, RF is a commonly applied algorithm for this timber tracing problem, achieving good performance. Rf is a supervised, decision-tree-based ensemble method (Breiman, 2001).

5.3.1 Tree-based methods

A decision tree partitions the predictor space by performing recursive binary splits on various features, starting from the root node and continuing until terminal nodes are reached (James et al., 2013). This approach is applicable to both regression and classification problems. Continuous features are split at optimal threshold values, while categorical features are split by determining whether a given sample belongs to a specific subset or not (Cutler et al., 2012), as illustrated in figure 5.1. At each non-terminal node, all possible splits across all predictors are evaluated. The optimal split is selected based on a splitting criterion. The non-terminal node splits into two descendant nodes, resulting in a tree-like structure. This process continues recursively until a stopping condition is met, such as a minimum node size (James et al., 2013).

An impurity measure is used as a splitting criterion. A frequently used impurity measure is the Gini index:

$$G = \sum_{k=1}^K \hat{p}_{mk} (1 - \hat{p}_{mk})$$

with $\hat{p}_{mk}$ being the proportion of observations in the m^{th} region that are part of the class k . The cross-entropy is an alternative with similar characteristics:

$$D = - \sum_{k=1}^K \hat{p}_{mk} \log \hat{p}_{mk}$$

Both impurity measurements tend to go to zero when $\hat{p}_{mk}$ is close to 0 or 1, indicating a good separation (James et al., 2013).

Decision trees are flexible and interpretable, handling both regression and classification with continuous and categorical features, and are robust to outliers and monotone transformations. However, individual trees tend to overfit and are therefore often used as components of ensemble methods such as RF (James et al., 2013; Cutler et al., 2012).

5.3.2 Random Forest Algorithm

RF is a machine learning algorithm, originally proposed by Breiman (2001), that constructs an ensemble of predictive decision trees and combines their outputs to form a final prediction. Each tree is trained on a bootstrapped version of the original training data. At each split only a randomly selected subset of the available features is considered. This is because each individual decision tree suffers from high variance, meaning small differences in the data set would significantly influence the resulting tree (James et al., 2013). A new subset of predictors is sampled

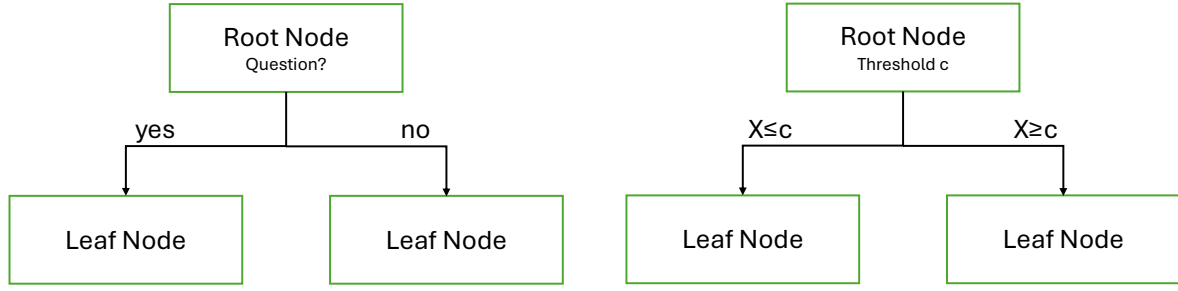


Figure 5.1: Example of two decision trees with 1 split, also called stumps. Left: the partitioning is done for a categorical feature by a yes/no question. Right: a continuous feature X is split at threshold c .

every time a split is performed, ensuring that different trees explore different predictor combinations (James et al., 2013; Cutler et al., 2012). In addition to being simple to use, the method is generally recognized for its accuracy and its ability to deal with small sample sizes and high-dimensional feature spaces (Zhu, 2020). At the same time, it has the potential to deal with large real-life systems (Biau and Scornet, 2016).

Cutler et al. (2012) described the construction and prediction mechanism of a RF model as following: Let $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$ denote the training set, where X_i are the input variables and Y_i the corresponding responses.

1. First, take a bootstrap sample $\mathcal{D}_j$ of size N from $\mathcal{D}$.
2. Then, using the bootstrap sample $\mathcal{D}_j$ as the training data, fit a tree using binary recursive partitioning (see 5.3.1):
 - (a) Start with all observations in a single node.
 - (b) Repeat the following steps recursively for each unsplit node until the stopping criterion is met:
 - i. Randomly select m predictors from the total set of p features.
 - ii. Determine the optimal binary split among all possible splits on these m predictors
 - iii. Split the node into two descendant nodes using the split from step ii.

Finally, the RF prediction is obtained by aggregating the predictions of all individual trees. For classification tasks, this aggregation corresponds to a majority vote over the tree predictions, whereas for regression it corresponds to their average (Gislason et al., 2004; Cutler et al., 2012).

The Random Forest (RF) classifier models predict discrete labels, which in our case correspond to countries. In contrast, the RF regressor models predict continuous

variables, namely geographic coordinates. However, these coordinates are constrained within their natural coordinate space (latitude and longitude), which motivates a transformation into a three-dimensional space. This transformation is defined as follows:

$$X = \cos(\text{lat}) \cdot \cos(\text{lon}), \quad Y = \cos(\text{lat}) \cdot \sin(\text{lon}), \quad Z = \sin(\text{lat}).$$

Using this representation, the regression model predicts precise geographic locations. The predicted coordinates are then compared to the true coordinates by computing the Euclidean distance in this three-dimensional space. Subsequently, this distance is converted into a great-circle distance on the Earth's surface using the haversine formula. The resulting distance, expressed in kilometers, is averaged across all predictions and serves as the evaluation metric.

5.4 Conformal prediction

As discussed in Section 3.2, CP is a technique that produces prediction sets with statistical guarantees (Romano et al., 2019; Angelopoulos and Bates, 2023). A useful feature of CP is that it can be built on top of almost any machine learning algorithm (Vovk et al., 2022), including this master's dissertation's RF models. The prediction sets must comply with two key aspects (Vovk et al., 2022; Romano et al., 2019).

1. **Validity:** in the long run, the frequency of error should not exceed α at any chosen confidence level $1-\alpha$.
2. **Efficiency:** prediction sets should be efficient, meaning the sets are as small as possible.

CP works for both regression as classification problems. In regression problems, the output is an interval of values. The length of the interval is a natural measure of efficiency of such a prediction. In classification problems, a natural measure of efficiency is the size of the prediction set, which includes possible labels. In general, smaller prediction sets are preferable, as they indicate higher certainty, whereas larger sets suggest greater uncertainty (Pourkamali-Anaraki, 2025).

5.4.1 Full versus split

As discussed in later in Section 5.4.2, a key component of CP is the use of calibration data and their conformity scores. The split CP approach separates model fitting from the construction of conformity scores through sample splitting (Lei et al.,

2018). Specifically, the regression function is trained on a training set, while a distinct calibration set is used to evaluate conformity scores and compute their empirical quantiles. Split CP is a computationally cheaper method (Romano et al., 2019; Lei et al., 2018), and was used during this master dissertation for both regression as classification tasks. Just as in Section 5.3.2, suppose that we are given n training samples

$$Z_i = (X_i, Y_i), \quad i = 1, \dots, n,$$

where x_i are the features and y_i are the corresponding responses.

The split conformal method begins by partitioning the training data into two disjoint subsets: a proper training set I_1

$$\{(X_i, Y_i) : i \in I_1\},$$

and a calibration set I_2

$$\{(X_i, Y_i) : i \in I_2\}.$$

5.4.2 Conformalized quantile regression

We first explain the CP algorithm wrapped around a regression model, which is in this master dissertation a RandomForestRegressor. In Section 5.4.3 we discuss the algorithm for a classification problem.

As discussed previously, in a regression problem, the output is continuous. We observe samples $Z_i = (X_i, Y_i) \in \mathbb{R}^d \times \mathbb{R}$, for $i = 1, \dots, n$.

Now, let $\mu(x)$ be the regression function. We aim to predict a new response Y_{n+1} corresponding to a new feature vector X_{n+1} , drawn from the same population, without imposing assumptions on μ (Lei et al., 2018). Formally, given an error rate $\alpha \in (0, 1)$, we seek to construct a prediction interval C based on (X_i, Y_i) such that the following holds (Angelopoulos and Bates, 2023):

$$\mathbb{P}(Y_{n+1} \in C(X_{n+1})) \geq 1 - \alpha$$

The prediction intervals can be constructed using different techniques. In this master dissertation, Conformalized Quantile Regression (CQR) is used, an algorithm capable of wrapping around any algorithm for quantile regression, including RF (Romano et al., 2019).

As discussed in Section 5.4.1, the split conformal method begins by partitioning the training data into two disjoint subsets: a training set I_1 and a calibration set I_2 . We apply CQR as described by Romano et al. (2019). After splitting the data, given any quantile regression algorithm $\mathcal{A}$, we fit two conditional quantile functions on the

training set,

$$\{\hat{q}_\alpha^{\text{lo}}, \hat{q}_\alpha^{\text{hi}}\} = \mathcal{A}(\{(X_i, Y_i) : i \in I_1\}),$$

corresponding to the lower and upper conditional quantiles, respectively.

In this work, the quantile regression algorithm $\mathcal{A}$ is a Quantile Random Forest (QRF) model. More specifically, the conditional quantiles are estimated using a random forest-based approach introduced by Meinshausen and Ridgeway (2006), which approximates the full conditional distribution of the response variable. The lower and upper quantile estimates $\hat{q}_\alpha^{\text{lo}}(x)$ and $\hat{q}_\alpha^{\text{hi}}(x)$ correspond to empirical quantiles of the conditional distribution $\hat{F}(Y | X = x)$ learned by the ensemble. The resulting prediction intervals are adaptive to local variability in the data, which is essential for constructing efficient conformal prediction sets.

In the key next step, we compute conformity scores that quantify the error made by the prediction interval $\hat{C}(x) = [\hat{q}_\alpha^{\text{lo}}(x), \hat{q}_\alpha^{\text{hi}}(x)]$. These scores are evaluated on the calibration set as

$$E_i = \max\{\hat{q}_\alpha^{\text{lo}}(X_i) - Y_i, Y_i - \hat{q}_\alpha^{\text{hi}}(X_i)\}, \quad i \in I_2,$$

where E_i measures either the extent to which the observed response Y_i lies outside the predicted interval at X_i or, when Y_i lies within the interval, the farthest distance between Y_i and the interval boundaries.

