

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/341228837>

Google Earth Engine for geo-big data applications: A meta-analysis and systematic review

Article in *ISPRS Journal of Photogrammetry and Remote Sensing* · May 2020

DOI: 10.1016/j.isprsjprs.2020.04.001

CITATIONS
56

READS
3,394

6 authors, including:



Haifa Tamiminia
State University of New York College of Environmental Science and Forestry
6 PUBLICATIONS 89 CITATIONS

[SEE PROFILE](#)



Bahram Salehi
State University of New York (SUNY) College of Environmental Science and Forestry...
100 PUBLICATIONS 1,587 CITATIONS

[SEE PROFILE](#)



Masoud Mahdianpari
C-CORE and Memorial University of Newfoundland
59 PUBLICATIONS 1,018 CITATIONS

[SEE PROFILE](#)



Lindi J Quackenbush
State University of New York College of Environmental Science and Forestry
60 PUBLICATIONS 2,228 CITATIONS

[SEE PROFILE](#)

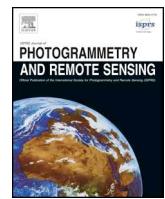
Some of the authors of this publication are also working on these related projects:



Masters of Science, ESF, Syracuse, New York [View project](#)



Remote Sensing for Wetland Mapping and Monitoring [View project](#)



Google Earth Engine for geo-big data applications: A meta-analysis and systematic review

Haifa Tamiminia^a, Bahram Salehi^a, Masoud Mahdianpari^{b,*}, Lindi Quackenbush^a, Sarina Adeli^a, Brian Brisco^c

^a Department of Environmental Resources Engineering, State University of New York College of Environmental Science and Forestry (ESF), NY 13210, USA

^b C-CORE and Department of Electrical and Computer Engineering, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada

^c The Canada Centre for Mapping and Earth Observation, Ottawa K1S 5K2, Canada



ARTICLE INFO

Keywords:

Google Earth Engine
Geo-big data
Cloud-based platform
Remote sensing
Planetary-scale
Geospatial
Machine learning
Environmental monitoring

ABSTRACT

Google Earth Engine (GEE) is a cloud-based geospatial processing platform for large-scale environmental monitoring and analysis. The free-to-use GEE platform provides access to (1) petabytes of publicly available remote sensing imagery and other ready-to-use products with an explorer web app; (2) high-speed parallel processing and machine learning algorithms using Google's computational infrastructure; and (3) a library of Application Programming Interfaces (APIs) with development environments that support popular coding languages, such as JavaScript and Python. Together these core features enable users to discover, analyze and visualize geospatial big data in powerful ways without needing access to supercomputers or specialized coding expertise. The development of GEE has created much enthusiasm and engagement in the remote sensing and geospatial data science fields. Yet after a decade since GEE was launched, its impact on remote sensing and geospatial science has not been carefully explored. Thus, a systematic review of GEE that can provide readers with the "big picture" of the current status and general trends in GEE is needed. To this end, the decision was taken to perform a meta-analysis investigation of recent peer-reviewed GEE articles focusing on several features, including data, sensor type, study area, spatial resolution, application, strategy, and analytical methods. A total of 349 peer-reviewed articles published in 146 different journals between 2010 and October 2019 were reviewed. Publications and geographical distribution trends showed a broad spectrum of applications in environmental analyses at both regional and global scales. Remote sensing datasets were used in 90% of studies while 10% of the articles utilized ready-to-use products for analyses. Optical satellite imagery with medium spatial resolution, particularly Landsat data with an archive exceeding 40 years, has been used extensively. Linear regression and random forest were the most frequently used algorithms for satellite imagery processing. Among ready-to-use products, the normalized difference vegetation index (NDVI) was used in 27% of studies for vegetation, crop, land cover mapping and drought monitoring. The results of this study confirm that GEE has and continues to make substantive progress on global challenges involving process of geo-big data.

1. Introduction

The term "big data" first emerged in scientific communities in the mid-1990s (Li et al., 2016) and gained popularity in 2006, a year after the creation of the second generation of the World Wide Web (Moed, 2012). In general, big data refers to a tremendous and complicated dataset that is difficult to store, manage, and process using traditional processing tools (Liu, 2015). Importantly, big data is characterized by three dimensions known as the 3Vs (Laney, 2001). First is volume, an inherent characteristic of big data that comprises a huge volume of data from various sources which poses challenges for storage and analysis

(Chi et al., 2016; Li et al., 2016). Second is variety, because big data typically comes in various types and formats, which may come to the user already combined in some fashion and/or may need to be combined by the user for a specific purpose. Accordingly, tremendous efforts have been conducted to manipulate different data types with complex structures. The third is velocity, which deals with the unprecedented speed of data streams emerging from different sources (Li et al., 2016). Over the past few years, big data analysis has drawn attention in different disciplines, such as business, health science, disaster management, and geoscience.

The growth of geospatial data has changed our perception and

* Corresponding author.

E-mail address: m.mahdianpari@mun.ca (M. Mahdianpari).

interaction with the planet. Given the massive volumes of existing geospatial data, its variety of origins and formats, and growing diversity and accessibility, it can be defined as big data (Lane, 2001; Mahdianpari et al., 2020). Geo-big data is collected from different sources such as ground surveying, remote sensing, geo-located sensors, and mobile mapping. When it comes to remote sensing big data, special intrinsic and extrinsic characteristics can be determined. Dynamic-state, multi-scale, and non-linear features are intrinsic characteristics of remote sensing big data (Liu, 2015). In particular, remote sensing big data reflects a dynamic state, as the Earth's surface changes continuously. Multi-scale features are related to resolution, time interval, spectral range, angle and polarization (Liu, 2015; Li et al., 2016). In addition, remote sensing big data is nonlinear, since time series data are typically non-linear and noisy. On the other hand, the multi-source, high-dimensional and isomer characteristics are extrinsic characteristics of remote-sensing big data (Liu, 2015). The reason behind the first two characteristics is the existence of different sensors and spectral/temporal dimensions of satellite data, respectively. The isomer characteristic reflects the variation in structure of available remote sensing data, such as raster or vector. These characteristics raise several challenges, including the acquisition, storage, searching, sharing, transferring, analysis, and visualization of big data (Liu, 2015). To overcome these difficulties, the need for novel methods is imperative.

To address the existing challenges in geo-big data analysis, two platforms with different system architectures are commonly used: cluster-based high performance computing (HPC) systems and cloud platforms (Ma et al., 2015). In cluster-based systems, huge computational problems are accomplished by the cooperation of multiple computers presenting a single-system image (Ma et al., 2015; Plaza and Chang, 2007). Although these cluster-based systems have a massive computational capacity, they suffer from loading and processing of tremendous volumes of data. Cloud platforms virtualize the supercomputing infrastructure like actual physical computers. However, compared to HPC systems, cloud platforms offer more accessibility, and affordability with flexible processors, memory and disk size. In particular, cloud computing presents a cloud storage for storing big data with accessible scalability (Ma et al., 2015). Furthermore, these cloud computing systems present infrastructure, platform, storage, and software as a la carte services. For example, Amazon EC2 provides

infrastructure as a service (IaaS), Microsoft Azure offers platform as a service (PaaS), and Google Earth Engine releases software as a service (SaaS) (Ma et al., 2015).

There are different cloud computing sources for geo-big data processing. Amazon web services (AWS), launched in 2006, offers a pay-as-you-go cloud computing platform for users to construct their own virtual data centers (Amazon, 2015). The AWS environment benefits from having access to the largest suite of machine learning and Artificial Intelligence (AI) services. AWS contains several types of satellite imagery data, for instance, Sentinel-1, Sentinel-2, Landsat 8 and the National Oceanographic and Atmospheric Administration Advanced (NOAA) High-Resolution Rapid Refresh (HRRR) Model. Google Cloud Platform (GCP), offered by Google in 2008, is a public cloud-based service for developing and hosting web applications in Google-managed data centers on an as-you-go basis. It provides a series of services such as data storage, data analysis, machine learning tools and enterprise mapping services (Krishnan and Gonzalez, 2015).

In 2010, Microsoft created a cloud computing platform called Azure for building, testing, deploying and managing applications and services through Microsoft-managed data centers (Wilder, 2012). From a remote sensing big data perspective, Azure contains machine learning services, a harmonized Landsat-Sentinel-2 product for North America from 2013 to the present, and satellite imagery from the Moderate Resolution Imaging Spectroradiometer (MODIS) since 2000. In 2011, the International Business Machines (IBM) cloud platform was launched to support storage and networking for either small organizations or large enterprise businesses (Lin et al., 2009). IBM cloud computing offers a combination of PaaS, IaaS and SaaS services through public, private and hybrid cloud delivery models.

Recently, Google Earth Engine (GEE) has been in the remote sensing big data processing spotlight. GEE is a cloud-based platform that enables parallelized processing of geospatial data on a global scale using Google's cloud (Gorelick et al., 2017). GEE is a free cloud platform and hosts petabyte scales of over 40 years of remotely-sensed data, such as Landsat, MODIS, National Oceanographic and Atmospheric Administration Advanced Very High Resolution Radiometer (NOAA AVHRR), Sentinel 1, 2, 3 and 5-P; and Advanced Land Observing Satellite (ALOS) data. (Gorelick et al., 2017). GEE also includes climate/weather and geophysical datasets. Additional ready-to-use products, such as the

Table 1
A summary of the algorithms and capabilities available in code editor-Google Earth Engine.

Package	Capabilities	Package	Capabilities
Machine learning	Supervised Classification Unsupervised Classification TensorFlow	Reducer	Image collection reductions Image Reductions Statistics of an image region
Image	Image Visualization RGB composites Color plates Masking Mosaicking Clipping Rendering categorical maps Thumbnail images Operations (mathematical, Boolean, morphological, convolutions, relational, conditional) Edge detection Texture Spatial Transformation Object-based Methods Registration Filtering Mapping Reducing Vector to raster Interpolation Landsat algorithms Sentinel-1 algorithms Resampling and Reducing Resolution	Charts	Statistics of image regions Statistics of image neighborhoods Statistics of feature collection columns Raster to vector conversion Vector to raster conversion Grouped reductions and zonal statistics Weighted reductions Linear Regression Time-series charts Histograms Image Regions charts Time-series in Image Regions Day-of-year Charts Charts by image classes
Geometry, Feature, Feature Collection		Image Collection	Filtering Mapping Reducing Composing and Mosaicking
Specialized algorithms			Iterating over an image collection

Table 2

Environmental studies that used GEE in different applications.

Application	Reference	Application	Reference
Forest mapping	B. Chen et al. (2017)	Lake and river mapping	F. Chen et al. (2017b)
Drought monitoring	Rembold et al. (2019)	Crop yield estimation	Chen et al. (2019)
Land use/land cover	Saah et al. (2019)	Evapotranspiration	Walker and Venturini (2019)
Fire	Parks et al. (2019)	Shoreline	Hagenaars et al. (2018)
Surface water detection	Wang et al. (2018)	Wetland	Mahdianpari et al. (2020)
Paddy rice	Zhang et al. (2018)	Albedo trend	Chrysoulakis et al. (2019)
Flood	Uddin et al. (2019)	Urban	Liu et al. (2018)
Snow	Snipir et al. (2019)	Soil	Castelli et al. (2019)
Mine mapping	Lobo et al. (2018)	Species habitat monitoring	Callaghan et al. (2018)
Disease risk mapping	Carrasco-Escobar et al. (2019)	Natural hazard management	Quintero et al. (2019)

Enhanced Vegetation Index (EVI) and the Normalized Difference Vegetation Index (NDVI), are also available (Kumar and Mutanga, 2018). In addition to the availability of a large repository of raw remotely sensed imagery, users have access to preprocessed, cloud-removed and mosaicked images in the GEE data catalog. Table A1, Appendix A, lists the available satellite and aerial imagery in the GEE platform.

The GEE platform leverages Google's computational infrastructure to enable parallel geospatial data processing to reduce computational time. APIs with full featured development environment for JavaScript and Python hosted on GitHub also facilitate requests to the Earth Engine servers. In addition, it provides a Git repository for storing, sharing and script versioning of users' codes that leads to more user collaboration (Gorelick et al., 2017). Another feature of GEE is code editor, which is available through a web-based Integrated Development Environment (IDE) and is designed for writing, developing and running complex scripts applying JavaScript API (Kumar and Mutanga, 2018). Importantly, the GEE code editor contains various algorithms that simplify script writing for both experts and non-experts. Several packages are available, for instance machine learning, image processing, image collection, geometry-feature, reducer, charts and specialized algorithms. Table 1 provides a comprehensive list of the capabilities of different GEE packages.

GEE also provides other ways to interact with users. The explorer is a straightforward web app for data catalog exploring, visualization, and basic analyses that allows users to run simple analyses. Additionally, timelapse is a video tool with zoom capabilities that allows tracking, measuring, and visualizing changes to the Earth's surface over the past 35 years (1984–2018). Users can also build their own timelapse dataset. The Earth Engine app provides access to a client-side user interface API for app enthusiastic developers to build and publish their own apps. Earth Engine apps are dynamic, accessible user interfaces and widely available for GEE analyses by experts and non-experts. Examples include LandTrendr (Kennedy et al., 2018), Split Panel, Mozaic editor, global population explorer, global forest change explorer (Hansen et al., 2013) and linked maps.

As discussed, the capability of GEE to process remote sensing big data has led researchers to harness this technology for various environmental applications (Table 2). Archives of geospatial datasets within GEE along with its coding, sharing, parallel processing, and visualization abilities have made it an unprecedented competitor among all existing cloud platforms. Although several survey studies have been carried out in geospatial big data analysis, they mainly focused on characteristics, existing tools and applications in geo-big data (Chi et al., 2016; Li et al., 2016; Liu, 2015). In remote sensing, Kumar and Mutanga (2018) presented the first review on the usage of GEE. They provided a review based on applications, data used, author's affiliated institution and study regions; however, a comprehensive investigation of sensor types, methods, features, region extent, and resolution associated with different applications is lacking. Thus, the aim of this review paper is to provide a comprehensive survey of publications to find out which dataset and algorithms were more successful and reliable for

environmental monitoring at large scale and produced more accurate results. Through a meta-analysis, we identified, categorized and analyzed the publications.

2. Materials and methods

The ISI Web of Science and Google Scholar databases were used on and up to October 13, 2019, for refereed journal articles using the keyword "Google Earth Engine" constrained to a time span from 2010 (launch of GEE) to 2019. Next, we followed the methodology Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher et al., 2009) for the selection of articles to be included in our analysis (Fig. 1). Of the 2,391 initial results of the search, 349 were eligible to include in the database with the following features: publication date, first author, journal name, citation, article type, application, sensor type, datasets used, single-date/multi-temporal, study area, region extent (km^2), date, resolution, pixel-based/object-based, feature/band/channel, method, and accuracy. We excluded from analysis those articles that cited GEE as a secondary source and did not contain all the descriptive features listed above.

