University of Groningen

Relative quantity and allocation disagreement measures for category-level accuracy assessment

Warrens, Matthijs J.

Published in:
International Journal of Remote Sensing

DOI:
[10.1080/01431161.2015.1110265](https://doi.org/10.1080/01431161.2015.1110265)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Warrens, M. J. (2015). Relative quantity and allocation disagreement measures for category-level accuracy assessment. *International Journal of Remote Sensing*, 36(23), 5959-5969.
<https://doi.org/10.1080/01431161.2015.1110265>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Relative quantity and allocation disagreement measures for category-level accuracy assessment

Matthijs J. Warrens

To cite this article: Matthijs J. Warrens (2015) Relative quantity and allocation disagreement measures for category-level accuracy assessment, International Journal of Remote Sensing, 36:23, 5959-5969, DOI: [10.1080/01431161.2015.1110265](https://doi.org/10.1080/01431161.2015.1110265)

To link to this article: <https://doi.org/10.1080/01431161.2015.1110265>

Published online: 23 Nov 2015.

Submit your article to this journal [↗](#)

Article views: 177

View related articles [↗](#)

View Crossmark data [↗](#)

Citing articles: 4 View citing articles [↗](#)

Relative quantity and allocation disagreement measures for category-level accuracy assessment

Matthijs J. Warrens*

Institute of Psychology, Methodology and Statistics, Leiden University, Leiden, Netherlands

(Received 17 June 2015; accepted 6 October 2015)

Accuracy assessment of classified imagery is an important task in remote sensing. A commonly used approach for assessing classification accuracy is based on the error matrix. The error matrix can be summarized by overall and category accuracy measures, e.g. the overall agreement and the user's and producer's accuracies. Two relatively new measures are the so-called quantity disagreement and allocation disagreement. In this article, relative versions of the category-level variants of these measures are defined. It is also shown how the new category measures are related to the overall quantity and allocation disagreement measures.

1. Introduction

Thematic land-cover maps are used in various applications, e.g. to describe the pattern and spatial distribution of land cover, and as the input of various environmental models (Liu, Frazier, and Kumar 2007). An important step in the construction of land-cover maps is the classification of remotely sensed data (Foody 2002). The usefulness of a map depends on the accuracy of the overall map and how the individual categories of the map have been classified. A popular approach for assessing the accuracy of a map is based on the error matrix (Foody 2002). An error matrix is a square cross-tabulation with numbers which express the proportion of sample units (e.g. pixels) assigned to a particular category relative to the actual category verified on the ground (Congalton 1991; Liu, Frazier, and Kumar 2007). The classification generated from the remotely sensed data is usually represented by the rows of the error matrix, while the columns indicate the reference data. The error matrix can be considered a population table if the image classifications and reference are available for all areas on the map (Stehman 1997).

The overall agreement, which is the proportion of units classified correctly, is commonly used for summarizing the overall accuracy in an error matrix (Stehman 1997; Liu, Frazier, and Kumar 2007). Furthermore, the user's and producer's accuracy are commonly used for summarizing the accuracy of a particular category. Pontius and Millones (2011) presented two new measures for accuracy assessment, called quantity disagreement and allocation disagreement. These two measures decompose the overall disagreement, which is the complement of the overall agreement, and provide additional information on the accuracy of the classification. The new measures have been applied in various research studies (see, e.g. Aldwaik, Onsted, and Pontius 2015; Hladik, Schalles, and Alber 2013; Müllerová, Pergl, and Pysek 2013; Navas, Telfer, and Ross 2012; Pandey, Joshi, and Seto 2013; Sohl et al. 2012).

*Email: warrens@fsw.leidenuniv.nl

If the overall disagreement is substantial it is useful to assess the disagreement on individual categories. Pontius and Millones (2011) proposed quantity and allocation disagreement measures for individual categories. Unlike user's and producer's accuracy, the category disagreement measures proposed by Pontius and Millones are absolute measures, because the measures do not take the totals, i.e. abundances, of the categories into account. Typically, absolute measures will have higher values if a category is more abundant. This makes their values difficult to compare across categories. Therefore, it would be useful to have relative category measures for quantity and allocation disagreement. Such measures are defined in this article. The values of the new category measures can be interpreted relative to the total abundance of a category.

