

# PROPOSITION DE SUJET DE STAGE 2011-2012

(Bac +5 ; 6 mois ; CSTB-Champs-sur-Marne)

**Titre : Traitement non-supervisé, mise en place d'un processus de sélection des variables**

## **Contexte et éléments de problématique du travail :**

L'Observatoire de la Qualité de l'Air Intérieur (OQAI) a réalisé entre 2003 et 2005 une campagne nationale dans les logements sur un échantillon de 567 logements représentatif du parc des 24 millions de résidences principales de la France continentale métropolitaine. Cette campagne visait à dresser un état de la pollution de l'air dans l'habitat afin de donner les éléments utiles pour l'estimation de l'exposition des populations, la quantification et la hiérarchisation des risques sanitaires associés, ainsi que l'identification des facteurs prédictifs de la qualité de l'air intérieur.

Il existe plusieurs domaines pratiques dans lesquels le nombre  $n$  d'individus est généralement plus petit ou de taille insuffisante par rapport à la dimension  $p$  des données (généralement le nombre de variables). En particulier dans le domaine environnemental, l'explication des phénomènes de pollution de l'air intérieur dépend d'un nombre élevé de variables mixtes inter-corrélées (qualitatives et quantitatives) pour un nombre modeste d'individus ce qui constitue une difficulté majeure pour l'analyse de ces données. La sélection des variables les plus informatives permet de réduire la dimension des données et de faciliter alors l'analyse.

Une méthode non-supervisée w-HMTM (Weighted Hierarchical Mixed Topological Map) basée sur les cartes topologiques permet d'obtenir une classification des individus par rapport à l'ensemble des variables. Un simple calcul des valeurs tests effectué sur l'ensemble des variables pour chacune des classes de la classification obtenue par application de w-HMTM permet de mettre en évidence les variables les plus informatives pour chaque classe. Cette méthode en deux étapes successives présente l'inconvénient de séparer le processus de sélection de celui de la recherche des classes.

## **Nature du travail et questions posées :**

L'objectif de ce stage est de mettre en place une méthode de classification dans laquelle la sélection des variables les plus informatives fait partie intégrante du processus. Elle sera appliquée sur les données de la campagne « logements » de l'observatoire de la qualité l'air intérieur.

## **Durée**

6 mois

## **Aptitudes souhaitées :**

Niveau Bac +5

Rigueur, initiative, esprit de synthèse, dynamisme

Capacité rédactionnelle

## **Déplacements :**

Eventuellement en région parisienne

## **Indemnités de stage :**

407 euros nets mensuels + 50 % des frais de transport en commun

## **Contacts :**

- Madame Ndèye Niang/ Madame Corinne Mandin/ Mory Ouattara
- Téléphone : 01 64 68 85 97/01 64 68 88 17
- Adresses e-mail : [corinne.mandin@cstb.fr](mailto:corinne.mandin@cstb.fr); [mory.ouattara@cstb.fr](mailto:mory.ouattara@cstb.fr)