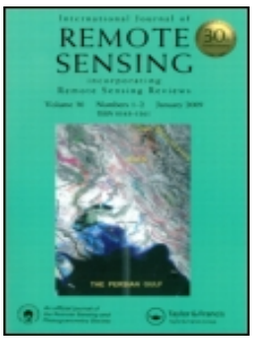

International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: http://www.tandfonline.com/loi/tres20

A framework for the geometric accuracy assessment of classified objects

Markus Möller, Jens Birger, Anthony Gidudu & Cornelia Gläßer

To cite this article: Markus Möller, Jens Birger, Anthony Gidudu & Cornelia Gläßer (2013) A framework for the geometric accuracy assessment of classified objects, International Journal of Remote Sensing, 34:24, 8685-8698, DOI: 10.1080/01431161.2013.845319

To link to this article: http://dx.doi.org/10.1080/01431161.2013.845319

International Journal of Remote Sensing, 2013
Vol. 34, No. 24, 8685–8698, http://dx.doi.org/10.1080/01431161.2013.845319

Taylor & Francis
Taylor & Francis Group

A framework for the geometric accuracy assessment of classified objects

Markus Möller[a]*, Jens Birger[b], Anthony Gidudu[c], and Cornelia Gläßer[a]

[a]*Department of Remote Sensing and Cartography, Institute of Geosciences and Geography, University of Halle-Wittenberg, 06120 Halle (Saale), Germany;* [b]*Environmental and Geodata Management GbR (UMGEODAT), 06108 Halle (Saale), Germany;* [c]*Department of Geomatics and Land Management, Makerere University, Kampala, Uganda*

(*Received 27 March 2013; accepted 2 August 2013*)

European initiatives for data harmonization and the establishment of remote-sensing-based services aim at the production of up-to-date land-cover information according to generally valid standards for the accurate qualification of thematic classification results. This is particularly true since new satellite systems provide data of high temporal and geometric resolution. While methods for point-related thematic accuracy assessment have already been established for years, there is a need for a commonly accepted framework for the geometric quality of tematic maps. In this study, an open and extendable framework for the geometric accuracy assessment is presented. The workflow begins with the definition of basic geometric accuracy metrics, which are based on differences in area and position between samples of classified and reference objects. The combination of user-defined metrics enables both a geometric assessment of single objects as well as the total data set. In an example of thematically classified agricultural fields in a German test site, we finally show how object relations between classified and reference objects can be identified and how they affect the global accuracy assessment of the total data set.

1. Introduction

The European Earth Observation Programme COPERNICUS aims at the establishment of remote-sensing-based services delivering customized and up-to-date geo-information to support the implementation and monitoring of environmental and security policies (Schreier et al. 2008). In this context, new satellite systems have been planned and launched to provide thematic geodata of fine spatio-temporal resolution. This ensures a more robust and long-term monitoring of dynamic land-surface phenomena (Berger et al. 2012).

The quality of thematic geodata can be considered as a desired goal characterized by their suitability for the intended use in decision-making and planning (Versic 2009). Quality represents the totality of features and characteristics of a product or service. To address different users' requirements, the International Organization for Standardization (ISO) has published specific group norms for cartographic products that help data producers objectively describe the quality of data and determine their quality using statistics rather than subjective judgements.

*Corresponding author. Email: markus.moeller@geo.uni-halle.de

Positional and thematic accuracies are among the most important ISO standard quality elements (Jakobsson and Giversen 2008). Thematic accuracy is particularly concerned with the correctness of class assignments. In the context of the accuracy of classified remote-sensing data, many point-related thematic accuracy metrics have been established (Stehman 1997; Foody 2002; Zhan et al. 2005; Liu, Frazier, and Kumar 2007; Olofsson et al. 2013). They are mostly related to the well-known confusion matrix, which 'presents a tabulated view of map accuracy' (Liu, Frazier, and Kumar 2007).

The (absolute) positional accuracy is defined as 'closeness of reported coordinate values to values accepted as or being true' (ISO 19138 2006). Coordinate values are related to, for example, points, lines, or two-dimensional polygons. With the emergence of object-based image analysis (OBIA) techniques in remote sensing in the last decade, the geometric accuracy assessment of two-dimensional objects has become more important and is identified as one of the 'hot' OBIA research topics (Blaschke 2010). In particular, post-classification comparison methods for classified object change detection (COCD) prevail, which were often applied for the updating of already existing maps or geographical information system (GIS) layers (Chen et al. 2012; Hussain et al. 2013).

In connection with COCD methods, several metrics to quantitatively compare the geometric correspondence of classified and reference objects were introduced in recent years. Geometric accuracy is related to the 'problem of matching objects' (Zhan et al. 2005). A complete matching can be assumed if there exists a *one-to-one* correspondence between objects accepted as or being true and objects which should be assessed (Clinton, Holt, and Gong 2010). Most common used metrics' variants are based on overlap relations between classified and reference objects for the quantification of over-segmentation and under-segmentation errors (e.g. Straub and Heipke 2004; Möller, Lymburner, and Volk 2007; Pontius et al. 2007; Clinton, Holt, and Gong 2010; Persello and Bruzzone 2010; Hernando et al. 2012; Sebari and He 2013). Under-segmentation describes the situation when a classified object is bigger than the corresponding reference object. In contrast, over-segmentation leads to a classified object which is too small compared to the reference object (Liu and Xia 2010). Some authors have also quantified differences between the locations of the centres of gravity of overlapping objects (e.g. Möller, Lymburner, and Volk 2007; Clinton, Holt, and Gong 2010; Wang, Jensen, and Im 2010; Sebari and He 2013). Persello and Bruzzone (2010) introduced the metrics *edge location, fragmentation error*, and *shape error.* The first index measures the precision of object edges, and the second one describes the degree of sub-partitioning single objects into different small regions. The *shape error* characterizes object differences regarding shape metrics such as compactness.

