COPERNICUS SENTINEL-2 IMAGERY: HOW MUCH DATA ARE AVAILABLE TO PERFORM TERRITORIAL MONITORING?

Flavio Lupia[1], Guido Bonati[2], Giuseppe Pulighe[3]

Abstract

The Copernicus Programme is the European effort to monitor the Earth and ecosystems by providing Earth Observation information and services to handle a broad range of challenges (e.g. climate change, transports, sustainable development, regional and local planning, security, health, agriculture, etc.). Copernicus information are collected through a dedicated constellation of satellites called “the Sentinels”. Satellite imagery are distributed freely contributing to open up a new economy with downstream services and applications suitable for a wide range of themes. Sentinel satellites, developed with state-of-the-art technologies, enact unprecedented powerful tools with near real-time capabilities to monitor changes for urban and rural planning and decision making. In this contribution, a first assessment of the amount of Earth Observation data from Sentinel-2 mission is performed by considering a an Italian region as reference. Data querying and discovery methods are applied on the Copernicus Open Access Hub, the node where all the satellite imagery are distributed with open policy. Indicators, tables and graphs are produced to describe the amount of data available and the temporal coverage, highlighting the increasing capabilities for territorial monitoring. Finally, we report some considerations about the vast amount of data and the need to address a Big data challenge where new options for data management and high speed data infrastructures need to be deployed.

Keywords: Copernicus Programme; Sentinel-2; Earth Observation; Big Data; Calabria.

1. Introduction

Monitoring the state of the environment is of chief importance for assessing and defining actions for earth surface changes driven by natural and man-made activities. Earth Observation (EO) techniques and tools, such as remote sensing, are by now a well-known approach for collecting data at different scale levels (from local to global scale). Land cover/land use mapping and monitoring are one of the most relevant applications where remote sensing is applied providing information to support decision making within fields such as spatial planning. Since 1970 several satellite-based programs have been launched with the aim of collecting information and understanding natural and anthropogenic processes. Each program has been always rated by its peculiar capabilities in terms of data collection and quality of information. Parameters such as spatial, temporal, spectral and radiometric resolution, determined by the evolution of sensors technology, are used to select the most suitable sensor for the specific monitoring application. The selection process of the most suitable satellite for a given application has often been constrained by costs, since access to data has been always charged for all missions, at least until the free distribution of the NASA Landsat Program imagery.

1 CREA – Centro di ricerca Politiche e Bioeconomia, Via Po, 14 – 00198 Roma, e-mail: flavio.lupia@crea.gov.it (corresponding author).
2 CREA – Centro di ricerca Politiche e Bioeconomia, Via Po, 14 – 00198 Roma, e-mail: guido.bonati@crea.gov.it
3 CREA – Centro di ricerca Politiche e Bioeconomia, Via Po, 14 – 00198 Roma, e-mail: giuseppe.pulighe@crea.gov.it

Copernicus is the programme launched by the European Commission and operated by the European Space Agency (ESA). It is by far the most relevant EO programme to date designed to produce accurate, timely and simply accessible information to support several applications in the domains of environment, climate change and civil security (ESA, 2018a). ESA coordinates the delivery of data from more than 30 satellites, which are complemented with a set of services for collecting of ground-based, airborne and seaborne sensor data. However, the chief advances are related to the development of a new family of satellites, called the Sentinels, equipped with state-of-the-art sensors for monitoring the Earth by using either active microwave sensors or passive multispectral sensors. Copernicus is both a data generator and a services provider combining EO and in-situ data to deduct information for terrestrial monitoring. For instance, the Copernicus Land Monitoring service (Copernicus Programme, 2018) provides georeferenced information on land cover/land use, vegetation and water cycle status supporting several applications (e.g. spatial planning, forest management, water management and agriculture).

The Programme is expected to provide new opportunities for EO and to push economic development of several sectors thanks to a key feature: the free and open access to data and tools for their management. An analysis drafted by PricewaterhouseCoopers in 2015 estimated market opportunities in several sectors between EUR 480 and EUR 3,135 million forecasting a total employment between 3,050 and 14,450 person years over the 2015-2020 period (European Commission, 2016).

