Geography and Sustainability 1 (2020) 25–32

Contents lists available at ScienceDirect

Geography and Sustainability

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

Article

Cloud services with big data provide a solution for monitoring and tracking sustainable development goals

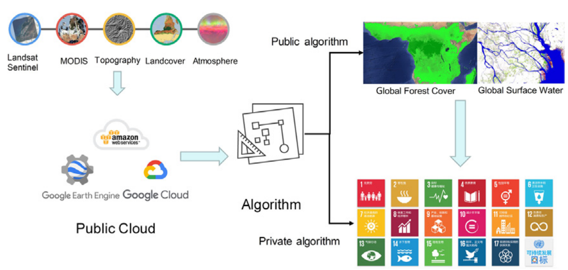
Bingfang Wu [a,b,*], Fuyou Tian [a,b], Miao Zhang [a], Hongwei Zeng [a,b], Yuan Zeng [a,b]

[a] *State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China*
[b] *University of Chinese Academy of Sciences, Beijing 10049, China*

HIGHLIGHTS

- Cloud services provide the cost-effective solutions for monitoring and tracking SDGs.
- Validation and quality control of public earth observation data is a key to SDGs.
- Crowdsourcing data is an alternative method of data collection on SDGs.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 8 February 2020
Received in revised form 12 March 2020
Accepted 13 March 2020
Available online 18 March 2020

Keywords:
Big data
Cloud services
SDGs
Monitoring

ABSTRACT

To achieve the Sustainable Development Goals (SDGs), high-quality data are needed to inform the formulation of policies and investment decisions, to monitor progress towards the SDGs and to evaluate the impacts of policies. However, the data landscape is changing. With emerging big data and cloud-based services, there are new opportunities for data collection, influencing both official data collection processes and the operation of the programmes they monitor. This paper uses cases and examples to explore the potential of crowdsourcing and public earth observation (EO) data products for monitoring and tracking the SDGs. This paper suggests that cloud-based services that integrate crowdsourcing and public EO data products provide cost-effective solutions for monitoring and tracking the SDGs, particularly for low-income countries. The paper also discusses the challenges of using cloud services and big data for SDG monitoring. Validation and quality control of public EO data is very important; otherwise, the user will be unable to assess the quality of the data or use it with confidence.

1. Introduction

The United Nations (UN) released “Transforming our World: The 2030 agenda for Sustainable Development”, which included 17 Sustainable Development Goals (SDGs) and 169 specific targets (United Nations, 2015, 2019), presenting an opportunity to transform lives for the better but a substantial challenge for the world as a whole. We are now in the process of implementing the SDGs; most of the instruments are in place to undertake policies aimed at accelerating progress and monitoring their results. To achieve the SDGs, high-quality data (relevant, timely, reliable, comparable) are needed to determine the breadth and depth of the problems and identify the affected populations to inform the formulation of policies and investment decisions, monitor progress towards the SDGs and evaluate the impacts of policies (Guo, 2017; Xu et al., 2020); this requirement has resulted in new challenges and opportunities for broad data collection efforts. SDG implementation is unlikely to be effective and the desired targets will not be achieved if there are no good data. Data are the “eyes and ears” of decision-makers and can maximize the efficiency and efficacy of intervention and bring us closer to achieving the SDGs.

Closing data gaps is essential for achieving sustainable development goals. SDGs data gaps are pervasive due to non-alignment of national/regional indicators with the global indicator framework (GIF), which comprises 232 unique indicators and low investments in SDGs

* Correspondence author.
E-mail address: wubf@radi.ac.cn (B. Wu).

https://doi.org/10.1016/j.geosus.2020.03.006

data production and use (Schmidt-Traub et al., 2017). Initially, it was expected that the statistical systems could measure and incentivize progress across the goals. There is an enormous need to develop capacity at the national level to ensure that countries can produce and report on most SDGs indicators. Currently, few countries use SDGs indicators systematically in their statistics. Only 22% of the environment-related SDGs are supported with available data, and 68% of the environment-related SDGs do not have sufficient data at the global level to assess progress (Campbell, 2019). In 12% of cases, the data for the 21 SDGs indicators under FAO custodianship are consistent with the global SDGs database. Seventy-two percent of countries require external assistance to produce one or more of the 21 SDGs indicators under FAO custodianship, and 62.4% required assistance in the analysis/interpretation of the SDGs indicators (FAO, 2019).

The cost of collecting environmental data is astonishing high (Castell et al., 2017). Environmental data collection requires a total of US$1 billion per annum to enable 77 of the world's low-income countries to catch up and implement statistical systems capable of supporting and measuring the SDGs (Schmidt-Traub et al., 2015). The total cost of the data needed to monitor the Tier I (satisfy all criteria) and Tier II (satisfy most criteria but data coverage is insufficient) indicators in all low- and middle-income countries (International Development Association eligible (IDA-eligible)) is likely to be on the order of $44 to $45 billion over the SDGs period (Schmidt-Traub et al., 2015). Compared to the estimates of implementing the full 2030 development agenda, which range from $700 billion to over $3 trillion, these costs are modest (Sustainable Development Solutions Network, 2015). However, significant additional costs will be incurred to implement data collection programmes for the Tier III (methodology still being developed) indicators, and further investments will be required in administrative systems that have not been included here. Other studies estimate that additional annual investment in statistics of USD 100-200 million is needed to monitor the SDGs (Sustainable Development Solutions Network, 2015). Nevertheless, these estimates will help to define the likely magnitude of the expenditures involved, although there are considerable differences. Few countries can afford the high cost. It is unlikely that a sudden increase in investment in statistics will be expected, which will encourage the exploration of alternative solutions to ensure that high-quality and disaggregated data for the SDGs are produced and used. A new cost-effective paradigm for data collection is critical to bridge data gaps without overburdening countries.

