

# ÉMULATEUR NEURONAL RAPIDE ET EVOLUTIF POUR DES PREVISIONS DE QUALITE DE L'AIR A HAUTE RESOLUTION SUR DE VASTES DOMAINES

M. Beauchamp<sup>1,3</sup>, V. Lécluse<sup>2</sup>, N. Thor<sup>2</sup>, J. Le Paih<sup>2\*</sup>

<sup>1</sup>Danish Meteorological Institute, NCKF, Sankt Kjelds Plads 11 - 2100 Copenhagen

<sup>2</sup>ATMO Grand Est, 5 Rue de Madrid, 67300 Schiltigheim

<sup>3</sup>IMT Atlantique, Lab-STICC OSE, INRIA Odyssey, 655 Av. du Technopole, 29280 Plouzané, France

\*Courriel de l'orateur : jerome.le-paih@atmo-grandest.eu

## TITLE

**Fast and scalable Air Quality neural emulator for high-resolution predictions on very large domains**

## ABSTRACT

SIRANet is a deep learning-based emulator designed to replicate air quality forecasts from the deterministic SIRANE model at fine resolution (25 m) across France's Grand Est region. Using UNet architectures conditioned on traffic, residential, and industrial emissions, it delivers accurate simulations while cutting computation time and costs. Validated against SIRANE results and observations, SIRANet enables operational forecasting and also offers new perspectives for analyzing pollution scenarios. This article focuses specifically on the results for particles.

## RESUME

L'outil SIRANet est un émulateur basé sur l'apprentissage profond, conçu pour reproduire les prévisions de qualité de l'air du modèle déterministe SIRANE avec une résolution fine (25 m) sur l'ensemble de la région Grand Est. Utilisant des architectures UNet conditionnées par les émissions liées au trafic, au résidentiel et à l'industrie, il fournit des simulations rapides et précises tout en réduisant le temps de calcul et les coûts. Validé par rapport aux résultats de SIRANE et aux observations, SIRANet permet des prévisions opérationnelles et offre également de nouvelles perspectives pour l'analyse de scénarios de pollution. Cet article s'intéresse spécifiquement aux résultats pour les particules.

**KEYWORDS:** UNet, emulator, air quality, forecasts / **MOTS-CLES :** UNet, émulateur, qualité de l'air, prévisions.

## 1. INTRODUCTION

La pollution atmosphérique constitue un enjeu majeur pour la santé publique et l'environnement, particulièrement dans les zones urbaines et industrialisées. Les modèles de dispersion déterministes tels que SIRANE sont largement utilisés pour simuler la dynamique des polluants à l'échelle urbaine. Bien que précis, ces modèles sont coûteux en temps de calcul et en ressources, ce qui limite leur utilisation pour des prévisions opérationnelles à haute résolution sur de grands domaines ou pour des analyses rapides de scénarios.

Pour répondre à ces contraintes, nous présentons SIRANet, un émulateur basé sur des réseaux de neurones profonds, conçu pour reproduire les sorties de SIRANE avec une précision élevée tout en réduisant drastiquement les temps de calcul. Ce travail s'inscrit dans le cadre du projet PREIPA (Prévision Régionale de l'Exposition Individuelle aux Polluants de l'Air), visant à fournir des prévisions à 25 mètres de résolution, horaires et quotidiennes pour la région Grand Est.

## 2. METHODOLOGIE

### 2.1. Le modèle SIRANE

Dans cette étude, nous avons développé un modèle de substitution basé sur l'apprentissage profond afin d'émuler le modèle de qualité de l'air SIRANE à une haute résolution spatiale de 25 mètres. Le modèle SIRANE repose sur une approche déterministe inspirée de la théorie des panaches gaussiens, adaptée aux environnements urbains. Il distingue deux zones : les rues canyon, où la dispersion est influencée par la géométrie urbaine et la dynamique locale, et les zones de fond, où s'appliquent des modèles gaussiens classiques (Souhac et al., 2011). Cette architecture permet une représentation réaliste des concentrations de polluants ( $\text{NO}_2$ ,  $\text{PM}_{10}$ ,  $\text{PM}_{2.5}$ ,  $\text{O}_3$ ) à des résolutions fines (25 m), mais au prix d'un coût computationnel élevé.

