EXPLOITATION OF ENVISAT ASAR AND SENTINEL-1 SAR DATA IN SUPPORT OF CARBON AND WATER CYCLE STUDIES

Maurizio Santoro[1], Oliver Cartus[1], Andreas Wiesmann[1], Urs Wegmüller[1], Josef Kellndorfer[2], Christiane Schmullius[3], Pierre Defourny[4], Olivier Arino[5], Marcus Engdahl[5], Frank Martin Seifert[5]

[1] GAMMA Remote Sensing AG, Gümligen, Switzerland

[2] Earth Big Data LLC, Woods Hole, USA

[3] Friedrich-Schiller University, Department of Earth Observation, Jena, Germany

[4] Université Catholique de Louvain, Earth and Life Science Institute, Louvain-la-Neuve, Belgium

[5] ESA ESRIN, Frascati, Italy

ABSTRACT

Spaceborne SAR dataset have been available for almost 30 years. C-band data records are the longest available; in particular data records from Envisat ASAR and Sentinel-1 are publically available, thus fostering the development of large-scale data products. In this paper, we review the processing behind the generation of SAR data stacks and related data products in support of applications in the domain of the carbon cycle and the water cycle. C-band SAR data were found to advance spatial detail and accuracy of estimates with respect to existing datasets; nonetheless, there are clear limitations that can be overcome with the synergy with other sources of observations from space.

Index Terms— SAR, Envisat ASAR, Sentinel-1, data preservation, data products.

1. INTRODUCTION

Synthetic Aperture Radar (SAR) data have been collected in regular manner during the last three decades. The independence of image acquisition from cloud cover and solar illumination allows accurate planning of frequency of observations in order to support mapping based on multi-temporal observations and monitoring to track dynamics. C-band satellites operated by ESA are the spaceborne SAR missions with the longest time record of observations (since 1991 with interruptions in 2001 and 2013). In addition, the data since 2002 are publically available, favoring the development of innovative mapping techniques and the continuation of existing mapping endeavors.

During the Envisat mission, the Advanced SAR (ASAR) was operated in multiple modes; in particular the ScanSAR mode provided repeated observations at moderate and coarse resolution (150 m and 1,000 m). For example, almost daily observations were collected over polar and boreal regions. The redundancy of observations has been exploited in studies supporting the assessment of variables of the carbon [1] and the water cycle [2, 3] at regional scale. The importance of repeated observations has been implemented in the observation strategy of the Sentinel-1 mission. With respect to the Envisat mission, Sentinel-1 observations are more targeted to exploit as much as possible the information contained in the SAR data with respect to the thematic applications deemed feasible with C-band.

In this paper, we report on achievements with the records of Envisat ASAR and Sentinel-1 observations in support of the carbon and water cycle. Such achievements are based on the exploitation of multi-year observations of the SAR backscattered intensity available through data archives at the European Space Agency and the Alaska Satellite Facility.

2. SAR DATA PROCESSING ON GRID AND CLOUD COMPUTING FACILITIES

Classification and retrieval algorithms based on SAR data require that these are calibrated, co-registered and filtered for speckle noise. As data provided by space agencies or data archives are not available in such format, a processing scheme starting from the original data is required. In addition, such processing scheme needs to be performing on large datasets (software and hardware requirement). With the GAMMA Software [4], it was possible to easily implement the processing scheme to achieve the required data products ensuring an overall fast processing turnaround with high quality data products certified by flags and processing statistics.

Back in 2010, the Grid Processing on Demand (G-POD) facility at ESA was the first of its kind to provide access to long-term archives of SAR images and a processing facility on which a processing sequence could be easily implemented. G-POD was set up to be a generic GRID-based operational environment allowing access to processing facilities closer to Earth Observation (EO) data stored locally. For this, software code could be implemented on the platform to easily access the data and run the code.

doi: 10.2760/383579

On G-POD, Envisat ASAR images acquired at moderate resolution (i.e., 75 m and 150 m) over all land masses except Antarctica between 2005 and 2012 were processed (more than 200,000 images) to form stacks of co-registered images of the SAR backscatter. The objective was to have at hand time series of observations to be used in multiple thematic applications. In a second exercise, all ASAR images acquired between October 2009 and February 2011 (30,000 images) were processed to 1,000 m pixel size for the scope of generating a map of forest biomass of the northern hemisphere representative for the year 2010.

