Anais XI SBSR, Belo Horizonte, Brasil, 05 - 10 abril 2003, INPE, p. 1427 - 1431.

CONTRIBUTION OF REMOTE SENSING AND GIS TO EVALUATE THE ACCURACY OF LAND USE/LAND COVER IN TRÊS PONTAS MUNICIPALITY, MINAS GERAIS STATE, BRAZIL

MAURÍCIO EDUARDO SALGADO RANGEL [1]
HERMAN JOHANN HEINRICH KUX [2]
GIL MENDES SALES [3]

[1] DEGEO/UFMA – Departamento de Geociências da Universidade Federal do Maranhão
Avenida dos Portugueses s/n – Centro de Ciências Humanas
Campus do Bacanga – São Luís/MA, Brasil
Tel/Fax.: +55 98 245-6474
mrangel_slz@yahoo.com.br

[2] INPE - Instituto Nacional de Pesquisas Espaciais
Caixa Postal 515 - 12201-970 - São José dos Campos/SP, Brasil
hermann@ltid.inpe.br

[3] GeoSystem – Consultoria
Rua Mitra, nº 10 – Sala 110 – Reanascença II – São Luís/MA, Brasil
Geosystem_consult@yahoo.com

Abstract. One of the most relevant questions for the use of information obtained from remote sensing systems, refers to its capacity to represent adequately those data obtained during field survey. The exactness of a map is based on a quantitative measurement of a thematic map that has been correctly classified. Under this assumption, the objective of this study is to analyze four statistics: *Kappa*, *Global*, *Tau* and *Sampling by Significance Level*. Therefore we used a TM-Landsat composite (RGB channels 4,5,3) that was manually analyzed. Samples for the evaluation of ground truth were obtained using a random procedure. Through *SAL* it was possible to determine the confidence of a land use/land cover map. Sampling by Significance Level indicated the need to express in advance statistically the truth of the reference data (the estimated trustworthy of the map), instead of using an index of agreement to estimate the exactitude. Referring to the estimators of agreement *G*, *K* and *T* we concluded that *K* did express best the estimate for accuracy of a thematic map obtained, when comparing to the other estimators.

Keywords: accuracy, sampling, estimators statistics.

1. Introduction

Most techniques used for the quantitative evaluation of the accuracy of mapping start with a confusion matrix. Based on such a matrix one can consider two forms of evaluation of the accuracy of a map. The first is based on the use of statistical estimators which describe mainly the percentage of agreement among a reference (considered as the truth) and a product (map), classified manually or digitally, i.e. there is an evaluation on how far the product generated approaches to the reference. Among those statistical methods that are being most used for the evaluation of accuracy of a map, one could mention: *global accuracy*, *kappa* and *tau*.

The second form can be considered as a methodology for the evaluation, based on a previously established statistics between the producer of the map and its' user, were both are looking for the best cost/benefit ratio, in order to obtain a precise map. In this frame Gnevan (1979) proposed a statistical method to estimate the accuracy of a map, which was later on also used by Aronoff (1982) and Valeriano (1985), based on a statistical branch known as Sampling by Acceptance Level (*SAL*). This method is commonly used for the quality control of industrial products, whose priority is to establish a minimum quantity of samples needed to define the confidence in a map.

The objective of this study is to use this methodology for the evaluation of a thematic map on land use/land cover in the municipality of Três Pontas, Minas Gerais State, Brazil, in order to have confidence on the survey of areas with coffee plantations. The need to use this methodology, is based on the fact that orbital remote sensing images with spectral and spatial resolutions that are compatible with such a study, would increase confidence on the results of a survey of coffee plantations, because it provides a clear overview of the culture, if compared with survey data, based on questionnaires which are subject to many errors. Another objective was to evaluate the performance of the statistical estimators *Global* (*G*), *Kappa* (*K*) and *Tau* (*T*), when related to the methodology proposed by Ginevan (1979).

2. Description of the Area Under Study

This study was made in the municipality of Três Pontas, located at southern Minas Gerais State, between geographical coordinates S21°15'00" to S21°30'00" and WGr.45°15'00" to WGr.45°45'00", encompassing totally 663 km^2. This municipality presents mainly 2 types of coverage: natural pasture and coffee plantations. The area is the largest coffee producer in Brazil, including 30 million coffee trees in an area of 30 million ha, producing per harvest approximately 350 thousand sacks, according to COCATREL (1999).

3. Materials and Methods

The manual interpretation of land use/land cover of the municipality Três Pontas was performed on an analogic TM/Landsat-5 image from July 22nd 1998, color composite 453 (*RGB*), at the scale 1:50,000. On the overlay obtained, a square grid with 5x5 mm cell size was placed. On this grid a simple random sampling was made of 274 samples (n), which were established on the risk assumed by the producer, to reject the map ($R_p = 0.20$), within a high accuracy value established by the producer ($P_p = 0.90$), for an user risk ($R_u = 0.05$) to accept a map with a quality below the users least acceptable quality ($P_u = 0.85$) according to Aronoff (1982), Ginevan (1979) and Valeriano (1985), where also the limit of erroneously accepted samples was defined ($x=31$).

Based on field survey, a confusion matrix was set up, where we determined *G*, *K* and *T*.

4. Results

Based on the results found, we could define the *Maximum Map Accuracy* (*MMA*) and the *real* R_p (Table 1), besides confirming that the map agreed with the statistical conditions established, i.e. it expressed confidence to obtain information referring to coffee plantation.

Table 1 – Confusion matrix obtained after field survey and of established sampling procedure.