## 2.2. Architectures neuronales pour la modélisation par substitution

L'émulateur SIRANet repose sur une architecture UNet modifiée (Ronneberger et al., 2015), adaptée aux données géospatiales grâce à sa capacité à capturer des caractéristiques multi-échelles. Il comprend trois sous-modèles dédiés aux principaux secteurs d'émission : trafic, résidentiel et industriel (ce dernier désactivé faute de données d'entraînement suffisantes). Ces sous-modèles sont combinés de manière additive pour estimer la contribution individuelle de chaque source aux concentrations de polluants.

Le trafic est traité à une résolution fine de 25 m, tandis que les émissions résidentielles et industrielles, disponibles à des échelles plus grossières, sont modélisées à 500 m pour réduire les coûts de calcul. Les sorties sont ensuite suréchantillonnées et intégrées spatialement, permettant des prédictions fines sans dépendre des inventaires haute résolution pour ces secteurs. Cette approche multi-résolution s'inscrit dans les tendances récentes des cadres hybrides combinant physique et données (Keller et al., 2021 ; Barre et al., 2022).

Pour améliorer la scalabilité et la précision sur de grands domaines, SIRANet adopte un cadre multi-échelle : une approximation grossière du champ de polluants est calculée à 1 km, puis les anomalies fines sont modélisées via un second réseau neuronal entraîné sur des patches haute résolution (25 m).

Enfin, SIRANet intègre des données météorologiques (vent, précipitations, température) et une concentration de fond constante, encodées sous forme de tenseurs spatialisés pour capturer les influences locales. Sa structure modulaire et additive, plutôt qu'un modèle monolithique, favorise l'interprétabilité et la flexibilité, des critères essentiels pour l'évaluation des politiques publiques et l'attribution des sources.

## 2.3. Données

Les simulations SIRANE reposent sur un inventaire des émissions et des profils temporels associés. Dans cette étude, l'inventaire le plus récent produit par Atmo Grand Est a été utilisé, nommé InventAIR-A2021-V2023, finalisé en 2023. Cet inventaire fournit des estimations pour l'année 2021 et respecte les normes réglementaires européennes ainsi que les conventions de formatage des données (Atmo Grand Est, 2025). Il couvre l'ensemble de la région Grand Est et distingue trois types de sources : ponctuelles (ex. cheminées industrielles), linéaires (ex. trafic routier), surfaciques (ex. zones résidentielles ou agricoles). En raison de son niveau de détail, cet inventaire nécessite un prétraitement et une agrégation en catégories opérationnelles avant intégration dans SIRANE. Les émissions de surface sont regroupées en trois classes : agricoles, résidentielles-tertiaires et diverses, tandis que les sources linéaires sont subdivisées en tronçons autoroutiers, urbains et autres segments.

Les polluants étudiés incluent les oxydes d'azote ( $\text{NO}$ ,  $\text{NO}_2$ ), les particules ( $\text{PM}_{10}$  et  $\text{PM}_{2.5}$ ) et l'ozone ( $\text{O}_3$ ), considérés comme essentiels pour l'évaluation de la qualité de l'air urbain.

Pour assurer la cohérence avec la dynamique à méso-échelle, un post-traitement est réalisé via la chaîne opérationnelle PREVEST Atmo Grand Est, qui s'appuie sur le modèle CHIMERE (Menut et al., 2024) et des paramètres météorologiques issus du modèle WRF calibré (Skamarock et al., 2008).

## 3. RESULTATS

### 3.1. Stratégie d'entraînement et validation

La stratégie d'apprentissage de SIRANet vise une généralisation robuste face à des conditions météorologiques et des régimes d'émissions variés. Le jeu d'entraînement comprend 50 jours simulés par SIRANE ( $\approx 1\,200$  h) couvrant des environnements urbains et ruraux du Grand Reims pour assurer la transférabilité au Grand Est. Cinq émulateurs spécialisés ont été développés pour  $\text{NO}$ ,  $\text{NO}_2$ ,  $\text{O}_3$ ,  $\text{PM}_{10}$  et  $\text{PM}_{2.5}$ , chacun entraîné sur 500 époques avec un GPU Tesla P100 ( $\approx 6$  h par modèle), démontrant la scalabilité.

