Big data in agriculture

Véronique Bellon-Maurel
and **Pascal Neveu**,
Alexandre Termier
& **Frédérick Garcia**,
Institut Convergences Agriculture Numérique #DigitAg

Abstract:
In the chain of agricultural production, data are proliferating as a result of a “technology-push” (connected devices, smartphones, sensors, precision agriculture, robots, drones, new satellites, plant detection, geolocaction systems, Internet connections and high-speed data transmission in rural areas, the social media, etc.) and a “market-pull” (tracking, traceability, decision-making systems, risk management, computerization of the paperwork required by the Common Agricultural Policy, etc.). Farming is now undergoing fast, exogenous changes (the climate, consumer demand, lower income, etc.) that force it to adapt by relying on new knowledge and on a granular optimization of the processes of production and sales. Big data, in association with deep learning, can make new knowledge emerge.

Studies on big data in agriculture are all very recent, and not yet numerous, a few dozen (KAMILARIS *et al.*, 2017, WOLFERT *et al.*, 2017). Although the adoption of big data is slower than in other sectors, their potential for agriculture is as immense as the issues and challenges spawned by them.[1]

Big data in agriculture.

The agricultural sector is undergoing major changes that are forcing it to adapt. Climate change and soil degradation are quickly altering climatic and pedological conditions; and farmers no longer control their production equipment. Public authorities and consumers have new demands with regard to agriculture: agroecology, animal well-being, pressures to stop using glyphosate, etc. All this requires changing the model of agriculture — a change that has to be rapid but is all the harder insofar as farmers are not necessarily familiar with the practices to adopt. Furthermore, the benchmarks for these practices do not yet exist owing to a lack of research. It is, therefore, indispensable to quickly build up our stock of knowledge in order to assist farmers in an ever more complicated production process that involves, all at once, trying to be more competitive, to obey regulations, to improve the farm’s financial statement and to follow up on changing practices. Our hypothesis is that the stock of data related to production in the agricultural sector is swelling to the point of becoming megadata (or big data) on which data analytics can be used to offer farmers decision-making tools.

[1] This article, has been translated from French by Noal Mellott (Omaha Beach, France).

The issues related to big data in agriculture are of several sorts. They affect production and marketing, but also involve public authorities for land management and policy-making.

With relation to production, data analytics is useful for making strategic decisions for the farm (Which crops? How to rotate them?) and tactical choices (technical interventions and maintenance). From such analyses, improved operational models can be inferred that take better account of local conditions, in particular for managing entrants. Seed companies talk about "prescriptive sowing", *i.e.*, advice on the density of sowing and the varieties to be sowed, a service they imagine selling to farmers. Likewise, we can learn how each individual animal eats or bears up under stress, and thus adjust its feed or other interventions. In precision agriculture for example, operational models can be inferred from the analysis of big data. The predictive possibilities of the knowledge drawn from big data are also important. In matters with limited knowledge, as in certain areas of agroecology or agroforestry, experiments can be conducted, including on farms, under various pedological and climatic conditions, in order to extract agronomic rules about the function of these new ecosystems and to provide advice to farmers on new crops itineraries. Information and communications technology (ICT) has opened the way toward crowdsourcing and citizen science (MINET *et al.* 2017), which will help diffuse innovations in agriculture.

For public authorities, the analysis of big data can serve to estimate yields (a valuable information for securing the food supply), monitor epidemics and adopt measures.

For index-based insurance, which has proven beneficial to small farmers, big data are indispensable for constructing the indices that provide real protection to farmers. This argument fully justifies increasing public investments in collecting the data that can help expand index insurance markets (CASTILLO *et al*. 2016).

Having described the issues, let us now examine how big data are put to use in agriculture.

The sources of data

There are many sources of big data for agriculture, and many more are emerging.

