Observational studies in agriculture: paradigm shift required

Longchamps, L., Tremblay, N., Panneton, B.

Agriculture and Agri-Food Canada, CRD St-Jean-sur-Richelieu, QC, Canada

A paper from the Proceedings of the
14th International Conference on Precision Agriculture
June 24 – June 27, 2018
Montreal, Quebec, Canada

Abstract. *There is a knowledge gap in agriculture. For instance, there is no way to tell with precision what is the outcome of cutting N fertilizer by a quarter on important outcomes such as yield, net return, greenhouse gas emissions or groundwater pollution. Traditionally, the way to generate knowledge in agriculture has been to conduct research with the experimental method where experiments are conducted in a controlled environment with trials replicated in space and time. While this method has proven its potential to generate knowledge, it has also shown limitations in terms of speed and amount of resources required. Indeed, at the current pace of agricultural impacts on the biosphere, it is likely that traditional experimental research won't be able to generate the knowledge required in a timely fashion. A paradigm shift is needed to shorten the time between the detection of a problem and the access to a reliable solution. One possible avenue is to use another scientific approach, notably the observational method, which relies on a large number of observations to draw conclusions. The advent of communication and information technologies in agriculture opens new possibilities, notably to conduct observational research with big datasets. By observing farm inputs and outputs contextualized with soil, climate, and weather data, there is a tremendous potential to improve farm input use efficiency by adjusting prescriptions to each and every location in every field of every farm. It is most likely that to keep up with the rapid pace of agricultural impacts on the environment, observational science needs to be implemented at a global scale.*

Keywords. *Observational studies, precision agriculture, big data, artificial intelligence.*

Introduction

Global agricultural crisis

The agricultural sector is facing a global multifold crisis impacting, among other factors: biodiversity, climate change, food production, soil degradation, water pollution and human health (Mazoyer and Roudart 2002). Farmers around the world are facing the colossal task of providing food, fuel and fiber in a sustainable way to a growing population while challenged to maintain profitability of their businesses. Agriculture is being performed with an industrial approach while it is taking place in nature, with all the variability and uncertainty implicit to natural environments. In other industrial production sectors, it is possible to evaluate the impacts of management on production with a high level of precision, thus enabling optimization of input use. In agriculture, farmers currently don't have the necessary tools to predict, for example, how reducing N fertilizer rate by one quarter will affect yield, net revenue, GHG emissions or ground water pollution (Fig. 1).

