

## OFFRE DE STAGE

### Détection et correction des anomalies dans les jeux de données de rendement issus de capteurs embarqués sur moissonneuse batteuse

*Type de contrat* : Stage de recherche - Niveau M1 ou M2

*Laboratoire d'accueil* : unité Mixte de recherche ITAP (Irstea/Montpellier SupAgro)

*Début du stage* : mars/avril 2017.

*Durée du stage* : 3 mois à 6 mois

*Localisation* : Montpellier

*Indemnisation* : Environ 500€/mois

*Conditions d'accueil* : Ordinateur à disposition. Possibilité de louer une chambre sur le campus de l'école.

*Contacts* : Corentin Leroux ([cleroux@smag-group.com](mailto:cleroux@smag-group.com)) – Tel : 06 47 04 08 77

### Contexte

Les données de rendement issues des capteurs embarqués sur moissonneuses batteuses (MB) ont été les premières informations disponibles en agriculture de précision (AP). En effet, ces capteurs équipent les MB depuis le début des années 90. Associés à la localisation par satellites (GNSS) des machines, ils permettent de réaliser des cartes de rendement intra-parcellaires à haute résolution spatiale. Ces données de rendement sont toutefois aujourd'hui sous exploitées voire pas utilisées par les professionnels (agriculteurs, conseillers, etc.). Une des raisons qui explique la faible valorisation de ces données est qu'elles arrivent à la fin du cycle de production sur une culture annuelle. Le potentiel d'aide à la décision pour l'année suivante n'est donc pas immédiat et aucune application pratique n'est proposée pour valoriser ces données. Du fait des faibles attentes de la profession, la chaîne de traitement des données de rendement n'a jamais été réellement fiabilisée. Les logiciels édités par les principaux équipementiers ont peu évolué depuis leur conception. Il s'avère que les algorithmes de traitement des données des logiciels proposés ne prennent pas en compte toute l'expertise acquise sur les facteurs susceptibles d'affecter la qualité des données de rendement mesurées. Par exemple :

- La prise en compte des passages lorsque la barre de coupe n'est pas en pleine coupe
- Les changements brusques de vitesse de la moissonneuse batteuse
- Les phénomènes de montée et descente en charge de la moissonneuse en entrée et sortie de parcelles
- Le temps de latence de la moissonneuse (le temps que met le grain à arriver jusqu'au capteur de rendement)
- les données d'humidité aberrantes (par rapport à la distribution des données),

Les données de rendement doivent donc nécessairement être fiabilisées avant de pouvoir être intégrées dans des processus de décisions pour améliorer la gestion des parcelles. Le stage viendra en appui à un travail de thèse actuellement en cours sur le traitement de données de rendement issus de capteurs embarqués sur moissonneuse batteuse. La thèse est le fruit d'une collaboration entre trois entités particulières : DEFISOL ([www.defisol.fr](http://www.defisol.fr)), SMAG (<http://www.smag-group.com/fr>), et l'unité mixte de recherche ITAP (<https://itap.irstea.fr/>). L'encadrement du stage sera assuré par le doctorant. Une méthode générale de pré-traitement de données spatiales a été proposée dans le cadre de la thèse pour fiabiliser les données de rendement mesurées. Pour l'instant, l'approche proposée consiste à identifier les anomalies et à les supprimer.

**Mots clés** : anomalies locales et globales, classification, filtrage, voisinages spatiaux, rendement

## Problématique

La suppression des observations jugées aberrantes dans les jeux de données de rendement entraîne une perte de données qui peut se révéler relativement importante. Plutôt que de supprimer les anomalies identifiées, ne serait-il pas plus pertinent de chercher à les corriger ? La correction de ces données entrainerait-elle un gain significatif sur la connaissance des caractéristiques globales et locales de rendement par rapport à une simple suppression de ces données ?

## Objectif du stage

Le stage proposé aura pour objectifs de :

- Proposer des méthodes pour (i) classifier les anomalies détectées par type d'erreur présente dans les jeux de données de rendement et (ii) corriger spécifiquement chaque type d'anomalie détectée. L'approche proposée dans le cadre de la thèse identifie pour l'instant les anomalies de manière générale mais ne les caractérise pas et donc ne permet pas de les corriger.
- Evaluer le gain de la correction des anomalies par rapport à une méthode plus directe de suppression des données
- Identifier et tester des pistes d'amélioration pour la méthode de détection d'anomalies actuellement proposée.

## Profil attendu

- Etudiant de niveau M1 ou M2 avec des bases solides en analyse de données et calcul scientifique
- Autonomie et curiosité
- Intérêt pour la recherche et le traitement de données
- Aisance à l'anglais, surtout pour la lecture d'articles scientifiques

## Compétences recherchées

- Analyse de données et traitements statistiques
- Calcul scientifique (la maîtrise de R sera particulièrement appréciée)
- Des connaissances sur le traitement de données spatialisées seront appréciées.
- La maîtrise de systèmes d'information géographique (notamment QGIS) serait un plus

## Candidature

CV et lettre de motivation à envoyer par mail à [cleroux@smag-group.com](mailto:cleroux@smag-group.com) avant le 03/03/2017

## Références

- Arslan, S., & Colvin, T. (2002). Grain yield mapping : yield sensing, yield reconstruction, and errors. CEUR Workshop Proceedings, 1225, 41–42.
- Chen, D., Lu, C-T., Kou, Y. & Chen, F. (2008). On Detecting Spatial Outliers. Geoinformatica, 12, 455-475
- Filzmoser, P., Ruiz-Gazen, A. & Thomas-Agnan, C. (2014). Identification of local multivariate outliers. Statistical Papers, 55(1), 29-47.
- Griffin, T., Dobbins, C., Vyn, T., Florax, R., & Lowenberg-DeBoer, J. (2008). Spatial analysis of yield monitor data: case studies of on-farm trials and farm management decision making. Precision Agriculture, 9(5), 269–283
- Lyle, G., Bryan, B., & Ostendorf, B. (2013). Post-processing methods to eliminate erroneous grain yield measurements: review and directions for future development. Precision Agriculture, 1–26.
- Simbahan, G.C., Dobermann, A., & Ping, J.L. (2004). Screening yield monitor data improves grain yield maps. Agronomy Journal, 96, 1091-1102
- Sudduth, K., & Drummond, S. T. (2007). Yield Editor : Software for Removing Errors from Crop Yield Maps. Agronomy Journal, 99(6), 1471.