Junior Management Science 11(3) (2026) 561-597

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Toward Reliable Emission Reporting: A Comparative Study of GHGRP Self-Reports and Climate TRACE Satellite Data

Felix Maximilian Kania

Technische Universität München

Abstract

Accurate and transparent greenhouse gas (GHG) emissions data is crucial for effective climate mitigation, yet existing reporting systems remain inconsistent and difficult to verify. Carbon accounting has emerged to give structure and legitimacy to these measurement efforts by mandating rules for affected GHG emitters through programs such as the Greenhouse Gas Reporting Program (GHGRP) by the United States government. Historically, these programs have relied on self-reporting, significantly limiting the verifiability of corporate-reported data. In contrast, emerging non-profit organizations such as Climate TRACE (CT) estimate facility-level GHG emissions using satellite-based remote sensing. This study quantifies facility-level discrepancies between these datasets and identifies their key drivers.

To do so, I systematically matched and compared facilities from both datasets based on reported emissions. I identified key predictors of these discrepancies using machine learning and Bayesian inference. My findings reveal that facility-specific effects drive most observed disparities, while parent-company and geographic influences play a secondary role. Industry-wide effects contribute minimally, with reporting year and total emissions volume having no significant impact. These results suggest that discrepancies stem from isolated inaccuracies rather than systemic errors, underscoring the need for hybrid verification frameworks that integrate self-reported (bottom-up) inventories with independent satellite-based monitoring (top-down) to enhance emissions transparency and accountability.

Keywords: greenhouse gas emissions; carbon accounting; emissions monitoring; satellite remote sensing; emissions verification

1 Introduction

Despite international climate agreements, global GHG emissions continue to rise, reaching a record high of over 50 billion metric tons of carbon dioxide equivalent (CO2e) in 2023, intensifying the risk of exceeding the 1.5°C warming threshold set by the Paris Agreement (UNFCCC, 2015). This escalation accelerates climate disruptions, including extreme weather, biodiversity loss, and economic instability (European Commission, 2024; IEA, 2024; IPCC, 2023). Given the urgency of addressing climate change (Fawzy et al., 2020), effective climate action depends on GHG reductions and the accuracy and transparency of emissions data, as reliable measurements are essential for tracking progress, enforcing regulations, and designing effective mitigation strategies. The growing emphasis on measuring and reporting emissions has raised critical concerns about different reporting systems' accuracy, consistency, and comparability (Huertas et al., 2012). However, discrepancies in carbon accounting - whether from self-reported corporate disclosures, regulatory inventories, or international emissions databases - pose a significant challenge to climate governance (Schaltegger & Csutora, 2012). Without reliable reporting, policymakers cannot design effective mitigation strategies; companies may misrepresent their carbon footprints, and global climate agreements risk being undermined by flawed emissions accounting (Klaaßen & Stoll, 2021). Carbon accounting methods have been developed to measure and verify emissions to address these challenges. While self-reported emissions inventories have been the dominant approach, they are susceptible to methodological inconsistencies and underreporting (Stechemesser & Guenther, 2012). In contrast, emerging satellite-based emissions tracking technologies offer independent, high-resolution verification, potentially transforming emissions accountability and climate policy enforcement (Kiemle et al., 2017).

While previous studies have compared different carbon accounting methods (Kennelly et al., 2019) including analyses of major public emissions databases (Moran & Wood, 2014) and satellite-based estimation techniques (Goetz et al., 2009), few have systematically examined discrepancies between facility-level self-reported emissions and independent satellite-based estimates. Research on Climate TRACE (CT) has primarily focused on sectoral trends or large-scale emissions monitoring, leaving questions about its reliability at the facility level (Gurney et al., 2024). This gap is particularly significant given the growing role of independent verification

DOI: https://doi.org/10.5282/jums/v11i3pp561-597

in climate governance and emissions trading schemes.

This study explores the convergence of two opposing approaches in carbon accounting: traditional self-reported inventories, which rely on standardized methodologies and industry disclosures, and emerging satellite-based monitoring, which offers real-time, observational data. While regulatory programs like the U.S. Environmental Protection Agencies' (EPA) GHGRP (Environmental Protection Agency, 2024e) are based on legally mandated self-reporting, independent satellite-based systems such as CT (Climate TRACE, 2024b) provide an alternative, potentially more objective means of emissions verification (J. He & Huang, 2023). However, discrepancies between these two approaches raise questions about their accuracy, consistency, and potential biases in reporting methods. Understanding these discrepancies is essential for improving emissions transparency and strengthening hybrid verification systems that integrate both approaches.

The primary objective of my study is to investigate discrepancies between self-reported emissions from GHGRP and independent satellite-based estimates from CT. Specifically, I aim to identify the key predictors of these discrepancies at the facility level, assess the extent and direction of reporting biases, and evaluate the broader implications for emissions accounting and policy enforcement. By systematically comparing these two data sources, I contribute to the growing field of hybrid carbon accounting methods that integrate self-reported and observational approaches.

To achieve these objectives, my study follows a central research question:

> What are the key predictors of discrepancies in facility-level emissions data between the self-reported GHGRP and independent satellite-based Climate TRACE system?

To further refine this inquiry, my study examines the following sub-questions:

1. To what extent do facility characteristics, such as size, ownership, and reported emissions volume, influence discrepancies?
2. How do sector classification (subparts) and county-level effects contribute to discrepancies in emissions reporting?
3. How consistent are these findings across different analytical approaches, including variations in matching thresholds and statistical models?

By addressing these questions, I systematically evaluate the drivers of emissions discrepancies, offering a deeper understanding of the accuracy and limitations of self-reported and satellite-based emissions inventories.

With this, I contribute across three key domains. First, my study provides new empirical insights into facility-level variations in emissions reporting, contributing to the growing body of research on hybrid verification systems. By systematically identifying predictors of discrepancies, it advances the understanding of how different carbon accounting methodologies compare in practice. Second, by uncovering the specific conditions under which self-reported and satellite-based emissions data diverge, this research offers evidence-based recommendations for improving regulatory frameworks. The findings emphasize the importance of independent verification mechanisms in compliance programs such as GHGRP and inform policymakers on strategies to enhance emissions transparency and accuracy. Lastly, my study introduces a structured, replicable approach for matching and comparing self-reported emissions data with independent observational sources. By refining methods for facility-level emissions verification, the research informs future applications of satellite-based monitoring in carbon accounting, supporting the development of more reliable hybrid verification systems.

Through the integration of self-reported and satellite-based approaches, my study underscores the potential for hybrid verification systems to strengthen global emissions reporting, ultimately supporting more effective climate policy and accountability mechanisms.

2 Literature review

2.1 Carbon Accounting: An overview

Adopting the Kyoto Protocol in 1997 marked the formal beginning of internationally coordinated efforts to measure, report, and regulate GHG emissions, laying the groundwork for modern carbon accounting practices by introducing legally binding reduction commitments and market-based mechanisms (United Nations, 1998). Following the definition by Tang (2017), carbon accounting is a "system that uses accounting methods to record and analyze climate-change information as well as account for and report carbon-related assets, liabilities, expenses, and income for decision-making purposes" (p. 2). In recent years, more and more studies have been conducted in the field of carbon accounting because it is an essential tool in fighting climate change (Schaltegger & Csutora, 2012), which demands holistic approaches for its management and the mitigation of negative impacts (Schaltegger & Csutora, 2012). This trend is also reflected in the increasing legal codification of carbon accounting, as more laws and regulations have been introduced to standardize and enforce reporting practices (Chapple et al., 2013). Thus far, a handful of nations have led the efforts to reduce emissions, which have been particularly successful in Europe, which grew its GDP and simultaneously reduced emissions between 1995 and 2019 (Schaltegger & Csutora, 2012). Upon further inspection, however, many of these successes can be attributed to shifting emission-heavy production to other countries, particularly those in Asia (Schaltegger & Csutora, 2012), underscoring the need for internationally coordinated efforts in carbon accounting to eventually prevent climate change. Historically, carbon accounting has evolved in four distinct stages: initially considered a subset of environmental management accounting, it later emerged as a separate field with growing attention from policymakers and businesses. Over time, its scope expanded to include supply chain and product-level footprints, ultimately progressing toward broader climate change accounting incorporating adaptation strategies and indirect impacts (Csutora & Harangozo, 2017).

In its basic form, carbon accounting "provides a mechanism through which carbon emissions can be measured and quantified and helps in knowing the emissions status of companies and making the necessary strategic decisions to achieve mitigation" (Hazaea, et al., 2023, p. 1). Thus, it is a decision-making tool at various institutional levels, including governments, corporations, and regulatory bodies. It enables informed policy development, business strategies, and sustainability initiatives (Schaltegger & Csutora, 2012). This is achieved through two basic approaches: documenting emissions and guiding mitigation efforts, underscoring its importance in measuring emissions and shaping proactive environmental strategies (Schaltegger & Csutora, 2012). Furthermore, carbon accounting is essential for integrating emissions data into policy frameworks and corporate decision-making processes. Establishing a quantitative basis for emissions allows organizations to assess their environmental impact and set reduction targets, offering transparency and accountability in sustainability reporting (Schaltegger & Csutora, 2012). Carbon accounting can be an emancipatory tool beyond its technical role in emissions tracking, challenging conventional accounting paradigms and integrating broader social and environmental concerns into corporate and policy decision-making (Ascui, 2014). The practice is central to climate governance and corporate responsibility by linking emissions data with regulatory compliance, financial performance, and strategic planning. Its ability to inform decision-making ensures that mitigation efforts are data-driven, allowing stakeholders to develop more effective policies and sustainable business models.

Carbon accounting follows the framework established by the GHG Protocol, the most widely used international standard (World Resources Institute, 2024). Emissions are categorized into three scopes: Scope 1 includes direct emissions from owned or controlled sources, such as company vehicles and boilers; Scope 2 covers indirect emissions from purchased electricity and other energy sources; and Scope 3 accounts for all other indirect emissions, including those from supply chains, transportation, and product usage. While Scopes 1 and 2 are relatively straightforward to measure, Scope 3 presents a more significant challenge due to its complexity and the need to include emissions along the entire value chain (Csutora & Harangozo, 2017). Beyond this, carbon accounting can be categorized into two significant methodologies: production-based accounting (PB), which attributes emissions to the geographic location of production, and consumption-based accounting (CB), which allocates emissions based on the end-user of goods and services (Marlowe & Clarke, 2022). The choice between PB and CB methodologies significantly impacts reported emissions, as studies have shown that CB often results in higher emission values due to the inclusion of embedded emissions in imported goods and services, which are otherwise excluded in PB calculations (Marlowe & Clarke, 2022). While traditional carbon accounting relies heavily on field sampling and land cover classifications, direct remote sensing methods offer a more spatially consistent way to monitor carbon stocks (Goetz et al., 2009). Recently, hybrid systems combining top-down (atmospheric measurements) and bottom-up (inventory reporting) emission measurements have been shown to significantly reduce the previously observed discrepancies between purely top-down and purely bottom-up (such as GHGRP) reporting systems (Chan et al., 2024).

In recent years, the transition from voluntary to mandatory carbon reporting has gained momentum, driven by the introduction of international reporting standards such as IFRS S1 and S2, which require companies to integrate carbon disclosures alongside traditional financial statements (Tang, 2024). While this shift enhances transparency and accountability, it presents significant challenges, including aligning carbon reporting boundaries with financial reporting structures and using standardized methodologies to ensure comparability across jurisdictions (Tang, 2024). While mandatory reporting aims to enhance transparency, companies may still have incentives to exaggerate carbon reduction claims, making regulatory oversight essential to mitigate greenwashing risks(Tang, 2024). Regulatory enforcement enhances carbon management effectiveness, which may, in turn, improve investor confidence and long-term financial stability (Sitompul et al., 2023). Several international initiatives, including the International Sustainability Standards Board's IFRS S1 and S2 standards and the EU's Corporate Sustainability Reporting Directive (CSRD), are refining disclosure requirements to enhance the quality, comparability, and consistency of reported emissions data. These regulations aim to bridge the gap between voluntary and mandatory disclosures by ensuring companies provide investors and regulators with standardized, audited, and comparable carbon information (Alfieri et al., 2024). The effectiveness of carbon reduction efforts varies by industry and regulatory environment. Research suggests that organizational capital plays a more decisive role in reducing emissions in carbon-intensive sectors. In jurisdictions with an Emissions Trading System (ETS), companies face more substantial financial and regulatory incentives to cut emissions (Papadopoulos, 2022).

Implementing carbon accounting extends beyond helping to reduce global carbon emissions; it can also enhance firms' financial performance by enhancing competition (Hazaea et al., 2023). A systematic review of more than 200 articles found that 59% of studies confirm a positive relationship between carbon management strategies and corporate financial performance, with 50% demonstrating a significant impact, reinforcing the financial incentives for sustainability efforts (Sitompul et al., 2023). However, some studies indicate that while carbon management enhances short-term financial performance through efficiency gains, long-term financial benefits depend on a firm's ability to offset initial investment costs and navigate regulatory complexities. The financial benefits of carbon management primarily arise from cost efficiency, enhanced productivity, product differentiation, and improved market positioning, highlighting the strategic advantages of emissions reductions beyond compliance (Sitompul et al., 2023). Nevertheless, the financial impact of carbon performance is not always linear. Evidence suggests a U-shaped relationship, where firms with superior or poor carbon performance benefit financially, while those in the middle face competitive disadvantages (Lewandowski, 2017). Empirical evidence from the EU-ETS suggests that carbon management does not continuously improve financial performance. Companies with environmental certifications face profitability declines when emissions rise, likely due to increased compliance costs or external scrutiny (Tuesta et al., 2020). A recent analysis of over 3800 firm-year observations of US publicly listed companies showed that businesses with higher organizational capital (specialized knowledge, skills, innovative technologies, and efficient logistics) tend to have lower GHG emissions (Provaty et al., 2024).

Additionally, while carbon emission mitigation can positively correlate with profitability, it can simultaneously lead to adverse investor reactions, as reflected in lower stock market valuations, potentially discouraging companies from ambitious emission reduction efforts (Lewandowski, 2017). Another study found that good carbon performance can reduce a firm's profitability and stock market valuation. Specifically, a 1% increase in carbon performance is associated with an average 0.055% decrease in return on assets (ROA), indicating that firms face financial trade-offs when improving their emissions profile (Busch et al., 2022).

Despite its growing importance, carbon accounting faces several limitations that impact its effectiveness and credibility. One key challenge is the lack of standardized methodologies, which creates inconsistencies in reporting and interpreting emissions data across different organizations and regulatory frameworks (Rogelj et al., 2019). One of the persistent challenges in carbon accounting is the significant measurement uncertainty associated with emissions estimates, which arises due to variations in data quality, reporting methodologies, and assumptions underlying emission factors (Marlowe & Clarke, 2022). This variation complicates efforts to compare emissions performance and undermines the reliability of carbon accounting as a decision-making tool (Hazaea et al., 2023). Additionally, the quality of carbon data remains a significant concern. An analysis of corporate disclosures revealed that emission data is often derived from a mix of reported figures and estimates. Only about 15% of data stems directly from company-reported sources, while the rest is derived from indirect indicators such as energy use and sustainability reports (Busch et al., 2022). A further significant limitation is the lack of comparability between reported emissions data across different firms and jurisdictions, making it difficult to assess performance, track progress, or enforce accountability (Marlowe & Clarke, 2022). The lack of harmonized global statistics on carbon emissions leads to data gaps and inconsistencies, as companies often rely on rough estimates. This affects firm-level and macro-level emissions reporting, complicating comparability and standardization across industries and jurisdictions (Alfieri et al., 2024).

Next, variations in how carbon performance is defined contribute to inconsistencies in financial impact assessments. Lewandowski (2017) highlights that some firms report carbon emissions primarily to establish a baseline comparison with industry peers, while others emphasize improvements in carbon performance over time to demonstrate progress. This discrepancy in disclosure strategies complicates the interpretation of carbon data, making it difficult to assess the impact of emission reductions on financial performance.

Determining reporting boundaries for carbon accounting presents another significant challenge, especially compared to traditional financial reporting frameworks (Rogelj et al., 2019). Control methods prescribed by accounting standards generally govern financial statements. In contrast, carbon accounting often allows for alternative boundary definitions, such as the equity method, which is not widely recognized in financial reporting (Tang, 2024). As carbon accounting becomes increasingly integrated into financial disclosures, aligning its reporting boundaries with financial statements is crucial to prevent inconsistencies and ensure clarity for investors and stakeholders (Tang, 2024). However, empirical evidence suggests that participation in emissions trading schemes like the EU-ETS does not necessarily translate into improved carbon performance. Firms within the EU-ETS do not exhibit better emissions management than non-participating firms, indicating that compliance alone does not drive proactive carbon reduction efforts. While organizational capital enhances a firm's ability to implement low-carbon strategies, financial constraints can undermine these efforts. High R&D and technology transition costs can prevent highly organized firms from fully realizing their emissions reduction potential (Provaty et al., 2024). As Schaltegger and Csutora (2012) emphasize, carbon "accounting must support managers to take the right decisions" (p. 1). However, if methodologies differ significantly, decision-makers may struggle to derive meaningful insights from data, potentially weakening emissions reduction efforts. Another concern is that carbon disclosure is often motivated by a desire to enhance corporate legitimacy rather than a genuine commitment to sustainability. Some companies use carbon accounting primarily as a public relations tool to maintain their image rather than to drive meaningful emissions reductions (Schaltegger & Csutora, 2012). This suggests that carbon reporting may sometimes be used defensively to manage external scrutiny rather than as a proactive emission reduction mechanism. Moreover, Scope 3 emissions measurement remains a particularly contested area, even in jurisdictions with mandatory carbon reporting (Tang, 2024). The complexity of supply chain emissions and data availability challenges create further barriers to reliable carbon accounting (Tang, 2024).

Furthermore, investor skepticism about the quality and reliability of carbon disclosures persists due to limited verification mechanisms. While investors increasingly push for transparency in emissions reporting, there is insufficient emphasis on validating the accuracy of reported data. This raises concerns about whether carbon accounting influences corporate decision-making and sustainability strategies or merely functions as a compliance exercise (Hazaea et al., 2023). Companies may selectively disclose favorable carbon data without stringent verification mechanisms while obscuring negative performance, underscoring the growing need for independent carbon assurance frameworks like ISAE 3410 (Tang, 2024). A growing concern in the field is the lack of expertise in carbon accounting, which has resulted in inconsistencies and misinterpretations of data. Many organizations lack trained professionals to assess and manage climate-related risks, leading to emissions calculations and reporting discrepancies (Tang, 2024). To address these issues, researchers emphasize the urgent need for robust policy frameworks and integrated accounting procedures that can enhance data reliability and improve the utility of carbon accounting for decision-making (Marlowe & Clarke, 2022). So far, standardization efforts have proven challenging, as carbon accounting inherently serves multiple, often conflicting purposes, including scientific measurement, political negotiation, financial reporting, and corporate disclosure, resulting in competing methodologies and frameworks (Ascui & Lovell, 2011).

In addition, the treatment of carbon credits and GHG offsetting remains ambiguous, further complicating carbon accounting. While carbon credits are expected to be disclosed in emissions reports, there is no consensus on whether they should be classified as assets, expenses, or financial instruments. This lack of regulation creates challenges in ensuring consistent and transparent reporting practices (Tang, 2024). These limitations underscore the need for stronger regulatory frameworks, standardized reporting methodologies, and independent verification mechanisms to enhance the credibility and impact of carbon accounting.

As carbon accounting continues to evolve, hybrid approaches integrating physical and monetary data are expected to play an increasingly important role. With the expansion of carbon accounting beyond direct emissions to include supply chains and product life cycles, researchers are developing new methodologies to improve accuracy and comparability (Schaltegger & Csutora, 2012). Multi-sensor approaches integrating satellite imaging, LIDAR, and radar measurements are expected to play a more significant role in verifying emissions data. Unlike traditional carbon accounting, which relies on financial reporting structures, hybrid verification systems combining physical and digital carbon tracking could enhance compliance verification, emissions trading, and corporate sustainability assessments (Goetz et al., 2009). Moreover, there is increasing pressure for international standardization in carbon accounting frameworks. Currently, emissions reporting varies significantly across jurisdictions, making it challenging to compare emissions between companies, sectors, and countries (Schaltegger & Csutora, 2012). Technological advancements, regulatory requirements, and growing investor demand for transparency will likely improve future carbon accounting.

Empirical evidence suggests that regulatory frameworks can mitigate the financial penalties associated with carbon reductions, reinforcing the need for stronger carbon pricing mechanisms and mandatory reporting requirements to align emissions reductions with corporate financial incentives (Busch et al., 2022). Innovations such as real-time emissions tracking, AI-driven data validation, and satellite-based verification systems are expected to enhance the accuracy and credibility of reported emissions. Blockchain technology can radically transform emissions data's transparency, reliability, and traceability in supply chains (Kaur et al., 2023). Emerging frameworks, such as the Dynamic Integrated Carbon Accounting (DICA) model, aim to address the complexity of modern carbon accounting by integrating evolving environmental, financial, and regulatory factors (Tang, 2024). Unlike traditional approaches, DICA incorporates forward-looking elements, financial carbon accounting, and management tools to enhance decision-making and climate risk assessment.

