MedGU Mediterranean Geosciences Union
Annual Meeting | Istanbul, 25–28 November 2021

Springer | Publishing Partner

The value of incremental inclusion of spectral indices on the accuracy of LULC mapping in a heterogeneous trans-boundary basin using machine learning and GEE

Blessing Kavhu [123], Zama Eric Mashimbye[1], Linda Luvuno[2]

[1]Department of Geography and Environmental Studies, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa.
[2]Centre for Sustainability Transition, Stellenbosch University, Stellenbosch 7600, South Africa.
[3]Zimbabwe Parks and Wildlife Management Authority, Headquarters, P. O. Box 140 Causeway, Harare.
Corresponding author: kavhublessing@gmail.com

Abstract. While Land use/cover (LULC) information is crucial for effective land-use planning and hydrological monitoring of drainage basins, evaluation of LULC change in heterogeneous transboundary basins is often highly complex thereby compromising the accuracy of LULC data. In this study, we investigate the value of incremental integration of spectral bands with ratio-based and orthogonal based indices to enhance the accuracy of multi-temporal land use/cover classification using machine learning approaches integrated on Google Earth Engine. Two experiments were set up where the first entails incremental integration of ratio-based spectral indices with individual bands, and the second involves integration with orthogonal spectral indices. Ready-to-use Landsat 5 TM and Landsat 8 OLI images, and the Random Forest classifier based on Google Earth Engine were used in this study. Supervised classification is conducted for the years 1996, 2004, 2013 and 2020. Average overall accuracy and Kappa coefficient were used to validate the results. The study found that incremental integration of spectral bands with indices improves the accuracy of land use/cover classification using machine learning. Integrating bands with orthogonal spectral indices yield higher accuracies in comparison to ratio-based spectral indices. Findings from this study facilitate reliable land use/cover monitoring in complex heterogenous transboundary basins with greater statistical confidence. Google Earth Engine is affirmed to be a dependable and cost-effective platform for mapping land use/cover and change across complex and transbοundary settings.

Keywords: Machine Learning, Ratio based indices, Orthogonal Indices, Remote Sensing, Landcover.

1. Introduction

Natural resources that cross international boundaries are often a source of regional conflicts across the globe[1]. There are 310 international transboundary river basins (TDBs) globally that cover approximately 47% of the world's land surface[2]. According to de Stefano et al. (2017), about half of the world's population live within TDBs. It is estimated that TDBs make up 80% of the global freshwater flow [3]. Despite their socio-ecological relevance, their natural resources are continually threatened by changes in land use/land cover (LULC) [4].

Remote sensing is central to producing LULC information for effective land-use planning and natural resource management [5]. However, the accuracy of LULC information is often an issue due to the complexity of TDBs [6]. Studies have explored the integration of spectral bands with spectral indices to enhance the discrimination of different features [7–9]. Categories of spectral indices include ratio-based spectral (RBS) and orthogonal spectral (OS) indices. RBS indices are based on the ratio between a pair of spectral bands [10], whilst the OS indices are based on the existence of a hyperplane in spectral space in which bare soils of varying brightness will lie with vegetation increasing along the hyperplane [11,12]. RBS indices include, amongst others, normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI). OS indices include the components of the Tasseled cap transformation such as the Tasseled cap wetness index (WTCAP), Tasseled cap greenness index (GTCAP) and Tasseled cap brightness index (BTCAP)[13,14].

Previous studies have reported that the integration of several spectral datasets refines the separation of LULC classes and enhances classification accuracy [8,15]. This has seen some authors integrating spectral bands with one or several RBS indices [15]. While research on the integration of spectral bands with RBS indices in heterogenous urban landscapes reported improved accuracy results, compared to those based on spectral bands only [9], previous studies have simply incorporated spectral indices. To the best of our knowledge, no studies have explored the value of incremental inclusion of spectral indices in improving LULC classification accuracy using different types of spectral indices (i.e OS and RBS indices). Understanding this could be crucial to the identification of the most ideal combination of bands and spectral indices to adopt when mapping LULC in complex and heterogeneous transboundary basins.

