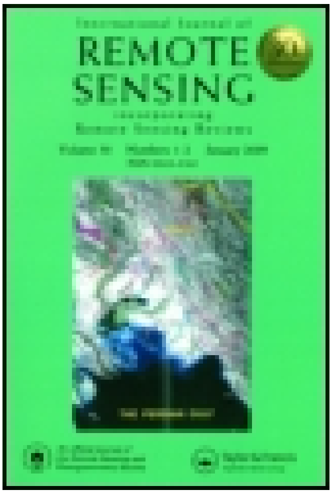
Accuracy assessment of object-based image classification: another STEP

Article *in* International Journal of Remote Sensing · August 2014

DOI: 10.1080/01431161.2014.943328

1 author:

Ivan Lizarazo
Universidad Distrital Francisco José de Caldas



International Journal of Remote Sensing
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tres20

Accuracy assessment of object-based image classification: another STEP

Ivan Lizarazo[a]
[a] Cadastral Engineering and Geodesy Department, Universidad Distrital Francisco José de Caldas, Bogota, Colombia
Published online: 22 Aug 2014.

To cite this article: Ivan Lizarazo (2014): Accuracy assessment of object-based image classification: another STEP, International Journal of Remote Sensing, DOI: 10.1080/01431161.2014.943328

To link to this article: http://dx.doi.org/10.1080/01431161.2014.943328

International Journal of Remote Sensing, 2014
http://dx.doi.org/10.1080/01431161.2014.943328

Taylor & Francis
Taylor & Francis Group

Accuracy assessment of object-based image classification: another STEP

Ivan Lizarazo*

Cadastral Engineering and Geodesy Department, Universidad Distrital Francisco José de Caldas, Bogota, Colombia

(*Received 23 March 2014; accepted 21 May 2014*)

This article proposes STEP, a novel object-based similarity matrix, for assessing both geometric and thematic accuracies of remote-sensing image classification. In contrast to the traditional error matrix, STEP uses samples of classified and reference objects rather than counts of pixels. Moreover, STEP provides four (4) similarity metrics for characterization of classified objects compared with reference objects: (i) shape similarity (*S*); (ii) theme similarity (*T*); (iii) edge similarity (*E*); and (iv) position similarity (*P*). Individual objects' similarity metrics are grouped by thematic class and expressed in the integrated STEP similarity matrix. The proposed approach is illustrated using both a hypothetical classification and a real urban land-cover classification obtained from high spatial resolution orthoimagery. Results show that the STEP indices and matrices are able to express meaningful information about thematic and geometric accuracies of object-based image classifications. It also yields area weighted aggregated-by-class error matrices that allow for calculating overall accuracy metrics.

1. Introduction

The potential for environmental applications of remote sensing is rapidly increasing as the techniques for both pixel and object-based data analysis are improving and becoming widely available. While this rise in interest is noteworthy, several issues may hinder the full potential use of remote-sensing products and complicate their integration with existing geographic datasets. Accuracy assessments determine the quality of the information derived from remotely sensed data (Congalton and Green 2008). However, accuracy assessments are a complex subject and numerous problems remain to be solved including: (i) the impossibility of specifying a single, all-purpose measure of classification accuracy; (ii) the definition of an appropriate sample size and sampling design as well as the specification and use of a measure of accuracy appropriate to the application in-hand; and (iii) the variety of thematic and non-thematic errors to measure and report (Foody 2002).

For object-based image analysis (OBIA), it is important to establish how well digital objects extracted from high-resolution images match existing geographic objects (Blaschke 2010; Lizarazo 2013). It has been suggested that objects, instead of pixels, should be used as sampling units for object-based classifications (Congalton and Green 2008). In several applications, object-based accuracy assessment should account for the thematic accuracy of the class labels as well as the spatial characteristics of represented objects (Stow et al. 2008). Furthermore, a number of authors have proposed several object-based metrics for assessing both geometric and thematic accuracies of object-based image classification (see, e.g.

*Email: ilizarazo@udistrital.edu.co

Persello and Bruzzone (2010), Hernando et al. (2012), Möller et al. (2013)). Although those metrics seem grounded, they have not yet been widely adopted by the OBIA community.

Radoux et al. (2011) reviewed the methods used in evaluating object-based land-cover classifications in a survey of 20 works reported in recent peer-reviewed papers. Radoux and co-workers found that only 40% of the papers used objects as sampling units. Such a trend is still relevant to more recent work. Note, for instance, that thematic accuracy of the recently released object-based Global Land Cover SHARE database (FAO 2014) was assessed using pixels as spatial units. Furthermore, non-thematic errors were not included in the accuracy assessment. It seems that much more work is needed to promote adoption of thematic and geometrics object-based accuracy assessment approaches.

This article proposes STEP, a novel object-centric approach, for assessing geometric and thematic accuracies of remote-sensing image classification. In contrast to the traditional error matrix, STEP uses samples of classified and reference objects rather than counts of pixels. Moreover, STEP provides similarity metrics for estimating correctness of classified objects in terms of shape, theme, edge, and position. The proposed approach is illustrated using an urban land-cover classification obtained from high spatial resolution orthoimagery.

2. Overview of the STEP approach

2.1. General procedure

Unlike recent OBIA-based accuracy assessment approaches, e.g. Persello and Bruzzone (2010), the STEP approach focuses on similarity metrics rather than on dissimilarity metrics. Similar to other approaches (Möller et al. 2013), it allows for comparison of one or many classified objects to reference objects. Furthermore, it attempts an integrated assessment of thematic and geometric accuracy. Moreover, to promote remote sensing and geographic information systems (GIS) integration, the STEP metrics build on standard GIS terms and definitions adopted by Open Geospatial Consortium (OGC 2011).

Similar to well-established methods that are commonly used in remote sensing, two samples are required in STEP for accuracy assessment: a set of reference objects that are considered a very accurate representation of ground conditions (i.e. existing geographic objects) and a set of classified objects generated from remote sensing. Geometric and thematic properties of reference objects represent existing ground conditions that can be determined either by field survey or by visual interpretation of aerial orthoimages (Congalton and Green 2008). Although there are concerns regarding the assumption that the reference data used in the assessment of classification accuracy are themselves an accurate representation of reality (Foody 2002), it is necessary that such an assumption holds for any accuracy assessment results to be meaningful.

This article proposes a four-step procedure for assessing the accuracy of object-based thematic classifications:

(1) selection of a random sample of reference objects;
(2) determination of the corresponding classified objects, i.e. selection of classified objects that intersect the sampled reference objects;
(3) matching each sample reference object to one or several classified objects; and
(4) calculation of thematic and geometric similarity metrics between each reference object and each corresponding classified object.

It should be noted that sampling objects introduce several issues that are not present for sampling pixels. First, a decision has to be made about what is the population to draw samples from (e.g. either the reference objects or the classified objects). Following Congalton and Green (2008), the spatial units used for accuracy assessment in this article were defined using reference polygons as they represent actual Earth surface features (in contrast to classified polygons that can be modified as the map is revised prior to completion of the final map product). Furthermore, using reference polygons instead of classified polygons allows for comparison of accuracies of image classifications obtained using different analysis techniques, which is often the case when conducting remote-sensing projects (Foody 2002).

To select the sample of reference objects, a number of points were randomly located in the region of interest (ROI) and then objects containing these random points were drawn. Clearly, this sampling approach preferentially selects objects with larger size (area). This is a valid sampling option because the probabilities of sampling different objects are known, but the subsequent analysis needs to take into account that objects are included in the sample with probability proportional to area of the object. This is done using a weighting technique that is explained further in Section 2.7.

