Computers and Electronics in Agriculture 143 (2017) 23–37

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Review

A review on the practice of big data analysis in agriculture

Andreas Kamilaris*, Andreas Kartakoullis, Francesc X. Prenafeta-Boldú

GIRO Joint Research Unit IRTA-UPC, Barcelona, Spain

ARTICLE INFO

Keywords:
Big data analysis
Agriculture
Survey
Smart Farming

ABSTRACT

To tackle the increasing challenges of agricultural production, the complex agricultural ecosystems need to be better understood. This can happen by means of modern digital technologies that monitor continuously the physical environment, producing large quantities of data in an unprecedented pace. The analysis of this (big) data would enable farmers and companies to extract value from it, improving their productivity. Although big data analysis is leading to advances in various industries, it has not yet been widely applied in agriculture. The objective of this paper is to perform a review on current studies and research works in agriculture which employ the recent practice of big data analysis, in order to solve various relevant problems. Thirty-four different studies are presented, examining the problem they address, the proposed solution, tools, algorithms and data used, nature and dimensions of big data employed, scale of use as well as overall impact. Concluding, our review highlights the large opportunities of big data analysis in agriculture towards smarter farming, showing that the availability of hardware and software, techniques and methods for big data analysis, as well as the increasing openness of big data sources, shall encourage more academic research, public sector initiatives and business ventures in the agricultural sector. This practice is still at an early development stage and many barriers need to be overcome.

1. Introduction

Population growth, along with socioeconomic factors have historically been associated to food shortage (Slavin, 2016). In the last 50 years, the world's population has grown from three billion to more than six, creating a high demand for food (Kitzes et al., 2008). As the (Food and Agriculture Organization of the United Nations, 2009) estimates, the global population would increase by more than 30% until 2050, which means that a 70% increase on food production must be achieved. Land degradation and water contamination, climate change, sociocultural development (e.g. dietary preference of meat protein), governmental policies and market fluctuations add uncertainties to food security (Gebbers and Adamchuk, 2010), defined as access to sufficient, safe and nutritious food by all people on the planet. These uncertainties challenge agriculture to improve productivity, lowering at the same time its environmental footprint, which currently accounts for the 20% of the anthropogenic Greenhouses Gas (GHG) emissions (Sayer and Cassman, 2013).

To satisfy these increasing demands, several studies and initiatives have been launched since the 1990s. Advancements in crop growth modeling and yield monitoring (Basso et al., 2001), together with global navigation satellite systems (e.g. GPS) (Aqeel ur et al., 2014) have enabled precise localization of point measurements in the field, so that spatial variability maps can be created (Pierce and N., 1999), a concept known as "*precision agriculture*" (Bell et al., 1995).

Nowadays, agricultural practices are being supported by biotechnology (Rahman et al., 2013) and emerging digital technologies such as remote sensing (Bastiaanssen et al., 2000), cloud computing (Hashem et al., 2015) and Internet of Things (IoT) (Weber and Weber, 2010), leading to the notion of "*smart farming*" (Tyagi, 2016; Babinet Gilles et al., 2015). The deployment of new information and communication technologies (ICT) for field-level crop/farm management extend the precision agriculture concept (Lokers et al., 2016), enhancing the existing tasks of management and decision making by context (Kamilaris et al., 2016), situation and location awareness (Karmas et al., 2016).

Smart farming is important for tackling the challenges of agricultural production in terms of productivity, environmental impact, food security and sustainability. Sustainable agriculture (Senanayake, 1991) is very relevant and directly linked to smart farming (Bongiovanni and Lowenberg-DeBoer, 2004), as it enhances the environmental quality and resource base in which agriculture depends, providing basic human food needs (Pretty, 2008). It can be understood as an ecosystem-based approach to agriculture, which integrates biological, chemical, physical, ecological, economic and social sciences in a comprehensive way, in order to develop safe smart farming practices

* Corresponding author.
E-mail addresses: andreas.kamilaris@irta.cat (A. Kamilaris), andreas.kartakoullis@irta.cat (A. Kartakoullis), francesc.prenafeta@irta.cat (F.X. Prenafeta-Boldú).

http://dx.doi.org/10.1016/j.compag.2017.09.037
Received 24 January 2017; Received in revised form 11 September 2017; Accepted 28 September 2017
Available online 10 October 2017
0168-1699/

that do not degrade our environment.

To address the challenges of smart farming and sustainable agriculture, as (McQueen et al., 1995) and (Gebbers and Adamchuk, 2010) point out, the complex, multivariate and unpredictable agricultural ecosystems need to be better analyzed and understood. The aforementioned emerging digital technologies contribute to this understanding by monitoring and measuring continuously various aspects of the physical environment (Sonka, 2016), producing large quantities of data in an unprecedented pace (Chi et al., 2016). This implies, as (Hashem et al., 2015) note, the need for large-scale collection, storage, pre-processing, modeling and analysis of huge amounts of data coming from various heterogeneous sources.

Agricultural "*big data*" creates the necessity for large investments in infrastructures for data storage and processing (Nandyala and Kim, 2016; Hashem et al., 2015), which need to operate almost in real-time for some applications (e.g. weather forecasting, monitoring for crops' pests and animals' diseases). Hence, "*big data analysis*" is the term used to describe a new generation of practices (Kempenaar et al., 2016; Sonka, 2016), designed so that farmers and related organizations can extract economic value from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and/or analysis (Waga and Rabah, 2014; Lokers et al., 2016).

Big data analysis is successfully being used in various industries, such as banking, insurance, online user behavior understanding and personalization, as well as in environmental studies (Waga and Rabah, 2014; Cooper et al., 2013). As (Kim et al., 2014) show, governmental organizations use big data analysis to enhance their ability to serve their citizens addressing national challenges related to economy, health care, job creation, natural disasters and terrorism.

Although big data analysis seems to be successful and popular in many domains, it started being applied to agriculture only recently (Lokers et al., 2016), when stakeholders started to perceive its potential benefits (Bunge, 2014; Sonka, 2016). According to some of the largest agricultural corporations, tailoring advice to farmers based on analyzing big data could increase annual global profits from crops by about US $20 billion (Bunge, 2014).

The motivation for preparing this survey stems from the fact that big data analysis in agriculture is a modern technique with growing popularity, while recent advancements and applications of big data in other domains indicate its large potential (Kim et al., 2014; Cooper et al., 2013). Current relevant surveys (Wolfert et al., 2017; Nandyala and Kim, 2016; Waga and Rabah, 2014; Wu et al., 2016) cover mostly theoretical aspects of this technique (e.g. conceptual framework, socioeconomics, business processes, stakeholders' network) or focus on particular sub-domains such as remote sensing (Chi et al., 2016; Liaghat and Balasundram, 2010; Teke et al., 2013; Ozdogan et al., 2010; Karmas et al., 2014) and geospatial analysis (Karmas et al., 2016). Thus, the main contribution of this survey is that it presents a more focused overview of the particular problems encountered in agriculture, compared to existing surveys, where data analysis is a key aspect and solutions are found inside the big data realm. Our survey highlights the (big) data used, the methods and techniques employed, giving specific insights from a technical perspective on the potential and opportunities of big data analysis, open issues, barriers and ways to overcome them.

2. Methodology

The bibliographic analysis in the domain under study involved three steps: (a) collection of related work, (b) filtering of relevant work, and (c) detailed review and analysis of state of the art related work. In the first step, a keyword-based search for conference papers and articles was performed from the scientific databases IEEE Xplore and ScienceDirect, as well as from the web scientific indexing services Web of Science (Thomson Reuters, 2017) and Google Scholar. As search keywords, we used the following query:

"Big Data" AND ["Precision Agriculture" OR "Smart Farming" OR "Agriculture"]

In this way, we filtered out papers referring to big data but not applied to the agricultural domain. Existing surveys (Wolfert et al., 2017; Nandyala and Kim, 2016; Waga and Rabah, 2014; (Chi et al., 2016; Wu et al., 2016) were also examined for related work. From this effort, 1330 papers were initially identified. Restricting the search for papers in English only, with at least two citations, the initial number of papers was reduced to 232. Number of citations was recorded based on Google Scholar. An exception was made for papers published in 2016–2017, where zero citations were acceptable.

In the second step, we checked these 232 papers whether they *made actual use of big data analysis* in some agricultural application. Use of big data analysis was quantified as satisfying some of its five "V" characteristics (Chi et al., 2016) (see Section 3). We primarily targeted the first three "V"s (i.e. volume, velocity and variety), since dimensions V4 and V5 (i.e. veracity and valorization) were more difficult to quantify. From the 232 papers, only 34 qualified according to our constraints. We were forced to discard (also) a small number of interesting efforts which did not qualify in terms of the data analysis employed and the solutions provided. In the final step, the 34 papers selected from the previous step were analyzed one-by-one, considering the problem they addressed, solution proposed, impact achieved (if measurable), tools, systems and algorithms used, sources of data employed and which "V" dimensions of big data they satisfied.

3. Big data in agriculture

Chi et al. (2016) characterize big data according to the following five dimensions:

- **Volume (V1)**: The size of data collected for analysis.
- **Velocity (V2)**: The time window in which data is useful and relevant. For example, some data should be analyzed in a reasonable time to achieve a given task, e.g. to identify pests (PEAT UG, 2016) and animal diseases (Chedad et al., 2001).
- **Variety (V3)**: Multi-source (e.g. images, videos, remote and field-based sensing data), multi-temporal (e.g. collected on different dates/times), and multi-resolution (e.g. different spatial resolution images) as well as data having different formats, from various sources and disciplines, and from several application domains.
- **Veracity (V4)**: The quality, reliability and potential of the data, as well as its accuracy, reliability and overall confidence.
- **Valorization (V5)**: The ability to propagate knowledge, appreciation and innovation.

