Anais XIV Simpósio Brasileiro de Sensoriamento Remoto, Natal, Brasil, 25-30 abril 2009, INPE, p. 6773-6780.

Analysis of sampling methods and its influence on image classification process of remotely sensed images through a qualitative approach

Leonardo Campos de Assis [1]
Rômulo Parma Gonçalves [2]
Carlos Antônio Oliveira Vieira [3]
Nerilson Terra Santos [4]

[1, 3, 4] Federal University of Viçosa
CEP 36570-000 - Viçosa - MG, Brazil
[1] leonardo.assis@ufv.br
[3] carlos.vieira@ufv.br
[4] nsantos@ufv.br

[2] Federal University of Alagoas
CEP 57072-970 - Maceió - AL, Brazil
romuloparma@yahoo.com.br

Abstract. The aim of this paper was to perform analysis due to the influence of the sampling schemes to the final quality of a classification product, through a purpose of 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 labeled pixels are evidenced. In addition, a new approach to determine Clustered Area Ratio imagery were generated permitting visualize the location of error pixels and quantify its grouping behavior under assumption of a normal Z distribution. Also was performed a statistical Z test consented to verify significant differences on a pair-comparison of the sample methods resulting insignificant difference between them. The conclusion about qualitative and quantitative analysis was that they are complementary in evaluating classification procedure of remotely sensed images.

Keywords: sampling schemes design; binary error imagery; error autocorrelation; clustered area ratio.

1. Introduction

Remotely sensed imagery has been applied on a large number of areas for various human applications, like precision agriculture, geology, control of burned areas, etc. 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 and 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.

The process of classification consists of two stages, the first is the recognition of categories of real-world objects, like water bodies, land cover types, etc., and the second step is the labeling of the entities to be classified (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 parameters 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 design are cluster sample, simple random sampling, stratified random sampling and systematic sampling.

This work proposes two main objectives. According to Gonçalves *et al.* (2007), the first is to compare four methods of feature extraction which better presents spectral classes, using the maximum likelihood classification method and evaluating the spatial error behaviour. In addition, a new method to verify spatial arrangement of pixel based on Binary Error Imagery and its degree of grouping is evaluated.

2. Methodology

The experimental area is localized at Serra do Salitre (MG), Brazil, with approximated geodetic coordinates, latitude = $19^{\circ}38'S$ and longitude = $46^{\circ}40'W$. A Ikonos image 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 different informational classes which compose the whole area, with one reference image. (Eastman, 2003).

2.1. Comparison of sampling methods

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.

To proceed with the signature feature extraction, the minimum number of pixel samples by class can be obtained by the equation (1), (Mather, 2004):

$$30 * NDV * NCl \tag{1}$$

Where, *NVD* is the number of discriminate variables and *NCl* is the number of classes. Considering the present study case, applying the Equation (1) ninety pixels were selected by class with seven informational distinct classes, which sums six hundred and thirty sampled pixels for the whole 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.

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).

The advantage of this method is when only one cluster is selected, the clusters should be pre-defined such that each of them covers the area as good as possible. And the disadvantage is because this type of designing does not produce any random repetition, no unbiased estimate of the sampling variance is available (Gruijter, 1999).

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 chosen 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, it could be generated an image presenting the occurrence of the mistaken classified pixels, since in these new imagery, the behavior of these pixels can be viewed in terms of their arrangement (cluster, disperse or random).

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. This test makes possible statistical 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). Accomplishing the analysis of the statistical differences between each sampling method, by applying the statistical Z test to the kappa coefficient value and its variance, a quantitative Z image is produced to visualize the grouped/ungrouped areas.

2.2. Error autocorrelation

Binary Error Imagery only describes the position of the wrongly labeled pixels, and one could argue that the Join Count Statistics could be applied to show the configuration of their arrangement. Instead of this, the method described here presents an alternative form to visualize and quantify where the error appears. As a matter of fact, the knowledge about the position of which pixels were wrongly labeled allows determining the type of autocorrelation that they occurred. The idea to explore a new approach was due to the instability of Moran's coefficient on handling large number of units (pixels).