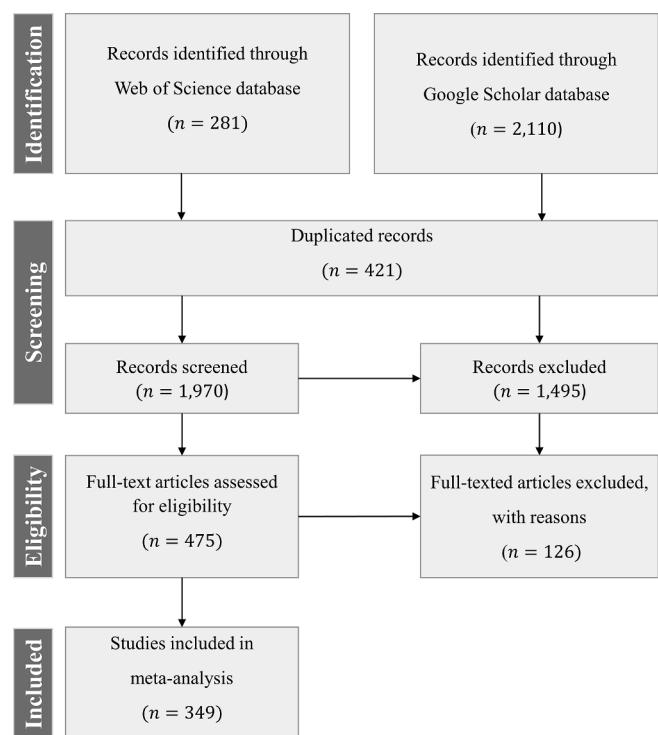


Fig. 1. PRISMA flow diagram for manuscript selection.

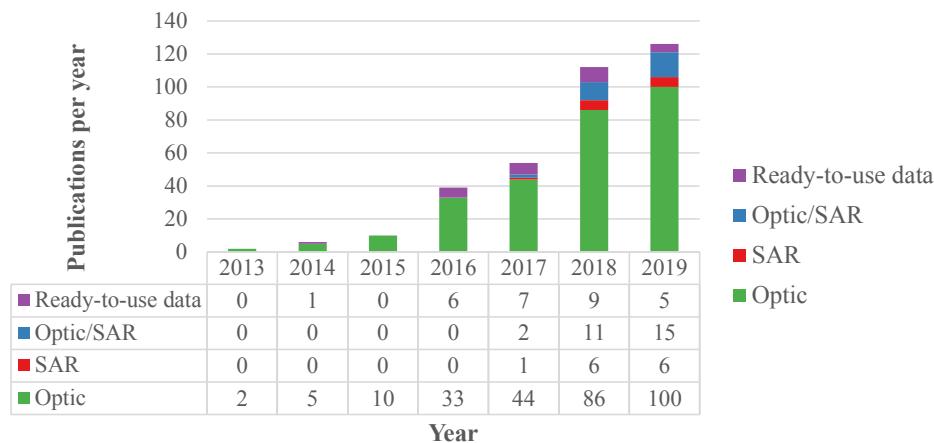


Fig. 2. Frequency of refereed journal publications using the Google Earth Engine platform and the primary data type(s) utilized. Results are based on a database search (date) for the period 2010–2019. No articles meeting our criteria were found prior to 2013.

3. Results

A total of 349 articles fit the criteria outlined in [Section 2](#). Based on the review of journal articles that used GEE in their applications, several categories of data have been extracted. This section presents a detailed review of the meta-analysis results. First, general characteristics about the articles, including the journals in which they were published, the disciplines that used GEE, the regions studied including the extent, and the topics of study are presented. Studies that use remote sensing data were further subdivided into those that used machine learning (i.e., classification and regression) and those that used other image processing techniques. Finally, the accuracies of studies that used machine learning techniques were assessed in detail.

3.1. General characteristics of studies using GEE

[Fig. 2](#) indicates the publication trends and data types among 349 articles that have been reviewed using PRISMA. Datasets are categorized into four groups, namely optical, Synthetic Aperture Radar (SAR), optical-SAR imagery and ready-to-use data products. While no peer-reviewed journal article was found from 2010 to 2012, [Fig. 2](#) shows a steadily increasing trend in the popularity of GEE since 2013. Following

the in-depth survey of publications in the last 7 years, it was revealed that only optical data were used in the early years. An increasing trend in the use of SAR and integration of optical-SAR data begins around 2017 with the availability of Sentinel-1. Ready-to-use data products, such as vegetation indices, land cover maps, digital elevation models, and soil moisture, have been used since 2016, but remain a small portion of the data used in published studies.

In total, the papers reviewed here were published in 146 different journals indicating the wide breadth of disciplines using GEE in their studies. Of these, 124 journals published only a single paper using GEE with various applications from climate change monitoring to archeology. Only journals that published more than five GEE articles are represented in [Fig. 3](#). As seen, *Remote Sensing*, *Remote Sensing of Environment* (RSE), and the *International Journal of Applied Earth Observation and Geoinformation* (JAG) are the top three journals to publish peer-reviewed papers on GEE applications.

As mentioned earlier, the 349 papers included in this study used GEE in a wide variety of remote sensing applications that can be categorized into 11 different groups. [Fig. 4](#) indicates that the largest number of papers were studies of crop mapping, including vegetation, rice paddy, and agricultural monitoring (74). This was followed by a considerable number of studies of water, such as surface water, lake,

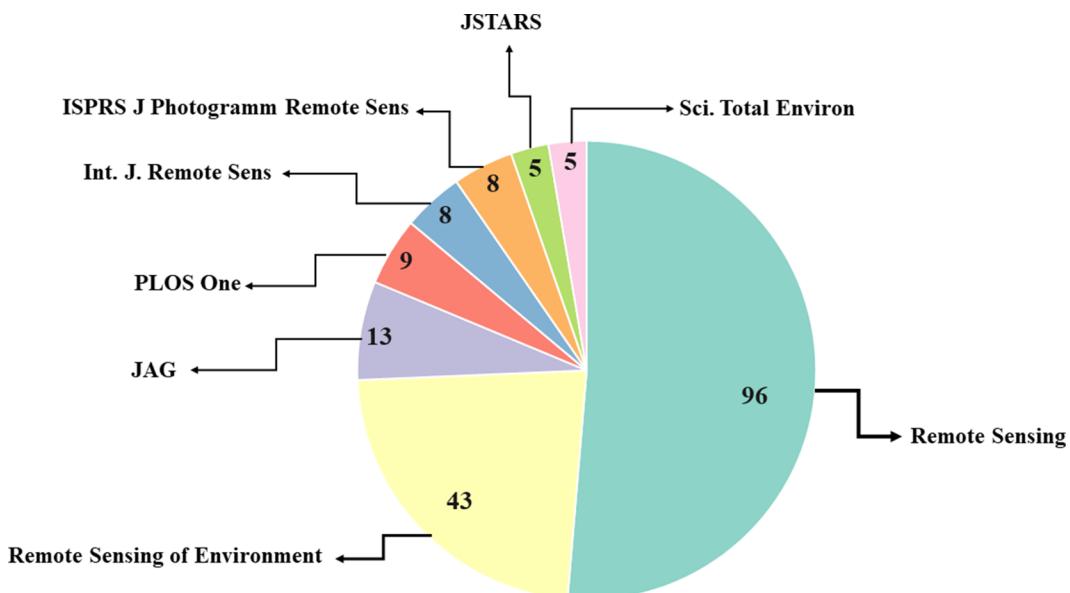


Fig. 3. Number of published GEE papers per journal.



Fig. 4. Categorization of GEE applications by discipline.

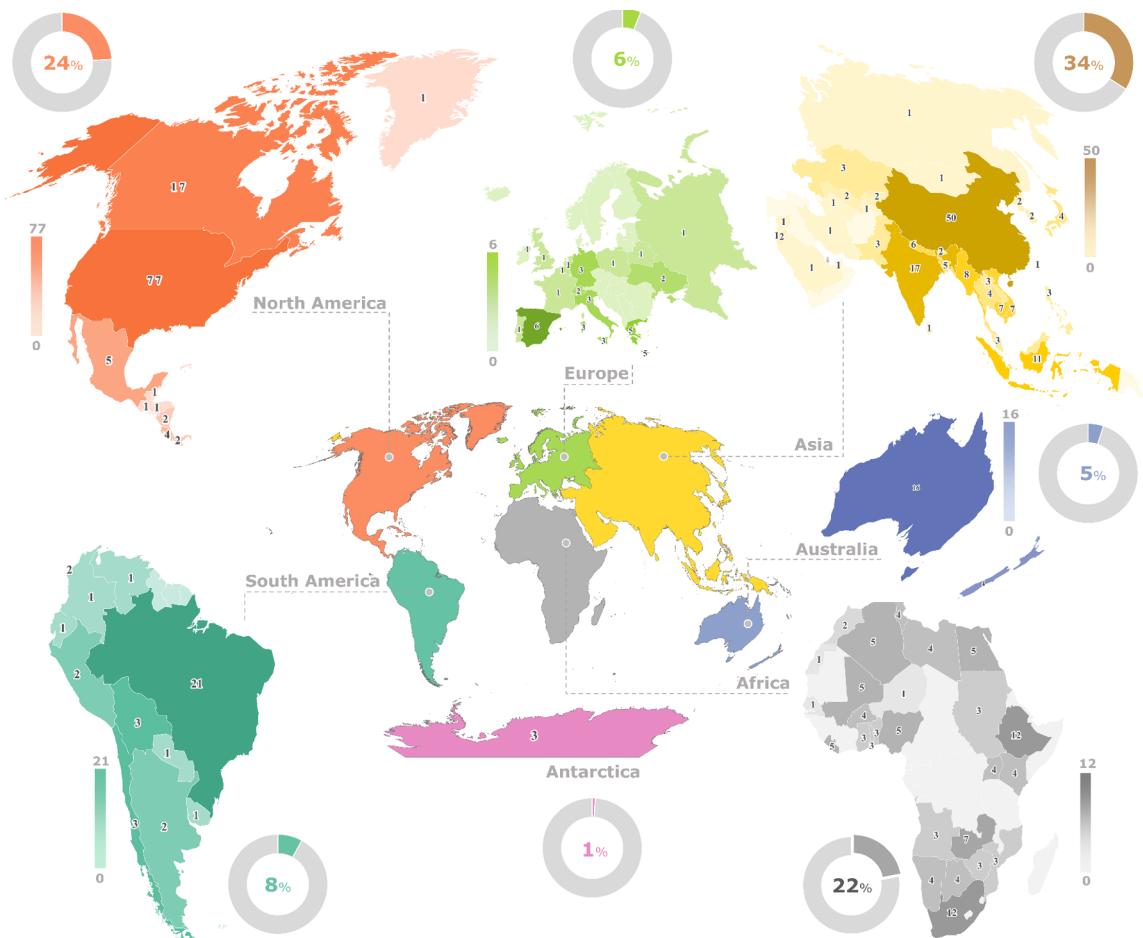
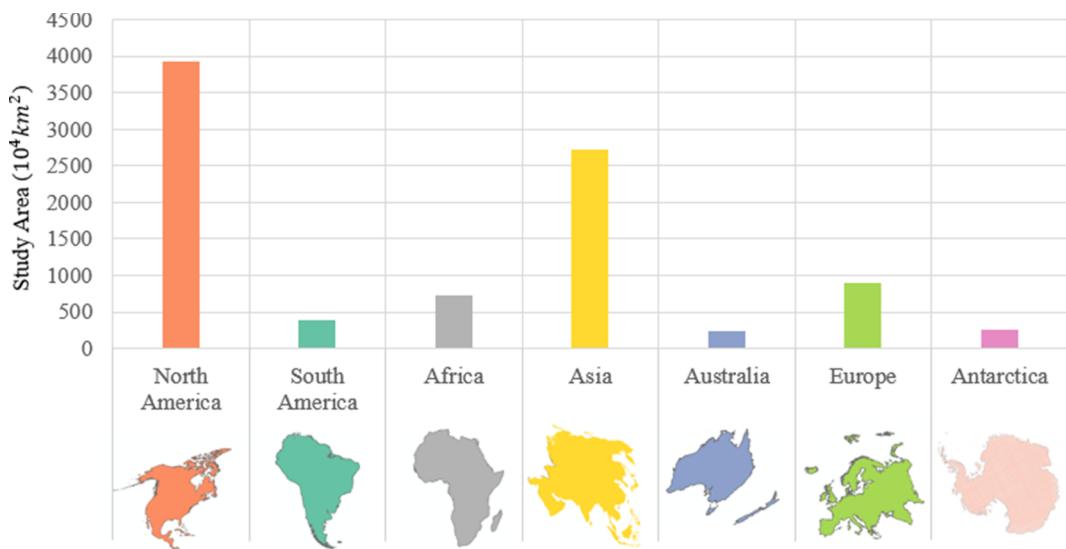


Fig. 5. Worldwide distribution of GEE studies.

Fig. 6. GEE continental study areas in km².

river, snow, glacial lake, algal bloom mapping, and shallow water bathymetry (62). Studies focusing on land use/land cover and forest mapping featured 56 and 30 articles, respectively. A total number of 28 articles focused on fire detection, flood, and drought monitoring in the disaster class. Climate change related applications, such as evapotranspiration estimation, atmosphere spectral characteristics retrieval, land and surface water temperature, albedo trends and heat island monitoring were investigated in 20 articles. Urban mapping studies were the subject of 15 research articles. Other topics were related to soil moisture and soil carbon sequestration (14), wetlands and mangroves (13), as well as data processing applications, including radiometric correction, mosaic image generation and cloud detection (12). Finally, a total of 25 studies that used GEE in a variety of applications were categorized in the “others” class, including archeology, mine and habitat mapping, geospatial monitoring and nuclear non-proliferation.

The worldwide distribution of study regions and their corresponding areas (in km²) are shown in Figs. 5 and 6, respectively. The GEE publications report on analysis in 104 countries from all 7 continents, which are illustrated in Fig. 5. As shown, most studies were conducted in the United States (77; including 3 studies for Alaska). China, with 50 articles, also represents a notable amount of studies. At a continental scale, the maximum numbers of studies belong to Asia, North America, and Africa, respectively; however, South America,

Europe, Australia, and Antarctica were the subject of a considerable number of studies.

Fig. 6 demonstrates the cumulative extent of studies in each continent (in km²). As indicated, North America has the greatest coverage area, whereas only limited areas of Australia have been studied using GEE. Importantly, a total number of 36 studies conducted at the global scale have not been included in these two figures.

Due to the wide spectrum of studies using different GEE capabilities, the meta-analysis was conducted by dividing articles into two major categories based on the data: remote sensing data and ready-to-use products (e.g., NDVI and land cover). Studies that utilized remote sensing data have been further classified by machine learning technique, including classification, regression, and an “Others” class consisting of methods, such as time-series, feature extraction, image composite, visual interpretation, and image pre-processing techniques. Fig. 7 demonstrates the classification of articles that utilized GEE with the number of studies in this review.

