ENSURING SPATIAL AND TEMPORAL CONSISTENCIES FOR THE TIME SERIES OF THE COPERNICUS LAND MONITORING PAN-EUROPEAN HIGH RESOLUTION LAYERS

Christophe Sannier, Sophie Villerot, Alexandre Pennec, Alice Lhernould, Clémence Kenner, Antoine Masse

SIRS - CLS Group, 27 rue du Caroussel, 59650 Villeneuve d'Ascq, France

ABSTRACT

Pan-European products assessing the sealed areas, spanning on more than a decade and still in production for 2018, the time series of High Resolution Layer Imperviousness has been making use of multiple sensors, whose data volume is still increasing, in particular with the introduction of Sentinel constellations, at multiple temporal and spatial scales. In this paper, we review the methodologies developed within the HRLs production and enhanced during the H2020 ECoLaSS project to ensure the coherence of this times series, whose updated areas from one date to the next lays within the accuracy specifications.

***Index Terms*—** High Resolution Layers, Sentinel, Copernicus, Change Detection, Time Series, ECoLaSS

1. INTRODUCTION

The urban population in 2014 represented 54% of the overall population, and is expecting to keep rising [1]. According to the World Health Organisation, the global urban population should grow approximately 1.84% per year from 2015 to 2020; this percentage slowly decreasing over the years to reach 1.44% per year between 2025 and 2030.

Despise the impression that the temporal change in urban area does not appear to be significant at the global scale, its impact on the neighbouring forests, agriculture, water systems, through consumption rise, can turned out to be critical. A close monitoring of the urban growth is necessary to ensure a sustainable development [2]. In most developing countries, urban growth is mainly driven by population growth. However, in Europe, population growth no longer increases substantially, but urban areas continue to expand, a phenomenon known as urban sprawl [3].

The Copernicus Land Monitoring (CLMS) is an effort coordinated by the European Environment Agency to produce land cover and land use information, through the CORINE Land Cover (CLC) dataset as well as the five High Resolution Layers (HRL) for each of the specific land cover characteristics: artificial areas, forest areas, grasslands, wetlands and water bodies, that should be soon complemented by a layer of small woody features. The imperviousness (IMP) products quantify the percentage of soil sealing in a status layer for a given year ($\pm$ 1 year) and capture the modifications from the previous status layer to the next into a change layer.

Those raster-based datasets are key to better inform policy makers on the spatial distribution, and extent of urban sprawl in particular for IMP, and are updated every three years.

Thanks to Research Executive Agency (REA) for funding, and to our partners (GAF, UCL, JR, DLR) for their contribution on the H2020 project ECoLaSS.

The H2020 project "Evolution of Copernicus Land Services based on Sentinel data" (ECoLaSS) [4] aims at developing and prototypically demonstrating selected innovative products and methods for future next-generation operational CLMS products of the pan-European and Global Components, based on a multi-temporal and multi-sensors approach. One of its objective is to lay out the feasibility of a higher update frequency for several products - namely, the imperviousness, forest (FOR) and grassland (GRA) layers - of the CLMS continental and global components, for mid-term (2018) and long-term (2020+) evolution. This increased update frequency implies the exploration of new methods to correctly identify the automated changes detected, to ensure the spatial and temporal coherence of those changes along the time series.

2. METHODS AND DATASETS USED TO GENERATE STATUS LAYERS

IMP is the HRL for which the longer time series is available. It consists of a series of 20m and 100m thematic raster status and change products derived from EO data for the 2006, 2009, 2012 and 2015 reference years. Up until 2015, the production of HRL Imperviousness degree (IMD) was based primarily on a combination of SPOT-4 and 5 and IRS LISSIII with a RapidEye coverage introduced in 2012, organized around two separate coverages at least 6 weeks apart during the vegetation growing season.

These coverages were serving multi-purposes: the CLC production as well as the one for 2006 IMD layers and other HRLs from 2012. However, the acquisition of a complete cloud free coverage has been problematic, having to rely on gap filling exercise at a final stage to ensure a near complete coverage. Therefore, the target to achieve complete coverage ($\pm$ 1 year) remains an operational difficulty. The ECoLaSS project will put to test the yearly Sentinel datasets (S1 and S2), through their incorporation in the processing chain.

