Assessing data limitations in ML-based LCLU

Angel Encalada-Davila[1], **Christian Tutiven**[1], **Jose Cordova-Garcia**[2]
[1]Faculty of Mechanical Engineering and Production Science
[2]Faculty of Electrical Engineering and Computer Science
ESPOL University
Guayaquil, Ecuador
`{angaenca, cjtutive, jecordov}@espol.edu.ec`

Abstract

This study addresses the accuracy challenge in Global Land Use and Land Cover (LULC) maps, crucial for policy making towards climate change mitigation. We evaluate two LULC products based on advanced machine learning techniques across two representative nations, Ecuador and Germany, employing a novel accuracy metric. The analysis unveils a notable accuracy enhancement in the convolutional neural network-based approach against the random forest model used for comparison. Our findings emphasize the potential of sophisticated machine learning methodologies in advancing LULC mapping accuracy, an essential stride towards data-driven, climate-relevant land management and policy decisions.

1 Introduction

Earth observation and satellite remote sensing enable the mapping of global land use land cover (LULC) in a spatially explicit manner that informs policy and land management decisions aimed at reaching sustainable development goals [1]. However, despite the advances in technology and techniques used to create these maps, the accuracy of the segmentation of certain geographic areas remains an issue [2, 3]. One of the main reasons is the quality and resolution of the satellite imagery used to create the maps. Also, some algorithms may not perform well in certain regions due to the complexity of the landscape or the presence of unique features that are not well represented in the training data used to develop the algorithms [4] [5]. For example, if the training data used to develop the algorithms is primarily collected from temperate regions and is then applied to a tropical region, the algorithms may not perform well in the tropical region due to the different types of vegetation and land cover present.

With the advances in machine learning (ML) algorithms and cloud computing platforms for Earth observation, like Google Earth Engine (GEE) [6] and openEO [7], the Sentinel satellites have enabled the mapping of LULC maps at a 10 m resolution [8]. In this study, we focus on two LCLU products, (Esri and WC), annually updated. Esri is developed from a convolutional neural network model trained over 5 billion manually annotated Sentinel-2 pixel batches from 24k individual image tiles (510 x 510 pixels each) distributed worldwide. On the other hand, WC comes from an random forest model trained on manually labeled pixels in 100 x 100 m grids at 141k unique locations distributed worldwide. WC involves both Sentinel-1 and Sentinel-2 data as predictors as well.

Given the importance of global LULC maps for several applications and the need for a full understanding of how accurate a LULC product can be, this work aims to address the following key points: i) Evaluate two different LCLU products based on ML selecting a suitable ground truth available for 2 different locations. , and ii) Propose a metric considering differences in ground truth due to location and highlight possible limitations of traditional reported accuracy.

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2023.

Table 1: Accuracy assessment of Esri and WC products and class of interest (crops).

Accuracy (%)		**South America**	**North America**	**Europe**	**Australia**	**Asia**	**Africa**
Crops	**Esri**	62-91	76-89	80-94	82-85	73-94	80-88
	WC	26-90	53-94	66-93	67-92	68-91	60-78

Table 2: Average IoMU of LULC maps regarding the assessed locations.

Location (Country)	**LULC Product**	
	WC	**Esri**
Ecuador	37 %	51 %
Germany	75 %	96 %

This research focuses on elevating the precision and efficiency of LULC mapping through ML, instrumental for enlightened land management and policy-making, thus facilitating climate change mitigation via enhanced carbon sequestration, biodiversity preservation, and sustainable agriculture.

2 Evaluating LCLU maps

One of the main challenges in assessing the accuracy of LULC maps is addressing potential bias in the data. Bias can arise due to factors such as differences in the quality or availability of data sources for different regions or land cover types. By selecting different geographic areas, researchers and map users can gain a more comprehensive understanding of the strengths and limitations of LULC maps and identify specific areas where improvements are needed particularly in regions that might have been underrepresented. In this study we select a representative country for the global north and south, Ecuador and Germany respectively. The Geo-Wiki (called before Global) [9] dataset is proposed to be the ground truth as it is a global reference dataset on cropland that was collected through a crowdsourcing.

