MAPPING SPATIAL ACCURACY OF FOREST TYPE CLASSIFICATION IN JAXA’s HIGH-RESOLUTION LAND USE AND LAND COVER MAP

N. Tsutsumida 1,*, S. Nagai2, P. Rodríguez-Veiga 3,4, J. Katagi5, K. Nasahara6, T. Tadono7

1,7 Graduate School of Global Environmental Studies, Kyoto University, Kyoto, Japan - naru@kais.kyoto-u.ac.jp
2 Department of Environmental Geochemical Cycle Research, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Kanagawa, Japan - nagais@jamstec.go.jp
3 Centre for Landscape and Climate Research, University of Leicester, Leicester, UK - prv4@leicester.ac.uk
4 NERC National Centre for Earth Observation (NCEO), Leicester, UK
5 Graduate School of Life and Environmental Sciences, University of Tsukuba, Ibaraki, Japan - kemiskkn37@gmail.com
6 Faculty of Life and Environmental Sciences, University of Tsukuba, Ibaraki, Japan - nasahara.kenlo.gw@u.tsukuba.ac.jp
7 Earth Observation Research Center, Japan Aerospace Exploration Agency (JAXA), Ibaraki, Japan - tadono.takeo@jjaxa.jp

Commission III, WG III/2

KEY WORDS: LULC, Geographically weighted, Accuracy assessment, Forest type classification

ABSTRACT:
Accuracy assessment of forest type maps is essential to evaluate the classification of forest ecosystems quantitatively. However, map users do not understand in which regions those forest types are well classified from conventional static accuracy measures. Hence, the objective of this study is to unveil spatial heterogeneities of accuracies of forest type classification in a map. Four forest types (deciduous broadleaf forest (DBF), deciduous needleleaf forest (DNF), evergreen broadleaf forest (EBF), and evergreen needleleaf forest (ENF)) found in the JAXA’s land use / cover map of Japan were assessed by a volunteered Site-based dataset for Assessment of Changing LAnd cover by JAXA (SACLAJ). A geographically weighted (GW) correspondence matrix was applied to them to calculate the degree of overall agreements of forest type classes (forest overall accuracy), and the degree of accuracy for each forest class (forest user’s and producer’s accuracies) in a spatially varying way. This study compared spatial surfaces of these measures with static ones of them. The results show that the forest overall accuracy of the forest map tends to be relatively more accurate in the central Japan, while less in the Kansai and Chubu regions and the northern edge of Hokkaido. Static forest user’s accuracy measures for DBF, DNF, and ENF are better than forest producer’s accuracy ones, while the GW approach tells us such characteristics vary spatially and some areas have opposite trends. This kind of spatial accuracy assessment provides a more informative description of the accuracy than the simple use of conventional accuracy measures.

1. INTRODUCTION

Land cover (LC) maps describe the Earth’s terrestrial surface, encompassing all attributes of the biosphere (International Panel on Climate Change, 2000). LC is a major part of the Earth system which physically interacts with climate, topography, human impacts, and their complex interactions. Remotely sensed (RS) imagery, is often used to produce thematic LC maps aiming to understand wide ranges of terrestrial environments from the properties of the land cover itself such as the forest / non-forest classification (Hansen et al., 2013; Shimada et al., 2014), the urban extent (Schneider et al., 2009), the cropland distribution (Xiong et al., 2017) to the application of biodiversity estimation (Tuanmu and Jetz, 2014), carbon stock estimation (Rodríguez-Veiga et al., 2017), and potential ecosystem services assessment (Andrew et al., 2014). It is hence important to make an accurate LC classification map for high-quality quantification of these properties.

The classification of plant functional types (PFTs) is one of the main and most challenging tasks toward accurate LC mapping. PFT mapping can be considered as a land cover classification in which plants are grouped with regard to their physiology and physiognomy (Moncrieff et al., 2016). In Japan, the forest type classification (i.e. deciduous broadleaf forest (DBF), deciduous needleleaf forest (DNF), evergreen needleleaf forest (ENF), and evergreen needleleaf forest (ENF)) is required for multiple purposes, as the tree cover rate is reported as approximately 68.5% of the entire of the territory (The World Bank, 2018). However, the accurate classification in the four forest types from RS data has not yet been achieved. Due to the similarity of spectral characteristics captured by satellite sensors, even the separability between Evergreen and Broadleaf forest types is difficult (Sharma et al., 2016).

