Research Article

A Suite of Tools for Assessing Thematic Map Accuracy

**Jean-François Mas,¹ Azucena Pérez-Vega,² Adrián Ghilardi,¹ Silvia Martínez,¹
Jaime Octavio Loya-Carrillo,¹ and Ernesto Vega³**

¹ Centro de Investigaciones en Geografía Ambiental, Universidad Nacional Autónoma de México,
Antigua Carretera a Pátzcuaro 8701, Colonia Ex-Hacienda de San José de La Huerta, 58190 Morelia, MICH, Mexico

² Dirección de la División de Ingenierías, Universidad de Guanajuato, Avenida Juárez 77, Zona Centro,
36000 Guanajuato, GTO, Mexico

³ Centro de Investigaciones en Ecosistemas, Universidad Nacional Autónoma de México, Antigua Carretera a Pátzcuaro 8701,
Colonia Ex-Hacienda de San José de La Huerta, 58190 Morelia, MICH, Mexico

Correspondence should be addressed to Jean-François Mas; jfmas@ciga.unam.mx

Received 14 March 2014; Revised 7 May 2014; Accepted 18 May 2014; Published 11 June 2014

Academic Editor: Laerte G. Ferreira

Copyright © 2014 Jean-François Mas et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Although land use/cover maps are widely used to support management and environmental policies, only some studies have reported their accuracy using sound and complete assessments. Thematic map accuracy assessment is typically achieved by comparing reference sites labeled with the “ground-truth” category to the ones depicted in the land use/cover map. A variety of sampling designs are used to select these reference sites. The estimators for accuracy indices and the variance of these estimators depend on the sampling design. However, the tools used to assess accuracy available in the main program packages compute the accuracy indices without taking into account the sampling and give inconsistent estimates. As an alternative, we present free user-friendly tools that enable users beyond the Geographic Information Science Community to compute accuracy indices and estimate corrected areas of given categories with their respective confidence intervals. The tool runs in Dinamica EGO, a free platform for environmental spatial modeling as well as a Q-GIS plugin and a R package. Additionally, a practical application example is described using a case study area in central-west Mexico.

1. Introduction

Thematic maps such as land use/cover maps are widely used to support management and environmental policies and therefore they should be supported by a statistically rigorous, credible accuracy assessment [1, 2]. Thematic accuracy is a measure of correctness that can be defined as the degree to which the attributes of a map agree with “truth” reference datasets. Accuracy assessment is typically based on a sample of reference sites to which the “true” land use/cover category is compared to the one in the map.

A variety of sampling designs can be used to select these reference sites (sample units). The objectives, the desirable criteria, and the resources of the assessment have to be taken into account to choose the sampling design. First of all, the sampling design should be a probability sampling design,

which means that the sample unit is selected randomly; the inclusion probability for each sample unit into the sample is known and must be greater than zero for all the units in the area under assessment. Probability sampling enables statistical inference allowing the computing of accuracy estimates along with their confidence intervals. Convenient procedures such as selecting training data used during supervised classification or by limiting the random sampling of reference sites to accessible sites or area covered by available high resolution images do not fulfill these requirements and cannot be considered as probability sampling procedures [3].

The most commonly used probability sampling designs are simple random sampling, systematic sampling, and stratified random sampling. In the simple random sampling design each sampling location is equally likely to be selected; that is, all the locations have the same inclusion probability (equal

probability sampling). The advantages of this design are its simplicity: the equations used to calculate the standard errors are less complex than in other designs and the sample size can be augmented or reduced easily. However this design may not produce appropriate sample sizes for rare categories to provide estimates with acceptable confidence intervals. Simple systematic sampling is achieved by selecting units using a systematic pattern, such as a grid. Systematic sampling is easy to carry out, gives a good spatial coverage, and is generally more precise than random to assess overall accuracy [3]. As in the simple random sampling design, rare categories will be rare in the sample because it is an equal probability sampling. Simple random and systematic sampling designs do not enable user to focus the sampling on a particular region or category.