The article is organized as follows. First, notation and definitions of various accuracy measures are presented. Next, the disagreement measures proposed in Pontius and Millones (2011) are discussed and category measures for relative quantity and allocation disagreement are defined. It is shown how the new measures are different from the absolute measures, and how the new measures are related to the overall quantity and allocation disagreement. The article ends with an illustration and a discussion.

2. Agreement measures

In this section the notation is introduced and several agreement measures are defined. We adopt some of the notation used in Pontius and Millones (2011), Liu, Frazier, and Kumar (2007), and Warrens (2015).

Let c denote the number of categories. Furthermore, let p_{ij} denote the proportion of the population units that is category i according to the classification information and category j according to the reference information, where $i, j = 1, \dots, c$. The square table with elements p_{ij} is called an error matrix. Table 1 is a hypothetical error matrix with three categories. The quantities p_{i+} and p_{+i} will be used to denote, respectively, the row and column total of category i . For the category totals we have the identities

$$\sum_{i=1}^c p_{i+} = 1 \quad \text{and} \quad \sum_{i=1}^c p_{+i} = 1. \quad (1)$$

In this article it is shown that various overall, e.g. map-level, accuracy measures are weighted averages of certain category measures. Suppose there are c category measures, $m_1, m_2, \dots, m_c$. Furthermore, let each category measure m_i have an associated weight, $w_i > 0$. The weighted average of the category measures using the w_i s as weights is then given by

Table 1. Hypothetical error matrix with three categories.

Classification	Reference			Row total
	Category 1	Category 2	Category 3	
Category 1	0.40	0.00	0.10	0.50
Category 2	0.00	0.30	0.05	0.35
Category 3	0.00	0.05	0.10	0.15
Column total	0.40	0.35	0.25	1.00

$$\frac{\sum_{i=1}^c w_i m_i}{\sum_{i=1}^c w_i} = \frac{w_1 m_1 + w_2 m_2 + \cdots + w_c m_c}{w_1 + w_2 + \cdots + w_c}. \quad (2)$$

If all the weights are identical, i.e. $w_1 = w_2 = \cdots = w_c$, average (2) is the ordinary average (or arithmetic mean) of $m_1, m_2, \dots, m_c$. Similar to the ordinary average, a weighted average is an average value of the various m_i -values. In addition, a weighted average takes the importance of each category measure m_i into account, where the importance of m_i is reflected in the associated weight, w_i . In this study, the weights are the category totals p_{i+} and p_{+i} or their sum $p_{i+} + p_{+i}$. The averages thus give more weight to category measures with higher corresponding category totals.

User's and producer's accuracy (Story and Congalton 1986) are two measures that are commonly used for assessing the accuracy of individual categories (Liu, Frazier, and Kumar 2007; Richards 1996). For category i the user's accuracy is given by

$$v_i = \frac{p_{ii}}{p_{i+}}. \quad (3)$$

In measure (3) the accuracy of category i is assessed with respect to only the classification information. The information in the reference totals is not important here. The producer's accuracy is defined as

$$\pi_i = \frac{p_{ii}}{p_{+i}}. \quad (4)$$

In measure (4) the accuracy of category i is assessed with respect to the reference information. The information in the classification totals is not important here. The values of the user's and producer's accuracy are 0 if there is no agreement on category i ($p_{ii} = 0$). The user's accuracy is 1 if $p_{ii} = p_{i+}$, while the producer's accuracy is 1 if $p_{ii} = p_{+i}$. Furthermore, Hellden (1980) proposed the accuracy measure given by

$$h_i = \frac{2p_{ii}}{p_{i+} + p_{+i}}. \quad (5)$$

Measure (5) is called the Dice measure in ecological biology (Dice 1945). In measure (5) the accuracy of category i is assessed with respect to both the classification and the reference information. Both types of information are equally important in (5). If $h_i = 1$ there is full agreement on category i , while if $h_i = 0$ there is no agreement on category i .

