# A Probabilistic Modification of the Decision Rule in the Skidmore/Turner Supervised Nonparametric Classifier 

Kim E. Lowell<br>University of Missouri, 1-31 Agriculture, Columbia, MO 65211


#### Abstract

Using a hypothetical example, the paper demonstrates how the original classification decision rule may cause the amount of land in each cover type to be misestimated. An alternative is proposed and discussed which should provide better area estimates for each type, though locational accuracy may be decreased. The aesthetic quality of the classified image is also reduced because of the "salt and peppering" that results from the use of the probabilistically based rule.


## INTRODUCTION

SKIDMORE AND TURNER (1988) presented and discussed a supervised nonparametric classifier which was used with multispectral SPOT data to improve classification accuracy compared to either a maximum-likelihood classifier or a Euclidean distance classifier for Monterey pine (Pinus radiata D. Don) plantations in Canberra, Australia. To utilize the classifier, a training field is selected from an unclassified image for each land-cover type of interest. For each training sample extracted, a unique N dimensional space is created in which each dimension represents one spectral band of data, and each "edge" of the space is dimensioned by the possible range of pixel brightness values. For example, 7 -band 8 -bit Thematic Mapper data would result in the creation of a series of seven-dimensional spaces with each dimension or "side" ranging from 0 to 255 . (The space, therefore, would have a total of 256 raised to the 7 th power cells, a number which may be reduced by using a "collapsing factor.") For each training sample, the number of pixels in each cell position of the appropriate N -dimensional space is tallied based on the brightness value of each band for each pixel. A probability is then assigned to each cell position in a particular N dimensional space relative to the number of pixels tallied in that cell position over all N -dimensional spaces. The image is then classified by assigning each unknown pixel to the cover type for which the probability is the greatest at the appropriate cell position.

However, because unknown pixels are not distributed to each cover type in the same proportion as they occur in the training samples, this decision rule may cause the amount of land in a particular type to be misestimated. To demonstrate this and describe an alternative decision rule, an intentionally simplistic example will be utilized so that laborious computations are minimized. The simplicity of the example does not decrease the validity of the points discussed, however.

In this example, a tract of land of 1000 hectares is available on which it is known that two cover types exist, with each type comprising exactly half -500 hectares - of the area. A cloudfree digital image of the area has been obtained using a sensor which recorded spectral reflectance values in two spectral bands using 1-bit data and a pixel size of 1 hectare. To classify this image using the nonparametric classifier, two training samples of 100 pixels each have been extracted from areas on the image which are known to be perfectly representative of Type 1 and Type 2, respectively. For each type, the pixels were tallied in two-dimensional space based on brightness values and spectral band numbers with the following results:


Utilizing the procedures described, the following probabilities result for each cell position of the two-dimensional spaces:


Given that the 200 pixels selected in the training fields are exactly representative of each type, the brightness values of the 1000 pixels can be expected to be distributed:

Band 1


Using the original decision rule, the 700 pixels/hectare in cell positions $(0,1),(1,0)$, and $(1,1)$ are assigned to Type 1 due to probability values of $0.55,0.57$, and 0.80 , respectively, compared to $0.45,0.43$, and 0.20 for Type 2 . The 300 pixels/hectare in cell position $(0,0)$ are assigned to Type 2 due to a probability of 0.83 compared to 0.17 for Type 1 .
However, it is known that 500 pixels/hectare are in each of Types 1 and 2. If 700 and 300 hectares of Types 1 and 2, respectively, are used as the estimates of the amount of each type in the area, then erroneous decisions having potentially severe consequences may be made by land managers. The degree of misestimation is dependent upon the highest type probability - as the highest probability decreases, the degree of misestimation increases (Figure 1). In case where the highest probability is 0.10 at a particular cell position, 10 times as much land