After selecting the sample of reference objects, a correspondence between sampled reference objects and classified objects was established as follows. First, the sample of reference objects was overlaid with the complete set of classified objects. Then, every classified object intersecting a sample reference object was selected. At this point, there is a set of sample reference objects and their corresponding classified objects. It is apparent that one reference object may match one or several classified objects. Furthermore, a classified object may match one or several reference objects. When the thematic category of a classified object is the same as a given reference object, such a classified object is considered to be a *correctly classified object* (CCO). Otherwise, the classified object is referred to as a *misclassified object* (MCO). It can be stated that a CCO object has a thematic similarity to a given reference object equal to the percentage of overlapping areas between such object and the reference object under evaluation. An MCO object, in turn, has a thematic similarity with the reference object equal to 0%.

Object-based image classification accuracy can be estimated after the sample is complete in both thematic and spatial (geometric) realms. Reliability of thematic attributes carried by classified objects was assessed in this article using a thematic similarity index. Reliability of objects' geometric properties was conducted using the following three indices: (i) shape similarity; (ii) edge similarity; and (iii) position similarity. These similarity indices are described in the next sections.

2.2. Shape similarity

Shape describes the geometric form of individual spatial objects (MacEachren 1985). Shape is often characterized through a compactness indicator, which describes the form of a given region based on how far it deviates from a specified norm (e.g. circle, square, or triangle) (Wentz 1997). A simple but meaningful shape indicator is the normalized perimeter index (NPI) (Angel, Parent, and Civco 2010). NPI is based on the equal area circle (*eac*), which is a circle with an area equal to the area of the object. The NPI was calculated using the following equation:

$$\mathrm{NPI} = p_{\mathrm{eac}}/p_{\mathrm{obj}}, \tag{1}$$

where p_{eac} is the eac's perimeter, and p_{obj} is the object's perimeter. A regular shape (i.e. a circular object) will have an NPI of 1.0, whereas less compact regions will typically have values less than 1.0 – the smaller the value, the further the shape is from being circular (Angel, Parent, and Civco 2010).

This article proposes a normalized shape similarity index to compare the geometric form of a classified object with that of the corresponding reference object(s) defined for accuracy assessment. The shape similarity (S) was computed using Equation (2):

$$S = r_{npi^k}, \tag{2}$$

where r_{npi} is the ratio of the classified object's NPI to the reference object's NPI, and k is given the value +1 when r_{npi} is less than or equal to 1.0, and the value –1 otherwise. S values range in the [0, 1] interval. A classified object that matches the reference object's form has a value of 1.0.

2.3. Theme similarity

Thematic properties of spatial objects refer to non-metric attributes such as land-cover or land-use categories. A theme similarity index describes how well classified objects represent categories assigned to reference objects. In this article, theme similarity (T) was evaluated by calculating the percentage of the reference object's area which intersects the classified object's area using Equation (3):

$$T = A_{\mathrm{int}}/A_{\mathrm{ref}}, \tag{3}$$

where A_{int} is the area of the geometric object representing the point set intersection (OGC 2011) between the classified object and the reference object, and A_{ref} is the area of the reference object's geometry. T returns the percentage of the reference object's area that overlaps the corresponding classified object(s). T values range in the (0, 1] interval. A classified object that completely covers the reference object and matches its thematic category has a value of 1.0. Any classified object that intersects the reference object but does not match its thematic category has a value of 0.0.

2.4. Edge similarity

The edge or boundary is defined as the set of line segments that represents the limit of an entity (OGC 2011). An edge similarity index considers both exterior and interior boundaries of corresponding objects. In this article, edge similarity (E) was evaluated by calculating the percentage of the reference object's boundary coincident with the classified object's boundary using Equation (4):

$$E = (l_{\mathrm{int}}/p_{\mathrm{ref}})^k, \tag{4}$$

where l_{int} is length of the boundary of the geometric object representing the point set intersection between the boundary of the classified object and the boundary of the reference object, p_{ref} is the perimeter of the reference object, and k is given the value +1 when l_{int} is less than or equal to p_{ref}, and the value −1 otherwise. Note that boundaries of reference objects have an inherent 'width' due to the uncertainty of the data sources and processing techniques (Skidmore and Turner 1992). The epsilon band is defined as a zone of uncertainty around an encoded line within which there is a certain probability of observing the 'actual' line. The epsilon distance can be considered as the maximum error to be tolerated. The epsilon distance can be defined based on the spatial accuracy of the data sources, i.e. equal to the reported ground distance (horizontal error) at a given confidence level (e.g. 95%). *E* returns the percentage of the reference object's boundary that overlaps the boundary of the corresponding classified object(s). *E* values range in the [0, 1] interval. A classified object with a coincident boundary has a value of 1.0.

2.5. Position similarity

The direct position of an object is described by a single set of coordinates within a coordinate reference system (OGC 2011). A position similarity index considers the centroid position of classified and reference objects. In this article, position similarity (*P*) was evaluated by first calculating the distance between corresponding centroids and then normalizing by the diameter of a combined area circle (cac), i.e. a circle whose area equals the sum of the reference object's area and the classified object's area, using Equation (5):

$$P = 1 - d_{\text{cent}}/d_{\text{cac}}, \tag{5}$$

where d_{cent} is euclidian distance between the centroid of a reference object and the centroid of the corresponding classified object(s), and d_{cac} is the diameter of the combined area circle. *P* returns a value describing the closeness of the corresponding set. *P* values range in the [0, 1] interval. A classified object with a correctly predicted position has a value of 1.0.

2.6. The STEP similarity matrix

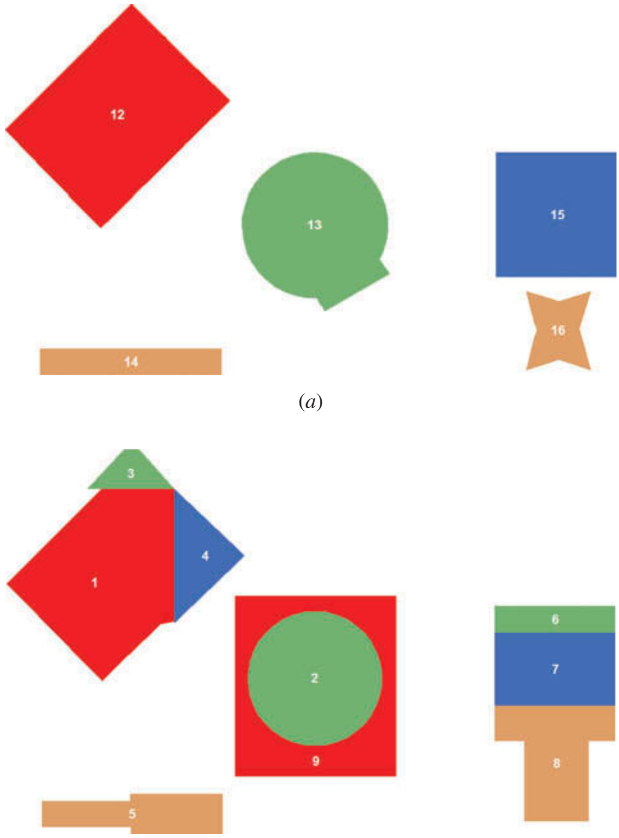
A STEP similarity matrix is a rectangular array set out in rows and columns that expresses the similarity between classified objects and their corresponding reference objects. The columns represent the classified objects whereas the rows indicate the reference objects (Table 1). The numbers in the STEP matrix are the similarity indices calculated by comparing thematic and geometric properties of classified objects with those of reference objects on a cell-by-cell basis. Each cell is split into quarters to store, in Z-order (i.e. a Morton order), the four similarity metrics for each classified object, i.e. *S, T, E*, and *P*. A similarity value in the [0, 1] interval is assigned to each classified object that matches (intersects) a given reference object. In case a classified object does not intersect a given reference object, there is no similarity to account for, and, thus, the corresponding cell value is empty.

