INTRODUCING AN ACCURACY INDICATOR BASED ON UNCERTAINTY RELATED MEASURES

S. B Fatemi [a] *, B. Mojaradi [a], M. Varshosaz [b]

[a] K.N.Toosi University, Geodesy and Geomatics Faculty, Valiasr Tehran, Iran - (sbfatemi, mojaradi)@yahoo.com
[b] . K.N.Toosi University, Geodesy and Geomatics Faculty, Valiasr Tehran, Iran – varshosazm@kntu.ac.ir

KEY WORDS: Analysis, Accuracy, Classification, Error, Indicators, Remote Sensing, Simulation

ABSTRACT:

Traditionally accuracy assessment of the classification results uses some collected reference data (ground truth). Ground truth collection is a time-consuming and money-swallowing activity and usually can not be done completely. Uncertainty is an important subject in remote sensing that can appear and be increased sequentially in a chain of remote sensing from data acquisition, geometric and radiometric processing to the information extraction. Conceptually the relation between uncertainty and accuracy is an inverse relation. This relation can aid us to construct a relation between accuracy measures and uncertainty related measures. In this paper we investigate this relation using the generated synthetic images (for the sake of the reliability of the obtained results) and try to find an uncertainty related measure that has a strong relationship with the accuracy parameters like overall accuracy.We have found that among the uncertainty measures the mean quadratic score has the strong and reliable relationship with the commonly used accuracy measures. This relationship can be a good basis for the future investigations that lead to the classification based accuracy measures and avoiding some problematic data related issued of ground truth data collection.

1. INTRODUCTION

Uncertainty is an important subject in remote sensing which has recently attracted the attention of many researchers. It can appear in a chain of remote sensing from data acquisition, geometric and radiometric processing to information extraction with its value increasing sequentially during image processing and image analysis. A thematic map produced by a different approach and various satellite images must be reliable to be used in GIS. Therefore the source causing uncertainty to increase must be defined and modeled or removed. Having extracted any information from satellite imagery, the presentation of uncertainty as an indicator is essential for users and it is important to define a measure to quantify uncertainty.
Goodchild (1995) argues that uncertainty is "generic and reasonably value-free, and implies nothing about sources or whether they can be corrected ". Stephanou and Sage (1987) said "uncertainty indicates lack of knowledge and is a concept to express the inability to be confident of and knowledgeable about the truth value of a particular data". The generic meaning of uncertainty implies that is if two individuals give the same answer to a question, one might be more certain than the other. A simple definition of uncertainty can then be "*the probability of error*".

In this research it we will try to answer the following questions:

- Is there any clean and formular relation between the uncertainty and the overall accuracy?
- If yes, is this relation independent from the source of uncertainty?
- Which uncertainty measure is more stable for showing accuracy?
- Which uncertainty measure isn't sensitive to the source of accuracy?
- Can we define a shift and a scale factor for the uncertainty of the overall accuracy?

2. UNCERTAINTY MEASURES

Regarding the main concept of the research, we need to investigate possible uncertainty measures. Based on the information theory an information source from set of symbols $\{a_1,a_2,..a_n\}$ generates a random of symbols . The probability of the event a_j that the source will produce is $P(a_j)$ and

$$\sum_{i=1}^{n} P(a_j) = 1 \qquad (1)$$

$$I(a_j)=Log(1/P(a_j))=-Log(P(a_j)) \qquad (2)$$

The amount of self information attributed to event a_j is inversely related to the probability of a_j .The base of the logarithm in equation (2) determines the unit used to measure information.

If k source symbols are generated, the low of large numbers stipulates that, for a sufficiently large value of k, symbol a_j will (on average) be output $k*p(a_j)$. Thus the average self information obtained from k outputs is:

$$-k\,P(a_1)\,Log\,P(a_1)-k\,P(a_2)L\,og\,P(a_2)-\ldots..-k\,P(a_n)\,Log\,P(a_n) = -K\sum_{i=1}^{n} P(a_i)LogP(a_i) \qquad (3)$$