Some LC products are available to classify PFT-based forest types in Japan, MODerate-resolution Imaging Spectroradiometer (MODIS) Land Cover Type Products (MCD12Q1) (Friedl et al., 2010) has a 500 m spatial resolution with the world standard geosphere-biosphere programme (IGBP) LC class definition and the 12 PFT classification scheme. Climate Change Initiative (CCI) LC data set developed by the European Space Agency (ESA) describes LC by 22 thematic classes at 300 m spatial resolution at a global scale (Bontemps et al., 2013). The Ministry of Environment Japan investigates detailed PFTs in Japan by the combination of field survey and RS image analyses and publishes a 1 km resolution grid map (Biodiversity Center of Japan, 2018). These well-known products are coarse and may stem the mixed pixel problem: pixels will typically contain a mix of LC classes and the propensity for this is greater with medium and coarse spatial resolution data, as every pixel represents the dominant category and neglect the presence of others, resulting in the failure of accurate representations of LC due to the coarse pixel size (Fisher and Pathirana, 1990). Recently, the Japan
Aerospace Exploration Agency (JAXA) has developed and published a High-Resolution Land Use and Land Cover (HRLULC) map (version 18.03) with 10 thematic classes at the spatial resolution of 30 m for the whole Japan territory (JAXA, 2018) and is expected to overcome such issue.

It is important for users to know how accurate the LC classification map is. Incorrect conclusions or decisions may be derived from the map if misclassification in the map is neglected (Daly, 2006). The accuracy assessment provides a guide to the quality of the data and their reliability (Foody, 2002). The accuracy of the classification map is usually reported using a reference (ground-truthing) sample. A confusion matrix, which is a cross-tabulation of classification and reference sample categories, is built to report measures of overall, user’s, and producer’s accuracy (Congalton, 1991, Tsutsumida et al, 2016). Although these accuracy measures are utilized in order to report the degree of correspondence of the classification against a reference sample, they are averaged summaries of the correspondence, meaning that spatial variation of accuracy is not taken into account. Spatial heterogeneity of the accuracy may exist especially when the classification model does not consider spatial configuration of data. The knowledge of such spatial variations of accuracy is required to inspect the map not only for understandings of the spatial distribution of forest types accurately, but also for minimizing error propagation when we analyze this map for further applications. Previous studies have demonstrated a means to represent spatial heterogeneity of accuracy for categorical raster data (Comber et al., 2012, 2017; Comber, 2013; Congalton, 1988; Foody, 2005). They focused on spatial dependency and / or non-stationarity of accuracy measures which are often found in RS-based LC classification map. Specifically, Comber et al. (2017) proposed a geographically weighted (GW) correspondence matrix which is able to build a local correspondence matrix with a spatially weighted pairwise of classification and reference sample data. The GW uses a distance-decayed moving window or kernel over space to calculate any types of models and statistical measures (Gollini et al., 2015; Lu et al., 2014).

Hence, the objective of this study is to unveil spatial heterogeneities of accuracies for four forest type classes in a LC classification map. We use the JAXA’s HRLULC map of Japan for the case study. In order to validate estimated classes in the map against the actual land cover situation, we used a ground truth reference data set, called site-based dataset for assessment of changing land cover by JAXA (SACLAJ). A GW correspondence matrix was applied to these data sets to analyze the degree of overall agreements of forest type classes and the degree of accuracy for each forest class in a spatially varying way. This allows to map spatial heterogeneities of such degrees over space, hidden in conventional accuracy measures. Results provide us insightful information for the usefulness of the JAXA’s HRLULC map with special attention to local accuracies of forest types against the actual land covers.

2. MATERIALS

2.1 JAXA’s High-Resolution Land Use and Land Cover map (HRLULC map)

The JAXA’s HRLULC map version 18.03 at the 30 m spatial resolution is used shown in Figure 1 (JAXA, 2018). This data describes LC in the entire territory of Japan during the period 2014-2016 with 10 thematic classes: water; urban; rice paddy; crop; grass; DBF; DNF; EBF; ENF; and bareland. This product uses multiple geospatial data sets for inputs such as Landsat 8, digital elevation model provided by Geospatial Information Authority of Japan, ALOS-2/PALSAR-2, Sumoi National Polar-orbiting Partnership (NPP), road geospatial data in OpenStreetMap, and ground truth sampling data in SACLAJ (details below). A kernel density estimation in a Bayesian inference was applied to obtain the class posterior probability and produce the JAXA’s HRLULC map (Hashimoto et al., 2014).

Figure 1. JAXA’s land use land cover map.

