

Sub-pixel confusion–uncertainty matrix for assessing soft classifications

J.L. Silván-Cárdenas^{a,b}, L. Wang^{c,*}

^a Texas State University-San Marcos, GIS Center-Department of Geography, 601 University Dr., ELA 139, San Marcos, TX 78666, USA

^b Centro de Investigación en Geografía y Geomática, Ing. “Jorge L. Tamayo”, Contoy 137, Lomas de Padierna, Mex. D.F., Mexico

^c University at Buffalo, the State University of New York, Department of Geography, 105 Wilkeson Quad., Buffalo, NY 14261, USA

Received 20 January 2007; received in revised form 19 July 2007; accepted 21 July 2007

Abstract

The prevailing concerns on ecological and environmental issues, occurring especially at regional to global scales, have prompted significant advances on the use of remote sensing data for the estimation of land cover information at sub-pixel level. However, the quality of such classification products, as well as the performance of the classification protocol employed, are difficult to quantify. This paper had the objectives of 1) reviewing the existing alternatives, while identifying major drawbacks and desirable properties, for sub-pixel accuracy assessment based on cross-comparison matrices, and 2) developing theoretical grounds, for a more general accuracy assessment of soft classifications, that account for the sub-pixel class distribution uncertainty. It was found that, for a sub-pixel confusion matrix to exhibit a diagonalization characteristic that allows identifying a perfect matching case, the agreement measure must be constrained at pixel level, whereas a disagreement measure can take into account the sub-pixel distribution uncertainty, leading to an underspecified problem termed *the sub-pixel area allocation problem*. It was demonstrated that the sub-pixel area allocation problem admits a unique solution if, and only if, no more than one class is either over- or underestimated at each pixel. In this case, the sub-pixel confusion can be uniquely determined. When no unique solution exists, the space of feasible solutions can be represented by confusion intervals. A new cross-comparison matrix that reports the confusion intervals in the form of a center value plus–minus maximum error was proposed to account for the sub-pixel distribution uncertainty. The new matrix is referred to as *sub-pixel confusion–uncertainty matrix* (SCM). Sub-pixel accuracy measures were also derived from this matrix. The practical use of the SCM and derived indices was demonstrated in assessing an invasive species detection method and a fuzzy classification of urban land use/land cover through remote sensing procedures.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Soft classification; Accuracy assessment; Error matrix; Uncertainty

1. Introduction

The significance of land cover as an environmental variable has made remote sensing one of the most attractive tools for the production of thematic maps of the earth's surface. However, in order for remote sensing to succeed as a valuable source of land cover information, more reliable accuracy measures are needed (Foody, 2002). In the past few decades, the prevailing concerns on ecological and environmental issues, occurring especially at regional to global scales, have prompted significant advances on the use of remote sensing data for the estimation of land cover information at sub-pixel level (Carpenter et al., 1999;

Cross et al., 1991; Fisher & Pathirana, 1990; Gutman & Ignatov, 1998; Roberts et al., 1993). However, the quality of such classification products, as well as the performance of the classification protocol employed, are difficult to quantify. Moreover, there is an increasing need for sub-pixel and super-pixel assessment of classification products made evident by recent remote sensing research (Latifovic & Olthof, 2004; Okeke & Karniele, 2006; Ozdogan & Woodcock, 2006; Shabanov et al., 2005). The assessment of the conventional (hard) allocation of image pixels to discrete classes has been standardized (to some extent) through the confusion matrix and some derived measures (Congalton, 1991; Congalton & Green, 1999; Stehman & Czaplewski, 1998). However, this method is appropriate only for hard classifications, where it is assumed that each pixel is associated with only one class in both the assessed and the reference datasets. For soft classifications,

* Corresponding author.

E-mail addresses: jlsilvan@txstate.edu (J.L. Silván-Cárdenas), lewang@buffalo.edu (L. Wang).

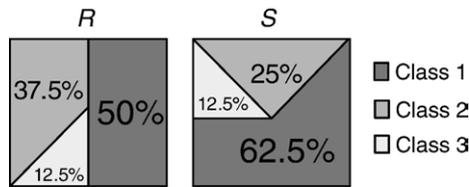


Fig. 1. Hypothetical reference (R) and assessed (S) sub-pixel partitions.

where multiple classes are assigned to a single pixel, a comparable standardized assessment procedure has not been established yet.

An argument is made here that a closer examination of the agreement–disagreement measures at sub-pixel level may allow new insights into the per-class accuracy and confusion of soft classifications. For example, consider the hypothetical reference and assessed sub-pixel partitions depicted in Fig. 1, where perfect co-registration is assumed. Table 1(a) shows the pairwise overlapping proportions that result from the overlay of the pixel partitions, wherein columns represent reference categories, and rows represent assessed categories. For instance, reference class 1 overlaps with assessed class 2 in 12.5%. Of course, the sub-pixel class boundaries are unknown in practice, and there is no way to determine the actual overlap among the classes based solely on the land-cover fractions. This problem is hereafter referred to as *sub-pixel area allocation problem*. One possible sub-pixel comparison approach can be based on the expected overlap among the classes. The expected overlap is determined from multiplication of the corresponding fractions (Table 1(b)). As we shall see below, a value so-computed provides a measure of overlap by chance between two classes, provided that the partitions have been independently produced. For instance, the expected overlap between reference class 1 (50%) and assessed class 2 (25%) is $0.5 \times 0.25 = 0.125$, or 12.5%. This value is shown in row 2 and column 1 of Table 1(b) and agrees with the actual value of Table 1(a); however, this is not always the case. A natural question is: What is the largest possible variation in the sub-pixel overlap? The answer to this question lies in determining the minimum and maximum possible overlap proportions. The minimum overlap occurs when the spatial distributions of the two classes are assumed as disjoint as possible within the pixel. This proportion is given by the excess, in respect to 100%, from the sum of the respective percentages in the reference and assessed pixels. For instance, the minimum overlap between reference class 1 and assessed class 2 is $\max(50 + 25 - 100, 0) = 0$. The minimum overlap percentages are presented in Table 1(c). On the other hand, the maximum overlap occurs when the spatial distributions of the two classes are assumed as coincident as possible within the pixel. This proportion is given by the minimum between the two percentages. For instance, the maximum possible overlap between reference class 1 and assessed class 2 is $\min(50, 25) = 25\%$. The maximum overlap percentages are presented in Table 1(d). At this point, the difficulties for sub-pixel comparison should be evident. Notice that the cross-comparison matrices given in Table 1(b)–(d) do not provide meaningful information at the pixel level. Specifically, the perfect matching

case could not be easily identified, nor would it be easy to tell which of two classifications has better accuracy based on this kind of analysis. In practice, the problem is further aggravated by the spatial misalignment and sampling issues. These issues, however, will not be addressed here. Instead, a negligible impact of registration accuracy on the sub-pixel confusion will be assumed in our analyzes.

For the evaluation of soft classifications in general, various suggestions have been made (Binaghi et al., 1999; Congalton, 1991; Foody, 1995; Gopal & Woodcock, 1994; Green & Congalton, 2004; Lewis & Brown, 2001; Pontius & Cheuk, 2006; Townsend, 2000), among which, the fuzzy error matrix (Binaghi et al., 1999) is one of the most appealing approaches, as it represents a generalization (grounded on the fuzzy set theory) of the traditional confusion matrix. In spite of its sounding theoretical basis, the fuzzy error matrix is not generally adopted as a standard accuracy report and statement for soft classifications. Some reasons for this have been highlighted as counterintuitive characteristics (Pontius & Cheuk, 2006). Specifically, for a cross-comparison to be consistent with the traditional confusion matrix, it is desirable that the cross-comparison results in a diagonal matrix when a map is compared to itself, and that its marginal totals match the total of membership grades. More importantly, a cross-comparison should convey readily interpretable information on the confusion among the classes. To date, the applicability of the fuzzy error matrix has been mainly concentrated on generating accuracy indices such as the overall accuracy, the user and producer accuracy, the kappa, and the conditional kappa coefficients (e.g., Binaghi et al., 1999; Okeke & Karnieli, 2006; Shabanov et al., 2005). Indeed, the derived indices do not

Table 1

Four cross-comparison matrices from sub-pixel partitions of Fig. 1(a) actual, (b) expected, (c) minimum, and (d) maximum possible sub-pixel overlapping area among reference (column label) and assessed (row label) class fractions

Class	1	2	3	Sum
<i>(a)</i>				
1	37.5	12.5	12.5	62.5
2	12.5	12.5	0.00	25.0
3	0.0	12.5	0.0	12.5
Sum	50.0	37.5	12.5	100.0
<i>(b)</i>				
1	31.3	23.4	7.8	62.5
2	12.5	9.4	3.1	25.0
3	6.3	4.7	1.6	12.5
Sum	50.0	37.5	12.5	100.0
<i>(c)</i>				
1	12.5	0.0	0.0	12.5
2	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0
Sum	12.5	0.0	0.0	12.5
<i>(d)</i>				
1	50.0	37.5	12.5	100.0
2	25.0	25.0	12.5	62.5
3	12.5	12.5	12.5	37.5
Sum	87.5	75	37.5	200.0

account for the off-diagonal cells of the matrix; rather, they are based only on the diagonal cells and the total grades from the reference and assessed datasets (Binaghi et al., 1999). Recently, a composite operator was proposed for computing a cross-comparison matrix that exhibits some of the aforementioned desirable characteristics (Pontius & Cheuk, 2006). Pontius and Cheuk (2006) showed how the composite operator can be used for a multi-resolution assessment of raster maps and compared it with other alternatives, including the traditional hardening of pixels, the minimum operator (Binaghi et al., 1999), and the product operator (Lewis & Brown, 2001). This composite operator was also suggested as a viable tool for the sub-pixel comparison of maps (Pontius & Connors, 2006). Although several desirable properties are found in the composite operator, its utility has been only demonstrated on the use of traditional accuracy indices (Kuzera & Pontius, 2004; Pontius & Cheuk, 2006; Pontius & Connors, 2006), and neither has the use of the off-diagonal cells been demonstrated, nor is their interpretation clear.

In light of the above, the objectives of this research are:

- (1) Review the existing alternatives, while identifying major drawbacks and desirable properties, for sub-pixel accuracy assessment based on cross-comparison matrices, and
- (2) Develop theoretical grounds, for a more general accuracy assessment of soft classifications, that account for the sub-pixel class distribution uncertainty.

