



## ARCUS E2D2

### Sujet de Thèse ARCUS E2D2 2017

**sélection de la thématique :** (cochez une ou plusieurs cases)

- Sp1 « Ville, Aménagement et Développement Durable » :
- Sp2 « Modélisation et Infrastructure pour l'Environnement » :
- Sp3 « Expertise et Traitement en Environnement » :
- Sp4 « Calcul scientifique »

**Partenaire proposant le Sujet :**

**Laboratoire d'accueil :** LISIC

**Responsable(s) :**

**André BIGAND (M.C. HDR), Emilie CAILLAULT (M.C.)**

**courriel :** [bigand@lisic.univ-littoral.fr](mailto:bigand@lisic.univ-littoral.fr)

**Tél. :** 33 321 465653

**LISIC, ULCO**

**50 Rue F.Buisson, BP 719**

**62228 CALAIS Cedex**

**Université d'accueil :** ULCO, Quai de l'Yser, 59, DUNKERQUE

**Partenaire potentiel pour la collaboration et la co-tutelle :**

- Si un partenaire est déjà identifié, veuillez compléter les informations suivantes (si disponible) : Y.Mohanna et O.Bazzi, Professeurs, Ecole Doctorale de l'Université Libanaise (UL), LIBAN

- Thématique : (1) qualité de l'air, (2) milieux aquatiques, (3) obésité, nutrition et activités sportives, (4) énergies propres et renouvelables, (5) gestion et le traitement des déchets, (6) Urbanisme

Laboratoire d'accueil :

Responsable(s) :

Université d'accueil: UL

**Mots clés :** Modèle de Markov caché, apprentissage semi-supervisé, complétion de données, efflorescences phyto-planctoniques.

**Points particuliers :**

(précisez les points particuliers que le candidat potentiel devra considérer, origine, langue, collaboration existante avec des équipes, compétences ....

Une bonne connaissance de la langue Française est souhaitable, ainsi qu'un goût certain pour la programmation. Un candidat de l'UL, Liban (convention UL-ULCO) sera préféré.

## **TITRE : MODÈLE DE MARKOV CACHÉ PAR APPRENTISSAGE SEMI-CONTRAIN. APPLICATION À LA CARACTÉRISATION DES EFFLORESCENCES PHYTO-PLANCTONIQUES.**

### **SUJET DE LA THESE**

Cette thèse a pour objet de proposer un apprentissage semi-supervisé d'une modélisation markovienne de phénomènes environnementaux à partir de données multidimensionnelles haute fréquence. Les aspects apprentissage automatique et apprentissage semi-supervisé sont des thématiques clés de l'Université Libanaise (UL, équipe des professeurs Y. Mohanna et O. Bazzi, Ecole Doctorale de l'Université Libanaise, LIBAN) et de l'université du Littoral Côte d'Opale (Laboratoire LISIC, équipe IMAP) dont les collaborations ont déjà été amorcées depuis 2011. Les collaborations du LISIC avec Ifremer (LER Boulogne-sur-Mer) existantes depuis 2008 permettront d'asseoir l'interprétation biologique des données et seront étendues avec l'UL.

Caillault et al., Rousseeuw et al. [1-2] ont montré théoriquement et expérimentalement qu'il est possible de générer un Modèle de Markov Caché (MMC) par apprentissage non supervisé. L'utilisation de la théorie associée à la classification spectrale a alors permis de mettre en avant une structure et caractérisation du MMC cohérente vis-à-vis des observations sans aucune intervention de paramétrage, sans aucun *a priori* sur la structure. Cette technique a été appliquée pour la première fois pour modéliser la dynamique des efflorescences phyto-planctoniques en Manche orientale et détecter des états particuliers à partir d'une base de données dite « haute résolution » dans le domaine de l'observation marine côtière (fréquence : 20 min.).

L'objectif de cette thèse est d'une part d'évaluer l'applicabilité des techniques d'apprentissage semi-supervisé, afin de déterminer un MMC fidèle à la connaissance existante à la fois en terme de structure et de temporalité et d'autre part proposer un système robuste à des événements/états extrêmes non appris. Ces états nécessitent la définition d'un rejet du système existant [3-4] et la mise en place d'un apprentissage dynamique. Une telle technique pourrait ainsi permettre de traiter un nombre important de données issues de stations instrumentés « haute fréquence » (fréquence infra-horaire) tout en insérant les connaissances acquises depuis des décennies via les approches conventionnelles dites « à basse résolution » (prélèvements mensuels ou bimensuels, rarement hebdomadaires). En Manche orientale, par exemple, il s'agira d'intégrer les données issues de la station MAREL Carnot (Station de mesures instrumentées autonomes, multi paramètres) avec les données des réseaux d'observation comme le REPHY/SRN (Réseaux Phytoplancton & Phycotoxines, et Suivi Régional des Nutriments) afin de mieux comprendre la dynamique phyto-planctonique dans le contexte du développement d'algues nuisibles, en insistant particulièrement sur les effets pressions/impacts (e.g. effets directs et/ou indirects des apports de nutriments) afin de contribuer aux enjeux des directives ou conventions de mers régionales. Cette thèse s'effectuera en collaboration avec l'UL et le laboratoire IFREMER, LER/Boulogne-sur-Mer, qui mettra à disposition les données, la plateforme logicielle et son expertise en écologie marine.

[1] E. Caillault, C. Viard-Gaudin "Mixed discriminant training of hybrid ANN/HMM systems for online handwriting word recognition" in International Journal of Pattern Recognition and Artificial Intelligence(IJPRAI), special issue: Advances in Graphonomics for Handwriting Analysis and Recognition Edited by A. Marcelli and C. De Stefano, Volume: 21, Issue: 1. February 2007, ISSN: 0218-0014

[2] Rousseeuw, K., Poisson-Caillault, E., Lefebvre, A. and Hamad, D. "Hybrid Hidden Markov Model for Marine Environment Monitoring", in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, doi 10.1109/JSTARS.2014.2341219. 20 août 2014

[3] Wacquet G., Poisson Caillault E., Hébert P.A., Semi-Supervised K-Way Spectral Clustering with Détermination of Clusters, *Computational Intelligence, Series Studies in Computational Intelligence, Springer, vol.465, pp. 317-330, isbn={978-3-642-35637-7}, 2013.*

[4] E. Caillault and A. Lefebvre. Towards Chl-a Bloom Understanding by EM-based Unsupervised Event Detection. *Oceans'17 MTS/IEEE conference, Aberdeen, Scotland, June 2017.*