

Type d'offre : Offre en laboratoire

Date de publication : 14.11.25

**Laboratoire IBISC (Université Evry
Paris-Saclay)**

Offre de stage "Heterogeneous IoT network with low-cost sensors for predicting pollutant concentrations"

Informations générales

Type de contrat : Stage

Durée du contrat : 5-6 mois

Contact :

[Vincent Vigneron](#) / [Hichem Maaref](#) / [Ayamane Souani](#)

Date de prise de poste : mer 01/04/2026 - 12:00

Métier : Technicien

Thématique : Analyse et traitement d'images

Laboratoire IBISC (Université Evry Paris-Saclay) :

Les recherches conduites au sein du laboratoire [IBISC](#) traitent de la modélisation, la conception, la simulation et la validation des systèmes complexes, qu'ils soient vivants ou artificiels. Le laboratoire est organisé en 4 équipes (AROBAS, COSMO, IRA2, SIMOB) permettant de définir 2 axes de recherche transverses : STIC & Vivant (biologie computationnelle, bio-informatique, assistance à la personne, signal et image pour le biomédical) et STIC & Smart System (systèmes autonomes et intelligents, systèmes ouverts et sûrs). IBISC possède non seulement des plateformes référencées et soutenues par Genopole : [EVR@](#) (Environnements Virtuels et de Réalité Augmentée) et la plateforme logicielle de bioinformatique [EvrRNA](#) mais aussi des plateformes diverses en lien avec les systèmes intelligents : véhicules 2 roues, drones, robots.

Détail de l'offre (poste, mission, profil) :

Porteurs scientifiques : Aymane Souani, Hichem Maaref et V. Vigneron (IBISC)

Partenaires : IBISC (Université d'Évry-Paris-Saclay), TMECOMESURE

Spécialisation IA et Data Science : théorie de l'apprentissage automatique, statistiques en grande dimension, incertitude, théorie de l'information, modèles génératifs

Durée : 5 à 6 mois, début entre janvier et avril 2026

Financement : gratification de stage ECOMESURE

Lieu : laboratoire IBISC

Domaine d'application : green tech

Mots-clés : deep learning, prédiction de séries temporelles, apprentissage faiblement supervisé, fusion de modalités

1. Contexte de l'offre

Cette proposition de stage vise à développer un système de prévision pour optimiser l'estimation des concentrations de polluants tels que PM2.5, PM10, NO2, O3, CO à partir de variables météorologiques locales (température, hygrométrie, pression, vitesse du vent) au sein du réseau propriétaire de capteurs [™]ECOMESURE ([™]Ecomzen, [™]Ecomlite, [™]Ecomtreck, [™]Ecomsmart).

L'entrepôt de données historique contient plus de 10⁹ observations collectées en environnements urbains, industriels et commerciaux.

[™]ECOMESURE opère un réseau croissant de capteurs IoT à bas coût capables de transmettre, en quasi temps réel (1-5 minutes), des mesures de PM2.5, PM10, NO2, O3, CO et de variables micro-météorologiques vers une plateforme SaaS sécurisée. Cette télémétrie dense permet déjà des services d'alerte et de reporting hyper-locaux. Pour transformer ce flux massif de données en information exploitable, il est nécessaire de :

1. maintenir une calibration dynamique face au bruit et à la dérive ;
2. fusionner ces signaux bas coût avec des sources de données hétérogènes ;
3. produire des prévisions multi-horizons fiables à 24 h, 72 h et 168 h [1].

De telles prévisions hyper-locales permettront d'optimiser la ventilation des bâtiments, d'améliorer l'information du public et de soutenir l'évaluation des politiques publiques.

Problématique

Cependant, exploiter un réseau IoT aussi dense et hétérogène pose plusieurs défis. Les capteurs bas coût sont sujets aux biais, à la sensibilité température-humidité et à la dérive à long terme, nécessitant une calibration régulière pour garantir la fiabilité des données. L'intervalle de transmission de 1-5 minutes génère des flux de données haute fréquence sujets à des manques, des valeurs aberrantes et des problèmes de synchronisation dus aux contraintes de communication ou d'alimentation.

De plus, les concentrations de polluants présentent une forte hétérogénéité spatio-temporelle due aux différences microclimatiques et aux variations d'émissions selon

les sites, nécessitant une modélisation adaptative et non stationnaire. À l'échelle système, la plateforme SaaS sécurisée doit ingérer et gérer de grands volumes de télémétrie multimodale tout en garantissant scalabilité et résilience.

Enfin, la prévision hyper-locale multi-horizons dans ces conditions requiert des modèles capables de capturer des dépendances complexes, de quantifier l'incertitude et de rester interprétables pour l'aide à la décision et les usages réglementaires.

