Title of the thesis	MUlti-scale processing of Spatio-temporal data applied to air Quality in Urban Areas
Acronym	MuSQUA
Reference number	036

Hosting institution	**Employer**
Université de Lille Website: https://www.univ-lille.fr/home/	Université de Lille Website: https://www.univ-lille.fr/home/
Hosting research unit 1	**Hosting research unit 2**
Name: Centre de Recherche en Informatique, Signal et Automatique de Lille Acronym: CRIStAL Identification number: UMR 9189 Address: Campus scientifique Bâtiment ESPRIT Avenue Henri Poincaré 59655 Villeneuve d'Ascq Website: https://www.cristal.univ-lille.fr/	Name: Laboratoire d'Optique Atmosphérique Acronym: LOA Identification number: UMR 8518 Address: UFR de Physique Bâtiment P5 Université de Lille 59655 Villeneuve d'Ascq Cedex Website: https://www-loa.univ-lille1.fr/
Principal supervisor	**Co-supervisor**
Name: Romain Surname: ROUVOY Email: romain.rouvoy@univ-lille.fr Phone: +33 359 358 777	Name: Suzanne Surname: CRUMEYROLLE Email: Suzanne.crumeyrolle@univ-lille.fr Phone: +33 3 20 43 44 72

Thesis information

Keywords	air pollution, distributed processing, particle matter, spatio-temporal model
Abstract	The presence of different pollutants (gaseous or particulate matter) within air in urban areas results in serious diseases and undesirable deterioration of our environment. While the political decision-makers have taken hold of the subject, they continue to stumble on the causes of these degradations, emitting various hypotheses on their origins. So far, the lack of widely-deployed *in situ* measurements, as well as the exploitation of inadequate mesoscale models, causes considerable difficulties in analyzing and predicting pollutant exposure evolution. Indeed, not only the pollutants but also a wide diversity of contextual factors (wind corridors, weather forecast, nearby constructions/roads, surrounding incidents, etc.) can contribute to the improvement or the degradation of air quality. This Ph.D. thesis, therefore, aims at revisiting the state-of-the-art approaches in distributed computing infrastructures to cope with the multi-scale challenge raised by modern air quality monitoring techniques. In particular, we aim at designing, implementing and deploying a multi-scale modeling approach that can leverage the volume and diversity of crowdsourced measurements to **recommend *in situ* deployments of measurement stations** (being mobile or stationary) in order to infer microscale propagation models (*e.g.*, streets), with accuracy. Then, this online propagation model will leverage artificial intelligence techniques, such as machine learning, to not only map the local propagation of particles, but also to **locate particle emitters and therefore deliver an open and continuously updated inventory of particle emissions**, which has never been provided at this resolution and therefore constitutes a strong innovation beyond the state of the art. To reach these ambitious objectives, we intend to leverage our experience in the design of large-scale distributed data processing systems. In

	particular, this Ph.D. thesis intends to benefit from the GreyCat machine learning database to support the implementation of our multi-scale modeling approach. GreyCat supports the storage and processing of complex time series as temporal graphs, which are a flexible data structure to model and reason on complex dynamic systems. The originality of GreyCat lies in its capability to embed machine learning algorithms (gaussian mixture models, recurrent neural networks, etc.) as nodes within the graph that can learn from the neighborhood. We believe that this approach offers an appropriate foundation to design and **build microscale propagation models that are composed to mesoscale ones**. This Ph.D. project will be conducted in close collaboration with experts from the DataThings company (Luxembourg) and the CSIRO research center (Australia). The impact of our joint research project intends not only to benefit computer science, but also a large spectrum of interdisciplinary domains involving technologies (electronics, sensors, algorithms), environments (physical models, atmospheric chemistry analysis), and humans (user studies, healthcare). Our results aim to be published in major conferences and journals of computer science, atmospheric physics and chemistry, and exposure science. Beyond the expected scientific advances, we also believe that our research activities can improve the quality of services and information delivered by local authorities and therefore favorably contribute to the reduction of particle emissions, thus improving the air quality in the large.
Expected profile of the candidate	The profile of the recruited Ph.D. candidates targets a talented Master student in Computer Science with advanced knowledge in machine learning algorithms and personal motivation for contributing to air quality modeling challenges.
Application procedure & Eligibility criteria	The application procedure and eligibility criteria are detailed on the European doctoral programme PEARL website www.pearl-phd-lille.eu. The funding is managed by the I-SITE ULNE foundation which is a partnership foundation between the University of Lille, Engineering schools, research organisms, the Institut Pasteur de Lille and the University hospital. The application file will have to be submitted before March 31, 2021 (10:00 AM - Paris Time) and emailed to the following address : international@isite-ulne.fr.
Net salary and Lump Sum	A net salary of about €1,600 + €530 per month to cover mobility, travel and family costs.