La fonction de perte combine la RMSE globale et locale, avec l'optimiseur Adam et des régularisations près des axes routiers. L'évaluation a été réalisée sur 7 jours répartis en 2021 (168 cartes SIRANE), en utilisant trois métriques par pixel : biais moyen, RMSE et corrélation de Pearson.

Les résultats montrent une forte concordance avec SIRANE, notamment pour  $\text{PM}_{10}$  et  $\text{PM}_{2.5}$ , avec une corrélation temporelle moyenne de 0,99, une RMSE de  $0,25\,\mu\text{g}/\text{m}^3$  ( $\text{PM}_{10}$ ) et  $0,53\,\mu\text{g}/\text{m}^3$  ( $\text{PM}_{2.5}$ ), et des biais très faibles. Le modèle reproduit fidèlement la dynamique temporelle et spatiale, y compris près des zones à fort trafic (corrélation  $>0,85$ ), bien qu'une légère sous-estimation apparaisse pour les concentrations les plus élevées ( $>40\,\mu\text{g}/\text{m}^3$  pour  $\text{PM}_{2.5}$ ).

Ces performances confirment que SIRANet constitue une alternative rapide et efficace pour la modélisation urbaine haute résolution.

### 3.2. Comparaison avec les observations

Pour évaluer la performance de l'émulateur, ses résultats ont également été comparés à ceux de SIRANE et aux observations de trois stations d'Atmo Grand Est (Bétheny, Doumer, Jean d'Aulan) dans le Grand Reims. Les analyses portent sur les  $PM_{2.5}$ , avec des indicateurs tels que la corrélation de Pearson, la RMSE normalisée et le biais moyen.

L'émulateur reproduit fidèlement la dynamique temporelle et les niveaux de pollution, mais hérite des biais de SIRANE puisqu'il a appris à le reproduire. Voici à titre d'illustration, les performances respectivement de SIRANE et de l'émulateur en  $PM_{2.5}$  au niveau de la station trafic Doumer : corrélation 0.79 / 0.81, RMSE 0.64 / 0.63, biais 1.6 / 2.01.

Cette limitation ouvre une perspective clé : intégrer des techniques d'assimilation de données (Asch et al., 2016) ou des approches end-to-end (Beauchamp et al., 2023 ; Martin et al., 2025) pour ajuster les prévisions selon les observations. Une telle intégration permettrait de rendre SIRANet cohérent avec les modèles méso-échelle comme PREVEST (Atmo Grand Est, 2017), améliorant la précision des prévisions urbaines et régionales.

### 3.3. Déploiement opérationnel de SIRANet

Depuis l'automne 2024, l'émulateur SIRANet est déployé par Atmo Grand Est pour fournir des prévisions horaires sur 48 heures à une résolution de 25 m sur toute la région Grand Est ( $\approx 57\,500\text{ km}^2$ ) comme en témoigne la Figure 1. Le domaine est divisé en 20 tuiles avec zones tampons pour garantir la continuité, représentant environ 95 millions de points par polluant et 20 milliards de valeurs par cycle.

Un post-traitement assure la cohérence avec le modèle méso-échelle PREVEST, avant l'assemblage des tuiles via des coefficients de transition. Ce système optimise mémoire et temps d'inférence, permettant une diffusion rapide pour l'information publique et les décisions sanitaires.

SIRANet est 89 fois plus rapide que SIRANE et réduit les coûts par 4 : les prévisions régionales sont générées en moins de 2 h avec 3 GPU, et pourraient passer à 20 min avec 20 GPU, offrant des prévisions quasi-temps réel.

Bien que le modèle ait été entraîné uniquement sur le Grand Reims, il se généralise efficacement à l'ensemble du Grand Est grâce à son architecture multi-échelle et à l'apprentissage par transfert, confirmant sa robustesse et son évolutivité.