Although these five "V"s can describe big data, big data analysis does not need to satisfy all five dimensions (Rodriguez et al., 2017). Big data is generally notorious for being less accurate and stable, usually compromising V4 (veracity). Another relevant "V" could be **visualization**, meaning the need of presenting complex data structures and rich information in an easy-to-understand way (Hashem et al., 2015; Karmas et al., 2016).

According to the above, as (Wolfert et al., 2017) explain, big data is less a matter of data volume than the capacity to search, aggregate, visualize and cross-reference large datasets in reasonable time. It is about the capability to extract information and insights where previously it was economically or technically not feasible to do so (Sonka, 2016).

In the following subsections, the most relevant research efforts, case studies and techniques in terms of solving agricultural problems by using big data analysis are discussed (Section 3.1), together with sources of big data (Section 3.2) and specific techniques employed for big data analysis (Section 3.3).

Table 1
Agricultural (general) areas and big data use. The authors performed an estimation of the first three "V"s of big data (volume, velocity, and variety), using simple indicators such as low (L), medium (M) or high (H). These estimations were based on the following: big data-related information were recorded from the 34 papers under review and then these papers were compared among them to create relative rankings, labeling each "V" dimension of each paper as L, M or H.

No.	Agri-area	No. of papers	V1 (Volume)	V2 (Velocity)	V3 (Variety)	Ref.
1.	Weather and climate change	4	M	M	H	Tripathi et al. (2006), Fuchs and Wolff (2011), Schnase et al. (2014), and Tesfaye et al. (2016)
2.	Land	5	H	L	M	Barrett et al. (2014), Schuster et al. (2011), Galford et al. (2008), Wardlow et al. (2007), and Thenkabail et al. (2007)
3.	Animals' research	4	M	H	L	McQueen et al. (1995), Kempenaar et al. (2016), Chedad et al. (2001), and Pierna et al. (2004)
4.	Crops	3	M	M	L	Waldhoff et al. (2012), Sakamoto et al. (2005), and Urtubia et al. (2007)
5.	Soil	2	M	L	L	Armstrong et al. (2007), and Meyer et al. (2004)
6.	Weeds	1	L	H	L	Gutiérrez et al. (2008)
7.	Food availability and security	4	M	L	M	Frelat et al. (2016), Jóźwiaka et al. (2016), Lucas and Chhajed (2004), and RIICE Partnership (2014)
8.	Biodiversity	1	M	L	H	Marcot et al. (2001)
9.	Farmers' decision making	2	H	M	H	Sawant et al. (2016), and Field to Market (2015)
10.	Farmers' insurance and finance	5	H	M	M	GSMA (2014), Syngenta Foundation for Sustainable Agriculture (2016), Global Envision (2006), Syngenta (2010), and Akinboro (2016)
11.	Remote sensing	3	H	M	M	Becker-Reshef et al. (2010), Nativi et al. (2015), and Karmas et al. (2014)

3.1. Applications of big data analysis in agriculture

As described in Section 2, 34 different applications of big data analysis in agriculture were selected for further analysis. This analysis covered the agricultural area concerned, the particular problem tackled, the solution and/or impact through the analysis performed, tools/algorithms used for addressing the problem, sources of data as well as an estimation (by the authors) of the first three "V"s of big data (i.e. volume, velocity, variety), using only simple indicators such as low (L), medium (M) or high (H). These estimations were based on the following: we first recorded all big data-related information from the papers we reviewed and secondly we compared these papers among them to create relative rankings, labeling each "V" dimension of each paper as L, M or H. As mentioned previously, dimensions V4 and V5 were not considered, as they are difficult to quantify. Our complete analysis for each of the 34 papers is provided in Appendix A.

Table 1 presents the general agricultural areas related with the papers identified in the survey. The third column of Table 1 shows the number of papers providing solutions in some area. Most of the studies deal with food availability and security, farmers' insurance and finance, weather and climate change, land management and animals-based research. From Table 1, columns 4–6 indicate the average rating by the authors of the first three "V"s of big data (volume, velocity and variety), as used at each agricultural area.

Most of the papers involve medium-to-high volumes of data, with medium-to-low velocity and variety. Exceptions are the animals- and weeds-related projects, which employ high velocity (considering that evidence of weeds and diseases require urgent actions). Also, exceptions are the weather and climate change-related efforts, biodiversity-based and farmers' decision making apps, which are characterized by a high variety of information (considering the need for various different data sources to forecast weather, model climate change, estimate biodiversity or effectively assist farmers' everyday tasks).

The highest volume of data appears in remote sensing applications, due to the large size of the images. The lowest velocity occurs in land-related projects, papers on food availability and security, biodiversity and soil analysis. Finally, the lowest variety appears in weeds-, soil-, crops- and animals-related research, and in remote sensing. This indicates that these areas do not require (or do not have access to) a variety of data, in order to address their particular problems.

While Table 1 listed the "V" dimensions in general agricultural research areas, Table 2 depicts which "V" characteristics are *highly* used in the *particular* agricultural applications of the papers under study. Applications related to estimations of crop production and yields, land mapping, weather forecasting and food security require large volumes of data. Recognition of animals' diseases and plants' poor nutrition require high velocity, as well as decisions on farmers' productivity, weather forecasting and safety/quality of food, which need to be taken in (near) real time. In these cases, the time horizon of the decisions involved requires operational or tactical planning, instead of longer-term strategic planning (i.e. lower velocity). Some applications such as insurance indexes and farmers' (sustainable) productivity improvement require a wide variety of data from heterogeneous sources. Finally, some applications require high reliability of the data involved, such as diseases of plants/animals, herd culling, farmers' improvement of productivity and estimations of yield.

In general, big data analysis in smart farming is still at an early development stage, and this can be inferred from the currently limited number of scientific publications and commercial initiatives. This fact is supported by the findings listed in Fig. 1, produced by Web of Science (Thomson Reuters, 2017), based on the search:

"Big Data" AND ["Agriculture" OR "Farm"]

returning 110 papers (only a fraction of which were actually relevant) and 140 citations in the last 4 years. In contrast to the rather profuse bibliography on big data, its subset on agriculture appears to be relatively recent and limited. When compared to the other research areas employing big data analysis, "*agriculture*" ranks at the 56th position in published papers. However, the findings indicate a rising trend.

3.2. Sources of big data

Considering the analysis performed, it is worth examining the sources of big data used by the different papers under study, which originate from various different sources, such as the farmers' field, i.e. from ground sensors (e.g. chemical detection devices, biosensors, weather stations, etc.) (Chedad et al., 2001; Kempenaar et al., 2016), (historical) data gathered by governmental and third-party organizations (e.g. statistical yearbooks, governmental reports, regulations[1] and guidelines from public bodies, alerts, etc.), distributed via online repositories and web services (McQueen et al., 1995; Tesfaye et al., 2016), data from airborne sensors (e.g. unmanned aerial vehicles, airplanes and satellites) (Becker-Reshef et al., 2010; Gutiérrez et al., 2008), real-time web data from private companies through online web services (GSMA, 2014; Syngenta, 2010), crowdsourcing-based

[1] See also the supplementary material of this publication.

Table 2
Big data use in different agricultural applications.

"V" Dimension	"V" Description	Agricultural (specific) applications
V1	High volume, large quantities of data	Weather forecasting (Tripathi et al., 2006), dairy herd culling (McQueen et al., 1995), crop identification (Waldhoff et al., 2012), farmers' productivity improvement (GSMA, 2014), small farmers' insurance and protection (Syngenta, 2010), farmers financing (Global Envision, 2006), crop production estimations (Becker-Reshef et al., 2010), food security estimations based on remote sensing (RIICE Partnership, 2014), land use and land cover changes classification (Wardlow et al., 2007), data sharing of earth observations (Nativi et al., 2015)
V2	High velocity, when (real) time becomes important and very relevant	Weather forecasting (Tripathi et al., 2006), wine fermentation (Urtubia et al., 2007), safety and quality of animal food (Pierna et al., 2004), weed discrimination (Gutiérrez et al., 2008), animal disease recognition (Chedad et al., 2001), farmers' productivity improvement (GSMA, 2014), farmers' financial transactions in remote areas (Akinboro, 2016), data sharing of earth observations (Nativi et al., 2015)
V3	High variety, when data come from various heterogeneous sources	Management zones identification (Schuster et al., 2011), food availability estimation in developing countries (Frelat et al., 2016), wildlife population evaluation (Marcot et al., 2001), small farmers' insurance and protection (Syngenta, 2010), farmers' productivity improvement (GSMA, 2014), farmers' understanding of sustainability performance and operational efficiency (Field to Market, 2015), crops' drought tolerance (Fuchs and Wolff, 2011; Tesfaye et al., 2016), climate science (Schnase et al., 2014)
V4	High veracity, when accuracy and reliability of data is of critical importance	Dairy herd culling (McQueen et al., 1995), safety and quality of animal food (Pierna et al., 2004), weed discrimination (Gutiérrez et al., 2008), animals' disease recognition (Chedad et al., 2001), food availability estimation in developing countries (Frelat et al., 2016), small farmers' insurance and protection (Syngenta, 2010), farmers' productivity improvement (GSMA, 2014), data sharing of earth observations (Nativi et al., 2015), evaluate wildlife population viability (Marcot et al., 2001)

techniques from mobile phones (e.g. transportation data, information about plants, crops, yields, weather conditions, etc.) (Akinboro, 2016; Global Envision, 2006), feeds from social media (e.g. mentioning of natural hazards happening, pests/diseases identified in various farms and fields) (Sawant et al., 2016), etc.