Recently, the data landscape has been changing rapidly, challenging the global community to find ways to utilize new technologies and to forge new partnerships. The unprecedented rate of innovation in data collection techniques and technologies and the capacity to distribute data widely and freely has expanded the horizon of possibility (Sustainable Development Solutions Network, 2015). These innovations will dramatically change the way we collect data and advance our ability to monitor the impact of development programmes. With the emergence of big data, artificial intelligence, and cloud-based calculations, reforms in environmental and resource monitoring are forthcoming in the domain (Guo et al., 2016), while will influence both official data collection processes and the operation of programmes they monitor. Big data, such as data from numerous individuals, is a new opportunity for data collection. All mobile devices collect large amounts of spatiotemporal information, including geographical locations, moving speeds, moving paths, migration and photos with geo-tags, forming a new situation in which everyone collects geospatial information. This provides a new solution for collecting data for the SDGs (Vinuesa et al., 2020). Recently, Google, Amazon and Alibaba introduced their cloud computing products, which have archived a large catalogue of relevant environmental and resource data, such as gauge station data and satellite imagery, in Google Earth Engine (GEE) and the Amazon Web Service (AWS) (Dong et al., 2016; Gorelick et al., 2017). The Alibaba Cloud has become the third most commonly used product globally and number 2 in Asia (Shah, 2019) and is actively used for environmental data. The report of *Big Earth data in support of the Sustainable Development Goals* (Chinese Academy of Science, 2019) compiles China's measures, progress, challenges and plans about its implementation of the 2030 Agenda. The report includes 27 case studies on targeted poverty alleviation, innovation-led development, and jointly building the Belt and Road initiative (Chinese Academy of Science, 2019). Furthermore, new methods allow data to be collected through mobile devices and GPS systems with location-based data for improved efficiency in data reporting, e.g., direct reporting without going through government layers, as well as very high potential for lowering the cost of SDGs monitoring over time.

This paper intends to explore the potential of big data, including crowdsourcing, public cloud data and services, to provide cost-effective, easy-sharing and timely data for monitoring and tracking the SDGs in general national environment management in particular.

2. Crowdsourcing Data and Services for SDG monitoring

Crowdsourcing geographic information is an effective way to acquire geographically located data that is mainly collected by non-professional users and submitted to servers, distributed databases or cloud platforms according to unified rules (Tulloch, 2014) and follows a standardized processing protocol. The data collected through this approach are called volunteered geographic information (VGI).

VGI has become a widespread data acquisition method in environmental and resource monitoring (Rollason et al., 2018; Stehman et al., 2018). The VGI is mainly collected automatically by people who do not possess special knowledge, such as that required for traditional data collection, and do not have a specific purpose. In 1890, the US National Weather Service established the "Cooperative Observer Program" before the invention of modern communication technology (Leeper et al., 2015). Many datasets obtained from this project have been widely used in scientific research, such as weather monitoring, extreme weather warning and climate change (Leeper et al., 2015). After that, the "North American Breeding Bird survey" project employed a VGI approach to monitor bird population activity over a long period of time and recorded the distributions of and quantitative changes in the populations of more than 400 species of birds (Sauer et al., 2013). With the support of VGI, remote sensing-based landcover monitoring has been assessed for accuracy (Stehman et al., 2018).

With the wide use of mobile phones, sensors in smartphones, such as GPS, accelerometers, gyroscopes, and photoreceptors, have become major platforms for VGI collection. For example, Fritz et al. created the Geo-Wiki website and released the "GEO wiki pictures" mobile app (http://www.geo-wiki.org/), which allows volunteers around the world to upload geo-tagged photographs as VGI for forest, grassland, farmland, and water body samples (Fritz et al., 2009). Using these landcover-related sample data, a global cropland map was validated, and its accuracy was improved (Fritz et al., 2015). In addition, "CrowdWater" (https://www.crowdwater.ch) is also a VGI platform that collects hydrological data and depends on the participation of the public and anyone who is interested in water. Photos of the water ruler can be used to retrieve the water level automatically, replacing the manual records currently used by hydrological departments worldwide. In contrast to most stream monitoring that focuses on larger streams, another VGI platform called "Stream Tracker" (https://www.streamtracker.org/) aims to focus on small streams by combining a network of citizen scientists, sensors, and satellite imagery to track when and where streams flow. "GIS cloud" (http://www.giscloud.com/), "Poimapper" (http://ww.poimapper.com/), "GeoODK collect" (http://geoodk.com/), "FieldMap" (http://maptext.com/) and many other mobile applications are also widely used in VGI collecting. Traditional geographic information system (GIS) facilitators such as ArcGIS have also developed applications for mobile terminals, including "Collector for ArcGIS" (http://www.esri.com/software/arcgis/smartphones/collectorapp), in

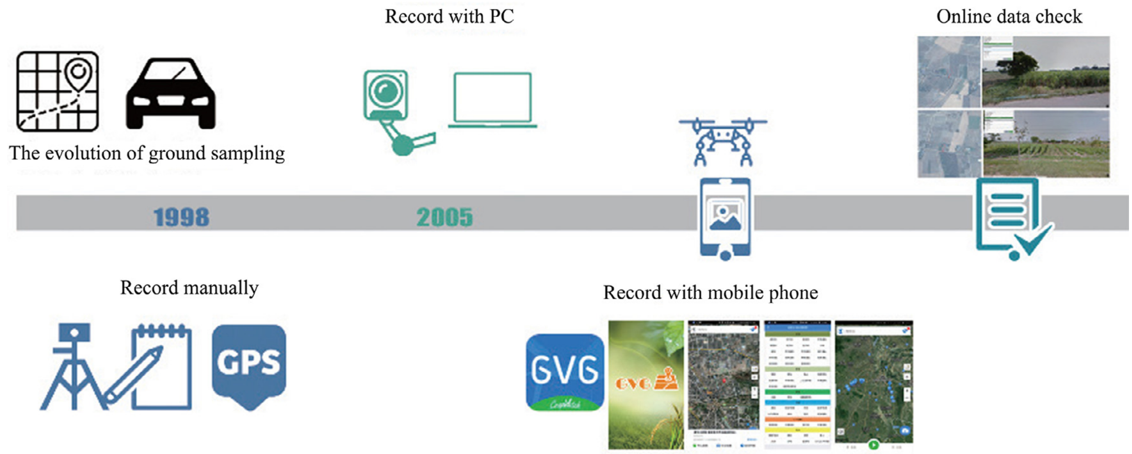


Fig. 1. Evolution of the GVG crowdsources geographic data collection app. In 1998, sample points were recorded manually. Sample points could be collected with PCs and GPS and video cameras starting in 2000 and on smartphones beginning in 2015.

which users can use a variety of functions such as "custom form", "precise positioning", "offline map", "mark plot" and "crop type".

