EXTENDING ACCURACY ASSESSMENT PROCEDURES OF GLOBAL COVERAGE LAND COVER MAPS THROUGH SPATIAL ASSOCIATION ANALYSIS

D. Oxoli [1, *], G. Bratic [1], H. Wu [2], M. A. Brovelli [1]

[1] Politecnico di Milano, Department of Civil and Environmental Engineering, Piazza Leonardo da Vinci 32, 20133 Milan, Italy - (daniele.oxoli,gorica.bratic, maria.brovelli)@polimi.it
[2] National Geomatics Center of China, Lianhuachi West Road 28, 100830 Beijing, China - wuhao@ngcc.cn

Commission IV, WG IV/4

KEY WORDS: High-Resolution Global Land Cover, GlobeLand30, Accuracy, Validation, Spatial Association

ABSTRACT:

High-resolution land cover maps are in high demand for many environmental applications. Yet, the information they provide is uncertain unless the accuracy of these maps is known. Therefore, accuracy assessment should be an integral part of land cover map production as a way of ensuring reliable products. The traditional accuracy metrics like Overall Accuracy and Producer's and User's accuracies - based on the confusion matrix - are useful to understand global accuracy of the map, but they do not provide insight into the possible nature or source of the errors. The idea behind this work is to complement traditional accuracy metrics with the analysis of error spatial patterns. The aim is to discover errors underlying features which can be later employed to improve the traditional accuracy assessment. The designed procedure is applied to the accuracy assessment of the GlobeLand30 global land cover map for the Lombardy Region (Northern Italy) by means of comparison with the DUSAF regional land cover map. Traditional accuracy assessment quantified the classification accuracies of the map. Indeed, critical errors were pointed out and further analyses on their spatial patterns were performed by means of the Moran's I indicator. Additionally, visual exploration of the spatial patterns was performed. This allowed describing possible sources of errors. Both software and analysis strategies were described in detail to facilitate future improvement and replication of the procedure. The results of the exploratory experiments are critically discussed in relation to the benefits that they potentially introduce into the traditional accuracy assessment procedure.

1. INTRODUCTION

Nowadays, the Land Cover (LC) data is entitled as key information to a number of pressing scientific and societal concerns connected to the environment. These include - among others - natural resources management (see e.g. Pielke, 2005; Cui et al., 2014), climate and biodiversity studies (see e.g. Turner et al., 2003; Feddema et al., 2005; Bontemps et al., 2013; Oxoli et al., 2018), urban planning (see e.g. Esch et al., 2010), etc.

The progress in Earth Observation (EO) techniques and platforms have significantly upheld the development of the LC maps. Over the last two decades, the improvement of the satellite imagery spatial, temporal, and radiometric resolution together with the increasing availability of computing power and the decreasing costs of these technologies have brought LC maps development to an unprecedented level (Tatem et al., 2008). Modern LC maps - in fact - provide often with global spatial coverage, frequent updates, and pixel resolution up to few meters thus enabling outstanding opportunities for analysing and modelling phenomena taking place on the Earth surface. Examples of the above are a number of high-resolution products released by different national and international EO programs. These include multi-class and global coverage LC datasets with a spatial resolution equal or lower than 30 m such as the GlobeLand30 (GL30) with 10 classes (Chen et al. 2015), and the Fine Resolution Observation and Monitoring of Global Land Cover (FROM-GLC) with 8 classes (Gong et al., 2013). Higher resolution (10m) thematic datasets are also available, among others the Global Human Settlement Layer (GHS) (Pesaresi et

al., 2016) and the Global Urban Footprint (GUF) (Esch et al., 2013).

On one hand, global LC maps provide users with consistent LC information, thus empowering reproducible LC-based applications at a global scale. On the other hand, this growing LC maps availability proposes additional burdens and concerns connected with their quality and - in turn - with the design of reliable accuracy assessment procedures. Accuracy assessment is key to determine the goodness of a map of representing the reality and, therefore, its suitability for any specific application. In fact, for applications at a smaller or local scale the accuracy of global LC maps does not always meet the users' requirements, thus making the use of regional or local land cover maps - where available - often preferred. In this context, the accuracy assessment of global LC maps by means of comprehensive validation procedures still features a pivotal task to provide a detailed description of the global LC maps quality as well as to promote their use into accuracy-demanding or local applications.

