Estimating The Corporate Carbon Footprint

A Meta-Model Machine Learning Approach

Dr. Ben McNeil
Dr. Nicholas Pittman
Adam Kitto
Harrison Greven

EMMI

7 NOVEMBER 2024

Summary

The majority of publicly listed companies don’t report all carbon emissions, leaving large gaps for investors and regulators in understanding climate risk within the corporate sector. We introduce a new ensemble machine learning approach for estimating Scope 1, 2 & 3 emissions of public companies using standard financial data. We assess 30,000 simulations using five different performance metrics, creating a 'Meta-Model' of the top 10 performing models. This approach allows for comprehensive robust emissions coverage of up to 45,000 listed companies.

We estimate the cumulative emissions for the listed equity universe to be 11.4 billion tonnes in 2023, a reduction of ~2.5% from 2022. About 28% of those emissions come from the top 100 companies in the Western world. Our results are validated by independent reported data, with a median error rate of 6-10% for the most carbon-intensive sectors.

For investors, we estimate our approach gives a median weighted error of 15% for Scope 1, 17% for Scope 2 and 19% for Scope 3 emissions across the major indices in the US, UK, Europe, Canada or Australia.

Our Meta-Model approach offers greater stability, robustness and accuracy in estimating carbon transition risk footprints, a crucial first step for investors to understand and report their disclosures.

Emmi provides emissions data and climate risk analysis across all major public and private asset classes. Our data and analytics are built to support climate-related reporting, and feed directly into investment management processes.

Contents

Contents

01

Introduction

Understanding corporate carbon footprints is the first component required to understand transition climate risk assessment and sustainable finance alignment in the context of transitioning towards a low-carbon economy.

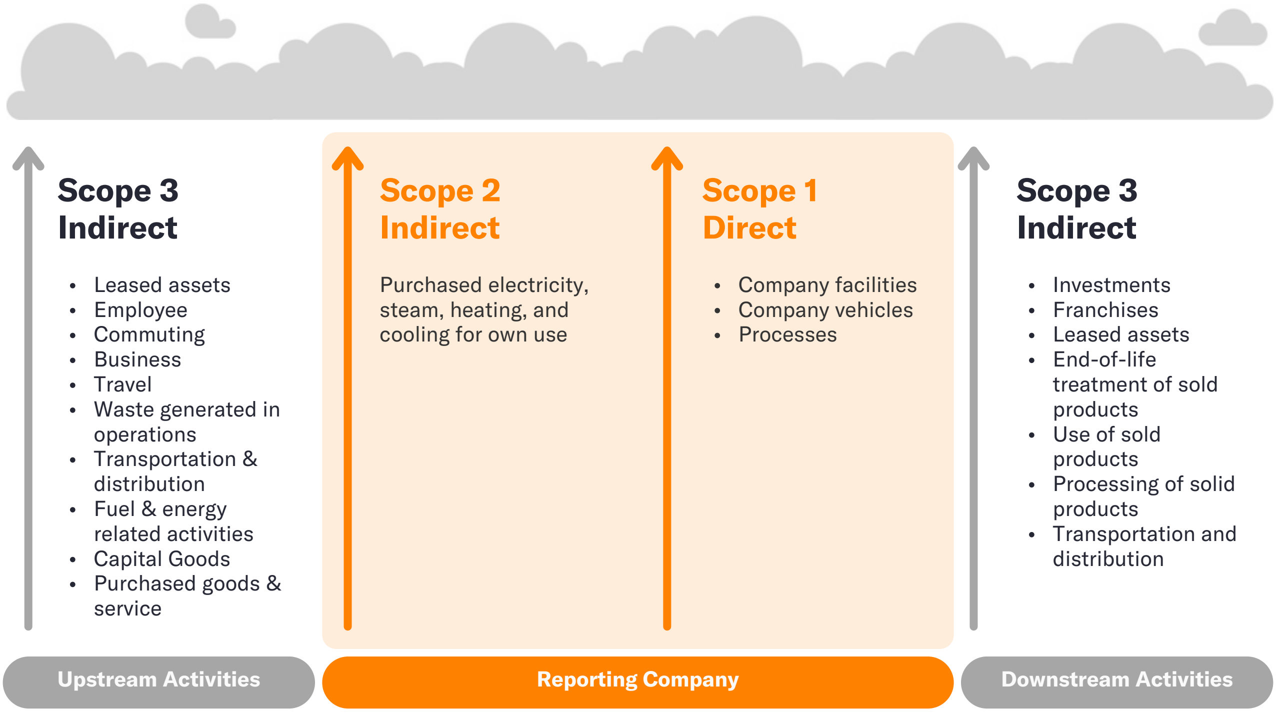
Reported Greenhouse Gas Emissions

Greenhouse gas emissions described and estimated in this study are reported in carbon dioxide equivalents (CO2-e). The vast majority of greenhouse gas emissions emitted within the industrial sector are carbon dioxide - so we refer to 'greenhouse gas emissions' as 'carbon emissions' in this study for brevity.

Carbon emission disclosure reporting is guided by the GHG Protocol which divides carbon emissions into three scopes:

- Scope 1 emissions are emissions which come directly from a company and its controlled entities
- Scope 2 emissions come from the generation of purchased energy
- Scope 3 emissions are all indirect (value chain) emissions - like investments or how customers use a product

Types of Emissions



Scope 3 Indirect

- Leased assets
- Employee
- Commuting
- Business
- Travel
- Waste generated in operations
- Transportation & distribution
- Fuel & energy related activities
- Capital Goods
- Purchased goods & service

Scope 2 Indirect

Purchased electricity, steam, heating, and cooling for own use

Scope 1 Direct

- Company facilities
- Company vehicles
- Processes

Scope 3 Indirect

- Investments
- Franchises
- Leased assets
- End-of-life treatment of sold products
- Use of sold products
- Processing of solid products
- Transportation and distribution

Chart 1: Different types of carbon emissions at the company level

The Importance of Scope 3 for Climate Risk

In the past decade, regulators have mostly focused on disclosures and reductions in a company's Scope 1+2 emissions (operational emissions). For instance, under Australia's climate laws, only the top 200 companies with high operational emissions are subject to emission caps. Companies that have high Scope 1 emissions include fossil-fuel electricity generation, steel production, or cement manufacturing - so regulators have focused on them. However, other companies and sectors in materials and manufacturing have most of their emissions coming from Scope 3.

Scope 3 emissions are difficult to accurately measure, and companies have little direct control over their reductions, since those emissions occur upstream and downstream of a company's products.

However as carbon pricing and regulation get sharper, the market will shift away from carbon-intensive products towards lower carbon products/alternatives. So understanding sectors that have high Scope 3 exposure is financially prudent to truly understand carbon risk. Take airlines and oil companies, for instance. They have a significant portion of their total emissions as Scope 3, making it essential to consider in terms of financial risk rather than just focusing on Scope 1 and 2 emissions.

From a climate risk perspective, we believe it is inadequate to only focus on Scope 1 and 2 emissions and any comprehensive model to estimate emissions must require Scope 3 as well.

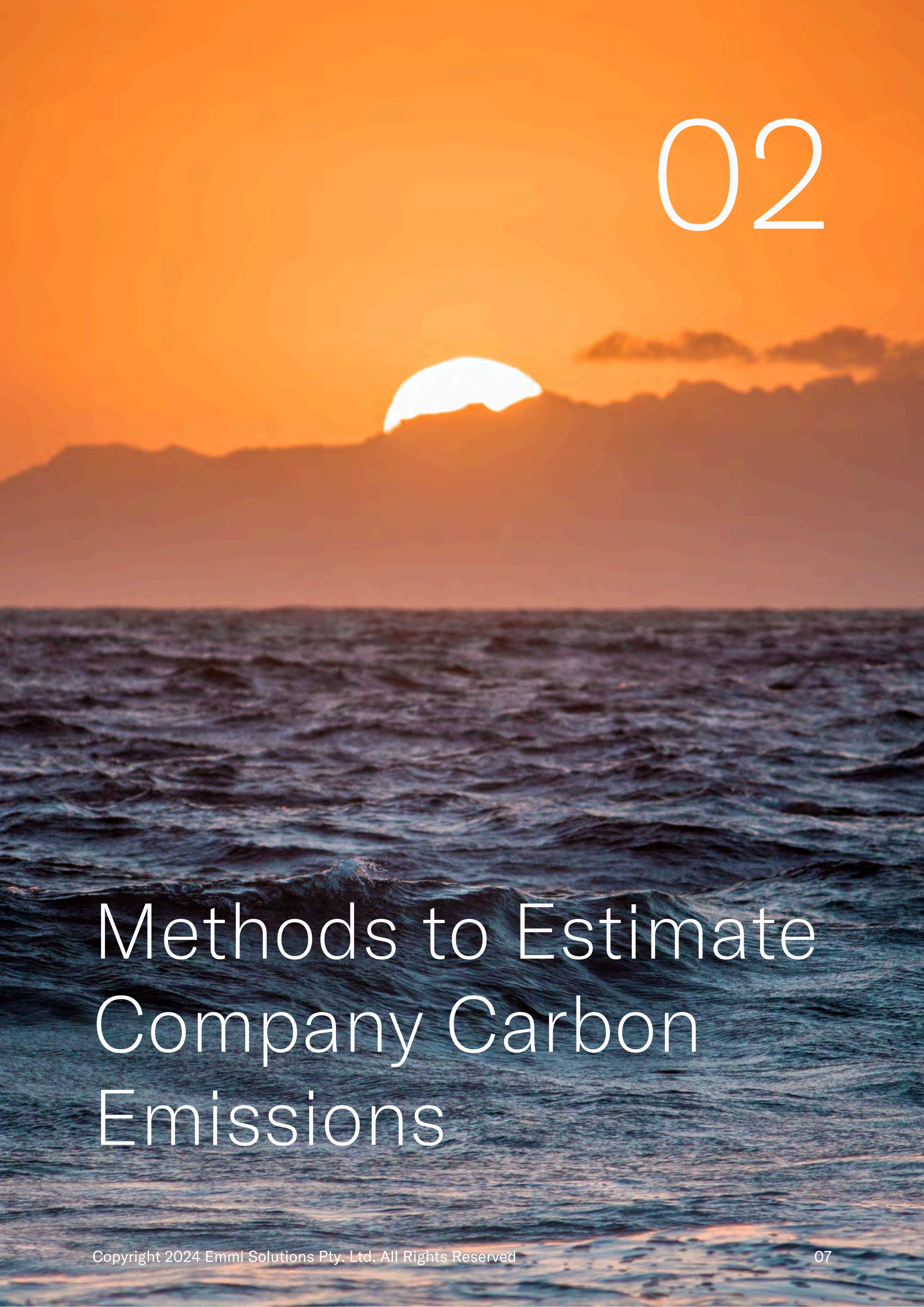
The Carbon Reporting Gap

> ~83% of publicly listed companies don't report their Scope 1, 2 & 3 emissions. ”

According to the Carbon Disclosure Project*, about 8,500 companies disclose all carbon emissions. Given there are about 50,000 publicly listed companies globally, the vast majority (~83%) don't report all of their carbon emissions. Meanwhile, even for companies that do disclose emissions, it's important to recognise reported emissions are typically 12-18 months old.