When users are interested in obtaining more detailed information of a particular subregion or specific category, then a stratified sampling should be used. In stratified sampling, the area under assessment is divided into various subregions (strata) and each stratum is sampled independently. For example, a simple random sampling is applied in each stratum (stratified random sampling). The categories of the map under assessment are often used to stratify the sampling. In that case, the stratification may be used to guarantee a minimum sampling size in each stratum and obtain more precise estimates for rare categories. This approach may also enable users to adapt the stratum sample size to the precision requirements of each category according to the objectives of the study. Sample size may be augmented for category of interest which requires a precise accuracy estimate and reduced for less important categories, improving cost-effectiveness. It is worth noting that in these cases the number of sampling units per category is not proportional to the category area (nonequal probability sampling design). Equations that needed to calculate the accuracy indices and their confidence intervals in stratified sampling are more complex than those used in simple random or systematic sampling. This is because estimates combining data across strata must weigh the unequal inclusion of probabilities that result from “forcibly” allocating sampling points in subrepresented areas that would seldom host validation plots when using a simple random or systematic sampling [4].

Cluster sampling is also a popular design that reduces the cost of collecting data by constraining unit samples to fall within a limited number of sites (clusters). However, it introduces a larger spatial correlation in the sample data, reducing the precision of the accuracy estimates [5]. We do not take into account this sampling design in the present study. A detailed review of the basic sampling designs can be found in Stehman [3, 6] and Stehman and Czaplewski [2].

Although the methods to carry out accuracy assessment are well established, few studies producing land use/cover or land use/cover change maps present sound and complete accuracy assessments [7] partly because it is not a straightforward procedure and because mainstream GIS or satellite image processing software programs only provide incomplete tools to carry out such assessment. For instance, GIS software often has built-in tools to carry out accuracy assessment, but these are limited to cases of equal probability

samplings designs where the number of sample units by category is proportional to the category area, for example, the simple random sampling. Equations used to estimate accuracy indices depend on the sampling design, which in most cases is a stratified sampling [7–9]. In case of providing data obtained by nonequal probability sampling, as stratified random sampling, estimates of accuracy provided by these tools are erroneous. Moreover, usually these tools do not provide information about the certainty of the estimates (confidence intervals) neither estimates of area adjusted to eliminate bias due to map classification errors.

This paper presents a set of free tools that are readily available to the public and which enable users to compute accuracy indices and to estimate a corrected area of a given category and construct confidence interval (CIs) for quantifying the uncertainty of estimates. The aim is to aid users beyond the Geographic Information Science Community to adopt statistically sound accuracy assessment methods as part of a routine practice. With land use/cover map applications growing sharply to support environmental management, policy strategies, and even scientific hypotheses, we expect that these free and easy to use tools will boost accuracy assessments in both the academic and policy sectors. The tools are implemented in Dinamica EGO, Q-GIS, and R.

In Section 2, we briefly describe the software programs we used and review the method behind the tools that produces a statistically rigorous report of accuracy of any given categorical map, including the estimation of correct areas within CIs. In Section 3 we apply the tool to assess the accuracy of a 2010 land use/cover map for central Mexico.

2. Material and Methods

2.1. Software Packages. Dinamica EGO (<http://www.csr.ufmg.br/dinamica/>) is a free platform for environmental spatial modeling that enables the design of complex dynamic spatial models [10, 11]. Examples of Dinamica EGO models include land use/cover change modeling [10, 12, 13], rent and opportunity costs [14], assessment of the cobenefits of REDD [15], and ROC analysis [16].

R (<http://www.r-project.org/>) is an open source language and environment for data manipulation, statistical analysis, and graphic elaboration. A large number of packages (collections of R functions and compiled code) are available for download and installation from the CRAN package repository (<http://cran.r-project.org/web/packages/>).

Q-GIS (<http://www.qgis.org/>) is a popular cross-platform open source desktop geographic information system (GIS) software program that provides both vector and raster data viewing, editing, and analysis capabilities. Plugins, written in Python, extend its capabilities. Q-GIS enables also users to run R scripts.

2.2. Accuracy Estimates. In order to assess the accuracy of a map with q categories, a sample of reference sites (e.g., pixels) is selected by systematic, simple random, or stratified random sampling (using the map categories as strata). A confusion matrix, also referred to as the error matrix, is constructed by

using the sample counts. The map categories ($i = 1, 2, 3, \dots, q$) and reference categories ($j = 1, 2, 3, \dots, q$) are represented by rows and columns, respectively, (Table 1).

For stratified sampling, the number of samples for each map category is not necessarily proportional to the area covered by each category. This lack of proportion should be taken into account when calculating accuracy indices. Prompted by the considerations suggested by Card [8], we adjusted the confusion matrix derived from a stratified sampling by weighing the number of sites using the area of each category on the map.