The Sentinel satellites fleet produce daily a wealth of data (10 Terabytes per day ca.) and information providing a wide range of options for territorial monitoring with a steady increase of the data volume expected with the future development of Copernicus. In this context, the amplitude of the potentiality for scientific analysis of terrestrial phenomena is counterbalanced by the difficulties to store, process and distribute large volume of data and EO-based services. Managing large amaunts of data has environmental and economic impacts related to the costs of hardware and energy consumption (Kempeneers and Soille, 2017). Remotely sensed data, such as those from the Copernicus Programme, are Big Data and belong to the Petabyte Age (Anderson, 2008), a scenario where new tools sift through huge data flows to mine information to be turned into knowledge and data-driven research supersedes the classical hypothesis-driven research.

The terrestrial monitoring potentiality of the Sentinel missions and the volume of data generated have been explored in different studies, some carried out before the operational deployment of the satellites. Such studies, often explore the monitoring capabilities of the integration of different satellites programs, a common approach where the capabilities of the different platforms are integrated to build a so-called “virtual constellation”, enabling to overcome the limitation of a single mission (e.g. temporal resolution can be increased when images are acquired from different platforms). Achieving high revisit times is significant when terrestrial observations need to monitor rapidly changing phenomena. Applications, such as agricultural monitoring, need frequent observations, for example to monitor vegetation phenology, and the availability of many acquisitions is fundamental especially in areas with a persistent cloud coverage when optical satellites are used.

A relevant example is the study of Li and Roy (Li and Roy, 2017) that analysed the volume of data acquired globally by Landsat-8, Sentinel-2A and Sentinel-2B by focussing on the revisit time of the single and combined platforms for monitoring areas worldwide. Similarly, Whitcraft et al. (Whitcraft et al., 2015) simulated the combination of different satellite missions with moderate spatial resolution (10 to 30 m pixel size) to assess the capability to meet the EO requirements for global agricultural monitoring (i.e. 8 days frequency of optical images acquisition).

In this paper, we used the global archived multispectral data of Sentinel-2A and 2B (the most relevant for terrestrial monitoring) to extract spatial and temporal metadata to explore the monitoring potentiality of the freely accessible Copernicus dataset. The approach uses archived data instead of simulated acquisitions as in the above mentioned studies. Metadata are used to produce basic metrics describing the volume of data and the revisit time in the selected study area. Computations are made without assessing the quality of the scenes (e.g. cloud coverage), the redundancy of data from overlapping scenes and their spatial coverage when acquisitions partially cover the study area.

2. Materials and methods

1.1 Copernicus Sentinel-2 missions

Sentinel-2 is a multi-spectral imaging mission with a wide swath (290 km), global coverage and high spatial and temporal resolution. The mission is made up of twin satellites (2A, launched on 23 June 2015 and 2B on 7 March 2017) flying in the same orbit but phased at 180° and designed for a revisit time of 5 days at the Equator and 2-3 days at mid-latitudes. Sentinel-2 collects radiometric information through 13 spectral bands: four bands at 10 m, six bands at 20 m and three bands at 60 m spatial resolution.

Data products generated are provided with different level of preprocessing where Level-1C and Level-2A are the most relevant for users. Level-1C products represent top of atmoshere reflectances, undergone radiometric and geometric corrections and are provided as granules (tiles) of 100 km^2 in cartographic geometry (UTM –WGS84). Each tile is coded following the Military Grid Reference System (MGRS), the standard coordinate system used by NATO and derived from the Universal Transverse Mercator grid system. Level-1C products are resampled with a constant ground sampling distance of 10, 20 and 60 m depending on the resolution of the different spectral bands. Level-2A products are derived from Level-1C by applying atmosferic corrections and represent bottom of atmosphere reflectances (ESA, 2015).

The acquisition of images with a large swath, 13 spectral bands, global coverage and high revisit frequency implies the creation of a large amount of raw compressed data per satellite unit (up to 1. 3 TB daily) that corresponds to an average continuously sustained raw data supply rate of 160 Mbps (ESA, 2018b).

Data distribution is performed with open access policy. Sentinel-2 products are fully accessible to registered users through the portal Copernicus Open Access Hub (https://scihub.copernicus.eu/). Several services are available for data discovery and access: preview of images and metadata, full-text search and saved serch queries, bach scripting and Application Programming Interface (API) for development of customized services.