In the example STEP matrix shown in Table 1, there are five hypothetical sampled reference objects representing four different land-cover categories. There are also nine corresponding classified objects. As can be seen in the first row, the reference object with ID = 13 is matched by two classified objects identified with ID equal to 2 and 9. Shape similarity (*S*) between the classified object identified as 2 and the reference object identified as 13 is 0.96. Theme similarity (*T*) is 0.82. Edge similarity (*E*) is 0.0. Position similarity (*E*) is 0.99. As can be seen in the fourth and fifth rows, the classified object identified as 8 matches the two reference objects identified as 15 and 16.

Figure 1 shows the hypothetical sampled reference objects along with their matching classified objects. Figure 1(*a*) shows five reference objects belonging to four thematic categories. Figure 1(*b*) shows nine classified objects that intersect the reference objects.

Table 1. Example of STEP similarity matrix for the sampled objects depicted in Figures 1–6. ID denotes each object's identifier. LC denotes each object's thematic category. Numbers in parentheses describe the cardinality between each reference object and its corresponding classified object(s). Values at each cell represent, in Z-order, shape similarity, theme similarity, edge similarity, and position similarity, respectively.

	Classified objects																	
Reference objects	ID = 5 LC = Y		ID = 1 LC = X		ID = 3 LC = W		ID = 4 LC = Z		ID = 2 LC = W		ID = 9 LC = X		ID = 7 LC = Z		ID = 6 LC = W		ID = 8 LC = Y	
ID = 13									0.96	0.82	0.41	0.00						
LC = W (1:2)									0.00	0.99	0.00	0.92						
ID = 12			0.99	0.84	0.87	0.00	0.84	0.00										
LC = X (1:3)			0.52	0.94	0.01	0.47	0.01	0.57										
ID = 14	0.95	1.00																
LC = Y (1:1)	0.56	0.91																
ID = 15													0.97	0.59	0.77	0.00	0.82	0.10
LC = Z (1:3)													0.30	0.99	0.36	0.68	0.10	0.42
ID = 16																	0.98	1.00
LC = Y (1:1)																	0.07	0.71



(*a*)

(*b*)

Figure 1. (*a*) Sampled reference objects for the hypothetical dataset in Table 1. (*b*) Corresponding classified objects. Polygon's label represents the object identifier. Each colour represents a different thematic attribute: green colour means class W, blue colour means class Z, red colour means class X, and khaki colour means class Y. The polygon labels are the numbers shown.

Figure 2(*a*) shows that the reference object identified as 13 holds a thematic attribute value of W. Figure 2(*b*) shows that two classified objects identified as 2 and 9 match the previously mentioned reference object. The classified object identified as 2 is a CCO object as it has a thematic attribute value of W. The classified object identified as 9 is an MCO object with a thematic attribute value of X. Figure 2(*c*) highlights the overlapping areas between the two classified objects and the reference object.

Figures 3–6 show: (a) the reference objects associated with the second, third, fourth, and fifth rows in Table 1; (b) their corresponding classified objects; and (c) the overlapping areas between classified objects and their corresponding reference objects. Note that the reference object with ID = 12 (second row) is matched by three classified objects identified with ID equal to 1, 3, and 4. Object 1 is a CCO object whereas 3 and 4 are MCO objects. The reference object with ID = 14 (third row) is matched by the CCO object with ID = 5. The reference object with ID = 15 (fourth row) is matched by one CCO object (ID equal to 7) and two MCO objects (ID equal to 6 and 8). Finally, the reference object with ID = 16 (fifth row) is matched by one CCO object (ID equal to 8).

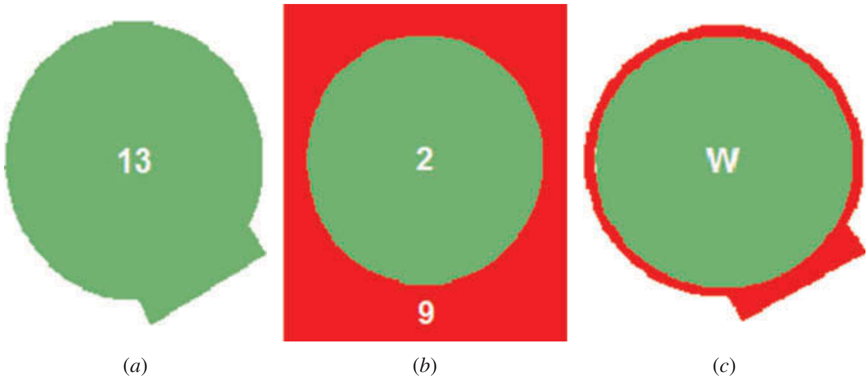


Figure 2. (*a*) Reference object identified as 13. (*b*) Corresponding classified objects identified as 2 and 9. (*c*) Overlapping areas between the given reference object and its corresponding classified objects. Green colour represents thematic category W. Red colour represents thematic category X. These objects correspond to the first row in the example STEP matrix shown in Table 1.

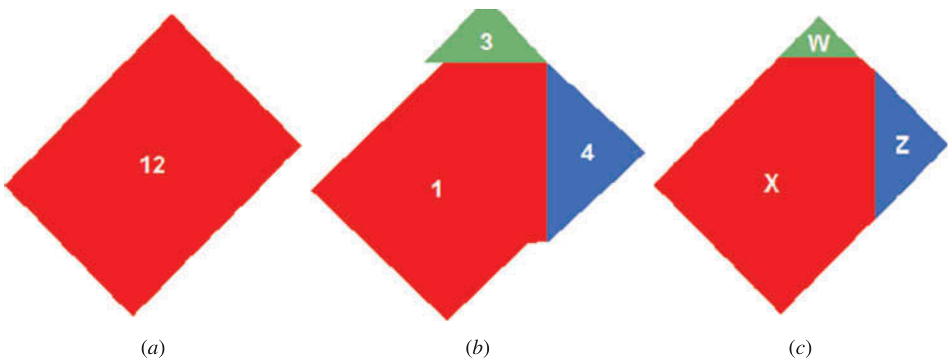


Figure 3. (*a*) Reference object identified as 12. (*b*) Corresponding classified objects identified as 1, 3, and 4. (*c*) Overlapping areas between the given reference object and its corresponding classified objects. Green colour represents thematic category W. Blue colour represents thematic category Z. Red colour represents thematic category X. These objects correspond to the second row in the example STEP matrix shown in Table 1.

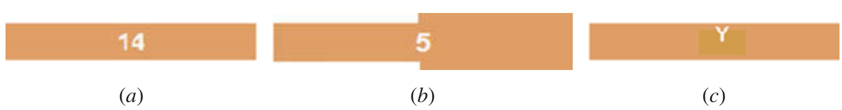


Figure 4. (*a*) Reference object identified as 14. (*b*) Corresponding classified object identified as 5. (*c*) Overlapping areas between the given reference object and its corresponding classified object. Khaki colour represents thematic category Y. These objects correspond to the third row in the example STEP matrix shown in Table 1.

Note also that the classified object with ID = 8 matches two reference objects identified as 15 and 16. In Figures 1–6, each colour represents a different thematic attribute: green colour means class W, blue colour means class Z, red colour means class X, and khaki colour means class Y. The thematic attribute of classified objects is used to label overlapping areas shown in Figures 2(*c*), 3(*c*), 4(*c*), and 5(*c*).