Measures (3), (4), and (5) assess the accuracy of a particular category i . Researchers are usually also interested in the overall accuracy of the classification. The proportion of units classified correctly, or the overall agreement, is given by

$$C = \sum_{i=1}^c p_{ii}. \quad (6)$$

Measure (6) is commonly used as an overall agreement measure (Warrens 2010). If $C = 1$ there is full agreement between the reference and comparison maps and the classification is perfect. If $C = 0$ there is no agreement between the maps.

There are three other ways to interpret the overall agreement. As pointed out by Richards (1996) and Liu, Frazier, and Kumar (2007), if we take a weighted average of the user's accuracies, using the category totals p_{i+} as weights, we obtain the overall agreement:

$$\frac{\sum_{i=1}^c (p_{i+})v_i}{\sum_{i=1}^c p_{i+}} = \frac{\sum_{i=1}^c p_{ii}}{1} = C. \quad (7)$$

Equation (7) shows that the overall agreement can be interpreted as an average user's accuracy. Furthermore, the overall agreement is also equivalent to the weighted average of the producer's accuracies using the totals p_{+i} as weights:

$$\frac{\sum_{i=1}^c (p_{+i})\pi_i}{\sum_{i=1}^c p_{+i}} = \frac{\sum_{i=1}^c p_{ii}}{1} = C. \quad (8)$$

Hence, the overall agreement is an average producer's accuracy. Finally, the overall agreement is also a weighted average of the category measures proposed by Hellden (1980) if we use the sum of the totals, i.e. $p_{i+} + p_{+i}$, as weights:

$$\frac{\sum_{i=1}^c (p_{i+} + p_{+i})h_i}{\sum_{i=1}^c (p_{i+} + p_{+i})} = \frac{\sum_{i=1}^c 2p_{ii}}{2} = C. \quad (9)$$

The overall agreement can thus be interpreted as an average of Hellden's category measures.

3. Absolute disagreement measures

In the previous section, measures were considered that can be used for analysing the overall and category agreement between the reference and a comparison map. Pontius and Millones (2011) presented measures that can be used for analysing disagreement between maps. The authors showed that an analysis of the disagreement may reveal valuable information on the classification accuracy.

The overall disagreement is defined as the complement of the overall agreement:

$$D = 1 - C. \quad (10)$$

Pontius and Millones (2011) proposed two types of disagreement measures to assess category and overall accuracy. The quantity disagreement of category i is given by

$$q_i = |p_{i+} - p_{+i}|. \quad (11)$$

Measure (11) reflects the level of difference between the reference map and a comparison map that is due to the less than perfect match in the category totals p_{i+} and p_{+i} (Pontius and Millones 2011). Measure (11) has a value of 0 if the category totals p_{i+} and p_{+i} are equal. The overall quantity disagreement is given by

$$Q = \frac{\sum_{i=1}^c |p_{i+} - p_{+i}|}{2}. \quad (12)$$

The overall quantity disagreement is obtained by summing (11) over all categories and dividing the result by 2. Measure (12) reflects the overall difference between the reference map and a comparison map that is due to the less than perfect match in the category totals.

Next, the allocation disagreement of category i is given by

$$a_i = 2\min(p_{i+}, p_{+i}) - 2p_{ii}. \quad (13)$$

Measure (13) has a value of 0 if either the user's or producer's accuracy is 1. The total allocation disagreement is given by

$$A = \sum_{i=1}^c \min(p_{i+}, p_{+i}) - C. \quad (14)$$

Measure (14) can be obtained by summing measure (13) over all categories and dividing the result by 2. The overall allocation disagreement reflects the difference between the reference map and a comparison map that is due to the less than maximum match in the spatial allocation of the categories, given the category totals of the maps (Pontius and Millones 2011). Warrens (2015) showed how the values of (11) and (12), and the values of (13) and (14), are related.