Fig. 1. Area overestimation for each decision rule. (A value of 1.0 indicates no overestimation. A factor of 10 indicates that 10 times too much land will be estimated for the type with the highest probability. The area in all other classes will be underestimated.)
as actually exists will be placed in a single type and zero area will be estimated to occur for the remaining classes, even though 90 percent of the pixels at this cell position in the training samples were not in this type. If accurate area estimates are to be obtained, therefore, one must hope that the errors of under- and overestimation for each type will compensate - rather than compound - over all cell positions
A possible alternative is a decision rule which would assign unknown pixels to land types in an unbiased fashion so that pixels are distributed according to the calculated probabilities. To do this, a range of values is associated with a particular type proportional to its probability at a given cell position, and a random number is generated to determine to which type the pixel is assigned. For cell position $(0,0)$ in the example presented, values 1 through 17 and 18 through 100 would be assigned to Types 1 and 2, respectively. For each unknown pixel having brightness values of 0 for both Bands 1 and 2, a random number between 1 and 100 inclusive would be generated with a value less than 18 resulting in the classification of the pixel as Type 1, and a larger random number resulting in the pixel being assigned to Type 2. In this way, 17 percent of the pixels in this cell position would be assigned to Type 1 and 83 percent would be assigned to Type 2, as similar pixels were distributed in the training samples. As a result, a better estimate of the amount of land in each type should result. In this example, the probability for each cell position can be multiplied by the expected number of pixels in each position for the entire scene to demonstrate that area estimates would better reflect true values:

| type 1 | Position |  |  |
| :---: | :---: | :---: | :---: |
| $300^{*} 0 . \overline{167}$ | $=$ | type 2 |  |
| $275^{*} 0.545$ | $=150$ | $(0,0)$ | $300^{*} 0.8 \frac{833}{3}$ |

While this decision rule clearly provides better estimates of the total area of each type, there are other factors that will affect which rule is preferable in individual cases. The primary weakness of the proposed rule is that of location. For example, if there are only two types represented at a cell position for which probabilities are 0.90 and 0.10 , respectively, the original decision rule will assign 100 percent of the unknown pixels to Type 1. Though this means that 10 percent of the unknown pixels
(which are actually Type 2) will be incorrectly assigned to Type 1 , and also that no pixels at this cell position will be correctly loacted for Type 2, 90 percent of all pixels (or 100 percent of the type 1 pixels) can be expected to be located correctly. Conversely, use of the proposed decision rule would result in 90 percent of the unknown pixels randomly placed in Type 1 being classified correctly (or 81 percent of the total) and 10 percent of these being classified incorrectly ( 9 percent of the total). For Type 2, 10 percent of the pixels placed in Type 2 would be classified correctly ( 1 percent of the total), but 90 percent of those placed in Type 2 would be classified incorrectly ( 9 percent of the total). Thus, the original rule could be expected to locate 90 percent of the unknown pixels correctly, whereas the proposed rule would locate only 82 percent correctly.

The proposed rule is not inferior for locational purposes over the entire probability range, however. In cases where the highest probability is near 1.0, and in cases where the probability for a cell position is nearly equal over all types, the two rules can be expected to locate approximately the same number of pixels correctly (Figure 2). Thus, both rules perform similarly when there is either near certainty, or great uncertainty, about the true type of a pixel at a particular cell position.

Considering the trade-off between accuracy and location, the choice of either decision rule depends on a number of factors. In the original article, Skidmore and Turner used the classifier to identify types - forest plantations of different ages - which are known to occur in blocks, rather than in a distributed pattern. In similar situations, the original decision rule is probably preferable because it may be undesirable to have the inherent "salt and peppering" which occurs if pixels are classified based on a random number. However, area estimates may be slightly incorrect using the original rule due to the mixed pixels which would occur at the edges of different plantations.

This also leads to the suggestion that the probability matrices should be examined for each type before selecting a decision rule. Figure 2 demonstrates that the two decision rules perform similarly for both area and location if great certainty or uncertainty exists concerning the actual type of a pixel. Thus, to classify a scene where types are either spectrally very similar (e.g., soy beans and corn) or spectrally very different (e.g., roads and forest), the choice of decision rule may make little difference.

