Revista Brasileira de Cartografia (2017), Nº 69/2: 361-373
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto
ISSN: 1808-0936

GEOGRAPHICAL SIMULTANEITY: A NEW INDEX TO VALIDATE RESULTS OBTAINED FROM DIGITAL IMAGE PROCESSING OF THEMATIC MAPS

Simultaneidade Geográfica: Um Novo Índice para Validar Resultados Obtidos de Processamento Digital de Imagens de Mapas Temáticos

Ardemirio de Barros Silva[1], Jocimara Souza Britto Lobão[1] & Edson Eyji Sano[2]

[1]Universidade Estadual de Feira de Santana - UEFS
Departamento de Ciências Exatas
Caixa Postal 44036-200 Feira de Santana, BA, Brasil
abarros@uefs.br, juci.lobao@gmail.com

[2]Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA
EMBRAPA Cerrados
BR-020 km 18 CEP: 73301-970 Planaltina, DF, Brasil
edson.sano@embrapa.br

Recebido em 22 de Abril, 2016/ Aceito em 28 de Dezembro, 2016
Received on April 22, 2016/ Accepted on December 28, 2016

ABSTRACT

New technologies have enabled the dissemination of maps. Maps are commonly produced either by sophisticated image processing techniques or by more conventional methods. Therefore, defining the limits of the use of maps by measuring their associated errors is essential. There are several sources of such errors, which can be related to temporal, positional and interpretation aspects, among others. Regardless of the source, maps must be validated. The main procedures to validate maps are the Kappa (KI) and the Tau (τ) indices. These indices often present inconsistencies due to the marginal variations. This study presents a new index, the Geographical Simultaneity (GS), which corrects these inconsistencies. Comparison of GS, KI and τ results involving both hypothetical data and data available in the literature demonstrated the effectiveness of the use of GS.

Keywords: Map Validation, Kappa Index, Errors, Tau.

RESUMO

As novas tecnologias possibilitaram a disseminação de mapas. Mapas podem ser produzidos seja por sofisticadas técnicas de processamento de imagens, seja por métodos convencionais. Definir os limites do uso de mapas através dos erros contidos neles é essencial. Existem numerosas fontes de erros os quais podem estar relacionados com data, posição e interpretação, entre outros. Independente da fonte dos mapas, eles precisam de validação. Os principais procedimentos para validação são o índice de Kappa (IK), a prevalência e índice Tau (τ). Muitas vezes estes índices apresentam inconsistências devidos a variações marginais. Este estudo apresenta um novo índice, a Simultaneidade Geográfica (SG) que corrige estas inconsistências. As comparações com SG, IK e τ envolveram tanto dados hipotéticos como dados reais disponíveis na literatura para comprovar o êxito do uso de SG.

Palavras chave: Validação de Mapas, Índice Kappa, Erros, Tau.

1. INTRODUCTION

Multimap analyses correspond to a set of cumulative and/or non-cumulative algebraic operations involving two or more maps to result in a final map that represents a synthesis. These operations generally correspond to a multivariate analysis. They take into account the relationships between spatial variables. The values of different variables must be obtained in the same geographic location, and these variables must be considered simultaneously and interdependently. According to Hubert & Arabie (1985), one of the challenges in geospatially based data analysis is the evaluation of the magnitude of spatial autocorrelation. When the results are expressed as maps, the challenge is to define the relationship between the measurements of association and the correlation between measurements of reference maps, i.e., maps produced based on data rather than on information and on interpreted data. Another situation can occur: mismatch between two thematic maps from the same geographical space with the same spatial resolution and describing the same variables; such a situation requires analysis of the similarities and the mismatches.

Currently, map production using the existing technological apparatus is becoming increasingly easier. Consequently, measuring the quality of map production is also becoming even more important. One of the most important sources of thematic maps are those derived from the classification of remotely sensed data. These maps are used to describe the spatial distribution and pattern of land covers, to determine the potential areas of mining exploration, to establish environmental models and temporal analysis and to understand the epidemiology of geographical patterns, among others. Errors among maps can result from many factors, which can or cannot be controlled, such as temporal and positional errors or misinterpretation. Regardless of the origin, the agreement error must be measured to define the limits of map applicability more appropriately.

Whenever the quality of a measurement is evaluated, it is very common to use the terms accuracy and precision. Accuracy is used to describe the closeness of a measurement to the true value, usually expressed in terms of the standard deviation or the variance of measurements. Accuracy is the degree of agreement between a measurement and the value set as a reference. The thematic accuracy is the degree of accuracy of the attributes generated from interpretations and from those found on the ground at the same geographical position and the same spatial resolution. Precision is the closeness of agreements among a set of results. One of the most effective methods to define the thematic accuracy is the one based on the analysis of the error matrix, also called the confusion matrix, the confusion table, or the contingency table. This matrix consists of a squared array that expresses the number of categories associated with the pixels arranged in rows and columns. Compared with the categories associated to the field, all pixels should have the same spatial resolution, from which we can derive estimates of the levels of agreement, which can be expressed in terms of individual categories or full agreement. Establishing the spatial distribution pattern of errors is essential to assess the quality of georeferenced databases.

Since 1960, a number of researchers have been involved in measuring map accuracies based on error matrices. Examples of such measurements include the Kappa index (KI) (COHEN, 1960; ROSENFIELD & FITZPATRICK-LINS, 1986); the Tau (τ) agreement measurements (KLECKA, 1980), which compare the results of classification with a random assignment of pixels to the classes; the overall accuracy (OA) (STORY & CONGALTON, 1986); the average accuracy from the producer's perspective (OAPP) (FUNG & LEDREW, 1988); the classification success index (CSI) (KOUKOULAS & BLACKBURN, 2001); the modified Kappa index (MKI) (AICKIN, 1990); the Kappa-like statistic alpha (KSA) (FOODY, 1992); the map-level normalized accuracy (MNA) (CONGALTON, 1991); the average mutual information (AMI) (FINN, 1993); the prevalence and bias adjusted Kappa (PABAK) (BYRT *et al.*, 1993); the normalized mutual information using the arithmetic mean of the entropies on the map and on the ground truth (NMI_1) (STREHL & GHOSH, 2002); and the normalized mutual information using the geometric mean of the entropies on the map and on the ground truth (NMI_2) (GHOSH *et al.*, 2002). Each procedure varies in specific ranges: 0 to 1 for OA, OAPP, MNA, NMI_1, NMI_2, KI, PABAK and τ; -1 to +1 for CSI, MKI and KSA; and 0 to $+\infty$ for AMI. According to Liu *et al.* (2007), these procedures can be classified into three groups: OAPP and CSI (first group); OA, KI, MKI, PABAK and τ (second group); and the others (third group).