This study is aimed to assess the value of incremental inclusion of RBS and OS indices to assess LULC change based on a ML classifier using GEE. The specific objectives of this study are to; 1) evaluate the effect of incremental integration of RBS indices on the accuracy of LULC classification, 2) evaluate the effect of OS indices on the accuracy of land cover classification.

2. Methods or Materials and Methods

Ready-to-use Landsat 5 and Landsat 8 OLI images captured during June for the years 1996, 2002, 2013 and 2020 in the Okavango basin were used. Spectral features used in the analysis comprises a combination of six bands RBS indices and OS indices (Figure 1). Sample points were sourced from the Okavango River Basin Water Commission

(OKACOM) geodatabase and National Geographic Okavango and Wilderness Project (NGOWP). Additional samples were generated from digitizing high-resolution satellite imagery from Google earth leading to a total of 5140 samples for eight LULC classes, namely bare land, built-up land, bushland, forest/woodland, grassland, cultivated land, water and wetland. Spectral indices were incrementally integrated with six spectral bands. The sequence for integrating spectral indices was premised on how frequent the indices were previously used in LULC studies starting with the most commonly used indices. The RF classifier was used in this study. The performance of the classifier was measured using the overall accuracy (OA) and the Kappa statistic (Kappa) calculated using the built-in accuracy evaluation functions in GEE. The workflow of the study design is depicted in Figure 1.

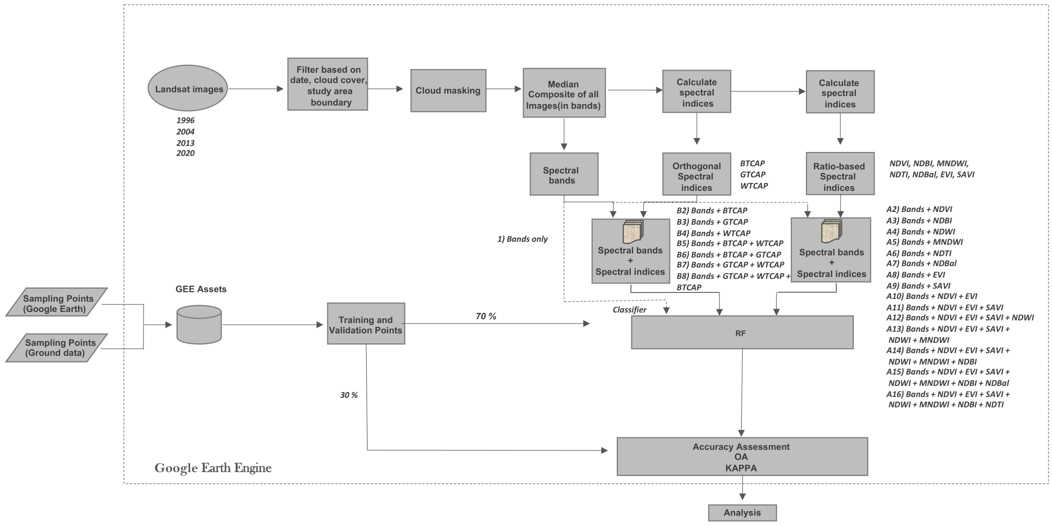

Figure 1. Study workflow

3. Results

Results of this study show that incrementally integrating spectral indices (RBS and OS) to bands increase the accuracy of LULC classification. The average OA for incremental inclusion of RBS indices ranges between 87.56 and 89.78 (Figure 2a). The highest OA of 89.78 (Kappa = 0.88) is based on six spectral bands and six RBS indices (NDVI, EVI, SAVI, NDWI, MNDWI and NDBI). Regarding incremental inclusion of OS indices, OA range from 87.56 to 92.73 (Figure 2b). The highest OA (OA = 92.73) is based on spectral bands and WTCAP. In comparison to RBS, incremental inclusion of OS indices yielded higher accuracies.

