A General Method for Assessing the Uncertainty in Classified Remotely Sensed Data at Pixel Scale

Yanchen Bo\textsuperscript{1,2} \footnote{Corresponding author. Tel.: +86-10-58802062; fax: +86-10-58805274 \E-mail address: boyc@bnu.edu.cn} and Jinfeng Wang\textsuperscript{3}

\textsuperscript{1}. State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and Institute of Remote Sensing Applications, CAS, \\
\textsuperscript{2}. Research Center for Remote Sensing and GIS, School of Geography, Beijing Key Laboratory for Remote Sensing of Environment and Digital Cities, Beijing Normal University, Beijing, 100875, China \\
\textsuperscript{3}. LREIS, Institute of Geographical Science and Natural Resources Research Chinese Academy of Sciences, Beijing, 100101, China

Abstract. The uncertainty assessment on the classification of remotely sensed data is a critical problem in both academic arena and applications. The conventional solution to this problem is based on the error matrix (i.e. confusion matrix) and the kappa statistics derived from the error matrix. However, no spatial distribution information of classification uncertainty can be presented by this method. A probability vector-based method has been developed for assessing classification at pixel-scale. However, the use of this method is severely limited because the probability vector can be derived only through the Bayesian classification. In practice, other classifiers such as Artificial Neural Network Classifier, Minimum Distance Classifier, Mahalanobis Distance Classifier and Fuzzy Classifier are wildly used for remote sensing data classification. To assess the uncertainty of thematic maps by these classifiers at pixel-scale, a general method is presented in this paper which extends the probability vector-based method to the assessment on the uncertainty classified by classifiers beyond Bayesian classifier. This extension is realized through a transformation method, which transforms the “Membership Vector” in various classifiers to the “transformed probability vector” so that it is comparable to the probability vector in Bayesian Classifier. The uncertainty measurements could be derived from the probability vector were evaluated and, the probability residual and entropy that derived from the extended probability vector are used as indicators to assess the absolute and relative uncertainty perceptively. The uncertainties by different classifiers are compared at pixel scale. Some examples of the uncertainty of maps from distance classifiers were presented and compared with that of maps from MLC classifier.

Keywords: scale, \textit{a posteriori} probability vector, uncertainty measure.

1. Introduction

Categorical spatial data set such as land use/land cover and vegetation type derived from remotely sensed data is the critical inputs in ecological and environmental modeling. As the classified maps from remotely sensed data are uncertain in nature, assessment on the accuracy and uncertainty of remotely sensed data classification is essential to the propagation and assessment on the uncertainty of models’ output and, furthermore, the risks of decisions based on the models outputs. An essential aspect of the increasing sophistication of ecological and environmental models is the use of spatially explicit inputs and outputs. Thus, the spatial distribution of the classified maps from remotely sensed data has become a new challenge.

The widely accepted method for assessing the accuracy of thematic maps from remotely sensed data has been the error matrix, or confusion matrix (Congalton and Green, 1999). A limitation of the error matrix is that it allows only one reference class for each reference site. This is problematic in sites where selection of
only one class is difficult or inappropriate. A solution to this problem is developed by using fuzzy set theory (Gopal and Woodcock, 1994; Woodcock and Gopal, 2000).

A common drawback of error matrix based method and fuzzy set theory based method is that it can not provide the spatial distribution of classification uncertainty. Assessing on the uncertainty of classification based on the MLC a posteriori probability vector was proposed by Goodchild et al. (1992). several uncertainty measures were derived from the MLC a posteriori probability vector by Shi (1998). However, the practicability of per-pixel assessment on classification uncertainty based on the a posteriori probability vector was limiting because the a posteriori probability vector can only be derived through MLC classification. For other widely used classifiers in remotely sensed data classification, such as Minimum Distance Classifier Mahalabous Distance Classifier, Artificial Neural Network Classifier and Fuzzy Classifier, how to assess the uncertainty of classification at pixel-scale is remain problematic. In addition, the various uncertainty measures were developed based on the MLC a posteriori probability vector. Evaluating and selecting appropriate uncertainty measures for uncertainty assessment on remotely sensed data classification at pixel-scale is important to the uncertainty assessment practices.