In views of the above, this work focuses on the accuracy assessment of the GL30 map that is the most frequently updated (2000, 2010, and 2015 announced) high-resolution global LC multiple-class map currently available. The validation is carried out at a regional scale for the Lombardy Region (Northern Italy) by means of comparison with higher resolution LC regional datasets. The DUSAF (Destinazione d'Uso dei Suoli Agricoli e Forestali) land cover vector map at a scale 1:10000 (ERSAF, 2010) is employed as the reference dataset for the classification accuracy. Indeed, it represents the highest resolution LC map

---

* Corresponding author

available for the study area. Traditional accuracy assessment procedures - based on the confusion matrix - are considered (Congalton, 2004) and enriched with data exploratory experiments for analysing spatial variability and patterns of classification errors by means of spatial association statistics. The analysis of spatial association to assess the accuracy of thematic maps has been proposed in the literature (see e.g. Campbell, 1981; Steele et al., 1998). In this work, a procedure to investigate also partial disagreements in the map classification by exploiting data at different resolutions is proposed. Results provide metrics allowing to describe global, intra-class, and inter-class classification accuracy. The preliminary outcomes of the errors spatial variability analysis outline interesting correlations among some classification error features and the errors underlying spatial patterns. This makes the introduction of this additional analysis component into the confusion matrix promising.

The paper proceeds as follows. In Section 2, the details of the LC dataset used in this work are presented. In Section 3, data processing strategies are described. Results are reported and discussed in Section 4. Conclusions and future directions for the work are included in Section 5.

2. DATASETS

The accuracy assessment presented in this work considers two LC datasets, namely the GL30 as target raster map and the DUSAF vector map as reference or ground truth. The main characteristics of these datasets are presented in the following.

2.1 GlobeLand30 (GL30)

The GL30 (www.globeland30.org) is an open-access 30 m resolution global LC raster data product developed by the National Geomatics Center of China. It is comprised of 10 classes, including Cultivated land, Forest, Grassland, Shrubland, Wetland, Water bodies, Tundra, Artificial surfaces, Bareland, Permanent snow and ice in the years 2000 and 2010. The GlobeLand30 for the Year 2010 is considered in this work. The definition of the GL30 classes is reported in Table 1.

Code	Class	Definition
10	Cultivated Land	Lands used for agriculture, horticulture and gardens, including paddy fields, irrigated and dry farmland, vegetation and fruit gardens, etc.
20	Forest	Lands covered with trees, with vegetation cover over 30%, including deciduous and coniferous forests, and sparse woodland with cover 10 - 30%, etc.
30	Grassland	Lands covered by natural grass with cover over 10%, etc.
40	Shrubland	Lands covered with shrubs with cover over 30%, including deciduous and evergreen shrubs, and desert steppe with cover over 10%, etc.
50	Water bodies	Water bodies in the land area, including river, lake, reservoir, fish pond, etc.
60	Wetland	Lands covered with wetland plants and water bodies, including inland marsh, lake marsh, river floodplain

		wetland, forest/shrub wetland, peat bogs, mangrove and salt marsh, etc.
70	Tundra	Lands covered by lichen, moss, hardy perennial herb and shrubs in the polar regions, including shrub tundra, herbaceous tundra, wet tundra and barren tundra, etc.
80	Artificial surfaces	Lands modified by human activities, including all kinds of habitation, industrial and mining area, transportation facilities, and interior urban green zones and water bodies, etc.
90	Bareland	Lands with vegetation cover lower than 10%, including desert, sandy fields, Gobi, bare rocks, saline and alkaline lands, etc.
100	Permanent snow and ice	Lands covered by permanent snow, glacier and ice cap.

Table 1. GlobeLand30 classes description

The datasets were extracted from the Landsat 7 (https://landsat.gsfc.nasa.gov/landsat-7) and the HJ-1 (https://earth.esa.int/web/eoportal/satellite-missions/h/hj-1) satellite imagery using the Pixel-Object-Knowledge (POK) method (Chen, et al. 2015). On 22 September 2014, Chinese government donated GlobeLand30 to United Nations (UN) as a contribution towards global sustainable development. Since the release of GlobeLand30 in 2014, it has been freely downloaded by more than 10,000 users from 130 countries.