Timely and deep coverage of emissions for companies does not yet exist, so it's important to develop accurate ways to estimate emissions for companies to fill the large gap for financial institutions, investors and regulators.

*) Carbon Disclosure Project: https://www.cdp.net/en/companies/cdp-2023-disclosure-data-factsheet



02

Methods to Estimate Company Carbon Emissions

Methods to Estimate Company Carbon Emissions

Without reported emissions disclosures, what methods can be used to estimate a companies emissions?

Method	Overview	Source	Strengths	Limitations
Production-Based Factors	Applies emissions intensity factors relative to a unit of production, specific to a given widget or fuel	Company-level production data	• Highly specialised and accurate • Best for deep-dive into single or few companies	• Very limited coverage • Factors updated infrequently • Impractical for investors covering hundreds to thousands of holdings
Emissions Factors using the Reported Emissions Database	Calculates median carbon intensity for individual sector 'peer groups' as defined by sector and region	Reported Emissions Database	• Uses real-world data, not theoretical models • Can give a clear breakdown across Scope 1, 2 & 3 • Granularity can be adjusted to specific sectors or regions	• Relies on accuracy of reported data • Poor data coverage across all sectors • Results are averages • Little understanding of uncertainties
Emission Factors via Environmentally-Extended Input-Output Models	Derives carbon intensities for individual business segments from Environmentally Extended Input-Output (EEIO) tables	EEIO tables (e.g. Exiobase)	• Transparent methodology and easily auditable • Generates nuanced estimates for complex, diversified firms with multinational exposures • Consistent boundary conditions for emissions estimates.	• Theoretical not real-world data with outputs highly dependent on EEIO table selected and quality of segment mapping, leading to large biases • No perspective on uncertainties • EEIO tables are infrequently updated and do not reflect year-on-year trends in industry emissions levels • No Scope 1, 2 or 3 breakdown. • Significant biases in Scope 3
Machine-learning using the Reported Emissions Database	Quantifies relationship between firm attributes (sector, multiple financial variables) and reported carbon intensity	Reported Emissions Database	• Up to 30% more accurate than traditional revenue factors approach • Maximum coverage for investors • Systematic statistical framework with clear uncertainties • Highly flexible, allowing users to include or omit predictive variables	• Relies on accuracy of reported data • Potential for overfitting • Not as accurate for a single company relative to production-based factors

Emission Factors via Environmentally-Extended Input-Output Models

Environmentally-Extended Input-Output Models (EEIO) are economic models built to link financial data to environmental indicators like greenhouse gas emissions from organisations like the US EPA or the EU. Using information on national or regional economies, EEIO produces standard estimates of carbon emissions levels per million dollars for each industry. These models can be used to estimate carbon via financial data in certain industries.

The Exiobase EEIO model is made up of 163 sectors across 44 countries and five "rest of the world" regions. Each combination of sector and country has an associated estimated economic output and total emissions.

For example, 'Agriculture' in the USA may be hypothetically modelled to have $100billion in economic output and 1 million tonnes of CO2-e emissions within the EEIO model. Therefore for 'USA-based Agriculture', the Scope 1 emission intensity factor is therefore 100 tonnes per $million of economic output.

These factors can then be applied at the company-level to estimate emissions. Let's say you're invested in a US-based public agriculture company 'Deer Co.' that generates $10million in revenue each year. Assuming revenue is equivalent to economic output, then using EEIO approach is to simply multiply the emission factor of 100 by 10 (10x per million), resulting in an estimated 1,000 tonnes of carbon emissions for Deer Co. using this approach.

Limitations in the EEIO approach

01

Results are averages

The EEIO approach uses sectoral-level emissions factors which are averages, meaning that all companies in the same sector receive the same emissions intensity factor per million of revenue.

02

No breakdown of Scope 1, 2 or 3 emissions

Since this approach uses an economic model, carbon estimates are an aggregate of Scope 1, 2 and upstream Scope 3 emissions. Regulators and investors require granular breakdowns of each to understand climate transition risk.

03

Large Scope 3 biases

Our research* reveals 'Downstream' categories like 'Use of Product' contribute 60% of Scope 3 emissions, while 'Upstream' categories only contribute around 10%. EEIO factors only consider upstream supply chain Scope 3 emissions and therefore cannot accurately assess climate transition risk. Some data providers try to address this issue with an 'uplift factor', but this introduces further biases.

04

Old data

These emission factors from EEIO tables are built on historical data, which can limit the accuracy of current emissions estimates.

05

Opaque uncertainties

There is no systematic quantification and understanding of emission factor uncertainties or how those uncertainties translate to portfolio-level climate transition risk.

*) Nguyen et al., 2023

Emission Factors using the Reported Emissions Database

The number of companies actively reporting their greenhouse gas emissions data has increased in recent years. This can be attributed to the heightened global awareness, the growing demand for transparency and accountability, as well as the recognition of the potential advantages and liabilities associated with climate change. Additionally, sharing emission details can provide a positive reflection on their reputation, draw in sustainable investors, and abide by increasing government mandates to disclose.

Established in 2000, the Carbon Disclosure Project (CDP) is the primary database where businesses report their greenhouse gas emissions. By 2023, about 8,510 companies disclose their Scope 1 + 2 + 3 emissions. This data-set allows for a direct approach for estimating emissions for public companies.

Instead of using EEIO models to come up with emission factors, this approach uses the reported database of emissions to produce emission factors for each industry/country sub-sector across Scope 1, 2 and 3. For example, Agriculture in the USA emission factors come directly from the reported data-set where available.

> 8,510 companies disclose their Scope 1, 2 & 3 emissions

Uses real-world data
not theoretical models with many different economic assumptions

Clear breakdown
Clear breakdown across Scope 1, 2 & 3

Up to date
these real-world factors are more frequently updated when reported data is disclosed, thereby better reflecting the real-world efficiency gains across sectors.

Limitations in the Emission Factors Approach

01 **Poor data coverage across all sectors**
Many sectors simply don’t have enough reported data for coverage, meaning coverage is sparse.

02 **Results are averages**
The results are averaged out using emission factors meaning that all companies in the same sector receive the same emissions intensity factor per million of revenue.

03 **Little understanding of uncertainties**
Using emission factors makes it difficult to quantify and understand emission uncertainties for each public company.

Machine Learning using the Reported Emissions Database

What is Machine Learning?

Machine learning (ML) in finance has a long history, starting in the 1960s with early use of statistical modelling and decision trees for automation. With advancements in computers, institutions began using ML algorithms to analyze data and improve trading methods. The turn of the 21st century brought even more growth with the emergence of big data and high-frequency trades. Now, institutions rely on ML models for risk assessment and fraud detection, leading to a more proactive approach to managing risks since the 2008 financial crisis.

Today, machine learning is a cornerstone of many elements in the finance industry, from robo-advisors, credit rating programs, market segmentation techniques, and fraud-detection networks. Natural language processing (NLP) makes it possible to analyze comments from financial news reports as well as social media content, helping investors make better choices. Furthermore, predictive data analytics are used widely to predict future stock market trends and improve portfolios.

Machine learning has revolutionised finance, evolving from early decision trees to today's advanced analytics.

Emmi developed nowcasted emissions estimates for all public companies, regardless of whether they have reported carbon emissions.

Introducing Machine Learning to Estimate Carbon Emissions

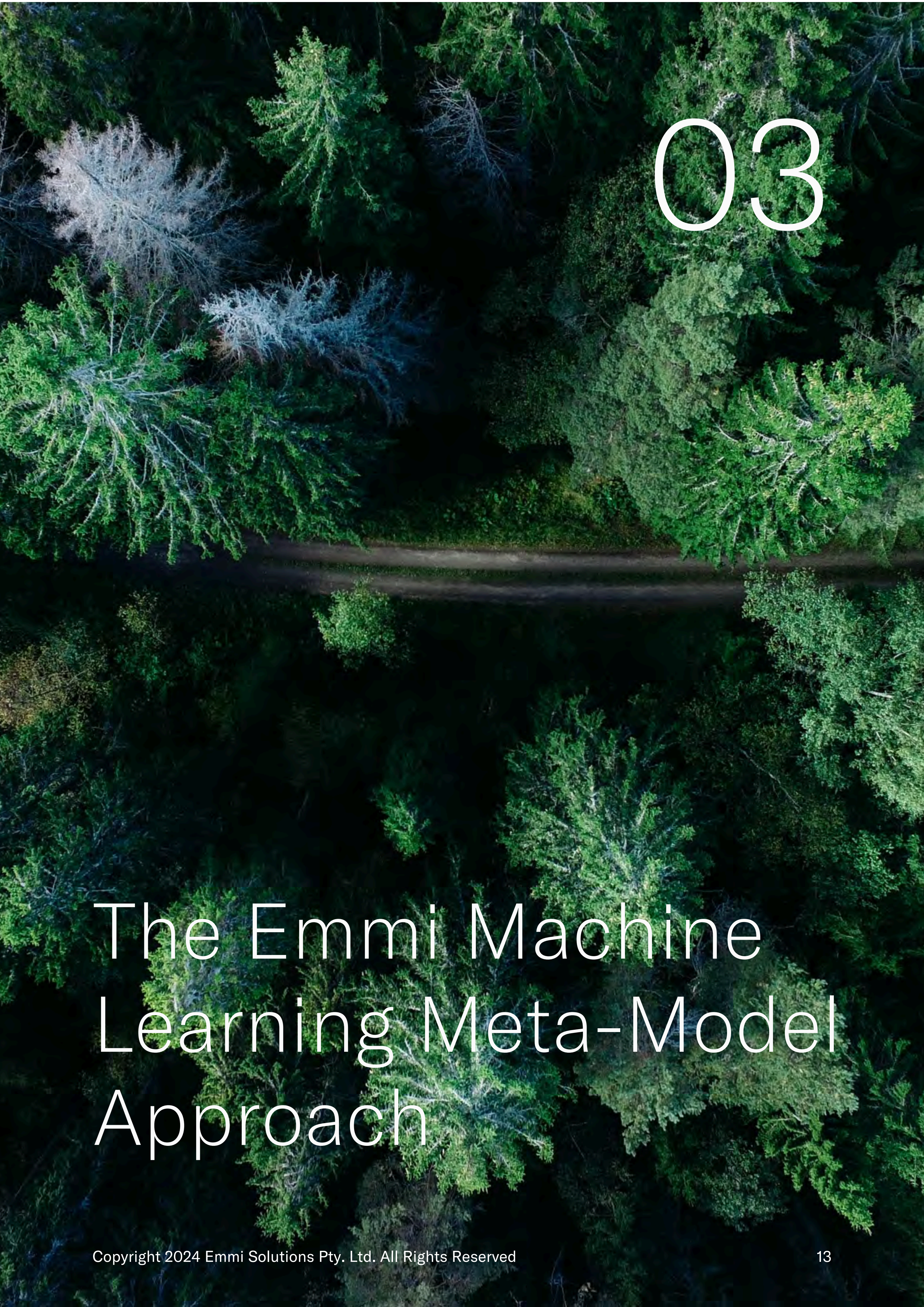
Nguyen et al (2021) was the first to present a new approach using supervised machine learning to estimate corporate greenhouse gas (GHG) emissions using publicly available information. Supervised machine learning involves an algorithm that learns from data with known labels, meaning the input data is matched with the correct output. The objective is to understand and establish a connection between input variables (known as features) and output variables (also known as labels or target values), in order to make accurate predictions on unfamiliar data.

