Report from the conference, 'identifying obstacles to applying big data in agriculture'

Emma L. White
Texas A&M University, emmaly.white@tamu.edu

J. Alex Thomasson
Texas A&M University

Brent Auvermann
Texas A&M AgriLife Research

Newell R. Kitchen
United States Department of Agriculture

Leland Sandy Pierson
Texas A&M University

See next page for additional authors

Precision Agriculture (2021) 22:306–315
https://doi.org/10.1007/s11119-020-09738-y

BRIEF COMMUNICATION

Report from the conference, 'identifying obstacles to applying big data in agriculture'

Emma L. White[1] · J. Alex Thomasson[1] · Brent Auvermann[2] · Newell R. Kitchen[3] · Leland Sandy Pierson[1] · Dana Porter[2] · Craig Baillie[4] · Hendrik Hamann[5] · Gerrit Hoogenboom[6] · Todd Janzen[7] · Rajiv Khosla[8] · James Lowenberg-DeBoer[9] · Matt McIntosh[10] · Seth Murray[1] · Dave Osborn[11] · Ashoo Shetty[12] · Craig Stevenson[13] · Joe Tevis[14] · Fletcher Werner[15]

Published online: 15 July 2020

Abstract
Data-centric technology has not undergone widespread adoption in production agriculture but could address global needs for food security and farm profitability. Participants in the U.S. Department of Agriculture (USDA) National Institute for Food and Agriculture (NIFA) funded conference, "Identifying Obstacles to Applying Big Data in Agriculture," held in Houston, TX, in August 2018, defined detailed scenarios in which on-farm decisions could benefit from the application of Big Data. The participants came from multiple academic fields, agricultural industries and government organizations and, in addition to defining the scenarios, they identified obstacles to implementing Big Data in these scenarios as well as potential solutions. This communication is a report on the conference and its outcomes. Two scenarios are included to represent the overall key findings in commonly identified obstacles and solutions: "In-season yield prediction for real-time decision-making", and "Sow lameness." Common obstacles identified at the conference included error in the data, inaccessibility of the data, unusability of the data, incompatibility of data generation and processing systems, the inconvenience of handling the data, the lack of a clear return on investment (ROI) and unclear ownership. Less common but valuable solutions to common obstacles are also noted.

Keywords Automation · Big data · Farm profitability · Food security

Introduction

Current state of Big Data issues according to experts

Global needs are driving production agriculture toward utilizing advances from new Big Data technologies developed in other industries. A fundamental global need is food

✉ Emma L. White
emmaly.white@tamu.edu

Extended author information available on the last page of the article security, which is threatened by an increasing population, decreasing arable land area, a changing climate, food waste and living standards that focus consumer preference for animal protein. Another major need is farm profitability, which is critical to incentivize food production. The costs of inputs including seed and fertilizer continue to increase (USDA Economic Research Service 2020a), while commodity prices tend to decrease in real dollars (USDA Economic Research Service 2020b). Thus, farmers must grow increasingly more food per unit land area to remain profitable. Food production must increase 25 to 70% by 2050 (Hunter et al. 2017) to meet world demand, and it is reasonable to conclude that per-hectare food output may need to double by the time world population plateaus around 2100 (United Nations 2019). Plant breeders and geneticists are developing higher yielding and more resilient crops through efforts increasingly dependent on Big Data technologies. Big Data deals with 'extensive datasets—primarily in the characteristics of volume, velocity and/or variability—that require a scalable architecture for efficient storage, manipulation and analysis' (NIST 2015). Volume refers to the number of data bytes, while velocity refers to the rate or frequency at which data are collected or updated, and variability refers to the irregularity of the data, which affects information content and compressibility of the data. Some of these Big Data technologies are also moving beyond research and into production agriculture.

Precision agriculture technology focuses on spatio-temporal variability in farm inputs and outputs, and has grown in sophistication over the past three decades (Lowenberg-DeBoer 2015). This concept took hold with the advent of global navigation satellite systems (GNSS) and soon progressed to geographic information system (GIS) mapping and analysis. Three main precision-agriculture practices—soil and yield mapping, GNSS guidance and variable-rate application—have gained significant traction among farmers, varying by crop and country (e.g., 59% of maize area planted with GNSS guidance in the U.S. as of 2016) (Lowenberg-DeBoer and Erickson 2019). Advanced sensors and autonomous sensing platforms including robots and drones, Big Data analytics, Internet of Things (IoT), wireless communications and cloud-based systems have begun to provide vast amounts of data and modeling tools for real-time control of in-field vehicles and implements.