VGI also provides many daily public services. The most representative example is a map service facilitator using the location information obtained from mobile phones to determine road congestion status. It can also provide information services, such as road condition judgement and optimal navigation paths that update in real time, providing convenient services for public travel. "OpenStreetMap" (OSM) from the UK is also based on the concept of crowdsourced data collection and has created a global map with free content that can be edited by anyone. Registered users can upload the GPS path that users found based on handheld GPS devices, aerial photos, satellite imagery or other methods and edit the vector map using the OSM editor or other software to collaboratively maintain an online map (Ramm et al., 2010). Recently, *Tianjin City, China* released a public reporting app for illegal parking, which was named the "Hexi branch of *Tianjin City, China* Public Security Bureau" app (https://apps.apple.com/cn/app/id1227573301). Using this app, users will obtain bonuses of 100 RMB for each report of illegal parking and a maximum of 1000 RMB per month. More than 17000 people registered as participants and reported 1775 reports of illegal parking on the first day. Additionally, some apps were released for pollution reporting, such as "China environment news". Dwellers volunteered to report photographs of both air and river pollution using the app, which enables timely and low-cost environmental monitoring.

By collecting geo-tagged photos, a great deal of information can be collected, such as the shape and path of features. Many VGI collection apps for mobile phones have been introduced around the world, and many applications have been carried out based on VGI. For example, the VGI data sources that have been used to validate land cover can also be employed for crop condition monitoring and to provide early warning for extreme weather and climate change. Since 2015, CropWatch, a global crop monitoring system (Wu et al., 2015; Wu et al., 2014) that upgraded the desktop agricultural sampling system "GVG software (GPS, video and GIS)", for use with a smartphone (Zhang, 2017a; Zhang, 2017b). It is freely available to the public in the smartphone application market and enables the collection of crop planting status photos at any time and in any place. The rapid acquisition of farmland photos and crop types has been successfully achieved with this method. More than 100,000 crop planting structure survey data have been obtained by different users every year, which greatly reduces the time and cost for obtaining crop planting structure information worldwide. These VGI data of crop types provide ground-based field data for crop acreage estimation and prediction (Fig. 1). In 2017, 758,084 ground samples of crop types, 1,128,000 photographs and 2.57 TB data in 1381 counties of China were quickly and efficiently captured with this app in 70 days; these data have been used to support the 2017 paddy field/dry land identification and other land cover mapping (Tian et al., 2019; Zhang et al., 2018). The upgraded GVG application uses a user-friendly format to provide non-professionals with different solutions for land cover type identification, which reduces uncertainty in the VGI collection process. The agricultural VGI collection application GVG has changed the way that highly rely on ground-based observations or data sharing and reduces the amount of ground observation work and the required human resources and financial input, providing a cost-effective solution for landcover samples. The GVG app is now freely downloaded from Google, Apple, Huawei and other application platforms (Zhang, 2017a; Zhang, 2017b). The app communicates with the cloud server in real-time; users only need to verify their accounts and then can use the app to collect land cover (SDG 15, Life on Land), crop samples (SDG 2, Zero Hunger), etc.

Furthermore, in addition to using smartphone photos, the sensors embedded in the smartphone can support advanced app development by using the distance or height estimated with the gyroscope in the smartphone or the noise detected with the sound sensor. With the emergence of smartphones with dual cameras, distances could be measured directly in the future. These sensors could extend the content in VGI software and make VGI information more accurate, especially with direct measurements of distance. The novel sensors in smartphones can collect additional types of VGI data to serve the SDGs. For example, the sound sensor could make the collection of noise in VGI software possible, which could directly affect the sleep, mood and health of residents (good health and well-being, SDG 3).

3. Public Earth Observation Data Products for the SDGs

Earth observation (EO) provides large-scale, high-quality and unbiased data on the physical, chemical and biological systems of the Earth. EO data and information can measurably enhance the quality of interventions and financial investments in the context of the SDGs. Earth observation data are a critical source for a number of SDG indicators, such as forest cover, mountain green cover (SDG 6.6; SDG 15.1) (Mondal et al., 2020), crop area and production (SDG 2.4) (Whitcraft et al., 2019; Wu et al., 2015), the sub-indicators of land degradation (SDG 15.3) (Giuliani et al., 2020) and agricultural sustainability (SDG 2.4), and subnational data disaggregation. In 2016, the European Space Agency (2019) launched the Earth Observation for Sustainable Development (EO4SD) initiative to develop indicators to monitor the progress of agriculture and rural development, water resource management, urban development, climate resilience, disaster risk reduction, marine resources, etc. Big earth data in support of the Sustainable Development Goals (Chinese Academy of Science, 2019) are the latest attempts to use Earth Observation data to develop indicators for monitoring the progress of the 6 SDGs of zero hunger, water, climate change, urban development, disaster, and territorial ecosystem.

In 2011, Google released the "Google Earth Engine" geographic data cloud computing platform. In addition, the AWS has opened a total of 61 datasets, including the “NASA Earth Exchange”, “Landsat Series”, “Sentinel Series Satellite”, “Next Generation Weather Radar (NEXRAD)” and “National Agriculture Imagery Program” (NAIP) and "digital elevation model dataset (DEM) ". With the help of cloud-stored observation data, users can easily conduct global resource and environmental monitoring in the cloud platform (Shao et al., 2012). For example, Google has developed a GEE cloud platform with global-scale PB-scale data processing capabilities for Earth observation big data, which greatly enhances the processing and information mining capabilities of big data for Earth observation (Gorelick et al., 2017). GEE’s built-in preprocessed long-term sequence of Landsat, MODIS, Sentinel and other data enables rapid monitoring of long-term, wide-ranging ecological dynamics (Hansen et al., 2013).

Japan also joined the data sharing process by providing free and open access to wide-swathe observation data from L-band radar satellites, such as ALOS (ALOS/AVINIR-2, PALSAR) and ALOS-2 (ALOS-2/ScanSAR), which is essential, especially in the tropics, where cloud cover hinders optical sensor observations (Rosenqvist et al., 2014). China announced the 16-m resolution optical data by GF-1 and GF-6 satellites available to the public (http://www.cnsageo.com/), which allows users to be able to redistribute, reproduce, and modify the data. China’s data policy is a very large step and a significant contribution to global data sharing. It opens a new era for the integration and synergy of global satellite data.