3.2. Satellite imagery in GEE

Of the 349 articles considered in this review, 312 processed satellite imagery in GEE. Of these, 265 used machine learning techniques, whereas 47 used other methods. Those that used machine learning

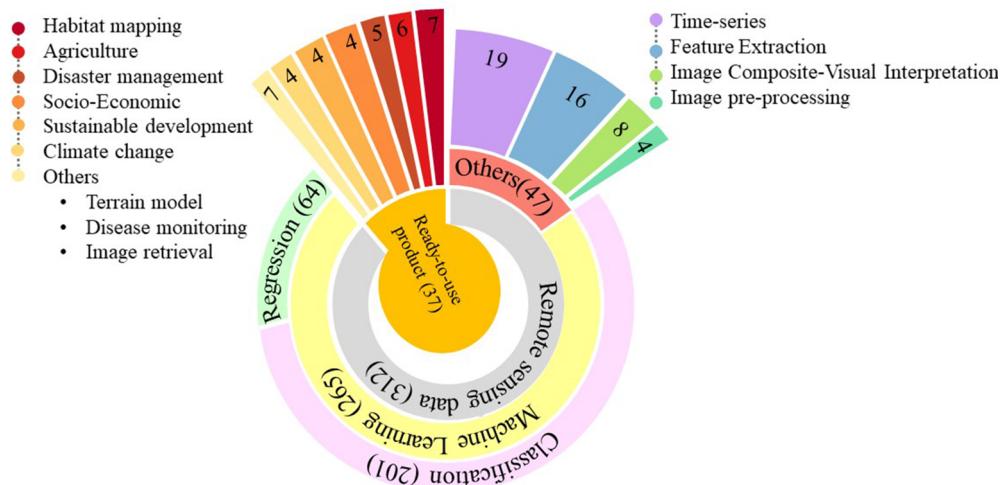


Fig. 7. Categorization of articles that utilized GEE.

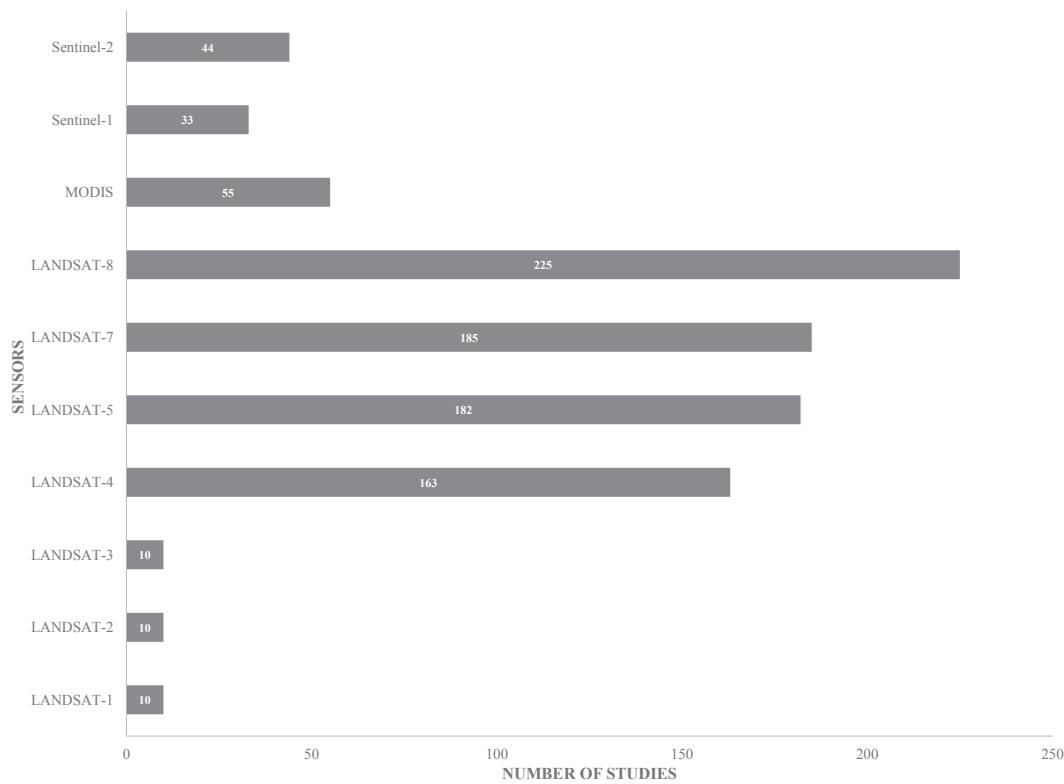


Fig. 8. Sensor types and number of studies.

techniques were further divided into image classification ($n = 201$) and regression ($n = 64$). The trends in these studies, and a detailed analysis of the accuracies of studies that use machine learning techniques are presented in the following sub-sections.

3.2.1. Sensor types and GEE

In terms of remote sensing data, 312 studies applied satellite and/or aerial imagery available in the GEE data catalogue. Sensor types that were included in at least 5 studies are shown in Fig. 8. As shown, optical images such as Landsat and MODIS are the most frequently used data source. Moreover, among all Landsat missions, Landsat-8 was most commonly adopted. Sentinel-2 and Sentinel-1 were also applied in 77 studies.

Many studies combine multiple datasets. Fig. 9 indicates the intersections of the top ten satellites that have been applied in GEE articles. As expected, Landsat missions placed first with 255 studies, followed by MODIS, Sentinel-2 and Sentinel-1 with 55, 44 and 33 studies, respectively. The combination of Landsat and MODIS in 18 studies shows the popularity of these two satellites, followed by Landsat-Sentinel-2 ($n = 13$), Landsat-Sentinel-1 ($n = 7$), Sentinel-1-Sentinel-2 ($n = 6$), and MODIS-Sentinel-1 ($n = 1$). A single study investigated the combination of data from four satellites.

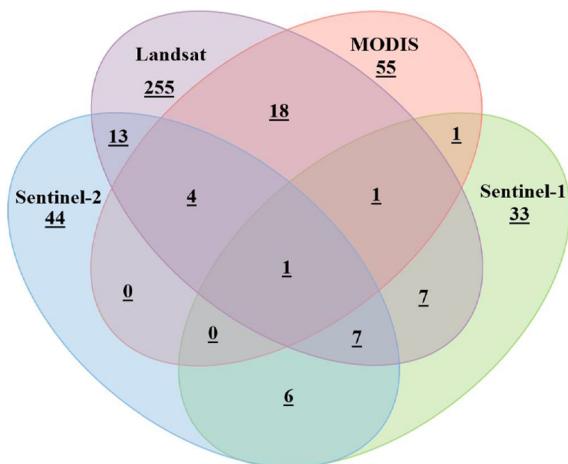
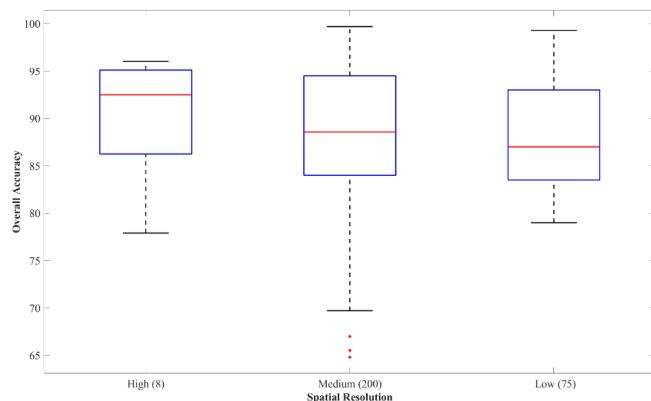
3.2.2. Classification and GEE

As seen in Fig. 7, the majority of studies included in this review used classification methods to process satellite images in GEE. As such, the following sub-sections assess the overall accuracies of these techniques through a consideration of different parameters, including spatial resolution, classifier type/method, data type, sensory type, and classification strategy.

3.2.2.1. Spatial resolution of satellite imagery processed in GEE. To investigate the impact of spatial resolution on classification performance, boxplot graphs were created (Fig. 10). As expected, data collected with high ($<4\text{m}$), medium ($4\text{ m to }30\text{ m}$), and low

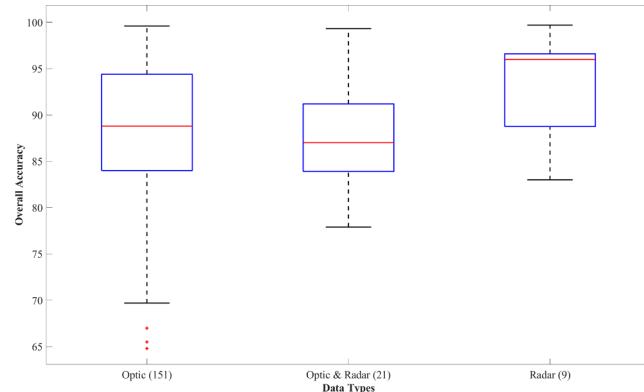
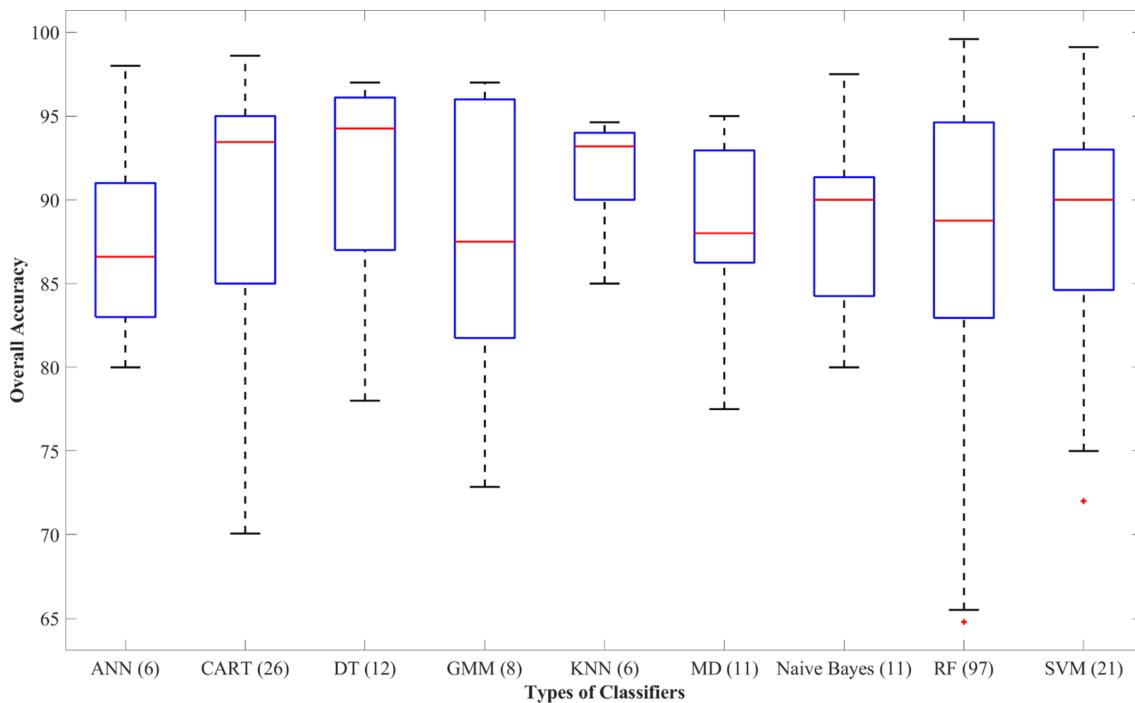
($>30\text{ m}$) spatial resolutions achieved the highest, intermediate, and lowest median in overall accuracies, respectively. Most studies ($n = 200$) applied medium resolution satellite imagery, including that from Landsat, Sentinel-1 and Sentinel-2, for their analysis. As illustrated, medium resolution data exhibits the highest maximum, yet its wider distribution with the lowest minimum shows less consistent results when compared to that of high- and low-resolution sensors. Studies focusing on high spatial resolution were least common ($n = 8$) although such data demonstrates more consistent results, as it obtains a more condensed shape compared to the others. The usefulness of low spatial resolution images for different applications was reported in 75 studies. However, it should be noted that spatial resolution is not the only parameter that affects the overall accuracy of remote sensing image classification. Other factors, such as data type, the nature of the classes, and classification method, are also important and should be considered.

3.2.2.2. GEE classification methods and their overall accuracies. Fig. 11 shows the classification accuracies of several approaches adopted within the GEE platform. As indicated, the median classification accuracies for all classifiers are greater than 85%. Fig. 11 illustrates that the highest median overall accuracy was achieved by the Decision Tree (DT) classifier, followed by the Classification and Regression Trees (CART) method. DT was superior compared to CART in terms of variation in overall accuracies (see the length of the boxplot in Fig. 11), yet the former was used less frequently relative to the latter. The K-nearest neighbor (KNN) method showed the least variation in overall accuracies compared to other classifiers with the interquartile range (IQR) of KNN significantly smaller than the other methods. The Support Vector Machine (SVM) and Naive Bayes classifiers show relatively equal strength in terms of overall accuracies. However, the variance and IQR of Naive Bayes were narrower compared to SVM. As shown, the Random Forest (RF) classification method was most commonly selected ($n = 97$) and exhibited both the minimum and maximum overall accuracies compared to other classifiers. This wide

**Fig. 9.** Combination of satellite imagery used in GEE studies.**Fig. 10.** Overall accuracy versus different spatial resolutions.

range of overall accuracy obtained with the RF classifier suggests that other parameters, such as the number of extracted features, are also influential factors on the classification performance. The median overall accuracy obtained with the minimum distance (MD) method is mediocre although it is superior compared to that of the Gaussian Mixture Model (GMM) and Artificial Neural Networks (ANN) classifiers. From a range point of view, the GMM classifier has the widest IQR range. Overall, the performance of the GMM classifier was not much different from those of other model types. Notably, the lowest median overall accuracy corresponds to ANN, yet 50% of accuracies obtained using this method are still greater than 87%.

