

Modélisation statistique pour la Qualité de l’Air

Elizabeth Fu, 5ème année GMM-N, Promotion 54

INSA Toulouse - Département de Mathématiques Appliquées

Introduction

Airparif met à disposition des cartographies des concentrations horaires de NO_2 , PM_{10} et $\text{PM}_{2.5}$ ($\mu\text{g}/\text{m}^3$) sur toute l’Île-de-France grâce à la chaîne de modélisation Horair (somme des concentrations de fond du modèle de chimie-transport Esmeralda et des concentrations de l’impact trafic du modèle de dispersion urbaine ADMS-Urban). Compte tenu des temps de calculs conséquents, Horair ne fonctionne pas encore en mode prévisionnel : le vecteur d’impact trafic comporte un grand nombre d de points récepteurs. Des techniques statistiques de l’état de l’art permettent de construire des métamodèles qui approchent les modèles déterministes de façon rapide et précise. L’objectif est de développer un mode prévisionnel (J+0 à J+2) pour Horair pour les concentrations de NO_2 .

Objectifs

- Développer plusieurs métamodèles ADMS.
- Développer plusieurs versions Horair (Esmeralda, MOCAF... \times métamodèle ADMS) et agréger séquentiellement ces versions pour fournir une prévision d’ensemble.

Méthodologie

Le métamodèle ADMS est composé d’un réducteur basé sur une technique de réduction de dimensionnalité : ACP, NMF (décomposition matricielle) ou AutoEncodeur (réseaux de Neurones), et d’un régresseur basé sur les Forêts Aléatoires [1]. À partir de quelques entrées ADMS (jour de l’année, heure de la journée et conditions météo-chimiques associées), le régresseur approxime un vecteur ADMS réduit, composé de N ($\ll d$) coefficients associés à N composantes. Ces composantes engendrent un sous espace d’approximation des vecteurs ADMS complets. Chaque coefficient quantifie l’importance de la composante associée dans la reconstruction du vecteur complet. La base d’apprentissage de ces techniques est composée de n heures (6 mois) échantillonnées de façon à ne pas sur-représenter certaines conditions. La performance des modèles est évaluée sur la reconstruction des vecteurs ADMS de l’année 2020 selon 3 scores : Biais, RMSE et Erreur maximale ($\mu\text{g}/\text{m}^3$).

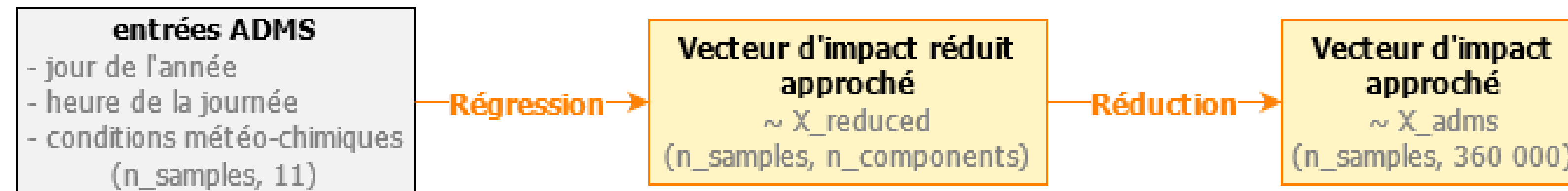


Figure 1: Utilisation opérationnelle du métamodèle ADMS

La prévision d’ensemble pour Horair est donnée par un modèle de régression linéaire Lasso (coefficient de régularisation α), qui pondère de façon optimale M versions de Horair en mode prévisionnel (coefficients α_i). La sortie \hat{y}_k d’une version Horair en mode prévisionnel est la somme des concentrations de fond calculées par un modèle de fond et les concentrations de l’impact trafic calculées par un métamodèle ADMS.

$$\text{forecast} = \alpha_0 + \sum_{k=1}^M \alpha_k \hat{y}_k \quad (1)$$

Les coefficients sont définis sur une fenêtre passée où les prédictions des versions Horair et les observations aux stations sont confrontées. Grâce à un paramètre β , les données passées les plus récentes (généralement meilleures) peuvent être considérées comme plus importantes.