For comparison, while not too different in area for Ecuador there are only 846 data points related to croplands, while in Germany there are 11635 data points. Moreover, each sample unit is a frame-pixel of roughly 300 m x 300 m subdivided into 25 grid cells. Further examining the dataset we identified a difference in resolution within the dataset. Germany uses a resolution of roughly 20 x 30 m , but in Ecuador, the resolution is roughly 33 x 33 m. On the other hand, for Esri map, the pixel (grid cell) resolution is around 5 m x 5 m, while for WC the resolution is 7 m x 7 m.Thus, to account for the differences in grid cells for the GT and LULC maps we propose a modification to the IoU metric used in computer vision, as follows:

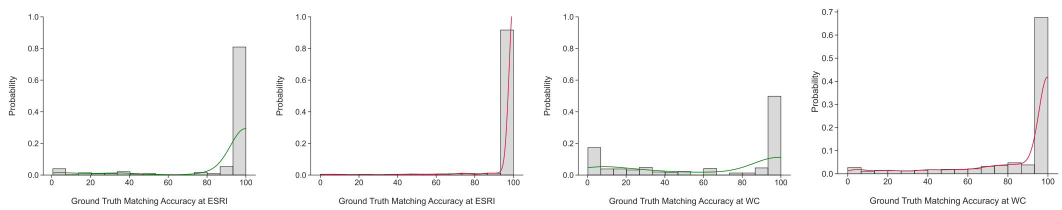
$$IoMU_k = \frac{\sum_{i=1}^{N} IA_{i,k}^{GT|LULC}}{A_k^{GT}}, \tag{1}$$

where $IA_{i,k}^{GT|LULC}$ is the intersection area of the LULC map's i_{th} pixel with the ground-truth's k_{th} grid cell, and A_k^{GT} is the area of the GT's k_{th} grid cell. The score could be in the range of 0 to 1, where 0 means that GT's grid cell has no match with any LULC map's pixel, and 1 means full correspondence between the GT's grid cell and the LULC map's predictions.

3 Results and Discussion

In global terms, the overall accuracy for both products is: Esri (75 %), and WC (65 %), while for crop class only it is: Esri (75-92 %), and WC (59-89 %). In contrast, we show in Tables 1 and 2 the assessment of accuracy for the region and country levels.

Regarding the LULC products, looking at the results by columns, it is possible to observe how the IoMU accuracy increases when the LULC product changes from WC (37% EC, 75% GE) to ESRI



(**a**) ESRI Ecuador. (**b**) ESRI Germany. (**c**) WC Ecuador. (**d**) WC Germany.

Figure 1: GT matching accuracy distribution for both LULC maps and selected locations.

(51% EC, 96% GE). This behavior or trend is general across the world, as well as for the regular accuracy metric shown in Table 1. The main reason behind this relies on the LULC products' ML strategy. Recalling that WC is an RF-based trained map, in contrast to ESRI which is a CNN-based trained map so the classification is performed pixel-by-pixel but considering the surrounding pixels. It means that, for any class, ESRI delivers classifications based on joint-pixel groups. So for any class, it is easier to match that predicted class even by chance (higher probability of matching). On the other hand, WC classifies with more granularity (e.g., one class can be attributed to one single pixel) which affects exactly matching any predicted class.

More specifically to to the assessed countries, the IoMU accuracy shown in the table significantly increases when it changes from Ecuador to Germany. As identified before, mainly data imbalance affects the capability of a trained model to generalize well.There is not a lot of data available for small or not well-known areas like Ecuador or South America (except for Brazil due to its large size and importance level in the world economy), for example, while for other locations like Germany or Europe in general, data availability is higher.

The results above-reported talk in average terms, however, they can be further studied. Figure 1 shows the matching accuracy distribution on GT (i.e., how each GT's grid cell matches with a LULC product's pixel area in terms of accuracy), considering the LULC product and location.

It is expected that all LULC map's cropland area fully covers the GT area, however, that is not the real case. Figures 1a and 1b explain the Esri case in Ecuador and Germany. In Ecuador, the probability of fully covering (100%) a GT grid cell is 0.8 (out of 1.0), while in Germany is much higher, 0.90. This matches with what was explained before as Esri performs better. Regarding the WC case, Figures 1c and 1d illustrate the results, where in Ecuador the probability of fully covering (100%) a GT grid cell is just 0.50 and the rest is distributed between 0 and 40% GT grid cell covering. In Germany, the probability of full covering reaches almost 0.7 while the rest is distributed around 80% GT grid cell covering.

4 Conclusions

In this work a case study about the data issues in LULC maps has been presented, considering two different locations, i.e., Ecuador and Germany, and two different LULC products, i.e., Esri and WC. When the assessment of LULC maps is the purpose, the GT selection is another challenge since besides being scarce data resolution may not match the classifiers.