2.2 SACLAJ

3,014 reference sample points in SACLAJ are used for validation in this study. Such sample points are randomly distributed across the entire territory of Japan and are captured during the period 2014-2016, which is the same as the target period of the JAXA HRLULC map. Approximately 300 data points are allocated in each class category of which definition is the same as the one in the JAXA’s HRLULC map. Scientists and engineers who visited field sites took geo-referenced pictures on the ground and interpreted them into one of the land cover categories (Nagai et al., accepted; Tadono et al., 2014). Such labels are also confirmed by very fine spatial resolution satellite images in geospatial visualization tools such as Google Earth. Since 2013, more than 50,000 data points with information on date, LC category, location, and size of the LC, have been recorded across Japan and all over the world.

3. METHODS

3.1 Correspondence matrix

For the purpose of the accuracy assessment of a classification map, it is common to build a correspondence matrix between the classified and reference values and to calculate some accuracy measures. At the point in total L sample points considered, a pair of the classified class (c_i) and the reference class (r_j) is obtained. A correspondence matrix (M) is built so that rows represents classified data, and columns represent reference data, and the number of correspondence pairs between c_i and r_j corresponding to the class i in the sample is stored at the ith row and jth column in the M. Similarly, the number of pairs between c_i corresponding to the class i and r_j corresponding to the class j in the sample is stored at the jth rows and the ith columns in the
Thus, the element $m_{ij}$ of the $M$ at the $i$th row and the $j$th column is:

$$m_{ij} = \sum_{k=1}^{L} a_{ij(\ell)}$$

$$a_{ij(\ell)} = \begin{cases} 1, & \text{if } c_\ell = i \text{ and } r_\ell = j, \\ 0, & \text{otherwise} \end{cases}$$

(1)

where $a_{ij(\ell)}$ is a binary response whether the pair of the classification class $c_\ell$ and the reference data $r_\ell$ at the $\ell$th of $L$ sample points are the class $i$ and the class $j$, respectively, or not. It is noted that a coordination $(u_\ell, v_\ell)$ at the $i$th sample point is not considered in the $m_{ij}$. In this way, $M$ represents a summary of the degree of agreement and disagreement among classes.

Generally, accuracy measures are calculated from the matrix $M$ to assess classification maps and to tell data users how the degree the correspondence between classified and reference class at the sample points is accurate. The overall accuracy describes the proportion of the number of pairs having the same class in the classified and reference data for all classes in total. User’s and producer’s accuracies are also often reported to represent how the degree the classification is accurate at a class level. User’s accuracy describes the proportion of the number of classified class at sample points that actually corresponds to the classification class, whereas producer’s accuracy is the proportion of the number of reference class at sample points that correctly corresponds to the classified class.

In this study, we would like to know how much correctly forest types are classified into four forest classes. We extend the concept of overall, user’s and producer’s accuracy measures to focus only on forest related classes. We define forest overall, forest user’s and forest producer’s accuracies as:

$$FO = \frac{\sum_{i=1}^{n} m_{ii}}{\sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij}}$$

(2)

$$FU_i = \frac{m_{ii}}{\sum_{i=1}^{n} m_{ii}}$$

(3)

$$FP_i = \frac{m_{ii}}{\sum_{i=1}^{n} m_{i*}}$$

(4)

where $FO$ is forest overall accuracy, and $FU_i$ and $FP_i$ are forest user’s and producer’s accuracies of the class $i \in f_1, \ldots, f_n$, respectively. Here the $f_n$ is the $n$th forest class and in this study, $n = 4$. $\sum_{i=1}^{n} m_{ii}$ denotes the marginal sum of $m$ for the classification class $i$ in $n$ forest classes, and $\sum_{i=1}^{n} m_{i*}$ denotes the marginal sum of $m$ for the reference class $i$ in $n$ forest classes.

These measures are averaged values with different aspects of map accuracy. Although the result is represented by a ‘value’ which describes how much degrees the overall agreement rate between the classified and the reference class, the classified class corresponds to the one in the reference class, or the reference class corresponds to the one in the classified class, these measures do not consider any spatial configurations in the sample points. There have been argued that they only function when errors are assumed to be spatial homogeneous, while the error in the classification map often represents spatial heterogeneity (Comber et al., 2017; Comber, 2013; Foody, 2005; Tsutsumida and Comber, 2015; Tsutsumida et al., 2019). It is important to tackle with this issue.