As an example, we can consider a confusion matrix with q columns and q lines as the matrix in Table 1, where each element of the array n_{ij} is replaced by $\hat{p}_{ij}$ which is an unbiased estimator of the proportion of area using

$$\hat{p}_{ij} = \frac{\pi_i n_{ij}}{n_{i+}}, \quad (1)$$

where π_i is the proportion of area of category i in the map, n_{ij} is the number of samples mapped as i and belonging to category j in the reference data, and n_{i+} is the number of samples mapped as i in the map.

In this new adjusted matrix, each cell element $\hat{p}_{ij}$ represents the probability that a randomly selected area is classified under category i in the image and under category j in the reference data. As a consequence, the sum of the cells of each row $\hat{p}_{i+}$ is equal to π_i , which is the proportion of category i in the map. Based on this matrix, the computing of the overall, user, and producer accuracy indices is carried out as described in (2), (3), and (4).

The overall accuracy $\hat{O}$ is the overall proportion of area correctly classified and calculated by adding the $\hat{p}_{ii}$ values of the diagonal matrix as follows:

$$\hat{O} = \sum_{k=1}^q \hat{p}_{kk}, \quad (2)$$

where q is the number of categories. User accuracy $\hat{U}_i$ and producer accuracy $\hat{P}_j$ are, respectively, calculated using (3) and (4). Producer accuracy, which is related to omission errors, shows that proportion of the reference sample of a particular category is correctly classified in the map. User accuracy, related to commission error, is the proportion of samples classified as a particular category in the map which are correctly classified:

$$\hat{U}_i = \frac{\hat{p}_{ii}}{\hat{p}_{i+}}, \quad (3)$$

$$\hat{P}_j = \frac{\hat{p}_{jj}}{\hat{p}_{+j}}. \quad (4)$$

For stratified sampling, the CIs for the overall, producer, and user accuracy estimates are calculated as follows [8]:

$$\text{HCI}_{\hat{O}} = z \sqrt{\sum_{i=1}^q \hat{p}_{ii} \frac{(\pi_i - \hat{p}_{ii})}{n_{i+}}}, \quad (5)$$

TABLE 1: Confusion matrix expressed in sample counts.

Map	Reference						Sum n_{i+}
	1	2	...	j	...	q	
1	n_{11}	n_{12}	...	n_{1j}	...	n_{1q}	n_{1+}
2	n_{21}	n_{22}	...	n_{2j}	...	n_{2q}	n_{2+}
...	...	...	...	...	...	...	...
i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{iq}	n_{i+}
...	...	...	...	...	...	...	...
q	n_{q1}	n_{q2}	...	...	...	n_{qq}	n_{q+}
Sum n_{+j}	n_{+1}	n_{+2}	...	n_{+j}	...	n_{+q}	

where $\text{HCI}_{\hat{O}}$ is the half-width CI of the overall accuracy and z corresponds to the percentile of the normal distribution (for 95% confidence, $z = 1.96$):

$$\text{HCI}_{\hat{U}_i} = z \sqrt{\frac{\hat{p}_{ii} (\pi_i - \hat{p}_{ii})}{\pi_i^2 n_{i+}}}, \quad (6)$$

where $\text{HCI}_{\hat{U}_i}$ is the half-width CI of the user accuracy for category i :

$$\text{HCI}_{\hat{P}_j} = z \left(\hat{p}_{jj} \hat{p}_{+j}^{-1} \left[\hat{p}_{jj} \left(\sum_{i \neq j}^q \frac{\hat{p}_{ij} (\pi_i - \hat{p}_{ij})}{n_{i+}} \right) + \frac{(\pi_j - \hat{p}_{jj}) (\hat{p}_{+j} - \hat{p}_{jj})^2}{n_{j+}} \right] \right)^{1/2}, \quad (7)$$

where $\text{HCI}_{\hat{P}_j}$ is the half-width CI of the producer accuracy for category j .

This method is also applicable to designs which include simple random and systematic sampling. When no stratification is applied to these sampling designs, the stratified estimator is referred to as a “poststratified” estimator to distinguish between using stratification in the sampling design (i.e., stratified sampling) and using stratification in the estimator (i.e., poststratified estimation). This improves the estimates of accuracy indices because simple random and systematic samplings do not guarantee that the proportion of the various categories among samples is exactly the same as the proportion for the map category areas.