The calculation of geometric accuracy metrics requires the existence of independent reference objects. Furthermore, each reference object has to be spatially assigned to at least one corresponding classified object (Persello and Bruzzone 2010). Such assignment is often realized by a GIS-based overlay of reference and classified objects. As a result, different types of hierarchical object relations emerge. In this study, we show how object relations can be identified (Section 2.3) and how they affect the global accuracy assessment of the total data set. Therefore, we use statistical metrics characterizing and comparing distributions (Section 2.4). This investigation is embedded in a framework for the geometric validation of thematic maps, which includes the definition of geometric accuracy metrics based on hierarchical object properties (Section 2.1) and their user-specific combination (Section 2.2) as well as the local and global accuracy assessment (Section 2.4).

The background of this study is the geometric disaggregation of the land-use class *agricultural land* by up-to-date field objects on which the geometric assessment framework is exemplified (Section 3). The geometric accuracy of field objects is important because of their impact on sediment transport and soil erosion modelling (Evans 2013). In addition,

field boundaries are legislative reference units for measures of soil conservation (Volk, Möller, and Wurbs 2010).

2. Methods

2.1. Geometric metrics

According to the requirements of ISO quality elements, the degree of matching of field objects is described by the hierarchical consideration of both the overlap and positions of the objects. In Figure 1(*a*), a reference object (R) overlaps a classified object (F). The intersection of R and F is S and corresponds to the Boolean AND operation (Figure 1(*b*); Equation (1)). The O^{R} metric is derived from the ratio of the metrics *Area of Intersection* (A^{S}) and *Area of the Reference Object* (A^{R}) as shown in Equation (2); on the other hand, O^{F} is derived from the ratio of A^{S} and the metric *Area of the Classified Object* (A^{F}):

$$S = F \cap R, \tag{1}$$

$$O^{X} = \frac{A^{S}}{A^{X}} \text{ with } X \in [R, F]. \tag{2}$$

Positional differences of two-dimensional objects can be expressed by distances between their gravity centres. In Figure 1, the gravity centres c^{R}, c^{F}, and c^{S} of the objects R, F, and S are shown. The distances between the c^{S} and c^{R} (dist(c^{S}, c^{R})) or F (dist(c^{S}, c^{F})) are referred to as the *Relative Position* metrics P^{R} and P^{F} (Equation (3)). As the normalization factor, the maximal distance between c^{S} and $c^{R*,max}$ (dist(c^{S}, $c^{R*,max}$)) or $c^{F*,max}$ is used (dist(c^{S}, $c^{F*,max}$)) whereas the objects R*, F*, $R^{*,max}$ and $F^{*,max}$ result from Boolean NOT

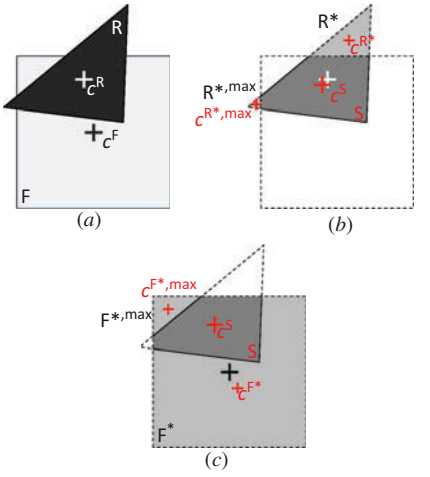

Figure 1. Visualization of object features for the calculation of geometric accuracy metrics: (*a*) overlap of R and F, (*b*) comparison of S and R, and (*c*) comparison of S and F (R, reference object; F, classified object; S, intersection of R and F; c^{F}, gravity centre of F; c^{R}, gravity centre of R; c^{S}, gravity centre of S; R*, relative complement of R and S; F*, relative complement of F and S; $R^{*,max}$, relative complement of R and S whose gravity centre has the farthermost distance from c^{S} within the extent of R; $F^{*,max}$, relative complement of F and S whose gravity centre has the farthermost distance from c^{S} within the extent of F; c^{F*}, gravity centre of F*; c^{R*}, gravity centre of R*; $c^{R*,max}$, farthermost gravity centre from c^{S} within the extent of R; $c^{F*,max}$, farthermost gravity centre from c^{S} within the extent of F).

