INTRODUCING A GLOBAL DATASET OF OPEN PERMANENT WATER BODIES

Maurizio Santoro [(1)], Céline Lamarche [(2)], Sophie Bontemps [(2)], Urs Wegmüller [(1)], Vasileios Kalogirou [(3,4)], Olivier Arino [(3)], Pierre Defourny [(2)]

[(1)] Gamma Remote Sensing, Worbstrasse 225, 3073 Gümligen, Switzerland, E-Mail: santoro@gamma-rs.ch, wegmuller@gamma-rs.ch
[(2)] Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium, Email: celine.lamarche@uclouvain.be, sophie.bontemps@uclouvain.be, pierre.defourny@uclouvain.be
[(3)] European Space Agency, Frascati, Italy, Email: olivier.arino@esa.int
[(4)] EU Satellite Centre, Spain, Email: Vasileios.kalogirou@satcen.europa.eu

ABSTRACT

This paper introduces a 300-m global map of open permanent water bodies derived from multi-temporal synthetic aperture radar (SAR) data. The SAR dataset consisted of images of the radar backscatter acquired by Envisat Advanced SAR (ASAR) in Wide Swath Mode (WSM, 150 m spatial resolution) between 2005 and 2010. Extended time series of WSM to 2012, Image Mode Medium resolution (IMM) and Global Monitoring Mode (GMM) data have been used to fill gaps. Using as input the temporal variability (TV) of the backscatter and the minimum backscatter (MB), a SAR-based indicator of water bodies (SAR-WBI) has been generated for all continents with a previously validated thresholding algorithm and local refinements. The accuracy of the SAR-WBI is 80%; a threshold of 50% has been used for the land/water fraction in the case of mixed pixels. Correction of inconsistencies with respect to auxiliary datasets, completion of gaps and aggregation to 300 m were applied to obtain the final global water body map referred to as Climate Change Initiative Land Cover Water Body (CCI-LC WB) Product.

1. INTRODUCTION

Poor characterization of inland water bodies in global land cover products [1, 2] or partial coverage in existing water body datasets (e.g., SRTM Water Body Dataset, SWBD) triggered an investigation that exploits synthetic aperture radar (SAR) data to provide up-to-date and reliable information on water bodies extent. To this scope, Envisat Advanced SAR (ASAR) data acquired in the moderate resolution (150 m) Wide Swath Mode (WSM) were considered. Up to daily observations were possible thanks to the strong overlap of swaths of adjacent orbital tracks. The high density of observations allows generating metrics of the temporal variability (TV) of the SAR backscatter and the minimum SAR backscatter (MB), which were found to have unique features (high TV and low MB) with respect to other land cover types [3]. A simple thresholding algorithm in the feature space of TV and MB allowed for accurate detection of open permanent water bodies with respect to land surfaces [3]. Overall Accuracy above 90% was obtained for pure pixels of land and water. When also mixed pixels were taken into account, the accuracy decreased. For a 50% threshold on water fraction, the accuracy was approximately 80%. The main advantage of the detection algorithm is the global classification rule, i.e., the independency from the local land cover types and, therefore, from a set of measurements necessary to calibrate the algorithm. Major limitations are the necessity of at least 10 SAR backscatter observations to guarantee that the TV is different over water and land surfaces, and the frequent omission of shorelines in the case of partial water fraction.

This paper presents a global dataset of open permanent water bodies, obtained from multi-temporal Envisat ASAR backscatter data and the proposed water detection algorithm. This dataset is used as the main source for the water class of the 300m spatial resolution global land cover maps produced in the Climate Change Initiative Land Cover (CCI-LC) project [4]. Fig. 1 shows the processing framework.

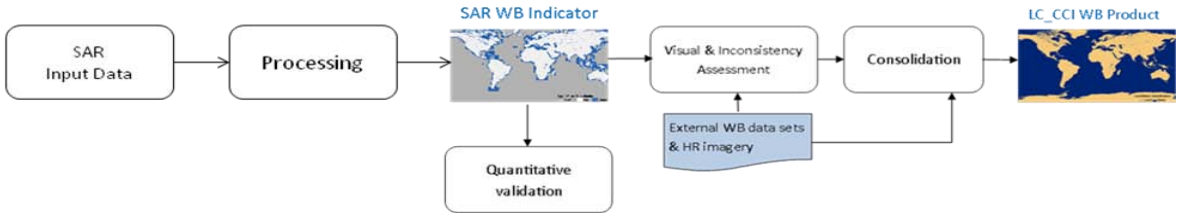

Figure 1. Processing chain to derive the global water body dataset from SAR data.