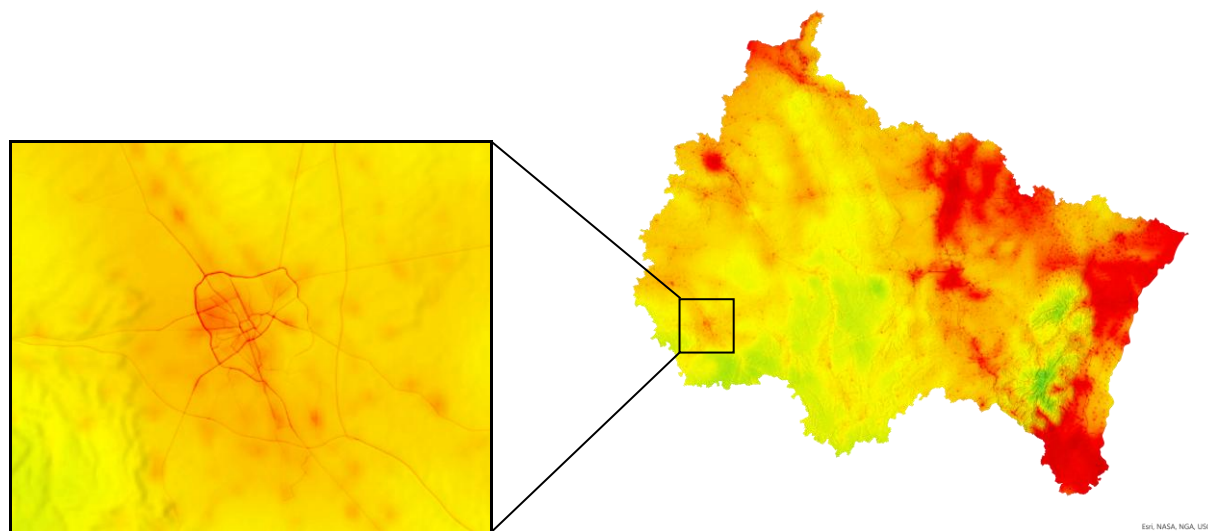


Figure 1. Pour une échéance donnée (5 mars 2025 18h) concentrations en  $PM_{2.5}$  sur la région Grand Est obtenues via SIRANet, avec zoom sur une zone d'intérêt (Agglomération de Troyes) permettant de mettre en évidence l'influence des émissions linéiques, ainsi que l'absence de discontinuité à l'échelle du la région.

### 3.4. Au-delà du cas d'utilisation opérationnel : évaluation rapide de scénarios d'émissions

L'émulateur SIRANet offre une fonctionnalité clé : la simulation rapide de scénarios d'émissions, essentielle pour tester des politiques de réduction ou des plans d'urgence (Relvas et al., 2022 ; Atkinson et al., 2022). En modifiant simplement les inventaires d'émissions (trafic, résidentiel, industriel), il prédit l'impact sur les concentrations de polluants avec une haute résolution spatiale, sans relancer SIRANE.

Chaque carte est générée en quelques secondes, permettant d'évaluer plusieurs scénarios en minutes plutôt qu'en heures ou jours. Pour cette étude, trois scénarios illustratifs ont été testés sur les  $PM_{2.5}$  : S1 (-50 % toutes émissions), S2 (-50 % trafic, -100 % résidentiel) et S3 (-100 % trafic, -50 % résidentiel).

Les résultats montrent une forte concordance avec SIRANE, y compris pour les réponses non linéaires aux réductions d'émissions. L'inspection visuelle confirme la plausibilité : suppression du trafic efface le réseau routier, tandis que la suppression des émissions résidentielles concentre visuellement la pollution sur les axes routiers.

Cette capacité fait de SIRANet un outil puissant pour la planification urbaine et l'évaluation des politiques environnementales.

#### 4. CONCLUSION ET DISCUSSION

Cette étude met en évidence le potentiel de SIRANet, un émulateur basé sur UNet, pour reproduire des simulations de qualité de l'air urbain à haute résolution générées par SIRANE. Il offre notamment une forte fidélité pour les particules PM<sub>10</sub> et PM<sub>2.5</sub> grâce à une architecture multi-échelle et modulaire.

L'émulateur permet aussi des évaluations rapides de scénarios en modifiant les émissions d'entrée, ce qui en fait un outil efficace pour la planification et la gestion de la qualité de l'air. Il peut produire des prévisions à 48h et 25m de résolution pour toute la région Grand Est en moins de deux heures, voire 20 minutes avec plus de GPU, rendant possible une utilisation en temps réel.