The most successful model applied in Nguyen et al., 2021 was a two-step process that utilised a meta-learner to combine predictions from six different base-learners: OLS, Elastic Net, Neural Network, K-Nearest Neighbours, Random Forest, and Extreme Gradient Boosting. They demonstrated machine learning was up to 30% more accurate than the traditional revenue-based emission factors in predicting total emissions (Scope 1 + Scope 2).

Since 2021, Emmi has partnered with these pioneering climate finance researchers from the University of Otago and Griffith University to extend and apply those machine learning methods to estimate corporate carbon emissions for all publicly-listed companies and public companies.

Nguyen et al (2023) was a follow-up study between Emmi, the University of Otago and Griffith University that researched the data quality and predictability of Scope 3 using machine learning methods. They showed significant under-reporting of Scope 3 reported data while machine learning techniques were found to improve predictions of Scope 3 by up to 25% from linear regression traditional approaches. This research formed the basis to develop and extend prediction capabilities that allowed Emmi to generate ‘nowcasted’ emissions estimates for all public companies - irrespective of whether they have reported carbon emissions.

Since Nguyen et al., 2021, other researchers have used machine learning to estimate company Scope 1 and 2 emissions (Heurtebize et al., 2022; Serafeim and Velez Caicedo, 2022).



03

The Emmi Machine Learning Meta-Model Approach

The Emmi Machine Learning Meta-Model Approach

Our approach extends the work of Quyen et al., (2021 and 2023) into a unique machine learning framework designed for the market around core pillars that include:

Quality

Ensure the approach has least bias in particular with regards to over-fitting when choosing a single machine learning technique

Flexibility

Ensure the highest quality of estimates for companies with limited data inputs

Reproducible

Ensure the methods are reproducible and interpretable for auditing and standards

Coverage

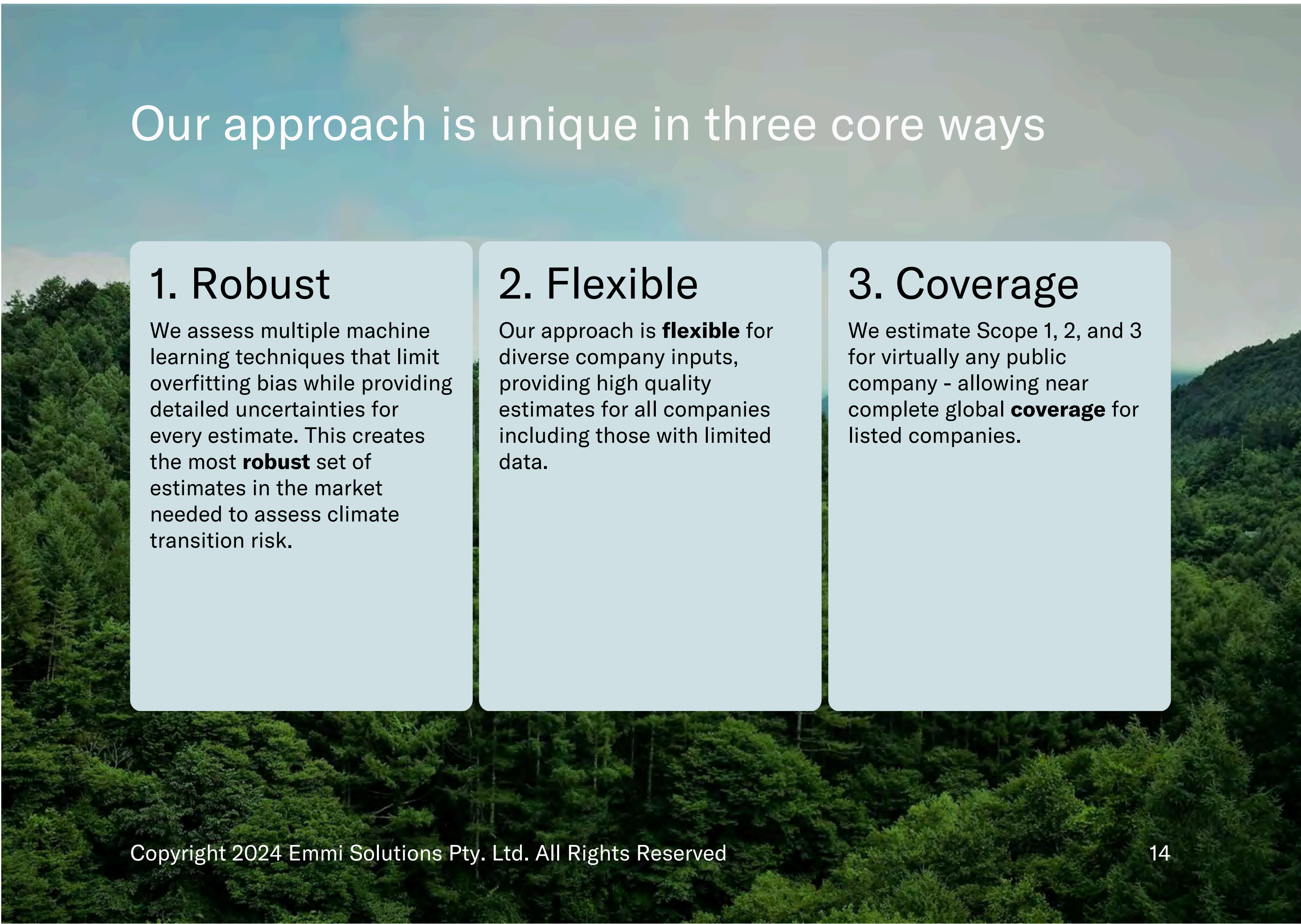
Ensure the largest coverage across public and private markets

Uncertainties

Ensure every estimate contains complete statistical knowledge of uncertainties

Timeliness

Ensure estimates are available as soon as financial results are disclosed



Our approach is unique in three core ways

1. Robust

We assess multiple machine learning techniques that limit overfitting bias while providing detailed uncertainties for every estimate. This creates the most **robust** set of estimates in the market needed to assess climate transition risk.

2. Flexible

Our approach is **flexible** for diverse company inputs, providing high quality estimates for all companies including those with limited data.

3. Coverage

We estimate Scope 1, 2, and 3 for virtually any public company - allowing near complete global **coverage** for listed companies.

Sources of Data

Financial Data

Each year, we gather financial and business information listed in Table 1 for all publicly listed companies which provides insight into the company size, profitability, assets and location of the company. Company data was extracted from 2019 to 2023 using the Factset Company Fundamentals database, totalling to 50,236 companies.

Feature	Information
Revenue	Financial Performance
Country	General Information
Employee Number	Workforce
GICS Sector	Classification
GICS Industry	Classification
Total Assets	Financial Position
Net Property, Plant, and Equipment (NPPE)	Assets
Gross Property, Plant, and Equipment (GPPE)	Assets
Capital Expense	Financial Performance
Gross Profit	Financial Performance
Operational Expense	Financial Performance
EBIT (Earnings Before Interest and Taxes)	Financial Performance
EBITDA (Earnings before Interest, Taxes, Depreciation, and Amortization)	Financial Performance
Net Income	Financial Performance
Total Debt	Financial Position
Current Liability	Financial Position
Current Assets	Financial Position

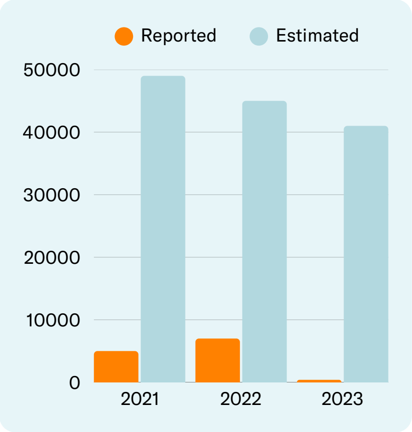
Sources of Data

Carbon and Financial Input Data

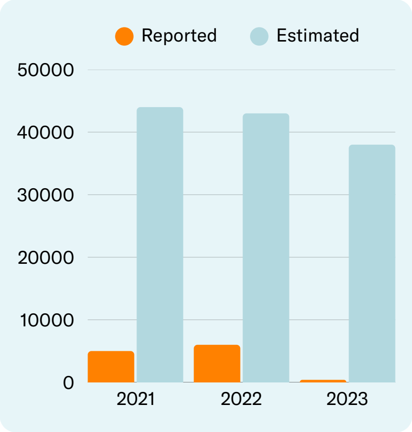
We source reported carbon emissions from over 6,000 companies from Factset ESG solutions for 2021 and 2022. The data was collected from companies' extra-financial communications and annual reports and merged with financial data using Factset.

