Big Data Methods for Resource and Environmental Data Generation (Postprint)

Authors: Wu Bingfang, Zhang Xin, Zeng Hongwei, Zhang Miao, Tian Fuyou

Date: 2023-03-19T00:00:00+00:00

Abstract

Resource and environmental monitoring serves as the cornerstone for achieving sustainable development goals for humanity, with ground surveys and remote sensing monitoring representing critical approaches in this domain. The widespread adoption of smartphones and the consequent emergence of crowdsourced geographic data have introduced novel methodologies and pathways for ground-based resource and environmental surveys. The exponential proliferation of cloud-based resource and environmental data products has significantly enhanced data transparency and reliability. The flourishing development of cloud-based professional services for both crowdsourced geographic data and cloud-based resource and environmental data products will supplant existing data acquisition channels, management paradigms, and analytical techniques with more efficient big data approaches. Public participation facilitates civic engagement in resource and environmental monitoring and governance, aligning with the fundamental principle that resources and environment should serve the common interests and welfare of all humankind, while concurrently mitigating the substantial investment requirements and high uncertainties associated with conventional operational data collection processes.

Full Text

Big Data Methods for Environmental Data Generation

WU Bingfang^{1,2,3*} **ZHANG Xin**^{1,2} **ZENG Hongwei**^{1,2} **ZHANG Miao**^{1,2} **TIAN Fuyou**^{1,2,3}

^{1} Institutes of Aerospace Information, Chinese Academy of Sciences, Beijing 100101, China
^{2} Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China
^{3} University of Chinese Academy of Sciences, Beijing 100049, China

Abstract

Resource and environmental monitoring have always been crucial for achieving sustainable development goals, with ground surveys and remote sensing serving as fundamental approaches. The proliferation of smartphones and the resulting crowdsourced geographic data offer new avenues for ground-based resource and environmental investigations. Meanwhile, the exponential growth of cloud-based resource and environmental data products has dramatically enhanced data transparency and reliability. The vigorous development of cloud-based professional services for crowdsourced geographic data and resource-environment data products will replace existing data acquisition pathways, management approaches, and analytical methods with more efficient big data techniques. Public participation enables citizens to engage in resource and environmental monitoring and management, aligning with the fundamental purpose of resource and environmental stewardship for the common good of humanity while avoiding the excessive investment and high uncertainty inherent in traditional survey methods.

Keywords: resource environment, cloud computing, crowdsourced geographic data, big data

---

1. Crowdsourced Geographic Data

Crowdsourced geographic information represents an effective means of rapidly acquiring large volumes of geospatial data, primarily collected by non-professionals and aggregated to servers, distributed databases, or cloud platforms according to specific standards. This approach yields what is known as volunteered geographic information (VGI). Such data has become a favored acquisition method in resource and environmental fields, primarily sourced from unconscious collection by the public without requiring specific objectives or traditional data collection expertise. As early as 1890, before modern communication technologies, the U.S. National Weather Service established the Cooperative Observer Program, whose datasets have been widely applied in weather monitoring, extreme weather warnings, and climate change research. Similarly, the North American Breeding Bird Survey has employed crowdsourcing for long-term transnational monitoring of avian populations, documenting distribution patterns and population changes for over 400 species. Since the 21st century, mobile internet technology has made smartphones indispensable daily devices, while diverse mobile applications have further propelled the development of crowdsourced geographic data.

The trend in crowdsourced data collection has accelerated rapidly, particularly with smartphone penetration providing powerful acquisition tools. In the mobile internet era, each terminal continuously generates rich spatiotemporal information resources, creating a new paradigm where everyone becomes an earth observation information collector and spawning massive volumes of crowdsourced geographic data encompassing location, velocity, trajectories, coverage areas, and photographs. Numerous mobile applications have been developed for geographic data collection, enabling various applications. For instance, Fritz et al. established the Geo-Wiki.org website and released the "GEO-wiki pictures" mobile application, allowing global volunteers to upload GPS-tagged photos to provide "crowdsourced data" on forests, grasslands, farmland, and water bodies, which was subsequently used to correct and improve global cropland distribution data quality. Other applications including "GIS Cloud," "Poimapper," "GeoODK Collect," and "FieldMap" have been widely used for crowdsourced geographic information collection. Traditional GIS tool providers like ARCGIS have also developed mobile applications such as "Collector for ArcGIS." OpenStreetMap (OSM), based on crowdsourcing principles, creates a freely editable global map where registered users can upload GPS tracks obtained from handheld devices, aerial photographs, satellite imagery, or other methods, and edit vector data using OSM's online editor or other software to achieve collaborative map maintenance.

