



## Offre de Post doctorat - ULCO 2023

**Intitulé du Projet** : Complétion de données et simulation pour la qualité de l'air

**Laboratoire d'accueil** : LISIC UR 4491 (Equipes OSMOSE et SPÉCIFI)

**Encadrants** : Cyril Fonlupt, Matthieu Puigt, Gilles Roussel, Sébastien Verel,

**Contacts** :

[sebastien.verel@uni-littoral.fr](mailto:sebastien.verel@uni-littoral.fr)

[gilles.roussel@univ-littoral.fr](mailto:gilles.roussel@univ-littoral.fr)

[cyril.fonlupt@univ-littoral.fr](mailto:cyril.fonlupt@univ-littoral.fr)

[matthieu.puigt@univ-littoral.fr](mailto:matthieu.puigt@univ-littoral.fr)

**Résumé**:

Le LISIC est partie prenante dans l'élaboration d'un sujet scientifique partenarial dans le cadre d'un projet local. Il s'insère dans le projet de laboratoire sur les *jumeaux numériques* et s'inscrit dans les objectifs proposés en réponse à l'AAP "Projet de recherche et innovation" du Dialogue stratégique et de gestion du rectorat. Une spécification d'un outil numérique intégré a été défini au sein du LISIC et repose sur l'idée d'un **jumeau numérique** appelé "Plateforme Numérique pour l'Urban" couvrant différentes dimensions de surveillance de la qualité de l'air, de la simulation, du diagnostic de sources, de services aux usagers ou aux collectivités urbaines. Fortement basée sur l'apprentissage, la modélisation, l'optimisation et l'estimation, cette plateforme doit être lancée sans attendre la connexion à la collecte de mesures. Pour cela, un simulateur (Sim-Urban) doit être construit sur la base d'outils existants en modélisation-simulation de mobilité urbaine (dont la plateforme MatSim) complétée par des fonctions d'émission et de dispersion des sources élémentaires de la scène. Ce simulateur permet ainsi d'émuler des mesures synthétiques issues de capteurs fixes et mobiles. Ces mesures constitueront les données d'entrées du module monitoring-Urban dont la fonction sera d'estimer une cartographie de concentration de pollution de la scène urbaine

en s'appuyant sur une approche originale d'étalonnage et complétion de données par une décomposition matricielle/tensorielle régularisée par des connaissances de types modèle et matériel. Le présent projet initie donc une plateforme qui se veut fédérative, tant en interne au LISIC qu'avec des futurs partenaires associatifs, privés ou institutionnels.

## **Projet de recherche :**

### **I Contexte scientifique:**

Le LISIC est partie prenante dans l'élaboration d'un sujet scientifique partenaires dans le cadre d'un projet local. Une spécification d'un outil numérique intégré a été défini au sein du LISIC et repose sur l'idée d'un **jumeau numérique** appelé "Plateforme Numérique pour l'Urbain" couvrant différentes dimensions de surveillance de la qualité de l'air, de la simulation, du diagnostic de sources, de services aux usagers ou aux collectivités urbaines. Une décomposition de l'outil a pu être mis en évidence sur la base de différents modules dont :

**Monitor-Urban** : module d'assimilation-complétion et de cartographie des différents types de champs scalaires : (vent, pollution, flux urbains).

**Sim-Urban** : module effectuant hors lignes sur la même plateforme, la simulation et la représentation de scénarios de pollutions urbaine.

D'autres modules en aval des deux précédents concernent davantage l'optimisation de la gestion urbaine (d'un point de vue mobile ou statique) ou des services vers l'individu ou l'industrie.

Le projet d'instrumentation permettant la mesure des informations in-situ n'étant pas mis en place à ce stade, il est proposé ici toutefois d'initier une partie de la plateforme de monitoring à partir de données synthétiques issues de modèles. Ce projet s'insère dans le projet de laboratoire sur les *jumeaux numériques* et s'inscrit dans les objectifs proposés en réponse à l'AAP "Projet de recherche et innovation" du Dialogue stratégique et de gestion du rectorat. Le but est de construire des modèles (jumeaux) que l'on peut simuler numériquement afin de mieux comprendre le système et de tester de nouveaux scénarios potentiels futurs.

Afin de participer au démarrage du projet, un chercheur Post Doc est sollicité afin de s'approprier les travaux existants et de les intégrer dans la plateforme spécifiée.

