COMPARISON OF SAMPLING METHODS TO CLASSIFICATION OF REMOTELY SENSED IMAGES.

Rômulo Parma Gonçalves[1], Leonardo Campos Assis[2], Carlos Antônio Oliveira Vieira[3]

IV Simpósio Internacional de Agricultura de Precisão, 23 a 25 de outubro de 2007, Viçosa-MG.

ABSTRACT: *Due to the influence of the sampling schemes to the final quality of a classification product, the present work purposes a qualitative method aggregated to the quantitative traditional approach on determining the general quality degree of a classification procedure based on non-spatial statistics parameters like Kappa coefficient. This consists of a binary error image in which incorrectly labelled pixels are evidenced. In addition, binary images from different sampling schemes permitted to know a frequency imagery which provided the location occurrence of erroneous pixels in percentiles. The statistical Z test values consented to verify significant differences on a pair-comparison of the sample methods resulting insignificant difference between them. Qualitative and quantitative analysis are complementary evaluating classification procedure of remotely sensed images.*

KEYWORDS: sampling schemes, classification, binary error image.

INTRODUCTION: Remotely sensed imagery has been applied on a large number of areas for various human applications. Conversion of spectral information on thematic information is usually referred to as imagery classification, in which spectral pattern recognition procedure of the spatial imagery produce thematic maps. According to Lillessand e Kiefer (1999), the classification has the objective to categorize all pixels of a determined imagery, assigning a label (informational class) to each of them through computer procedures (Mather, 2004).

Performing spatial data analysis on data of unknown accuracy will result in a product with low reliability. The data quality is a function of its inherent and its intended use (Vieira, 2001). Common statistics used to validate remotely sensed data include overall, individual class, user and producer accuracy, Kappa coefficient and others.

Vieira (2000) reports about the importance of the sampling strategy applied to statistical analysis. The most commonly used sampling schemes are cluster sample, simple random sampling, stratified random sampling and systematic sampling.

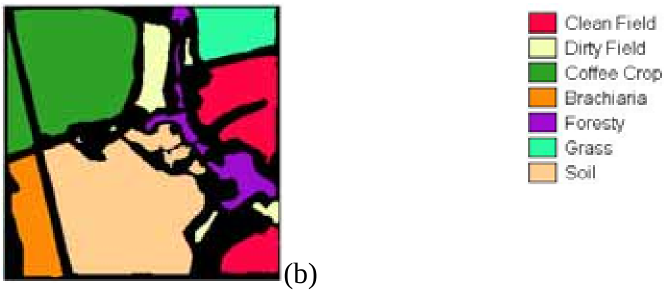
This work proposes to compare four methods of feature extraction which better presents spectral classes, using the maximum likelihood classification method and evaluating the spatial error behaviour. A frequency image of error pixels make this possible. It also aims to analyse the statistical differences between each sampling method and to verify if there is, or not, superiority of one of them applying the statistical *Z* test to the kappa coefficient value and its variance.

METHODOLOGY: The experimental area is localized on Serra do Salitre (MG), Brazil. A Ikonos image (Figure 1(a)) consisting of three bands (2, 3 and 4) that presents spectral classes diversity (seven ones) to perform the supervised classification approach. Between the various steps through the classification process, the first is the definition and representation of the object to identification of

[1] Survey Engineer, Post graduating in Spatial Information - UFV, Substitute Professor of the Civil Engineering Department, UFV. E-mail: romuloparma@yahoo.com.br
[2] Bachelor in Information Systems, Post graduating in Spatial Information – UFV. E-mail: assis_leonardo@yahoo.com.br
[3] Survey Engineer, M.Sc, Ph.D. P.Ph.D., Associate Professor of the Civil Engineering Department, UFV. E-mail: carlos.vieira@ufv.br

different informational classes which compose the whole area, with one reference image (Figure 1(b)) (Eastman, 2003).


(a)



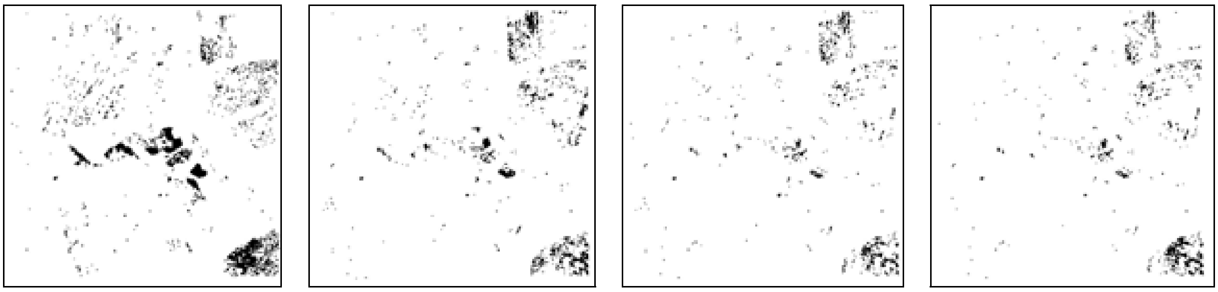
(b)

Figure 1: Three bands false color composition image (a); thematic reference image of the study area (b).