The indices that have been widely used in the scientific community since the decades of 1970 to 1980 are the KI and τ. These indices take into account either the total number of categories or the total number of validations. These indices have been adapted to the necessity that became more prominent when geotechnologies emerged, bringing new demands. These indices often have inconsistencies due to changes in the marginal proportions. This study aims to present a new index, the Geographical Simultaneity (GS), which corrects these inconsistencies. The effectiveness of GS was demonstrated by an example that illustrates the results obtained by GS, KI and τ.

2. THEORY

A clear distinction is found in the relationship between the association of the measurements and the agreement of the measurements. To have agreement between two or more measurements, the measurements must correspond to identical categories, while two or more measurements will be associated when one of them can be predicted from the knowledge of the other categories. In other words, because agreement is a special case of association, we can have different alternatives, for example, low agreement and high association or high agreement and high association. The concepts of agreement and combination can be better clarified using the analysis of Table 1a, which indicates that the data have a perfect association and no agreement, while Tables 1b and 1c demonstrate that the data have different levels of association and agreement. These examples show, according to the terminology of Roy & Mitra (1956), marginal variations. The measurements may correspond to an unlimited number of observations exhibiting asymmetric marginal variations. This possible existence of asymmetry can create difficulties in the interpretation of the agreement, and many indices cannot clearly establish the existing asymmetries.

The comparison of the results obtained by the interpretations with the actual measurements is of fundamental importance in the analyses of products derived from remote sensing and geospatial data. Historically, measurements and indices of agreement have been proposed by a number of authors to define the quality of the mappings (e.g., FINLEY, 1884; EBEL, 1951; CARTWRIGHT, 1956; HAGGARD, 1958; COHEN, 1960; EVERITT, 1968). The KI is the most popular index to compare maps having the same variables. Galton (1892), in his work on fingerprints, presented the fundamentals and theoretical bases of the KI. Cohen (1960), in his research in medicine, defined, formalized and implemented the KI. The KI became very popular for two main reasons: a) the tremendous advance in the use of Geographical Information Systems (GIS) and Digital Processing of Remote Sensing Images (DPRSI) in several areas of knowledge; and b) the available number of public and commercial GIS and DPRSI software packages that perform the calculation of the KI (VISSER & DE NIJS, 2006; ERDAS, 2008; NETELER & MITASOVA, 2008; EASTMAN, 2009).

Table 1: Representation of cross-tabulated data with marginal variations, indicating situations with perfect association and no agreement (a) and different levels of association and agreement [(b) and (c)]. Tot = Total

		Yes	No	Tot
Real	Yes	0	60	60
	No	0	0	0
	Tot*	0	60	60

(a)

		Yes	No	Tot
Real	Yes	60	20	80
	No	20	0	20
	Tot	80	20	100

(b)

		Yes	No	Tot
Real	Yes	20	60	80
	No	0	20	20
	Tot	20	80	100

(c)

Several researchers, such as Congalton *et al.* (1983), Monserud & Leemans (1992), Congalton & Green (2009), Smits *et al.* (1999) and Wilkinson (2005), among others, recommended the use of the KI. Articles, theses and dissertations developed in several Brazilian and international universities have noted the use of the KI as the best estimate of the accuracy of the results (CORREIA *et al*., 2007; LIU *et al*., 2007; ABUBAKER *et al*., 2013; NERY *et al*., 2013). There are many texts that make reference to the use of Kappa in many areas of knowledge. Table 2 presents the results of a survey conducted in September 2015 with the Brazilian CAPES (Higher Education Personnel Training Coordination) software, which provides reference management tools. The keyword used was the "Kappa index". CAPES is a foundation of the Ministry of Education (MEC) of Brazil that plays a key role in the expansion and consolidation of graduate studies (M.Sc. and Ph.D.) in all Brazilian states. The results of searches conducted in Google Scholar and Web of Science are also presented in this table.

Table 2: Number of articles (journals and proceedings), reviews, theses and dissertations that cited the Kappa index in the CAPES, academic Google and Web of Science until September 2015

Time Period	CAPES	Academic Google	Web of Science
Before 1973	4,515	1,320	502
1974 - 1983	33,183	1,310	997
1984 - 1993	29,759	4,150	8,065
1994 - 2004	118,898	15,300	61,941
After 2004	288,483	16,600	135,828
TOTAL	474,838	38,680	207,333

In contrast, other researchers, such as Foody (1992, 2002), Stehman & Czaplewski (1998), Ma & Redmond (1995), Fielding & Bell (1997), Stehman (1992), Turk (2002), Pontius Jr. (2000, 2002), Pontius Jr. *et al.* (2008) and Pontius Jr. & Millones (2011), have made severe critiques of the widespread use of the KI. Congalton & Green (2009) recognized some of these critiques, but noted that Kappa "must still be considered vital accuracy assessment measure." However, there are many cases in which the KI indicates the correct proportion or inaccuracy when the reality may present contradictions, revealing the inaccuracy of the KI. Another index that has been used for data validation is the τ correlation coefficient (MA & REDMOND, 1995), which is based on the expected agreement (Pr) or a priori probability (KLECKA, 1980). The Pr is expressed by 1/n, where n is the number of categories or classes. Table 3 presents the evolution of the use of this coefficient.

Usually, evaluation of the accuracy between two thematic maps is obtained by indices that are calculated from a cross-tabulation or from a confusion matrix. The matrix is formed by a rectangular array in which the rows and columns express the number of simultaneity or categories that occupy the same geographic position in two existing maps. If we have N hypothetical pixels having the same spatial resolution and the same geographic position and that have been classified as n categories, then we can build a generic confusion matrix. Table 4 presents a generic confusion matrix. Typically, the columns represent the reference map, which is compared with the results of the interpretation categories that are generally represented along lines. The data in the diagonal indicate the level of accuracy.

