

FROM THE GUEST EDITORS

BY PIER GIORGIO MARCHETTI, PIERRE SOILLE, AND LORENZO BRUZZONE

A Special Issue on Big Data from Space for Geoscience and Remote Sensing

Big data from space refers to Earth- and space observation data that are collected by spaceborne and ground-based sensors. Whether for Earth or space observation, they qualify for being called *big data* given the sheer volume of sensed data (i.e., archived data reaching the exabyte scale), their high velocity (new data are acquired on a continuous basis and at an increasing rate), their variety (data are delivered by sensors acting over various frequencies of the electromagnetic spectrum in passive and active modes), as well as their veracity (sensed data are associated with uncertainty and accuracy measurements, and extracted information needs to be validated). Last but not least, the value of big data from space depends on our capacity to extract information and meaning from them.

In the case of Earth observation (EO), the first data revolution occurred in 2008 when the U.S. Geological Survey decided to release its Landsat archive to the public for free, with that archive being the world's largest historical collection of Earth imagery. The Copernicus program of the European Union, with its fleet of Sentinel twin-satellite constellations operated by the European Space Agency (ESA), is further revolutionizing EO. It is indeed making EO truly enter the big data era by generating unprecedented amounts of free, full, and open EO data, producing up to 10 TB of data per day when all six twin-satellite constellations are at full operational capacity. The Sentinel satellite series was initiated in April 2014 with the launch of the Sentinel-1A Synthetic Aperture Radar mission. It was followed by the successful launches of the Sentinel-2A for high-resolution optical imagery on 29 June 2015, the Sentinel-3A multiple-sensing instruments on 4 February 2016, and the Sentinel-1B, the second unit of the Sentinel-1 constellation, on 25 April 2016.

Digital Object Identifier 10.1109/MGRS.2016.2586852
Date of publication: 20 September 2016

Furthermore, on 29 January 2016, the first European Data Relay Satellite (EDRS-A) was launched. Dubbed the *Space Data Highway*, EDRS is advancing satellite communications as Europe's first optical space communications network, capable of relaying user data in near-real time at an unprecedented 1.8 Gb/s. This Space Data Highway will play a key role in relaying the Sentinel data to ground, given the sheer amount of data that are being and will be generated by the Sentinel satellites operated by the ESA in the framework of the Copernicus program that is funded and managed by the European Commission.

Other recent initiatives for open access to satellite and satellite-derived data include the U.S. National Oceanic and Atmospheric Administration's big data partnership with industry, the free availability since April 2016 of all Earth imagery from the prolific Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Japanese remote sensing instrument operating aboard NASA's Terra spacecraft since late 1999, the provision by the Japanese Space Agency at no charge of the global Advanced Land-Observing Satellite "DAICHI" (ALOS-1) arcsec global digital surface model, and open access to the data acquired by Brazil from the China–Brazil Earth Resources Satellite series. The investments made by public funds have been complemented in the last few years by additional initiatives from the private sector, building and launching constellations of small satellites dedicated to imagery. Besides satellites, numerous other sources of digital geospatial data produced by countless devices as well as by citizens are becoming available and can further increase the volume, velocity, variety, and veracity of the data flows of interest to the field of remote sensing and geosciences.

These revolutions in EO, paralleled by similar developments in space observation, triggered the rise of a new community under the big data from space umbrella that was initiated in 2013 with the highly successful Big

Data from Space event organized by the ESA in Frascati, Italy, which was followed by two equally successful conferences co-organized by the ESA, the Joint Research Center of the European Commission, and the European Union Satellite Center in 2014 and 2016, respectively. The goal of Big Data from Space is to foster research, science, innovation, and technological developments that tackle the big data challenges in EO to increase and promote our capacity to extract meaningful information.

THESE FIVE ARTICLES PROVIDE A SAMPLING THAT HIGHLIGHTS THE DIVERSITY OF THE BIG DATA ISSUES EMERGING IN EO.

For this special issue on big data from space for geoscience and remote sensing, we have collected five articles that present a snapshot of the recent developments and advances. These articles were selected through a two-step submission/review process. The resulting five contributions cover some of the key challenges and leading-edge developments in the field of big data from space for the field of remote sensing and geoscience.