Finally, independent verification mechanisms, such as ISAE 3410, are expected to improve data credibility (Tang, 2024). To improve emissions tracking and minimize errors from multiple-counting, the E-liability approach has been proposed as a novel methodology that applies financial accounting principles, such as double-entry bookkeeping, to carbon tracking (Alfieri et al., 2024). This method allows firms to record direct emissions, account for emissions transferred from suppliers, and allocate emissions to products more accurately, reducing reliance on rough estimates and assumptions. By integrating real-time, auditable carbon data at each value chain stage, E-liability offers a potential solution for enhancing carbon reporting transparency and ensuring a more precise measurement of cumulative emissions (Alfieri et al., 2024). As carbon accounting moves beyond compliance, integrated and adaptive frameworks will support businesses' transition to net-zero operations.

2.2 GHGRP: The EPA's Greenhouse Gas Reporting Program

The US GHGRP originated from growing concerns over the effects of GHG emissions on climate and the environment within various US agencies in the 1990s, including the US Environmental Protection Agency (EPA). In response, the United States government set goals to limit their emission of greenhouse gases, thus establishing the need for accurate and transparent GHG accounting (Schrock et al., 1994). Initial voluntary reporting frameworks provided the basis for establishing the mandatory GHGRP in October 2009 (codified at 40 CFR Part 98 (Environmental Protection Agency, 2009)). The program was further influenced by previous successful cap-and-trade programs in the US, like the SO2 trading program (Kruger, 2005).

GHGRP aims to enhance transparency and facilitate data-driven US climate policies through its detailed GHG emissions accounting (Miller et al., 2024). Additionally, its "data can be used by businesses and others to track and compare facilities' greenhouse gas emissions, identify opportunities to cut pollution, minimize wasted energy, and save money" (Environmental Protection Agency, 2024, p. 1).

To do so, GHGRP "requires reporting of greenhouse gas (GHG) data and other relevant information from large GHG emission sources, fuel and industrial gas suppliers, and CO2 injection sites in the United States" (Environmental Protection Agency, 2024, p. 1). It includes CO2 and non-CO2 gases, mainly methane, nitrous oxide, and fluorinated gases, with tailored methodologies to capture sector-specific emissions (Hanle et al., 2012). "Direct-emitting facilities report emissions from each source category (rule subpart) included in the GHGRP, and these emissions can generally be categorized as either combustion or process emissions. Combustion emissions comprise CO2, CH4, and N2O emitted from fossil fuel combustion or biomass feedstocks. Process emissions generally include emissions from chemical transformations of raw materials and fugitive emissions" (Environmental Protection Agency, 2017, p. 1). Significant sources are defined as emitting at least 25,000 tons of CO2e annually, while smaller sources, sinks of greenhouse gases, and specific sectors like agriculture are excluded (Environmental Protection Agency, 2024b). It encompasses 46 industrial categories and recently expanded to include Direct Air Capture facilities, thus allowing for comprehensive tracking of CO2 flows from capture to final sequestration (Miller et al., 2024). Both direct GHG emitters and upstream suppliers are required to report. Direct emissions alone account for around half of US GHG emissions and fall under Scope 1. Here, reporting is done on the facility level, with parent company information also being collected (Environmental Protection Agency, 2024b). "Suppliers report the amount of CO2e that would be released if the products they produce, import, or export (e.g., fossil fuels) were released, combusted, or oxidized. These are reported at the corporate level and fall under Scope 3 emissions" (Environmental Protection Agency, 2024, p. 1). Together, these cover 85-90% of US GHG emissions (Environmental Protection Agency, 2024b). As the program is federal, the location within the United States does not directly affect the reporting rules (U.S. National Archives and Records Administration, 2025). Emissions from fuel combustions, e.g., burning biomass for energy or heat production, are measured through a continuous emission monitoring system (CEMS), which is considered the most accurate approach, or by applying default emissions factors. An emissions factor is an average value that estimates pollutant emissions per unit of activity, typically expressed as the weight of emissions per unit of fuel burned or industrial output (Environmental Protection Agency, 2024a). On the other hand, process emissions (emissions released from chemical reactions or transformations of raw materials) can be accounted for by either CEMS, a mass balance approach, or site-specific or default emission factors (Environmental Protection Agency, 2017).

Firms are required to submit annual reports electronically by 31 March of the following year. Reporters are provided with a series of online training sessions to provide them with the necessary information to comply with the program's reporting requirements, depending on the industry sector of their facilities. The program's online reporting tool employs built-in data validation checks to ensure accuracy (Environmental Protection Agency, 2024e). GHGRP improves upon earlier programs, such as the Inventory of U.S. Greenhouse Gas Emissions and Sinks (Environmental Protection Agency, 2025a), by requiring direct measurements for specific sectors like methane recovery, reducing uncertainties and enhancing robustness in landfill emissions tracking (Bronstein et al., 2012). Generally, the program follows a flexible approach, balancing accurate reporting on the one hand with the need for cost-effectiveness and not overburdening firms with unnecessarily complex bureaucracy. For example, facilities can choose between direct measurement methods or applying EPA-approved emission factors to site-specific activity data in most industries (Hanle et al., 2012).

A previous study on the program showed that electric power plants affected by its reporting regulations consequently reduced CO2e emissions by 7% compared to non-reporting facilities (Yang et al., 2024). This reduction effect is even more prominent for companies with prior experience in pollution prevention, demonstrating how GHGRP builds upon pre-existing environmental knowledge systems (Lee & Bi, 2020). Furthermore, it leads to firms significantly improving their emission intensity, defined as Scope 1 specific emissions divided by total assets, even without substantially reducing absolute emissions. Through its mandatory and standardized reporting, the program also reduces impression management opportunities, meaning the ability of organizations to influence how they are perceived by others, thus increasing the credibility of reported data (Bauckloh et al., 2022). Recently, the program's introduction has also been linked to companies' significant reduction in greenwashing behavior by increasing their motivation to improve ESG performance and enhancing transparency and accountability, especially for large firms worried about public perception (Nguyen et al., 2025). GHGRP has further been shown to shape corporate behavior as firms mandated to report under its rules are 68% less likely to face environmental lawsuits, driven by increased community stakeholder pressure from enhanced reporting transparency (Huang et al., 2023). In general, the standardized and easily accessible data provided by the EPA allows for large-scale analyses of greenhouse gases, which offer significant opportunities for diverse scientific inquiries (Jain et al., 2021).

Several limitations of GHGRP have been evaluated so far. As mentioned, the program might not necessarily reduce absolute CO2e emissions (Bauckloh et al., 2022), as it is not directly linked to reduction mechanisms such as carbon pricing. Therefore, the program may incentivize operational efficiency without necessarily aligning with the broader climate goal of reducing total emissions. GHGRP also demonstrates methodological limitations by relying on outdated emission factors and thresholds, excluding significant emission sources (Subramanian et al., 2015). Here, the authors stress the need for direct measurements of gases to guarantee accurate reporting, challenging indirect measurement systems like those satellite-based systems introduced in the following chapter. The program's cost-effectiveness has also been under scrutiny, arguing that its contribution to emission reductions is negligible for its high implementation costs (Sanchez et al., 2012). Looking at the facilities required to report, the associated incremental costs are as high as $1.42 per metric ton of CO2e, not considering potential tax credits that might exceed this amount (Godec et al., 2022). However, it should be noted that this study was released before the data used in this thesis started. Since then, the program has provided open access to its data, addressing one of the authors' main concerns. Due to data confidentiality concerns, facility-level data in specific sectors, such as suppliers of carbon dioxide, is withheld, limiting data consistency (Miller et al., 2024). Further criticism has focused on certain methodological flexibilities, which facilities are offered to compensate for their diverse operational and physical characteristics, which might lead to the systematic underestimation of the US methane inventory (Stark et al., 2024). More reporting inaccuracies have been attributed to GHGRP's often static and overly simplistic assumptions, as demonstrated through the discovery of the significant effect of rainfall on methane emissions from landfills, which is not considered by the program's emission factors (Jain et al., 2021). It should be noted that EPA continues to update its program, regularly adding new features to its tools and expanding its datasets (Environmental Protection Agency, 2024c).

2.3 Climate TRACE: Satellite-based emissions monitoring

Climate TRACE is a non-profit coalition of various organizations with the stated goal of making "meaningful climate action faster and easier by harnessing technology to track greenhouse gas (GHG) emissions with unprecedented detail and speed, delivering information that is relevant to all parties working to achieve net-zero global emissions" (Climate TRACE, 2025, p. 1). As of the writing of this paper, over 100 organizations, including Google.org, NASA, the US EPA, the European Space Agency, and various universities like Carnegie Mellon, are collaborating on the project. It was first started in May 2019 with funding from Google.org and the vision to use artificial intelligence to monitor power plant emissions from space (Climate TRACE, 2025a). Since then, the coalition has expanded to cover all GHG emissions and improve its technology, releasing its first emissions inventory in September 2021. In their second facility-level inventory, released in December 2023, CT covered over 352 million emissions sources worldwide (Climate TRACE, 2025a). This is also the inventory I used in the analysis of this paper. In November 2024, their third inventory was released, covering more than 660 million individual sources of GHG emissions (Climate TRACE, 2025a). The collected data is publicly available through their interactive website, API, and download packages.

Contrary to most other GHG accounting projects, such as the GHGRP discussed before, the non-profit started providing emissions inventories that were not based on self-reported data. Its primary purpose is to enable and facilitate "the creation of policies, programs, and campaigns aimed at limiting global temperature rise to 1.5°C as agreed to under the Paris Climate Agreement" (Climate TRACE, 2025, p. 1).

CT's data collection is based on over 300 satellites orbiting the Earth. The first step, thus, is identifying the locations of significant sources of emissions around the globe. These include power plants, factories, mines, airports, etc. Next, ground truth - direct on-site measurement - samples are collected for some of these, meaning precise emission measurements. These can then be used to train machine-learning algorithms together with satellite imagery. The resulting models can then be extrapolated to all sources of emissions that do not have ground truth samples (Climate TRACE, 2025a). This approach is not unique, with various other projects like Landsat 8 by NASA (NASA, 2025a) or PlanetScope by ESA (The European Space Agency, 2025) providing the satellite imagery necessary to estimate carbon emissions remotely (Couture et al., 2024).

The resulting facility-level data is categorized into 10 sectors (Figure 1), in decreasing order of total emissions: Power (the leading source with around 25% of global emissions), Manufacturing, Fossil Fuel Operations, Transportation, Agriculture, Buildings, Waste Fluorated Gases, Mineral Extraction, and Forestry And Land Use, as well as 67 sub-sectors. The project's database contains information on over 500 billion tons of CO2e between January 2015 and December 2024 (Climate TRACE, 2025b).

Source: (Climate TRACE, 2025a)

For power plants specifically, CT employs a dual-method approach to estimate facility-level emissions. The first method, applied to 12.5% of global power plants, utilizes machine learning (ML) and remotely sensed water vapor plumes to derive time-dependent capacity factors (CFs) at the facility level. The second method, which accounts for 87.5% of facilities, relies on national-average power plant properties to estimate emissions (Gurney et al., 2024). This distinction is crucial, as the ML-based approach offers higher precision by incorporating real-time satellite observations. In contrast, the national average method introduces more significant uncertainty due to its reliance on broad assumptions (Gurney et al., 2024).

Another advantage of satellite-based sensing is its potential to identify and locate previously unknown or underestimated emissions hotspots, such as methane-emitting solid waste landfills (Dogniaux et al., 2024). Generally, hybrid approaches combining activity-based and atmospheric measurements can "produce more complete and accurate estimates of GHG emissions and sinks" (National Academies of

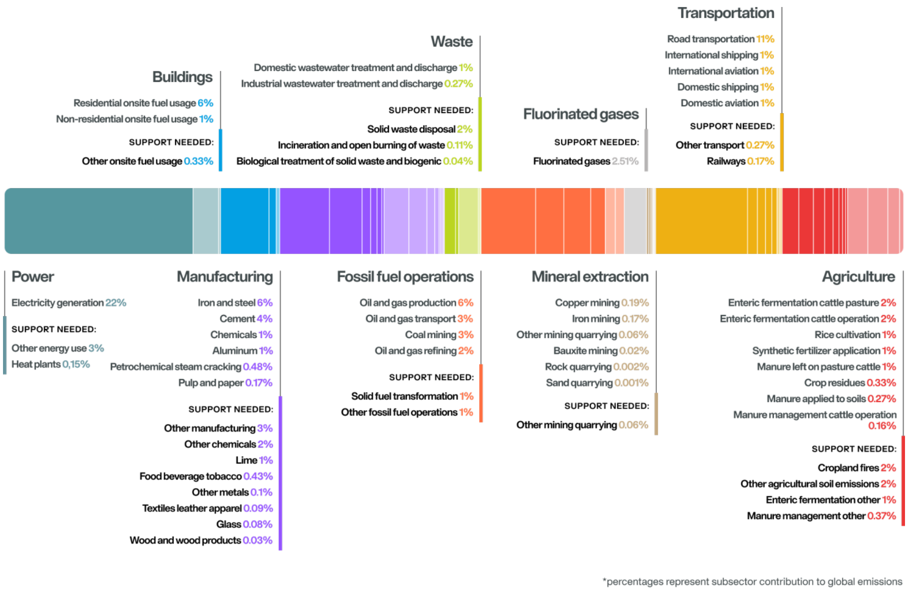

Figure 1: Climate Trace global emissions per category 2024

Sciences, Engineering, and Medicine, 2022) than individual approaches.

A 2024 study compared Climate TRACE's power plant emissions estimates with the Vulcan-power v1.1 dataset, revealing notable discrepancies. While CT's AI-based satellite detection produced relatively accurate estimates (mean relative difference = -1.1%, SD: 46.4%), its national-average capacity factor (CF) - the typical ratio of actual energy output to maximum possible output for a given energy source across a country, averaged over time - approach led to large underestimations (MRD = -50.0%, SD: 117.7%) (Gurney et al., 2024). The study highlights that CT's use of national-mean CFs does not account for plant-level variations, leading to significant biases. In contrast, emissions estimates derived directly from AI-based satellite observations showed far better agreement with Vulcan-power v1.1, suggesting that expanding direct satellite measurements rather than relying on generalized assumptions could improve CT's accuracy (Gurney et al., 2024). A – yet-to-be-peer-reviewed study - conducting a direct comparison between GHGSat-detected emissions (HGSat Inc., 2025) - another satellite-based GHG emissions monitor - and CT estimates for landfills found no correlation between the two datasets, raising concerns about the reliability of modeled emissions estimates (Dogniaux et al., 2024). This suggests that CT's landfill emissions estimates may diverge significantly from direct satellite observations, highlighting potential methodological limitations. While CT's approach integrates machine learning and global facility data, its reliance on modeled emissions factors rather than direct satellite plume detection could explain the observed discrepancies.

Furthermore, the study found substantial differences in magnitude between CT and GHGSat estimates, with CT overestimating or underestimating emissions (Dogniaux et al., 2024). Another point of criticism centers around the organization's transparency regarding its data and source code, which are mostly not publicly available. The National Academy of Sciences has categorized the website as Low on Usability and timeliness because there is little evidence that it is used by relevant decision-makers, Low on Information Transparency because of the mentioned limited access to data and source code, and Low on Evaluation and transparency resulting from few peer-reviewed evaluations of its data (National Academies of Sciences, Engineering, and Medicine, 2022).

These findings reinforce a key limitation of Climate TRACE: while its dataset provides unprecedented global coverage, discrepancies arise when compared to direct remote sensing measurements.

Looking ahead, the launch of ever more numerous and sophisticated satellites like the Copernicus CO2 Monitoring mission (Deutsches Zentrum für Luft- und Raumfahrt e.V., 2024) scheduled for 2026 is expected to further improve remote emissions monitoring (Couture et al., 2024), directly improving CT's data coverage and quality.

2.4 Existing research on emissions discrepancies

There have already been several other studies focused on understanding emissions discrepancies between different data sources. Recent - not yet peer-reviewed - research has shown that discrepancies in landfill methane emissions are particularly pronounced, with satellite observations revealing emissions up to six times higher than reported to the GHGRP (Balasus et al., 2025). This divergence is primarily driven by the systematic selection of a reporting methodology that yields lower emissions estimates, as landfills increasingly opt for the recovery-first model, which estimates emissions from recovered methane rather than total generation, often overestimating collection efficiency and underreporting actual emissions trends (Balasus et al., 2025). Discrepancies have also been observed at national inventory levels, as a recent study comparing CT and Brazil's SEEG inventory found emissions estimates differing by up to 1 gigaton of CO2e in some years, primarily due to methodological differences in the forestry and land-use sectors (da Costa et al., 2025). Additionally, CT appears to better capture seasonal variability in emissions and removals, whereas traditional inventory methods rely on static emissions factors, potentially leading to underestimations of dynamic carbon sinks (da Costa et al., 2025). A further high-resolution satellite-based analysis confirms systematic underreporting of landfill methane emissions to GHGRP, with TROPOMI-based estimates showing landfill emissions are 77% larger than reported. Independent aircraft-based validation supports these findings, further reinforcing self-reported landfill methane data reliability concerns (Nesser et al., 2024). Another study compared CO2 emissions estimates from the ODIAC satellite-based dataset with city-reported inventories across 14 global cities, revealing significant differences. In some cities, such as Cape Town (+148%) and São Paulo (+43%), satellite-based estimates were much higher than reported data, likely due to industrial emissions being misallocated in developing countries (Chen et al., 2020). Further research on U.S. city-level emissions has found that self-reported inventories underestimate emissions by an average of 18.3%, with some cities misreporting emissions by up to 145.5%. If applied to all U.S. cities, this underreporting would exceed California's total annual emissions by 23.5%, highlighting the scale of uncertainty in urban carbon accounting (Gurney et al., 2021).

2.5 Comparing GHGRP and Climate TRACE

Accurate reporting of greenhouse gas emissions is critical for climate policy and corporate accountability. However, as discussed, discrepancies persist between self-reported emissions data, such as the U.S. Environmental Protection Agency's (EPA) Greenhouse Gas Reporting Program (GHGRP), and independent satellite-based estimates from sources such as Climate TRACE (CT). These discrepancies stem from methodological differences, operational limitations, and data verification challenges (Balasus et al., 2025; Chen et al., 2020; da Costa et al., 2025; Gurney et al., 2021; Nesser et al., 2024).

GHGRP relies on a bottom-up inventory-based approach in which facilities self-report emissions using standardized estimation techniques, such as fuel-based emissions factors, mass balance methods, or direct measurement via Continuous Emissions Monitoring Systems (CEMS) (Environmental Protection Agency, 2024b). While this framework allows for industry-specific reporting flexibility, it inherently depends on the accuracy of facility-reported data and the assumptions underlying emissions calculations (Hanle et al., 2012).

CT employs a "top-down" approach integrating satellite observations, remote sensing techniques, and artificial intelligence (AI) models to estimate facility-level emissions (Climate TRACE, 2025a). This methodology can independently verify carbon stock estimates and enhance emissions monitoring accuracy, potentially identifying discrepancies in self-reported inventories (Goetz et al., 2009). However, CT's reliance on model extrapolation for facilities without direct observations introduces uncertainty, particularly in cases where ground-truth verification is limited (Gurney et al., 2024).

A key distinction between the two methodologies is their ability to capture transient emissions events, meaning short-term, high-intensity releases of pollutants or greenhouse gases that occur intermittently rather than continuously. GHGRP, which requires facilities to submit annual emissions reports, provides estimates based on standardized methodologies rather than real-time measurement (Environmental Protection Agency, 2024b). As a result, short-term fluctuations due to operational changes, infrastructure failures, or meteorological conditions may not be fully reflected (Subramanian et al., 2015). In contrast, satellite-based monitoring can detect short-lived but significant emissions surges, such as methane "burps" from landfills or unintentional fugitive emissions from oil and gas facilities (Cusworth et al., 2021). This distinction is particularly relevant for methane accounting, as episodic emissions events can represent a significant portion of total emissions but remain undetected in self-reported inventories (Duren et al., 2019).

Beyond methodological differences, facility-specific factors are crucial in reporting discrepancies in emissions (Cusworth et al., 2021; Gurney et al., 2024; Nesser et al., 2024). GHGRP provides multiple methodologies for estimating emissions, including direct measurement, mass balance calculations (which track inputs and outputs to infer emissions), and default emissions factors, depending on the sector (Environmental Protection Agency, 2017). However, sectors relying on modeled emissions factors, such as landfills and oil and gas facilities, are prone to underreporting, as assumptions about operational efficiency and emissions capture rates do not always align with real-world conditions. This aligns with broader criticism that self-reported emissions inventories tend to underestimate non-combustion sources, particularly fugitive emissions (Subramanian et al., 2015).

In contrast, like other airborne spectrometer-based methods, CT's remote sensing approach can effectively detect fugitive methane emissions by directly observing plumes. While studies have primarily focused on oil and gas infrastructure, similar methods could be applied to landfills to improve emissions detection and characterization (Cusworth et al., 2021). This is supported by a growing body of research advocating for hybrid verification approaches that combine self-reported data with independent observational methods to improve emissions accuracy (Chan et al., 2024).