Following the feature subset extraction procedure which better distinguishes the object, that is synonymous of sample training sites, the sample collect can be doing based on various choice criterion, as the number of pixels as the number of samples by class, until the spatial positioning of them in the image. The ideal number of samples depends on the spatial variation of the area and the number of classes, therefore, is recommended at least 5 to 10 sample (Gruijter, 1999).
In relation to the pixel sampling methods, which is the focus of this work, four common ones was selected, the cluster, the simple random, the stratified random, and the systematic (Gruijter, 1999).
Validity of statistical estimates depends upon the number of variables (spectral bands) and properties, whose statistical parameters are to be estimated, and the representativeness of the sample (Mather, 2004).
The fundamentals of the clustered sampling is based on statistical characterization of the reflectance on each informational class data (signature analysis) draw from know examples digitised on screen in the reference image (training sites) (Eastman, 2003).
Simple random sample method has no restrictions on the randomization. This means that all sample points are selected with equal probability and independently from each other.
The stratified random sampling is restricted on randomization. It works dividing areas in sub-areas called 'strata', and, in each of which the simple random sampling method is applied with sample sizes chosen beforehand.
Systematic sampling method is a regular square grid of points that require the total number of samples to have an integer square root. In this study, this number was six hundred and twenty-five (closest to six hundred and thirty of the other methods).
The decision rule choose was the maximum likelihood algorithm (maxlike). This method takes into account that the frequency distribution of the class membership can be approximated by the multivariate normal (Gaussian) probability distribution with a density probability function (Mather, 2004).
After the classification procedure, the next step aims to evaluate the quality of the generated products determining the degree of error (pixels incorrectly labelled) associated with them. This requires a new collection of testing samples, where about a half of the number of points used for training samples. The sample subsets to test of the classified images originated from the methods are: simple random, stratified random and systematic. Hence, the classification by the cluster method was also evaluated using one of the three method earlier mentioned instead a cluster sample site.
In order to evaluate the classification procedure, 324 points were generated for each of the methods. In sequence, these pixels were overlaid with the reference image and the classified images (by maxlike algorithm). The results produced the Kappa coefficient value.
Before assessing the quantitative aspects of the error, is interesting to observe its qualitative aspects. One manner to achieve it, is through the generation of a Binary or Boolean Error Image obtained by the direct cross comparison between the reference image and the images produced from classification. In this image, for each pixel only two possible values can be assigned, 0 (correctly labelled) or 1(incorrectly labelled). It was made four Binary Error Images, one for each sampling method. From these methods, could be generated an image that presents the number of occurrence of the mistake classified pixels, since in these new image, the comportment this pixels can be viewed in terms of frequency of occurrence.

Mather (2004), reports that the most common used method to represent the degree of accuracy of a classification is to build an error matrix. .This matrix provides elements used to obtain the Overall and Kappa coefficient of accuracy. The Kappa coefficient (*Ka*) is a suitable measure of the accuracy of a thematic classification because contemplates the full confusion matrix (Congalton, 1991).
Moreover, Vieira (2000) suggests other test to determine the significant difference between independent Kappa coefficient values, and therefore, between confusion matrices. The test makes possible statistically comparison of two different algorithms or methods. The determination of the normal distributed Z value is obtained by the ratio among the difference value of two distinct Kappa coefficients and the difference of the respective variance of them (Skidmore, 1999).

RESULTS: The products (images) originated from the classification algorithm for each method were crossed against the reference image. Thus, it permitted to generate an image with all the possible combination of the classes present on the both images. From the crossed images was possible to obtain the binary error images, which described the spatial behaviour and positional location of the error pixels (incorrectly labelled). These images (Figure 2) are presented all together to permit a better comparison.



(a) (b) (c) (d)

Figure 2: Illustrations of four Binary Error Images corresponding to the incoherencies of the maximum likelihood classification algorithm for each scheme of sampling: cluster (a), simple random (b), stratified random (c) and systematic (d).