CIs can also be estimated by bootstrap stratified resampling [17, 18]. In order to carry out bootstrapping, the tool enables the replication sample sets by resampling with replacement from the original sample set. It uses stratification to insure that each sample has the same proportion of each category as in the original sample. The computing of accuracy indices is performed on each replicated sample. Then CIs are estimated using the bootstrap percentile interval method, which uses the empirical quantiles of the bootstrap replicates.

2.3. Area Estimates. Due to asymmetrical classification errors, the area of a particular category directly obtained from

the map (e.g., by pixels counting) is likely to be biased [4]. For example, the area of a category systematically affected by an omission error will be underestimated whereas the area of a category mainly affected by a commission error will be overestimated in the map. Therefore, areas obtained from the map should be adjusted to eliminate bias due to map classification errors and these error-adjusted area estimates have to be accompanied by confidence intervals that reflect their uncertainty. This method enables users to estimate error-adjusted areas using information from the accuracy assessment sample to correct the bias in area estimates [7, 8, 19]. In the error-adjusted matrix, the sum of the elements of column p_{+j} is an unbiased estimator of the proportion of the area of category j . Therefore, the area of category j , $\hat{A}_j$ is calculated by

$$\hat{A}_j = A_{\text{tot}} \cdot \hat{p}_{+j}, \quad (8)$$

where A_{tot} is the total area.

Equation (9) gives the estimated half-width confidence interval for the estimated area proportion p_{+j} :

$$\text{HCl}_{\hat{p}_{+j}} = z \sqrt{\sum_{i=1}^q \left(\frac{\hat{p}_{ij} n_{i+}}{n_{ij}} \right)^2 \frac{n_{ij}/n_{i+} (1 - n_{ij}/n_{i+})}{n_{i+} - 1}}. \quad (9)$$

2.4. How Do the Tools Work? Dinamica EGO models are designed as workflows that execute sequences of geoprocessing operations and are constructed by dragging and connecting data “functors” (data operators) in a model diagram displayed in the graphic interface. Models can be saved as submodels and stored as new functors in the functor library, thus helping users to better organize and share models [11, 16]. For this, new library called “Accuracy Assessment” composed of five submodels was created. It enables a user to carry out various operations related to accuracy assessments using maps in raster format. These operations include the construction of the confusion matrix, the bias-adjustment of this matrix using Card’s method [8], computing estimates of accuracy indices, and the error-corrected area estimates along with their CIs (Table 2). The confidence intervals are estimated by using estimations described in the previous section or by bootstrapping. The library, along with the application data and submodels, which integrate the tool, is available for downloading at <http://www.ciga.unam.mx/ciga/images/proyectos/vigentes/modelos/images/AccAssess.zip>. A brief and concise user’s manual based in this paper is also available.

The Q-GIS plugin “AccurAssess” enables the user to carry out the accuracy assessment using vector or raster inputs map. The tool computes the bias adjusted of this matrix, the estimates of accuracy indices, and the error-corrected area estimates along with their CIs. The R package “MapAccurAssess” has to be fed with the raw matrix and with a two-column text table which give for each reference site the mapped and the true categories. This kind of table is easily obtained through map overlay within a GIS. It enables user to calculate bias-adjusted matrix, the estimates of accuracy

indices, and the error-corrected area estimates along with their CIs.

3. Tool Application for a Case Study Area in Central-West Mexico

We applied the tool on a 2010 land use/cover map for the *Ayuquila* basin (411,500 ha) in central-west Mexico (Figure 1). The map was produced by visual interpretation (monoscopic) of SPOT5 images projected over a computer screen at 1:40,000. Six SPOT5 scenes were used with a processing level of 2A, which were acquired on November 11, December 12, and December 17, 2010. All six scenes were mosaicked followed by a fusion between the panchromatic (2.5 m) and color (10 m) bands for spatial enhancement. A first order polynomial transformation was conducted for rectifying the images, using nearest neighbor resampling and a threshold value of 2.5 m for residuals (i.e., less than image resolution). Finally, two band arrangements: 1, 2, 3 and 4, 1, 5 were used to “bring forward to the eye” different characteristics of land use/cover classes. Various image features were considered by the interpreter such as color, texture, shade, and tone. Classified polygons (i.e., vectors) were converted into a 100 m resolution raster map.