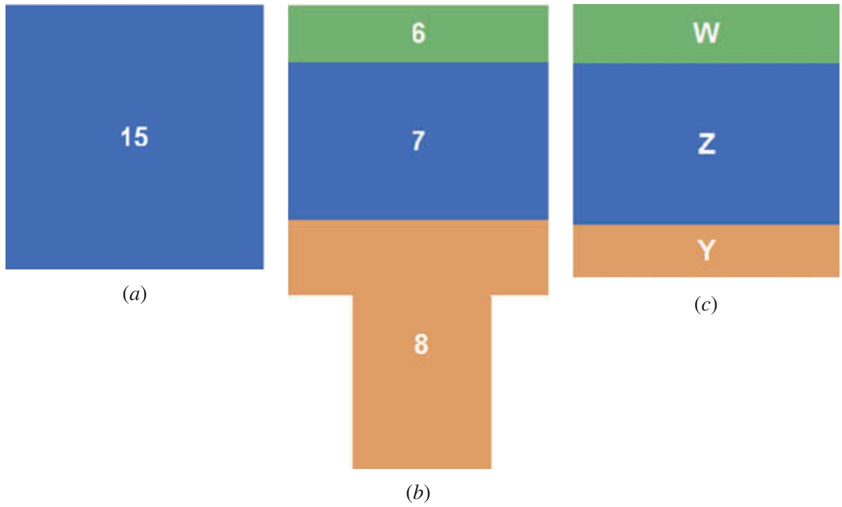


Figure 5. (*a*) Reference object identified as 15. (*b*) Corresponding classified objects identified as 6, 7, and 8. (*c*) Overlapping areas between the given reference object and its corresponding classified objects. Green colour represents thematic category W. Blue colour represents thematic category Z. Khaki colour represents thematic category Y. These objects correspond to the fourth row in the example STEP matrix shown in Table 1.

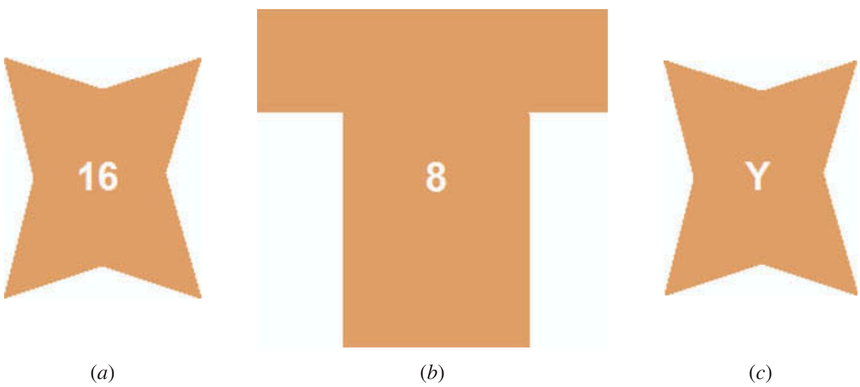


Figure 6. (*a*) Reference object identified as 16. (*b*) Corresponding classified object identified as 8. (*c*) Overlapping areas between the given reference object and its corresponding classified object. Khaki colour represents thematic category Y. These objects correspond to the fifth row in the example STEP matrix shown in Table 1.

2.7. Aggregated similarity metrics

The individual object-based STEP matrix allows for setting up a thematically grouped STEP similarity matrix. This new STEP matrix is a square array that expresses similarity between reference objects (rows) and classified objects (columns) aggregated by thematic category. The number of rows and columns is equal to the number of thematic classes. Each cell value stores a set of four thematically grouped similarity metrics. In order to build this new matrix from the individual object STEP matrix, the following procedure was followed: (i) aggregation of column values (i.e. aggregation by reference object); and (ii) aggregation of row values (i.e. aggregation by thematic class).

The aggregation of column values for shape similarity was conducted using Equation (6):

$$S_j = \sum_{i=1}^{n} (a_i S_i), \tag{6}$$

where S_j is the aggregated S value for reference object j, a_i is the reference area percentage occupied by classified object i, and S_i is the S value for the ith object.

The aggregation of row values was performed by assigning a weight to sample reference objects based on its probability of selection within each category. Let a_j denote the area of reference object j, and A_c the total area sampled in the object's category. Then, the overall probability of selection of a reference object p_j is given by Equation (7):

$$p_j = \frac{a_j}{A_c}. \tag{7}$$

Therefore, the weight of a sampled reference object w_j is given by Equation (8):

$$w_j = \frac{1}{p_j}. \tag{8}$$

Equation (9) was used to calculate the aggregated-by-class shape similarity metric:

$$S_{ag} = \frac{\sum_{j=1}^{n} (w_j S_j)}{\sum_{j=1}^{n} (w_j)}, \tag{9}$$

where S_{ag} is the aggregated S value for a given thematic category, w_j is the weight given to reference object j, and S_j is the aggregated S value for the jth object.

Similar equations to Equations (6) and (9) were used to calculate aggregated T, E, and P similarity values. The class-aggregated STEP similarity matrix corresponding to the reference and classified objects shown in Figures 1–6 is presented in Table 2. In this particular example, there are only two reference objects with the same thematic category, i.e. objects with ID = 14 and ID = 16 are allocated to category Y. Thus, aggregation of row values takes place only for thematic category Y. Note that, for a given row, theme similarity (T) values should add up to 1.0.

Considering that thematic accuracy of categorical maps created using an OBIA approach should be weighted by the area of the sampled reference units, an error matrix that incorporates area rather than counts of polygons into each cell has been recently suggested (MacLean and Congalton 2012). This article proposes an area weighted error matrix, in which the individual cells reflect the weighted area of the reference units that fall into that cell. Using a weighted area instead of the actual area allows for compensation of the unequal probability sampling proposed here. A given cell value, denoted as A_{kl}, represents the total area of the reference objects in reference data class k and map data class l multiplied by the normalized weight of class k. The simple weight of each category w_k was calculated using Equation (10):

Table 2. Example of a class-aggregated STEP similarity matrix for the sampled objects depicted in Figures 1–6. Values in each cell represent, in Z-order, shape similarity, theme similarity, edge similarity, and position similarity, respectively. Cell values in the major diagonal are in bold. The total area of the reference sample is 120,575. The area of the reference objects is given in arbitrary units. Numbers in parenthesis denote proportional area.

		Classified objects							
Reference objects	Reference area	Class W		Class X		Class Y		Class Z	
Class W	31,935 (0.26)	**0.79**	**0.82**	0.07	0.18	0.00	0.00	0.00	0.00
		0.00	**0.81**	0.00	0.16	0.00	0.00	0.00	0.00
Class X	45,447 (0.38)	0.03	0.04	**0.82**	**0.84**	0.00	0.00	0.11	0.13
		0.00	0.02	**0.48**	**0.79**	0.00	0.00	0.00	0.07
Class Y	15,475 (0.13)	0.00	0.00	0.00	0.00	**0.96**	**1.00**	0.00	0.00
		0.00	0.00	0.00	0.00	**0.35**	**0.83**	0.00	0.00
Class Z	27,718 (0.23)	0.31	0.41	0.00	0.00	0.00	0.00	**0.57**	**0.59**
		0.15	0.28	0.00	0.00	0.00	0.00	**0.18**	**0.59**

$$w_k = \frac{A_t}{a_k}, \tag{10}$$

where a_k denotes the total area of reference objects within category k, and A_t the total area sampled. The normalized weight of each category was obtained by dividing the simple weight by the sum of all category weights.

The STEP approach allows for compiling this OBIA thematic error matrix from the area values of the geometric objects representing the point set intersection between the classified objects and the reference objects, i.e. the A_{int} values calculated using Equation (3). The area weighted aggregated-by-class OBIA thematic error matrix for the example in this article is shown in Table 3. The next to last column in this OBIA thematic error matrix is the weighted area of each reference object. The next to last row in this OBIA thematic error matrix is the weighted area of the corresponding classified objects allocated to a given class. Using the above OBIA thematic error matrix, overall accuracy was computed, similar to how overall accuracy is computed in the traditional error matrix (MacLean and Congalton 2012). User's and producer's accuracies were also computed from the OBIA error matrix, and are shown in Table 3.

Three additional aggregated-by-class error matrices were calculated from the STEP matrix: the OBIA shape error matrix, the OBIA edge error matrix, and the OBIA position error matrix. These area weighted error matrices are shown in Tables 4–6. Overall accuracy indices were estimated from the four aggregated-by-class similarity matrices.