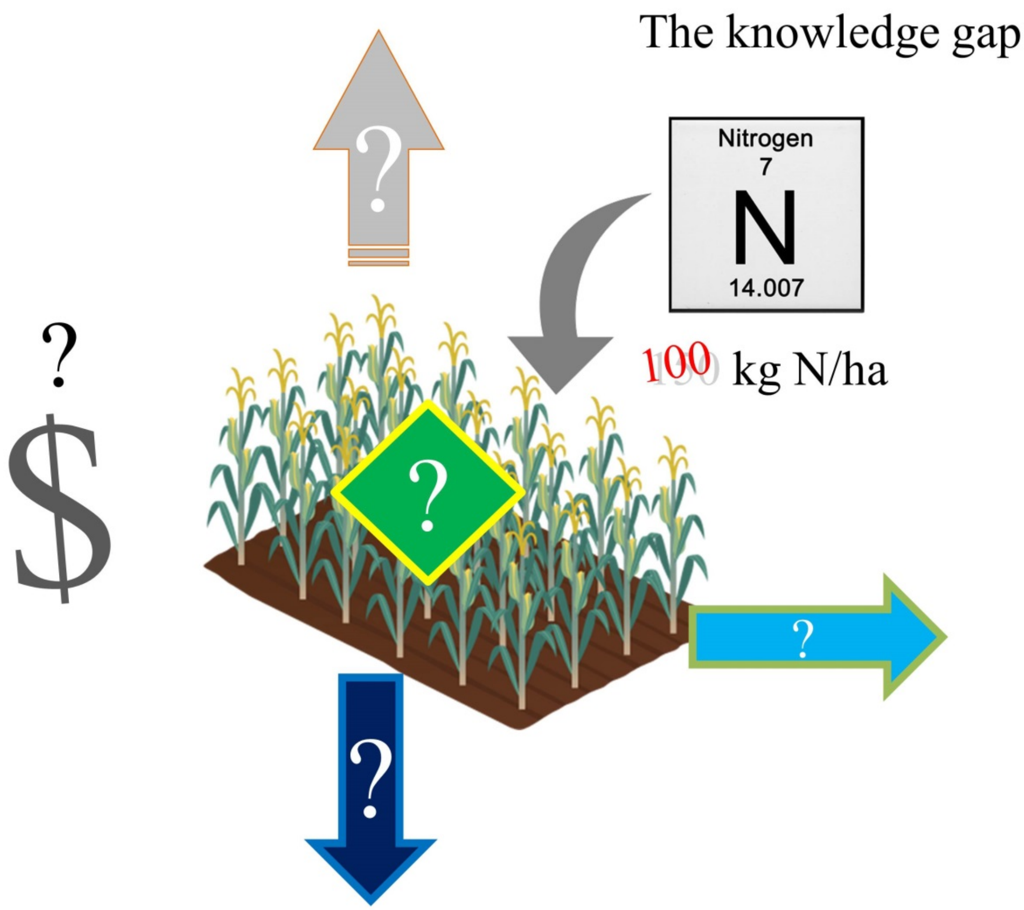

Fig 1. Currently, farmers and agronomists do not know the precise implication of, for instance, cutting N rate by 50 kg/ha on yield, net return, water or air pollution.

A similar challenge applies for a large majority of farm inputs. Without all the information required to manage farm inputs with precision, we tend to overcompensate and use higher amounts of inputs than necessary, thus contributing to the current agricultural crisis in which we have crossed multiple planetary boundaries below which, we maintain a safe operating space for humanity (Rockström et al. 2009). Furthermore, there is a lack of accountability for these actions, as well as a decoupling in space and time of management practices effects. In this regard, a practice can be beneficial from the farmer's business perspective but can have harmful, and potentially unknown, consequences later on the farm or elsewhere (e.g. US Midwest maize production impacting fisheries in the Gulf of Mexico).

In order to provide required tools for farmers to develop production systems that are sustainable in an environmental perspective, farmers, practitioners, businesses and scientists are developing techniques and technologies that have substantial impacts on agriculture production. Research in agronomy and related disciplines such as bioengineering, Information and Communication Technologies or mechanical engineering in the recent decades have enable agricultural production to greatly increase in efficiency (Barker 2007). This translates into a lower environmental footprint per unit of food produced that is the fruit of scientific progress and outstanding achievements in research and development. However, it appears that the environmental impacts of agricultural production outpace the scientific progress (Fig. 2). At this pace, there is a risk that we may not be able to provide the right tools to farmers before it is too late.