In this context, one hypothesis, complementary, seems more interesting to be done. That would be about how to know which regions wrongly labeled were less reliable. The present work make the assumption that: maybe, by assigning different weights to the clustered areas, could bring good answers. The first step is to individualize the "true" pixels (those valued as one), or group of pixels, of the Binary Error Imagery by assigning a unique numeric value identification *(Id)* to them. In order to compute the group pixel arrangement identification, a better choice is to consider only the first-lag pixel value, left, right, above, below and diagonals than to include only the diagonals units (Hook's case). Once the identification is done, the next step is to compute the number of cells at each *Id*. Resulting image is almost identical to the Id images (individual or grouped pixels), however, the number of cells of each Id is used to replace the Id value itself. By extracting mean and standard deviation descriptive statistics from these images, in a next step, the assumption about a *Z* distribution under normal curve is made. In a per pixel operation, calculus is done to determine the Clustered Area Ratio (CAR) image and Z_{CAR} image by the following equations (2 and 3):

$$CAR = \frac{Idc_i}{\sum_{i=1}^{n} Idc_i} \tag{2}$$

$$Z_{CAR} = \frac{CAR - m_{CAR}}{\sigma_{CAR}} \tag{3}$$

where *Id*c is the number of cells of *ith* Id, which values vary from one to *n*. m_{CAR} and σ_{CAR} are the average and standard deviation of CAR, respectively. The idea of CAR is, the lower the value, trending to zero, more dispersed is the arrangement. Otherwise, the higher the value, near to one, more clustered it is. Values much closer to zero could be interpreted as a disperse arrangement. Comparing to Binary Error Imagery, the advantage of this approach is that, once the CAR values are determined on a per pixel operation, the image produced presents *"where"* are the critical areas of erroneous clustered pixels instead a simple value indicating if the arrangement is or not clustered. Image Z_{CAR} is an image where pixels values corresponding to *Z* values. Considering the analogous assumption of this approach to the join count statistics (Ebdon, 1985), the following considerations about hypothesis testing must be made:

1) The null hypothesis:
    a. Free sampling
    b. Non-free sampling;
2) The alternative hypothesis:
    a. Directional
    b. Non-directional

In order to determine what is the null hypothesis tested, a decision about to consider that the reference image is from all the evaluated area (non-free sampling) or if it is only a portion (or sample) of a larger area (free-sampling). Following, the null hypothesis frequently relates to randomness arrangement condition. So, the other assumption to be made is about the alternative hypothesis, being directional or non-directional. A directional one will be able to specify whether the arrangement is thought to be clustered or disperse. A non-directional will point only if the arrangement is random or not. One tailed test evaluates the directional alternative hypothesis and a two-tailed tests the non-directional.

To visualize the comparison of the methods, a normalization process between the individual CAR imagery must be done, considering all scenarios together. This can is obtained by the estimative of mean and standard deviation parameters of a CAR image for the combinatory scenario (an image about the wrongly labeled pixels by all the methods analyzed). These parameters are employed by the equation (4):

$$NCAR_m = \frac{CAR_m - M_{CAR}}{\Sigma_{CAR}} \tag{4}$$

where $NCAR_m$ is the normalized clustered area ratio image for each method (*m*), considering the CAR_m image of the respective method, the mean (M_{CAR}) and the standard deviation (Σ_{CAR}) of the combinatory scenario (all methods). This imagery (*NCARm*) produced is used as visual approach to a comparative analysis between methods.

3. Results

The products (imagery) 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 imagery, which described the spatial positional location of the error pixels. These imagery (Figure 1) are presented all together to permit a better comparison.