Proc. 'ESA Living Planet Symposium 2013', Edinburgh, UK
9–13 September 2013 (ESA SP-722, December 2013)

The input SAR data served to generate a first map of water bodies referred to as SAR Water Body Indicator (SAR-WBI). This underwent quantitative validation and a thorough screening for gaps and inconsistencies. With aid of additional water body datasets and high spatial resolution imagery, the SAR-WBI was improved, completed and resampled to generate the final global water body product, herewith referred to as CCI-LC WB Product.

2. SAR DATASET

The primary dataset consisted of images of the radar backscatter acquired by Envisat ASAR in the Wide Swath Mode (WSM) with a spatial resolution of 150 m. The time span of the dataset was maximized to limit gaps in the coverage and obtain everywhere a time series of measurements for reliable estimation of the multi-temporal SAR metrics. As a result, all images acquired over land surfaces between 2005 and 2010 were used to generate a single epoch map of water bodies. The data was available through ESA's Grid Processing on Demand (G-POD) platform (http://gpod.eo.esa.int/). While the coverage of the northern latitudes (> 60°N) and Europe in WSM was outstanding with at least weakly observations, other areas of the world were characterized by patchy acquisitions. Gaps over Japan and Southeast Asia could be filled with WSM data acquired through to the end of the Envisat mission (April 2012). Further gaps (central US, central Asia, China) were filled with Envisat ASAR Image Mode Medium resolution data (IMM, 75 m resolution). Remaining gaps, primarily over South America and Australia, could only be filled with Global Monitoring Mode (GMM, 500 m resolution) data. These data were oversampled to 150 m during the processing. The 500 m resolution of ASAR GMM was sub-optimal in our mapping efforts; still, this dataset allowed complete coverage of all continents. Isolated isles or groups of isles (e.g., in Oceania), south Panama and west Mexico remained unmapped because there are practically no ASAR observation.

The WSM and IMM datasets were processed on G-POD from Level 1P (radar intensity in ground range geometry) to obtain geocoded backscatter measurements at 150 m resolution. GMM gap fillers were processed on local machines depending on the gap to be filled and the quantity of backscatter observations required to achieve a reliable estimate of the multi-temporal metrics.

SAR processing consisted of absolute calibration using factors published by ESA in the image metadata, automated terrain geocoding [5], image tiling, speckle filtering based on a multi-channel approach [6] and normalization of the backscatter to reduce the effect of sloped terrain and different viewing geometries [7]. For optimal management of computing resources, the data were tiled according to a 1° × 1° tiling system. The SAR processing approach is further described in [8]. Major efforts had to be spent on the selection of the ASAR data because of occasional multiple entries of the same dataset in the G-POD archives and the limited flexibility of the G-POD platform to handle very large datasets.

More than 11 Gbyte of SAR backscatter data were produced from approximately 200,000 WSM and IMM images available on G-POD. The GMM dataset consisted of 3 Gbyte of gap fillers. For each pixel, the number of SAR observations, the start and end date of the ASAR dataset were also computed (one layer for WSM+IMM and one layer for GMM data).

3. WATER BODY DETECTION ALGORITHM

The SAR data was fed to the water body detection algorithm on a tile-by-tile basis (Fig. 2). From the multi-temporal SAR backscatter measurements, single multi-year datasets of TV and MB were obtained. The classification followed the thresholding rules in [3]:

- Water if $MB < 3.5 \cdot TV - 28$
- Land if MB > -16 dB or TV < 1.5 dB
-

Automatic labelling as land was applied if the local terrain slope was greater than 10°. The local terrain slope was derived from a global Digital Elevation Model (DEM) based on several DEM datasets [9-12].

The output of the first classification consisted of a map of potential water bodies. Four typologies of incorrect classification were identified at this stage due either to specific land-cover types or to environmental conditions and the distribution in time of the SAR data with respect to these. In turn, area specific refinements were set up (see Fig. 2). The refinement rules were based on careful observation of water commission errors (more frequent) and omission errors (less frequent). Combination of refinement methods in a cascade approach was also considered if a single refinement still presented residual mis-classifications. If none of the refinement methods listed below were successful, manual corrections of large incorrect features were applied based on matching with Google Earth imagery. Only in the case of permanent sea ice in the Arctic, no correction could be applied at this stage.