The aforementioned sources are mostly heterogeneous and the data differs in volume and velocity. They are represented in different types and formats, while access to the data varies as well (e.g. web services, static repositories, live feeds, files and archives etc.).

Table 3 lists the different sensors and data sources (column 3) employed at each agricultural area (column 2). Each agricultural application requires different sources of big data to address the problem it tackles. Almost in all agricultural areas, information from static databases and datasets is being used, while geospatial data and data from satellite-based remote sensing are quite popular. Crop-, soil- and animal-related research utilize ground sensors deployed at the field, while weather stations are employed in weather and climate change applications, land mapping, farmers' decision making, insurance and finance. Except from a few approaches (Tripathi et al., 2006; Chedad et al., 2001; Meyer et al., 2004; Armstrong et al., 2007; Urtubia et al., 2007; Jóźwiaka et al., 2016; GSMA, 2014), most of the reviewed papers combine data from a variety of sources to address their particular problems.

3.3. Techniques and tools for big data analysis

Table 3 also presents the particular techniques and approaches (column 4) employed in the different agricultural areas which are considered in the papers under review (column 2). All the 34 studies analyze big data according to some particular technique or combination of methods, such as those listed in (Vitolo et al., 2015; Mucherino et al., 2009). As Table 3 shows, machine learning (used in 13 papers), cloud-based platforms (9), image processing (8), modeling and simulation (7), statistical analysis (6) and NDVI vegetation indices (6) are the most commonly used techniques, while some approaches employ online services (e.g. publish/subscribe messaging, online portals, decision support) (5) and geographical information systems (GIS) (4). Except from soil applications, the rest employ a combination of techniques, especially land applications with remote sensing (Barrett et al., 2014; Wardlow et al., 2007; Becker-Reshef et al., 2010).

Machine learning tools (Ma et al., 2014; Mucherino et al., 2009) are used in prediction (Tripathi et al., 2006; Kempenaar et al., 2016), clustering (Pierna et al., 2004) and classification problems (Armstrong et al., 2007; Meyer et al., 2004) while image processing (Chi et al.,

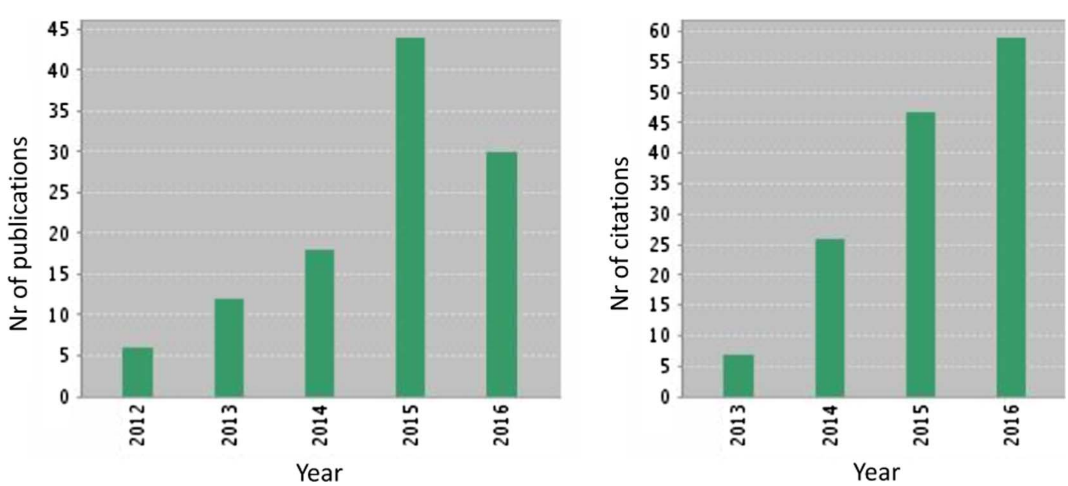

Fig. 1. Research publications related to "big data in agriculture" (left) and number of citations from these publications (right).
Source: Web of Science.

Table 3
Sources of big data and techniques for big data analysis per agricultural area.

No.	Agricultural area	Big data sources	Techniques for big data analysis	Ref.
1.	Weather and climate change	Weather stations, surveys, static historical information (weather and climate data, earth observation data), remote sensing (satellites), geospatial data	Machine learning (scalable vector machines), statistical analysis, modeling, cloud platforms, MapReduce analytics, GIS geospatial analysis	Tripathi et al. (2006), Fuchs and Wolff (2011), Schnase et al. (2014), and Tesfaye et al. (2016)
2.	Land	Remote sensing (satellites, synthetic aperture radar, airplanes), geospatial data, historical datasets (land characterization and crop phenology, rainfall and temperature, elevation, global tree cover maps), camera sensors (multispectral imaging), weather stations	Machine learning (scalable vector machines, K-means clustering, random forests, extremely randomized trees), NDVI vegetation indices, Wavelet based filtering, image processing, statistical analysis, spectral matching techniques, reflectance and surface temperature calculations	Barrett et al. (2014), Schuster et al. (2011), Galford et al. (2008), Wardlow et al. (2007), and Thenkabail et al. (2007)
3.	Animals' research	Historical information about soils and animals (physiological characteristics), ground sensors (grazing activity, feed intake, weight, heat, milk production of individual cows, sound), camera sensors (multispectral and optical)	Machine learning (decision trees, neural networks, scalable vector machines)	McQueen et al. (1995), Kempenaar et al. (2016), Chedad et al. (2001), and Pierna et al. (2004)
4.	Crops	Ground sensors (metabolites), remote sensing (satellite), historical datasets (land use, national land information, statistical data on yields)	Machine learning (scalable vector machines, K-means clustering), Wavelet based filtering, Fourier transform, NDVI vegetation indices	Urtubia et al. (2007), Waldhoff et al. (2012), and Sakamoto et al. (2005)
5.	Soil	Ground sensors (salinity, electrical conductivity, moisture), cameras (optical), historical databases (e.g. AGRIC soils)	Machine learning (K-means clustering, Farthest First clustering algorithm)	Armstrong et al. (2007), and Meyer et al. (2004)
6.	Weeds	Remote sensing (airplane, drones), historical information (digital library of images of plants and weeds, plant-specific data)	Machine learning (neural networks, logistic regression), image processing, NDVI vegetation indices	Gutiérrez et al. (2008)
7.	Food availability and security	Surveys, historical information and databases (e.g. CIALCA, ENAR, rice crop growth datasets), GIS geospatial data, statistical data, remote sensing (synthetic aperture radar)	Machine learning (neural networks), statistical analysis, modeling, simulation, network-based analysis, GIS geospatial analysis, image processing	Frelat et al. (2016), Jóźwiaka et al. (2016), Lucas and Chhajed (2004), and RIICE Partnership (2014)
8.	Biodiversity	GIS geospatial data, historical information and databases (SER database of wildlife species	Statistics (Bayesian belief networks)	Marcot et al. (2001)
9.	Farmers' decision making	Static historical information and datasets (e.g. US government survey data), remote sensing (satellites, drones), weather stations, humans as sensors, web-based data, GIS geospatial data, feeds from social media	Cloud platforms, web services, mobile applications, statistical analysis, modeling, simulation, benchmarking, big data storage, message-oriented middleware	Sawant et al. (2016), and Field to Market (2015)
10.	Farmers' insurance and finance	Web-based data, historical information, weather stations, humans as sensors (crops, yields, (financial transactions data)	Cloud platforms, web services, mobile applications	GSMA (2014), Syngenta (2010), Global Envision (2006), Syngenta Foundation for Sustainable Agriculture (2016), Akinboro (2016)
11.	Remote sensing	Remote sensing (satellite, airplane, drones), historical information and datasets (e.g. MODIS surface reflectance datasets, earth land surface dataset of images, WMO weather datasets, reservoir heights derived from radar altimetry, web-based data, geospatial data (imaging, maps)	Cloud platforms, statistical analysis, GIS geospatial analysis, image processing, NDVI vegetation indices, decision support systems, big data storage, web and community portals, MapReduce analytics, mobile applications, computer vision, artificial intelligence	Becker-Reshef et al. (2010), Nativi et al. (2015), and Karmas et al. (2014)

2016; Manickavasagan et al., 2005) is used when data originates from images (i.e. cameras (Meyer et al., 2004) and remote sensing (Sakamoto et al., 2005)). Image processing includes algorithms for Fourier and harmonic analysis, wavelet decomposition and curve fitting (Galford et al., 2008; Sakamoto et al., 2005). These tools are used together with remote sensing (Mucherino et al., 2009) and are often combined, e.g. to use the image processing output as input to a machine learning model (Vibhute and Bodhe, 2012).

Cloud platforms (together with MapReduce) (Hashem et al., 2015) offer possibilities for large-scale storing (Nativi et al., 2015), pre-processing, analysis and visualization of data (Becker-Reshef et al., 2010), while GIS (Lucas and Chhajed, 2004) are used in geospatial problems (Barrett et al., 2014; RIICE Partnership, 2014). Big datasets are appropriate for the storing of large volumes of heterogeneous information, using database management systems (DBMS) that implement the array data model and NoSQL database management platform (Karmas et al., 2016). NoSQL platforms store and manage large unstructured data. Array DBMS are built specifically for serving big raster datasets.