Finally, given a new input X_{n+1} , we construct the prediction interval for Y_{n+1} as

$$C(X_{n+1}) = [\hat{q}_\alpha^{\text{lo}}(X_{n+1}) - Q_{1-\alpha}(E; I_2), \hat{q}_\alpha^{\text{hi}}(X_{n+1}) + Q_{1-\alpha}(E; I_2)],$$

where

$$Q_{1-\alpha}(E; I_2) \text{ equals the } (1 - \alpha) \left(1 + \frac{1}{|I_2|}\right)\text{-th empirical quantile of } \{E_i : i \in I_2\}$$

with $|I_2|$ the number of calibration samples (Romano et al., 2019).

It is important to emphasize that the quantile regression model is trained solely on the training set I_1 , while the calibration set I_2 is used exclusively for estimating the correction term. The quantile estimates themselves are not updated during calibration. This strict separation between training and calibration is a defining characteristic of split conformal prediction and is essential for obtaining finite-sample coverage guarantees.

Overall, the CQR procedure combines the flexibility of quantile regression with the rigorous coverage guarantees of conformal prediction. The quantile regression model provides locally adaptive interval estimates, while the conformal calibration step ensures that these intervals achieve the prescribed marginal coverage level, regardless of potential model misspecification.

Lastly, continuous prediction intervals are transformed to prediction set, including labels (countries), in order to enable comparison between the two different methods. The model first generated prediction intervals for latitude and longitude, which define a rectangular geographic region. This region is then intersected with country region, and each country is assigned a score proportional to the fraction of the predicted area it covers. These scores are normalized to form a probability distribution, and countries are ranked according to their relative overlap with the predicted region. Prediction sets are then formed by accumulating countries in order of decreasing overlap until a total probability of at least $1 - \alpha$ is reached, which is similar to the APS approach, discussed in the following section.

5.4.3 Classification

In this section, we focus on classification problems. Here, prediction sets aren't intervals, but a collection of multiple possible labels (Vovk et al., 2022). However, general principles, such as validity and efficiency as discussed in Section 5.4, remain. The algorithm implemented and adapted to our context, is the one described by Angelopoulos and Bates (2023).

Here, we start with a model that predicts the inputs into classes, which we will call the classifier $\hat{f}$. In this master dissertation we use `RandomForestClassifier`. Again, a split CP approach is applied. We refer to Section 5.4.1 for more explanation. The classifier $\hat{f}$ must be able to predict estimated class probabilities for each class K , i.e., $\hat{f}(x) \in \Delta^{K-1}$, where $\Delta^{K-1} = \{p \in [0, 1]^K : \sum_{k=1}^K p_k = 1\}$ denotes the probability simplex. First, we perform a calibration step with the calibration dataset. We get the non-conformal scores by doing one minus the softmax output of the true class. This simple non-conformity score is also called the Least Ambiguous set-valued Classifier (LAC) and defines how unusual a sample is based on the softmax score of its true label. We then define $\hat{q}$ to be the $[(n+1)(1-\alpha)]/n$ quantile ($\hat{q}$ is basically the $1-\alpha$ quantile, but with a small correction). Finally, for a new test data point (where X_{test} is known but Y_{test} isn't), create a prediction set

$$C(X_{test}) = \{y : \hat{f}(X_{test})_y \geq 1 - \hat{q}\},$$

The prediction set for the test point $C(X_{test})$ contains now all possible labels y with a sufficiently high softmax output, greater than or equal to $1 - \hat{q}$. This algorithm satisfies validity, as also stated by Angelopoulos and Bates (2023).

5.4.3.1 Adaptive prediction sets

According to Angelopoulos and Bates (2023), a possible improvement on the classification algorithm described above, is classification with adaptive prediction sets

(APS). APS works very similar as standard LAC-based classification conformal prediction. Here, the main difference is that this approach utilizes the softmax outputs of all classes, rather than relying solely on the probability assigned to the true class when constructing conformity scores. The score is now defined as

$$s(x; y) = \sum_{j=1}^k \hat{f}(x)_{\pi_j(x)}, \quad \text{where } y = \pi_k(x).$$

with $\pi(x)$ being the permutation of $\{1, \dots, K\}$ that sorts $\hat{f}(X_{test})$ from most likely to least likely. The next step is again to set a quantile $\hat{q}$ and then form the prediction set: $\{y : s(x, y) \leq \hat{q}\}$, modified slightly to avoid zero-size sets:

$$\mathcal{C}(x) = \{\pi_1(x), \dots, \pi_k(x)\}, \quad \text{where } k = \sup \left\{ k' : \sum_{j=1}^{k'} \hat{f}(x)_{\pi_j(x)} < \hat{q} \right\} + 1.$$

In other words, we greedily include classes in our set until we reach the true label, then we stop. By adding +1 to the term k , we assure that at least one class is included, namely the top-ranked class $\pi_1(x)$. The main difference in output is thus that by doing this small modification, this more conservative method is unable to return zero-size sets (Angelopoulos and Bates, 2023).

6. Results

Before answering our research question, we summarize the results. We report results of the data exploration, RF models and the applied CP methods. We mention findings, optimal parameters and appropriate evaluation metrics. Here and there some extra explanation is given, but the majority of the discussion is done in Section 7 Discussion.

6.1 Results Data exploration

Data exploration is part of any project when working with data, and often the first step. We screen the dataset for irregularities and report (1) missing values, (2) incorrect country assignments (3) duplicated IDs, and (4) Outliers. First, the data was checked for missing values. For $\delta^{15}N_{\text{lipidextracted}}$ around 140 samples were missing, while $\delta^{34}S_{\text{lipidextracted}}$ has three samples without a value. Other features ($\delta^{13}C_{\text{lipidextracted}}$, $\delta^{18}O_{\text{lipidextracted}}$, $\delta^2H_{\text{lipidextracted}}$) contained values for every sample. This was also checked for the ME dataset. Three samples did not have any values for any ME feature. These three were dropped during the pre-processing of the data. Otherwise, only two elements (Br and Ba) included missing values for the remaining samples. We refer also to Table 6.1 and Table 6.2 for summaries of these results.

Another critical aspect of every sample is the unique WFID Identifier. WFID Identifiers have a structure of XXXYYY with X being a capital letter, and Y being a number. 81 samples do not match this pattern. 79 of these samples are samples from Russia and Kazakhstan. The Russian and Kazakhstan samples were not collected by World Forest ID, as such they had different identifiers. The chemical measurement protocol was the same however. The two other samples are from Austria. All, but one ID were unique.

Table 6.1: Total amount of missing values (NaN) for the features with missing values and the fraction of missing values for the SIR data.

Column name	Sum NaN	Percentage
$\delta^{15}N_{\text{lipidextracted}}$	140	11.137629
$\delta^{34}S_{\text{lipidextracted}}$	3	0.2386635

Table 6.2: Total amount of missing values (NaN) for the features with missing values and the fraction of missing values for the ME data.

Column name	Sum NaN	Percentage
Br	34	2.943516
Ba	204	16.46778

Each sample was assigned a country, and this assignment was validated by verifying that the sample's coordinates fell within the corresponding country. They were compared to countries borders from a GeoPandas dataset. 26 samples were potentially misassigned to a country, either by not detecting a country or by detecting a different country. However, after manual checking, all country assignments were correct, including the 26 potential mistakes.

Lastly, the data was checked for potential outliers. This was done according to Tukey's rule as stated in Hoaglin et al. (1986)

$$\text{Lower bound} = Q_1 - 1.5 \text{ IQR}$$

$$\text{Upper bound} = Q_3 + 1.5 \text{ IQR}$$

with $\text{IQR} = Q_3 - Q_1$, and with $Q_1 = \text{P25}$ and $Q_3 = \text{P75}$.

Results of this outlier detection are summarized in Figure 6.1 for both SIRA and ME data. As most detected outliers were only marginally outside the defined thresholds, and given that Random Forest models are generally robust to outliers, no additional measures were taken to address them.

6. Results

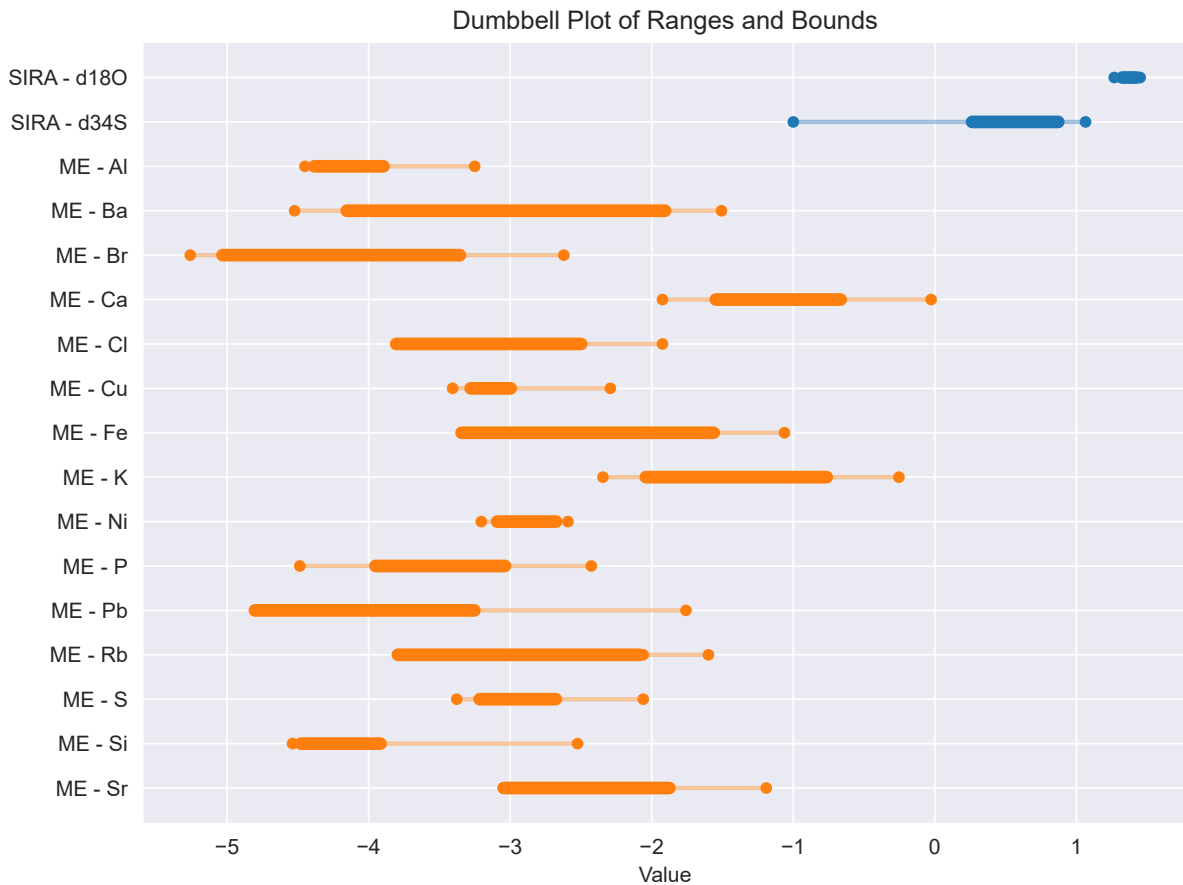


Figure 6.1: Ranges and Tukey-derived bounds for variables in the SIRA and ME datasets. Dumbbell plot of summary statistics for variables in the SIRA and ME datasets. Thin line segments represent the full range (minimum to maximum), while thicker segments indicate the interval between the lower and upper bounds calculated using Tukey's rule. Endpoints denote the extreme values. The blue colored lines are for the two SIRA features, while orange represents ME features.