Scope	Type of Emissions	Number of Companies		
		Training	Testing	Validation
Scope 1	Direct Emissions	3,636	545	1,815
Scope 2	Indirect Emissions (Electricity)	3,515	527	1,903
Scope 3	Other Indirect Emissions Upstream & Downstream	2,269	340	714

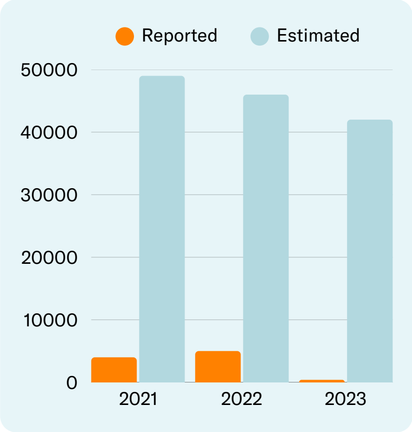
Emissions Coverage by Scope



Scope 1 Emissions

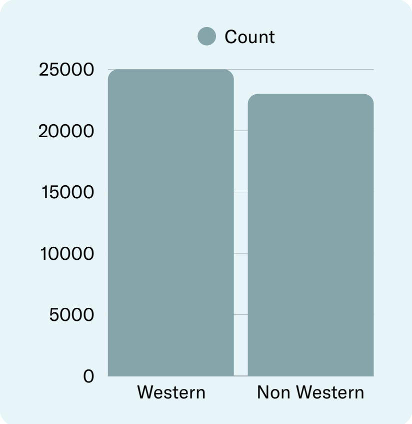


Scope 2 Emissions

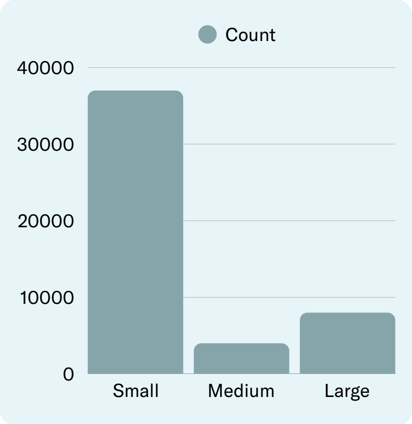


Scope 2 Emissions

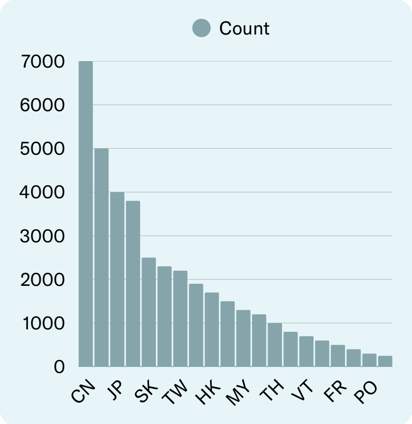
Scope Category Counts



Region Counts



Company Size Counts



Top 20 Country Counts

Data Cleaning & Training

We cleaned and validated reported carbon emissions and financial input data to train our models with the highest quality data. To identify anomalies, we checked year-to-year differences and excluded or cross-checked any data with 100% changes. This resulted in excluding about 5% or 300 companies from the training data-set.

Climate reporting standards are more mature in advanced Western countries*. In order to use the highest quality of reported data for training, we only use data from 25 advanced countries.

> We cleaned and validated data, excluding 5% of companies with anomalous year-over-year changes in reported emissions.

Model Evaluation

Rather than relying on a single machine learning technique, our innovative approach utilizes a diverse range of techniques to ensure improved robustness, stability, and performance.

We developed performance metrics and a comprehensive model evaluation infrastructure to assess thousands of different learners and data combinations commonly used in the market.

We optimised and trained over 30,000 unique models for Scopes 1, 2, and 3 which involved testing a variety of classification models and input financial data (eg revenue, NPPE, total liabilities, etc.), country, sector/industry, and emissions. Additionally, we fine-tuned the hyperparameters of these models to ensure their robustness and employed a variety of machine learning techniques.

The types of machine learning techniques we used included decision trees, gradient boosting, random forest, linear tree and ordinary least squares.

Decision Trees

A decision tree is a hierarchical structure that outlines a sequence of decisions. Each node corresponds to a decision based on a characteristic, and each leaf represents the final predicted outcome. Data is divided into subsets at each node to increase homogeneity, until a stopping condition is reached. The leaves hold the predicted outcomes for their subsets, making the model easy to interpret.

Decision trees are easy to interpret and can handle various types of data, making them scalable for large datasets. They capture nonlinear relationships without complex transformations or feature engineering. However, deep and complex trees may lead to overfitting and sensitivity to training data variations. They also struggle with capturing complex relationships and favor features with more levels in the splitting process.

*) Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Israel, Italy, Japan, Luxembourg, Netherlands, New Zealand, Norway, Poland, Portugal, Singapore, South Korea, Spain, Sweden, Switzerland, United Kingdom, United States

> Our innovative approach combines multiple machine learning techniques to enhance robustness, stability, and overall performance in emissions modeling.

Gradient Boost

Gradient boosting attempts to improve existing models by minimising a loss function that iteratively tests different sets of model input parameters known as hyperparameters, such as how deep a tree should go. It involves combining the predictions of multiple weak learners, typically via decision trees, to create a strong predictive learner model.

Gradient boosting is a popular machine learning method known for its accuracy and resistance to overfitting. But it's important to carefully choose and evaluate models, as hyperparameter tuning can be difficult and computation-heavy.

Random Forest

A random forest is a type of classification algorithm made up of a collection of several decision tree models that work together to make predictions. Each decision tree in the random forest is built from a random sample of the data, and at each step of building the tree, a random subset of the features is considered for splitting the data. This means that each tree in the forest is different and is able to capture different aspects of the data. The final prediction is made by averaging the predictions of all the trees in the forest. This combination of many decision trees results in a model that is more robust and less prone to overfitting than an individual decision tree.

Linear Tree

Linear Trees combine the learning ability of decision trees with the predictive and interpretive power of linear models. Data is split using decision rules, similar to tree-based algorithms, and the quality of the splits is measured by fitting linear models at each node. This means that the models in the leaves of the tree are linear, as opposed to constant predictions as in traditional decision trees.

Ordinary Least Squares (OLS)

OLS regression is commonly used to model the linear relationship between independent and dependent variables. Its pros include ease of use, unbiased estimates, interpretability, widespread application, and flexibility. However, it relies on assumptions and can be sensitive to outliers, multicollinearity, and overfitting, which may limit its predictive power and generalization performance.

Techniques used and differences

Criteria	Decision Trees	Gradient Boosting	Random Forest	Linear Trees	Ordinary Least Squares (OLS)
Model Complexity	Simple to Moderate	High	Moderate to High	Moderate	Simple
Interpretability	High	Low	Moderate	Low to Moderate	High
Training Time	Fast	Slow	Moderate	Moderate	Very Fast
Handling Non-Linearity	Good	Excellent	Excellent	Excellent	Poor
Overfitting Risk	High	Moderate	Low	Moderate	High
Use Case	Easy to interpret models, quick decisions	Complex problems requiring high accuracy	Problems with high variance	Situations needing both non-linearity and interpretability	Simple, linear relationships

Model and Feature Selection

The art of building machine learning models is to find the right balance between enhancing the model's accuracy through model/feature selection, while limiting the chances of overfitting which produces a less reliable and robust model. Overfitting occurs when the model is too well-trained on the training data and begins to simply remember it instead of learning the underlying patterns. Because of this, the model may not function as accurately with unseen data or be generalised to real-world scenarios.

Overfitting can occur when there are too many parameters compared to the amount of available training data, causing specialised fitting and poor generalization for new data. Those over-fit models become overly sensitive to noise or irrelevant features present in the training set, meaning small changes could lead to big anomalies in predictions. In order to optimise for both model and feature selection - it's necessary to develop model performance evaluation metrics that can be used to judge the best models.

> Building effective machine learning models means balancing accuracy with the risk of overfitting to ensure good generalization.

Model Performance Evaluation

We split the reported data where 71% of the data was used in training, 4% in testing and 25% to independently validate the results (hold-out). Our model performance evaluation criteria was split into four categories - Accuracy, Predictability, Feature Selection and Robustness.

Accuracy

We assess accuracy using the absolute percentage error (APE) for a given company (i):

$$\mathrm{APE}_i = \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100$$

Where: y_i is the actual value

$\hat{y}_i$ is the predictad value

Mean Absolute Percentage Error (MAPE), calculated as the mean of APE across all data is often used to quantify accuracy across larger data-sets. However, MAPE can give a misleading representation of the typical errors if the distribution is skewed and sensitive to outliers.

We found MAPE to be heavily biased given the significant outliers within the reported carbon data-set. Instead, we used the Median Absolute Percentage Error (MdAPE) to assess mode accuracy across the reported data-set, giving a more reliable and robust measure of the overall performance of the model across the data-set.

Predictability

To assess how well a given model and feature set explains the variance in the data we use R-squared, which is a statistical measure that represents the proportion of the variance in the dependent variable that is predictable from the independent variable(s). R-squared indicates the goodness of fit of the model. It ranges from 0 to 1, where 0 indicates that the model does not explain any variance in the dependent variable, and 1 indicates that the model perfectly explains the variance.

It provides a straightforward interpretation of model performance and easy comparison between models for performance however assumes linearity between features or provides information on whether the model's features importance or how stable the model is.

> A model's goodness of fit indicates its effectiveness in predicting outcomes based on the provided input variables.

Robustness

Robustness is the ability of a model to maintain accuracy and performance when faced with variations, noise, or unexpected inputs that were not present during training. This is important for several reasons:

Handling Real-World Data Variability

In real-world applications, models will encounter data that differs from the training set. A robust model can generalise well to new, unseen data and maintain performance across a range of inputs.

Prevention of Overfitting

Overfitting occurs when a model learns the noise or specific details of the training data rather than the underlying patterns. A non-robust model may perform well on the training data but fail on new data due to overfitting. Robustness helps in building models that generalise well and perform consistently on different datasets.

