Science Societies

Data science in agriculture

Part II: Applications, advancements, and future prospects

By Eeswaran Rasu, Rachael Bernstein

June 18, 2020



Sensors can be used alone or mounted on unmanned aerial vehicles (left), robot-assisted imaging platforms (right), or cable-suspended camera systems to take real-time images and measurements throughout the selected time duration. Photos courtesy of Flickr/Wil C. Fry (left) and Flickr/PETROSS, TERRA-MEPP & WEST (right).

Editor's note: This article is Part II of a two-part series. See last month's article at **https://doi.org/10.1002/csan.20145**.

In the last issue, we introduced data science in the context of agriculture and highlighted its importance in precision agriculture. In this issue, we discuss some of the insights into the applications, advancements, and future of data science in agricultural systems.

Breeding and Screening of Crop Varieties

Data science is increasingly used in plant breeding. High-throughput phenotyping, which measures certain plant traits from the molecular to whole-plant level, has benefitted from the use of remote and proximal sensing to guide breeders through accurate trait selection (Singh et al., **2016**). This type of non-destructive data collection has been adopted to crop species to identify and evaluate tolerant traits to abiotic (e.g., drought and flood) and biotic (pests, weeds, and diseases) stresses throughout different growth stages. Conventional phenotyping was particularly constrained, being a laborious, destructive, and time-consuming process (Beauchêne et al., **2019**).

There have been several advancements in the types of sensors used to image, such as hyperspectral, fluorescence, thermal, and visible to infrared. These sensors can be used alone or as mounted on unmanned aerial vehicles (UAVs), robot-assisted imaging platforms, or cable-suspended camera systems to take real-time images and measurements throughout the selected time duration. Unmanned aerial vehicles are particularly useful for low-latitude, high-resolution aerial imaging to evaluate crop emergence, vigor, and yield potential of row crops (Sankaran et al., **2015**; Shi et al., **2016**). Moreover, big data and machine-learning techniques have been applied in the areas of molecular biology and biotechnology, including genomics, transcriptomics,

proteomics, and systems biology (Ma et al., 2014) and to evaluate the yield performance of genetically modified crops (Johnson et al., 2019).

Crop Protection



To improve pesticide recommendations, researchers have been using data science to accurately calculate the rate and timing of application of these agro-chemicals. Photo courtesy of Flickr/Tamina Miller.

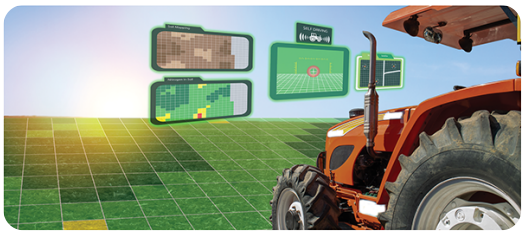
There can be monetary and environmental costs that accompany pesticide applications currently required for pest management. While also costly, excess use of pesticides has been shown to increase the prevalence of pesticide resistance and negative effects on non-target species (Mallet, 1989). To improve pesticide recommendations, researchers have been using data science to accurately calculate the rate and timing of application of these agro-chemicals. This approach considers data collected about the pest, such as life cycle and biology, along with environmental factors such as natural predators, growth stages of the target plant, and weather conditions during the growing season (Meisner et al., 2016). Additionally, big data and machine-learning methods have become very useful to control weeds (Ip et al., 2018), detect foliar diseases (Ferentinos, 2018), classify field crop pests (Xie et al., 2018), and detect stored grain pests (Shen et al., 2018).

Remote Sensing

Remote sensing is another area where data science is intensively used. The common applications of remote sensing in agriculture are monitoring agricultural land use, estimating soil moisture, generating crop planting/intensity maps, classifying crops,

assessing within- and between-field variabilities, forecasting crop yields, estimating evapotranspiration and irrigation requirements, monitoring crop health for input optimization, and evaluating ecosystem services (Weiss et al., 2020). Satellite missions such as Landsat and Sentinel have substantially improved their spatial, spectral, temporal, and radiometric resolutions in the past (Wulder et al., 2019). These high-resolution images are utilized to calculate various vegetative indices to evaluate the performances of different agricultural systems (Hatfield et al., 2019). For example, Xu et al. (2018) demonstrated how remote sensing can be used to estimate nitrogen uptake and cover crop biomass following winter, which could be useful in determining management practices for early spring.