3.2.2.3. Data type and overall accuracy. To determine the effect of different data types on overall accuracy, boxplots of this parameter are presented in Fig. 12. As seen, the median overall accuracies for all types of data is more than 85%. A total of 151 papers utilized optical imagery for their studies. Although the median overall accuracy of optical data is about 88%, the range between minimum and maximum extends from 70% to 99% (Fig. 12). The popularity of using these data for a wide

**Fig. 12.** Overall accuracy versus data types.**Fig. 11.** Overall accuracy of different classification methods.

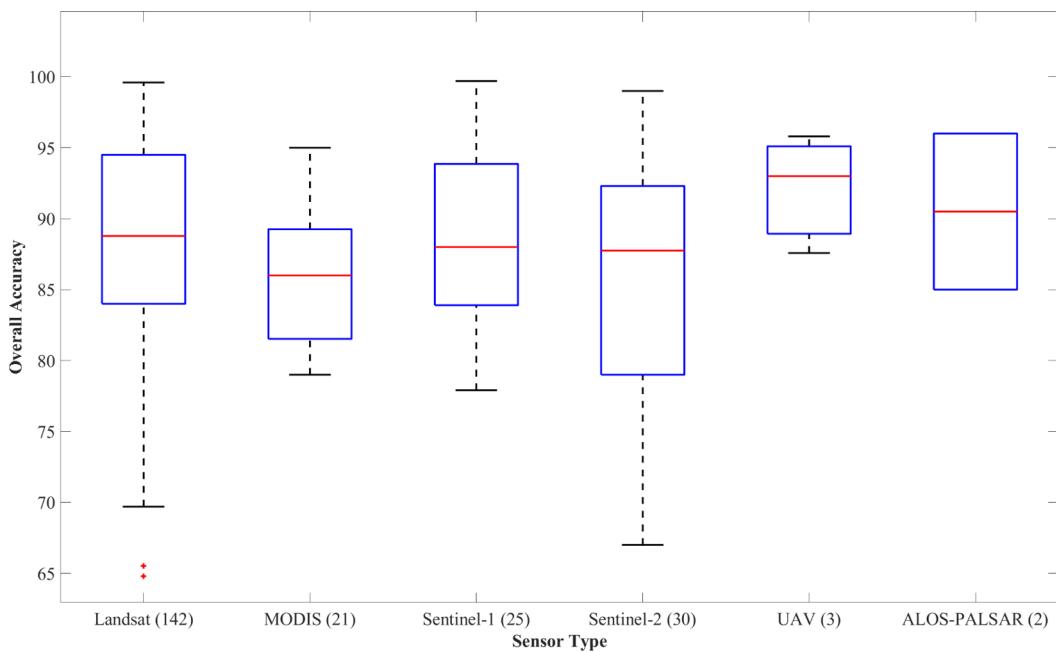


Fig. 13. Overall accuracy versus sensor type.

spectrum of applications has made the results less consistent and their accuracy strongly depends on the classifier type, study region and application. The integration of optical and radar data in 21 studies illustrated better results than optical data with a smaller IQR range. Nine studies utilized radar data alone, which produced the best results with a median of about 96% and a small variance.

3.2.2.4. Overall accuracy versus sensor type. Boxplots in Fig. 13 represent the overall accuracies of 6 sensors types. Although three studies used Unmanned Aerial Vehicle (UAV) data in their publications, the classification methods were implemented in GEE code editor. UAV, with spatial resolution less than 5 m, shows the best median with the lowest IQR range change and more reliable results (see Fig. 13). The use of radar data (found to produce the most accurate classification in Section 3.2.2.3) obtained using Sentinel-1 with a spatial resolution of 10 m also results in a high overall classification accuracy. Both Sentinel-2 and Landsat have a similar range of accuracy, which indicates their results are consistent, as expected. Landsat has been widely used in GEE papers and the wide spectrum of disciplines that use this sensor type has likely considerably influenced the range of results. Finally, MODIS has the low spatial resolution of the sensor types considered here, and expectedly has the lowest median overall accuracy.

Researchers have taken advantage of recent advances in remote sensing technologies and the increasing availability of high-resolution satellite imagery. High-resolution images offer potentially more accurate classification and analysis, which will greatly improve overall classification accuracy (See Fig. 10). Although GEE does not provide high resolution (<4m) satellite imagery, 4% of studies considered in this review uploaded high resolution imagery and carried out subsequent analysis in GEE. Table 3 provides the satellite datasets that do not exist in GEE but were used in these studies. Some of these studies also combined Landsat datasets with these high-resolution images.

3.2.2.5. Comparison of overall accuracy and different image processing strategies. Fig. 14 illustrates four types of implementation strategies for remote sensing data classification. The majority of papers selected a pixel- rather than object-based approach for classification. The IQR range and the difference between the minimum and maximum overall accuracies of the object-based boxplot is smaller than that of the pixel-based (Fig. 14). However, the pixel-based boxplot includes a wider

Table 3

Satellite images used in GEE publications that do not exist in GEE data catalogue.

Satellite	Spatial resolution (m)	Satellite	Spatial resolution (m)
Pleiades-1A satellite	0.5	Quick Bird	0.61
RADARSAT-1	10, 30, 100	WorldView-2	0.5
RADARSAT-2	1, 100	WorldView-3	0.31
ALOS_PALSAR	10, 100	Hyperion	30
ALOS-2_PALSAR-2	1, 3	SPOT	1.5
TanDEM-X	12	ASTER	15
Formosat-2	2, 4	AVHRR	1000
COSMO-Sky Med-1	1	UAV	2

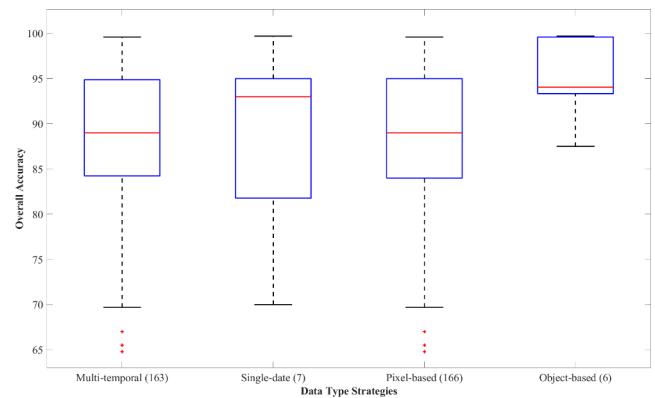


Fig. 14. Overall accuracy of different remote sensing strategies.

range of overall accuracies from 70% to 99% and its median lies below 90%. While the object-based methods used in 6 studies produced more accurate overall classifications, this may reflect the small number of studies to use the more complex object-based approach, whereas the larger range of the pixel-based approach may reflect its use in 166 papers.

From a temporal point of view, 163 papers used time-series whereas only 7 articles that used single-date datasets have been published. As seen in Fig. 14, there is little difference in the ranges of the boxplots.

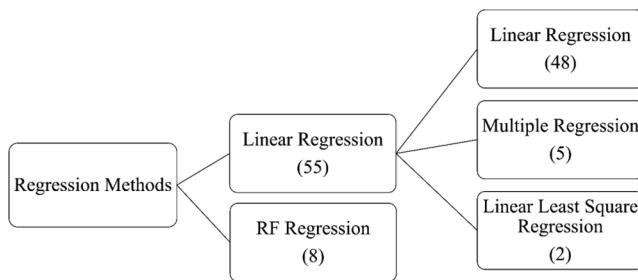


Fig. 15. Regression types that have been used in GEE.

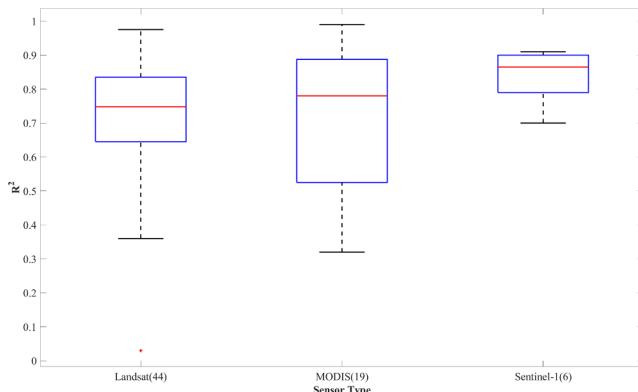


Fig. 16. Distribution of R^2 for different data types used in GEE.

Although the median of single-date boxplot lies higher than multi-temporal, the variance and the IQR range of time-series datasets change less. Therefore, studies with multi-temporal datasets presented more consistent results.

3.2.3. Regression and GEE

In terms of remote sensing data analysis, 64 studies used regression methods for their analysis. Fig. 15 shows different categories of regression types and the number of studies.

Fig. 16 illustrates boxplots representing the R^2 of different regression-based studies divided by sensor type. Sentinel-1 shows the highest median and the lowest IQR range in comparison with other sensors. This means that analyzing radar data with regression methods provided more correlated and consistent results. The range of Landsat is similar to that of MODIS; however, Landsat, the most used dataset, provided the lowest median and its IQR range is smaller. Thus, Landsat produced more correlated results than MODIS.

3.2.4. Others satellite image processing methods in GEE

A total of 47 papers were categorized as “others”, since different methods were used in their analysis instead of machine learning. The others class included the following techniques: time-series analysis, feature extraction, image pre-processing, image composite and visual interpretation. Image mosaic, cloud and error detection were incorporated into the image pre-processing category.

3.3. Ready-to-use products in GEE

As previously indicated, GEE provides various types of ready-to-use products derived from satellite imagery. A total of 37 papers included in this review utilized these ready-to-use products. These articles have been classified into 6 groups: habitat mapping, GIS agricultural land suitability, disaster management, socio-economic, sustainable development and “others”. The “others” category includes topographic modeling, disease monitoring and image retrieval (see Fig. 7). The most frequently used datasets are illustrated in Fig. 17. Among all available

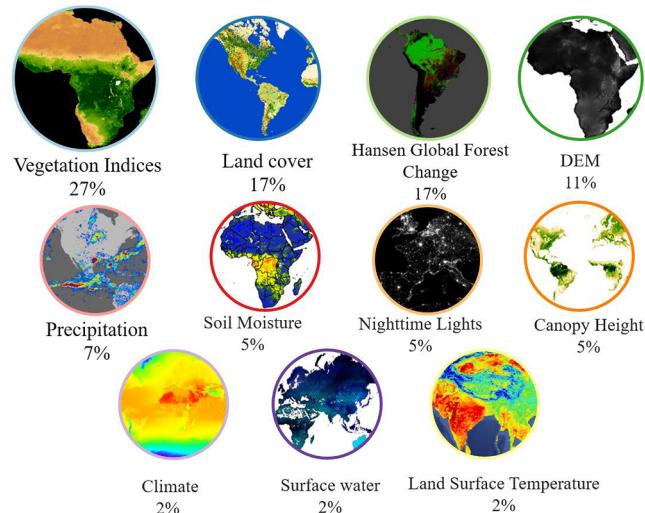


Fig. 17. A summary of ready-to-use products applied in GEE.

ready-to-use products in the GEE data catalogue, vegetation indices, land cover, Hansen global forest change and Digital Elevation Models (DEMs) have been most commonly applied (Hansen et al., 2013; Kong et al., 2019; Koskinen et al., 2019; Xiong et al., 2017).

4. Discussion

4.1. GEE capabilities and big data analysis

This paper presents a comprehensive review of articles using GEE from a meta-analysis point of view. Following the trend of publications, it can be argued that over time the popularity of GEE has increased, as has the number of published papers that utilized this geospatial analysis platform. In particular, a considerable number of studies have been published since 2017. In recent years advances in satellite imagery technologies have increased demand for remote sensing and geospatial applications, thus GEE is steadily becoming more widespread. GEE is a cloud computing platform that provides free satellite imagery with a Git repository for storing, sharing and geo-big data processing facilities (Gorelick et al., 2017).

Thanks to the GEE capabilities in dealing with big data analysis challenges, such as handling tremendous amount of data at unprecedented speed in a timely manner, many studies have performed geospatial analysis using remote sensing data in various applications at local and global scales. As previously mentioned, remote sensing big data analysis has intrinsic and extrinsic characteristics. The results of reviewing 349 studies that used GEE as a geo-big data analysis tool revealed that the number of research that employed multi-temporal data to monitor the desired phenomena is more than 24 times of single-date and thus, putting emphasis on the intrinsic characteristics of geo-big data. In addition, a large number of studies have used the combination of satellite, aerial, and geospatial datasets, indicating the multi-scale feature of the big data. In regard to extrinsic characteristics, GEE offers ready-to-use-products and remote sensing imagery with different spatial, spectral, and temporal resolutions. Therefore, GEE has broadened the potential of analyzing and processing geospatial and remote sensing big data.

As mentioned earlier, big data deals with various issues, such as capturing data, searching, sharing, storing, transferring, visualizing, querying, and updating information. With respect to remote sensing and geospatial data analysis, this is of particular concern. For example, the volume of Landsat-8 (only for one scene) is 1 GB for compressed file and 2 GB for the uncompressed one. With the advent of cloud platforms such as GEE, the NASA Earth Exchange, AWS, and Microsoft Azure, new opportunity for large-scale geospatial analyses has been opened (Hird

et al., 2017). Therefore, choosing a suitable cloud platform which has the capability to support and process those datasets is of paramount importance. Since each cloud computing platform has its own pros and cons, choosing an appropriate platform greatly depends on application. Overall, for remote sensing studies, the accessibility to satellite data and machine learning algorithms can make a big difference in the usage of cloud computing platforms. To compare the top three cloud platforms, GEE, AWS, and Azure, the availability of the open access satellite data along with the machine learning techniques and image processing tools has made GEE the most popular platform among the others. For example, GEE hosts a comprehensive archive of Landsat data since 1972, while AWS and Azure offer only Landsat 8 dataset. Sentinel-2 data is offered by all of them, while Sentinel-1 data is only available in GEE and AWS. As a result, although AWS and Azure platforms are older than GEE, most researchers tend to use GEE to get access to broad satellite imagery archives and geospatial datasets. This is because GEE is specifically designed for processing and analyzing geospatial dataset, making it advantageous for remote sensing applications compared to AWS and Azure. Researchers in GEE have access to a vast amount of publicly available data including a remote sensing archive with petabytes of data stored in the cloud. Furthermore, researchers have the option to upload and analyze their own images. This survey identified that 10% of studies carried out analyses with user-supplied images, ready-to-use products and GEE code editor API. One of the great advantages of GEE relative to Azure and AWS is the price of the platform. In particular, GEE is a free cloud platform, while AWS and Azure follow pay as you go services in terms of per hour and per minute, respectively. However, both AWS and Azure provide a free tier services that belong to the categories of free for a year and always free. The processing time is another difference between three platforms. When users have a paid account for AWS and Azure, they can estimate the processing time, whereas in GEE the processing time depends on the server and there is no control over it.

Data availability along with computing infrastructure and programming capabilities, such as JavaScript code editor, and API reference documentation as well as a Git-based script manager have motivated both professionals and non-professionals with low programming skills to consider GEE for their applications. In addition to scripting and code sharing tools, GEE presents a computation engine based on parallelizing processes and analyses on many CPUs in Google's data center. Unprecedented speed that reduces processing time and accessibility to an online platform with petabytes of remote sensing data, algorithms, advanced raster processing tools, and intensive computational infrastructure are the unique characteristics of GEE.

In terms of global distribution, the peer-reviewed publications suggest that GEE has been successful in terms of providing public equal access to geospatial data. The wide extent of study areas in all continents, and especially, in less-developed countries puts emphasis on the potentials of GEE. For instance, allowing users to upload their own data, apply pre-processed images, and utilize data storage capability have created a potential avenue for researchers to take advantage of GEE for studies across the world (Kumar and Mutanga, 2018).

4.2. GEE and data type

As Fig. 12 illustrates, more than 150 studies used optical imagery, which is more accessible and familiar to users. Thanks in large part to the free 40-year archive of Landsat imagery, optical remote sensing data remain the most frequently employed data source. The GEE data catalogue provides optical satellite imagery from 1972 to the present, enabling researchers to conduct earth monitoring studies. Moreover, the wide range of applications of GEE users suggest that optical images

are easier to process and interpret for non-remote sensing experts. This is one of the reasons why GEE has become popular throughout the world across a wide spectrum of scientific fields. Although SAR data alone was utilized in only nine studies reviewed here, more studies have concluded that integrating optical and SAR data can significantly increase the accuracy of classification, especially for multi-temporal analyses (Mahdianpari et al., 2017a; McNairn et al., 2009; Mohammadimanesh et al., 2018c).