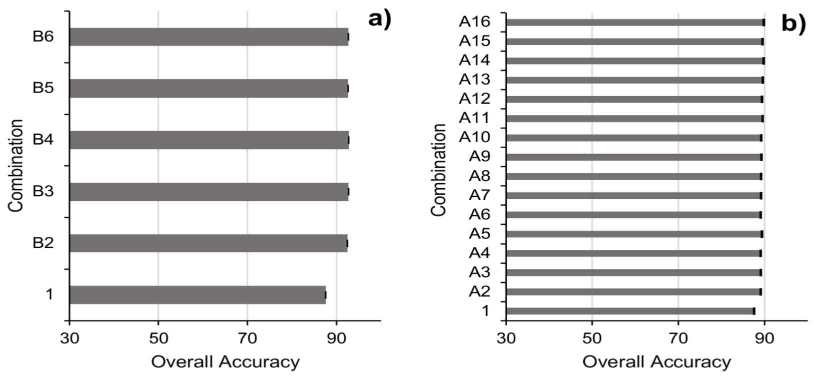

Figure 2. Comparison of accuracy to the integration of spectral bands with different combinations of Orthogonal spectral indices(a) and ratio-based spectral indices(b).

4. Discussion

Results show that incremental integration of spectral bands with indices improve the accuracy of LULC in the Okavango Basin. OS indices improve accuracy more than RBS indices. This is highly likely due to OS indices being derived from six spectral bands which give them enhanced abilities to distinguish class boundary in comparison to RBS indices which are derived from pairs of spectral bands [13]. Additionally, OS indices possess complementary soil adjustment factors which assist in correcting soil background noise thereby enhancing LULC accuracy markedly in dry ecosystems such as semi-arid regions [16]. Furthermore, the fact that OS indices are fewer than RBS indices but remained superior is probably due to model parsimony [17,18]. Parsimonious models have high predictive power even with fewer variables. The results of this study are consistent with those of Huete and Jackson (1988) who reported OS indices to be superior in the detection of vegetation cover than RBS indices. In contrast, Lawrence and Ripple (1998) assert that RBS indices are superior for distinguishing vegetation classes in mountainous landscapes. Variation in methods used, scale, landscape and targeted features on which their investigation was based most likely play a role in contrasting results of this study and those of Lawrence and Ripple (1998).

Mushore et al (2017) evaluated the inclusion of RBS indices on an urban landscape and found that the inclusion of indices yielded an increase in OA from 82.65 to 89.33. Results for this study yielded higher accuracy values as compared to previous studies. OA values for RBS indices in this study ranged from 87.56 to 92.73. This could be due to variation in the ML classifier used and the number of training samples and the scale at which the study is conducted. Notwithstanding that, their findings concurred with our findings that spectral indices edify the definition of LULC classes. That said, Zeng et al (2020) assert that when RBS indices are used together with spectral bands in LULC classification, they have little effect on accuracy as the addition of spectral indices only yielded an increase in OA from 76.43 to 76.55. Their study only used RBS indices. The strength of this study rests in that it incorporated OS indices on a highly heterogeneous transboundary basin. Although previous studies have reported the incorporation of RBS indices in enhancing LULC maps [19], the comparative effect of incremental inclusion of OS and RBS indices in a heterogeneous landscape is novel. This study is amongst the first to provide evidence for the influence of incremental integration of spectral bands with indices to enhance the accuracy of LULC maps, particularly in a TDB setting.

5. Conclusions

Importantly, the study demonstrates the superiority of including OS indices over RBS indices when evaluating LULC classification. Incremental inclusion of OS improves LULC accuracy based on a ML classifier. A conclusion is drawn from this study that, incremental inclusion of OS indices yields results with greater statistical confidence. It is hoped that others, who wish to perform multitemporal analysis using ML classifiers in heterogeneous landscapes, may now consider the inclusion of OS indices. The use of GEE based on analysis-ready data is affirmed by the usefulness of cloud-based analysis for mapping LULC on complex transboundary settings.

References

[1] M. Giordano, A. Drieschova, J. A. Duncan, Y. Sayama, L. de Stefano, and A. T. Wolf, "A review of the evolution and state of transboundary freshwater treaties," *International Environmental Agreements: Politics, Law and Economics*, vol. 14, no. 3, pp. 245–264, 2014.

[2] M. McCracken and C. Meyer, "Monitoring of transboundary water cooperation: Review of Sustainable Development Goal Indicator 6.5.2 methodology," *Journal of Hydrology*, vol. 563, pp. 1–12, Aug. 2018, doi: 10.1016/j.jhydrol.2018.05.013.

[3] I. UNEP and W. R. I. TNC, "Green infrastructure: guide for water management," 2014.