However, EO data are only raw materials for producing data products (Wu et al., 2019). Cloud-based computing and machine learning could facilitate the transition from publicly accessible EO data to public data products for monitoring purposes. For example, Gong et al. used the “Google Cloud” to produce 30 m global land cover products (Gong et al., 2013). The EU Joint Research Center completed monitoring of 30 m resolution land surface waters from 1984 to 2015 (Pekel et al., 2016). Since land surface waters are the most intuitive reflection of regional water resources, long-term, comparable time-series datasets provide valuable information for diagnosing the degree of water stress and its changes in arid ecosystems. The Joint Research Centre of the European Union implemented global habitat monitoring in 1975, 1990, 2000, and 2014 (Corbane et al., 2019). Maryland University of the USA completed a global forest cover change dataset from 2000 to 2016 with a 30 m resolution (Hansen et al., 2013); the USGS announced the mapping of global cropland at 30 m resolution in 2015 using the GEE platform and high-performance computing (Xiong et al., 2017); The Institute of Remote Sensing and Digital Earth of the Chinese Academy of Sciences published a global burned area data with a 30 m resolution in 2015 (Long et al., 2019) using the GEE platform. ESA developed the “S2ToolBox” tool for the vegetation leaf area index (LAI), the fraction of absorbed photosynthetically active radiation (FAPAR), and the fraction of vegetation cover (FCOVER) with a spatial resolution of 20 m (Baret et al., 2007).

Cloud computing with the support of machine learning is also the major methodology used for ChinaCover (Wu, 2017; Zhang et al., 2014). By using crowdsourcing data to train the decision tree and validate the accuracy, classification and change detection were performed on GEE to generate the landcover change area, while data verification and validation were performed on the Alibaba Cloud. The land trend (Kennedy et al., 2010) and change vector analysis algorithm were employed to detect the landcover change area, while a random forest (Pal, 2005) algorithm was used to classify the landcover type. Multi-source and time series remote sensing data were used for classification, including Landsat 8 and Sentinel-2. A total of 118316 VGI points were used in the training decision tree method to classify land cover and validate the final result. Quality control was performed using an online verification system based on Alibaba Cloud. Based on VGI data and massive remote sensing data, including ZY-3, GF-2, and Sentinel-2 data, China’s cropland was identified at 10-m resolution on the cloud platform. In other words, there is a new pattern in the data life cycle: VGI data could be used as training and validation data. Increasingly high-resolution satellite data can be effortlessly shared on the cloud platform (Fig. 2). EO data can be used to monitor crop health and, in connection with machine learning and drones, can build models to help farmers maximize their yields while reducing their environmental impact (Independent Group of Scientists appointed by the Secretary-General, 2019) to address SDG 2.

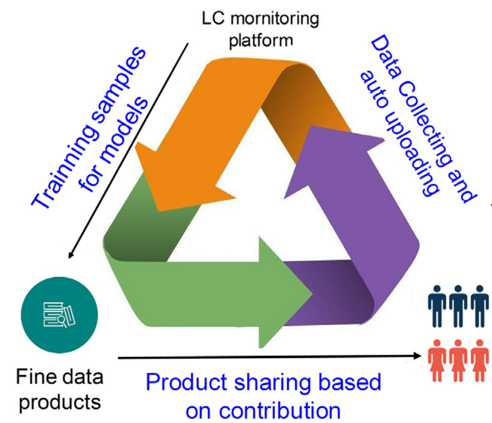


Fig. 2. Paradigm of the new pattern of data sharing. Each user of a VGI app, such as GVG or Geo-wiki, can upload the sample data. Based on adequate VGI training samples, fine-resolution production could be produced.

Compared with traditional personal computers and servers, the cloud platform has the characteristics of high computing efficiency, high performance, flexible expansion, large storage capacity, low price, and data security. It is suitable for processing and computing massive geographic data. With the popularity of geographic data cloud platforms, employing data through cloud platforms has become a dominant method for resource and environment monitoring (Shao et al., 2012; Wu et al., 2019). Using high-speed computing devices on the cloud platform, data can be processed efficiently in the cloud without downloading large amounts of data to local computers, and the final result of the analysis and processing can be extracted or downloaded, thereby greatly enhancing the efficiency of processing resource and environment data. It is also possible to conduct environmental resource monitoring and assessment for longer periods and at higher spatial scales, eliminating the limitations of computing and storage capabilities.

In the past, low spatial resolution data products were dominant in the public domain, with free access for the scientific community, such as Modis products (Savtchenko et al., 2004), GLASS products (Zhao et al., 2013), VGT/Prob-V products (Van Achteren et al., 2012), and global precipitation and soil moisture products (Liu et al., 2018). Recently, the resolution of the data products has increased from several kilometres to 100 metres to ten metres as global 10-metre landcover products are becoming available. With increased reliance on the high-performance cloud-based computing and open data resources, high-resolution data products will gradually become the main data sources. The high-resolution data products generated by the scientific community, which mainly generates coarse data, can be used or supplemented by businesses and government bodies. The boundary between scientific research and business will be greatly diluted by using EO data products. Earth observation data will be a significant step forward in making useful information available for water, food and land supply. However, statistics systems must use these publicly available data to provides support for monitoring and tracking the SDGs (Mondal et al., 2020).

4. Cloud Services for the SDGs

Data are needed to explore the interlinkages across thematic areas together with social and economic information to produce insights. There is a need not only for national-level statistics but also for geospatial data that can be disaggregated for vulnerable populations. To achieve this, new and existing crowdsourcing and public EO data must be integrated with in situ, survey, statistical, transactional and other forms of data. However, downloading these products to the user's own machine and processing them locally is tedious and time-consuming (He et al., 2013). However, it is not just downloading that is time-consuming; data needs to be disaggregated by income, gender, age, race, ethnicity, immigration status, disability, geographic location and other characteristics relevant in the national contexts. This requires the development and maintenance of an analysis system, which is hardly affording for low-income countries, both technically and financially (Sustainable Development Solutions Network, 2015; Schmidt-Traub et al., 2015).

How can public EO data be used, especially for monitoring and tracking the SDGs? Progress towards the SDGs is monitored at the local, national, regional, and global levels. The focus of SDG monitoring will be at the national level. Complementary monitoring will occur at regional and global levels. Moreover, each major thematic community, such as the agricultural community, will mobilize, analyse, and communicate data on progress towards achieving its objectives; this will require a wide range of services, such as those services that can be provided by EO data products, including food security risk management through agricultural production (SDG 2), irrigation management (SDG 6), rural infrastructure, agricultural ecosystem services, land degradation, the impact of commodities on deforestation, and environmental and social safeguards (SDG 15) (Mondal et al., 2020) (Fig. 3). Such thematic monitoring and reviewing will be an important complement to official monitoring and reviewing at national, regional, and global levels (Schmidt-Traub et al., 2015).