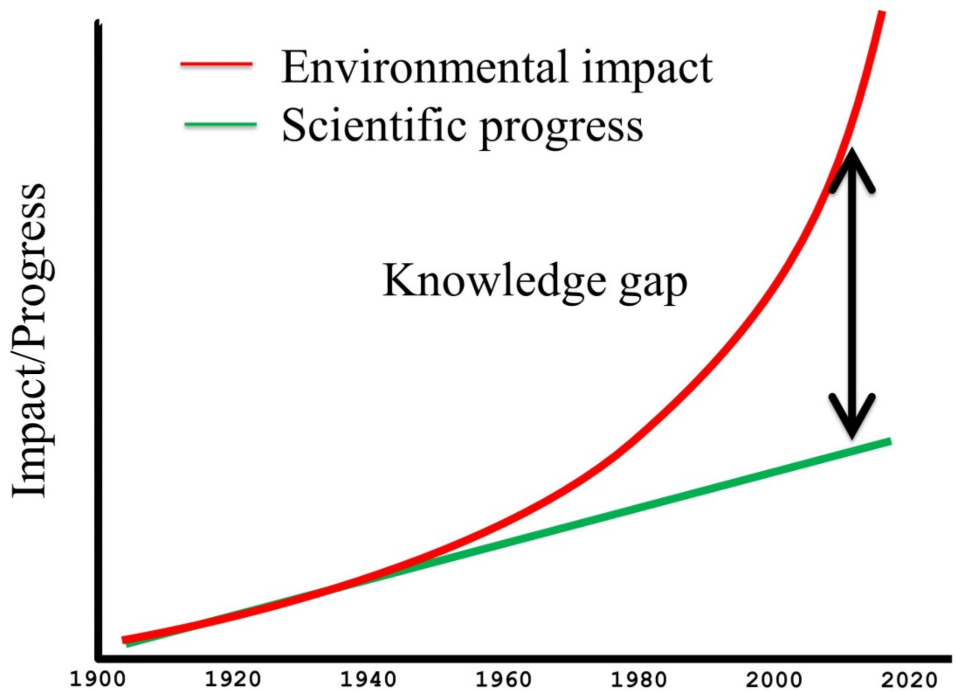

Fig 2. Schematic of the trend of agricultural impact on the environment that is faster than the speed of scientific progress.

Knowledge and innovation gap in agriculture

Crop and livestock production involve complex systems interacting in ways that are largely beyond the reach of our current management ability. Conceptually, one can summarize these interactions as part of the G x E x M (Genetics x Environment x Management) production context (Hatfield and Walthall 2015).

Traditional research in agronomy implies that to acquire knowledge, scientists need to conduct experiments in a controlled environment, replicating trials in space and time. Use of the traditional experimental study's methodology is problematic and unpractical for conducting experiments in a context where so many uncontrolled variables and unknown interactions exist; this leads to large experimental errors that often hide the effect of controlled variables. This knowledge generating process cannot keep up with the current acceleration of degradation of our planet due to agricultural practices. Therefore, we need to think of new ways to generate decision support tools for agricultural production systems. Further, traditional research methods have largely failed in delivering management practices that are optimal for the diversity of circumstances that are inherent to agricultural production, leading to a mix of low adoption, sub-optimal resource use and undesirable environmental consequences. For instance, global nitrogen use efficiency is about 47%, which is of great concern as this major farm input exhibits very high environmental toxicity (Lassaletta et al. 2014). The current crisis emerged, to a large extent, because agricultural practices have neglected to adapt to nature's variability due to a lack of available techniques and technologies to do so. There was a time when there was little hope of improving research and development models, technologies and data management options. Today, digitalization of agriculture is leading to new opportunities that are much better for addressing the multiple factors and variability that exist in nature. However, our current science, development, transfer and adoption system lives with a legacy of habits that limit our ability to generate, assemble, analyze and derive recommendations that will be applicable and trusted in diverse circumstances to produce desirable outcomes.

Potential of observational studies and big data to help bridge that gap

As agricultural research scientists, how can we do better? How can we improve on approaches using controlled experiments and making recommendations based on 'averaged' data? In simple systems, the best route is to build a theoretical model that takes into account all significant factors so that one can predict outcomes from a set of controlled and uncontrolled inputs and state variables. This is a powerful approach since it has been clearly demonstrated that strong theories can predict unobserved outcomes: Higgs boson was predicted from theory in the 1960s and was only recently observed. But this is the realm of theoretical physics for which there is no equivalent in agriculture… Why? A comprehensive system for describing crop growth in a natural environment would be very complex and would involve many uncontrolled random variables. Existing crop growth models rely heavily on empirical formulae and parameters (Di Paola et al. 2016).

In agricultural sciences, actionable knowledge is acquired by testing hypotheses with planned experiments or by developing regressions (empirical models) from observations. Today, tools are available to perform observational studies "on steroids". These fall under the realm of big data. There are good reasons why large-scale studies with fine-scale observations can result in much better working models. It is possible to gather the results of thousands of uncontrolled experiments happening year after year on a variety of soils for a large array of crops under varying climatic conditions, along with the significant accompanying variables (imagery, sensors, tracers, weather data, production data, and soil parameters), to create big data sets.

But data alone in a digital storage system is nothing more than record. This is where Artificial Intelligence (AI) must come into play. By applying AI tools to big agricultural data sets, it is expected that it will be possible to identify patterns where the right amount of farm inputs are delivered with precision at the right time and the right location to maximize yield while minimizing environmental impacts. This should naturally lead to the formulation of new hypotheses that could be tested under controlled experiments in order to refine our understanding and modeling capabilities. On paper, this is all valid. But how far we are from this paradigm and what actions can be taken to move into that direction?