operations according to Equations (4) and (5). In the majority of cases, more than one relative complement arises as a result of an overlay operation (see Trimble 2012; Wang, Jensen, and Im 2010). In Figures 1(*a*) and (*b*), in each case two complements emerged. The calculation of the P^{R} or P^{F} metrics only considers the distance for normalization which is farthermost from c^{S} within the extent of R or F. However, the normalization factor can cause inaccuracies when the complement is characterized by an irregular shape. For instance, the gravity centres of a horseshoe-shaped river or road objects can be situated outside of the polygons. In such situations, a normalization factor such as the square root of the *Area of Intersect* (A^{S}) is more suitable:

$$P^{\mathrm{X}} = 1 - \frac{\mathrm{dist}\left(c^{\mathrm{S}}, c^{\mathrm{X}}\right)}{\mathrm{dist}\left(c^{\mathrm{S}}, c^{\mathrm{X}^*,\mathrm{max}}\right)} \text{ with } \mathrm{X} \in [\mathrm{R}, \mathrm{F}], \tag{3}$$

$$\mathrm{R}^* = \mathrm{R}/\mathrm{F}, \tag{4}$$

$$\mathrm{F}^* = \mathrm{F}/\mathrm{R}. \tag{5}$$

2.2. Combination of basic accuracy metrics

All basic metrics (B) represent a value range between 0 and 1. While the value 0 stands for no match, the value 1 indicates a complete correspondence of S with R or F regarding area or position. The normalized value ranges enable an easy combination of basic metrics. There is a wide variety of statistical methods averaging values (Crawley 2007). The often used *root mean square* and *quadratic mean* (Clinton, Holt, and Gong 2010) leads to a stronger emphasis of high summand values. However, geometric quality is mainly determined by minimal overlaps and maximal positional distances of classified and reference objects (see Figure 1). Thus, combined metrics (C) based on the geometric and harmonic mean are more appropriate in order to smooth the effect of high values. In this study, geometric means were calculated according to Equation (6). A precondition is that each factor should be greater than zero, which is ensured by the definitions of the basic metrics (Equations (2) and (3)). This is especially true for the metrics P^{F} and P^{R} where the normalized distance between gravity centres is subtracted from the value 1. Thus, non-existing distances are represented by the value 1:

$$C = \left(\prod_{i=1}^{n} B_i\right)^{\frac{1}{n}} \tag{6}$$

$$\text{with } C \in \left[O, P, G^{\mathrm{R}}, G^{\mathrm{F}}, G\right] \text{ and } B \in \left[O^{\mathrm{R}}, P^{\mathrm{R}}, O^{\mathrm{F}}, P^{\mathrm{F}}\right],$$

where n is the number of used metrics. All combined metric variants C are listed in Table 1. The combination of O^{R} and O^{F} as well as of P^{R} and P^{F} results in the metrics O or P, which show the total mismatching of the objects regarding area and position. In doing so, two hierarchical points of view are considered, meaning that the total geometric inaccuracy of an object results from the inaccuracies of both the reference and classified object.

The geometric mean of P and O is considered the *Overall Geometric Accuracy* (G). The combinations of O^{F} and P^{F} as well as of O^{R} and P^{R} represent other points of view. The resulting metrics reflect the *Classified-Object-Related* (G^{F}) and *Reference-Object-Related Geometric Accuracy* (G^{R}) regarding area and position.

Table 1. Combined metrics (C; see Equation (6)) and associated basic metrics (B).

C	Description	B
O	Combined relative area	O^{R}, O^{F}
P	Combined relative position	P^{R}, P^{F}
G^{F}	Classified-object-related geometric accuracy	O^{F}, P^{F}
G^{R}	Reference-object-related geometric accuracy	O^{R},P^{R}
G	Overall geometric accuracy	O^{R}, P^{R}, O^{F}, P^{F}

2.3. *Hierarchical object relations*

As already mentioned in Section 1, the overlay of reference and classified objects leads to different types of hierarchical object relations.

(1) Figure 2(*a*) represents a *one-to-one* relation where the intersected object (S) is related unambiguously to one reference or classified object (X).
(2) Figure 2(*b*) demonstrates *one-to-many* relations between X and the intersected objects S1, S2, S3, and S4, whereas the objects can vary in size.
(3) *One-to-one* and *one-to-many* relations are characterized by unique object identifier combinations. This is in contrast to *many-to-many* relations where spatially disconnected reference and classified objects are labelled by the same identifier (Figure 2(*c*)). Such objects are often sliver polygons, which occur as a typical side effect of overlay operations (Chen et al. 2012).

One-to-one, *one-to-many*, and *many-to-many* relations are detectable by the analysis of object identifier combinations. Using the `R` functions `ddply` and `count` (Wickham 2011), a stepwise filtering procedure was applied.

(1) *One-to-one*, relations are characterized by the fact that each reference and classified object identifier as well as their combination is unique.
(2) *One-to-many* relations can be analysed considering two points of view which reflect possible users' requirements and affect geometric assessment results.
   (a) The first variant uses unique object identifier combinations but allows duplicates of reference or classified object identifiers. In Figure 2(*b*), this is true for all X:S relations.
   (b) *One-to-many* relations can also be filtered according to the highest overlap within each reference or classified object (Persello and Bruzzone 2010). After this operation, *one-to-one* relations remain. In Figure 2(*b*), only the relation X:S1 would be taken into account.

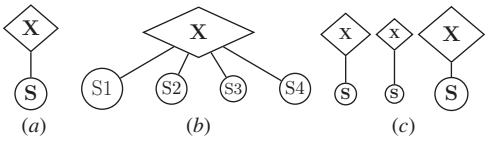

Figure 2. Schematic representations of (*a*) *one-to-one,* (*b*) *one-to-many*, and (*c*) *many-to-many* relations between classified or reference objects (X) and their corresponding intersection(s) (S). The S objects' diameters stand for different overlap sizes.

(3) Similar to variant (2)(b), the application of the majority function on *many-to-many* relations leads to *one-to-one* relations. Objects with the greatest overlap would be detected. For instance, this is true for the right-hand object relation in Figure 2(*c*).