For the processing, a Computing Element consisting of 10 processing nodes and based in ESRIN was made available. The GAMMA software was installed on the platform and processing chains consisting of commands of the software arranged in batch scripts were implemented on G-POD. A user interface allowed selection of data to be processed and managing the processing queue. Pre-processing consisted of data ingestion, calibration, terrain geocoding to a geographic coordinate system, tiling in windows of 1° × 1° (for 150 m data) or 2° × 2° (for 1,000 m) and multi-channel speckle filter. Because of limited resources available at that time, a processing task could at most handle 100 ASAR moderate resolution images or 1,000 ASAR coarse resolution images. Processing of the global ASAR dataset at moderate resolution resulted in over 2,000 tasks. Data selection for each tasks was extremely cumbersome because of redundancies in the ASAR archives at ESA, requiring manual de-selection of copies of an original image dataset. The pre-processed data consisted of almost 40 Terabytes of data including images of the SAR backscatter, the local incidence angle and the pixel area. Data were stored locally on external hard disks and shipped to GAMMA once full.

The G-POD exercise served to lift such a system to cope with the demands of big data processing, teaching a number of lessons to both service providers and users.

With the increased availability of grid and cloud computing facilities responding to the need of the larger throughput of data, e.g., by spaceborne remote sensing platforms, we utilized Earth Big Data's Cloud Processing system to process time series data of Sentinel-1 images covering West Africa, acquired between October 2014 and May 2017 (approximately 6,300 images). Earth Big Data LLC (EBD) has developed a fully automated processing solution that ties SAR processing software (e.g., GAMMA) and geospatial open source processing tools (like GDAL, python, R, Octave, etc.) into a scalable cloud processing environment utilizing Amazon Web Services (AWS).

Backbone of the processing is the EBD Software for Earth Big Data Processing, Prediction Modeling and Organization (SEPPO). At its core, SEPPO features a flexible "recipe processor" that can ingest a range of data from various cloud resource endpoints at run time and can be coupled with the appropriate machine size of the AWS elastic compute instance fleet. Recipes are stand-alone processing algorithms that assume an input data stack to be available in a working directory in which the recipe performs all computations on the data. At conclusion of a recipe run, simple rules for transfer of the desired end products to permanent or cached cloud storage is performed. Data input and transfer to the final storage is implemented to work with standardized URL prefix strings that determine where input and output data are located including transfers via scp or ftp protocols. For example the "s3://" prefix determines to retrieve from or put data to the AWS simple storage system s3 with AWS tools. The "http(s)://" prefix determines endpoints using the transfer protocols and software, including credential verification to obtain data from the respective URL location.

For the processing of the 6,300 raw Sentinel-1 GRD frames for this task, we designed a suite of recipes that would first perform radiometric terrain correction (RTC) based on the GAMMA software. This recipe obtained data from the Alaska Satellite Facility (ASF) obtaining the appropriate 'http://' URL's from a spatial query of EBD's cloud-based Spatial Metadata Database (SMDDB). The SMDDB is updated nightly with holdings of data in the ASF archive and allows flexible query and intersections with region of interest data sets. Each of the 6,300 scenes was thus coupled with the appropriate processing recipe and required machine configuration, and submitted to the SEPPO implemented cloud processing queuing system. After queuing, 30 compute instances were launched in parallel to consume the processing queue. We limited the number of compute instances to 30 after consultation with the ASF to optimize data access rates. Recent trials suggest that parallel access with as many as 100 processors become quite feasible. As ASF is in the process of moving data holdings to the AWS s3 store in the US-East-1 availability zone, we were able to achieve for the bulk of the data transfer rates of up to 80MB/s, which allowed us to access full Sentinel-1 GRD scenes of about 1GB in size in less than 20 seconds for a scene. RTC processing was queued to progress from West to East. All RTC products were stored in a AWS s3 scratch space.