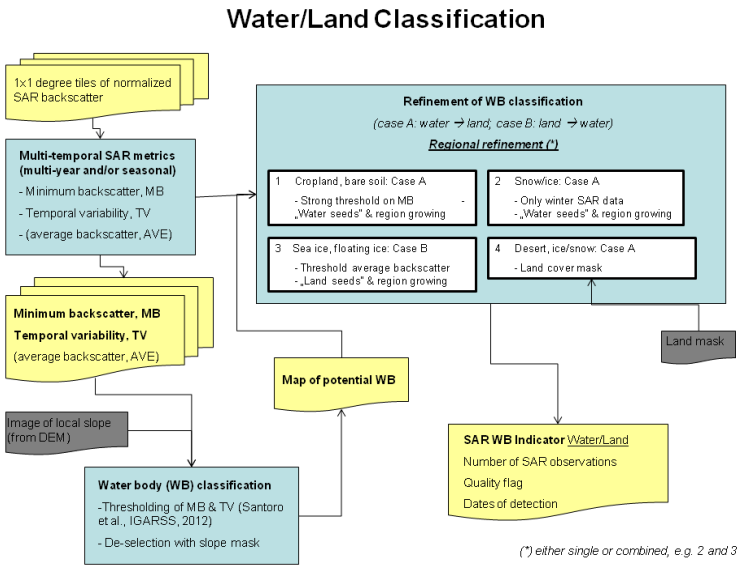

Figure 2. Flowchart of water body classification from multi-temporal SAR data.

1. *Cropland, bare soil*. Frequent coverage of these land cover types during wet periods rather than during dry periods caused false detection of land as water. Classification was refined by setting a more stringent threshold on the MB; pixels labelled here as water then served as seeds to expand the detection with a region growing approach. This served to correct for false detections of land as water.
2. *Snow covered areas*. When the time series of ASAR measurements included dates with wet snow conditions, TV and MB behaved similarly to water. Classification was refined by restricting observations to winter periods only. If needed, pixels labelled here as water then served as seeds to expand the detection with a region growing approach. This served to correct for false detections of land as water.
3. *Desert and cold or arid regions*. The backscatter in such areas is characterized by significant variability because of the large range of incidence angles in WSM data. Restricting to data acquired with a single incidence angle was not possible because of the sparse ASAR WSM datasets of backscatter over desert and arid regions. Classification could only be improved by re-labelling water as land according to desert or bare soil classes in the GlobCover 2005 dataset.
4. *Sea ice*. Areas with long periods of sea ice (polar regions) presented features of the TV and MB metrics over the multi-year dataset similar to land. However, the average backscatter was different over sea ice and land. The classification was therefore refined by applying a threshold of -14 dB on the average backscatter. Pixels labelled here as land then served as seeds to expand the detection with a region growing approach. This served to correct for false detections of water as land.

Fig. 3 shows an example of refinement for the part of the west coast of Greenland. Snow-covered areas were detected as water body because of wet snow conditions that caused high TV and very low MB, typical of water. Restricting the classification to winter data removed cases of wet snow conditions thus avoiding false detection of land as water.

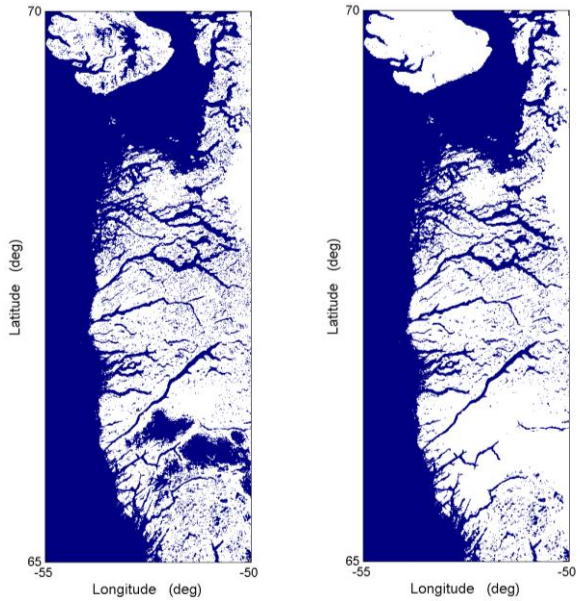

Figure 3. Map of potential water bodies (left) and SAR-WBI (right) for the west coast of Greenland (blue: water, white: land).