Pontius and Millones (2011) showed that the overall disagreement can be decomposed into the quantity disagreement and the allocation disagreement:

$$D = Q + A. \quad (15)$$

If $p_{i+} = p_{+i}$ for all categories there is no quantity disagreement. In this case all disagreement is allocation disagreement, i.e. $D = A$. Furthermore, if $p_{ii} = \min(p_{i+}, p_{+i})$ for all categories there is no allocation disagreement. In this case all disagreement is quantity disagreement, i.e. $D = Q$.

4. Relative disagreement measures

Measures (11) and (13) are absolute accuracy measures. Typically, the measures will have higher values if a category is more abundant, which makes problematic comparison of the values of the absolute measures of different categories. In this section we formulate relative measures for assessing category disagreement. These new measures have similar properties to measures (3), (4), and (5).

4.1. Relative quantity disagreement

A measure of the relative quantity disagreement of category i can be defined as

$$\phi_i = \frac{q_i}{p_{i+} + p_{+i}} = \frac{|p_{i+} - p_{+i}|}{p_{i+} + p_{+i}}. \quad (16)$$

Measure (16) is a ratio of measure (11) to the sum of the category totals. Since the sum $p_{i+} + p_{+i}$ reflects the abundance of category i , measure (16) quantifies the same information of category i as measure (11), but now with respect to the abundance of

category i . The minimum value of measure (16) is 0, which is obtained if the category totals p_{i+} and p_{+i} are equal. Its maximum value is 1, which is obtained if p_{i+} or p_{+i} is equal to 0.

Category measure (16) is related to overall measure (12) in a similar way that measures (3), (4), and (5) are related to the overall agreement. If we take a weighted average of the relative quantity disagreements, using the sums $p_{i+} + p_{+i}$ as weights, we obtain the overall quantity disagreement:

$$\frac{\sum_{i=1}^c (p_{i+} + p_{+i})\phi_i}{\sum_{i=1}^c (p_{i+} + p_{+i})} = \frac{\sum_{i=1}^c |p_{i+} - p_{+i}|}{2} = Q. \quad (17)$$

The overall quantity disagreement is thus an average of relative quantity disagreement.

Next, we consider two examples to illustrate some differences between measures (11) and (16). Table 2 presents four hypothetical error matrices, labelled (a)–(d). Each matrix has three categories, but we will focus on the first two categories of each matrix. Consider matrix (a) first. For measure (11) we have $q_1 = 0.20$ and $q_2 = 0.10$, which shows that the absolute quantity disagreement on category 1 is twice as high as the absolute quantity disagreement on category 2. However, category 1 is more abundant, so a higher value is to be anticipated. For measure (16) we have $\phi_1 = \phi_2 = 0.33$. The relative quantity disagreement on categories 1 and 2 is thus equal. If we take the category totals into account, the quantity disagreement on category 1 is no higher than the quantity disagreement on category 2.

Next, consider matrix (b) in Table 2. For measure (11) we have $q_1 = q_2 = 0.10$. Thus, the absolute quantity disagreement on categories 1 and 2 is equal. However, the totals of category 1 are also higher. For measure (16) we have $\phi_1 = 0.20$ and $\phi_2 = 0.33$. Thus, if we take the category totals into account, the quantity disagreement on category 1 is less than that on category 2.

Measures (11) and (16) have a value of 0 if the category totals p_{i+} and p_{+i} are equal. It is unusual for measure (11) to produce high values, because high values require that

Table 2. Four hypothetical error matrices ((a), (b), (c), and (d)) with three categories.

(a)					(b)				
	Reference					Reference			
Classification	1	2	3	Total	Classification	1	2	3	Total
Category 1	0.20	0.00	0.20	0.40	Category 1	0.20	0.00	0.10	0.30
Category 2	0.00	0.10	0.10	0.20	Category 2	0.00	0.10	0.10	0.20
Category 3	0.00	0.00	0.40	0.40	Category 3	0.00	0.00	0.50	0.50
Total	0.20	0.10	0.70	1.00	Total	0.20	0.10	0.70	1.00