Rarely, however, does a scene exist where all classes of interest are either extremely similar or different. Therefore, the selection of either decision rule must be based upon the in-


FIG. 2. Locational accuracy. (Based on two classes only. The graph indicates the percent of all pixels which will be located correctly with either decision rule.)
tended use of the imagery, and the composition of the land cover within the scene. In cases in which estimates of area are of greatest importance, then the alternative decision rule is superior. Conversely, the original decision rule will decrease the amount of "salt and peppering" to produce a more aesthetically pleasing classified image in which a greater number of features are located correctly.

## REFERENCES

Skidmore, B.J., and B.J. Turner, 1988. Forest mapping accuracies are improved using a supervised nonparametric classifier with SPOT data. Photogrammetric Engineering and Remote Sensing, 54(10):14151421.
(Received 23 November 1988; revised and accepted 17 February 1989)

## Response

# Why Areal Accuracy Is Not Correctly Estimated Using Lowell's Modification of the Supervised Nonparametric Classifier 

TWO TYPES of thematic mapping error are (a) locational, where a pixel is incorrectly classified according to some ground truth criteria, and (b) areal, where the area of a class on an image does not equal the area of the class according to the ground truth.

The concern of Lowell is that areas for classes may be poorly estimated using the supervised nonparametric classifier. This problem is common to all probabilistic classifiers, including the maximum likelihood classifier.

At the outset, it should be emphasized that spectrally discrete classes will yield images with a higher spatial and areal accuracy, independent of the probabilistic classifier used. As discussed by Skidmore and Turner (1988), the image (Figure 4 on page 1418) of the empirical probability of correct classification (i.e., $\mathrm{P}(i \mid \mathbf{X})$ ) clearly differentiates classes which are spectrally similar from classes which are spectrally discrete. Thus, spectrally similar classes that will have low locational and areal accuracies can be identified. If there is co-occurrence at vector position ( $\mathbf{X}$ ) then information cannot be created from the confusion. The advantage of the supervised nonparametric classifier becomes apparent when two adjacent vector positions ( $\mathbf{X}$ ) each contain a separate class. In such a situation, parametric classifiers (such as the maximum likelihood classifier) may parameterize the adjacent vector positions as having the two classes co-occurring, when in fact the classes do not co-occur.

Lowell's example of a binary data set with two features is unrealistic as it is not characteristic of remotely sensed data. As discussed by Skidmore and Turner (1988), data sets such as the one proposed by Lowell could be considered to be in a severe state of "collapse" from the original 6- or 8-bit data. Using such a data set, a classifier will behave unpredictably due to classes being merged into (i.e., co-occurring in) a common vector position. High mapping accuracies are obtained when classes do not co-occur in vector positions, which is hopefully the situation when using a less severely collapsed data set. With more vector spaces, the area estimates of the supervised nonparametric classifier would be closer to the true values.

Furthermore, Lowell's modification will not correctly estimate the area of a class, where the estimated a priori probability for a class (over the area being classified) does not equal the true proportion of the class.

Following the notation of Skidmore and Turner (1988), the supervised nonparametric classifier is used to compute $(\mathrm{P}(i \mid \mathbf{X})$ for $i=1,2$, where $\mathrm{P}(i \mid \mathbf{X})$ is the probability of class i occurring
at vector position X. Lowell's allocation of pixels to class 1 can be succinctly stated as follows:

$$
S_{1}=\sum_{x=1}^{n} P(1 \mid \mathbf{X}) T(\mathbf{X})
$$

where $S_{1}$ is the proportion of the image classified as class 1 , $T(X)$ is the total number of image pixels occurring at vector space $(\mathbf{X})$, and $n$ is the total number of pixels in the image. Note that it is not necessary to perform list sampling or create a classified image as proposed by Lowell.
$\mathrm{P}(i \mid \mathbf{X})$ may be calculated using Bayes' Theorem (see Skidmore and Turner, 1988). Using Lowell's example, $\mathrm{P}(1)=0.5$, $\mathrm{P}(2)=0.5$, the area of class 1 is 500 pixels, the area of class 2 is 500 pixels, and there are two features. It is then simple to calculate $\mathrm{P}(1 \mid \mathbf{X}), \mathrm{P}(2 \mid \mathbf{X}), \mathrm{S}_{1}$, and $\mathrm{S}_{2}$ at each vector position in two-dimensional feature space, as shown in Table 1.
So when $\mathrm{P}(1)=\left(\mathrm{S}_{1} / \mathrm{n}\right)$, Lowell's modification works. In other words, the modification works when the a priori probabilities equal the actual proportions of class areas in the image.
However, if the a priori probabilities are not known exactly and (as is usually the case) estimated incorrectly, then the areas of the classes are also incorrectly estimated. For example, the effect of changing $\mathrm{P}(1)$ to 0.45 and $\mathrm{P}(2)$ to 0.55 on the estimated areas can be seen in Table 2. Therefore, using Lowell's modi-

Table 1.

| Vector <br> position | $\mathrm{T}(\mathbf{X})$ | $\mathrm{P}(1 \mid \mathbf{X})$ | $\mathrm{P}(2 \mid \mathbf{X})$ | $\mathrm{P}(1 \mid \mathbf{X}) * \mathrm{~T}(\mathbf{X})$ | $\mathrm{P}(2 \mid \mathbf{X}) * \mathrm{~T}(\mathbf{X})$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 0,0 | 300 | 0.167 | 0.833 | 50.1 | 249.9 |
| 0,1 | 275 | 0.545 | 0.455 | 149.9 | 125.1 |
| 1,0 | 175 | 0.571 | 0.429 | 99.9 | 75.1 |
| 0,0 | 250 | 0.800 | 0.200 | 200.0 | 50.0 |
|  |  |  |  | $\mathrm{~S}_{1}=500.0$ | $\mathrm{~S}_{2}=500.0$ |

Table 2.

|  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Vector |  |  |  |  |  |
| position | $\mathrm{T}(\mathbf{X})$ | $\mathrm{P}(1 \mid \mathbf{X})$ | $\mathrm{P}(2 \mid \mathbf{X})$ | $\mathrm{P}(1 \mid \mathbf{X})^{*} \mathrm{~T}(\mathbf{X})$ | $\mathrm{P}(2 \mid \mathbf{X}) * \mathrm{~T}(\mathbf{X})$ |
| 0,0 | 300 | 0.141 | 0.859 | 42.2 | 257.8 |
| 0,1 | 275 | 0.495 | 0.505 | 136.2 | 138.8 |
| 1,0 | 175 | 0.522 | 0.478 | 91.3 | 83.7 |
| 0,0 | 250 | 0.766 | 0.234 | 191.5 | 58.5 |
|  |  |  |  | $\mathrm{~S}_{1}=461.0$ | $\mathrm{~S}_{2}=539.0$ |

fication the area of class 1 is incorrectly estimated as 461 and the area of class 2 is incorrectly estimated as 539.

The errors in estimating class areas become larger as the estimate of the a priori probabilities deviates further from the true values. For example, when $\mathrm{P}(1)=0.2, \mathrm{~S}_{1}=250$; and when $\mathrm{P}(1)=0.75, \mathrm{~S}_{1}=700$ (remember, the true value should be 500).

The following question then arises. If we know a priori the exact area of class 1 and class 2, then why are we bothering to calculate the areas? If we do not know the exact areas, then an incorrect estimation of the a priori probabilities will lead to an incorrect estimation of the class areas.

Even using Lowell's pathological example, the supervised nonparametric classifier decision rule gives an equal or better estimate of area at some a priori probabilities (e.g., between $\mathrm{P}(1)$ $=0.75$ and $P(1)=0.8$ ), even though the spectral resolution (i.e., a binary two-feature data set) has little similarity with remotely sensed data (which normally has three to seven features and a spectral resolution per channel of 6 to 8 bits), and such a discretised data set accentuates errors in the area estimates of the supervised nonparametric classifier because there are only four vector spaces at which the classifier makes a decision.

Lowell's modification will increase locational errors in the classified range at all a priori probabilities (except $1.0,0.5$, and 0 ) by introducing noise through randomly allocating pixels to classes according to the proportions of $\mathrm{P}(1 \mid \mathbf{X})$ and $\mathrm{P}(2 \mid \mathbf{X})$ (see Lowell's Figure 2).