[4] V. Cabello, B. Willaarts, M. Aguilar, L. del M. I.-E. and Society, and undefined 2015, "River basins as social-ecological systems: linking levels of societal and ecosystem water metabolism in a semiarid watershed," *JSTOR*, Accessed: Jul. 16, 2021. [Online]. Available: https://www.jstor.org/stable/26270237

[5] S. T. Azgin and F. D. Celik, "Evaluating Surface Runoff Responses to Land Use Changes in a Data Scarce Basin: a Case Study in Palas Basin, Turkey," *Water Resources*, vol. 47, no. 5, pp. 828–834, 2020.

[6] R. Manandhar, I. O. A. Odeh, and T. Ancev, "Improving the accuracy of land use and land cover classification of Landsat data using post-classification enhancement," *Remote Sensing*, vol. 1, no. 3, pp. 330–344, 2009.

[7] M. Mohajane *et al.*, "Land Use/Land Cover (LULC) Using Landsat Data Series (MSS, TM, ETM+ and OLI) in Azrou Forest, in the Central Middle Atlas of Morocco," *Environments*, vol. 5, no. 12, p. 131, Dec. 2018, doi: 10.3390/environments5120131.

[8] F. Evrendilek and O. Gulbeyaz, “Boosted decision tree classifications of land cover over Turkey integrating MODIS, climate and topographic data,” *International journal of remote sensing*, vol. 32, no. 12, pp. 3461–3483, 2011.

[9] T. D. Mushore, O. Mutanga, J. Odindi, and T. Dube, “Assessing the potential of integrated Landsat 8 thermal bands, with the traditional reflective bands and derived vegetation indices in classifying urban landscapes,” *Geocarto international*, vol. 32, no. 8, pp. 886–899, 2017.

[10] C. J. Tucker, “Red and photographic infrared linear combinations for monitoring vegetation,” *Remote sensing of Environment*, vol. 8, no. 2, pp. 127–150, 1979.

[11] C. D. Elvidge and R. J. P. Lyon, “Influence of rock-soil spectral variation on the assessment of green biomass,” *Remote Sensing of Environment*, vol. 17, no. 3, pp. 265–279, 1985.

[12] R. L. Lawrence and W. J. Ripple, “Comparisons among vegetation indices and bandwise regression in a highly disturbed, heterogeneous landscape: Mount St. Helens, Washington,” *Remote Sensing of environment*, vol. 64, no. 1, pp. 91–102, 1998.

[13] E. P. Crist and R. C. Cicone, “A physically-based transformation of Thematic Mapper data—The TM Tasseled Cap,” *IEEE Transactions on Geoscience and Remote sensing*, no. 3, pp. 256–263, 1984.

[14] A. R. Huete and R. D. Jackson, “Soil and atmosphere influences on the spectra of partial canopies,” *Remote Sensing of Environment*, vol. 25, no. 1, pp. 89–105, 1988.

[15] K. Adepoju, S. Adelabu, and O. Fashae, “Vegetation Response to Recent Trends in Climate and Landuse Dynamics in a Typical Humid and Dry Tropical Region under Global Change,” *Advances in Meteorology*, vol. 2019, 2019.

[16] S. P. Healey, W. B. Cohen, Y. Zhiqiang, and O. N. Krankina, “Comparison of Tasseled Cap-based Landsat data structures for use in forest disturbance detection,” *Remote sensing of environment*, vol. 97, no. 3, pp. 301–310, 2005.

[17] X. Llora, D. E. Goldberg, I. Traus, and E. Bernadó, “Accuracy, parsimony, and generality in evolutionary learning systems via multiobjective selection,” in *International Workshop on Learning Classifier Systems*, 2002, pp. 118–142.

[18] F. J. Martinez-de-Pison, R. Gonzalez-Sendino, A. Aldama, J. Ferreiro-Cabello, and E. Fraile-Garcia, “Hybrid methodology based on Bayesian optimization and GA-PARSIMONY to search for parsimony models by combining hyperparameter optimization and feature selection,” *Neurocomputing*, vol. 354, pp. 20–26, 2019.

[19] A. K. Kuzucu and F. B. Balcik, “Testing the potential of vegetation indices for land use/cover classification using high resolution data,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 4, p. 279, 2017.