ELSEVIER

Procedia
Computer Science

www.elsevier.com/locate/procedia

COMP 3610-01 Database Systems
Fall 2021

The Rise of Big Data in Agriculture: Applications, Opportunities, and Challenges

Vu Hoang Anh Nguyen

Abstract

The agriculture sector is of strategic importance for our society and economy. With the growth of the human population comes the constantly rising demand for agricultural products. Due to the sector's complexity, agricultural operators have to manage many different and heterogeneous repositories and sources of information, so there is an arising strong need to secure big data. To tackle the increasing challenges of agricultural production, the complex agricultural ecosystems need to be better understood. In this paper, I provide a review of the research dedicated to applications of data analysis techniques in relevant agricultural systems. Big data technologies and high-performance computing create new opportunities for data-intensive decision-making, enabling producers to improve productivity and show potential to support several agricultural domains. A review on works in agriculture that employ the practice of big data analysis to solve various problems is performed, and opportunities and promising areas of use are revealed. However, big data analysis has not yet been widely adopted in agriculture. The high volume and complexity of the data produced nowadays pose challenges in successfully implementing precision agriculture. This paper presents research trends and barriers that need to be overcome to realize the potential of big data analysis to revolutionize agriculture.

Keywords: Agriculture; Big Data; Data-Driven; Data Analytics; Precision Agriculture; Machine Learning; Internet of Things

1. Introduction

By 2050, the world's population is expected to be 34% larger than today (FAO, 2009). The United Nations estimates that there is a need to increase global food production to 60 percent by 2050 to keep up with the rising population. Agriculture is essentially a primary source of food for the population and raw material for a large number of industries (Silva, 2012). This poses the challenge of improving agricultural productivity while lowering its environmental footprint. With the need to produce more food using fewer inputs, agriculture is seeking new products, practices and technologies. Research activities centring on genomics, bioinformatics and computational biology of plants and animals enable scientists and organizations to better feed the world and improve the quality of food and animal crops.

Agriculture big data is playing an important role by incorporating Machine Learning (ML), the scientific field that gives machines the ability to learn without much programming. In order to obtain better productivity, people are using agricultural big data. The mechanism used in big data-based agriculture is the emergence of big data technologies and machine learning, which creates new opportunities to ease, quantify and understand data-intensive processes in agricultural operational environments as well as provides more accurate solutions. The farmers are using data to calculate harvest yields, fertilizer demands, cost savings and even to identify optimization strategies for future crops. Since technologies and sensors on farms and farm data are growing in quantity and scope, farming processes is becoming increasingly data-driven and data-enabled (Gopal & Chintala, 2020). Ultimately, my objectives of this work are:

i. to provide a comprehensive look at the applications of big data in agriculture;
ii. to enhance the awareness for the potential implications of big data analytics in agriculture, presenting existing opportunities and promising areas;
iii. to shed light on the factors that delay big data adoption in agriculture, providing future directions, open issues, and research trends to speed up adoption.

2. Methods

The primary purpose of this systematic review article is to identify pertinent research in the field of study. The research methodology process used in this systematic literature review consists of planning, implementing, and investigation result. The initial stage involves formulation of the review, recognizing its requirements, and outlining its rules including a) research questions, b) paper extraction, c) and selection of relevant papers for review. The second stage comprises extracting the relevant information from the selected papers. Lastly, present discussion, conclusion, and future work. The study might assess all the available literature from 2010 to 2021. The period is taken into consideration because the initiation of big data has been a relatively new phenomenon. Some important keywords like "big data”, “data-driven”, “machine learning”, “Internet of Things” (IoT), “farming”, “smart agriculture”, and “precision agriculture" are used. I have identified some well-known digital libraries and web sources that are used to extract the relevant works. Table 1 presents the digital libraries and web sources. Table 2 contains some keywords and some concepts related to our field of study and used certain connectors to build search strings.

The study follows four inclusions criteria: availability of full article, article in the English language, relevant research objectives, and big data technology related to agricultural management. The study also uses three exclusion criteria: no full text, no relevancy of objectives, and article highlighting only the technical side of big data technology.

