

Le Big Data en agriculture

Par Véronique BELLON-MAUREL

ITAP, Irstea, Montpellier Supagro, Université de Montpellier

Pascal NEVEU

MISTEA, INRA, Montpellier Supagro, Université de Montpellier

Alexandre TERMIER

Université de Rennes, Inria, CNRS, IRISA

et Frédérick GARCIA ⁽¹⁾

MIAT, UR875, Université de Toulouse, INRA

Les travaux sur le Big Data agricole sont tous très récents, et encore peu nombreux puisqu'on n'en dénombre que quelques dizaines (Kamilaris *et al.*, 2017 ; Wolfert *et al.*, 2017). Cependant, même si l'adoption est plus lente que dans d'autres secteurs, il n'en demeure pas moins que le potentiel du Big Data en agriculture est énorme, tout autant que les enjeux et les challenges à relever.

Les enjeux du Big Data en agriculture

L'agriculture vit des transformations majeures qui lui imposent de s'adapter. Le changement climatique et la dégradation des sols font rapidement évoluer les conditions pédoclimatiques et l'agriculteur ne maîtrise plus son outil de production. Les pouvoirs publics et les consommateurs ont des exigences nouvelles par rapport à la production agricole : agro-écologie, bien-être animal, arrêt du glyphosate... Tout ceci impose un changement de modèle agricole, qui doit être rapide mais qui est d'autant plus difficile que les agriculteurs ne connaissent pas forcément les pratiques à mettre en œuvre, et que les référentiels n'existent pas encore, par manque de recherche. Il est donc indispensable de construire rapidement une connaissance pour accompagner l'agriculteur dans un processus de production toujours plus complexe, combinant recherche de compétitivité, respect des réglementations, recherche d'une meilleure valorisation financière et évolution des pratiques. Nous faisons l'hypothèse que les données en lien avec la production agricole sont sur une trajectoire de massification qui permettra de constituer des méga-données – ou Big Data – et peuvent être utilisées dans des analyses de Big Data pour fournir à l'agriculteur des outils d'aide à la décision.

Les enjeux du Big Data agricole sont de plusieurs ordres : ils touchent la production agricole et la commercialisation, mais également les pouvoirs publics pour la gestion de l'espace et la mise en place de politiques publiques. Concernant la production agricole, l'analyse des données massives est utile à la fois dans les choix stratégiques de l'exploitation (quelles espèces ? quelles rotations ?) et dans les choix tactiques (interventions dans l'itinéraire technique). Des modèles de fonctionnement plus précis, car intégrant mieux les conditions locales, sont déduits de l'analyse, en particulier pour la gestion des intrants ; les semenciers parlent de « semis prescriptif », à savoir de conseils sur la densité des semis et les variétés à semer, un service qu'ils envisagent de vendre aux agriculteurs. De la même manière, pour les animaux, on pourra mieux connaître la manière dont chaque individu s'alimente ou supporte des stress, et ainsi ajuster son alimentation ou les interventions. Ainsi, dans ces domaines de l'agriculture et de l'élevage de précision, des modèles de fonctionnement plus

(1) Les auteurs appartiennent à l'Institut Convergences Agriculture Numérique #DigitAg, qui bénéficie d'une aide de l'État gérée par l'Agence nationale de la Recherche au titre du programme d'Investissements d'avenir portant la référence ANR-16-CONV-0004.

précis peuvent être inférés de l'analyse du Big Data. L'aspect prédictif des connaissances issues du Big Data est aussi un enjeu. Par exemple, dans certains domaines dans lesquels la connaissance est limitée, comme pour certaines approches de l'agro-écologie ou de l'agroforesterie, il s'agit de mettre en place des expériences, y compris chez les agriculteurs, dans des conditions pédoclimatiques très variées, pour en extraire les règles agronomiques de fonctionnement de ces nouveaux écosystèmes et pouvoir conseiller les agriculteurs sur ces nouveaux itinéraires culturaux. Le numérique facilite cette approche de *crowdsourcing* et de sciences participatives (Minet *et al.*, 2017), ce qui permettra aussi une meilleure diffusion des innovations agricoles. Du côté des pouvoirs publics, l'analyse des données massives permet d'estimer les rendements des cultures, information précieuse pour garantir la sécurité alimentaire, de suivre les épisodes épidémiques et de prendre des mesures sanitaires. Enfin, dans le cadre des services d'assurance indicielle, qui s'avèrent être bénéfiques aux petits agriculteurs, le Big Data est le socle indispensable pour construire des indices effectifs apportant une réelle protection aux agriculteurs ; cela justifie pleinement l'augmentation des investissements publics dans la collecte des données qui peuvent faciliter l'expansion des marchés de l'assurance indicielle (Castillo *et al.*, 2016).