The majority filtering operation applied on *one-to-many* (type (2)(b)) and *many-to-many* relations (type (3)) leads to a reduction of sliver polygons. This is not true for the first *one-to-many* filtering operation (type (2)(a)) where all intersected objects within a reference or classified object remain. The determination of object size thresholds which are considered as meaningful geometric changes depends on user-specific needs.

2.4. *Local and global assessment*

The local validation aims at single objects. A relative comparison of objects can be realized by using the combined metrics *O, P*, or *G*. Their relations to classified- and reference-object-related metrics express the kind and strength of mismatching of local objects (M^{l}). In this study, we use the expression $O^{F} - O^{R}$ for the characterization of the classified-object-related under-segmentation ($M^{l}(O) < 0$) and over-segmentation ($M^{l}(O) > 0$). The additional consideration of positional differences leads to the expression $G^{F} - G^{R}$. The strength of mismatching is revealed by the positive or negative value distance from zero, whereas zero represents a total correspondence of *one-to-one* objects (Equation (7)):

$$M^{l}(Y) = Y^{F} - Y^{R} \text{ with } Y \in [O, P, G]. \tag{7}$$

Means ($x(O, P, G)$) and medians ($\tilde{x}(O, P, G)$) of the distributions of combined metrics are used for the global assessment of the total data set. In addition, cumulative distribution functions of basic and combined metrics $f(C, B)$ allow a global assessment of mismatching (M^{g}) regarding strength and kind similar to the local assessment approach. But instead of using metrics related to single objects, normalized distances D between distributions of classified-object-related metrics (O^{F}, P^{F}, G^{F}) and distributions of their corresponding reference-object-related metrics (O^{R}, P^{R}, G^{R}) are calculated. This is realized by the non-parametric Kolmogorov-Smirnov (KS) goodness-of-fit test, which quantifies whether two distributions are the same as or significantly different from each other (Thas 2010). Applying the `ks.test` function from the `R` package `dgof` (Arnold and Emerson 2011), the distances of the function $f(O^{R},P^{R},G^{R})$ left (D^{-}) and right D^{+} from the function $f(O^{F},P^{F},G^{F})$ are measured (Equation (8); Figure 3):

$$\begin{aligned} D^{+} &= \max^{+} |f\left(Y^{F}\right) - f\left(Y^{R}\right)|, \\ D^{-} &= \max^{-} |f\left(Y^{F}\right) - f\left(Y^{R}\right)|, \\ &\text{with } Y \in [O, P, G]. \end{aligned} \tag{8}$$

Finally, M^{g} results from the difference between $D^{-}(O^{F},O^{R})$ and $D^{+}(O^{F},O^{R})$ (Equation (9)). The expression $M^{g} < 0$ stands for a data set which mainly contains under-segmented classified objects with or without the consideration of positional differences. The condition $M^{g} > 0$ characterizes the opposite case of over-segmentation:

$$M^{g}\left(Y^{F},Y^{R}\right) = D^{-}\left(Y^{F},Y^{R}\right) - D^{+}\left(Y^{F},Y^{R}\right) \text{ with } Y \in [O,P,G]. \tag{9}$$

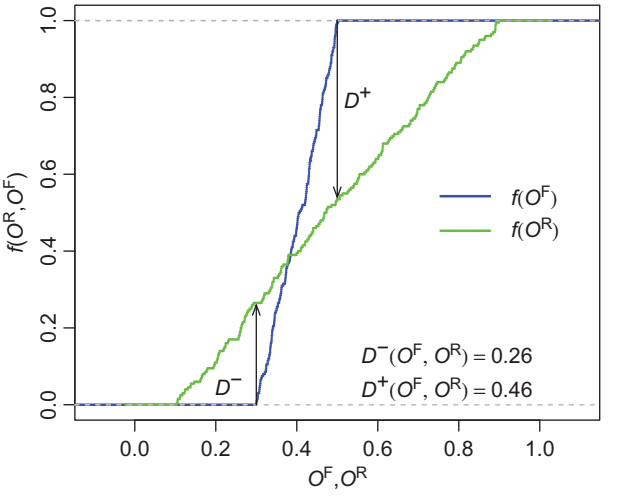

Figure 3. Principle for the calculation of D^+ and D^- values on the example of the cumulative distributions functions of the metrics, $f(O^F)$ and $f(O^R)$. D^- is calculated from the left-hand side and D^+ from right-hand side of the function $f(O^F)$.

2.5. *Sampling of reference objects*

The introduced geometric accuracy assessment approach is part of a validation process chain where thematically verified maps and classifications are geometrically assessed. The thematic validation should 'include a clear description of the sampling design [. . .], an error matrix, the area or proportion of area of each category according to the map, and descriptive accuracy measures such as user's, producer's, and overall accuracy' (Olofsson et al. 2013, 1). Especially, an appropriate sample size – characterized by specifications such as significance level and power – is of importance for the interpretation and comparability of mapping results (Foody 2009). Radoux et al. (2011) introduced an object-based sampling framework which enables the statistically sound definition of a minimal sample size. In this study, all classified objects are spatially associated with corresponding reference objects. Their survey can be, for example, the result of a manual on-screen digitization of remotely sensed data such as ortho-photographs and satellite imagery, which should fit to the thematic target scale (Möller, Lymburner, and Volk 2007).

3. Application

3.1. *Study site and input data*

The accuracy assessment procedure is applied on a study site with an area of 263 km^2 (Figure 4). The study site is situated in the south of the German Federal State of Saxony-Anhalt. For this area, a remote-sensing-based land-use classification exists, which was carried out in 2011 for a target scale of 1:25,000. The geometric accuracy assessment is exemplified on 277 classified agricultural field objects; 324 reference objects were digitized by a visual interpretation of aerial photographs.