1.2 Data discovery and database creation

We built a database containing references to all the images acquired by the twin missions Sentinel-2A and Sentinel-2B for the time range 1 July 2017 – 30 April 2018 (10 months). The date range was selected to ensure the contemporary operativity of the twin missions since Sentinel-2B launch occurred more than one year apart from Sentinel-2A. We considered only products processed at Level-1C composed of 100x100 km^2 tiles (ortho-images in UTM/WGS84 projection) projected in cartographic geometry by using a Digital Elevation Model (DEM).

The database was populated with data from the catalogue of Sentinel-2 imagery acquired worldwide in text format (comma-separated values - CSV) by ESA. These data, available as from 4 September 2017, are a catalogue view of the products published on the Copernicus Open Access Hub and the CSV files are updated routinely every month. The CSV files are published online with a hierarchical structure with folder and subfolders (i.e. from parent to child: MISSION (e.g. S2A), YEAR (e.g. 2017), MONTH (e.g. 01)) with a full disaggregation of the catalogue with one file for each day of the year (e.g. the single file S2A_20170101_OPENHUB_catalogue_20180228235959.csv contains all the imagery acquired globally by the mission Sentinel-2A on 1 January 2017).

The version updated on 7 May 2018 was downloaded, it contains the snapshot of the whole catalogue with products up to 30 April 2018. All the possible daily CSV files for the whole time range for the two missions were downloaded through a batch script with the free software package GNU Wget (https://www.gnu.org/software/wget/) obtaining a total of 779 files later joined together to create a single file with a script written in R language (R Development Core Team, 2018). The merged information of the catalog was later imported in a database to be queried and analyzed.

Among the available attributes of the catalogue the following fields were retained to allow the analysis foreseen by this study: NAME (product filename); CONTENTLENGTH (size in bytes of the .ZIP file of the product); INGESTIONDATE (date and time of the ingestion of the product in the data hub system) and CONTENTDATE:START (sensing start date and time of the product). In the database only Level-1C products

from the two Sentinel missions were retained originating a list of 2,443,488 different products with a worldwide coverage for the 10 months time range. Later on, a subset of the product was extracted by querying on the six tiles covering Calabria region (Figure 1) amounting to a total of 739 products for the defined time range.

Figure 1- The study area, Calabria region, with the overlaps of the Sentinel-2 tiles boundaries. Tiles structure and coding derive from the Military Grid Reference System (MGRS). Each Sentinel-2 product is provided as tile (granule) that is a 100 km^2 ortho-images in UTM/WGS84 projection. Tile boundaries are derived from the KML file distribute by European Space Agency.

3. Results and discussions

A set of metrics were computed by considering the single Sentinel-2 mission (2A and 2B) and their combination as constellation (2A+2B) for the 10 month time range: cumulative number of acquisitions, mean revisit time and data volume.

3.1 Cumulative number of acquisitions

Figure 2 shows the monthly number of products acquired for the single Sentinel-2 missions and for the constellation with the cumulative number for the period of analysis (1 July 2017 -30 April 2018). The amount of products available for the time range is computed for the six tiles as whole covering the study area: 370 (S2A), 369 (S2B) and 739 (S2A+S2B). The monthly mean values are: 37 (S2A), 36.9 (S2B) and 73.9 (S2A+S2B). The monthly number of acquisitions are comparable between S2A and S2B since the missions are part of the same constellation and have similar orbit configuration.

The number of acquisition available for each tile is reported in Table 1. The mean number of products per tile is 62 for either S2A and S2B and 123 for S2A+S2B. Slithly variation in the number of products among tile is due to the overlap between swaths from adjacent orbits and to different viewing conditions. In addition, the archive may miss some products due to acquisition errors or platforms maintenance.