In image classification for a pixel, viewed as a statistical variable C, the uncertainty in class C_i is defined as:

$$-Log_2^{P(C=C_i/X)} \tag{4}$$

for $i = 1,...,n$, where X denotes the available data; the uncertainty is measured in bits. Generally, the true class of the pixel is not known and, as a consequence, the amount of information required revealing the pixel's class is unknown. The entropy of the pixel is therefore defined as the expected information content of a piece of information that would reveal its true class. To this end, the entropy measure combines the uncertainties in the various classes of the pixel by weighting them by their probabilities:

$$-\sum_{i=1}^{n} P(C = C_i / X) * Log_2^{P(C = C_i / X)} \tag{5}$$

As another measure of weighted uncertainty, the *quadratic score* (Glasziou and Hilden, 1989) is briefly discussed here. The quadratic score is built on the notion of confirmation. The uncertainty in a single class for a pixel is the amount of *probability* required to establish this class with complete accuracy. The uncertainty in class Ci isdefined as 1-P(C=C_i/X), where X once more denotes the available data. The quadratic score of the pixel is then:

$$\text{QS} = \sum_{i=1}^{n} (1 - P(C = C_i / X)) * P(C = C_i / X) \tag{6}$$

This measure exhibits the same behavior in its minimum and maximum values as does the entropy measure. The two measures differ, however, in their slopes as is shown in Figure(1). The slope of the entropy measure is steeper than the slope of the quadratic score. As a result, the entropy measure for example more strongly weighs small deviations from probabilities equal to zero or one than the quadratic score.

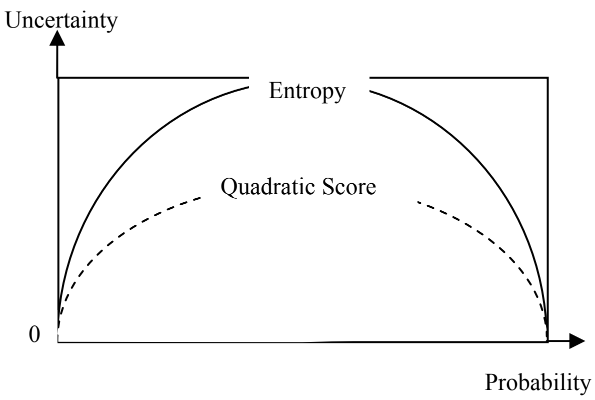

Figure 1. Relation between quadratic score and entropy

As can be seen in Figure(1) we can argue that when entropy is increased we have a lot of chaos information (many objects are in a given pixel) and we are not certain about the labeling (class) of the pixel. This means we have considerable radiometric overlap between classes and vice versa. When uncertainty is decreased we have less chaos and we are more certain about labeling of desired pixels. Therefore radiometric overlap between classes is low; as a result they are separated from each other.

We can use this concept to design a relation between uncertainty and accuracy of classified images. Thus as will be shown (section 3.2) the amount of uncertainty is a good indicator to investigate the accuracy of a map. Traditional approaches for accuracy assessment of thematic maps use ground truth. However, usually this ground truth is usually inherently unreliable. Hence, it is not a good idea to compare the extracted information (with a specific level of uncertainty) with a reference data set that is uncertain itself.

Ground truth could be non-representative (i.e. only partly covering the general characteristics of a particular land cover class), insufficient, incomplete (overlooked classes) or even outdated and thus lay an unstable foundation for accuracy assessment. Additionally the collection of this data is often a time-consuming and money-swallowing activity which in order to get rid of which, it is simply replaced by a visual inspection of some cartographic document or the image itself.

3. TESTS

Regarding the mentioned questions in the section 2, we have investigated the inverse relation between uncertainty and accuracy. To this end we have produced some synthetic images and (using some well known ground truth) and have classified them. Finally some accuracy and uncertainty related measures (URMs) have been calculated. Relation between these parameters is the major theme of the experiment.

3.1 Generation of the Synthetic Images

In this case study some synthetic images are used generated by a simple algorithm. For each image 3 spectral bands have been generated. Firstly in order to simulate the imaging process and generation of these bands in each case, we generate a ground truth map. This is used to generate the spectral bands of the synthetic images and in addition to evaluate the actual accuracy of the classification results. The general ground truth map has 10 spectral classes with the various radiometric overlaps between them. This ground truth map can be generated automatically or manually. In this case study this map has been generated manually and regarding the real world it was tried to include various shapes of the possible objects [Figure 2.A].