3.2. *Local accuracy assessment*

The local accuracy assessment is exemplified on two *one-to-one* objects (Figure 5(*a*) and (*b*)) and one *one-to-many* classified object (Figure 5(*c*)). The classified objects are

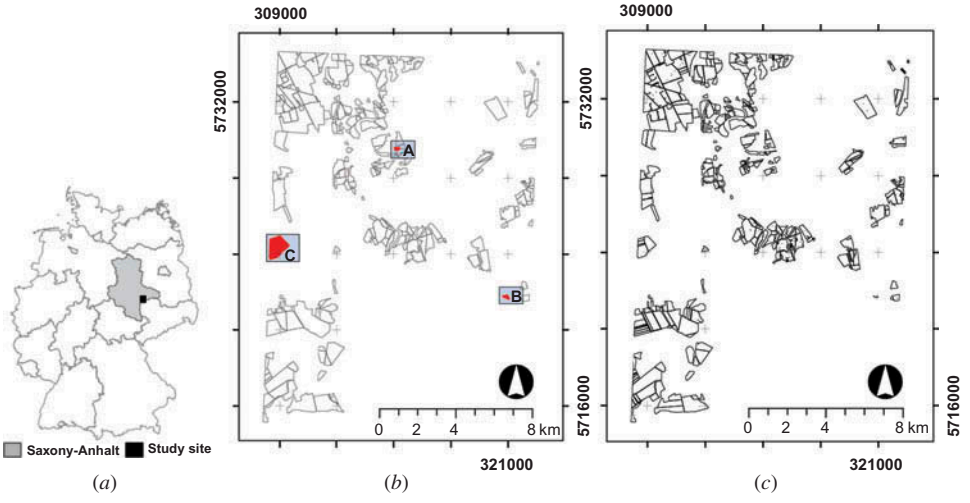

Figure 4. Test site location within Germany and Saxony-Anhalt (*a*), classified data set (*b*), and reference data set (*c*). The characters A, B, and C indicate example objects (Projection: EPSG code 32633; see Spatialreference 2013).

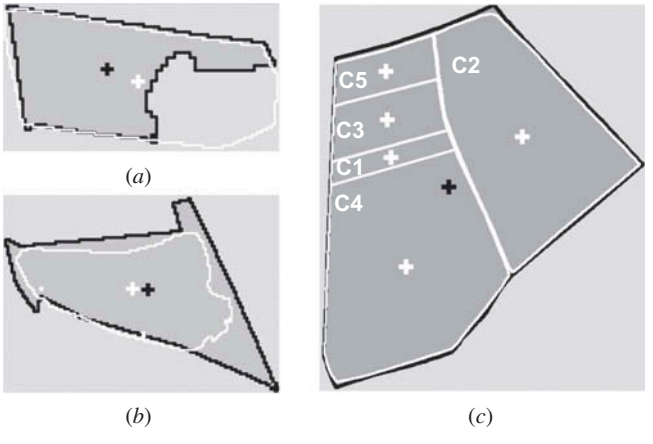

Figure 5. Visualization of the example objects A (*a*), B (*b*), and C (*c*). Their locations are shown in Figure 4. The classified objects are characterized by black-coloured boundaries and cross-marked gravity centres. The reference objects have white-coloured boundaries and gravity centres. The labels C1–C5 represent reference objects of a *one-to-many* relation.

black-framed and can be distinguished from the reference objects, which have white-coloured boundaries.

The corresponding basic and combined accuracy metrics are listed in Table 2. Accordingly, the G values of the examples A and B show a similar accuracy, whereas the classified object B is more accurate ($G = 0.88$) than object A ($G = 0.81$). The same can be seen in the O and P values. However, both examples differ in the corresponding G^{R} and G^{F} values. Their relation points out the direction of mismatching. Example A represents an over-segmented classified object which is smaller than the reference object. This is indicated by a positive $M^{1}(G)$ value. Example B shows the opposite

Table 2. Local geometric accuracy results of the example objects A, B (*one-to-one* relations), and C (*one-to-many* relation; see Figure 5). The bold emphasized columns indicate *one-to-one* relations (objects A and B) as well as a majority filtering result (object C4) according to relation (2)(b) (see Section 2.3).

Metrics	A	B	C1	C2	C3	C4	C5
O^R	**0.65**	**0.97**	1.00	1.00	1.00	**1.00**	1.00
O^F	**0.95**	**0.70**	0.04	0.37	0.08	**0.42**	0.06
O	**0.78**	**0.82**	0.19	0.61	0.28	**0.64**	0.24
P^R	**0.73**	**0.99**	1.00	1.00	1.00	**1.00**	1.00
P^F	**0.98**	**0.88**	0.66	0.54	0.53	**0.54**	0.32
P	**0.85**	**0.93**	0.82	0.73	0.73	**0.73**	0.57
G^R	**0.69**	**0.98**	1.00	1.00	1.00	**1.00**	1.00
G^F	**0.97**	**0.78**	0.15	0.45	0.21	**0.47**	0.14
G	**0.81**	**0.88**	0.39	0.67	0.45	**0.69**	0.37
$M^1(O)$	**0.31**	**−0.27**	−0.96	−0.63	−0.92	**−0.58**	−0.94
$M^1(P)$	**0.25**	**−0.10**	−0.34	−0.46	−0.47	**−0.46**	−0.68
$M^1(G)$	**0.28**	**−0.19**	−0.85	−0.55	−0.79	**−0.52**	−0.86

case of under-segmentation ($M^1(G) < 0$), which is also true for all reference objects of example C.

