## General Disclaimer One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)



7.9-1 0.0.2.2<br>CR. 157911

(E79-10022) STATISTICAL SEPARABIIITY AND<br>N79-12533<br>CLASSIPICATION OF LAND JSE CLASSES USING<br>IMAGE-100 (Instituto de Pesquisas Espaciais,<br>Sao Jose) 17 p HC $\AA 02 / \mathrm{MF} \mathrm{A01}$ CSCL 05B<br>G3/43 00022

## ORIGINAL PAGE IS OF POOR QUALITY

## $\cdots$

STAMISTLCAL SEPAMBILITY AND CLASSHFICATLON OF LAND USE CLASSLS USING J:ACE-10U
R. Kunar* and M.Nierom

Instituto de Posquisas Espaciais (INPL:)
Conselho Nacional de Desenvolvimento Científico e Teenolögico (CNPq)
12.200 - S. J. dos Campos, S.P., Brasil.

The purpose of this study was so compare the classification accuracy of land use classes ar S.J. dos Campos, S.P., Brasil, using the different options of classification available in the Image-100 of General Electric Co. and the sample classifier. In addition, the statistical separability of land use elasses in the subsets of one to four spectral channels fras investigated. With the help of ground observations and aerial photography, the multispectral scanner (MSS) data of LANDSAT were analysed using the Image-100. In the subsets of one to three spectral channels (ch. $4=0.5$ to 0.6 m , ch. $5=$ 0.6 to 0.7 . $\mu_{\mathrm{m}}$, ch. $6=0.7$ to $0.8 \mu \mathrm{~m}$, ch. $7=0.8$ to $1.1 \mu \mathrm{~m}$ ); channel 4, channels $4 \& 7$, channels $4,5 \& 7$ were found to be the best choices. For the single-cell option of the Inage-joo, the errors of omission varied from $5 \%$ for the class industrial to $46 \%$ for the class institutional. The errors of commission varied from $11 \%$ for the class comercial to $39 \%$ for the class industrial. As expected, the multi-cell option increased the crrors of omission and decreased the errors of commission. however, considering both the errors of omission and commission, this option considerably decreased the percentage of correct classification as compared to the single-cell. option, On the Whole, the sample classifier gave considerably more accurate results as compared to the single-cell or multi-cell option.
$\qquad$ ..
Research Scientist in the Space Information Science Division.
** Graduate Research Assistant in the Department of. Remote Sensing of Natural Resources:


* Research Scientist in the Space Information Science Division.
** Graduate Research Assistant in the Department of Remote Sensing of Natural Resources.



## Introduction

the purpose of this study was to comate the chasificam tion aecuraty of land use dasses using the different options of chassification availablo in the Image 100 (Imperatoo is a data processing system matheded by General Plectric Co., that extracts themate information thon matispectral imagery, entances the imate, vec.). In addition, the atatistical separability of land use chasses in the subsets of one to four spectal chmmels was investigated.

Cloud free multispectral scamer data from l.ANDSAlt, of reasonable quality, over sĩo Jose dos campos ( $23^{\circ} 10^{\prime} \mathrm{s}, 45^{\circ}$ $50^{\prime}$ W), Stio Paulo, Beasil, acquired on September 8, 1972, were available. In addition, atrial photography and ground observations were avaibale to assist the amalys of the data, sao Jose dos Campos was selected because it $i$ s one of the most industrialiged and fastest growing small-siad towns of Buasil and the authors are well fami liar with it. Many of the problens of this town are similar to the problems of much larger urban centers.

With the halp of ground obstrvations and aerial photography, a map of sato Jose dos Gampos showing the following land use classes was obtained; residential areas, multi-family residential areas, comercial areas, industrial areas, institutional areas, agricultural arwa and unoceupied areas.

The specific objectives of the study are stated as follows:

1. To determine what combinations of one through three spectral chamels out of four available chanmels give the greatest overall statistical separability of the above seven limd use classes.
2. To compare the classification atcuracy of land use classes using 'singte-cell signature acquisition' and 'multi-cell signature acquisition' options of elassification available in the Inage-100 and a sample classifier on-minemode in the Image-100.