Reports suggest that investments in data-centric agricultural technologies have been increasing at a high rate annually since 2012 (Walker et al. 2016). A survey of over 1500 farmers demonstrated high rates of data collection—mainly because it is inherent in modern equipment—but low rates of data usage (Brooks 2017; Roberts 2017). An abundance of optimism has developed around data-driven improvements in agricultural productivity and profitability, but a thorough review on the potential for Big Data to enable "smart-farming" demonstrates that farmers struggle to use the data for decision making (Wolfert et al. 2017).

Conference description

In light of the foregoing, the USDA-NIFA funded a conference called, "Identifying Obstacles to Applying Big Data in Agriculture." Texas A&M University, the recipient of the grant, hosted the conference held in Houston, TX, on August 20–21, 2018. Attendees (58) represented diverse interests, including industry (12), growers (6), academia (30) and government organizations (10). Industries represented included the following: machinery and equipment, broadband services, seed products and services, data services, satellite data products, cloud-based support services, analytical services, legal services and venture capital services. Grower representation included grains and cotton as well as swine and cattle production. The research and teaching fields of the academic participants included agricultural engineering, agricultural economics, agronomy, breeding and genetics, crop physiology, entomology and plant pathology, biology, mechanical engineering, and electrical and computer engineering. Government participants represented the research-funding arm and the statistics arm of the U.S. Department of Agriculture. Countries represented included the U.S., Canada, Australia, and the U.K.

The conference was opened with the following popular definition: "Big data is data sets that are so big and complex that traditional data-processing application software are inadequate to deal with them. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy and data source. There are a number of concepts associated with big data: originally there were 3 concepts: volume, variety, velocity. Other concepts later attributed [to] big data are veracity (i.e., how much noise is in the data) and value" (Wikipedia 2018). Therefore, it was made clear that agricultural Big Data is a much more complex topic than simply the need for more and better data. A deeper level of discussion of the underlying science was thus held on the first day, when the speakers covered fundamental scientific issues as well as generation of, utilization of, and practical issues in agricultural Big Data. With this information presented as a backdrop, working-group discussions of on-farm utilization were held on the second day and generated the principal outcomes of the conference.

Outcomes

Working groups were tasked with identifying two to three common scenarios in which Big Data can improve farm profitability or environmental risk mitigation, creating a list of obstacles to using Big Data, and determining potential solutions to overcome the obstacles. Conference participants were selectively grouped together to form four working groups, enabling a breadth of knowledge to be considered in each different working group. In general, the groups tended to be oriented towards field crops, but one group was particularly oriented towards animal production. Ten scenarios were generated by the working groups, and obstacles and solutions were identified for each scenario. Two of the ten are presented here as a representation of overall key findings, and the rest are listed in Table 1 and available in detail in the full conference report under the “Outcomes” tab on the conference website www.agbigdataobstacles.com.

Scenario: “in-season yield prediction for real-time decision-making”

The setting

Accurate yield predictions are useful to row crop farmers in marketing their commodities and informing their crop management decisions. To generate accurate and precise yield predictions, farmers need to make use of historical yield data and previous field inputs including tillage, seed type and spacing, irrigation and fertilizer, as well as soil properties, remote sensing data and weather data.

Table 1 Obstacles identified in specific agricultural Big Data scenarios. Column headings described in text below

Scenario	Obstacle						
	Error in the data	Inaccessibility	Unusability	Incompatibility	Inconvenience	Lack of ROI	Unclear ownership
Sow lameness	x	x		x	x		
Irrigation in cotton management		x		x			x
Mid-season yield prediction for real-time decision-making		x	x	x		x	
In-season decision making	x		x		x		x
Policy maker perspective	x			x	x		x
Cropping selection system	x	x		x		x	x
Business analytics for agriculture		x	x	x		x	
Grower's perspective	x	x	x	x		x	
Consumer perspective	x			x			x
Benchmarking scenario—comparing individual grower yields to modeled outputs based on other people's data	x	x	x		x		

The obstacles

A lack of data interoperability prevents the integration and unified analysis of data collected by multiple sensors and platforms. Lack of rural bandwidth often makes data transmission, particularly in large datasets including images, impossible. With data coming from numerous sensors, calibration, ground truth data and data standards are needed to support accurate model predictions. Some standards do exist for collecting well-attributed data, but they require a higher level of management. Models that predict profit and specific well-defined value propositions are also needed to determine return on investment (ROI) in decision making. Finally, crop growth models need to be tied to better representations of soil properties instead of proxies, and weather forecasting needs to be more specific to individual farms and fields.