A set of 110 reference plots (one hectare each) were distributed following a stratified design based on land use/cover categories. This was done in order to compensate for less representative categories such as bare land or riparian forests. A second independent interpreter classified all reference plots using the same imagery and approach described above but at scale 1:5,000. This was possible given the resolution of the SPOT5 imagery used (2.5 m). Finally, all reference plots were labeled according to the land use/cover category covering the plot. In cases where more than one category was present, the category covering the majority of the plot was selected.

The raw confusion matrix (Table 3) presents the number of reference samples. Rows correspond to land use/cover categories in the classified raster map (Figure 1) and columns correspond to land use/cover categories used for labeling reference plots when classified at 1:5,000, (i.e., “true” category).

As shown in Table 4, the number of samples by category is not proportional to the category’s area in the map due to stratification. Categories such as 3, 5, 6, 8, 9, 11, 13, and 17 are better represented in the sample set than in the map. The extreme case is category 17 with 4.5% of the reference plots belonging to this category, which covers only 0.01% of the map. As a consequence this bias has to be corrected before computing accuracy indices.

Table 5 represents the bias adjusted matrix based on (1) using the Dinamica submodel *Generate Confusion Matrix*, the Q-GIS plugin, or the R package.

Accuracy indices along with their respective half confidence intervals were calculated using the submodel *Calculate Accuracy Indices and CI*, the Q-GIS plugin, or the R package (Table 6). Overall accuracy resulted in 0.915 ± 0.052 , which represents a CI of (0.863–0.967). It is worth noting that if we compute the accuracy indices directly from the raw matrix, the estimates are different. For instance, the value of overall

TABLE 2: Main characteristics of the five submodels for the accuracy assessment Dinamica tool.

Submodel name	Function	Input(s)	Output(s)
Generate Confusion Matrix	Constructs the raw and bias-adjusted confusion matrices	Thematic map under assessment, map of reference samples	Raw confusion matrix, bias-adjusted confusion matrix, map categories area, number of samples per map category
Calculate Accuracy Indices And CI	Calculates bias-adjusted accuracy indices along with their CI as well as error-adjusted estimates of the proportion area of the different categories with their CIs	Tables: raw confusion matrix, bias-adjusted confusion matrix, area per mapped category, number of samples per category	Overall, user, and producer accuracies with their respective half confidence intervals. Error adjusted proportion of categories estimated with its CI
Adjust Matrix	Applies bias adjustment to the input raw matrix	Tables: map category's area, raw confusion matrix, number of samples per map category	Bias-adjusted confusion matrix
Bootstrap	Elaborates a replicate of the input raw matrix by sorting samples with repetition	Raw confusion matrix	Bootstrapped replicate of the input matrix
Bootstrap CI	Calculates confidence intervals (CI) around an index at a given significance level (by default 5%) with bootstrapped estimates of this index	Table of the index values derived from a large number of replicates	Value of the average and half confidence interval

FIGURE 1: Land use/cover map obtained through the classification of SPOT5 imagery by visual interpretation at 1:40,000 (Ayuquila basin, central-west Mexico).

accuracy is 0.89 and the values of producer accuracy are 0.83 and 0.80 for categories 5 and 15, respectively. These values are not valid because they are biased by the sampling design. However, it is worth noting that this approach is often an overoptimistic estimate of accuracy, since when there are no off diagonals in a given column or row, the category accuracy will be 1, and the CI will be zero, suggesting that there is no uncertainty about this estimate. In these cases, the half CI has to be considered with caution and may be better represented as not available. Other approaches such as Bayesian analysis could be considered allowing users to combine prior information in the error matrix analysis and improve the precision of accuracy indices [20]. In many cases, due to the low number of sampling units per category, the estimates of accuracy present high uncertainty. For instance, the CI of user accuracy for categories 10 and 13 is (0.17–1.00). In these cases, sample size should be augmented to improve precision of the estimate.

Table 7 shows the area for each category derived directly from the map (pixel counting) along with the estimate of

the area adjusted for the error and its confidence interval. Despite the high accuracy of the map, in some cases, the error-adjusted area is rather different from the area directly obtained from the map. For instance, the area for category 2, which presents commission errors, was overestimated by the map whereas the area for category 12 was underestimated by the map due to omission errors.

4. Discussion

According to Stehman [9], accuracy indices directly interpretable as probabilities of encountering certain types of misclassification errors should be preferred to measures not interpretable as such. Overall accuracy is the probability for a randomly selected location in the map to be correctly classified. User accuracy for category i is the conditional probability that an area classified as category i in the map is classified as category i in the reference data. Producer accuracy for category j is the conditional probability that an area classified as category j in the reference data is classified

TABLE 3: Raw confusion matrix.