## Witerature Review

Many investigators have malysed the multispectral scanner (NSS) data of $A, A N D S A$ satellite for applications to land use classification. For exareple, Tode and Bamgarther (1973) analysed ANDSAT MSS data, htained over Marion Comaty (Indianapolis.), Indiana, by compute-implemented techniques, to

## Introduction

The purpose of this study was to compare the classification accuracy of lad use classes using the different options of classification available in the Image-100 (Image-100 is a deta processing system marketed by General Electric Co., that extracts thematic information fron multispectral imagery, enhances the imare, etc.). In addition, the statistical separability of land use classes in the subsets of one to four spectral channels was investigated.

Cloud free multispectral scanner data from LaNDSAT, of reasonable quality, over São José dos Campos $\left(23^{\circ} 10^{\prime} \mathrm{S}, 45^{\circ}\right.$ $50^{\prime}$ W), São Paulo, Brasil, acquired on September 8, 1972,were available, In addition, aerial photography and ground observations were available to assist the analysis of the data. São Jose dos Canpos was selected because it is one of the most industrialized and fastest growing small-size towns of Brasil add the authors are well familiar with it. Many of the problems of this town are similar to the problems of much larger urban centers.

With the help of ground observations and aerial photography, a map of Säo José dos Canpos showing the following land use classes was obtained; residential areas, multi-family residential areas, conmercial areas, industrial areas, institutional areas, agricultural areas and unoccupied areas.

The specific objectives of the study are stated as follows:

1. To determine what combinations of one through three spectral chamels out of four available channels give the greatest overall statistical separability of the above seven land use classes.
2. To compare the classification accuracy of land use classes using 'single-cell signature acquisition' and 'malti-cell signature acquisition' options of classification available in the lnage-100 and a sample classifier on-line-mode in the Image-100.

## Literature Review

Many investigators have analysed the maltispectral scanner (MSS) data of LANDSAT satellite for applications to land use classification. For example, Todd and Baumgardner ${ }^{1}$ (1973) analysed IANDSAT MSS data, btained over Marion County (Indianapolis), Indiana, by computer-implemented techniques, to

## ORIGINAL PAGLE IS of pour quality

evaluate the utility of satellite data for urban land use clasaification. Several land use classes, such as commerce/ir dustry, single-iamily (Newer) residential, trees, and water exhbited spectrally separable characteribtics and were identified with greater than 90 pereent accuracy. Fllefsen et a ${ }^{2}$. (1973) did computer-aided amalysis of landsit mis data of the san loancisco bay area. Emith et nl. ${ }^{3}$ (197i) have given the application of spatial features to satellite land-use analysis. Ellefson et al. ${ }^{4}$ (1974) have given new techiques in mapping urban land use and monitaring change for selucted U.S. metropolitam areas. They analysed hannsat ass data using automatic pattern recognition techniques for classilication. Kumar and Silvab (1977) have analysed the statistical separability of agricultural cover types in much detail, data quantity and depth in the subset: of one to twelve spectral chanels.

Economy et a. $0^{6}(1974)$ have given the classificatien accuracy of the Image- 100 using single cell option (described later in the report). They reported an accuracy of $70 \pm 12 \%$. Factors whin limited the classification accuracy wete spatial resolution of the ERTS-1 (now called LANDSAT) scanner, the land-use classificarion system, and inaccuracies in the eround truch. Goldberg et. 11.7 (1975) have described methods and procedures which outside investigators may use with the automated processing equipment of the Canada Centre for Remote Sensing (CCRS) for the purpose of natural resource exploration and mapping. They have compared the accuracies of unsupervised and supervised methods on the basis of the confusion matrices gencrated by classifying exactly the same area.

## Method of Amalysis

Multispectral scamer data of computer compatible tapes of LeANSSAT were analysed using Image-100. With the aid of land use map of Säo Jose dos Campos mentioned above, rectangular areas of each of the above seven land use classes vere selected, avoiding the boundaries of classes on the display of the Image-100. The areas of each of these classes were selected carefully so that they could be considered to be representative of the respective land use classes. Assuming that each of these classes has a multivariate gaussian distribution, the B-distance, based on Bhattacharyya coefficient, was calcilated between all possible pairs of chese classes in all possible combinations of one, two, three and four spectral chanacls using the feature selection algorithm of the bratilian Institute of Space Research (INPE) on-line-mode with the Image-100g9 For each value of b-distance, che probability of correct classification was reasonably estimated from the curve of Swain