Agricultural System Models and Decision Support Systems



Data science is intensively used in remote sensing and decision support platforms based on web GIS. Photo courtesy of Adobe Stock/Monopoly919.

Agricultural system models require a vast number of data inputs to simulate target outputs (e.g., crop development, water balance, nutrient dynamics, soil carbon, yield, net return, etc.) at farm and landscape scales (Antle et al., 2017). Moreover, advanced data science approaches such as deep learning have been used to train crop model inputs and outputs to evaluate the impacts of irrigation amount and timing on crop yield (Saravi et al., 2019). Wireless sensor networks (WSN), in combination with internet of things (IoT) and data analytics, aid in collection of various agro-environmental data, so these models can be integrated into decision support platforms. For example, Ellenburg et al. (2019) introduced Regional Hydrological Extremes Assessment System (RHEAS), an integrated modeling framework, to estimate onset, severity, recovery, and duration of regional droughts and

expected crop yield outlooks. Moreover, decision support platforms based on web GIS are useful to connect local farmers to regional and global agriculture to improve demand-and-supply-based input provision, marketing, and agricultural policies (Delgado et al., **2019**).

Combating Climate Variability and Change



Climate data enables researchers to evaluate the impacts of climate change on major crops. Photo courtesy of Adobe Stock/scharfsinn86.

Climate data enables researchers to evaluate the impacts of climate change on major crops and formulate and implement adaptation strategies such as climate-smart agriculture (Rosenzweig et al., **2014**; Hassani et al., **2019**). Data science together with the advancement in information and communication technology (ICT) is used in forecasting weather and seasonal climate, which could help farmers to deal with future climate variability (Klemm and McPherson, **2017**). Improved understanding of the upcoming season would give producers time to make necessary changes to farm management. Long-term crop, soil, climate, and topographic data from large commercial farms have demonstrated the potential to design site-specific adaptations to climate variability (Martinez-Feria and Basso, **2020**). Furthermore, artificial intelligence, Bayesian models, and neural networks have shown their potential to enhance climate warnings to extreme weather (Huntingford et al., **2019**). Recently, Newlands et al. (**2019**) explored the applicability of deep learning to climate risk management options such as agricultural insurance. It was more accurate than the other approaches typically used and demonstrated its potential in lowering insurance coverage costs, among others.

Development of Smallholder Farms

Datasets that evaluate the production performance of major crops such as legumes (Cernay et al., 2016) and spatial databases like RiceAtlas (Laborte et al., 2017) can be utilized to enhance the food and nutritional security of smallholder farmers in developing countries. These datasets are particularly useful to identify the interaction effects of crop species × location × growing season × treatment combinations on yield, nitrogen content, and water use efficiency under different management practices (Cernay et al., 2016). Utilizing this information can allow farmers to predict issues surrounding environmental variability and to increase production efficiency across different agro-ecological regions. Data science also offers many opportunities such as increasing the effectiveness of agricultural development projects and generating better information for policy decisions. This involves creating systems that respond to local problems and can be utilized in future predicaments. However, it requires technological development in extension services, careful governance, and public investment in order to avoid a few growers monopolizing available space and reducing further development (van Etten et al., 2017).

Marketing and Supply Chain Management

Pre-season yield forecasting can assist on-farm decision making regarding product harvesting, storage, and marketing (Oliveira et al., 2018). Data science is widely used in tracking of food supply chains with the use of Radio Frequency Identification (RFID), intelligent supply–demand forecasting tools, and information management systems to ensure quantity, quality, and food system transparency while minimizing food losses along the supply chain (Tzounis et al., 2017).