Les enjeux étant posés, intéressons-nous aux spécificités du Big Data agricole.

Le Big Data agricole, de la donnée à l'usage

Les sources de données du Big Data agricole

Les sources de données du Big Data agricole sont multiples et prolifèrent. Les premières sont les images satellitaires, dispositifs qui datent de plus de quarante ans, mais qui connaissent aujourd'hui un déploiement inédit avec le lancement de nouveaux satellites. Par exemple, la constellation Sentinel délivre des images gratuitement à une très haute fréquence temporelle (tous les cinq jours), ce qui offre des opportunités totalement nouvelles à la recherche et aux entreprises, en permettant par exemple de fonder un conseil agronomique. À cela s'ajoutent des dispositifs de mutualisation comme le dispositif GeoSud THEIA (<https://www.theia-land.fr/fr/projets/geosud>), qui permet l'accès à moindre coût à des images satellitaires plus sophistiquées, multispectrales et susceptibles d'être programmées. Un autre facteur de prolifération des données est la multiplication des objets connectés en agriculture (Tzounis *et al.*, 2017) : stations météorologiques, pièges à insectes, capteurs d'humidité du sol et compteurs d'eau connectés en irrigation, mais aussi divers capteurs installés sur les animaux pour évaluer leur état (état de santé, présence de chaleurs...), sur les robots de traite (quantité et qualité du lait) ou sur les automates d'alimentation. Les agro-équipements sont aussi de plus en plus fréquemment équipés de capteurs, pour l'agriculture de précision (nécessité de connaître les besoins de la plante pour lui apporter exactement ce dont elle a besoin), mais aussi pour la maintenance prédictive. Les exigences de traçabilité des productions agricoles se traduisent aujourd'hui par des systèmes de lecture automatisée, avec les puces RFID, ou par la saisie manuelle des interventions agricoles à partir de smartphones avec transmission directe vers les applications des éditeurs de logiciels, ce qui remplace les cahiers de culture en papier. Le défi est d'automatiser l'acquisition de données afin qu'elle n'ait virtuellement aucun coût (Wolfert *et al.*, 2017). Enfin, les méthodes de phénotypage rapide, incontournables pour raccourcir le cycle de production de nouvelles semences, sont également sources de données massives à mettre en relation avec les données génotypiques (Halewood *et al.*, 2018). La section suivante montre en quoi ces données massives peuvent être qualifiées de Big Data.

Les 6 « V » du Big Data agricole

Dans le domaine des sciences de données, le Big Data se caractérise par les «V» et le Big Data de l'agriculture n'échappe pas à cette règle mais avec certaines spécificités.

- Le «Volume» toujours croissant des données est lié au développement des technologies, à l'usage du web et, au-delà de la captation de la donnée, à la réduction des coûts de stockage de l'information qui permet de la réutiliser.
- La «Variété» est la caractéristique forte du Big Data agricole. Au-delà des différentes natures des informations (texte, mesure, image, spectre, simulation...), le domaine de l'agriculture exige de prendre en compte des sources de données nombreuses et extrêmement diverses, provenant :
 - des échelles allant du territoire à la génomique en passant par différents niveaux tels que l'exploitation, la population ou l'individu (plante ou animal) ;
 - des informations spatiales et cartographiques pouvant décrire des zones, des trajectoires, des opérations culturales ou des sites hétérogènes d'acquisition ; des interactions et échanges qui se produisent à toutes ces échelles avec l'environnement (sol, eau, climat, bioagresseurs...) ;
 - des évolutions (stades de croissance, modes de conduite) ;
 - des transformations qui vont de la ferme à l'assiette ;
 - des impacts sociaux, économiques, environnementaux, sanitaires.

