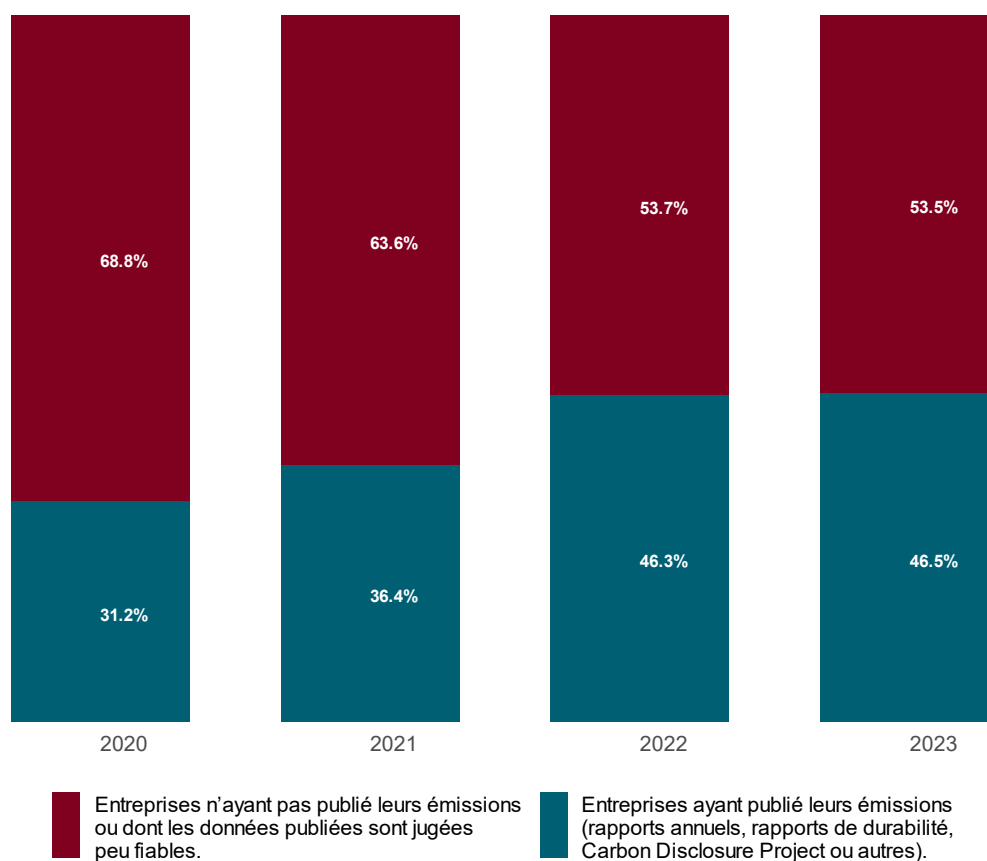


Estimation par intelligence artificielle des émissions carbone des entreprises

Par Maxime Barthe, Thomas Choquet, [Tristan Jourde](#)

Lorsqu'aucune donnée d'émissions carbone n'est publiée par les entreprises, cette information peut être estimée à l'aide de modèles d'apprentissage automatique, dont les performances prédictives surpassent celles des méthodes classiques. Complétés par une expertise humaine, ces modèles permettent de combler les lacunes en matière de données et d'affiner l'évaluation des risques de transition.



Graphique 1 :
Part des entreprises qui publient leurs émissions carbone

Source : ISS Stoxx, calculs Banque de France

Note : échantillon d'entreprises couvertes par ISS, toutes zones géographiques confondues.

Pour les banques centrales, les données relatives aux émissions carbone des entreprises sont devenues essentielles à l'exercice de leurs missions, c'est à dire la politique

monétaire, la stabilité financière et les services à l'économie et à la société. Ces données sont nécessaires à l'évaluation de l'exposition du système financier et de l'économie au risque de transition, qui émerge des nouvelles politiques climatiques, des ruptures technologiques ou des changements de préférences des consommateurs. Elles sont aussi requises pour piloter le verdissement des portefeuilles financiers, y compris ceux des banques centrales. Par exemple, la Banque de France a récemment annoncé l'inclusion de critères environnementaux, intégrant notamment les données d'émissions des entreprises, dans son cadre opérationnel de politique monétaire, afin de se protéger d'une « éventuelle diminution de la valeur des garanties en cas de chocs de transition défavorables liés au climat » (voir [Communiqué de presse du 29 Juillet 2025](#)). Elle a aussi adopté une stratégie d'investissement responsable pour ses portefeuilles en compte propre (voir le [Rapport sur l'action climat de la Banque de France et de l'ACPR, 2025](#)).