GHGRP and CT offer distinct advantages and limitations that shape their effectiveness for monitoring facility emissions. GHGRP provides legally mandated, structured reporting that aligns with federal regulations and compliance mechanisms. This ensures that all significant GHG emitters in the U.S. must adhere to a standardized methodology, improving industry comparability (Environmental Protection Agency, 2024b). The program also tailors emissions calculation methods to specific industries, allowing for detailed sectoral tracking (Hanle et al., 2012). Another strength is its continuous dataset dating back to 2010, enabling longitudinal analysis of emissions trends at the facility level (Environmental Protection Agency, 2024b). On the other hand, emissions estimates are only as accurate as the reporting facility's methodology and compliance efforts, raising concerns about potential underreporting or misclassification (Subramanian et al., 2015). Furthermore, GHGRP excludes certain smaller facilities and unregulated emissions sources, potentially omitting significant emission contributions (Environmental Protection Agency, 2017).

CT offers an external validation mechanism for facility-reported emissions, reducing reliance on self-reported estimates (Climate TRACE, 2025b). Its satellite-based monitoring enables continuous emissions tracking, potentially detecting previously unknown sources and emissions spikes (Goetz et al., 2009). Finally, unlike GHGRP, which is limited to the U.S., CT provides global emissions data across multiple sectors, facilitating international climate monitoring efforts (Climate TRACE, 2025b; Environmental Protection Agency, 2017). While some facilities benefit from direct satellite observations, many CT estimates rely on machine-learning extrapolations, which introduce uncertainty in emission calculations (Gurney et al., 2024). CT's facility-level methane emission estimates show significant discrepancies from national reporting programs and independent satellite observations, with no strong correlation found. This suggests potential accuracy limitations, particularly in industrial sectors where site-level validation is critical (Dogniaux et al., 2024).

Given the strengths and limitations of both systems, a hybrid verification approach combining self-reported data (GHGRP) with independent observational data (Climate TRACE and other remote sensing systems) might offer the best path forward for improving emissions accuracy.

Despite their differences, GHGRP and CT provide facility-level emissions data that can be directly compared to enhance emissions verification efforts. GHGRP's structured reporting framework allows consistent facility-level tracking, while CT offers an independent validation mechanism using satellite-based estimates (Climate TRACE, 2024b; Environmental Protection Agency, 2024d). The availability of comparable facility-level emissions data from both sources enables cross-validation, offering a promising pathway for improving emissions monitoring and policy enforcement. The availability of comparable facility-level emissions data from both sources enables the comparative analysis conducted in this paper, providing a foundation for assessing discrepancies and improving emissions monitoring approaches.

2.6 Identified gaps in the literature

Despite growing efforts to improve emissions accounting, significant gaps remain in reconciling self-reported and independent emissions estimates (Dogniaux et al., 2024; Subramanian et al., 2015). Studies have highlighted systematic underreporting of methane and CO2 emissions across different sectors and geographic scales, yet the drivers of these discrepancies remain poorly understood (Balasus et al., 2025; Nesser et al., 2024; Subramanian et al., 2015). While previous research has compared self-reported inventories and independent estimates, these analyses often lack facility-level resolution, making it challenging to pinpoint systematic biases in emissions reporting (Dogniaux et al., 2024; Gurney et al., 2021). Furthermore, the effectiveness of hybrid verification approaches, which integrate bottom-up self-reported data with top-down satellite-based estimates, remains underexplored at the facility level (Chan et al., 2024; Goetz et al., 2009).

To address these gaps, I conducted a facility-level comparative analysis of this study's GHGRP self-reported dataset and CT satellite-based estimates. By systematically matching individual facilities across the two datasets and applying machine learning and Bayesian statistical methods, this research identifies key predictors of emissions discrepancies and assesses potential biases in facility-level reporting.

3 Methodology

3.1 Research design

In this study, I performed a comparative analysis to evaluate discrepancies between self-reported emissions data from the GHGRP and independent satellite-based estimates from CT. My primary objective was to identify key predictors of emissions discrepancies and assess systematic biases that may exist in facility-level GHG reporting. I systematically matched individual facilities across the two datasets to do so, directly comparing facility levels. With this approach, I provide new insights into patterns of over- or underestimation and the factors driving these discrepancies. For the complete analysis, see Appendix A.

To ensure a robust and reproducible analysis, I followed a structured four-step workflow:

1. Since GHGRP and CT do not share a common identifier, I developed a multi-step facility-matching algorithm based on geographical proximity, name similarity, and industry classification.

2. I measured emissions differences using absolute discrepancies in tons of CO2e, allowing me to perform a nuanced analysis of over- and under-reporting trends.

3. Given the skewed distribution of emissions discrepancies, I applied an appropriate statistical transformation to ensure reliable modeling.

4. To determine the predictors of discrepancies, I used two complementary approaches, namely Random Forest regression, to identify key predictors through machine learning and Bayesian Hierarchical Modelling to capture structured variability at multiple levels, such as facility, corporate ownership, and geographic region.

By integrating self-reported and independent satellite-based data with this study, I contribute to improving our understanding of reported emissions differences and furthering our understanding of the validity of satellite-based emissions accounting. The findings provide practical insights for policymakers, industries, and researchers, supporting efforts to enhance the accuracy and reliability of global emissions monitoring.

3.2 Data collection and matching process

3.2.1 GHGRP

I downloaded the GHGRP data from the EPA's Greenhouse Gas Reporting Program's website's Facility Level Information on GreenHouse Gases Tool (FLIGHT) (Environmental Protection Agency, 2024d) in CSV format, covering 2015-2022 and containing around 7,600 unique facilities. The resulting dataset held 61,358 rows of facility-level information on CO2e emissions, including an ID, the facility's name, latitude, longitude, county, parent company, industry, and emissions in tons. Next, I set out to preprocess the data, which included handling the 9.62% of rows missing a county. Addressing this was critical for later statistical modeling. To do so, I downloaded a dataset containing a lookup between US Zip Codes and counties from Kaggle (Ofer, 2018) and used this to fill in the missing counties, as all the GHGRP rows contained ZIP Code information. Following this approach, I reduced the percentage of missing counties down to 1.46% of total rows. Finally, I harmonized naming conventions for parent companies and counties using fuzzy matching (rapidfuzz, 2025), as many slight variations of the same names existed, which would have caused problems later.

3.2.2 Climate TRACE

The CT dataset I used in this analysis also covered 2015-2022, and I downloaded it in CSV format from the organization's website (Climate TRACE, 2024a). Like the first dataset, each row contained information on a facility's CO2e emissions for a single year, an ID, the facility's name, latitude, longitude, and industry sector. In total, the data frame contained 363,514 rows, covering the whole USA. Of these, 1.83% were missing emissions quantity information. As this information is critical for my analysis, I decided to omit the affected rows. Another challenge was that around 8,200 rows of data did not display yearly but monthly emissions data, so I needed to aggregate these to later compare them with the GHGRP dataset. Around 20,000 unique facilities were covered annually, significantly more than the GHGRP dataset because CT covers more industries than GHGRP and does not impose the 25,000 tons of CO2e threshold mentioned before.

3.2.3 Facility-level data matching

Now, both datasets were ready to be matched. The goal here was to find the CT equivalent for each row of GHGRP data so that the same facility information would be available for the same year. As no standard ID linked the two datasets, I needed to develop a matching algorithm based on multiple steps, including geographic information through the facility coordinates, similarity between facility names, and industry overlap. Another study has already validated a similar approach connecting the CT data to other facility-level emissions data sets (Gurney et al., 2024). This approach was designed to ensure accurate matches while minimizing false positives:

1. Coordinate-based matching: I used the facilities' latitude and longitude as key identifiers with a k-d tree algorithm (SciPy, 2024) to efficiently query facilities in the CT dataset within 5 meters of a given GHGRP facility. I chose this distance as a conservative threshold, as this would allow for variability while minimizing false positives.

2. Temporal filtering: Then, I restricted the resulting matches to the same reporting year to maintain temporal alignment between the datasets.

3. Category overlap filtering: Facilities were filtered further by comparing their broad emission categories. A match was only considered valid if the categories between the two datasets overlapped or the GHGRP facility was classified as "Other". I introduced this step as the industry sector in the two datasets was initially not aligned. The exact categorization can be seen in Table 1. This step reduced mismatches where geographically close facilities belonged to different industries, ensuring that only functionally comparable facilities were matched.

4. Name similarity scoring: Up to this point, multiple CT facilities could be valid matches for a GHGRP facility. To address this, I calculated name similarity scores using the fuzz.ratio function from the rapidfuzz library (rapidfuzz, 2025) for facilities that met the spatial and category overlap. This score, ranging from 0 to 100, quantified the textual similarity between facility names across datasets.

5. Combined scoring and ranking: I then integrated the spatial proximity and name similarity scores into a single combined score, calculated as:

$$\begin{aligned} Combined\ Score = 0.5 \times \left(1 - \frac{Distance}{5\,m}\right) \\ + 0.5 \times \left(\frac{Name\ Similarity}{100}\right) \end{aligned} \tag{1}$$

Where Distance is the distance in meters between possible matches. This formula equally weighted geographic proximity and textual similarity to prioritize accurate matches.

6. Selecting the best match: As a final step, for each GHGRP facility, I ranked all potential matches by

Table 1: Broad category mapping between GHGRP subparts and CT subsector

Broad category	GHGRP subparts	CT subsector
Manufacturing	Iron and Steel Production, Glass Production, Cement Production, Pulp and Paper Manufacturing, Lime Manufacturing, Ammonia Manufacturing, Petrochemical Production, Hydrogen Production, Aluminum Production, Electronics Manufacturing, Silicon Carbide Production, Soda Ash Manufacturing, Titanium Dioxide Production, Manufacture of Electric Transmission and Distribution Equipment, Lead Production, Phosphoric Acid Production, Nitric Acid Production, Ferroalloy Production, Zinc Production, Magnesium Production, Adipic Acid Production	other-manufacturing, aluminum, petrochemicals, steel, cement
Mineral Extraction	Underground Coal Mines	coal-mining, copper-mining, iron-mining
Power	Electricity Generation	electricity-generation
Transportation	Road Transportation, Domestic Aviation, International Aviation	road-transportation, international-aviation, domestic-aviation, domestic-shipping, international-shipping
Waste	Municipal Solid Waste Landfills, Industrial Waste Landfills, Industrial Wastewater Treatment	solid-waste-disposal, wastewater-treatment-and-discharge
Fossil fuel operations & Fluorinated Gases	Petroleum and Natural Gas Systems	oil-and-gas-production-and-transport, oil-and-gas-refining
Other	Geologic Sequestration of Carbon Dioxide, Use of Electric Transmission and Distribution Equipment, Importers and Exporters of Equipment Pre-Charged with Fluorinated GHGs or Containing Fluorinated GHGs in Closed-Cell Foams, HCFC-22 Production and HFC-23 Destruction, Injection of Carbon Dioxide, Fluorinated Gas Production, Petroleum Refineries, Suppliers of Natural Gas and Natural Gas Liquids, Suppliers of Petroleum Products, General Stationary Fuel Combustion Sources, Suppliers of Coal-based Liquid Fuels, Suppliers of Carbon Dioxide, Miscellaneous Uses of Carbonate, Suppliers of Industrial Greenhouse Gases	

Source: Own illustration modified from (Environmental Protection Agency, 2024f)

their combined score and selected the highest-ranking match as the most likely pair. This way, I could employ a transparent, scalable, empirically validated method for linking GHGRP and CT facilities while minimizing false matches.

The resulting data frame now contained 17,762 rows, which meant I found matches for 28.9% of all GHGRP rows. To further improve this number, I employed an additional fuzzy name-matching algorithm for the unmatched data:

1. Integrating ownership data: The GHGRP dataset held information for a facility's parent company. However, the original CT dataset did not have this information for all categories. Since ownership information can help resolve ambiguous matches, I merged ownership data from a separate CT dataset from the organization's website (Climate TRACE, 2024a).

2. Fuzzy name matching: The actual matching was done by applying fuzzy matching (rapidfuzz, 2025) using a facility's name and its parent company's name. This resulted in a text similarity score between 0 and 100 for all unmatched facilities of the same reporting year. I balanced the importance of facility name and parent company name equally, using the following formula:

$$Combined\ Score = \frac{Facility\ Name\ Similarity + Ownership\ Similarity}{2} \quad (2)$$

This way, I could account for facilities with similar names belonging to different companies. I considered two facilities a potential match if the score was above 70.

3. Extending potential matches with distance calculations: For the identified potential matches, I next added the distance between the facilities, using their Haversine distance – the distance between two points on Earth using their latitude and longitude - (PyPi, 2024) through the coordinates, as I did in the coordinate-based matching before. This allowed me to evaluate whether fuzzy-matched facilities were also geographically reasonable.

4. Manually selecting valid matches: The previous steps identified 38 potential matches. After manually checking each of these, considering their similarity score and distances, I used the GHGRP FLIGHT tool (Environmental Protection Agency, 2024d) and the interactive CT emissions map (Climate TRACE, 2024b) to inspect them. Doing so resulted in only 10 more valid matches, combined with the 17,762 rows from before.

3.2.4 Validation checks on matched data

I employed several validation checks on the final matched data to understand how well my matching process worked. An overview of the relevant metrics can be seen in Table 2 below.

I suspect the reason for the relatively low percentage of original GHGRP rows that found a match (28.96%) in the CT dataset is the different scopes of the two data sources ex-

Table 2: Overview of evaluation metrics of matched data

Number of matched rows	Fraction of rows matched	Fraction of unique facilities matched	Mean distance between facilities	Median distance between facilities	Mean facility name similarity	Median facility name similarity
17,772	28.96%	36.40%	0.87m	0.00m	92.39	100.00

Table 3: Percentage of matches per year

Year	Number of GHGRP rows	Fraction of rows matched (%)	Fraction of unique facilities matched (%)
2015	7,983	2.43	14.90
2016	7,643	2.41	14.79
2017	7,569	2.39	14.69
2018	7,676	2.42	14.92
2019	7,683	4.47	27.80
2020	7,643	4.38	27.27
2021	7,615	5.24	32.66
2022	7,546	5.22	32.53

Source: Own illustration

plained before and the conservative choice of using a maximum allowed distance between potential facilities of only 5 meters. I will address this tradeoff in a later sensitivity analysis exploring different distance thresholds. My primary goal for the principal analysis was to minimize the prevalence of false matches. I concluded that I succeeded by looking at the low distances between matches and the high similarity between matched facility names.

Further investigation of matches per year, shown in Table 3, reveals that while the number of GHGRP facilities per year is stable, more recent years resulted in a higher percentage of matches. To ensure this increase does not indicate flaws in my matching process, I investigated further to identify potential explanations for the trend.

I came up with several hypotheses for this increase:

1. Improved data coverage by CT over time: The organization improved its technology and coverage in more recent years, supported by its first emissions inventory released in 2021 (Climate TRACE, 2025a). For later years, this could mean more complete reporting, including more facilities covered, leading to better facility matching.

2. Improved industry representation in CT: If specific industries were underrepresented in earlier CT datasets but later improved, this would explain the increasing number of matches.

3. Changes in matching accuracy: If name similarity or location accuracy changes over time, this could lead to higher rates later.

Testing this hypothesis revealed several drivers for the increase in matches in later years:

1. The Facility coverage over time graph (Figure 2) shows that CT increased its number of tracked facilities significantly from 2019 onwards. Thus, it is only logical that the percentage of matches increases as the number of GHGRP facilities remains stable.

2. The Industry match rates over time graph (Figure 3) further supports this by showing shifts in the distribution of matched industries, suggesting better industry representation over time. This means that from 2019 onwards, more industry sectors were present in the CT dataset, leading to increased matches.

3. The increased distance between matched facilities (Figure 4) and the decrease in name similarity suggests a slight decline in match quality, likely leading to more matches as the criteria for determining a match became less strict.

In conclusion, there is data to support all three of my hypotheses for the increase in matches in later years. The primary reason for the increase is the expansion of the CT dataset after 2018. However, this is not evidence of a fundamental issue or a significant reduction in the quality of the matched data.

3.3 Calculation of GHG differences

My next goal was to quantify the discrepancies in reported emissions between matched facilities. Both datasets report metric tons of CO2e emissions using a Global Warming Potential (GWP) of 100 years(Climate TRACE, 2024a; Environmental Protection Agency, 2025b), making unit conversion unnecessary. I calculated absolute discrepancies in CO2e emissions using the following formula:

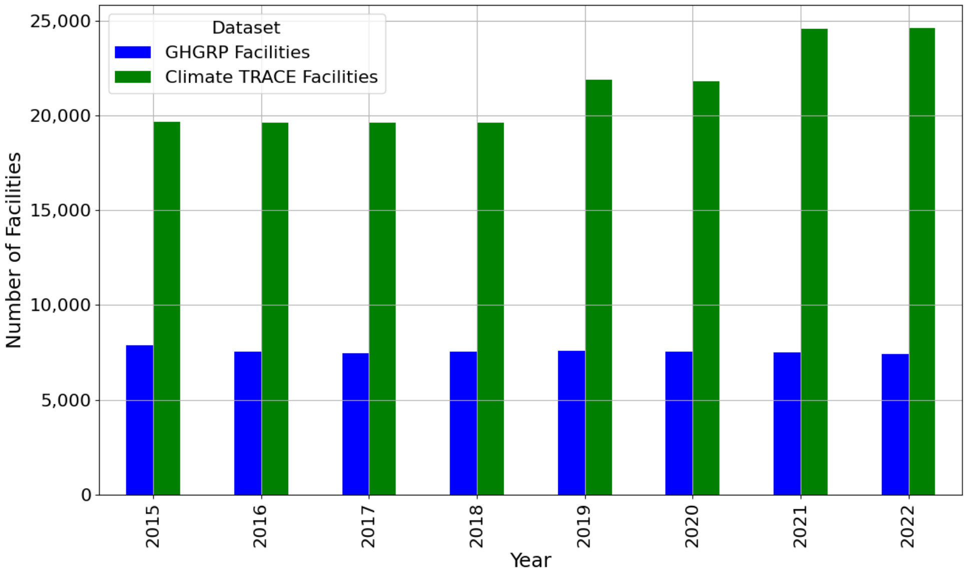

Figure 2: Facility coverage over time

Source: Own illustration

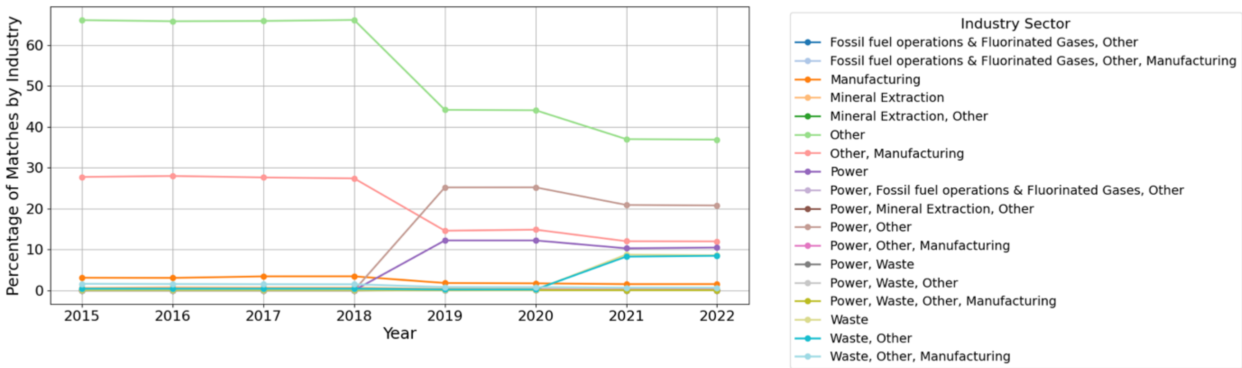

Figure 3: Industry match rates over time

Source: Own illustration

$$Emissions\ Discrepancy = GHGRP\ Emission - CT\ Emissions \tag{3}$$

resulting in direct differences in metric tons of CO2e. Using an absolute difference is consistent with previous comparative emissions studies (Gurney et al., 2024) to quantify over- or underestimation trends. I decided to use the emissions reported under GHGRP as the ground truth, as it is a legally mandated reporting system, following standardized methodologies, as explained before. CT is a relatively new initiative subject to fewer studies than GHGRP, so I assumed its data to be less reliable.

To enable additional analyses, I also calculated relative discrepancies, expressed as percentage deviations from GHGRP emissions, to normalize facility-scale effects:

$$Percentage\ Discrepancy = \left(\frac{Emissions\ Discrepancy}{GHGRP\ Emissions}\right) \times 100 \tag{4}$$

allowing for comparability across facilities with varying emission magnitudes. It should be noted that significant Emissions Discrepancies for low-emission facilities will disproportionately affect these results.

Of the 17,767 matched facility rows, 275 rows reported zero emissions discrepancies. I decided to exclude these from further calculations as they do not contribute to the analysis of discrepancies and could skew statistical interpretations by artificially lowering the observed variance in emissions differences.