The research objectives of this paper are twofold. Firstly, select the appropriate measures for uncertainty assessment on remotely sensed data classification at pixel scale. Secondly, develop a general method for assessing the uncertainty of classification at pixel scale so that the uncertainty measures derived MLC a posteriori probability vector can be used to the assessment on the uncertainty of maps classified by classifiers beyond MLC at pixel scale. In section 2, the MLC a posteriori probability vector was introduced, the various uncertainty measures derived from a posteriori probability vector were evaluated and appropriate measures were selected. In section 3, the extended probability vector method were developed to extend the uncertainty measures at pixel scale from the assessment on uncertainty of MLC classification to the assessment on the uncertainty of ANN classification, Distance classification and Fuzzy classification. Section 4 gives some examples of uncertainty assessment at pixel using measures from a posteriori probability vector on the maps classified by classifiers beyond MLC. Section 5 summarizes the whole paper and discussions on further studies were given.

2. The a Posterior Probability Vector and the Uncertainty Measures

2.1. The a Posterior Probability Vector

In the MLC classification, the decision of assigning a pixel to a specific class is based on the a posteriori probability vector which presents the conditional probability the pixel belongs to every pre-defined class in the classification scheme. Most often, the pixel was assigned to the class with the maximum probability value in the probability vector.

Let the spectral classes for an image be represented by

\[ \omega_i, \quad i = 1...M \]  

(1)

where \( M \) is the total number of classes. According to Richard et al. (1998), to determine the class to which a pixel at a location \( x \) belongs, the conditional probabilities are of interest.

\[ p(\omega_i | x), \quad i = 1...M \]  

(2)

where the position vector \( x \) is a column vector of brightness values for the pixel, it describes the pixel as a point in multispectral space with coordinates defined by the brightness. The probability \( p(\omega_i | x) \) gives the likelihood that the correct class is \( \omega_i \) for a pixel at position \( x \). Classification is performed according to

\[ x \in \omega_i \quad \text{if} \quad p(\omega_i | x) > p(\omega_j | x) \quad \text{for all} \quad j \neq i \]  

(3)

i.e., the pixel at \( x \) belongs to class \( \omega_i \) if \( p(\omega_i | x) \) is the largest.
Often, the $p(\omega_i | x)$ in (1) is unknown. Suppose that sufficient training data is available for each ground cover type, the probability distribution for a cover type can be estimated. The desired $p(\omega_i | x)$ and the $p(x | \omega_i)$ estimated from training data are related by Bayes’ theorem:

$$p(\omega_i | x) = p(x | \omega_i) p(\omega_i) / p(x)$$

(4)

where $p(\omega_i)$ is the probability that class $\omega_i$ occurs in the image. $p(x)$ is the probability of finding a pixel from any class at location $x$ and can be expressed as:

$$p(x) = \sum_{i=1}^{M} p(x | \omega_i) p(\omega_i)$$

(5)

Where the $p(\omega_i)$ are called *a priori* probability and $p(\omega_i | x)$ are *a posteriori* probability. Using (2), the classification rule of (1) is:

$$x \in \omega_i, \text{ if } p(x | \omega_i) p(\omega_i) > p(x | \omega_j) p(\omega_j) \text{ for all } j \neq i$$

(6)

where the $p(x)$ has been removed as a common factor. For mathematical convenience, define:

$$g_i(x) = \ln \{p(x | \omega_i) p(\omega_i)\} = \ln p(x | \omega_i) + \ln p(\omega_i)$$

(7)

then (6) is reformed as:

$$x \in \omega_i, \text{ if } g_i(x) > g_j(x) \text{ for all } j \neq i$$

(8)

This is the decision rule used in maximum likelihood classification. $p(x | \omega_i)$ are referred to as discriminate functions. Assume that the probability distributions for the classes are of multivariate normal distribution, then the discriminant function can be restated as:

$$g_i(x) = \ln p(\omega_i) - \ln |\Sigma_i| - (x - m_i)^T \Sigma_i^{-1} (x - m_i)$$

(9)

where $m_i$ and $\Sigma_i$ are the mean vector and covariance matrix of the data in class $\omega_i$. They were estimated from training data.