La couverture des données sur les émissions carbone des entreprises s'améliore mais reste incomplète

Malgré leur importance croissante, les données disponibles sur les émissions carbone des entreprises demeurent partielles ([Grisey, 2022](#)). Si leur publication a progressé entre 2020 et 2023, cette tendance reste insuffisante, incertaine et inégalement répartie. Seules 47% des entreprises ont publié leurs émissions en 2023, sur un échantillon de plusieurs dizaines de milliers d'entités (*Graphique 1*). Ce chiffre est d'autant plus partiel qu'il exclut les petites structures, majoritairement non soumises à des obligations de publication. Avec les initiatives européennes de simplification réglementaire, telle que la directive Omnibus, de nombreuses entreprises européennes pourraient continuer à ne pas publier leurs émissions.

Dans ce contexte, et pour servir les cas d'usage cités plus haut, une estimation des émissions carbone est donc indispensable pour mesurer les risques de transition. Cela peut passer par une analyse des entreprises à dire d'expert, ou par le recours à des fournisseurs de données commerciaux. Bien que les évaluations d'experts offrent des résultats généralement fiables, elles sont coûteuses en ressources et en temps. Par ailleurs, les modèles utilisés par les prestataires commerciaux souffrent souvent d'un manque de transparence, tant sur la méthodologie que sur leur performance prédictive.

Ce billet vise dès lors à proposer une voie pour combler le déficit d'information autour des entreprises qui ne publient pas leurs émissions. À cette fin, un ensemble de méthodes d'apprentissage automatique (*Machine Learning*) est mobilisé, en vue d'estimer ces émissions de manière automatisée, transparente et à grande échelle. Une comparaison de la performance prédictive de ces algorithmes est menée, par rapport aux estimations produites par des méthodes économétriques classiques. Cette comparaison permet d'identifier le modèle le plus approprié pour évaluer les émissions non publiées par les entreprises, en termes de précision et de robustesse. Grâce à leur flexibilité, les modèles d'apprentissage automatique peuvent analyser et intégrer des relations plus complexes entre les variables prédictives et les émissions carbone que les modèles économétriques classiques.

Ce billet s'inscrit dans une lignée de travaux académiques sur la modélisation des émissions carbone des entreprises. [Goldhammer et al. \(2017\)](#) ont été ainsi parmi les premiers à proposer une approche quantitative, en recourant à des modèles de régression linéaire. Dans leur prolongement, [Nguyen et al. \(2021\)](#) ont approfondi cette démarche en mobilisant

des modèles d'apprentissage automatique. Ils ont montré que ces méthodes permettent d'améliorer considérablement la précision des estimations d'émissions.