For instance, the integration of SAR and optical satellites allow researchers to overcome difficulties and issues regarding cloud shadows (Stroppiana et al., 2015). The effect of cloud cover and low albedo surfaces (e.g. mangroves and wetlands), particularly in tropical regions, with persistent cloud cover and areas with forest or areas deforested by fires can greatly affect the performance of optical imagery (Mahdianpari et al., 2017b; McNairn et al., 2009; Mohammadimanesh et al., 2018a). Consequently, the integration of optical and SAR data can improve classification accuracy and provide more information for detecting changes to the Earth's surface (Mohammadimanesh et al., 2019a). Thus, the integration of optical datasets (Landsat, MODIS, Sentinel-2) with SAR data from Sentinel-1 have been used in 21 articles included in this review, particularly with application to wetland and forest monitoring (B. Chen et al., 2017).

SAR sensors can acquire data regardless of weather and illumination conditions (Mahdianpari et al., 2019b). Moreover, the combination of wavelength and polarization provides more useful information about the Earth's surface (Mahdianpari et al., 2019a; McNairn et al., 2004; Mohammadimanesh et al., 2018d). As shown in Fig. 12, SAR data had a higher classification accuracy than studies that used optical and optical/SAR data. However, only nine studies reviewed here applied SAR imagery, which may be because speckle noise reduces the effective spatial resolution of these images (Choi and Jeong, 2019). Also, SAR image processing and interpretation is difficult for non-remote sensing experts. SAR signals are sensitive to vegetation geometric structure (e.g., leaves and stalks), dielectric properties of the canopies, and soil characteristics yet only two studies reviewed here used SAR imagery to map crops and forests. This is likely because several researchers have emphasized the performance of L-band signals for crop classification (Lee et al., 2001; Lee and Pottier, 2009; Silva et al., 2009), and L-band SAR is not available within the GEE archive. Eight of the studies reviewed here demonstrated that SAR imagery is suited for rice mapping and can provide multi-temporal images throughout the growing season and eliminate cloud cover in rainy environments. Multiple studies have also focused on the use of C-band SAR imagery for shoreline detection and flood monitoring (Uddin et al., 2019; Wong et al., 2019). In addition, Kim et al. (2007) reported that SAR imagery with shorter wavelengths (e.g. C- or X-band) produce more accurate results in extracting shorelines than longer wavelength (e.g., L- or P-band) (Hagenaars et al., 2018; Kim et al., 2007; Marghanay et al., 2011).

4.3. GEE and sensor type

In this review paper, Landsat was the prevailing sensor used in the reviewed papers (see Figs. 8, 9, and 13). Landsat is considered an important remote sensing data source since it provides continuous images of the Earth's surface (Wulder et al., 2019). The Landsat 9 satellite will be launched in 2020 with the purpose of pursuing the Landsat program's critical role in monitoring the Earth's resources. Having access to the more than 40-year record of Landsat data in GEE allows long-term land cover change research to be conducted at both regional and global scales (Liu et al., 2020). The studies reviewed in this paper demonstrate that Landsat is useful for a broad range of applications, including vegetation, forest, crop, land cover/land use changes, fire,

urban, lake, river, water surface, wetland, and climate change monitoring.

Sentinel-1 is another popular source of satellite imagery in the studies reviewed here ($n = 33$) and achieved very accurate classification results. Sentinel-1 consists of a constellation of two satellites, Sentinel-1A and Sentinel 1-B launched in 2014 and 2016, with a spatial resolution of 10 m and revisit time of 6 days (Torres et al., 2012). It is equipped with a dual-polarization C-band SAR sensor which provides data in all-weather, day and night conditions. Sentinel-1 data was used in 33 studies (Fig. 8) across many topics and disciplines, including marine monitoring, shoreline detection, and mapping land cover, climate change, rice fields, and disasters such as flood monitoring.

The Sentinel-2 mission comprises a constellation of two satellites, Sentinel-2A (launched in 2015) and Sentinel-2B (launched in 2017), which provide optical imagery with a spatial resolution of 10 m, 20 m, and 60 m, and an approximate temporal resolution of 5 days (Drusch et al., 2012; Mahdianpari et al., 2019c). This review identified 44 studies conducted using this medium-resolution free imagery in applications related to agricultural, such as crop and vegetation monitoring, coastal zones observations, and land cover classification.

The Moderate Resolution Imaging Spectroradiometer (MODIS) was launched in 1999 on board the Terra satellite, and in 2002 on board the Aqua satellite (Engel-Cox et al., 2004). Researchers have access to MODIS data in GEE in 36 spectral bands and three varying spatial resolutions (250 m, 500 m, and 1 km) with 1-day revisit time. MODIS imagery was applied in 55 of the studies reviewed here. Even though the spatial resolution of MODIS is low, its high temporal resolution allows researchers to monitor short and long-term global environmental changes (dynamics). Examples of such applications using the GEE platform are snow cover, coastline, flood, fire, climate change, vegetation time-series, and evapotranspiration monitoring.

Several studies have emphasized the importance of multi-temporal satellite imagery in remote sensing applications (Clement et al., 2018; McNairn and Shang, 2016; Tomer et al., 2015). For instance, mapping crops with multi-temporal images offers researchers the opportunity to efficiently trace the phenology of crops throughout the growing season. Therefore, time-series data provides more information about the study region and presents a more accurate classification and long-term analysis. GEE multi-temporal satellite imagery was employed in 96% ($n = 163$) of the reported studies and produced more consistent overall accuracies compared to single-date data, which was utilized in only 4% ($n = 7$) of the articles reviewed. Finally, pixel-based analyses were used in 98% of studies while only 2% of papers used object-based approach for their analyses. This is because users do not have access to object-based analysis functions in the GEE platform (Kumar and Mutanga, 2018).

4.4. Remote sensing data analysis

4.4.1. Machine learning techniques

Machine learning, a subset of artificial intelligence, deals with the design of algorithms to train models to make decisions or predictions (Huang and Jensen, 1997). Machine learning methods can be divided into two major groups: parametric and non-parametric (Holloway and Mengersen, 2018). Parametric machine learning algorithms use a fixed number of parameters or assumptions; however, they are independent of the number of training samples. Although parametric algorithms are faster, assumptions about the data can greatly affect and limit the learning process. On the other hand, non-parametric algorithms use a flexible number of parameters. Non-parametric methods can become slower and more complex with increasing amounts of data; however,

they make fewer assumptions about the distribution of the data. Machine learning methods have been effectively adopted for remote sensing data processing (Schulz et al., 2018). Classification, clustering, regression, and dimension reduction are four main analytic categories of machine learning algorithms (Holloway and Mengersen, 2018).

4.4.1.1. Classification. DT, CART, KNN, non-linear SVM, RF, and ANN are non-parametric algorithms, since the number of parameters grows with the size of the training set. In this review, 186 studies used non-parametric models because they are more flexible, and they can process a large amount of data with no prior knowledge. Furthermore, due to the inherent characteristics of remote sensing datasets, they may not be normally distributed and non-parametric algorithms are assumption-free models (Holloway and Mengersen, 2018). In the GEE studies reviewed, researchers preferred to utilize non-parametric algorithms. This provides particular advantage when combining noisy data from multiple sources that may have differing or unknown distributions.

Among classification methods, the DT classifier had the best overall accuracies with 50% of results over 94% (see Fig. 11). DT works based on hierarchical associations between input variables and provides a set of rules that are easy to interpret. It also does not require an extensive design or training and it is computationally efficient. Some facts may restrict its use: it uses hyperplane classification boundaries parallel to the feature axes, which causes challenges in the case of mixed pixels. It also becomes complex when various values in the output are correlated. Therefore, this supervised algorithm has only been utilized in 12 studies reviewed here.

The CART algorithm was the second most popular classification method. CART has become more popular due to its efficiency and simplicity in solving a wide range of problems not only in remote sensing applications but also in engineering, agriculture, and other fields. (Steinberg and Colla, 2009). A number of studies reported that CART performed well for their applications (Bittencourt and Clarke, 2003). One of the strongest advantages of using CART is that it divides a complex problem into simpler sub-problems (Bittencourt and Clarke, 2003). As such, 26 studies exploited CART classifier in this review.

The KNN classification requires few assumptions and parameters to tune, it is easy to implement, and robust regarding the search space, and is therefore suitable for non-linear separable datasets (Li and Cheng, 2009). Moreover, it can be adopted for classification, regression and search applications. The main disadvantages of this algorithm are that it is sensitive to noise and unbalanced data, which results in less meaningful distance numbers (Blanzieri and Melgani, 2008).

In this investigation, 21 studies utilized the SVM classifier. Although the median overall accuracy for SVM is about 89%, the high variance of results can be explained by the benefits and limitations of this method. There are three important characteristics of SVM that are discussed here. First, the SVM approach maps the input data into a higher feature space which increases the separability pattern among the data. Second, using a convex cost function and dealing with quadratic problems allows SVM to find the global minimum and consecutively obtain the optimal solution. Finally, in the case of remote sensing data and ground reference issues, SVM works well with a limited amount of training data (Mountrakis et al., 2011). Being sensitive to feature selection and tuning several parameters such as kernel function and width are the most challenging aspects of SVMs. Moreover, mislabeled pixels or outliers can greatly affect the performance of the classification, because SVMs cannot deal with noisy data (Holloway and Mengersen, 2018; Mountrakis et al., 2011).

RF is the most commonly used classification method in the GEE platform with 97 studies reviewed here (Hu et al., 2018; Teluguntla

et al., 2018; Xie et al., 2019). There are several advantages that likely led to this level of popularity. The RF algorithm is robust, easy to train, less sensitive to the quality of training data, and there are fewer parameters to tune in comparison with other non-parametric classifiers (Belgiu and Drăguț, 2016; Mahdianpari et al., 2017c; Mohammadi manesh et al., 2018b). Even though complexity will increase with the number of trees and training data, this approach can significantly increase the classification accuracy (Mahdianpari et al., 2018a). Belgiu and Drăguț (2016) reported that the RF approach increases classification accuracy, especially for high dimensional input data such as hyperspectral imagery. In some cases, low quantity training data cause misclassification, which explains the large variance of the RF boxplot in Fig. 11.

ANN is a data driven, self-adaptive technique that can manage noisy data efficiently. In this survey, 6 studies used the ANN classifier with the lowest median of overall accuracy (see Fig. 11). A minimum number of studies used ANN because it is not supported in the GEE built-in functions and it suffers from high computation rates, time-consuming training, difficulties in choosing the type of network architecture (Bischof et al., 1992), and problems with local minimum in training (Benediktsson et al., 1993). Since GEE does not include any neural network-based models, users train and run neural network and even deep neural network (DNN) algorithms in Google Colab with data from GEE.

GMM, MD, and Naïve Bayes are some examples of parametric machine learning methods that several articles reported applying through the GEE code editor. The results of these algorithms are highly dependent on the initial assumptions. On the condition that the assumptions are correct, the algorithm may work well, but the reverse is also true. Naïve Bayes is highly scalable, and it requires less training data. Thus, Naïve Bayes performed better among parametric methods with a higher median of overall accuracy (Voight et al., 2019).

Clustering is an unsupervised learning method that tries to combine objects into clusters based on similarity criteria of input variables without training data (Holloway and Mengersen 2018). Although clustering algorithms do not require training data for implementation, absence of ground truth data for evaluating the results may oblige scientists to use classification methods instead. Due to issues with clustering assessment, in most unsupervised algorithms the number of clusters should be identified before implementation. Therefore, only two studies reviewed here used K-means unsupervised classification for remote sensing data analysis. Because it is fast and has few computations, K-means is one of the most well-known clustering algorithms. However, it also starts with initial random cluster centers and therefore it may yield different clustering results on different runs of the algorithm. Thus, the results may not be repeatable and lack consistency (Chen et al., 2005).

4.4.1.2. Regression.

Regression is a supervised machine learning method that aims to estimate or predict output variables based on a set of covariates (Holloway and Mengersen, 2018). Like classification, regression trains the model based on input variables, but the output variables are numerical (continuous). Linear models such as linear regression and logistic regression are common examples of parametric methods. While linear regressions are not considered as a machine learning technique, here they have been categorized under the regression section for consistency.

The results of this review found 55 of 63 regression-focused studies utilized linear regression and its extensions, such as linear least square and linear multiple regression (MLR) methods. Linear regression was used in 48 papers because it is easy to implement, particularly for

studies covering a large area. Another positive aspect of linear regression is its fast-computational speed, which is an important factor in geobig data analysis. Forkuor et al. (2017) reported that MLR models suffer from managing non-linear relationships between dependent and independent variables during prediction (Forkuor et al., 2017). As a result, only five studies reviewed here utilized the MLR method. Eight studies utilized random forest regression (RFR) in GEE. The running time and the complexity of the algorithm are the main disadvantages of RFR algorithms (Forkuor et al., 2017). Although RFR is robust against non-linearity and it requires no assumption regarding probability distribution of the target predictors, it requires tuning of some parameters, such as the number of trees and randomly selected predictors.

4.4.2. Other GEE image processing functions

GEE hosts several built-in image processing functions that users can apply for remote sensing data analysis. In this regard, time-series analysis, feature extraction, image color composite- visual interpretation, and image pre-processing techniques were applied to satellite imagery instead of machine learning methods.

The Earth's surface is changing at an unprecedented rate, so time-series analysis of satellite imagery is of paramount importance for following trends, changes, and detecting patterns to develop models and predict changes (de Oliveira et al., 2016). GEE was used in 19 studies for time-series analysis since it supports large volumes of data for applications working with high resolution imagery or at a global scale. Moreover, GEE provides time-series charts that allow users to follow and evaluate changes to Earth's surface features. For time-series analysis a sequence of images with mostly uniform time intervals is needed. With the comprehensive archive, e.g. of Landsat imagery, GEE is a valuable imagery source to monitor changes to the Earth's surface. This survey concluded that most of the studies in shoreline detection, and coastal and water resource monitoring have used time-series analysis (Fang et al., 2019; Nguyen et al., 2019; Wang et al., 2018). In addition, some research has been conducted to track deforested areas, land-use changes, and the role of temperature in climate change (Hu et al., 2018; Jamei et al., 2019; Workie and Debella, 2018).

Several studies utilized feature extraction and band ratio techniques in different domains, including vegetation, drought, and land cover change (Ravanelli et al., 2018; Rembold et al., 2019; Saah et al., 2019; Scherler et al., 2018; Sidhu et al., 2018; Waller et al., 2018; Weissmann et al., 2017). Feature extraction is an image processing technique that uses image spectral and geometrical characteristics such as color, shape, and context to identify mutual relationships between image regions (Momm and Easson, 2011). The process of feature extraction is important because it can reduce the resources needed for processing without losing important information.

As previously stated, visual interpretation of satellite imagery is the type of analysis for both experts and non-remote sensing experts in studies that require less accuracy or apply to small areas. Eight papers with applications in land use/land cover monitoring used this method. In this case, remote sensing images are displayed in a color composite to extract meaningful information from the imagery (Sader and Winne, 1992). Accurate results from visual interpretation require an interpreter with specialized knowledge of the subject, the geographic region under study and training in the application of remote sensing.