2.2 DUSAF

DUSAF is the official LC map of the Lombardy Region, produced by means of photo-interpretation of high resolution (50 cm) aerial imagery. Multiple versions of the DUSAF are available related to different reference years, namely 1999, 2007, 2009, 2012, and 2015, which can be freely downloaded from the Lombardy Region Geoportal (http://www.geoportale.regione.lombardia.it). The data is distributed under Italian Open Data License (IODL, https://www.dati.gov.it/content/italian-open-data-license-v20). The DUSAF 4.0 related to the year 2012 is considered in this work. The DUSAF is available in vector format with a declared rendered scale of 1:10.000. DUSAF classification consists of a five-level hierarchical structure starting with coarse LC classes whereas each sub-level represents its parent-level class with a higher detail. Table 2 describes the 3rd level of DUSAF classification, that was of interest for this paper. Looking at the first digit of the class code, one can also understand the 1st level classification (1. Artificial areas, 2. Agricultural areas, 3. Forest and semi-natural areas, 4. Wetlands, 5. Water bodies).

DUSAF 3rd level class code	Description
111	Continuous urban fabric
112	Discontinuous urban fabric
121	Industrial, commercial, public and private units
122	Road and rail networks and associated land
124	Airports and helipads
131	Mineral extraction sites
132	Dump sites
133	Construction sites
134	Non-used and non-vegetated degraded areas
141	Green urban areas

142	Sport and leisure facilities
211	Non-irrigated arable land
213	Rice fields
221	Vineyards
222	Fruit trees and berry plantations
223	Olive groves
224	Arboriculture of wood
231	Pastures
311	Broad-leaved forest
312	Coniferous forest
313	Mixed forest
314	Recent afforestation
321	Natural grassland at high altitudes
322	Bushes and shrubland
324	Transitional woodland / shrub
331	Beaches, dunes and sand planes
332	Bare rock and debris accumulation
333	Sparsely vegetated areas
335	Glaciers and perpetual snow
411	Inland marshes and peat bogs
511	Water courses
512	Water bodies

Table 2. Third level of DUSAF classification legend

3. DATA PROCESSING

Generally speaking, the accuracy assessment of LC raster maps requires the involved datasets to be harmonized in terms of classification legend, coordinate reference system, projection, and resolution. This implies that one among the target map and the reference map needs to be resampled at a convenient resolution such as the reference map - conventionally with a higher resolution - is downsampled at the lower resolution of the target map. The downsampling, coupled with reclassification and cropping of the two maps on the same area, is considered to carry out traditional accuracy assessments, based on the confusion matrix. Details of the latter are reported in Section 3.1. To better investigate disagreements between the target and the reference map, in this work an alternative procedure enabling sub-pixel classification errors detection by preserving the original higher resolution of the reference map is proposed and tested. The processing steps required by this alternative procedure are described in Section 3.2.

3.1 Data Preprocessing

Regarding the GL30 preprocessing, the raster map was cropped over the Lombardy Region extent and thus aligned the exact DUSAF map region. No further preprocessing of GL30 was needed to accomplish the accuracy assessment here presented.

The DUSAF vector map required additional preprocessing steps. The most critical one was the selection of the pixel size to perform rasterization of the vector dataset. Rasterization is integral to perform both confusion matrix-based accuracy assessment as well as to the proposed sub-pixel errors detection procedure. The reason is better explained in the next sections. Rasterization step is a requirement due to the selection of the reference map (vector) for this work. The pixel size for rasterizing DUSAF was arbitrary set equal to 5 m to take into account the following factors. i) the theoretical minimum mappable unit that at the DUSAF scale (1:10000) is equal to 2 m. ii) the alignment of the output reference raster with the target map (30 m) being its pixel size an integer multiple of the selected one. iii) the output raster map dimension in memory to ease

further processing operations with limited computational and storage resources.

DUSAF class (level 3)	Target GL30 class
211, 213, 221, 222, 223, 224, 231	10
311, 312, 313, 314	20
321	30
322, 324	40
411	50
511, 512	60
111, 112, 121, 122, 124, 131, 132, 133, 134, 141, 142	80
331, 332, 333	90
335	100

Table 3. Reclassification rules adopted in the conversion of DUSAF classes to GL30 classes

The rasterization of the DUSAF was performed by considering the 3rd level of DUSAF classification hierarchy. Information about the DUSAF classes considered is reported in the Table 2. At this level, the description of the DUSAF classes is enough comprehensive to make the link to the classification legend of the GL30. Afterwards, DUSAF was reclassified to match the classes of GL30 according to reclassification rules reported in Table 3.

The 5 m resolution reclassified map was adopted for the sub-pixel classification errors analysis. The preprocessing of both target and reference map was carried out using the Free and Open Source Software (FOSS) GRASS GIS (https://grass.osgeo.org).