Table 7 summarizes overall similarity metrics for the hypothetical classification used in this article. The standard deviation, s, of the overall similarity was estimated by Equation (11) (Rossiter 2014):

$$s = \sqrt{\frac{p(1-p)}{n}}, \tag{11}$$

Table 3. Example of a weighted theme error matrix, which gives credit for unequal probability sampling, for the sampled objects depicted in Figures 1–6. The value in each cell represents the weighted overlapping area, aggregated by classified thematic category, between reference objects and the corresponding classified objects. Cell values in the major diagonal are in bold. Overall theme similarity is 77.9%.

	Classified objects					
Reference objects	Class W	Class X	Class Y	Class Z	Total	Producer's accuracy (%)
Class W	**5332**	816	0	0	6148	86.7
Class X	360	**5423**	0	1338	7121	76.2
Class Y	0	0	**6492**	0	6492	100.0
Class Z	1210	0	2271	**3830**	7312	52.4
Total	6902	6239	8763	5168	**27,072**	
User's accuracy (%)	77.2	86.9	74.0	74.1		**77.9**

Table 4. Example of a weighted shape error matrix, which gives credit for unequal probability sampling, for the sampled objects depicted in Figures 1–6. The value in each cell represents the weighted overlapping area, aggregated by classified thematic category, between reference objects and the corresponding classified objects, multiplied by shape similarity in given Table 2. Cell values in the major diagonal are in bold. Overall shape similarity is 81.7%.

	Classified objects					
Reference objects	Class W	Class X	Class Y	Class Z	Total	Producer's accuracy (%)
Class W	**5119**	334	0	0	5453	93.8
Class X	313	**5369**	0	1124	6806	78.9
Class Y	0	0	**6250**	0	6250	100.0
Class Z	932	0	1862	**3715**	6509	57.1
Total	6364	5704	8112	4839	**25,018**	
User's accuracy (%)	80.4	94.1	77.0	76.8		**81.7**

Table 5. Example of a weighted edge error matrix, which gives credit for unequal probability sampling, for the sampled objects depicted in Figures 1–6. The value in each cell represents the weighted overlapping area, aggregated by classified thematic category, between reference objects and the corresponding classified objects, multiplied by edge similarity in given Table 2. Cell values in the major diagonal are in bold. Overall edge similarity is 90.3%.

	Classified objects					
Reference objects	Class W	Class X	Class Y	Class Z	Total	Producer's accuracy (%)
Class W	**0**	0	0	0	0	0
Class X	5	**5966**	0	19	5989	99.6
Class Y	0	0	**4504**	0	4504	100.0
Class Z	871	0	476	**2336**	3685	63.4
Total	876	5966	4981	2355	**14,179**	
User's accuracy (%)	0	100.0	90.4	99.2		**90.3**

Table 6. Example of a weighted position error matrix, which gives credit for unequal probability sampling, for the sampled objects depicted in Figures 1–6. The value in each cell represents the weighted overlapping area, aggregated by classified thematic category, between reference objects and the corresponding classified objects, multiplied by position similarity in given Table 2. Cell values in the major diagonal are in bold. Overall position similarity is 85.0%.

	Classified objects					
Reference objects	Class W	Class X	Class Y	Class Z	Total	Producer's accuracy (%)
Class W	**5278**	751	0	0	6029	87.6
Class X	169	**5098**	0	763	6030	84.5
Class Y	0	0	**5358**	0	5358	100.0
Class Z	823	0	954	**3792**	5569	68.1
Total	6271	5849	6312	4554	**22,986**	
User's accuracy (%)	84.1	87.1	84.9	83.3		**85.0**

Table 7. Example of overall thematic and geometric accuracy predictors for the hypothetical data in Figures 1–6.

Component	Overall accuracy	95% CI
Shape	0.82	[0.38, 1.00]
Theme	0.78	[0.32, 1.00]
Edge	0.90	[0.54, 1.00]
Position	0.85	[0.44, 1.00]

Note: CI, confidence interval.

where p is overall similarity, and n is the sample size (n = 5 in this example).

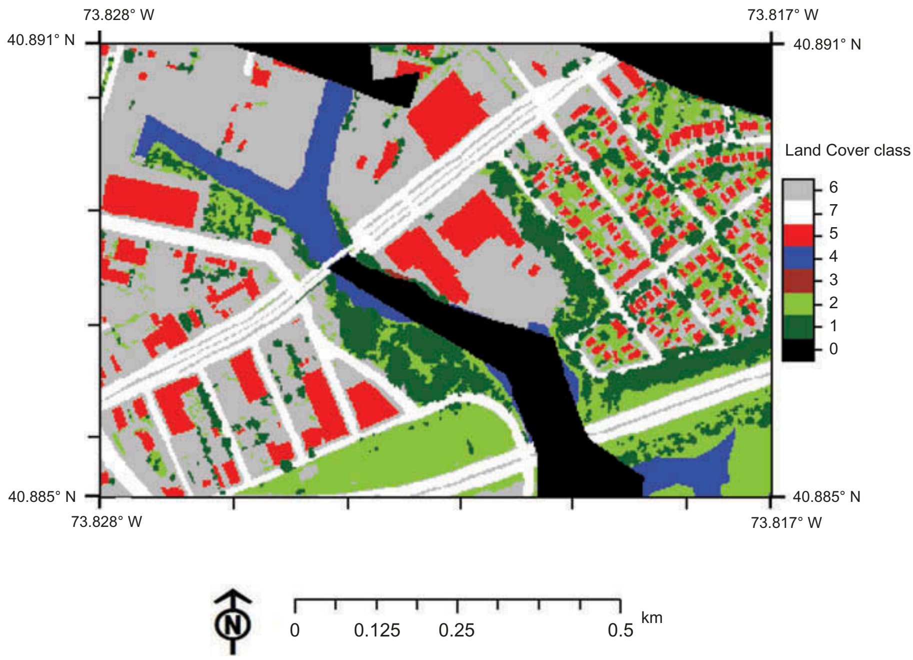
The confidence interval (CI) of similarity estimates was calculated using Equation (12) (Rossiter 2014):

$$\mathrm{CI} = p \pm \left[sZ_{1-\alpha} + \frac{1}{2n} \right], \tag{12}$$

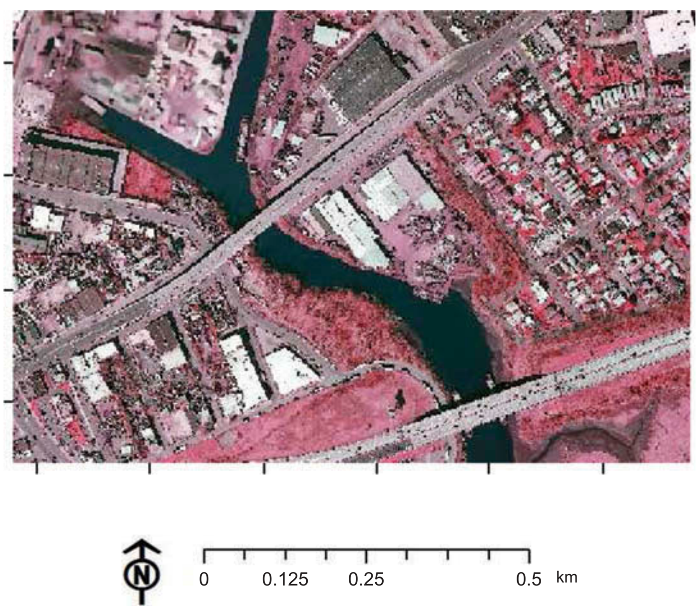
where $Z_{1-\alpha}$ is the two-tailed score for the probability of Type I error α. The expression $1/2n$ is the small-sample correction.