The remaining of the paper is organized as follows. In Section 2, a theoretical framework for generalized cross-comparison matrices is discussed, and various potential cross-comparison matrices for sub-pixel accuracy assessment are compared within this framework. In Section 3, a new cross-comparison matrix that reports the confusion interval in the form of a center value plus–minus maximum error is introduced. The new matrix is referred to as *sub-pixel confusion–uncertainty matrix* (SCM). Sub-pixel accuracy measures are also derived and illustrated in this section. In Section 4, two application examples are described and the results presented in Section 5. A thorough discussion with concluding remarks is presented in Section 6.

2. Theoretical background

2.1. Notation and definitions

The following symbols are used throughout the text:

N	number of pixels in the reference and assessed datasets,
K	number of categories or classes,
n	pixel index, where $n = 1, \dots, N$,
k, l	class indices, where $k, l = 1, \dots, K$,
s_{nk}	grade of membership of pixel n to class k assigned by the assessed dataset,
r_{nl}	grade of membership of pixel n to class l assigned by the reference dataset,

s_{+k}	total grade of class k from the assessed dataset, $s_{+k} = \sum_n s_{nk}$,
r_{+l}	total grade of class l from the reference dataset, $r_{+l} = \sum_n r_{nl}$,
s'_{nk}	overestimation error of class k at pixel n , $s'_{nk} = \max(s_{nk} - r_{nk}, 0)$,
r'_{nl}	underestimation error of class l at pixel n , $r'_{nl} = \max(r_{nl} - s_{nl})$,
p_{nkl}	agreement–disagreement between membership grades from assessed class k and reference class l at pixel n ; it is called <i>agreement</i> when $k=l$ and <i>disagreement</i> (or <i>confusion</i>) when $k \neq l$,
P_{kl}	overall agreement–disagreement between assessed class k and reference class l , $P_{kl} = \sum_n p_{nkl}$,
P_{k+}	marginal row sum of P_{kl} for class k , $P_{k+} = \sum_l P_{kl}$,
P_{+l}	marginal column sum of P_{kl} for class l , $P_{+l} = \sum_k P_{kl}$,
P_{++}	total sum of P_{kl} , $P_{++} = \sum_k \sum_l P_{kl}$,

The grade of membership have various interpretations throughout the text, including possibility, probability and sub-pixel fractions of land-cover. In any case, they are constrained so that $0 \leq r_{nk} \leq 1$, $0 \leq s_{nk} \leq 1$ and $\sum_k r_{nk} = \sum_k s_{nk} = 1$ hold for all the pixels. The agreement–disagreement at pixel n , p_{nkl} , is computed using a comparison operator of the form $C(s_{nk}, r_{nl})$. The notations p_{nkl}^C and P_{kl}^C may be used to specify the comparison operator, C , employed for computing the per-pixel and overall agreement–disagreement, respectively. A confusion matrix is a cross-tabulation (see Table 2(a)) formed by the overall agreement–disagreement, P_{kl} , where row and column labels of the matrix represent assessed categories and reference categories, respectively. Thus, the agreement values correspond to the diagonal cells, whereas the disagreement values correspond to the off-diagonal cells.

Table 2
Traditional error matrix for K classes (a) and common accuracy indices (b)

(a)					
Class	Reference				Row total
	Class 1	Class 2	...	Class K	
Class 1	P_{11}	P_{12}	...	P_{1K}	P_{1+}
Class 2	P_{21}	P_{22}	...	P_{2K}	P_{2+}
⋮	⋮	⋮	⋮	⋮	⋮
Class K	P_{K1}	P_{K2}	...	P_{KK}	P_{K+}
Col. Tot.	P_{+1}	P_{+2}	...	P_{+K}	P_{++}

(b)	
Accuracy index	Formula ^a
Overall accuracy, OA	$\frac{\sum_k P_{kk}}{P_{++}}$
k -th User Accuracy, UA(k)	$\frac{P_{kk}}{P_{k+}}$
k -th Producer Accuracy, PA(k)	$\frac{P_{kk}}{P_{+k}}$
Expected proportion of agreement, P_c	$\frac{\sum_k P_{k+} P_{+k}}{P_{++}^2}$
Kappa coefficient of agreement, κ	$\frac{(P_{++} - P_c)}{1 - P_c}$

^a P_o is the observed proportion of agreement=OA.

2.2. A generalized cross-comparison framework

Hard classifications are commonly assessed through the so-called confusion matrix (also known as error matrix) and a series of derived indices (Congalton, 1991; Congalton & Green, 1999; Stehman & Czaplewski, 1998). Once generated, the confusion matrix can be used for a series of descriptive and analytical techniques, such as those based on accuracy indices. Table 2(a) shows the general structure of the confusion matrix and Table 2(b) lists the most common accuracy indices derived from the confusion matrix. Details on the definitions and uses of these accuracy indices for hard classifications can be found in Congalton (1991), Congalton and Green (1999), Stehman and Czaplewski (1998), to list just a few. Naturally, many researchers have tried to generalize the confusion matrix for soft classifications (Binaghi et al., 1999; Latifovic & Olthof, 2004; Lewis & Brown, 2001; Pontius & Cheuk, 2006; Townsend, 2000; Woodcock & Gopal, 2000). While all these efforts have some value for various remote sensing applications, the theoretical background behind each of them does not generally consider a set of desirable characteristics inherited from the confusion matrix, and thus are difficult to interpret (at least in the sense of a traditional confusion matrix).

In the search for the fundamental characteristics of a generalized confusion matrix for soft classifications, it is sometimes suggested that the matrix should fulfill two characteristics:

- i. *Diagonalization*. The matrix should be diagonal if, and only if, the assessed data match perfectly the reference data.
- ii. *Marginal sums*. Marginal sums should match the total grades from the reference and assessed data.

The first characteristic is desirable for the matrix to be useful in identifying a perfect matching case; nevertheless, it does not constrain the matrix characteristic under slight deviation from the perfect match. Therefore, many alternatives could be envisaged that lead to a unique diagonal matrix for the perfect matching case, but to rather different matrices when non-perfect match occurs. The second characteristic is desirable (although it may not be necessary) for the matrix to be useful in deriving accuracy indices such as those listed in Table 2(b). For hard classifications, accuracy indices are customarily written in terms of row and column totals, provided that these marginal sums correspond to the class proportions in the assessed and reference datasets, respectively (i.e., $P_{+l}=r_{+l}$, $P_{k+}=s_{k+}$, and $P_{++}=N$). For soft classifications, however, marginal sums not matching the class proportions are often ignored, and class proportions are used instead for the computation of accuracy indices (e.g., Binaghi et al., 1999; Okeke & Karniele, 2006; Shabanov et al., 2005).

Instead of looking at the properties of a generalized confusion matrix, we seek to establish a number of fundamental properties on the agreement–disagreement measures that can lead to meaningful matrix entries. These properties are loosely described below and then formalized in mathematical terms. First, a meaningful agreement measure does not consider whether the assessed pixel membership is above or below the

reference pixel membership, i.e., does not depend on the over- or underestimation errors. In contrast, the sense and amount of discrepancy are important for defining a disagreement measure. An overestimation of the reference pixel membership by the assessed pixel membership leads to errors of commission type. These commission errors appear in the off-diagonal cells along the row of the class. Conversely, an underestimation of the reference value by the assessed value leads to errors of omission type. These omission errors appear in the off-diagonal cells along the column of the class. Second, agreement and disagreement are, in some sense, complimentary yet non-negative measures. This is also stated by the constrained marginal sums characteristic. Consequently, when the agreement for a given class achieve its maximum (e.g., in the case of a perfect match), the overall disagreement (sum of off-diagonal cells) for that class must be minimum (zero). Conversely, if the overall disagreement is maximum, then the agreement is minimum.

In formal grounds, one requires the agreement–disagreement measure to conform to Eq. (1), where A and D denote agreement and disagreement operators, respectively, which satisfy the properties outlined in Table 3, and s'_{nk} and r'_{nl} denote the over- and underestimation errors at pixel n . Notice that the expressions given in Eqs. (2) and (3) for the over- and underestimation errors, respectively, are mathematically equivalent to those given in Section 2.1.

$$C(s_{nk}, r_{nl}) = \begin{cases} A(s_{nk}, r_{nl}) & \text{if } k = l \\ D(s'_{nk}, r'_{nl}) & \text{if } k \neq l \end{cases} \quad (1)$$

$$s'_{nk} = s_{nk} - \min(s_{nk}, r_{nk}) \quad (2)$$

$$r'_{nl} = r_{nl} - \min(s_{nl}, r_{nl}) \quad (3)$$

The *Commutativity* property expresses a symmetric characteristic of agreement–disagreement measures respect to its arguments. It ensures that under- and overestimation of a membership grade are equally considered. The *Positivity* ensures closure over the positive space of membership values and, together with the *Nullity* property, constrains non-null comparison values to pixels with non-null membership values. The *Upper bound* property implies that the comparison operator measures essentially the degree of similarity, as contrasted to dissimilarity or distance measures, between two membership values. The *Homogeneity* property states that the agreement–disagreement values can be denormalized in proportion to

Table 3
Basic properties for general agreement and disagreement measures. $C(s,r)$ denotes a comparison (agreement or disagreement) measure between grades s and r , and a is a positive number

Property	Definition	Agreement	Disagreement
i. Commutativity	$C(s,r)=C(r,s)$	Yes	Yes
ii. Positivity	$s > 0 \vee r > 0 \Rightarrow C(s,r) > 0$	Yes	Yes
iii. Nullity	$s = 0 \wedge r = 0 \Rightarrow C(s,r) = 0$	Yes	Yes
iv. Upper bound	$C(s,r) \leq C(r,r)$	Yes	No
v. Homogeneity	$C(as,ar) = aC(s,r)$	Yes	Yes

denormalized grade values. This property is desirable when the accuracy assessment is inserted in a multi-resolution framework (see below). In sum, two notable differences between agreement and disagreement measures are established: 1) an agreement value depends on the original assessed and reference values, whereas a disagreement value depends on the over- and under estimation errors, and 2) an agreement value has an upper bound at perfect match, whereas disagreement values do not share an upper bound at null agreement.