Equation (3) implies that while the pixel at position $x$ was assigned to class $\omega_i$, it was not 100% sure that pixel at $x$ belongs to class $\omega_i$. Thus, this is an uncertain decision. The posteriori probability in the discriminate function is an indicator of the uncertainty in classification decision. In equation (4), the essence in classification decision is to estimate the conditional probability at which the pixel at $x$ belongs to class $\omega_i$, $i = 1,...,M$. The posteriori conditional probability pixel at $x$ belongs to every class composed a posteriori probability vector:

$$[p(\omega_1 | x), p(\omega_2 | x),..., p(\omega_i | x),..., p(\omega_M | x)]^T$$

(10)

Remove the elements of zero value in the posteriori probability vector and rank the elements from large to small, a ranked posteriori probability vector can be formed (Shi, 1998):

$$PV(x) = [p(\omega_{i1} | x), p(\omega_{i2} | x),..., p(\omega_{ik} | x),..., p(\omega_{kj} | x)]^T$$

(11)

where $k$ is the number of non-zero elements in equation (10). In equation (11), $p(\omega_i | x) \geq p(\omega_j | x)$ if $i < j$.

### 2.2. Uncertainty measures derived from a posteriori probability vector

Various uncertainty measures are derived from posteriori probability vector. Shi (1998) defined 4 measures based on the ranked posteriori probability vector.

- **Absolute Uncertainty**
Absolute Uncertainty was defined as:

\[ U_A(x) = \frac{p(\omega_i \mid x)}{[1 - p(\omega_i \mid x)]} \]  

(12)

in equation (12), \( U_A \) ranges in a interval \((0, +\infty)\). The larger the \( U_A \), the more certain the class \( \omega_i \) was assigned to the pixel at \( x \).

- Relative Uncertainty

Relative uncertainty describes the misclassification of pixel at \( x \) between class \( \omega_i \) and class \( \omega_j \). The relative uncertainty was defined as:

\[ U_R(x, i, j) = \left| \frac{p(\omega_i \mid x) - p(\omega_j \mid x)}{p(\omega_i)} \right| \]  

(13)

\( U_R(x, i, j) \) in equation (13) ranges in a interval of \([0, 1]\). The larger the value of \( U_R(x, i, j) \), the easier the pixel at \( x \) to be discriminated between class \( \omega_i \) and \( \omega_j \).

- Mixture degree of pixel

Mixture degree of pixel measures the degree that a pixel is a mixed pixel. It is defined as:

\[ M = \sum_{j=1}^{n} \frac{\sum_{i=1}^{M} | p(x \in \omega_j) - p(x \in \omega_i) |}{p(x \in \omega_i)} \text{ for all } i \neq j \]  

(14)

Where \( p(x \in \omega_i) \) is the probability that pixel at \( x \) belongs to class \( \omega_i \). The similar the elements in the posteriori probability vector, the smaller the \( M \) value, and the more possibility that the pixel \( x \) is a mixed pixel.

- Evidence Incompleteness

In the process of remotely sensed data classification, some pixels are hard to determine their classes and were labeled as “unclassified” because the evidence for classification is incomplete. The evidence incompleteness (\( \Theta \)) was defined as:

\[ \sum_{j=1}^{k} p(\omega_j \mid x) + \Theta = 1 \]  

(15)

Where the value of \( \Theta \) range in the interval of \([0, 1]\). Large value of evidence incompleteness indicates high possibility that the pixel was labeled as “unclassified”.

- Relative Maximum Probability Deviation (Clark Lab, 2001)

There is another uncertainty measure developed by Clark Lab (2001) that we call the Relative Maximum Probability Deviation. The Relative Maximum Probability Deviation was defined as:

\[ U = 1 - \frac{\max \left\{ \sum \right\} - \frac{1}{M}}{1 - \frac{1}{M}} \]  

(16)

Where \( \max \) stand for the maximum value in probability vector. \( \sum \) stands for the sum of all elements in probability. \( M \) is the number of elements in probability vector.