The solutions

Standards or guidelines for farmers on how to collect "good" data should be developed, so that collected data are usable in multi-platform systems, and are underpinned by ground-truth data and proper sensor calibration. These standards should be deployed through existing channels such as extension and agricultural suppliers. Funding should be made available for studies focused on determining ROI for data-intensive endeavors including yield prediction. Increased communication between local co-operatives, consultants, farmers, researchers, policy makers, etc. can encourage farmers to adopt Big Data practices for achieving useful yield predictions. Finally, researchers should use available farmer-supplied data and machine learning to derive better representations of soils for crop modeling.

Scenario: "sow lameness"

The setting

On hog farms worldwide, sow lameness is a principal cause of mortality and poor productivity. Lameness in sows is a deviation from their normal gait (Anderson 1994) that may result from infection, heredity, environment and nutrition and is used as a criterion for culling. Incorrect identification of lameness adds significant cost to swine operations, including animal death, reduced value, diagnosis and treatment, and extra labor (Rowles 2001). Early detection of lameness through integration of imaging and analytical software would improve decision-making relative to sow lameness and consequently would improve profitability.

The obstacles

A major issue shown in research is that animals can change their behavior during data collection because of human interference, so personally observable lameness data may not be trustworthy. Also, the practicality of currently available systems, integration of software across systems and on-site expertise present a major challenge to reducing premature mortality and animal suffering.

The solutions

Real-time, image-based data sets of a large number of animals that have been calibrated and validated could be analyzed in order to classify sows into categories with respect to lameness; artificial intelligence (AI) or Bayesian analysis methods would improve these estimates. Bayesian analysis is a method in which probability for a hypothesis (e.g. this sow has lameness) is updated as more data (additional sow observations) become available. Genetic markers for predisposition to lameness could be developed from these data including animal weight distribution and biomechanics including animal gait and gross behavior. RFID (radio-frequency identification) tags can be used to inform these algorithms for long-term selection and cull decisions. Avoiding false identification of sow lameness may require human validation. Research and practice in related industries including bovine, equestrian and human medicine should also be explored for other appropriate solutions. Research and development efforts need to be multi-disciplinary to enable solutions that impact farmers. Funding models for this type of research could include business start-ups, due to the high marketability and potential returns on resulting prediction systems. Researchers and farmers must engage with the private sector to develop infrastructure and expertise for operation, maintenance and development of sensing and analysis systems.

Common obstacles and solutions identified

The discussions that generated all the scenarios revealed several common obstacles. These obstacles are discussed briefly here and listed in Table 1 alongside the scenarios in which they were identified.

Error in the data

Data have inherent error, and the accuracy level is commonly unknown. Farmers generally do not have the time or resources to ensure that collected data are accurate and precise, especially when the ROI on this data collection remains unclear. Currently many sensor calibration systems and methods, when available, tend to be tedious and lack obvious value to the farmer. Automated calibration of sensing systems and a clear understanding of measurement accuracy need to be developed in partnership with farmers, manufacturers, researchers and data scientists.

Inaccessibility

Data are often not easily accessible to farmers or service providers working for them. Physical accessibility to data and actionable information derived from them are limited by the volume of the data and the lack of communications bandwidth in rural areas and particularly in remote farm fields. Farm-based data, chiefly large data files, are generally difficult to transport to farm-based and cloud-based analytical systems. Major improvements in rural broadband are seen as part of the solution. Of the 24 million people in the U.S. without access to broadband, 80% are in rural areas (USDA 2019), and nearly 40% of rural residents do not have access to broadband that is 25 Mbits/s or faster (ITIF 2017). While

5G wireless technology is expanding to rural towns due in part to USDA's Rural Broadband Initiative, the remoteness of a vast number of farm fields suggests that real-time data transmission as well as transmission of large data files like images collected with drones will be impractical for the foreseeable future. Easier data transfer systems would also be useful, and edge computing of collected data would reduce the overall data volume to be transferred over long distances.