A feature of the RTC processing recipe was the generation of quicklook data at reduced resolution. Throughout the RTC processing, quicklook mosaics of entire Sentinel-1 strips were generated automatically and visually inspected for potential problems (Figure 1). Except for a handful of inaccessible GRD products at ASF, virtually no issues with the RTC processing was found. In parallel to the runs for RTC processing, we deployed a developed recipe for Sentinel-1 time series processing that included subsetting of Sentinel-1 scenes into 1° × 1° geographic tiles of 0.000181818 arcsecond resolution (ca. 20 meters) covering the target region. This recipe performed multi-temporal speckle filtering on the subsetted data, incidence angle map generation for each distinct Sentinel-1 path

doi: 10.2760/383579

covering the tile, and output of the filtered time series and incidence angle stack to a zipped s3 archive file.

A total of 328 tiles covered the region of interest (Figure 2). For tile processing, larger AWS compute instances were required and requested via SEPPO's queuing system. The consumption of the tile processing queue was coordinated via SEPPO's implemented process completion dependency checks with the processing of the RTC products. As such, once all RTC products pertaining to the coverage of a 1° × 1° tile (as determined from SMDDB spatial intersection queries) were completed, tile processing of a respective tile kicked in.

To match the progress of RTC processing with tile processing, we deployed 10 tile recipe processors in parallel, each storing results in a zip data archive on AWS s3. After completion of the 328 tiles, dedicated data access URLs were generated and securely shared with the collaborators at Gamma Remote Sensing for data download. Total production time from RTC, processing, quicklook inspection, and tile processing was about 88 hours. Total input raw data volume of GRD scenes was about 6.2 Terabytes. Volume of final tile products was about 2.5 Terabytes.

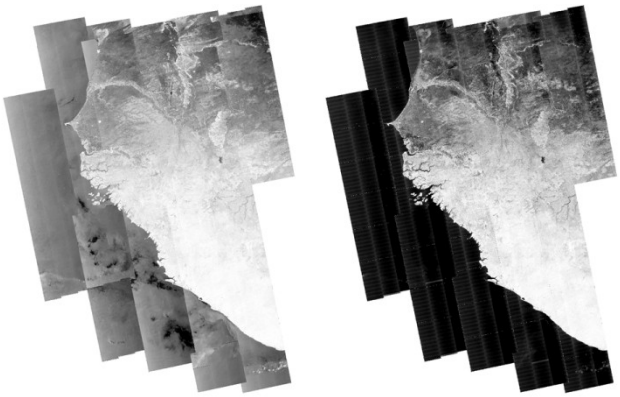


Figure 1. Strips of Sentinel-1 RTC Quicklook data for quick visual inspection of data quality. Left: VV polarization; Right: VH polarization.

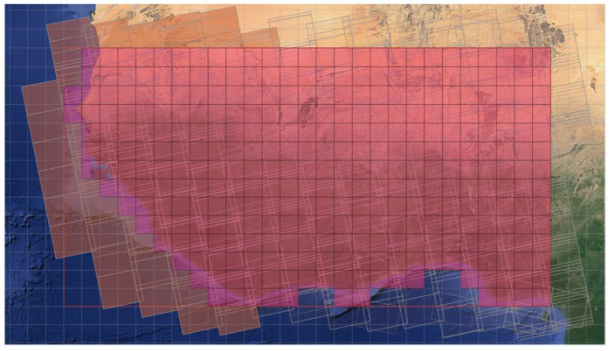


Figure 2. Tiles (red) in the target region and Sentinel-1 frame coverage (gray outlines). Orange colors show the progress of processing from east to west.