It was assumed that the statistical distribution of the image data (pixel values) is a multi dimensional normal distribution. This assumption doesn't affect the final results and just simplify the band generation and avoid the wrong assumption of the distribution of the data that is used in the maximum likelihood (MLH) classification. For generation of the images we have to consider some values for mean and variance vectors. Therefore we have a mean and variance value for each class per band (totally 30 values for means and 30 values for the variances). Covariances between all of the bands were assumed to be zero for the sake of simplicity and the little effect of them.

Regarding these assumptions, spectral bands for the various case studies presented in this paper were generated. Algorithm firstly considers the class number of the pixels gathered from the ground truth map. Then it searches the mean and variance matrices and selects the corresponding values for mean and variance vectors considering the class number. After that using these signatures the pixel value per band is generated using a function that returns a vector of random numbers having the normal distribution, this algorithm repeats for each pixel until all of the pixels have their appropriate values in the 3 bands. Figure 2 shows a sample ground truth map and generated color composite image.

3.2 Investigating Relationship between Accuracy and Uncertainty

The mentioned algorithm (Section 3.1) was used to produce the desired images to perform the experiments on them. Some constraints and conditions were applied on the all of the experiments. Size all of the images is 512 × 512 pixels and have been generated using the algorithm explained at section 3.1. For the classification of the images it was decided to use maximum likelihood classification because of the relative powerful ability to classify images, also this method is available at the most image processing softwares. Additionally the results of this method are per pixel probabilities and labels which permit us to evaluate and calculate pixel by pixel quadratic score and accuracy values. All of the cases were done based on the equal prior probability assumption of the classes.

As Masseli and et.al (1994) have noted and the authors have investigated the mean of entropy values has not a straightforward and certain relationship with accuracy measures. Therefore we choose the mean quadratic score (MQS) which shows a strong linear regression between the overall accuracy (OA) and kappa coefficient (K) (as the accuracy measures) and this uncertainty measure.

In order to show the strong (inverse) relationship between classification uncertainty related measure and accuracy of the classification some images were generated by changing the radiometric overlap between the various classes. This was done simply by altering the mean and variance values. When two classes have more similar values then radiometric overlap between them also increased accordingly. The closer mean vectors the higher radiometric overlap. Also using the large values for the variances can lead to the more radiometric overlap.

Regarding this logical assumption the mean and variance values of the classes were changed 11 times and then 11 data sets were obtained. Having applied the maximum likelihood classification on the data sets; 11 overall accuracies and corresponding mean quadratic scores and kappa coefficients were calculated. Figure 3 shows the approximately linear relationship between overall accuracies and kappa coefficient and corresponding mean quadratic scores.

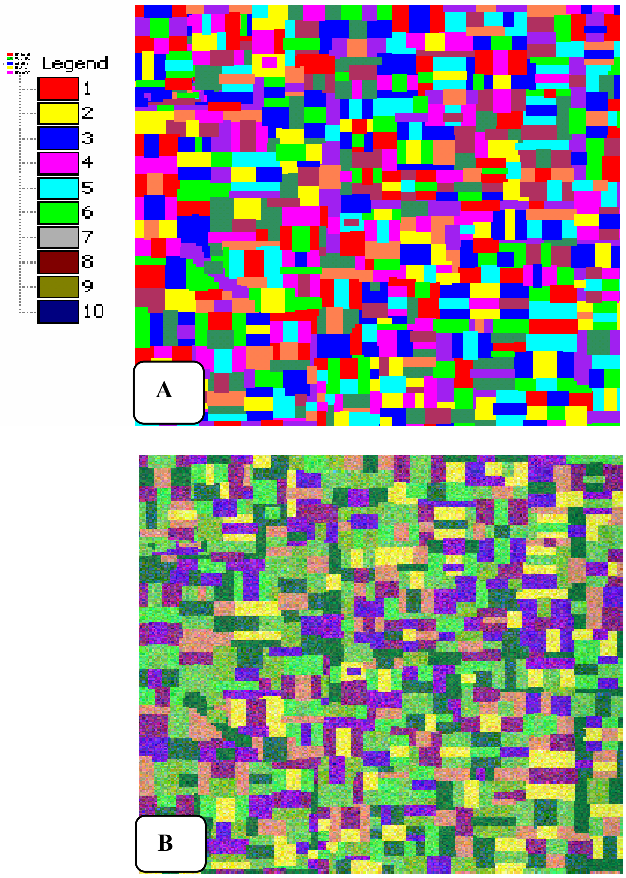

Figure 2. The generated ground truth map (A) and the corresponding sample synthetic image (B).