Au-delà des données, c'est la variété des acteurs, qui les produisent et qui les manipulent, qui est en soi un défi en agriculture. Ces acteurs (agronomes, généticiens, biologistes, statisticiens, industriels, distributeurs ...) ont par essence des vocabulaires, des sémantiques et des modes de production de données très hétérogènes. Cette forte variété rend difficile l'intégration des données. Cela demande donc de nouvelles approches pour gérer et analyser les données en prenant en compte les dimensions spatiales, temporelles mais aussi organisationnelles (filiales, circuits courts, normes sanitaires).

- La «Vélocité» dans l'agriculture correspond à de nouvelles générations d'outils pour une aide à la décision où la vitesse d'acquisition des données et la vitesse pour fournir un résultat aux utilisateurs sont cruciales. Par exemple, ces outils peuvent être liés à l'exploitation d'un grand nombre de capteurs équipant des animaux, à l'utilisation du *crowdsourcing*, aux plateformes de phénotypage à haut-débit, ou encore à l'agriculture de précision. Un des enjeux est de pouvoir gérer et analyser de grandes masses de données en temps réel.
- D'autres «V» peuvent être mentionnés. La «Véracité» et la «Validité» qui caractérisent la qualité de sources de données ; la «Visualisation» qui est une problématique pour de grands ensembles de données complexes ; la «Visibilité» de faits ou d'éléments pertinents qu'il faut extraire de gros volumes de données ; enfin l'analyse de toutes ces données peut poser des problèmes de respect de la «Vie privée» d'exploitants agricoles.

Enfin, le véritable enjeu du Big Data, c'est la Valeur qu'on peut en obtenir. Cet enjeu dépend de notre capacité à structurer le Big Data. Ceci permettra de réutiliser et d'agréger des données de différentes façons mais aussi de pouvoir les lier avec d'autres données. Autrement dit, l'innovation autour du Big Data agricole sera indispensable pour créer de la valeur et cela passe par la structuration des données et le développement de nouvelles approches d'analyse.

L'analyse du Big Data agricole

L'analyse peut avoir pour but de fournir aux différents utilisateurs, comme les agriculteurs ou les conseillers agricoles, une aide pour prendre la bonne décision au bon moment, par des indicateurs, des alertes ou des explications. Elle peut aussi avoir pour but de découvrir de nouvelles connaissances, aussi bien générales, par la masse des données collectées, que localisées sur une exploitation, une parcelle ou un animal particulier grâce à un suivi fin. Un domaine historique est l'analyse d'images, que ce soient des images satellitaires (ex : comprendre l'utilisation des sols et la dynamique des cultures) ou des images prises depuis un aéronef ou un drone (ex : suivre l'évolution de la croissance des feuilles et les éventuelles maladies). Des approches d'apprentissage automatique

ou de fouille de données peuvent être appliquées pour corrélérer les grandeurs mesurées avec des variables d'intérêt ou effectuer des prédictions (Kamilaris *et al.*, 2017). Par exemple, dans les exploitations laitières, il est désormais possible de suivre la température d'un animal et son activité au long de la journée : un enjeu est d'analyser ces données pour prédire au plus tôt les périodes de fertilité, ou les maladies (Steenefeld *et al.*, 2015).

Une difficulté est l'adaptation des méthodes de fouille de données aux caractéristiques multi-scalaires (temporelles ou spatiales) des données agricoles. D'autres difficultés sont la gestion de l'incertitude dans les données et la réduction du nombre de paramètres dans les méthodes utilisées afin de simplifier leur utilisation.

Les acteurs du Big Data agricole

Aujourd'hui de nombreux acteurs se positionnent : les agro-équipementiers avec leurs outils connectés, l'agrofourniture qui ambitionne de vendre des services plutôt que des intrants, les grosses coopératives qui génèrent de la donnée, mais aussi des acteurs exogènes au monde agricole comme des financiers (capital-risqueurs) et des spécialistes de la gestion de données par les technologies *Cloud* (Google, IBM, Fujitsu...) (Wolfert *et al.*, 2017). Par exemple Google Venture a investi 15 millions de dollars dans une société capable de mettre en œuvre du Big Data pour aider les exploitants à mieux vendre leurs produits⁽²⁾. De nouveaux jeux d'acteurs et de nouvelles relations de pouvoir se mettent en place dans la chaîne de valeur de l'agro-alimentaire, au travers de la maîtrise du Big Data. Les pouvoirs publics doivent jouer un rôle dans cet espace : mettre à disposition des infrastructures, favoriser l'égalité des territoires par la connectivité, veiller à éviter les situations de monopoles...

