

Type d'offre : Laboratory offer

Post date : 14.11.25

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

Internship Offer "Heterogeneous IoT network with low-cost sensors for predicting pollutant concentrations"

Informations générales

Contract type : Stage

Contract length : 5-6 months

Contact :

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

Starting date : Wed 01/04/2026 - 12:00

Trade : Technicien

Topic : Analyse et traitement d'images

Laboratoire IBISC (Université Evry Paris-Saclay) :

Research conducted at the [IBISC](#) laboratory focuses on the modeling, design, simulation, and validation of complex systems, whether living or artificial. The laboratory is organized into four teams (AROBAS, COSMO, IRA2, SIMOB), enabling two cross-disciplinary research areas to be defined: ICT & Life Sciences (computational biology, bioinformatics, personal assistance, signals and images for biomedicine) and ICT & Smart Systems (autonomous and intelligent systems, open and secure systems). IBISC not only has platforms referenced and supported by Genopole: [EVR@](#) (Virtual and Augmented Reality Environments) and the [EvryRNA](#) bioinformatics software platform, but also various platforms related to intelligent systems: two-wheeled vehicles, drones, robots.

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

Scientific Supervisors: Aymane Souani, Hichem Maaref, and V. Vigneron (IBISC)

Partners: IBISC (University of Évry-Paris-Saclay), TM ECOMESURE

Specialized AI and Data Science: machine learning theory, high-dimensional statistics, uncertainty, information theory, generative models

Duration: 5 to 6 months, starting between January and April 2026

Funding: ECOMESURE internship grant

Location: IBISC laboratory

Application domain: green tech

Keywords: deep learning, time-series prediction, weakly supervised training, modality fusion

1. Context

This internship aims to develop a forecasting system to optimize the estimation of pollutant concentrations such as PM2.5, PM10, NO2, O3, and CO from local

meteorological variables (temperature, humidity, pressure, wind speed) across [™]ECOMESURE's proprietary sensor network ([™]Ecomzen, [™]Ecomlite, [™]Ecomtreck, [™]Ecomsmart).

The historical data warehouse contains more than 10^9 observations collected in urban, industrial, and commercial settings.

[™]ECOMESURE operates an expanding network of low-cost IoT sensors capable of transmitting, in near real-time (1–5 min), measurements of PM_{2.5}, PM₁₀, NO₂, O₃, CO, and micro-meteorological variables to a secure SaaS platform. This dense telemetry already supports hyper-local alerting and reporting services. To transform this massive data stream into actionable intelligence, it is necessary to:

1. maintain dynamic calibration against noise and drift;
2. fuse these low-cost signals with heterogeneous data sources;
3. produce reliable multi-horizon forecasts at 24 h, 72 h, and 168 h [1].

Such hyper-local predictions will optimize building ventilation, improve citizen information, and support public policy evaluation.

Problem Statement

Operating such a dense and heterogeneous IoT network presents multiple challenges. Low-cost sensors are prone to bias, temperature-humidity sensitivity, and long-term drift, making regular calibration essential to ensure reliable data. The 1–5 min transmission interval generates high-frequency data streams subject to gaps, outliers, and synchronization issues due to communication or power constraints.

Moreover, pollutant concentrations exhibit strong spatio-temporal heterogeneity driven by micro-climatic conditions and emission differences across sites, requiring adaptive, non-stationary modeling. At the system level, the secure SaaS platform

must ingest and manage large volumes of multimodal telemetry while maintaining scalability and resilience.

Finally, hyper-local multi-horizon forecasting under such conditions requires models capable of capturing complex dependencies, quantifying uncertainty, and remaining interpretable for decision-making and regulatory use.

2. Methods / Modeling Approach

To address these challenges, we propose a self-supervised learning framework designed to exploit the large volumes of unlabeled data generated continuously by heterogeneous low-cost sensor (LCS) networks.

The method performs pre-training on multi-source environmental datasets using:

- masked-sequence reconstruction,
- contrastive representation learning.

This enables the model to capture invariant temporal and cross-variable dependencies across diverse locations and device types [2].

A domain adaptation strategy is then applied to align the latent representations of the pre-trained model with the specific distribution of [™]ECOMESURE sensors, reducing the need for local calibration or labeled data. This transfer process combines adversarial feature alignment with distributional regularization to ensure consistency across pollutant and meteorological modalities.

The resulting model can be fine-tuned with minimal supervision to forecast multi-horizon air-quality quantiles, improving generalization under sensor drift and environmental variability. By coupling self-supervised pre-training with robust domain adaptation [3], the proposed approach aims to reduce prediction errors and maximize transferability across the expanding [™]ECOMESURE network.

Data Pipeline and Calibration

The dataset comprises 12 months of collocated measurements from EcomSmart sensors and Atmo-France reference stations, enabling joint calibration and validation.

Raw signals underwent:

- outlier detection,
- quantile normalization,
- temporal fusion at 5-minute resolution to ensure consistency.

An initial neural network calibration corrected sensor biases and environmental drift. Next, a multi-platform domain adaptation strategy aligned latent embeddings to stabilize first- and second-order statistics across heterogeneous sensor domains.

The resulting forecasting model was distilled into a lightweight, edge-deployable version [4], providing multi-horizon (1–168 h) air-quality predictions across the ECOMESURE network.

3. Internship Supervision and Scientific Environment

Candidate Profile

We are looking for highly motivated candidates:

- (i) with a background in mathematics, physics, computer science, or engineering;
- (ii) with strong foundations in linear algebra, analysis, probability and statistics, machine learning, and deep learning;
- (iii) with solid programming skills in a scientific language, preferably Python.

Knowledge of sensors—particularly pollutant sensors—is not required but is a strong plus.

Knowledge of basic optimization theory is also appreciated.

Practical Information

The intern will be primarily hosted at the UFR Sciences and Technology (40 rue du Pelvoux), close to the city center. Some periods may also be spent at ECOMESURE. The monthly internship stipend is approximately €1000.

Application Procedure

Send a motivation letter, a CV, and your academic transcript to:

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

What We Offer

- Hands-on experience with cutting-edge AI techniques for sensor control
- Work on real-world, high-impact green tech applications using deep learning
- Close mentorship from experienced researchers at the IBISC laboratory
- Opportunities to co-author publications and present your work at conferences
- Possibility to continue into PhD studies

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>