### 3.2 Geographically weighted correspondence matrix

A GW correspondence matrix has been introduced by Brunsdon et al. (2016) and Comber et al. (2017). It builds locally weighted correspondence matrix at each location by a moving-window distance-decayed kernel. Again, we consider the equation (1) but deal with the spatial structure of the sample points by applying GW framework. Here at the location $k$ on a grid with the spatial resolution of $r$ covered in the study area and $w_k$ is a bisquared kernel defined as:

$$w_k = \begin{cases} \left(1 - \frac{|d_k|}{b}\right)^2, & \text{if } |d_k| < b, \\ 0, & \text{otherwise} \end{cases}$$

(5)

where $b$ is the bandwidth size, which is arbitrary determined. Larger $b$ tends to describe more spatial homogeneity (stationary) closer to the global (conventional) measures of accuracy. The total number of grids $K$ is determined by the spatial resolution $r$ in the study area. The weighted element $m_{ij}$ of the local matrix $M_k$ at the location $k$ is now replaced by the $\rho_{ij(k)}$:

$$\rho_{ij(k)} = \frac{\sum_{k=1}^{K} a_{ij(k)} w_k}{\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij(k)} w_k}$$

(6)

Sample points found within the distance $b$ from the location $k$ are only used for the calculation of the GW correspondence matrix. As this way, at location $k$, the GW forest overall ($GWFO_{(k)}$), and GW forest user’s, and GW forest producer’s accuracy measures for the class $i$ ($GWFU_{i(k)}$ and $GWFP_{i(k)}$, respectively) are:

$$GWFO_{(k)} = \frac{\sum_{i=1}^{n} \rho_{ii(k)}}{\sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij(k)}}$$

(7)

$$GWFU_{i(k)} = \frac{\rho_{ii(k)}}{\sum_{i=1}^{n} \rho_{ii(k)}}$$

(8)

$$GWFP_{i(k)} = \frac{\rho_{ii(k)}}{\sum_{i=1}^{n} \rho_{i*}(k)}$$

(9)

where the class $i, j \in f_1, \ldots, f_n$ $\sum_{i=1}^{n} \rho_{ii(k)}$ denotes the marginal sum of $\rho$ for the classification class $i$ in $n$ forest classes, and $\sum_{i=1}^{n} \rho_{i*}(k)$ denotes the marginal sum of $\rho$ of the reference class $i$ in $n$ forest classes. In this study, $b$ and $r$ are defined as 200 km and 5 km, respectively, to show regional variations of accuracy measures well with relatively less calculation time.

### 4. RESULTS

In order to confirm the whole-map accuracy at first, the conventional correspondence matrix with JAXA’s HRLULC map and SACLAJ reference sample was built (Table 1). The conventional overall accuracy is 82.2% for 10 classes. This table suggests relatively better classifications for water (user’s accuracy: 95.3%), producer’s accuracy: 97.7%), while some mis-classifications are also found such as between urban and bareland, and between crop and grass. When focusing on the classification between aggregated forest and non-forest classes, 97.4% of the overall agreement rate is achieved, suggesting the high performance of the forest/non-forest distinction. From this result, our interests of accuracy in this map move forwards to the performance of the four forest types classification.
Table 2 represents a summary of FU and FP. All values in FU for four forest types except the ENF class are larger than those in FP, suggesting that these forest classes in the classification map tend to be identified at the reference points correctly well, compared to that in the reference against the classification map. The ENF class shows the opposite pattern. The FO is 82.2%.

<table>
<thead>
<tr>
<th>Forest Type</th>
<th>FU (Accuracy)</th>
<th>FP (Accuracy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBF</td>
<td>83.2%</td>
<td>86.3%</td>
</tr>
<tr>
<td>DNF</td>
<td>84.2%</td>
<td>76.4%</td>
</tr>
<tr>
<td>EBF</td>
<td>81.8%</td>
<td>84.8%</td>
</tr>
<tr>
<td>ENF</td>
<td>73.4%</td>
<td>88.7%</td>
</tr>
</tbody>
</table>

Table 2. Forest user’s and producer’s accuracy for deciduous broadleaf forest (DBF), deciduous needleleaf forest (DNF), evergreen broadleaf forest (EBF), and evergreen needleleaf forest (ENF).

As expected, such measures inform us the degree of correspondence of classes in the whole map, however it lacks spatial information on accuracy. Thus, GW correspondence matrix was applied to show spatial heterogeneities in accuracy measures. Figure 2 shows a result map of GWFO. It clearly shows spatial heterogeneity of the FO hidden in the conventional overall accuracy measure. The GWFO varies around the FO: 82.2% and on the whole, but it tends to be more accurate in central Japan while less in the Kansai (western mainland) and the northern edge of the main island and Hokkaido (north island).

