



Ecole Nationale de la
Météorologie
Direction des Etudes
42 avenue Gaspard Coriolis
BP 45712

ECOLE NATIONALE DE LA
METEOROLOGIE

STAGE D'APPROFONDISSEMENT

INGENIEURS DE L'ECOLE NATIONALE DE LA METEOROLOGIE

FICHE DE PROPOSITION DE SUJET

IENM

Thème du sujet :

Utilisation du machine learning pour valoriser un réseau d'instruments opérationnels.
Comparaison par rapport aux meilleures méthodes existantes sur un problème connu.

Titre du sujet proposé :

Machine learning pour l'estimation de hauteur de couche limite sur un réseau de lidars aérosols

Organisme ou service proposant le sujet :

(nom et adresse précise du laboratoire ou du service où se déroulera le stage et où pourront être adressés les courriers du responsable principal et de l'élève).

Météo-France, 42 avenue Gaspard Coriolis, 31000 Toulouse, France.
DSO/DOA/IED
CNRM (UMR3589) GMEI/LISA

Responsable(s) scientifique(s) du stage :

Responsable principal (le responsable principal est l'interlocuteur direct de l'Ecole. C'est à lui, en particulier, que seront adressés les courriers ultérieurs) :

NOM : RIEUTORD
téléphone : 05.61.07.96,53

Prénom : Thomas
Email : thomas.rieutord@meteo.fr

Autres responsables :

NOM(S) : AUBERT
téléphone : 05.61.07.95,60

Prénom(s) : Sylvain
Email : sylvain.aubert@meteo.fr

Le stage peut-il être effectué en "binôme" ?

oui non (rayer la mention inutile)

Ecole Nationale de la Météorologie

Téléphone : 05.61.07.80.80 - Télécopie : 05.61.07.96.30 - <http://www.enm.meteo.fr>
Météo-France, Etablissement Public sous la tutelle du ministère chargé des Transports
L'ENM et Météo-France sont certifiés ISO 9001 par Bureau Veritas Certification



L'étude présente-t-elle un caractère de confidentialité ?

~~non~~ (rayer la mention inutile)

~~oui~~ (à préciser)

(La notion de confidentialité pourra être révisée en cours de stage).

Possibilités éventuelles de logement sur le lieu du stage :

Oui, mais non garantie

(en cas de classement confidentiel, le responsable de cycle en fixera les limites avec l'encadrement : possibilité de soutenance orale publique, restriction de la diffusion et de la consultation du compte-rendu, durée de la confidentialité).

Les rubriques suivantes doivent être obligatoirement explicitées sur feuille séparée en tenant compte des recommandations indiquées pour chacune d'elle.



1) Objectifs scientifiques, techniques poursuivis (une demi-page minimum)

La couche limite atmosphérique est la partie de l'atmosphère en contact avec le sol et qui subit directement son influence, à l'échelle de l'heure. C'est une zone d'intérêt pour divers acteurs (associations de qualité de l'air, modélisateurs, aéronautique légère...) qui se distingue par divers aspects (présence de turbulence, accumulation d'aérosols et de polluants...). Néanmoins, fixer une frontière à cette couche reste un problème car sa définition « par l'intérieur » devient floue à la frontière. La concentration en aérosols est un des traceurs les plus pertinents de la couche limite atmosphérique. Par conséquent, plusieurs équipes de recherche ont concentré leurs efforts sur la détection du sommet de la couche limite depuis les lidars aérosols et/ou les télémètres.

Certains de ces algorithmes commencent à faire consensus dans la communauté de la télédétection : l'algorithme STRAT-2D (SIRTA, MeteoSwiss), l'algorithme CABAM (UK Met Office). Le stage sera l'occasion de déployer ces algorithmes sur le réseau de lidar aérosols opérationnels de Météo-france.

Dans de nombreux domaines, le machine learning et plus généralement l'intelligence artificielle sont sources de progrès spectaculaires (reconnaissance d'objets dans des images, web sémantique, profilage d'individus...). Fortes de ces réussites, ces technologies très polyvalentes laissent penser qu'elles peuvent résoudre des problèmes jusqu'alors difficiles. Le stage aura pour but d'évaluer l'opportunité apportée par de telles techniques pour le problème de détection du sommet de la couche limite.

Pour départager les algorithmes existants et celui utilisant du machine learning, les résultats seront confrontés aux données de radiosondages colocalisés avec les lidars du réseau (sites de Brest et Trappes).

L'objectif du stage sera d'implémenter un algorithme de détection de sommet de couche limite basé sur du machine learning et de le comparer aux algorithmes existants STRAT-2D et CABAM. Les forces et faiblesses de chaque méthode seront évaluées en comparaison avec les radiosondages.