Trust & Reliability

Users are more likely to trust and rely on models that are robust because they know that the model will perform well even in less-than-ideal conditions. Additionally, robust models provide more consistent and reliable outputs, which is important, especially for investment decision-making processes.

To assess a model's robustness we introduce a noise injection test.

Noise Injection Test (MdAPE*)

To evaluate a model's performance and avoid overfitting, we add 30% noise to our feature datasets and analyze its ability to handle variations. This allows for sensitivity analysis, generalization testing, and assessment of feature importance.

If the model maintains stable accuracy despite the added noise, it suggests that it can effectively generalize to new datasets. On the other hand, if the added noise decreases accuracy substantially, it likely indicates overfitting to the training data.

Additionally, this metric provides insight into how well a model can generalize beyond its training data by evaluating performance with added noise. Features that maintain their significance even with added noise are considered robust and have consistent predictive power.

> Stable accuracy under noise indicates resilience, while significant sensitivity suggest overfitting and highlight feature importance.

Feature Importance

Feature selection is crucial in machine learning for better performance, less overfitting, faster training, and easier deployment. We tested 64 different financial feature sets to predict emissions for listed companies. Machine learning models can be powerful but are often considered 'black-boxes' with unclear important features. We used metrics that offer transparency and understanding of features for each model.

Shapley Values

Originally used in game theory, Shapley values (SHapley Additive exPlanations, or SHAP) have been adapted for machine learning to explain the impact of features on a model's predictions. They provide a clear understanding of feature importance and increase transparency by breaking down predictions into individual contributions from features. This promotes trust and understanding of complex models.

The SHAP value measures the average impact of a feature in all potential combinations of features but is computationally expensive since it essentially varies a feature while keeping all others constant in order to determine the impact on the model. Although feature importance is not included in our model performance, it offers important insights into the interpretability of each model.

Gini Coefficient on SHAP Values

The Gini coefficient, a statistical tool commonly used in economics, measures the distribution of wealth and inequality. When applied to SHAP values within a model, it reveals any disparities in feature importance - indicating whether a small subset of features hold the majority of predictive power, or if it is more evenly spread across numerous features. This provides valuable insight into the significance of each feature on the overall accuracy and reliability of the model.

A SHAP value Gini coefficient of 0 represents complete balance, where all features make an equal contribution to the model's predictions. A coefficient of 1 signifies significant inequality, with one or a few features carrying most of the weight in determining the predictions while others have minimal or no impact.

In general, a SHAP value Gini coefficient below 0.5 is considered ideal for models, as it shows a balanced distribution of feature importance, while 0.5-0.7 is considered acceptable. If the Gini coefficient exceeds 0.7, the model may need re-tuning to ensure that it is not overly dependent on only a handful of features unless this reliance is unavoidable to the application of the model..

> The Gini coefficient measures inequality and can be calculated from SHAP values of feature importance within each model.

Meta-Model Approach

By utilising model performance metrics that measure accuracy, robustness, and predictability (as outlined above), we carefully selected a collection of models (known as Meta-Model A and Meta-Model B) based on data availability inputs in the market.

The two set of meta-models utilise historical emissions where reported emissions data is available for the previous year, and financial-only meta-models, where only financial data is available.

Emmi Meta-Model A
Historical Emissions Models

For publicly listed companies that report their emissions within FactSet, we use a historical emissions meta-model. This meta-model uses historical financial and emissions data as one of the major predictors for the current-year emissions, and has been shown to be the most accurate type of emissions model (Nguyen et al., 2021). In general, the historical emissions meta-models have a median accuracy of ~20% for all Scopes 1-3 and 5-15% for those most carbon-intensive sectors. However the coverage of this model is about 7% since it relies on companies reporting historical emissions.

Emmi Meta-Model B
Financial Only Models

For public companies in the FactSet Universe without reported emissions, we use financial-only models. These can (but aren’t limited to) include models using last year's historical financial information to understand the change over time. Financial-only models have an accuracy of 50-70% depending on the Scope, with Scope 3 typically being the worst (primarily due to systematic under-reporting of material Scope 3 categories). Since this meta-model relies on financial information only - coverage in the market is near 100%.

> Our Meta-Model approach combines multiple models for more robust predictions.

Assessing the model performance metrics across over 30,000 different combinations of models for each scope, we select up to 10 top-performing models, which we refer to as our ‘meta-model’ approach. This ‘ensemble’ approach has long been used by climate researchers to make better, more robust predictions for climate scenarios in the IPCC scientific process.

By using the median of several best estimations, the meta-model reduces the impact of outliers and the ability of any single individual model to bias the outcome. This approach is more reliable and robust than relying on a single model, as it can help reduce overfitting and increase the generalisation and performance of the model. Additionally, this meta-model approach helps to capture different aspects of the data, leading to a more comprehensive understanding of the underlying patterns in the data.

Results - Scope 1

Reported vs Estimated

One way to assess the model performance is to compare reported emissions versus model predictions for a given year. Although reported emissions themselves have inherent biases and uncertainties that are unknown - systematic biases can often be shown when comparing the differences between reported and estimated.

Scope 1

To help illustrate reported vs predicted comparisons, we plot all available data for the year 2022 along with the 100 companies with the highest reported emissions within Western countries. For all available reported data, our estimates show a high degree of predictability with a Median Absolute Percentage Error (MdAPE) of 13% for 3,493 companies who report Scope 1 emissions from Western nations.

The highest 100 emitters of Scope 1 in Western countries cumulatively emit 3.2 billion tonnes of greenhouse gas emissions, which is 27% of the entire industrial footprint. Of those top 100, our approach has an MdAPE of 10.7%.

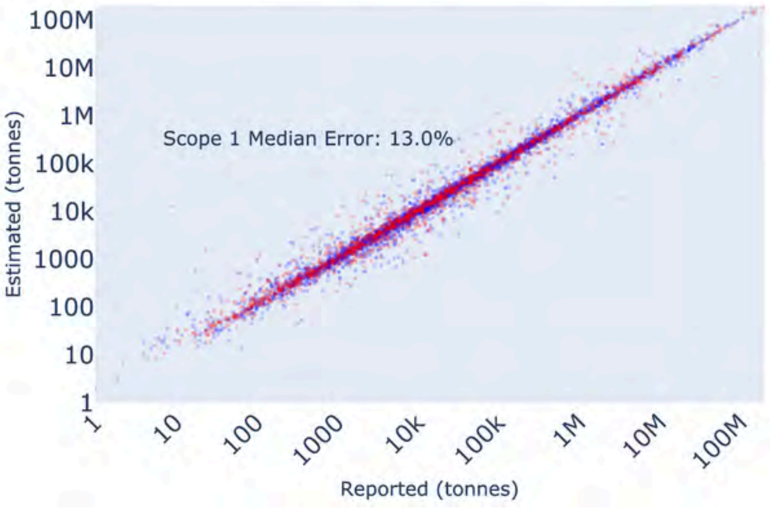


Chart: Reported emissions (tonnes) for Western & Non-Western Regions

Legends:
Western
Non-Western

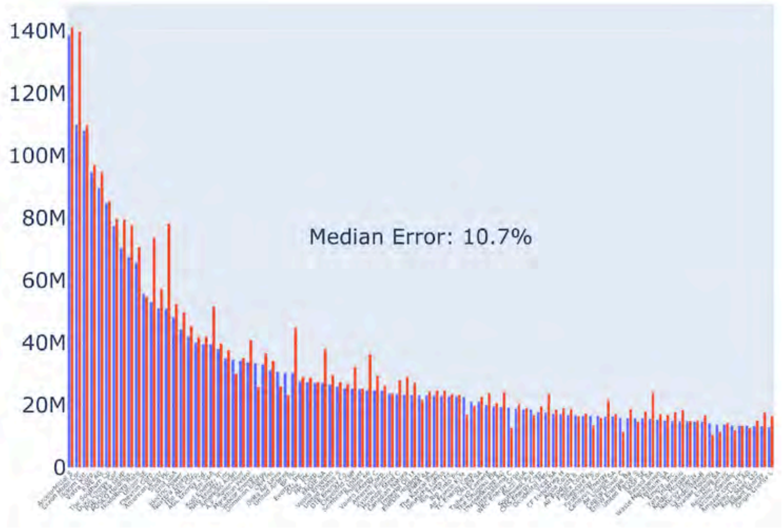


Chart: Scope 1 Top 100 Western Companies (Reported & Estimated)

Legends:
Western
Non-Western

Results - Scope 2

Scope 2

Scope 2 emissions are largely indirect emissions from electricity needs for companies. They are more variable than Scope 1 emissions, dependent on whether a company chooses to report equity-share vs operational control or location-based or market-based emissions.

Companies can quickly reduce their Scope 2 emissions by switching to renewable electricity. This change has been seen in many companies recently, while others have yet to make the shift. These step-changes in Scope 2 are common for companies. However, models trained on historical data cannot pick up this company Scope 2 'event' easily, since it has nothing to do with companies fundamentals - it is simply a reporting decision.

These events across the market make model prediction more difficult for Scope 2, which is reflected in our reported vs estimated comparisons.

For all available reported Scope 2 data, our MdAPE for our estimates is 18.7% for 3,398 companies who report Scope 2 emissions from Western nations.

The highest 100 emitters of Scope 2 in Western countries cumulatively emit 2.1 billion tonnes of greenhouse gas emissions, which is ~20% of the entire industrial footprint of Scope 2. Of those top 100, our approach has an MdAPE of 12.8%.