For a more complete accuracy assessment, a multi-resolution approach has been set forth (Pontius, 2002; Pontius & Cheuk, 2006), which consists of the use of derived agreement measures at many aggregation levels. Coarser pixel membership are determined through spatial aggregation of membership from the finest resolution pixels. Because the number of aggregated finest-resolution pixels may not be always the same (e.g., around the image boundary), a weighted average of the contributions to the agreement–disagreement must be used. The weight of each coarse-pixel is set to the number of fine-resolution pixels that constitute it. The normalized matrix entry at a given resolution is computed as in Eq. (4), where w_n denotes the weight for each pixel.

$$P_{kl} = \frac{\sum_n w_n P_{nkl}}{\sum_n w_n} \quad (4)$$

Of course, this multi-resolution approach can be adopted regardless of the type of operator employed. However, the method to compute the matrix entries, P_{kl} , should remain consistent across the resolutions. This is ensured only if the operator satisfies the homogeneity property (Table 3). The homogeneity property ensures that the weighted comparisons in Eq. (4) results in normalized comparisons of weighted membership values.

2.3. Operators for assessing sub-pixel classifications

2.3.1. The sub-pixel ontology

According to Pontius and Cheuk (2006), each cross-comparison operator is rooted on a specific ontology of the pixel. More specifically, a cross-comparison depends on how the pixel–class relationship is defined and quantified. Two major definitions for the pixel–class relationship that admit multiple memberships have been broadly used in the land-cover classification research. The first definition conceives this relationship as uncertain and formalizes it through the probability theory. The second definition conceives this relationship as vague or ambiguous and formalizes it through the fuzzy set theory. In both cases, the hard classification is covered as a special case, that is when no uncertainty nor ambiguity exists. A third pixel–class relationship, that has received less attention (at least from the accuracy assessment point of view, with exception of Latifovic and Olthof (2004), relates pixel to class through a fractional land cover. This pixel–class relationship definition implies the existence of unknown crisp boundaries among the classes at sub-pixel level. It should be noted that the kind of uncertainty a sub-pixel classification

represents can be related only to the positional accuracy resolved by the sensor (pixel resolution). Soft classifications emphasizing the thematic uncertainty, which are linked to the impossibility of uniquely identifying a land cover category regardless of the sensor resolution, are not being considered here. If the uncertainty represented by a soft classification describes vague concepts, which are characteristic of the human reasoning, then other alternatives may be pursued (see for instance Gopal & Woodcock, 1994).

For the sub-pixel ontology we consider that:

- (1) The pixel–class relationship is defined through the sub-pixel fraction of class coverage, and
- (2) The agreement–disagreement is quantified as the proportion of area overlap among the classes at sub-pixel level.

2.3.2. Basic operators

Various operators have been developed under rather distinct pixel ontologies, some of which are listed in Table 4. Some have been considered previously within a multi-resolution framework (Kuzera & Pontius, 2004; Pontius & Cheuk, 2006), and their sub-pixel interpretation has been also discussed (Pontius & Connors, 2006). However, the major focus has been so far on the accuracy indices at multiple resolutions. In the present review, we investigate their suitability for assessing sub-pixel classifications within the context of the generalized cross-comparison framework introduced above.

The *minimum* operator (MIN) is the classic fuzzy set *intersection* operator. This operator has been suggested as the natural choice for producing cross-comparison matrices for fuzzy classifications (Binaghi et al., 1999). In the traditional ontology of fuzzy classifications, the pixel–class relationship describes the admission of the possibility (given by a so-called membership function) that the pixel is a member of a class. This pixel–class relationship definition is useful for handling the imprecision of meaning of concepts that are characteristic of much of the human reasoning (Gopal & Woodcock, 1994). The area estimation by map users is generally difficult under this ontology (Woodcock & Gopal, 2000). In the sub-pixel fraction ontology, the MIN operator measures the maximum sub-pixel overlap among the classes, as demonstrated in the introductory example. Therefore, if membership values are (linearly) related to sub-pixel land cover areas (see e.g., Shabanov et al., 2005), the fuzzy set intersection operator corresponds to the maximum sub-pixel overlap between the classes. The minimum operator

Table 4
Four basic operators

Operator ID	Operator of the form ^a $C(s_{nk}, r_{nl})$	Traditional interpretation	Sub-pixel interpretation
MIN	$\min(s_{nk}, r_{nl})$	Fuzzy set intersection	Maximum overlap
SI	$1 - \frac{ s_{nk} - r_{nl} }{s_{nk} + r_{nl}}$	Similarity index	Normalized maximum
PROD	$s_{nk} \times r_{nl}$	Joint probability	Expected overlap
LEAST	$\max(s_{nk} + r_{nl} - 1, 0)$	Minimum overlap	Minimum overlap

^a s_{nk} and r_{nk} denote assessed and reference grades of class K at pixel n .

satisfies all the properties outlined in Table 3. However, the MIN matrix can overestimate the actual sub-pixel agreement–disagreement and, consequently, the marginal sums can be greater than the sub-pixel fractions. Also, even in the case of a perfect match, non-null degrees of mismatch are obtained for the off-diagonal cells. These characteristics generally limit the matrix’s utility for drawing a conclusion about the confusion among the classes.

A variant of the MIN operator is sometimes used as a *similarity index* (SI) for comparing soft classifications (see e.g., Townsend, 2000). This variant results after normalizing the MIN operator by the sum of the grade values, and can be expressed as shown in Table 4. The SI operator is also meaningful for sub-pixel comparison, as it corresponds to a normalized maximum sub-pixel overlap. Nevertheless, it does not satisfy the homogeneity property, as it is invariant under scaling of the grade values. A cross-comparison matrix based on the SI operator does not satisfy the diagonalization and marginal sums characteristics outlined above.

The *product* operator (PROD) arises from a pure probabilistic view of the pixel–class relationship. In the traditional probabilistic ontology, the pixel–class relationship represents the probability that a pixel (entirely) belongs to a class, and the PROD operator gives the *joint probability* that the reference and assessed pixels belong to two given classes, provided that the pixels have been independently classified. In the sub-pixel fraction ontology, the PROD operator measures the expected class overlap by chance between the reference and assessed sub-pixels partitions. More specifically, consider a randomly drawn point from the space spanned by pixel n . Since all the points within the pixel have the same probability to come out, then the joint probability that the point belongs to class k in the assessed partition and to class l in the reference partition is given by the product $s_{nk} \times r_{nl}$, provided that the land-cover fractions s_{nk} and r_{nl} have been generated by independent processes. A cross-comparison matrix based on the PROD operator has marginal sums that agree with the per-class areas. However, non-null disagreement values can result from the perfect matching case. In fact, it does not satisfy the upper-bound and homogeneity properties of Table 3. The latter, however, could be fulfilled if the operator is properly normalized (see the MIN-PROD composite operator below). The use of this operator for the assessment of soft classifications has been demonstrated in (Lewis & Brown, 2001), and its counterintuitive characteristics have been illustrated in Pontius and Cheuk (2006).

A LEAST operator was recently incorporated in the discussion of sub-pixel comparison of maps (Pontius & Connors, 2006). The expression for the LEAST operator is given in Table 4. As demonstrated in the introductory example, this operator measures the *minimum possible sub-pixel overlap* between two classes. Even though this operator is meaningful for sub-pixel accuracy assessment, it may be of little use for other contexts, as it has even more counterintuitive characteristics than the PROD operator. Specifically, this operator fails to fulfill all but the commutativity and nullity properties from Table 3. As with the PROD operator, the homogeneity property could be met by a modified LEAST operator that relaxes the

sum-to-unit constraint (see the MIN-LEAST composite operator below).

2.3.3. Composite operators

None of the basic operators above satisfy the diagonalization characteristic discussed in Section 2.2. Indeed, in order for an operator to exhibit the diagonalization characteristic, it must conform to Eq. (1). This type of operator is referred to as *composite*. The formalism in Table 3 is then useful for guiding the selection of potential composite operators for general soft classifications. For example, the only operator from Table 4 that satisfies all the basic properties in Table 3 for an agreement measure is the MIN operator. The uniqueness of the MIN operator as an agreement measure is also evidenced in Eqs. (2) and (3), where over- and underestimation errors are given in terms of the MIN operator. Here, we consider only three composite operators that use the MIN operator as agreement measure. They are referred to as MIN-PROD, MIN-MIN and MIN-LEAST, respectively. These are defined in Table 5.

The MIN-PROD composite operator was recently proposed by Pontius and Cheuk (2006). It uses the minimum operator for the diagonal cells and a normalized product operator for the off-diagonal cells, thus combining the fuzzy set view with the probabilistic view. Expressions for the agreement and disagreement from this composite operator are presented in Table 5. This operator satisfies the basic properties of Table 3. In addition, the MIN-PROD matrix satisfies the diagonalization and marginal sums characteristics. The interpretation of the composite operator in the context of sub-pixel agreement–disagreement is aligned with an assumption of maximum overlap between corresponding categories (diagonal cells), followed by the allocation of the residual sub-pixel fractions onto the other categories (off-diagonal cells). The disagreement measure corresponds to the expected overlap by chance constrained to the unmatched sub-pixel fraction. Specifically, the disagreement between two membership values, s_{nk} and r_{nl} , corresponds to the joint probability that a randomly drawn point within the space spanned by the unmatched fraction, $1 - \sum_i \min(s_{ni}, r_{ni})$, of pixel n , belongs to classes k and l in the residual class fractions $s_{nk} - \min(s_{nk}, r_{nk})$ and $r_{nl} - \min(s_{nl}, r_{nl})$ of the assessed and reference pixels, respectively.

The MIN-MIN composite operator uses the minimum operator for both agreement and disagreement. However, it differs from the MIN operator in that it assigns agreement in a

Table 5
Three composite operators

Operator ID	Agreement ^a	Disagreement ^b ($k \neq l$)	Sub-pixel confusion
MIN-PROD	$\min(s_{nk}, r_{nk})$	$s'_{nk} \times r'_{nl} / \sum_i r'_{ni}$	Constrained expected
MIN-MIN	$\min(s_{nk}, r_{nk})$	$\min(s'_{nk}, r'_{nl})$	Constrained maximum
MIN-LEAST	$\min(s_{nk}, r_{nk})$	$\max(s'_{nk} + r'_{nl} - \sum_i r'_{ni}, 0)$	Constrained minimum

^a s_{nk} and r_{nk} denote the assessed and reference grades for class K at pixel n .