Maps of both GWFU and GWFP shown in Figure 3 and 4 demonstrate clear spatial variations of accuracy. The highest variation can be found in the EBF for both of the FU and the FP (the standard deviations are 25.3 and 30.5, respectively). Surfaces of accuracy in some areas in the Japanese territory lacks due to insufficient number of classification and reference sample points, implying that there is no distribution of the class found in such areas. Each map indicates potential spatial clusters of high/low accuracy. For example for the DNF, both FU and FP represent a cluster with high accuracy (over 90%) in the central mainland. It is reasonable to find highly accurate clusters in such regions as the DNF is distributed in cooler regions such as Hokkaido and high-elevated mountainous areas in Japan (mostly found in the central Japan) (Oshida et al., 2009). In the meantime, the DNF are relatively less found in other areas, meaning the...
Accurate forest type classification is challenging. To understand the quality of a LC classification map, data users have to inspect the accuracy of the map in which regions are well described or the quality of a LC classification map, data users have to inspect areas which have opposite trends from global results. For example, FP is larger than FU for the DBF, in the central mainland and the northern edges of Hokkaido for the DBF, and in the southern Hokkaido and the north mainland for the DNF, and in the western mainland for the EBF. In contrast, FP is larger than FU for the ENF according to Table 2, however, some areas with larger FU than FP can be found in the middle of Hokkaido and the north-western mainland. These spatial characteristics of accuracy are hidden in the global accuracy diagnostics. Such maps of spatial heterogeneities of accuracy enable data users to pay special attention to the regional difference of the quality of the JAXA’s HRLULC map and to areas which have opposite trends from global results.

There are some technical discussions which help build bridges to further potential studies. Firstly, the bandwidth size is arbitrary determined (in this study, 200 km). This was explored preliminarily and determined so that the kernel enables to cover sufficient reference sample points over space. For a conventional GW regression (GWR) and other types of GW regression models such as generalized GWR, some optimization approaches for the bandwidth size have been developed. One of the popular methods is a leave-one-out cross validation (LOOCV) (Gollini et al., 2015; Lu et al., 2014). The LOOCV approach can also be applicable to GW summary statistics (Brunsdon et al., 2002; Fotheringham et al., 2002). However, there have not yet proposed such optimization approaches for the GW correspondence matrix. Even if the bandwidth can be optimized for the GW correspondence matrix, it does not produce optimized GW accuracy measures, because the LOOCV calculation requires a statistic which represents a nature of the matrix, and it is not optimal for accuracy measures. Furthermore, the LOOCV can be applicable to each of GW accuracy measures, however optimized bandwidth would not be consistent through measures, resulting in the failure of their comparisons. The topic on bandwidth optimization is highly linked with the question which spatial scale level is appropriate to investigate local accuracy in the map. Hence new developments for the bandwidth optimization are expected. Similar discussion can be found in Tutsuimda et al. (2019). Secondly, the quality of the labelling in reference samples is not considered in this study. Ultimately, any results of accuracy assessments are relative to the reference sample. The SACLAJ reference data set is well organized collected from volunteers who visited field sites and/or from very fine spatial resolution satellite imagers. However, no reports on the quality are found. Previous studies have argued the impact of the use of imperfect reference sample to accuracy assessments (Foody, 2011; Foody et al., 2016; Zhao et al., 2017) and such findings would be helpful. Furthermore, potential studies can focus on the uncertainty between a classification map and imperfect reference sample when these reliabilities are available, by applying GW fuzzy differences (Comber et al., 2012) and/or GW mean absolute error (Tutsuimda et al., 2019). Lastly, our approaches evaluated local accuracy but did not contribute to the improvement of the classification. Future studies will focus on the usefulness of such information on spatial accuracy to improve the classification. A spatial backpropagation approach based on the GW error measures would be worth to be investigated toward minimizing the spatial heterogeneity of accuracy.
6. CONCLUSION

The forest type classification, a major part of the PFT, is essential to estimate aboveground biomass, carbon cycle, biodiversity, and many other environmental studies. Accurate forest type classification maps are required for these applications, however such map has not yet been achieved. This study demonstrated how to estimate the spatial accuracy of classification in four forest types in Japan to inform users the reliability of the classification map. Three types of accuracy measures tell different aspects of local accuracy: the degree of overviews of the correspondence of the classification to reference sample; the degree of how much reference sample points are correctly classified in the predicted map; and the degree of how much classified sample are found in the reference sample. Particularly, this study applied geographically weighted approaches to the forest type classification in the JAXA’s HRLULC map and highlighted spatial heterogeneities of these three types of accuracy measures. The spatial accuracy provides a more informative description of accuracy than the simple use of some conventional accuracy measures.

ACKNOWLEDGEMENTS

This research is funded by ROIS-DS-JOINT (006RP2018) and joint research program of CERES, Chiba university (2018).

REFERENCES


