

1. [Accueil](#)
2. Entity Print

Document de travail

[Finance verte](#)

+ 1 -

[Covid-19](#)

Les données satellite de pollution de l'air prédisent-elles la production industrielle ?

18 Novembre 2021

Auteurs : [Jean-Charles Bricongne](#), Baptiste Meunier, Thomas Pical

Document de travail n°847. La crise de la Covid-19 a illustré le rôle des données haute-fréquence pour mesurer en temps réel l'activité économique. Dans cet esprit, nous étudions si les données satellite de pollution au dioxyde d'azote (NO₂, un polluant émis principalement par l'activité industrielle) dans la troposphère permettent de prédire la production industrielle. Nous montrons d'abord que ces données doivent être corrigées par les facteurs météorologiques qui affectent la qualité des données et le niveau de pollution. Nous utilisons des techniques de machine learning pour tenir compte des non-linéarités et interactions entre variables. Nous montrons ensuite que les prévisions en temps réel ont de meilleures performances en utilisant les données de pollution par rapport à des modèles basés sur les enquêtes PMI ou des termes autorégressifs. L'analyse fait aussi apparaître des hétérogénéités avec une contribution plus significative des données de pollution pendant les épisodes de crise, et avec une élasticité de la pollution au NO₂ par rapport à l'activité plus importante dans les pays où la part de l'industrie dans la valeur ajoutée est la plus forte. Disponible quotidiennement, à un niveau de précision élevée, et une couverture uniforme de tous pays – y compris ceux aux systèmes statistiques peu développés – cette étude montre le potentiel des données de satellite pour améliorer le suivi en temps réel de l'économie.

[DT_847.jpg](#)

Le choc soudain de la crise du Covid-19 - certains pays s'étant presque entièrement arrêtés en quelques jours - a mis un nouvel accent sur les données à haute fréquence. Les données hebdomadaires, quotidiennes, voire horaires, ont été largement utilisées pour évaluer en temps réel l'impact de la pandémie. En particulier, les mesures satellitaires de la pollution atmosphérique ont été mises en avant à plusieurs reprises pour montrer l'effet important supposé de la fermeture des usines sur la pollution atmosphérique.

Dans ce contexte, nous évaluons si les données satellitaires de la pollution troposphérique peuvent aider à prévoir la production industrielle. Nous nous concentrons sur le dioxyde d'azote (NO₂), un polluant principalement émis par l'activité industrielle. Par rapport aux indicateurs officiels ou alternatifs, ces données présentent les avantages d'une fréquence quotidienne et disponibles rapidement (le jour suivant la mesure), d'une couverture mondiale - y compris dans les pays en développement où les statistiques officielles sont limitées, de la granularité et de la gratuité. Si une piste de recherche future pourrait consister à comparer les performances prédictives des données satellitaires sur le NO₂ avec d'autres indicateurs à haute fréquence, la couverture mondiale et uniforme de ces données apparaît comme un avantage essentiel.

Les données satellitaires brutes sont cependant loin d'être prêtes à l'emploi. Notre première étape consiste à récupérer les données et à faciliter leur traitement : nous sélectionnons et agrégeons les données pertinentes au niveau du code postal, ce qui permet de passer d'un téléchargement quotidien de 4 Go en plusieurs fichiers à un seul fichier csv de 20 Mo. Comme la qualité des données satellitaires peut être altérée pour les zones nuageuses ou enneigées, nous nettoyons également les points aberrants. Il en résulte cependant un grand nombre de points manquants – au niveau local – qui peuvent entraîner des effets de composition indésirables lors de l'agrégation au niveau national. Nous interpolons donc ces données manquantes grâce à une technique d'apprentissage automatique (l'algorithme des k-plus proches voisins) qui tient compte des corrélations spatiales et temporelles. Enfin, la pollution au NO₂ dépend fortement des facteurs météorologiques (température, vent, humidité). Etant donné que leur effet est non-linéaire et comporte des interactions entre les variables, nous utilisons une forêt aléatoire (random forest) pour corriger des effets météorologiques sur la pollution. Les données sont ensuite agrégées au niveau national afin de correspondre à la granularité des statistiques officielles, notamment la production industrielle que nous souhaitons prévoir.

Dans un deuxième temps, nous vérifions la pertinence de la pollution au NO₂ du point de vue de la prévision. Nous nous appuyons sur une régression en panel sur 17 pays émergents et 16 pays avancés pour compenser la faible période disponible (depuis déc. 2018 seulement). Nous trouvons qu'un modèle basé sur des données quotidiennes de pollution au NO₂ surpasse les modèles de référence basés sur des données d'enquête (PMI) ou des termes autorégressifs (AR). En imitant une configuration en temps réel de mars 2020 à décembre 2020, nous constatons que cette surperformance se maintient pour tous les jours du mois : le modèle basé sur les données quotidiennes de pollution au NO₂ surpasse les modèles de référence pour chaque jour. Ces gains prédictifs sont plus importants à mesure que le mois avance et que davantage de points quotidiens deviennent disponibles (voir figure 1). On constate des gains de précision supplémentaires lorsqu'on s'appuie sur une approche MIDAS (*MIXed DATA Sampling*) - en utilisant un modèle MIDAS en panel récemment introduit dans la littérature.

Nous trouvons enfin des preuves d'hétérogénéités. Premièrement, les gains de précision sont plus importants pour les pays qui ont été plus touchés par la crise de la Covid-19, ce qui suggère que la contribution des données à haute fréquence est plus importante pendant les épisodes de "crise" que pendant les périodes "normales". Deuxièmement, l'élasticité de la pollution à la production industrielle semble être plus grande pour les pays dont la part de l'industrie manufacturière dans la valeur ajoutée est plus importante. Pour finir, nous travaillons sur la détection de point d'inflexion dans le cycle économique. En nous appuyant sur un modèle Markov-switching, nous constatons que les données quotidiennes sur la pollution au NO₂ permettent de détecter plus rapidement les points de retournement que les données officielles mensuelles, avec un gain d'environ 2,5 mois pour les premières par rapport aux secondes.

Télécharger l'intégralité de la publication

[Document de travail n°847. Les données satellite de pollution de l'air... \(PDF - 2 Mo\)](#)