Table 1. Digital Libraries and Web Sources

Name	**Digital Library**	**Web Source**
ScienceDirect (Elsevier)	×	
Springer	×	
Taylor & Francis	×	
IEEE Xplore	×	
Google Scholar		×
Research Gate		×

Table 2. Keywords utilized to extract relevant papers

Area	**Keyword(s)**	**Correlated Concepts**
Agriculture	Agriculture, Farming, Farms	Smart Agriculture, e-Agriculture, Smart Farming, Precision Agriculture, Green Evolution
Big Data	Big Data, Massive Data	Big Data, Data Mining, Data-Driven, Data-Enabled, Data Analytics
Machine Learning	Machine Learning, Deep Learning	Machine Learning, Deep Learning, AI, ANN
Internet of Things	Internet of Things	Internet of Things

3. Big Data Ontology

3.1. The 10 V-based characterizations of big data

Recent technological development led to the automation of several processes in various domains like agriculture, health care, and fraud detection, which in turn led to the generation of humungous data. Although the term “big data” is ubiquitous in various academic and non-academic publications, there is not a formal definition for the concept. According to Mauro et al., big data is the information asset characterized by high *volume*, *velocity* and *variety*, and it requires specific technology and analytical methods to be transformed into value (2016). The agricultural landscape rightly fits the big data landscape in terms of volume (more than 570 million farms in the world), velocity (tractors are now recording 100 s of data points at 5 Hz on every operation), and variety (crop types, soil-health data, weather, disease, and data formats). Other scholars specify more characteristics of big data, resulting in 10 V-based characterizations which are illustrated in Table 1. These V-based characterizations represent ten different challenges associated with the main tasks involving big data: capture, cleaning, curation, integration, storage, processing, indexing, search, sharing, transfer, mining, analysis, and visualization (Gopal & Chintala, 2020).

Table 3. The 10 V’s of Big Data

Volume	There is a large amount of data is generated from various sources.
Variety	Data have complex structures, different data types, and formats depending on the data sources.
Velocity	In real-time processing, the rate of data flowing into and out of the systems is high.
Veracity	Due to the different data sources being of many different origins, data quality is not all verifiable.
Validity	The data should have quality, governance, master data management (MDM) on massive, diverse, distributed and heterogeneous.
Value	To derive significant value from high volumes of data with a low-value density is not straightforward.
Variability	The data source should be dynamic, evolving, spatio-temporal, time series, seasonal, and any other types of non-static behaviour.
Venue	Data sources are distributed, heterogeneous data from multiple platforms, from different owners’ systems, with different access and formatting requirements,

	private vs. public cloud.
Vocabulary	It relates new concepts, schema, data models, semantics, ontologies, taxonomies, and other content- and context-based metadata that describes the data's structure, syntax, content, and provenance.
Vagueness	It relates to the confusion over the meaning of big data and overall developments around big data.

3.2. Some tools of big data analytics

These days, big data processing techniques also include cloud computing and machine learning. These technologies help in reducing manual inputs and oversight by automating many processes and tasks. The development of big data computing technologies and tools has eased the process of exploring big data for various applications. There are many industry-standard BD analytics tools such as Hadoop, MapReduce, HDFS, HIVE, and HBase (Himesh, 2018).

- Hadoop is an open-source multi-task software for data storage and running analytics.
- MapReduce is a programming model for big data processing with parallel and distributed algorithms on a cluster.
- HDFS is a Java-based file system that is modular, scalable and reliable.
- HIVE is capable of analyzing large data sets stored in Hadoop's HDFS.
- HBase is a distributed database designed to handle large tables with billions of rows and millions of columns and can provide real-time read-write access to big data.