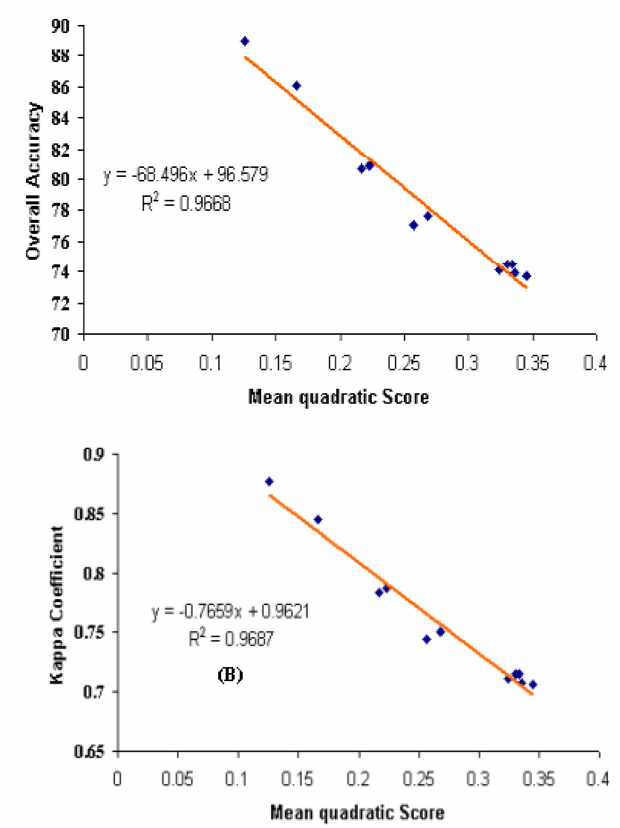

Figure 3. Relationship between MQS and OA (A) and relationship between QMS and K(B). See the strong linear relationship between accuracy and uncertainty parameters.

This relationship can be seen between various data sets and have been investigated by the authors. However the obtained regression formula changes for various cases but the strong linear relationship between MQ and accuracy measures is preserved. This is the predicted relationship at previous section and confirms the mentioned relation between uncertainty and accuracy.

4. DISCUSSION

The obtained linear relationship between mean quadratic score and overall accuracy is a good representation of the famous inverse relation between accuracy and uncertainty. In the other hand this relation can help us to guess the probable value of the accuracy parameters (which need to collect some reference data) without any field observations. Although we can't define the exact value of these parameters but the estimated approximate values will be near the real values to some extent. This is caused from this fact that some environmental and procedural parameters influence the estimated linear relationship and then in some cases slop and intercept of the regression line can be the other values.

Generally all of the factors involved in the accuracy assessment process can affect the final estimated accuracy values and therefore it may compute some different values for a given classification. Some of these factors have been listed by Congalton (1991) as: ground data collection, classification schema, spatial autocorrelation, sample size and sampling scheme.

Two major aspects of the ground data collection are sampling schema, and sample size. These two issues affect the overall accuracy estimation and therefore can lead to bias estimation of the accuracy. Thus if we change any parameter that have a major influence on the estimated accuracy (e.g. sampling schema); we have a different value for accuracy and therefore MQS can not have a fixed relationship with the all of these different accuracies that are for a particular classification. Considering this problem we investigated the relation between the accuracy and the sample size and concluded that a sample size between 70-100 pixels per class can lead to a reliable accuracy assessment. However, generally this depends on some environmental aspects [Congalton, 1991].

