Sentinel-2 land cover product comparison: South African National land cover 2020 vs ESRI Global land cover 2020

Luncedo Ngcofe [a,*], Kgaugelo Nkoana [b]

[a] Department of Agriculture, Land Reform and Rural Development, South Africa - Luncedo.ngcofe@dalrrd.gov.za,
[b] Department of Forestry, Fisheries and Environment, South Africa – knkoana@dffe.gov.za

* Corresponding author

Abstract:

Remote Sensing makes land cover mapping over large areas easily achievable when compared to field-based mapping, due to its synoptic coverage capabilities. Remotely sensed land cover mapping can be generated at local, national, regional, and global scales. Land cover mapping is amongst the required information for wide variety of applications such as scientific research, policies, spatial planning, and environmental management investigations. However land cover mapping products are not considered perfect and therefore need to be subjected to statistically rigorous accuracy assessment before being used for any application and / or decision-making process (Stehnman & Czaplewski, 1998; Congalton & Green, 1999; Brown et al., 1999). The land cover product inaccuracies limit their potential for diverse range of applications by multiple users, organizations and / or countries (Grekousis et al., 2015). This study conducted an independent and comparative accuracy assessment for South African National land cover 2020 (SANLC 2020) vs ESRI Global land cover 2020 (ESRI GLC 2020) products, both generated from Sentinel 2 imagery. This is to establish the usability of the ESRI Global land cover 2020 dataset to the South African environment.

Both land cover products (SANLC 2020 & ESRI GLC 2020) were generated through automated cloud-based computing method with different processing techniques. The ESRI GLC 2020 has 10 classes (water, trees, grass, flooded vegetation, crops, shrubs, built area, bare ground snow/ice and clouds). While SANLC 2020 has 73 hierarchal classes which could be grouped into 9 major classes (waterbodies, forest land, grassland, wetland, cultivated, shrubland, built-up, barren land, mines & quarries). The comparison of these land cover products was based on the same accuracy point database generated independently from the team who produced these products. The accuracy points were generated through a collaboration of Department of Agriculture, Land Reform and Rural Development (DALRRD), Department of Forestry, Fisheries and Environment (DFFE) and GeoTerra Image company. This was through merge, edit and update of generated databases from the different organisations to one single database. DALRRD maintains the database by visual point edit and / or new point capture using available 2020 Sentinel RGB false colour band composite (432), aerial imagery (25cm & 50cm) together with Google Earth imagery. This resulted in a total number of 6517 accuracy assessment points covering South Africa.

The accuracy assessment of the land cover products was through the error matrix which is also known as confusion matrix. The error matrix method distinguishes between error of commission and omission through producer's accuracy and user's accuracy analyses. It also provides the overall accuracy assessment analysis (Congalton &Green, 1999; Senseman et al., 1995). The SANLC 2020 had grassland with least producer and user's accuracy of 58% and 68.6% respectively, while waterbodies had highest accuracy of 98.9% and 95.8%. The overall accuracy achieved was 91.5%. The ESRI GLC 2020 had grassland with least producer's accuracy of 26.3%; with flooded vegetation as the least user's accuracy of 11.9%. The water class had the most producer and user's accuracy with 96.7% and 90.1% respectively. The overall accuracy achieved was 73.1%.

Grekousis et al., (2015) argue that global land cover products rarely reach 80% when such datasets are independently validated. They vary between 10% to 50% overall accuracy. It is a challenge to strike a balance between the global land cover generation products that are locally relevant and globally consistent. The local needs might be of high accuracy on certain class features which global land cover products cannot attain. Hence Mitchell et al., (2018) argues that the land cover products overall accuracy should not be analysed independently of producer and user's accuracy as the combination of the results provides an in-depth outcome of the product. The SANLC 2020 & ESRI GLC 2020 have both high producer and user's accuracy for water class feature of 90% and above. Therefore, using water class of both land cover products independently would yield similar results. However, for grassland and flooded vegetation, extra caution needs to be

applied. Wardlow & Egbert, (2003) further argue that the scope of mapping activities also needs to be considered when comparing the local land cover products with the global products. The SANLC 2020 was generated to provide detailed land cover class features for South Africa which resulted in 73 classes while the ESRI GLC focused on mapping only 10 land cover class features for the world. Due to the high number of detailed land cover class features for SANLC 2020, the country's ancillary data was used to assist with automated spectral separability of land class features, hence its high accuracy when compared to ESRI GLC 2020. However, for a global land cover project, sourcing the ancillary datasets for each country would be tedious and time-consuming process. Furthermore, some countries especially the under-developed might not have the required ancillary datasets. The independent statistic accuracy assessment applied in this study provide the strength and limitations of each land cover product (SANLC 2020 & ESRI GLC 2020) with the conclusion that they both have valuable qualities and can be utilised for the South African environment.

Acknowledgements

References

Brown, J.F., Loveland, T.R., Ohlen, D.O., Zhu, Z., 1999. The global land-cover characteristics database: the user's perspective. *Photogrammetric Engineering and Remote Sensing*, 65:1069-1074.

Congalton, R.G., Green, K., 1999. Assessing the accuracy of remotely sensed data: Principles and Practices. Lewis Publishers, Boca Raton.

Grekousis, G., Mountrakis, G., Kavouras, M., 2015. An overview of 21 global and 43 regional land cover mapping products. *International Journal of Remote Sensing*, 36 (21): 5309 – 5335.

Mitchell, P.J., Downie, A.L., Diesing, M., 2018. How good is my map? A tool for semi-automated thematic mapping and spatially explicit confidence assessment. *Environmental Modelling and software Journal*, 108: 111 – 122.

Senseman, G.M., Bagley, C.F., Tweddale, S.A., 1995. Accuracy assessment of discrete classification of remotely-sensed digital data for land cover mapping. *USACERL, Technical Report* EN-95/04.

Stehnman, S.V., Czaplewski, R.L., 1998. Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sensing of the Environment*, 64: 331-344.

Wardlow, B.D., Egbert, S.L., 2003. A state-level comparative analysis of the GAP and NLCD land cover data sets. *Photogrammetric Engineering and Remote Sensing*, 69 (12): 1387 – 1397.