Photographs can yield substantial information for determining object shapes and pathways. Since 2015, the Global Crop Monitoring team has upgraded its original "GVG (GPS, VIDEO, and GIS) crop sampling system" desktop software to a smartphone application, available in major app stores, enabling 随时随地 collection of crop cultivation status photos. This method has successfully facilitated rapid collection of farmland photos and crop type information, obtaining over 100,000 annual data records from different users conducting global crop planting structure surveys, significantly reducing the time and financial costs of acquiring global crop planting structure information while providing ground observation big data support for crop area estimation and forecasting. In 2017, leveraging this application, over 750,000 ground sample records covering 1,381 county-level administrative units across China were efficiently acquired within 70 days, supporting regional extraction of paddy fields, dryland, and other land cover types in the 2017 national land cover classification.

The upgraded "GVG Crowdsourced Geographic Data Collection" application employs fixed forms to provide non-professionals with solutions for identifying different land cover types, reducing uncertainty in the crowdsourced data collection process. The application has transformed previous work models requiring global ground observations or data sharing, substantially reducing ground observation workload, manpower, and financial investment while providing an economical solution for land cover classification sample acquisition. Currently available for free download on Google, Apple, and Huawei platforms, the application communicates with cloud servers in real-time; after simple account verification, users can collect land cover and crop sample information. As user numbers increase, the data collected through this application assumes ever-greater big data characteristics.

However, current smartphone photos remain underutilized, with applications remaining superficial and deeper applications yet to be developed. Moreover, the standardized approach limits unconscious public behavior, transforming unconscious data collection into conscious data gathering, leading to the "niche-ization" of crowdsourced data—representing the greatest challenge facing crowdsourced data collection.

2. Cloud-Based Resource and Environmental Data

The 21st century has witnessed explosive growth in earth observation data and its widespread application, enabling multi-scale, comprehensive three-dimensional earth observation. Massive multi-source remote sensing data has brought tremendous convenience to resource and environmental monitoring. However, conventional computing models can no longer satisfy the rapid processing requirements for such large volumes of long time-series remote sensing data. Cloud storage and computing technologies have developed rapidly in recent years. Compared with traditional personal computers and servers, cloud platforms offer high computational efficiency, strong performance, elastic scalability, large storage capacity, low cost, and data security—characteristics highly suitable for processing massive geographic datasets. Geographic data cloud platforms have increasingly become research and production platforms for resource and environmental data, enabling efficient data processing in the cloud without downloading massive datasets locally, thereby dramatically improving analysis efficiency and enabling examination of longer time-series and larger spatial scales without computational or storage limitations. This has spawned substantial volumes of cloud-based resource and environmental data.

In 2011, Google released the Google Earth Engine geographic data cloud computing platform, while the Australian Geoscience Agency simultaneously proposed and developed the "Data Cube" cloud-based geographic data processing solution running on Australia's supercomputing platform, achieving consistent data management frameworks for remote sensing, meteorological, and ground station data across Australia. In 2016, Australia open-sourced "Data Cube," fully disclosing all data frameworks and application algorithms while supporting users in building their own data management and computing systems, thereby gaining widespread global user support. Additionally, AWS has opened earth observation data, providing 61 datasets including NASA Earth Exchange, global Landsat series, Sentinel series satellites, NEXRAD weather radar, NAIP, and digital elevation models (DEM), enabling scientists to conveniently conduct global resource and environmental monitoring in the cloud.

Cloud computing and machine learning have substantially promoted the transition of resource and environmental monitoring toward high-resolution element monitoring. Chinese scientists have produced 30 m global land cover products using Google Cloud. The EU Joint Research Centre completed global-scale 30 m resolution surface water spatial distribution monitoring from 1984–2015 using Google Earth Engine. Since surface water most intuitively reflects regional water resource conditions, long-term comparable sequential datasets provide valuable information for diagnosing water resource stress and changes in arid ecosystems. The EU Joint Research Centre has also conducted global human settlement monitoring for 1975, 1990, 2000, and 2014. The University of Maryland completed global 30 m resolution forest cover change monitoring from 2000–2016. The USGS used Google Earth Engine and supercomputers to produce the first global 30 m resolution cropland planting spatial distribution map for 2015. The Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences completed global 30 m resolution burned area remote sensing monitoring for 2015. The European Space Agency developed the S2ToolBox for Sentinel datasets, enabling monitoring of vegetation leaf area index (LAI), fraction of absorbed photosynthetically active radiation (FAPAR), and vegetation cover (FCOVER) at 20 m resolution, substantially integrating high-resolution type and function monitoring.