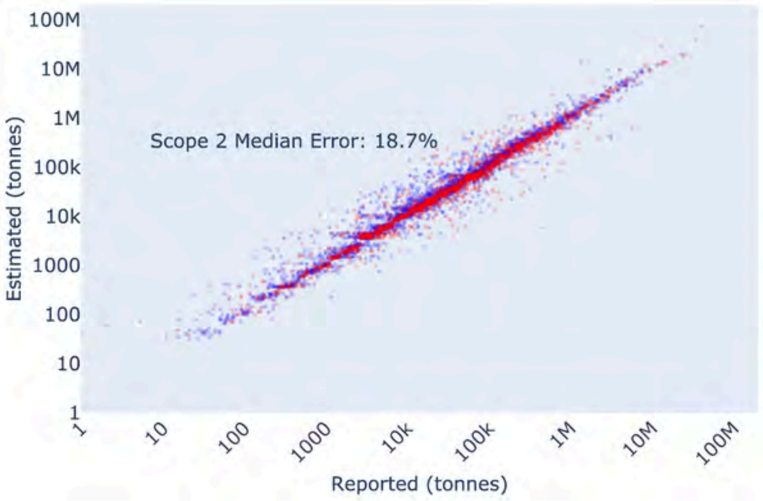


Chart: Reported emissions (tonnes) for Western & Non-Western Regions

Legends:
Western
Non-Western

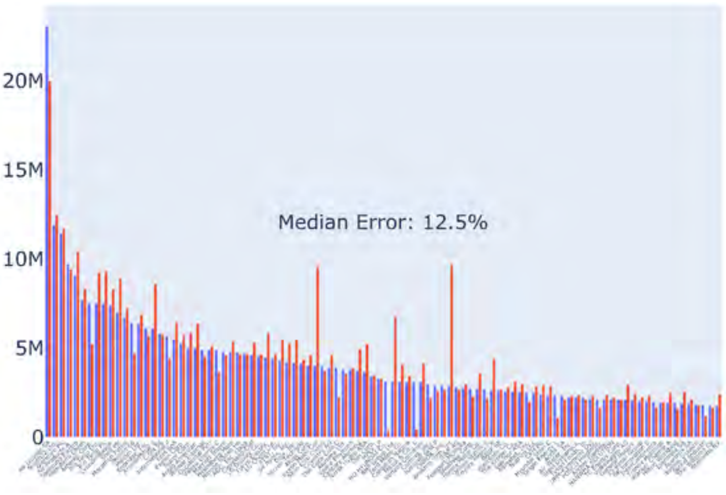


Chart: Scope 2 Top 100 Western Companies

Legends:
Western
Non-Western

Results - Scope 3

The upstream and downstream emissions for a company (Scope 3) are difficult to quantify due to complex supply chains and diverse use of sold products. When companies do report Scope 3 emissions, they are often under-reported across the more material 15 categories.

Banks & Financial Institutions

Banks and financial institutions in particular, have traditionally under-reported their Scope 3 emissions significantly by not disclosing their loans and investments footprints for Scope 3. Since our models are trained off the under-reported emissions data for banks, this would significantly bias the Scope 3 estimates in the financial sector.

Therefore for banks and financial institutions we do not utilise the machine learning approach. We take a different approach for Scope 3 estimates where we utilise global institutions who do report good coverage across their loans and investments with total assets to benchmark other banks. All the data points where Scope 3 estimates are considerably higher than reported in the chart below are due to this approach for banks and financial institutions.

For all available reported Scope 3 data, our MdAPE for our estimates are 19.2% for 1,993 companies who report Scope 3 emissions from Western nations.

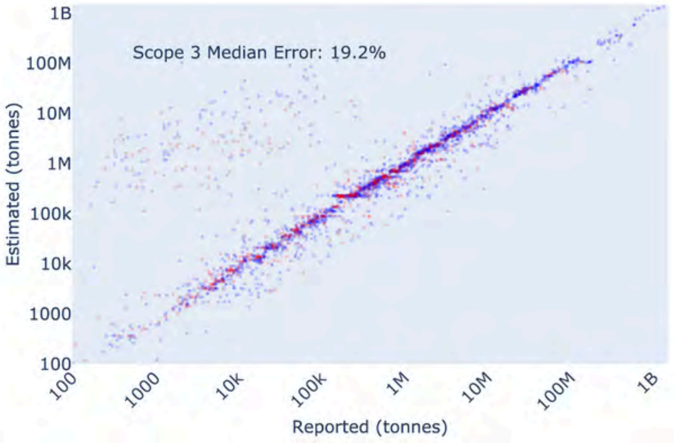


Chart: Reported emissions (tonnes) for Western & Non-Western regions

Legends:
Western
Non-Western

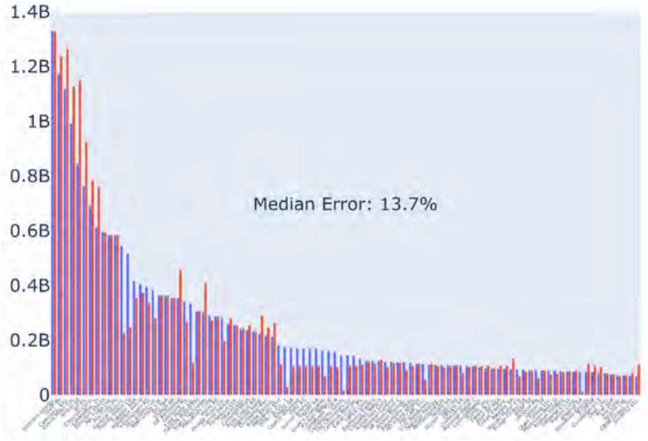


Chart: Scope 3 Top 100 Western Companies

Legends:
Western
Non-Western

The highest 100 emitters of Scope 3 in Western countries are very different to the highest Scope 1 emitters - since ‘Use of Sold Products’ dominates the Scope 3 footprints for companies. Cumulatively the top 100 Scope 3 emitters in the West are responsible for 23.1 billion tonnes of greenhouse gas emissions, which is ~34% of the entire industrial footprint of Scope 3. Of those top 100, our approach has an MdAPE of 13.7%.

Assessing Model Performance and Uncertainties

The median absolute percentage errors (MdAPE) was used to assess accuracy between predicted and actual values, as described earlier. Our models are applied at the company level so MdAPE assesses uncertainties in the context of predicting any individual company's carbon emission footprint.

Meta-Model Performance for Companies

Emissions Type	Meta-Model	MdAE (tonnes)	MdAPE (%)	MdAPE* (%)	R2	Gini Coeff of SHAP
Scope 1	A	3,546	15.6	20.5	0.97	0.65
Scope 1	B	15,863	65.5	66.2	0.32	0.45
Scope 2	A	5,240	16.1	24.7	0.59	0.88
Scope 2	B	15,281	53.9	56.7	0.51	0.48
Scope 3	A	60,609	17.1	25.3	0.93	0.75
Scope 3	B	355,143	70.8	79.3	0.71	0.33

Robustness and Feature Selection

The random noise test was important to assess model stability and robustness (MdAPE*). This performance metric evaluates a model's ability to handle variations in data, and detect potential overfitting issues and sensitivity analysis to give an assessment of generalisation abilities and feature importance. If the model maintains stable accuracy despite the added noise, it suggests that it is robust and can effectively generalise patterns in new datasets, without the risk of over-fitting.

We found some models, in particular random forest, where MdAPE increased 100% when adding 30% noise to the hold-out data. Random forest models were the least robust, subject to significant over-fitting and not selected for our best Meta-Models. On the other hand, linear forest and gradient boost models were found to be the most robust and optimal for performance. For most of the final set of financial only models (Meta-Model B), the MdAPE varied by 2-12% when adding 30% noise (see Table).

For Meta-Model A, MdAPE increased 31-44% when adding 30% noise. This pattern is likely due to lower Gini Coefficients of SHAP for Meta-Model B, which implies a better spread of predictors than for Meta-Model A, where historical emissions is a dominant feature.

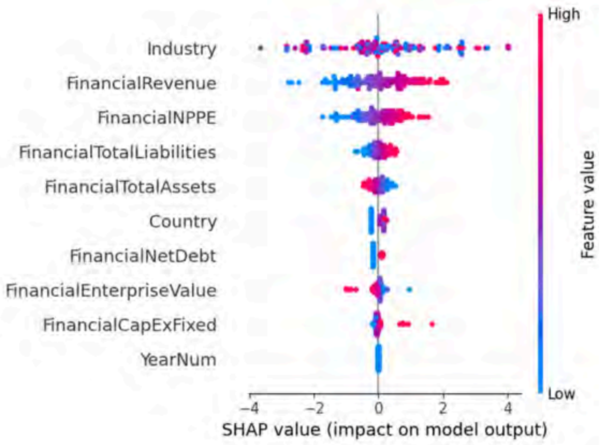


Chart: Example of SHAP value output for a model - predictors and colours indicate feature importance and values.

How does region and company size impact the accuracy?

Where a company is located and its size is likely to have an impact on the accuracy of both reported carbon emissions and the predicted estimates that train off the reported data-set.

Region

Countries have diverse regulations, data availability, and industry composition, all of which affect emissions reporting accuracy. Western nations with longer and more strict climate mandates should lead to more precise reporting, while non-Western with less oversight may have more variability. Better infrastructure and access to accurate data also allow for more accurate measurement tools.

Size

Big corporations have more resources to accurately track emissions, while smaller businesses may rely on simpler methods. Government regulations and available tools can also impact the accuracy of reporting for both types of companies.

To investigate, we looked at the prediction error for Scope 1 as a function of both company size and region, as defined by Western vs Non-Western companies. For company size, we sliced the universe into small cap (market cap <$2bil), mid-cap (market cap $2-10bil) and large cap (market cap >$10bil) cohorts.

In general larger companies located in Western countries had the lowest errors. This trend was replicated for Scope 2 and Scope 3.

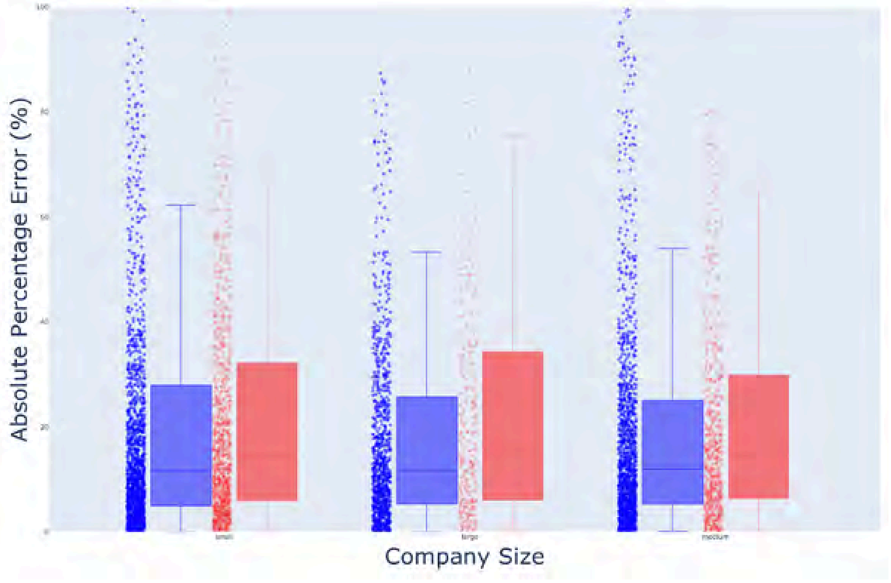


Chart: Absolute Percentage Error (%) by Company Size

Legends:
Western
Non-Western

Meta-Model Performance for Portfolios

For investors, MdAPE is not the best way to understand uncertainties since it treats all portfolio companies or holdings equally, regardless of their size or importance. In reality financial portfolios have diverse individual companies and assets that have significantly higher weights or materiality of risk, leading to an inaccurate assessment of the overall portfolio's uncertainty when using MdAPE.