The HRL IMP production has been largely focused on the creation of a reliable built-up mask, which is then combined with Normalized Difference Vegetation Index (NDVI) data to derive the IMD [5], from 0% to 100%, as displayed in Figure 1. A threshold set at 33% of IMD is then used to create the binary status layer between urban and non-urban areas. Most of the error sources for both layers (status and change) are attributable to the correctness of this input built-up mask.

2.1. Status layer for IMP 2015

Optical datasets from various sources e.g. Landsat-8, SPOT-5, Resourcesat-2, and S-2A, all resampled at a 20m resolution, have been used to generate the 2015 built-up mask. Biophysical variables such as the NDVI and additional parameters such as the Normalized Difference Built-up Index (NDBI) [6] have been computed and time

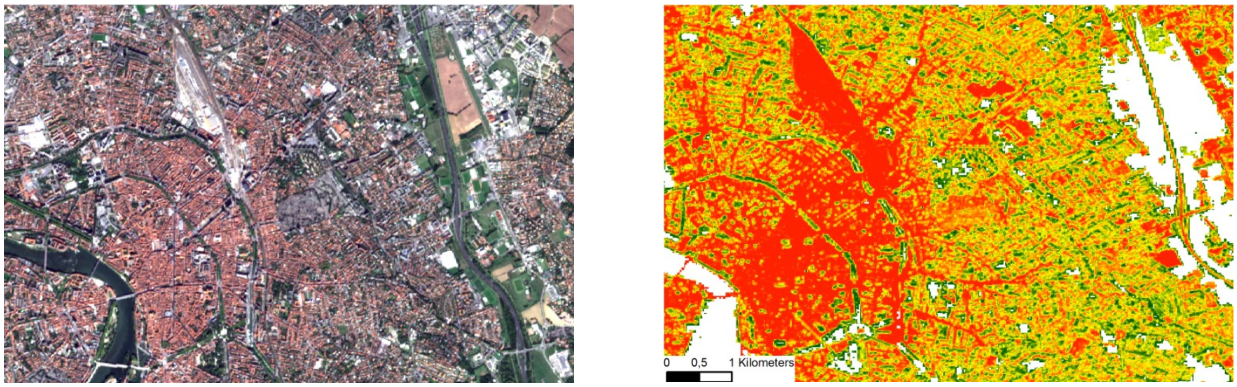

Fig. 1. On the left, S-2 optical image taken above Toulouse, France. On the right, matching HRL IMD layer for 2017, with 100% (fully impervious) in red down to 0% (no sealing) of IMD in green.

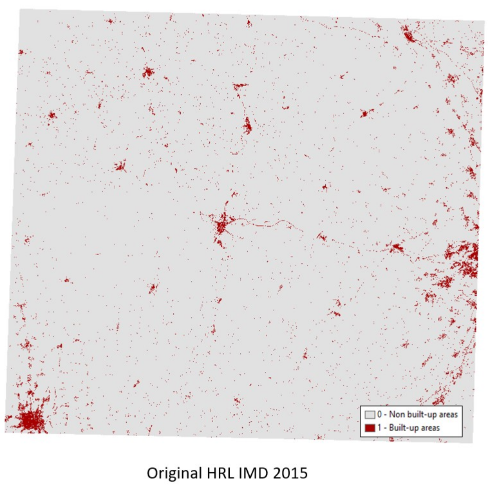

Fig. 2. Status layer for HRL IMP 2015, on 2 S-2 tiles in the South-West of France, at a 20m resolutions.

