VALIDATION METHODS FOR REGIONAL RETROSPECTIVE HIGH RESOLUTION LAND COVER FOR UKRAINE

Article · July 2016

6 authors, including:


Mykola Lavreniuk
National Academy of Sciences of Ukraine


Nataliia Kussul
National Academy of Sciences of Ukraine


Andrey Yu. Shelestov
National Technical University of Ukraine Kiev P…


Bohdan Yailymov
Space Research Institute NAS Ukraine and SSA …

VALIDATION METHODS FOR REGIONAL RETROSPECTIVE HIGH RESOLUTION LAND COVER FOR UKRAINE

Mykola Lavreniuk[1,3], Nataliia Kussul[1,2], Andrii Shelestov[2,1], Bohdan Yailymov[1], Tamara Oliinyk[1,3], Alexander Kosteckyi[1]

[1]Space Research Institute NASU-SSAU, Kyiv, Ukraine
[2]National Technical University of Ukraine "Kyiv Polytechnic Institute", Kyiv, Ukraine
[3]Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

ABSTRACT

Many applied Earth observation problems are based on land cover and land use maps, derived from satellite data. That is why it is important to assess their accuracy. We have developed retrospective regional 30 meter resolution land cover maps for Ukraine based on Landsat data for 1990, 2000 and 2010. As there is no reference data for validating retrospective periods, validation of the maps could be done only with photo-interpretation. In this paper we investigate two different sampling schemes for reference samples selection: pseudo-random (purposeful) samples selection (first approach) and systematic on regular grid (second approach). With systematic samples selection we receive the lower accuracy of classification (overall, user and producer), then with pseudo-random. Nevertheless we consider the validation results with the systematic sampling scheme (the second approach) to be more reliable comparing to the first one, because the second sampling scheme is less subjective. Moreover, samples proportion within the second approach better corresponds to the statistics.

Index Terms — land cover map, validation, sampling scheme.

1. INTRODUCTION

A lot of applied satellite monitoring problems are solved using land cover and land use (LCLU) maps [1]–[4]. That is why, it is extremely important to assess the maps' accuracy and reliability. We have built high resolution land cover maps for the whole territory of Ukraine based on the Landsat-4/5/7 images for three decades: 1990s, 2000s and 2010s [5]. Land cover maps validation is performed by comparing classification results with the reference data (validation set) using the confusion matrix, user's and producer's accuracies [6]–[7]. In this study, we consider two most common methods for reference data generation: pseudo-random (purposeful) sampling and systematic sampling on a regular grid [8]–[9]. When using a pseudo-random approach, an expert selects samples that can be interpreted by him with minimal errors. In such a case, the accuracy of the map could be overestimated. Systematic sampling approach is more objective for reference data selection, but might be more difficult and resource consuming for photo-interpretation. Taking into account the impact of human subjectivity, two independent experts participated in reference data collecting within the second approach [10]. Within photo-interpretation, they provided a linguistic measure of reliability along with identified classes [10]. Then a more experienced expert ("chief analyst") determined the final value of reference class for each sample based on two experts' results. This technique allows us to provide independent validation for land cover map and to compare it with the results based on pseudo-random selection of reference samples. Regular and qualitative land cover maps made it possible to assess land cover changes. These maps allow us to estimate the general trends of LCLU in Ukraine and the impact on the climate changes and economic indicators. The proposed approach for land cover maps validation is also suitable for land cover changes maps validation.

2. DATA DESCRIPTION

We developed high resolution (30 m) land cover maps for the whole territory of Ukraine for three decades: 1990s, 2000s and 2010s (UALandCover30). For this, atmospherically corrected time-series of Landsat-4/5/7 images were classified using a neural network ensemble [11]–[12]. These maps contain six main land cover classes of the European Land Use and Cover Area frame Survey (LUCAS) nomenclature: artificial surface, cropland, grassland, forest, bare land and water. We used a vector mask of inhabited localities for the whole territory of Ukraine in order to exclude from the consideration the territory of cities and villages. Validation was also performed for the GlobeLand30-2010 map that is also available for Ukraine at 30 m spatial resolution [13].

Land cover maps were filtered to remove isolated groups of pixels [14]. It allowed us to decrease “salt and pepper noise” in land cover changes maps.