3.2 Data Processing for Sub-pixel Errors Analysis

In order to investigate spatial patterns of sub-pixel classification errors, a specific procedure was developed to enable counting the number of reference map pixels actually in disagreement or agreement - in terms of classification - with every single pixel of the target map. This allowed to exploit the higher resolution of the target map to obtain information also on partial disagreements and to arrange the results in a tabular format. The obtained data is used as input to analyse and map the disagreement spatial patterns. To accomplish the goals mentioned above, a new raster layer named ID was introduced to store information about the position of each GL30 pixel. ID raster was an exact copy of the GL30 raster in terms of grid size and resolution whereas each pixel value was substituted with a unique identifier *i,* e.g. an integer index (Figure 1a). The computation of the sub-pixel disagreements table is illustrated in Figure 1.

The ID, GL30 and DUSAF rasters were overlaid in such a way that each pixel of the GL30 and the ID (30 m) overlapped to 36 pixels of the DUSAF (Figure 1a). This number simply derives from the selected DUSAF pixel size that is 5 m. The overlay statistics were computed using the r.stats module of the GRASS GIS (https://grass.osgeo.org/grass70/manuals/r.stats). The DUSAF pixels "under" each ID/GL30 pixel were aggregated according to the class they belong to, keeping also track of their counts per class (Figure 1b). The introduction of the ID raster was crucial for the computation of per pixel counts. The GRASS GIS r.stats module provides only the possibility to compute cumulative pixel counts between classes of two or more raster maps at a different resolution. The artefact of introducing the ID raster allowed treating each GL30 pixel as a unique class thus enabling to compute DUSAF pixel counts at each GL30 pixel. The raw r.stats output table was processed using Python to

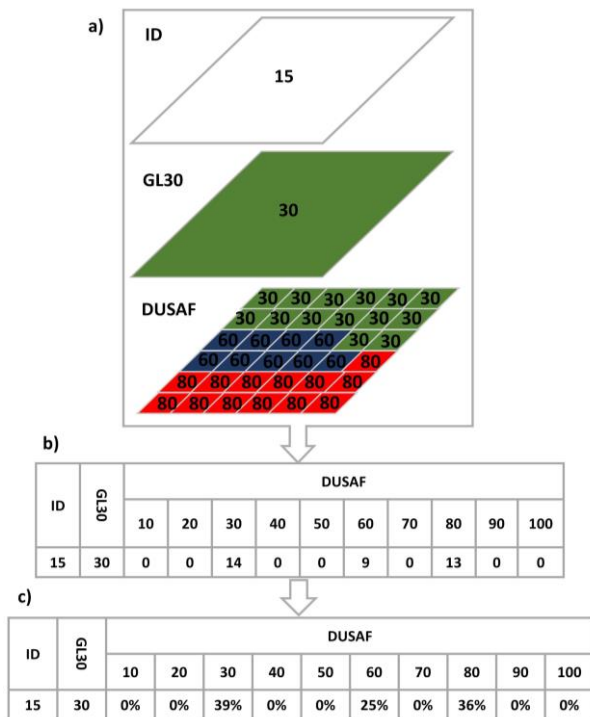compute per pixel counts (Figure 1b) and to replace them by percentages (Figure 1c).

Figure 1. Schematic of the data overlay procedure to compute sub-pixel error on a sample target map pixel i=15

By considering the percentage for the DUSAF class identical to the GL30 class, the accuracy of the GL30 pixel is obtained. On the other way round, the error for a single GL30 pixel can be computed as the sum of the percentages of DUSAF classes that differs from the class of the considered GL30 pixel. Intra-class and inter-class errors can be computed by selecting and combining table columns to describe disagreements for a specific class or a combination of classes. The computed error percentages (or counts) can be mapped as numerical spatial variables by tracing back pixel coordinates using the ID raster. This feature enables to explore disagreement spatial patterns as well as to compute spatial association statistics to describe them. Finally, the table with pixel counts (Figure 1b) can be used to compute the confusion matrix, so that traditional accuracy indexes can be estimated. The confusion matrix was computed by summing column wise all the elements of the table for all unique GL30 class values.

The number of entries for the full table - computed as in Figure 1 - was larger than 26 million that corresponds to the number of GL30 pixels covering the Lombardy Region (~ 23868 km²). Hence, operations on such a large file (> 10 GB) required a careful selection of best software tools to handle these data with limited computational and storage resources. This is the case of this work where a standard desktop machine was used. Tabular data processing in Python is commonly performed using the Pandas library (http://pandas.pydata.org). However, to cope with large tabular data the Dask library (http://docs.dask.org) was adopted which allows extending common interfaces like Pandas to larger-than-memory datasets as well as supports multithreading computation for data analysis. On the other hand, the use of standard commands of GRASS GIS proved to be sufficient to perform the processing of the raster maps involved in the analysis.