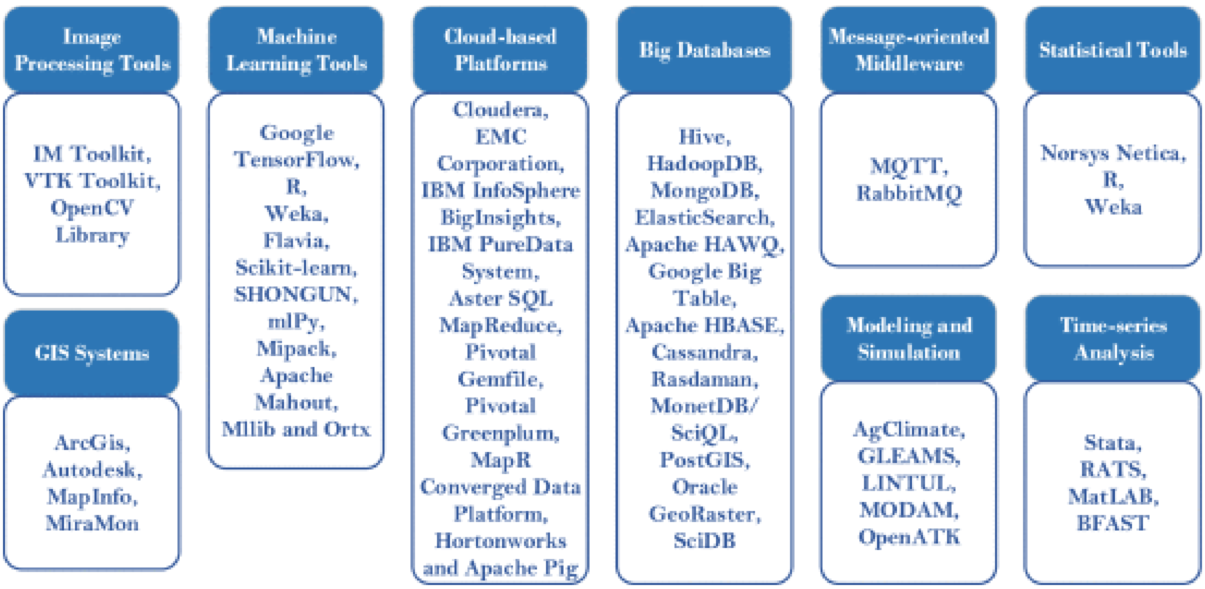

Figure 1. Software tools employed in agriculture big data analysis

4. Results

4.1. Applications of Big Data Analysis in Agriculture

Big data and the Internet of Things (IoT) are poised to revolutionize farming practices and operations of the agriculture sector. Applications of big data is a key tool to digitalize the agriculture sector. Big data applications in agriculture include smart farming, data-driven agriculture, precision farming, sensor deployment and analytics, and predictive modelling.

4.1.1. Smart Farming

The integration of big data in the agricultural sector leads to the concept of smart farming or smart agriculture. The concept of smart farming refers to farm managing using data analytics, communications systems, IoT, ICT, sensors, GPS, satellite, robots, and drones. With all of these technologies, farmers can collect data, monitor the field conditions without physically going to the field, and make strategic decisions. Smart farming is a major tool to handle and manage the threats, challenges, risks, diseases, and pest attacks and ensure sustainability (Sadiku et al., 2020). The basic management functions in smart farming are:

- Sensing and monitoring: measurement of the actual performance of the farm processes. This can be done manually by a human observer or automated by using sensing technologies such as sensors or satellites. In addition, external data can be acquired to complement direct observations.
- Analysis and decision making: compares measurements with the norms that specify the desired performance, signals deviations and decides on the appropriate intervention to remove the signalled disturbances.
- Intervention: plans and implements the chosen intervention to correct the farm processes' performance (Wolfert et al., 2017).