The methods analysed generate an image that shows frequency in percentiles of error caused by the mistake classifieds pixels. Combining the binary error images, one can verify (Figure 3) that the pixels labelled with 25% implies that it has one occurrence in the image, with 50% it has two occurrences, with 75%, three occurrences and with 100%, indicates that the pixels was wrongly labelled for all the methods.

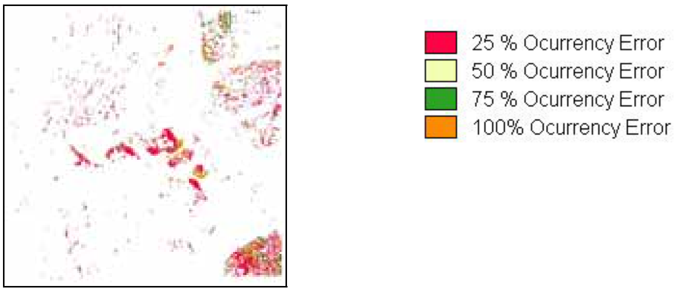


Figure 3: Illustration of erroneous pixels frequency in percentiles.

With the purpose to present the accuracy assessment obtained, the Kappa coefficient of the classified data and the test data were disposal on the Table 1.

Table 1: Kappa coefficient values of the classified data and the test data.

	Cluster	Simple Random	Stratified Random	Systematic
$\hat{K}$	0.9127	0.9495	0.9650	0.9638
$\hat{K}(test)$	0.9157	0.9492	0.9620	0.9474

Based on Table 1 data, there is no statistical difference between the Kappa coefficient values, where $\hat{K}(test)$ is the Kappa values obtained for test images and $\hat{K}$ is the Kappa value obtained for the classified images, in other words, the classification is valid for all methods.

After the validation of the effectiveness of the classification, the next step consisted to compare statistically the four methods at a significance level. The values of Kappa coefficient and their variance applied to estimate the *Z* values. These values are shown in the Table 2.

Table 2: Values of Kappa coefficient and variance of Kappa for each sample method

Methods	$\hat{K}(test)$	$\sigma^2(\hat{K})$
Cluster	0.9157	0.000437
Simple Random	0.9492	0.000272
Stratified Random	0.9620	0.000231
Systematic	0.9474	0.000291

Assuming for the *Z* test, given the null hypothesis $H_0 : K_1 = 0$, and the alternative hypothesis $H_1 : K_1 \neq 0$, the *Ho* hypothesis is rejected if *Z* value is grater or equal to *1.64* of the two-tailed *Z* test and the degrees of freedom are assumed to be infinity.

At the *90.0* confidence level, and the value of the test *Z* statistic is grater than *1.64*, the result is significant for all the evaluated methods. The Table 3 data shows the Z values estimated.

Table 3: *Z* values of the significant difference between sample methods

Methods	Simple Random	Stratified Random	Systematic
Cluster	1.258119	0.682582	1.174880
Simple Random		0.570724	0.075861
Stratified Random			0.639025

CONCLUSION: Assuming the presented analyses conditions and data available, over a qualitative and quantitative approaches, the results permit to conclude that even with a good classification supported on the statistical Kappa coefficient value, it could be still have subsets into image with high incoherence. Such incoherence can be verified through binary error imagery efficiently.

All the analysed sample methods used in the present work shown the equivalence on the statistical answers provided by the Z test at a ninety percent confidence level.

REFERENCES

CONGALTON, R. G. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment, 49(12), p.1671-1678. 1991

EASTMAN, J.R. Idrisi Kilimanjaro Guide to Gis and Image Processing, Clark Labs. Clark University. Worcester, MA. USA. 306 p. 2003

GRUIJTER, J. Spatial sampling schemes for remote sensing. In: A. STEIN et. al. (eds.), Spatial statistics for remote sensing. p 57-81. Netherlands: Kluwer Academic Publishers. 1999.

LILLESSAND, T. M.; KIEFER, R. W. Remote sensing and image interpretation. 4 Ed. New York: John Wiley & Sons, 724 p. 1999.

MATHER, P. M. Computer processing of remotely-sensed images: An introduction. 3 Ed. Chichester: John Wiley & Sons, 324 p, 2004.

SKIDMORE, A. K. Accuracy assessment of spatial information. In: A. STEIN et. al. (eds.), Spatial statistics for remote sensing. p 197-209. Netherlands: Kluwer Academic Publishers. 1999.

VIEIRA, C. A. O. Accuracy of remotely sensing classification of agricultural crops: a comparative study. 327p. Thesis (Doctor of Philosophy) – University of Nottingham, Nottingham. 2000.

VIEIRA, C. A. O; MATHER, P. M. Validation of remotely sensed data. GIM International Feature Magazine, February. 2001.