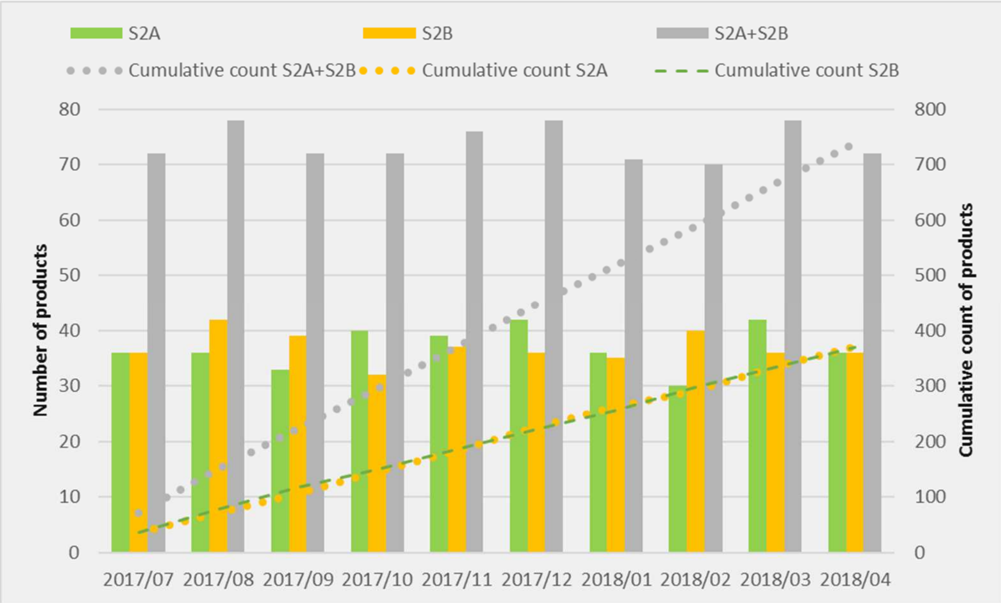

Figure 2- Monthly number of products available from the single Sentinel-2 missions and their combination (S2A+S2B) and cumulative number of products for the time range of analysis (1 July 2017-30 April 2018).

3.2 Mean revisit time

Table 1 reports, for each mission and their combination, the number of products and the mean revisit time (days) by considering each one of the six tiles covering the study area. The mean revisit time is computed as ratio between the total number of days in the time range (300) and the number of product per tile. Both S2A and S2B have a mean revisit time lower than than 5 days per tile, while the constellation (S2A+S2B) have a revisit time of about 2.5 days per tile. These values are lower than the nominal revisit frequency of the constellation (5 days) at the latitude of the study area due to overlap of acquisitions over the tiles.

Table 1 - Average revisit time for the time range of analysis (1 July 2017-30 April 2018) for the single Sentinel-2 missions and for their combination (S2A+S2B) computed for each tile of the study area.

Tile	Number of products			Mean revisit time (days)		
	S2A	S2B	S2A+S2B	S2A	S2B	S2A+S2B
33SWC	61	62	123	4.9	4.8	2.4
33SWD	63	61	124	4.8	4.9	2.4
33SXC	61	61	122	4.9	4.9	2.5
33SXD	63	62	125	4.8	4.8	2.4
33TWE	61	61	122	4.9	4.9	2.5
33TXE	61	62	123	4.9	4.8	2.4

3.3 Data volume

Figure 3 depicts the storage size required by the total number of products available monthly and for for the whole time range of analysis within the study area. Here the constellation is considered (S2A+S2B) and data volume is expressed in Gigabytes. Total size of the available products slithly changes across months ranging from 32 to 38 GB ca. The cumulated size of all the products is about 352 GB. Monthly variation of the size depends on the different number of products available and on the extent of the area covered by each acquisition

since some of them cover partially the tiles. These figures provide a first assessment of the volume of the data to be managed for monitoring an Italian region taking the study area as example.

Since terrestrial monitoring is concerned with the analysis of changes over time to understand the various phenomena (e.g. land use/cove change), a better appraisal of the data volume to be handled over time is useful. To this end, a projection of the numer of products acquired over the study area and their total size is done by considering the constellation (S2A+S2B). Figure 4 depicts the figures for the current exercise (10 months) and for three additional time ranges: 1, 2 and 5 years. The relevant amount of information to be managed is clearly visible from the graph where for a monitoring range of 5 years the products to be managed nearly reach the threshold of 8,000 (7,981) with a total size of 2,112 GB.