(c)					(d)				
	Reference					Reference			
Classification	1	2	3	Total	Classification	1	2	3	Total
Category 1	0.30	0.10	0.00	0.40	Category 1	0.15	0.10	0.05	0.30
Category 2	0.10	0.10	0.00	0.20	Category 2	0.10	0.10	0.00	0.20
Category 3	0.00	0.00	0.40	0.40	Category 3	0.05	0.00	0.45	0.50
Total	0.40	0.20	0.40	1.00	Total	0.30	0.20	0.50	1.00

the absolute difference between the category totals is substantial. For example, if $p_{i+} = 0.10$ and $p_{+i} = 0.90$ we have $q_i = 0.80$. However, it is unlikely that this will happen with real-life maps. Measure (16) can obtain high values. These occur if the difference between the category totals is large compared with their sum. For example, if $p_{i+} = 0.09$ and $p_{+i} = 0.01$ we have $\phi_i = 0.80$.

Finally, it should be noted that the value of measure (16) is exclusively determined by the ratio of the totals p_{i+} and p_{+i} . If we let $p_{i+}/p_{+i} = d_i$, where d_i is a positive number, measure (16) is given by

$$\phi_i = \frac{|1 - d_i|}{1 + d_i}. \quad (18)$$

For example, if one of the category totals is one and a half times larger than the other, i.e. $d_i = 1.50$ or $d_i = 0.66$, we have $\phi_i = 0.20$. Furthermore, if one of the totals is three times larger than the other, i.e. $d_i = 3$ or $d_i = 0.33$, we have $\phi_i = 0.50$. For example, if $p_{i+} = 0.03$ and $p_{+i} = 0.01$ we have $\phi_i = 0.50$, and if $p_{i+} = 0.05$ and $p_{+i} = 0.15$ we again have $\phi_i = 0.50$.

4.2. Relative allocation disagreement

A measure of the relative allocation disagreement of category i can be defined as

$$\alpha_i = \frac{a_i}{p_{i+} + p_{+i}} = \frac{2\min(p_{i+}, p_{+i}) - 2p_{ii}}{p_{i+} + p_{+i}}. \quad (19)$$

Measure (19) is a ratio of measure (13) to the sum of the category totals. Measure (19) quantifies the same information for category i as does measure (13), but measure (19) does this with respect to the abundance of category i . The minimum value of measure (19) is 0, which is obtained if $\min(p_{i+}, p_{+i}) = p_{ii}$. Furthermore, the maximum value is 1, which is obtained if $p_{ii} = 0$ and $p_{i+} = p_{+i}$. We anticipate the value of α_i to be low in general, i.e. closer to 0 than 1, because if p_{i+} and p_{+i} are high then p_{ii} is usually high as well.

Measure (19) is related to overall measure (14) in a way similar to measure (16) being related to overall measure (12). If we take a weighted average of the relative allocation disagreements, using the sums $p_{i+} + p_{+i}$ as weights, we obtain the overall allocation disagreement:

$$\frac{\sum_{i=1}^c (p_{i+} + p_{+i})\alpha_i}{\sum_{i=1}^c (p_{i+} + p_{+i})} = \frac{\sum_{i=1}^c (2\min(p_{i+}, p_{+i}) - 2p_{ii})}{2} = A. \quad (20)$$

The overall allocation disagreement is thus an average of relative allocation disagreement.

Next, we consider two examples to illustrate some differences between measures (13) and (19). First, consider the first two categories of matrix (c) of Table 2. For measure (13) we have $a_1 = a_2 = 0.20$, from which we conclude that the absolute allocation disagreement on the categories is identical. Since there is more agreement on category 1 than category 2, the disagreement on category 2 could be higher. For measure (19) we

have $\alpha_1 = 0.25$ and $\alpha_2 = 0.50$. Thus, if we take the category totals into account, the allocation disagreement on category 2 is twice as high as the disagreement on category 1.

Next, consider the first two categories of matrix (d) in Table 2. For measure (13) we have $a_1 = 0.30$ and $a_2 = 0.20$, which shows that absolute allocation disagreement on category 1 is one and a half times as high as the absolute allocation disagreement on category 2. For measure (19) we have $\alpha_1 = \alpha_2 = 0.50$. Thus, in terms of relative disagreement, the allocation disagreement on the categories is equal.