Moreover, related to remote sensing applications, vegetation indices (VI) are frequently used for crops/soils mapping (Sakamoto et al., 2005; Barrett et al., 2014), defined as combinations of surface reflectance at two or more wavelengths designed to highlight a particular property of vegetation. The most popular one is the normalized difference vegetation index (NDVI), which is a graphical indicator used to analyze remote sensing measurements, and to assess whether the target being observed contains green vegetation or not (Wardlow et al., 2007; Thenkabail et al., 2007).

Finally, some online services involve message-oriented middleware, which serve in event-based systems where fast notifications and alerts are necessary, e.g. when some natural hazard is happening (GSMA, 2014; Global Envision, 2006).

4. Discussion

As Section 3 illustrated, a large variety of agricultural issues are currently approximated by the use of big data analysis, employing a variety of different algorithms, approaches and techniques (see Table 3). Table 4 presents the specific, most common software used in the revised papers, mapped according to the type of analysis employed. A wide variety of software available for big data analysis exist, in all different types of analysis.

A recent practice is to approximate agricultural problems by employing image analysis (Karmas et al., 2014; Teke et al., 2013), using images originating from remote sensing (Atzberger, 2013), either from airborne (Gutiérrez et al., 2008; Barrett et al., 2014; Wardlow et al., 2007) or satellites (Schnase et al., 2014; Waldhoff et al., 2012; Sakamoto et al., 2005; RIICE Partnership, 2014; Becker-Reshef et al., 2010). This is in line with recent statistics showing a tremendous increase in the use of Landsat satellite data after 2008, when the archive became available at no cost (Wulder et al., 2012). Remote sensing has several advantages when applied to agriculture (Teke et al., 2013), being a well-known, non-destructive method to collect information systematically over very large geographical areas. A modern application of remote sensing in agriculture, as observed from the surveyed papers, is on the delivery of operational insurance products such as insurance from crop damage (de Leeuw et al., 2014; Global Envision, 2006), flood and fire risk assessment (de Leeuw et al., 2014), or from drought and excess rain (Syngenta, 2010).

For large areas such as countries, the primary sources of data have typically been coarse-resolution satellites with wide-area coverage (Atzberger, 2013), such as AVHRR, MODIS, MERIS and SPOT-VEGETATION. From the surveyed papers, relevant agricultural applications include weather forecasting (Schnase et al., 2014; Becker-Reshef et al., 2010) and characterization/mapping of crops (Galford et al., 2008), soil and land (Barrett et al., 2014; Schuster et al., 2011). On the other hand, data from high spatial resolution satellites like Landsat and SPOT have been used in support of local- and regional-scale applications requiring increased spatial detail (Ozdogan et al., 2010), such as farmers' decision making support (Sawant et al., 2016). Finally, some papers use airplanes and drones to achieve their goals, focusing on weeds' identification (Gutiérrez et al., 2008), or grassland inventories (Barrett et al., 2014). Combining remote sensing with ancillary data (e.g. GIS data, historical data, field sensors, etc.) significantly improves the analysis performed, especially when it includes some form of prediction, i.e. crop identification (Waldhoff et al., 2012) or accuracy of distinguishing grasslands (Barrett et al., 2014).

The relatively low identification of related work, mostly produced in the last 8–10 years, indicates that big data in smart farming is still at an early development stage, an observation made also by Lokers et al. (2016) and Bunge (2014), however with increasing adoption, use and application in various agricultural areas, as discussed by Sonka (2016). The increasing number of scientific peer-reviewed publications indicate the large potential of big data analysis applied in the domain of agriculture. The field presents a high degree of dynamics with a diverse stakeholders' network and new players entering continuously the market, mostly in the form of high-tech companies and start-ups, such as (SenseFly, 2012; aWhere Inc., 2015; PEAT UG, 2016; Blue River Technology, 2011), taking existing and creating new roles in agricultural big data analysis and management. As Hashem et al. (2015) and Kempenaar et al. (2016) point out, high-tech companies, together with large investments in cloud platforms for larger storage and computation capabilities, could create new services and business analytics at a new scale and speed, inventing new business models.

4.1. Open problems of big data in agriculture

The application of big data analysis in agriculture has not been beneficial in all cases, as it has created (or is expected to create) some problems too. We list below the problems, as identified and mentioned

Table 4
Common software tools used for big data analysis in agriculture.

No.	Category	Software tools
1.	Image processing tools	IM toolkit, VTK toolkit, OpenCV library
2.	Machine learning (ML) tools	Google TensorFlow, R, Weka, Flavia, scikit-learn, SHOGUN, mlPy, Mlpack, Apache Mahout, Mllib and Oryx
3.	Cloud-based platforms for large-scale information storing, analysis and computation	Cloudera, EMC Corporation, IBM InfoSphere BigInsights, IBM PureData system for analytics, Aster SQL MapReduce, Pivotal GemFire, Pivotal Greenplum, MapR converged data platform, Hortonworks and Apache Pig
4.	GIS systems	ArcGIS, Autodesk, MapInfo, MiraMon
5.	Big databases	Hive, HadoopDB, MongoDB, ElasticSearch, Apache HAWQ, Google BigTable, Apache HBASE, Cassandra, Rasdaman, MonetDB/SciQL, PostGIS, Oracle GeoRaster, SciDB
6.	Message-oriented middleware	MQTT, RabbitMQ
7.	Modeling and simulation	AgClimate, GLEAMS, LINTUL, MODAM, OpenATK
8.	Statistical tools	Norsys Netica, R, Weka
9.	Time-series analysis	Stata, RATS, MatLab, BFAST

in the revised papers:

- From a sociopolitical perspective, creation of large monopolies in the agri-food industry and dependence of the farmers on large corporations about their farming operations becomes possible (Sykuta, 2016). Big data concentrated in the hands of big agri-businesses limits the potential of this technology, only reinforcing the capacities and business advantages of a few corporations (Carbonell, 2016).
- Privacy issues are raised, in respect to who owns the data and who can monetize it (Nandyala and Kim, 2016). Farmers are concerned about the potential misuse of information related to their farming activities (Shin and Choi, 2015), by seed companies or competitor farms (Carolan, 2016). (Schuster, 2017) warns that hedge funds might use real-time data at harvest time from a large number of sources (e.g. weather data, yields predictions, remote sensing, data from machinery such as combines, etc.) to speculate in commodity markets.
- The practice of big data collection and analytics has raised questions over its security, accuracy and access, as discussed in (Nandyala and Kim, 2016) and (Sykuta, 2016).
- Moreover, the use of big data differs in developed vs. developing countries, according to Kshetri (2014) and Rodriguez et al. (2017). RIICE Partnership (2014) and Syngenta (2010) believe a digital divide exists between developed and developing economies, due to unbalanced access to technology (i.e. computing power, internet bandwidth and sophisticated software), and lack of skilled analysts in the developing world. Especially in respect to volume and variety, big data in the developing world is smaller-scale and less diverse (Rodriguez et al., 2017), as the surveyed papers suggest (Tesfaye et al., 2016; Frelat et al., 2016; Sawant et al., 2016; GSMA, 2014; Akinboro, 2016). Big data collection efforts mainly benefit big, well-educated farmers who have the means and the expertise to collect it successfully and accurately (Kshetri, 2014; Oluoch-Kosura, 2010).
- From a technical perspective, product developers have only limited access to ground truth information (Atzberger, 2013), an issue that has been observed in many of the revised papers (Armstrong et al., 2007; Waldhoff et al., 2012; Sakamoto et al., 2005; Jóźwiaka et al., 2016; Frelat et al., 2016). Ground truth information is necessary for evaluating products and services under various settings and physical or weather conditions (Capalbo et al., 2016). Also, visualization of large data volumes is still difficult (Schnase et al., 2014; Karmas et al., 2016).

4.2. Barriers for wider adoption of big data analysis

Related work indicated various barriers hindering the wider use of big data analysis, such as lack of human resources and expertise (Sawant et al., 2016) and limited availability of reliable infrastructures to collect and analyze big data (Akinboro, 2016; Syngenta Foundation for Sustainable Agriculture, 2016). Frelat et al. (2016) note that accurate and actionable data requires considerable technical skills to handle data mining and analysis methods, while infrastructures are needed for efficient data storage, management and processing of multi-modal and high-dimensional datasets, including provisioning for real-time processing in many critical geospatial applications (Karmas et al., 2016).

Further, there is generally a lack of structure and governance related to agricultural big data, as pointed out by Nandyala and Kim (2016) and Nativi et al. (2015), as well as identified and addressed in some of the revised papers (Schnase et al., 2014; Marcot et al., 2001; Becker-Reshef et al., 2010). Kempenaar et al. (2016) suggest that business models are needed that are attractive enough for solution providers, enabling at the same time a fair share between the different stakeholders. In addition, Lokers et al. (2016) consider that the general absence of well-defined semantics complicates big data understanding and reuse by other researchers and organizations.

As observed in this study (see Table 1), much of the attention on big data has focused mainly on large volumes (i.e. applications in weather and climate change, land identification, farmers' decision-making, insurance and finance, remote sensing). This has led to a skewed and narrow perspective of the value of big data to organizations and the society since aspects of data velocity, variety, veracity and valorization are equally important, as pointed out by Shin and Choi (2015) and Capalbo et al. (2016).

Moreover, technical challenges of remote sensing systems for farm management still exist (Zhang and Kovacs, 2012), such as the collection and delivery of images in a timely manner (Galford et al., 2008), sampling errors and the lack of high spatial resolution data (Nativi et al., 2015), image interpretation and data extraction issues (Karmas et al., 2014), the influence of weather conditions (Barrett et al., 2014), etc. Finally, common barriers involve the absence of data itself (or part of it) and its limited reliability, variety or time relevance as observed and discussed in some of the revised papers (Fuchs and Wolff, 2011; Schnase et al., 2014; Schuster et al., 2011; Armstrong et al., 2007; Frelat et al., 2016; RIICE Partnership, 2014; Marcot et al., 2001).