3. METHOD FOR LAND COVER MAPS VALIDATION

For the purpose of land cover classification, we formed reference data sets using a photo-interpretation method with uniform spatial distribution over the territory of interest and proportional representation of all classes [8], [15]. As a result, reference data consisted of 14,261, 13,492 and 13,575 polygons for 1990, 2000 and 2010, respectively. To deal with Big data processing we used high performance computation techniques [16]–[17]. These polygons have been randomly divided into training (50%) and validation (test) (50%) sets. Training set was used for land cover classification and test set was used for independent validation only. Since the location of validation polygons is selected by an expert, it is pseudo-random (purposeful), because the expert tries to select polygons that are easiest to photo-interpret. Therefore, we propose the second approach to provide independent validation which is systematic sampling on the regular grid. Within each administrative region, we selected test points on the regular grid with the 10 km distance between each neighborhood points (Fig. 1, bottom). The experts provided reference classes for each point by the photo-interpretation. In some cases, it was difficult to determine the reference class with photo-interpretation. To decrease the uncertainty in the reference data, two experts and “the chief analyst” took part in the second reference data collecting approach [10]. The resulting reference class for each sample was determined by the “chief analyst” by the averaging two experts’ results.

4. QUALITY ESTIMATION AND ACCURACY ASSESSMENT

To estimate the accuracy of land cover maps using the first approach we used the validation sets with 5,353, 6,111 and 6,426 polygons for 1990, 2000 and 2010, respectively, for all the territory of Ukraine (Fig. 1, top). Percentage of each class in the training and validation set roughly corresponded to the class percentage at land cover for each oblast according to the statistics.

Fig.2 illustrates the distribution of different land cover classes for Odessa region in validation set, selected within the first pseudo-random sampling scheme (a pie chart at the left); according to statistics for the region (centre) and systematically selected on the regular grid (at the right). It can be seen that the second (systematic) sampling approach is more statistically correct, because it better reflects statistical distribution of different land cover classes.

In Table 1, we provide accuracy assessment of our UALandCover30 maps for 1990, 2000 and 2010 and compare them to GlobeLand30-2010 land cover in terms of producer’s accuracy (PA), user’s accuracy (UA) and overall accuracy (OA) using the first validation approach.

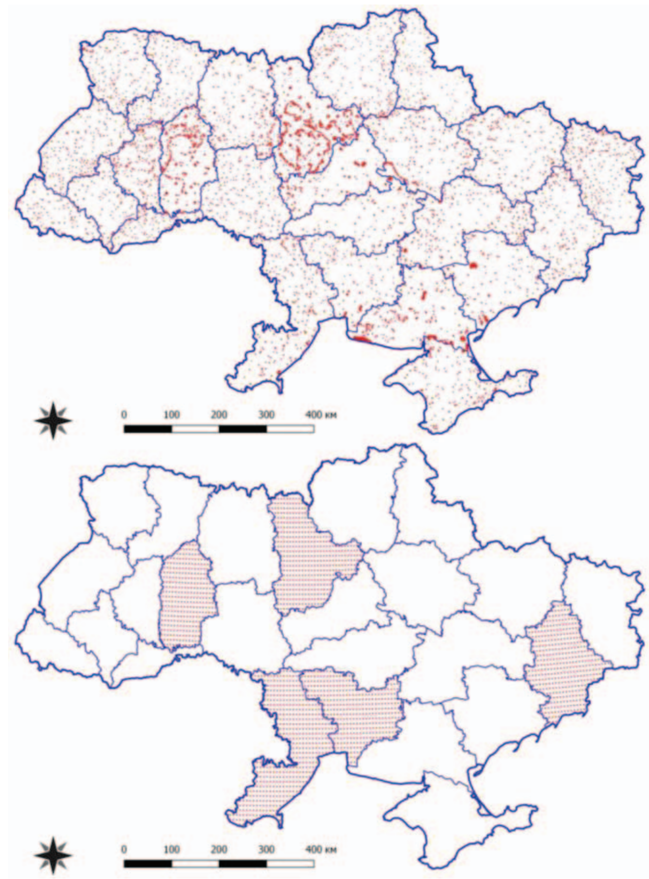

Fig. 1. Samples distribution using pseudo-random first (top) and systematic second (bottom) sampling schemes in 2010.