4. DATA ANALYSIS AND RESULTS

4.1 Traditional Accuracy Assessment

The GL30 accuracy with respect to the DUSAF has been already published (Bratic, et al. 2018a). In this former work, the approach used for computing the confusion matrix was slightly different than the one reported here. Namely, in Bratic et al. (2018a) the confusion matrix was computed using DUSAF raster map downsampled at 30 m. The exact confusion matrix derivation adopted in this work is described in section 3.2. It is equivalent to the confusion matrix that would be derived as if both reference and classified datasets had 5 m resolution. Nevertheless, the different approach did not incur significant changes in results (up to 1%). Table 4 includes the confusion matrix normalized by column (i.e. divided by the total number of pixels in each GL30 class). In Table 4, it can be observed that the agreement (diagonal values) of class 40 (Shrubland) is the lowest, and that the highest confusion (extra-diagonal values) is between class 40 and class 20 (Forest). This confusion is also evident from traditional accuracy indexes, e.g. the Producer's accuracy (PA) and the User's accuracy (UA) (Congalton, 2004) shown in Table 5. Due to the similar physical properties of these two classes, the error may be caused by the classification algorithm used for producing the map. In order to better investigate and describe errors features, spatial patterns of disagreements between class 40 and 20 is analysed. The next section focuses on this example to test the discuss the potential benefit of coupling the traditional accuracy assessment accuracy to exploratory and statistical spatial patterns analysis. Nevertheless, the same can be applied to the analysis of any intra-class and inter-class classification accuracy.

Class		GlobeLand30								
		10	20	30	40	50	60	80	90	100
DUSAF	10	90	11	1	20	35	9	30	1	0
	20	5	79	13	42	16	3	2	7	0
	30	1	3	51	13	4	0	1	7	0
	40	0	3	10	14	0	0	0	5	0
	50	0	0	0	0	35	1	0	0	0
	60	0	0	0	0	8	83	0	0	0
	80	3	2	0	2	1	2	66	0	0
	90	0	1	24	8	1	1	0	79	19
	100	0	0	0	0	0	0	0	0	81

Table 4. Normalized confusion matrix [%]

Class	PA	UA
10	90	82
20	79	79
30	51	40
40	14	23
50	35	39
60	83	94
80	66	82
90	79	79
100	81	88

Table 5. Producer's and User's accuracy [%]

Accuracy for the whole map is estimated by Overall Accuracy (OA) index (Congalton, 2004), and it is equal to 79%. The result for the map accuracy is satisfactory, but exploiting Grassland,

Shrubland or Wetland class exclusively is not recommended due to their low PA and UA. The results for the Lombardy Region are aligned to the ones reported in previous accuracy assessments of GL30 over larger regions (Bratic et al., 2018a; Brovelli et al., 2015; Chen et al., 2015)

4.2 Error Spatial Patterns Investigation

As a first experiment, a map of the global error of each GL30 pixel is created from the error full table computed in Section 3.2. Results provide visual insight into the general spatial patterns of agreements (pixel accuracy = 1), disagreements (pixel accuracy = 0), and partial disagreement (0 < pixel accuracy < 1) as shown in Figure 2. From the global error map (Figure 2c) is it possible to appreciate the global error spatial pattern that highlights a disagreement around peculiar map features such as rivers as well as forest areas (Figures 2a,b). A strong positive spatial association affecting the global error emerged with the presence of clusters of similar values (i.e. full/partial disagreements and full agreements) across the map (Figure 2c). To quantify this visually detected spatial patterns, the global Moran's I index (Moran, 1950) is introduced. This index allows for assessing at which degree similar values of a spatial variable - i.e. the pixel error - are also neighbours in space. The neighbouring relationship among pixels is here defined by means of a K-nearest neighbour spatial weights matrix (Getis, 2009) with K=8. The global Moran's I value ranges from -1 to 1. A negative value depicts negative association, a value close to 0 indicates randomness or no association, whereas a positive value denotes positive association. Both the spatial weight matrix as well as the global Moran's I index were computed exploiting functionalities of the PySAL Python library (https://pysal.org). The global Moran's I value computed for the global error map is 0.80 that confirms the marked positive spatial association visually detected on the global error map (Figure 2c).
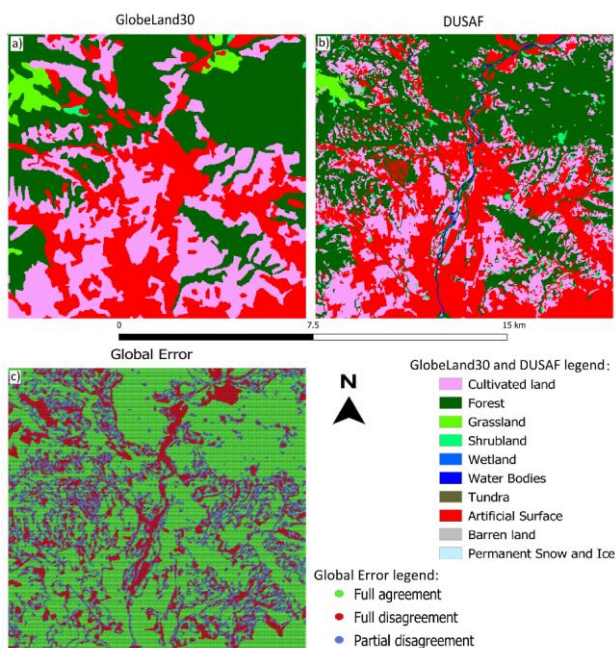