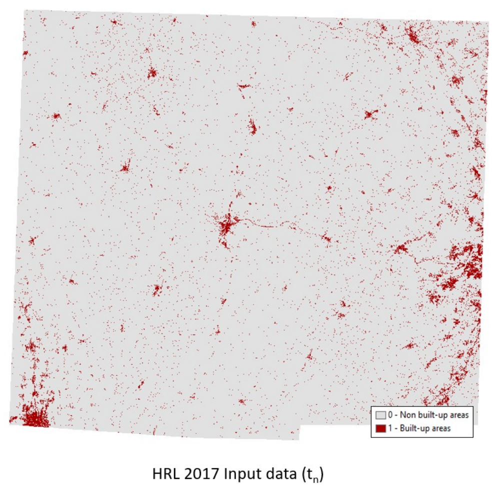

Fig. 3. Status layer for ECoLaSS prototype HRL IMP 2015, on the same area, deduced from the full dataset of S-2 images, cloud-free, with all the spectral bands available.

series statistics on the seasonal and yearly mean, median, maximum, minimum, standard deviation as well as seasonal and yearly range have been used to take full advantage of the cross-sensor seasonal time series of the top of atmosphere images.

Textural features, e.g. the angular second moment of the gray-level co-occurrence matrix, are also added as classification inputs to highlight the inherent heterogeneity of man-made building-structures [7].

All those variables have then been ingested in a semi-automatic classification approach, based on supervised trees, to identify the 2015 built-up areas, whose resulting classification can be seen on Figure 2.

2.2. Status layer for IMP 2017

In the framework of ECoLaSS, a new status layer for the year 2017, has been generated on a selected testing site (matching S-2 tiles 31TCJ, 30TYP) in South-West of France - yielding a change of spatial resolution for the layer at 10m from the previous 20m.

Classifications have been produced image-by-image by a fully automated processing chain, based on a random forest algorithm applied on a subset of the S-2 optical best scenes (pre-processed to get bottom of atmosphere reflectances) and several spectral and textural indices e.g. NDVI and NDBI. A support vector machine classifier is used to obtain classification results from Sentinel-1 SAR datasets. The resulting stack of classified layers (results from optical and SAR images) has then been merged using a Dempster-Shafer algorithm, with the overall precision as metric.

The final classification accuracies are slightly lower than the actual specifications of the HRL IMP (at 90% user and producer accuracies) but this could be easily improved by manual enhancement. The result can be seen on Figure 3. Please also note that Sentinel-1 classification contribution has been studied and quantified with a 3 points improvement for the global accuracy.

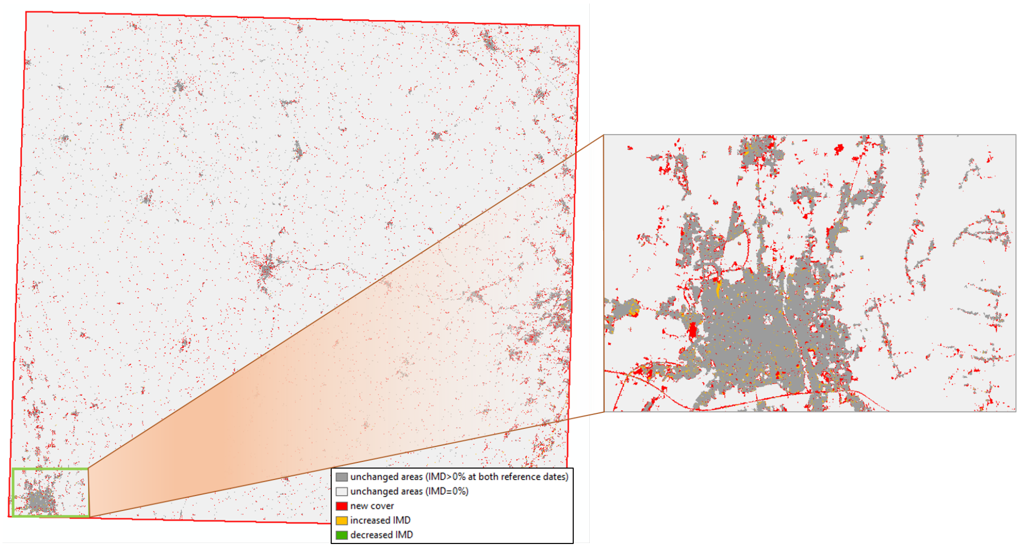

Fig. 4. Final IMP change layer between 2015 and 2017.