ORIGINAL PAGE is OF POOR QUALITY

evaluate the utility of satellite data for urban land use classification. Several land use classes, such as commerce/im dustry, single-iamily (Newer) residential, trees, and water exhibited spectrally separable characteristics and were identified with greater than 90 percent accuracy, Fllefsen et $a 1^{2}$. (1973) did computer-aided analysis of LANDSAT MSS data of the San Prancisco bay area. Smith et al. ${ }^{3}$ (19/4) have given the application of spatial features to satellite land-use analysis. Ellefson et a1.4 (1974) have given new techniques in mapping urban laad use and monitoring change for selected U.S, metropolitan areas. They analysed LAADDAT Mss data using automatic pattern recognition techniques for classification. Kumar and Silvas (1977) have analysed the statistical separability of agricultural cover types in much detail, data quantity and depth in the subsets of one to twelve spectral chamels.

Economy et $a: .{ }^{6}(1974)$ have given the classification accuracy of the Image-100 using single cell option (described later in the report). They reported an accuracy of $70 \pm 12 \%$. Factors which limited the classification accuracy were spatial resolution of the ERTS-1 (now called LANDSAT) scanner, the land-use classification system, and inaccuracies in the eround truth. Goldberg et. $11 .^{7}(1975)$ have described methods and procedures which outside investigators may use with the automated processing equipment of the Canada Centre for Remote Sensing (CCRS) for the purpose of natural resource exploration and mapping. They have compared the accuracies of unsupervised and supervised methods on the basis of the confusion matrices gencrated by classifying exactly the sane area.

## Method of Analysis

Multispectral scanner data of computer compatible tapes of LANDSAT were analysed using Image-100. With the aid of land use map of São José dos Compos mentioned above, reetangular areas of each of the above seven land use classes were selected, avoiding the boundaries of classes on the display of the Image-100. The areas of each of these classes were selected carefully so that they could be considered to be representative of the respective land use classes. Assuming that each of these classes has a multivariate gaussian distribution, the B-distance, based on Ehattacharyya coefficient, was calculated between all possible pairs of these classes in all possible combinations of one, two, three and four spectral chantels using the feature selection agorithm of the Brazilian Institute of Space Research (INPE) on-line-mode with the Image-100,9 For each value of B-distance, the probability of correct classification was reasonably estimated from the curve of Swain

ल
and King (1973) 10. The B-distance for two probability dentitite; $P_{1}(x)$ and $r_{2}(x)$ is given liy

$$
\begin{equation*}
18 u \int_{x}\left[\sqrt{p_{1}(x)-\sqrt{p_{2}}(x)}\right]^{2} d x \tag{1}
\end{equation*}
$$

If $P_{1}(x)$ and $p_{2}(x)$ are maltivariate gansian distributions as nbove, then

$$
\begin{equation*}
B=2\left(1-e^{-\alpha}\right), \tag{2}
\end{equation*}
$$

where
where $U_{1}$ and $U_{2}$ are mean vectors of classes one and two respectively; whereas $\Sigma_{1}$ and $\Sigma_{2}$ are the covariance matricus of classes one and two respectively.

$$
\begin{equation*}
\Sigma \boxminus \frac{1}{2}\left[\Sigma_{1}+\Sigma_{2}\right] \text { and } T \text { denotes cranspose } \tag{4}
\end{equation*}
$$

B-distance i.s defined for two distributions. Remote sensing usually involves more than two classes. One strategy is to compute the average b-distance over all pairs of classes and select the subset of features for which the average B-distance is maximum, i.e., maximize, with respect te ail n-tuples

$$
\begin{equation*}
B_{A V E}\left(c_{1}, \bar{c}_{2}, \ldots c_{n} j-\frac{2}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} 3\left(i, j \mid c_{1}, c_{n}, \ldots c_{n}\right)\right. \tag{5}
\end{equation*}
$$

where

$$
m \text { number of classes }
$$

$B\left(i, j \mid C_{1}, C_{2}, \ldots, C_{n}\right)=B-d i s t a n c e$ between classes $i$ and $j$ in the channeis $C_{1}, C_{2}^{\text {n }}, \ldots, C_{n}$.
and King (1973)10. The B-distance for two probability densities $p_{1}(x)$ and $P_{2}(x)$ is given by

$$
\begin{equation*}
B=\int_{x}\left[\sqrt{p_{1}(x)}-\sqrt{p_{2}(x)}\right]^{2} d x \tag{1}
\end{equation*}
$$