We proposed assessing maps using the IoMU metric that compares and measures the matching between the LULC products' predictions and the GT data, considering in this study the focus on cropland class. Comparing the location, it is noticed that the selected global south representative has a much lower accuracy in both LULC maps. This highlights the need for taking a closer look into these kind of products when using them in climate change applications such as assessing the climate impact on crops.

References

[1] Argyro Kavvada, Graciela Metternicht, Flora Kerblat, Naledzani Mudau, Marie Haldorson, Sharthi Laldaparsad, Lawrence Friedl, Alex Held, and Emilio Chuvieco. Towards delivering on the sustainable development goals using earth observations. *Remote Sensing of Environment*, 247:111930, September 2020.

[2] Christophe F. Randin, Michael B. Ashcroft, Janine Bolliger, Jeannine Cavender-Bares, Nicholas C. Coops, Stefan Dullinger, Thomas Dirnböck, Sandra Eckert, Erle Ellis, Néstor Fernández, Gregory Giuliani, Antoine Guisan, Walter Jetz, Stéphane Joost, Dirk Karger, Jonas Lembrechts, Jonathan Lenoir, Miska Luoto, Xavier Morin, Bronwyn Price, Duccio Rocchini, Michael Schaepman, Bernhard Schmid, Peter Verburg, Adam Wilson, Paul Woodcock, Nigel Yoccoz, and Davnah Payne. Monitoring biodiversity in the anthropocene using remote sensing in species distribution models. *Remote Sensing of Environment*, 239:111626, March 2020.

[3] M.A.K. Sydenham, Z.S. Venter, K. Eldegard, S.R. Moe, M. Steinert, A. Staverløkk, S. Dahle, D.I.J. Skoog, K.A. Hanevik, A. Skrindo, and G.M. Rusch. High resolution prediction maps of solitary bee diversity can guide conservation measures. *Landscape and Urban Planning*, 217:104267, January 2022.

[4] Tyler J. Lark, Ian H. Schelly, and Holly K. Gibbs. Accuracy, bias, and improvements in mapping crops and cropland across the united states using the USDA cropland data layer. *Remote Sensing*, 13(5):968, March 2021.

[5] Pontus Olofsson, Giles M. Foody, Martin Herold, Stephen V. Stehman, Curtis E. Woodcock, and Michael A. Wulder. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148:42–57, May 2014.

[6] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202:18–27, December 2017.

[7] Matthias Schramm, Edzer Pebesma, Milutin Milenković, Luca Foresta, Jeroen Dries, Alexander Jacob, Wolfgang Wagner, Matthias Mohr, Markus Neteler, Miha Kadunc, Tomasz Miksa, Pieter Kempeneers, Jan Verbesselt, Bernhard Gößwein, Claudio Navacchi, Stefaan Lippens, and Johannes Reiche. The openEO API–harmonising the use of earth observation cloud services using virtual data cube functionalities. *Remote Sensing*, 13(6):1125, March 2021.

[8] Darius Phiri, Matamyo Simwanda, Serajis Salekin, Vincent Nyirenda, Yuji Murayama, and Manjula Ranagalage. Sentinel-2 data for land cover/use mapping: A review. *Remote Sensing*, 12(14):2291, July 2020.

[9] Juan Carlos Laso Bayas, Myroslava Lesiv, François Waldner, Anne Schucknecht, Martina Duerauer, Linda See, Steffen Fritz, Dilek Fraisl, Inian Moorthy, Ian McCallum, Christoph Perger, Olha Danylo, Pierre Defourny, Javier Gallego, Sven Gilliams, Ibrar ul Hassan Akhtar, Swarup Jyoti Baishya, Mrinal Baruah, Khangsembou Bungnamei, Alfredo Campos, Trishna Changkakati, Anna Cipriani, Krishna Das, Keemee Das, Inamani Das, Kyle Frankel Davis, Purabi Hazarika, Brian Alan Johnson, Ziga Malek, Monia Elisa Molinari, Kripal Panging, Chandra Kant Pawe, Ana Pérez-Hoyos, Parag Kumar Sahariah, Dhrubajyoti Sahariah, Anup Saikia, Meghna Saikia, Peter Schlesinger, Elena Seidacaru, Kuleswar Singha, and John W Wilson. A global reference database of crowdsourced cropland data collected using the geo-wiki platform. *Scientific Data*, 4(1), September 2017.