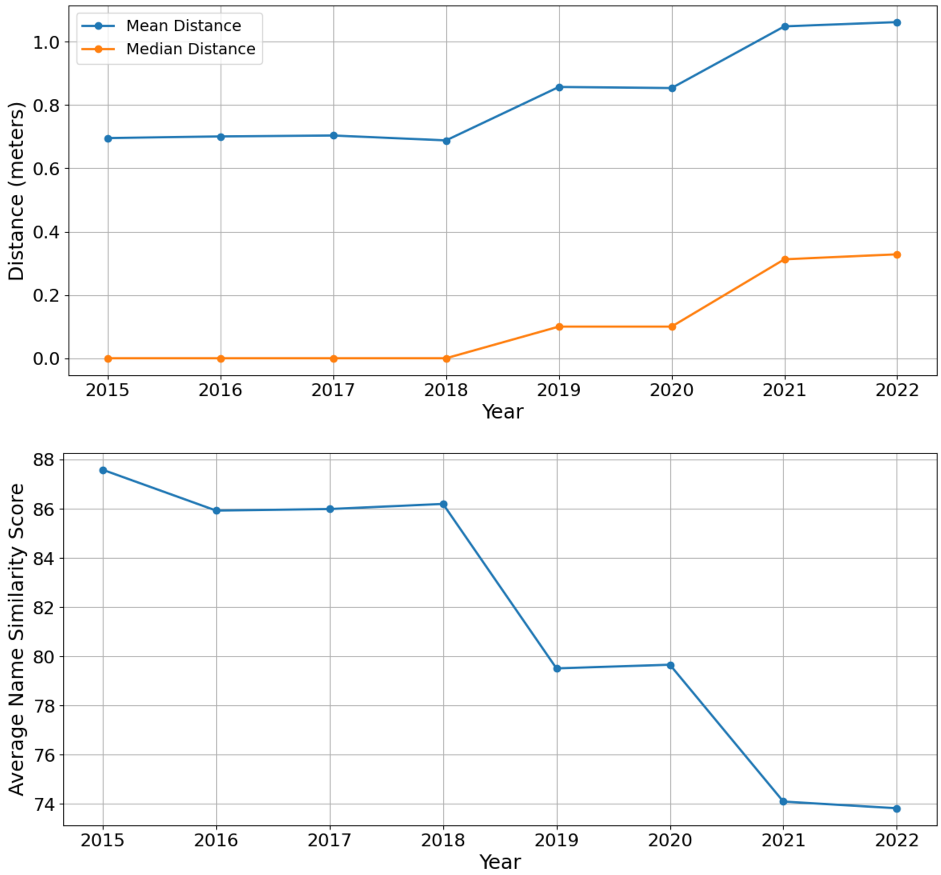

Figure 4: Distance between matched facilities (top) and name similarity over time (bottom)

Source: Own illustration

3.4 Outlier investigation

As a final data validation step, I manually reviewed the top 1% highest and top 1% lowest emission discrepancies. This was crucial in investigating the validity of these outliers, where discrepancies were extreme and potentially resulted from incorrect matches rather than genuine differences in reported emissions. I identified 15 facility-year pairs - less than 0.1% of the total matched data - that needed to be excluded from the analysis. My decision was based on manually comparing their GHGRP and CT facility names. I removed these mismatches to prevent outliers from disproportionately influencing statistical trends. This step enhanced the validity of subsequent analyses by ensuring that discrepancies are primarily driven by systemic reporting factors rather than data inconsistencies. Their low number speaks to the quality of the matching process, even before removing these false positives.

3.5 Variables and measures

Consistent with prior studies examining emission differences (Gurney, et al., 2024; Gurney, et al., 2021; He, et al., 2024; Papadopoulos, 2022), I used the absolute difference in reported emissions between the GHGRP and CT data sets, expressed as Emissions Discrepancies as the dependent variable of this paper. It is an absolute number of metric tons of CO2e, showing the extent of reporting deviations.

To explain these differences, I included a set of independent variables. The facility's GHGRP ID remains consistent across reporting years and, together with the Year of the report, enables the identification of within-facility variations over time. This allowed me to analyze temporal trends in discrepancies, particularly for facilities that report emissions in multiple years. Additionally, the reported CO2e emissions of a facility in the GHGRP dataset (Emission Tons) served as a proxy for its emission intensity, reflecting the scale of its GHG output. Each facility is classified into one or more industry categories (Subparts) according to the GHGRP classification system to capture potential industry-specific patterns. The County Name, representing the U.S. county where the facility is located, was included for potential geographic effects, such as environmental and climate conditions or infrastructure variations that may influence emissions reporting discrepancies. Finally, the parent company to which the facility belongs (Parent Company) is recorded to account for corporate-level influences on emissions reporting and potential systematic discrepancies across different organizations.

3.6 Data transformation and preprocessing

Before conducting statistical modeling, I applied multiple transformations to the dataset to address key data distribution issues, ensuring robustness in the analysis. To validate the appropriateness of parametric modeling approaches, I first examined the distribution of the dependent variable and its residuals (Figure 5). Tests for normality, including the Shapiro-Wilk test (González-Estrada, 2019) and D'Agostino's K-squared test (Statext LLC, 2024), revealed significant deviations from normality ($p < 0.001$). Additionally, the Breusch-Pagan (Bobbitt, 2020) test for heteroscedasticity indicated a significant variance instability ($p < 0.001$), suggesting that raw emissions discrepancy values were not homoscedastic across predictor variables.

These results indicated standard parametric models would be biased if applied to untransformed data. To address these issues of non-normality and heteroscedasticity, I explored several transformations on the dependent variable, including Box-Cox transformation, Log transformation, Quantile transformation, and Yeo-Johnson transformation. Each transformation was evaluated based on its impact on skewness, kurtosis, and normality test p-values. The resulting distributions can be seen in Figure 6 below.

Table 4 delivers further insights into these transformations.

Table 4: Transformation metrics

Transformation	Skewness	Kurtosis	Shapiro-Wilk (p)	Breusch-Pagan (p)
Original	0.37	49.50	2.19e-111	0.00
Box-Cox	1.16	44.96	3.16e-111	0.00
Log	-9.32	318.96	3.21e-116	0.00
Quantile	0.04	0.11	1.03e-04	0.00
Yeo-Johnson	-1.25	60.43	1.42e-111	0.00

The results of the transformation analysis, as presented in Table 4, demonstrate that the Quantile Transformation yields the most effective normalization of the emissions discrepancy variable. Compared to the original distribution, which exhibits high skewness (0.37) and extreme kurtosis (49.50), the Quantile Transformation reduces skewness to near zero (0.04). It drastically lowers kurtosis (0.11), resulting in a distribution that closely approximates normality. While alternative transformations such as Box-Cox and Yeo-Johnson improve distributional properties to some extent, they do not achieve the same level of symmetry and kurtosis reduction. The log transformation introduces extreme negative skewness (-9.32) and inflates kurtosis (318.96), making it unsuitable for this analysis. Furthermore, despite the Quantile Transformation's Shapiro-Wilk p-value (1.03e-04) indicating a statistically significant deviation from perfect normality, it remains superior to other transformations in reducing skewness and achieving a well-behaved distribution. This transformation ensured better model validity by improving adherence to the assumptions of parametric statistical methods while maintaining interpretability.

3.7 Modeling approach

Unlike previous studies on emissions discrepancies, which relied on conventional statistical techniques such as linear regression (Wang et al., 2022; Zhou et al., 2021), I implemented modern machine learning and Bayesian approaches to explain my dependent variable. My motivation is their ability to handle non-linear relationships and complex interactions that traditional regression methods may not capture ("Predictive Species and Habitat Modeling in Landscape Ecology", 2011). Additionally, as demonstrated in the previous section, parametric assumptions such as normality and homoscedasticity are violated in the raw emissions discrepancy data, so I chose methods less reliant on strict distributional assumptions (Kern et al., 2019).

A Random Forest Model is a non-parametric ensemble learning method that leverages multiple decision trees to improve predictive accuracy and reduce overfitting (Breiman, 2001). Unlike linear models, a Random Forest can automatically capture non-linear relationships between independent variables and emissions discrepancies. The method offers feature importance ranking, which allowed me to assess which predictors contribute most significantly to explaining discrepancies (Breiman, 2001). Furthermore, the model is robust to outliers and missing data, which is relevant to the data used in this paper, as identified earlier (Roy & Larocque, 2012).

I trained the model using Randomized Search Cross-Validation through the RandomizedSearchCV module from the sklearn library (scikit-learn developers, 2025), using the following hyperparameters:

- Number of trees (n_estimators = 500)
- Maximum depth of trees (max_depth = 50, None)
- Minimum samples per split (min_samples_split = 2, 5)
- Minimum samples per leaf (min_samples_leaf = 2, 3)

I employed a standard 70% training and 30% testing data split to evaluate the model's generalizability and used Mean Squared Error (MSE), R^2 score, and Mean Absolute Error (MAE) for model evaluation. After trying and comparing multiple different hyperparameter configurations, I saved the best model with the parameters above for further analysis.

Bayesian modeling allows for explicit uncertainty quantification in parameter estimates (Helleckes et al., 2022), which is particularly relevant to this paper's case of emissions accounting, where discrepancies may stem from multiple sources. Unlike frequentist models, Bayesian inference incorporates prior information, enabling more robust estimation with limited data in specific facility categories (Li et al., 2016). Using hierarchical effects allowed me to capture facility-specific, corporate-level, and regional influences on reporting discrepancies. Specifically, my model included:

- Facility-level effects to account for systematic biases in reporting at the facility level

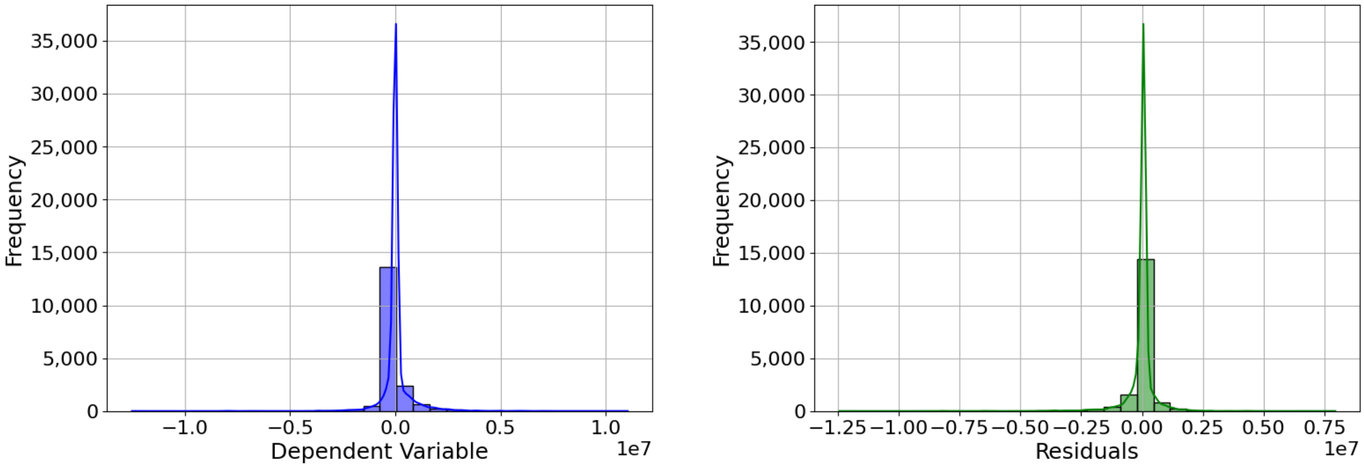

Figure 5: Original dependent variable and residual distributions

Source: Own illustration

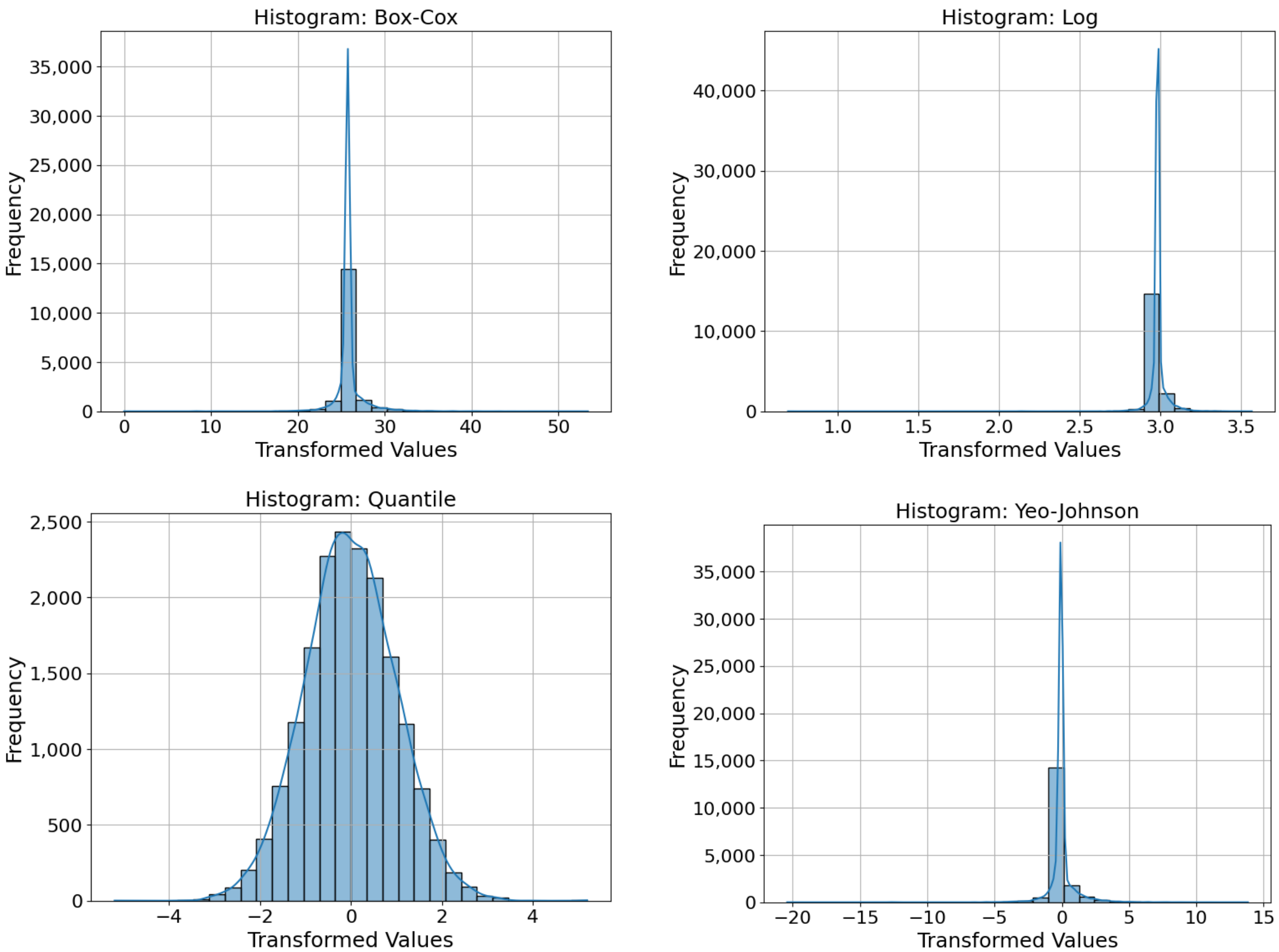

Figure 6: Overview of different transformations on the dependent variable

Source: Own illustration

- Parent company effects, as corporate structures may influence how emissions are reported

- County-level effects, capturing geographic influences on discrepancies
- Industry (Subpart) effects, accounting for sector-specific variations

I chose a Student-T likelihood function instead of a Gaussian likelihood because of the non-normal data distribution discussed earlier to account for potential heavy-tailed distributions in emissions discrepancies (Vogel et al., 2024). I then performed Markov Chain Monte Carlo (MCMC) sampling using the No-U-Turn Sampler (Hoffman & Gelman, 2014) with 1,500 posterior draws and 3,000 tuning steps, running four chains in parallel to ensure convergence, with diagnostics such as the Gelman-Rubin statistic (R-hat) checked for stability (Vehtari et al., 2021).

Comparing the two models, both offer unique advantages, further adding to the robustness of their insights. While Random Forest provides a black-box predictive model, Bayesian inference allows for interpretable parameter estimates and uncertainty quantification. By incorporating facility-specific priors, the Bayesian model can adjust for structural biases in reporting that the purely data-driven Random Forest model might not capture. Finally, I evaluated and compared both models' predictive accuracy. My Random Forest model identified key predictors of emissions discrepancies, while the Bayesian Hierarchical model provided a structured, interpretable framework for understanding uncertainty and variability in reporting discrepancies. This combined approach allowed me to develop a more comprehensive understanding of emissions discrepancies.

4 Results

4.1 Preliminary data exploration and descriptive statistics

In this chapter, I present an in-depth exploration of the dataset, focusing on patterns, trends, and key insights into discrepancies between GHGRP self-reported emissions and CT satellite-based estimates. My goal here is to describe the key statistical properties of the dataset before applying modeling techniques. This includes visualizations of the distribution of emissions discrepancies to identify patterns and systematic biases in different sectors, facility types, and geographic regions.

First, looking at the distribution of matched facilities over the distance between them (Figure 7) indicates that most matches have close to no space (0.000m) to one another.

This strongly supports the quality of the matched data. Next, the scatterplot comparing distance and emissions discrepancies (Figure 8) shows high variance at near-zero distances, suggesting discrepancies are not purely due to location and facility mismatches.

Variability decreases with distance, but discrepancies persist, indicating distance alone is not a key driver, showing no strong trend, and implying other factors drive the discrepancies. I thus concluded that the observed differences are not due to false matches.

As mentioned before, my primary variable of interest is the difference in reported emissions between GHGRP and CT, expressed in absolute metric tons of CO2e. For deeper insights, I also calculated percentage discrepancies to provide a relative difference, calculated as the absolute discrepancy divided by GHGRP-reported emissions. Finally, I computed weighted percentage discrepancies, which provide a measure accounting for the influence of facility size by weighting discrepancies based on each facility's total emissions, ensuring that large-emission facilities do not dominate the analysis disproportionately. The results for the matched data set are shown in Table 5 below.

The median absolute discrepancy of -2,175 tons CO2e suggests that at least half of the matched facilities exhibit only minor differences between the two data sources. This difference appears even more minor when considering that the median annual emissions per facility are 76,610 tons of CO2e for GHGRP facilities and 85,413.65 tons of CO2e for CT facilities. However, the mean absolute discrepancy is substantially higher (56,865 tons), indicating that extreme outliers skew the distribution. The median percentage discrepancy is close to zero (-0.0044%), but the mean percentage discrepancy is highly negative (-2,328.71%), reinforcing the influence of facilities with extreme percentage deviations. The median weighted percentage discrepancy is near zero (-3.99e-7%), while the mean weighted percentage discrepancy is -1.14%, revealing that extreme outliers strongly impact unweighted percentage calculations. The divergence between median and mean suggests that while most facilities have relatively small discrepancies, a subset of high-emission facilities accounts for disproportionately large absolute differences. The weighted percentage discrepancy further indicates that, when properly adjusted for facility size, the overall discrepancies are much more minor, suggesting that the extreme percentage deviations in the unweighted metric are not representative of the dataset. In the following sections, I explore these patterns in more depth.

I started by examining the distribution of emissions discrepancies. The histogram in Figure 9 - which I restricted to the 1st–99th percentile range of the emissions discrepancy to improve visual clarity - shows a slightly right-skewed distribution, meaning that while most discrepancies are minor, a few facilities exhibit extreme deviations. This further supports the findings from the raw discrepancy figures discussed before.

I concluded that most facilities report discrepancies close to zero but with a long tail in both positive and negative directions. Figure 10's Kernel Density Estimate plot provides a smoothed density representation of absolute discrepancies, further confirming the skewed distribution.

When looking at total emissions per data source, Figure 11 shows that GHGRP consistently reports higher total emissions per year for the matched facilities than CT. This is likely due to GHGRP's reliance on direct measurements and standardized reporting methodologies. In contrast, CT often extrapolates emissions using sector-wide models and satellite-derived estimates, which may not fully cap-

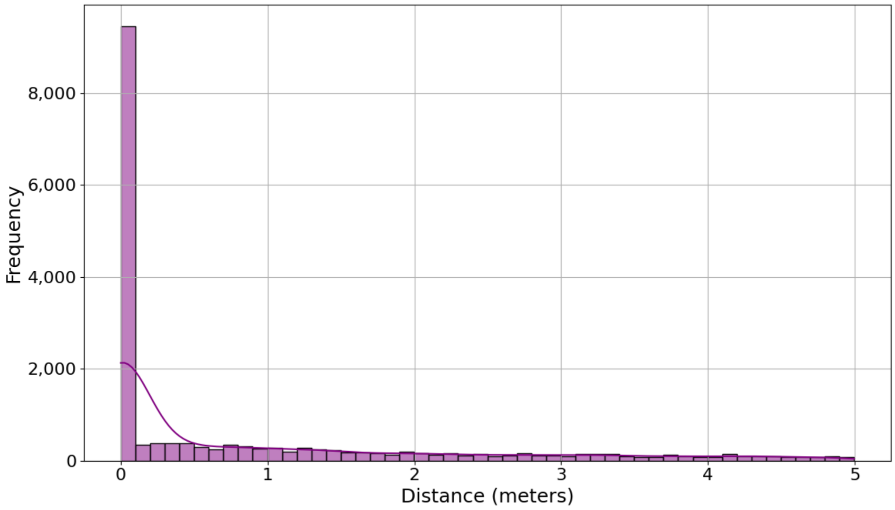

Figure 7: Distribution of matches by distance between them

Source: Own illustration

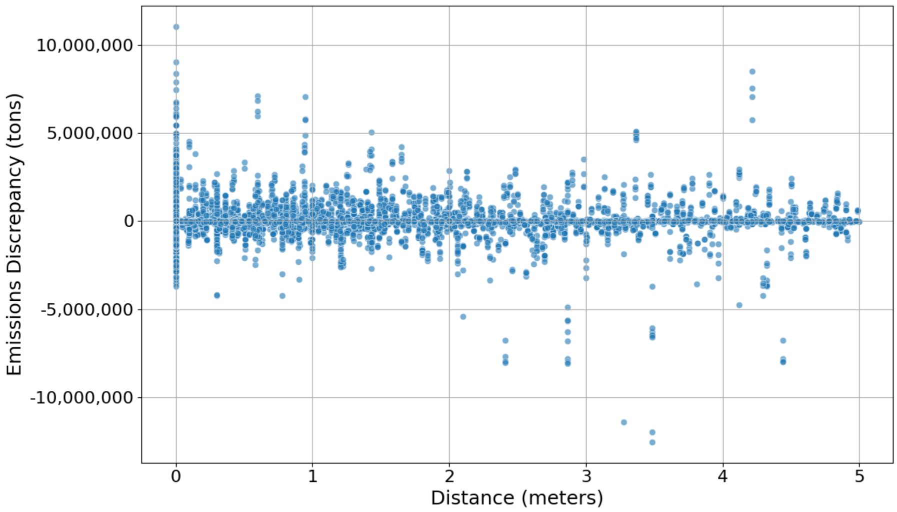

Figure 8: Distance vs. emissions discrepancy

Source: Own illustration

Table 5: Matched data emissions discrepancies

Statistic	Absolute discrepancy	Percentage Discrepancy	Weighted Discrepancy
Median	-2,175 tons	-0.0044%	-3.99e-7%
Mean	56,865 tons	-2328.71%	-1.14%

Source: Own illustration

ture specific facility-level emissions (Environmental Protection Agency, 2024c; Gurney et al., 2024). GHGRP also includes some fugitive and supplier-based emissions, which CT may not comprehensively account for, further contributing to the discrepancy (Dogniaux et al., 2024; Hanle et al., 2012).