6.2 Results Random Forest

A common theme throughout this dissertation is that timber origin prediction is approached using both discrete and continuous methods. In this section, results are presented for both a country-level classification approach and a regression-based approach, both employing a RF model.

6.2.1 Classification Random Forest Model

First a classifier RF model using all data (i.e. a singular model for all genera) was created. Hyperparameters of the RF model were tuned using a validation set in combination with cross-validation. Hyperparameters are parameters that can be changed to an optimal configuration in order to optimize the performance of the model. Parameters that were tuned and differ from the default values are `criterion`, which is the mathematical function used to measure the quality of a split at each node, set to "log_loss" and a `max_depth`, which is the maximum number of splits, of 20 (corresponding to the number of unique countries in the dataset). The `n_estimators` parameter is not primarily considered a hyperparameter, as increasing the number of trees in a Random Forest generally improves performance. However, for computational efficiency, this parameter can still be tuned. The `n_estimators` parameter was equal to 105 for this RF model. The overall accuracy achieved is 69.73%. The results are also visualized in Figure 6.2, where sample locations are purposely ambiguous due to a large marker and a slight offset, to preserve the exact locations. A summary of the results per country is provided in Table 6.3.

It is also important to note that among all samples for which the country was incorrectly predicted, 46.7% were assigned to a neighboring country, 34.4% to a country with one country in between, 16.0% to a country with two countries in between, and 2.9% to a country located with more than two countries in between.

Furthermore, as discussed in Chapter 3, incorporating genus or species information into the model may improve predictive performance. To evaluate this, the same RF classification pipeline was applied, but now with dummy variables representing the different genera incorporated. Here, three extra binary columns (one per genus minus one) hold genus information. This provides the model with additional info, resulting in a slight performance improvement. This overall accuracy increased to 72.79%, corresponding to an improvement of 3.06%.

Lastly, we also made a separate RF model for each genus. The overall accuracy for *Betula* is 72.29%, for *Fagus* is 67.49%, for *Quercus* is 62.75%, and 90.34% for *Pinus*. The overall accuracy, calculated as a weighted sum based on the relative fraction of each genus, is 71.60%.

6. Results

Table 6.3: Accuracy per country. The third column shows how many samples per country are in the dataset.

Country	Correct percentage	Total amount of samples
Belarus	45.31	64
Bosnia and Herzegovina	0.00	7
Croatia	83.13	166
Czech Republic	0.00	5
Estonia	78.33	60
Finland	90.91	66
France	0.00	2
Hungary	56.10	82
Kazakhstan	88.00	25
Latvia	75.00	60
Lithuania	82.65	98
Moldova	80.39	51
Poland	10.71	56
Romania	78.50	200
Russia	97.10	69
Slovakia	12.96	54
Spain	0.00	6
Turkey	50.00	8
Ukraine	67.98	178

6.2.2 Regression Random Forest Model

Just as with the classifier, we can opt for an approach with one RF model for all the genera, one RF model for all the genera with dummies and a separate RF model per genus. The first approach gives an overall error of 354.21 km from the true locations. With the inclusion of dummy columns, this changes to 331.41 km. These results appear realistic when compared with values reported in the literature; see also Section 3.1.2. Lastly, the approach using a separate RF model for each genus results in the following results: the overall score for *Betula* is 432.71 km, for *Fagus* is 190.05 km, for *Quercus* is 293.77 km, and 282.96 km for *Pinus*. The overall score, calculated as a weighted sum based on the relative fraction of each genus, is 304.45 km.

6.3 Results Conformal Prediction

The results are now presented for the CP pipeline. Two modeling approaches are considered: one using a single Random Forest (RF) model, and another using a separate RF model for each genus. Both approaches are evaluated for the discrete and regression pipelines.

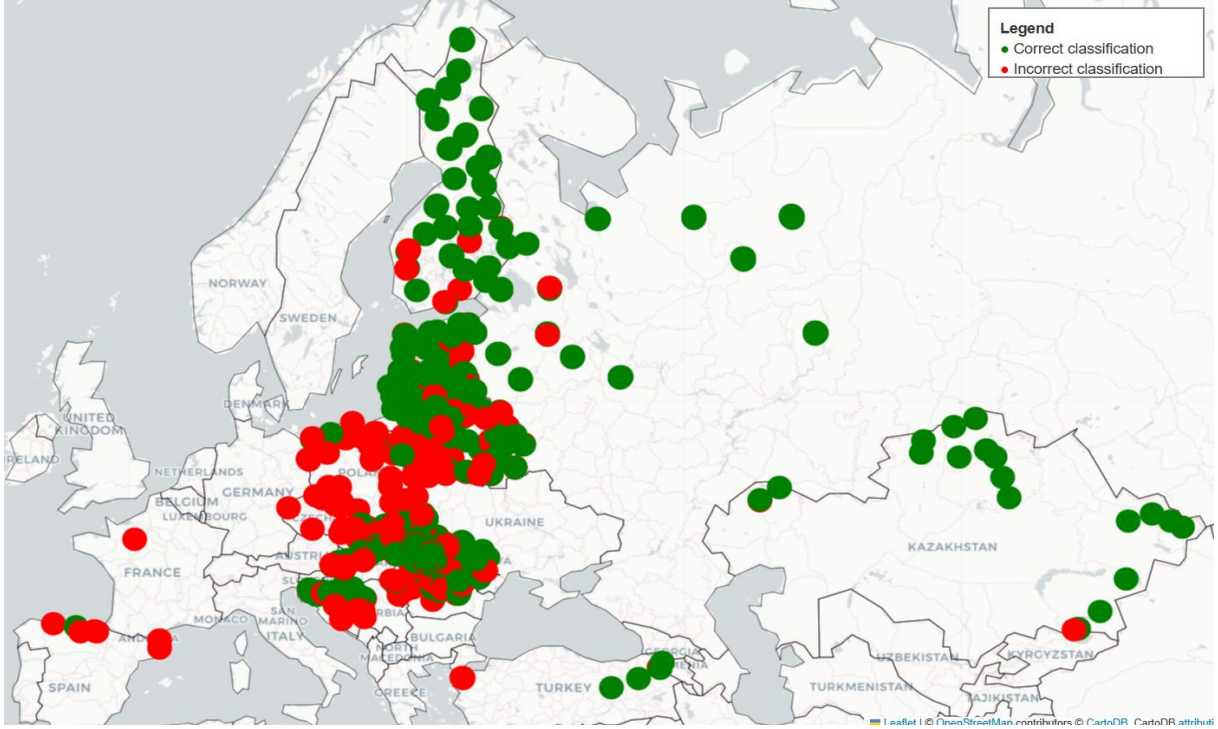


Figure 6.2: Visual representation of the correctly (green) and incorrectly (red) predicted samples for the classification pipeline of RF at country level. The samples are shown on the map with an offset to withhold exact sample locations.

6.3.1 Results LAC Conformal Prediction

First, the results of the LAC CP approach are reported. As discussed in Section 5.4.3, for a user-chosen error rate α , a coverage of at least $1 - \alpha$ is expected. We evaluate the method for $\alpha \in \{0.025, 0.05, 0.075, 0.10\}$. The resulting coverage levels and average prediction set sizes are presented in Table 6.4.

Table 6.4: Results LAC CP, all genera: For different alphas, the different coverages and average predictions set size are reported. Additionally, the average set size for prediction sets including (correct) and excluding (incorrect) the true country are reported. Lastly, the average rank of the true country in the correct prediction sets are shown.

Alpha	Coverage	Set size	Size correct	Size incorrect	Avg. rank true country
0.025	98.71	6.22	6.18	8.63	3.63
0.05	95.43	4.07	4.05	4.85	2.50
0.075	91.93	2.77	2.73	3.33	1.81
0.1	89.43	2.38	2.34	2.69	1.63

6.3.2 Results APS Conformal Prediction

Angelopoulos and Bates (2023) state that Classification with APS is an improvement over the standard approach, as it guarantees non-empty prediction sets. In our re-

6. Results

sults, the increased conservativeness of the APS method is seen in both the higher achieved coverage levels and the more conservative, average prediction set sizes, also shown in Table 6.5.

Table 6.5: Results APS CP, all genera: For different alphas, the different coverages and average predictions set size are reported. Additionally, the average set size for prediction sets including (correct) and excluding (incorrect) the true country are reported. Lastly, the average rank of the true country in the correct prediction sets are shown.

Alpha	Coverage	Set size	Size correct	Size incorrect	Avg. rank true country
0.025	99.07	7.39	7.36	9.43	1.68
0.05	97.14	5.18	5.15	6.15	1.61
0.075	96.71	4.79	4.76	5.71	1.61
0.1	96.29	4.66	4.63	5.38	1.60

6.3.3 Results Conformalized Quantile Regression

CP was also applied using a continuous approach, specifically conformalized quantile regression (CQR), in which a geographic area is predicted rather than a discrete set of countries. To enable comparison between both approaches, the predicted area was mapped to countries by ranking them according to the fraction of the area they cover. In particular, countries covering a larger portion of the predicted area are considered more likely and are therefore ranked higher in the prediction set. This is treated more critically in Section 7.2, where its limitations are discussed. Results are reported for four values of α , namely 0.025, 0.05, 0.075, and 0.1, as shown in Table 6.6.