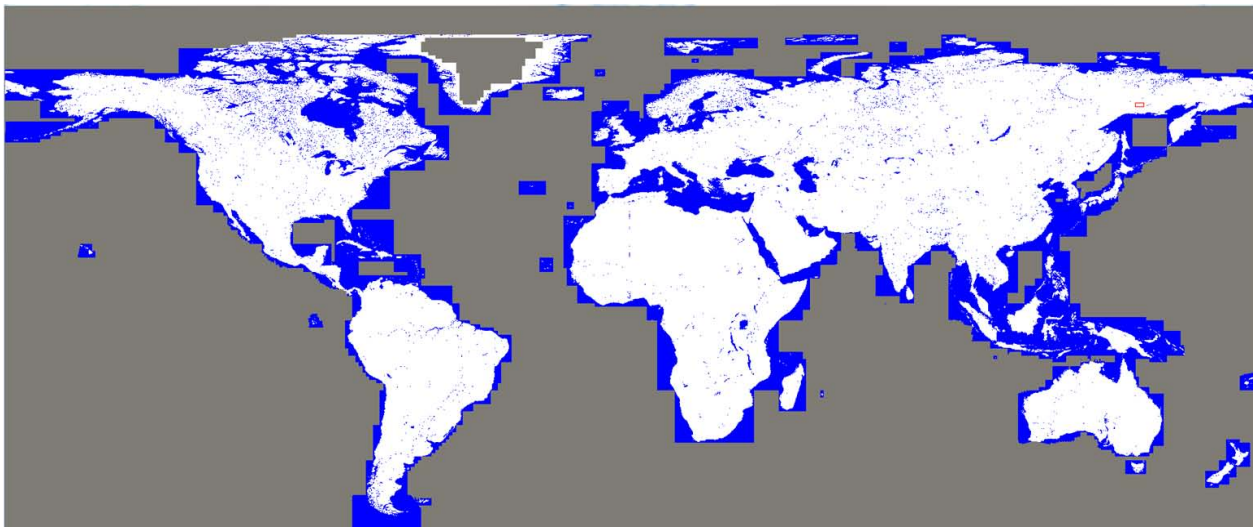

Figure 4. SAR-WBI (blue: water, white: land, grey: no data).

The refinements were applied on a tile-by-tile basis and were chosen depending on the type of error characterizing each single tile. As a consequence, the SAR-WBI had still errors when we failed in identifying the reason for an omission or a commission error.

4. SAR WATER BODY INDICATOR

The SAR-WBI is a static map of stable open water bodies at 150m spatial resolution with a nearly global extent between 84°N and 60°S on the basis on the selected SAR data. The SAR-WBI is shown in Fig. 4. It covers all continents; small gaps appear in correspondence of south Panama and north Australia because of too few observations available. Several isolated isles could not be mapped because of too few observations (e.g., Polynesia, Kuril and Aleutian isles). The central part of Greenland and Antarctica were not mapped because it was assumed there are no major open permanent water bodies. Because of permanent sea ice, the northernmost latitudes of North America and Greenland were not included in the SAR-WBI because of sea ice-covered areas classified as land.

Visual assessment of the SAR-WBI consisted in a global scan to identify remaining macroscopic inconsistencies followed by comparisons to imagery with high spatial resolution (Bing Maps and Google Earth) and other water bodies products. Finally, a comparison to global or continental ancillary datasets compiled together to form a maximum extent water layer (MOD44W [13], SWBD [14], GLWD 1 and 2 [15], the permanent water bodies of the 250m continental g2_Biopar water bodies detections over Africa and the commercial Global Insight water layer) helped to reveal more subtle inconsistencies. The latter comparison resulted in a "layer of potential inconsistencies" which revealed various types of inconsistencies, described as potential commission and omission errors.

The quality of the SAR-WBI was high in areas characterized by large number of backscatter observations (Europe and north of 60°N). Fig. 5 shows an example for the Eskimo Lakes, northwest Canada. Water bodies larger than twice the ASAR pixel size (300 m) were well detected. Conversely MODIS 250 m land-water mask (MOD44W) and the Global Lakes and Wetlands database (GLWD) either overestimated or underestimated the area covered by water bodies.

For other regions, the SAR-WBI was mostly correct with a few sporadic errors related to the amount of ASAR data available, their temporal distribution and the presence of features for which TV and MB would behave similarly to water.