The sharp increase in total emissions in both datasets from 2018 onwards can be explained through the combination of an increase in the number of matched facilities from 1494 in 2015 to 2,738 in 2019 and finally, 3,200 facilities in 2022, as well as the increase in the average emissions per facility, shown in Figure 12.

Looking at how discrepancies fluctuate over time, see Fig-

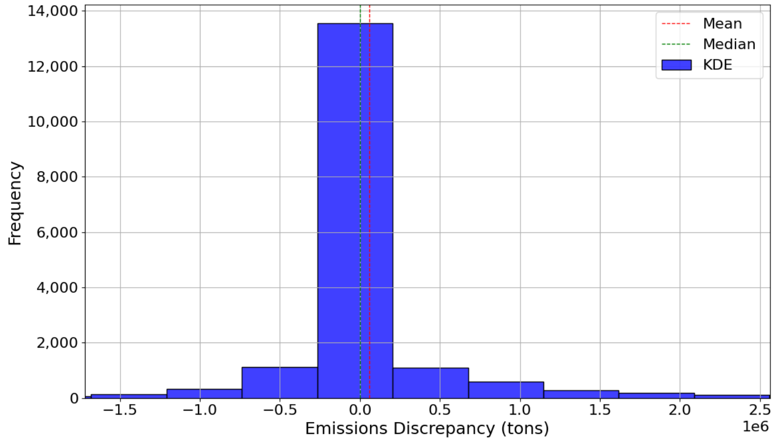

Figure 9: Histogram of absolute emissions discrepancies (1st to 99th percentile)

Source: Own illustration

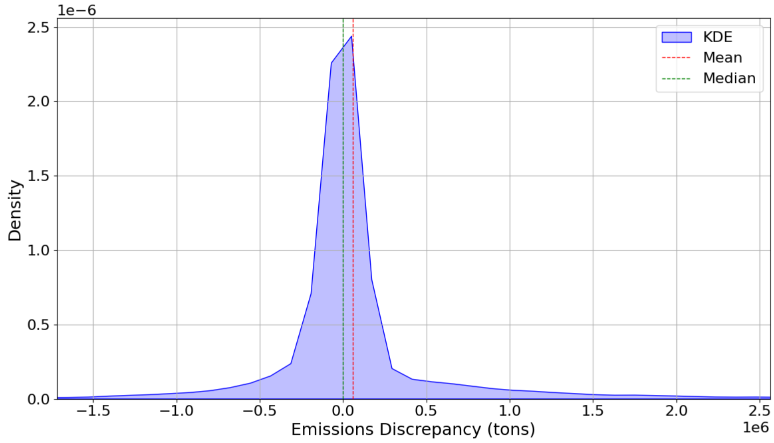

Figure 10: Kernel Density Estimate plot

Source: Own illustration

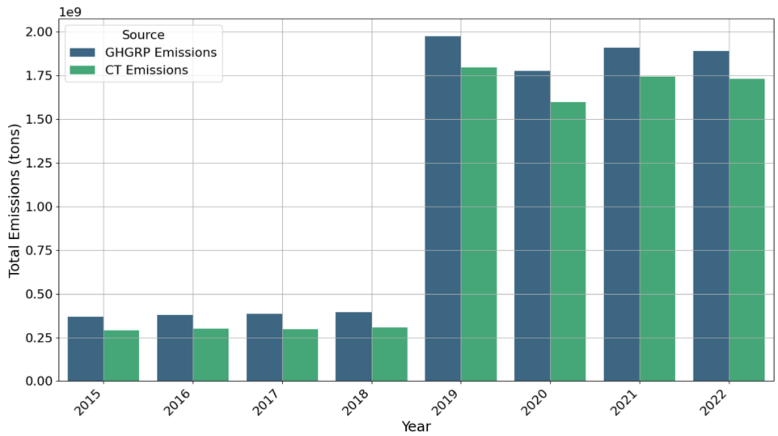

Figure 11: Total emissions by year (GHGRP vs Climate TRACE)

Source: Own illustration

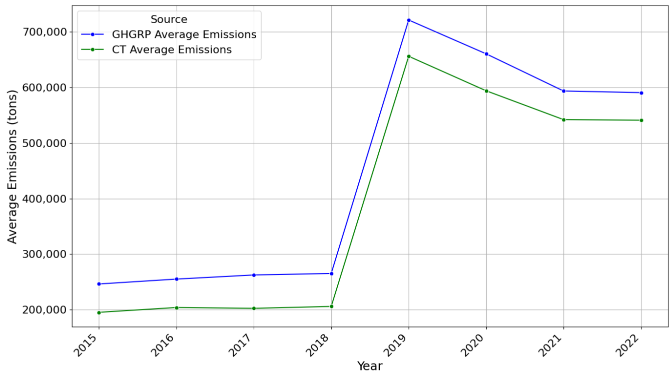

Figure 12: Average facility emissions per year (GHGRP vs Climate TRACE)

Source: Own illustration

ure 13. For absolutes - the left graph - shows a steady trend from 2015 to 2016, followed by a sharp increase until a peak in 2020. From then on, discrepancies declined in 2020 to below even in 2015. The discrepancies are always positive, in line with the total emissions per data source, showing higher emissions for GHGRP on average. However, this flips for relative discrepancies, and the percentage discrepancies become negative. This further shows how extreme outliers strongly impact unweighted percentage calculations.

Figure 14 displays the distribution of emissions discrepancies for each year to further examine how discrepancies evolved. While the previous trend analysis highlighted mean shifts, this visualization provides deeper insights into the dispersion and presence of outliers across different years. The boxplots reveal that while the median discrepancy remains relatively stable, the spread of discrepancies remains vast, with strong outliers in both directions across all years.

This distributional view strengthens prior observations that a subset of facilities consistently exhibits extreme discrepancies rather than random fluctuations affecting all facilities equally.

The next step was to examine how weighing emissions discrepancies by facility size affects these patterns (Figure 15). By adjusting for facility size, I could better assess whether the observed patterns hold across all emitters or if a subset of facilities primarily drives them.

This adjustment confirms that the extreme percentage discrepancies are primarily driven by small facilities, where even minor absolute differences lead to large relative deviations. When accounting for facility size, the overall trend in discrepancies remains more stable, indicating that the observed variations are less pronounced among larger emitters.

Finally, the impact of high-emissions facilities is made more evident by the direct relationship between discrepancies and the emissions reported under GHGRP (Figure 16).

A clear pattern emerges here: more emission-intensive facilities tend to have more significant absolute discrepancies. Notably, some high-emission facilities report extreme underestimations or overestimations.

I concluded that the divergence between absolute and percentage discrepancies stems from facility size effects. Smaller facilities often exhibit extreme percentage discrepancies despite minor absolute differences, while larger-emitting facilities dominate absolute discrepancies but show more minor relative errors. The weighted percentage discrepancy corrects for this imbalance, revealing that extreme unweighted percentage errors are largely statistical distortions. When adjusted for facility size, the overall discrepancies are much more minor, indicating that the high negative mean percentage discrepancy is driven by a few small emitters rather than systemic issues.

This showed that the differences in reported emissions are not static and have evolved significantly over time. As I discussed before, this increase coincided with an expansion of CT's dataset, suggesting that improvements in satellite-based monitoring may be influencing discrepancies.

Next, I explored if geographic patterns emerge, which might help understand the observed discrepancies.

The graphs in Figure 17 clearly show significant differences between the states, with certain states (e.g., North Dakota, Wyoming) exhibiting substantial positive discrepancies, suggesting GHGRP may systematically underreport emissions or CT may overestimate them. Other states (e.g., Georgia and Puerto Rico) show negative discrepancies, indicating potential over-reporting in GHGRP or under-detection by CT. This observation further stresses the need to consider the geographical location of facilities, which I did in the later statistical modeling.

At this point, the facilities' industry, or subsector, remained the last unexplored data parameter. To better understand its potential impact on emissions discrepancies, see Figure 18.

The first key observation is that the total amount of emissions reported is not equal between the different sectors, with a few industries like energy generation, manufacturing, or cement production dominating the reported emissions.

The plot in Figure 19 displays the relative percentage discrepancies across sectors.

International and domestic aviation exhibit the most negative discrepancies, suggesting GHGRP underreports emis-

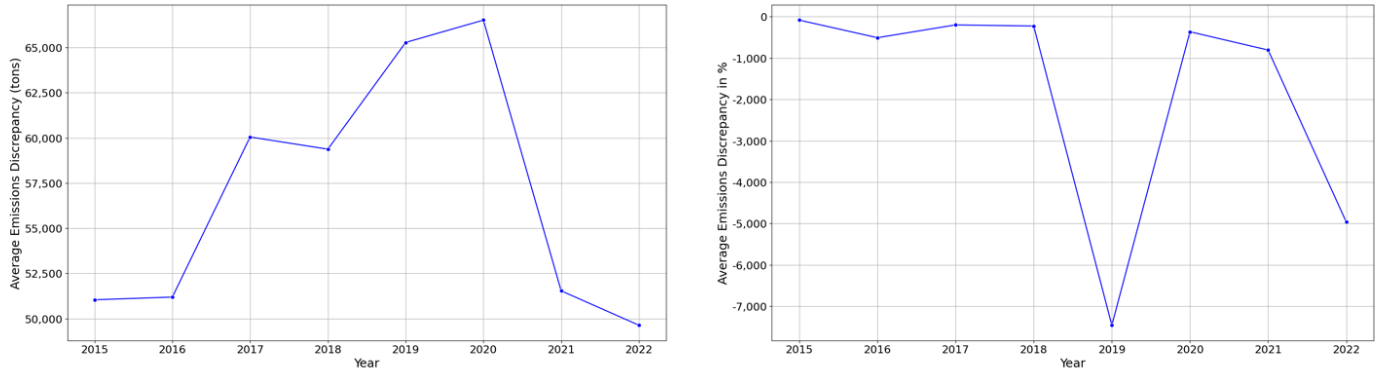

Figure 13: Temporal trend in emissions discrepancies (left: absolute, right: relative)

Source: Own illustration

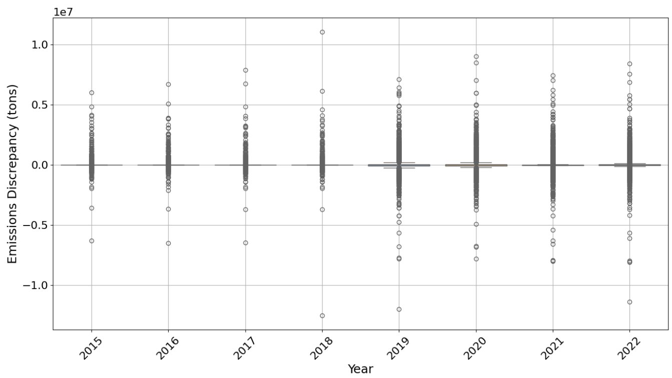

Figure 14: Distribution of emissions discrepancies by year

Source: Own illustration

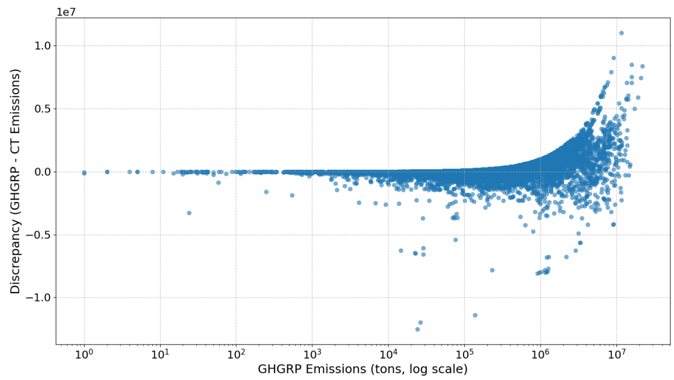

Figure 16: Emissions discrepancies vs facility size

Source: Own illustration

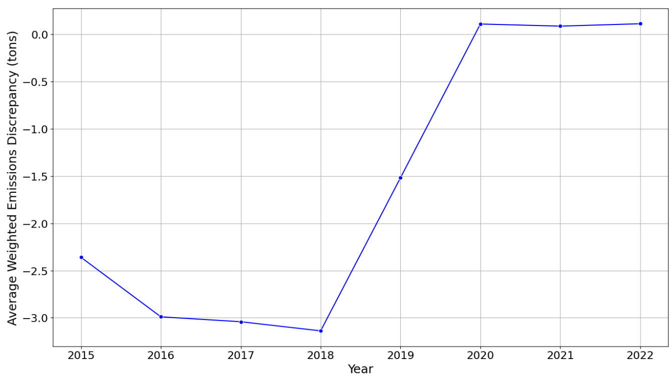

Figure 15: Temporal trends in weighted emissions discrepancies

Source: Own illustration

sions or CT overestimates them. As these percentages are unweighted and Figure 18 shows that these sectors have little total emissions, their significant percentage gaps dominate the plot, even though the total discrepancies are minimal. This point is further underscored by looking at the outliers for these sectors in Figure 20 below.

Electricity generation and international aviation also show substantial negative discrepancies, while sectors like aluminum and domestic shipping have relatively small percentage deviations. These differences highlight sector-specific biases that must be accounted for in further analysis.

Figure 21 shows the top 10 highest emitting facilities by average GHGRP emissions as a final insight.

Interestingly, all these facilities belong to the energy generation subsector.

In summary, this preliminary data exploration showed that emissions discrepancies, with sectoral and geographic differences, are not random. Furthermore, extreme outliers strongly influence mean values, necessitating careful statistical treatment, which I performed in the following section.

4.2 Main model results

This section presents the results from my Random Forest and Bayesian Hierarchical modeling. Both were applied to assess the key predictors of discrepancies between GHGRP self-reported and CT satellite-based emissions. As mentioned, the models differ in their approach: Random Forest captures non-linear interactions and produces substan-

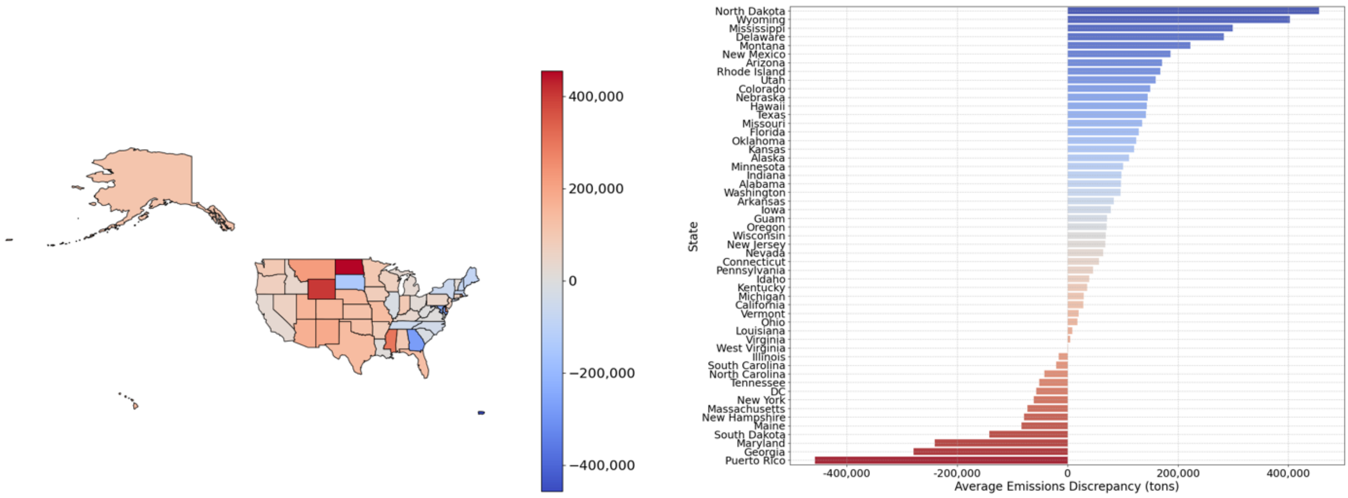

Figure 17: Emissions discrepancies by US State

Source: Own illustration

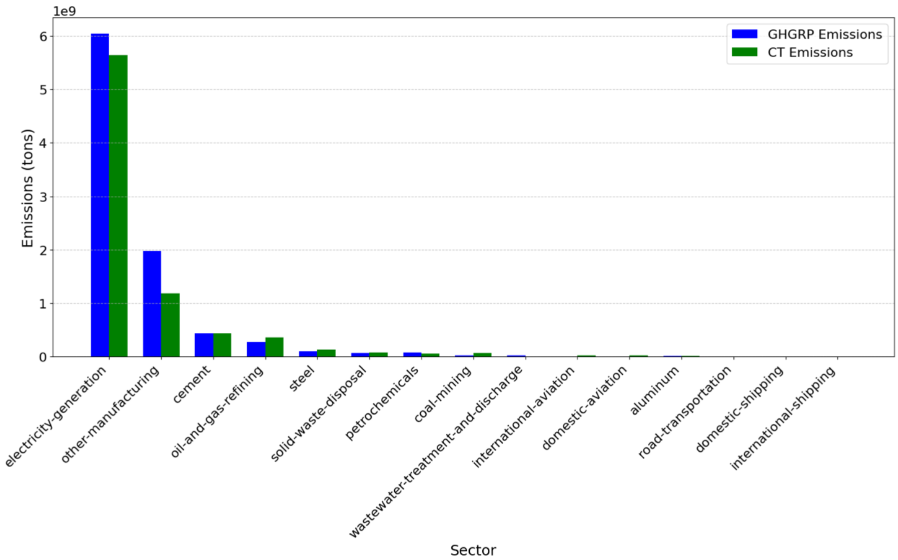

Figure 18: Sector-wise emissions comparison

Source: Own illustration

tial predictive accuracy, while Bayesian Hierarchical Modeling provides uncertainty quantification and variance decomposition. I first present each model individually, followed by a comparative analysis of their performance and insights.

The Random Forest model was evaluated using R^2, RMSE, and MAE to assess predictive accuracy. Cross-validation results were also analyzed to confirm robustness. The model exhibits strong predictive performance (Train $R^2 = 0.8926$, Test $R^2 = 0.8677$). Cross-validation across five folds confirms the model's generalizability (Mean $R^2 = 0.8410$) with minimal overfitting. The model also predicts discrepancies with a relatively small error (Test MAE = 0.2821); see Table 6 below.

These results suggest that the key predictors identified by the model generalize well beyond the training data, making this approach reliable for understanding emissions discrepancies.

Figure 22 presents the residuals vs. predicted values (left) and a Q-Q plot (right) to assess model performance beyond summary metrics.