3. FOREST BIOMASS RETRIEVAL

The C-band SAR backscatter is not the most suited observable to retrieve forest biomass given its weak sensitivity in high biomass. Nonetheless, it was shown that combining hundreds of observations it is possible to improve the retrieval accuracy beyond what achieved with a single observation [1]. The BIOMASAR retrieval algorithm was therefore developed to implement such approach so that ultimately biomass is estimated from a set of hyper-temporal C-band SAR observations [1]. The algorithm was applied to the Envisat ASAR dataset acquired on a 17-months' time period about the year 2010 and processed on G-POD to retrieve forest biomass of the northern hemisphere [5]. This data product (Figure 3) is unique in terms of coverage, spatial detail (1,000 m spatial resolution) and thematic consistency, allowing assessments of biophysical processes of boreal and temperate forests [6]. The retrieval has recently being extended to tropical and sub-tropical forests to support the generation of a global dataset of forest biomass for the year 2010.

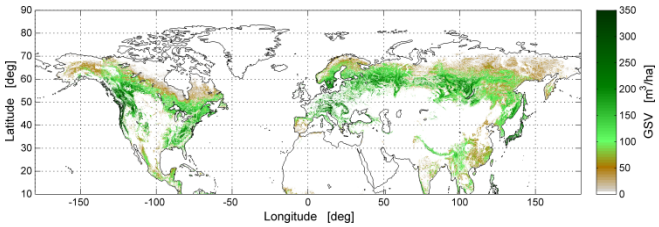


Figure 3. Estimates of forest biomass expressed as growing stock volume (GSV; unit m^3/ha) obtained from a 17 months dataset of Envisat ASAR backscatter observations around the year 2010 (reproduced from [5]).

4. MAPPING AND MONITORING OF INLAND WATER BODIES

Detection of water bodies in C-band SAR images is rather straightforward because of the very low backscattered signal. Nonetheless, this feature is not unique to water surfaces as sand dunes, very arid terrain and wet snow present the same signature. To reduce the water commission error, additional features need to be exploited. The time series of observations processed for the period 2005-2012 (only high- and moderate-resolution modes) allowed the generation of multi-temporal metrics with better separability between water and other land surface types. Ultimately, such SAR dataset was used to generate an indicator of water bodies [7] by means of a simple thresholding algorithm. The water bodies indicator with a spatial resolution of 150 m was selected as major dataset to support the generation of a global dataset of inland water bodies representative for the year 2010 [8]. The Climate Change Initiative (CCI) Land Cover (LC) water bodies product is a self-standing dataset of the CCI-LC climate data package (Figure 4).

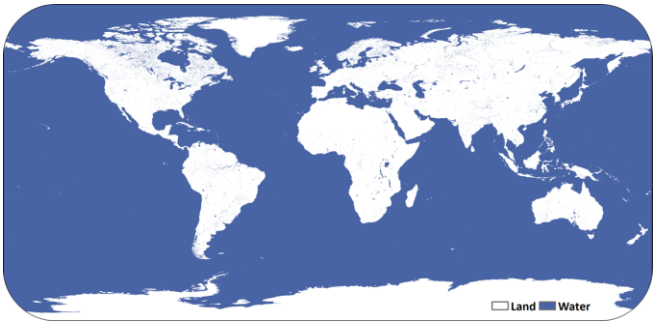


Figure 4. CCI Water Body dataset available at http://maps.elie.ucl.ac.be/CCI/viewer [8].

The entire time series of ASAR observations (from high- to low-resolution) has recently been used to generate an additional data product related to the water cycle. A climatology of water occurrence for 2005-2012 was generated with a spatial resolution of 1,000 m (Figure 5). The dataset portrays variability of water cover as seen by the ASAR instrument and is currently being assessed in terms of its thematic accuracy.



Figure 5. Occurrence of open standing water for the last week of May for the time period 2005-2012 in North America.

The data record of Sentinel-1 observations over West Africa has a unique feature of one observation every 12 days ensuring the possibility of building on the experience of mapping water bodies with Envisat ASAR data. Again, multi-temporal metrics have been exploited but a more sophisticated approach has been selected (bagged decision trees) to cope with the variability of the backscattered signal throughout the year. The Sentinel-1 water bodies dataset has a spatial resolution of 20 m and provides a complete representation of open water bodies of West Africa, not yet achieved with optical imagery (Figure 6).