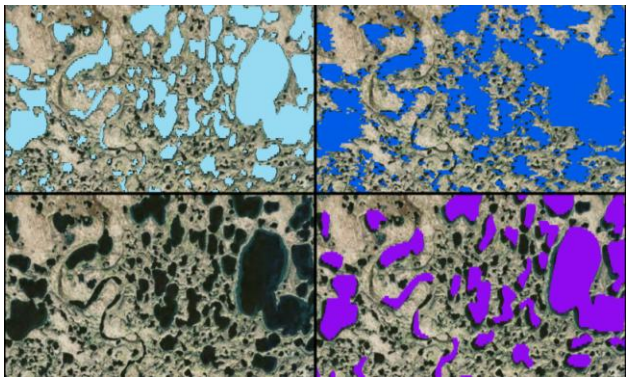

Figure 5. SAR-WBI (top left), MODIS Land/Water Mask (top right), Bing Maps (bottom left and background image in other panels) and the Global Lakes and Wetlands Database (bottom right) for Eskimo Lakes, Canada.

Commission errors typically occurred in correspondence of:

- unprocessed areas
- coastal areas (where GMM data was used)
- inundated areas
- deserts/salars
  (sand dunes and bare surfaces presented TV and MB similar to water because of strong dependency of backscatter on look direction)
- irrigated croplands
  (rice paddies when SAR data was mostly acquired during flooding and growing season)
- sloped terrain in mountainous areas.

Omission errors occurred for

- long-lasting sea ice
- coastal areas (where GMM data was used
- few SAR backscatter observations
- low spatial resolution of the SAR data.

Fig. 6 shows an example of commission and omission errors along coastlines. This was typical of areas where the SAR dataset consisted mostly of GMM data (South America and Australia).

Fig. 7 shows an example of a desert area detected as water. Here, by mistake the regional refinement involving the use of a land cover map to remove correct for false detections of water in deserts and arid areas was not applied.

Quantitative validation of the SAR-WBI consisted of a comparison against data samples (footprints of 150 × 150 m^2) interpreted in high resolution imagery in Google Earth. The 2,232 samples were selected using a stratified random sampling approach based on the requirements of uniform spatial distribution and equal representation of consistencies and inconsistencies with respect to the layer of inconsistencies described above. Experts could base their evaluation of the land/water classification of the sample on the footprints of the pixel. Each footprint was labelled either as “land”, “water > 50%” or “water <50%”. A confusion matrix (Table 1) was obtained based on 2,079 footprints. 75 footprints were removed because they were situated in unprocessed areas of the SAR-WBI and 79 could not be evaluated by the experts in Google Earth. Assuming that the "land" and the "water < 50%" could be grouped under one class, the corresponding classification accuracy was 80%. This result is in line with the assessment done at local sites when using a threshold of 50% on the water cover fraction within a pixel [3]. The major issue with the SAR-WBI appeared to be water omissions in correspondence of coastlines and shorelines while the commission error was more limited.

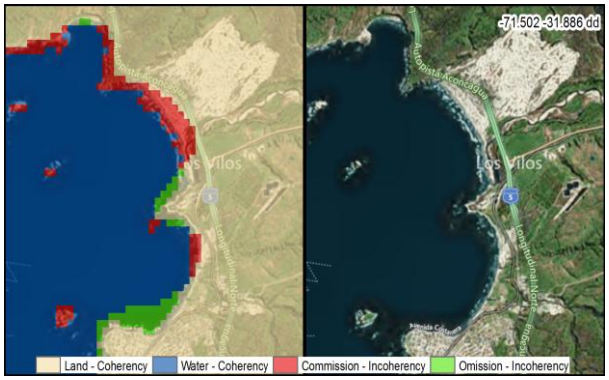

Figure 6. Local inconsistencies in the delineation of coastlines in Chile. Left: SAR-WBI; right: Bing Maps Hybrid.

Figure 7. Illustration of a desert detected as water (commission error in purple) in Peru. Left: SAR-WBI; right: Bing Maps Hybrid.

Table 1. Contingency matrix for "water > 50%" and "no water or water < 50%" class.