Table 6.6: Results CQR, all genera: For different alphas, the different coverages and average predictions set size are reported. Additionally, the average set size for prediction sets including (correct) and excluding (incorrect) the true country are reported. Lastly, the average rank of the true country in the correct prediction sets are shown.

Alpha	Coverage	Set size	Size correct	Size incorrect	Avg. rank true country
0.025	98.71	14.9	14.84	16.03	5.93
0.05	95.71	9.41	9.35	10.01	3.99
0.075	90.77	7.2	7.12	7.81	3.29
0.1	89.27	5.74	5.68	6.07	2.78

6.3.4 Comparison of the different methods

We compared both classification-based methods with the CQR approach using a fixed value of $\alpha = 0.1$. First, the LAC CP method was compared with the CQR approach. Among all samples, 69 were not covered by either method, 135 were covered only by the LAC CP method, and 79 were covered only by the CQR method.

Next, the APS method was compared with CQR for $\alpha = 0.1$. In this case, 35 samples were not covered by either method, 169 were covered only by the APS method, and 17 were covered only by the CQR method. Finally, it was observed that some countries are easier to predict than others. The per-country coverage results for the three methods are presented in Figure 6.3.

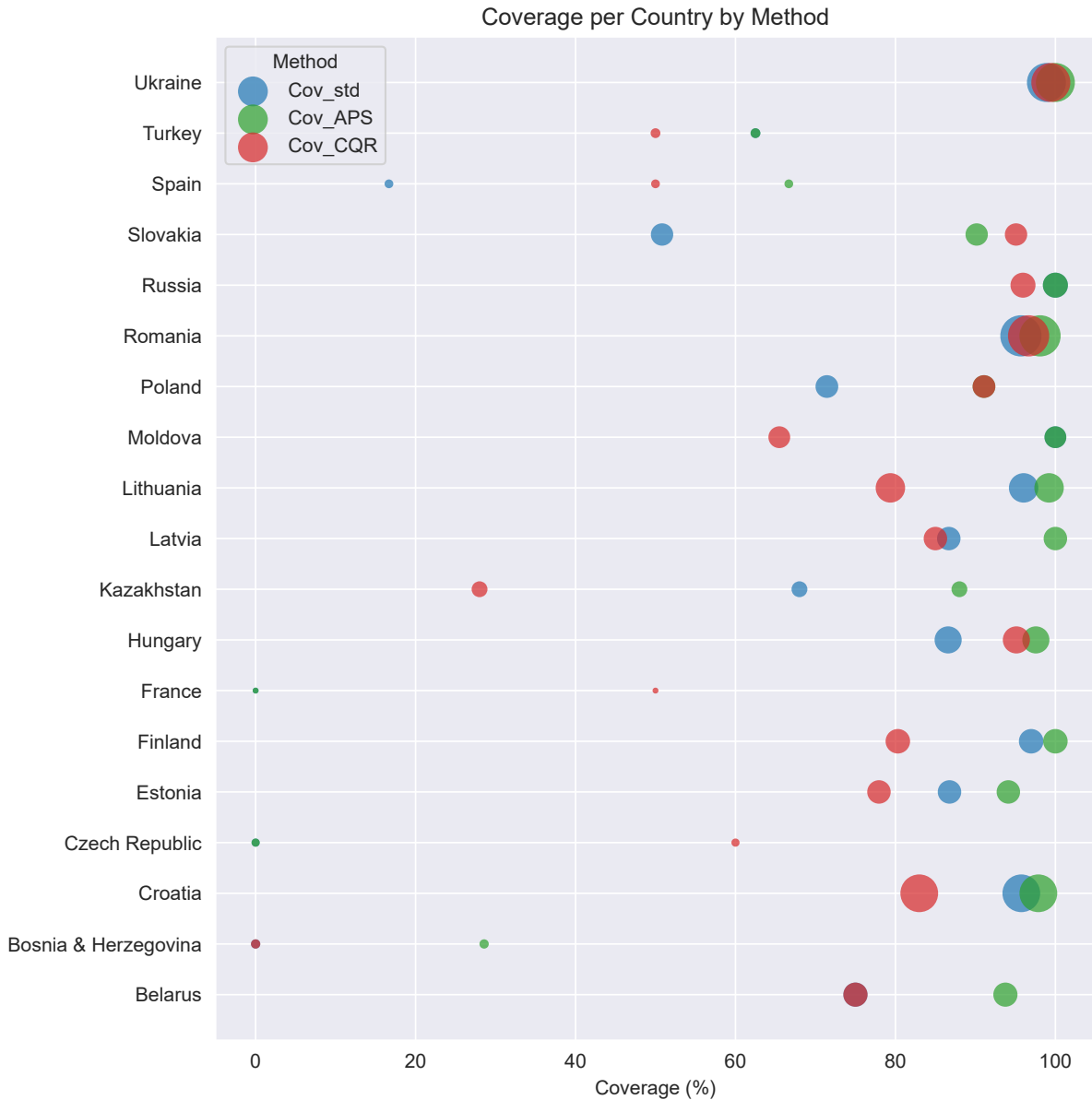


Figure 6.3: Coverage of three methods (std or LAC, APS, and CQR) across countries for $\alpha = 0.1$. Dots represent coverage values (%) per method, with color indicating the method and the marker size is proportional to the number of samples present in the country.

6.3.5 Results Per Genus

In this section, results are reported for the same CP pipelines (LAC, APS and CQR), but now with a separate RF model for each genus. *Fagus* has a total amount of 366, distributed over 6 different countries. Results for the CP methods are shown in table 6.7. *Quercus* has a total amount of 298, distributed over 15 different countries. Results for the CP methods are shown in table 6.8. *Betula* has a total amount of 386, distributed over 13 different countries. Results for the CP methods are shown in table 6.9. *Pinus* has a total amount of 207, distributed over 6 different countries. Results for the CP methods are shown in table 6.10.

Table 6.7: Results CQR, for *Fagus*: For different alphas, the different coverages and average predictions set size are reported. Additionally, the average set size for prediction sets including (correct) and excluding (incorrect) the true country are reported. Lastly, the average rank of the true country in the correct prediction sets are shown.

Method	Alpha	Cov.	Set size	Size correct	Size incorrect	Avg. rank true country
LAC						
	0.025	98.91	3.63	3.64	3.25	2.26
	0.05	98.91	3.63	3.64	3.25	2.26
	0.075	97.81	3.28	3.28	3.00	2.05
	0.1	93.72	2.55	2.53	2.87	1.66
APS						
	0.025	100.0	5.17	5.17	–	1.58
	0.05	100.0	5.17	5.17	–	1.58
	0.075	99.73	3.93	4.73	5.00	1.56
	0.1	97.81	3.93	3.93	4.13	1.49
CQR						
	0.025	100.0	7.96	7.96	–	2.46
	0.05	97.54	6.12	6.14	5.38	2.15
	0.075	95.08	4.89	4.90	4.58	1.94
	0.1	94.26	4.04	4.05	4.00	1.80

Table 6.8: Results CQR, for *Quercus*: For different alphas, the different coverages and average predictions set size are reported. Additionally, the average set size for prediction sets including (correct) and exluding (incorrect) the true country are reported. Lastly, the average rank of the true country in the correct prediction sets are shown.

Method	Alpha	Cov.	Set size	Size correct	Size incorrect	Avg. rank true country
LAC						
	0.025	97.31	7.37	7.37	7.25	3.79
	0.05	97.31	7.37	7.37	7.25	3.79
	0.075	92.60	5.30	5.32	5.09	2.96
	0.1	91.60	4.75	4.73	4.96	2.75
APS						
	0.025	98.66	9.33	9.33	9.25	2.05
	0.05	98.66	9.33	9.33	9.25	2.05
	0.075	95.63	6.51	6.50	6.61	1.87
	0.1	95.30	6.19	6.20	6.14	1.86
CQR						
	0.025	100.0	10.29	10.29	–	6.44
	0.05	94.95	6.07	6.01	6.40	3.48
	0.075	94.95	5.33	5.26	5.68	2.98
	0.1	92.93	4.46	4.41	4.64	2.49

Table 6.9: Results CQR, for *Betula*: For different alphas, the different coverages and average predictions set size are reported. Additionally, the average set size for prediction sets including (correct) and exluding (incorrect) the true country are reported. Lastly, the average rank of the true country in the correct prediction sets are shown.

Method	Alpha	Cov.	Set size	Size correct	Size incorrect	Avg. rank true country
LAC						
	0.025	97.93	4.92	4.93	4.38	3.07
	0.05	97.93	4.92	4.93	4.38	3.07
	0.075	95.85	3.77	3.77	3.88	2.41
	0.1	93.00	2.82	2.81	2.93	1.90
APS						
	0.025	100.0	6.26	6.26	–	1.61
	0.05	100.0	6.26	6.26	–	1.61
	0.075	97.67	4.56	4.56	4.44	1.51
	0.1	97.15	4.19	4.19	4.36	1.49
CQR						
	0.025	98.44	13.83	13.91	11.36	7.35
	0.05	98.44	9.68	10.06	6.75	4.56
	0.075	96.88	7.49	7.56	7.00	3.49
	0.1	96.88	6.19	6.21	6.07	3.15

Table 6.10: Results CQR, for *Pinus*: For different alphas, the different coverages and average predictions set size are reported. Additionally, the average set size for prediction sets including (correct) and excluding (incorrect) the true country are reported. Lastly, the average rank of the true country in the correct prediction sets are shown.

Method	Alpha	Cov.	Set size	Size correct	Size incorrect	Avg. rank true country
LAC						
	0.025	98.07	1.55	1.54	1.75	1.25
	0.05	98.07	1.55	1.54	1.75	1.25
	0.075	98.07	1.55	1.54	1.75	1.25
	0.1	98.07	1.55	1.54	1.75	1.25
APS						
	0.025	100.0	3.40	3.40	–	1.13
	0.05	100.0	3.40	3.40	–	1.13
	0.075	100.0	3.40	3.40	–	1.13
	0.1	100.0	3.40	3.40	–	1.13
CQR						
	0.025	98.55	11.22	11.20	12.50	5.60
	0.05	98.55	7.33	7.29	8.43	4.04
	0.075	94.20	5.12	5.03	6.50	3.18
	0.1	91.30	4.37	4.30	5.12	2.66

7. Discussion

In this section, the results are examined in greater detail and discussed in relation to the previously defined research questions: how model performance differs when using a continuous versus a discrete formulation of geographical space, and whether conformal prediction (CP) can provide more robust and trustworthy predictions of timber harvest origin.