^b s'_{ni} and r'_{ni} denote the over- and underestimation errors of class i at pixel n .

first step and then, in a second step, it computes the disagreement based on the over- and underestimation errors. Table 5 shows the expressions for agreement and disagreement from this composite operator. The MIN-MIN composite operator satisfies all the properties outlined in Table 3. In addition, it leads to a cross-comparison matrix that satisfies the diagonalization property. However, it does not warrant the marginal sum characteristic. Marginal totals of a MIN-MIN matrix will, generally, overestimate the class proportions from the reference and assessed datasets because the MIN operator, used for computing the off-diagonals cells, accounts for the maximum possible overlapping area among the residual fractions at sub-pixel level. In this sense, the disagreement measure from the MIN-MIN operator provides an upper bound for the possible sub-pixel overlap constrained to the unmatched sub-pixel fraction.

The MIN-LEAST composite operator uses the MIN operator for the diagonal cells and a re-normalized LEAST operator for the off-diagonal cells. Table 5 shows the expressions for agreement and disagreement from this composite operator. While the agreement measure satisfies all the properties of Table 3, the disagreement measure does not satisfy the required positivity property. The MIN-LEAST operator produces a diagonal matrix for a perfect matching case. However, sub-pixel areas from the reference and assessed datasets are generally underestimated by marginal totals. This is because the disagreement measure corresponds to the minimum possible sub-pixel overlap constrained to the unmatched sub-pixel fraction. Specifically, the re-normalized LEAST operator determines the excess of the sum of two residual class fractions, $s_{nk} - \min(s_{nk}, r_{nk})$ and $r_{ni} - \min(s_{ni}, r_{ni})$, respect to the unmatched pixel fraction, $1 - \sum_i \min(s_{ni}, r_{ni})$.

3. The sub-pixel confusion–uncertainty matrix

3.1. Sub-pixel confusion intervals

The preceding review of potential cross-comparison matrices for assessing sub-pixel classifications has shown that: 1) a composite operator is necessary to warrant the diagonalization characteristic, and 2) the MIN operator is the most appropriate candidate for agreement measure. It is worth noting that the use of a MIN operator for allocating sub-pixel proportions along the diagonal cells accounts only for the agreement at pixel level, i.e., the possible spatial distribution of classes within the pixel is not taken into account, but only the sub-pixel area proportions are matched. In contrast, the sub-pixel disagreement can take into account the possible spatial distribution of classes within the pixel. Nevertheless, there is no unique way to exactly allocate the remaining sub-pixel proportion into the off-diagonal cells. Specifically, the sub-pixel area allocation problem remains underspecified, as there are more unknowns ($K^2 - K$ off-diagonals elements) than equations ($2K$ constraints from column and row totals). Some exceptions occur, for example, if $K=2$ or $K=3$. One possibility is to use the statistical center of possible confusions, as given by the MIN-PROD composite operator. However, the sub-pixel distribution

uncertainty could not be accounted in this way. An alternate solution is proposed here that uses the *confusion intervals*, $[P_{kl}^{\text{MIN-LEAST}}, P_{kl}^{\text{MIN-MIN}}]$, formed by the MIN-LEAST and MIN-MIN operators. These intervals express the possible confusions among the classes. As demonstrated below, if there is a unique solution to the area allocation problem, then these intervals are tight (i.e., the lower and upper bounds of each confusion interval have the same value). In this case, the three composite operators of Table 5 lead essentially to the same confusion matrix.

3.1.1. A simple example

In order to illustrate the use of the confusion intervals, consider a reference pixel belonging to classes 1, 2, 3, and 4 with membership values $r_1=0.4, r_2=0.3, r_3=0.2$ and $r_4=0.1$, respectively. Consider also the following three cases of assessed pixels:

- (a) *Perfect matching*: $s_1=0.4, s_2=0.3, s_3=0.2, s_4=0.1$
- (b) *Non-perfect matching 1*: $s_1=0.3, s_2=0.2, s_3=0.4, s_4=0.1$
- (c) *Non-perfect matching 2*: $s_1=0.3, s_2=0.1, s_3=0.1, s_4=0.2$

The fundamental difference between the cases (b) and (c) is in the number of overestimated classes (one and two overestimated classes, respectively). The sub-pixel confusion intervals for cases (a)–(c) are shown in Table 6(a)–(c), respectively. Notice that the sub-pixel area allocation problem is uniquely determined for cases (a) and (b), as the maximum and minimum confusions are the same. However, it cannot be

Table 6
Sub-pixel confusion interval matrices for three cases: perfect matching (a), underestimation with unique solution (b), and underestimation without unique solution (c)

Class	Reference				CPA
	1	2	3	4	
<i>(a)</i>					
1	[0.4,0.4]	[0,0]	[0,0]	[0,0]	0.4
2	[0,0]	[0.3,0.3]	[0,0]	[0,0]	0.3
3	[0,0]	[0,0]	[0.2,0.2]	[0,0]	0.2
4	[0,0]	[0,0]	[0,0]	[0.1,0.1]	0.1
CPR	0.4	0.3	0.2	0.1	1.0
<i>(b)</i>					
1	[0.3,0.3]	[0,0]	[0,0]	[0,0]	0.3
2	[0,0]	[0.2,0.2]	[0,0]	[0,0]	0.2
3	[0.1,0.1]	[0.1,0.1]	[0.2,0.2]	[0,0]	0.4
4	[0,0]	[0,0]	[0,0]	[0.1,0.1]	0.1
CPR	0.4	0.3	0.2	0.1	1.0
<i>(c)</i>					
1	[0.3,0.3]	[0,0]	[0,0]	[0,0]	0.3
2	[0,0]	[0.1,0.1]	[0,0]	[0,0]	0.1
3	[0,0.1]	[0.1,0.2]	[0.2,0.2]	[0,0]	0.4
4	[0,0.1]	[0,0.1]	[0,0]	[0.1,0.1]	0.2
CPR	0.4	0.3	0.2	0.1	1.0

CPR and CPA are class proportion from reference and assessed pixels, respectively.

Table 7
MIN-PROD matrix for case (c) of Table 6

Class	Reference				Row total
	1	2	3	4	
1	0.30	0.00	0.00	0.00	0.30
2	0.00	0.10	0.00	0.00	0.10
3	0.07	0.13	0.20	0.00	0.40
4	0.03	0.07	0.00	0.10	0.20
Col. tot.	0.40	0.3	0.2	0.10	1.00

Values are rounded to two decimals.

uniquely resolved for case (c), where the minimum and maximum confusions of class 1 and class 2 with class 3 are not the same. What this matrix says is, for instance, that class 2 is confounded with class 3 in at least 0.1 ([0.1,0.2]), whereas class 1 might not be confounded at all with class 3 ([0,0.1]). Table 7 shows the MIN-PROD matrix for case (c). As pointed out before, the confusion values provided by the MIN-PROD operator must be considered as the expected confusion by chance, i.e., it represents a statistical center for all possible confusions; whereas the confusion interval defines the uncertainty associated to the sub-pixel confusion.

3.1.2. Tight confusion intervals

Since a confusion interval involves the notion of uncertainty on the confusion, it is natural to inquire under which circumstances the confusion intervals would be tight for an arbitrary number of classes. Here it is shown that if no more than one class is either over- or underestimated at each single pixel, then the sub-pixel area allocation problem can be uniquely resolved. Not surprisingly, this unique-solution scenario includes any of the following cases: 1) there is a perfect match, 2) there are no more than three classes, 3) at least one of the datasets is crisp.

Consider the n -th reference and assessed pixels with membership values r_{nk} and s_{nk} , respectively, of belonging to class k , for $k=1,\dots,K$, where $K>1$. If no class is underestimated nor overestimated, then there is a perfect match, and the proof for tight intervals is straightforward. The analysis when only one class is underestimated follows. Assume underestimation for class i at pixel n , so that $s_{ni}<r_{ni}$ and $s_{nk}\geq r_{nk}$ for $k\neq i$. The contributions to the diagonal elements, from both the MIN-LEAST and MIN-MIN operators, will be r_{nk} , at row $k\neq i$, and s_{ni} , at row i . Then, the contribution to the confusion intervals for

Table 8
Contribution of the n -th pixel to the upper and lower bounds of the confusion intervals for K classes, when only class i is underestimated

Class	1	2	...	i	...	$K-1$	K	Row total
1	r_{n1}	0	...	$s_{n1}-r_{n1}$...	0	0	s_{n1}
2	0	r_{n2}	...	$s_{n2}-r_{n2}$...	0	0	s_{n2}
...
i	0	0	...	s_{ni}	...	0	0	s_{ni}
...	0	...
$K-1$	0	0	...	$s_{nK-1}-r_{nK-1}$...	r_{nK-1}	0	s_{nK-1}
K	0	0	...	$s_{nK}-r_{nK}$...	0	r_{nK}	s_{nK}
Col. tot.	r_{n1}	r_{n2}	...	r_{ni}	...	r_{nK-1}	r_{nK}	1

Table 9
Sub-pixel confusion–uncertainty matrix, where marginal totals have been replaced by the class proportions from the reference (CPR) and assessed (CPA) datasets

Class	Reference				CPA
	Class 1	Class 2	...	Class K	
Class 1	P_{11}	$P_{12}\pm U_{12}$...	$P_{1K}\pm U_{1K}$	s_{+1}
Class 2	$P_{21}\pm U_{21}$	P_{22}	...	$P_{2K}\pm U_{2K}$	s_{+2}
...
Class K	$P_{K1}\pm U_{K1}$	$P_{K2}\pm U_{K2}$...	P_{KK}	s_{+K}
CPR	r_{+1}	r_{+2}	...	r_{+K}	N

columns $l\neq i$ becomes zero. Whereas the contributions to the lower and upper bounds of the confusion interval, at column i row $k\neq i$, become $\max(s_{nk}-r_{nk},0)=s_{nk}-r_{nk}$ and $\min(s_{nk}-r_{nk},r_{ni}-s_{ni})=s_{nk}-r_{nk}$, respectively. The latter equality can be concluded from the sum-to-unity constraint. Table 8 shows the form of the contribution by the n -th pixel to both the upper and lower bounds of the matrix. A similar matrix can be obtained when only one class is overestimated, wherein only one row has non-null elements in the off-diagonal positions. Therefore, the confusion intervals are tight if at most one class is either overestimated or underestimated on a per-pixel basis.