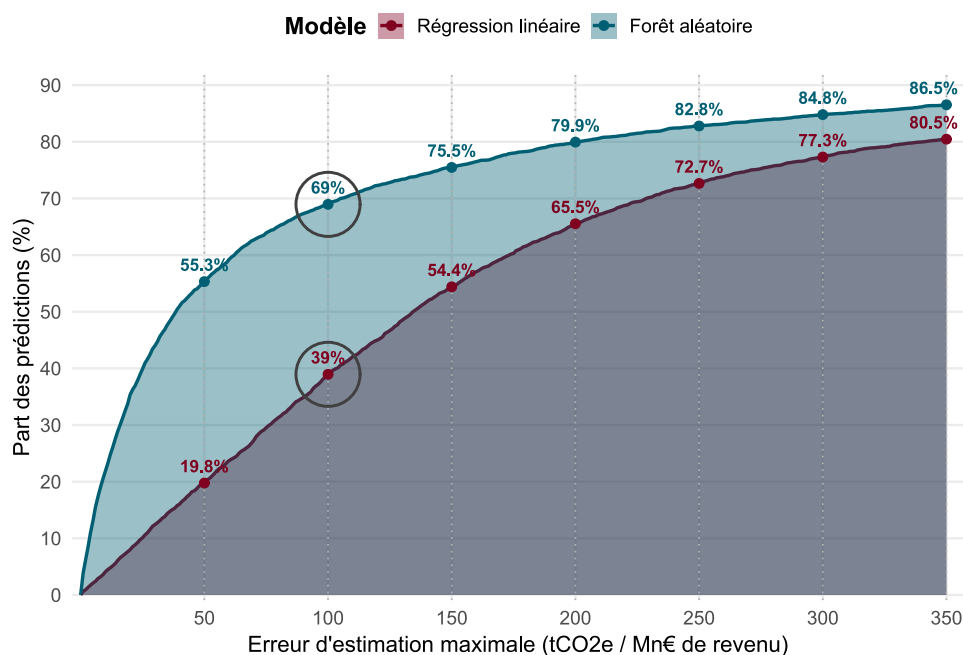
Des émissions estimées précisément et à grande échelle grâce à l'apprentissage automatique

Pour combler ces données manquantes, la première étape consiste à calibrer plusieurs modèles à partir des données de plus de 7 000 entreprises cotées, issues d'une centaine de pays à travers le monde. Ces modèles visent à prédire aussi précisément que possible l'intensité carbone des entreprises, c'est-à-dire la quantité de gaz à effet de serre émis directement par l'entreprise (Scope 1) et indirectement, principalement via leur consommation d'électricité (Scope 2), pour chaque million d'euros de chiffre d'affaires. Cet indicateur permet de rapporter les émissions à la production de l'entreprise (en valeur monétaire). Une alternative pourrait consister à prédire les émissions en niveau absolu (tCO₂e).

Les variables mobilisées pour réaliser les prédictions couvrent les aspects environnementaux de l'entreprise (ses objectifs de décarbonation, son secteur, etc.) et ses caractéristiques financières (taux d'investissement, risque de marché, etc.). Dans un second temps, le modèle présentant les meilleures prédictions au sein de l'échantillon d'entraînement est sélectionné pour estimer les intensités carbone des entreprises qui ne publient pas leur émissions.

Le modèle le plus performant est un modèle de type forêt aléatoire, développé par [Breiman \(2001\)](#). Il parvient à prédire près de 70% des intensités carbone des entreprises hors échantillon avec une erreur inférieure à 100 tCO₂e/million d'euros de chiffre d'affaires (*Graphique 2*), soit environ un quart de la valeur moyenne des émissions publiées (370 tCO₂e/million d'euros de chiffre d'affaires). La corrélation élevée (0,78) entre les valeurs estimées et les valeurs publiées confirme la fiabilité du modèle (*Graphique 3*). Ces performances sont cohérentes avec celles obtenues dans d'autres travaux, à l'image de ceux de l'Institut Louis Bachelier ([Barreau et al., 2024](#)), et nettement supérieures à celles d'un modèle de régression linéaire classique.

Pour estimer les intensités carbone, les variables les plus pertinentes sont, d'après notre modèle, les classifications sectorielles, la localisation géographique et, dans une moindre mesure, la capitalisation boursière et la part des actifs immobilisés. Les objectifs de décarbonation annoncés par les entreprises jouent un rôle limité dans l'estimation de leur intensité carbone.