Weighted Median Absolute Percentage Error

Weighted Uncertainty Formula
The total weighted uncertainty (U_{total}) can be calculated using this formula

$$U_{total} = \sqrt{\left(\frac{U_1 \times E_1}{E_{total}}\right)^2 + \left(\frac{U_2 \times E_2}{E_{total}}\right)^2 + \cdots}$$

Where: U_1 is the uncertainty for each emission value

E_1 is each emission value

E_{total} is the sum of all emissions

Weighted Median Absolute Percentage Error (WMAPE) is a more accurate reflection of portfolio uncertainties as it offers advantages in a variety of ways:

Reflects the Materiality of Different Emission Sources

In a portfolio, carbon footprints are highly skewed and vary in magnitude/materiality. For example, transport companies contribute more to overall portfolio emissions than technology companies. WMAPE provides a more accurate representation by giving greater relative weights to reflect the actual diversity of footprints across the portfolio.

Minimizing Bias from Small, Less Significant Sources

Portfolios tend to contain greater volume of companies that have low emission profiles yet higher uncertainties. High uncertainties in these small companies can disproportionately affect the overall portfolio error metric when using MdAPE. By assigning lower relative weights to smaller emission companies, WMAPE prevents them from distorting the overall picture.

Prioritizing Opportunities for Significant Reductions

WMAPE helps identify the most impactful areas for reducing carbon footprints and risk. Instead of focusing on minor sources, WMAPE allows investors to understand where errors in estimates are most material, directing better understanding toward the most significant opportunities for reducing risk.

Listed Universe Uncertainties

In order to better assess the weighted uncertainties of our approach for portfolios we aggregate and slice the listed universe into small cap (market cap <$2bil), mid-cap (market cap $2-10bil) and large cap (market cap >$10bil) cohorts.

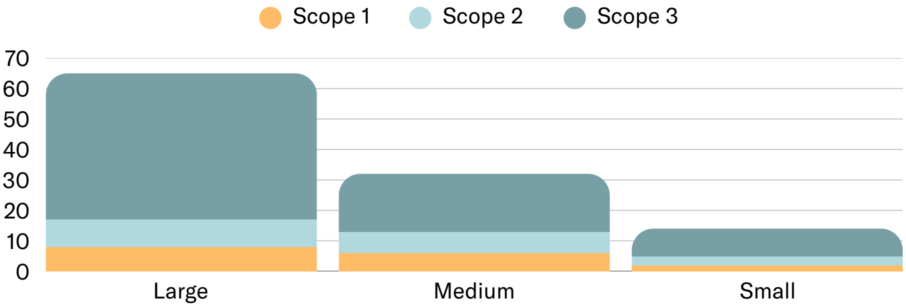


Chart: Cumulative Emissions Footprint (Billion Tonnes) of Listed Universe by Company Size

For all data in 2022, Scope 1 and Scope 2 are found to have a 21% median weighted error, while 32% for Scope 3. These uncertainties are lower for large caps (18-27%), while much higher for small caps (32-56%) - see table.

Company Size	Scope	Footprint (tonnes)	Weighted Median Error (%)
All Data	**Scope 1**	**11,854,692,675**	**21.5**
All Data	**Scope 2**	**2,167,672,491**	**21.3**
All Data	**Scope 3**	**59,968,274,491**	**32.1**
Small Caps	Scope 1	1,580,564,762	35.6
Small Caps	Scope 2	360,995,686	31.9
Small Caps	Scope 3	6,663,333,953	56.1
Mid Caps	Scope 1	4,113,391,534	20.5
Mid Caps	Scope 2	644,194,639	20.7
Mid Caps	Scope 3	12,618,149,875	36.3
Large Caps	Scope 1	6,160,736,379	18.5
Large Caps	Scope 2	1,162,482,166	18.3
Large Caps	Scope 3	40,686,790,662	26.8

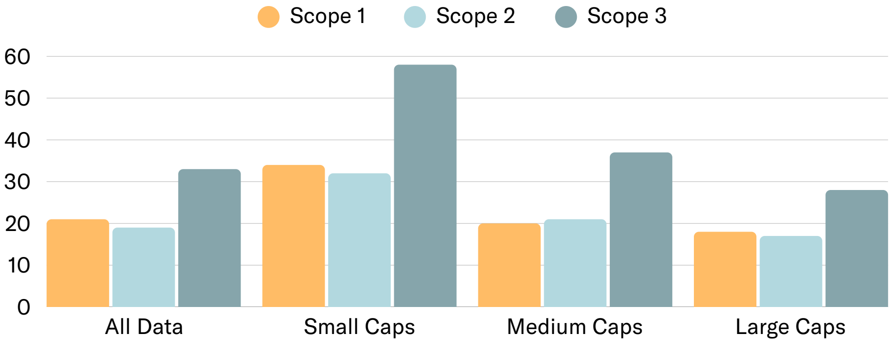


Chart: Overall Median Weighted Error (%) by Company Size and Scope

Global Index Funds

To help interpret the accuracy of our approach for investors, we have calculated the reported versus estimated errors for some of the major index funds. Median weighted errors were found to be ~15% for Scope 1, ~17% for Scope 2 and ~19% for Scope 3.

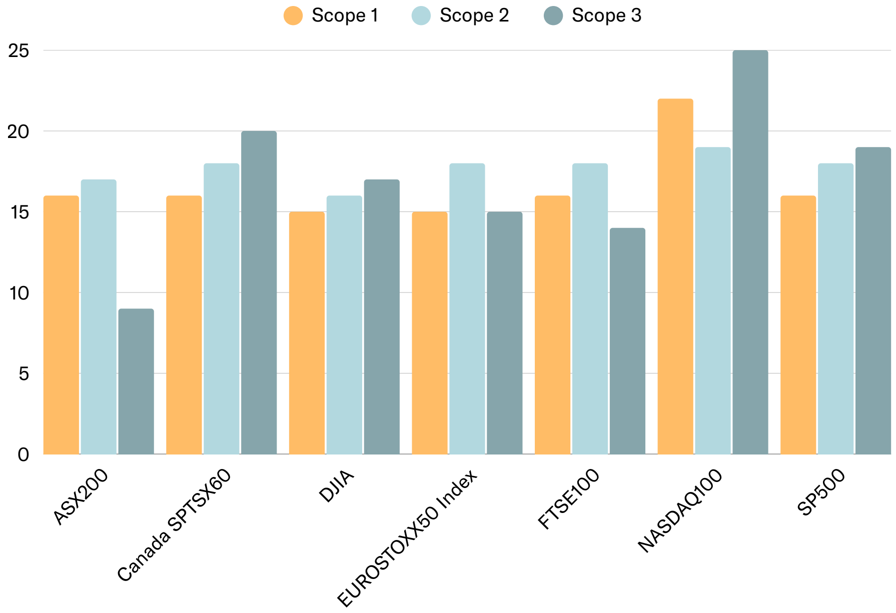


Chart: Overall Median Weighted Error (%) by Index and Scope

Most Carbon-Intensive Sectors

When focussing on the most material emitting sectors for Scope 1, the models perform well, with MdAPE ranging from 6-20%, while the carbon intensive sectors within 5-10%. Similar results are found for Scope 2 and 3.

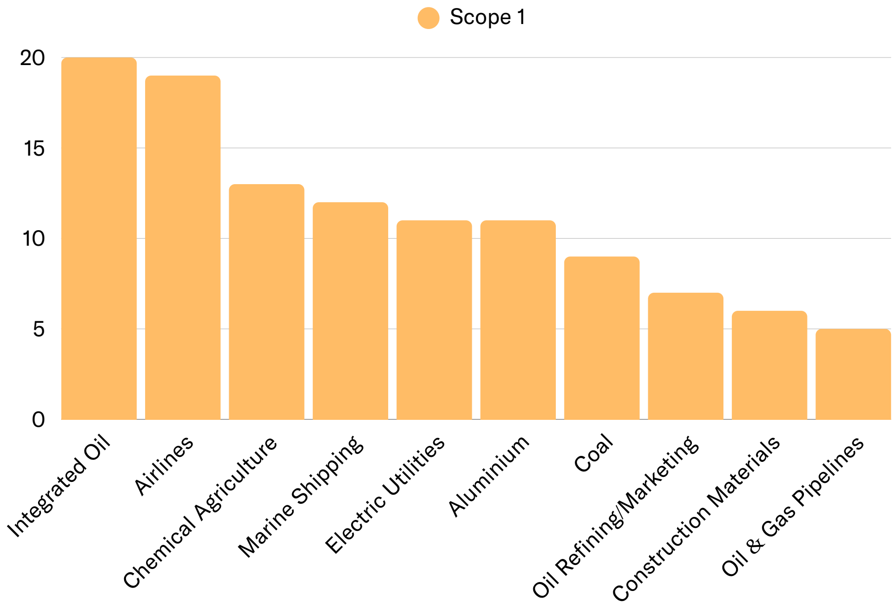


Chart: Scope 1 - MdAPE by Sector for Western Nations Showing Material Scope 1 Sectors

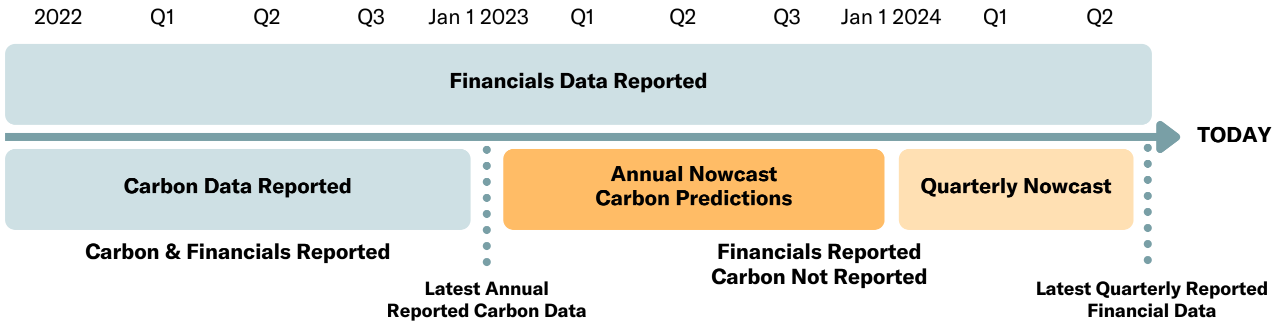
Emmi Nowcasted Emissions Estimates for 2023

What is Nowcasting?