4.3. Addressing open problems and overcoming barriers

From a sociopolitical view, many farmers from around the world started to mobilize and organize themselves (e.g. in cooperatives, online communities), increasing their power in terms of sharing of know-how and experiences, and big data understanding (Farm Hack, 2010). Shin and Choi (2015) believe that the data-driven economy has the potential to create suitable knowledge for the users of data ecosystems, such as the one of agriculture.

Moreover, formation of a policy framework on data ownership is required Sykuta (2016)), which will protect owners' copyrights and control user access (Shin and Choi, 2015). Policies for data management and security are needed (Kshetri, 2014), towards the democratization of big data, broadening its potential impact and value through the adoption and accessibility of appropriate support tools (World Economic Forum, 2012).

From a technical aspect, investments in cloud infrastructures are essential for large-scale storage, analysis and visualization of agricultural data (Hashem et al., 2015), supporting business analytics in high scale and speed (Kempenaar et al., 2016). The infrastructures should be easily accessible by non-technical personnel and not be expensive (Hashem et al., 2015). Techniques such as data aggregation, data reduction and proper analysis can contribute towards more user-friendly platforms (Karmas et al., 2014). Various ways of technically managing high volumes of widely varied data, addressing the "V"s dimensions of big data effectively are discussed in Karmas et al. (2016) and Nativi et al. (2015).

Also, well-defined and commonly accepted technologies for data semantics (e.g. RDF, OWL, SPARQL and Linked Data) and ontologies (e.g. AGROVOC, Agricultural Ontology Service and AgOnt) can be used as common terminologies towards data interoperability, as proposed in Lokers et al. (2016), Cooper et al. (2013), Kamilaris et al. (2016). Karmas et al. (2016) suggest that open standards need to be adopted and agreed for data integration, such as OGC.

Furthermore, open-source software tools and libraries would be useful, such as crop type maps and calendars (Sakamoto et al., 2005), biophysical measures and vegetation indices (Wardlow et al., 2007), yield models (RIICE Partnership, 2014), crop area estimates (Becker-Reshef et al., 2010) and seasonal weather forecasts (Tripathi et al., 2006). These tools should be easily mergeable with other platforms to support large-scale, highly-varied data analysis, a strategic goal of the GEOSS platform (Nativi et al., 2015) and MERRA Services (Schnase et al., 2014).

More big datasets should become publicly available (Carbonell, 2016), and there is a growing trend towards this direction already (OADA, 2014; GODAN, 2015; AgGateway, 2005). Besides, numerous organizations on the web have started to provide various large-scale datasets covering a wide spectrum of agricultural areas.[2]

4.4. Potential areas of application of big data analysis

This subsection lists potential areas of applying big data analysis for addressing various agriculture-related problems in the future. These areas have not been covered adequately (or not covered at all) by the existing research and papers under study (according to the authors' opinion), and include the following possibilities:

1. Platforms enabling supply chain actors to have access to high-quality products and processes (RIICE Partnership, 2014; Sawant et al., 2016), enabling crops to be integrated to the international supply chain, according to the global needs (Cropster, 2007; Syngenta Foundation for Sustainable Agriculture, 2016).
2. As farmers are sometimes not able to sell harvests due to over-supply or not getting the planned harvest (Frelat et al., 2016; Syngenta Foundation for Sustainable Agriculture, 2016), tools for better yield and demand predictions must be developed (Tesfaye et al., 2016; Kempenaar et al., 2016; Becker-Reshef et al., 2010).
3. Providing advice and guidance to farmers based on their crops' responsiveness to fertilizers is likely to lead to a more appropriate management of fertilizer use (Giller et al., 2011). This could apply as well to better use of herbicides and pesticides (i.e. (Gutiérrez et al., 2008; Sawant et al., 2016; aWhere Inc., 2015)).
4. Scanning equipment in plants, shipment tracking and retail monitoring of consumers' purchases creates the potential to enhance products' traceability through the supply chain (Armbruster and MacDonell, 2014), increasing food safety (Jóźwiaka et al., 2016; Lucas and Chhajed, 2004). Prevention of foodborne illnesses is an issue that requires international collaboration and investment by local/global organizations and governments (Grace and McDermot, 2015), to ensure safer food (Chedad et al., 2001; RIICE Partnership, 2014). Moreover, since the agricultural production is prone to deterioration after harvesting (Wari and Zhu, 2016), optimization procedures are essential to minimize losses and maximize quality (Pierna et al., 2004). Promising optimization techniques already being applied to food processing involve (meta-) heuristics and genetic algorithms (Wari and Zhu, 2016), as well as neural networking (Erenturk and Erenturk, 2007).
5. Remote sensing for large-scale land/crop mapping will be critical for monitoring the impacts of various countries and areas in respect to measuring and achieving their productivity and environmental sustainability targets (Barrett et al., 2014; Waldhoff et al., 2012; Schuster et al., 2011; Becker-Reshef et al., 2010).
6. More advanced and complete scientific models and simulations for environmental phenomena could provide a basis for establishing platforms for policy-makers, assisting in decision-making towards sustainability of physical ecosystems (Schnase et al., 2014; Nativi et al., 2015; Song et al., 2016).
7. High-throughput screening methods that can offer quantitative analysis of the interaction between plants and their environment, with high precision and accuracy (Furbank and Teste, 2011; Karmas et al., 2014).
8. Self-operating agricultural robots could revolutionize agriculture and its overall productivity, as they may automatically identify and remove weeds (Gutiérrez et al., 2008; Blue River Technology, 2011), identify and fight pests (PEAT UG, 2016), harvest crops (Waldhoff et al., 2012), etc.
9. Fully automatized and data-intensive closed production systems (i.e. greenhouses and other indoor led-illuminated aeroponics) (Love et al., 2014; Anon., 2016), would be on the rise within the framework of the circular economy (e.g. less use of pesticides, water and nutrient recycling, proximity to the consumer, etc.).
10. Precise genetic engineering, known as "genome editing", would make it possible to change a crop or animal's genome down to the level of a single genetic "letter" (Hartung and Schiemann, 2014). As (González-Recio et al., 2015) comment, this could be more acceptable to consumers, because it simply imitates the process of mutation on which crop breeding has always depended, and it does not imply the generation of transgenic plants or animals. This technology would supplement existing research in epigenetics (McQueen et al., 1995; Tesfaye et al., 2016).

The majority of the aforementioned potential applications would produce large amounts of (big) data, which could be used by future policy-makers to balance offer and demand (applications #1 and #2), reduce the negative impact of agriculture on the environment (applications #3, #5 and #9), raise food safety (applications #2, #4 and #6) and security (applications #2, #7, #8 and #10), increase productivity (applications #8, #9 and #10). The potential open access of this data to the public could create tremendous opportunities for research and development towards smarter and more sustainable farming.

5. Conclusion

This paper performed a review of big data analysis in agriculture, mostly from a technical perspective. Thirty-four research papers were identified and analyzed, examining the problem they addressed, the solution proposed, tools/techniques employed as well as data used. Based on these projects, the reader can be informed about which types of agricultural applications currently use big data analysis, which characteristics of big data are being used in these different scenarios, as well as which are the common sources of big data and the general methods and techniques being employed for big data analysis. Open problems have been identified, together with barriers for wider adoption of the big data practice. Various approaches for addressing these problems and mitigating barriers have been discussed.

As we saw in this survey, the availability and openness of hardware and software, techniques, tools and methods for big data analysis, as well as the increasing availability of big data sources and datasets, shall encourage more initiatives, projects and start-ups in the agricultural sector, either addressing some of the problems which are now being addressed by the 34 examples we presented, or focusing on some of the emerging future application areas we have identified, or even creating radical-new services and products applied in new agricultural areas.

This increasing availability of big data and big data analysis techniques, well described through common semantics and ontologies, together with adoption of open standards, have the potential to boost even more research and development towards smarter farming, addressing the big challenge of producing higher-quality food in a larger scale and in a more sustainable way, protecting the physical ecosystems and preserving the natural resources.

Acknowledgments

We would like to thank the reviewers for their valuable feedback, which helped us to reorganize the structure of the survey more appropriately, as well as to improve its overall quality. This research has been supported by the P-SPHERE project, which has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No 665919.

[2] Some example websites providing for free large datasets related to agriculture are listed in the supplementary material of this publication.

Appendix A

List of projects and studies employing big data analysis techniques in agriculture.