Table 1. The classification accuracy for the territory of Ukraine for 1990, 2000 and 2010 years using our UALandCover30 (I) for 1990, 2000 and 2010 and their comparison to GlobeLand30-2010 (II)

	2010 year				2000 year		1990 year	
	I		II		I		I	
Class	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**
Artificial	100	87.8	3.4	79.5	100	84	89.6	73.6
Cropland	93.5	96.2	85.3	99.4	95	98.1	94.5	96.6
Forest	95.4	96.2	95.9	89.9	97.3	97.8	95.2	97.7
Grassland	81.4	71.2	60.5	34.4	88.3	75	82.7	70
Bare land	91.7	96.4	57.1	0.4	96.2	88.3	93.2	92.5
Water	99.5	99.6	99.9	96.6	99.4	99.8	99.3	99.6
OA, %	94.7		89.7		96.4		95.1	

The overall accuracy of the regional UALandCover30 map calculated on the pseudo-randomly selected validation set was approximately 95%. Accuracies for each individual class were higher than 70%. It was 5% higher than overall accuracy of the GlobeLand30-2010 map for the same validation set. At the same time, the PA and UA accuracies of grassland classification in our maps were +10% and +45% better than in GlobeLand30-2010, respectively. It was the lowest accuracy value among the main six land cover classes, since it was difficult to separate grassland from cropland.

In Table 2, we excluded class of artificial objects for UALandCover30 because test samples are located in villages and cities, that are excluded from classification using vector mask of settlements. In the second approach, two experts have provided the reference classes and a linguistic value for each point on the regular 10 km grid. To investigate the impact of human subjectivity on the result, the "chief analyst" revised these results and provided the final reference data. Validation was performed for five regions in Ukraine: Kyiv, Khmelnitskyi, Odessa, Mykolaiv and Donetsk. Accuracy comparison using final reference data shows that the overall accuracy of UALandCover30 at the systematically selected validation set is more accurate then GlobeLand30-2010 by +6.9%, +17%, +15.2%, +1.4%, and +1.7% for 5 regions (Table 2).

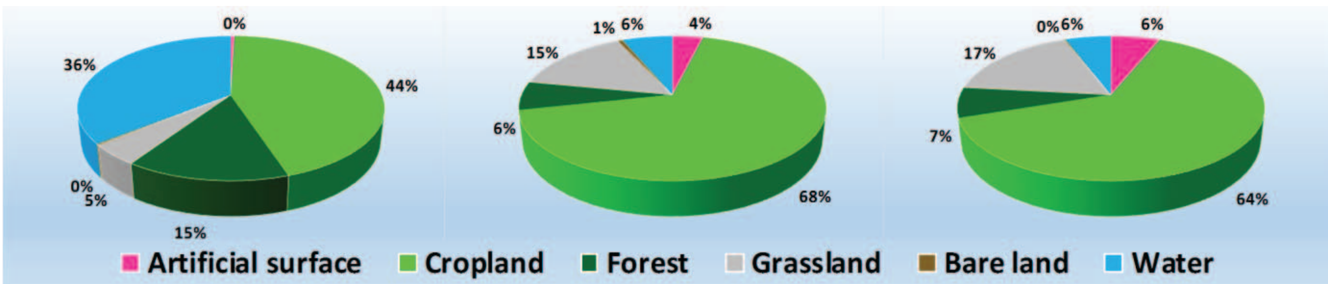

Fig. 2. The samples area proportion using first (left), second (right) approaches and statistics (centre) in Odessa region (2010).