2) Méthodologie envisagée (une demi-page minimum)

En particulier, on devra décrire les outils théoriques nécessaires, la nature dominante du travail proposé (expérimentation, informatique, ...), les moyens de calcul disponibles, la nature et l'état des données éventuelles devant être utilisées.

Au cours du stage, l'étudiant devra s'approprier la problématique de la détection du sommet de la couche limite. Il devra aussi prendre en main les algorithmes existants utilisés dans de nombreux services météorologiques nationaux.

Ce stage nécessitera d'être à l'aise avec la programmation scientifique en langage Python 3.5 (en particulier les bibliothèques Scipy, Numpy et Matplotlib), qui sera l'outil de travail privilégié de l'étudiant retenu. Pour l'élaboration des figures, du rapport et du support présentation, l'étudiant pourra bénéficier d'une formation aux outils Latex, Inkscape et Beamer.

Côté théorie, le stage requiert de la curiosité pour les techniques d'apprentissage automatique, en particulier non-supervisée (K-means, expectation-maximisation, classification hiérarchique...). Là aussi, une bonne place sera laissée à la formation et à l'exploration, le choix de la méthode restant ouvert.

Les données sont issues du réseau de lidar aérosols de Météo-France, opérationnel depuis 2016. Elles seront disponibles sous le format en usage à Météo-France. L'étudiant aura à sa disposition un ordinateur et un accès au calculateur de moyenne performance du CNRM.

3) Plan de travail détaillé mois par mois, pour les cinq mois de stage

Mois 1 : Appropriation du problème, des données, des algorithmes et étude bibliographique.

Mois 2 : Manipulation des données du réseau opérationnel, tests des algorithmes déjà existants.

Mois 3 : Choix et codage de l'algorithme de machine learning.

Mois 4 : Conduite des expériences de comparaison entre les différents algorithmes.

Mois 5 : Rédaction du rapport et élaboration des présentations pour les soutenances IENM



4) Etat d'avancement actuel des travaux sur le sujet, éventuellement lien avec les autres activités de recherche au sein du laboratoire

Les aérosols atmosphériques sont l'objet de nombreuses recherches (au CNRM comme ailleurs) en raison de leur implication dans la formation des nuages (source d'incertitude la plus importante dans les scénarios climatiques du GIEC) et dans la qualité de l'air. Les lidars aérosols permettent de les observer et ils s'imposent dans de nombreux services météorologiques comme des instruments opérationnels de télédétection. Le calcul de la hauteur de couche limite est un sujet de recherche encore actif car ce paramètre sert dans beaucoup de paramétrisations physiques. Il est aussi déterminant pour les modèles de qualité de l'air utilisés par les associations de qualité de l'air (AASQA).

Le machine learning, bien que recouvrant des techniques très variées, est aujourd'hui plébiscité. A Météo-France aussi, l'intérêt pour ces techniques est croissant et leur usage sur des problèmes nouveaux est encouragé. Le problème posé ici semble avoir toutes les qualités pour se prêter à une bonne résolution grâce à ces techniques. En proposant de faire l'implémentation de bout en bout d'un algorithme de machine learning, ce stage participe à la montée en compétence dans ce thème là.

Enfin, l'algorithme choisi à l'issue du stage a vocation à être utilisé sur des données opérationnelles. Il fait donc l'interface entre DSO/DOA et le CNRM.

5) Autres remarques jugées utiles

6) Documents joints à la présente proposition et/ou références bibliographiques sur le sujet

Sur l'algorithme STRAT-2D :

Morille, Y., Haeffelin, M., Drobinski, P., & Pelon, J. (2007). STRAT: An automated algorithm to retrieve the vertical structure of the atmosphere from single-channel lidar data. *Journal of Atmospheric and Oceanic Technology*, 24(5), 761-775.

Haeffelin, M., Angelini, F., Morille, Y., Martucci, G., Frey, S., Gobbi, G. P., ... & Wastine, B. (2012). Evaluation of mixing-height retrievals from automatic profiling lidars and ceilometers in view of future integrated networks in Europe. *Boundary-Layer Meteorology*, 143(1), 49-75.

Sur l'algorithme CABAM :

Kotthaus, S., & Grimmond, C. S. B. (2018). Atmospheric Boundary Layer Characteristics from Ceilometer Measurements Part 1: A new method to track mixed layer height and classify clouds. *Quarterly Journal of the Royal Meteorological Society*.

Sur le machine learning :

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1, No. 10). New York, NY, USA.: Springer series in statistics.

7) Ce sujet pourrait-il être prolongé par une thèse ?

préciser la nature de la prolongation et de ses enjeux.

Non