5. ACKNOWLEDGMENTS

The work here presented was undertaken within the BIOMASAR (ESRIN contract No. 21892/08/I-EC), the GlobBiomass (ESRIN contract No. 4000113100/14/I-NB) and the Climate Change Initiative Land Cover (CCI-LC) (ESRIN contract No. 4000101774/10/I-LG) projects, all sponsored by the European Space Agency (ESA). Access to the G-POD processing facilities was possible through ESA's Category 1 Project ID 9209 "G-POD processing of ASAR Wide Swath imagery for multi-purpose applications".

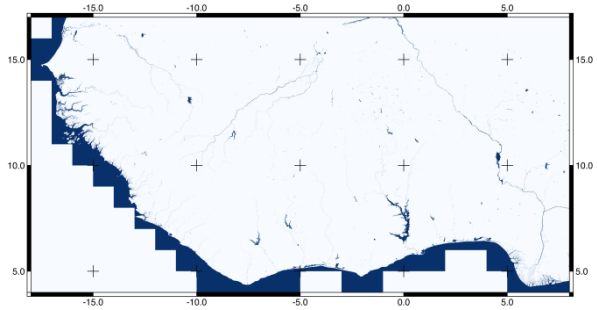


Figure 6. Map of water bodies of West Africa generated from time series of Sentinel-1 SAR backscatter observations.

6. REFERENCES

[1] M. Santoro, C. Beer, O. Cartus, C. Schmullius, A. Shvidenko, I. McCallum, U. Wegmüller, and A. Wiesmann, "Retrieval of growing stock volume in boreal forest using hyper-temporal series of Envisat ASAR ScanSAR backscatter measurements," *Remote Sens. Environ*., vol. 115, pp. 490-507, 2011.

[2] A. Bartsch, R. Kidd, C. Pathe, W. Wagner, and K. Scipal, "Satellite radar imagery for monitoring inland wetlands in boreal and sub-arctic environments," *Journal of Aquatic Conservation: Marine and Freshwater Ecosystems*, vol. 17, pp. 305-317, 2007.

[3] D. O'Grady, M. Leblanc, and A. Bass, "The use of radar satellite data from multiple incidence angles improves surface water mapping," Remote Sens. Environ., vol. 140, pp. 652-664, 2014.

[4] C. Werner, U. Wegmüller, T. Strozzi, and A. Wiesmann, "GAMMA SAR and interferometric processing software," *Proc. ERS-Envisat Symposium*, Gothenburg, 16-20 October, 2000.

[5] M. Santoro, A. Beaudoin, C. Beer, O. Cartus, J. E. S. Fransson, R. J. Hall, C. Pathe, D. Schepaschenko, C. Schmullius, A. Shvidenko, M. Thurner, and U. Wegmüller, "Forest growing stock volume of the northern hemisphere: spatially explicit estimates for 2010 derived from Envisat ASAR data," *Remote Sens. Environ.,* vol. 168, pp. 316-334, 2015.

[6] M. Thurner, C. Beer, M. Santoro, N. Carvalhais, T. Wutzler, D. Schepaschenko, A. Shvidenko, E. Kompter, B. Ahrens, S. R. Levick, and C. Schmullius, "Carbon stock and density of northern boreal and temperate forests," *Global Ecol. Biogeogr*., vol. 23, pp. 297-310, 2014.

[7] M. Santoro, U. Wegmuller, C. Lamarche, S. Bontemps, P. Defourny, and O. Arino, "Strengths and weaknesses of multi-year Envisat ASAR backscatter measurements to map permanent open water bodies at global scale," *Remote Sens. Environ*., vol. 171, pp. 185-201, 2015.

[8] C. Lamarche, M. Santoro, S. Bontemps, R. d'Andrimont, J. Radoux, L. Giustarini, C. Brockmann, J. Wevers, P. Defourny, and O. Arino, "Compilation and validation of SAR and optical data products for a complete and global map of inland/ocean water tailored to the climate modeling community". *Remote Sens.*, vol. 9, 36, 2017.