### **II Le sujet**

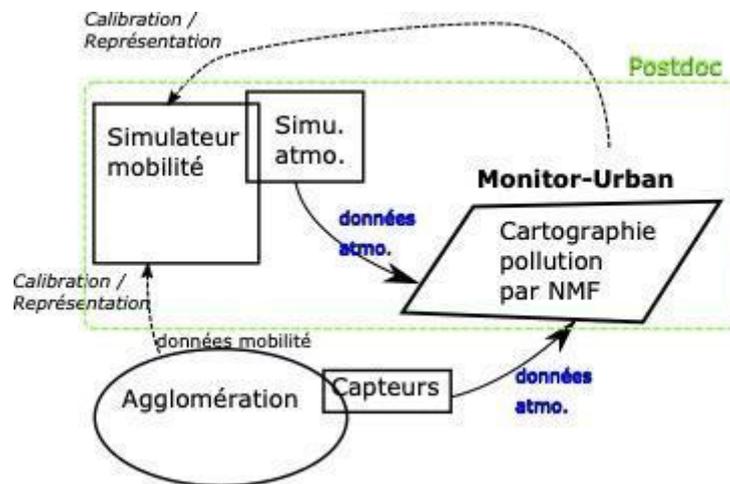
Le présent sujet de Post Doctorat s'intéresse aux développements scientifiques portant à la fois sur le module Monitor-Urban et Sim-Urban de la plateforme décrite précédemment. Concernant Monitor-Urban, le travail sera concentré sur l'assimilation par complétion de données à partir de mesures synthétiques irrégulièrement distribuées, spatialement et temporellement, sur un secteur géographique urbain. La plateforme a pleinement vocation à se connecter à une base de données alimentée par la collecte des mesures. Mais en attente de cette disponibilité, les développements théoriques et algorithmiques (nombre de capteurs, qualité des données, convergence des algorithmes, etc.) doivent être anticipés en vue de la phase d'exploitation des données réelles. C'est donc à partir de données synthétiques maîtrisées et paramétrables que nous proposons de s'appuyer pour alimenter le module Monitor-Urban. Afin de générer les données synthétiques, il est donc proposé d'utiliser un simulateur de pollution urbaine reprenant un simulateur de flux routiers associé à un modèle de sources [1,2]. Dans un premier temps, le simulateur envisagé est MatSim qui est

déjà utilisé dans l'équipe de recherche [3] : il permet à partir de données ouvertes sur les déplacements de simuler la mobilité urbaine et d'obtenir une estimation des sources de pollution (HBEFA factors).

Monitor-Urban a pour objectif de réaliser la cartographie de la qualité de l'air à partir de mesures issues de capteurs aux qualités hétérogènes. Plusieurs verrous scientifiques sont alors à résoudre :

- Les capteurs sont potentiellement non-étalonnés, du fait de la dérive des réponses des capteurs les plus bas-coût. Cela peut engendrer des analyses erronées puisque les mesures ne correspondent plus aux grandeurs réellement mesurées.
- Les capteurs (de vent ou de qualité de l'air) peuvent présenter des valeurs aberrantes, par exemple dues à des problèmes de communication. Cela signifie que ponctuellement, des mesures étonnamment basses ou élevées peuvent être réalisées. Ne pas détecter ces valeurs aberrantes comme tel peut alors engendrer des erreurs d'analyse.
- Le modèle de la qualité de l'air doit intégrer ces mesures de qualité diverse et aussi des mesures locales de vitesse et direction de vent. En effet, nous pensons qu'un modèle précis de la qualité de l'air doit tenir compte des mouvements d'air qui déplacent une partie de la pollution. Ces mouvements sont très complexes à modéliser en milieu urbain et le déploiement des capteurs de vent bas-coût présentés dans la partie "verrou technologique" nous permettra de mieux appréhender cette modélisation.

Pour répondre à ces trois questions de manière conjointe, nous proposons une modélisation de type *data-driven* basée sur le couplage d'une technique d'apprentissage par factorisation matricielles des données à faible rang, régularisées par des connaissances de type modèles et matérielles en même temps qu'une technique de complétion de données, elles aussi inhérente à une modélisation structurale de la décomposition matricielle.