Map	Reference																	Sum
1	7																	7
2		7							1		3							11
3			5															5
4				6														6
5					5													5
6						4					1							5
7							5											5
8								5										5
9									5									5
10										3		1		1				5
11											4			1				5
12												13						13
13					1							1	3					5
14														13				13
15														1	4			5
16															1	4		5
17																	5	5

Note: (1) Irrigated agriculture (40,448 ha); (2) rainfed agriculture (62,962 ha); (3) urban (3,787 ha); (4) oak forest (37,599 ha); (5) fir forest (7,453 ha); (6) pine forest (275 ha); (7) pine-oak forest (47,647 ha); (8) montane cloud forest (12,783 ha); (9) water bodies (1,506 ha); (10) subtropical shrubland (26,916 ha); (11) cultivated grassland (3,323 ha); (12) induced grassland (66,727 ha); (13) high-mountain prairies (722 ha); (14) tropical dry forest (98,306 ha); (15) tropical semideciduous forest (336 ha); (16) riparian forest (607 ha); (17) bare land (48 ha).

TABLE 4: Sampling intensity among map categories.

Map category	1	2	3	4	5	6	7	8	9
Number of samples	7	11	5	6	5	5	5	5	5
Proportion of sample (%)	6.36	10.00	4.55	5.45	4.55	4.55	4.55	4.55	4.55
Proportion of area (%)	9.83	15.30	0.92	9.14	1.81	0.07	11.58	3.11	0.37
Ratio proportion of sample/proportion of area	0.65	0.65	4.94	0.60	2.51	68.01	0.39	1.46	12.42
Map category	10	11	12	13	14	15	16	17	sum
Number of samples	5	5	13	5	13	5	5	5	110
Proportion of sample (%)	4.55	4.55	11.82	4.55	11.82	4.55	4.55	4.55	100
Proportion of area (%)	6.54	0.81	16.22	0.18	23.89	0.08	0.15	0.01	100
Ratio proportion of sample/proportion of area	0.69	5.63	0.73	25.90	0.49	55.66	30.81	389.63	

as category j in the map. After applying the bias-adjustment proposed by Card [8], our tool provides accuracy indices that possess such probabilistic interpretation. The tools do not calculate the Kappa index because it does not fulfill this requirement due to the adjustment for hypothetical chance agreement [9]. Moreover, this index has been strongly criticized [21].

However, to avoid biased results it is important to avoid nonprobability sampling by convenient procedures including selecting training data used during supervised classification, limiting the random sampling of reference sites to accessible or homogeneous areas. These procedures will conduce to nonrepresentative samples and generally to optimistically biased estimates of accuracy. When preparing the sampling design, it is also crucial to clearly identify the sampling unit

as well as the evaluation and labeling protocol used to assign a category to the sample unit [2].

Finally, accuracy assessments should be applied to transitions (or change categories) in land use/cover change analyses. In particular, a confidence interval should be provided in order to quantify the uncertainty of the land use/cover change area estimates [7]. This is particularly relevant when reporting critical transitions such as deforestation processes. The set of reference plots can be selected using popular sampling such as stratified random, simple random, and systematic designs. In studies aimed at estimating the area of land use/cover change, the estimation of accuracy is generally based on a stratified random sampling because the categories of interest (the change areas) present a much smaller area than the areas of permanence. Stratification should be based on transitions'

TABLE 5: Adjusted confusion matrix (values were multiplied by a factor of 100 for clarity).

Map	Reference																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	9.83																
2		9.74								1.39		4.17					
3			0.92														
4				9.14													
5					1.81												
6						0.05						0.01					
7							11.58										
8								3.11									
9									0.37								
10										3.93		1.31		1.31			
11											0.65			0.16			
12												16.22					
13					0.04							0.04	0.11				
14														23.89			
15														0.02	0.07		
16															0.03	0.12	
17																	0.01

TABLE 6: Accuracy indices along with their CI obtained by (3) to (7). Bootstrapping with 1000 replicates gave very similar results.