First, the overall performance of both RF models aligns with expectations and is consistent with findings reported in the literature, even though they are slightly lower than those reported in similar studies such as Boeschoten et al. (2023) and Mortier et al. (2024). For the classification approach (Section 6.2.1), northern countries appear generally easier to classify. The Baltic states (Estonia, Latvia, and Lithuania) show strong performance, as do Finland and Russia. Other countries with relatively high accuracy include Moldova, Croatia, Romania, and Ukraine. Notably, Romania, Ukraine, and Croatia are also the countries with the largest number of samples. This highlights a well-known limitation: models require good coverage of reference data across the study area, and are currently limited by the small and fragmented reference datasets available (Low et al., 2022).

Countries with less than 10 samples unsurprisingly perform bad with only Turkey having an accuracy different than 0 (50.00%). Poland and Slovakia only achieve an accuracy of around 10%, despite having over 50 samples. For both Poland and Slovakia, over 50% of the predicted countries correspond to either Ukraine or Romania. Notably, Ukraine and Romania are the countries with the largest available datasets. These findings indicate that, particularly in Central-Eastern Europe, the model is affected by data imbalance and limited geographical representativeness, failing to distinguish less-represented countries.

Historically, sampling has often been based on convenience rather than optimal coverage for model performance. Given the considerable time and financial costs associated with collecting and analyzing samples, as well as geopolitical constraints (e.g., safety issues in conflict regions and restrictions under international agreements such as the Nagoya Protocol), these data limitations are unlikely to be resolved in the short term. A possible short-term solution is described in Section 7.3, where isotopic landscapes (isoscapes) are shortly discussed. These isoscapes have shown promising potential, as demonstrated by multiple authors (Truszkowski et al., 2025; Watkinson et al., 2020).

Furthermore, incorporating genus-level dummy variables into the model leads to improved performance. By feeding the model this extra information, we saw an improvement of 3.06% for the classification approach, and -22.80 km for the regression approach. When the focus would lie solely on improving the performance of one predictive model for all genera, our results indicate that genera dummies are crucial to enhance predictive performance. We advise to test whether dummies as species-level might have an even more substantial impact. However, in this master dissertation the focus was on making the outputs more robust, safe and reliable. The main emphasis is placed on the CP framework, which is discussed in the following Section 7.2.

Lastly, we look at the performances of the different RF models for each genus. *Pinus* clearly outperforms the others. This is easily explained when we take a more detailed look into the nature of the *Pinus* dataset. 207 samples are distributed across only six countries, namely Croatia, Finland, Lithuania, Hungary, Ukraine, and Russia, which are geographically well separated. In contrast, the other genera exhibit a spatial distribution more similar to the overall dataset, and their performance therefore more closely matches that of the RF model trained on all genera combined. When taking the weighted sum based on the relative fraction of each genus, the found 71.60% is therefore an expected overall result. Nevertheless, the strong performance of the *Pinus* RF model suggests that SIR and ME data differ sufficiently across these six countries to enable effective discrimination by the model. This again illustrates that clear geographically separated groups plays an important part in the ability to separate between geographical areas. The 90.34% of the *Pinus* RF model approaches the near-perfect results of more simplistic (i.e. binary classification) experiments, such as Boeschoten et al. (2022) and Wang et al. (2023).

When considering the approach of training one model per genus, but now using an RF regression model, *Fagus* performs notably better, whereas *Betula* performs substantially worse. Once again, this can largely be explained by the nature of the available data. *Fagus* samples are mainly concentrated in Central-Eastern countries such as Poland, Slovakia, Croatia, Hungary, Romania, and Ukraine. In other words, the samples are relatively clustered, making it easier to predict using a regression approach. In contrast, *Betula*'s samples originate from a much larger area, including 25 samples originating from Kazakhstan. These geographical outliers have a strong impact on the overall results, as they are systematically predicted further east, closer to the majority of *Betula* samples. Specifically, these 25 samples exhibit an average prediction error of 1539.19 km. Excluding them reduces the overall error substantially to 356.18 km. Similarly, *Quercus* contains 8 clear geographical outliers, situated in Spain and France. These samples are predicted 957.28 km away from their true location, however, due to their small number, they have only a limited effect on the overall error, which would be 275 km without the inclusion of

these 8 samples. *Pinus* showed the best performance in the classification pipeline. As discussed in the previous paragraph, these samples are distributed over a large geographical area. In particular, samples from Russia are predicted relatively far from their true locations (418.64 km on average). Nevertheless, despite this wide spatial distribution—ranging from northern Finland to Croatia—the overall performance for *Pinus* remains reasonably good.

Pinus scoring well for the classification and average for the regression despite having samples spread over a large area, is interesting and warrants further analysis. When looking at feature importance for the *Pinus* RF regression model, we find that the three most important features are all SIR features, namely: $\delta^{18}\text{O}_{\text{lipidextracted}}$ ($= 0.536919$), $\delta^2\text{H}_{\text{lipidextracted}}$ (0.238237), and $\delta^{34}\text{S}_{\text{lipidextracted}}$ (0.056081). For comparison, $\delta^{18}\text{O}_{\text{lipidextracted}}$ is also the most important feature across the other genus-specific RF models, but its relative importance differs: 0.304048 for *Betula*, 0.257813 for *Quercus*, and only the fourth most important feature for *Fagus* (0.074318). The unique thing about *Pinus*, compared to the other 3 genera present, is the fact that it is a conifer, unlike the other genera, which are angiosperms. The two clearly most important features ($\delta^{18}\text{O}_{\text{lipidextracted}}$ ($= 0.536919$), $\delta^2\text{H}_{\text{lipidextracted}}$ (0.238237)) are both heavily related to water-uptake and usage. For the classification RF model, the same two features are the most important, (respectively 0.188954 , and 0.160636). As discussed in Sections 2.1.2.2 and 2.1.2.3, fractionation of O and H isotopes strongly depends on water sources and evapotranspiration in leaves. Stomatal conductance, the rate at which gases enter and water vapor exits through stomata, differs substantially between conifers (*Pinus*) and angiosperms (*Fagus*, *Betula*, and *Quercus*), with values being more than 30% higher in angiosperms (Lusk et al., 2003). A possible explanation of why O and H isotopes are more important for conifers could be that higher stomatal conductance in angiosperms leads to faster evapotranspiration, meaning the residence time of water is longer in conifers than in angiosperms (Meinzer et al., 2006). Consequently, kinetic fractionation processes have less time to act in angiosperms, leading to less enrichment, whereas conifers exhibit relatively more enriched SIR values, resulting in heavier isotopes on average. For *Pinus* the mean values for $\delta^{18}\text{O}_{\text{lipidextracted}}$ and $\delta^2\text{H}_{\text{lipidextracted}}$ are respectively 0.28 and 1.14 , whereas for the other three angiosperms these are -0.012 and -0.025 . This observation supports our hypothesis that O and H isotopes are more enriched, most likely due to the different water usage of conifers. This difference in water use, together with the resulting longer residence time, may be contributing to greater stability in the values and a stronger expression of geographical variability. Consequently, this could facilitate more accurate timber origin assignment. However, it is important to note that data for *Pinus* is highly unbalanced, and thus that making conclusions should be done with caution.

7.1 Discussion Conformal Prediction performance

We first discuss results of CP with a singular RF model as underlying model, followed by the discussion of the CP pipeline with a separate RF model for each genus.

One of the main advantages of CP is that it provides guaranteed statistical coverage. The algorithm is designed to aim at a coverage higher than $1 - \alpha$. In our results, all three approaches, LAC CP, APS CP, and CQR, satisfy this requirement. We also see that, for all three methods, the average prediction set size for sets containing the true country is smaller than for sets that do not contain the true country. This behavior is expected and indicates that the CP framework operates as intended: when the underlying model is less certain, larger prediction sets (or areas) are produced to maintain coverage. In general, we find that the prediction set size highly depends on the non-conformity function (besides also the error rate), as similarly reported by Dhillon et al. (2024).

An α of 0.1 is commonly used in literature and can be considered a standard choice. For this value, the LAC CP method yields an average prediction set size of 2.38. Given that the RF classification model achieves an accuracy of 69.73%, this implies that adding approximately 1.38 extra countries (in practice one or two) increases the accuracy to 90%, while adding around 5.22 countries increases accuracy to 98.71%.

APS is clearly a more conservative approach. The average rank of the true country in the prediction set lies between 1.6 and 1.7 for α values ranging from 0.025 to 0.1. This is lower than for the LAC CP. The lower average rank of the true country under APS arises because APS constructs prediction sets by accumulating probability mass in descending order of class likelihood. As a result, highly probable labels are included first, and the true label, assuming a well-calibrated model, is typically assigned a high rank. This leads to prediction sets that are concentrated among top-ranked labels. This also explains why the coverage decreases only slowly as α increases. As α grows, smaller prediction sets are allowed; however, since the true country is typically ranked near the top, removing lower-ranked countries has limited impact on coverage. Combined with the fact that APS is inherently a more conservative approach, which produces larger prediction sets, makes it that coverage tend to consistently stay high. It is important to note that this seemingly unusual behavior is not wrong, as coverage is guaranteed to be at least $1 - \alpha$ or higher. Two studies comparing LAC and APS classification CP on images found very similar results. APS is by definition a more conservative approach and produces larger sets, which can lead to "overcoverage", as also found by (Pourkamali-Anaraki, 2025; Angelopoulos et al., 2020).

The CQR's approach returns even larger average prediction set size. However, this is partly due to how these sets are made. Because the predicted region is contin-

uous while country boundaries are discrete, multiple countries can partially overlap with the region. As a result, even countries with small overlaps receive non-zero probability and may be included in the ranking. This often leads to larger prediction sets, as probability can be distributed across several partially overlapping countries. Therefore, comparisons of average set size with classification-based CP should be interpreted with caution. A possible improvement for future research is defining a different method of making this ranking, as the middle of the predicted area is inherently more likely than the borders of this area, which is not taken into account currently. Incorporating a spatial weighting scheme that reflects higher probability density near the center could improve ranking accuracy and reduce prediction set sizes. Additionally, the average rank of the true country is higher (i.e. worse) for CQR compared to classification-based methods. This again results from the translation of predicted areas to countries. Unfortunately, no examples in literature were found where the prediction sets with a regression pipeline were transformed to a prediction set with labels. Guha et al. (2024) translated continuous variables beforehand and then performed classification CP. Applied to our problem, this would mean we cluster samples which are closely spatially located, disregarding borders, creating new labels. We could then apply one of the classification CP methods and get prediction sets with these newly defined labels.