Current state of big data in agriculture

Big data is the subject of much excitement in agricultural production these days. The reality is that very few, if any, sources of agricultural data actually qualify as "big data". Interpolated weather data and satellite imagery may be the only publicly available data sources that can currently be labeled as such. Yield maps would also qualify only if they would be better calibrated, standardized and made more accessible. Indeed, yield maps may exist in the cloud but not necessarily in an open manner and with a right of use for research purposes. It is likely that other anticipated big data sources would be limited either in spatial density (*in situ* sensors, IoT devices, genomics samples) or temporal intensity (UAV imagery). Another significant limitation for the implementation of big data and AI in agriculture is the difficulty of accessing records for the diversity of management practices implemented by the farmers in their day-to-day operations or as they conduct their own on-farm experimentation. A greater adoption of traceability technologies and electronic logbooks would allow for bridging this gap. This data would be invaluable in the calibration and validation of artificial intelligence algorithms.

Specific to data, two concepts are fundamental. First, as it is the case for most systems, the quality of the output depends on the quality of the input. Secondly, we can only optimize the management of what we can measure. Within that context, multiple sources of data are important in order to generate valuable actionable information. For instance, in order to optimize crop input use, management data such as input type, rate and timing as well as production data such as yield value and input costs are required. This data also needs to be contextualized using local soil, weather and farm records to be meaningful for each specific site. Because we can only optimize the management of what we measure, it is essential to access environmental data such as greenhouse gas emission, soil quality or water pollution. Otherwise, optimizing only yield and profitability may lead to even greater damages to the environment. Currently, scientists have access to data such as satellite imagery, weather data and national soil surveys, but do not have access to crop management and production data, which are held by farmers. Environmental data are sparse and inexistent at the scale required.

Farmers are aware that data is considered the new oil. Data is now the world's most valuable resource because of the tremendous potential for management improvement that can be achieved. However, data is worthless without an infrastructure organizing, diffusing and processing them into actionable information. Therefore, without such an infrastructure, it is difficult for farmers to estimate the value of their data and even more, fully appreciate the value of sharing their data. It is thus in developing and demonstrating the value of their data on their farm that resides the key for the involvement of farmers. Traditionally, agronomists rely on knowledge generated through small plot experiments to advise farmers on crop management. Farmers usually prefer to rely on knowledge generated from studies that were conducted close to their farm because they know that agricultural practices are specific to each and every setting where they are implemented. Therefore, it is by providing farmers with contextualized answers to their questions that they will appreciate the value of their data. In that matter, the most important data to scale the potential of big data and observational studies are the farm management and production data.

It should also be stated that the rise of big data in agriculture is also changing relationships within the industry. Suppliers are no longer simply product providers. Data and analytics have become new currencies in this relationship. Additionally, a new generation of agronomic service providers has emerged – each with their own data system and basis from which to provide both agronomic and economic advice. These changes are also impacting the role and relevance of research whether this is provided publically or privately.

The paradigm shift

Multiple artificial intelligence methods and especially one of the most powerful ones such as Deep Learning, involve the use of "black boxes". These techniques use "black boxes" by nature rather than by design. Indeed, the Deep Learning process involves hidden layers of interpretation to produce a result (Castelvecchi 2016). In this case, AI is able to provide an answer, a pretty reliable one at it (Ferentinos 2018), but cannot provide explanation on why this is the answer. In order to understand the "why" one would have to conduct experimental research and establish the causality. However, as it was mentioned above, conducting experimental research is costly and slow. Thus, the question is, *Can we manage farms using recommendations for which we don't understand the "why"*? Another way to put it is, *Can we afford anymore to understand everything that we implement*? These are important questions that will require answers and that may lead to a major paradigm shift in agronomy.

The role of scientists

There is a very small probability (2.1 %) that soil and plant scientists be replaced by big data and AI when these become main stream for farm management (Frey and Osborne 2017). Indeed, AI is not intended to invent new techniques and technologies, but can optimize what we already use on a large scale better than other processes currently used. For instance, AI could not come up with band applied fertilizer if all fertilizer is broadcasted. Scientists would have to think of new approaches, test it in controlled environment and implement it on the farm on a large scale before AI would recognize that this approach is more efficient. Another fundamental reason why the role of scientists cannot be replaced by AI is that it does not provide explanations to the answers it provides and there is always a certain margin of error. The task of understanding causality and better prepare for the rare cases when AI is mistaken, scientists will have to conduct experimental science to better understand the patterns detected by AI. Scientists will also have the role of determining the important data to consider and find ways to collect that data to be provided to AI algorithms in the appropriate volume and velocity. The role of scientists will thus be synergetic with AI rather than be replaced by it.