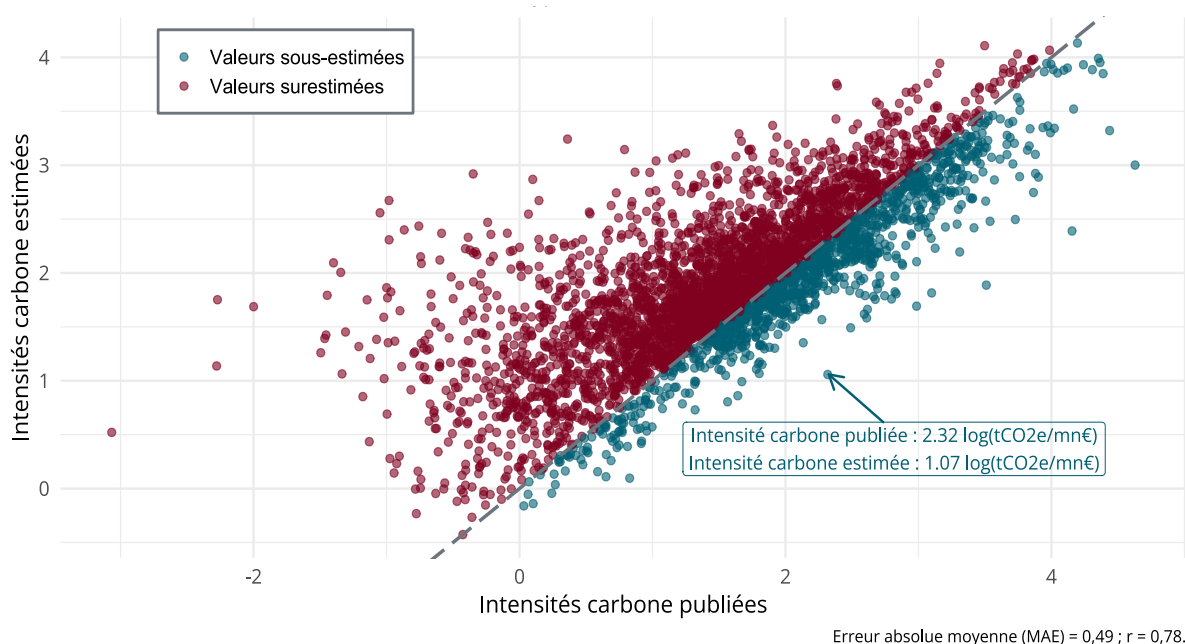
Graphique 2 : Comparaison des performances de deux modèles selon leur précision d'estimation

Source : ISS Stoxx, Eikon Refinitiv, calculs Banque de France.

Note de lecture : Le modèle de forêt aléatoire produit des prédictions dont l'erreur est inférieure à 100 tCO₂e/Mn€ de chiffre d'affaires dans 69% des cas, contre 39% pour la régression linéaire.

Les modèles d'IA doivent être complétés par une expertise humaine

Une limite du modèle est sa tendance à surestimer les intensités carbone, notamment pour les entreprises peu polluantes (*Graphique 3*), et, à l'inverse, à sous-estimer les intensités carbone des entreprises les plus émettrices. Pour ces cas singuliers, il est nécessaire de recourir à une analyse d'expert, afin d'apprécier les émissions carbone au regard des spécificités individuelles et sectorielles de chaque entreprise. Le développement de sous-modèles spécifiques à certains groupes d'entreprises, intégrant de nouvelles variables comme l'usage d'énergies fossiles, pourrait également améliorer la capacité prédictive globale.



Graphique 3 : Intensités carbone estimées par IA versus publiées par les entreprises

Source : ISS Stoxx, Eikon Refinitiv, calculs Banque de France.

Note : Les intensités carbone correspondent aux émissions carbone directes et indirectes divisées par le chiffre d'affaires. Les estimations ont été obtenues à l'aide d'un modèle d'apprentissage automatique de type forêt aléatoire. Les valeurs sont exprimées en logarithme décimal.

En conclusion, la qualité et la disponibilité des données sur les émissions carbone constituent un enjeu clé, tant pour les banques centrales que pour les institutions financières. Face à l'insuffisance des données publiées sur les émissions carbone, les modèles d'apprentissage automatique constituent une solution efficace pour les estimer à grande échelle. Ces estimations enrichissent la compréhension des risques de transition, de leur effet sur la politique monétaire et la stabilité financière, ainsi que de l'exposition de l'Eurosystème à ces risques. Sans se substituer à l'expertise humaine, ces modèles constituent un outil complémentaire précieux. En outre, alors que ce billet cherche à estimer les données manquantes, une extension de ce travail pourrait consister à prédire les émissions futures des entreprises, un enjeu clé pour anticiper la trajectoire de décarbonation des entreprises.