Category	1	2	3	4	5	6	7	8	9
User accuracy	1.00	0.64	1.00	1.00	1.00	0.80	1.00	1.00	1.00
Half CI	0.00	0.28	0.00	0.00	0.00	0.35	0.00	0.00	0.00
CI lower bound	1.00	0.35	1.00	1.00	1.00	0.45	1.00	1.00	1.00
CI upper bound	1.00	0.92	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Prod accuracy	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	1.00
Half CI	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00
CI lower bound	1.00	1.00	1.00	1.00	0.95	1.00	1.00	1.00	1.00
CI upper bound	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Category	10	11	12	13	14	15	16	17	
User accuracy	0.60	0.80	1.00	0.60	1.00	0.80	0.80	1.00	
Half CI	0.43	0.35	0.00	0.43	0.00	0.35	0.35	0.00	
CI lower bound	0.17	0.45	1.00	0.17	1.00	0.45	0.45	1.00	
CI upper bound	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Prod accuracy	0.74	1.00	0.75	1.00	0.94	0.69	1.00	1.00	
Half CI	0.39	0.00	0.16	0.00	0.09	0.39	0.00	0.00	
CI lower bound	0.35	1.00	0.59	1.00	0.86	0.30	1.00	1.00	
CI upper bound	1.00	1.00	0.90	1.00	1.00	1.00	1.00	1.00	

categories instead of land use/cover categories. Finally, they have to be labeled accordingly as land use/cover transitions. The same procedure described above is then applied to assess accuracy and improve area estimates. For example, Olofsson et al. [7] assessed the accuracy of a deforestation map and found that the error-adjusted area estimate of deforestation was about two times larger than the mapped area due to a large error of omission of deforested areas. The uncertainty of the change area estimate, expressed through the CI, can be used to assess the uncertainty of estimates based on

land change area as input, such as carbon release due to deforestation.

5. Conclusions

The confusion matrix provides users with information on the magnitude and patterns of the classification errors. However, in order to calculate accuracy indices and their associated uncertainty, the type of sampling design used to select the verification sites should be taken into account. The confusion

TABLE 7: Estimates of category's area (ha) directly derived from the map (by simple pixel count) and by using error adjustment along with their confidence interval. As for Table 5, half CI of zero has to be considered with caution.

Category	Map	Adjusted	Half CI	CI lower bound	CI upper bound
1	40448	40448	0	40448	40448
2	62962	40067	18772	21294	58839
3	3787	3787	0	3787	3787
4	37599	37599	0	37599	37599
5	7453	7597	283	7314	7880
6	275	220	108	112	328
7	47647	47647	0	47647	47647
8	12783	12783	0	12783	12783
9	1506	1506	0	1506	1506
10	26916	21873	17113	4761	38986
11	3323	2658	1303	1356	3961
12	66727	89481	20334	69147	109815
13	722	433	347	87	780
14	98306	104421	10632	93789	115053
15	336	390	272	118	662
16	607	486	238	248	724
17	48	48	0	48	48
Total	411445	411445			

matrix enables also users to carry out an adjustment of the area estimator and avoids the possible measurement bias associated with the area obtained directly from the map (e.g., pixel counting). Unfortunately, accuracy assessments often fail in correctly computing accuracy indices and providing the required information to use data. The tools presented in this paper enable users to carry out accuracy assessments of thematic categorical maps. As shown in the case study, the tools enable users to compute the overall, user, and producer accuracy estimates along with two types of confidence intervals and provide an error-adjusted area estimator. When reporting accuracy, it is recommended to report both user and producer accuracies as well as the full error matrix and sampling design [9]. We believe that these tools will provide a wide range of users worldwide with user-friendly programs to carry out statistically rigorous accuracy assessments and complete reports, without ample expertise in GIS and statistics. Given that Dinamica EGO, Q-GIS, and R enables users to build their own tools, it is possible to improve and modify these tools or complement it with new elements. Thus we hope this study will provide users with tools useful to design and implement thematic map accuracy assessment following “good practice” recommendations and that these tools will evolve in order to follow the improvements in technology and the development of new methods in geographical sciences.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

Financial support for this research was made possible by SEP-CONACyT (project no. 178816). They thank the Dinamica EGO team for providing assistance and helpful advice during the development of the tool. Comments and constructive suggestions of reviewers greatly contributed to the improvement of this paper.