Table 1. Classification results for the IMP change layer 2015-2017

Total change areas	For the first calibration	For the second calibration
New built-up 2017	9%	9.64%
Omission: undetected built-up 2015	58%	76.65%
Commission: false built-up 2017	33%	13.71%

3. PRODUCTION OF THE CHANGES LAYER

Specifications only focus on the accuracy of status layers for which the target is set at 90% for both producer and user accuracy, but the results from the previous epochs show that for the IMP change layers, this level of accuracy is still above the expected level of change over the current 3-year period.

The imperviousness change detection relies on two input data described previously: the reference layer (t_0), the HRL IMP 2015, and the new status layer (t_n), the protoype HRL IMP for the year 2017.

The post-classification comparison, to attain full spatial and temporal consistencies, can be decomposed into three main steps:

– A spatial and temporal comparison based on the reference data (t_0) whose purpose is to enforce a geometrical harmonization between the different epochs to prevent problems related to an image-to-image approach;

– A post-processing filtering to remove a significant portion of noise due to small aggregated groups of pixels, which are most likely misclassifications;

– A contextual analysis based on change probability, such as discussed in [8] - this final step will consider the impervious pixels in the 2015 built-up mask to establish a probability map of changes. The analysis describes each pixel's relationship or membership to their neighboring pixels.

The assumption made using this final analysis is that urbanized areas spread more than they appear randomly in the landscape: the resulting urban membership estimates allows the isolation of change areas.

Errors can be present in the reference layer, and new errors could have appeared in the detection of change between two time epochs. They can be linked to:

– Omissions of change new urban areas that appear between 2015 and 2017 were not detected;

– Technical changes due to commission errors added for the new period, i.e. areas falsely flagged as new urban zones, as well as omission errors detected for the previous period, i.e. urban areas, already present in 2015, that were not then detected as such, but have now been flagged as urban areas in the 2017 layer.

A first validation based on a stratified ground truth collection is executed, and the first statistics can be found the second column of the Table 1. The relative magnitude of actual change is then estimated to 9% of the total change areas detected. Thus, errors concerning the remaining 91% of the change areas detected are related to the omission and commission errors detailed above. Regarding the omission errors from the previous epoch, the 2015 production was mostly based on Landsat-8 data whereas the 2017 built-up was produced from S-2 resulting in a nine-fold improvement in spatial resolution, since a Landsat pixel is characterized by nine S-2 pixels, explaining most of the omission errors origin.

This procedure relies heavily on the reference dataset for the statistical calibration of changes described above, which is used to produce statistics from which the estimate areas of each of the three categories in the change stratum will be interfered. Those areas provide then a basis to fine-tune the targets of re-processing, whose objective is to extract the real change areas. This step is achieved by adopting

Table 2. Final classification results for HRL TCD change layer 2015-2018 on testing sites near Avignon, France.

TCD change layer		
	Gain stratum	**Loss stratum**
In 2018	0.5% of gain	17.09% of loss
Omissions	undetected tree: 59% in 2015	undetected tree: 24.62% in 2018
Commissions	false tree detection: 40.5% in 2018	false tree detection: 58.29% in 2015

Table 3. Final classification results for HRL GRA change layer 2015-2018, on testing sites near Arles, France.

GRA change layer		
	Gain stratum	**Loss stratum**
In 2018	2.5% of gain	14% of loss
Omissions	undetected grassland: 46.5% in 2015	undetected grassland: 24.5% in 2018
Commissions	falsely labeled GRA: 51% in 2018	falsely labeled GRA: 61.5% in 2015

a re-classification approach linking the three categories with suitable training data between 2015 and 2017 imagery.

The statistical computation is then reiterated on the reclassified change stratum, and the results, which can be found in the third column of Table 1, confirm the first rough outcome. Based on the reprocessing, of the total area initially detected as changed, only 10% effectively represent new built-up areas while the remaining 90% are mostly omissions undetected in 2015 (76.7%) and new commission errors introduced by the 2017 new built-up mask (13.7%). Most of the omission errors concern small and isolated built-up features and roads, which is mostly attributable to resolution change between Landsat-8 and S-2. Regarding the commissions from 2017, mostly usual errors like small gardens, bare soils in the neighborhood of impervious scattered areas were found. The original change layer represented a nearly 50% increase of the artificial area in the test area which is unrealistic, considering that in fact over 75% of the detected changes were omission from 2015. In the re-classified layer, new built-up areas represent a 4% increase which appear more realistic and already represents a substantial increase over a 2-year period.