Future Prospects

Data science has been identified as one of the breakthroughs to advance food and agricultural research by 2030 by the National Academies of Sciences, Engineering, and Medicine. However, challenges exist regarding quality, ownership, privacy, and security of data; integration of data types; data processing and analytics; and uncertainties in algorithms (Wolfert et al., **2017**). Moreover, data science and scientific principles should be utilized together to explore data insightfully and to interpret the results accurately. This theory-guided data science will require interdisciplinary collaboration in research. With increasing open sources of big data, computational infrastructures, innovative methodologies, and public involvement in data collection such as citizen science (Ryan et al., 2019), data science has the potential to increase food security and environmental sustainability while promoting public- and private-sector initiatives and business ventures in future agriculture.



Datasets can be utilized to enhance the food and nutritional security of smallholder farmers in developing countries. Photo courtesy of Flickr/CIGAR Research Program.

References

Antle, J.M., Basso, B., Conant, R.T., Godfray, H.C.J., Jones, J.W., Herrero, M., … Tittonell, P. (2017). Towards a new generation of agricultural system data, models and knowledge products: Design and improvement. *Agricultural Systems*, **155**, 255–268.

Beauchêne, K., Leroy, F., Fournier, A., Huet, C., Bonnefoy, M., Lorgeou, J., … Cohan, J.P. (2019). Management and characterization of abiotic stress via PhénoField®, a high-throughput field phenotyping platform. *Frontiers in Plant Science*, **10**, 904.

Cernay, C., Pelzer, E., & Makowski, D. (2016). A global experimental dataset for assessing grain legume production. *Scientific Data*, **3**(1), 1–20.

Delgado, J., Short, N.M., Roberts, D.P., & Vandenberg, B. (2019). Big data analysis for sustainable agriculture. *Frontiers in Sustainable Food Systems*, **3**, 54.

Ellenburg, W.L., Ndungu, L.W., Mishra, V., Oware, M., Miller, S., Mithieu, F., Wahome, A.M., & Mugo, R.M. (2019). Development of a drought and yield assessment system in Kenya. Poster presented at the AGU Fall Meeting, 9–13 Dec. 2019, San Francisco, CA.

Ferentinos, K.P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, **145**, 311–318.

Hatfield, J.L., Prueger, J.H., Sauer, T.J., Dold, C., O'Brien, P., & Wacha, K. (2019). Applications of vegetative indices from remote sensing to agriculture: Past and future. *Inventions*, **4**(4), 71.

Hassani, H., Huang, X., & Silva, E. (2019). Big data and climate change. *Big Data and Cognitive Computing*, **3**(1), 12.

Huntingford, C., Jeffers, E.S., Bonsall, M.B., Christensen, H.M., Lees, T., & Yang, H. (2019). Machine learning and artificial intelligence to aid climate change research and preparedness. *Environmental Research Letters*, **14**(12), 124007.

Ip, R.H., Ang, L.M., Seng, K.P., Broster, J.C., & Pratley, J.E. (2018). Big data and machine learning for crop protection. *Computers and Electronics in Agriculture*, **151**, 376–383.

Johnson, P.M., Bennartz, R., & Camp, J.V. (2019). Using machine learning to quantify the impacts of genetically modified crops on US Midwest corn yields. *Applied Geography*, **110**, 102058.

Klemm, T., & McPherson, R.A. (2017). The development of seasonal climate forecasting for agricultural producers. *Agricultural and Forest Meteorology*, **232**, 384–399.

Laborte, A.G., Gutierrez, M.A., Balanza, J.G., Saito, K., Zwart, S.J., Boschetti, M., … Koo, J. (2017). RiceAtlas, a spatial database of global rice calendars and production. *Scientific Data*, **4**, 170074.

Ma, C., Zhang, H.H., & Wang, X. (2014). Machine learning for big data analytics in plants. *Trends in Plant Science*, **19**(12), 798–808.

Mallet, J. (1989). The evolution of insecticide resistance: have the insects won ? *Trends in Ecology & Evolution*, **4**(11), 336–340.

Martinez-Feria, R.A., & Basso, B. (2020). Unstable crop yields reveal opportunities for site-specific adaptations to climate variability. *Scientific Reports*, **10**(1), 1–10.