Resultats

10 réducteurs à 30 ou 40 composantes : 1 ACP, 4 NMF et 5 Autoencodeurs et 10 régresseurs (Forêt Aléatoire à 100 arbres) ont été construits. Les réducteurs sont globalement peu biaisés, le RMSE est inférieur à $2 \mu\text{g}/\text{m}^3$ et l’Erreur Maximale est inférieure à $175 \mu\text{g}/\text{m}^3$ et certains surpassent même le modèle d’ACP considéré comme référence. Comparé aux observations, les 10 métamodèles ADMS performant aussi bien que ADMS (Table 1). Globalement, ils montrent une surestimation des concentrations inférieure à $0.5 \mu\text{g}/\text{m}^3$ et une erreur quadratique inférieure à $4 \mu\text{g}/\text{m}^3$ en moyenne.

Modèle	Bias	RMSE	Erreur max
ADMS	-43.307	47.657	100.309
acp30 + rf	-41.275	46.057	96.792
nmf30 + rf	-40.379	45.444	96.972
nmf30_bis + rf	-40.310	45.336	97.118
nmf30_nonsparse + rf	-40.480	45.528	97.077
nmf30_nonsparse_bis + rf	-40.560	45.565	97.301
ae40_nonneg + rf	-39.516	44.525	95.156
ae40_nonneg_bis + rf	-40.135	45.105	96.373
ae30_elu + rf	-42.014	46.896	98.765
ae30_elu_bis + rf	-41.467	46.413	98.166
ae40_elu_nonneg + rf	-40.130	45.178	96.481

Table 1: Performance de ADMS et des 10 métamodèles pour la station N2AUT (trafic à Paris). Les concentrations de NO_2 prédites par les modèles sont interpolées aux stations.

Le temps de calcul des concentrations horaires en NO_2 est de l’ordre d’une dizaine de ms, et d’environ 100 ms pour les autoencodeurs profonds non-linéaires sur un seul coeur de calcul, contre 1 à 2 minutes sur 32 coeurs de calculs pour ADMS-Urban.

20 versions de la chaîne Horair en mode prévisionnel ont été construites à partir de 2 modèles de fond (Esmeralda et MOCAF) et des 10 métamodèles ADMS. La prévision d’ensemble à quelques stations a été évaluée sur une fenêtre future antérieure à la date de test, ce qui permet d’extraire les observations aux stations sur l’horizon de prévision. Une prévision d’ensemble spécifique (paramétrisation de α et β) à chaque station a été donnée jusqu’à J+2 en considérant les données passées jusqu’à J-3 (exemple de la station N2A1 en Fig 2). Les résultats suggèrent que la prévision d’ensemble est meilleure que les prévisions individuelles des différentes versions Horair et est satisfaisante jusqu’à J+1.