New data sources continue emerging, and new classification, recognition, and analysis methods—particularly artificial intelligence technologies—are evolving rapidly. Cloud-based resource and environmental data products are growing exponentially, encompassing high, medium, and low resolutions, with increasing medium- and high-resolution remote sensing data types and resolutions progressing from kilometer-level to hectometer-level and decameter-level, with meter-level products inevitably emerging. Global-scale, higher-resolution remote sensing product production will become mainstream and a competitive frontier in earth observation. Following the 普及 of high-resolution cloud-based resource and environmental data, boundaries between scientific research and operational production will substantially blur. Leveraging cloud computing power and open data resources, producing high-resolution data products will gradually replace the functions of professional departments traditionally dedicated to data production.

3. Cloud-Based Professional Services for Resources and Environment

Massive data storage in the cloud and convenient information extraction provide new pathways for discovering resource and environmental problems and delivering professional services. Since 2013, the CropWatch team has analyzed differentiated information needs for crop monitoring and early warning among different users, establishing a cloud service platform for diverse agro-information requirements. This platform provides different agricultural information services for different users, creating the CropWatch-Cloud participatory global self-service crop monitoring platform based on public cloud. The platform comprises four modules: CropWatch Pro (global crop online production system), CropWatch Explorer (global crop online browsing), CropWatch Analysis (global crop remote sensing bulletin online analysis), and CropWatch Bulletin (global crop online publishing), achieving seamless chain integration of cloud data extraction and integration, monitoring model cloud processing, data information transparency, open analysis participation, and open monitoring results.

Cloud-Based Water Resources and Environment Monitoring: Integrating the EU Joint Research Centre's water surface monitoring dataset from 2000–2015 stored on Google Cloud and using Google Earth Engine, we analyzed changes in Beijing's permanent water surface area from 2000–2017. The results show Beijing's permanent water surface area declined rapidly from 515 km^2 in 2000 to 197 km^2 in 2015, a cumulative reduction of 318 km^2 (61.7%). The area rebounded from 2015–2017, reaching 244 km^2 in 2016 and 285 km^2 in 2017, representing increases of 47 km^2 and 85 km^2 respectively compared to 2015. Analysis of Tropical Rainfall Measuring Mission (TRMM) precipitation data stored on GEE revealed fluctuating upward trends in precipitation intensity from 2000–2017. The completion and water supply of the South-North Water Transfer Project's central route represents the primary cause of Beijing's water surface area rebound. According to the Beijing Water Resources Bulletin, the project supplied 1.063 billion cubic meters to Beijing in 2016 alone. Changes in permanent water surface area comprehensively reflect regional water resource abundance under combined natural and anthropogenic influences, demonstrating that cloud data alone can monitor and evaluate water resource management effectiveness.

Cloud-Based Agricultural Resources Monitoring: CropWatch Pro achieves fully automated operational monitoring of global agricultural meteorological conditions, crop growth status, and global food supply situations through cloud data extraction and aggregation. Simultaneously, the platform opens data and computing capabilities to users, allowing them to run and test their own crop monitoring algorithms. CropWatch Explorer combines vector maps, raster maps, and dynamic charts to display multiple agro-meteorological, crop condition, and food production indicators in browsers, enabling near-real-time global crop information services. CropWatch Analysis includes functions for creating analysis tasks, task allocation, online analysis and submission, and report publishing, allowing users to complete crop remote sensing monitoring and analysis for regions of interest anytime and anywhere via internet devices, achieving remote participation by global experts in monitoring and analysis work and collaboratively producing multi-language versions of global crop bulletins, thereby substantially enhancing global participation. Using containerization technology, CropWatch encapsulates all crop monitoring algorithms in its system, meeting users' needs for customized crop monitoring systems through microservices. This service has provided customized crop monitoring systems for countries along the Belt and Road Initiative, substantially improving their crop monitoring capabilities.

Cloud computing has increasingly transformed current data storage and processing methods and concepts. Compared with traditional geographic information systems, cloud-based professional service systems significantly reduce establishment, customization, updating, and maintenance costs. Cloud platform characteristics eliminate the need for users to waste time downloading and processing data, enabling resource and environmental monitoring and analysis of any region of interest anytime and anywhere, thereby breaking previous restrictions of national borders, geographic regions, and domains and allowing humanity to collectively address common resource and environmental challenges for the first time.