3.2. Sub-pixel confusion–uncertainty matrix

In practice, it is convenient to express each confusion interval in the form $P_{kl}\pm U_{kl}$, where P_{kl} and U_{kl} are the interval center and the interval half-width, respectively. These are computed as indicated by Eqs. (5) and (6), respectively.

$$P_{kl} = \frac{P_{kl}^{\text{MIN-MIN}} + P_{kl}^{\text{MIN-LEAST}}}{2} \tag{5}$$

$$U_{kl} = \frac{P_{kl}^{\text{MIN-MIN}} - P_{kl}^{\text{MIN-LEAST}}}{2} \tag{6}$$

This notation is preferred, as it provides a center value and allows documenting explicitly its associated uncertainty,¹ which in turn is necessary for further error propagation analysis. By extension to our definitions, row marginal sum, column marginal sum, and total sum from uncertainty values are defined as $U_{k+}=\sum_l U_{kl}$, $U_{+l}=\sum_k U_{kl}$, $U_{++}=\sum_k \sum_l U_{kl}$, respectively.

Eq. (5) defines an operator that satisfies all the basic properties in Table 3. This operator leads to a matrix that exhibits the diagonalization characteristic. However, it does not warrant the marginal sum characteristic. A typical way to circumvent this inconvenience has been the use of the total membership grades from the reference and assessed datasets in place marginal totals (Binaghi et al., 1999). Examples of this kind of cross-tabulations are shown in Table 6(a)–(c) and, more formally, in Table 9. In this way, the accuracy indices of Table 2 (b) are readily generalized, where row and column totals are simply replaced by the corresponding total membership grade.

¹ The half-width of the confusion interval is termed the uncertainty, as it reflects the uncertain nature of the sub-pixel distribution of classes.

Table 10
General structure of the SCM (a) and derived sub-pixel accuracy–uncertainty indices (b)

(a)					
Class	Reference				Row total
	Class 1	Class 2	...	Class K	
Class 1	P_{11}	$P_{12} \pm U_{12}$...	$P_{1K} \pm U_{1K}$	$P_{1+} \pm U_{1+}$
Class 2	$P_{21} \pm U_{21}$	P_{22}	...	$P_{2K} \pm U_{2K}$	$P_{2+} \pm U_{2+}$
⋮	⋮	⋮	⋮	⋮	⋮
Class K	$P_{K1} \pm U_{K1}$	$P_{K2} \pm U_{K2}$...	P_{KK}	$P_{K+} \pm U_{K+}$
Col. tot.	$P_{+1} \pm U_{+1}$	$P_{+2} \pm U_{+2}$...	$P_{+K} \pm U_{+K}$	$P_{++} \pm U_{++}$

(b)		
Sub-pixel accuracy index	Center	Uncertainty
Overall accuracy, OA_s	$\frac{P_{++} \sum_k P_{kk}}{P_{++}^2 - U_{++}^2}$	$\frac{U_{++} \sum_k P_{kk}}{P_{++}^2 - U_{++}^2}$
k-th User Accuracy, $UA_s(k)$	$\frac{P_{kk} P_{k+}}{P_{k+}^2 - U_{k+}^2}$	$\frac{P_{kk} U_{k+}}{P_{k+}^2 - U_{k+}^2}$
k-th Producer Accuracy, $PA_s(k)$	$\frac{P_{kk} P_{+k}}{P_{+k}^2 - U_{+k}^2}$	$\frac{P_{+k} U_{+k}}{P_{+k}^2 - U_{+k}^2}$
Kappa coefficient, κ_s	$\frac{(P_0 - P_c)(1 - P_c) - (*U_0 + U_c)U_c}{(1 - P_c)^2 - U_c^2}$	$\frac{*(1 - P_c)U_c + (1 - P_c)U_0}{(1 - P_c)^2 - U_c^2}$

Row and column totals of the SCM are determined as sum of center values (P_{ki}) plus–minus sum of uncertainty values (U_{ki}). The observed proportions of agreement ($P_o \pm U_o$) correspond to the overall accuracy (OA_s), whereas the expected proportion of agreement ($P_e \pm U_e$) is given by Eqs. (7) and (8).

* = sign of $(1 - P_o - U_o)(1 - P_e - U_e)$.

Unfortunately, the accuracy indices so-derived cannot reflect the uncertainty of the confusion as they do not depend on the off-diagonal cells. Notice that diagonal cells in Table 9 does not provide U -values. Since the U -values reflect the sub-pixel distribution uncertainty, which is not considered for the agreement, these are zeros for the diagonal cells.

Another possibility, which is pursued here, is to consider column and row totals as intervals (Table 10(a)). These intervals can be used to derive intervals of accuracy indices that reflect the uncertain nature of classes sub-pixel distribution. Table 10(a) shows the general structure of the proposed sub-pixel confusion–uncertainty matrix (SCM).

3.3. Derived accuracy–uncertainty indices

Table 10(b) shows the expressions of derived accuracy–uncertainty indices based exclusively on values from Table 10 (a). These expressions represent generalizations from traditional single-valued accuracy indices to intervals, which are expressed in the form of a center value plus–minus maximum deviation (or uncertainty). For instance, the definition of the kappa coefficient interval (κ_s) is based on the definition of the traditional kappa coefficient of agreement (Cohen, 1960, see Table 2(b)), where the overall accuracy interval (OA_s) was considered as the observed proportion of agreement; this is specified in the form $P_o \pm U_o$. Likewise, the expected proportion of agreement was determined in terms of marginal totals and overall total from the SCM and specified in the form $P_e \pm U_e$. The explicit expressions for the expected proportion’s center value and uncertainty are given in Eqs. (7) and (8), respectively. The uncertainties from both the observed and expected

proportions of agreement are propagated onto the kappa coefficient, which results in an interval of kappa coefficients specified through a center value and its associated uncertainty, as given in Table 10(b).

$$P_e = \sum_k \frac{(P_{++}^2 + U_{++}^2)(P_{+k}P_{k+} + U_{+k}U_{k+}) - 2P_{++}U_{++}(U_{+k}P_{k+} + P_{+k}U_{k+})}{(P_{++}^2 - U_{++}^2)^2} \tag{7}$$

$$U_e = \sum_k \frac{2P_{++}U_{++}(P_{+k}P_{k+} + U_{+k}U_{k+}) - (P_{++}^2 + U_{++}^2)(U_{+k}P_{k+} + P_{+k}U_{k+})}{(P_{++}^2 - U_{++}^2)^2} \tag{8}$$

The new accuracy–uncertainty indices must be interpreted as approximations to the traditional ones. Such approximations are in the order of the associated uncertainties. As a matter of fact, in the absence of uncertainty, the expressions for their center values are similar to those of Table 2(b). Interestingly, in the limit when uncertainty tends to zero, the overall accuracy (OA_s) can be interpreted as a complimented distance between membership values, as stated by Eq. (9).²

$$OA_s = 1 - \frac{1}{2N} \sum_k \sum_n |r_{nk} - s_{nk}| \tag{9}$$

3.3.1. A simple example

The following example illustrates the utility of the accuracy–uncertainty indices. As before, consider a reference pixel belonging to classes 1, 2, 3, and 4 with membership values $r_1=0.4$, $r_2=0.3$, $r_3=0.2$ and $r_4=0.1$, respectively. This time, we want to compare the accuracy of the following two classified pixels:

- (a) No uncertainty: $s_1=0.2, s_2=0.3, s_3=0.4, s_4=0.1$
- (b) Uncertainty: $s_1=0.3, s_2=0.4, s_3=0.1, s_4=0.2$

Notice that the maximum classification error committed in case (b) is lower than in case (a). Therefore, one should expect higher accuracy for case (b) than for case (a). Furthermore, in case (a) the errors are concentrated in two classes (class 1 and class 3), whereas in case (b) the errors are evenly distributed among the four classes. Since the source of errors can be attributed to a larger number of sub-pixel confusions, one should expect the error uncertainty (and thus the accuracy uncertainty) to be higher for case (b) than for case (a). However, these observations could not be revealed through the traditional indices. Specifically, if the assessment is based on the MIN-PROD operator³, an overall accuracy of 80% is obtained for both case (a) and case (b). Indeed, the kappa values obtained for case (a) (0.7297) and case (b) (0.7222) are even counter-intuitive. In contrast, if the SCM is applied, an overall fuzzy accuracy of 80%±0% ($\kappa_s=0.7297 \pm 0$) results for case (a) and

² This expression can be derived by applying the identity $\min(a,b)=(a+b-|a-b|)/2$.

³ The same result is obtained if the assessment is based on Table 9.

83.33%±16.67% ($\kappa_s=0.7778\pm0.2222$) for case (b), thus agreeing with the reasoning above. In sum, the accuracy–uncertainty indices from the SCM are able to differentiate between sub-pixel classifications having distinct error distributions, even in the case they have the same overall accuracy. Evidently, the larger the uncertainty of an index is, the less useful the center value will be.

4. Application of the SCM to real datasets

The SCM was used for assessing two datasets derived from remote sensing techniques. The first example involves the multi-resolution assessment of an invasive species detection method by hyper-spectral remote sensing. The second example illustrates the use of the SCM and derived indices for assessing sub-pixel proportions of urban land use/land cover from fuzzy

classifications of landsat imagery. In both cases, we have used the MIN-PROD operator as a benchmark of the proposed SCM.