2. Méthodes / Approche de modélisation

Pour relever ces défis, nous proposons un cadre d'apprentissage auto-supervisé conçu pour exploiter les grands volumes de données non annotées produits en continu par les réseaux de capteurs bas coût hétérogènes (LCS).

La méthode effectue un pré-entraînement sur des jeux de données environnementales multi-sources via :

- la reconstruction de séquences masquées,
- l'apprentissage contrastif de représentations.

Cela permet au modèle de capturer des dépendances invariantes temporelles et inter-variables à travers des localisations et des types de dispositifs variés [2].

Une stratégie d'adaptation de domaine est ensuite appliquée afin d'aligner les représentations latentes du modèle pré-entraîné avec la distribution spécifique des capteurs TMECOMESURE, réduisant ainsi le besoin de calibration locale ou de données annotées. Ce transfert combine alignement adversarial des caractéristiques et régularisation distributionnelle pour assurer la cohérence entre les modalités de polluants et de variables météorologiques.

Le modèle résultant peut être affiné avec une supervision minimale pour prédire des quantiles de qualité de l'air à plusieurs horizons, améliorant la généralisation malgré la dérive des capteurs et la variabilité environnementale. En combinant pré-entraînement auto-supervisé et adaptation de domaine robuste [3], l'approche vise à réduire les erreurs de prédiction et maximiser la transférabilité au sein du réseau TMECOMESURE.

Pipeline de données et calibration

Le jeu de données comprend 12 mois de mesures co-localisées issues des capteurs EcomSmart et des stations de référence Atmo-France, permettant calibration et validation conjointes.

Les signaux bruts ont subi :

- détection d'anomalies,
- normalisation par quantiles,
- fusion temporelle à 5 minutes pour garantir la cohérence.

Une première calibration par réseau de neurones a corrigé les biais capteurs et dérives environnementales. Ensuite, une stratégie d'adaptation multibanc a aligné les représentations latentes afin de stabiliser les statistiques d'ordre 1 et 2 entre capteurs hétérogènes.

Le modèle de prévision obtenu a été condensé dans une version légère déployable en edge [4], fournissant des prévisions multi-horizons (1-168 h) sur tout le réseau Ecomesure.

3. Encadrement du stage et environnement scientifique

Profil recherché

Nous recherchons des candidats fortement motivés :

- (i) issus d'une formation en mathématiques, physique, informatique ou école d'ingénieurs ;
- (ii) ayant de solides bases en algèbre linéaire, analyse, probabilités, statistiques, machine learning et deep learning ;
- (iii) maîtrisant un langage scientifique, de préférence Python.

La connaissance des capteurs, notamment des capteurs de polluants, n'est pas requise mais constitue un atout.

La connaissance des bases de l'optimisation est également appréciée.

Informations pratiques

Le stage se déroulera principalement à l'UFR Sciences et Technologies (40 rue du Pelvoux), proche du centre-ville. Des périodes pourront également être effectuées chez ECOMESURE.

La gratification mensuelle est d'environ 1000 €.

Procédure de candidature

Envoyer une lettre de motivation, un CV et vos relevés de notes à :

[Vincent Vigneron](#) / [Hichem Maaref](#) / [Ayamane Souani](#)

Ce que nous offrons

- Une expérience pratique avec des techniques d'IA de pointe pour le pilotage de capteurs
- La possibilité de travailler sur des solutions green tech à fort impact
- Un encadrement rapproché par des chercheurs expérimentés du laboratoire IBISC
- Des opportunités de co-signer des publications et de présenter vos travaux en conférence
- Une possible continuation en thèse

References

- [1] G. Chen, S. Chen, D. Li, and C. Chen. A hybrid deep learning air pollution prediction approach based on neighborhood selection and spatio-temporal attention. *Scientific Reports*, 15(1), 2025.
- [2] C. Malings, K. E. Knowland, N. Pavlovic, J. G. Coughlin, D. King, C. Keller, S. Cohn, and R. V. Martin. Air quality estimation and forecasting via data fusion with uncertainty quantification: Theoretical framework and preliminary results. *Journal of*

Geophysical Research: Machine Learning and Computation, 1(4), 2024.

[3] K. Niresi, I. Nejjar, and O. Fink. *Efficient unsupervised domain adaptation regression for spatial-temporal air quality sensor fusion*, 2024.

[4] P. Wang, H. Zhang, J. Liu, F. Lu, and T. Zhang. *Efficient inference of large-scale air quality using a lightweight ensemble predictor*. *International Journal of Geographical Information Science*, 39(4):900–924, 2025.

URL de l'offre :

https://www.dataia.eu/sites/default/files/sujet_stage-Aymane.pdf

Lien vers l'offre sur le site dataia.eu :<https://da-cor-dev.peppercube.org/node/1459>