Count	Reference			
SAR WBI	no water or water < 50%	water > 50%	SUM	User Acc.
no water	1193	267	1460	82 %
water	148	470	618	76 %
SUM	1341	737	2078	
Prod. Acc.	89 %	64 %		80 %

5. CCI-LC WB PRODUCT

To obtain a truly global water body dataset from the SAR-WBI, the latter was consolidated using an independent dataset made of a combination of existing global water bodies products. This consolidation step consisted in completing unprocessed areas, improving coastline delineations, removing false detection of land as water (commission errors) and replacing false detection of water as land (omission errors).

The location of the corrections was guided by the above-mentioned visual assessment and the layer of potential inconsistencies. The zones of inconsistencies indicate a mismatch between land/water classifications but cannot be considered systematically as errors. In addition, true errors are not systematically related to a particular land cover. This implied a comparison of each discrepancy to high-resolution imagery. To this scope, a web interface allowing quick superimpositions of various raster layers in Google Earth and interactive edition of polygons in a shapefile was built.

Missing islands in the SAR-WBI were added from the SWBD dataset. Antarctica was added from the Antarctic Digital Database Version 6.0 from the Scientific Committee on Antarctic Research (SCAR) (http://www.add.scar.org/index.jsp). As the main source of omission errors was the lack of sufficient SAR observations, a systematic correction was applied when the number of WSM+IMM $\leq$ 15 acquisitions. In this case, the water bodies from the SWBD whose area exceeded 150 meters square were added to the SAR-WBI. Fig. 8 shows some examples of consolidation of the SAR-WBI.

The consolidated SAR-WBI was finally aggregated from 150 m to 300 m to form the final CCI-LC WB Product (Fig. 9). The CCI-LC WB product includes two classes (land and water), giving the repartition of open and permanent water bodies (inland water and oceans) at 300m spatial resolution at global scale. The Coordinate Reference System is a geographic coordinate system based on the WGS84 reference ellipsoid.