No.	Agri area	Problem description	Solution/impact	Application, tools, systems, algorithms used	Sources of data	Big data V dimensions			Ref.
						V1	V2	V3	
1.	Weather and climate change	How to forecast weather changes?	A technique based on machine learning showing that forecasting is superior to conventional downscaling.	Machine learning (scalable vector machines)	Weather stations	H	H	L	Tripathi et al. (2006)
2.		How to protect small farmers against droughts?	Development of a rainfall-based index insurance. The area covered by the program grew from around 100,000 hectares in 2003 to 12 million hectares in 2013 with positive and significant effect of 6% on maize yields.	Statistical analysis, modeling	Surveys, static historical information, weather stations	M	M	H	Fuchs and Wolff (2011)
3.		How to address the big data challenges of climate science, especially storing, analyzing and visualizing data and information?	MERRA Analytic Services (MERRA/AS) is a Climate Analytics-as-a-Service (CAaaS) platform. It offers high performance, data proximal analytics, management, software appliance virtualization, adaptive analytics and a domain-harmonized API.	MERRA/AS cloud-based platform, modeling, MapReduce-based analytics	Supports any type of weather and climate data, including imaging from satellites and other earth observational data	H	M	H	Schnase et al. (2014)
4.		How to consider the opportunities of using drought-tolerant maize varieties in Southern Africa?	Development of a geospatial crop modeling approach for assessing the impact of different varieties.Results show that new DT varieties could give a yield advantage of 5–40% across drought environments	(Crop) modeling and simulation, geospatial analysis	UEA climate database, CIMMYT database, static historical information, weather stations	L	L	H	Tesfaye et al. (2016)
5.	Animals research	How to perform dairy herd culling more productively?	95% correct culling decisions for cows based on the overall productivity and the potential of the animals	Machine learning (decision trees)	Static historical information about animals from database systems, animals' current physiological characteristics	H	M	L	McQueen et al. 1995)
6.		How much feed does a cow consume in a certain time period at a specific parcel and how does this relate to the milk production in that period?	Development of a learning model to predict the roughage intake of cows, with a precision of approximately 92.4%	Machine learning (artificial neural networks)	Sensory measurements of grazing activity, feed intake, weight, heat and milk production of individual cows, static historical information about cows' registration data	M	M	M	Kempenaar et al. (2016)

7.		How to recognize animal diseases?	Detection of pigs' coughing with more than 90% accuracy.	Machine learning (neural networks)	Sensory measurements of sound	L	H	L	Chedad et al. (2001)
8.		How to assure the safety and the quality of feedstuffs for animals?	Discrimination between vegetable and meat/bone meal with 78% accuracy.	Machine learning (scalable vector machines)	Multi-spectral camera (spectroscopic images)	M	H	L	Pierna et al. (2004)
9.	Soil	How to characterize soil and plants effectively?	Classification with varying accuracy depending on plant/soil type (10–95%)	Machine learning (K-means algorithm)	Optical camera (color images)	M	L	L	Meyer et al. (2004)
10.		How to classify and profile soil?	Classification of soils with varying accuracy (20–50%)	Machine learning (Farthest First clustering algorithm)	AGRIC soils historical database	M	L	L	Armstrong et al. (2007)
11.	Crops	How to effectively monitor the wine fermentation process?	Detection of over 70% of the problematic fermentations within 72 h	Machine learning (K-means algorithm)	Sensory measurements of metabolites	M	H	L	Urtubia et al. (2007)
12.		How to properly identify crops to study crop rotation?	Crop classification with accuracies of 89–92%	Machine learning (scalable vector machines)	Remote sensing (satellite images), land use historical datasets	H	M	M	Waldhoff et al. (2012)
13.		How to determine phenological stages of paddy rice?	Creation of map of rice phenology. Results showed less than 12 days root mean square errors of the estimated phenological dates (planting, heading, harvesting, growing) against the statistical data	Wavelet based filter for determining crop phenology (WFCP), fourier transform, vegetation indices (NDVI)	Remote sensing (MODIS satellite), static historical information (national land information), statistical data (MAFF Japan)	M	L	L	Sakamoto et al. (2005)
14.	Land	How to create accurate inventories of grasslands using remote sensing?	Inventories of grasslands with accuracy of 89–98%	Machine learning (scalable vector machines, random forests, extremely randomized trees), vegetation indices (NDVI)	Remote sensing (synthetic aperture radar images), GIS geospatial data, land characterization and crop phenology datasets	M	L	M	Barrett et al. (2014))
15.		How to identify management zones for cotton production?	Identification of management zones for cotton with a near optimal number of zones.	Machine learning (K-means algorithm)	Remote sensing (multi-spectral imaging), static historical information	M	L	H	Schuster et al. (2011)
16.		How to detect expansion and intensification of row-crop agriculture?	Determination of characteristic phenology of single and double crops. Estimated that over 3200 sq. km were converted from native vegetation and pasture to row-crop agriculture from year 2000–2005 in a study area encompassing 40,000 sq. km total	90% power wavelet transform	Remote sensing (MODIS satellite), Static historical information, GPS data	M	M	L	Galford et al. (2008)
17.		How to classify land use and land cover changes, as well as map crops?	MODIS time-series at 250 m ground resolution had sufficient temporal and radiometric resolution to discriminate major crop types and crop-related land use practices. Most crop	Image processing, distance-based classification, statistical analysis, vegetation indices (NDVI)	Remote sensing (MODIS satellite, Landsat ETM + imagery), FSA static database of aerial photos of crop types and practices	H	L	M	Wardlow et al. (2007)

			classes were separable during the growing season based on their phenology-driven spectral-temporal differences						
18.		How to determine land-use/land-cover (LULC) and irrigated areas through remote sensing?	A study conducted in the Krishna river basin (India) managed to correctly identify and label areas based on qualitative spectral matching techniques. The total irrigated area during years 1982–1985 was calculated as 2,975,800 hectares. Production of a global irrigated area map (GIAM).	Spectral Matching Techniques, vegetation indices (NDVI), reflectance and surface temperature calculations	Remote sensing (satellite – AVHRR continuous time series, NASA GSFC reflectance data, JERS-1 SAR data), monthly rainfall and temperature, elevation, global tree cover map	H	M	M	Thenkabail et al. (2007), and Thenkabail et al. (2009)
19.	Weeds	How to identify weeds in sunflower crops to minimize the impact of herbicide?	Weed discrimination with accuracies between 99.2% and 98.7%.	Machine learning (artificial neural networks), logistic regression, image processing, vegetation indices (NDVI)	Remote sensing (airplane), static historical information	L	H	L	Gutiérrez et al. (2008)
20.	Food availability and security	How to estimate food availability in sub-Saharan Africa?	An indicator of food availability which could be used to fight poverty. Results indicated that bridging yield gaps is important, but improving market access is essential	Machine learning (artificial neural networks), statistical analysis	Surveys, various static databases, CIALCA dataset	M	L	H	Frelat et al. (2016)
21.		How to support food chain safety measures for cattle holdings?	Development of a network-based assessment methodology suitable for risk-based planning and of simulation of epidemiological situations. This work helped to determine the most vulnerable parts of a cattle holding network	Modeling and simulation, network-based analysis	ENAR database for cattle	L	L	L	Jóźwiaka et al. (2016)
22.		How to find spatial equilibrium and optimal locations of agricultural facilities?	Use of planar, discrete, spatial equilibrium and network flow modeling to simulate the system and solve the relevant optimization problems (sum of time/distance traveled, cost of building/operating new facilities, number of new facilities needed, etc.)	Modeling and simulation, network analysis (Benders' decomposition, mixed-integer programming, branch and bound, enumeration, heuristics), geospatial analysis	Static historical information, various static databases and datasets, GIS geospatial data, statistical data	M	L	M	Lucas and Chhajed (2004)
23.		How to better target food security programs in areas that are most likely to be affected by damaged crops?	Development of an accurate yield forecasting model based on remote sensing. Assistance to stakeholders involved	Geographical information systems, (rice) crop growth simulations and modeling, image processing	Remote sensing (synthetic aperture radar), GIS geospatial data, rice crop growth historical datasets	H	M	M	RIICE Partnership (2014)

			in rice production to better manage the risks involved						
24.	Biodiversity	How to evaluate fish and wildlife population viability under land management alternatives?	Modeling of species' influences under different land variations	Statistics (Bayesian belief networks)	SER database of wildlife species, GIS geospatial data	M	L	H	Marcot et al. (2001)
25.	Farmers' decision making	How to provide farmers in India with access to agricultural inputs, scientific practices and market intelligence?	Development of a farm management platform that provides personalized agricultural advice on how to optimize costs, increase productivity and access markets. Results showed 64% increase in productivity in first year, 112% in the second year	PRIDE cloud-based platform and mobile application	Static historical information, humans as sensors, web-based data	M	M	M	Sawant et al. (2016)
26.		How to better understand how management choices affect sustainability performance and operational efficiency?	Development of an online calculator that estimates field-level performance based on various sustainability indicators.	Statistical analysis, modeling and simulation, benchmarking	Various static databases and datasets, US government survey data	H	L	H	Field to Market (2015)
27.	Farmers' insurance and finance	How to improve the productivity and income of smallholder farmers?	Mobile services with information on financial services, supply chain solutions, technical assistance and best practices.	mAgri cloud-based platform and mobile application, web services	Web-based data	H	H	H	GSMA (2014)
28.		How to provide fair and enticing insurance and finance for farmers?	Development of index insurance for farmers to promote investment in quality seeds and fertilizers, and access to agricultural loans. Insured farmers invested 20% more in their farms and earned 16% more income than their uninsured neighbors	Cloud-based platform	Static historical information, web data, weather stations, humans as sensors	H	M	H	(Syngenta, 2010)
29.		How to facilitate farmers' financing and easier payments?	Development of a platform that supports data, payment and settlement mechanisms between agricultural financiers, service providers, markets and farmers	Agrilife cloud-based platform and mobile application, web services	Static historical information, humans as sensors	H	M	M	Global Envision (2006)
30.		How to connect smallholders (fruit and vegetable growers) with export markets?	Development of a platform that enhances farmers' ability to meet export market standards/ certifications while at the same time ensuring a more stable and predictable supply of good quality for exporters	FarmForce cloud-based platform and mobile application	Static historical information, humans as sensors, web data, weather stations	M	M	M	Syngenta Foundation for Sustainable Agriculture (2016)