If $P_{1}(x)$ and $P_{2}(x)$ are multivariate gaussian distributions as above, thea

$$
\begin{equation*}
B=2\left(1-e^{-\alpha}\right), \tag{2}
\end{equation*}
$$

where

$$
a=\frac{1}{8}\left(U_{1}-U_{2}\right)^{T} \Sigma^{-1}\left(U_{1}-U_{2}\right)+\frac{1}{2} \log _{e}\left[\frac{\operatorname{det} \Sigma}{\sqrt{\operatorname{det} \Sigma_{1} \cdot \operatorname{det} \Sigma_{2}}}\right](3
$$

where $U_{1}$ and $U_{2}$ are mean vectors of classes one and two respectively; whereas $\Sigma_{1}$ and $\Sigma_{2}$ are the covariance matrices of classes one and two respectively.

$$
\begin{equation*}
\Sigma=\frac{1}{2}\left[\Sigma_{1}+\Sigma_{2}\right] \quad \text { and } T \text { denotes transpose } \tag{4}
\end{equation*}
$$

b-distance is defined for two distributions. Remote sensing usually involves more than two classes. One strategy is to compute the average $B$-distance over all pairs of classes and select the subset of features for which the average $B$-distance is maximum, i.e., maximize, with respect to ail n-tuples

$$
\begin{equation*}
B_{A V E}\left(c_{1}, c_{2}, \ldots c_{n}\right)=\frac{2}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} B\left(i, j \mid c_{1}, C_{2}, \ldots C_{n}\right) \tag{5}
\end{equation*}
$$

where
$\mathrm{m}=$ number of classes
$B\left(i, j \mid C_{1}, C_{2}, \ldots, C_{n}\right)=B$-distance between classes $i$ and $j$ in the channels $C_{1}, C_{2}^{\mathrm{n}}, \ldots, C_{n}$.

ORLGINAI, PAU.
OE POOR SUALITY

BAVE was computed for all possible subsets of one, two, three amd four spectral chamels, out of available four channels.

Bach of these land use elasses was divided into two independent sets: training fields and test fields. Using training fields of residential areas, test fields of each of the above seven classes and training, fields of each of the classes except residential areas were classified, using the singlowcell signature acquisition option of liage-100. This option creates a four-dimensional rectangular parallelepiped, each of the sides, of which correspond to the signature limits of the training. areas in each channel.the number of pixels classified as residential areas by the computer inside the test fields of each of these seven classes were determined. An identieal analysis was repeated for cach of the other six land use classes. Thus, a confusion matrix showing the total number of pixels (picture elements) of each class classified correctly as vell as classified incorrectly into each of the other chasses was obtained.

This whole procedure was repeated for the multicell signature acquisition option of the Image-l00. In the multicell signature aequisition, the parallelepiped of spectral signature is subdivided into cells each of unit volume and the number of pixels in each of these unit cells is counted. These cell combs are thus measures of the probability distribution of the spectral cluster. By raising or lowering the threnhold on the cell counts, one can vary the size and four dimensional probability distribution of the spectral cluster, by deleting or adding cells with counts greater than a variable threshold. In the interactive signature modification option, the user performs training on the misclassilied area adding the errors of omission and subtracting the errors of comatssionuntil satisfied with the results. The preliminary results showed that this option did not seen to improve much the classification accuracy of the 'single celi' option, because the basic problem was the overlap between the classes in the four-dimensional spectral space.

The same craining fields of each class were used to classify the test fields using the sample classifier based on B-distance. b-distance was computed between a test ficld and: each of the seven training classes and was classified into the class for which the B-distance was minimum.
${ }^{\mathrm{B}}$ AVE was computed for all possible subsets of one, two, three and four spectral chanals, out of available four channels.

Each of these land use classes was divided into two independent sets: training fields and test fields. Using training fields of residential areas, test fields of each of the above seven classes and training fields of each of the classes except residential areas were classified, using the single-cell signature acquisition option of Image-100. This option creates a four-dimensional rectangular parallelepiped, each of the sides of which correspond to the signature limits of the training areas in each channel. The number of pixeis classified as residential areas by the compucer inside the test fields of each of these seven classes were determined. An identical analysis was repeated for each of the other six land use classes. Thus, a confusion matrix showing the total number of pixels (picture elements) of each class classified correctly as well as classified incorrectly into each of the other ciasses was obtained.