Table 3: Total number of articles (journals and proceedings), reviews, theses and dissertations that cited the Tau index in the CAPES, academic Google and Web of Science until September 2015

Time Period	CAPES	Academic Google	Web of Science
Before 1973	15	102	19
1973 - 1983	30	106	35
1984 - 1993	47	121	422
1994 - 2004	103	150	2,327
After 2004	347	373	3,117
TOTAL	542	852	5,920

Table 4: General representation of a confusion matrix

	Reference Map						Line total
	Category	1	2	3	...	*n*	
Interpret-ed Map	1	x_{11}	x_{12}	x_{13}	...	x_{1n}	x_{1+}
	2	x_{21}	x_{22}	x_{23}	…	x2n	x_{2+}
	...	…	…	…	…	…	…
	n	x_{n1}	x_{n2}	x_{33}	…	x_{3n}	x_{n+}
Row total		x_{+1}	x_{+2}	x_{+3}	…	x_{+n}	*N*

With the data obtained from the confusion matrix, different indices can be used to assess their accuracy, including Kappa and τ. The Kappa index can be used as a measure of the agreement between the model predictions and reality (CONGALTON, 1991) or to determine if the values contained in an error matrix represent the results significantly better than the corresponding random value (JENSEN, 1996). The Kappa index is computed as given in equation 1:

$$K = \frac{P_o - P_e}{1 - P_e} = \frac{N\sum_{i=1}^{r}\chi_{ii} - \sum_{i=1}^{r}(\chi_{i+} * \chi_{+i})}{N^2 - \sum_{i=1}^{r}(\chi_{i+} * \chi_{+i})} \quad (1)$$

where N is the total number of sites in the matrix; r is the number of rows in the matrix; x_{ii} is the number from row i and column I; x_{+i} is the total for row I; and x_{i+} is the total for column i. Around the value of the KI, confidence intervals can be calculated using the sampling variance and the fact that the statistical distribution of the KI is usually asymptotic.

In an attempt to address the fact that the KI can provide, by chance, overestimated levels of agreement, including the actual agreement, Ma & Redmond (1995) proposed another method of assessing the thematic accuracy, the τ index. The Tau index measures the agreement by comparing the classification with a random assignment of pixels to the classes (KLECKA, 1980). MA & REDMOND (1995) introduced the Tau index for remote sensing data analysis, as defined according to equation 2.

$$\tau = \frac{P_o - P_c}{1 - P_c}$$

$$\tau = \frac{P_o - P_c}{1 - P_c}$$

$$P_o = \frac{1}{N}\sum_{i=1}^{n}\chi_{ii}$$

$$P_c = \sum_{i=1}^{n}\frac{\chi_{+1}}{N} * \frac{\chi_i}{N} = \frac{1}{N^2}\sum_{i=1}^{n}\chi_{+1}\,\chi_i$$

$$N = \sum_{i=1}^{n}\chi_{1+} = \sum_{i=1}^{n}\chi_{+1} \quad (2)$$

where χ_{+1} is the row total and χ_i is the diagonal value for category i (i.e., the number of correct assignments for category i).

Both the τ and the Kappa indices have similar concepts and are calculated from the marginal distributions of the reference data. The critical and important difference between the two coefficients is that τ is based on a priori probabilities of group membership, whereas the Kappa index uses a posteriori probabilities.

The interpretation of the KI, and the τ are as follows: value = 1 indicates that the maps resulting from the interpretations have perfect thematic accuracy; values close to 0 (zero) indicate that the thematic accuracy leads to nearly useless results. To calculate these indices, the cross-tabulation or confusion matrices must be established between the reference map (R) and the maps resulting from interpretation (A, B, C, D, E, F, G and H, Figure 1). Figure 2 shows these cross-tabulations, while Table 5 presents the results of the calculations of the KI, and τ, based on the reference map and the eight interpreted maps shown in Figure 2.

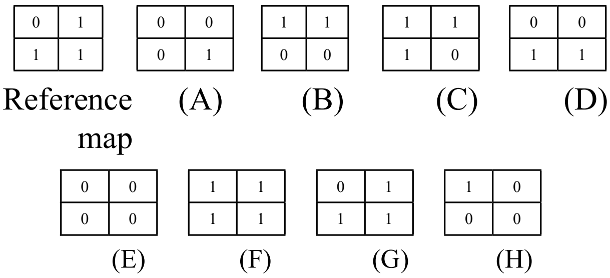

Fig. 1 - Reference map and eight possible maps (A, B, C, D, E, F, G e H) resulting from interpretation.

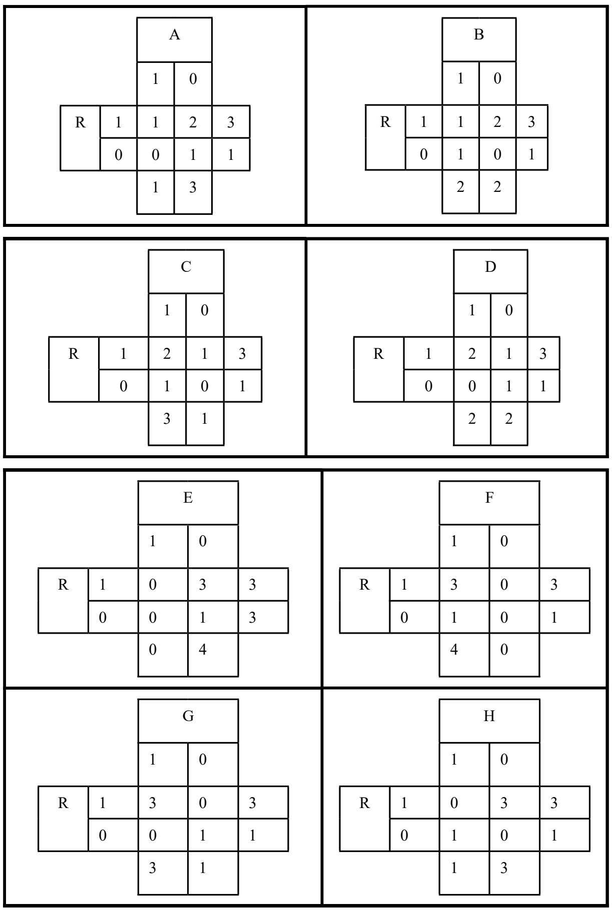

Fig. 2 - Cross-tabulation between the reference map (R) and the eight maps (A, B, C, D, E, F, G e H) resulting from interpretation.