When analyzing the coverage per country, we see similar trends to the RF classifier results. Countries such as the Baltic states, Croatia, Moldova, Russia, and Ukraine exhibit high coverage, whereas countries with limited sample availability tend to show lower coverage. Nevertheless, the coverage obtained with the CQR approach appears to be less irregular across countries. In particular, no country achieves full (100%) coverage, and only one country exhibits zero coverage. Countries located further from the most densely sampled regions, such as Kazakhstan, Turkey, Finland, and Spain, generally score worse. This difference can be explained by the underlying modeling assumptions. The classification approach does not explicitly account for geographical proximity, treating all countries independently. As a result, if sufficient reference samples are available, the model can effectively distinguish even geographically isolated countries. In contrast, the continuous CQR approach does capture the spatial structure, making predictions for countries that are geographically further away from the central, densely sampled regions more challenging.

Based on the previously discussed result, we conclude that all CP have merits and limitations. CQR depends less on the amount of reference data available for a country, although a more radical translation of predicted area to a country ranking might give different results. When samples are located further away geographically, a classification CP pipeline could be a better approach. APS guarantees zero empty prediction sets, but in return appears to be slightly less efficient (i.e. larger prediction sets on average).

Now, we discuss the results for the different CP methods, using a separate model for each genus. A first general observation is that, for small values of α , the average set size of prediction sets excluding the true country tends to be slightly smaller than that of sets including the true country, although the difference is minimal. Additionally, for small α , the coverage approaches 100%, implying that only a limited number of sets exclude the true country, and thus the corresponding average set size is based on very few observations. As α increases, more prediction sets exclude the true country, and the average set size of these sets tends to exceed that of the sets including the true country, which is the expected behavior. Another general impression is that coverages tend to be slightly higher than expected. This is not wrong, as CP methods guarantee of at least $1 - \alpha$, rather than exactly $1 - \alpha$.

Another key reported result is efficiency, expressed in terms of the average set size. For $\alpha = 0.1$, *Pinus* consistently performs better. Overall, efficiency closely mirrors the performance of the underlying RF models. For the classification-based CP methods, *Pinus* clearly achieves the best results, while *Quercus* performs the worst. In contrast, for the CQR method, *Fagus* performs best and *Betula* performs worst. These observations are consistent with those obtained from the RF classification and regression models, respectively. Compared to the CP pipeline based on a single RF model, *Quercus* performs worse in the classification-based methods, whereas *Pinus* shows improved efficiency across all three methods. Because *Pinus* scores so well in classification pipelines, recall the accuracy of the RF classifier for *Pinus* being 90.34%, the predictions for all α 's tested are the same for the LAC and APS method. Only when we would make α even larger, meaning allowing more room for error, would differences in the resulting prediction sets begin to emerge.

7.2 Discussion Conformal Prediction merits

Robust and safe predictions are seen as increasingly important in machine learning, particularly in contexts where predictions have legal implications, as is the case with timber origin assignment. CP's prediction sets gives possible alternatives to the most likely answer, which can be crucial. By providing a finite set of possibilities, they can guide decision makers to focus their attention on a limited number of possibilities. Furthermore, the easy to understand end-result of CP makes it highly attractive, as it offers transparent uncertainty estimates and gives users the flexibility to adapt the desired confidence level. Another advantages is the fact that it works without assuming a specific data distribution and can provide reliable guarantees.

Beside the useful and informative output, CP also contains a big advantage in the fact that coverage is guaranteed, regardless of the underlying model. If the origin prediction would be much more difficult, resulting in much lower uncertainties, the

prediction set size would just increase. CP wraps around the underlying model, and ensures coverage according to a certain α . Combined with the fact that CP works for both classification as regression problems, this makes it a useful technique for many applications.

CP's key advantage is hence that it provides valid coverage guarantees independently of the underlying model. This makes it robust to model flaws and broadly applicable across different machine learning settings. Even though that this guarantee comes at the cost of potentially larger prediction sets, especially when the underlying model is weak, still in many settings this advantage is a crucial one.

RF models are well suited for handling large, high-dimensional datasets with missing values and do not rely on strong statistical assumptions. However, they may struggle with high-cardinality categorical variables or unbalanced data and can be sensitive to hyperparameter choices (Zhu, 2020). As discussed in Section 3.1.2, Gaussian Process (GP) models are another commonly used approach in origin assignment. However, GP models can be challenging to apply in practice due to the need for careful design choices, such as kernel selection, and their computational cost for large datasets (Schulz et al., 2018). Poor choices in these aspects may lead to misleading predictions.

Wrapping the model with conformal prediction is beneficial because it adds reliable uncertainty guarantees on top, regardless of these modeling difficulties or possible model flaws. Given the aforementioned advantages of CP, we believe that CP can offer substantial added value in timber origin assignment, particularly in terms of improving both the interpretability and robustness of predictions.

7.3 Possible improvements

CP is promising and should be explored further. We also already look for possible improvement on the current application.

A particularly promising direction, already explored to some extent, is the development of isoscapes and elementscapes. As discussed previously, and as clearly reflected in the results of this thesis, a major limitation remains the lack of reference data. Machine learning offers a way to make traceability methods more predictive and efficient by integrating atmospheric, meteorological, and environmental drivers, enabling geospatial models to estimate SIR and ME values beyond sampled sites and reducing reference sampling needs. While recent efforts in the literature attempt to address the lack of SIR reference data by predicting isoscapes with promising results (Truszkowski et al., 2025; Watkinson et al., 2020), less focus has been on elementscapes.

Next, another potential way to improve the performance of machine learning models is the integration of additional multimodal data sources; for example, combining SIR, ME and genetic data. While such combinations have shown synergistic gains (Boeschoten et al., 2025; Sarkar et al., 2025), and spatial models of genetic timber data exist (Degen et al., 2025), there is currently no comprehensive framework that jointly models SIR, ME, and genetic data within a unified probabilistic spatial hierarchy. In addition, satellite imagery represents another promising data source for improving assignment accuracy. Incorporating land cover classifications, or even more advanced approaches, such as detecting signs of selective logging using radar satellite data, offers unique opportunities. Specifically, these high-resolution remote-sensing data can be used to detect the occurrence and scale of canopy disturbances and opening of logging roads to establish the likelihood that selective logging or land conversion to agricultural land has taken place. These developments enable the creation of verification tools that assess whether logging or land conversion has likely occurred at a claimed origin, and allow for the integration of dynamic probabilities into assignment models.

Lastly, a possible improvement for the CP algorithm could be the implementation of a more hierarchical CP approach. Recent research has introduced hierarchical conformal prediction methods, which explicitly account for class hierarchies in both the construction and interpretation of prediction sets (Mortier et al., 2025). Integrating such methods into traceability applications could offer significant advantages, as they provide uncertainty-aware predictions that align with the underlying spatial structure. This would enable a reduction in the predicted overall area, since prediction sets could be restricted to specific regions within countries rather than the full country area.

Conclusion

This dissertation demonstrates that classic machine learning models, such as RF, have certain limitations when it comes to timber origin assignment. The main limitation is due to the lack of reference data available. CP can make predictions more robust and safe, via returning prediction sets, rather than a singular label.

In general, we found that both a classification model and a regression model are viable options as underlying models in a CP for timber origin assignment. Results vary based on data available and possibly on genus as well. For the RF regressor, we demonstrated that geographical outliers can have a severe impact on overall score, as is the case with *Betula*. On the contrary, for the RF classifier, *Pinus* performed clearly better than the other genera. Whether this is due to the nature of the reference data, the different water usage, or a combination of both, future research has to clarify. Overall, the results for the three CP approaches closely reflect the performance of the underlying model. In particular, within the pipeline that employs a separate model per genus, the average set sizes (i.e., the efficiency of the CP method) are small when the underlying model is confident, and larger when the model exhibits greater uncertainty.

In addition, by converting a prediction area to a set of countries, a comparison between a classification and a regression method becomes possible. Overall, the classification-based methods appear to be more efficient, as reflected by a smaller average set size. However, a key aspect is the transformation from the prediction area to the final prediction set with countries. In this thesis, a relatively simple approach is used, where countries are assigned based on the proportion of their area overlapping with the predicted region. This approach can be further improved in future work, as locations closer to the center of the prediction area should intuitively carry more weight than those near its boundaries. Therefore, caution is warranted when directly comparing discrete and continuous approaches, given the influence of this transformation step.

Nevertheless, CQR does appear to have some merits compared to the classification methods. Due to the on average larger prediction set size, CQR is more likely to include countries that are underrepresented in the dataset. Additionally, the results indicate that the efficiency of CQR improves substantially when the available data are more densely clustered, as is the case for *Fagus*. The fact that CQR does not adhere to country borders is also an important advantage. While a classification-based approach may yield confident predictions, it remains constrained by predefined na-

tional boundaries. As a result, two predictions may appear identical at the country level, even though the underlying model may be substantially more certain for one sample than the other. In contrast, CQR is not restricted by such boundaries and can capture this difference in uncertainty through the size of the predicted area, allowing for more nuanced and informative outputs.

Beside the performance, CP offers clear advantages in the context of timber origin assignment. The EUDR follows a due diligence framework, in which the ability of operators to explicitly control the acceptable level of error is highly beneficial. Providing a set of plausible alternatives, rather than a single most likely prediction, is particularly valuable in legal contexts such as origin verification. This intuitive and interpretable output makes conformal prediction especially attractive, as it enables transparent uncertainty quantification while allowing users to adjust the desired confidence level.