This whole procedure was repeated for the multicell signature acquisition option of the Image-100. In the multicell signature acquisition, the parallelepiped of spectral signature is subdivided into cells each of unit volume and the number of pixels in each of these unit cells is counted. These cell counts are thus measures of the probability distribution of the spectral cluster. By raising or lowering the threshold on the cell counts, one can vary the size and four dimensional probability distribution of the spectral cluster, by deleting or adding cells with counts greater than a variable threshold. In the interactive signature modification option, the user performs training on the misclassified area adding the errors of omission and subtracting the ertors of comaissionuntil satisfied with the results. The preliminary results showed that this option did not seen to improve mach the elassification accuracy of the 'single cell" option, because the basic problem was the overlap between the classes in the four-dimensional spectral space.

The same training fields of each class were used to classify the test fields using the sample classifier based on B-distance. B-distance was computed between a test field and. each of the seven training classes and was classified into the class for which the B-distance was minimum.

## $|\cdots \cdots|$

Results and Conclusions
Table 1 shows the distribution of pisels of the raining and the test fields of each class. Tables 2 to 5 give the values of $\mathrm{B}_{\mathrm{y}}$ in all possible combinations of one, two, three and four chanticls respectively out of four available chamels. As one would expect, the probability of correct classification increases with an incrense in the number of channelis. In the bubsets of one to three spectral channels: chamel 4 , chamels ( \& 7 (one in the visible and one in the near infrared), channels 4,5 \& 7 (two in the visible and one in the near infrared hre found to be the best choices. Table 2 shows that in the subbet of two channels, chamels 4 and 5 (visible wavelength region) give higher probability of correct classification than channels of 8 . (near infrared wavelength region). The authors believe chat each wavelength region - visible, near infrared, cifdde infeared and thermal infrated has valuable spectral information. Thus, in the subset of two spectral channese, ane chamel in the visible and one channel in the near infrired wavelength region are found to be the best choice. Kumar (1977) ${ }^{11}$ has analysed the aircraft MSS data in much detail, data quantity and depth in the subsets of one to tivelve spectral channels, to evaluate each spectral channel as well as possible combinations of wavelength regions, for statistical separability of agricultural cover types.

Using the training fields of a class, the number of pixels classified by the Inage- 100 into each of the seven classes (training + test fields, except the class being used to train the computer in which case only test ficlds) are given in Table 6. From Table 6 , the errors of omission (while using traifing fields of residential, number of pizels of test fields of residential not classified as residential constitute the errors of omission, etc.) and the errors of commission (while using training fields of residential, number of pixels of classes other than residential but which are actually classifjed les residential) ware calculated and are shown in Thable 7. Similarly, the errors of omission and comassion using the multicell signature acquisition ( $m=1, m=2$ and $m=3$ ) for the lsame training and test fields of each class were calculated and are given in Table 7. The option $m=1$ means that all the unit cells in the four dimensional spectral space which had.. less than one pixel were deleted from the spectral signature fof the training fields for doing classification. Similarly, the option $\mathrm{m}=2$ means that all the unit cells in the four dimensional spectral space which had less than two pixels were deIleted from the spectral sipnature of the craining fields for doing classificetion etc. trable 7 shows that for the singledeell option, the errors of omission vary from $5 \%$. Cor the class

## Buantrpala Results and Conclusions

Table 1 shows the distribution of pizels of the training and the test fields of each class. Tables 2 to 5 give the values of $B$, in all possible combinations of one, two, three and four chamiels respectively out of four available channels. As one would expect, the probability of correct classification increases with an increase in the number of channels. In the subsets of one to three spectral channels: channel 4 , channels 487 (one in the visible and one in the near infrared), channels 4,$5 ; 7$ (two in the visible and one in the near infrared are found to be the best choices.Table 2 shows that in the subset of two channels, chamels 4 and 5 (visible wavelength region) give higher probability of correct classification than channels $6 \& 7$ (near infrared wavelength region). The authors believe that each wavelength region -- visible, near infrared, aiddle infrared and thermal infrared has valuable spectral infommation. Thus, in the subset of two spectral channels, ene chanel in the visible and one channel in the near infrared wavelength region are found to be the best choice. Kumar (1977) ${ }^{11}$ has analysed the aircraft MSS data in much detail, data quantity and depth in the subsets of one to tweive spectral channels, to evaluate each spectral channe 1 as well as possible combinations of wavelength regions, for statistical separability of agricultural cover types.