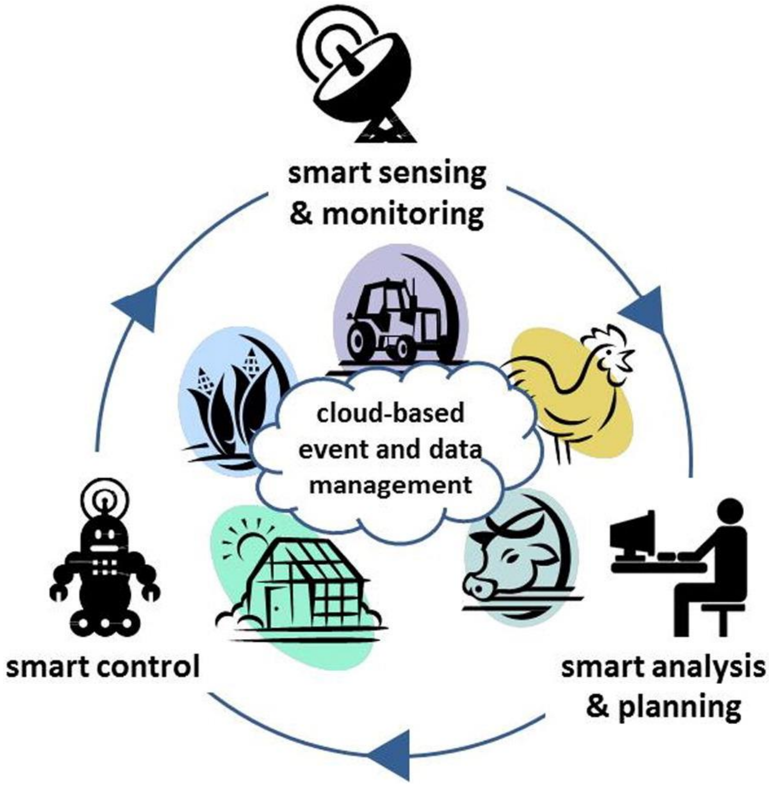

Figure 2.

The management cycle of smart farming enhanced by cloud-based event and data management

4.1.2. Precision Agriculture

The advent of Global Positioning Systems (GPS) and Global Navigation Satellite Systems (GNSS) enabled the practice of precision agriculture, which is a management strategy that uses information technologies to bring data from multiple sources to bear on decisions associated with crop production (Tantalaki et al., 2019). Precision agriculture makes a precise plant and cattle treatment possible, providing more accurate farming techniques for planting and growing crops. Figure 2 represents a precision agriculture system. The objective field data acquisition is done by deploying IoT devices, remote sensing, and other sensor networks. The data collected about soil, crops, weather, and ambient from the IoT sensor networks is stored on local or cloud storage. Where machine learning-based big data algorithms are used to abstract vital information and help incorrect decision-making by the farmer. Finally, the required action recommended by the decision system is executed physically by the advanced machinery used based on the decision received through an intelligent control system (Bhat & Huang, 2021).