Data from crowdsourced and public EO data are not explicitly designed for SDGs monitoring purposes. It is essential to tailor these products to the needs of the users to facilitate decision-making and downstream applications (Buontempo et al., 2019). Cloud services allow anyone to view, process, and analyse crowdsourcing and public EO data with a few simple lines of command to implement regional, national, intercontinental, and even global-scale analysis (Wu et al., 2019). Cloud services for thematic areas can be developed to monitor, trace and assess SDG progress, particularly for rapid assessments (http://unstats.un.org/sdgs/) and to provide real-time decision-making support through natural resource liability tables and natural resource audits.

Based on a large available dataset, Copernicus services will provide essential information for six main domains: ocean, land and atmosphere monitoring, emergency response, security and climate change. In each thematic area, the service provides information including global, European and regional indicators, essential climate variables (ECVs) and tools/scripts (workflows) that run on the toolbox to develop user-driven, sector-specific applications through the datasets and tools in the Copernicus infrastructure. The services provides a working example of how the data and the tools available on the Copernicus infrastructure could be used in user-specific contexts to engage with the users to determine and document what they need and to provide examples of good practices in the development of climate services (https://www.copernicus.eu/en/services) (Buontempo et al., 2019).

Another example of a cloud-based service is the CropWatch Cloud, which was designed for global crop condition monitoring (SDG 2). The CropWatch Cloud (http://cloud.cropwatch.com.cn), which is an Alibaba Cloud-based crop monitoring platform (Wu et al., 2015; Wu et al., 2014), provides an agro-climate, agronomic information service paradigm and a unique solution for developing countries aspiring to conduct their own crop monitoring to promote leapfrog development to achieve the UN SDGs of eliminating poverty (SDG 1) and zero hunger (SDG 2) using publicly available data from around the world. The fundamental principle of the CropWatch Cloud is to move from the current paradigm of providing users with data and tools to the paradigm of providing information services to end-users anytime and anywhere in the global food security community.

In addition to data integration and continuous expansion, the tools or functionalities to retrieve, process, analyse, query and visualize data are important components in allowing a vast community of users to manage data for their own requirements with a fully established workflow (Buontempo et al., 2019). The CropWatch Cloud collects and converges crowdsourcing and public EO data with sophisticated analytics and modelling tools, which allows decision-makers and policy-makers to make more well-informed decisions based on access to unbiased quantified information at previously unavailable scales, resolution and frequencies (Wu et al., 2015). The Earth Observation for Sustainable Development (EO4SD) Agriculture and Rural Development Cluster project demonstrated how EO-based information and services could support agricultural monitoring and management tasks for SDG 2 (European Space Agency, 2019).

5. Opportunities and Challenges for the use of Cloud Services with Big Data for the SDGs

This paper explores the potential and possibility of integrating crowdsourcing and public EO data to monitor and track the SDGs. They need to be integrated with in situ, survey, statistics, transactional and other forms of data through cloud services with the support of analysis

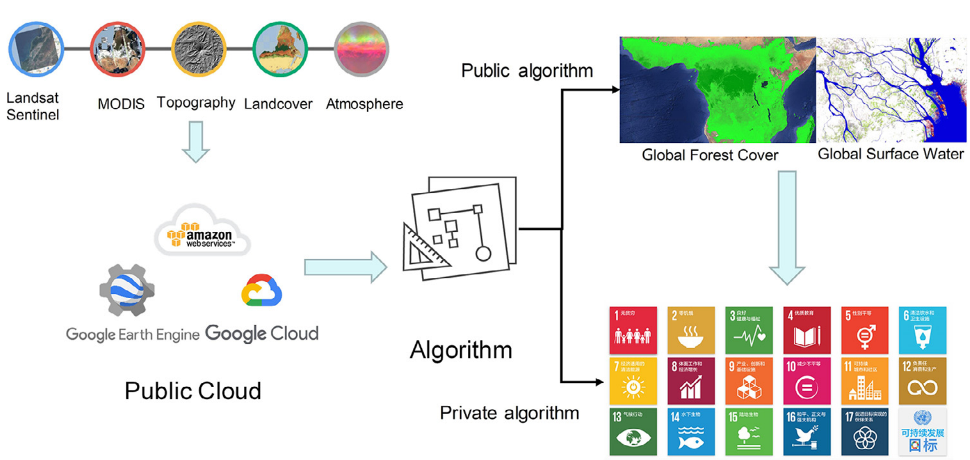


Fig. 3. Cloud-based services. Similar to GEE, many geographic data can be obtained from Google Cloud. A user can apply their own algorithm without downloading any data. This will be more convenient for SDG monitoring.

functions and models for the SDGs' thematic areas, such as agricultural (SDG 2), water (SDG 14), health (SDG 6), and poverty (SDG 1).

The changing data landscape is obvious everywhere; this paper only discusses the collection of smartphone-based data and publicly available EO data. Social media platforms also provide low-cost and widely available new data sources for SDGs monitoring, such as for determining real-time market prices for agricultural products (SDG 2) (Killian et al., 2019), (UNCTAD 2017) and assessing the non-native tree distribution in the city (Vaz et al., 2019), etc.

5.1. Data manipulation and quality control

There are considerable challenges and high costs involved in collecting high-quality data for monitoring or tracking the SDGs. For example, there are more than 20,000 hydrological gauge stations (China Meteorological Administration, 2019) and 400,000 ground sampling points for forest resource monitoring throughout China. Therefore, few countries around the world could afford this huge budget. In addition, during the collection process, several departments conduct the same work at the same time, which may lead to duplicated observations, data redundancy and inconsistencies. Additionally, the timeliness of data collection is important for decisionmaking but cannot be promised due to limited budgets and other factors.

New data collection and monitoring technologies are rapidly becoming available (IEAG, 2014). Everyone from all countries enable access to VGI and publicly available cloud data for free, which is considered an effective alternative method for monitoring the SDGs. Some of these innovations have considerable cost-saving potential(Castell et al., 2017). The popularity of smart phone-based data collection has the potential to reduce the time and cost of data collection after streamlining integration with other information sources and open up new possibilities to complement official statistics with information, enabling greater disaggregation of traditional statistics and improving timeliness. There are many sensors in smartphones that can support advanced app development for SDGs monitoring.

In contrast to the traditional methods for tracking SDGs progress, which are very expensive, the use of new technology for monitoring the SDGs with big data is more economical, effective, practical and even fair. Like the tracking of poverty, we need to conduct interviews and questionnaires in a conventional way, which will be time-consuming and labour-intensive. Now, we can assess the poverty rate with freely available earth observation data (Jean et al., 2016) and mobile phone data (Blumenstock et al., 2015; Steele et al., 2017) because the house roofs of poor people can be identified and the social network of the poor community can be recognized. This is one example of the strength of the new technology.