The relative strength of mismatching results from the comparison of absolute M^1 values: the more different is the value from zero, the more inaccurate is the object. Examples C1 and C5 are the most inaccurate objects with $M^{1,C1}(G) = -0.85$ and $M^{1,C5}(G) = -0.86$.

The joint consideration of O, P, and G values reveals which kind of inaccuracy dominates. For instance, example C1 is characterized by the smallest overlap leading to the lowest O value ($O^{C1} = 0.19$). The overlap of C5 is higher ($O^{C5} = 0.24$), but the gravity centres' distance is longer, which leads to a less positional accuracy ($P^{C1} = 0.82$, $P^{C5} = 0.57$).

3.3. *Global accuracy assessment*

3.3.1. *Effect of object size thresholds*

The overlay of the 324 reference and 277 classified data set resulted in 2617 objects. Hence, the global accuracy assessment requires the selection of meaningful objects. As mentioned in Section 2.3, a simple way to reduce the object number is the determination of a user-specific object size threshold. Figure 6 shows for the study area the relation between thresholds of object sizes as well as corresponding object numbers (N) and the overall accuracy metrics (G). Accordingly, the threshold 0.1 ha leads to the most significant change of N and G. Applied on the data set, the object number would be reduced to 361 objects.

3.3.2. *Effect of object relations*

The overlay data set of 2617 objects was filtered according to (1) *one-to-one*, (2) *one-to-many*, and (3) *many-to-many* relations (Section 2.3). Two filtering variants were applied on *one-to-many* relations which differ in regard to the consideration of all *one-to-many* relations (type (2)(a)) and objects which are characterized by the maximal overlap of spatially

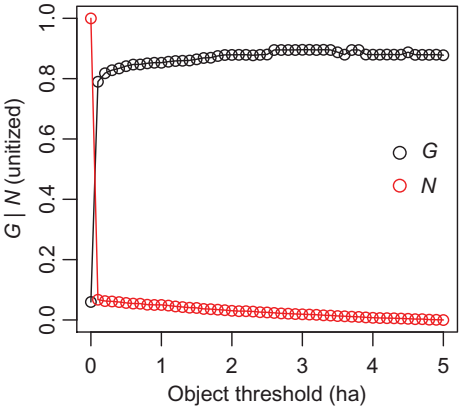

Figure 6. Relation between thresholds of object sizes, object numbers (N), and overall geometric accuracies (G). N has been normalized with zero minimum using the `R` function `data.Normalization` (Dudek and Walesiak 2012).

Table 3. Global geometric accuracy assessment results based on means (x), medians ($\tilde{x}$), and cumulative distribution distances (D^{+}, D^{-}) of the metrics O, P, and G for the relations (1), (2)(a), (2)(b), and (3) (see Section 2.3; N – object number).

Relation type	Metric	N	X	$\tilde{x}$	D^{-}	D^{+}	M^{g}
(1)	O	87	0.92	0.95	0.00	0.68	−0.68
(2)(a)	O	211	0.55	0.58	0.04	0.53	−0.49
(2)(b)	O	145	0.81	0.89	0.01	0.62	−0.61
(3)	O	165	0.23	0.01	0.04	0.42	−0.38
(1)	P	87	0.96	0.98	0.00	0.52	−0.52
(2)(a)	P	211	0.55	0.58	0.04	0.53	−0.49
(2)(b)	P	145	0.87	0.94	0.01	0.47	−0.46
(3)	P	165	0.41	0.43	0.01	0.37	−0.36
(1)	G	87	0.94	0.97	0.00	0.66	−0.66
(2)(a)	G	211	0.61	0.67	0.04	0.52	−0.48
(2)(b)	G	145	0.85	0.92	0.01	0.60	−0.59
(3)	G	165	0.28	0.09	0.01	0.41	−0.40

connected reference and classified objects (type (2)(b)). The last relation type is exemplified in Table 2, where the reference object C4 shows the greatest match within the classified object. This is indicated by a maximal O value of 0.64.

In Table 3, for each relation the means (x) and medians ($\tilde{x}$) as well as the cumulative distribution distances of the metrics *O, P,* and *G* are listed. Accordingly, 87 *one-to-one* and 165 *one-to-many* relations were detected. As was to be expected, there exist more (2) (a) than (2)(b) relations. Each relation type shows different x and $\tilde{x}$ values. The *one-to-many* relations (type (1)) are characterized by the best accuracies followed by the relations (2) and (3).

Assessing the total data set, the selection of the *one-to-many* filtering variant is crucial for the overall accuracy. The accuracy metrics contained in Figure 7 illustrates the effect: the combination of the relations '(1) + (2)(a) + (3)' resulted in an accuracy which is significantly lower than of the combination '(1) + (2)(b) + (3)'. Both combinations also show different object numbers of 463 and 397, which both differ from the object number