		CLASSES OBSERVED IN THE FIELD																
		C 1	C 2	M	G	CA	CP	P	MN	SA	AP	AQ	SE	A	CH	AU	TOTAL	% INCLUSION
CLASSES MAPPED	C 1	39							2								41	5
	C_2		51	1				1		2	1						56	9
	M			2				1									3	33
	G				0												0	0
	CA					1											1	0
	CP						0										0	0
	P		1					97					2				100	3
	MN							1	44								45	2
	AS									0							0	0
	AP		1								3						4	25
	AQ											0					0	0
	SE			1				8					11				20	43
	A											1		0			1	100
	CH														0		0	0
	AU															3	3	0
	TOTAL	39	53	4	0	1	0	108	46	2	4	1	13	0	0	3	274	
	% OMISSION	0	4	50	0	0	0	10	4	100	25	100	15	0	0	0		

Nr of samples (*n*)	**274**
Nr. Of points admitted as erroneously classified (*x*)	**31**
Nr. Of points observed as erroneously classified (*x'*)	**23**
Users Risk (R_u)	**5%**
Users least acceptable accuracy (P_u)	**85%**
Producers risk (R_p) for accuracy established by the producer ($P_p = 0.90$)	**20%**
Real R_p based in *x'*	**14,8%**
Maximum Map Accuracy (*MMA*) based in *x'*	**89%**

LEGEND:

- C_1 = *COFFEE_1*
- C_2 = *COFFEE_2*
- M = *CORN*
- G = *SUNFLOWER*
- CA = *SUGAR CANE*
- CP = *GREENFODDER*
- P = *PASTURE*
- MN = *PRIMARY FOREST*
- SA = *SECUNDARY VEGETATION*
- AP = *HERBACEOUS*
- AQ = *BURNED AREA*
- SE = *BARE SOIL*
- A = *WATER*
- CH = *FLOODED AREA*
- AU = *URBAN AREA*

Table 2 presents the calculated values for *MMA* using *SAL* and other methods according to Ma and Redmond (1995).

Table 2 – Estimated values based on the confusion matrix obtained.

Statistical Methods	Estimated value in	
	Dec. Point Pointimal	Percentageal
MMA (*SAL*)	0,89000002	89%
G	0,91605839	92%
K	0,89008459	89%
T	0,91046228	91%

As observed, all methods present a higher performance than the value of precision required by the user (Table. 1). Nevertheless for the case of *SAL*, the high precision value established by the producer (P_p) was the upper limit of accuracy for the map which were surpassed by both *G* and *T* estimators, i.e. they overestimated the real accuracy of the map (*MMA = 0.89*) and the highest exactitude value accepted by the producer at his risc (R_p = *0.20*) while stating that the map would have a P_p *=0.90*.

G and *T* overestimated the accuracy is probably due to the fact that *G* considers only the main diagonal for its' calculation and *T* is an average value between real and casual agreement.

The *K* estimator became equivalent to *MMA* and remained within the limit of the map. Being so, here *K* presented a higher confidence to represent the accuracy estimation of the map. For this case, those considerations made by Ma and Redmond (1995), referring to the representative value of an improved estimation of *T* as related to *K* was different than what was found in this work.

One should consider that the indices mentioned, except for *G*, can be considered good estimators of accuracy of a map, because they consider casual agreement, even though in this study, the performance of *K* was better than *T*.

A relevant result of this work was the use of *SAL* for the definition of map accuracy, because it allowed the exact estimation of the map, and besides that, it can be considered as a reference for comparison of performance from the other estimators (Table 2).

5. Conclusion

SAL is important to emphasize the need to express statistically a priori the truth of reference data (the estimated map confidence), i.e., based on previously established statistical parameters (P_u, R_u, P_p, R_p, *n* and *x*) instead of using only an agreement index to estimate the accuracy of some type of classification only due to convenience or frequent use.

Through *SAL* it was possible to determine that the land use/land cover map presents confidence and consequently it was possible to obtain more accurate information of the coffee plantation area in the municipality of Três Pontas.

Referring to the agreement estimators, *G*, K and *T* one concludes that *K* expressed best the accuracy of the thematic map obtained if compared to others. Nevertheless both *K* and *T* can be considered in the evaluation of map accuracy, because they don't disdain the causal agreement, different than *G*.

Acknowledgements

We acknowledge help received from the Cooperative of Coffee Planters of Três Pontas (COCATREL). for the information delivered, as well as to the local office from EMATER-MG for data delivery and support during field survey. INPE provided funds for field survey.

References

Aronoff, S. "The map accuracy report: a user's view," Photogrammetric Engineering and Remote Sensing, vol. 48, n. 8, pp. 1309-1312, August 1982.

Aronoff, S. "Classification accuracy: a user approach," Photogrammetric Engineering and Remote Sensing, vol. 48, n. 8, pp. 1299-1307, August 1982.

Cooperative of Coffee Planters of Três Pontas (COCATREL), July 1999.

Ginevan, M. "Testing land use map accuracy: another look," Photogrammetric Engineering and Remote Sensing, vol. 45, n. 10, pp. 1371-1377, 1979.

Ma, Z., Redmond, R. L. "Tau coefficients for accuracy assessment of classification of remote sensing data," Photogrammetric Engineering and Remote Sensing, vol. 61, n. 4, pp. 435-439, April 1995.

Valeriano, D. M. "Processamento digital de dados do MSS-Landsat aplicado ao mapeamento da cobertura da terra da planície costeira do Rio Tubarão, SC: metodologia e estimativa de exatidão de classificação," MSc. Thesis, INPE-3455-TDL/188, 87 p., São José dos Campos, March 1985.