In economics, nowcasting is the process of forecasting the current or recently passed state of an economic indicator. It is commonly used in financial markets to predict upcoming GDP releases or quarterly financial data for businesses. This is typically done by analysing past financial information and applying it to forecasts.

Why ‘Nowcast’ for Carbon?

Public companies often report their financials and carbon emissions with a large mismatch. The delay in carbon reporting, typically >12-18 months, makes it difficult for investors to understand and act on their investments.


2022
Q1
Q2
Q3
Jan 1 2023
Q1
Q2
Q3
Jan 1 2024
Q1
Q2
Financials Data Reported
TODAY
Carbon Data Reported
Annual Nowcast Carbon Predictions
Quarterly Nowcast
Carbon & Financials Reported
Latest Annual Reported Carbon Data
Financials Reported Carbon Not Reported
Latest Quarterly Reported Financial Data

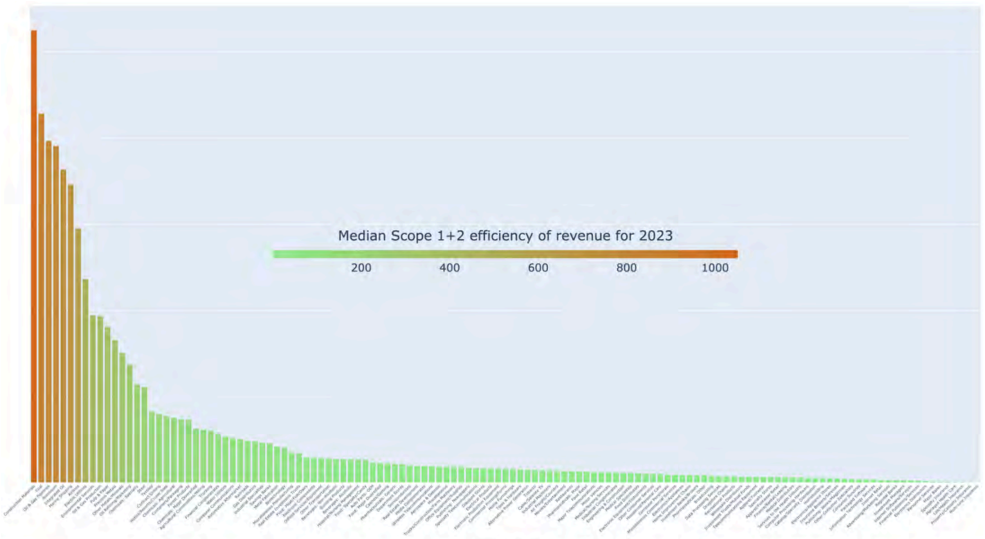
> At Emmi, our approach allows us to predict carbon emissions when financial data is reported, fixing the mismatch between financial reporting and carbon reporting.

Overall, up to 45,000 companies' emissions can be nowcasted as long as they have reported financials. The benefit of our approach is that investors can get a timely estimate with known uncertainties for emissions up to the latest financial disclosures.

u

At the time of writing this research paper, 2023 reported emissions had not been disclosed while financial disclosures were available for 2023. Once reported emissions are disclosed, the models are retrained to take into account the latest reported emissions across the market.

Globally we estimate the listed corporate footprint to be 11.4 billion tonnes in 2023, a decrease by 2.5% from 2022.



Summary

01

45,000 Companies

Emmi's 'Meta-Model' emission estimates cover Scope 1, 2 and 3 greenhouse gas emissions for up to 45,000 companies

02

Award-Winning Research

The Meta-Model approach is an extension of our award-winning university research collaboration, making it even more comprehensive and robust to other approaches in the market

03

Nowcasted Estimates

Nowcasted estimates are available for the latest quarter and financial year when data is available via Factset

04

Portfolio Uncertainties

For a typical investor, we demonstrate portfolio-level uncertainty of those estimates at

- ~15% for Scope 1
- ~17% for Scope 2 and
- ~19% for Scope 3

05

Financial Data Input

Our approach can be used on limited financial data input with uncertainties reflected and disclosed for reporting purposes

06

Factset / APIs

Emmi offers these products through Factset or APIs. For more information on how we can help provide climate data for your reporting and investment management

About Emmi

Emmi provides financed emissions data and climate risk analysis across all major public and private asset classes. These support climate-related reporting, and analysis that feeds into investment management processes.

We use proprietary machine-learning models and algorithms to do this. Our tools translate emissions into financial implications, based on climate and pricing scenarios. This gives our clients actionable insights about their carbon exposure.

This diagnostics ‘toolkit’ is backed by our team of climate and finance experts.

Emmi believes that a low carbon economy is possible, and that properly incentivising and mobilising capital is the fastest and most cost-effective way to reach Net Zero and beyond.

Incorporating the cost of carbon into every decision will enable the finance sector, and its customers, to efficiently allocate resources towards this goal, which will accelerate decarbonisation.

To achieve this, and to meet regulatory requirements, there is a need for a broad spectrum of quality carbon emissions data and climate risk analysis. We have built Emmi to solve that problem.

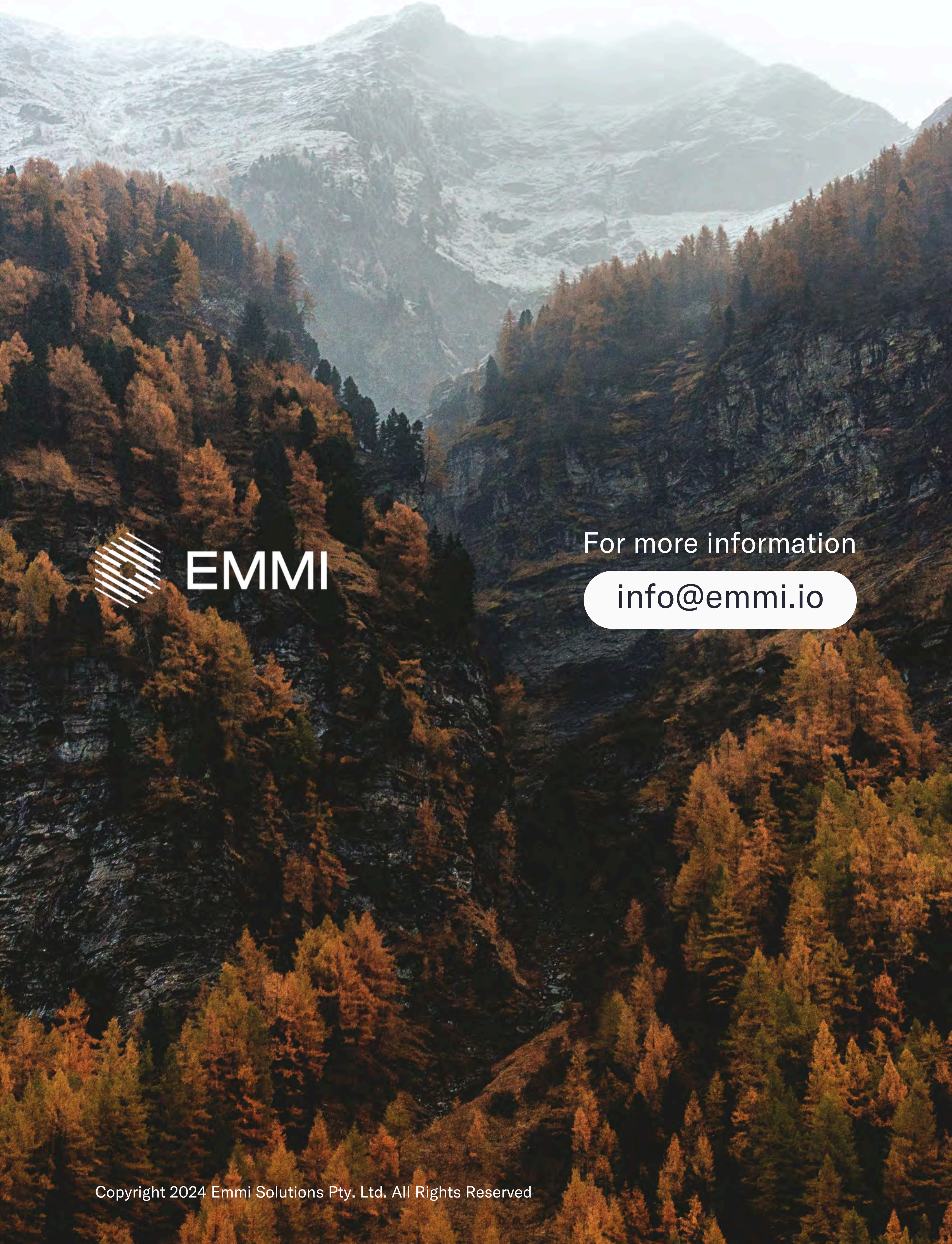
References

Heurtebize, Thibaut & Chen, Frederic & Soupe, Francois & Leote de Carvalho, Raul. (2022). Corporate Carbon Footprint: A Machine Learning Predictive Model for Unreported Data. Available at SSRN.

Nguyen Q, Diaz-Rainey I, Kuruppuarachchi D, (2021) Predicting corporate carbon footprints for climate finance risk analyses: A machine learning approach, Energy Economics, Volume 95, 105129.

Nguyen Q, Diaz-Rainey I, Kitto A, McNeil BI, Pittman NA, Zhang R, (2023) Scope 3 emissions: Data quality and machine learning prediction accuracy. PLOS Clim 2(11): e0000208. https://doi.org/10.1371/journal.pclm.0000208

Serafeim G, Velez Caicedo G. (2022) Machine Learning Models for Prediction of Scope 3 Carbon Emissions. Available at SSRN.



EMMI

For more information

info@emmi.io