Conclusion

Malgré l'importance des enjeux en termes de développement et viabilité économiques, l'adoption des innovations liées au numérique et au Big Data par les acteurs du monde agricole reste freinée par de nombreuses barrières (Kamilaris *et al.*, 2017), dont les principales sont techniques : manque d'expertise et de ressources humaines dans le domaine des technologies de l'information appliquées à l'agriculture, offre limitée et caractère difficilement accessible des infrastructures de données offrant à la fois des services de stockage de l'information, mais également d'analyse des Big Data, et, lorsque ces offres existent, difficulté de coupler entre elles ces différentes sources de données, du fait de leur grande hétérogénéité (le « V » de Variété) et de l'absence d'ontologies largement partagées. Par ailleurs, le développement du Big Data en agriculture est également freiné par l'absence d'un modèle économique clair et attractif pour l'ensemble de la chaîne de la donnée, intégrant les agriculteurs, les transformateurs, les distributeurs, ainsi que les consommateurs, les citoyens et les services publics également dans un rôle de producteurs de données. Il s'agit d'inventer ici un fonctionnement économique en accord avec les principes actuellement mis en place en France et Europe autour du droit, de la propriété intellectuelle et de l'ouverture des données et, plus globalement, des produits numériques qui autorise un partage équitable de la plus-value créée par le Big Data et qui favorise la création et le développement d'innovations numériques en termes d'outils et d'usages pour le domaine agricole.

Au-delà de ces obstacles, il est également nécessaire de prévenir certains risques associés au développement du numérique en agriculture, comme la perte potentielle d'autonomie des agriculteurs avec la possible hyper-technologisation du domaine, et l'hypercentralisation de la chaîne de la donnée par quelques acteurs globaux, ou encore l'augmentation de la fracture numérique entre les agricultures des pays développés et en développement.

(2) <https://venturebeat.com/2015/05/19/google-ventures-leads-15m-investment-in-big-data-for-farmers/>

Tous ces sujets, techniques et sociaux, font aujourd’hui l’objet de nombreuses recherches, en particulier au sein de l’Institut Convergences Agriculture Numérique #DigitAg.

Références

- CASTILLO M.J., BOUCHER S. and CARTER M. (2016), “Index Insurance: Using Public Data to Benefit Small – Scale Agriculture”, *International Food and Agribusiness Management Review*, Special Issue – Vol.19 Issue A, 93-114.
- COOPER J., NOON M., JONES C., KAHN E. and ARBUCKLE P. (2013) “Big Data in Life Cycle Assessment”, *J Ind. Ecology* 17 (6), 796-799.
- HALEWOOD M., CHIURUGWI T., SACKVILLE HAMILTON R., KURTZ B., MARDEN E., WELCH E., MICHIELS F., MOZAFARI J., SABRAN M., PATRON N., KERSEY P., BASTOW R., SHAWN DORIUS S., DIAS S., McCOUCH S. and POWELL W. (2017), “Plant genetic resources for food and agriculture: opportunities and challenges emerging from the science and information technology revolution”, *New Phytologist* 217, 1407-1419.
- KAMILARIS A., KARTAKOULLIS A. and PRENAFETA-BOLDU F.X. (2017), “A review on the practices of big data analysis in agriculture”, *Computers and Electronics in Agriculture*. 143, 23-37.
- MINET J., CURNEL Y., GOBIN A., GOFFART J.P., MÉLARD F., TYCHON B., WELLENS J. and DEFOURNY P. (2017), “Crowdsourcing for agricultural applications: A review of uses and opportunities for a farmsourcing approach”, *Computers and Electronics in Agriculture* 142, 126-138.
- STEENEVELD W. and HOGVEEN H. (2015), “Characterization of Dutch dairy farms using sensor systems for cow management”, *Journal of Dairy Science*, 709-717.
- TZOUNIS A., KATSOULAS N., BARTZANAS, T. and KITTAS C. (2017), “Internet of things in agriculture, recent advances and future challenges”, *Biosystems Engineering*, 164, 31-48.
- WOLFERT S., GE L., VEROUW C. and BOGAARDT M.J. (2017), “Big Data in smart farming – A review”, *Agricultural Systems*, 153, 69-80.