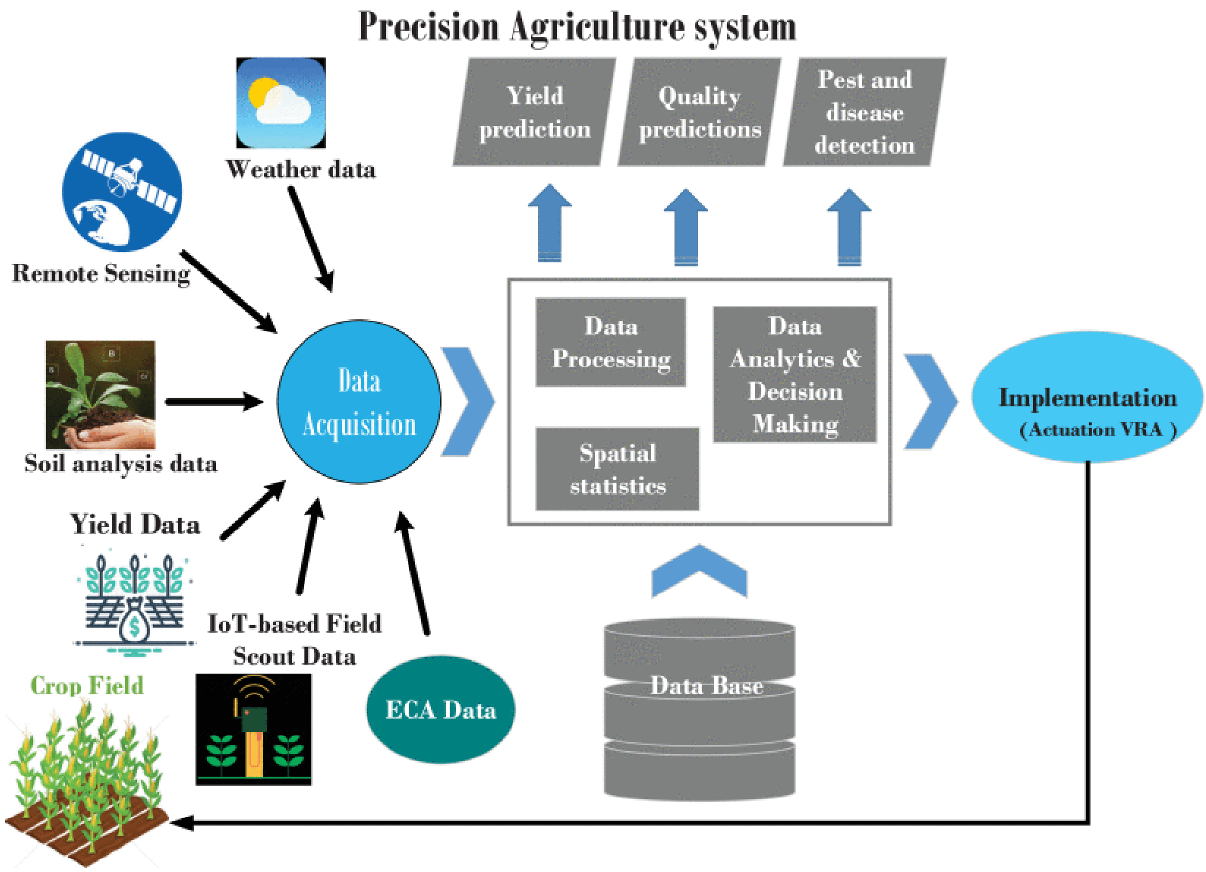

Figure 3. Big data-based precision agriculture system representation

4.2. Agricultural Big Data Systems

The agricultural field is one of the most urgent and important development areas of big data and IoT technology. A large number of sensor nodes form a variety of monitoring networks in the agricultural field through the IoT technology, and with the help of various sensors to collect information, it helps farmers to find the troubles in time and determine the location of the troubles accurately. Operators, growers, or managers can adjust their planting plans according to various data of the agricultural IOT system to maximize the profits of agricultural production. The data is collected, extracted, preprocessed, and converted to knowledge.

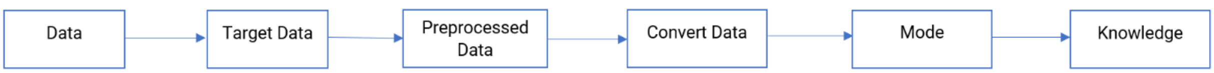

Figure 4. Data chain

4.2.1. Categories of Agricultural Big Data Systems

Agricultural big data systems can be divided into three categories. Specifically, the systems use advanced machine learning techniques and IoT technologies but have different approaches and different scopes.

- **Advanced sensor technology systems** – refers to systems that use data collection to characterize spatial and temporal variability in the production progress and determine actions to be taken in field. In these systems, prescriptive analysis is conducted to determine necessary interventions.
- **Risk management systems** – refers to systems that manage crop risk using advanced analytic techniques. These systems manage risk specifically to field location, soil type, and desired yields. They also help assess the most probable risks on a given farm. Common matters of interest in such systems are weather and climate change adaptation and mitigation.
- **Agricultural management systems** - refers to systems that provide smart farming solutions. They address farm needs such as accounting, food market access and traceability, and wireless linking to provide support for better management practices (Tantalaki et al., 2019).

5. Discussion

5.1. Challenges of Big Data Adoption in Agriculture

Big data could potentially be very useful for non-industrial farming practices, but emerging moral and ethical questions about access, cost, and support should be addressed to realize this benefit. During this initial phase, benefits from data are not so large for the farmers. Concerns are held among growers, that the benefits and risks of big data-related developments will be unevenly distributed.

- *Limited data storage and preservation*: The increasing volume, variety and velocity of agricultural big data sets demand excessive computing power and computational resources to manage and analyze. High-resolution datasets can be collected at high temporal resolution via ground-based sensors and remote sensing satellites. It has become increasingly difficult to store and maintain the data without significant investment in big data platforms such as high-performance computing.
- *Data sharing barriers*: Data privacy is a major concern of agricultural big sharing, and private data owners may be reluctant to share the data. Even if private data is shared, it is not

certain that one private dataset will be compatible with another private dataset or with other public datasets.

- *Insufficient data documentation*: To support big data analytical methods in agriculture, such as data mining, it is increasingly common for satellite imagery to be supplemented with more and more field data. Although original owners may be fully aware of their data's properties, it is often difficult for others to reuse the data or combine it with other data due to a lack of this sort of documentation (Shekhar et al., 2018).
- *Data collection challenges*: In agriculture applications, big data comes from various sources. Combining data from a variety of sources raises concerns about matters of data quality and data fusion, and the access to collected big data raises concerns about security and privacy (Tantalaki et al., 2019).