4.1. Invasive species detection data

The first study site is located in the central transect of the Rio Grande River near Presidio, Texas (Fig. 2-left). The Rio Grande River is the second longest river system in the United States and southern Texas’ major source of water. One of the region’s critical problems has been the invasion of water ways by the noxious species *Tamarix chinensis* Laur and *Tamarix Ramo-sissima* Ledeb, commonly referred to as tamarisk or saltcedar (Baum, 1967; Everitt & DeLoach, 1990; Hart et al., 2005). Saltcedar is a deciduous shrub (up to 9 m in height when mature) that has invaded riparian sites of the southwestern United States and northern Mexico. This invasive shrub brings

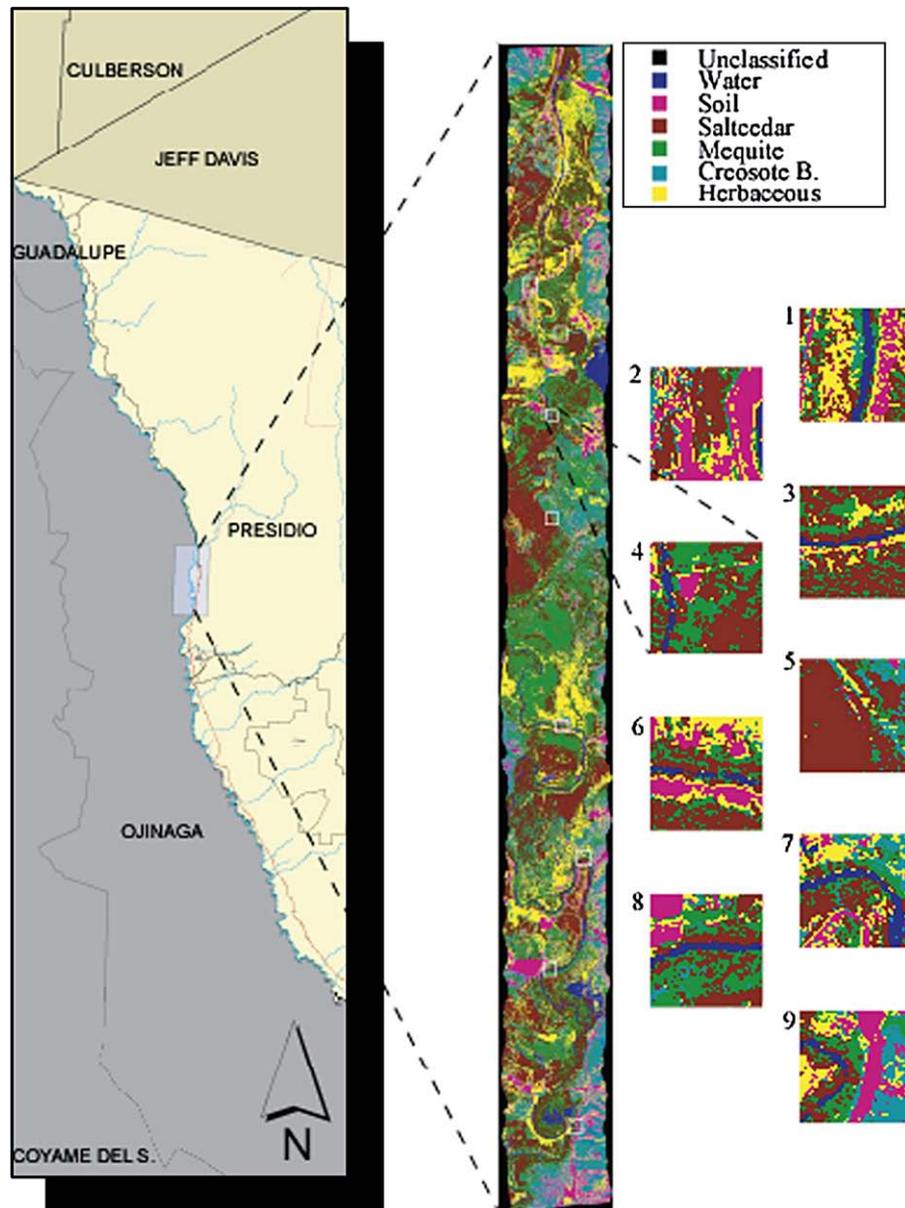


Fig. 2. Study site along the middle Rio Grande River (left), classification map (center) and reference plots used for the accuracy assessment (right).

several critical problems to invaded ecosystems. It grows faster than many native plants and can quickly form near monoculture, out-competing native plants, such as cottonwood (*Populus* spp.) and willow (*Salix* spp.) communities for sunlight, moisture, and nutrients.

An image of the study site was acquired by means of the airborne hyperspectral imaging system AISA in December, 2005. The acquisition date corresponds to the season when saltcedar's foliage turns a yellow-orange to orange-brown color before leaf drop, and thus, can be more easily distinguished from native species such as honey mesquite and false willow (Everitt & DeLoach, 1990). The AISA data used here consists of 61 bands with spectral bandwidth of around 10 nm, wavelength centers evenly distributed from 401 to 981 nm, and spatial resolution of 1 m. A feature selection strategy, based on linear discriminant analysis, was applied to reduce the total number of bands to seven. The reduced dataset was then classified through a supervised maximum likelihood classification that was trained with GPS points collected on the field for

up to 16 land cover types. The classified image was finally aggregated to 6 categories (Fig. 2-center), where the label of the most representative category was kept.

A reference dataset for the multiresolution assessment was generated through manual delineation of up to 16 land cover types over nine plots of 128-by-128 m (Fig. 2-right). The delineation of plot 4 was supported with intensive field work, whereas the delineation of other plots was based mainly on visual interpretation of true-color and visible-infrared composite displays. Nevertheless, the uncertainty introduced by the visual interpretation was assumed negligible given the high spatial and spectral quality of the image employed. Class merging was also carried out in order to match the reference data to the classification system of the assessed image.

4.2. Urban land use-land cover data

A second study site was selected in the eastern metropolitan area of Houston, Texas (Fig. 3), covering an area of 1200 km².

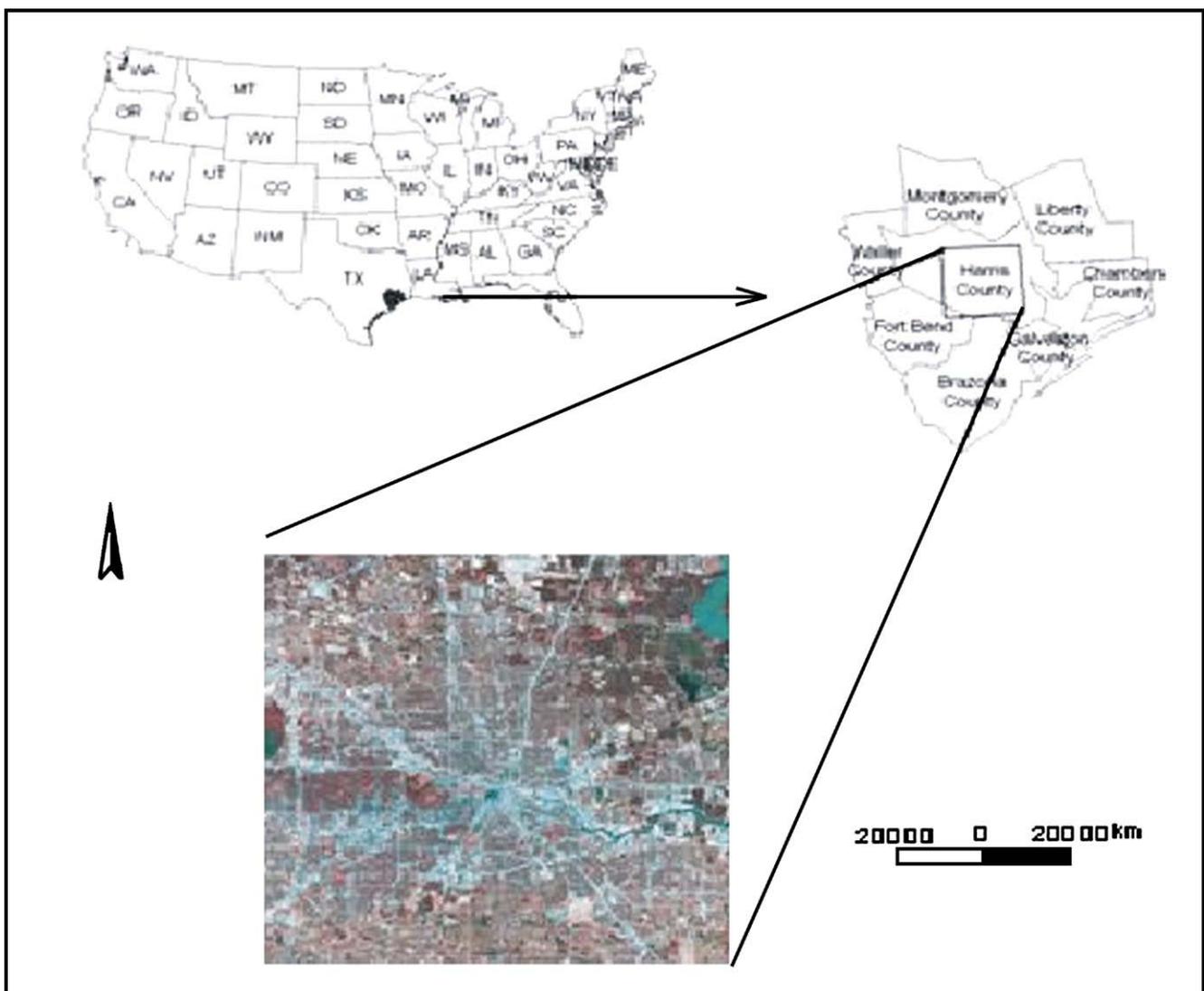


Fig. 3. Study site in Houston, Texas and Landsat ETM+ imagery acquired on January 2, 2003.

The typical land use/land cover classes in this study site include: residential area, commercial/industrial area, transportation, woodland, grassland, and barren/soil. A subset image from Landsat 7 ETM+ (path 25, row 39) acquired on January 2, 2003 was employed in this study. A bundled IKONOS image, comprising a 1-meter panchromatic and a 4-meter multispectral image, that were acquired on January 2, 2002, were adopted as references for choosing the training samples for the different classes as well as the test samples of known fractions from Landsat ETM+. In order to reduce errors due to spatial

misalignment, the original Landsat imagery was geo-referenced using the IKONOS geometry, which had better spatial accuracy. A registration accuracy under half pixel (as given by the RMSE) was attained for the Landsat dataset.