The first article, “Big Data Challenges in Climate Science,” distinguishes itself by striking a balance between data analytics and data storage within a cyberinfrastructure for climate science. The article proposes a new service model, i.e., Climate Analytics-as-a-Service, where adaptive analytics services coexist with a climate data service that is based on an evolution of the Open Archive Information System model while securing interoperability with the Earth System Grid Federation (ESGF). This type of new model is urgently needed in view of the research community’s collective work that is expected to generate up to 100 PB of data in the framework of the next Intergovernmental Panel on Climate Change Assessment Report, which is scheduled for 2020.

The second article, “Managing Big, Linked, and Open Earth-Observation Data,” addresses the challenge posed by the many applications that require access to data originating from different data silos. Indeed, linked data is a new paradigm that studies how data following the Resource Description Framework can be interconnected to increase their value. The developed stack of tools is demonstrated in the context of wildfire-monitoring services. Particular attention is given to exploration and visualization as well as discovering services that allow the selection of spatial and temporal relations.

The third article, “Supporting Earth-Observation Calibration and Validation,” introduces an emerging field exploiting the vast amounts of data provided by citizens in the form of in-situ data as well as visually interpreted and manually digitized EO data sets. The article presents a review of the existing crowdsourcing initiatives of relevance to EO and details a series of case studies before expanding on the challenges and barriers limiting the adoption of citizen science and crowdsourcing in support of EO.

The fourth article, “Big Data Management in Earth Observation,” emphasizes that big data management in EO is not just an information technology issue and that close collaboration and interaction between operations engineers, data managers, system and software engineers, and scientists are the key to the successful exploitation of satellite data archives.

The fifth and last article, “A Multi-GPU Program for Uncertainty-Aware Drainage Basin Delineation,” addresses the timely processing of massive digital elevation models (DEMs) for river network delineation thanks to the use of graphics processing units (GPUs) combined with Message Passing Interface via the host central processing units. The proposed solution is scalable and will cope with the data volumes originating from future countrywide DEMs at a spatial resolution down to 2 m.

These five articles provide a sampling that highlights the diversity of the big data issues emerging in EO. The velocity at which the data volumes and variety of data sources is increasing will call for further advances and innovative approaches. Further effort will be needed, particularly in metadata processing, to ensure that it will be possible to select, in an efficient and effective way, the smallest subset of the available (linked) data needed to provide the desired information while bringing the uncertainty to an acceptable level. In addition, big data in geosciences and remote sensing, unlike in other fields faced with big data issues, cannot be reduced to data-driven science by simply looking for correlations and patterns in the input data. Indeed, expert knowledge in geosciences needs to be combined with advanced data science to maximize the extraction of valuable information in an effective and efficient way. At a technical level, it will become unsustainable to keep bringing the data to geoscientists, for the evolution of the bandwith of network connections is not keeping up with the pace of data volume increase. Therefore, it will become more and more important to develop seamless ways to bring the processing close to the data, including such innovative approaches as the ESGF developed in the first article of this special issue.

AUTHOR INFORMATION

Pier Giorgio Marchetti received the Laurea (M.Sc.) degree in electronic engineering from the University of Rome “La Sapienza,” Italy, in 1977. From 1979 to 1985, he served as researcher with the Fondazione Ugo Bordoni, Rome, Italy, where he carried out research projects on microwave atmospheric propagation, digital signal processing, digital radio communication, and cochannel interference for radio relay and satellite links. He participated in the cochannel interference experiment between the Italian SIRIO and the ESA’s Orbital Test Satellite with small receiving stations in Italy and China. In 1978, he served as a junior engineer in the Radio Laboratory of Autovox Spa (a subsidiary of Motorola Inc.), Rome, Italy. From 1986 to 2000, he served as a staff member of the ESA in various areas, including software development, maintenance, operations, and line management. He ran technology research projects on hypermedia and intelligent interfaces and contributed to American National Standards Institute Z39.50 standardization activities. From 2000 to 2010, he was with the ESA's Earth Observation Ground Segment Department, running studies and managing projects on ground segment technology development, data and information mining, the harmonization and standardization of ground segment interfaces, and service-oriented infrastructures encompassing both Earth observation and geospatial data. He coauthored the book *Heterogeneous Missions Accessibility, ESA TM-21*. Since June 2010, he has been head of the Research and Ground Segment Technology Section, Earth Observation Programme, European Space Agency (ESA), Frascati, Italy. He was cochair of the Big Data from Space conference in 2014 and 2016. He has coauthored more than 80 technical reports and conference and journal papers. His current research interests include big data from space and processes for research and development.