Meisner, M.H., Rosenheim, J.A., & Tagkopoulos, I. (2016). A data-driven, machine learning framework for optimal pest management in cotton. *Ecosphere*, **7**(3), e01263.

Newlands, N., Ghahari, A., Gel, Y.R., Lyubchich, V., & Mahdi, T. (2019). *Deep learning for improved agricultural risk management*. In Proceedings of the 52nd Hawaii International Conference on System Sciences (pp. 1033–1042), Grand Wailea, Maui, Hawaii.

Oliveira, I., Cunha, R.L., Silva, B., & Netto, M.A. (2018). A scalable machine learning system for pre-season agriculture yield forecast. *arXiv*, 1806.09244.

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Müller, C., Arneth, A., … Neumann, K. (2014). Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proceedings of the National Academy of Sciences*, **111**(9), 3268–3273.

Ryan, S.F., Adamson, N.L., Aktipis, A., Andersen, L.K., Austin, R., Barnes, L., … Cooper, C.B. (2018). The role of citizen science in addressing grand challenges in food and agriculture research. *Proceedings of the Royal Society B*, **285**(1891), 20181977.

Sankaran, S., Khot, L.R., Espinoza, C.Z., Jarolmasjed, S., Sathuvalli, V.R., Vandemark, G.J., … Pavek, M.J. (2015). Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. *European Journal of Agronomy*, **70**, 112–123.

Saravi, B., Nejadhashemi, A.P., & Tang, B. (2019). Quantitative model of irrigation effect on maize yield by deep neural network. *Neural Computing and Applications*, **31**, 1–14.

Shen, Y., Zhou, H., Li, J., Jian, F., & Jayas, D.S. (2018). Detection of stored-grain insects using deep learning. *Computers and Electronics in Agriculture*, **145**, 319–325.

Shi, Y., Thomasson, J.A., Murray, S.C., Pugh, N.A., Rooney, W.L., Shafian, S., … Rana, A. (2016). Unmanned aerial vehicles for high-throughput phenotyping and agronomic research. *PloS One*, **11**(7).

Singh, A., Ganapathysubramanian, B., Singh, A.K., & Sarkar, S., 2016. Machine learning for high-throughput stress phenotyping in plants. *Trends in Plant Science*, **21**(2), 110–124.

Tzounis, A., Katsoulas, N., Bartzanas, T., & Kittas, C. (2017). Internet of Things in agriculture, recent advances and future challenges. *Biosystems Engineering*, **164**, 31–48.

van Etten, J., Steinke, J., & van Wijk, M. (2017). How can the Data Revolution contribute to climate action in smallholder agriculture ? *Agriculture Development*, **30**, 44.

Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, **236**, 111402.

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M.J. (2017). Big data in smart farming–a review. *Agricultural Systems*, **153**, 69–80.

Wulder, M.A., Loveland, T.R., Roy, D.P., Crawford, C.J., Masek, J.G., Woodcock, C.E., … Dwyer, J. (2019). Current status of Landsat program, science, and applications. *Remote Sensing of Environment*, **225**, 127–147.

Xie, C., Wang, R., Zhang, J., Chen, P., Dong, W., Li, R., Chen, T., & Chen, H. (2018). Multi-level learning features for automatic classification of field crop pests. *Computers and Electronics in Agriculture*, **152**, 233–241.

Xu, M., Lacey, C.G., & Armstrong, S.D. (2018). The feasibility of satellite remote sensing and spatial interpolation to estimate cover crop biomass and nitrogen uptake in a small watershed. *Journal of Soil and Water Conservation*, **73**(6), 682–692.

Siddhu Jayaraman, Sathiyabhama Balasubramaniam, A machine learning paradigm for sustainable agronomy, INTERNATIONAL CONFERENCE ON GREEN COMPUTING FOR COMMUNICATION TECHNOLOGIES (ICGCCT – 2024), 10.1063/5.0263555, (020121), (2025).

More news & perspectives

Back to issue

Back to home

Text © . The authors. CC BY-NC-ND 4.0. Except where otherwise noted, images are subject to copyright. Any reuse without express permission from the copyright owner is prohibited.