A final problem not discussed by Lowell is that training areas may not be representative of cover classes, because of natural
variability in cover classes or pixels containing more than one cover class (i.e., a mixed pixel). If it is assumed, for the example Lowell presents, that training areas are not representative of the classes, the areal (and of course locational) estimates would be incorrect. This is because the $\mathrm{P}(1 \mid \mathbf{X}), \mathrm{P}(2 \mid \mathbf{X})$, and $\mathrm{T}(\mathbf{X})$ calculated from the training areas would not equal to the $\mathrm{P}(1 \mid \mathbf{X})$, $P(2 \mid \mathbf{X})$, and $T(X)$ for the whole image.

Methods for improving mapping accuracies include merging similar classes into one class, selecting realistic training areas, introducing collateral data, considering the spatial context of features in the remotely sensed data, or improving the spatial, spectral, and temporal resolution of the remotely sensed data.

## ACKNOWLEDGMENT

We would like to acknowledge the assistance of Dr. D. Jupp of the C.S.I.R.O., Canberra, for discussing the material in this comment.

## REFERENCE

Skidmore, A.K., and B.J. Turner, 1988. Forest mapping accuracies are improved using a supervised nonparametric classifier with SPOT data. Photogrammetric Engineering and Remote Sensing 54(10):14151421.

-A. K. Skidmore and B. J. Turner<br>Department of Forestry<br>Australian National University<br>G. P. O. Box 4<br>Camberra, A. C. T. 2601, Australia

## Erratum

In the paper, "Forest Mapping Accuracies Are Improved Using a Supervised Nonparametric Classifier with SPOT Data," by Andrew K. Skidmore and Brian J. Turner (PE\&RS, October 1988, page 1416), Equation 2 should read as follows:

$$
\mathrm{P}(i \mid \mathbf{X})=\frac{\left(\mathrm{F}_{j} / \mathrm{F}_{i}\right) \mathrm{F}_{i}(\mathbf{X}) \mathrm{P}(i)}{\Sigma\left(\mathrm{F}_{j} / \mathrm{F}_{i}\right) \mathrm{F}_{j}(\mathbf{X}) \mathrm{P}(j)}
$$

## New Sustaining Member

## PCI Remote Sensing Corp.

2300 M Street, NW, Suite 130, Washington, DC 20037<br>Telephone 202-785-8281; fax 202-785-8278

PCI REMOTE SENSING CORP. develops and supports image analysis software and provides integrated computer workstations. Hundreds of PCI systems are used around the world to more effectively monitor and map crops, terrain, forests, oceans, and the atmosphere.

PCI's software, called EASI/PACE, converts remotely-sensed imagery and digitized aerial photography into vivid, color-enhanced images, allowing users to readily analyze and extract meaningful information. It's one of the most comprehensive and best-supported image analysis systems; users benefit from the flexibility of the software, as it was designed for easy networking and operates on a wide variety of hardware. The benefits of EASI/PACE continue to be discovered for use in resource management and conservation; exploration; teaching and research; mapping, charting, and geodesy (MC\&G); and for intelligence data fusion.

The company was established in 1982 and is driven by one of Canada's leading founders of remote sensing, Dr. Murray Strome. Dr. Strome assisted in the design and implementation of the country's first Landsat processing system and now presides as Chairman of the Board.

The PCI team includes professionals with strong education and experience. One team member is a geologist with ten years' service with a large mining company; another member spent six years as a researcher and manager of a remote sensing research project with the Canadian Forestry Service. Others have university degrees in geophysics, geography, and engineering; several have backgrounds in computers and business. Because of the team's collective strengths, PCI is able to provide extensive education, system support, and consulting services.

EASI/PACE continues to be upgraded with new ideas and technology; this emphasis on software enhancements has contributed to the company's success, both at home and abroad. PCI's head office is located in Toronto, Canada; a network of representatives and distributors are located outside North America to help build PCI's leading reputation.