4.3. Relative disagreement

Definitions (16) and (19) can be used to define a category measure that combines relative quantity and allocation disagreement. Since quantity and allocation disagreement are two distinct types of disagreement, a (general) relative disagreement measure of category i is given by

$$\delta_i = \phi_i + \alpha_i. \quad (21)$$

Equation (21) shows that, similar to the overall disagreement in (15), the category disagreement can be decomposed into relative quantity and allocation disagreement. For the totals p_{i+} and p_{+i} we have the identity

$$p_{i+} + p_{+i} = 2\min(p_{i+}, p_{+i}) + |p_{i+} - p_{+i}|. \quad (22)$$

Identity (22) is used in Pontius and Millones (2011) to prove identity (15). Using equality (22), measure (21) is equivalent to

$$\delta_i = \frac{p_{i+} + p_{+i} - 2p_{ii}}{p_{i+} + p_{+i}} = 1 - \frac{2p_{ii}}{p_{i+} + p_{+i}} = 1 - h_i. \quad (23)$$

Thus, the disagreement measure defined in Equation (21) is the complement of Hellden's agreement measure presented in (5). If we take a weighted average of the relative disagreements, using the sums $p_{i+} + p_{+i}$ as weights, we obtain the overall disagreement:

$$\frac{\sum_{i=1}^c (p_{i+} + p_{+i})\delta_i}{\sum_{i=1}^c (p_{i+} + p_{+i})} = \frac{2 - 2 \sum_{i=1}^c p_{ii}}{2} = D. \quad (24)$$

The overall disagreement is thus an average category disagreement if disagreement is assessed with measure (21).

5. Discussion

In this article we present category measures for assessing relative quantity and allocation disagreement. The new measures can be used as complements to the quantity and allocation disagreement originally proposed by Pontius and Millones (2011). The latter measures are measures of absolute disagreement, and are usually easier to interpret. However, since absolute measures have typically higher values if a category is more abundant, their values are difficult to compare across categories. These new relative measures assess disagreement

with respect to the category totals. In certain cases this can be more informative than assessing absolute disagreement. However, since the new measures are ratios, the measures may also go awry. For example, when the value of a ratio is high it may not be clear whether the reason is a large numerator or a small denominator. A rare category is usually not important if one is interested in the major sources of disagreement. Nevertheless, the new measures may produce high values for rare categories.

It turns out that the overall quantity disagreement is an average of the category measures of relative quantity disagreement, and that the overall allocation disagreement is an average of the category measures of relative allocation disagreement. We may use the data in Table 1 to illustrate how the new measures complement the measures of quantity and allocation disagreement proposed by Pontius and Millones (2011). Table 1 may also be used to illustrate identities (9), (17), and (20) presented in this study. The values of various accuracy measures for the data in Table 1 are presented in Table 3.

The overall agreement in Table 1 is $C = 0.80$, which indicates there is high agreement between the maps. Hellden's measure indicates that the agreement is high on category 1 ($h_1 = 0.89$) and category 2 ($h_2 = 0.86$), but only moderate on category 3 ($h_3 = 0.50$). The value $C = 0.80$ is an average of the agreement on the categories ($h_1 = 0.89$, $h_2 = 0.86$, $h_3 = 0.50$). The overall disagreement between the maps is $D = 1 - C = 0.20$, which is quite substantial. The overall disagreement is equally divided over the quantity disagreement ($Q = 0.10$) and allocation disagreement ($A = 0.10$). An analysis of both types of category disagreement seems therefore useful (Pontius and Malizia 2004).