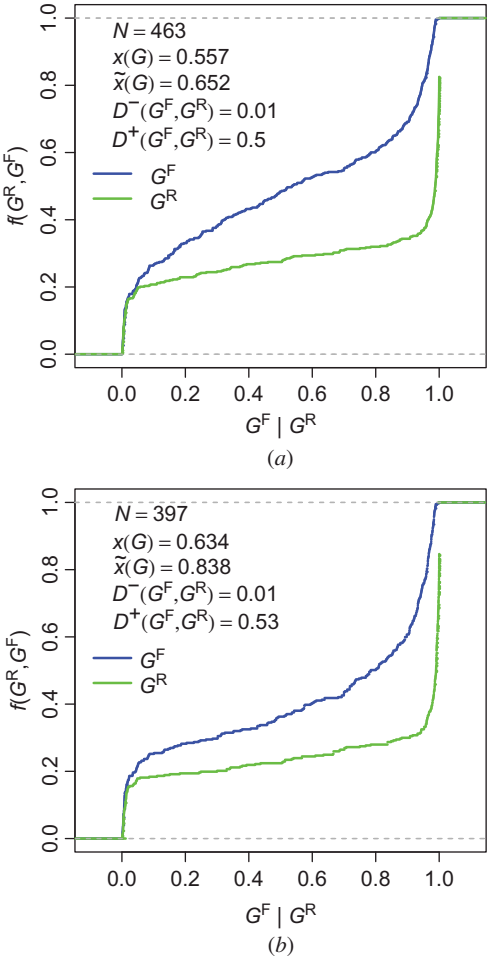

Figure 7. Visualization of global geometric accuracy assessment results based on means (x), medians ($\tilde{x}$), and cumulative distribution distances (D^+, D^-) of the metric G for the combined relations '(1) + (2)(a) + (3)' (*a*) and '(1) + (2)(b) + (3)' (*b*) (see Section 2.3; N – object number).

resulting from the threshold-based filtering ($N = 361$; Section 3.3.1). This means that both combinations contain objects which are smaller than 0.1 ha. Their elimination would lead to different accuracy values.

For all relations, the KS test resulted in small D^- and higher D^+ values. The corresponding negative M^g values indicate that the majority of the classified objects are mainly under-segmented, which is also true after the consideration of positional mismatching. Figure 7 exemplifies the situation for both combinations of the total data set, where the cumulative distribution distances (D^+, D^-) of the metrics G^F and G^R are shown. According to the M^g values listed in Table 3, *one-to-one* relations present the highest strength of mismatching followed by the relation types *one-to-many* and *many-to-many.* This means that *one-to-one* relations contain more under-segmented classified objects than other relation types in spite of the fact that their geometric accuracies – expressed by x and $\tilde{x}$ values – are the highest. In Figure 7, the strength of mismatching is slightly smaller for the combination '(1) +(2)(a) + (3)'.

4. Discussion and conclusion

European efforts to harmonize and provide geodata lead to increased demands regarding the thematic and geometric quality of classification results. While standard techniques for the point-related thematic accuracy assessment have already been established, there is a deficit in commonly accepted protocols for the geometric assessment of two-dimensional classified objects.

In this study, a framework for the geometric validation of thematic maps is presented. Framework means that the underlying methods and geometric metrics can be adapted to user-specific requirements (Pontius et al. 2007). The framework is based on accuracy metrics whose calculation corresponds to a hierarchical comparison of the properties of reference and classified objects. Here, basic metrics based on differences regarding objects' areas and gravity centre positions were calculated. Their combination to aggregated metrics enables a user-specific identification and assessment of geometric inaccuracies which are related to overlapping and positional mismatching. Independent of the combination approach (here, geometric mean), the metrics make both possible a relative geometric accuracy comparison of single objects as well as an assessment of the kind (over- and under-segmentation) and strength of mismatching.

The global accuracy assessment was realized by statistical metrics characterizing distributions. Means and medians of combined accuracy metrics describe the overall accuracy. The kind and strength of mismatching of data sets is assessed by relations between cumulative distribution distances.

The framework includes the detection of *one-to-one*, *one-to-many*, and *many-to-many* object relations, which result from the overlay of classified and reference objects. Thus, users can control which objects should go into the global accuracy assessment. This concerns in particular *one-to-many* object relations. The assessment can contain all relations or only objects which are filtered according to the highest overlap of reference and classified objects. On the example of a German test site, we could show that their selection is crucial. Both accuracy assessment results could be considered as domain of uncertainty.

The accuracy assessment framework can also be adapted to users' requirements by weighting of metrics (e.g. according to object areas) and by the integration of metrics, characterizing other object properties within a value range of zero and one.

The geometric accuracy results can be coupled with a point-related thematic accuracy assessment (Persello and Bruzzone 2010). In this respect, the introduced approach can be considered as an in-depth assessment of thematically tested objects. In doing so, the samples used for the thematic validation also have to mark the locations of the corresponding reference objects.

Reference objects should fit to the scale of the classified data set which is to be assessed. However, there often exists a scale-related gap between reference and classified objects meaning that the delineated reference objects are 'too accurate' or 'too coarse'. In principle, the proposed framework offers levers to deal with this issue. For instance, users can define which minimal object size is relevant for the accuracy assessment (see Section 3.3.1). Another option is the definition of tolerance levels based on the local mismatching metric (M^1; Section 2.4). For instance, the strength of mismatching is an expression of the difference in area of objects. A tolerance level could define which difference is acceptable and reflects the scale-specific gap. This issue is the subject of further investigations.

Acknowledgements

This study was supported by the German Ministry of Economics and Technology and managed by the German Aerospace Center (contract no.: FKZ 50EE0915 and FKZ 50EE1230). We would like to thank the editor Timothy Warner as well as three anonymous reviewers who helped to improve the article.

References

Arnold, T. A., and J. W. Emerson. 2011. "Nonparametric Goodness-of-Fit Tests for Discrete Null Distributions." *The R Journal* 3: 34–39.