These innovations will dramatically advance our ability to monitor the environment and resources, to assess the impacts of government programmes and interventions, the wellbeing of people and to forecast future social, economic, and environmental changes (Fritz et al., 2009). However, streamlining is essential to ensure that smart phone-based data collection provides consistent data coverage in spatial and temporal dimensions.

On the other hand, high-resolution satellite data products will influence the way we generate data and the way it is used to help deliver sustainable development (Gong et al., 2013; Hansen et al., 2013; Long et al., 2019; Pekel et al., 2016; Wu, 2017; Xiong et al., 2017). For example, the cost of high-resolution satellite data products is falling with the increased availability of cloud computing and automated processing. In addition, these data can be disaggregated for vulnerable populations, which is essential for upholding the Agenda 2030's commitment to inclusivity and to ensuring that the most vulnerable people are reached. In this regard, using geospatial data is essential for ensuring that no one is left behind (IEAG, 2014), which is one of three signature elements of the SDGs.

However, while cloud computing and machine learning make data product generation from Earth observation data easier than before, there is another even more important issue that deservces our attention. Who controls public data cloud quality? If there is no quality control, how can the users choose the correct dataset? For example, there are 4 recently published landcover/cropland datasets of the African continents available publicly, and a recent study showed that they were only 63.8% consistent (Nabil et al., 2020). Even those pixels with the same class had only 70% accuracy as validated by using in situ data. Then, which dataset should users choose? Can findings from those datasets be trusted? If no one uses these datasets, then why develop them? In this regard, quality control and global validation are very important for public EO data products, which are large challenges for the global scientific community; otherwise, cloud data are simply an enjoyable diversion for scientists, but business users and government bodies will not have confidence to use them. High-resolution data products will then be useless for SDGs monitoring. Regardless, it is national agencies' responsibility to generate and report national data.

5.2. Public participation and customization

An exponentially growing number of data products stored in the cloud are available for public use, which is contrary to the current popular practices, in which most data products are stored in various laboratories and buildings. The data will no longer be hidden in the producer's hard disks or department archives. Cloud platforms make resource environment data accessible (Wu et al., 2019). The easy-to-contribute, easy-to-obtain and easy-to-use features of cloud data have greatly improved the transparency and confidence of data, which is essential for public participation in achieving the SDGs.

Cloud services have increasingly changed the strategy and philosophy of current data storage and processing and analysis (Buontempo et al., 2019; Kalia et al., 2017). Compared with the traditional information system, the cost of establishing, customizing, updating and maintaining the cloud professional service system is significantly lower (IEAG, 2014). At the same time, the characteristics of the cloud platform mean that users no longer need to waste time downloading and processing data. Monitoring and tracking the SDGs in any region of interest in the world could be feasible at any time and any place, thus breaking through the boundaries of previous borders, regions and fields and enabling humans to jointly manage the environmental resource problem that human beings are facing for the first time.

Designing free, reusable and customizable cloud reporting services for the SDGs is one of the key instruments for SDGs implementation and review (Biggeri et al., 2019). There are many applications for integrating such data for multiple goals, such as predicting harvests, disaster response, and food security situations (Buontempo et al., 2019); monitoring geographic patterns and likely transmission corridors of diseases; measuring population density and the spread of new settlements; and mapping and assessing the impacts of transportation infrastructure. Some of these applications are now cloud-based, which enables stakeholders worldwide to access and complete their own monitoring and assessment and promote public participation in monitoring and tracking the SDGs. These applications need to be enhanced for use as predictive and anticipative services.

Demonstrations show that adoption of cloud services is attractive while the relevant user capacity is being developed (Buontempo et al., 2019; Wu et al., 2015). However, on the one hand, countries, particularly institutions in developing countries, require continuous data and tools to support their own business to expand and accelerate the implementation of the 2030 Agenda (Independent Group of Scientists appointed by the Secretary-General, 2019). On the other hand, cloud services face a challenge in providing complementary support to existing statistical systems for monitoring and tracking the SDGs. Specific customization is needed for a particular country or topics by tailoring data and tools for users' requirements and specific applications.

6. Conclusions

Data are essential for monitoring and tracking the SDGs to ensure that the desired targets are achieved by 2030. To close the existing data gap in serving the SDGs, a novel and effective way to collect and use data for the SDGs is urgent to save the unaffected cost of traditional statistical methods. We discussed the potential of crowdsourced and public EO data for providing cost-effective, shareable and timely data for monitoring and tracking the SDGs and proposed the development of free, reusable and customizable reporting cloud services for the SDGs. Before this can be done, there are some bottlenecks, including streamlining continuous crowdsourcing data for specific purposes, global efforts on quality control and validation of public EO data and to customizing cloud services to support national or international organizations in monitoring and tracking the SDGs. Crowdsourcing, public EO data and cloud services allow public participation in and improve transparency in monitoring and tracking the SDGs.

Declaration Competing of Interest

The authors declare no conflict of interest.

Acknowledgements

This study was funded by the National Key Research and Development Program of China (Grant No. 2016YFA0600304) and the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDA19030201).

References

Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Niño, F., Weiss, M., Samain, O., Roujean, J.L., 2007. LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION: Part 1: Principles of the algorithm. Remote Sens. Environ. 110 (3), 275–286.

Biggeri, M., Clark, D., Ferrannini, A., Mauro, V., 2019. Tracking the SDGs in an 'integrated' manner: A proposal for a new index to capture synergies and trade-offs between and within goals. World Dev. 122, 628–647.

Blumenstock, J., Cadamuro, G., On, R., 2015. Predicting poverty and wealth from mobile phone metadata. Science 350 (6264), 1073–1076.

Buontempo, C., Hutjes, R., Beavis, P., Berckmans, J., Cagnazzo, C., Vamborg, F., Thepaut, J., Bergeron, C., Almond, S., Amici, A., 2019. Fostering the development of climate services through Copernicus Climate Change Service (C3S) for agriculture applications. Weather Climate Extremes, 100226.

Campbell, J., 2019. We Lack Data for 68% of SDG Indicators. https://unspbf.org/editorial/we-lack-data-for-68-of-sdg-indicators-closing-data-gaps-essential-to-achieving-sdgs/ (accessed 19 Mar 2020).