The results presented in Table 5 reveal several paradoxes: a) negative values of the KI that are not originally foreseen (LANDIS & KOCH 1977); b) the KI is null and τ is positive; d) map F showed 50% accuracy, but the KI was 0; c) in map H, the KI presented negative and high values, while τ was null.

These contradictions are not resolved using equations that do not prevent variations in the marginal proportions.

Table 5: Results of the KI and τ calculations

	P_o	P_e	P_c	KI	τ
A	0.50	0.38	0.25	0.20	0.33
B	0.25	0.50	0.19	-0.50	0.08
C	0.50	0.63	0.38	-0.33	0.20
D	0.75	0.50	0.44	0.50	0.56
E	0.25	0.25	0.06	0	0.20
F	0.75	0.75	0.56	0	0.43
G	1.00	0.63	0.63	1.00	1.00
H	0	0.38	0	-0.60	0

3. GEOGRAPHICAL SIMULTANEITY (GS): PROPOSAL OF A NEW INDEX

When assessing the ability of a resulting map derived from digital processing images from remote sensing to be helpful to understand the landscape, it is important that its interpretation is not a product of guesswork.

To avoid and correct possible flaws in interpretations of the KI and τ, especially when they must assess the agreement between maps, even for those having a large number of pixels, this paper proposes a new index: the Geographical Simultaneity (GS).

The concept of GS considers both the amount of geographical correlation between existing pixels in two maps and the positioning of pixels, which eliminates the inconsistencies resulting from changes in the marginal proportions, making this a quite reliable index.

The GS is composed of two parts. The first is the exclusion, that is, the percentage of errors made between a particular attribute of a reference map and the errors made by this attribute in an interpreted map. The second is the extent, that is, the percentage of errors made between a particular attribute of an interpreted map and the errors made by this attribute in a reference map. The sum of the exclusion and the extent represents the GS, which can vary between 0 (no correlation) and 2 (total correlation). Table 6 shows the general confusion matrix between the reference (R) and interpreted (I) maps.

Table 6: Confusion matrix between a reference map (r) and an interpreted map (i) describing the same geographical position

		Interpreted Map (I)		
Reference Map (R)		1	Others	
	1	χii	$\chi i+ - \chi ii$	$\chi i+$
	Others	$\chi +i - \chi ii$	$N – [\chi ii + (\chi i+ - \chi ii) = (\chi +i - \chi ii)]$	$\{N – [\chi ii + (\chi i+ - \chi ii) + (\chi +i - \chi ii)]\} = (\chi +i - \chi ii)$
		$\chi +i$	$\{N – [\chi ii + (\chi i+ - \chi ii) + (\chi +i - \chi ii)\} + (\chi i+ - \chi ii)$	N

The extent and the exclusion errors are calculated by the equations (3) and (4), respectively.

$$Extent = 1 - \frac{(\chi_{i+} - \chi_{ii})}{\chi_{i+}} \tag{3}$$

$$Exclusion = 1 - \frac{(\chi_{+i} - \chi_{ii})}{\chi_{+i}} \tag{4}$$

where χ_{ii} is the total concordance category, (χ_{i+} - χ_{ii}) represents the total concordance category with the other, χ_{i+} is the total of the line, χ_{+i} represents the total column, (χ_{+i} - χ_{ii}) is the column total less the total concordance of the category and N is the total pixels. The GS is obtained using the following relation: GS = Exclusion + Extent.

The GS can be calculated in all cases, i.e., maps having *n* number of categories. However, the GS is valid only when those involved maps represent exactly the same geographical area and have the same spatial resolution, which means having the same number of rows and columns. The GS varies from 0 to 2, corresponding to the sum of the exclusion and the extent. When the GS tends to 2, the result is consistent, and when it tends to zero, the result should not be used. If there are more than two categories, then the GS can be calculated for each category, and the overall GS will be calculated as the arithmetic sum of the standard GS indices. Given the variation of GS, a simplified hierarchy of values is proposed here to assist the interpretation of the correspondence between two maps, as shown in Table 7, where the amplitudes of the data scale assign intermediate degrees to accomplish the various objectives. According to Table 7, if there is geographical similarity, then the result should be ≥ 1.5. A very good degree of accuracy is constituted by data ranging from 1.5 to 1.7. The degree of accuracy is excellent when GS ≥ 1.7.

Table 7: Degrees of accuracy for the interpretation of GS

Accuracy	Ranking
1.7 to 2.0	Excellent
1.5 to 1.7	Very good
1.2 to 1.5	Good
0.9 to 1.2	Regular
0.6 to 0.9	Poor
< 0.6	Unacceptable

4. EXAMPLES OF APPLICATIONS OF GS

To understand the use of GS, we formulated a hypothetical study of the usefulness of interpreted maps derived from digital processing of remote sensing images. This study is based on scenarios (Figures 3, 4 and 5) which involve random variation of the intersections, the same geographical positions and a total discrepancy of the data.

We considered three scenarios with two maps, each having 16 pixels. In scenario 1, a clear match was established between the same attributes. In scenario 2, a discrepancy exists between the total intersections, while in the third scenario, there is a random variation of the intersections.

Scenario 1 expresses a hypothetical situation in which the data are highly correlated, as shown in Figure 4, revealing the existence of seven pixels for which the first attribute occupies the same position in the two geographical maps; four pixels with attribute 2 in the same geographic locations; and five pixels with attribute 3 in the same geographical locations.

Referenced map (R)

1	1	3	2
2	3	1	1
1	3	3	2
3	1	1	2

Interpreted map (I)

1	1	3	2
2	3	1	1
1	3	3	2
3	1	1	2

Cross-tabulation map (MAP I)

MAP R		1	2	3
	1	7	0	0
	2	0	4	0
	3	0	0	5

Fig. 3 - Cross-tabulation between paired maps showing all correspondences.