Berger, M., J. Moreno, J. A. Johannessen, P. F. Levelt, and R. F. Hanssen. 2012. "ESA's Sentinel Missions in Support of Earth System Science." *Remote Sensing of Environment* 120: 84–90.

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

Chen, G., G. J. Hay, L. M. T. Carvalho, and M. A. Wulder. 2012. "Object-based Change Detection." *International Journal of Remote Sensing* 33: 4434–4457.

Clinton, N., A. Holt, and P. Gong. 2010. "Accuracy Assessment Measures for Object-based Image Segmentation Goodness." *Photogrammetric Engineering & Remote Sensing* 76: 289–299.

Crawley, M. J. 2007. *The R Book*. Chichester: Wiley.

Dudek, M., and A. Walesiak. 2012. *clusterSim: Searching for Optimal Clustering Procedure for a Data Set*. R Package Version 0.41-8. Accessed July 22, 2013. http://CRAN.R-project.org/package=clusterSim

Evans, R. 2013. "Assessment and Monitoring of Accelerated Water Erosion of Cultivated Land – When Will Reality Be Acknowledged?" *Soil Use and Management* 29: 105–118.

Foody, G. M. 2002. "Status of Land Cover Classification Accuracy Assessment." *Remote Sensing of Environment* 80: 185–01.

Foody, G. M. 2009. "Sample Size Determination for Image Classification Accuracy Assessment and Comparison." *International Journal of Remote Sensing* 30: 5273–5291.

Hernando, A., 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.

Hussain, M., D. Chen, A. Cheng, H. Wei, and D. Stanley. 2013. "Change Detection from Remotely Sensed Images: From Pixel-Based to Object-Based Approaches." *ISPRS Journal of Photogrammetry and Remote Sensing* 80: 91–106.

ISO 19138. 2006. *Geographic Information: Data Quality Measures*. Geneva: International Organization for Standardization.

Jakobsson, A., and J. Giversen. 2008. *Guidelines for Implementing the ISO 19100 Geographic: Information Quality Standards in National Mapping and Cadastral Agencies*. Brussels: EuroGeographics.

Liu, C., P. Frazier, and L. Kumar. 2007. "Comparative Assessment of the Measures of Thematic Classification Accuracy." *Remote Sensing of Environment* 107: 606–616.

Liu, D., and F. Xia. 2010. "Assessing Object-Based Classification: Advantages and Limitations." *Remote Sensing Letters* 1: 187–194.

Möller, M., L. Lymburner, and M. Volk. 2007. "The Comparison Index: A Tool for Assessing the Accuracy of Image Segmentation." *International Journal of Applied Earth Observation and Geoinformation* 9: 311–321.

Olofsson, P., G. M. Foody, S. V. Stehman, and C. E. Woodcock. 2013. "Making Better Use of Accuracy Data in Land Change Studies: Estimating Accuracy and Area and Quantifying Uncertainty Using Stratified Estimation." *Remote Sensing of Environment* 129: 122–131.

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.

Pontius, R. G., R. Walker, R. Yao-Kumah, E. Arima, S. Aldrich, M. Caldas, and D. Vergara. 2007. "Accuracy Assessment for a Simulation Model of Amazonian Deforestation." *Annals of the Association of American Geographers* 97: 677–695.

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.

Schreier, G., S. Dech, E. Diedrich, H. Maass, and E. Mikusch. 2008. "Earth Observation Data Payload Ground Segments at DLR for GMES." *Acta Astronautica* 63: 146–155.
Sebari, I., and D.-C. He. 2013. "Automatic Fuzzy Object-Based Analysis of VHSR Images for Urban Objects Extraction." *ISPRS Journal of Photogrammetry and Remote Sensing* 79: 171–184.
Spatialreference. 2013. Catalogs of Spatial Reference Systems. Accessed July 22, 2013. http://spatialreference.org.
Stehman, S. V. 1997. "Selecting and Interpreting Measures of Thematic Classification Accuracy." *Remote Sensing of Environment* 62: 77–89.
Straub, B. M., and C. Heipke. 2004. "Concepts for Internal and External Evaluation of Automatically Delineated Tree Tops." *International Archives of Photogrammetry and Remote Sensing* 26: 62–65.
Thas, O. 2010. *Comparing Distributions*. Springer Series in Statistics. 1st ed. Heidelberg: Springer.
Trimble. 2012. *eCognition Developer 8.8 Reference Book*. Munich: Trimble Germany GmbH.
Versic, A. 2009. "Spatial Data Quality Control Process Based on ISO 19113 and ISO 19114." In *24th International Cartographic Conference (ICC 2009)*, Santiago, November 15–21.
Volk, M., M. Möller, and D. Wurbs. 2010. "A Pragmatic Approach for Soil Erosion Risk Assessment Within Policy Hierarchies." *Land Use Policy* 27: 997–1009.
Wang, Z., J. R. Jensen, and J. Im. 2010. "An Automatic Region-based Image Segmentation Algorithm for Remote Sensing Applications." *Environmental Modelling & Software* 25: 1149–1165.
Wickham, H. 2011. "The Split-Apply-Combine Strategy for Data Analysis." *Journal of Statistical Software* 40: 1–29.
Zhan, Q., M. Molenaar, K. Tempfli, and W. Shi. 2005. "Quality Assessment for Geo-spatial Objects Derived from Remotely Sensed Data." *International Journal of Remote Sensing* 26: 2953–2974.