Castell, N., Dauge, F.R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Broday, D., Bartonova, A., 2017. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates. Environ. Int. 99, 293–302.

China Meteorological Administration. 2019. Data sharing from more than 20,000 hydrological and meteorological stations helps flood prevention in the Yangtze River. http://www.cma.gov.cn/2011xzt/2019zt/20190927/2019092706/201909/t20190930_536784.html (accessed 19 Mar 2020).

Chinese Academy of Science. 2019. Big Earth Data in Support of the Sustainable Development Goals. https://www.fmprc.gov.cn/mfa_eng/topics_665678/2030kcxfzyc/P020190924800116340503.pdf (accessed 19 Mar 2020).

Corbane, C., Pesaresi, M., Kemper, T., Politis, P., Florczyk, A.J., Syrris, V., Melchiorri, M., Sabo, F., Soille, P., 2019. Automated global delineation of human settlements from 40 years of Landsat satellite data archives. Big Earth Data 3, 140–169.

Dong, J., Xiao, X., Menarguez, M.A., Zhang, G., Qin, Y., Thau, D., Biradar, C., Moore III, B., 2016. Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. Remote Sens. Environ. 185, 142–154.

European Space Agency. 2019. EO4SD Agriculture and Rural Development cluster. https://www.eo4idi.eu/sites/default/files/eo4sd_agri_finalreport_en_2019-final.pdf (accessed 19 Mar 2020).

FAO. 2019. Investing in data for the SDGs: Why good numbers matter. http://www.fao.org/partnerships/resource-partners/news/news-article/en/c/1200471/ (accessed 19 Mar 2020).

Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., Kraxner, F., Obersteiner, M., 2009. Geo-Wiki. Org: The use of crowdsourcing to improve global land cover. Remote Sens. 1, 345–354.

Fritz, S., See, L., McCallum, I., You, L., Bun, A., Moltchanova, E., Duerauer, M., Albrecht, F., Schill, C., Perger, C., 2015. Mapping global cropland and field size. Global Change Biol. 21 (5), 1980–1992.

Giuliani, G., Mazzetti, P., Santoro, M., Nativi, S., Lehmann, A., 2020. Knowledge generation using satellite earth observations to support sustainable development goals (SDG): A use case on Land degradation. Int. J. Appl. Earth Observ. Geoinform. 88, 102068.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., 2013. Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. Int. J. Remote Sens. 34 (7), 2607–2654.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.

Guo, H., 2017. Big Earth data: A new frontier in Earth and information sciences. Big Earth Data 1 (1–2), 4–20.

Guo, H., Liu, Z., Jiang, H., Wang, C., Liu, J., Liang, D., 2016. Big Earth Data: a new challenge and opportunity for Digital Earth's development. Int. J. Digital Earth 10 (1), 1–12.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S., Tyukavina, A., Thau, D., Stehman, S., Goetz, S., Loveland, T.R., 2013. High-resolution global maps of 21st-century forest cover change. Science 342 (6160), 850–853.

He, B., Wu, J., Lü, A., Cui, X., Zhou, L., Liu, M., Zhao, L., 2013. Quantitative assessment and spatial characteristic analysis of agricultural drought risk in China. Nat. Hazards 66 (2), 155–166.

IEAG, 2014. A World that Counts–Mobilising the Data Revolution for Sustainable Development. The United Nations Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development.

Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., Ermon, S., 2016. Combining satellite imagery and machine learning to predict poverty. Science 353 (6301), 790–794.

Kalia, A.C., Frei, M., Lege, T., 2017. A Copernicus downstream-service for the nationwide monitoring of surface displacements in Germany. Remote Sens. Environ. 202, 234–249.

Kennedy, R.E., Yang, Z., Cohen, W.B., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—Temporal segmentation algorithms. Remote Sens. Environ. 114 (12), 2897–2910.

Killian, S., Lannon, J., Murray, L., Avram, G., Giralt, M., Oriordan, S., 2019. Social Media for Social Good: Student engagement for the SDGs. Int. J. Manage. Educ. 17 (3), 100307.

Leeper, R.D., Rennie, J., Palecki, M.A., 2015. Observational Perspectives from US Climate reference network (USCRN) and cooperative observer program (COOP) network: Temperature and precipitation comparison. J. Atmos. Oceanic. Technol. 32 (4), 703–721.

Liu, Q., Reichle, R.H., Bindlish, R., Cosh, M.H., Crow, W.T., de Jeu, R., De Lannoy, G.J., Huffman, G.J., Jackson, T.J., 2018. The contributions of precipitation and soil moisture observations to the skill of soil moisture estimates in a land data assimilation system. J. Hydrometeorol. 12 (5), 750–765.

Long, T., Zhang, Z., He, G., Jiao, W., Tang, C., Wu, B., Zhang, X., Wang, G., Yin, R., 2019. 30 m Resolution Global Annual Burned Area Mapping Based on Landsat Images and Google Earth Engine. Remote Sens. 11 (5), 489.

Mondal, P., Mcdermid, S.S., Qadir, A., 2020. A reporting framework for Sustainable Development Goal 15: Multi-scale monitoring of forest degradation using MODIS, Landsat and Sentinel data. Remote Sens. Environ. 237, 111592.

Nabil, M., Zhang, M., Bofana, J., Wu, B., Stein, A., Dong, T., Zeng, H., Shang, J., 2020. Assessing factors impacting the spatial discrepancy of remote sensing based cropland products: A case study in Africa. Int. J. Appl. Earth Observ. Geoinform. 85, 102010.

Pal, M., 2005. Random forest classifier for remote sensing classification. Int. J. Remote Sens. 26 (1), 217–222.

Pekel, J.F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. Nature 540 (7633), 418–422.

Ramm, F., Topf, J., Chilton, S., 2010. OpenStreetMap: Using, and Contributing to, the Free World Map. UIT Cambridge, Cambridge.

Rollason, E., Bracken, L.J., Hardy, R.J., Large, A.R.G., 2018. The importance of volunteered geographic information for the validation of flood inundation models. J. Hydrol. 562, 267–280.

Rosenqvist, A., Shimada, M., Suzuki, S., Ohgushi, F., Tadono, T., Watanabe, M., Tsuzuku, K., Watanabe, T., Kamijo, S., Aoki, E., 2014. Operational performance of the ALOS global systematic acquisition strategy and observation plans for ALOS-2 PALSAR-2. Remote Sens. Environ. 155, 3–12.