Figure 2. Portions of a) GlobLand30 and b) DUSAF raster map and c) global error map within the Lombardy Region

In principles, also local spatial association can be analysed by using the global error of each GL30 pixel and the same spatial weight matrix adopted above. Local Indicators of Spatial Association (LISA) such as the local version of the Moran's I

(Anselin, 1995) allows to describe spatial association at each location of the dataset, i.e. pixels in this case study. Outcome of LISA can be mapped to assess whether a pixel belong to a cluster (positive spatial association), an outlier (negative spatial association), or is not affected by significant spatial association (random). The local Moran's I map was computed for the global the global error of each GL30 pixel. However, the obtained results were not adding substantial information on the errors spatial patterns. A marked positive spatial association emerged as a result of diffuse clustering affecting the error along the whole region. This was also described by the global Moran's I value. The clusters pattern actually retraced the pattern of the full disagreement pixels. In views of the above, the LISA maps were not further considered for this case study.

Focusing on the lowest detected classification accuracy, i.e. the one of class 40, the analysis was repeated by using the errors between the highest and the lowest confusions for this class which are class 20 (Figure 3a) and class 80 (Figure 3b) respectively. Once again, it was possible to observe a marked cluster activity of the error across the maps which is stronger for the error between classes 40 and 20. The global Moran's I values are equal to 0.82 for classes 40 and 20 whereas equal to 0.62 for classes 40 and 80. This confirms the outcome of the visual assessment and provides credit to the outputs of the traditional accuracy assessment.
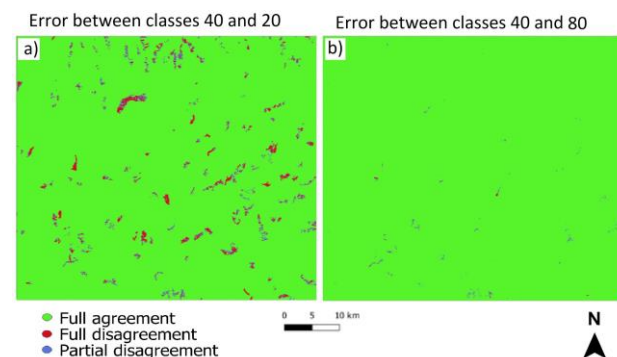

Figure 3. Example of inter-class error pattern maps

Finally, the visual exploration of the error maps pointed out additional local features in the disagreements as reported in Figure 4. By looking at the partial disagreements, it was encountered across the map that many patches of full disagreement pixels were surrounded by none partial disagreement pixels (Figure 4a).

Partial disagreements are expected along transitions between LC classes as well as within areas with heterogeneous LC characteristics such as where pixels of the target map overlaps multiple pixels of the reference map that partially belongs to the same class (Figures 4d,e,f). In principle, these patterns represent reasonable errors that are due to the generally lower representation quality of the target map with respect to the reference one. The fact that none of the partial disagreement pixels surrounds a full disagreement pixels patch (Figure 4a) may indicate instead relevant missing objects or underlying issues in the reference map reclassification. An example is reported in Figures 4a,b,c which addresses the case of an urban park. According to the GL30 classification legend (Table 1), urban parks should be classified as Artificial Surface (class 80). The DUSAF reclassification was performed accordingly (Figure 4b). The sharp transition between full agreement and full disagreements spotted on the global error map (Figure 4a) is

