



**UNIVERSITE DE CORSE - PASCAL PAOLI**  
**ECOLE DOCTORALE ENVIRONNEMENT ET SOCIETE**  
**UMR CNRS 6134 (SPE)**



**Thèse présentée pour l'obtention du grade de**  
**DOCTEUR EN AUTOMATIQUE, SIGNAL,**  
**PRODUCTIQUE, ROBOTIQUE**  
**Mention : Génie informatique, automatique et**  
**Traitement du signal**

**Soutenu publiquement par**  
**Wani Théo TAMAS**  
Le 17 novembre 2015

---

**Prévision statistique de la**  
**qualité de l'air et d'épisodes de**  
**pollution atmosphérique en Corse**

---

**Directeurs :**

M. Gilles Notton, Dr-HDR, Université de Corse  
M. Christophe Paoli, Dr-HDR, Université de Corse

**Rapporteurs :**

M. Matthias Beekmann, DR, Université Paris Est Créteil  
M. Pierre-Charles Maria, PREM, Université de Nice Sophia Antipolis

**Jury :**

M. Matthias Beekmann, DR, Université Paris Est Créteil  
M. Dominique Lambert, Dr-HDR, Université Toulouse III - Paul Sabatier  
M. Pierre-Charles Maria, PREM, Université de Nice Sophia Antipolis  
M. Jean-François Muzy, DR, Université de Corse  
M. Gilles Notton, Dr-HDR, Université de Corse  
M. Christophe Paoli, Dr-HDR, Université de Corse

## Résumé

L'objectif de ces travaux de doctorat est de développer un modèle prédictif capable de prévoir correctement les concentrations en polluants du jour pour le lendemain en Corse. Nous nous sommes intéressés aux PM10 et à l'ozone, les deux polluants les plus problématiques sur l'île. Le modèle devait correspondre aux contraintes d'un usage opérationnel au sein d'une petite structure, comme Qualitair Corse, l'association locale de surveillance de la qualité de l'air.

La prévision a été réalisée à l'aide de réseaux de neurones artificiels. Ces modèles statistiques offrent une grande précision tout en nécessitant peu de ressources informatiques. Nous avons choisi le Perceptron MultiCouche (PMC), avec en entrée à la fois des mesures de polluants, des mesures météorologiques, et des sorties de modèles de chimie-transport (CHIMERE via la plate-forme AIRES) et de modèles météorologiques (AROME).

La configuration des PMC a été optimisée avant leur apprentissage automatique, en conformité avec le principe de parcimonie. Pour en améliorer les performances, une étude de sélection de variables a été au préalable menée. Nous avons comparé l'usage d'algorithmes génétiques, de recuits simulés et d'analyse en composantes principales afin d'optimiser le choix des variables d'entrées. L'élagage du PMC a été également mis en œuvre.

Nous avons ensuite proposé un nouveau type de modèle hybride, combinaison d'un classifieur et de plusieurs PMC, chacun spécialisé sur un régime météorologique particulier. Ces modèles, qui demandent un large historique de données d'apprentissage, permettent d'améliorer la prévision des valeurs extrêmes et rares, correspondant aux pics de pollution. La classification non-supervisée a été menée avec des cartes auto-organisatrices couplées à l'algorithme des k-means, ainsi que par classification hiérarchique ascendante. L'analyse de sensibilité a été menée grâce à l'usage de courbes ROC.

Afin de gérer les jeux de données utilisés, de mener les expérimentations de manière rigoureuse et de créer les modèles destinés à l'usage opérationnel, nous avons développé l'application « Aria Base », fonctionnant sous Matlab à l'aide de la Neural Network Toolbox.

Nous avons également développé l'application « Aria Web » destinée à l'usage quotidien à Qualitair Corse. Elle est capable de mener automatiquement les prévisions par PMC et de synthétiser les différentes informations qui aident la prévision rendues disponibles sur internet par d'autres organismes.

**Mots-clés** : Prévision ; Perceptron Multicouche ; Réseaux de Neurones Artificiels ; PM10 ; Ozone

## **Abstract**

The objective of this doctoral work is to develop a forecasting model able to correctly predict next day pollutant concentrations in Corsica. We focused on PM10 and ozone, the two most problematic pollutants in the island. The model had to correspond to the constraints of an operational use in a small structure like Qualitair Corse, the local air quality monitoring association.

The prediction was performed using artificial neural networks. These statistical models offer a great precision while requiring few computing resources. We chose the MultiLayer Perceptron (MLP), with input data coming from pollutants measurements, meteorological measurements, chemical transport model (CHIMERE via AIRES platform) and numerical weather prediction model (AROME).

The configuration of the MLP was optimized prior to machine learning, in accordance with the principle of parsimony. To improve forecasting performances, we led a feature selection study. We compared the use of genetic algorithms, simulated annealing and principal component analysis to optimize the choice of input variables. The pruning of the MLP was also implemented.

Then we proposed a new type of hybrid model, combination of a classification model and various MLPs, each specialized on a specific weather pattern. These models, which need large learning datasets, allow an improvement of the forecasting for extreme and rare values, corresponding to pollution peaks. We led unsupervised classification with self organizing maps coupled with k-means algorithm, and with hierarchical ascendant classification. Sensitivity analysis was led with ROC curves.

We developed the application “Aria Base” running with Matlab and its Neural Network Toolbox, able to manage our datasets, to lead rigorously the experiments and to create operational models.

We also developed the application “Aria Web” to be used daily by Qualitair Corse. It is able to lead automatically the prevision with MLP, and to synthesize forecasting information provided by other organizations and available on the Internet.

**Keywords:** Forecasting; Multilayer Perceptron; Artificial Neural Network; PM10; Ozone