Using the training fields of a class, the number of pixels classified by the Inage-100 into each of the seven classes (training + test fields,except the class being used to train the computer in which case only test fields) are given in Table 6. From Table 6, the errors of omission (while using training fields of residential, number of pixels of test fields of residential not classified as residential constitute the errors of omission, etc.) and the errors of commission (while using training fields of residential, number of pixels of classes other than residential but which are actually classified as residential) were calculated and are shown in Table 7. Similarly, the errors of omission and comaission using the multicell signature acquisition $(m=1, m=2$ and $m=3$ ) for the same training and test fields of each chass were calculated and are given in Table 7. The option $m=1$ means that all the unit cells in the four dimensional spectral space which had. less than one pixel were deleted from the spectral signature of the training fields for doing classification. Similarly, the option $n=2$ means that all the unit cells in the four dinensional spectral space which had less than two pixels were deleted from the spectral sisnature of the training fields for doing classification ete. Table 7 shows that for the singlecell option, the errors of onission vary from $5 Z$ for the class


ORIGINAL PAGE IS
OF POOR OUALITY

Table 2. Valucs of $\mathrm{B}_{\mathrm{AV} \text { : }}$ in subsct of one channe i

| Chame 1 | ${ }^{\text {A }}$ AV: | Estimated percentage probability of correct classification (swain and King, 1973) |
| :---: | :---: | :---: |
| 4 | 0.7787 | 82.4 |
| 5 | 0.7250 | $\because 81.0$ |
| 6 | 0.4267 | $\cdots 172.8{ }^{\text {a }}$ |
| 7 | 0.3510 | 169.6 |

Table 3. Values of BAVE in subset of two channels


Table 4. Values of $B_{A V E}$ in subset of three chamelis.

| Channels | $B_{\text {AVE }}$ | Estimated.... | (Swain and King, 1973) |  |
| :---: | :---: | :---: | :---: | :---: |
| $4-5-6$ | 0.9920 | $\therefore$ | 87.2 | . |
| $4-5-7$ | 1.0449 | 88.0 | . |  |
| $4-6-7$ | 1.0186 | 87.2 | . |  |
| $5-6-7$ | 0.9504 | 85.7 |  |  |

Table 5. Values of ${ }^{B}$ AVE in the available four channels

| Chame Is | ${ }^{1}$ NVE | Estimated. . . (Swain and King, 1973) |  |
| :---: | :---: | :---: | :---: |
| $4-5-6-7$. | 1.0940 | 89.0 |  |

Note: $B_{A V E}$ is defined in eq. (5).

Table 2. Values of $\mathrm{B}_{\mathrm{AVE}}$ in subset of one channel

| Chamel | BAVE | Letimated perceatage probability of cor- |  |
| :---: | :---: | :---: | :---: |
| rect classification (Swain and King, 1973) |  |  |  |
| 4 | 0.7787 | 82.4 |  |
| 5 | 0.7250 | 81.6 |  |
| 6 | 0.4267 | 72.8 |  |
| 7 | 0.3510 | 69.6 |  |

Table 3. Values of BAVE in subset of two channels


Table 4. Values of $\mathrm{B}_{\mathrm{AVE}}$ in súbset of three channels

| Channels | $B_{\text {AVE }}$ | Estimated.... (Swain and King, 1973) |
| :---: | :---: | :---: |
| $4-5-6$ | 0.9920 | 87.2 |
| $4-5-7$ | 1.0449 | 88.0 |
| $4-6-7$ | 1.0186 | 87.2 |
| $5-6-7$ | 0.9504 | 85.7 |

Table 5. Values of $\mathrm{B}_{\mathrm{AVE}}$ in the available four channels

| Channe 1 s | ${ }^{3}$ AVE | Estimated. . . (Swain and King, 1973) |
| :---: | :---: | :---: |
| $4-5-6-7$ | 1.0940 | 39.0 |