- Probability Entropy

In addition to the uncertainty measures defined by Shi (1998), probability entropy derived from posteriori probability entropy was wildly used as a uncertainty measures of remotely sensed data classification (Maselli, 1994; Foody, 1996). Probability entropy is a measure of uncertainty and information formulated in terms of probability theory, which expresses the relative support associated with mutually exclusive alternative classes. When two or more alternative classes have non-zero probabilities
associated with them then each probability is in conflict with the others. When there is a finite set of alternative classes the expected value of conflict is given by the probability entropy (Foody, 1996). This may be used to describe the variations in class membership probabilities associated with each pixel. The probability entropy indicates further required information content to assign a pixel to a class with 100% confidence.

The probability entropy was expressed as:

\[ H(p) = -\sum_{i=1}^{M} p(\omega_i | x) \log_2 p(\omega_i | x) \]  

(17)

For the convenience of visual presentation, the relative probability entropy was developed by Maselli et al. (1994) and was used in practice. The relative probability entropy was defined as:

\[ H_r(p) = \frac{H(p)}{H_{\text{max}}(p)} \times 100 \]  

(18)

3. Selecting Uncertainty Measures and Extending Posteriori Probability Vector

3.1. Selecting measures for uncertainty assessment at pixel scale

To make the assessment on the uncertainty of remotely sensed data classification at pixel scale be comprehensive and comparable, appropriate uncertainty measures have to be selected.

The absolute uncertainty defined by Shi (1998) and the Relative Maximum Probability Deviation present similar uncertainty information. In the definitions of these two measures, the element with maximum value in the probability vector is emphasized. These two measures are simple and straightforward in describing the absolute uncertainty in assigning a pixel to a specific class. The first drawback of these two measures is that the relationships between elements in probability vector were not considered. Most often, relationship between elements in probability vector, especially the relationship between the maximum one and the second maximum one, provide much more uncertainty information. Take a classification with 3 classes as an example, two posteriori probability vectors \([0.5, 0.3, 0.2]\) and \([0.5, 0.4, 0.1]\) present quite different uncertainty information. Though the absolute uncertainty and the relative maximum probability deviation of these two probability vector are the same, the former implies less uncertainty than the later one. Another drawback of these two uncertainty measures is that values range in a unlimited interval between \((0, +\infty)\), which make it hard to be presented by a gray scale image.

The relative uncertainty compares the possible misclassification of pixel between every two classes. In relative uncertainty, relationship between elements in probability vector is considered. However, because the differences of posteriori probability between every two classes have to be calculated, the relative uncertainty was presented by many posteriori probability differences together. Take a classification with 5 classes as an example, 10 relative uncertainty values need to be calculated, and for a classification with 10 classes, 45 relative probability values have to be calculated!

The mixture degree of pixel (M) describes the degree to which the pixel being classified is a mixed classification. While the differences between elements in probability vector are large enough, the M value is sensitive to the mixture degree of pixel. However, if the maximum of elements in two probability vector are the same, M is not sensitive. Take the two probability vector \([0.6, 0.4, 0.0]\) and \([0.6, 0.2, 0.2]\) again as example, both M values of these two probability vector are 1.33.

The evidence incompleteness is subject to the posteriori probability threshold set in the MLC classification. In the MLC classification, a probability threshold was set. If the maximum value of the probability vector of a pixel is less than the threshold, then there is no sufficient evidence to assign this pixel to any classes. Thus, different posteriori probability threshold comes up with different mixture degree for same pixel.

The probability entropy was wildly used as an uncertainty measure. As all the elements in the probability vector were accounted in the calculation of probability entropy, all the information in the probability vector was fully exploited. Because probability entropy implies the variation and dispersion of the elements in
probability vector (Maselli, 1994), both the information expressed by relative uncertainty and the mixture degree of pixel were contained in the probability entropy. An important merit of probability entropy is that it expresses the uncertainty contained in probability vector using a single value. Though the probability entropy is not sensitive to the class that pixel was finally assigned (Zhu, 1997), because the pixels to be classified are always assigned to the class with maximum probability value in the probability vector, thus the absolute uncertainty and the relative uncertainty information were both implied in the probability entropy. The probability entropy has been accepted as a good measure of classification uncertainty (Zhu, 1997; Foody, 1992; Maselli, 1994).