There is some quantity disagreement on category 1 ($q_1 = 0.10$) and 3 ($q_3 = 0.10$), but no quantity disagreement on category 2 ($q_2 = \phi_2 = 0.00$). Thus, the totals of categories 1 and 3 differ by 0.10, while the totals of category 2 are equal. A difference of 0.10 is more substantial for category 3 since this category is less abundant. This is reflected in the relative quantity disagreement. The relative quantity disagreement on category 1 is low ($\phi_1 = 0.11$), whereas it is moderate on category 3 ($\phi_3 = 0.25$). Furthermore, the overall quantity disagreement ($Q = 0.10$) is an average of the relative quantity disagreement on the categories ($\phi_1 = 0.11$, $\phi_2 = 0.00$, $\phi_3 = 0.25$).

Next, there is some allocation disagreement on categories 2 ($a_2 = 0.10$) and 3 ($a_3 = 0.10$), but no allocation disagreement on category 1 ($a_1 = \alpha_1 = 0.00$). However,

Table 3. Values of accuracy measures for the data in Table 1.

Overall measures	Symbol			
Overall agreement	C	0.80		
Overall disagreement	D	0.20		
Quantity disagreement	Q	0.10		
Allocation disagreement	A	0.10		
Category measures	Symbol	Category 1	Category 2	Category 3
Hellden's measure	h_i	0.89	0.86	0.50
Relative disagreement	δ_i	0.11	0.14	0.50
Quantity disagreement	q_i	0.10	0.00	0.10
Relative quantity disagreement	ϕ_i	0.11	0.00	0.25
Allocation disagreement	a_i	0.00	0.10	0.10
Relative allocation disagreement	α_i	0.00	0.14	0.25

category 2 is more abundant ($p_{2+} + p_{+2} = 0.70$) than category 3 ($p_{3+} + p_{+3} = 0.40$). If the category totals are taken into account, the allocation disagreement is moderate on category 3 ($\alpha_3 = 0.25$) but only low on category 2 ($\alpha_2 = 0.14$). Furthermore, the overall allocation disagreement ($A = 0.10$) is an average of the relative allocation disagreement on the categories ($\alpha_1 = 0.00$, $\alpha_2 = 0.14$, $\alpha_3 = 0.25$). Tables 1 and 3 show that the relative disagreement measures provide information different to the absolute disagreement measures.

The overall and category measures of quantity and allocation disagreement may also be useful tools in other fields of science. For example, in medical, social, and behavioural sciences researchers are frequently interested in assessing the intensity of a behaviour or a disease on a categorical scale. To assess the accuracy of the ratings, the classification task is usually performed by two observers. Standard tools for assessing agreement between observers are Cohen's kappa coefficient (Warrens 2010) and weighted kappa coefficient (Warrens 2013), but quantity disagreement and allocation disagreement can also be used in agreement studies.

Acknowledgement

The author thanks Professor Timothy Warner and an anonymous referee for their helpful comments and valuable suggestions on an earlier version of this article.

Disclosure statement

No potential conflict of interest was reported by the author.

References

- Aldwaik, S. Z., J. A. Onsted, and R. G. Pontius, Jr. 2015. "Behavior-Based Aggregation of Land Categories for Temporal Change Analysis." *International Journal of Applied Earth Observation and Geoinformation* 35: 229–238. doi:10.1016/j.jag.2014.09.007.
- Congalton, R. G. 1991. "A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data." *Remote Sensing of Environment* 37: 35–46. doi:10.1016/0034-4257(91)90048-B.
- Dice, L. R. 1945. "Measures of the Amount of Ecologic Association between Species." *Ecology* 26: 297–302. doi:10.2307/1932409.
- Foody, G. M. 2002. "Status of Land Cover Classification Accuracy Assessment." *Remote Sensing of Environment* 80: 185–201. doi:10.1016/S0034-4257(01)00295-4.
- Hellden, U. 1980. *A Test of Landsat-2 Imagery and Digital Data for Thematic Mapping Illustrated by an Environmental Study in Northern Kenya*. Report, Vol. 47. Lund: Lund University Natural Geography Institute.
- Hladik, C., J. Schalles, and M. Alber. 2013. "Salt Marsh Elevation and Habitat Mapping Using Hyperspectral and LIDAR Data." *Remote Sensing of Environment* 139: 318–330. doi:10.1016/j.rse.2013.08.003.
- Liu, C., P. Frazier, and L. Kumar. 2007. "Comparative Assessment of the Measures of Thematic Classification Accuracy." *Remote Sensing of Environment* 107: 606–616. doi:10.1016/j.rse.2006.10.010.
- Müllerová, J., J. Pergl, and P. Pyšek. 2013. "Remote Sensing as a Tool for Monitoring Plant Invasions: Testing the Effects of Data Resolution and Image Classification Approach on the Detection of a Model Plant Species *Heracleum Mantegazzianum* (Giant Hogweed)." *International Journal of Applied Earth Observation and Geoinformation* 25: 55–65. doi:10.1016/j.jag.2013.03.004.
- Navas, J. M., T. C. Telfer, and L. G. Ross. 2012. "Separability Indexes and Accuracy of Neuro-Fuzzy Classification in Geographic Information Systems for Assessment of Coastal