Sampling scheme also can have a notable effect on the accuracy assessment. Congalton (1988) notes that it is the spatial complexity of a given environment which dictates the appropriate sampling scheme(s) to be used for creating error matrices necessary to assess the accuracy of maps generated from remotely sensed data. Thus each strategy for sampling and ground truth gathering can affects the overall accuracy and consequently the relationship between MQ and OA.

Some of the objects properties have influences on the uncertainty and accuracy derived from the classification results. Geometric properties (e.g. objects size), spectral properties (e.g spectral similarity) of the objects are two major aspects that influence both of accuracy and uncertainty measures. Although these object properties present in the uncertainty and accuracy relation but have not the same effect on the accuracy and uncertainty. Therefore they prevent establishing a robust relationship between uncertainty and accuracy measures. As a consequence of this problem, we can not propose a valid fixed formula that gets uncertainty measure and gives the accuracy value for all cases.

As a consequence we can use the mean quadratic score as a cost free parameter that can tell us how much the classification is reliable without any need to collect the ground truth data. In comparing the individual classification results that have the same classification algorithm but have been done by different persons this parameter can be used. The smaller MQS the more accurate result. In the other way if we classify some data and after that perform some modifications on the entered data (or the other parameters) and then perform a new classification thus we can see the results of these modifications by estimating the MQS for both of the classifications and comparing them. Again that classification which gives the smaller value for the MQ can be selected as the better classification.

5. CONCLUSION

In this paper a linear relation between an uncertainty measure and an accuracy parameter has been investigated. The uncertainty measure that used the mean quadratic score with the overall accuracy and kappa coefficient was chosen as the commonly used accuracy measures. The famous inverse relationship between uncertainty and accuracy has been confirmed by this experiment and a strong relation between an averaged uncertainty value (MQS) and an averaged accuracy value (OA) have been found.

Although we have mentioned that these parameters are influenced by the various factors but we can use the MQS in comparing different classifications (not classifiers!). In fact this is the MQS that can be used to compare the reliability and performance of the classifications and the obtained relation (between OA and MQS) can not be used to predict the exact accuracy of the classification result. This is caused from this fact that both of the accuracy and uncertainty are influenced by some various factors that can alter the parameters of the linear relation. Among these effective factors the sampling scheme, sample size, classification procedure, and objects properties are some the most important effective parameters on the accuracy assessment and uncertainty analysis process.

In this study maximum likelihood classifier was used as a common procedure in the classification literatures. This procedure is able to produce probability vectors that are used to calculate the quadratic score. Therefore if any classifier that can not produce such information is used then we can not compute an uncertainty measure. Using another classifier such as minimum distance or artificial neural networks we should define an appropriate uncertainty measure and then test it whether it has any straight relation with the accuracy measures. This is a topic for the future investigations but as a general consequence it is anticipated that the linear relationship between MQS an OA will be remain.

We have found that among the uncertainty measures the mean quadratic score has a strong and reliable relationship with the commonly used accuracy measures. This relationship can be a good basis for the future investigations that can lead to the classification based accuracy measures and avoiding some problematic data related issues of ground truth data collection. The other uncertainty measures can be tested to define whether they have any stronger and more stable relation than the one we have found?

6. REFERENCES

Congalton R.,G., 1991. A review of assessing the of classifications of remotely sensed data. Remote Sensing of Environment, 37:35-46

Congalton R. G., 1988. A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. Photogrammetric Engineering and Remote sensing, Vol. 54, No. 5 May 1988

Glasziou, P. , J. Hilden,1989. Test selection measures. Medical decision making. Vol.9. pp. 133 - 141.

Gonzalez, R.C., Woods, R., 1993. Digital Image Processing. Addison Wesley Company , New York.

Goodchild, M.F., 1995. Attribute accuracy. In S.C. Guptill & J.L. Morrison (eds.): Elements of spatial data quality. pp. 59 - 79. International Cartographic Association / Pergamon, Oxford.

Stephanou, H.E. and Sage A.P., 1987. Perspectives on imperfect information processing. ieee transactions on systems, man and cybernetics, vol. 17, no. 5. pp. 780 - 798.

Maselli F., Conse C., Petkov L., 1994.Use of probability entropy for the estimation and graphical representation of the accuracy of maximum likelihood classifications. Isprs Journal of Photogrammetry and Remote Sensing 49(2)