The results of the cross-tabulation for scenario 1 indicate the developments presented in Table 8. In this case, for the three categories, the sum of the exclusion and the extent for each attribute results in GS = 2. This result proves the theorem established, in which, when the data set has full match, the GS of all categories is equal to 2; the result also confirmed by the KI = 1 and τ = 1. To calculate the total GS, the data are normalized for each category, and then, the arithmetic sum is performed as shown in equation 5.

$$GS_t = \frac{2}{3} + \frac{2}{3} + \frac{2}{3} = 2 \tag{5}$$

Table 8: Calculation of the Geographical Simultaneity between the attributes of the reference and interpreted maps for data with total correspondence

Cross-tabulation 1 1		Map I		Geographical Simultaneity = 2	
		Others	Exclusion	Extent	
Map R	Attribute 1	7	0	$1-\frac{0}{7}=1$	$1-\frac{0}{7}=1$
	Others	0	9		

Cross-tabulation 2 2		Map I		Geographical Simultaneity = 2	
		Others	Exclusion	Extent	
Map R	Attribute 2	4	0	$1-\frac{0}{4}=1$	$1-\frac{0}{4}=1$
	Others	0	12		

Cross-tabulation 3 3		Map I		Geographical Simultaneity = 2	
		Others	Exclusion	Extent	
Map R	Attribute 3	5	0	$1-\frac{0}{5}=1$	$1-\frac{0}{5}=1$
	Others	0	11		

In scenario 2, the data did not exhibit any correlation, as shown in Figure 4.

Referenced map

1	1	3	2
3	3	2	1
2	2	3	2
2	1	3	2

Interpreted map

2	3	1	1
2	1	1	3
1	3	2	1
3	2	1	1

Cross-tabulation map

MAP R \ MAP I	1	2	3
1	0	2	2
2	5	0	2
3	3	2	0

Fig. 4 - Comparison between paired maps that have no correspondence.

The results of the cross-tabulation for scenario 2 have the consequences presented in Table 9. The result of the calculation of the GS referring to scenario 2 indicates that, in this case, for the three categories, the sum of the exclusion and the extent is zero, which supports the mismatch between the data, as we sought to confirm. In this case, the KI has negative values. The values of τ were 0.25 for category 1, 0.32 for category 2 and 0.44 for category 3, indicating little correlation, when, in fact, there is no correlation. To calculate the total GS, normalization of the data for each category must be performed, followed by calculating the arithmetic sum:

$$GS_t = \frac{0}{3} + \frac{0}{3} + \frac{0}{3} = 0 \tag{6}$$

In Scenario 3, we note that there are geographical areas of intersection between the data, as shown in Figure 5.

Referenced map

1	1	3	2
2	3	1	1
1	3	3	2
3	1	1	2

Interpreted map

1	1	3	2
2	1	1	3
1	3	2	1
3	2	1	1

Cross-tabulation map (MAP I)

MAP R	1	2	3
1	5	1	1
2	2	2	0
3	1	1	3

Fig. 5 - Cross-tabulation between paired data that show little correlation.

Table 9: Calculation of the geographical simultaneity between the attributes of the reference and interpreted maps, for data with no matches

Cross-tabulation 1		Map I		Geographical Simultaneity = 0	
1		Others	Exclusion	Extent	
Map R	Attribute 1	0	8		
	Others	4	4		
Cross-tabulation 2		Map I		Geographical Simultaneity = 0	
2		Others	Exclusion	Extent	
Map R	Attribute 2	0	4		
	Others	7	5		
Cross-tabulation 3		Map I		Geographical Simultaneity = 0	
3		Others	Exclusion	Extent	
Map R	Attribute 3	0	4		
	Others	5	7		

The results of the cross-tabulation related to scenario 3 (Table 10) indicate that, in this case, for the three categories, thee exclusion and the sum are different from zero, and the same applies to extent, with the GS index in the range between 0 and 2 (1.32, 1.00, 0 95), which supports the partial match between the data. The KI ranged between 0.33 and 0.64 and τ ranged between 0.63 and 0.80. To calculate the total GS, the data for each category must be normalized, followed by calculation of the arithmetic sum:

$$GS_t = \frac{1.32}{3} + \frac{1}{3} + \frac{0.95}{3} = 1.09 \tag{6}$$

If we return to Figure 1 and calculate the GS for the eight interpreted maps, we see that the GS corrects the errors encountered when the KI calculations are performed, as shown in Table 11.

Table 11 reveals that the GS corrects the results for all maps, indicating that, in the case of the E and F maps presented, the KI values of 0 (zero) and negative values were corrected to positive values. If we perform a number of simulations (Table 12), varying the number of intersections using a binary connectivity table and comparing the GS, KI and τ indices, then we can verify the consistency of the GS use, eliminating the potential paradox existing in the calculation of the KI and τ, which is usually derived from the variations of the marginal proportions.

Table 10: Calculation of the geographical simultaneity between the attributes of the reference and interpreted maps, for data with little correspondence

Cross-tabulation 1		Map I		Geographical Simultaneity = 1.32	
		Others	Exclusion	Extent	
Map R	Attribute 1	5	3	$1-\frac{2}{7}=0.71$	$1-\frac{3}{8}=0.62$
	Others	2	6		
Cross-tabulation 3		Map I		Geographical Simultaneity = 1.00	
		Others	Exclusion	Extent	
Map R	Attribute 2	2	2	$1-\frac{2}{4}=0.50$	$1-\frac{2}{4}=0.50$
	Others	2	10		
Cross-tabulation 3		Map I		Geographical Simultaneity = 0.95	
		Others	Exclusion	Extent	
Map R	Attribute 3	3	1	$1-\frac{2}{5}=0.40$	$1-\frac{1}{4}=0.75$
	Others	2	10		

Table 11: Comparison of the KI and GS

Maps	τ	KI	GS
A	0.33	0.20	1.33
B	0.08	-0.50	0.83
C	0.20	-0.33	1.33
D	0.56	0.5	1.67
E	0.20	0	0
F	0.43	0	1.75
G	1	1	2
H	0	-0.60	0

To demonstrate the existence of errors arising from the interpretation of the KI and, by contrast, emphasize the importance of the results from the calculation of the GS, Table 12 presents, in column A, the variation from perfect results (seven matches between category 1 of a given reference map and a map interpreted from the same geographical area that both have the same spatial resolution) to total disagreement (correspondence between category 1 of a given reference map and a map interpreted from the same geographical area that both have the same spatial resolution). These correspondences between the reference map and the eight interpreted maps have a linear decrease.