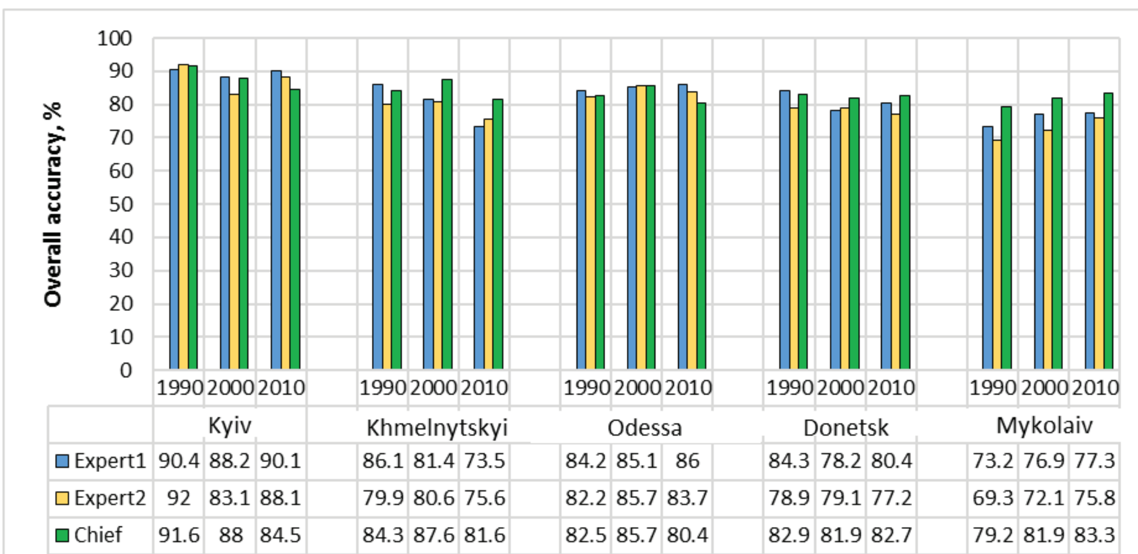

Fig. 3. Two experts and chief results within the second approach on UALandCover30 for 1990, 2000 and 2010.

Table 2. Accuracy comparison of UALandCover30-2010 (I) and GlobeLand30-2010 (II) using sampling on the regular grid

Product	Donetsk				Khmelnytskyi				Kyiv				Odessa				Mykolaiv			
	I		II		I		II		I		II		I		II		I		II	
Class	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**
Artificial	-	-	51.4	94.7	-	-	40	62.5	-	-	65.3	86.5	-	-	69	66.7	-	-	57.9	84.6
Cropland	95.3	80	76.8	98.3	97.3	71.8	64.8	96.9	92.9	80.7	60.8	98	94.1	83.8	78.2	97	97.1	81.6	82.9	98.8
Forest	70	87.5	91.7	68.8	100	91.5	91.4	62.7	84.5	94.6	84.8	79.6	81	54.8	86.7	40.6	87.5	43.8	50	11.1
Grassland	64.5	87.6	82.9	30.1	59.1	95.6	16.7	1.2	74.8	80.3	76.3	21	47.7	79.2	75	28.6	54.2	96.7	88.5	35.4
Bare land	80	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Water	85.7	100	85.7	100	66.7	100	50	66.7	95.2	100	95.5	100	100	96.3	100	87.5	100	100	100	100
OA, %	82.7		75.8		81.6		64.6		84.5		69.3		80.4		78.7		83.3		81.9	

5. DISCUSSION AND CONCLUSIONS

This paper presented two methods for the retrospective land cover maps validation for the territory of Ukraine based on Landsat data at 30 m resolution. With the first pseudo-random sampling approach, the overall accuracy is approximately 95% for UALandCover30 for three different time periods (1990, 2000 and 2010). High accuracy is explained by the way this dataset was formed: the experts tend to select polygons that can be easily interpreted with minimal errors. We think this result is more relevant (objective) due to regularity of grid and more independent selection of validation set. Moreover, classes distribution in the regular grid approach is more close to statistics (real world distribution of land cover classes) (Fig. 2). So, the validation results with the systematic sampling scheme are more reliable comparing to the pseudo-random. Systematically selected validation set showed that UALandCover30-2010 is more accurate than GlobeLand30-2010 by +6.9%, +17%, +15.2%, +1.4%, and +1.7% for 5 select regions of Ukraine.

6. REFERENCES

[1] N. Kussul, A. Shelestov, R. Basarab, S. Skakun, O. Kussul, and M. Lavreniuk, "Geospatial intelligence and data fusion techniques for sustainable development problems," *CEUR Workshop Proceedings*, 1356, pp. 196–203, 2015.