The fuzzy supervised classification was derived from the traditional supervised classification by softening the output of conventional hard classifiers (Richards & Jia, 1999). The training and test samples for the six classes were selected from the Landsat ETM+ image by referring to the IKONOS image with the help of field checks. Each class has thirty training plots

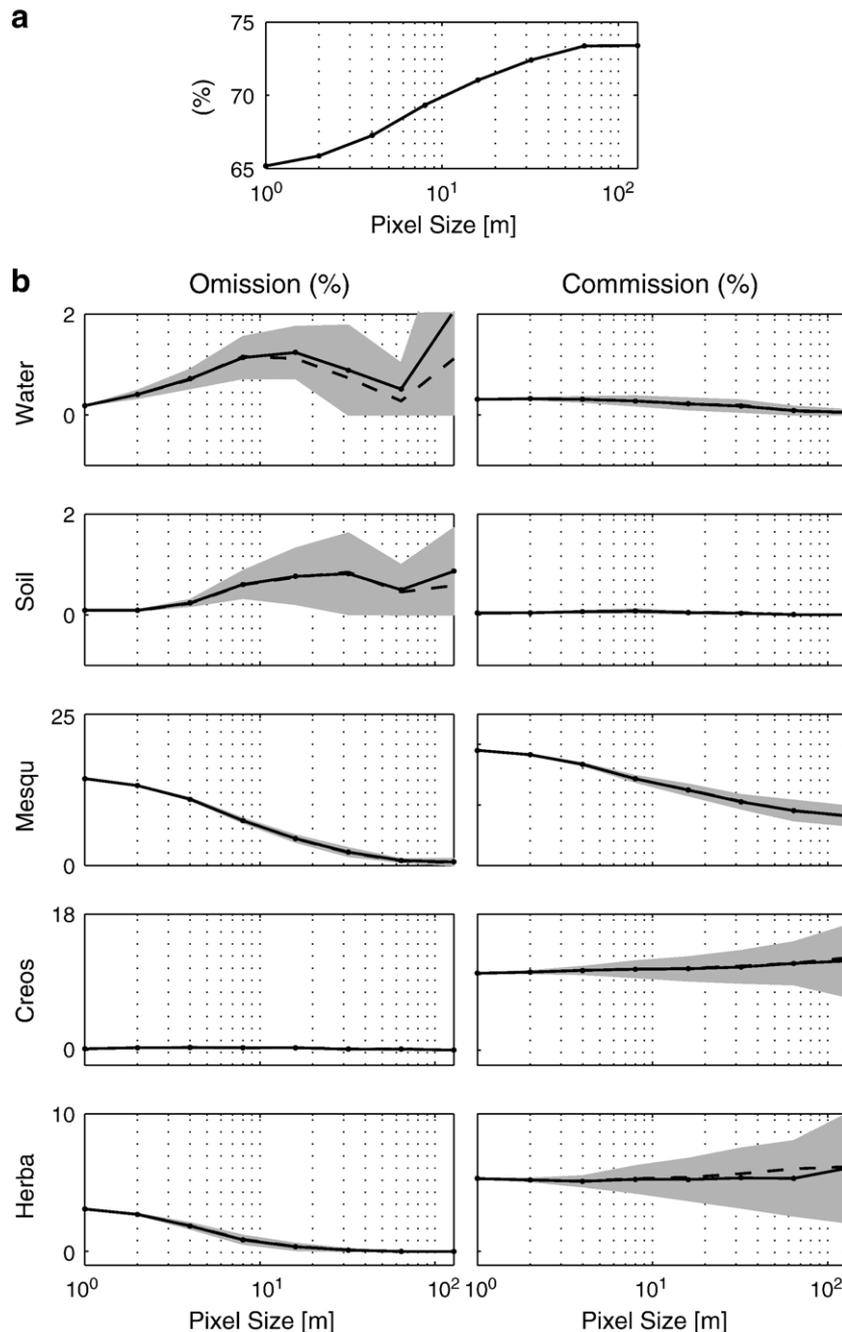


Fig. 4. Multi-resolution agreement (a) and disagreement (b) plots for Saltcedar class. Shaded area represents uncertainty, the solid line represents interval center and dashed line represents the MIN-PROD confusion.

and every training plot covers $90 \times 90 \text{ m}^2$. This sampling unit is comprised by a 3-by-3 pixel window of the Landsat image, and was chosen in place of individual pixels to reduce the influence of the positional accuracy of the pixel boundaries. Considering the statistical requirement and the size of our study area, around 200 random sample plots in $90 \times 90 \text{ m}^2$ were selected from the IKONOS image using ERDAS Imagine accuracy assessment module. For each sample unit, the corresponding *actual* land use classification was acquired through digitizing the IKONOS image. The actual fractions of each landscape class in the sample units were obtained through dividing the class area by the total area of one sample plot.

5. Results

5.1. Multi-resolution assessment of an invasive species detection method

The SCM was built from land-cover fractions within pixels of increasing size, where pixel sizes of 1, 2, 4, 8, 16, 32, 64, and 128 m were considered. We extracted the row and column that correspond to the Saltcedar category from each cross-comparison matrix. Plots of the agreement and disagreement values for Saltcedar class are shown in Fig. 4(a), and (b), where the pixel size is in logarithmic scale. The agreement oscillates from 65.2% at the finest resolution to 73.4% at the coarsest resolution. The disagreement is presented as per-class confusion of two types (Fig. 4(b)): Omission errors (off-diagonals on the column) and Commission errors (off-diagonals on the row). The omission curves correspond to the area percentage of Saltcedar that was classified as other categories, whereas the commission curves correspond to the area percentage of other categories that were classified as Saltcedar. The possible confusions defined by the confusion intervals across the resolutions are represented by the shaded areas, and the expected disagreement as given by the composite operator is also represented for comparison purposes.

Not surprisingly, the disagreement curves reveal that the most problematic categories for saltcedar detection are of vegetation type. The confusions of Saltcedar class with Water and Soil classes, as well as the confusions of Water and Soil classes with Saltcedar class, remain low across all the resolutions (below 2% in most cases). The most problematic category for saltcedar detection seems to be the labeled as Mesquite. This category includes honey mesquite, poverty weed, false willow and some other bushes and weeds. Both omission and commission errors are high (with highest values around 14% and 19%, respectively), yet with low relative uncertainty across the resolutions (under 1.4% and 3.3%, respectively). A drop in the confusion (especially of omission type) with the Mesquite category toward coarser resolutions implies that misclassifications are more related to location errors than to quantity errors (Pontius, 2002). These location errors could be due to mixed pixels of Saltcedar and Mesquite categories, perhaps due to the influence of under story vegetation. The confusion with Creosote Bush and Herbaceous categories is mostly of commission type, indicating a substantial

Table 11
Urban LULC classification assessment: SCM (a), MIN-PROD matrix (b), and per-class accuracy measures based on each matrix (c)

Class	Residential	Commercial/ Industrial	Transport	Other	Row total
<i>(a)</i>					
Res	13.78±0.00	0.63±0.63	4.83±0.63	1.73±0.00	20.98±1.27
C/I	0.03±0.03	5.54±0.00	1.87±0.03	0.22±0.00	7.66±0.06
Tra	0.04±0.04	0.74±0.04	2.20±0.00	0.00±0.00	2.98±0.08
Oth	13.70±0.07	11.91±0.67	37.24±0.66	61.05±0.00	123.90±1.41
CT	27.55±0.14	18.83±1.35	46.14±1.33	63.00±0.00	155.52±2.82
<i>(b)</i>					
Res	13.78	0.52	4.94	1.73	20.98
Com	0.02	5.54	1.88	0.22	7.66
C/I	0.02	0.77	2.20	0.00	2.98
Oth	13.73	12.00	37.12	61.05	123.90
CT	27.55	18.83	46.14	63.00	155.52
<i>(c)</i>					
	SCM		MIN-PROD		%RMSE
	Producer Accuracy	User Accuracy	Producer Accuracy	User Accuracy	
Res	50.0±0.3	65.9±4.0	50.0	65.7	16.5
Com	29.6±2.1	72.3±0.6	29.4	72.3	15.6
Tra	4.8±0.1	73.8±2.0	4.8	73.7	28.4
Oth	96.9±0.0	49.3±0.6	96.9	49.3	40.3
	OA _s =53.11±0.96%		OA=53.09%		
	κ _s =26.89±2.29%		κ=26.90%		

Matrix entries represent areas in hectares.

contribution to quantity errors. The reason for this commission error is that the spread of saltcedar reflectance during the dormant phenological stage tends to be higher than that of mesquite and false willow (Everitt & DeLoach, 1990), thus causing certain dry vegetation, such as johnson grass, creosote bush and marsh weed, to be mistaken as saltcedar. The increasing uncertainty toward coarser resolutions indicates that the overestimation of saltcedar’s fractions become equally explained by the underestimation of that category (i.e., Creosote Bush or Herbaceous) and the underestimation from remaining categories.

5.2. Assessment of an urban land use soft classification

Table 11(a) and Table 11(b) show the SCM and the MIN-PROD operator, respectively, where categories Woodland, Grassland and Barren/Soil have been grouped in a single (Other) category to save space. Proportions were multiplied by the total area of one sample plot (in hectares) and rounded to two decimals to reflect the actual area while facilitating readability. Table 11(c) presents the summary accuracy measures based on both the SCM and the MIN-PROD operator. For comparison purposes, the root mean square error (RMSE) was also computed on a per-class basis (Eq. (10)) and multiplied by 100%. These are also shown in Table 11(c).

$$RMSE_k = \sqrt{\frac{1}{N} \sum_n (r_{nk} - s_{nk})^2} \tag{10}$$

There is great similitude in the confusion values conveyed by both matrices. The SCM shows the associated sub-pixel confusion uncertainties, which are relatively low (the overall uncertainty of 2.82 ha represents just around 3.5 sample plots). The information conveyed by each uncertainty must be interpreted as sub-pixel land cover proportions. For example, of the total area observed for Commercial/Industrial land use type through IKONOS, around 0.63, 0.74 and 11.91 ha were misclassified as Residential, Transportation and Other land uses, respectively, when the fuzzy classification of Landsat was adopted. However, the maximum errors that can be committed with these estimations are 0.63, 0.04 and 0.67 ha, respectively. It is also noticeable the great similarity of user and producer accuracies determined from both matrices. The producer accuracy for category Other is both the highest and the least uncertain. This result seems to contradict the fact that this class also exhibits the highest RMSE. The issue is resolved if the user accuracy is also taken into account. Notice that the overall classification performance, as given by the SCM-based kappa (26.89%), is similar to that determined from the MIN-PROD operator (26.90%), yet the uncertainty associated with the former (2.29) reveals the relatively low uncertainty on the sources of error.

6. Discussion and conclusions

Determining land cover information accurately from remote sensing is crucial to understand several ecological and environmental processes occurring at a range of scales. Since the spatial pattern of land cover information can be smaller than the sensor footprint, soft classifications offer a flexible way to infer sub-pixel land cover information. However, the accuracy assessment of these representations has been recognized to be far more difficult than traditional crisp classifications (Foody, 2002). A great variety of measures derived from the traditional error matrix exists for describing the accuracy of crisp classifications (Congalton, 1991; Congalton & Green, 1999; Stehman & Czaplewski, 1998). These measures in turn have been adapted and used to assess soft classification (Binaghi et al., 1999; Latifovic & Olthof, 2004; Okeke & Karniele, 2006; Pontius & Cheuk, 2006). In the crisp case, most of these measures have a statistical interpretation, but the implications of their generalization to fuzzy domains is rarely discussed. Moreover, in spite of the substantial number of suggestions for assessing soft classifications reported in the literature (Binaghi et al., 1999; Foody, 1995; Gopal & Woodcock, 1994; Latifovic & Olthof, 2004; Lewis & Brown, 2001; Pontius & Cheuk, 2006; Pontius & Connors, 2006; Townsend, 2000), the utility and interpretation of the interclass confusion in the context of sub-pixel land cover information extraction had not been addressed explicitly.