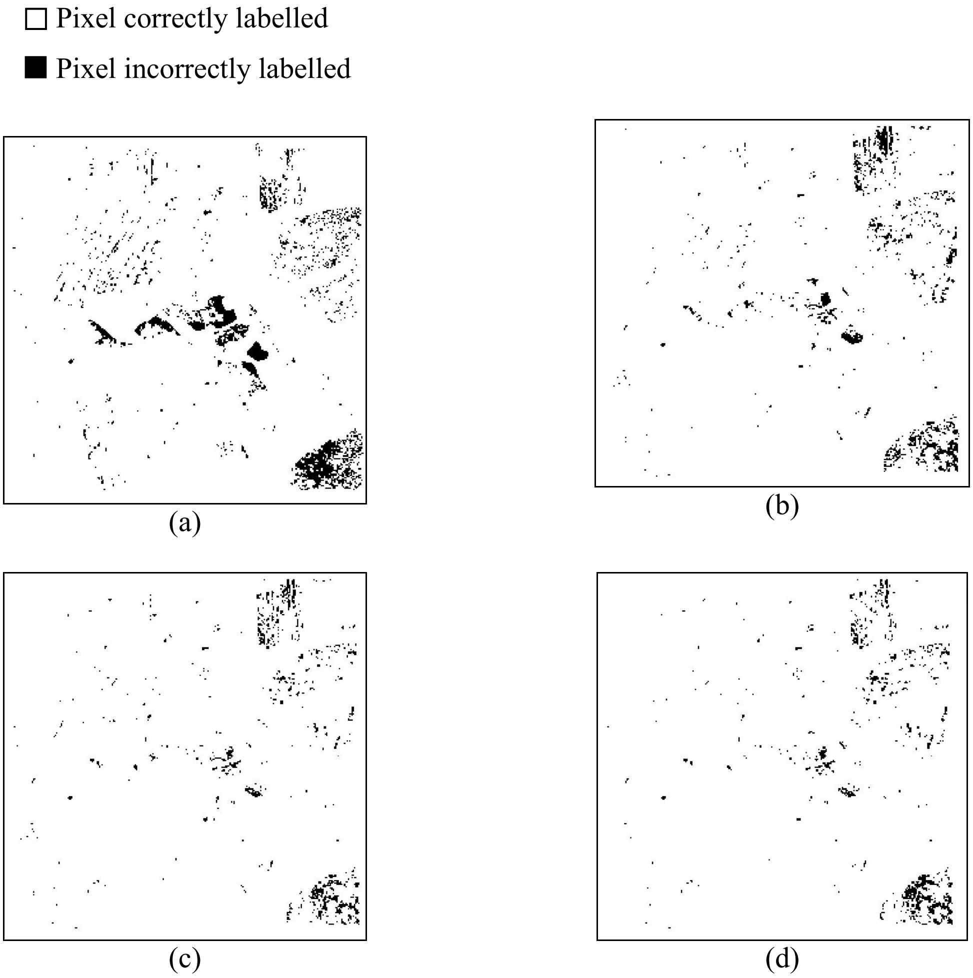

Figure 1. Illustrations of four Binary Error Images corresponding to the incoherencies of the maximum likelihood classification algorithm for each scheme of sampling: cluster (a), simple random (b), stratified random (c) and systematic (d).

Once the evaluated area is only a portion of a large one, the null hypothesis assumed is about free-sampling, and alternative hypothesis is non-directional. Figure 2 presents NCAR imagery, visually evidenced the degree of clustering that each sampling method produced. By the referred figure illustration of each method and respective legend values, stratified random exhibited on letter (c) presents a better (less clustering, or random) distribution scenario, followed by simple random (b). Systematic and cluster sampling schemes produced the two worst results.

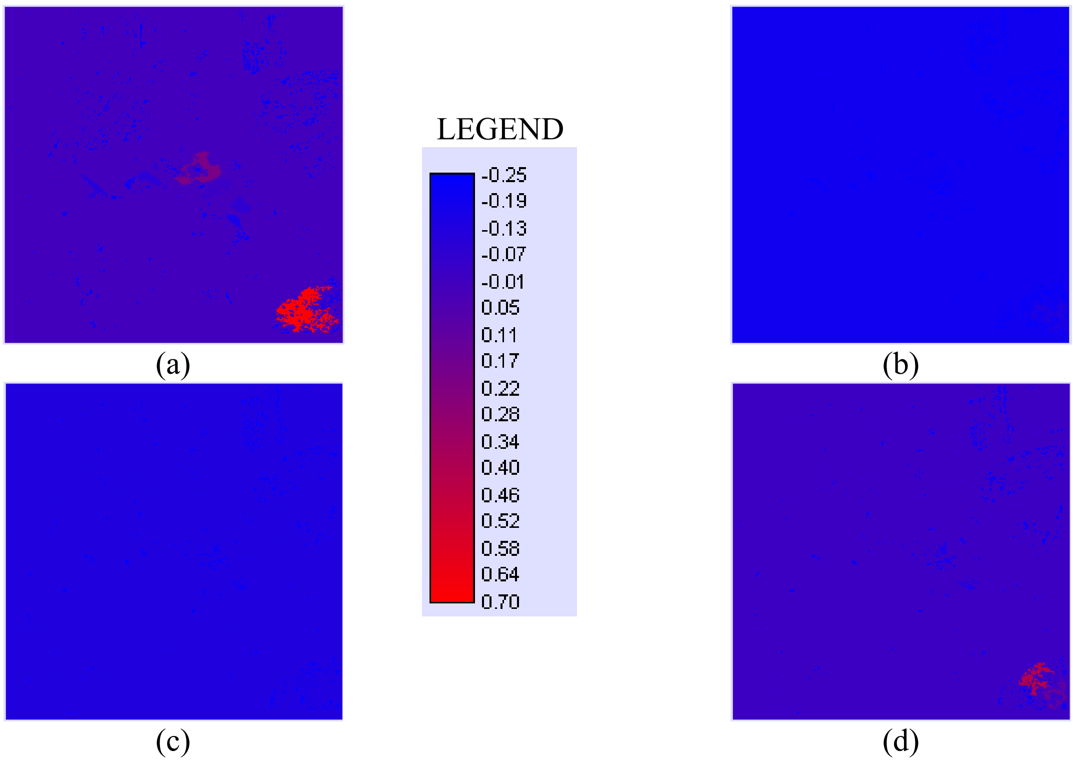

Figure 2. NCAR imagery derived from Binary Error Images representing each sampling method: clustered (a), simple random (b), stratified random (c) and systematic (d).