In summary, the absolute uncertainty and maximum relative probability deviation both indicate the absolute uncertainty information contained in probability vector, but no relative uncertainty information implied because the relationships between elements in probability vector were not accounted in. Take the uncertainty measures developed by Shi (1998) as a whole, they present both absolute and relative uncertainty information contained in probability vector, however need many measures simultaneously. Probability entropy expresses the uncertainty information contained in the probability vector using a single value. Because the relationships between elements in probability vector were accounted in probability entropy, the relative uncertainty information contained in probability vector was implied in probability entropy. Thus, we use the relative maximum probability deviation and the probability entropy to present the absolute and relative uncertainty contained in probability vector respectively.

3.2. Extending the a posteriori probability vector

As was discussed in section 2, the posteriori probability vector could only be derived through MLC classification. In practice, many others classifiers including ANN classifier, Fuzzy classifier and various distance classifiers were widely used in remotely sensed data classification. In these classifiers, there are also measures that present the memberships pixel belongs to every class (Foody, 1996; 2000), and can be taken as an approximation of posteriori probability that pixel belongs to given class.

Another type of classifier wildly used in remotely sensed data classification is spectral distance based classifiers, including Minimum Distance Classifier and Mahalanobis Distance Classifier (Richards and Jia, 1998). Like in the MLC classification where the memberships of a pixel to every class were presented by the posteriori probabilities of the pixel belong to every class, the memberships of a pixel to every class in distance classifiers were presented by the Euclidean distances or Mahalanobis distances from the feature of pixel to the features of classes estimated from training data. Thus, the memberships can also compose a “Distance Vector”

$$[d(\omega_1 \mid x), d(\omega_2 \mid x), \ldots, d(\omega_i \mid x), \ldots, d(\omega_d \mid x)]^T$$

(19)

In the classification decision of distance classifiers, pixels were assigned to the class with the minimum distance in the “distance vector”. Like the probability vector, the distance vector contains the uncertainty information of the distance-based classification. However, because the value of elements in “distance vector” ranges in an unlimited interval, rather than in the interval of [0,1] in probability vector, distance vectors from different classifications are not comparable, and the measures of uncertainty discussed in section 2 can not derived directly from the distance vector. Take three classes classification as an example. Given the means vector estimated from the training data is [30, 60, 90], the spectral DN value of pixel A is 70, and pixel B 100. Based on the minimum distance classification, the distance vector of pixel A and pixel B are [40, 10, 20] and [70, 50, 10] respectively. Clearly it is unreasonable to draw a conclusion that this two distance vectors present same absolute uncertainty just because they have same minimum element.

Thus, it is necessary to convert the elements in the distance vector to the interval of [0,1] while the relative relationship between elements are preserved. The conversation was made using the following equation:
\[
p_{c}(\omega_{i} | x) = \frac{1/d(\omega_{i} | x)}{\sum_{i=1}^{M} 1/d(\omega_{i} | x)}
\]

(20)

Applying the equation (20) to the example above, we can see that the conversed “probability vector” for pixel A is \([0.571, 0.143, 0.286]\), and pixel B the \([0.534, 0.333, 0.1333]\). From these two “probability vector”, it is reasonable to draw a conclusion that the absolute uncertainty of pixel A is less than that of pixel B.

By the conversion, uncertainty measures from posteriori probability vector in MLC can be extended to the assessment on uncertainty of classification from neural network classifier and distance classifiers at pixel scale. Thus, we call the “activation level vector” and the conversed “distance vector” together the “Extended probability vector”.

4. Example

The land cover mapping by multispectral remotely sensed data classification was used here to illustrate the uncertainty assessment at pixel scale presented in this paper.