Table 12: Simulation of different amounts of intersections of pixels that occupy the same geographic position

Maps	A	B	C	D	E	F	G
	1 and 1	1 and others	others and 1	Others and others	Tau	GS	KI
1	7	0	0	0	1.00	2	Impossible
2	6	1	1	7	0.76	1.71	0.73
3	5	2	2	6	0.58	1.43	0.46
4	4	3	3	5	0.43	1.14	0.20
5	3	4	4	4	0.30	0.86	-0.07
6	2	5	3	5	0.30	0.69	-0.09
7	1	6	2	6	0.29	0.48	-0.11
8	0	7	1	7	0.29	0	-0.13

Note that in analyzing Table 12, when there is perfect agreement (map 1), only the GS and Tau (τ) can identify the agreement; when there is total disagreement (map 8), only the GS identifies the disagreement; in maps 2, 3 and 4, a correlation exists between the results for τ and GS. Only the results for the GS exhibit a reduction of consistently linear thematic accuracy from map 1 to map 8, corresponding to an expected correlation between the reference and the interpreted maps.

The existence of different patterns in the behavior of the indices of thematic accuracy is revealed in Figure 7. This figure shows the variations between perfection and disagreement for the interpretation of the maps obtained with the data in columns E, F and G which represent the values obtained for τ, GS and KI, respectively. Notice that the GS values are linearly distributed, showing a regular pattern while τ and KI show discontinuities.

We also illustrate the application of the GS in a specific case. Sano *et al.* (2009) analyzed the potential of ALOS PALSAR radar images to map the land use and the land cover classes of the Federal District (Brazil). The amplitude images obtained in the L-band and the HH, HV and VV polarizations were converted into backscatter coefficients and processed through the image segmentation technique by growing region. The following thematic classes were discriminated: consolidated urban areas; urban areas in consolidation; croplands; pasturelands; reforestation; grasslands; Cerrado shrubland; indiscriminated forest; gallery forest; and reservoirs. The mapping accuracy provided by the τ index was 70%. A set of 86 points were gathered in the field to generate the confusion matrix (Table 13).

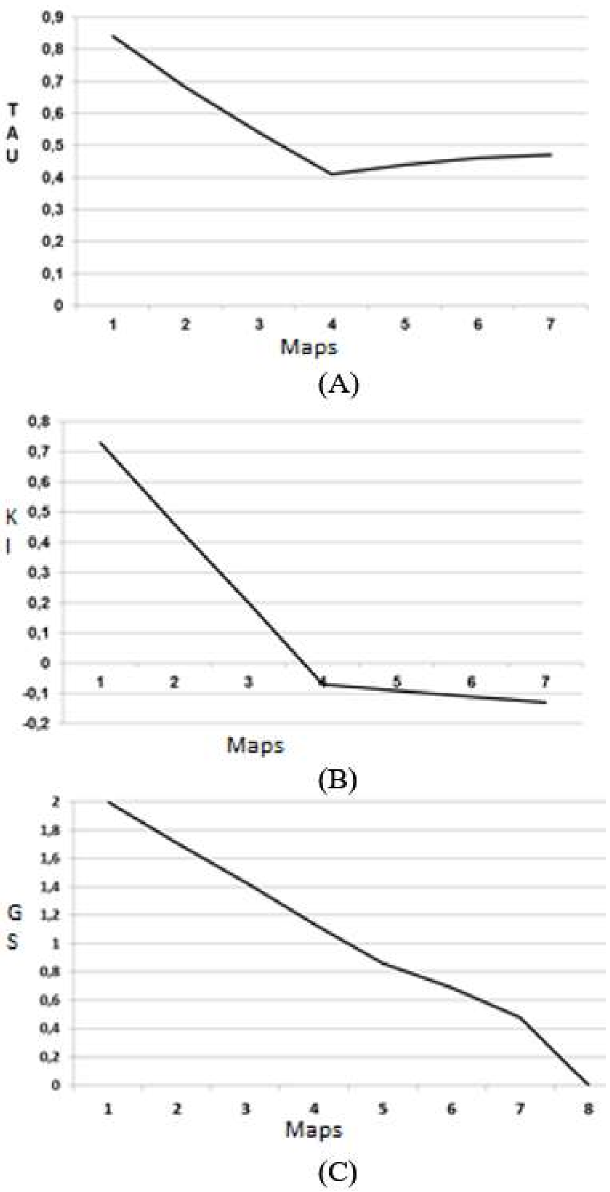

Fig. 7 - Graphics representing the calculated values and showing the lack of a standard linear pattern between perfection ($\tau = 1$) and total disagreement ($\tau = 0$) (A); linear decrease between perfection (GS = 2) and total disagreement (GS = 0) (B); the lack of a standard linear pattern between perfection (KI = 1) and total disagreement (KI = 0) (C).

Table 13: Confusion matrix of the land use and land cover mapping errors of the study area (Federal District, Brazil). AUC = consolidated urban areas; AUE = urban areas in consolidation; CUL = croplands; PAS = planted pasturelands; REF = reforestation; CAM = grasslands; CTI = Cerrado shrubland; MIN = indiscriminated forests; MGA = gallery forest; RES = reservoirs; EO = omission errors (Congalton, 1991); EC = commission errors (Congalton, 1991); GS = geographical simultaneity; GS (n) = normalized geographical simultaneity.