Case study: Thresholds for integrated pest management in berry production

In the province of Quebec, integrated pest management (IPM) has lost traction among berry growers who question the validity of crop injury thresholds (CIT) that have been developed in the 1990s with climate conditions, crop practices, insecticides and crop varieties that were different than the ones used nowadays. Nevertheless, these CITs required significant investments in resources and time to develop. Indeed, the implementation and replication of large factorial design experiments that would cover several crop varieties and be conducted in multiple berry growing regions for the most important pest species can hardly be envisaged anymore. Moreover, the reliability in time of these CITs could be questioned because of the climate that is changing. Efficient and reliable ways of determining CITs using modern technologies are yet to be developed, which would facilitate the use of CITs in an IPM strategy blending in the current cropping practices. Without reliable CITs, more insecticides are used than necessary, leading to depletion of natural biological control agents and pollinators as well as pollution of the biosphere. The synergetic effect of depleting biological control agents and increased insecticide applications greatly reduces berry production profitability and acceptability by consumers.

Another possibility for developing CITs is harnessing the workforce that is already collecting large amounts of data relative to pest management in berries. The development of CIT requires multiple pieces of information notably: pest identification and counts, pest growth stage, crop growth stages, crop variety, date of the year, location of the sample, weather data, type and date of pesticide applied and yield (crop injury) harvested among others. All this information is being collected routinely by growers and pest management consultants, but it is often scattered and disorganized and therefore cannot be analyzed. There is a need to gather and organize the data collected by growers and pest management consultants to develop a bank of data that would constantly be updated with new data collection. Using AI and metadata analysis, CITs could be determined in a dynamic way, constantly evolving with new data in a web-based tool that would be made available to all berry growers.

Conclusion

In conclusion, because of the speed at which agriculture is impacting the planet and the time and resources required to improve the situation, the *status quo* in soil and plant science is not acceptable anymore. The use of big data and artificial intelligence in agriculture has the potential to accelerate the access to optimal recommendations for each and every grower. However, there are risks and pitfalls to avoid such as optimizing only productivity and profitability and omitting environmental sustainability. There is an opportunity to be ceased and soil and plant scientists better hop in the wagon to the risk of becoming obsolete.

Acknowledgements

We want to acknowledge the work of Éléonore Tremblay for her help in the literature review. The funding was provided by Agriculture and Agri-Food Canada.

References

Barker, A. V. (2007). Enhancing the Efficiency of Nitrogen Utilization in Plants. HortScience, 42(6), 1501–1501.

Castelvecchi, D. (2016). Can we open the black box of AI? Nature News, 538(7623), 20. doi:10.1038/538020a

Di Paola, A., Valentini, R., & Santini, M. (2016). An overview of available crop growth and yield models for studies and assessments in agriculture: Overview of crop models for agriculture. Journal of the Science of Food and Agriculture, 96(3), 709–714. doi:10.1002/jsfa.7359

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 145, 311–318. doi:10.1016/j.compag.2018.01.009

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? Technological Forecasting and Social Change, 114, 254–280. doi:10.1016/j.techfore.2016.08.019

Hatfield, J. L., & Walthall, C. L. (2015). Meeting Global Food Needs: Realizing the Potential via Genetics × Environment × Management Interactions. Agronomy Journal, 107(4), 1215–1226. doi:10.2134/agronj15.0076

Lassaletta, L., Billen, G., Grizzetti, B., Anglade, J., & Garnier, J. (2014). 50 year trends in nitrogen use efficiency of world cropping systems: the relationship between yield and nitrogen input to cropland. Environmental Research Letters, 9(10), 105011. doi:10.1088/1748-9326/9/10/105011

Mazoyer, M., & Roudart, L. (1997). Histoire des agricultures du monde, *Du néolithique à la crise contemporaine* Paris, France, Le Seuil.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J. & Nykvist, B. (2009). A safe operating space for humanity. Nature, 461(7263), 472–475. doi:10.1038/461472a