31.		How can farmers living in remote areas of developing countries carry out financial transactions in an easy way?	E-wallets is a micro-payments ecosystem for smallholder farmers. The mobile wallet network extends to tens of thousands of villages and some eight million farmers	Cellulant cloud-based platform and mobile application	Humans as sensors (financial transactions data), web-based data	H	H	L	Akinboro (2016)
32.	Remote sensing	How to enhance agricultural monitoring and crop production estimations using satellite observations?	GLAM is a global agricultural monitoring system that provides timely, easily accessible, scientifically validated remotely sensed data and derived products as well as data analysis tools, for crop condition monitoring and production assessment. Particular applications include a global croplands map, a near real-time surface reflectance product, a enhanced vegetation index and global lake levels estimations	GLAM cloud-based platform, image processing, vegetation indices (NDVI), decision support system	Remote sensing (MODIS satellite), MODIS surface reflectance datasets, earth land surface historical dataset of images, WMO weather datasets, reservoir heights derived from radar altimetry	H	M	L	Becker-Reshef et al. (2010)
33.		How to achieve global and multidisciplinary data sharing in the earth observation realm?	GEOSS is a global and flexible network of content providers allowing decision makers to access an extraordinary range of data and information in a broker-based cloud platform. Provides access to more than 80 million single datasets characterized by a large heterogeneity.	GEOSS Cloud-based platform offering various services for platform and infrastructure as a service, big data storage, retrieval and analysis, web and community portals	Supports any type of earth observation-related data from any web-based virtual source	H	H	H	Nativi et al. (2015)
34.		How to exploit big data on earth observation (EO) systems and datasets?	RemoteAgri web GIS system has been used for various agricultural applications, such as crop monitoring (water stress), precision farming (canopy estimation), and creation of accurate agricultural maps (vegetation detection)	Image processing, statistical analysis	Remote sensing (Landsat 8 satellite multi-spectral, multi-temporal dataset)	H	L	L	Karmas et al. (2014)

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.compag.2017.09.037.

References

AgGateway, 2005. [Online] Available at: < http://www.aggateway.org/Home.aspx > (accessed 2017).

Akinboro, B., 2016. *Bringing Mobile Wallets to Nigerian Farmers.* [Online] Available at: < http://www.cgap.org/blog/bringing-mobile-wallets-nigerian-farmers > (accessed 2017).

Anon., 2016. *AeroFarms.* [Online] Available at: < http://aerofarms.com/ > (accessed 2017).

Aqeel ur, R., Abbasi, A., Islam, N., Shaikh, Z., 2014. A review of wireless sensors and networks' applications in agriculture. Comput. Stand. Interfaces 36 (2), 263–270.

Armbruster, W.J., MacDonell, M.M., 2014. Informatics to Support International Food Safety. s.l., s.n., pp. 127–134.

Armstrong, L., Diepeveen, D., Maddern, R., 2007. The application of data mining techniques to characterize agricultural soil profiles. s.l., s.n.

Atzberger, C., 2013. Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs. Remote Sens. 5 (2), 949–981.

aWhere Inc., 2015. [Online] Available at: < http://www.awhere.com/ > (accessed 2017).

Babinet, Gilles et al., 2015. The New World economy, s.l.: Report addressed to Ms Segolene Royal, Minister of Environment, Sustainable Development and Energy, working group led by Corinne Lepage.

Barrett, B., Nitze, I., Green, S., Cawkwell, F., 2014. Assessment of multi-temporal, multi-sensor radar and ancillary spatial data for grasslands monitoring in Ireland using machine learning approaches. Remote Sens. Environ. 152 (2), 109–124.

Basso, B., et al., 2001. Spatial validation of crop models for precision agriculture. Agric. Syst. 68 (2), 97–112.

Bastiaanssen, W., Molden, D., Makin, I., 2000. Remote sensing for irrigated agriculture: examples from research and possible applications. Agric. Water Manage. 46 (2), 137–155.

Becker-Reshef, I., et al., 2010. Monitoring global croplands with coarse resolution earth observations: the Global Agriculture Monitoring (GLAM) project. Remote Sens. 2 (6), 1589–1609.

Bell, J., Butler, C., Thompson, J., 1995. Soil-terrain modeling for site-specific agricultural management. Site-Specific Management for Agricultural Systems, American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, pp. 209–227.

Blue River Technology, 2011. [Online] Available at: < http://www.bluerivert.com/ > (accessed 2017).

Bongiovanni, R., Lowenberg-DeBoer, J., 2004. Precision agriculture and sustainability. Precision Agric. 5 (4), 359–387.

Bunge, J., 2014. Big data comes to the farm, sowing mistrust: seed makers barrel into technology business, s.l.: Wall Street Journal (Online).

Capalbo, S.M., Antle, J.M., Seavert, C., 2016. Next generation data systems and knowledge products to support agricultural producers and science-based policy decision making. Agric. Syst.

Carbonell, I., 2016. The ethics of big data in big agriculture. Internet Policy Rev. 5 (1), 1–13.

Carolan, M., 2016. Publicising food: big data, precision agriculture, and co-experimental techniques of addition. Soc. Ruralis.

Chedad, A., et al., 2001. AP – animal production technology: recognition system for pig cough based on probabilistic neural networks. J. Agric. Eng. Res. 79 (4), 449–457.

Chi, M., et al., 2016. Big data for remote sensing: challenges and opportunities. Proc. IEEE 104 (11), 2207–2219.

Cooper, J., et al., 2013. Big data in life cycle assessment. J. Ind. Ecol. 17 (6), 796–799.

Cropster, 2007. [Online] Available at: < https://www.cropster.com/ > (accessed 2017).

de Leeuw, J., et al., 2014. The potential and uptake of remote sensing in insurance: a review. Remote Sens. 6 (11), 10888–10912.

Erenturk, S., Erenturk, K., 2007. Comparison of genetic algorithm and neural network approaches for the drying process of carrot. J. Food Eng. 78 (3), 905–912.

Farm Hack, 2010. Farm Hack. [Online] Available at: < http://farmhack.org > (accessed January 2017).

Field to Market, 2015. Fieldprint Calculator. [Online] Available at: < https://www.fieldtomarket.org/fieldprint-calculator/ > (accessed 2017).

Food and Agriculture Organization of the United Nations, 2009. How to Feed the World in 2050., Rome: s.n.

Frelat, R., et al., 2016. Drivers of household food availability in sub-Saharan Africa based on big data from small farms. Proc. Natl. Acad. Sci. 113 (2), 458–463.

Fuchs, A., Wolff, H., 2011. Drought and retribution: evidence from a large scale rainfall index insurance in Mexico. s.l., s.n., pp. 13–14.

Furbank, R.T., Teste, M., 2011. Phenomics – technologies to relieve the phenotyping bottleneck. Trends Plant Sci. 16 (12), 635–644.

Galford, G.L., et al., 2008. Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. Remote Sens. Environ. 112 (2), 576–587.

Gebbers, R., Adamchuk, V., 2010. Precision agriculture and food security. Science 327 (5967), 828–831.

Giller, K., et al., 2011. Communicating complexity: integrated assessment of trade-offs concerning soil fertility management within African farming systems to support innovation and development. Agric. Syst. 104 (1), 191–203.

Global Envision, 2006. *Unleashing Ugandan farmers' potential through mobile phones.* [Online] Available at: < https://www.mercycorps.org/research-resources/unleashing-ugandan-farmers-potential-through-mobile-phones > (accessed 2017).

GODAN, 2015. Global Open Data for Agriculture and Nutrition (GODAN) initiative. [Online] Available at: < http://www.godan.info/ > (accessed 2017).

González-Recio, O., Toro, M., Bach, A., 2015. Past, present and future of epigenetics applied to livestock breeding. Front. Genet. 6, 305.

Grace, D., McDermot, J., 2015. Reducing and Managing Food Scares, Washington, DC: International Food Policy Research, 2014–2015 Global Food Policy Report.

GSMA, 2014. mAgri Programme. [Online] Available at: < http://www.gsma.com/mobilefordevelopment/programmes/magri/programme-overview > (accessed 2017).

Gutiérrez, P., et al., 2008. Logistic regression product-unit neural networks for mapping Ridolfia segetum infestations in sunflower crop using multitemporal remote sensed data. Comput. Electron. Agric. 64 (2), 293–306.

Hartung, F., Schiemann, J., 2014. Precise plant breeding using new genome editing techniques: opportunities, safety and regulation in the EU. Plant J. 78 (5), 742–752.

Hashem, I., et al., 2015. The rise of "big data" on cloud computing: review and open research issues. Inform. Syst. 47, 98–115.

Jóźwiaka, Á., Milkovics, M., Lakne, Z., 2016. A network-science support system for food chain safety: a case from hungarian cattle production. Int. Food Agribusiness Manage. Rev. Special Issue 19(A).

Kamilaris, A., Gao, F., Prenafeta-Boldú, F.X., Ali, M.I., 2016. Agri-IoT: A Semantic Framework for Internet of Things-Enabled Smart Farming Applications. Reston, VA, USA, s.n.

Karmas, A., Karantzalos, K., Athanasiou, S., 2014. Online analysis of remote sensing data for agricultural applications. s.l., OSGeo's European conference on free and open source software for geospatial.

Karmas, A., Tzotsos, A. & Karantzalos, K., 2016. Geospatial Big Data for Environmental and Agricultural Applications. In: s.l.: Springer International Publishing, pp. 353–390.

Kempenaar, C., et al., 2016. Big Data Analysis for Smart Farming, vol. 655 Wageningen University & Research, s.l.

Kim, G.-H., Trimi, S., Chung, J.-H., 2014. Big-data applications in the government sector. Commun. ACM 57 (3), 78–85.