Table 4. Opportunities, challenges and cost-benefit analysis big data adoption in agriculture

Opportunities	**Benefits**
- Characterize spatial and temporal variability in soil, crop, and environmental characteristics on precise scales. - Determine growing conditions and identify needs and threats in real-time. - Predict yield, weather, and threats accurately and on time. - Explore hidden structures and extract common features on farms across large regions and time scales. - Identify wanted traits and dissect their genetics.	- Automation of agricultural procedures. - Accurate and timely decision-making. - Better personalized on-farm management practices. - Improved food access and supply chain management. - Precise and effective genome editing for plant breeding.
Challenges	**Costs**
- Data quality issues - Data heterogeneity - Data availability - Data security holes and privacy concerns - Voluminous datasets	- Data quality management techniques required - Data preparation, fusion and representation techniques required - Data initiatives and producers cultural

- Data interpretation	change needed - Laws and regulations needed - Parallel/distributed infrastructure in the cloud required - Advanced visualization techniques required

6. Conclusion

Big data is an emerging trend in agricultural applications. The agriculture sector is undergoing a new revolution driven by big data, IoT, cloud and sensor technologies. The ultimate goal of establishing big data systems for agriculture is to promote the optimization of the agricultural economy by providing more effective support and tools. Although there are some major bottlenecks in the developing world to fully exploit agriculture driven by big data, the digital revolution in agriculture is set in motion which is irreversible, and farm-to-food sector is witnessing a paradigm shift. Further research is needed to investigate how to tackle the challenges of big data adoption in agriculture.

References

Silva, J. G. D. (2012, June). *Feeding the world sustainably*. United Nations. Retrieved December 14, 2021, from https://www.un.org/en/chronicle/article/feeding-world-sustainably

Bhat, S. A., & Huang, N.-F. (2021). Big Data and Ai Revolution in Precision Agriculture: Survey and challenges. *IEEE Access*, *9*, 110209–110222. https://doi.org/10.1109/access.2021.3102227

Food and Agriculture Organization of the United Nations. (2009, October 12). *How to feed the world in 2050*. 2050 High-Level Experts Forum: The Forum. Retrieved December 14, 2021, from https://www.fao.org/wsfs/forum2050/wsfs-forum/en/

Gopal, M. P. S., & Chintala, B. R. (2020, January 1). *Big Data Challenges and opportunities in agriculture*. International Journal of Agricultural and Environmental Information Systems (IJAEIS). Retrieved December 14, 2021, from https://www.igi-global.com/article/big-data-challenges-and-opportunities-in-agriculture/244147

Himesh, S. (2018). Digital Revolution and big data: A new revolution in agriculture. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources*, *13*(021). https://doi.org/10.1079/pavsnnr201813021

Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). A review on the practice of Big Data Analysis in agriculture. *Computers and Electronics in Agriculture*, *143*, 23–37. https://doi.org/10.1016/j.compag.2017.09.037

Mauro, A. D., Greco, M., & Grimaldi, M. (2016). A formal definition of big data based on its essential features. *Library Review*, *65*(3), 122–135. https://doi.org/10.1108/lr-06-2015-0061

Sadiku, M. N. O., Ashaolu, T. J., & Musa, S. M. (2020, September 2). *Big Data in Agriculture*. International Journal of Scientific Advances. Retrieved December 15, 2021, from https://www.ijscia.com/?p=966

Shekhar, S., Schnable, P., LeBauer, D., Baylis, K., & VanderWaal, K. (2018, January 3). *Agriculture Big Data (AgBD) Challenges and Opportunities From Farm To Table: A Midwest Big Data Hub Community† Whitepaper*. Midwest Big Data Hub. Retrieved December 16, 2021, from https://midwestbigdatahub.org/agbd_whitepaper/

Tantalaki, N., Souravlas, S., & Roumeliotis, M. (2019). Data-driven decision making in precision agriculture: The rise of big data in Agricultural Systems. *Journal of Agricultural & Food Information*, *20*(4), 344–380. https://doi.org/10.1080/10496505.2019.1638264

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M.-J. (2017). Big Data in smart farming – a review. *Agricultural Systems*, *153*, 69–80. https://doi.org/10.1016/j.agsy.2017.01.023