The residual plot suggests a pattern of heteroskedasticity, where prediction errors tend to be more significant for extreme values, indicating that the model struggles more with high discrepancy cases. This suggests that the model may not fully capture some important factors influencing discrepancies. Additionally, the Q-Q plot shows deviations from a normal distribution, particularly at the tails, further confirming the presence of outliers that the Random Forest

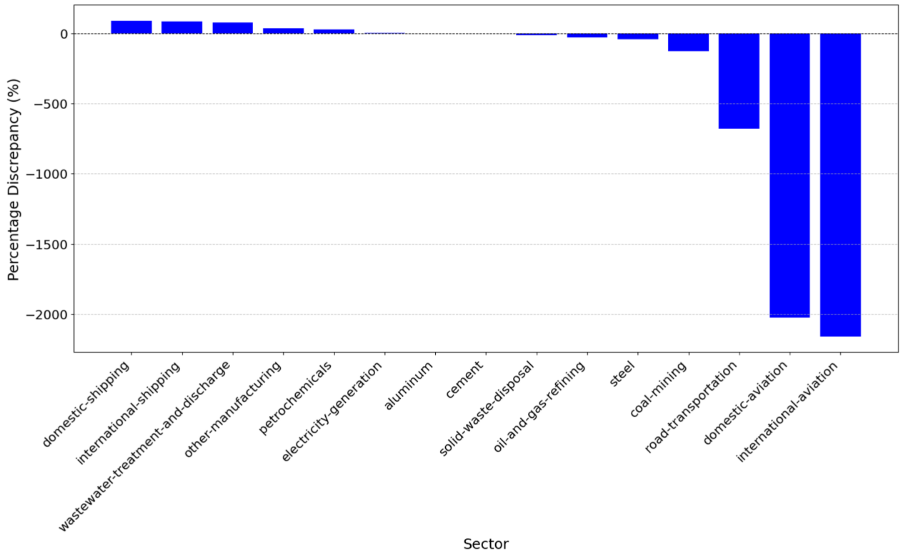

Figure 19: Percentage discrepancies by sector

Source: Own illustration

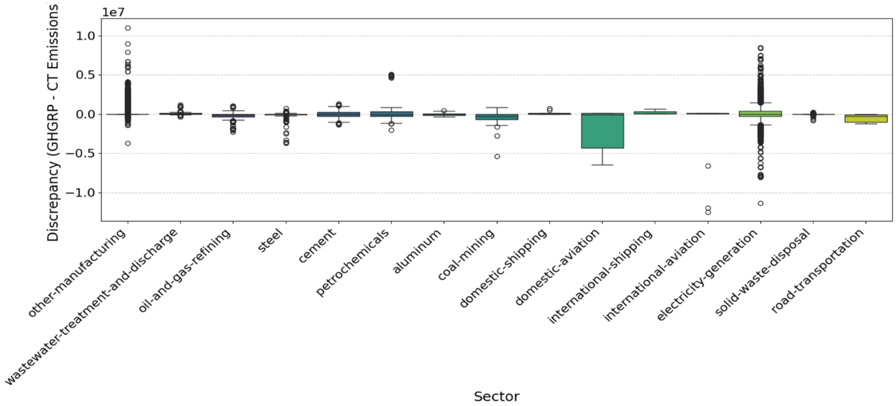

Figure 20: Discrepancies in emissions by sector

Source: Own illustration

model does not fully account for. These findings indicate that while the model performs well overall, it may be less reliable for predicting extreme discrepancies and could benefit from additional feature engineering or alternative modeling approaches. These diagnostics highlight the need for a robust statistical treatment of outliers and uncertainty quantification, which I later addressed with my Bayesian model.

I conducted a bootstrapped permutation importance analysis – which estimates feature importance by repeatedly shuffling the outcome variable and using bootstrap sampling to correct biases and improve reliability (Altmann et al., 2010) - to understand what drives emissions discrepancies according to my model, summarized in Table 7.

Facility-specific effects (GHGRP ID) dominate (40.78%) the explained differences in reported emissions between GHGRP and CT. This suggests that emission discrepancies depend highly on the individual facility rather than purely sectoral or geographic trends. Differences in reporting practices, operational factors, or data quality at specific facilities likely play a significant role. County and Parent Company Effects (24.23% and 22.56%) also contribute significantly, hinting at geographic or climate effects and corporate-level report-

Table 6: Random Forest performance metrics

Metric	Value
Train R^2	0.8926
Test R^2	0.8677
Cross-val. Mean R^2	0.8410
Cross-val. R^2	0.8419
Test MAE	0.2821

Source: Own illustration

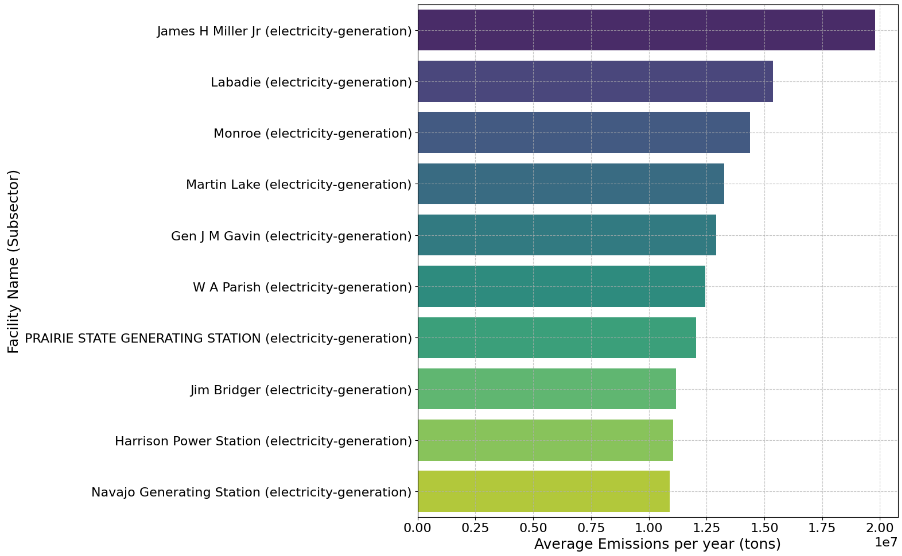

Figure 21: Top 10 emitters by average GHGRP emissions

Source: Own illustration

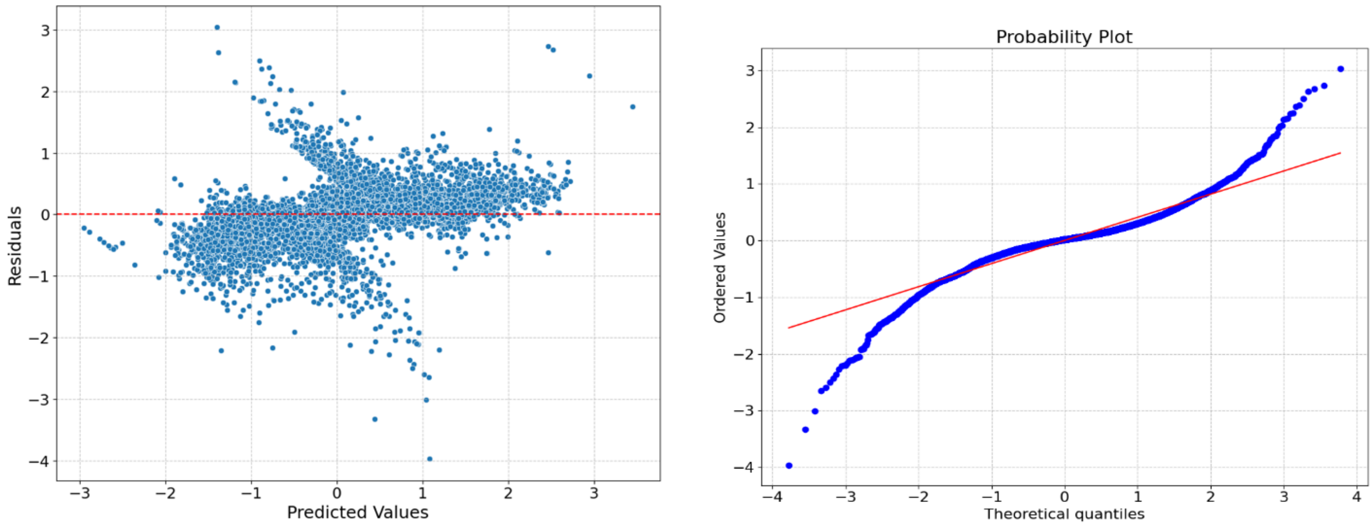

Figure 22: Residuals vs. predicted values (left) and Q-Q plot of residuals (right)

Source: Own illustration

ing strategies affecting emissions data. The industry Subpart (10.33%) to which a facility belongs also contributes considerably, indicating that specific industries or their methodologies may lead to systematic discrepancies. The reporting Year (2.11%) plays a minor role, showing that discrepancies do not change significantly over time, contrary to potential expectations that improved satellite monitoring might reduce them. Interestingly, after accounting for all other factors, the total Emissions Volume (0%) does not directly contribute to discrepancy predictions in this model. This contradicts prior assumptions that absolute emissions alone would drive discrepancies.

In summary, the Random Forest Model is highly predictive, but residual diagnostics reveal systematic variance in extreme cases, necessitating additional modeling approaches. Observed facility-level characteristics dominate the variance, while county and parent-company factors also play a decisive role. Emission volume is not a meaningful predictor, reinforcing the need to consider facility-specific biases and reporting practices. These findings suggest that discrepancies are not random but instead driven by structured facility, corporate, and regional factors - highlighting the importance of multi-level modeling.

I employed the Bayesian model to complement the Random Forest analysis by incorporating uncertainty quantification and variance decomposition. This approach enabled me to better understand the contribution of different hierarchical levels to the observed discrepancies.

Table 7: Random Forest model results

Normalized group contributions	Mean importance	Contribution (%)
Facility-Specific Effects (GHGRP ID)	0.16	40.78
County-Level Effects	0.09	24.23
Parent Company Effects	0.09	22.56
Subparts (Industry)	0.04	10.33
Year of Reporting	0.00	2.11
Emission Volume (tons)	0.00	0.00

Source: Own illustration

The Bayesian model exhibited strong predictive performance, with an overall Bayesian R^2 of 0.8660 and a test R^2 from cross-validation of 0.8659 ± 0.0141, closely aligning with the Random Forest model. The RMSE and MAE were 0.429 and 0.176, respectively, indicating relatively low error levels; see Table 8 below.

Table 8: Bayesian performance metrics

Metric	Value
Bayesian R^2	0.8660
RMSE	0.4290
MAE	0.1760
Cross-Validation R^2	0.8659 ± 0.0141
R-hat (Convergence)	≤ 1.01 (Stable)

Source: Own illustration

To assess the reliability of the estimates, I examined convergence diagnostics (R-hat values) and adequate sample sizes (ESS). Most parameters demonstrated good convergence (R-hat ≈ 1.0), but some parameters exhibited slight deviations, suggesting minor instability. Furthermore, I detected two divergences during sampling, indicating that future refinements (e.g., adjusting the step size or target acceptance rate) may improve model stability.

Figure 23 presents the residual distribution across cross-validation folds, indicating that residuals are symmetrically distributed around zero, with some variation at the tails. Compared to the Random Forest residuals, which exhibited clear heteroskedasticity, the Bayesian model's residuals appear more stable, suggesting that the hierarchical structure helps capture variance at different levels more effectively.

The variance decomposition analysis revealed that facility-level effects dominate, explaining approximately 93.95% of the variance, significantly higher than the 40.78% found in the Random Forest model. This suggests that the Bayesian model captures a more hierarchical structure, reinforcing that emissions data discrepancies are highly facility-dependent.

Table 9 summarizes the variance contributions across different hierarchical levels.

These findings confirm that individual facilities account for most observed discrepancies (93.95%), while County (3.67%) and Parent Company (1.6%) effects play secondary roles. Interestingly, the Subpart (0.01%) contribution is negligible, indicating that systematic discrepancies are more facility-specific than sector-wide. I omitted the emissions variable altogether, as it previously proved insignificant, and drastically reduced the model's computational burden.

Figure 24 presents the posterior distributions of key parameters, including uncertainty intervals for beta (fixed effects) and variance components.

The relatively narrow spread of these distributions suggests model stability.

Having analyzed both models separately, I now compared their performance and interpretability.

Both models exhibit similar R^2 values (∼0.86), suggesting they are equally effective at capturing the key predictors of emissions discrepancies. However, the Bayesian model achieves slightly lower RMSE and MAE, indicating marginally better predictive accuracy. Table 10 presents a side-by-side comparison of factor contributions and an average between them.

Table 10: Random Forest and Bayesian model comparison

Group contributions	Random Forest (%)	Bayesian Model (%)	Mean (%)
Facility-Specific Effects (GHGRP ID)	40.78	93.95	67.37
County-Level Effects	24.23	3.67	13.95
Parent Company Effects	22.56	1.60	12.08
Subparts (Industry)	10.33	0.01	5.17
Year of Reporting	2.11	-	1.05
Emission Volume (tons)	0.00	-	0.00

Source: Own illustration

Table 9: Bayesian variance contributions

Variance contribution	Mean %	Lower CI (2.5%)	Upper CI (97.5%)
Facility Effect	93.95	79.53	99.04
County Effect	3.67	0.45	17.74
Parent Effect	1.60	0.002	7.90
Subparts Effect	0.01	0.00	0.07
Residual	0.77	0.65	0.84

Source: Own illustration

While the reporting year is a minor factor in both models, this comparison highlights key differences:

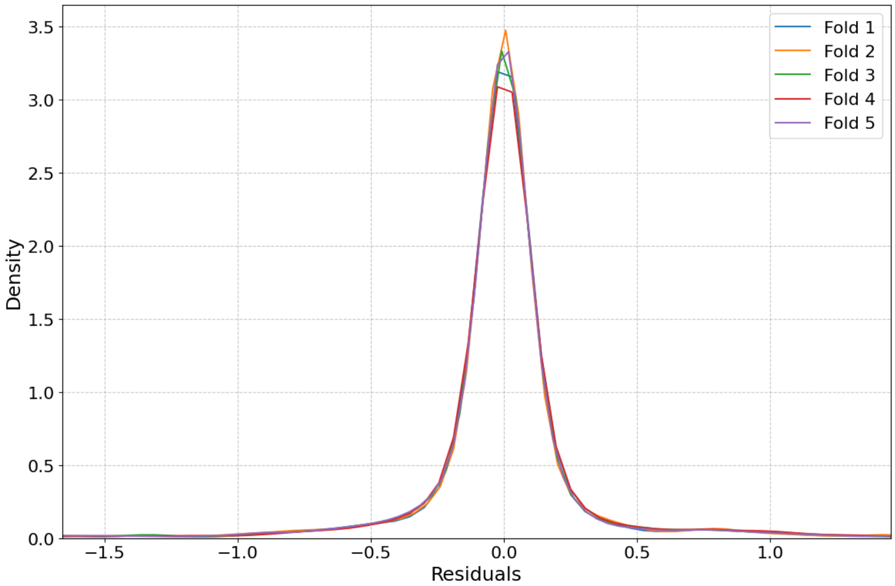

Figure 23: Bayesian residual distribution across folds

Source: Own illustration

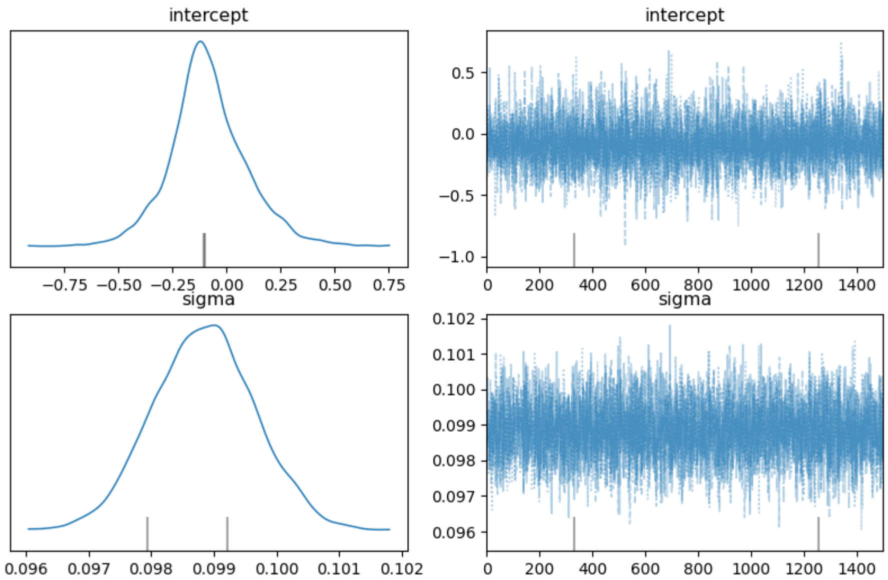

Figure 24: Bayesian variance contributions

Source: Own illustration

- The Bayesian model attributes most variance to facility-specific effects (93.95%), while the Random Forest model suggests that county- and parent-level effects also play a prominent role (24.23% and 22.56%, respectively).
- Industry (subpart) effects are negligible in the Bayesian model but contribute 10.33% in Random Forest, indicating that some industries may have systematic reporting differences.

However, both models identified facility-specific (GHGRP ID) effects as the major contributor, followed by County and Parent Companies. The Subpart (industry) of the facility, the Year of the report, and the Emissions volume are negligible.

4.3 Sensitivity analyses

To validate the results of this paper and ensure they are robust to alternative data and model configurations, I performed several sensitivity analyses.

4.3.1 Using different distance thresholds for coordinate-based matching

As mentioned, I decided on a 5-meter radius for the coordinate-based matching during my primary analyses, as this distance represents a conservative approach. The primary rationale behind this choice was to reduce the number of false-positive matches to a minimum while allowing for reasonable distances between potential matching facilities. However, this was still an individual selection and not empirically validated, so I set out to find the computationally optimal matching distance for the dataset. First, I reran the coordinate-based matching between GHGRP and CT data on multiple distance thresholds. Next, I performed a similarity evaluation of the names of the matched facilities of the two datasets using the token_set_ratio function from the rapidfuzz module, which returns a similarity score between 0 and 100 (rapidfuzz, 2025). I did this assuming that, on average, a higher name similarity between the two datasets' facility names implied a better match. Finally, I calculated the distances between matched facilities for the different thresholds, again assuming that the closer the two matched facilities are, the higher the likelihood of a correct match. An overview of the matching results can be seen in Table 11.

Table 11: Results of different distance thresholds on coordinate matching

Thresholds (m)	% Unique facilities	% Total rows	Median facility distance (m)	Median facility name similarity
5	36.40	28.96	0.00	100.00
10	42.04	33.07	0.28	100.00
15	45.52	36.11	0.56	100.00
20	47.60	38.17	0.78	100.00
25	49.19	39.91	0.95	100.00
30	50.76	41.80	1.20	100.00
35	52.36	43.60	1.43	100.00
40	53.94	45.60	1.80	100.00
45	55.02	47.00	2.09	100.00
50	56.04	48.40	2.43	100.00
75	60.68	55.05	4.35	100.00
100	63.32	58.69	6.35	84.21
500	77.27	72.62	25.35	54.55
1000	85.23	79.56	44.64	46.15

Source: Own illustration

I combined the three factors (percentage of matched rows, distance between facilities, and name similarity) into one score to calculate the optimal distance thresholds. First, I normalized them between 0 and 1 to compare them on the same scale. Then, I assigned weights to decide how important each factor was. For example, one set of weights might assign 40% to matched percentage, 40% to median distance, and 20% to name similarity. All weight combinations always added up to 1. For each weight combination:

1. I calculated an overall score for each threshold by multiplying each normalized factor by weight and adding them.

2. I identified the best threshold for that specific weight combination (the one with the highest score).

Figure 25 shows the effect of the different weight combinations on the threshold score.

Finally, I calculated the average overall score for each threshold by averaging its scores across all weight combinations. As shown in Figure 26, this allowed me to rank the thresholds and identify the one that performed best.

For 14 of 18 possible weight combinations, the threshold of 75m delivered the highest score. Since it also resulted in the highest average score, I concluded that this distance threshold delivers the best possible matches.

This new threshold increased the number of total matches by 90.07% from 17,772 to 33780 rows. Compared to the 5m data, the mean distance between matches increased from 0.87m to 16.84m, and the mean facility name similarity decreased from 92.39 to 72.01.

To assess if this increased maximum distance between facilities would change the results of my analysis, I reran the Random Forest Model used before on the new dataset. To ensure comparability with the 5-meter dataset, I did not change any of the model's hyperparameters. Table 12 shows the results of this analysis.