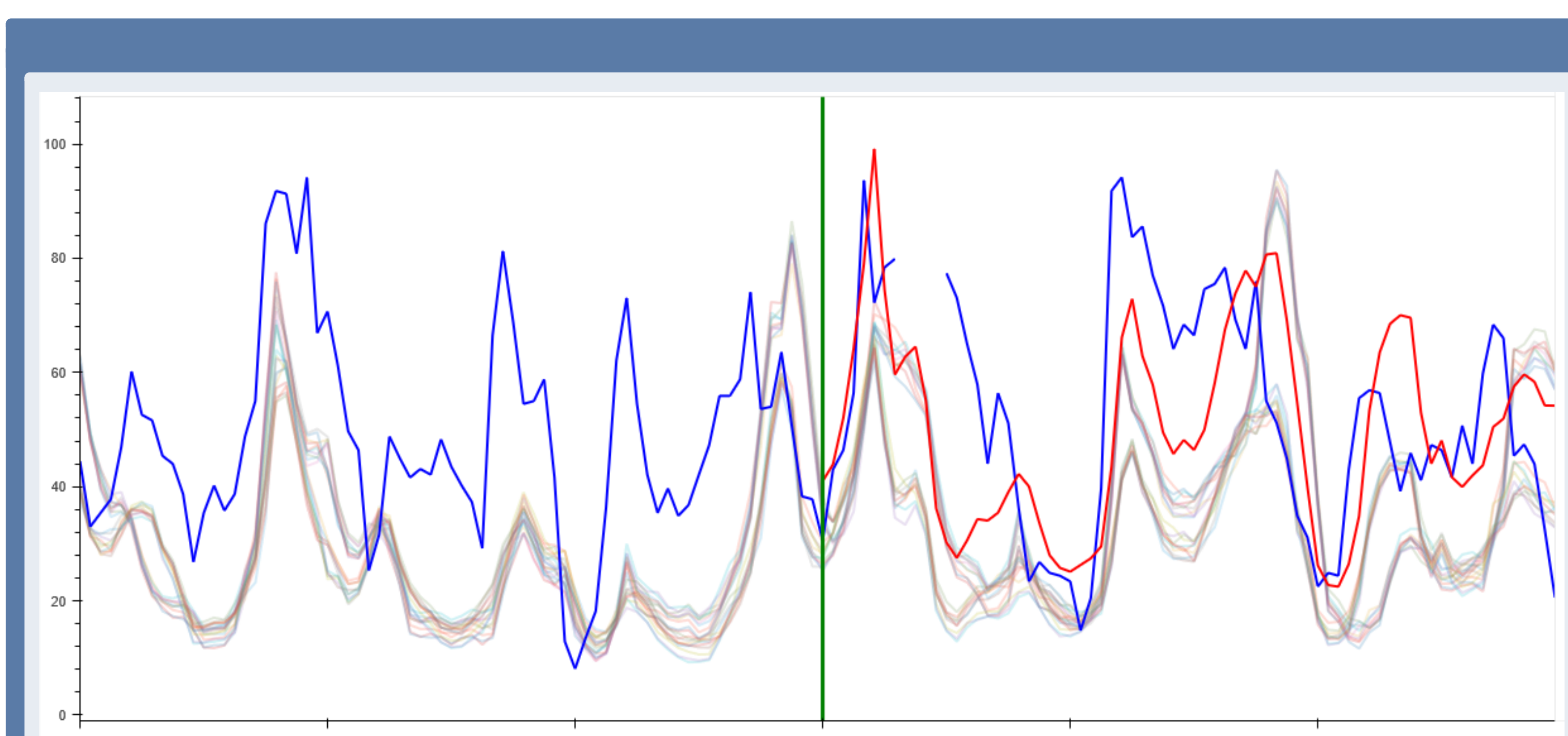


Figure 2: Observations (courbe bleue) et prévision d’ensemble (courbe rouge) du 11 mai 2021 (ligne verticale verte) jusqu’à J+2 pour la station N2A1 (trafic). Les courbes grisées représentent les prédictions individuelles des membres de l’ensemble. Modèle Lasso avec $\alpha = 10^2$, $\beta = 10^1$

Discussion

L’objectif était de développer un mode prévisionnel pour la chaîne de modélisation Horair pour les concentrations en NO_2 . Pour cela, différentes versions de la chaîne Horair en mode prévisionnel ont été construites par couplage de plusieurs métamodèles ADMS-Urban et de fonds déterministes. Les métamodèles proposés dans cette étude utilisent des techniques de réduction de dimensionnalité et de régression, et sont capables de reproduire le comportement général d’ADMS-Urban sur l’ensemble de l’IDF et sur une année entière. La prévision d’ensemble pour Horair est fournie par une agrégation séquentielle des différentes versions Horair jusqu’à J+2 et est globalement meilleure que les prédictions individuelles et satisfaisante jusqu’à J+1.

Améliorations futures

- Amélioration/ajout de métamodèles ADMS pour le NO_2 avec d’autres techniques de réduction (ICA, Sparse-PCA) et de régression (Boosted Trees, krigeage, etc.)
- Amélioration de la prévision d’ensemble Horair pour le NO_2 en considérant plus de modèles de fond.
- Déploiement d’une prévision d’ensemble pour les particules fines en suivant la même démarche.

References

- [1] Vivien Mallet, Anne Tilloy, David Poulet, Sylvain Girard, and Fabien Brocheton. Meta-modelling of adms-urban by dimension reduction and emulation. *Atmospheric Environment, Volume 184, Pages 37-46*, 2018.

Acknowledgements

Ce stage a été réalisé au service Emissions et Modélisation d’Airparif sous l’encadrement de Cyril Joly, ingénieur en mathématiques appliquées. Un grand merci pour leur bienveillance et l’intérêt qu’ils ont porté à mon travail.

Contact Information

- Email: elizabeth.wt.fu@gmail.com