Cloud-based resource and environmental data products are growing exponentially. Unlike current data mostly stored in individual laboratories, data will no longer remain hidden on producers' hard drives or departmental archives. Cloud data makes resource and environmental data accessible to everyone, substantially improving data transparency and reliability through easy acquisition, comparison, and use.

Conclusion

Big data methods for acquiring resource and environmental data based on crowdsourced geographic data will gradually replace traditional operational methods, more effectively serving industry departments while enabling public participation in resource and environmental monitoring and management. This aligns with the fundamental principle that resource and environment constitute common human interests. These approaches will replace existing resource and environmental data acquisition methods, management approaches, and information value mining with more efficient crowdsourced and big data methods, gradually 淡化 the data production functions of specialized institutions, thereby substantially improving data collection efficiency and information value extraction while significantly reducing monitoring costs.

References

1. Wang Juanle, Sun Jiulin. Considerations on gridded resource and environment comprehensive scientific investigation. *Journal of Geo-Information Science*, 2015, 17(7): 758-764.
2. Li Deren. Prospects for geospatial informatics in the big data era. *Acta Geodaetica et Cartographica Sinica*, 2016, 45(4): 379-384.
3. Tulloch D. Crowdsourcing geographic knowledge: volunteered geographic information (VGI) in theory and practice. *International Journal of Geographical Information Science*, 2014, 28(4): 847-848.
4. Leeper R D, Rennie J, Palecki M A. Observational Perspectives from US Climate Reference Network (USCRN) and Cooperative Observer Program (COOP) Network: Temperature and Precipitation Comparison. *Journal of Atmospheric and Oceanic Technology*, 2015, 32(4): 703-721.
5. Sauer J R, Link W A, Fallon J E, et al. The North American breeding bird survey 1966–2011: summary analysis and species accounts. *North American Fauna*, 2013, 79: 1-32.
6. Fritz S, McCallum I, Schill C, et al. Geo-Wiki. Org: The use of crowdsourcing to improve global land cover. *Remote Sensing*, 2009, 1(3): 345-354.
7. Fritz S, See L. Mapping global cropland and field size. *Global Change Biology*, 2015, 21(5): 1980-1992.
8. Ramm F, Topf J, Chilton S. *OpenStreetMap: using and enhancing the free map of the world*. Cambridge: UIT, 2010.

9. Wu B, Gommes R, Zhang M, et al. Global crop monitoring: a satellite-based hierarchical approach. *Remote Sensing*, 2015, 7(4): 3900-3933.
10. Zhang M, Wu B, Zeng H, et al. CropWatch: a global crop monitoring system with sentinel-1 SAR and sentinel-2 optical data. *International Journal of Digital Earth*, 2014, 7(2): 113-137.
11. Zhang M. GVG for iPhone. 2017. [2018-08-14]. https://itunes.apple.com/us/app/apple-store/id1244686128?mt=8.
12. Zhang M. GVG for Android. 2017. [2018-08-14]. https://play.google.com/store/apps/details?id=com.sys
13. Zhang M, Wu B, Zeng H, et al. Estimating forest leaf area index using digital camera: a case study in Maoershan forest farm of Northeast China. *Journal of Central South University of Forestry & Technology*, 2016, 36(7): 81-86.
14. Dong Jinwei, Kuang Wenhui, Liu Jiyuan. Continuous dynamic monitoring of global land cover supported by remote sensing big data. *Scientia Sinica (Terrae)*, 2018, 48(2): 259-260.
15. Shao Y, Di L, Bai Y, et al. Geoprocessing on the Amazon cloud computing platform—AWS. *2012 First International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*. Shanghai: IEEE, 2012: 1-6.
16. Gong P, Wang J, Yu L, et al. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing*, 2013, 34(7): 2607-2654.
17. Pekel J F, Cottam A, Gorelick N, et al. High-resolution mapping of global surface water and its long-term changes. *Nature*, 2016, 540(7633): 418.
18. Pesaresi M, Ehrlich D, Ferri S, et al. Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014. Luxembourg: Publications Office of the European Union, 2016.
19. Hansen M C, Potapov P V, Moore R, et al. High-resolution global maps of 21st-century forest cover change. *Science*, 2013, 342(6160): 850-853.
20. Xiong J, Thenkabail P S, Gumma M K, et al. Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2017, 126: 225-244.
21. Baret F, Weiss M, Lacaze R, et al. GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part 1: Principles of development and production. *Remote Sensing of Environment*, 2013, 137: 299-309.

Note: Figure translations are in progress. See original paper for figures.