Bibliography

- Abraham, U. I. (2020). Deforestation, air quality degradation and increased cardiopulmonary diseases. *SRMS JOURNAL OF MEDICAL SCIENCE*, 5(02).
- Aitchison, J. (1983). Principal component analysis of compositional data. *Biometrika*, 70(1):57–65.
- Alnuaimi, A. F. and Albaldawi, T. H. (2024). An overview of machine learning classification techniques. In *BIO Web of Conferences*, volume 97, page 00133. EDP Sciences.
- Angelopoulos, A., Bates, S., Malik, J., and Jordan, M. I. (2020). Uncertainty sets for image classifiers using conformal prediction. *arXiv preprint arXiv:2009.14193*.
- Angelopoulos, A. N. and Bates, S. (2023). Conformal prediction: A gentle introduction. *Foundations and Trends in Machine Learning*, 16(4):494–591.
- Aparicio-Muriana, M. M., Hong, Y., Chilaka, C. A., Quinn, B., Montes-Niño, A. M., Birse, N., and Elliott, C. T. (2025). Mapping the global origins of soybean: a study using icp-ms and chemometrics. *npj Science of Food*, 9(1):263.
- Baccini, A., Goetz, S. J., Walker, W., Laporte, N., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P., Dubayah, R., Friedl, M., et al. (2012). Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature climate change*, 2(3):182–185.
- Biau, G. and Scornet, E. (2016). A random forest guided tour. *Test*, 25(2):197–227.
- Boeschoten, L. E., Rocha Venancio Meyer-Sand, B., Boom, A., Bouka, G. U. D., Ciliane-Madikou, J. C., Engone Obiang, N. L., Guieshon-Engongoro, M., de Groot, A., Loumeto, J. J., Mbika, D.-m. M., et al. (2025). Combined genetic and chemical methods boost the precision of tracing illegal timber in central africa. *Communications Earth & Environment*, 6(1):789.
- Boeschoten, L. E., Sass-Klaassen, U., Vlam, M., Comans, R. N., Koopmans, G. F., Meyer-Sand, B. R. V., Tassiamba, S. N., Tchamba, M. T., Zanguim, H. T., Zemtsa, P. T., et al. (2022). Clay and soil organic matter drive wood multi-elemental composition of a tropical tree species: Implications for timber tracing. *Science of the Total Environment*, 849:157877.

- Boeschoten, L. E., Vlam, M., Sass-Klaassen, U., Rocha Venâncio Meyer-Sand, B., Adzkie, U., Bouka, G. D., Ciliane-Madikou, J. C., Engone Obiang, N. L., Guieshon-Engongoro, M., Loumeto, J. J., et al. (2023). A new method for the timber tracing toolbox: applying multi-element analysis to determine wood origin. *Environmental Research Letters*, 18(5):054001.
- Boner, M., Sommer, T., Erven, C., and Förstel, H. (2007). Stable isotopes as a tool to trace back the origin of wood. In *Proceedings of the international workshop "Fingerprinting methods for the identification of timber origins*, pages 3–5.
- Booksh, K. S., Celani, C. P., Ralbovsky, N. M., and Smith, J. P. (2025). Assessing classification models of pharmaceuticals with conformal prediction. *Journal of Chemometrics*, 39(3):e70017.
- Bowen, G. J., Wassenaar, L. I., and Hobson, K. A. (2005). Global application of stable hydrogen and oxygen isotopes to wildlife forensics. *Oecologia*, 143(3):337–348.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- Camin, F., Boner, M., Bontempo, L., Fauhl-Hassek, C., Kelly, S. D., Riedl, J., and Rossmann, A. (2017). Stable isotope techniques for verifying the declared geographical origin of food in legal cases. *Trends in Food Science & Technology*, 61:176–187.
- Cheadle, C., Vawter, M. P., Freed, W. J., and Becker, K. G. (2003). Analysis of microarray data using z score transformation. *The Journal of molecular diagnostics*, 5(2):73–81.
- Craine, J. M., Brookshire, E., Cramer, M. D., Hasselquist, N. J., Koba, K., Marin-Spiotta, E., and Wang, L. (2015). Ecological interpretations of nitrogen isotope ratios of terrestrial plants and soils. *Plant and Soil*, 396(1):1–26.
- Cresswell, J. C., Sui, Y., Kumar, B., and Vouitsis, N. (2024). Conformal prediction sets improve human decision making.
- Cutler, A., Cutler, D. R., and Stevens, J. R. (2012). Random forests. In *Ensemble machine learning*, pages 157–175. Springer.
- da Silva Araújo, M. G., Amorim, D. J., Almeida, R. F., Mazzi, E., Batista, A. C. G., Souza-Silva, I. M., Brunello, A. T., Nardoto, G. B., Bowen, G. J., and Martinelli, L. A. (2025). Evaluating an isotopic model for tracing the geographic origin of soybeans in brazil. *Food Chemistry*, page 147586.
- DeFries, R. S., Rudel, T., Uriarte, M., and Hansen, M. (2010). Deforestation driven by urban population growth and agricultural trade in the twenty-first century. *Nature geoscience*, 3(3):178–181.

-
- Degen, B., Yanbaev, Y., and Müller, N. A. (2025). Machine learning techniques for continuous genetic assignment of geographic origin of forest trees. *Plos one*, 20(6):e0324994.
- DeNiro, M. J. and Epstein, S. (1979). Relationship between the oxygen isotope ratios of terrestrial plant cellulose, carbon dioxide, and water. *Science*, 204(4388):51–53.
- Dhillon, G. S., Deligiannidis, G., and Rainforth, T. (2024). On the expected size of conformal prediction sets. In *International Conference on Artificial Intelligence and Statistics*, pages 1549–1557. PMLR.
- Duncker, P. S., Raulund-Rasmussen, K., Gundersen, P., Katzensteiner, K., De Jong, J., Ravn, H. P., Smith, M., Eckmüllner, O., and Spiecker, H. (2012). How forest management affects ecosystem services, including timber production and economic return: synergies and trade-offs. *Ecology and Society*, 17(4).
- EarthSight (2023). “Conflict timber” worth over \$1bn sold in US since Ukraine invasion. Press release.
- Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., and Barcelo-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical geology*, 35(3):279–300.
- Ehleringer, J. R., Casale, J. F., Lott, M. J., and Ford, V. L. (2000). Tracing the geographical origin of cocaine. *Nature*, 408(6810):311–312.
- EU (2023). Commission implementing regulation. *Official Journal of the European Union*, L 207:77–82. C/2023/5564.
- European Commission (2026). Regulation on deforestation-free products. https://environment.ec.europa.eu/topics/forests/deforestation/regulation-deforestation-free-products_en. Accessed: 2026-03-16.
- Finch, K., Espinoza, E., Jones, F. A., and Cronn, R. (2017). Source identification of western oregon douglas-fir wood cores using mass spectrometry and random forest classification. *Applications in Plant Sciences*, 5(5):1600158.
- Flores, B. M., Montoya, E., Sakschewski, B., Nascimento, N., Staal, A., Betts, R. A., Levis, C., Lapola, D. M., Esquivel-Muelbert, A., Jakovac, C., et al. (2024). Critical transitions in the amazon forest system. *Nature*, 626(7999):555–564.
- Food and Agriculture Organization of the United Nations (2024). *The State of the World’s Forests 2024: Forest-Sector Innovations towards a More Sustainable Future*. FAO, Rome.
- Getirana, A., Libonati, R., and Cataldi, M. (2021). Brazil is in water crisis—it needs a drought plan. *Nature*, 600(7888):218–220.

- Gislason, P. O., Benediktsson, J. A., and Sveinsson, J. R. (2004). Random forest classification of multisource remote sensing and geographic data. In *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium*, volume 2, pages 1049–1052. IEEE.
- Gomes, M. D. G., Costa, F. J. V., Nascentes, C. C., Martinelli, L. A., and Nardoto, G. B. (2025). Multi-elemental analysis for the determination of the geographic origin of tropical timber from the brazilian legal amazon. *Forests*, 16(8):1284.
- Gomes, V. H., Vieira, I. C., Salomão, R. P., and ter Steege, H. (2019). Amazonian tree species threatened by deforestation and climate change. *Nature Climate Change*, 9(7):547–553.
- Gori, Y., Deklerck, V., and Camin, F. (2025). Stable isotope ratio analysis to determine the geographical origin of timber: A review. *TrAC Trends in Analytical Chemistry*, page 118448.
- Greenacre, M. (2021). Compositional data analysis. *Annual Review of Statistics and Its Application*, 8(1):271–299.
- Guha, E., Natarajan, S., Möllenhoff, T., Khan, M. E., and Ndiaye, E. (2024). Conformal prediction via regression-as-classification. *arXiv preprint arXiv:2404.08168*.
- Hoaglin, D. C., Iglewicz, B., and Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 81(396):991–999.
- Hoefs, J. (2009). Variations of stable isotope ratios in nature. In *Stable Isotope Geochemistry*, chapter 3, pages 93–227. Springer, Berlin, Heidelberg.
- INTERPOL (2016). Uncovering the risks of corruption in the forestry sector. Published 9 December 2016.
- Jahren, A. H., Sternberg, L. S. L., et al. (2002). Eocene meridional weather patterns reflected in the oxygen isotopes of arctic fossil wood. *GSA today*, 12(1):4–9.
- James, G., Witten, D., Hastie, T., Tibshirani, R., et al. (2013). *An introduction to statistical learning: with applications in R*, volume 103. Springer.
- Kumar, A., Dhanka, S., Sharma, A., Sharma, A., Nain, M., Kumar, P., Gupta, A. R., Bansal, J., Saxena, N. K., and Pant, R. (2025). A comprehensive review of bias in ai, ml, and dl models: Methods, impacts, and future directions. *Archives of Computational Methods in Engineering*, pages 1–31.
- Kyere-Boateng, R. and Marek, M. V. (2021). Analysis of the social-ecological causes of deforestation and forest degradation in ghana: Application of the dpsir framework. *Forests*, 12(4).

-
- Köthke, M., Lippe, M., and Elsasser, P. (2023). Comparing the former eutr and upcoming eudr: Some implications for private sector and authorities. *Forest Policy and Economics*, 157:103079.
- Leal Filho, W., Dinis, M. A. P., Canova, M. A., Cataldi, M., da Costa, G. A. S., Enrich-Prast, A., Symeonakis, E., and Brearley, F. Q. (2025). Managing ecosystem services in the brazilian amazon: the influence of deforestation and forest degradation in the world's largest rain forest. *Geoscience Letters*, 12(1):24.
- Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411.
- Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., and Wasserman, L. (2018). Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111.
- Leung, R. (2021). Empirical observations on the effects of data transformation in machine learning classification of geological domains. *arXiv preprint arXiv:2106.05855*.
- Low, M. C., Schmitz, N., Boeschoten, L. E., Cabezas, J. A., Cramm, M., Haag, V., Koch, G., Meyer-Sand, B. R., Paredes-Villanueva, K., Price, E., et al. (2022). Tracing the world's timber: the status of scientific verification technologies for species and origin identification. *Iawa Journal*, 44(1):63–84.
- Lusk, C. H., Wright, I., and Reich, P. B. (2003). Photosynthetic differences contribute to competitive advantage of evergreen angiosperm trees over evergreen conifers in productive habitats. *New Phytologist*, 160(2):329–336.
- Maes, J., TELLER, A., ERHARD, M., CONDE, S., VALLECILLO, R. S., BARREDO, C., PARACCHINI, M.-L., MALAK, D. A., TROMBETTI, M., VIGIAK, O., et al. (2021). Eu ecosystem assessment.
- Maor, R., Truszkowski, J., Ablett, F., Jennings, H., Walker, H., Dunn, J., Norman, M., Carrasco, R., Jaime-Arteaga, M., Miles-Bunch, I., et al. (2026). High-resolution soybean tracing for deforestation-free supply chains. *Communications Earth & Environment*, 7(1):310.
- Marques, L. (2024). Crossing tipping points in the amazon rainforest: the decisive decade. *The Highlander Journal*, 3(2):4–33.
- Mavrellis, C. (2017). Transnational crime and the developing world. Published March 27, 2017.
- McKinney, C. R., McCrea, J. M., Epstein, S., Allen, H., and Urey, H. C. (1950). Improvements in mass spectrometers for the measurement of small differences in isotope abundance ratios. *Review of Scientific Instruments*, 21(8):724–730.