Unusability

Data are often not readily usable. Data are often being collected in high volume and complexity, making utilization difficult. Data are also often stored with different units, formats, metadata, time and space intervals, etc. Farm-training through extension education can be part of the solution, but technical solutions including rapid simplification of large data sets (e.g., images) as well as standardized data formats for agriculture that enable integration with IoT are also needed.

Incompatibility

Software packages used to collect, store and analyze data in various applications are commonly incompatible with one another. Developing software systems (including community data repositories) with better interoperability and functionality can improve the feasibility and efficiency of data analyses.

Inconvenience

Capturing, storing, analyzing and using data are commonly too difficult or time-consuming to be considered worthwhile by agricultural users. Automating procedures from data capture, data cleaning, through data analysis, as well as automating various decisions, like whether to sell crop quantities at a certain time, would help to overcome these obstacles.

Lack of ROI

Possibly the greatest impediment to on-farm use of data-intensive technologies is the lack of a clear ROI. Research studies focused on determining ROI in common, important, decision-making applications in broadly grown crops or animal industries with local specificity are critical.

Unclear ownership

Concerns about data ownership, privacy and security are major constraints on the growth of data-intensive technologies in agriculture. Farmers sense that their data are valuable and that many companies may wish to extract value from those data without compensation. Farmers are also concerned about having their data used against them in insurance claims, litigation and regulatory enforcement. Consistent corporate standards, regulations and/or laws need to be put in place to prevent abusive behavior and encourage utilization of data and related decision-making systems. Also, incentives need to be clear and engagement needs to be simple for farmers to participate in common data repositories. Local government or co-op owned data clearinghouses may provide additional trust.

Uncommon but valuable solutions to common obstacles

Some solutions to obstacles were infrequently mentioned and unique, yet exhibited particular insight:-

(1) AI should be exploited for its capabilities in (a) creating actionable information from Big Data, (b) solving calibration problems, and (c) deriving representative data for use in crop modeling, for example.
(2) While traceability is desirable for supply chain management, food safety and sustainability metrics, traceability in many crops is difficult to achieve, especially because of the diffuse stakeholders involved, unlike consumer retail where a few companies dominate. It would require “data linking” across field operations, business interests and relevant data standards.
(3) Special attention should be paid to the Big Data needs of regional and small crop farmers who may not receive the same level of access to Big Data (e.g., remote-sensing data) or data processing that large farmers of major crops receive.
(4) Agricultural researchers should engage with the private sector in public–private partnerships and determine ROI to collaboratively develop data-centric agricultural technologies. Such engagement helps to ensure that researchers are solving real-world problems and the private sector is appropriately dealing with issues of science.

Post-conference analysis

It is worth noting in hindsight that some of the obstacles and solutions identified in the conference relate to agricultural data in general and are not specific to agricultural Big Data. For example, the error inherent in data can be an issue regardless of the volume, variety or velocity of the data. Furthermore, the incompatibility of software packages used to collect, store and analyze data can be a problem with large and small data files. Additionally, data ownership, privacy and security concerns exist even with small data files.

Summary

In summary, the technologies associated with agricultural Big Data were discussed at the conference to provide background information for identifying obstacles to applying Big Data in agriculture. Seven common obstacles were identified for production agriculture; these include error in the data, inaccessibility, unusability, incompatibility, inconvenience, lack of ROI and unclear ownership. Specific technological issues contained in these obstacles include the “size” of Big Data and lack of broadband communications in remote areas. Overcoming these obstacles is a complex problem, but solutions were proposed which could potentially be employed to improve farm profitability, mitigate environmental risk and help meet the global need for food security in the near and far future.

Acknowledgements This work was supported by a United States Department of Agriculture—National Institute of Food and Agriculture, Food and Agriculture Cyberinformatics and Tools (USDA NIFA FACT) initiative conference Grant Number 2018-67021-28692.