3. A real-world data set for applying the STEP approach

The classified dataset is a high-resolution land-cover dataset for New York City (NYC). The original dataset includes seven land-cover classes: (1) tree canopy, (2) grass/shrub, (3) bare Earth, (4) water, (5) buildings, (6) roads, and (7) other paved surfaces. The minimum mapping unit for the delineation of features was set at 0.28 m^2. The primary sources used to derive this land-cover layer were a 2010 lidar dataset and a 2008 4-band orthoimage. The land-cover dataset was obtained using OBIA techniques to extract land-cover information using the best available remotely sensed and vector GIS datasets. Overall accuracy assessed using the traditional error matrix was reported to be 96% (NYC Open Data 2013). For this article's accuracy assessment, the land-cover subset shown in Figure 7(*a*) was used. It comprises 992 pixels × 670 pixels with 0.90 m pixel size. The classified



(*a*)



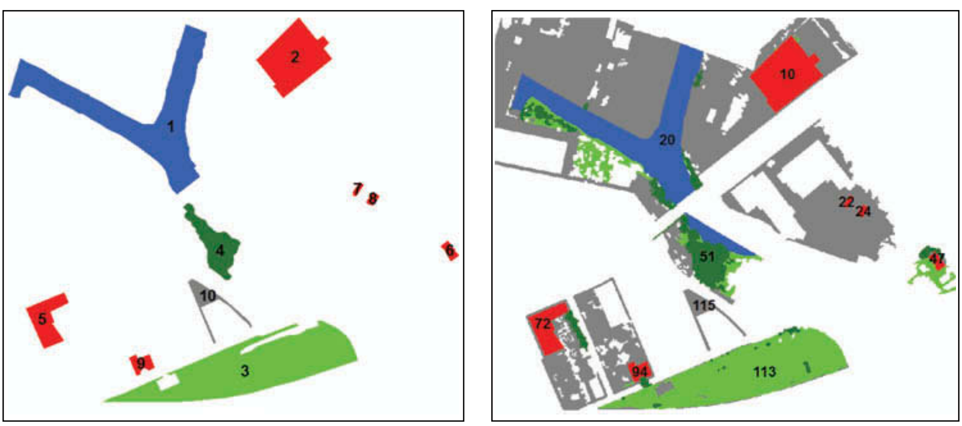
(*b*)

Figure 7. (*a*) 2010 New York City land-cover classification; (*b*) red, green, and blue 432 colour composition of the corresponding 2009 orthoimagery.

dataset's coverage is 0.556 km^2. There are 2272 classified polygons with sizes ranging from 0.84 to 44,440 m^2. Mean size, μ, is 186.54 m^2. Standard deviation, σ, is 1445.4 m^2. The coefficient of variation for the classified objects' area, given by $c_v = \sigma/\mu$, is 7.75.

The classified subset was selected as its extent coincides with the coverage of a publicly available digital orthoimage with a pixel size of 0.3048 m (NYS GIS Clearinghouse 2013). The orthoimage size is 3000 pixels × 2000 pixels, its type is 4-band, red, green, and blue & near-infrared, and its spatial reference is North American Datum 1983 StatePlane New York East FIPS 3101 Feet. The orthoimage's horizontal accuracy is within 1.22 m at the 95% confidence level as established by the US National Standard for Spatial Data Accuracy (NSSDA) (FGDC 1998). This means that users should be 95% confident that the actual position of any feature is within the area specified by an epsilon distance of 1.22 m. As this orthoimage, shown in Figure 7(*b*), is an independent source of high accuracy, it was visually interpreted to obtain reference data for accuracy assessment of the classified land cover.

For this NYC dataset, classification results were evaluated using a sample of reference objects belonging to classes (1) tree canopy, (2) grass/shrub, (4) water, (5) buildings, and (7) other paved surfaces. This sample was obtained by randomly sampling ten point locations within the area covered by the orthoimage and then selecting reference objects containing those locations. For the sake of illustration, the sample being investigated is deliberately quite small. While the selection of a small number of sampled objects considerably increases the uncertainty of the overall accuracy metrics, it is assumed that the sample can be used for accuracy assessment of the classified dataset. Figure 8(*a*) shows the sampled reference objects obtained through visual interpretation of the available orthoimage. Sampled reference objects represent about 10% of area of the land-cover classification subset. Each polygon's label represents its identifier. Figure 8(*b*) shows the sampled classified objects. Similarity metrics were evaluated using vector data structures. Raster classified objects were first converted to vector.



(*a*) (*b*)

Figure 8. (*a*) Sampled reference objects for the NYC land-cover classification subset; (*b*) Corresponding classified objects. Each polygon's label represents the object identifier. Dark green colour represents tree canopy. Light colour represents grass/shrub. Blue colour represents water. Red colour represents buildings. Grey colour represents other paved surfaces.

4. Results and discussion

Similarity metrics were evaluated for the sampled classification objects selected from the New York City's land-cover dataset. Figures 9–13 show the individual STEP metrics for classified objects allocated to land-cover classes: tree canopy, grass/shrub, water, buildings, and other paved surfaces, respectively. Dark green colour represents tree canopy. Light green colour represents grass/shrub. Blue colour represents water. Red colour represents buildings. Grey colour represents other paved surfaces.

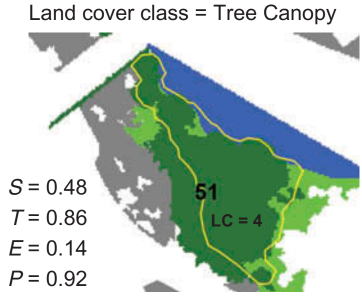


Figure 9. STEP metrics for classified objects allocated to land-cover class tree canopy. The label represents the object's classified ID. The reference boundary is yellow. LC denotes land-cover category.

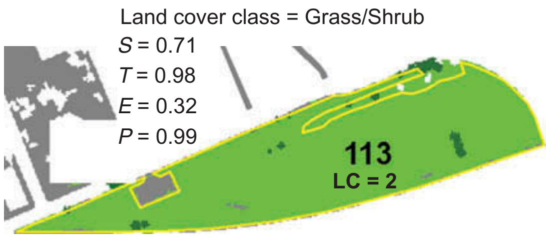


Figure 10. STEP metrics for classified objects allocated to land-cover class grass/shrub. The label represents the object's classified ID. The reference boundary is yellow. LC denotes land-cover category.

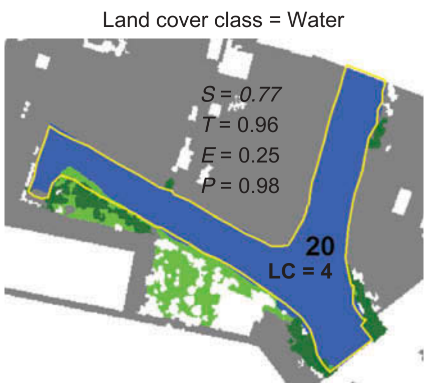


Figure 11. STEP metrics for classified objects allocated to land-cover class water. The label represents the object's classified ID. The reference boundary is yellow. LC denotes land-cover category.