Kitzes, J., et al., 2008. Shrink and share: humanity's present and future ecological footprint. Philos. Trans. Royal Soc. B: Biol. Sci. 363 (1491), 467–475.

Kshetri, N., 2014. The emerging role of Big Data in key development issues: opportunities, challenges, and concerns. Big Data Soc. 1 (2).

Liaghat, S., Balasundram, S.K., 2010. A review: the role of remote sensing in precision agriculture. Am. J. Agric. Biol. Sci. 5 (1), 50–55.

Lokers, R., et al., 2016. Analysis of Big Data technologies for use in agro-environmental science. Environ. Model. Software 84, 494–504.

Love, D.C., et al., 2014. An international survey of aquaponics practitioners. PLoS ONE 9 (7).

Lucas, M.T., Chhajed, D., 2004. Applications of location analysis in agriculture: a survey. J. Operational Res. Soc. 55 (6), 561–578.

Ma, C., Zhang, H.H., Wang, X., 2014. Machine learning for Big Data analytics in plants. Trends Plant Sci. 19 (12), 798–808.

Manickavasagan, A., Jayas, D.S., White, N., Paliwal, J., 2005. Applications of thermal imaging in agriculture–a review. Can. Soc. Eng. Agric. Food Biol. Syst. 5 (2).

Marcot, B.G., et al., 2001. Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. For. Ecol. Manage. 153 (1), 29–42.

McQueen, R., Garner, S., C.G., N.-M., Witten, I.H., 1995. Applying machine learning to agricultural data. Compt. Electron. Agric., 12(1).

Meyer, G., Neto, J.C., Jones, D.D., Hindman, T.W., 2004. Intensified fuzzy clusters for classifying plant, soil, and residue regions of interest from color images. Comput. Electron. Agric. 42 (3), 161–180.

Mucherino, A., Papajorgji, P., Pardalos, P., 2009. Data Mining in Agriculture, vol. 34. Springer Science & Business Media.

Nandyala, C.S., Kim, H.-K., 2016. Big and meta data management for U-agriculture mobile services. Int. J. Software Eng. Appl. (IJSEIA) 10 (1), 257–270.

Nativi, S., et al., 2015. Big data challenges in building the global earth observation system of systems. Environ. Model. Software 68 (1), 1–26.

OADA, 2014. Open Agriculture Data Alliance. [Online] Available at: < http://openag.io/ > (accessed 2017).

Oluoch-Kosura, W., 2010. Institutional innovations for small-holder farmers' competitiveness in Africa. African J. Agric. Resource Econ. 5 (1), 227–242.

Ozdogan, M., Yang, Y., Allez, G., Cervantes, C., 2010. Remote sensing of irrigated agriculture: opportunities and challenges. Remote Sens. 2 (9), 2274–2304.

PEAT UG, 2016. Plantix. [Online] Available at: < http://plantix.net/ > (accessed 2017).

Pierce, F.J., & N., P., 1999. Aspects of precision agriculture. Advances in agronomy, vol. 67, pp. 1–85.

Pierna, J.A., et al., 2004. Combination of support vector machines (SVM) and near-infrared (NIR) imaging spectroscopy for the detection of meat and bone meal (MBM) in compound feeds. J. Chemom. 18 (7–8), 341–349.

Pretty, J., 2008. Agricultural sustainability: concepts, principles and evidence. Philos. Trans. Royal Soc. London B: Biol. Sci. 363 (1491), 447–465.

Rahman, H., Sudheer Pamidimarri, D., Valarmathi, R., Raveendran, M., 2013. Omics: Applications in Biomedical, Agriculture and Environmental Sciences, s.l. CRC PressI Llc.

RIICE Partnership, 2014. Remote sensing-based Information and Insurance for Crops in Emerging economies. [Online] Available at: < http://www.riice.org/ > (accessed 2017).

Rodriguez, D., et al., 2017. To mulch or to munch? Big modelling of big data. Agric. Syst. 153, 32–42.

Sakamoto, T., et al., 2005. A crop phenology detection method using time-series MODIS data. Remote Sens. Environ. 96 (3), 366–374.

Sawant, M., Urkude, R., Jawale, S., 2016. Organized data and information for efficacious agriculture using PRIDE model. Int. Food Agribusiness Manage. Rev. 19(A).

Sayer, J., Cassman, K., 2013. Agricultural innovation to protect the environment. Proc. Natl. Acad. Sci. U.S.A. 110 (21), 8345–8348.

Schnase, J.L., et al., 2014. MERRA analytic services: meeting the big data challenges of climate science through cloud-enabled climate analytics-as-a-service. Comput. Environ. Urban Syst.

Schuster, E.W. et al., 2011. Infrastructure for data-driven agriculture: identifying management zones for cotton using statistical modeling and machine learning techniques. s.l., IEEE.

Schuster, J., 2017. Big data ethics and the digital age of agriculture. Am. Soc. Agric. Biol. Eng. 24 (1), 20–21.

Senanayake, R., 1991. Sustainable Agriculture: definitions and parameters for measurement. J. Sustain. Agric. 1 (4), 7–28.

SenseFly, 2012. Drone applications in agriculture. [Online] Available at: < https://www.sensefly.com/applications/agriculture.html > (accessed 2017).

Shin, D.-H., Choi, M.J., 2015. Ecological views of big data: perspectives and issues.

Telematics Inform. 32 (2), 311–320.
Slavin, P., 2016. Climate and famines: a historical reassessment. Wiley Interdiscipl. Rev.: Clim. Change 7 (3), 433–447.
Song, M.-L., Fisher, R., Wang, J.-L., Cui, L.-B., 2016. Environmental performance evaluation with big data: theories and methods. Ann. Oper. Res. 1–14.
Sonka, S., 2016. Big data: fueling the next evolution of agricultural innovation. J. Innovation Manage. 4 (1), 114–136.
Sykuta, M.E., 2016. Big data in agriculture: property rights, privacy and competition in Ag data services. Int. Food Agribusiness Manage. Rev. Special Issue 19(A).
Syngenta Foundation for Sustainable Agriculture, 2016. FarmForce. [Online] Available at: < http://www.farmforce.com/ > [accessed 2017].
Syngenta, 2010. Syngenta Foundation for Sustainable Agriculture: Kilimo Salama – An agricultural insurance initiative. [Online] Available at: < https://kilimosalama.wordpress.com/about/ > (accessed 2017).
Teke, M., et al., 2013. A Short Survey of Hyperspectral Remote Sensing Applications in Agriculture. IEEE, s.l, pp. 171–176.
Tesfaye, K., et al., 2016. Targeting drought-tolerant maize varieties in southern Africa: a geospatial crop modeling approach using big data. Int. Food Agribusiness Manage. Rev. 19(A), 1–18.
Thenkabail, P., et al., 2009. Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. Int. J. Remote Sens. 30 (14), 3679–3733.
Thenkabail, P., et al., 2007. Spectral matching techniques to determine historical land-use/land-cover (LULC) and irrigated areas using time-series 0.1-degree AVHRR Pathfinder datasets. Photogram. Eng. Remote Sens. 73 (10), 1029–1040.
Thomson Reuters, 2017. Web of Science. [Online] Available at: < http://www.webofknowledge.com/ > (accessed 2017).
Tripathi, S., Srinivas, V.V., Nanjundiah, R.S., 2006. Downscaling of precipitation for climate change scenarios: a support vector machine approach. J. Hydrol. 330 (3), 621–640.
Tyagi, A.C., 2016. Towards a second green revolution. Irrigation Drainage 65 (4), 388–389.
Urtubia, A., Perez-Correa, J., Soto, A., Pszczolkowski, P., 2007. Using data mining techniques to predict industrial wine problem fermentations. Food Control 18 (1), 1512–1517.
Vibhute, A., Bodhe, S.K., 2012. Applications of image processing in agriculture: a survey. Int. J. Comput. Appl. 52 (2).
Vitolo, C., et al., 2015. Web technologies for environmental Big Data. Environ. Model. Software 63 (1), 185–198.
Waga, D., Rabah, K., 2014. Environmental conditions' big data management and cloud computing analytics for sustainable agriculture. World J. Comput. Appl. Technol. 2 (3), 73–81.
Waldhoff, G., Curdt, C., Hoffmeister, D., Bareth, G., 2012. Analysis of multitemporal and multisensor remote sensing data for crop rotation mapping. Int. Arch. Photogram., Remote Sens. Spatial Inform. Sci. 25 (1), 177–182.
Wardlow, B.D., Egbert, S.L., Kastens, J.H., 2007. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains. Remote Sens. Environ. 108 (3), 290–310.
Wari, E., Zhu, W., 2016. A survey on metaheuristics for optimization in food manufacturing industry. Appl. Soft Comput. 46, 328–343.
Weber, R.H., Weber, R., 2010. Internet of Things. Springer, New York, NY.
Wolfert, S., Ge, L., Verdouw, C., Bogaardt, M., 2017. Big data in smart farming – a review. Agric. Syst. 153, 69–80.
World Economic Forum, 2012. Big Data, Big Impact: New Possibilities for International Development. World Economic Forum, Geneva, Switzerland.
Wu, J., Guo, S., Li, J., Zeng, D., 2016. Big Data Meet Green Challenges: Big Data Toward Green Applications, s.l.: s.n.
Wulder, M., et al., 2012. Opening the archive: how free data has enabled the science and monitoring promise of Landsat. Remote Sens. Environ. 122 (1), 2–10.
Zhang, C., Kovacs, J.M., 2012. The application of small unmanned aerial systems for precision agriculture: a review. Precision Agric. 13 (6), 693–712.