Thanks to the GEE image processing tools and remote sensing images with various resolutions, users have leveraged GEE for image pre-processing techniques such as image mosaicking, cloud, and error detection. The GEE platform provides original raw images, but the primary challenge that needs to be faced in applying optical data is often cloud cover (Mateo-García et al., 2017). Users have access to

ancillary layers (e.g., Landsat Collection 1 Level-1 Quality Assessment Band) or algorithms (e.g., Simple Cloud Score and F-Mask) that can be used for cloud detection. Therefore, GEE has presented a baseline for the development of image pre-processing algorithms as well as environmental studies. As a result, remote sensing data catalogue and built-in functions can be considered as a powerful tool for enhancing image pre-processing techniques.

4.5. GEE and ready-to-use products

Based on this review, remote sensing data is more popular for developers and scientists with most papers ($n = 312$ articles) published in remote sensing related journals using imagery over ready-to-use products ($n = 37$ articles). Researchers mostly utilized vegetation indices for their analysis, of which the Normalized Difference Vegetation Index (NDVI) was predominant (Kong et al., 2019). As the name suggests, NDVI is calculated as a normalized of difference of near-infrared (which vegetation strongly reflects) and red (which vegetation absorbs) wavelengths (Karnieli et al., 2010). Thus, NDVI is very useful for environmental monitoring and particularly in crop, vegetation, and forest mapping (Xiong et al., 2017). NDVI is a good way to investigate healthy vegetation (chlorophyll), vegetation changes over time, irrigated vegetation, biomass estimation, forest supply, and leaf area index (Xie et al., 2019). Moreover, climate change monitoring studies leveraged the power of several products such as NDVI, Land Surface Temperature (LST), land cover type, and precipitation. NDVI and LST are indicators of drought as water deficiency and temperature changes limit the growth of vegetation which can be used in climate change investigations (Karnieli et al., 2010). Precipitation information available in GEE also allows scientists to focus on water supplies and soil moisture. Studies that focused on habitat and disease monitoring used land cover, NDVI, precipitation, and the Hansen et al. (2013) global forest change indices. GEE presents the Global Forest Change dataset based on time-series analysis of Landsat images between 2000 and 2014, which was used by Hansen et al. (2013) to produce forest extent and change at global scale. Furthermore, access to Shuttle Radar Topography Mission DEMs in 30 m resolution has enabled researchers to use this data for topographic modelling across wide extents. In summary, thanks to the ready-to-use products available in GEE, various applications in regional and global scale have been conducted.

4.6. Future work

GEE has become one of the most popular platforms for geospatial and big data analysis; nevertheless, there are still areas for expansions. In particular, GEE has provided an easy to use and free platform for geospatial processing, yet users do not have control over the details of parallel processing environment which can be source of some computational issues. This means that GEE manages every aspect of computation, such as source allocation, parallelism, data distribution, and retries, while users are unable to influence Do-It-Yourself (DIY) parallelization (Gorelick et al., 2017). In general, GEE limitations can be classified into three main categories: computation, dataset, and algorithms.

Notably, GEE has some computational limitations, including time, memory, and storage. In respect to time limit issue, there are two computation modes in GEE: on-demand and batch processing. The former deals with limited number of runs, while the latter may run as long as the code is running. Hence, it is reasonable to use the batch process for huge computations, since tasks are run in on-demand mode.

Moreover, in some cases, GEE might run into memory issues when processing is performed on a huge amount of datasets. Although scripts may be valid JavaScript without logical errors, sometimes users face internal errors in parallelizing and executing the computations, such as internal service error, computation timed out, user memory limit exceeded, and too many concurrent aggregations. These errors, known as scaling errors, might occur when outputs are too big, numerous, or take long to compute. Regarding storage, users can save their results in Google Drive, Google Cloud, and GEE assets. Nevertheless, the restricted amount of 250 GB capacity for saving data in the GEE asset should be taken into account. There are also several limitations on the size and shape of GEE table assets.

Although GEE includes an extensive archive of imagery, but for many studies, historical and high-resolution data are of limited value. Specifically, 13% of GEE studies focused on disaster mapping and particularly in drought monitoring. GEE can also provide reliable information for monitoring other disasters, such as earthquake and flood mapping. Moreover, GEE currently provides high resolution imagery, including National Agriculture Imagery Program (NAIP) and Planet Skysat in both RGB and multi-spectral collection. The US NAIP offers aerial image data in 1 m resolution. These images are available in GEE since 2003–2018 with 3- and 5-year cycles. GEE also provides Planet Skysat imagery in RGB and multispectral/Pan collection from 2014 to 2016. The RGB imagery is offered in 0.8 m resolution (1 m off-nadir images), while R, G, B, and Near-IR bands have approximately 2 m resolution and Pan band has 0.8 resolution (1 m for off-nadir). Thus, the need for high spatial and temporal resolution imagery is necessary.

Sentinel-1 imagery has been the only available SAR data within GEE so far; however, new avenues have opened by adding L-band data collected from ALOS PALSAR for several applications, demanding longer wavelengths, such as crop mapping. Compared to C-band, which has a moderate penetration capability and mainly interacts with upper section of canopies, L-band has a deeper penetration capability and could interact with stems and branches, thus making it advantageous for several applications, for example.

Implementation of new algorithm could be challenging in GEE. Over the past few years, deep learning methods have drawn attention in the remote sensing community for several applications, given their superiority compared to conventional machine learning tools (Mahdianpari et al., 2018b; Mohammadimanesh et al., 2019b; Nogueira et al., 2017; Rezaee et al., 2018; Sun and Wang, 2018; Zhang et al., 2018; Zhu et al., 2017). In particular, deep neural networks have been extensively used in image classification tasks and they showed promising results in terms of classification accuracy (Maggiori et al., 2016). However, deep learning algorithms are not yet supported directly by GEE. When it comes to deep learning, the choice of an open-source framework is of paramount importance. TensorFlow is the most popular framework among open-source deep learning frameworks, such as Caffe, Microsoft CNTK, MXNet, Facebook Torch, Deeplearning4j, and Theano, (Zhu et al., 2017). Although deep learning classifiers are not yet directly supported, GEE has recently been linked to TensorFlow (since September 2019). In particular, users now have access to packages, allowing them to interact with the TensorFlow's saved model format hosted on the Google AI platform (DeLancey et al., 2020). GEE API provides an opportunity to import/export imagery, training, and testing datasets in TFRecord format. TFRecord format can handle huge volumes of data and it allows users to run the classifier as a batch processing without the need to store all the data. Thus, this capability undoubtedly manages the challenges in analyzing big data.

According to the best performance of GEE in pixel-based

classification, as the result shows most researches have been done using pixel-based approach. However, this would give rise to an issue of finding high quality reference data over large areas. Moreover, GEE could be more persuasive if complex unsupervised classification algorithms have been implemented successfully. For instance, at present only K-means, X-means, LVQ, and Cobweb are supported in GEE code editor. Due to the inherent characteristics of remote sensing imagery such as non-normal distributed data and existence of mixed pixels, implementing complex algorithms such as ISODATA, fuzzy K-means, Probabilistic K-means, and kernel-based clustering methods is of paramount significance. Kernel-based clustering algorithms can transform input data into a higher dimensional space where clusters are more distinguishable.

Overall, GEE demonstrated to be better suited for image analysis compared to vector-based processing. Furthermore, analysis based on pixel spatial relations of pixels are harder to be completed, as the processing is carried out using multiple CPUs. Moreover, image segmentation and hydrologic modeling options are limited. Accordingly, improvements to object-based image analysis, heavy duty vector operations, and providing high resolution satellite imagery for recent years are recommended.

5. Conclusions

This meta-analysis confirms the use of GEE has become more widespread across the world for geospatial data analyses in different domains. According to the review of 349 peer-reviewed articles, GEE has played a vital role in conducting geo-big data analysis for a variety of remote sensing applications across the world over the last seven years. A database, including typical fields related to geo-big data processing, was created using information extracted from this systematic review. Consequently, several conclusions were drawn:

- (1) The GEE cloud computing platform, with petabyte of satellite imagery, massive computational capabilities, high-level application programming interface (API), and available machine learning algorithms, enables researchers to process huge volume of geo-big data in various remote sensing applications. In particular, crop mapping (22%) followed by water resource applications (20%) and LULC mapping (16) were the top-three remote sensing studies, which have used the GEE capabilities for their analysis.
- (2) The use of GEE for various remote sensing applications has significantly increased since 2017. Specifically, 13%, 25%, and 29% of 349 reviewed papers have been published in 2017, 2018, and 2019, respectively. This growing rate in the application of GEE clearly highlights its popularity in the remote sensing community.
- (3) Although most GEE studies have been conducted in the United States (77 studies, 22%), this cloud computing platform has offered new opportunities for accessing to and processing of geo-big data across the world, even in the developing countries. As such, Asia has most GEE users among remote sensing community featuring 119 studies.
- (4) GEE can be considered as a multi-disciplinary tool, as the reviewed articles in this paper have been published in 146 different peer-reviewed journals in a variety of fields.

Appendix A

See [Table A1](#).

- (5) Satellite and aerial imagery products were used in 90% of articles, whereas ready-to-use products were only employed in 11% of studies. Vegetation indices, land cover, Hansen global forest change, and DEM were the most ready-to-use products.
- (6) From data point perspective, optical data (90%) and, in particular, Landsat imagery (82%) were the most commonly used datasets. Notably, a combination of medium spatial resolution (Landsat) and high temporal resolution (MODIS) data (6%) was the most popular integration between satellite imagery. Due to the predominance of medium resolution imagery available in the GEE data catalogue, most studies used data with spatial resolutions between 5 m and 30 m (71%).
- (7) Machine learning methods have been extensively used for remote sensing data processing in GEE. In particular, RF (49%) was the most popular classifier followed by CART (13%) and SVM (11%).
- (8) Among different spectral, contextual, and textural features, NDVI was the most frequently extracted feature for image classification, as it represents vegetation vigor and is an indicator of drought and water deficiency. Thus, it has been often used in studies related to vegetation and crop mapping, as well as climate change related research.

GEE offers a novel platform for conducting geo-big data analyses in various environmental applications. While there is always room for further development of the platform, the emergence of GEE with its unique computational capability, free availability of satellite imagery, and scripting tools have increased researcher's ability to perform geospatial analysis across a diverse areas. Thanks to these unique features, GEE has been superior compared to other competitor platforms, yet a few of its weaknesses should be alleviated. For example, a limited number of algorithms on object-based image analysis and clustering methods are currently available in GEE. Notably, the implementation of advanced segmentation and clustering algorithms significantly contribute to delineation of accurate ground truth data. This is of particular importance for successful application of machine learning tools for large scale land cover mapping. In this paper, different GEE applications were compared in a joint analysis to clarify GEE potentials in geo-big data management. Accordingly, an investigation of GEE articles with a narrow subject (e.g., each application and study area) could provide a detailed understanding in different research domains.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors would like to thank Dr. Fariba Mohammadimanesh and Dr. Colin M. Beier who greatly improved the quality of the manuscript with their invaluable suggestions during the study. The authors also appreciate the time, effort, and insight offered by three anonymous reviewers and members of the editorial team.

Table A1
A summary of the main satellite imagery available in Google Earth Engine.

Sensor	Datasets	Characteristics	Spatial Resolution (m)	Temporal Resolution (Day)	Data availability in GEE	Dataset Provider
Sentinel-1	Sentinel-1 SAR GRD	C-band Synthetic Aperture Radar Ground Range Detected, log scaling	10	6	2014-10-3_present	European Union/ESA/Copernicus
Sentinel-2 MSI: Multispectral Instrument	Surface Reflectance	Level-2A orthorectified atmospherically corrected surface reflectance.	10, 20, 60	5	2017-03-28,present	European Union/ESA/Copernicus
Sentinel-3	Top-of-Atmosphere Reflectance	Level-1C orthorectified top-of-atmosphere reflectance.	10, 20, 60	5	2015-06-23,present	European Union/ESA/Copernicus
Sentinel-3 OLCI EFR Resolution	OLCI EFR: Ocean and Land Color Instrument Earth Observation Full Resolution	300	2	2016-10-18,present	European Union/ESA/Copernicus	
Sentinel-5P TROPOMI	Sentinel-5P UV Aerosol Index	A measure of the prevalence of aerosols in the atmosphere	0.01 arc degrees	1	2018-07-04,present	European Union/ESA/Copernicus
Sentinel-5P Cloud	Cloud characteristics including fraction, height and pressure for base and top, optical depth, and surface albedo.	0.01 arc degrees	1	2018-07-04,present	European Union/ESA/Copernicus	
Sentinel-5P Carbon Monoxide Concentrations of Carbon monoxide (CO) and water vapor.	0.01 arc degrees	1	2018-06-28,present	European Union/ESA/Copernicus		
Sentinel-5P Formaldehyde Formaldehyde concentration.	0.01 arc degrees	1	2018-10-02,present	European Union/ESA/Copernicus		
Sentinel-5P Nitrogen Dioxide Total, tropospheric, and stratospheric nitrogen dioxide concentration.	0.01 arc degrees	1	2018-06-28,present	European Union/ESA/Copernicus		
Sentinel-5P Ozone Total atmospheric column ozone concentration.	0.01 arc degrees	1	2018-07-10,present	European Union/ESA/Copernicus		
Sentinel-5P Sulphur Dioxide Atmospheric sulphur dioxide (SO ₂) concentration.	0.01 arc degrees	1	2018-07-10,present	European Union/ESA/Copernicus		
Sentinel-5P Methane Atmospheric methane (CH ₄) concentration.	0.01 arc degrees	1	2019-02-08,present	European Union/ESA/Copernicus		
Landsat-1 MMS	Landsat-1 MSS Collection 1 Tier 1 and Tier 2 Raw Scenes	Landsat-1 MSS Collection 1 DN values, representing scaled, calibrated at-sensor radiance.	30, 60	16	1972-07-23_1978-01-07	USGS
Landsat-2 MSS	Landsat-2 MSS Collection 1 Tier 1 and Tier 2 Raw Scenes	Landsat-2 MSS Collection 1 DN values, representing scaled, calibrated at-sensor radiance.	30, 60	16	1975-01-22_1982-02-26	USGS
Landsat-3 MSS	Landsat-3 MSS Collection 1 Tier 1 and Tier 2 Raw Scenes	Landsat-3 MSS Collection 1 DN values, representing scaled, calibrated at-sensor radiance.	30, 60	16	1978-03-05_1983-03-31	USGS
Landsat-4	Landsat 4 TM/MSS Collection 1 Tier 1 and Tier 2 Raw Scenes	Landsat 4 TM and Landsat 4 MSS Collection 1 DN values, representing scaled calibrated at-sensor radiance.	30, 60	16	1982-08-1993-12	USGS
	Landsat 4 TM Collection 1 Tier 1 TOA Reflectance	Landsat 4 TM Collection 1 calibrated top-of-atmosphere (TOA) reflectance.	30, 60	16	1982-08_1993-12	USGS
	USGS Landsat 4 Surface Reflectance Tier 1	Atmospherically corrected surface reflectance from the Landsat 4 ETM sensor.	30, 60	16	1982-08-1993-12	USGS
Landsat-5	Landsat 5 TM/MSS Collection 1 Tier 1 and Tier 2 Raw Scenes	Landsat 5 TM and Landsat 5 MSS Collection 1 DN values, representing scaled calibrated at-sensor radiance.	30, 60	16	1984-03_2012-05	USGS
	Landsat 5 TM Collection 1 Tier 1 and Tier 2 TOA Reflectance	Landsat 5 TM Collection 1 calibrated top-of-atmosphere (TOA) reflectance.	30, 60	16	1984-03_2012-05	USGS
	Landsat 5 Surface Reflectance Tier 1 and Tier 2	Atmospherically corrected surface reflectance from the Landsat 5 ETM sensor.	30, 60	16	1999-01_present	USGS
	Landsat 7 Collection 1 Tier 1 and Tier 2/Tier 1 + Real-time Raw Scenes	Landsat 7 Collection 1 images, representing scaled, calibrated at-sensor radiance.	30, 60	16	1999-01_present	USGS
	Landsat 7 ETM + Collection 1 Tier 1 and Tier 2/Tier 1 + Real-time TOA Reflectance	Landsat 7 ETM + Collection 1 calibrated top-of-atmosphere (TOA) reflectance.	30, 60	16	1999-01_present	USGS
	Landsat 7 ETM + Surface Reflectance Tier 1 and Tier 2	Atmospherically corrected surface reflectance from the Landsat 7 ETM + sensor.	30, 60	16	1999-01_present	USGS