[2] N. Kussul, S. Skakun, A. Shelestov, and O. Kussul, "The use of satellite SAR imagery to crop classification in Ukraine within JECAM project," in: IGARSS 2014, pp. 1497–1500, 2014.

[3] N. Kussul, S. Skakun, A. Shelestov, O. Kravchenko, J.F. Gallego, and O. Kussul, "Crop area estimation in Ukraine using satellite data within the MARS project," in: IGARSS 2012, 22-27 July 2012, Munich, Germany, pp. 3756–3759.

[4] N. Kussul, S. Skakun, A. Shelestov, O. Kravchenko and O. Kussul "Crop classification in Ukraine using satellite optical and SAR images," *Models&Analyses,* vol. 2, no. 2, pp. 118–128, 2013.

[5] M. Lavreniuk, N. Kussul, S. Skakun, A. Shelestov, and B. Yailymov, "Regional retrospective high resolution land cover for Ukraine: methodology and results," in: IGARSS 2015, pp. 3965–3968, 2015. DOI: 10.1109/IGARSS.2015.7326693.

[6] M. Story and R. G. Congalton, "Accuracy Assessment: A User's Perspective," *Photogrammetric Engineering & Remote Sensing*, vol. 52, pp. 397–399, 1986.

[7] M. Lavreniuk, S. Skakun, A. Shelestov, B. Yalimov, S. Yanchevskii, D. Yaschuk, and O. Kostetsky, "Large Scale Classification of Land Cover on the Base of Retrospective Satellite Data," *Cybernetics Syst. Anal.*, vol. 52, no. 1, pp. 137-149, 2016.

[8] P. Olofsson, G.M. Foody, M. Herold, S.V. Stehman, C.E. Woodcock, and M.A. Wulder, "Good Practices for Estimating Area and Assessing Accuracy of Land Change," *Remote Sensing of Environment*, vol. 148, pp. 42–57, 2014.

[9] S. V. Stehman, "Comparison of systematic and random sampling for estimating the accuracy of maps generated from remotely sensed data," *Photogrammetric Engineering and Remote Sensing*, vol. 58, no. 9, pp. 1343-1350, 1992.

[10] P. Sarmento, C. C. Fonte, J. Dinis, S. V. Stehman and M. Caetano, "Assessing the impacts of human uncertainty in the accuracy assessment of land-cover maps using linguistic scales and fuzzy intervals," *International Journal of Remote Sensing*, vol. 36, no.10, pp. 2524-2547, 2015.

[11] N. Kussul, S. Skakun, A. Shelestov, M. Lavreniuk, B. Yailymov, and O. Kussul, "Regional Scale Crop Mapping Using Multi-Temporal Satellite Imagery," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XL-7/W3*, pp. 45–52, 2015.

[12] S. Skakun, N. Kussul, A. Y. Shelestov, M. Lavreniuk, O. Kussul, "Efficiency Assessment of Multitemporal C-Band Radarsat-2 Intensity and Landsat-8 Surface Reflectance Satellite Imagery for Crop Classification in Ukraine," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015, DOI: 10.1109/JSTARS.2015.2454297.

[13] Geoportal Openlandservice. Electronic resource: http://www.globallandcover.com

[14] N. Kussul, G. Lemoine, J. Gallego, S. Skakun and M. Lavreniuk, "Parcel based classification for agricultural mapping and monitoring using multi-temporal satellite image sequences," in: IGARSS 2015, pp. 165-168, 2015.

[15] F. J. Gallego, N. Kussul, S. Skakun, O. Kravchenko, A. Shelestov, and O. Kussul, "Efficiency assessment of using satellite data for crop area estimation in Ukraine," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 29, pp. 22–30, 2014.

[16] A.Yu. Shelestov, A.N. Kravchenko, S.V Skakun, S.V. Voloshin, and N.N. Kussul, "Geospatial information system for agricultural monitoring," *Cybernetics and System Analysis*, vol. 49, no. 1, pp. 124–132, 2013.

[17] N. Kussul, A. Shelestov and S. Skakun "Grid and sebsor web technologies for environmental monitoring," *Earth Science Informatics*, vol. 2, no. 1-2, pp. 37–51, 2009.