- Mehrtens, H., Bucher, T., and Brinker, T. J. (2023). Pitfalls of conformal predictions for medical image classification. In *International workshop on uncertainty for safe utilization of machine learning in medical imaging*, pages 198–207. Springer.
- Meinshausen, N. and Ridgeway, G. (2006). Quantile regression forests. *Journal of machine learning research*, 7(6).
- Meinzer, F. C., Brooks, J., Domec, J.-C., Gartner, B. L., Warren, J., Woodruff, D., Bible, K., and Shaw, D. (2006). Dynamics of water transport and storage in conifers studied with deuterium and heat tracing techniques. *Plant, cell & environment*, 29(1):105–114.
- Mensah, P., Pimenta, A. S., de Melo, R. R., Amponsah, J., Tuo, G., Chakurah, I., Ampadu, S. D., Buckman, I., Nikoi, M., Minkah, E., et al. (2025). The global supply chain of wood products: A literature review. *Forests*, 16(7):1036.
- Moreira, G. R., Fernandes, E. A. D. N., Bacchi, M. A., Gonzaga, C. L., and de Lima, R. C. (2024). Elemental profile of ipê species to support timber tracing in amazon rainforest. *Trees, Forests and People*, 18:100689.
- Mortier, T., Javanmardi, A., Sale, Y., Hüllermeier, E., and Waegeman, W. (2025). Conformal prediction in hierarchical classification with constrained representation complexity. *arXiv preprint arXiv:2501.19038*.
- Mortier, T., Truszkowski, J., Norman, M., Boner, M., Buliga, B., Chater, C., Jennings, H., Saunders, J., Sibley, R., Antonelli, A., et al. (2024). A framework for tracing timber following the ukraine invasion. *Nature Plants*, 10(3):390–401.
- Muccio, Z. and Jackson, G. P. (2009). Isotope ratio mass spectrometry. *Analyst*, 134(2):213–222.
- Nadras, S., Mazlan, R., Hussain, H., and Md Shah, I. (2024). The european union deforestation-free regulation (eudr): assessing impacts and strategies for malaysian and the global oil palm industry. *Journal of Sustainability Science and Management*, 19(6):54–74.
- Nallapati, R. (2004). Discriminative models for information retrieval. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 64–71.
- Nellemann, C. (2012). Green carbon, black trade: illegal logging, tax fraud and laundering in the world’s tropical forests.
- Osburn, M. R., Dawson, K. S., Fogel, M. L., and Sessions, A. L. (2016). Fractionation of hydrogen isotopes by sulfate-and nitrate-reducing bacteria. *Frontiers in microbiology*, 7:1166.

-
- Pendrill, F., Persson, U. M., Godar, J., and Kastner, T. (2019a). Deforestation displaced: trade in forest-risk commodities and the prospects for a global forest transition. *Environmental Research Letters*, 14(5):055003.
- Pendrill, F., Persson, U. M., Godar, J., Kastner, T., Moran, D., Schmidt, S., and Wood, R. (2019b). Agricultural and forestry trade drives large share of tropical deforestation emissions. *Global environmental change*, 56:1–10.
- Pereira, N. S., Sial, A. N., Pinheiro, P. B., Freitas, F. L., and Silva, A. M. (2021). Carbon and oxygen stable isotopes of freshwater fish otoliths from the são francisco river, northeastern brazil. *Anais da Academia Brasileira de Ciências*, 93(1):e20191050.
- Pourkamali-Anaraki, F. (2025). Aerial image classification in scarce and unconstrained environments via conformal prediction. *arXiv preprint arXiv:2504.17655*.
- Reddington, C., Butt, E., Ridley, D., Artaxo, P., Morgan, W., Coe, H., and Spracklen, D. (2015). Air quality and human health improvements from reductions in deforestation-related fire in brazil. *Nature Geoscience*, 8(10):768–771.
- Roden, J., Saurer, M., and Siegwolf, R. T. (2022). Probing tree physiology using the dual-isotope approach. In *Stable isotopes in tree rings: Inferring physiological, climatic and environmental responses*, pages 463–479. Springer.
- Romano, Y., Patterson, E., and Candes, E. (2019). Conformalized quantile regression. *Advances in neural information processing systems*, 32.
- Rousseeuw, P. J. and Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical association*, 88(424):1273–1283.
- Sachse, D., Billault, I., Bowen, G. J., Chikaraishi, Y., Dawson, T. E., Feakins, S. J., Freeman, K. H., Magill, C. R., McNerney, F. A., Van Der Meer, M. T., et al. (2012). Molecular paleohydrology: interpreting the hydrogen-isotopic composition of lipid biomarkers from photosynthesizing organisms. *Annual Review of Earth and Planetary Sciences*, 40(1):221–249.
- Sarkar, S., Yousuf, R. B., Wang, L., Mayer, B., Mortier, T., Deklerck, V., Truszkowski, J., Simeone, J. C., Norman, M., Saunders, J., et al. (2025). Chasing the timber trail: Machine learning to reveal harvest location misrepresentation. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 4796–4805.
- Sassen, M., Sheil, D., Giller, K. E., and ter Braak, C. J. (2013). Complex contexts and dynamic drivers: Understanding four decades of forest loss and recovery in an east african protected area. *Biological Conservation*, 159:257–268.
- Schulz, E., Speekenbrink, M., and Krause, A. (2018). A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of mathematical psychology*, 85:1–16.

- Sharp, Z. (2017). Principles of stable isotope geochemistry.
- Siegwolf, R. T., Savard, M. M., Grams, T. E., and Voelker, S. (2022). Impact of increasing CO_2 , and air pollutants (NO_x , SO_2 , O_3) on the stable isotope ratios in tree rings. In *Stable isotopes in tree rings: inferring physiological, climatic and environmental responses*, pages 675–710. Springer.
- Song, X.-P., Hansen, M. C., Potapov, P., Adusei, B., Pickering, J., Adami, M., Lima, A., Zalles, V., Stehman, S. V., Di Bella, C. M., et al. (2021). Massive soybean expansion in south america since 2000 and implications for conservation. *Nature sustainability*, 4(9):784–792.
- Sturchio, N. C., Beloso Jr, A., Heraty, L. J., Wheatcraft, S., and Schumer, R. (2014). Isotopic tracing of perchlorate sources in groundwater from pomona, california. *Applied geochemistry*, 43:80–87.
- Tcherkez, G. and Tea, I. (2013). $^{32}\text{S}/^{34}\text{S}$ isotope fractionation in plant sulphur metabolism.
- Tilman, D., Clark, M., Williams, D. R., Kimmel, K., Polasky, S., and Packer, C. (2017). Future threats to biodiversity and pathways to their prevention. *Nature*, 546(7656):73–81.
- Tipple, B. J., Berke, M. A., Doman, C. E., Khachatryan, S., and Ehleringer, J. R. (2013). Leaf-wax n-alkanes record the plant–water environment at leaf flush. *Proceedings of the National Academy of Sciences*, 110(7):2659–2664.
- Truszkowski, J., Maor, R., Yousuf, R. B., Biswas, S., Chater, C., Gasson, P., McQueen, S., Norman, M., Saunders, J., Simeone, J., et al. (2025). A probabilistic approach to estimating timber harvest location. *Ecological Applications*, 35(1):e3077.
- Van Nguyen, T. T., Heuchenne, C., and Tran, K. P. (2022). Machine learning for compositional data analysis in support of the decision making process. In *Machine Learning and Probabilistic Graphical Models for Decision Support Systems*, pages 184–215. CRC Press.
- Vovk, V., Gammerman, A., and Shafer, G. (2022). *Algorithmic Learning in a Random World*. Springer, Cham, 2 edition.
- Wang, H., Liu, H., Lu, B., Ma, M., Chen, J., and Nie, J. (2023). Origin authentication of european and american ash (*fraxinus* spp.) based on stable isotope ratio and elemental characteristics combined with chemometrics methods. *Chemosensors*, 11(10):536.
- Wang, T. and Dobriban, E. (2026). Optimal decision-making based on prediction sets. *arXiv preprint arXiv:2602.00989*.

-
- Watkinson, C. J., Gasson, P., Rees, G. O., and Boner, M. (2020). The development and use of isoscapes to determine the geographical origin of quercus spp. in the united states. *Forests*, 11(8):862.
- Watkinson, C. J., Rees, G. O., Hofem, S., Michely, L., Gasson, P., and Boner, M. (2022). A case study to establish a basis for evaluating geographic origin claims of timber from the solomon islands using stable isotope ratio analysis. *Frontiers in Forests and Global Change*, 4:645222.
- Weisse, M., Goldman, E., and Carter, S. (2024). Tropical forest loss drops steeply in brazil and colombia, but high rates persist overall. *Global Forest Watch*, 4.
- Wiedenhoeft, A. C., Simeone, J., Smith, A., Parker-Forney, M., Soares, R., and Fishman, A. (2019). Fraud and misrepresentation in retail forest products exceeds us forensic wood science capacity. *PloS one*, 14(7):e0219917.
- World Forest ID (2025). Unlocking transparency, exposing material risk: Key findings from timber market study. Insight Report, April 2025.
- Zhu, T. (2020). Analysis on the applicability of the random forest. In *Journal of Physics: Conference Series*, volume 1607, page 012123. IOP Publishing.