In this paper, we have shown that the fuzzy confusion thicket can be unravelled when membership values correspond to land cover fractions, and the agreement and disagreement are defined in terms of the amount of sub-pixel overlap among the reference and assessed pixels. For the cross-comparison report to be useful for identifying a perfect match between the reference and assessed data, it was necessary to constrain the agreement measure at the pixel level. In the most general case, it was shown

that there is no analytical way to determine uniquely the actual confusion based solely on the information of land cover fractions. This problem was termed the sub-pixel area allocation problem. A recently introduced MIN-PROD composite operator (Pontius & Cheuk, 2006) seemed meaningful for assessing sub-pixel classifications, however, it provides one of (possibly) infinite number of solutions to the sub-pixel area allocation problem. That solution corresponds to the expected sub-pixel class overlap constrained to the unmatched sub-pixel fraction. Two new composite operators (MIN-LEAST and MIN-MIN) were introduced to provide the minimum and maximum possible sub-pixel class overlap constrained to the unmatched sub-pixel fraction. The intervals defined by these operators are arranged within a matrix, in the form of a center value plus-minus its uncertainty, termed the sub-pixel confusion-uncertainty matrix (SCM). We showed that all the confusion intervals are tight (i.e., no confusion uncertainty exists) when at most one class is either under- or overestimated at each pixel. Only in these circumstances, the SCM results in the MIN-PROD composite operator-based cross-comparison matrix. This is certainly the case when at least one of the compared sets is crisp, or when the number of classes is less than four. Therefore, uncertainty-free matrices can result often provided that there are many remote sensing methods for producing soft classifications in which only three classes may suffice for describing a wide variety of land cover characteristics (Carpenter et al., 1999; Milton, 1999; Roberts et al., 1993; Small, 2004). This result is also convenient for the common practice of assessing continental and global products through moderated resolution images. In this case, crisp classification from coarse resolution images are assessed using fractions derived from moderate resolution images (Latifovic & Olthof, 2004). It is also relevant for applications where reference data cannot be acquired, as in the case of historical data. In such cases, the hardened version of a fuzzy classification can be assessed using the fuzzy values (Okeke & Karniele, 2006). Indeed, the SCM results in the traditional confusion matrix if both datasets are crisp.

Traditional accuracy indices were also generalized to account for the sub-pixel distribution uncertainty. The practical use of the new cross-comparison matrix and the derived indices was demonstrated in assessing the invasive species detection through remote sensing procedures and the fuzzy classification of urban land use/land cover. In the former case, a multi-resolution approach was used to infer the error distribution within pixels of increasing sizes. The multi-resolution approach matches naturally the interpretation of membership values as sub-pixel land cover fractions, and is useful to describe how well the classes are clustered spatially (Pontius & Cheuk, 2006). Thus, it incorporates a spatial component to the traditional single-scale analysis made through the confusion matrix, and enables the potential to discern among distinct error distributions, an issue that has been risen as a major research challenge in the remote sensing field (Foody, 2002).

We believe that the SCM could be potentially useful in assessing soft classifications of a more general nature, but the analyst must be cautious when membership values cannot be interpreted as land cover fractions. Evidently, when applied to a more general case, the interpretation of the SCM entries in the

context of sub-pixel comparison are irrelevant. One should notice that the result for tight confusion intervals (Section 3.1.2) does not presuppose interpreting membership values as sub-pixel fractions, and its validity holds as long as membership grades satisfy the sum-to-one constraint. The user of the SCM should be also aware of the implications of co-registration inaccuracies. In the examples shown here, we have strived for reducing the effects of misalignment between the reference and assessed datasets; however, the impact of the positional accuracy on the confusion intervals remains an issue.

Given the remarkably higher interest on classification agreement, as compared to classification disagreement, we make a call for considering the confusion information, as well as the uncertainty information, provided by the SCM for improving soft classification products. While map accuracy is a central issue from a user perspective, map errors are the key for improving production methods. For instance, Carpenter et al. (1999) adopted the well-known root mean square error (RMSE), between predicted and actual cover proportions of vegetation types, as a measure of performance for three prediction methods tested. Based on the RMSE values alone, they conclude that the ARTMAP neural network was superior to maximum likelihood classifier and linear mixture models. Studies like this might benefit from the SCM by providing detailed information on the sub-pixel confusion and uncertainty.

Acknowledgements

The study was supported by the funding to Le Wang from US Department of Agriculture CSREES Award # 2004-38899-02181. The authors would like to thank Dr. Russel Congalton for his valuable comments on an earlier version of this paper. The urban LULC data was graciously shared graduate student Junmei Tang at Texas State University-San Marcos from her dissertation research. Thanks also to the three anonymous reviewers, and especially to the first reviewer, for the very helpful comments.

References

- Baum, B. R. (1967). Introduced and naturalized tamarisks in the United States and Canada. *Bayleya*, 15, 19–25.
- Binaghi, E., Brivio, P. A., Ghezzi, P., & Rampini, A. (1999). A fuzzy set-based assessment of soft classification. *Pattern Recognition Letters*, 20, 935–948.
- Carpenter, G. A., Gopal, S., Martens, S., & Woodcock, C. E. (1999). A neural network method for mixture estimation for vegetation mapping. *Remote Sensing of Environment*, 70, 138–152.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37, 35–46.
- Congalton, R. G., & Green, K. (1999). *Assessing the accuracy of remotely sensed data: Principles and practices*. Boca Raton, FL: Lewis.
- Cross, A. M., Settle, J. J., Drake, N. A., & Paivinen, R. T. M. (1991). Subpixel measures of tropical forest cover using AVHRR data. *International Journal of Remote Sensing*, 12, 1119–1129.
- Everitt, J. H., & DeLoach, C. J. (1990). Remote sensing of chinese tamarisk (*Tamarix chinensis*). *Weed Science*, 38, 273–278.
- Fisher, P., & Pathirana, S. (1990). An evaluation of fuzzy memberships of land cover classes in the suburban zone. *Remote Sensing of Environment*, 34, 121–132.
- Foody, G. M. (1995). Cross-entropy for the evaluation of the accuracy of a fuzzy land cover classification with fuzzy ground data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 80, 185–201.
- Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80, 185–201.
- Gopal, S., & Woodcock, C. E. (1994). Theory and methods for accuracy assessment of thematic maps using fuzzy sets. *Photogrammetric Engineering and Remote Sensing*, 60(2), 181–188.
- Green, K., & Congalton, G. (2004). An error matrix approach to fuzzy accuracy assessment: The NIMA geocover project. In R. S. Lunetta & J.G. Lyon (Eds.), *Remote sensing and GIS accuracy assessment* (pp. 163–172). Boca Raton: CRC Press.
- Gutman, G., & Ignatov, A. (1998). The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. *International Journal of Remote Sensing*, 60(2), 181–188.
- Hart, C. R., White, L. D., McDonald, A., & Sheng, Z. (2005). Saltcedar control and water salvage on the Pecos river, Texas, 1999–2003. *Journal of Environmental Management*, 75, 399–409.
- Kuzera, K., & Pontius, R. G., Jr. (2004). Categorical coefficients for assessing soft-classified maps at multiple resolutions. *Proc. of the joint meeting of The 15th Annual Conference of the International Environmental Society and the 6th Annual Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. Portland ME, 28 June–1 July 2004, Portland Maine (St. Paul, Minnesota: Spatial Accuracy Symposium)*.
- Latifovic, R., & Olthof, I. (2004). Accuracy assessment using sub-pixel fractional error matrices of global land cover products derived from satellite data. *Remote Sensing of Environment*, 90, 153–165.
- Lewis, H. G., & Brown, M. (2001). A generalized confusion matrix for assessing area estimates from remotely sensed data. *International Journal of Remote Sensing*, 22, 3223–3235.
- Milton, E. J. (1999). Image endmembers and the scene model. *Canadian Journal of Remote Sensing*, 25(2), 112–120.
- Okeke, F., & Karnieli, A. (2006). Methods for fuzzy classification and accuracy assessment of historical aerial photographs for vegetation change analyses. Part I: Algorithm development. *International Journal of Remote Sensing*, 27(1–2), 153–176.
- Ozdogan, M., & Woodcock, C. E. (2006). Resolution dependent errors in remote sensing of cultivated areas. *Remote Sensing of Environment*, 103, 203–217.
- Pontius, R. G., Jr. (2002). Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. *Photogrammetric Engineering and Remote Sensing*, 68(10), 1041–1049.
- Pontius, R. G., Jr., & Cheuk, M. L. (2006). A generalized cross-tabulation matrix to compare soft-classified maps at multiple resolutions. *International Journal of Geographical Information Science*, 20(1), 1–30.
- Pontius, R. G., Jr., & Connors, J. (2006). Expanding the conceptual, mathematical and practical methods for map comparison. *Proc. of the Spatial Accuracy Meeting 2006. Lisbon, Portugal 16 pp.*, (available from www.clarku.edu/~rpontius).
- Richards, J. A., & Jia, X. (1999). *Remote sensing digital image analysis*, 3rd edition Berlin: Springer-Verlag.
- Roberts, D. A., Smith, M. O., & Adams, J. B. (1993). Green vegetation, non-photosynthetic vegetation and soils in AVIRIS data. *Remote Sensing of Environment*, 44, 255–269.
- Shabanov, N. V., Lo, K., Gopal, S., & Myneni, R. B. (2005). Subpixel burn detection in Moderate Resolution Imaging Spectroradiometer 500-m data with ARTMAP neural networks. *Journal of Geophysical Research*, 110, 1–17.
- Small, C. (2004). The landsat ETM+ spectral mixing space. *Remote Sensing of Environment*, 93, 1–17.
- Stehman, S. V., & Czaplewski, R. L. (1998). Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sensing of Environment*, 64, 331–344.
- Townsend, P. A. (2000). A quantitative fuzzy approach to assess mapped vegetation classifications for ecological applications. *Remote Sensing of Environment*, 72, 253–267.
- Woodcock, C. E., & Gopal, S. (2000). Fuzzy set theory and thematic maps: Accuracy assessment and area estimation. *International Journal of Geographical Information Science*, 14(2), 153–172.