### III Etat du sujet dans le laboratoire et l'équipe d'accueil :

Les travaux sur la mobilité urbaine ont été initiés au LISIC lors de la thèse de Florian Lepretre (2017 - 2020) en collaboration avec l'université de Shinshu à Nagano au Japon [3]. Ils ont permis de mettre au point à l'aide du simulateur MatSim des méthodes d'optimisation adaptative pour le réglage des feux de signalisation. Les modèles de déplacement urbains sont au cœur de 3 thèses qui ont débuté au LISIC en octobre 2021 qui développent des modèles de simulation à trois échelles (micro, méso, macro). Ces travaux développent un

modèle à événement discret original (DEVS) pour l'étude des communications inter-véhicules, ou pour l'étude de scénario urbains (calibration automatique sur des données issues d'une entreprise), ou encore utilisent le simulateur macro-échelle MatSim pour l'optimisation multi-objective des réseaux de transport publique.

Dans les travaux de la thèse de Clément Dorffer [5-7], nous avons proposé une méthode d'étalonnage à distance des réponses de capteurs basée sur le formalisme de factorisation matricielle. L'idée est d'apprendre à la fois la cartographie de la qualité de l'air et les paramètres d'étalonnage de chaque capteur. Pour cela, l'approche suppose que les capteurs soient mobiles et peuvent être en *rendez-vous*, c'est-à-dire être dans un voisinage spatio-temporel qui leur permet de mesurer le même phénomène. L'approche permet à la fois la cartographie et l'étalonnage des capteurs, en nécessitant beaucoup moins de *rendez-vous* que les approches de la littérature. Elle prend aussi en compte des mesures fixes beaucoup mieux étalonnées. Il est à noter que ces travaux ont été relevés par d'autres auteurs comme particulièrement novateurs (voir Section IV.D de [8]).

Les méthodes développées souffrent de plusieurs défauts qui ne les rendent pas applicables pour surveiller en temps réel la qualité de l'air :

- Elles nécessitent une optimisation des algorithmes actuels pour gérer le déluge de données qui seront fournies dans le cadre de la mesure de la qualité de l'air et il faut donc accélérer ces calculs. Des travaux ont été déjà menés dans ce sens [9,10]. Ils consistent à distribuer les calculs et à compresser les données par "projections aléatoires", c'est-à-dire de réaliser les calculs sur une esquisse "de petite taille" de la matrice originale à traiter. L'intérêt réside dans le fait que cette esquisse possède les mêmes propriétés et permette le même traitement que la matrice d'origine tout en étant beaucoup plus rapide à traiter, du fait de sa taille réduite ;
- L'approche d'assimilation dynamique a fait ses preuves en termes de finesse et de qualité par le mécanisme prédiction-mesures-correction souvent utilisé dans l'assimilation de données traditionnelles [11]. Cette approche est très gourmande en termes de calcul. Elle est néanmoins utilisée pour l'assimilation des champs de vent [12,13]. Une option proposée pour l'assimilation des champs de pollution dans l'approche par factorisation / complétion tensorielle a été d'associer un modèle de dispersion stationnaire local de la qualité de l'air. Dans ce projet, il s'agirait de coupler les techniques de factorisation proposées à un modèle d'assimilation de données du champ de vent car les polluants se comportent comme un fluide qui suit les mouvements de l'air. Elle fait apparaître toutefois un deuxième niveau de calcul. La modélisation d'un champ de vent urbain est généralement complexe à réaliser. Les techniques uniquement basées sur la prédiction par un modèle physique sont difficiles à gérer, du fait d'un grand nombre de paramètres à ajuster. Dans Monitor-Urban, les approches dites d'assimilation devront permettre d'estimer les champs de vent à partir des données mesurées. Compte tenu de la grande masse de données supposée disponible, il peut être intéressant de ne chercher à utiliser que les mesures vraiment utiles au modèle. Habituellement, restreindre le nombre de capteurs revient à n'utiliser que des capteurs placés de manière optimale, en fonction des incertitudes du modèle assimilé.

#### IV. Travail et échéancier proposés :

Tâche	Durée	Période
Bibliographie	1 mois	De Mois-1 à Mois-12
Combinaison de simulateurs et tests	4 mois	De Mois-2 à Mois-6
Génération de mesures synthétiques	1 semaine	De Mois-5 à Mois-12
Travaux algorithmiques de assimilation/complétion de données	5 mois	De Mois-7 à Mois-11
Rapport de fin de projet	3 semaines	Mois-12
congés	5 semaines	De mois-2 à Mois-12
<b>Total</b>	<b>12 mois</b>	<b>De Mois-1 à Mois-12</b>

#### V Collaborations prévues

Ce sujet fait l'objet d'un travail inter-équipe en mettant en commun les avancées en termes de simulation ou en termes de complétion de données. Ici il impliquerait le suivi de l'étude par des chercheurs de l'équipe OSMOSE et de l'équipe SPÉCIFI. Par ailleurs, le LISIC mettra en actions des demandes d'accès à des données partagées pour des tests sur des données in-situ.