The standard false color composed imagery of Landsat Thematic Mapper in Lanier Lake, U.S. in Figure 1. Land cover types in the imagery include agricultural field, bare ground, grass land, conifers, deciduous, water, urban and the developed. Because there are some cloud and the shade of cloud in this image, the classes in classification include the cloud and the shade the imagery was classified by MLC classifier and Minimum Distance classifier respectively using same training samples. The classified land cover maps were shown in figure 1. The posteriori probability vector was saved in the MLC classification. The distance vector was saved in the minimum distance classification and was conversed to the “probability vector”using equation (28). The Maximum Relative Probability Deviation and the Probability Entropy for both MLC classification and Minimum Distance classification were calculated and presented in Figure 2.

![Fig. 1. The Landsat TM imagery and the land cover maps classified using Maximum Likelihood Classifier and Minimum Distance Classifier: (a) The Landsat TM imagery in Lanier Lake, U.S.A (b) Land cover map classified using MLC (c) Land cover map classified using Minimum Distance Classifier (d) Land cover legend](image-url)
The absolute uncertainty of landcover map classified using MLC was presented using the maximum relative probability deviation in Figure 2 (a). The pixels with uncertainty larger than 0.5 mainly locate in the urban and the developed, bare ground, grass and agriculture field and the boundary of water. These spatial distribution of absolute uncertainty was caused by the spectral ambiguity between urban and the bare ground, and between grass land and agriculture field. Pixels with uncertainty between [0.3, 0.5] locates in the conifers and the deciduous. This spatial distribution was caused by the mixed pixels of the conifers and the deciduous. The absolute uncertainty decreases from the objects center to their boundaries. The relative uncertainty of land cover map classified using MLC was presented using probability entropy in Figure 2 (b), in which the pixels with high probability entropy locate in the boundary of the water, the clouds, the shades, the conifers and the deciduous. The uncertainties of these pixels were mainly caused by pixel mixture.

The absolute uncertainty and relative uncertainty of land cover map classified using the minimum distance classifier were presented in Figure 2 (c) and figure 2 (d) respectively. The maximum relative probability deviation and the probability entropy were derived from the “probability vector” converted from the “distance vector” produced by the minimum distance classifier. Due to the characteristics of the minimum distance classifier, the classes with similar spectral characteristics, such as urban areas, clouds and bare ground, showed high uncertainty. The water showed less uncertainty as its spectral characteristics has distinctive difference with other classes. Figure 2 showed that the uncertainty of maps classified using different classifiers are quite different even using same training samples.

5. Conclusions
The uncertainty assessment on the classification of remotely sensed data is a critical problem in both academic arena and applications. The conventional solution to this problem is based on the error matrix (i.e. confusion matrix) and the kappa statistics derived from the error matrix. However, no spatial distribution information of classification uncertainty can be presented by this method. A probability vector-based method.
has been developed for assessing classification at pixel-scale. However, the use of this method is severely limited because the probability vector can be derived only through the maximum likelihood classification and because lack of the comprehensive and comparable uncertainty measures derived from the posteriori probability. The researches in this paper focused on the solutions to overcome these two drawbacks. Firstly, the uncertainty measures from the probability vector were evaluated. The maximum relative probability deviation and the probability entropy were selected to present the absolute uncertainty and relative uncertainty respectively. Secondly, the “activation level vector” produced in the feedforward neural network with backpropagation algorithm classification and the “distance vector” produced in the distance classifiers were converted to “probability vector” so that all elements in it are in an interval of [0,1] and are comparable to the probability vector produced in the ML classifier. By this conversion, uncertainty measures for uncertainty assessment on remotely sensed data classification using MLC at pixel scale can be used in the uncertainty assessment on classification using neural network classifier and distance classifiers at pixel scale. The examples of land cover mapping by classifying Landsat TM multispectral remotely sensed data showed that the maximum relative probability deviation and the probability entropy can present the spatial distribution of absolute uncertainty and relative uncertainty well. The spatial distributions of uncertainty of maps classified using different classifiers are quite different.

6. Acknowledgements

This research were supported by the grants from National Basic Research Program of China (Project No.2007CB714402-10), State Key Lab. of Remote Sensing Science, NSFC (40301033),and the Program for Changjiang Scholars and Innovative Research Team in University(PCSIRT, No.IRT0409)

7. References