References

- [1] R. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, CRC/Taylor & Francis, Boca Raton, Fla, USA, 2nd edition, 2009.
- [2] S. V. Stehman and R. L. Czaplewski, “Design and analysis for thematic map accuracy assessment: fundamental principles,” *Remote Sensing of Environment*, vol. 64, no. 3, pp. 331–344, 1998.
- [3] S. V. Stehman, “Basic probability sampling designs for thematic map accuracy assessment,” *International Journal of Remote Sensing*, vol. 20, no. 12, pp. 2423–2441, 1999.
- [4] S. V. Stehman, “Thematic map accuracy assessment from the perspective of finite population sampling,” *International Journal of Remote Sensing*, vol. 16, no. 3, pp. 589–593, 1995.
- [5] J. D. Wickham, S. V. Stehman, J. H. Smith, T. G. Wade, and L. Yang, “A priori evaluation of two-stage cluster sampling for accuracy assessment of large-area land-cover maps,” *International Journal of Remote Sensing*, vol. 25, no. 6, pp. 1235–1252, 2004.
- [6] S. V. Stehman, “Statistical rigor and practical utility in thematic map accuracy assessment,” *Photogrammetric Engineering & Remote Sensing*, vol. 67, no. 6, pp. 727–734, 2001.
- [7] P. Olofsson, G. M. Foody, S. V. Stehman, and C. E. Woodcock, “Making better use of accuracy data in land change studies:

- estimating accuracy and area and quantifying uncertainty using stratified estimation,” *Remote Sensing of Environment*, vol. 129, pp. 122–131, 2013.
- [8] D. H. Card, “Using known map category marginal frequencies to improve estimates of thematic map accuracy,” *Photogrammetric Engineering & Remote Sensing*, vol. 48, no. 3, pp. 431–439, 1982.
 - [9] S. V. Stehman, “Comparing estimators of gross change derived from complete coverage mapping versus statistical sampling of remotely sensed data,” *Remote Sensing of Environment*, vol. 96, no. 3–4, pp. 466–474, 2005.
 - [10] B. Soares-Filho, H. Rodrigues, and M. Follador, “A hybrid analytical-heuristic method for calibrating land-use change models,” *Environmental Modelling and Software*, vol. 43, pp. 80–87, 2013.
 - [11] B. S. Soares-Filho, H. Rodrigues, and W. L. S. Costa, *Modeling Environmental Dynamics with Dinamica EGO*, 2009.
 - [12] R. M. Almeida and E. E. N. MacAu, “Stochastic cellular automata model for wildland fire spread dynamics,” *Journal of Physics: Conference Series*, vol. 285, no. 1, Article ID 012038, 2011.
 - [13] G. Cuevas and J. F. Mas, “Land use scenarios: a communication tool with local communities,” in *Modelling Environmental Dynamics*, M. Paegelow and M. T. Camacho, Eds., pp. 223–246, Springer, Berlin, Germany, 2008.
 - [14] R. Giudice, B. S. Soares-Filho, F. Merry, H. O. Rodrigues, and M. Bowman, “Timber concessions in Madre de Dios: are they a good deal?” *Ecological Economics*, vol. 77, pp. 158–165, 2012.
 - [15] F. Nunes, B. Soares-Filho, R. Giudice et al., “Economic benefits of forest conservation: assessing the potential rents from Brazil nut concessions in Madre de Dios, Peru, to channel REDD+ investments,” *Environmental Conservation*, vol. 39, no. 2, pp. 132–143, 2012.
 - [16] J. F. Mas, B. Soares Filho, R. G. Pontius Jr., M. Farfan Gutierrez, and H. Rodrigues, “A suite of tools for ROC analysis of spatial models,” *ISPRS International Journal of Geo-Information*, vol. 2, pp. 869–887, 2013.
 - [17] B. Efron and R. Tibshirani, *The Bootstrap Method for Assessing Statistical Accuracy*, Stanford University, Stanford, Calif, USA, 1985.
 - [18] K. T. Weber and J. Langille, “Improving classification accuracy assessments with statistical bootstrap resampling techniques,” *GIScience & Remote Sensing*, vol. 44, no. 3, pp. 237–250, 2007.
 - [19] S. V. Stehman, “Estimating area from an accuracy assessment error matrix,” *Remote Sensing of Environment*, vol. 132, pp. 202–211, 2013.
 - [20] R. Denham, K. Mengersen, and C. Witte, “Bayesian analysis of thematic map accuracy data,” *Remote Sensing of Environment*, vol. 113, no. 2, pp. 371–379, 2009.
 - [21] R. G. Pontius Jr. and M. Millones, “Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment,” *International Journal of Remote Sensing*, vol. 32, no. 15, pp. 4407–4429, 2011.