(continued on next page)

Table A1 (continued)

Sensor	Datasets	Characteristics	Spatial Resolution (m)	Temporal Resolution (Day)	Data availability in GEE	Dataset Provider
Landsat-8	Landsat 8 Collection 1 Tier 1 and Tier 2/Tier 1 + Real-time Raw Scenes	Landsat 8 Collection 1 DN values, representing scaled, calibrated at-sensor radiance.	30, 60	16	2013-04-present	USGS
	Landsat 8 Collection 1 Tier 1 and Tier 2/Tier 1 + Real-time TOA Reflectance	Landsat 8 Collection 1 calibrated top-of-atmosphere (TOA) reflectance	30, 60	16	2013-04-present	USGS
	Landsat 8 Surface Reflectance Tier 1 and Tier 2	Atmospherically corrected surface reflectance from the Landsat 8 OLI/TIRS sensors	30, 60	16	2013-04-present	USGS
	MCD43A4.006 MODIS Nadir BRDF	Adjusted Reflectance, Daily 500 m	500	1	2000-02-18-Present	NASA LP DAAC at the USGS EROS Center
	MCD43A3.006	MODIS Albedo Daily 500 m	500	1	2000-02-18-Present	NASA LP DAAC at the USGS EROS Center
	MCD43A2.006 MODIS BRDF	Albedo Quality Daily 500 m	500	1	2000-02-18-Present	NASA LP DAAC at the USGS EROS Center
	MOD09GQ.006	Terra Surface Reflectance Daily Global 250 m	250	1	2000-02-24-Present	NASA LP DAAC at the USGS EROS Center
	MOD10A1.006	Terra Snow Cover Daily Global 500 m	500	1	2000-02-24-Present	NSIDC
	MOD11A1.006	Terra Land Surface Temperature and Emissivity Daily Global 1 km	1000	1	2000-03-05-Present	NASA LP DAAC at the USGS EROS Center
	MOD09GA.006	Terra Surface Reflectance Daily I2G Global	500, 1000	1	2000-02-24-Present	NASA LP DAAC at the USGS EROS Center
MODIS	MODOCGA.006	Terra Ocean Reflectance Daily Global	1000	1	2000-02-24-Present	NASA LP DAAC at the USGS EROS Center
	MOD14A1.006	Terra Thermal Anomalies & Fire Daily Global	1000	1	2000-02-18-Present	NASA LP DAAC at the USGS EROS Center
	MOD43A1.006	MODIS BRDF-Albedo Model Parameters Daily	500	1	2000-02-18-Present	NASA LP DAAC at the USGS EROS Center
	High-resolution Imagery	Planet SkySat Public Ortho Imagery, Multispectral	This data from Planet labs Inc. SkySat satellites was collected for the experimental “Skybox for Good Beta” program in 2015, as well as for various crisis response events and a few other projects. The data is available in both a 5-band Multispectral/Pan collection, and a Pansharpened RGB collection.	0.8	N/A	2014-07-03_2016-12-24
	168		This data from Planet labs Inc. SkySat satellites was collected for the experimental “Skybox for Good Beta” program in 2015, as well as for various crisis response events and a few other projects. The data is available in both a 5-band Multispectral/Pan collection, and a Pansharpened RGB collection.	N/A	2014-07-03_2016-12-24	Planet Labs Inc.
	NAIP: National Agriculture Imagery Program	The National Agriculture Imagery Program (NAIP) acquires aerial imagery during the agricultural growing seasons in the continental U.S. NAIP projects are contracted each year based upon available funding and the FSA imagery acquisition cycle. Beginning in 2003, NAIP was acquired on a 5-year cycle. 2008 was a transition year, and a three-year cycle began in 2009.	1	N/A	2003-01-01_2018-01-01	USDA Farm Service Agency

References

- Amazon, E.C., 2015. Amazon web services. Available <Httpaws Amaz. Comesec2November 2012>.
- Belgiu, M., Drăguț, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31.
- Benediktsson, J.A., Swain, P.H., Ersoy, O.K., 1993. Conjugate-gradient neural networks in classification of multisource and very-high-dimensional remote sensing data. *Int. J. Remote Sens.* 14, 2883–2903.
- Bischof, H., Schneider, W., Pinz, A.J., 1992. Multispectral classification of Landsat-images using neural networks. *IEEE Trans. Geosci. Remote Sens.* 30, 482–490.
- Bittencourt, H.R., Clarke, R.T., 2003. Use of classification and regression trees (CART) to classify remotely-sensed digital images. In: IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No. 03CH37477). IEEE, pp. 3751–3753.
- Blanzieri, E., Melgani, F., 2008. Nearest neighbor classification of remote sensing images with the maximal margin principle. *IEEE Trans. Geosci. Remote Sens.* 46, 1804–1811.
- Carrasco-Escobar, G., Manrique, E., Ruiz-Cabrejos, J., Saavedra, M., Alava, F., Bickersmith, S., Prussing, C., Vinetz, J.M., Conn, J.E., Moreno, M., 2019. High-accuracy detection of malaria vector larval habitats using drone-based multispectral imagery. *PLoS Negl. Trop. Dis.* 13, e0007105.
- Castelli, G., Oliveira, L.A.A., Abdelli, F., Dhaou, H., Bresci, E., Ouessa, M., 2019. Effect of traditional check dams (jessour) on soil and olive trees water status in Tunisia. *Sci. Total Environ.* 690, 226–236.
- Callaghan, C.T., Major, R.E., Lyons, M.B., Martin, J.M., Kingsford, R.T., 2018. The effects of local and landscape habitat attributes on bird diversity in urban greenspaces. *Ecosphere* 9, e02347.
- Chen, B., Jin, Y., Brown, P., 2019. Automatic mapping of planting year for tree crops with Landsat satellite time series stacks. *ISPRS J. Photogramm. Remote Sens.* 151, 176–188. <https://doi.org/10.1016/j.isprsjprs.2019.03.012>.
- Chen, B., Tai, P.C., Harrison, R., Pan, Y., 2005. Novel hybrid hierarchical-K-means clustering method (HK-means) for microarray analysis. In: 2005 IEEE Computational Systems Bioinformatics Conference-Workshops (CSBW'05). IEEE, pp. 105–108.
- Chen, B., Xiao, X., Li, X., Pan, L., Doughty, R., Ma, J., Dong, J., Qin, Y., Zhao, B., Wu, Z., Sun, R., Lan, G., Xie, G., Clinton, N., Giri, C., 2017a. A mangrove forest map of China in 2015: Analysis of time series Landsat 7/8 and Sentinel-1A imagery in Google Earth Engine cloud computing platform. *ISPRS J. Photogramm. Remote Sens.* 131, 104–120. <https://doi.org/10.1016/j.isprsjprs.2017.07.011>.
- Chen, F., Zhang, M., Tian, B., Li, Z., 2017b. Extraction of glacial lake outlines in Tibet Plateau using Landsat 8 imagery and Google Earth Engine. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 10, 4002–4009.
- Chi, M., Plaza, A., Benediktsson, J.A., Sun, Z., Shen, J., Zhu, Y., 2016. Big data for remote sensing: Challenges and opportunities. *Proc. IEEE* 104, 2207–2219.
- Choi, H., Jeong, J., 2019. Speckle Noise Reduction Technique for SAR Images Using Statistical Characteristics of Speckle Noise and Discrete Wavelet Transform. *Remote Sens.* 11, 1184.
- Clement, M.A., Kilsby, C.G., Moore, P., 2018. Multi-temporal synthetic aperture radar flood mapping using change detection. *J. Flood Risk Manag.* 11, 152–168.
- Chrysoulakis, N., Mitraka, Z., Gorelick, N., 2019. Exploiting satellite observations for global surface albedo trends monitoring. *Theor. Appl. Climatol.* 137, 1171–1179.
- de Oliveira, S.S.T., de Castro Cardoso, M., dos Santos, W.M., Costa, P.C., Vagner, J., Martins, W.S., 2016. A new platform for time-series analysis of remote sensing images in a distributed computing environment. In: GeolInfo, pp. 128–139.
- DeLancey, E.R., Simms, J.F., Mahdianpari, M., Brisco, B., Mahoney, C., Kariyeva, J., 2020. Comparing Deep Learning and Shallow Learning for Large-Scale Wetland Classification in Alberta, Canada. *Remote Sens.* 12, 2.
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* 120, 25–36.
- Engel-Cox, J.A., Holloman, C.H., Coutant, B.W., Hoff, R.M., 2004. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmos. Environ.* 38, 2495–2509.
- Fang, Y., Li, H., Wan, W., Zhu, S., Wang, Z., Hong, Y., Wang, H., 2019. Assessment of Water Storage Change in China's Lakes and Reservoirs over the Last Three Decades. *Remote Sens.* 11, 1467.
- Forkuor, G., Hounkpatin, O.K., Welk, G., Thiel, M., 2017. High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. *PLoS One* 12, e0170478.
- 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.
- Hagenars, G., de Vries, S., Luijendijk, A.P., de Boer, W.P., Reniers, A.J., 2018. On the accuracy of automated shoreline detection derived from satellite imagery: A case study of the sand motor mega-scale nourishment. *Coast. Eng.* 133, 113–125.
- Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., 2013. High-resolution global maps of 21st-century forest cover change. *Science* 342, 850–853.
- Hird, J.N., DeLancey, E.R., McDermid, G.J., Kariyeva, J., 2017. Google Earth Engine, open-access satellite data, and machine learning in support of large-area probabilistic wetland mapping. *Remote Sens.* 9, 1315.
- Holloway, J., Mengersen, K., 2018. Statistical machine learning methods and remote sensing for sustainable development goals: A review. *Remote Sens.* 10, 1365.
- Hu, Y., Dong, Y., Batunacun, 2018. An automatic approach for land-change detection and land updates based on integrated NDVI timing analysis and the CVAPS method with GEE support. *ISPRS J. Photogramm. Remote Sens.* 146, 347–359. <https://doi.org/10.1016/j.isprsjprs.2018.10.008>.
- Huang, X., Jensen, J.R., 1997. A machine-learning approach to automated knowledge-base building for remote sensing image analysis with GIS data. *Photogramm. Eng. Remote Sens.* 63, 1185–1193.
- Jamei, Y., Rajagopalan, P., Sun, Q.C., 2019. Time-series dataset on land surface temperature, vegetation, built up areas and other climatic factors in top 20 global cities (2000–2018). *Data Brief* 23.
- Karnieli, A., Agam, N., Pinker, R.T., Anderson, M., Imhoff, M.L., Gutman, G.G., Panov, N., Goldberg, A., 2010. Use of NDVI and land surface temperature for drought assessment: Merits and limitations. *J. Clim.* 23, 618–633.
- Kennedy, R.E., Yang, Z., Gorelick, N., Braaten, J., Cavalcante, L., Cohen, W.B., Healey, S., 2018. Implementation of the LandTrendr algorithm on google earth engine. *Remote Sens.* 10, 691.
- Kim, D., Moon, W.M., Park, S.-E., Kim, J.-E., Lee, H.-S., 2007. Dependence of waterline mapping on radar frequency used for SAR images in intertidal areas. *IEEE Geosci. Remote Sens. Lett.* 4, 269–273.
- Kong, D., Zhang, Y., Gu, X., Wang, D., 2019. A robust method for reconstructing global MODIS EVI time series on the Google Earth Engine. *ISPRS J. Photogramm. Remote Sens.* 155, 13–24. <https://doi.org/10.1016/j.isprsjprs.2019.06.014>.
- Koskinen, J., Leinonen, U., Vollrath, A., Ortmann, A., Lindquist, E., d'Annunzio, R., Pekkarinen, A., Käyhkö, N., 2019. Participatory mapping of forest plantations with Open Foris and Google Earth Engine. *ISPRS J. Photogramm. Remote Sens.* 148, 63–74. <https://doi.org/10.1016/j.isprsjprs.2018.12.011>.
- Krishnan, S.P.T., Gonzalez, J.L.U., 2015. Building Your Next Big Thing with Google Cloud Platform: A Guide for Developers and Enterprise Architects. Springer.
- Kumar, L., Mutanga, O., 2018. Google Earth Engine applications since inception: Usage, trends, and potential. *Remote Sens.* 10, 1509.
- Laney, D., 2001. 3D data management: Controlling data volume, velocity and variety. *META Group Res. Note* 6, 1.
- Lee, J.-S., Grunes, M.R., Pottier, E., 2001. Quantitative comparison of classification capability: Fully polarimetric versus dual and single-polarization SAR. *IEEE Trans. Geosci. Remote Sens.* 39, 2343–2351.
- Lee, J.-S., Pottier, E., 2009. Polarimetric radar imaging: from basics to applications. CRC Press.
- Li, S., Dragicevic, S., Castro, F.A., Sester, M., Winter, S., Coltekin, A., Pettit, C., Jiang, B., Haworth, J., Stein, A., 2016. Geospatial big data handling theory and methods: A review and research challenges. *ISPRS J. Photogramm. Remote Sens.* 115, 119–133.
- Li, Y., Cheng, B., 2009. An improved k-nearest neighbor algorithm and its application to high resolution remote sensing image classification. In: 2009 17th International Conference on Geoinformatics. IEEE, pp. 1–4.
- Lin, G., Fu, D., Zhu, J., Dasmalchi, G., 2009. Cloud computing: IT as a service. *IT Prof.* 10–13.
- Liu, X., Hu, G., Chen, Y., Li, X., Xu, X., Li, S., Pei, F., Wang, S., 2018. High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine Platform. *Remote Sens. Environ.* 209, 227–239.
- Liu, D., Chen, N., Zhang, X., Wang, C., Du, W., 2020. Annual large-scale urban land mapping based on Landsat time series in Google Earth Engine and OpenStreetMap data: A case study in the middle Yangtze River basin. *ISPRS J. Photogramm. Remote Sens.* 159, 337–351.
- Liu, P., 2015. A survey of remote-sensing big data. *Front. Environ. Sci.* 3, 45.
- Lobo, F. de L., Souza-Filho, P.W.M., Novo, E.M.L. de M., Carlos, F.M., Barbosa, C.C.F., 2018. Mapping Mining Areas in the Brazilian Amazon Using MSI/Sentinel-2 Imagery. *Remote Sens.* 10, 1178. <https://doi.org/10.3390/rs10081178>.
- Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A., Jie, W., 2015. Remote sensing big data computing: Challenges and opportunities. *Future Gener. Comput. Syst.* 51, 47–60.
- Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Trans. Geosci. Remote Sens.* 55, 645–657.
- Mahdianpari, M., Mohammadimanesh, F., McNairn, H., Davidson, A., Rezaee, M., Salehi, B., Homayouni, S., 2019a. Mid-season Crop Classification Using Dual-, Compact-, and Full-Polarization in Preparation for the Radarsat Constellation Mission (RCM). *Remote Sens.* 11, 1582.
- Mahdianpari, M., Motagh, M., Akbari, V., Mohammadimanesh, F., Salehi, B., 2019b. A Gaussian random field model for de-speckling of multi-polarized Synthetic Aperture Radar data. *Adv. Space Res.* 64, 64–78.
- Mahdianpari, M., Salehi, B., Mohammadimanesh, F., 2017a. The effect of PolSAR image de-speckling on wetland classification: introducing a new adaptive method. *Can. J. Remote Sens.* 43, 485–503.
- Mahdianpari, M., Salehi, B., Mohammadimanesh, F., Brisco, B., 2017b. An assessment of simulated compact polarimetric SAR data for wetland classification using random Forest algorithm. *Can. J. Remote Sens.* 43, 468–484.
- Mahdianpari, M., Salehi, B., Mohammadimanesh, F., Brisco, B., Homayouni, S., Gill, E., DeLancey, E.R., Bourgeau-Chavez, L., 2020. Big Data for a Big Country: The First Generation of Canadian Wetland Inventory Map at a Spatial Resolution of 10-m Using Sentinel-1 and Sentinel-2 Data on the Google Earth Engine Cloud Computing Platform: Mégadonnées pour un grand pays: La première carte d'inventaire des zones humides du Canada à une résolution de 10 m à l'aide des données Sentinel-1 et Sentinel-2 sur la plate-forme informatique en nuage de Google Earth Engine™. *Can. J. Remote Sens.* 1–19.
- Mahdianpari, M., Salehi, B., Mohammadimanesh, F., Homayouni, S., Gill, E., 2019c. The first wetland inventory map of Newfoundland at a spatial resolution of 10 m using sentinel-1 and sentinel-2 data on the google earth engine cloud computing platform. *Remote Sens.* 11, 43.
- Mahdianpari, M., Salehi, B., Mohammadimanesh, F., Larsen, G., Peddle, D.R., 2018a.