INTERPRETED

REFERENCE	AUC	AUE	CUL	PAS	REF	CAM	CTI	MIN
AUC	5	0	0	0	0	1	0	0
AUE	0	5	0	1	0	0	1	0
CUL	0	0	4	0	0	0	0	0
PAS	0	0	2	18	0	2	5	1
REF	0	0	0	0	5	0	0	0
CAM	0	0	1	0	0	8	2	0
CTI	0	0	1	0	0	3	5	0
MIN	0	0	0	0	0	0	2	3
MGA	0	0	0	0	0	0	0	1
RES	0	0	0	0	0	0	0	0
	5	5	8	19	5	14	15	5

INTERPRETED

REFERENCE	MGA	RES	TOT	%EO	%EC	GS	GS(n)
AUC	0	0	6	0.17	0.00	1.83	0.18
AUE	0	0	7	0.29	0.00	1.71	0.17
CUL	0	0	4	0.00	0.50	1.5	0.15
PAS	0	0	28	0.36	0.05	1.59	0.16
REF	0	0	5	0.00	0.00	2	0.2
CAM	0	0	11	0.27	0.43	1.3	0.13
CTI	0	0	9	0.44	0.67	0.89	0.09
MIN	0	0	5	0.40	0.40	1.2	0.12
MGA	5	0	6	0.17	0.00	1.83	0.18
RES	0	5	5	0.00	0.00	2	0.2
	5	5	86				1.59

5. CONCLUDING REMARKS

A variety of approaches have been developed to assess the accuracy of thematic maps. Various measurements have been developed to compare the accuracy of maps, but the extent to which these measures are consistent is questionable. The overall accuracy of a final digital or analogical map dictates the value of the data for any application. It is well-known that map accuracy measurements allow map producers to analyze the sources of error and the weaknesses of final maps.

Among several existing indices in the literature to measure accuracy, the KI and τ are the most widely used indices. However, the KI is quite vulnerable to changes in the marginal proportions, increasing the appearance of paradoxes that are often not distinguishable, which can cause unacceptable errors in the validation of the results and their consequent interpretations. The origin of the errors is based on the three indices designed to analyze extremely dependent data; alternatively, the cause and effect are visible and interdependent, for example, epidemiological, psychological and educational measurements, among others. In the case of thematic maps, where the variables are primarily independent, changes in the marginal proportions are evident, often causing errors in the interpretation of the KI and they should not be used to validate maps. τ, in certain circumstances, can be contaminated by the marginal proportions, leading to an incorrect interpretation.

The GS corrects and eliminates the possibility of interference of the changes in the marginal proportions and can be used at multiple scales and for maps with large amounts of pixels, provided that the following fundamental conditions are met: the same spatial resolution and an even geographic space are used. The calculation of the GS ensures that the contents will always have positive values and that the variation range is always between 0 and 2, where 0 indicates no thematic accuracy and 2 indicates overall thematic accuracy. The distribution of GS values, points out a distribution shown by straight which demonstrates a regular pattern, easy to understand.

The interpretation of the case study considered in this study led to the following conclusions: the sum of GS and the two class specific indices defined by Congalton (1991) EO and EC will always be equal to 2; when GS is equal to 0 (zero), EO and EC are equal to 1; when GS is equal to 2, there are no EO and EC; GS can be calculated for individual categories; the total GS can be set, i.e., the total subject accuracy will be the algebraic sum of all the normalized GS(n); and the joint use of EO, EC and GS enables identification, in each category, of the error introduced in the interpretation that was added to the map and that exists in the field as well as calculation of the error due to the non-inclusion of data on the map that exists on the ground and determination of the estimated accuracy, that is, the thematic accuracy. Furthermore, the calculations of GS are quite simple and can be adapted to data from various PDISR and GIS software.

REFERENCES

ABUBAKER, H. M. A.; ELHAG, A. M. H.; SALIH, A. M. Accuracy assessment of land use & land cover classification (LU/LC). Case study of Shomadi area - Renk County - Upper Nile State, South Sudan. **International Journal of Scientific and Research Publications**, v. 3, n. 5, pp. 1-6, 2013.

AICKIN, M. Maximum likelihood estimation of agreement in the constant predictive probability model and its relation to Cohen's kappa. **Biometrics**, v. 46, n. 2, pp. 293-302, 1990.

BYRT, T.; BISHOP, J.; CARLIN, J. B. Bias, prevalence and Kappa. **Journal of Clinical Epidemiology**, v. 46, n. 5, pp. 423-429, 1993.

CARTWRIGHT, M. L. **Integral Functions** (Cambridge Tracts in Mathematics and Mathematical Physics, 44). Cambridge: Cambridge University Press, 1956, 143 p.

COHEN, J. A. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, v. 20, n. 1, pp. 37- 46, 1960.

CONGALTON, R. G. A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, v. 37, n. 1, pp. 35-46, 1991.

CONGALTON, R. G.; GREEN, K. **Assessing the Accuracy of Remotely Sensed Data. Principles and Practices.** Boca Raton: CRC Press, 2nd ed., 2009, 183 p.

CONGALTON, R. G.; ODERWALD, R. G.; MEAD, R. A. Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. **Photogrammetric**

Engineering and Remote Sensing, v. 49, n. 12, pp. 1671-1678, 1983.

CORREIA, V. R. M.; MONTEIRO, A. M. V.; CARVALHO, M. S.; WERNECK, G. L. Uma aplicação do sensoriamento remoto para a investigação de endemias urbanas. **Caderno de Saúde Pública**, v. 23, n. 5, pp. 1015-1028, 2007.

EASTMAN, J. R. IDRISI 16: **The Taiga Edition**. Worcester: Clark University, 2009, 342 p.

EBEL, R. L. Estimation of the reliability of ratings. **Psychometrika**, v. 16, n. 4, pp. 407-24, 1951.

ERDAS. **ERDAS Field Guide**. Norcross: Erdas, Inc. 2008, 852 p.

EVERITT, W. N. Some positive definite differential operators. **Journal of the London Mathematical Society**, v. 43, n. 1, pp. 465-473, 1968.

FIELDING, A. H.; BELL, J. F. A review of methods for the assessment of prediction errors in conservation presence/absence models. **Environmental Conservation**, v. 24, n. 1, pp. 38-49, 1997.

FINLEY, J. P. Tornado predictions. **American Meteorological Journal**, v. 1, pp. 85-88, 1884.