A first source is satellite photography, which dates back more than forty years but is now proliferating as never before thanks to the launching of new satellites. For example, the Sentinel constellation is delivering for free images at a very high frequency (every five days). This opens completely new opportunities for research and for firms (by providing information for formulating agronomic advice). In addition, there are projects like GeoSud THEIA, which pools efforts for providing access at a low cost to sophisticated satellite imagery (multispectral and programmable).[2]

The proliferation of connected devices in agriculture is another factor contributing to this abundance of data (TZOUNIS *et al*. 2017): weather stations and insect traps, as well as soil humidity sensors and connected water meters for irrigation, not to mention various sensors on animals (to assess their state of health or detect whether they are in heat), on milking robots (to measure the milk's quantity and quality) and on feeder robots. Farm machines and implements are increasingly equipped with sensors for precision agriculture (to know a plant's needs so as to provide exactly what it needs) and, too, for predictive maintenance. The traceability of produce can be done by using systems for automatically reading radio-frequency identification (RFID) tags or for manually entering data via a smartphone that are then directly transmitted to software applications (thus replacing notebooks and sheets of paper). The challenge is to automate data collection so that it will cost virtually nothing (WOLFERT *et al*. 2017).

Finally, real-time phenotyping, indispensable for shortening the production cycle of new crops, is also a source of massive data that can be related to data on genotypes (HALEWOOD *et al*. 2018).

[2] https://www.theia-land.fr/en/projects/geosud.

The six *v*'s of big data in agriculture

The data sciences describe big data in terms of four *v*'s. Big data in agriculture falls under this description but with a few traits of its own.

The ever growing VOLUME of data is tied to technological developments, the use of the Web and, once the data are captured, the reduction of the costs of storing them for reuse.

VARIETY is a major trait of big data in agriculture. Not only are the data varied (text, measurements, images, simulations, etc.), but also agriculture requires tapping numerous and varied sources of data:

- data on various scales ranging from local areas to genomics and from the population (plant or animal) to the individual;
- geographical and cartographical information describing zones, itineraries, growing conditions and locations as well as the interactions and exchanges that take place at all levels of the environment (soil, water, climate, pests, etc.);
- trends, stages of growth, behavior patterns;
- food processing from the farm to the consumer's plate; and
- social, economic, environmental and health effects.

Besides the variety of data, there is, for agriculture, the variety of the parties who produce and manipulate the data. Agronomists, geneticists, biologists, statisticians, manufacturers, distributors, etc. have, by nature, quite different vocabularies, semantics and methods of data production; and this makes it hard to consolidate data. To mine and analyze data, new approaches must be worked out that take into account the dimensions not only of space and time but also of organizations (supply chains, direct distribution, health standards) (COOPER *et al*. 2013).

In agriculture, VELOCITY corresponds to a new generation of tools for assisted decision-making. The speed of acquiring data and of providing results to farmers is crucial. These tools might, for example, be hooked up to a large number of sensors on animals, or to platforms for crowdsourcing, high-speed phenotyping or precision agriculture. At stake are the management and analysis of large masses of data in real time.

Other *v*'s might be mentioned. VERACITY and VALIDITY, related to the quality of data sources. VISUALIZATION is a problem in the case of big sets of complex data. Also important is the VISIBILITY of the relevant facts or elements extracted from big data, or the VIOLATION OF PRIVACY that might occur when analyzing big data.

Finally, there is the VALUE to be gleaned from big data. At stake here is our ability to structure big data so as to aggregate them in different ways, reuse them and, too, link them to other data. Innovations involving big data are indispensable to create value in agriculture. This means curating and consolidating the data, and developing new analytics.

Data analytics in agriculture

Data analytics in agriculture might have the goal of assisting users, whether farmers or extension workers, with making the right decision at the right time with the help of indicators, alerts or explanations. Another goal might be to discover new knowledge, whether general (owing to the mass of fine-grained data collected) or local (specific to a farm, a field or an animal, thanks to monitoring). The analysis of images from satellites (*e.g.*, to know how the land is used and crops are planted) or drones (*e.g.*, to monitor leaf growth or pathologies) has already been put to use. Automatic learning procedures or data mining can help correlate measurements with variables for making predictions (KAMILARIS *et al*. 2017). In dairy farms for example, it is now possible to monitor an animal's temperature and activities throughout the day, but the problem is to analyze these data in order to predict as soon as possible periods of fertility, or pathologies (STEENEVELD *et al*. 2015).