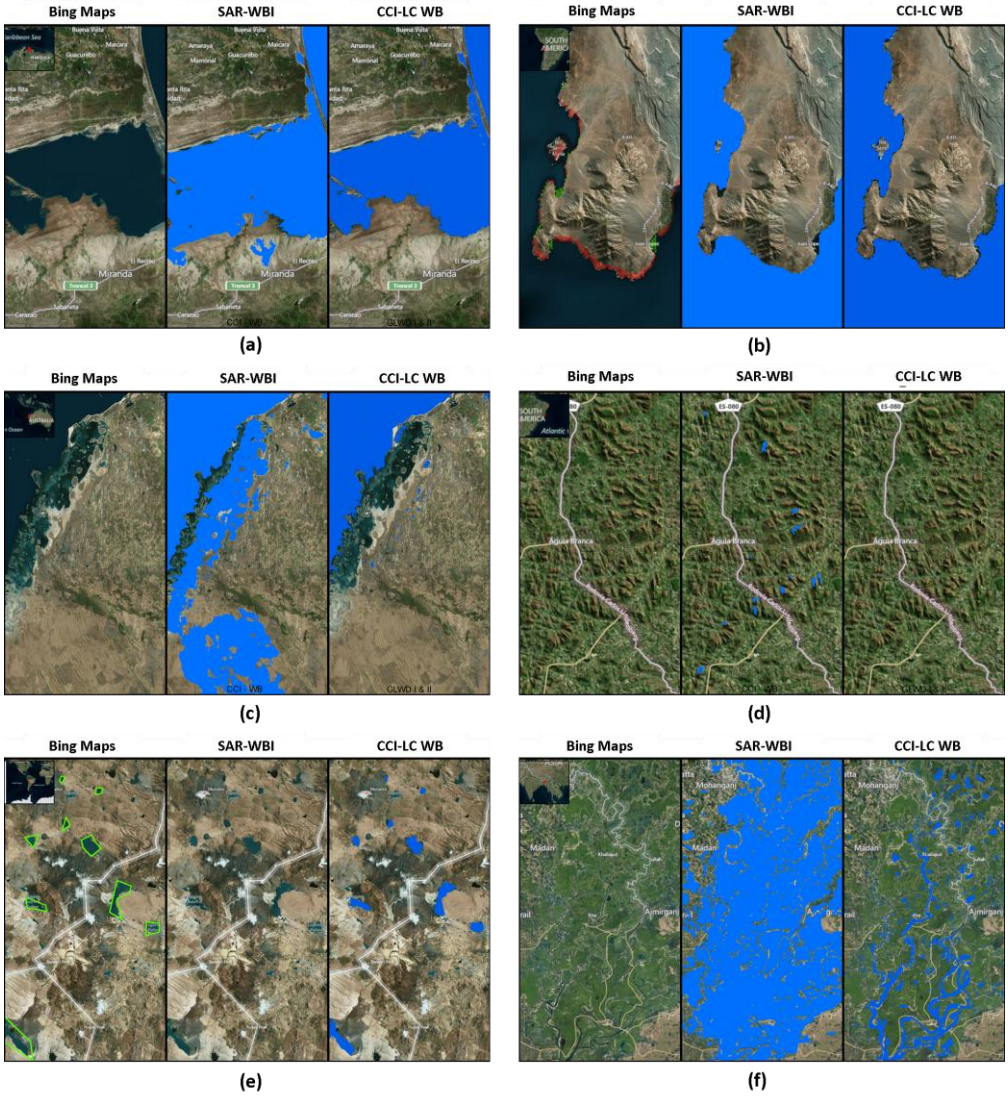

Figure 8. Illustration of various consolidation examples. The SAR-WBI (centre) and the CCI-LC WB product (right) are compared to HR imagery in Bing Maps (left). Figures (a, b, c) show improvements in coastlines delineation, (d) removal of false detections in mountain areas, (e) addition of omitted water bodies and (f) removal of false detections in flooded areas.

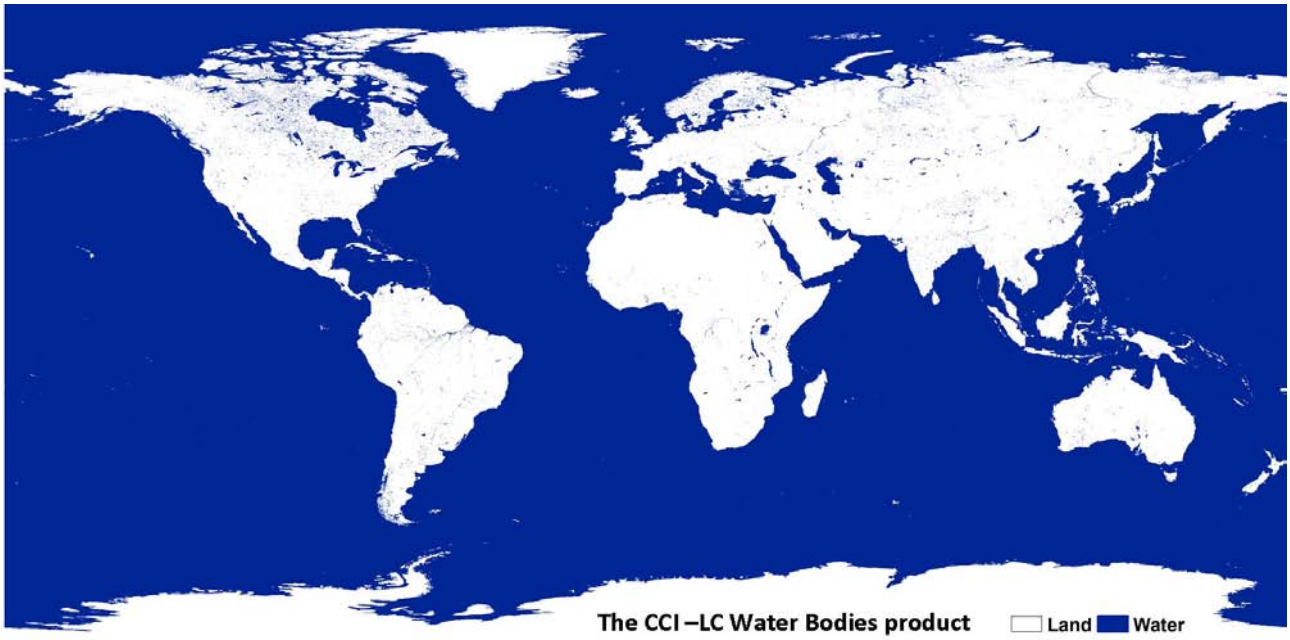

Figure 9. The CCI-LC WB Product.

6. CONCLUSIONS

This paper described the generation of a global water body dataset based on Envisat ASAR multi-temporal images. A more detailed validation protocol of the CCI-LC WB Product is ongoing with the aim of better understanding the sources of uncertainty associated with the product. Various factors that may affect the accuracy of the water body product (e.g. climatology, geophysical characteristics, signal-based information, etc.) will be identified and retrieved over a sample of points globally distributed. Factors will then be compared to the accuracy of the product and univariate and multivariate statistical analyses will be carried out to estimate the dependence of type I and type II errors on the indices. The results of the analysis will also allow contributing to the current SAR knowledge for water detection.

