

DATASCIENCE ET CAUSALITÉ : EXEMPLE DES FORÊTS ALÉATOIRES

par Emmanuelle **GOURDAIN** (Arvalis-Institut du végétal)

Dans le milieu agricole, les usages liés aux méthodes de « Datasciences » restent encore du domaine de la recherche et du développement (R&D) mais les résultats sont prometteurs. Ils mettent en évidence les potentialités de ces méthodes pour la détection et la reconnaissance de plantes ou de maladies à partir de photos, l'identification des stades de développements ou encore l'estimation des rendements. Ces méthodes, couplées à des robots, des machines agricoles ou des outils d'aide à la décision, devraient engendrer un saut technologique important. Les instituts techniques s'emparent largement de ces méthodes pour accélérer l'acquisition de références et le développement de modèles prédictifs.

La facilité d'usage des méthodes de Datascience à travers des logiciels comme R ont largement démocratisés les usages par les agronomes. Cependant, de plus en plus, ce type de modèle est utilisé également dans un but explicatif, en vue de mettre en évidence des relations de cause à effet entre variables. En particulier, les forêts aléatoires semblent de bonnes candidates pour expliquer un phénomène en proposant des sorties comme l'importance des variables et les partials plot, qui peuvent être interprétées dans but explicatif.

Même si les méthodes de l'intelligence artificielle apportent une aide importante pour analyser ces grandes quantités de données qu'il serait impossible d'appréhender par l'homme en un temps raisonnable et pour faire des prédictions, leur usage dans un but explicatif, à savoir mettre en évidence des relations causales entre variables, reste à démontrer.

À notre connaissance, seuls les principes de base de l'expérimentation (randomisation, répétition, blocking) offrent le cadre rigoureux permettant la mise en évidence de relations de cause à effet. Afin d'avancer sur ces questionnements, ARVALIS a entrepris des travaux de recherche qui tendent à montrer qu'il est loin d'être évident que ces méthodes identifient des liens causaux.

Mini cv :

Emmanuelle GOURDAIN (e.gourdain@arvalis.fr)

Après une formation en agronomie, j'ai rejoint ARVALIS-Institut du végétal en 2004 en tant que chef de projet sur la qualité sanitaire des céréales et plus particulièrement sur la contamination des grains de blé par les mycotoxines. Avec l'ambition d'aider les agriculteurs à réduire les risques en cours de campagne, je me suis orientée en 2011 vers la modélisation du complexe fusarien puis, plus largement vers la modélisation des pathosystèmes.

En 2016, j'ai pris l'animation de l'équipe de modélisation et aide à la décision dont la mission principale est de développer des modèles prédictifs performants en vue d'être déployer dans

des outils d'aide à la décision tactique sur la protection contre les maladies, les apports d'azote, etc.

En 2019, j'élargis mon spectre d'action en prenant la direction du service Innovations Digitales, Méthodologiques et Matériels d'expérimentation, qui mène des activités transversales autour des statistiques, de la modélisation et des calculs scientifiques, de la gestion des données et du matériel d'expérimentation. Cette unité regroupe une vingtaine de collaborateurs.