The illustration of the Figure 3 presents Z_{CAR} imagery under 3 significance levels: 0.1,0.05, and lower, for each sampling method being evaluated. This quantity measures the grouping pixels spatially distributed and presents the degree of arrangement of the pixels (error pixels in this study)

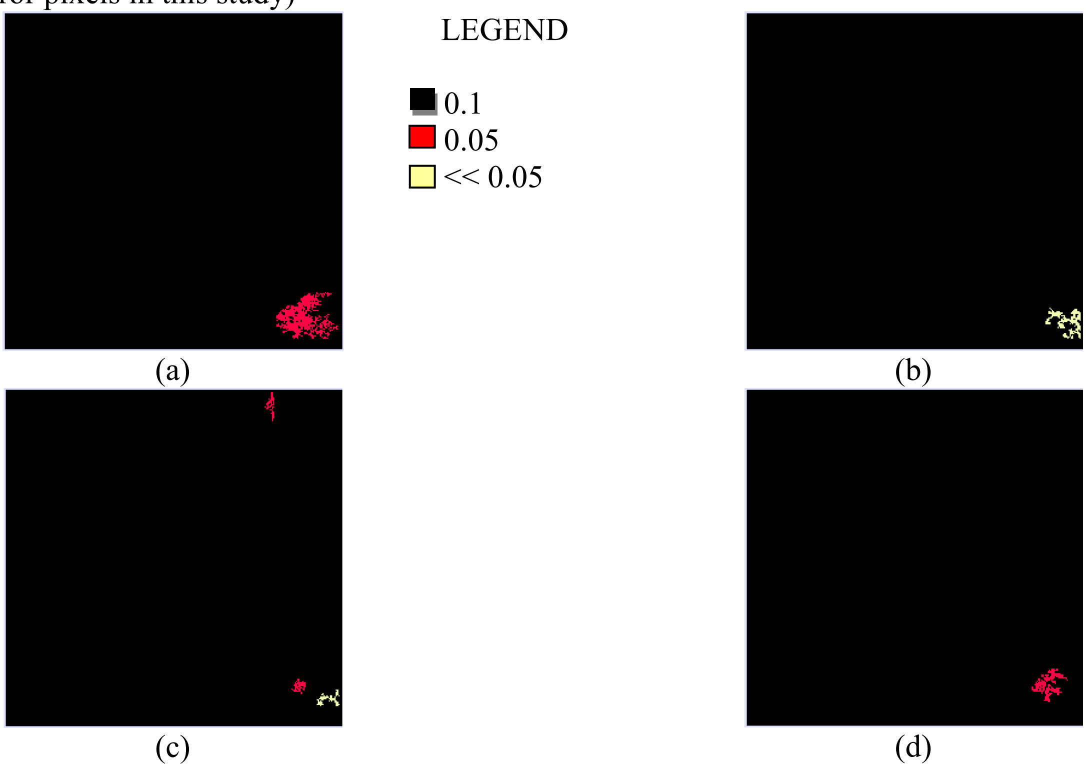

Figure 3. Z_{CAR} imagery presenting Z test spatially applied at three significance levels, 0.1, 0.05 and lower, from CAR imagery.

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

4. Conclusion

Considering the presented analyses conditions, the data set available and hypothesis assumptions, over 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. From these Binary Error Imagery, another approach to deal with the pixels arrangement was presented. Despite the equivalence provided by the Z test at a ninety percent confidence level between sample methods, through the CAR imagery methodology employed, it could be verified that stratified random sampling method presented better results considering a random arrangement of error instead a grouping one. It also was able to determine Z test imagery, permitting to visualize different grouping scenarios quantitatively determined. It could be concluded that Z_{CAR} imagery is an alternative purpose to verify spatial arrangements of clustered areas yonder the traditional methods like Moran's coefficient and Join Count Statistics, where these last methods do not fits well (on large *n* analysis).

Acknowledgments

The authors want to thanks Prof. Vicente Paulo Soares at Forestry Department in Federal University of Viçosa for the data set imagery supplied. They also are grateful the financial support obtained from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

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.

Ebdon, D. **Statistics in geography**. 2nd. Ed. Malden: Blackwell Publishers., 232 p. 1985. ISBN 0-631-13688-6.

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

Gonçalves, R.P., Assis, L.C., Vieira, C.A.O. Comparison of sampling methods to classification of remotely sensed images. **IV Simpósio Internacional de Agricultura de Precisão.** 2007.

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.