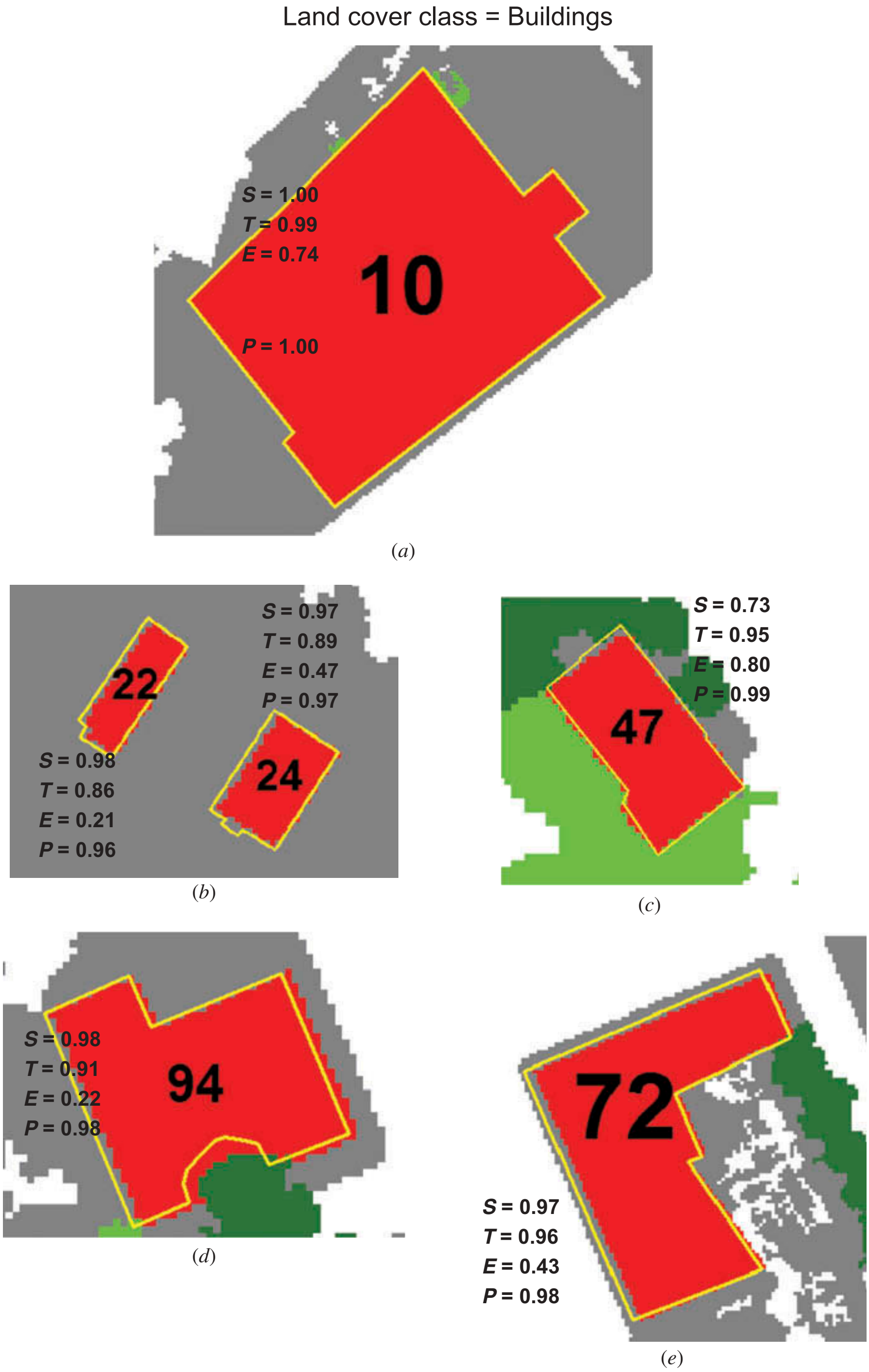


Figure 12. STEP metrics for classified objects allocated to land-cover class buildings. The label represents the object's classified ID. The reference boundary is yellow.

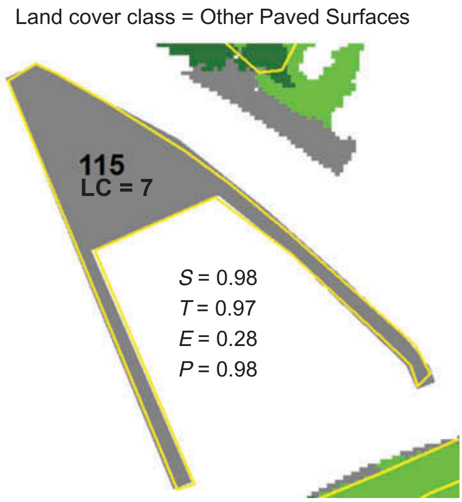


Figure 13. STEP metrics for classified objects allocated to land-cover class other paved surfaces. The label represents the object's classified ID. The reference boundary is yellow.

Table 8 shows the thematically grouped STEP matrix. Table 9 shows the area weighted OBIA theme error matrix. Table 10 shows the area weighted OBIA shape error matrix. Table 11 summarizes overall STEP similarity metrics for the NYC land-cover dataset.

It can be seen in Table 11 that due to the use of a very small sample, the uncertainty of the overall accuracy metrics is too high. The bounds of this CI would not be acceptable for practical applications. It is advisable that a pragmatic application of the proposed approach uses a larger number of samples in order to decrease the variance of accuracy predictors. In the case of the NYC classification subset used in this article, a number of 1000 sampled polygons would be needed to provide a 95% CI equal to 0.10. A sample size like that could be easily beyond most users' reach. This suggests the need of using a much more efficient OBIA accuracy assessment framework such as the recently proposed approach by Radoux and Bogaert (2014).

The STEP approach, as introduced here, aims to establish how well ground objects can be mapped, but not how well properties of classified objects are really what they mean. This lack of symmetry is one limitation of the proposed method. In fact, the reader may have noticed that the individual STEP indices proposed here match the producer's perspective in the traditional error matrix (i.e. the STEP indices are related to errors of omission). While the STEP view seems to overlook the alternative user's perspective, it was shown in this article that the STEP approach allows for the creation of aggregated-by-class area weighted similarity matrices that can also be examined from the point of view of errors of commission. While such an option could be of interest for many applications, it could also further complicate interpretation of accuracy metrics. An additional limitation is that the matrix format of the individual object-based STEP matrix seems not to work for a sample of more than 10–15 objects. However, in such cases, the accuracy assessment procedure could split the sampled objects using tiles (or even sub-tiles) in order to report individual objects' STEP metrics organized by tiles.

Regarding the accuracy assessment indices obtained for the illustrated dataset, results suggest that the most problematic aspect of land-cover classification is the poor

Table 8. Class-aggregated STEP similarity matrix created from the individual object's STEP similarity matrix for the classified objects depicted in Figure 8(*b*). Values in each cell represent, in Z-order, shape similarity, theme similarity, edge similarity, and position similarity, respectively. Cell values in the major diagonal are in bold.

	Classified objects									
Reference objects	Class 4 Water		Class 5 Buildings		Class 2 Grass/Shrub		Class 1 Tree canopy		Class 7 Other paved	
Class 4 Water	**0.74**	**0.96**	0.00	0.00	0.01	0.02	0.01	0.01	0.00	0.02
	0.18	**0.94**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Class 5 Buildings	0.00	0.00	**0.66**	**0.88**	0.00	0.00	0.00	0.00	0.03	0.12
	0.00	0.00	**0.53**	**0.85**	0.00	0.00	0.00	0.00	0.00	0.00
Class 2 Grass/Shrub	0.00	0.00	0.00	0.00	**0.71**	**0.98**	0.01	0.01	0.01	0.01
	0.00	0.00	0.00	0.00	**0.22**	**0.97**	0.00	0.01	0.01	0.00
Class 1 Tree canopy	0.02	0.04	0.00	0.00	0.04	0.09	**0.43**	**0.87**	0.00	0.00
	0.00	0.02	0.00	0.00	0.00	0.03	**0.09**	**0.79**	0.00	0.00
Class 7 Other paved	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	**0.95**	**0.97**
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.27**	**0.95**

Table 9. Weighted thematic error matrix, which gives credit for unequal probability sampling, for the NYC classification data subset. The value in each cell represents the weighted overlapping area, aggregated by classified thematic category, between reference objects and the corresponding classified objects. Cell values in the major diagonal are in bold. Overall theme similarity is 87.6%.

	Classified objects						
Reference objects	Class Water	Class Building	Class Grass	Class Tree	Class Other paved	Total	Producer's accuracy (%)
Water	**8814**	0	156	324	1699	10,993	80.2
Building	0	**8934**	2	16	2088	11,041	80.9
Grass	0	0	**9030**	720	962	10,712	84.3
Tree	67	0	140	**8066**	9	8283	97.4
Other paved	0	0	0	0	**8965**	8965	100
Total	8882	8934	9329	9127	13,722	**49,994**	
User's accuracy (%)	99.2	100.0	96.8	88.4	65.3		**87.6**

Table 10. Weighted shape error matrix, which gives credit for unequal probability sampling, for the NYC classification data subset. The value in each cell represents the weighted overlapping area, aggregated by classified thematic category, between reference objects and the corresponding classified objects, multiplied by shape similarity in Table 8. Cell values in the major diagonal are in bold. Overall shape similarity is 91.6%.