Consistent with earlier findings, facility-specific effects (GHGRP ID) remain the dominant contributors to the explained variance in the model. The relative contributions of predictors show some minor shifts when comparing the 5-meter and 75-meter datasets. GHGRP Parent Companies, for instance, increased from 22.6% to 24.1%, while GHGRP ID decreased slightly from 40.8% to 37.7%. Similarly, County

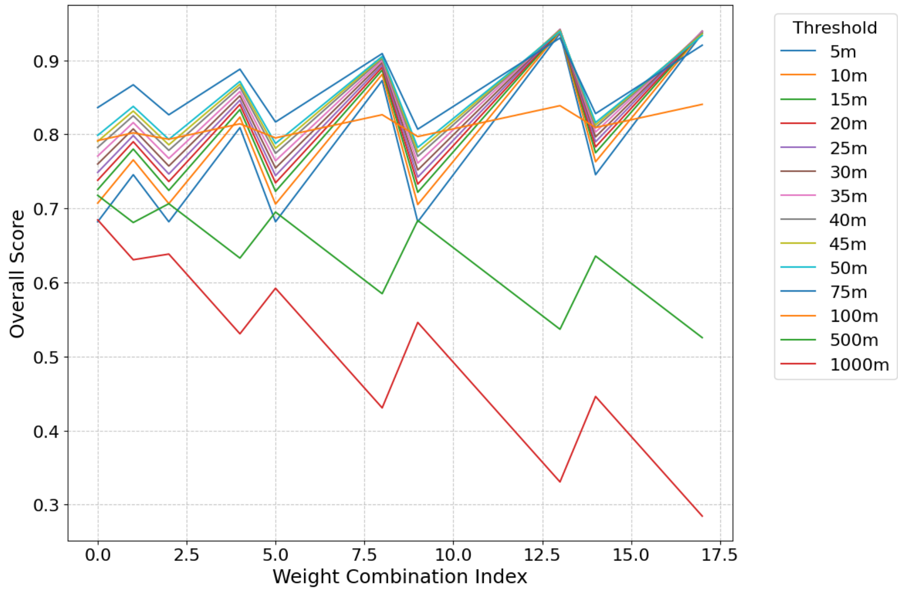

Figure 25: Variation in overall scores across weight combinations

Source: Own illustration

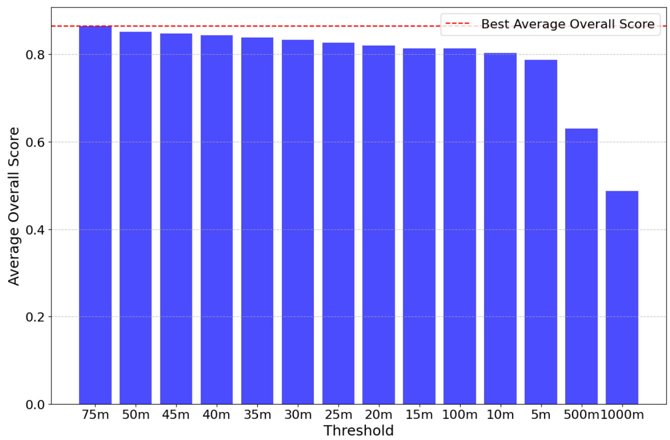

Figure 26: Average overall score across all weight combinations

Source: Own illustration

Table 12: Comparison of Random Forest results: 5m and 75m radius data

	5m	75m
Train R^2	0.8926	0.8779
Test R^2	0.8677	0.8498
CV Mean R^2	0.8410	0.8292
Mean Std Dev R^2	0.0419	0.0408
Contributions to explained variance:		
GHGRP ID	40.8%	37.7%
County Name	24.2%	20.1%
Parent Companies	22.6%	24.1%
Subparts	10.3%	14.8%
Year	2.1%	3.2%
Emissions tons	0.0%	0.0%

Source: Own illustration

Name decreased from 24.2% to 20.1%, while Subparts increased from 10.3% to 14.8%. These shifts likely reflect the inclusion of additional matches in the larger dataset, which introduces a broader range of facility and industry characteristics.

However, it is essential to note that these changes do not significantly alter the conclusion that facility-specific effects still dominate the explained variance. The contributions of shared effects, such as geographic location (County Name) and industry-specific characteristics (Subparts), remain secondary and show only marginal adjustments in their relative importance.

Moreover, the performance metrics of the Random Forest model remain robust. Train R^2 and Test R^2 show slight declines, from 0.8926 to 0.8779 and 0.8677 to 0.8498, respectively, while the cross-validated mean R^2 drops marginally from 0.8410 to 0.8292. The consistency of these metrics, even when the dataset size nearly doubles, demonstrates that the model's predictions are stable and not highly sensitive to changes in the spatial matching radius.

These results reinforce the robustness of my earlier findings. The slight shifts in predictor contributions do not undermine the central conclusion: emissions discrepancies are primarily driven by facility-specific factors, with shared effects such as geographic and industry-level characteristics contributing to a lesser extent.

4.3.2 Excluding matches with minor emission differences

Figure 27 shows the differences between the emissions reported by GHGRP and CT in absolutes.

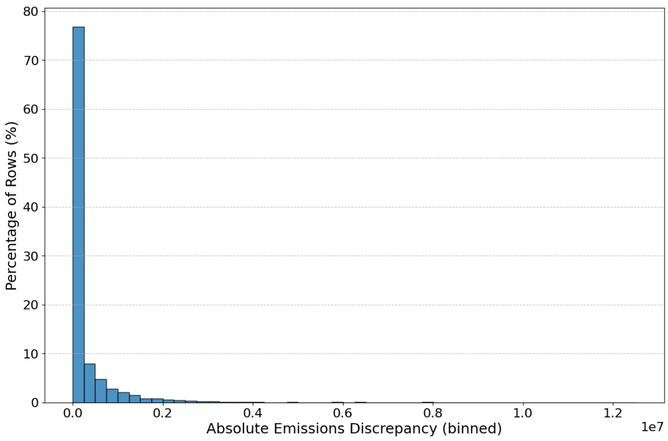

Figure 27: Distribution of absolute emissions discrepancies

Source: Own illustration

Here, most (70.23%) of the matches - the first bin - have an absolute difference of between 0 and 125,147.31 tons, which is only 0.009% of the maximum difference of 12,514,731.42 tons of CO2e. Since my earlier analyses identified facility-specific effects as the main drivers of the emissions discrepancies, I tested if this trend upholds when excluding matches with minimal emission differences from the dataset. I hypothesized that excluding matches with minimal differences may reveal more apparent patterns in the predictors of the dependent variable, as these minor differences might mask the proper drivers of variance in the dataset. To do so, I ran two additional Random Forest models, using the same hyperparameters again but excluding the data either ±1 or ±2 standard deviations around the mean of the above distribution. I chose this as it offers a data-driven approach to excluding minimal discrepancies, avoiding the arbitrariness of setting a fixed threshold, and ensuring consistency. The resulting two distributions are displayed in Figure 28 below.

This reduced the number of matches from 17,772 to 1,979 and 801 rows, respectively. Table 13 compares the results of the Random Forest models on the complete data with the results on the two reduced datasets.

Excluding matches with minimal discrepancies led to some shifts in the results, but the core findings remain robust. I observed increases in Train R^2, Test R^2, and Cross-Validation Mean R^2, with the Test R^2 rising from 0.8677 in the entire dataset to 0.9117 and 0.9118 in the ±1 and ±2 standard deviation datasets, respectively. This suggests that removing minor discrepancies reduces noise, allowing the model to better capture meaningful patterns. Since the gap between train and test R^2 did not significantly increase, overfitting did not either.

The contributions of predictors also shifted. GHGRP ID's importance decreased slightly (from 40.8% to 39.6% and 35.6%), while County Name's contribution increased significantly (24.2% to 31.0% and 33.0%). Parent Companies remained relatively stable, with a slight rise in the ±2 dataset (24.6%). Subparts decreased noticeably, from 10.3% in the whole dataset to below 5% in both reduced datasets. Year and Emissions Tons showed only minimal changes, with the latter remaining negligible.

These shifts suggest that removing minor discrepancies amplifies the role of shared factors like geography (County Name) while slightly reducing the dominance of facility-specific effects (GHGRP ID). However, the overall conclusions remain consistent: facility-specific effects still drive discrepancies in emissions reporting, with County Name and Parent Companies playing complementary roles. The robustness of the Random Forest model across datasets, as shown by stable performance metrics, supports the reliability of my earlier findings.

4.3.3 Using traditional models: Comparing results with Random Forest and Bayesian Models

As a final test of robustness, I applied two more traditional modeling approaches: ANOVA and Elastic Net regression. These models provide alternative perspectives on the drivers of emissions discrepancies and allow for a comparison with the results obtained from the primary analysis models. While Random Forest models excel at capturing complex, nonlinear relationships and Bayesian models offer probabilistic insights, ANOVA and Elastic Net offer interpretability and insight into variable significance through linear and variance decomposition frameworks (Breiman, 2001; Hans, 2012; Helleckes et al., 2022; Ståhle & Wold, 1989).

The ANOVA model explains 91.0% of the variance in emissions discrepancies (R^2 = 0.910, F(4732, 24298)=1.114*10**6, $p < 0.001$). Consistent with my primary analysis results, facility-specific effects dominated the explained variance. GHGRP ID accounted for 84.6% of the variance, significantly higher than all other predictors. Parent Companies contributed 14.5%, while County Name accounted for 0.9%. Subparts and Year's contributions were negligible, explaining less than 0.1% of the variance. Emissions Tons had no significant contribution. Table 14 shows these results.

The ANOVA results align closely with my main findings, confirming that facility-specific effects drive emissions discrepancies. These results underscore the robustness of the primary analysis and highlight the consistency of the insights across different modeling approaches.

Elastic Net regression, optimized with an alpha of 0.00047 and an l1_ratio of 1.0 (indicating a purely LASSO approach), further supported these findings. The model achieved a Train R^2 of 0.7715 and a Test R^2 of 0.7237, corresponding Mean Squared Errors (MSE) of 0.3822 and 0.4517, respectively. Cross-validation results indicated stability, with a mean R^2 of 0.6909 and a standard deviation of 0.0420, see Table 15.

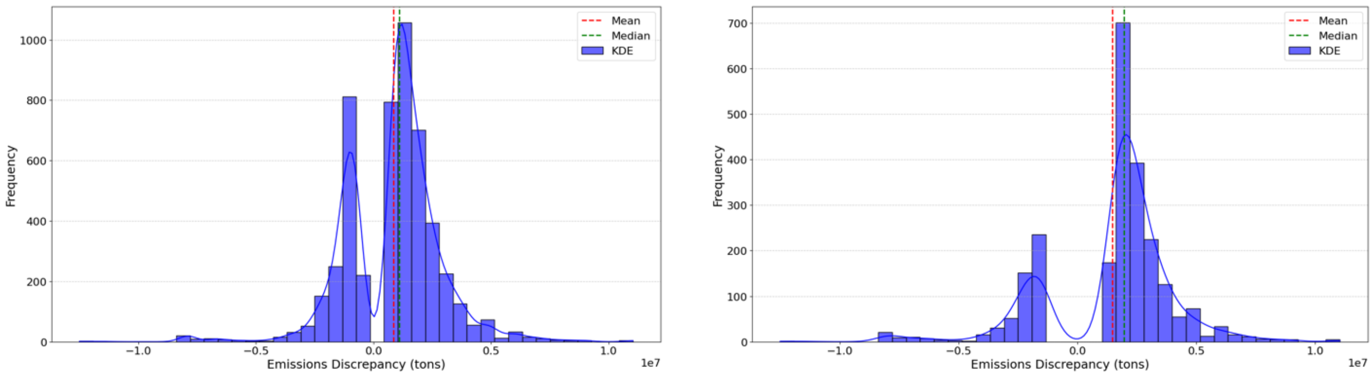

Figure 28: Distributions of emissions discrepancies with 1 (left) and 2 (right) SDs excluded

Source: Own illustration

Table 13: Random Forest models, original and minimal discrepancies removed

	Full data	±1 Std Dev removed	±2 Std Dev removed
Train R^2	0.8926	0.9405	0.9437
Test R^2	0.8677	0.9117	0.9118
CV Mean R^2	0.8410	0.8766	0.8527
Mean Std Dev R^2	0.0419	0.0380	0.0663
Contributions to explained variance:			
GHGRP ID	40.8%	39.6%	35.6%
County Name	24.2%	31.0%	33.0%
Parent Companies	22.6%	22.3%	24.6%
Subparts	10.3%	4.4%	3.9%
Year	2.1%	2.6%	2.9%
Emissions tons	0.0%	0.0%	0.0%

Source: Own illustration

Table 14: ANOVA results

Predictor	Sum of Squares	df	F	p-value	Partial Eta Squared
GHGRP ID	181,347,700	3483	350,239.38	< 0.001	0.846
County Name	1,864,718	985	12,734.54	< 0.001	0.009
Parent Companies	31,165,200	2250	93,173.79	< 0.001	0.145
Subparts	1.88	36	0.35	1.000	< 0.001
Year	12.22	7	11.74	< 0.001	< 0.001
Emissions tons	66.39	1	446.57	< 0.001	< 0.001

Source: Own illustration

Table 15: Elastic Net model results

Predictor	Total Contribution	Percentage (%)	Significant Predictors	Significance (%)
GHGRP ID	214.21	70.83	68	1.92
County Name	46.83	15.48	33	3.35
Parent Companies	39.23	12.97	47	2.05
Subparts	1.82	0.60	9	25.00
Year	0.01	0.003	1	100.00
Emissions tons	0.00	0.00	0	0.00

Source: Own illustration

Regarding predictor contributions, GHGRP ID again emerged as the dominant factor, accounting for 70.83% of the explained variance. Parent Companies contributed 12.97%, and County Name added 15.48%. Subparts and Year contributed minimally, explaining 0.60% and 0.34%, respectively, while Emissions Tons again had no significant impact.

The Elastic Net results confirm the dominance of facility-specific effects while providing additional nuance to the roles of shared factors. Although the contribution of County Name (15.48%) was slightly higher compared to the Random Forest results of my primary analysis, the overall pattern remained consistent: emissions discrepancies are primarily determined by facility-specific characteristics, with geographic and organizational factors playing supporting roles.

ANOVA and Elastic Net results align closely with the Random Forest and Bayesian models, consistently showing that facility-specific effects (GHGRP ID) dominate the explained variance. At the same time, Parent Companies, County Name and Subparts exert a weaker influence. The negligible impact of Year and Emissions Tons is also consistent across models.

The two linear models confirm the robustness of the more complex machine learning models, reinforcing the reliability of my conclusions. These additional analyses confirm that the main findings are not model-dependent but reflect authentic patterns in the data, further supporting their generalizability.

5 Discussion

In this study, I examined discrepancies between self-reported emissions from the Greenhouse Gas Reporting Program (GHGRP) and satellite-based estimates from Climate Trace (CT), identifying key predictors of these disparities. Understanding these differences is critical for improving emissions accounting accuracy, informing climate policy, and advancing hybrid verification frameworks integrating self-reported and independent monitoring approaches.

My analysis showed that facility-specific factors are the primary driver of discrepancies between GHGRP self-reported emissions and CT satellite-based estimates. County-level geographic factors and corporate (parent company) influences also contribute, but to a lesser extent. Notably, industry-wide effects are negligible, suggesting discrepancies stem from individual facility-level reporting practices rather than sector-wide methodological differences. Furthermore, the facility's total emissions volume and the report's year do not influence GHG discrepancies, which remained relatively stable over time. The sustained increase in emissions discrepancies from 2018 to 2020 suggests an influence of CT dataset expansion, possibly due to satellite monitoring and data coverage improvements.

These findings align with existing literature on GHG emissions; for example, they are consistent with Gurney et al. (2024), who found that facility-level biases significantly influence reported emissions. They also support research by Subramanian et al. (2015), which showed self-reported inventories underestimating fugitive emissions, particularly in oil and gas. My results align with those of Cusworth et al. (2021), who found no strong industry-wide effects, emphasizing reporting variability at the facility level. This is similarly supported by other research suggesting high facility-level variability in methane emissions (Papadopoulos, 2022).

However, my results also contradict some existing research. They contrast Nesser et al. (2024), who found systematic underreporting in landfill methane emissions, whereas this study finds both positive and negative discrepancies. While my study and Dogniaux et al. (2024) identify discrepancies between satellite-based methane estimates and facility-reported emissions, my study finds that CT's model-based approach introduces additional uncertainties, mainly due to extreme outliers and facility-specific factors. In contrast, Dogniaux et al. (2024) found no significant correlation between GHGSat satellite observations and reported emissions at all.

My analysis could not directly answer why facility-level factors dominate the observed discrepancies between reported GHG emissions from the two data sources. However, based on the reviewed literature and analysis results, the dominance of facility-level factors in explaining emissions discrepancies might stem from variability in emissions estimation methodologies, operational conditions, and regulatory flexibility. Facilities employ different calculation approaches, ranging from direct measurement (e.g., Continuous Emissions Monitoring Systems) to model-based estimation using default emissions factors, which inherently introduces variability. This methodological flexibility is a key limitation of carbon accounting systems like GHGRP, as facilities often select the approach that minimizes reported emissions, leading to systematic biases (Stark et al., 2024; Subramanian et al., 2015). Beyond methodology, operational conditions - such as equipment maintenance, process efficiency, and calibration practices - may also impact emissions measurement. For example, landfill methane emissions are significantly underestimated under "recovery-first" models, which assume optimistic gas capture rates despite empirical evidence suggesting substantial fugitive emissions(Nesser et al., 2024).

Additionally, local environmental factors, such as temperature and humidity, can influence emissions variability, particularly for landfills and petrochemicals, where atmospheric conditions affect methane oxidation and volatile organic compound emissions (Jain et al., 2021). This could explain my finding that a facility's geographic location (county) does influence the observed discrepancies. While corporate governance and geographic factors also contribute, their effects are secondary to facility-level variability. I suspect parent company policies and sustainability strategies shaping reporting transparency, as some firms engage in more rigorous voluntary disclosures than others. Similarly, county-level regulatory enforcement and state policies - although GHGRP is federally mandated - might indirectly influence reporting behaviors.

In contrast, industry-wide effects are minimal, suggesting that standardized sectoral methodologies do not eliminate intra-sectoral variability. While carbon accounting frameworks such as GHGRP provide industry-specific reporting guidelines, facilities might implement them differently, leading to inconsistent reporting practices within the same sector. As mentioned, this is supported by research indicating that discrepancies in reported emissions are primarily facility-driven rather than sector-wide trends (Cusworth et al., 2021; Gurney et al., 2024).

My findings underscore the need for hybrid verification systems, integrating bottom-up facility reporting with top-down satellite-based estimates, to reduce reliance on self-reported emissions data and enhance the accuracy of carbon accounting frameworks (Chan et al., 2024).

This study is not without limitations. First, I could only match 28.96% of GHGRP facilities to CT, meaning most discrepancies remain unaccounted for. Expanding the spatial matching threshold showed increased coverage but may introduce false positives. Thus, a more refined matching algorithm could increase the coverage and expand to the newest datasets from GHGRP and CT, including emission inventories for 2023, which were released after I finished my analysis. Second, as my study focuses solely on U.S.-based facilities, its generalizability to other countries using different reporting standards is limited. These might exhibit different patterns of discrepancies between their self-reporting and satellite-based emission data. Third, as the industry classifications between the two data sources do not perfectly align, I needed to introduce a manual mapping between them - see Table 1. I based this new categorization only on the existing category names, potentially introducing mismatches in their methodologies. This is especially relevant since earlier studies have found systematic underreporting in specific sectors (Nesser et al., 2024).

My study offers broad implications for the scientific literature, policy regulations, and practical applications. First, I contribute to scientific discussions on the need for hybrid verification approaches (Chan et al., 2024; Dogniaux et al., 2024; Goetz et al., 2009; National Academies of Sciences, Engineering, and Medicine, 2022) and show the potential of combining self-reported inventories with independent satellite-based monitoring. Second, my findings indicate that regulators should prioritize facility-level audits rather than industry-wide assessments, as discrepancies are facility-driven. Integrating third-party verification into programs like GHGRP, using satellite-based sensing approaches like CT, could enhance real-time detection of emissions events and improve data accuracy (Balasus et al., 2025; Cusworth et al., 2021; Duren et al., 2019). Finally, facility operators should use insights from third-party programs such as CT to refine their emissions monitoring practice and validate their data when creating reports, potentially lowering their risk of inaccurate reporting.

Building on the findings of my study, future research should expand and refine methodologies for assessing discrepancies between self-reported and satellite-based emissions data. A significant limitation of this study was the relatively low facility-matching rate (28.96%). Future research should refine matching algorithms to improve coverage while maintaining accuracy. This could involve integrating more detailed facility metadata, including geospatial attributes like temperature, humidity, and precipitation, and extending GHGRP and CT to include their newest inventories. Future research should additionally expand the comparison between CT and self-reported emissions to include other regulatory frameworks, such as the European Union Emissions Trading System (EU ETS), which covers over 11,000 power plants and industrial facilities (European Commission, 2025), Brazil's SEEG Inventory, which provides sectoral and facility-level emissions estimates (SEEG Community, 2025) and other national emissions reporting frameworks, such as China's Corporate GHG Inventory (UNFCCC, 2024) and Canada's Greenhouse Gas Reporting Program (Government of Canada, 2025). Likewise, comparing CT's emission data to other remote sensing systems, such as GHGSat (HGSat Inc., 2025) and NASA's OCO-2 and OCO-3 satellites (NASA, 2025a), could provide deeper insights into the reliability and consistency of satellite-based emissions monitoring. Beyond this, a thorough investigation into why specific sectors, namely aviation and road transport, exhibited far higher relative discrepancies in my analysis could result in a better understanding of potential weaknesses in CT's estimation algorithms. Finally, another promising avenue for future research is linking emissions discrepancies to corporate financial data. This would enable an exploration of whether financial incentives or governance structures influence emissions reporting accuracy. Specifically, this might include financial variables of facility-matched parent companies such as revenue, profitability, R&D investment, and ESG ratings. By expanding the analytical scope in these directions, future research can enhance the robustness of emissions verification methodologies, inform climate policy, and support the development of more reliable hybrid emissions monitoring frameworks.

6 Conclusion

With this study, I provide the first large-scale facility-level comparison of GHGRP self-reported emissions and CT satellite-based estimates. My analysis shows that facility-specific factors dominate emissions discrepancies, with geographic (county) and corporate (parent company) effects playing secondary roles. Industry-wide differences were minimal and remained stable over time, suggesting inconsistencies arise from individual reporting practices rather than systemic methodological flaws.

My findings underscore the limitations of self-reported emissions inventories and highlight the need for independent satellite-based verification to improve emissions monitoring. The results support hybrid verification systems integrating self-reported and remote-sensing data to improve accountability. Policymakers and facility operators should prioritize facility-level audits and verification mechanisms leveraging satellite-based insight to refine reporting accuracy and compliance strategies.