With the SAR dataset available, it was not possible to achieve globally the same classification quality and fall-back on additional data layers was necessary to ensure a gap-free and consistent result. At this stage, the CCI-LC WB Product appears to be complementary to the SWBD. Nonetheless, we expect further improvement at the level of the SAR-WBI when revisiting the refinement rules, which were more based on an empirical approach rather than on a sound and consistent procedure.

7. ACKNOWLDEGMENTS

This work was supported by the European Space Agency within the Climate Change Initiative Land Cover (CCI-LC) project. 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". R. Cuccu, S. Pinto and J. Farres at ESA/ESRIN are acknowledged for implementation of SAR processing chain on G-POD and support during processing.

8. REFERENCES

1. Bicheron, P., Defourny, P., Brockmann, C., Schouten, L., Vancutsem, C., Huc, M., Bontemps, S., Leroy, M., Achard, F., Herold, M., Ranera, F. & Arino, O. (2008). GLOBCOVER – Products Description and Validation Report. Media France.
2. Bartholomé, E. & Belward, A.S. (2005). GLC2000: A new approach to global land cover mapping from Earth Observation data. *Int. J. Remote Sens.* **26**(9), 1959-1977.
3. Santoro, M. & Wegmüller, U. (2012). Multi-temporal SAR metrics applied to map water bodies. In Proc. IGARSS'12, IEEE Publications, Piscataway, NJ, pp 5230-5233.
4. Defourny, P., Bontemps, S., Radoux, J., Van Bogaert, E., Lamarche, C., Achard, F., Mayaux, P., Boettcher, M., Brockmann, C., Kirches, G., Zülkhe, M., Kalogirou, V. & Arino, O. (2013). Consistent global land cover maps for climate modeling communities: Current achievements of the ESA's Land Cover CCI. In Proc. ESA

Living Planet Symposium 2013 (Eds. L. Ouwehand), ESA Communication, DGC-CBE, ESA/ESTEC, Nordwijk, the Netherlands, ESA SP-722 (CD-ROM).

5. Wegmüller, U. (1999). Automated terrain corrected SAR geocoding. In Proc. IGARSS'99, IEEE Publications, Piscataway, NJ, pp 1712-1714.
6. Quegan, S. & Yu, J.J. (2001). Filtering of multichannel SAR images. *IEEE Trans. Geosci. Remote Sensing* **39**(11), 2373-2379.
7. Wiesmann, A., Wegmüller, U., Santoro, M., Strozzi, T. & Werner, C. (2004). Multi-temporal and multi-incidence angle ASAR Wide Swath data for land cover information. In Proc. 4th International Symposium on Retrieval of Bio- and Geophysical Parameters from SAR Data for Land Applications, CD-ROM.
8. Santoro, M., Beer, C., Cartus, O., Schmullius, C., Shvidenko, A., McCallum, I., Wegmüller, U. & Wiesmann, A. (2011). Retrieval of growing stock volume in boreal forest using hyper-temporal series of Envisat ASAR ScanSAR backscatter measurements. *Remote Sens. Environ.* **115**(2), 490-507.
9. Rabus, B., Eineder, M., Roth, A. & Bamler, R. (2003). The Shuttle Radar Topography Mission - A new class of digital elevation models acquired by spaceborne SAR. *ISPRS Journal of Photogrammetry & Remote Sensing* **57**, 241-262.
10. de Ferranti, J. (2009). Digital Elevation Data. http://www.viewfinderpanoramas.org/dem3.html. Accessed on 28 February 2012.
11. Anon. (2007). Canadian Digital Elevation Data. http://www.geobase.ca/geobase/en/index.html. Accessed on 28 February 2012.
12. Anon. 7.5-Min DEM Native Format of the United States. http://www.webgis.com/. Accessed on 28 February 2012.
13. Carroll, M.L., Townshend, J.R., DiMiceli, C.M., Noojipady, P. & Sohlberg, R.A. (2009). A new raster water mask at 250 m resolution. *International Journal of Digital Earth* **2**(4), 291-308.
14. NASA/NGA (2003). SRTM Water Body Data Product Specific Guidance.
15. Lehner, B. & Döll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology* **296**(1-4), 1-22.