- Environmental Vulnerability.” *Ecological Informatics* 12: 43–49. doi:[10.1016/j.ecoinf.2012.06.006](https://doi.org/10.1016/j.ecoinf.2012.06.006).
- Pandey, B., P. K. Joshi, and K. C. Seto. 2013. “Monitoring Urbanization Dynamics in India Using DMSP/OLS Night Time Lights and SPOT-VGT Data.” *International Journal of Applied Earth Observation and Geoinformation* 23: 49–61. doi:[10.1016/j.jag.2012.11.005](https://doi.org/10.1016/j.jag.2012.11.005).
- Pontius, R. G., Jr., and N. R. Malizia. 2004. “Effect of Category Aggregation on Map Comparison.” In *GIScience*, Lecture Notes in Computer Science 3234, edited by M. J. Egenhofer, C. Freksa, and H. J. Miller, 251–268. Berlin: Springer-Verlag.
- Pontius, R. G. Jr., and M. Millones. 2011. “Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment.” *International Journal of Remote Sensing* 32: 4407–4429. doi:[10.1080/01431161.2011.552923](https://doi.org/10.1080/01431161.2011.552923).
- Richards, J. A. 1996. “Classifier Performance and Map Accuracy.” *Remote Sensing of Environment* 57: 161–166. doi:[10.1016/0034-4257\(96\)00038-7](https://doi.org/10.1016/0034-4257(96)00038-7).
- Sohl, T. L., B. M. Sleeter, K. L. Saylor, M. A. Bouchard, R. R. Reker, S. L. Bennett, R. R. Sleeter, R. L. Kanengieter, and Z. Zhue. 2012. “Spatially Explicit Land-Use and Land-Cover Scenarios for the Great Plains of the United States.” *Agriculture, Ecosystems & Environment* 153: 1–15. doi:[10.1016/j.agee.2012.02.019](https://doi.org/10.1016/j.agee.2012.02.019).
- Stehman, S. V. 1997. “Selecting and Interpreting Measures of Thematic Classification Accuracy.” *Remote Sensing of Environment* 62: 77–89. doi:[10.1016/S0034-4257\(97\)00083-7](https://doi.org/10.1016/S0034-4257(97)00083-7).
- Story, M., and R. G. Congalton. 1986. “Accuracy Assessment: A User’s Perspective.” *Photogrammetric Engineering and Remote Sensing* 52: 397399.
- Warrens, M. J. 2010. “Inequalities Between Kappa and Kappa-Like Statistics for $k \times k$ Tables.” *Psychometrika* 75: 176–185. doi:[10.1007/s11336-009-9138-8](https://doi.org/10.1007/s11336-009-9138-8).
- Warrens, M. J. 2013. “Conditional Inequalities between Cohen’s Kappa and Weighted Kappas.” *Statistical Methodology* 10: 14–22. doi:[10.1016/j.stamet.2012.05.004](https://doi.org/10.1016/j.stamet.2012.05.004).
- Warrens, M. J. 2015. ‘Properties of the Quantity Disagreement and the Allocation Disagreement.’ *International Journal of Remote Sensing* 36: 1439–1446. doi:[10.1080/01431161.2015.1011794](https://doi.org/10.1080/01431161.2015.1011794).