Future research should expand on my facility-matching methodology to improve dataset coverage, compare CT to other emissions inventories outside the US, assess discrepancies using additional remote sensing datasets, and investigate financial drivers of reporting discrepancies by linking emissions data with corporate financial performance.

Strengthening independent emissions verification is crucial for ensuring the accuracy of emissions accounting and advancing effective climate policies. My study highlights the need for integrated monitoring approaches, leveraging self-reported and satellite-based data to improve transparency, accountability, and climate action.

References

Alfieri, A., Holloway, L., von Kalckreuth, U., Moll, S., Schmieder, C., & Tissot, B. (2024). Empowering Carbon Accounting: From Data to Action. *Bank for International Settlements*. https://ideas.repec.org/p/bis/bisifr/16.html

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation Importance: A Corrected Feature Importance Measure. *Bioinformatics*, 1340–1347.

Ascui, F. (2014). A Review of Carbon Accounting in the Social and Environmental Accounting Literature: What Can It Contribute to the Debate? *Social and Environmental Accountability Journal*, 6–28.

Ascui, F., & Lovell, H. (2011). As Frames Collide: Making Sense of Carbon Accounting. *Accounting, Auditing and Accountability Journal*, 978–999.

Balasus, N., Jacob, D. J., Maxemina, G., Jenks, C., Nesser, H., Maasakkers, J. D., Cusworth, D. H., Scarpelli, T. R., Varon, D. J., & Wang, X. (2025). Satellite Monitoring of Annual US Landfill Methane Emissions and Trends. https://acmg.seas.harvard.edu/sites/projects.iq.harvard.edu/files/balasus-landfill-2024.pdf

Bauckloh, T., Klein, C., Pioch, T., & Schiemann, F. (2022). Under Pressure? The Link Between Mandatory Climate Reporting and Firms' Carbon Performance. *Organization & Environment*, 126–149.

Bobbitt, Z. (2020). Statology. https://www.statology.org/breusch-pagan-test/

Breiman, L. (2001). Random Forests. *Machine Learning*, 5–32.

Bronstein, K., Coburn, J., & Schmeltz, R. (2012). Understanding the Inventory of U.S. Greenhouse Gas Emissions and Sinks and the Greenhouse Gas Reporting Program for Landfills: Methodologies, Uncertainties, Improvements, and Deferrals. *International Emission Inventory Conference*.

Busch, T., Bassen, A., Lewandowski, S., & Sump, F. (2022). Corporate Carbon and Financial Performance Revisited. *Organization & Environment*.

Chan, E., Vogel, F., Smyth, S., Barrigar, O., Ishizawa, M., Kim, J., Neish, M., Chan, D., & Worthy, D. (2024). Hybrid Bottom-Up and Top-Down Framework Resolves Discrepancies in Canada's Oil and Gas Methane Inventories. *Nature*.

Chapple, L., Clarkson, P. M., & Gold, D. L. (2013). The Cost of Carbon: Capital Market Effects of the Proposed Emissions Trading Scheme (ETS). *Abacus*.

Chen, J., Zhao, F., Zeng, N., & Oda, T. (2020). Comparing a Global High-Resolution Downscaled Fossil Fuel CO2 Emission Dataset to Local Inventory-Based Estimates Over 14 Global Cities. *Carbon Balance and Management*.

Climate TRACE. (2024a). Climate TRACE. https://climatetrace.org/data

Climate TRACE. (2024b). Climate TRACE. https://climatetrace.org/explore

Climate TRACE. (2025a). Climate TRACE. https://climatetrace.org/about

Climate TRACE. (2025b). Climate TRACE. https://climatetrace.org/sectors

Couture, H., Alvara, M., Freeman, J., Davitt, A., Koenig, H., Kargar, A., O'Connor, J., Söldner-Rembold, I., Ferreira, A., Jeyaratam, J., Lewis, J., McCormick, C., Nakano, T., Dalisay, C., & Lewi. (2024). Estimating Carbon Dioxide Emissions from Power Plant Water Vapor Plumes Using Satellite Imagery and Machine Learning. *Remote Sensing*.

Csutora, M., & Harangozo, G. (2017). Twenty Years of Carbon Accounting and Auditing – A Review and Outlook. *Society and Economy*.

Cusworth, D. H., Duren, R. M., Thorpe, A. K., Olson-Duvall, W., Heckler, J., Chapman, J. W., Eastwood, M. L., Helmlinger, M. C., Green, R. O., Asner, G. P., Dennison, P. E., & Miller, C. E. (2021). Intermittency of Large Methane Emitters in the Permian Basin. *Environmental Science & Technology Letters*.

da Costa, L. M., Davitt, A., Volpato, G., Costa de Mendonça, G., Panosso, A. R., & La Scala Jr, N. (2025). A Comparative Analysis of GHG Inventories and Ecosystems Carbon Absorption in Brazil. *Science of The Total Environment*.

Deutsches Zentrum für Luft- und Raumfahrt e.V. (2024). CO2M - Copernicus Anthropogenic Carbon Dioxide Monitoring. https://www.d-copernicus.de/daten/satelliten/satelliten-details/news/co2m-copernicus-anthropogenic-carbon-dioxide-monitoring/

Dogniaux, M., Maasakkers, J., Girard, M., Jervis, D., McKeever, J., Schuit, B., Sharma, S., Lopez-Noreña, A., Varon, D., & Aben, I. (2024). Satellite Survey Sheds New Light on Global Solid Waste Methane Emissions. *EarthArXiv*.

Duren, R. M., Thorpe, A. K., Foster, K. T., Rafiq, T., Hopkins, F. M., Yadav, V., Bue, B. D., Thompson, D. R., Conley, S., Colombi, N. K., Frankenberg, C., McCubbin, I. B., & Eastwood, M. L. (2019). California's Methane Super-Emitters. *Nature*.

Environmental Protection Agency. (2009). Mandatory Reporting of Greenhouse Gases; Final Rule [Federal Register, 74(209)].

Environmental Protection Agency. (2017). EPA. https://www.epa.gov/sites/default/files/2017-12/documents/ghgrp_methodology_factsheet.pdf

Environmental Protection Agency. (2024a). Basic Information of Air Emissions Factors and Quantification. https://www.epa.gov/air-emissions-factors-and-quantification/basic-information-air-emissions-factors-and-quantification

Environmental Protection Agency. (2024b). EPA. https://www.epa.gov/ghgreporting/what-ghgrp

Environmental Protection Agency. (2024c). EPA. https://www.epa.gov/ghgreporting/ghgrp-newsletters

Environmental Protection Agency. (2024d). EPA FLIGHT. https://ghgdata.epa.gov/ghgp/main.do

Environmental Protection Agency. (2024e). GHGRP Overview Fact Sheet. https://www.epa.gov/system/files/documents/2023-08/ghgrp-overview-fact-sheet.pdf

Environmental Protection Agency. (2024f). Revisions and Confidentiality Determinations for Data Elements Under the Greenhouse Gas Reporting Rule. https://www.epa.gov/system/files/documents/2024-04/ghgrp-final-preamble-and-rule-april-2024.pdf

Environmental Protection Agency. (2025a). Inventory of U.S. Greenhouse Gas Emissions and Sinks. https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks

Environmental Protection Agency. (2025b). Overview of Greenhouse Gases. https://www.epa.gov/ghgemissions/overview-greenhouse-gases

European Commission. (2024). GHG Emissions of All World Countries. https://edgar.jrc.ec.europa.eu/report_2024

European Commission. (2025). EU Emissions Trading System (EU ETS). https://climate.ec.europa.eu/eu-action/eu-emissions-trading-system-eu-ets_en

European Commission, Joint Research Centre. (2023). JRC SCIENCE FOR POLICY REPORT GHG EMISSIONS OF ALL WORLD COUNTRIES. https://edgar.jrc.ec.europa.eu/booklet/GHG_emissions_of_all_world_countries_booklet_2023report.pdf

Fawzy, S., Osman, A. I., Doran, W. J., & Rooney, D. (2020). Strategies for Mitigation of Climate Change: A Review. *Environmental Chemistry Letters*, 2069–2094.

Godec, M. L., de Figueiredo, M. A., & Miller, M. L. (2022). Assessment of the Costs Associated with Monitoring, Reporting and Verification for Greenhouse Gas Reporting of Geologic Sequestration of Carbon Dioxide. *16th International Conference on Greenhouse Gas Control Technologies*.

Goetz, S. J., Baccini, A., Laporte, N. T., Johns, T., Walker, W., Kellndorfer, J., Houghton, R. A., & Sun, M. (2009). Mapping and Monitoring Carbon Stocks with Satellite Observations: A Comparison of Methods. *Carbon Balance and Management*.

González-Estrada, E. (2019). Shapiro–Wilk Test for Skew Normal Distributions Based on Data Transformations. *Journal of Statistical Computation and Simulation*, 3258–3272.

Government of Canada. (2025). Greenhouse Gas Reporting Program (GHGRP) - Facility Greenhouse Gas (GHG) Data. https://open.canada.ca/data/en/dataset/a8ba14b7-7f23-462a-bdbb-83b0ef629823

Gurney, K. R., Aslam, B., Dass, P., Gawuc, L., Hocking, T., Barber, J. J., & Kato, A. (2024). Assessment of the Climate Trace Global Powerplant CO2 Emissions. *Environmental Research Letters*, *19*(11).

Gurney, K. R., Liang, J., Roest, G., Song, Y., Mueller, K., & Lauvaux, T. (2021). Under-Reporting of Greenhouse Gas Emissions in U.S. Cities. *Nature Communications*.

Hanle, L., Kocchi, S., & Ragnauth, S. (2012). Developing a Facility-Level Reporting Programme for Sources of Non-CO2 Gases: The United States Experience. *Greenhouse Gas Measurement and Management*, 33–42.

Hans, C. (2012). Elastic Net Regression Modeling With the Orthant Normal Prior. *Journal of the American Statistical Association*, 1383–1393.

Hazaea, S., Al-Matari, E. M., Alosaimi, M., Farhan, N., Abubakar, A., & Zhu, J. (2023). Past, Present, and Future of Carbon Accounting: Insights from Scholarly Research. *Frontiers in Energy Research*, 1.

He, J., & Huang, B. (2023). Direct Estimation of Global Anthropogenic CO2 Emissions Using Satellite Data. *EGUsphere*.

He, M., Ditto, J., Gardner, L., Machechsky, J., Hass-Mitchell, T., Chen, C., Khare, P., & Sahin, B. (2024). Total Organic Carbon Measurements Reveal Major Gaps in Petrochemical Emissions Reporting. *Science*, 426–432.

Helleckes, L. M., Osthege, M., Wiechert, W., von Lieres, E., & Oldiges, M. (2022). Bayesian Calibration, Process Modeling and Uncertainty Quantification in Biotechnology. *Computational Biology*.

HGSat Inc. (2025). GHGSat. https://www.ghgsat.com/en/

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 1593–1623.

Huang, C., Patsika, V., Triantafylli, A., & Zhang, Y. (2023). Mandatory Greenhouse Gas Emissions Reporting and Firm Environmental Litigation Risk. *Accounting Forum*, 249–277.

Huertas, J. I., Huertas, M. E., & Díaz, J. (2012). Assessing Precision and Accuracy of Atmospheric Emission Inventories. *International Journal of Environmental Science and Technology*, 195–202.

IEA. (2024). CO2 Emissions in 2023. https://iea.blob.core.windows.net/assets/33e2badc-b839-4c18-84ce-f6387b3c008f/CO2Emissionsin2023.pdf

IPCC. (2023). CLIMATE CHANGE 2023 Synthesis Report Summary for Policymakers. https://www.ipcc.ch/report/ar6/syr/downloads/report/IPCC_AR6_SYR_SPM.pdf

Jain, P., Wally, J., Townsend, T., Krause, M., & Tolaymat, T. (2021). Greenhouse Gas Reporting Data Improves Understanding of Regional Climate Impact on Landfill Methane Production and Collection. *PLoS ONE*.

Kaur, R., Patsavellas, J., Haddad, Y., & Salonitis, K. (2023). The Concept of Carbon Accounting in Manufacturing Systems and Supply Chains. *Energies*.

Kennelly, C., Berners-Lee, M., & Hewitt, C. (2019). Hybrid Life-Cycle Assessment for Robust, Best-Practice Carbon Accounting. *Journal of Cleaner Production*, 35–43.

Kern, C., Klausch, T., & Kreuter, F. (2019). Tree-Based Machine Learning Methods for Survey Research. *Surv Res Methods*, 73–93.

Kiemle, C., Ehret, G., Amediek, A., Fix, A., Quatrevalet, M., & Wirth, M. (2017). Potential of Spaceborne Lidar Measurements of Carbon Dioxide and Methane Emissions from Strong Point Sources. *Remote Sensing*.

Klaaßen, L., & Stoll, C. (2021). Harmonizing Corporate Carbon Footprints. *Nature Communications*.

Kruger, J. (2005). From SO2 to Greenhouse Gases: Trends and Events Shaping Future Emissions Trading Programs in the United States. http://www.rff.org/

Lee, S., & Bi, X. (2020). Can Embedded Knowledge in Pollution Prevention Techniques Reduce Greenhouse Gas Emissions? A Case of the Power Generating Industry in the United States. *Environmental Research Letters*.

Lewandowski, S. (2017). Corporate Carbon and Financial Performance: The Role of Emission Reductions. *Business Strategy and the Environment*.

Li, B., Sun, Z., He, Q., Zhu, Y., & Qin, Z. (2016). Bayesian Inference with Historical Data-Based Informative Priors Improves Detection of Differentially Expressed Genes. *Bioinformatics*, 682–689.

Marlowe, J., & Clarke, A. (2022). Carbon Accounting: A Systematic Literature Review and Directions for Future Research. *Green Finance*.

Miller, M., Clark, T., Hadden, O., & Konovalchik, N. (2024). United States Greenhouse Gas Reporting Framework for CCS. *17th International Conference on Greenhouse Gas Control Technologies, GHGT-17*.

Moran, D., & Wood, R. (2014). Convergence Between the Eora, WIOD, EXIOBASE, and OpenEU's Consumption-Based Carbon Accounts. *Economic Systems Research*.

NASA. (2025a). Landsat 8. https://landsat.gsfc.nasa.gov/satellites/landsat-8/

NASA. (2025b). Orbiting Carbon Observatory-2. https://ocov2.jpl.nasa.gov/

National Academies of Sciences, Engineering, and Medicine. (2022). *Greenhouse Gas Emissions Information for Decision Making: A Framework Going Forward*. The National Academies Press.

Nesser, H., Jacob, D. J., Maasakkers, J. D., Lorente, A., Chen, Z., Lu, X., Shen, L., Qu, Z., Sulprizio, M. P., Winter, M., Ma, S., Bloom, A. A., Worden, J. R., Stavins, R. N., & Randles, C. A. (2024). High-Resolution US Methane Emissions Inferred from an Inversion of 2019 TROPOMI Satellite Data: Contributions from Individual States, Urban Areas, and Landfills. *Atmospheric Chemistry and Physics*.

Nguyen, H. L., Chau, L., Hiep, N. L., & Nguyen, D. T. (2025). Does Mandatory Greenhouse Gas Emissions Reporting Program Deter Corporate Greenwashing? *Journal of Environmental Management*.

Ofer, D. (2018). Kaggle. https://www.kaggle.com/datasets/danofer/zipcodes-county-fips-crosswalk

Papadopoulos, G. (2022). Discrepancies in Corporate GHG Emissions Data and Their Impact on Firm Performance Assessment. *SSRN Electronic Journal*.

Predictive Species and Habitat Modeling in Landscape Ecology. (2011). In *Predictive Species and Habitat Modeling in Landscape Ecology* (pp. 140–144). Springer.

Provaty, S., Hasan, M., & Luo, L. (2024). Organization Capital and GHG Emissions. *Energy Economics*.

PyPi. (2024). PyPi Haversine. https://pypi.org/project/haversine/

rapidfuzz. (2025). RapidFuzz 3.11.0 Documentation. https://rapidfuzz.github.io/RapidFuzz/Usage/fuzz.html

Rogelj, J., Huppmann, D., Krey, V., Riahi, K., Clarke, L., Gidden, M., Nicholls, Z., & Meinshausen, M. (2019). A New Scenario Logic for the Paris Agreement Long-Term Temperature Goal. *Nature*.

Roy, V., & Larocque, D. (2012). Robustness of Random Forests for Regression. *Journal of Nonparametric Statistics*, 993–1006.

Sanchez, M., Matthews, S., & Fischbeck, P. (2012). How Much Is United States Greenhouse Gas Emissions Certainty Worth? *Energy Policy*, 259–263.

Schaltegger, S., & Csutora, M. (2012). Carbon Accounting for Sustainability and Management. Status Quo and Challenges. *Journal of Cleaner Production*, 1–16.

Schrock, D., Stoops, J., Meier, A., Vine, E., & Solomon, B. (1994). Developmental Issues in Environmental Reporting Protocols. *American Council for Energy-Efficient Economy 1994 Summer Study on Energy Efficiency in Buildings*.

scikit-learn developers. (2025). RandomizedSearchCV. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html

SciPy. (2024). cKDTree. https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.cKDTree.html

SEEG Community. (2025). SEEG. https://seeg.eco.br/en/home/

Sitompul, M., Suroso, A. I., Sumarwan, U., & Zulbainarni, N. (2023). Revisiting the Impact of Corporate Carbon Management Strategies on Corporate Financial Performance: A Systematic Literature Review. *Economies*.

Ståhle, L., & Wold, S. (1989). Analysis of Variance (ANOVA). *Chemometrics and Intelligent Laboratory Systems*, 259–272.

Stark, B. M., Tian, K., & Krause, M. J. (2024). Investigation of U.S. Landfill GHG Reporting Program Methane Emission Models. *Waste Management*, 86–93.

Statext LLC. (2024). Normality Test. https://www.statext.com/practice/NormalityTest04.php

Stechemesser, K., & Guenther, E. (2012). Carbon Accounting: A Systematic Literature Review. *Journal of Cleaner Production*, 17–38.

Subramanian, R., Williams, L. L., Vaughn, T. L., Zimmerle, D., Roscioli, J. R., Herndon, S. C., Yacovitch, T. I., Floerchinger, C., Tkacik, D. S., Mitchell, A. L., Sullivan, M. R., Dallmann, T. R., & Robinson, A. L. (2015). Methane Emissions from Natural Gas Compressor Sta-

tions in the Transmission and Storage Sector: Measurements and Comparisons with the EPA Greenhouse Gas Reporting Program Protocol. *Environmental Science & Technology*, 3252–3261.
Tang, Q. (2017). Framework for and the Role of Carbon Accounting in Corporate Carbon Management Systems: A Holistic Approach. *SSRN Electronic Journal*.
Tang, Q. (2024). Managing the Shift from Voluntary to Mandatory Climate Reporting: The Role of Carbon Accounting. *SSRN*.
The European Space Agency. (2025). PlanetScope. https://earth.esa.int/eogateway/missions/planetscope
Tuesta, Y., Soler, C. C., & Feliu, V. R. (2020). Carbon Management Accounting and Financial Performance: Evidence from the European Union Emission Trading System. *Business Strategy and the Environment*.
UNFCCC. (2015). Paris Agreement. https://unfccc.int/sites/default/files/english_paris_agreement.pdf
UNFCCC. (2024). The People's Republic of China Fourth Biennial Update Report on Climate Change. https://unfccc.int/sites/default/files/resource/The%5C%20People%5C%E2%5C%80%5C%99s%5C%20Republic%5C%20of%5C%20China%5C%20Fourth%5C%20Biennial%5C%20Update%5C%20Report%5C%20on%5C%20Climate%5C%20Change.pdf
United Nations. (1998). Kyoto Protocol to the United Nations Framework Convention on Climate Change. https://unfccc.int/resource/docs/convkp/kpeng.pdf
U.S. National Archives and Records Administration. (2025). Code of Federal Regulations, Title 40, Chapter I, Subchapter C, Part 98, Subpart A. https://www.ecfr.gov/current/title-40/chapter-I/subchapter-C/part-98/subpart-A
Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bärkner, P.-C. (2021). Rank-Normalization, Folding, and Localization: An Improved $\hat{R}$ for Assessing Convergence of MCMC. *Bayesian Analysis*.
Vogel, R. M., Papalexiou, S. M., Lamontagne, J. R., & Dolan, F. (2024). When Heavy Tails Disrupt Statistical Inference. *The American Statistician*.
Wang, X., Chen, L. W., Lu, M., Ho, K.-F., Lee, S.-C., Ho, S. S. H., Chow, J., & Watson, J. G. (2022). Apportionment of Vehicle Fleet Emissions by Linear Regression, Positive Matrix Factorization, and Emission Modeling. *Atmosphere*.
World Resources Institute. (2024). Greenhouse Gas Protocol. https://www.itf-oecd.org/sites/default/files/1.2._ghg_protocol.pdf
Yang, L., Muller, N. Z., & Liang, P. J. (2024). The Real Effects of Mandatory CSR Disclosure on Emissions: Evidence from the Greenhouse Gas Reporting Program. *National Bureau of Economic Research (NBER)*. https://www.nber.org/system/files/working_papers/w28984/w28984.pdf
Zhou, Y., Zhang, J., & Hu, S. (2021). Regression Analysis and Driving Force Model Building of CO2 Emissions in China. *Scientific Reports*.