showing that the urban park is classified as Grassland instead as Artificial surface on the GL30.
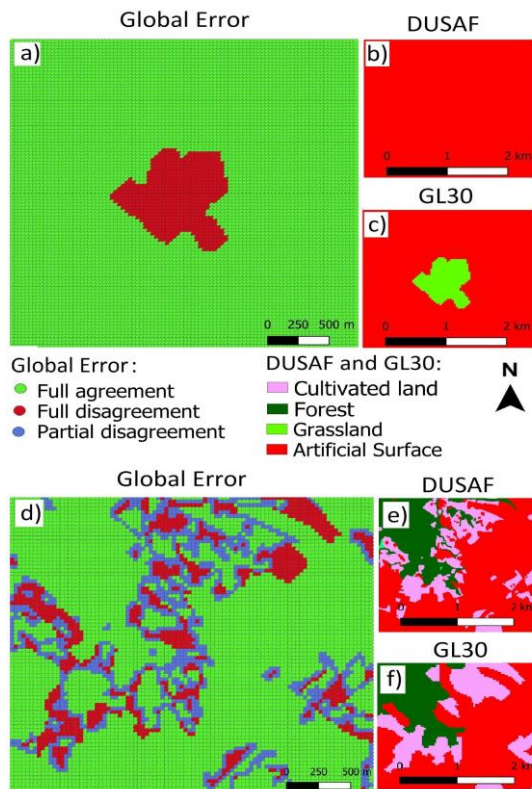

Figure 4. Example of visual exploration experiment on error maps

This ambiguity introduces an artificial error that cannot be directly traced back to the quality of the target map. The visual data exploration here proposed helps the analyst to better understand the dimension of the above undesired situation in performing accuracy assessment procedures.

4.3 Results Discussion

The traditional accuracy assessment analysis provided with robust indicators to describe the global accuracy of the GL30 map for the Lombardy Region. However, these indicators do not provide insight into the spatial distribution of errors. The visual inspection of the errors spatial patterns as well as their analysis by means of spatial association statistics introduced this capability into the accuracy assessment workflow. Results from the spatial association analysis - such as the Moran's I - allowed to uncover underlying patterns features providing alternative metrics to describe and quantify them by qualitatively linking the clustering of errors to the classification accuracies (Steele et al., 1998), as for the example described in Figure 3.

Despite the significant computational requirements, the sub-pixel disagreement table (Section 3.2) provides with a flexible and comprehensive input dataset to detailed accuracy assessments enabling - at the same time - local and global error patterns analysis. This can be considered the most valuable outcome of the presented work and it represents the starting point for future improvement of the proposed analysis.

Finally, the enabled possibility to nimbly display and explore classification errors on a map assists the analyst to critically review the accuracy assessment procedure that may result in more reliable judgment of the actual quality of the LC map.

5. CONCLUSIONS AND FUTURE WORK

This paper presented an experimental accuracy assessment procedure. The procedure was designed to take advantage of traditional accuracy assessment techniques and couple them with exploratory experiments on the data. The procedure was employed to perform the accuracy assessment of the GL30 global LC map for the Lombardy Region by means of comparison with the DUSAF regional LC map. Traditional procedures provided robust estimates of the global accuracy of the GL30 map. Exploratory experiments addressed the investigation of the errors spatial patterns by outlining interesting features of maps disagreement that may be later adopted to improve the quantitative accuracy assessment.

The experimental procedure required original and intensive data processing which was here addressed with the exclusive use of FOSS. The maturity, completeness, and reliability of such technology for spatial data management and analysis was verified during this work. FOSS provide the analysis with a potential to be empowered, replicated, and improved combined with sustainability in technology costs (Brovelli et al., 2017). Therefore, the use of FOSS technology is key to the future of this work as well as to the possible integration of the analysis pipeline e.g. into a GIS software module.

In views of the above, the future directions for the work will mainly focus on a critical review of the proposed accuracy assessment procedure in order to systematically integrate outcomes of the exploratory analysis on errors spatial pattern into the traditional accuracy assessment.

REFERENCES

Anselin, L., 1995. Local indicators of spatial association - LISA. *Geographical Analysis*, 27(2), 93-115.

Bontemps, S., Defourny, P., Radoux, J., Van Bogaert, E., Lamarche, C., Achard, F., Mayaux, P., Boettcher, M., Brockmann, C., Kirches, G. Zülkhe, M., 2013. Consistent global land cover maps for climate modelling communities: current achievements of the ESA's land cover CCI. In: *Proceedings the ESA Living Planet Symposium*, 9-13.