#### VI Retombées scientifiques et économiques attendues :

Ce travail est une première étape pour l'élaboration de la plateforme de jumeau numérique pour la surveillance de la qualité de l'air avec des approches novatrices que le LISIC souhaitait mettre en place.

#### VII Références :

- [1] Hülsmann, F., Gerike, R. and Ketzel, M., 2014. Modelling traffic and air pollution in an integrated approach—the case of Munich. *Urban Climate*, 10, pp.732-744.
- [2] Armas R, Aguirre H, Daolio F, Tanaka K (2017) Evolutionary design optimization of traffic signals applied to Quito city. *PLOS ONE* 12(12): e0188757.
- [3] Leprêtre, F., Fonlupt, C., Verel, S., Marion, V., Armas, R., Aguirre, H. and Tanaka, K., 2019. Fitness landscapes analysis and adaptive algorithms design for traffic lights optimization on SIALAC benchmark. *Applied Soft Computing*, 85, p.105869.
- [4] Azeez, O.S., Pradhan, B., Shafri, H.Z., Shukla, N., Lee, C.W. and Rizeei, H.M., 2019. Modeling of CO emissions from traffic vehicles using artificial neural networks. *Applied Sciences*, 9(2), p.313.

- [5] Clément Dorffer, Méthodes informées de factorisation matricielle pour l'étalonnage de réseaux de capteurs mobiles et la cartographie de champs de pollution, *thèse de doctorat de l'Université du Littoral Côte d'Opale, soutenue à Calais, le 13/12/2017* [\(NNT: 2017DUNK0507\)](#).
- [6] C. Dorffer, M. Puigt, G. Delmaire, Gilles Roussel, Informed Nonnegative Matrix Factorization Methods for Mobile Sensor Network Calibration, *IEEE Transactions on Signal and Information Processing over Networks (IF 2020 = 3,153)*, vol. 4, no. 4, pp. 667-682, Dec. 2018, <https://doi.org/10.1109/TSIPN.2018.2811962>
- [7] C. Dorffer, M. Puigt, G. Delmaire, G. Roussel, Outlier-Robust Calibration Method for Sensor Networks, 2017 IEEE International Workshop of Electronics, Control, Measurement, Signals and their application to Mechatronics (ECMSM), pp 1-6, 24-26 May 2017, San Sebastian, Spain
- [8] B. Maag, Z. Zhou and L. Thiele, "A Survey on Sensor Calibration in Air Pollution Monitoring Deployments," in *IEEE Internet of Things Journal*, 2018. doi: 10.1109/JIOT.2018.2853660
- [9] F. Yahaya, M. Puigt, G. Delmaire, G. Roussel, How to Apply Random Projections to Nonnegative Matrix Factorization with Missing Entries?, in 2019 27th IEEE European Signal Processing Conference (EUSIPCO) , 10.23919/EUSIPCO.2019.8903036 , 2-6 Sept. 2019, <hal-02151521>
- [10] Farouk Yahaya, Compressive informed (semi-)non-negative matrix factorization methods for incomplete and large-scale data, with application to mobile crowd-sensing data, *thèse de doctorat de l'Université du Littoral Côte d'Opale, soutenue à Calais le 21/11/2021*.
- [11] Asch, M., Bocquet, M., Nodet, M., 2016. *Data Assimilation: Methods, Algorithms, and Applications*. SIAM.
- [12] L. Bourgois, G. Roussel, M. Benjelloun, G. Delmaire, Estimation of a semi-physical GLBE model using dual EnKF learning algorithm coupled with a sensor network design strategy: application to air field monitoring, *Elsevier, Information Fusion*, 14 (2013), pp. 335-348, <http://dx.doi.org/10.1016/j.inffus.2013.03.001>
- [13] Laurent Bourgois, Automates cellulaires et estimation états-paramètres pour la modélisation semi-physique : application à l'assimilation de données environnementales, *thèse de doctorat de l'université du Littoral Côte d'Opale, soutenue à Calais, le 6/7/2010*.