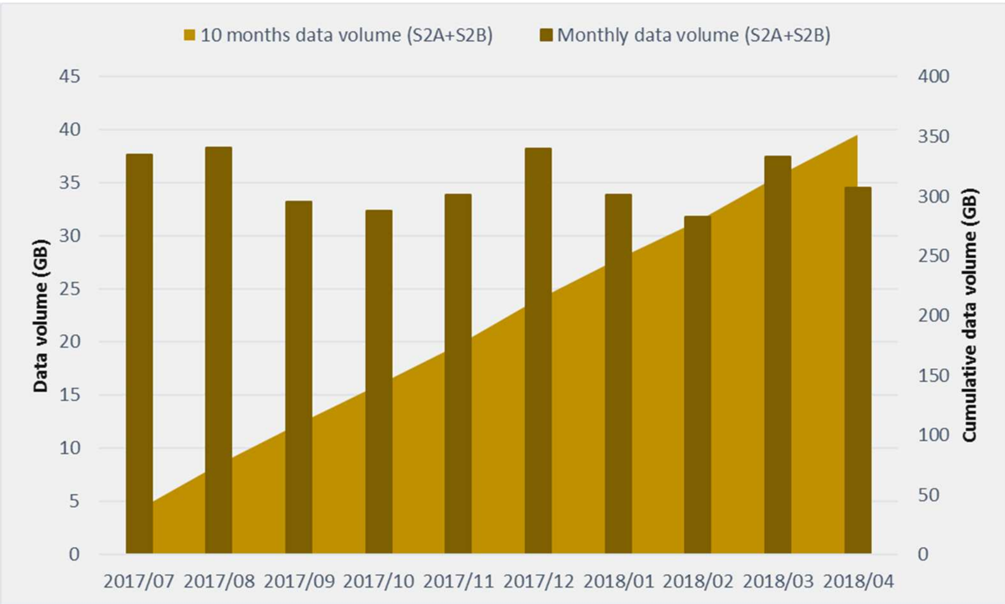

Figure 3- Monthly size of the products and cumulated size over the 10 moths time range of the constellation (S2A+S2B) within the study area.

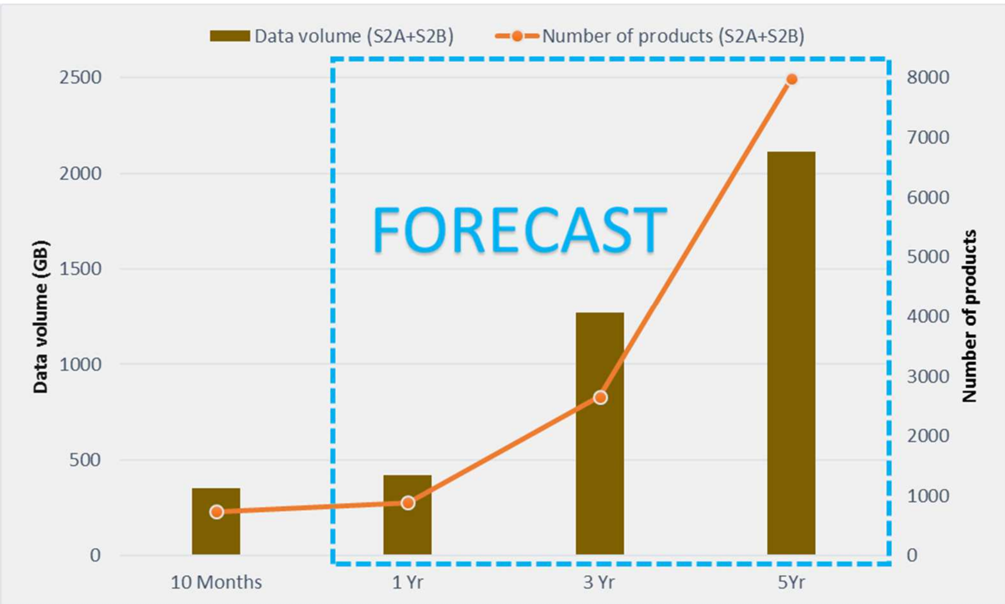

Figure 4- Number of products and size (Gigabytes) acquired by the constellation (S2A+S2B) within the study area. Data are relative to the 10 months time range and to three time range projections: 1, 2 and 5 years.

The metrics used highlight clearly the amount of data that are available for monitoring the study area. It appears clearly how the number of images and therefore the revisit time increase when the twin Sentinel-2 satellites are considered.

The minimum revisit time allows certainly to perform powerful terrestrial monitoring by capturing the dynamics of several phenomena (e.g. vegetation changes at different development stages). Monitoring capabilities are at the same time relevant due to the intrinsic characteristics of the platform: high spatial resolution (up to 10 m) and a high spectral resolution (13 different spectral bands from visible to short wave infrared).