Bratic, G., Brovelli, M. A., Molinari, M. E., 2018a. A free and open source tool to assess the accuracy of land cover maps: implementation and application to Lombardy Region (Italy). *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* Volume XLII-3, 87-92.

Bratic, G., Molinari, M. E., Brovelli, M. A., 2018b. Validation of the global high-resolution GlobeLand30 land cover map in Europe using land cover field survey database - LUCAS. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* Volume XLII-4, 51-59.

Brovelli, M.A., Minghini, M., Moreno-Sanchez, R., Oliveira, R., 2017. Free and open source software for geospatial applications (FOSS4G) to support Future Earth. *International Journal of Digital Earth*, 10(4), 386-404.

Brovelli, M.A., Molinari, M., Hussein, E., Chen, J., Li, R., (2015). The first comprehensive accuracy assessment of GlobeLand30 at a national level: Methodology and results. *Remote Sensing*, 7(4), 4191-4212.

Campbell, J.B., 1981. Spatial correlation effects upon accuracy of supervised classification of land cover. *Photogrammetric Engineering and Remote Sensing*, 47(3), 355-363.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M. and Zhang, W., 2015. Global land cover mapping at 30m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103, 7-27.

Congalton, R., 2004. Putting the map back in map accuracy assessment. *CRS Press*.

Cui, L.B., Zhu, L., Springmann, M., Fan, Y., 2014. Design and analysis of the green climate fund. *Journal of Systems Science and Systems Engineering*, 23(3), 266-299.

Ente Regionale per i Servizi all'Agricoltura e alle Foreste (ERSAF) 2010. Uso del suolo in Regione Lombardia. I dati DUSAF. *http://www.geoportale.regione.lombardia.it/en* (29 March 2019)

Esch, T., Taubenböck, H., Heldens, W., Thiel, M., Wurm, M., Geiss, C., Dech, S., 2010. Urban Remote sensing – How can Earth observation support the sustainable development of urban environments?. In: *Proceeding of the ISOCARP Congress,* 1-11.

Esch, T., Marconcini, M., Felbier, A., Roth, A., Heldens, W., Huber, M., Schwinger, M., Taubenböck, H., Müller, A., Dech, S.W., 2013. Urban footprint processor - fully automated processing chain generating settlement masks from global data of the TanDEM-X mission. *IEEE Geoscience and Remote Sensing Letters*, 10, 617–1621.

Feddema, J.J., Oleson, K.W., Bonan, G.B., Mearns, L.O., Buja, L.E., Meehl, G.A., Washington, W.M., 2005. The importance of land-cover change in simulating future climates. *Science*, 310(5754), 1674-1678.

Getis, A. 2009. Spatial weights matrices. *Geographical Analysis*, 41(4), 404-410.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., Li, X., Fu, W., Liu, C., Xu, Y., Wang, X., Cheng, Q., Hu, L., Yao, W., Zhang, H., Zhu, P., Zhao, Z., Zhang, H., Zheng, Y., Ji, L., Zhang, Y., Chen, H., Yan, A., Guo, J., Yu, L., Wang, L., Liu, X., Shi, T., Zhu, M., Chen, Y., Yang, G., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z., Chen, J. and Chen, J., 2013. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing,* 34(7), 2607–2654.

Moran, P.A., 1950. Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2), 17-23.

Oxoli, D., Ronchetti, G., Minghini, M., Molinari, M., Lotfian, M., Sona, G., Brovelli, M.A., 2018. Measuring Urban Land Cover Influence on Air Temperature through Multiple Geo-Data—The Case of Milan, Italy. *ISPRS International Journal of Geo-Information*, 7(11), 421.

Pesaresi, M., Ehrlich, D., Ferri, S., Florczyk, A. J., Freire, S., Halkia, M., Julea, A. M., Kemper, T., Soille, P., Syrris, V., 2016. Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014. *Publications Office of the European Union*.

Pielke, R.A., 2005. Land use and climate change. *Science*, *310*(5754), 1625-1626.

Steele, B.M., Winne, J.C., Redmond, R. L., 1998. Estimation and mapping of misclassification probabilities for thematic land cover maps. *Remote Sensing of Environment*, 66(2), 192-202.

Tatem, A.J., Goetz, S.J., Hay, S.I., 2008. Fifty years of Earth observation satellites: Views from above have lead to countless advances on the ground in both scientific knowledge and daily life. *American Scientist*, 96(5), 390.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steininger, M., 2003. Remote sensing for biodiversity science and conservation. *Trends in Ecology & Evolution*, 18(6), 306-314.