Cependant, il hérite des biais du modèle SIRANE, ce qui souligne la nécessité d'intégrer des techniques d'assimilation de données pour améliorer la précision. Les perspectives incluent l'extension à des modèles chimiques complets, l'intégration de scénarios extrêmes, l'ajout de champs météorologiques haute résolution et l'utilisation de réseaux neuronaux graphiques (GNN) pour mieux représenter la topographie urbaine. Actuellement, seules les sorties SIRANE servent de cibles, mais à l'avenir des contributions sectorielles explicites pourraient être intégrées.

#### 5. REMERCIEMENTS

Ce travail a bénéficié du soutien financier de l'Agence Régionale de Santé du Grand Est et de la Collectivité du Grand Reims. Il a également profité des capacités de calculs haute performance (HPC) du supercalculateur ROMEO, hébergé par l'Université Reims Champagne-Ardenne.

Nous remercions également l'équipe SIRANE (Laboratoire de Mécanique des Fluides et d'Acoustique (LMFA) de l'École Centrale de Lyon) pour son logiciel, utilisé dans ce travail pour construire le jeu de données d'entraînement, nécessaire au développement d'un émulateur performant de qualité de l'air.

#### 6. REFERENCES

- Asch, M., Bocquet, M., & Nodet, M. (2016). Data Assimilation. Society for Industrial and Applied Mathematics. Retrieved 2018-01-22, from <https://doi.org/10.1137/1.9781611974546> doi: 10.1137/1.9781611974546
- Atkinson, W., Eastham, S. D., Chen, Y.-H. H., Morris, J., Paltsev, S., Schlosser, C. A., & Selin, N. E. (2022). A tool for air pollution scenarios (taps v1.0) to enable global, longterm, and flexible study of climate and air quality policies. *Geoscientific Model Development*, 15 (20), 7767–7789. Retrieved from <https://gmd.copernicus.org/articles/15/7767/2022/> doi: 10.5194/gmd-15-7767-2022
- Atmo Grand Est. (2025). Données - Observatoire de la qualité de l'air du grand est. Récupéré à partir de <https://observatoire.atmo-grandest.eu/donnees/>
- Barre, J., et al. (2022). High-resolution air quality mapping with deep learning from fixed and mobile sensors. *Nature Sustainability*, 5 , 365–373. doi: 10.1038/s41893-022-00861-y
- Beauchamp, M., Febvre, Q., Georgenthum, H., & Fablet, R. (2023). 4dvarnet-ssh: end to-end learning of variational interpolation schemes for nadir and wide-swath satellite altimetry. *Geoscientific Model Development* , 16 (8), 2119–2147. Retrieved from <https://gmd.copernicus.org/articles/16/2119/2023/> doi: 10.5194/gmd-16-2119-2023
- Martin et al., 2025
- Menut, L., Cholakian, A., Pennel, R., Siour, G., Mailler, S., Valari, M., . . . Meurdesoif, Y. (2024). The chimere chemistry-transport model v2023r1. *Geoscientific Model Development* , 17 , 5431–5457. doi: 10.5194/gmd-17-5431-2024
- Keller et al., 2021
- Relvas, H., Lopes, D., Ferreira, J., Silva, A., Rafael, S., Lopes, M., . . . Miranda, A. I. (2022). Scenario analysis of strategies to control air pollution. *Urban Climate*, 44, 101201. Retrieved from <https://www.sciencedirect.com/science/article/pii/S2212095522001195> doi: <https://doi.org/10.1016/j.uclim.2022.101201>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention – miccai 2015* (Vol. 9351, pp. 234–241). Springer. doi: 10.1007/978-3-319-24574-4\_28
- Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., . . . Powers, J. G. (2008). A description of the advanced research wrf version 3 (NCAR Technical Note NCAR/TN-475+STR). Boulder, Colorado: National Center for Atmospheric Research. (Version 3; widely used reference for WRF citations)
- Soulhac, L., Perkins, R. J., & Salizzoni, P. (2011). The model sirane for atmospheric urban pollutant dispersion; part i, presentation of the model. *Atmospheric Environment* , 45 (39), 7379–7395. doi: 10.1016/j.atmosenv.2011.05.065