	Classified objects						
Reference objects	Class Water	Class Building	Class Grass	Class Tree	Class Other paved	Total	Producer's accuracy (%)
Water	**6787**	0	68	264	833	7952	85.3
Building	0	**6483**	1	10	472	6967	93.1
Grass	0	0	**6592**	532	686	7810	84.4
Tree	41	0	61	**3946**	4	4052	97.3
Other paved	0	0	0	0	**8785**	8785	100
Total	6828	6484	6722	4752	10,782	**35,567**	
User's accuracy (%)	99.3	100.0	98.1	83.0	81.5		**91.6**

Table 11. Overall thematic and geometric accuracy predictors for the NYC classification data subset.

Component	Overall accuracy	95% CI
Shape	0.92	[0.75, 1.00]
Theme	0.88	[0.67, 1.00]
Edge	0.43	[0.12, 0.73]
Position	0.99	[0.92, 1.00]

Note: CI, confidence interval.

coincidence between classified and reference boundaries. This issue confirms the need for post-processing classified objects before they can be integrated with existing GIS databases (Lizarazo and Elsner 2009). The similarity boundary analysis presented here assumes an uncertain reference boundary commensurate with the positional accuracy of the data source used for visual interpretation. Artificial boundaries such as building class may reach high levels of similarity due to the use of lidar data. Objects with natural boundaries, however, may not have crisp boundaries and are usually affected by boundary blur that results in classification error (Grunblatt 1987; Lizarazo and Elsner 2009).

5. Conclusions

The STEP similarity matrix allows an alternative method for comparison of reference and classified objects with object instances occurring in a variety of scales and locations. STEP is an object-centric approach to assessing classification accuracy for one-to-one, one-to-many, and many-to-one relationships between ground land-cover objects and digitally obtained land-cover objects. Noticeably, the STEP similarity matrix introduces a suite of meaningful metrics for comprehensively evaluating the results of geographic objects' extraction and classification. All in all, the STEP approach takes classification accuracy assessment one small step forward by proposing geometric and thematic similarity indices using standard GIS terms and definitions.

Acknowledgements

The orthoimagery used in this research was obtained from New York Statewide Digital Orthoimagery Program (NYSDOP) at http://www.orthos.dhses.ny.gov/. The high-resolution raster land-cover dataset for New York City was obtained from NYC OpenData at https://nycopendata.socrata.com/. The author acknowledges the thoughtful comments from Dr Timothy Warner and two anonymous reviewers, which helped improve the quality of this article greatly.

References

Angel, S., J. Parent, and D. L. Civco. 2010. "Ten Compactness Properties of Circles: Measuring Shape in Geography." *The Canadian Geographer/Le Géographe canadien* 54: 441–461. doi:10.1111/j.1541-0064.2009.00304.x.

Blaschke, T. 2010. "Object Based Image Analysis for Remote Sensing." *ISPRS Journal of Photogrammetry and Remote Sensing* 65: 2–16. doi:10.1016/j.isprsjprs.2009.06.004.

Congalton, R. C., and K. Green. 2008. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. 2nd ed. Boca Raton, FL: CRC Press.

FGDC (Federal Geographic Data Committee). 1998. *Geospatial Positioning Accuracy Standards, FGDC-STD-007.2-1998*. Reston, VA: FGDC.

FAO (Food and Agriculture Organization of the United Nations). 2014. *Global Land Cover (GLC-SHARE) Beta-Release 1.0 Database, Land and Water Division, John Latham, Renato Cumani, Ilaria Rosati and Mario Bloise*. Rome: FAO.

Foody, G. M. 2002. "Status of Land Cover Classification Accuracy Assessment." *Remote Sensing of Environment* 80: 185–201. doi:10.1016/S0034-4257(01)00295-4.

Grunblatt, J. 1987. "An MTF analysis of Landsat Classification Error at Field Boundaries." *Photogrammetric Engineering and Remote Sensing* 53: 639–643.

Hernando, G. M., D. Tiede, F. Albrecht, and S. Lang. 2012. "Spatial and Thematic Assessment of Object-based Forest Stand Delineation Using an OFA-matrix." *International Journal of Applied Earth Observation and Geoinformation* 19: 214–225. doi:10.1016/j.jag.2012.05.007.

Lizarazo, I. 2013. "Meaningful Image Objects for Object-oriented Image Analysis." *Remote Sensing Letters* 4: 419–426. doi:10.1080/2150704X.2012.744485.

Lizarazo, I., and P. Elsner. 2009. "Fuzzy Segmentation for Object-Based Image Classification." *International Journal of Remote Sensing* 30: 1643–1649. doi:10.1080/01431160802460062.

MacEachren, A. M. 1985. "Compactness of Geographic Shape: Comparison and Evaluation of Measures." *Geografiska Annaler. Series B. Human Geography* 67: 53–67. doi:10.2307/490799.

MacLean, M. G., and R. Congalton. 2012. “Map Accuracy Assessment Issues When Using an Object-Oriented Approach.”American Society of Photogrammetry & Remote Sensing 2012 Annual Conference, Sacramento, CA, March 19–23.
Möller, M., J. Birger, A. Gidudu, and C. Glässer. 2013. “A Framework for the Geometric Accuracy Assessment of Classified Objects.” *International Journal of Remote Sensing* 34: 8685–8698. doi:10.1080/01431161.2013.845319.
NYC Open Data. 2013. *Landcover Raster Data (2010)*. Accessed December 2013. https://data.cityofnewyork.us
NYS GIS Clearinghouse. 2013. *New York State High Resolution Digital Orthoimagery 2000–2012*. Accessed December 2013. http://gis.ny.gov/
Open Geospatial Consortium. 2011. *OGC 06-103r4, Opengis Implementation Standard for Geographic Information – Simple Feature Access – Part 1: Common Architecture*. Wayland: OGC.
Persello, C., and L. Bruzzone. 2010. “A Novel Protocol for Accuracy Assessment in Classification of Very High Resolution Images.” *IEEE Transactions on Geoscience and Remote Sensing* 48: 1232–1244. doi:10.1109/TGRS.2009.2029570.
Radoux, J., and P. Bogaert. 2014. “Accounting for the Area of Polygon Sampling Units for the Prediction of Primary Accuracy Assessment Indices.” *Remote Sensing of Environment* 142: 9–19. doi:10.1016/j.rse.2013.10.030.
Radoux, J., P. Bogaert, D. Fasbender, and P. Defourny. 2011. “Thematic Accuracy Assessment of Geographic Object-based Image Classification.” *International Journal of Geographical Information Science* 25: 895–911. doi:10.1080/13658816.2010.498378.
Rossiter, D. G. 2014. *Technical Note: Statistical Methods for Accuracy Assesment of Classified Thematic Maps*. Department of Earth Systems Analysis University of Twente, Faculty of Geo-Information Science & Earth Observation (ITC). Enschede: ITC.
Skidmore, A. K., and B. J. Turner. 1992. “Map Accuracy Assessment Using Map Line Intersect Sampling.” *Photogrammetric Engineering & Remote Sensing* 58: 1453–1457.
Stow, D., Y. Hamada, L. Coulter, and Z. Anguelova. 2008. “Monitoring Shrubland Habitat Changes through Object-Based Change Identification with Airborne Multispectral Imagery.” *Remote Sensing of Environment* 112: 1051–1061. doi:10.1016/j.rse.2007.07.011.
Wentz, E. 1997. “Shape Analysis in GIS.” *Auto-Carto* 13: 204–213.