Author contributions Conceptualized and developed conference: ELW, JAT; Wrote manuscript: ELW, JAT; Contributed technical material: CB, HH, GH, TJ, RK, JL-DB, MM, SM, DO, AS, CS, JT, JAT, FW; Organized scenarios: BA, NRK, LSP, DP; Revised: BA, LSP, DP, HH, GH, JL-DB, SM, JT, JAT.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Anderson, D. (1994). *Dorland's illustrated medical dictionary* (28th ed.). Philadelphia, PA, USA: J.B. Saunders.
Brooks, R. (2017). Farmer survey highlights need for agronomic data management services. *Ag Professional*. Retrieved November 21, 2019 from https://www.agprofessional.com/article/farmer-survey-highlights-need-agronomic-data-management-services.
Hunter, M. C., Smith, R. G., Schipanski, M. E., Atwood, L. W., & Mortensen, D. A. (2017). Agriculture in 2050: Recalibrating targets for sustainable intensification. *BioScience, 67*(4), 386–391.
Information Technology and Innovation Foundation. (2017). Defining and mapping broadband coverage in America. Retrieved August 10, 2019 from https://www2.itif.org/2017-defining-mapping-broadband.pdf.
Lowenberg-DeBoer, J. (2015). The precision agriculture revolution. *Foreign Affairs, 94*(3), 105–112.
Lowenberg-DeBoer, J., & Erickson, B. (2019). Setting the record straight on precision agriculture adoption. *Agronomy Journal, 111*(4), 1552–1569.
Chang, W. L., & Grady, N. (2015). NIST Big data interoperability framework: Volume 1, definitions. *NIST Special Publication 1500-1*.
Roberts, O. (2017). Survey shows gap between precision data collection and use. *Real Agriculture*.
Rowles, C. (2001). Sow lameness. *Journal of Swine Health and Production, 9*(3), 130–131.
United Nations. (2019). World population prospects. *United Nations Department of Economic and Social Affairs, Population Division*. Retrieved November 21, 2019 from https://population.un.org/wpp/.
U.S. Department of Agriculture. (2019). A case for rural broadband. Retrieved August 15, 2019 from https://www.usda.gov/sites/default/files/documents/case-for-rural-broadband.pdf.
USDA Economic Research Service. (2020a). Commodity costs and returns. Retrieved May 16, 2020 from https://www.ers.usda.gov/data-products/commodity-costs-and-returns/.
USDA Economic Research Service. (2020b). Inflation-adjusted price indices for corn, wheat, and soybeans show long-term declines. Retrieved May 16, 2020 from https://www.ers.usda.gov/data-products/chart-gallery/gallery/chart-detail/?chartId=76964.
Walker, D., Kurth, T., Van Wyck, J., & Tilney, M. (2016). Lessons from the frontlines of the Agtech revolution. *Boston Consulting Group (BCG) Perspectives*. Retrieved November 21, 2019 from https://www.bcg.com/en-us/publications/2016/process-industries-building-materials-strategy-lessons-frontlines-agtech-revolution.aspx.
Wikipedia. (2018). Big data. Retrieved January 20, 2018 from https://en.wikipedia.org/wiki/Big_data.
Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big data in smart farming: A review. *Agricultural Systems, 153*, 69–80.

Affiliations

Emma L. White[1] · J. Alex Thomasson[1] · Brent Auvermann[2] · Newell R. Kitchen[3] · Leland Sandy Pierson[1] · Dana Porter[2] · Craig Baillie[4] · Hendrik Hamann[5] · Gerrit Hoogenboom[6] · Todd Janzen[7] · Rajiv Khosla[8] · James Lowenberg-DeBoer[9] · Matt McIntosh[10] · Seth Murray[1] · Dave Osborn[11] · Ashoo Shetty[12] · Craig Stevenson[13] · Joe Tevis[14] · Fletcher Werner[15]

1 Texas A&M University, College Station, TX, USA

2 Texas A&M AgriLife Research and Extension Center at Amarillo, Amarillo, TX, USA

3 USDA-ARS Cropping Systems and Water Quality Research, Columbia, MO, USA

4 University of Southern Queensland, Toowoomba, QLD, Australia

5 IBM T.J. Watson Research Center, Yorktown Heights, NY, USA

6 University of Florida, Gainesville, FL, USA

7 Janzen Agricultural Law LLC, Indianapolis, IN, USA

8 Colorado State University, Fort Collins, CO, USA

9 Harper Adams University, Land, Farm and Agribusiness Management, Newport, Shropshire, UK

10 Journalist and Communications Professional, Mc Communications, Wheatley, ON, Canada

11 VTX1 Companies, Raymondville, TX, USA

12 Amazon Web Services, Dallas, TX, USA

13 BASF Canada Agricultural Solutions, Calgary, AB, Canada

14 Vis Consulting Inc, Minnetonka, MN, USA

15 The Climate Corporation, St. Louis, MO, USA