Pierre Soille received the agricultural engineering degree in 1988 from the University of Louvain, Louvain-la-Neuve, Belgium; the Ph.D. degree in agronomical sciences in 1992 from the University of Louvain, in collaboration with the Center for Mathematical Morphology of the École des Mines de Paris, France; and the habilitation degree in computer science in 1997 from Montpellier II University, France. After completing his Ph.D. on mathematical morphology applied to digital elevation models and satellite-image analysis, he pursued research on image analysis and mathematical morphology at the Commonwealth Scientific and Industrial Research Organization's Mathematical and Information Sciences in Sydney, Australia; the Center for Mathematical Morphology of the École des Mines de Paris, France; and the Pattern Recognition of the Fraunhofer-Institut IPK in Berlin, Germany. During 1995–1998, he was an assistant professor with the École des Mines d'Alès and EERIE, Nîmes, France. In 1999, he was appointed as a senior researcher with the Biotechnology and Biological Sciences Research Council, United Kingdom. He has been with the Joint Research Center (JRC) of the European Commission, Ispra, Italy, since December 1999. His current research activities are dedicated to automatic information retrieval from high- to very-high-resolution satellite-image data for the generation of new global geospatial data sets. He is currently leading a JRC pilot project that is focused on the development of big data technologies in the broad context of geospatial data, with an emphasis on the exploitation of datastreams originating from the European Union's Copernicus program and its fleet of Sentinel satellites. Together with Pier Giorgio Marchetti, he acted as program cochair for the first two Big Data from Space conferences (BiDS 2014 and BiDS 2016). He has authored and coauthored numerous and highly cited image-analysis and pattern-recognition algorithms and methods published in peer-reviewed journal and conference papers, as well as book chapters. He has also edited various books and is author of the reference monograph *Morphological Image Analysis: Principles and Applications* (Springer, 1999 and 2004).

Lorenzo Bruzzone received the Laurea (M.S.) degree in electronic engineering (summa cum laude) and the Ph.D. degree in telecommunications from the University of Genoa, Italy, in 1993 and 1998, respectively. He is currently a full professor of telecommunications at the University of Trento, Italy, where he teaches remote sensing, radar, pattern recognition, and electrical communications. He is the founder and director of the Remote Sensing Laboratory in the Department of Information Engineering and Computer Science, University of Trento. His current research interests are in the areas of remote sensing, radar and synthetic aperture radar, signal processing, and pattern recognition. He promotes and supervises research on these topics within the frameworks of many national and international projects. Among others, he is the principal investigator for the Radar for Icy Moon Exploration instrument in the framework of the Jupiter Icy Moons Explorer mission of the European Space Agency. He is the author or coauthor of 161 scientific publications in referred international journals (111 in IEEE journals), more than 220 papers in conference proceedings, and 17 book chapters. He is editor or coeditor of 16 books and conference proceedings and one scientific book. His papers are highly cited, as proven by the total number of citations (more than 15,300) and the value of his h-index, which is 63 (source: Google Scholar). He was invited as a keynote speaker at 30 international conferences and workshops. Since 2009, he has been a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society. He came in first in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium in Seattle, Washington, July 1998. Since that time, he has been a recipient of many international and national honors and awards. He was a guest coeditor of different special issues of international journals. He is cofounder of the IEEE International Workshop on the Analysis of Multi-Temporal Remote Sensing Images (MultiTemp) series and is currently a member of the Permanent Steering Committee of this series of workshops. Since 2003, he has been chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. Since 2013, he has been the (founding) editor-in-chief of *IEEE Geoscience and Remote Sensing Magazine*. Currently, he is an associate editor of *IEEE Transactions on Geoscience and Remote Sensing* and the *Canadian Journal of Remote Sensing*. In 2012, he was appointed a Distinguished Speaker of the IEEE Geoscience and Remote Sensing Society. He is a Fellow of the IEEE.

GRS