Sauer, J.R., Link, W.A., Fallon, J.E., Pardieck, K.L., Ziolkowski Jr, D.J., 2013. The North American breeding bird survey 1966–2011: summary analysis and species accounts. N. Am. Fauna 79, 1–32.

Savtchenko, A., Ouzounov, D., Ahmad, S., Acker, J., Leptoukh, G., Koziana, J., Nickless, D., 2004. Terra and Aqua MODIS products available from NASA GES DAAC. Adv. Space Res. 34 (4), 710–714.

Schmidt-Traub, G., De la Mothe Karoubi, E., Espey, J., 2015. Indicators and a monitoring framework for the Sustainable Development Goals: Launching a data revolution for the SDGs. Sustainable Development Solutions Network.

Schmidt-Traub, G., Kroll, C., Teksoz, K., Durand-Delacre, D., Sachs, J.D., 2017. National baselines for the Sustainable Development Goals assessed in the SDG Index and Dashboards. Nat. Geosci. 10 (8), 547–555.

Shah, M. 2019. Alibaba Cloud Ranked Third Globally. https://www.techfunnel.com/information-technology/alibaba-cloud-ranked-third-globally/ (accessed 19 Mar 2020).

Shao, Y., Di, L., Bai, Y., Guo, B., Gong, J., 2012. Geoprocessing on the Amazon cloud computing platform—AWS. In. 2012 first international conference on agro-geoinformatics (agro-geoinformatics). IEEE, pp. 286–291.

Steele, J.E., Sundsøy, P.R., Pezzulo, C., Alegana, V.A., Bird, T.J., Blumenstock, J., Bjelland, J., Engø-Monsen, K., de Montjoye, Y.-A., Iqbal, A.M., Hadiuzzaman, K.N., Lu, X., Wetter, E., Tatem, A.J., Bengtsson, L., 2017. Mapping poverty using mobile phone and satellite data. J. R. Soc. Interface 14, 20160690.

Stehman, S.V., Fonte, C.C., Foody, G.M., See, L., 2018. Using volunteered geographic information (VGI) in design-based statistical inference for area estimation and accuracy assessment of land cover. Remote Sens. Environ. 212, 47–59.
Sustainable Development Solutions Network, 2015. Data for development: A needs assessment for SDG monitoring and statistical capacity development.
Tian, F., Wu, B., Zeng, H., Zhang, X., Xu, J., 2019. Efficient Identification of Corn Cultivation Area with Multitemporal Synthetic Aperture Radar and Optical Images in the Google Earth Engine Cloud Platform. Remote Sens. 11 (6), 629.
Tulloch, D., 2014. Crowdsourcing geographic knowledge: volunteered geographic information (VGI) in theory and practice. Int. J. Geog. Inf. Sci. 28 (4), 847–849.
United Nations, 2015. Transforming our world: The 2030 agenda for sustainable development.
United Nations, 2019. The Sustainable Development Goals Report 2019 (accessed 19 Mar 2020).
Van Achteren, T., Dierckx, W., Sterckx, S., Livens, S., Saint, G., 2012. Proba-V, a SPOT-VGT Successor Mission, Product Definition and Specifications. In: 32nd EARSEL Symposium Proceedings–Advances in Geosciences, Mykonos Island.
Vaz, A.S., Goncalves, J., Pereira, P., Santarem, F., Vicente, J.R., Honrado, J., 2019. Earth observation and social media: Evaluating the spatiotemporal contribution of non-native trees to cultural ecosystem services. Remote Sens. Environ. 230, 111193.
Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S.D., Tegmark, M., Nerini, F.F., 2020. The role of artificial intelligence in achieving the Sustainable Development Goals. Nat. Commun. 11 (1), 233.
Whitcraft, A.K., Beckerreshef, I., Justice, C.O., Gifford, L., Kavvada, A., Jarvis, I., 2019. No pixel left behind: Toward integrating Earth Observations for agriculture into the United Nations Sustainable Development Goals framework. Remote Sens. Environ. 235, 111470.
Wu, B., Gommes, R., Zhang, M., Zeng, H., Yan, N., Zou, W., Zheng, Y., Zhang, N., Chang, S., Xing, Q., 2015. Global crop monitoring: a satellite-based hierarchical approach. Remote Sens. 7 (4), 3907–3933.
Wu, B., Meng, J., Li, Q., Yan, N., Du, X., Zhang, M., 2014. Remote sensing-based global crop monitoring: experiences with China's CropWatch system. Int. J. Digital Earth 7 (2), 113–137.
Wu, B.Q., Jinkai, Zeng, Yuan, 2017. Land Cover Atlas of the People's Republic of China (1: 1 000 000). SinoMaps Press, Beijing.
Wu, H., Zhang, L., Zhang, X., 2019. Cloud Data and Computing Services Allow Regional Environmental Assessment: A Case Study of Macquarie-Castlereagh Basin, Australia. Chin. Geogr. Sci. 29 (3), 394–404.
Xiong, J., Thenkabail, P.S., Gumma, M.K., Teluguntla, P., Poehnelt, J., Congalton, R.G., Yadav, K., Thau, D., 2017. Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. ISPRS J. Photogramm. Remote Sens. 126, 225–244.
Xu, Z., Chau, S.N., Chen, X., Zhang, J., Li, Y., Dietz, T., Wang, J., Winkler, J.A., Fan, F., Huang, B., 2020. Assessing progress towards sustainable development over space and time. Nature 577 (7788), 74–78.
Zhang, M. 2017a. GVG for Android. https://play.google.com/store/apps/details?id=com.sysapk.gvg (accessed 19 Mar 2020).
Zhang, M. 2017b. GVG for IOS. https://itunes.apple.com/us/app/apple-store/id1244686128?mt=8 (accessed 19 Mar 2020).
Zhang, L., Li, X., Yuan, Q., Liu, Y., 2014. Object-based approach to national land cover mapping using HJ satellite imagery. J. Appl. Remote Sens. 8, 083686.
Zhang, X., Wu, B., Ponce-Campos, G., Zhang, M., Chang, S., Tian, F., 2018. Mapping up-to-Date Paddy Rice Extent at 10 M Resolution in China through the Integration of Optical and Synthetic Aperture Radar Images. Remote Sens. 10 (8), 1200.
Zhao, X., Liang, S., Liu, S., Yuan, W., Xiao, Z., Liu, Q., Cheng, J., Zhang, X., Tang, H., Zhang, X., 2013. The Global Land Surface Satellite (GLASS) remote sensing data processing system and products. Remote Sens. 5 (5), 2436–2450.