Data reduction approaches can be applied to addrees the issue related to data storage. For instance, by spatial and spectral subsets of the images suitable for a given application or with approaches for finding the least amount of datasets maximizing the information required by the analysis to be carried out. Algorithms can be devised and applied to optimize the selection of datasets to be acquired decreasing the processing and data transfer requests. Data reduction based on clear sky conditions is another example. In this case, the useless images with a significant cloud coverage are removed from the dataset reducing strongly the amount of data to be treated. Studies reported nearly 20 times the rate of reduction of the average number of images per MGRS globally for the Sentinel-2A acquisitions for a 9 months period (Kempeneers et al., 2017)

Other initiative have been launched recently to address the Big Data challenge connected with the vast amount of Copernicus data. The Copernicus Data and Information Access Services (DIAS) is one of the most relevant that will allow users to access easely to the Copernicus data and stimulate innovation and creation of new business models based on EO data. DIAS, launched by European Commission in 2017, is a cloud-based architecture that will deliver data and information access and processing tools allowing a better user uptake and resolving the issues related to download and storage of data.

3.4 Future developments

There are some caveats to notice in the analysis performed herein that should require further evaluation to achieve a better understanding of the monitoring capabilities:

1. Cloud coverage was not considered. Further insights can be derived by considering in the analysis the percentage of cloud coverage of each product, the information can be derived from the products metadata.
2. Extracting the comprehensive set products metadata would allow to gather more information about the actual territorial coverage of each scene. In fact, each MGRS tile have several acquisition with full and partial coverage depending on the orbit of the satellites. In addition, other parameters could be retrieved for a deeper analysis such as cloud coverage, level of quality of the acquisition, etc.
3. Only products with processing Level-1C were extracted from the catalogue. A full accounting of the data volume distributed by the Copernicus Open Access Hub can be carried out by considering the additional products with atmospheric correction (Level-2A).

4. Conclusions

This research reported some evaluation on the current terrestrial monitoring capabilities provided with the new Sentinel-2 satellites from the Europen Copernicus Programme. The analysis carried out considered the amount of multispectral data available for an Italian region acquired by the twin satellites S2A and S2B. The revisit time computed, with some simplifications, shows that it is possible to use several images per week achieving monitoring capabilities suitable for capturing phenomena with high temporal dynamics. Such capabilities are very useful for different domains such as agriculture where will be possible to monitor several processes and to evaluate specific indicators related to the crop development and the sustainable use of input resources.

Novel and powerful applications are expected to be developed in the near future thanks also to the spatial, spectral and radiometric characteristics of Sentinel-2 coupled with the open access policy of the data. On the

other hand, the marked terrestrial monitoring opportunities are challenged by a never-ending flow of data, as it is happening in many sectors, and we are facing with a data deluge to be addressed by finding the best way to tap it (Economist, T., 2010).

Acknowledgements: This research was funded by the project Rete Rurale Nazionale 2014-2020 Scheda "Banda Ultralarga".

5. References

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. Wired magazine, 16(7), 16-07.

Copernicus Open Access Hub, (2018). https://scihub.copernicus.eu/

Copernicus Programme, (2018). Copernicus Land Monitoring Service. https://land.copernicus.eu/

Kempeneers, P., & Soille, P. (2017). Optimizing Sentinel-2 image selection in a Big Data context. Big Earth Data, 1(1-2), 145-158.

Economist, T. (2010). The data deluge. Special Supplement.

ESA (2015). Sentinel-2 User Handbook; Revision 2; ESA Standard Document; ESA: Paris,

France, 2015; 64p.

ESA (2018a). About Copernicus. https://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus

ESA (2018b) - Sentinel High Level Operations Plan (HLOP). Retrieved from: https://sentinels.copernicus.eu/documents/247904/685154/Sentinel_High_Level_Operations_Plan

Li, J., & Roy, D. P. (2017). A Global Analysis of Sentinel-2A, Sentinel-2B and Landsat-8 Data Revisit Intervals and Implications for Terrestrial Monitoring. Remote Sensing, 9(9), 902.

European Commission (2016). Study to examine the socio-economic impact of Copernicus in the EU. DOI: 10.2873/01661. https://publications.europa.eu/en/publication-detail/-/publication/97a5cf70-aa5f-11e6-aab7-01aa75ed71a1

R Development Core Team (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Whitcraft, A. K., Becker-Reshef, I., Killough, B. D., & Justice, C. O. (2015). Meeting earth observation requirements for global agricultural monitoring: An evaluation of the revisit capabilities of current and planned moderate resolution optical earth observing missions. *Remote Sensing*, *7*(2), 1482-1503.