A difficulty is to adapt data-mining methods to the multiscale nature (time or space) of agricultural data. Other difficulties have to do with managing the uncertainty of data and reducing the number of parameters so as to make it simpler to use these methods.

The economic agents active in big data in agriculture

Nowadays, several operatives are staking out positions: farm implement businesses providing connected devices, farm supply businesses who want to sell services rather than entrants, the big farming cooperatives that generate data and, too, players outside the farm world, such as financiers (venture capitalists) and specialists in data management using cloud-related technology (Google, IBM, Fujitsu…) (WOLFERT *et al*. 2017). For instance, Google Venture has invested $15 million in a big data firm capable of helping farmers sell their produce.[3]

The game is changing, and forces are, through their control over big data, realigning along the value chain in agribusiness. Public authorities must play a role in all this by seeing to it that infrastructures are made available, that local areas have equal access in terms of connectivity, that monopolistic situations are staved off…

Conclusion

Despite the stakes in terms of development and economic viability, several barriers are still keeping stakeholders in the agricultural sector from adopting innovations related to ICT and big data (KAMILARIS *et al.* 2017). The principal barriers are technical. For one thing, there is a lack of expertise and human resources for applying ICT to agriculture. For another, there are not many data platforms (and accessing them is not easy) that provide services for both storing and analyzing big data. When such services are offered, it is hard to couple various data sources, given the wide variety (one of the *v*'s) of the data and the absence of widely shared ontologies. Another impediment to developing big data in agriculture is the absence of a clear business model attractive to players all along the data chain: farmers, food-processors, distributors and even the consumers, citizens and public services that also produce data. The objective is to invent in agriculture an economy in compliance with the principles currently established in France and Europe in the laws on intellectual property rights, the opening of data and, in general, digital technology — this economy should fairly share the value added by big data among stakeholders and boost the creation and development of digital innovations, tools and uses.

Beyond these barriers, it is also necessary to ward off certain risks stemming from the development of digital technology in agriculture, such as farmers potentially losing their autonomy owing to "overtechnologization", as the chain of data comes to be overcentralized in the hands of a few global firms, or as the digital gap widens between farmers in developed and developing countries.

All these technical and social topics are now under research, in particular at the Institut Convergences Agriculture Numérique #DigitAg.[4]

[3] https://venturebeat.com/2015/05/19/google-ventures-leads-15m-investment-in-big-data-for-farmers/
[4] http://www.hdigitag.fr/fr/

References

CASTILLO M.J., BOUCHER S. & CARTER M. (2016) "Index insurance: Using public data to benefit small-scale agriculture", *International Food and Agribusiness Management Review*, 19(A), pp. 93-114.

COOPER J., NOON M., JONES C., KAHN E. & ARBUCKLE P. (2013) "Big data in life cycle assessment", *Journal of Industrial Ecology*, 17(6), pp. 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. & POWELL D.W. (2017) "Plant genetic resources for food and agriculture: opportunities and challenges emerging from the science and information technology revolution", *New Phytologist*, 217, pp. 1407-1419.

KAMILARIS A., KARTAKOULLIS A. & PRENAFETA-BOLDU F.X. (2017) "A review on the practices of big data analysis in agriculture", *Computers and Electronics in Agriculture*, 143, pp. 23-37.

MINET J., CURNEL Y., GOBIN A., GOFFART J.P., MÉLARD F., TYCHON B., WELLENS J. & DEFOURNY P. (2017) "Crowdsourcing for agricultural applications: A review of uses and opportunities for a farmsourcing approach", *Computers and Electronics in Agriculture* 142, pp. 126-138.

STEENEVELD W. & HOGEVEEN H. (2015) "Characterization of Dutch dairy farms using sensor systems for cow management", *Journal of Dairy Science*, pp. 709-717.

TZOUNIS A., KATSOULAS N., BARTZANAS T. & KITTAS C. (2017) "Internet of things in agriculture, recent advances and future challenges", *Biosystems Engineering*, 164, pp. 31-48.

WOLFERT S., GE L., VERDOUW C. & BOGAARDT M.J. (2017) "Big data in smart farming, a review", *Agricultural Systems*, 153, pp. 69-80.