Note: $B_{A V E}$ is defined in eq. (5).
ORIGINAL PAGE IS OF POOR QUALITY
Table 6 Number of pixels classified into each class using 'single cell' option
Using training $\quad$ Considering training and test fields of all classes except training
areas of number of pixels classified by lnage-100 into the ciass minus diagonal element 2
2163
2763
829
3471
2575
1422
1638




408
497
203
615
507
258
361

| Using training areas of | Considering training and test fields of all classes except training fields of the class being used to train the computer for classification, number of pixels classified by Image-100 into the class |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Res. | Nult. | Commi | Ind. | Ins. | Ag. | Uno. | Sumation of minusdiagon i |
| Res. | 876 | 48 | 270 | 430 | 408 | $17^{\circ}$ | 114 | 2163 |
| Mulc. Res. - 1 | 1206 | 128 | 137 | 872 | 497 | 0 | 51 | 2763 |
| Comm. | 160 | 45 | 288 | 230 | 203 | 0 | 190 | 829 |
| Ind. 1 | 1364 | 180 | 219 | 1231 | 615 | 6 | 67 | $3: 71$ |
| Ins. - 1 | 1219 | 88 | 240 | 589 | 507 | 49 | 390 | 2575 |
| Ag. | 8 | 13 | 31 | 93 | 258 | 866 | 1019 | 1422 |
| Uno. | 78 | 28 | 76 | 175 | 361 | 970 | 995 | 1638 |
| Sumnation of $\begin{array}{lllllllll}\text { column ainus } & 4911 & 403 & 973 & 2389 & 2342 & 1042 & 1852\end{array}$ |  |  |  |  |  |  |  |  |
| ciagonal element 0.63 0.05 |  |  |  |  |  |  |  |  |
| Previous row - | 0.63 | 0.05 | 0.13 | 0.38 | 0.35 | 0.16 | 0.29 |  |
| (test and rraining |  |  |  |  |  |  |  |  |
| classes except the |  | Res. = Residential |  |  |  |  | Comm. - CommercialUno. $=$ Unoccuried |  |
| class of diagonal |  |  |  |  |  |  | Ind. | ustrial |
| element) |  | Muti.Res. $=$ Multifamily residential |  |  |  |  |  |  |


lable $\%$. percentage errors of omigsion and comission. in 'aingle cell' and 'multiceli' signature acquisition
(A) Percentage urrors of mission

| Class | Single-cell | $\begin{gathered} \text { Muluicell } \\ m=1 \end{gathered}$ | $\underset{m=2}{\text { Nultirel }}$ | $\begin{gathered} \text { Multi-cell } \\ m=3 \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: |
| Residential | 26.8 | 63.21 | 73.91 | 83.61 |
| NuLti-fimily Residential | 33.3 | 80.21. | 88.54 | 94.27 |
| Commercial | 16.3 | 74.71 | 78.49 | 83.43 |
| Industrial | 5.0 | 68.44 | 80.00 | 90.00 |
| Institutional | 46.3 | 59.32 | 82.56 | 87.50 |
| Agricultural | 19.4 | 73.46 | 80.26 | 86.69 |
| Unoccupied | 16.9 | 77.46 | 82.30 | 85.40 |

(B). Yerenatage errors of commission

| Class | Single cell | $\begin{aligned} & \text { Muliti-cell } \\ & m=1 \end{aligned}$ | $\begin{gathered} \text { Multi-cell } \\ m=2 \end{gathered}$ | $\begin{gathered} \text { Multi-ce } 11 \\ m=3 \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: |
| Residential | 25.28 | 5.32 | 3.38 | 2.43 |
| Multi-family Hesidential | 35.64 | 0.92 | 0.27 | 0.01. |
| Cormercial | 10.99 | 1.75 | 0.81 | 0.41 |
| Industrial | 39.00 | 11.08 | 1.86 | 1.59 |
| Insticutional | 38.00 | 26.65 | 4.48 | 2.57 |
| Agriculcural | 21.41 | 5.20 | 4.02 | 3.20 |
| Unaccupied | 25.99 | 2.97 | 1.58 | 1.23 |

Note: The option $m=1$ meat.; that all the unit cells in the four dimensional spectral space which had less than one pixel were deleted from the apectral signature of the training fields for doing classification.