This methodology has been successfully implemented on test sites for two other HRLs. It is crucial that no substantial inbalance between omission and commission errors in the HRL change layers remains. Contrary to the IMP change layer, the Tree Cover Density (TCD) change layers and the Grassland (GRA) change layers are composed of two layers each, related to gain and loss, whose results for the reclassification can be found in Tables 2 and 3. The loss of impervious soils being extremely limited, only gain are presented in the change layer for IMP.

Regarding the GRA change strata, they represent together 9.5% of the total study area with losses representing an area 1.5 larger than gains, while the TCD change strata also amount to roughly 10% of the total study area with gains representing an area twice as large as losses. For this HRL, further steps need to be taken to ensure the thematic consistency, regarding the Dominant Leaf Type product, which can be either broadleaved or coniferous, but shouldn't switch between the two typologies over the course of different epochs.

4. CONCLUSION

The slow spatial progression of the sealed areas at European scale present a particular challenge in the frame of Copernicus HRLs. In this paper, the latest developments related to the Sentinel processing as well as the time series reanalysis are presented.

Through ECoLaSS, demonstration has been made of the added values of both Sentinel datasets, from Sentinel-1 (A and B) and Sentinel-2 (A and B) to orient the production toward a yearly release, all the while maintaining the integrity of HRL time series based on heterogeneous datasets coming from multiple sensors and tackling the increase volume of data, using temporal metrics in time-efficient automated algorithms.

New challenges are expected to be the focus of the 2018 production for the HRLs, such as the creation of a building footprint mask, that could be used as a future "backbone", opening new potential tools ensuring the spatial consistency of the time series, as well as in the second phase of the H2020 ECoLaSS project, where new prototypes and their robustness for operational roll-out will be tested.

REFERENCES

[1] UN Habitat, "Global Report on Urban Health equitable, healthier cities for sustainable development", 2016.

[2] Wilson B. and Chakraborty, A., "The Environmental Impacts of Sprawl: Emergent Themes from the Past Decade of Planning Research. Sustainability", Volume 5, pp. 3302-3327, 2013.

[3] "Urban sprawl in Europe, The ignored challenge", 56 pp, ISBN 92-9167-887-2, 2006.

[4] Moser, L., Probeck, M., Ramminger, G., Sannier, C., Descle B., Schardt, M., Gallaun, H., Deutscher, J., Defourny, P., Blaes X., Klein, I., Keil, M., Hirner, A., and Esch, T. "Sentinel-based Evolution of Copernicus Land Services on Continental and Global Scale", 2017, see: https://www.ecolass.eu/

[5] Carlson, T. N. and Arthur S. T., "The impact of land use land cover changes due to urbanization on surface microclimate and hydrology: a satellite perspective" Global and Planetary Change, Volume 25, pp. 49-65, 2000, doi:https://doi.org/10.1016/S0921-8181(00)00021-7.

[6] Zha, Y., Gao, J., and Ni, S., "Use of Normalized Difference Built-Up Index in Automatically Mapping Urban Areas from TM Imagery." International Journal of Remote Sensing, Volume 24, pp. 583-594, 2003.

[7] Pesaresi, M., Ehrlich, D., Caravaggi, I., Kauffmann, M., and Louvrier, C., "Toward global automatic built-up area recognition using optical VHR imagery." Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, Volume 4, pp. 923-934, 2011.

[8] Lefebvre, A., Sannier, C., and Corpetti, T., "Monitoring Urban Areas with Sentinel-2A Data: Application to the Update of the Copernicus High Resolution Layer Imperviousness Degree", Remote Sensing, Volume 8, pp. 1-21, 2016, doi:https://doi.org/10.3390/rs8070606.