FINN, J. T. Use of the average mutual information index in evaluating classification error and consistency. **International Journal of Geographical Information Systems**, v. 7, n. 4, pp. 349-366, 1993.

FOODY, G. M. On the compensation for chance agreement in image classification accuracy assessment. **Photogrammetric Engineering and Remote Sensing**, v. 58, n. 10, pp. 1459-1460, 1992.

FOODY, G. M. Status of land cover classification accuracy assessment. **Remote Sensing of Environment**, v. 80, pp. 185-201, 2002.

FUNG, T.; LeDREW, E. The determination of optimal threshold levels for change detection using various accuracy indices. **Photogrammetric Engineering and Remote Sensing**, v. 54, pp. 1449-1454, 1988.

GALTON, F. **Finger Prints**. London: Macmillan, 1892, 216 p.

GHOSH, J.; STREHL, A.; MERUGU, S. A consensus framework for integrating distributed clusterings under limited knowledge sharing. In: **Proceedings of NSF Workshop on Next Generation Data Mining (NGDM´02),** Baltimore, 2002, pp. 99-108.

HAGGARD, E. A. **Intraclass Correlation and the Analysis of Variance**. New York: Dryden Press, 1958.

HUBERT, L.; ARABIE, P. Comparing partitions. **Journal of Classification**, v. 2, pp. 193-218, 1985.

JENSEN, J. R. **Introductory Digital Image Processing. A Remote Sensing Perspective** (2nd ed.). Upper Saddle River: Prentice Hall, Inc., 1996, 318 p.

KLECKA, W. R. **Discriminant Analysis** (Quantitative Applications in the Social Sciences, 19.). Newbury: SAGE Publication, Inc., 71 p., 1980.

KOUKOULAS, S.; BLACKBURN, G. A. Introducing new indices for accuracy evaluation of classified images representing semi-natural woodland environments. **Photogrammetric Engineering and Remote Sensing**, v. 67, pp. 499-510, 2001.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, pp. 159-174, 1977.

LIU, C. R.; FRAZIER, P.; KUMAR, L. Comparative assessment of the measures of thematic classification accuracy. **Remote Sensing of Environment**, v. 107, n. 4, pp. 606-616, 2007.

MA, Z.; REDMOND, R. L. Tau coefficients for accuracy assessment of classification of remote sensing data. **Photogrammetric Engineering and Remote Sensing**, v. 61, pp. 435-439, 1995.

MONSERUD, R. A.; LEEMANS, R. Comparing global vegetation maps with the Kappa statistics. **Ecological Modelling**, v. 62, n. 4, pp. 275-293, 1992.

NERY, C. V. M.; OLIVEIRA, D. B.; ABREU, L. H. G. Estudo comparativo entre os índices NDVI obtidos a partir dos sensores Landsat 5 - TM e Resourcesat - LISS III. **Caminhos de Geografia**,

v. 14, n. 46, pp. 160-167, 2013.

NETELER, M.; MITASOVA, H. **Open Source GIS. A GRASS GIS Approach** (3rd ed.). New York: Springer, 2008, 420 p.

PONTIUS Jr., R. G. Quantification error versus location error in comparison of categorical maps. **Photogrammetric Engineering and Remote Sensing**, v. 66, n. 8, pp. 1011-1016, 2000.

PONTIUS Jr., R. G. Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. **Photogrammetric Engineering and Remote Sensing**, v. 68, n. 10, pp. 1041-1049, 2002.

PONTIUS Jr., R. G.; MILLONES, M. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. **International Journal of Remote Sensing**, v. 32, n. 15, pp. 4407-4429, 2011.

PONTIUS Jr., R. G.; BOERSMA, W.; CASTELLA, J. C.; CLARKE, K.; de NIJS, T. et al. Comparing the input, output, and validation maps for several models of land change. **Annals of Regional Science**, v. 42, n. 1, pp. 11-47, 2008.

ROSENFIELD, G. H.; FITZPATRICK-LINS, K. A coefficient of agreement as a measure of thematic classification accuracy. **Photogrammetric Engineering and Remote Sensing**, v. 52, n. 2, pp. 223-227, 1986.

ROY, S. N.; MITRA, S. K., 1956, An introduction to some nonparametric generalizations of analysis of variance and multivariate analysis. **Biometrika**, v. 43, n. 3/4, pp. 361-376, 1956.

SANO, E. E.; SANTOS, E. M.; MENESES, P. R. Análise de imagens do satélite ALOS PALSAR para o mapeamento de uso e cobertura da terra do Distrito Federal. **Geociências**, v. 28, n. 4, pp. 441-451, 2009.

SMITS, P. C.; DELLEPIANE, S. G.; SCHOWENGERDT, R. A. Quality assessment of image classification algorithms for land-cover mapping: A review and a proposal for a cost-based approach. **International Journal of Remote Sensing**, v. 20, n. 8, pp. 1461-1486, 1999.

STEHMAN, S. V. Comparison of systematic and random sampling for estimating the accuracy of maps generated from remotely sensed data. **Photogrammetric Engineering and Remote Sensing**, v. 58, n. 9, pp. 1343-1350, 1992.

STEHMAN, S. V.; CZAPLEWSKI, R. L. Design and analysis for thematic map accuracy assessment: Fundamental principles. **Remote Sensing of Environment**, v. 64, n. 3, pp. 331-344, 1998.

STORY, M.; CONGALTON, R. G. Accuracy assessment. A user's perspective. **Photogrammetric Engineering and Remote Sensing**, v. 52, n. 3, pp. 397-399, 1986.

STREHL, A.; GHOSH, J. Cluster resembles. A knowledge reuse framework for combining multiple partitions. **Journal of Machine Learning Research**, v. 3, pp. 583-617, 2002.

TÜRK, G. Letter to the editor: Map evaluation and chance correction. **Photogrammetric Engineering and Remote Sensing**, v. 68, pp. 123-129, 2002.

VISSER, H.; de NIJS, T. The map comparison kit. **Environmental Modeling & Software**, v. 21, pp. 346-358, 2006.

WILKINSON, G. G. Results and implications of a study of fifteen years of satellite image classification experiments. **IEEE Transactions on Geoscience and Remote Sensing**, v. 43, n. 3, pp. 433-440, 2005.