- Mapping land-based oil spills using high spatial resolution unmanned aerial vehicle imagery and electromagnetic induction survey data. *J. Appl. Remote Sens.* **12**, 036015.
- Mahdianpari, M., Salehi, B., Mohammadimanesh, F., Motagh, M., 2017c. Random forest wetland classification using ALOS-2 L-band, RADARSAT-2 C-band, and TerraSAR-X imagery. *ISPRS J. Photogramm. Remote Sens.* **130**, 13–31.
- Mahdianpari, M., Salehi, B., Rezaee, M., Mohammadimanesh, F., Zhang, Y., 2018b. Very deep convolutional neural networks for complex land cover mapping using multi-spectral remote sensing imagery. *Remote Sens.* **10**, 1119.
- Mahdianpari, M., Salehi, B., Mohammadimanesh, F., Homayouni, S., Gill, E., 2019. The first wetland inventory map of newfoundland at a spatial resolution of 10 m using sentinel-1 and sentinel-2 data on the google earth engine cloud computing platform. *Remote Sens.* **11**, 43.
- Marghany, M., Hashim, M., Cracknell, A., 2011. Simulation of shoreline change using AIRSAR and POLSAR C-band data. *Environ. Earth Sci.* **64**, 1177–1189.
- Mateo-García, G., Muñoz-Marí, J., Gómez-Chova, L., 2017. Cloud detection on the Google Earth engine platform. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, pp. 1942–1945.
- McNairn, H., Champagne, C., Shang, J., Holmstrom, D., Reichert, G., 2009. Integration of optical and Synthetic Aperture Radar (SAR) imagery for delivering operational annual crop inventories. *ISPRS J. Photogramm. Remote Sens.* **64**, 434–449.
- McNairn, H., Hochheim, K., Rabe, N., 2004. Applying polarimetric radar imagery for mapping the productivity of wheat crops. *Can. J. Remote Sens.* **30**, 517–524.
- McNairn, H., Shang, J., 2016. A review of multitemporal synthetic aperture radar (SAR) for crop monitoring. In: *Multitemporal Remote Sensing*. Springer, pp. 317–340.
- Moed, H., 2012. The Evolution of Big Data as a Research and Scientific Topic: Overview of the Literature, 2012, ResearchTrends.
- Mohammadimanesh, F., Salehi, B., Mahdianpari, M., Brisco, B., Gill, E., 2019a. Full and simulated compact polarimetry sar responses to canadian wetlands: Separability analysis and classification. *Remote Sens.* **11**, 516.
- Mohammadimanesh, F., Salehi, B., Mahdianpari, M., Brisco, B., Motagh, M., 2018a. Wetland Water Level Monitoring Using Interferometric Synthetic Aperture Radar (InSAR): A Review. *Can. J. Remote Sens.* **1**–16. <https://doi.org/10.1080/07038992.2018.1477680>.
- Mohammadimanesh, F., Salehi, B., Mahdianpari, M., Brisco, B., Motagh, M., 2018b. Multi-temporal, multi-frequency, and multi-polarization coherence and SAR back-scatter analysis of wetlands. *ISPRS J. Photogramm. Remote Sens.* **142**, 78–93.
- Mohammadimanesh, F., Salehi, B., Mahdianpari, M., English, J., Chamberland, J., Alasset, P.-J., 2018c. Monitoring surface changes in discontinuous permafrost terrain using small baseline SAR interferometry, object-based classification, and geological features: a case study from Mayo, Yukon Territory, Canada. *GIScience Remote Sens.* **1**–26.
- Mohammadimanesh, F., Salehi, B., Mahdianpari, M., Gill, E., Molinier, M., 2019b. A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem. *ISPRS J. Photogramm. Remote Sens.* **151**, 223–236.
- Mohammadimanesh, F., Salehi, B., Mahdianpari, M., Motagh, M., Brisco, B., 2018d. An efficient feature optimization for wetland mapping by synergistic use of SAR intensity, interferometry, and polarimetry data. *Int. J. Appl. Earth Obs. Geoinformation* **73**, 450–462.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann. Intern. Med.* **151**, 264–269.
- Momm, H., Easson, G., 2011. Feature extraction from high-resolution remotely sensed imagery using evolutionary computation. INTECH Open Access Publisher.
- Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **66**, 247–259.
- Nguyen, U.N., Pham, L.T., Dang, T.D., 2019. An automatic water detection approach using Landsat 8 OLI and Google Earth Engine cloud computing to map lakes and reservoirs in New Zealand. *Environ. Monit. Assess.* **191**, 235.
- Nogueira, K., Penatti, O.A., dos Santos, J.A., 2017. Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognit.* **61**, 539–556.
- Parks, S.A., Holsinger, L.M., Koontz, M.J., Collins, L., Whitman, E., Parisien, M.-A., Loehman, R.A., Barnes, J.L., Bourdon, J.-F., Boucher, J., 2019. Giving ecological meaning to satellite-derived fire severity metrics across North American forests. *Remote Sens.* **11**, 1735.
- Plaza, A.J., Chang, C.-I., 2007. High performance computing in remote sensing. CRC Press.
- Quintero, N., Viedma, O., Urbieto, I.R., Moreno, J.M., 2019. Assessing Landscape Fire Hazard by Multitemporal Automatic Classification of Landsat Time Series Using the Google Earth Engine in West-Central Spain. *Forests* **10**, 518.
- Ravanelli, R., Nascenti, A., Cirigliano, R.V., Di Rico, C., Leuzzi, G., Monti, P., Crespi, M., 2018. Monitoring the impact of land cover change on surface urban heat island through Google Earth Engine: Proposal of a global methodology, first applications and problems. *Remote Sens.* **10**, 1488.
- Reboldi, F., Meroni, M., Urbano, F., Csak, G., Kerdiles, H., Perez-Hoyos, A., Lemoine, G., Leo, O., Negre, T., 2019. ASAP: A new global early warning system to detect anomaly hot spots of agricultural production for food security analysis. *Agric. Syst.* **168**, 247–257.
- Rezaee, M., Mahdianpari, M., Zhang, Y., Salehi, B., 2018. Deep convolutional neural network for complex wetland classification using optical remote sensing imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **11**, 3030–3039.
- Saah, D., Johnson, G., Ashmall, B., Tondapu, G., Tenneson, K., Patterson, M., Poortinga, A., Markert, K., Quyen, N.H., San Aung, K., 2019. Collect Earth: An online tool for systematic reference data collection in land cover and use applications. *Environ. Model. Softw.* **118**, 166–171.
- Sader, S.A., Winne, J.C., 1992. RGB-NDVI colour composites for visualizing forest change dynamics. *Int. J. Remote Sens.* **13**, 3055–3067.
- Scherler, D., Wulf, H., Gorelick, N., 2018. Global Assessment of Supraglacial Debris-Cover Extents. *Geophys. Res. Lett.* **45**, 11–798.
- Schulz, K., Hänsch, R., Sörgel, U., 2018. Machine learning methods for remote sensing applications: an overview. In: *Earth Resources and Environmental Remote Sensing/GIS Applications IX*. International Society for Optics and Photonics, pp. 1079002.
- Sidhu, N., Pebesma, E., Câmara, G., 2018. Using Google Earth Engine to detect land cover change: Singapore as a use case. *Eur. J. Remote Sens.* **51**, 486–500.
- Silva, W.F., Rudorff, B.F., Formaggio, A.R., Paradella, W.R., Mura, J.C., 2009. Discrimination of agricultural crops in a tropical semi-arid region of Brazil based on L-band polarimetric airborne SAR data. *ISPRS J. Photogramm. Remote Sens.* **64**, 458–463.
- Steinberg, D., Colla, P., 2009. CART: classification and regression trees. *Top Ten Algorith. Data Min.* **9**, 179.
- Stroppiana, D., Azar, R., Calò, F., Pepe, A., Imperatore, P., Boschetti, M., Silva, J., Brivio, P., Lanari, R., 2015. Integration of optical and SAR data for burned area mapping in Mediterranean regions. *Remote Sens.* **7**, 1320–1345.
- Snipir, B., Momblanch, A., Jain, S.K., Waine, T.W., Holman, I.P., 2019. A method for monthly mapping of wet and dry snow using Sentinel-1 and MODIS: Application to a Himalayan river basin. *Int. J. Appl. Earth Obs. Geoinformation* **74**, 222–230.
- Sun, W., Wang, R., 2018. Fully Convolutional Networks for Semantic Segmentation of Very High Resolution Remotely Sensed Images Combined with DSM. *IEEE Geosci. Remote Sens. Lett.* **15**, 474–478.
- Teluguntla, P., Thenkabail, P.S., Oliphant, A., Xiong, J., Gumma, M.K., Congalton, R.G., Yadav, K., Huete, A., 2018. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS J. Photogramm. Remote Sens.* **144**, 325–340. <https://doi.org/10.1016/j.isprsjprs.2018.07.017>.
- Tomer, S., Al Bitar, A., Sekhar, M., Zribi, M., Bandyopadhyay, S., Sreelash, K., Sharma, A.K., Corgne, S., Kerr, Y., 2015. Retrieval and multi-scale validation of soil moisture from multi-temporal SAR data in a semi-arid tropical region. *Remote Sens.* **7**, 8128–8153.
- Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., 2012. GMES Sentinel-1 mission. *Remote Sens. Environ.* **120**, 9–24.
- Uddin, K., Matin, M.A., Meyer, F.J., 2019. Operational Flood Mapping Using Multi-Temporal Sentinel-1 SAR Images: A Case Study from Bangladesh. *Remote Sens.* **11**, 1581.
- Voight, C., Hernandez-Aguilar, K., Garcia, C., Gutierrez, S., 2019. Predictive Modeling of Future Forest Cover Change Patterns in Southern Belize. *Remote Sens.* **11**, 823.
- Walker, E., Venturini, V., 2019. Land surface evapotranspiration estimation combining soil texture information and global reanalysis datasets in Google Earth Engine. *Remote Sens. Lett.* **10**, 929–938.
- Waller, E.K., Villarreal, M.L., Poitras, T.B., Nauman, T.W., Duniway, M.C., 2018. Landsat time series analysis of fractional plant cover changes on abandoned energy development sites. *Int. J. Appl. Earth Obs. Geoinformation* **73**, 407–419.
- Wang, C., Jia, M., Chen, N., Wang, W., 2018. Long-term surface water dynamics analysis based on Landsat imagery and the Google Earth Engine platform: A case study in the middle Yangtze River Basin. *Remote Sens.* **10**, 1635.
- Weissmann, H., Kent, R., Michael, Y., Shnerb, N.M., 2017. Empirical analysis of vegetation dynamics and the possibility of a catastrophic desertification transition. *PLoS One* **12**.
- Wilder, B., 2012. Cloud architecture patterns: using microsoft azure. O'Reilly Media Inc.
- Wong, B.A., Thomas, C., Halpin, P., 2019. Automating offshore infrastructure extractions using synthetic aperture radar & Google Earth Engine. *Remote Sens. Environ.* **233**, 111412.
- Workie, T.G., Debella, H.J., 2018. Climate change and its effects on vegetation phenology across ecoregions of Ethiopia. *Glob. Ecol. Conserv.* **13**, e00366.
- Wulder, M.A., Loveland, T.R., Roy, D.P., Crawford, C.J., Masek, J.G., Woodcock, C.E., Allen, R.G., Anderson, M.C., Belward, A.S., Cohen, W.B., 2019. Current status of Landsat program, science, and applications. *Remote Sens. Environ.* **225**, 127–147.
- Xie, Y., Lark, T.J., Brown, J.F., Gibbs, H.K., 2019. Mapping irrigated cropland extent across the conterminous United States at 30 m resolution using a semi-automatic training approach on Google Earth Engine. *ISPRS J. Photogramm. Remote Sens.* **155**, 136–149. <https://doi.org/10.1016/j.isprsjprs.2019.07.005>.
- 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. <https://doi.org/10.1016/j.isprsjprs.2017.01.019>.
- Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., Atkinson, P.M., 2018. An object-based convolutional neural network (OCNN) for urban land use classification. *Remote Sens. Environ.* **216**, 57–70.
- Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: a comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **5**, 8–36.