Table 7. Percentage errors of omission and comission. in 'single ce11' and 'multice11' signature acquisition
(A) Percentage errors of omission

Class
Residential
Multi-family Residential
Comercial
Industrial
Institutional
Agricultural
Unoceupied


Mu
Ma
a
as $=1$
63.21
Multi-ce11
m $=2$
73.91
Multi-ce 11
26.8 63.21
80.21
88.54
$\mathrm{m}=3$
83.61
33.3
74.71
78.49
94.27
16.3
68.44
86.00
83.43
5.0
46.3
59.32
82.56
87.50
19.4
16.9
73.46
77.46
80.26
86.69
(B) Percentage errors of commission

Class
Residential
Multi-family kesidential
Cormercial
Industrial
Institutional
Agricultural
Unoccupied

Single cell
25.28
35.64
10.99
39.00
38.00
21.41
25.99

Multi-cell Multi-cell
$\mathrm{n}=1$
5.32
0.92
1.75
11.08
26.65
5.20
2.97

0.27
0.81
1.38
4.48
4.02
ii $=2 \quad$ mi $=3$
$3.38 \quad 2.43$
0.01
0.41
1.59
2.57
3.20

Note: The option $m=1$ mean; that all the unit cells in the four dimensional spectral space which had less than one pixel were deleted from the spectral signature of the training ficlds for doing classification.
induatrial to $46 \%$ for the chaw institutional. The errors of comission vary from ill\% for the class comane cial to $39 \%$ for the class.industifal. It shows that classification accuracy for all the chasses is rather poor except the class comercial where the pereentage of errors are reasonably small (errors of omiscion $=16.3 \%$, comatission $=11 \%$ ). This is because of small values of standard deviation (and hence less overlap) of this class in each of the spectral chamels especially in the channels one ( 0.5 to $0.6 \mu \mathrm{~m}$ ) and [our ( 0.8 to $1.1 \mu \mathrm{~m}$ ). In general, an increase fin the standard deviations of a class in the spectral channels tends to reduce the errors of onission and increase the errors of comission. it was found that taking into account both the errors of omission as well as the errors of commission, the classification accuracy generally reduces wi.th an increase in the standard deviations. The sum of standard deviations for the class industrial in the four spectral channels were much higher as compared to the sum of standard deviations for any of the other chasses. The intervals of its spectral response in eaci of the four spectral chanmels were very large and that is why it has low etrors of omission. The errors of onission as well as commission for the class institutional are very large. An institution can have ofifice buildings,parking areat, recreation areas, housing for amployees, workshops, trees, etc. Thus, it would require' a large numbe: of training areas for them to be truly representative of the chass institutional. The authors believe that the training areas of the class institutional were perhaps not truly representative of the elass institutional and hence, the errors of omission as well as commission were found to be rather large.

The first element of the last row of table 6 shows the cotal number of pizels that have been incorrectly classified as class residential divided by the total number of training and test fields of all the classes except the class residential. In this calculation, it should be noted that many pixels may have been classified incorrectly thore than once as a class residential. Similarly, the other elements of the last row are self esplanatory.

Table 7 shows as expected, that the multicell option increases the errors of omission and decreases the errors of conmission. Considering the errors of omission as well as. : atrors of commssion, multicell option for $m=1$ considerably decruases the percentage of corvect classification for each of the classes. This is because the number of pixels used for training of each class were relatively thall for statistical purpose.Thus, the unit cells in the four dimensional spectral spece ware sparsely populated. Thus, thote may be many cells

ORIGINAL PAGE IS OF POOR QUALITY

industrial to $46 \%$ for the class institutional. The errors of comission vary from $11 \%$ for the class comancrial to $39 \%$ for the class indust:al. It shows that classification accuracy for all the elasses is rather poor except the class comercial where the percentage of errors are reasonably small (errors of omission $=16.37$, comission $=11 \%$ ). This is because of small values of standard deviation (and hence less overlap) of this class in each of the spectral channels especially in the channels one ( 0.5 to $0.6 \mu \mathrm{~m}$ ) and four ( 0.8 to $1.1 \mu \mathrm{~m}$ ). In general, an increase in the standard deviations of a class in the spectral channels tends to reduce the errors of omission and increase the errors of comission. It was found that taking into account both the errors of omission as well as the errors of comaission, the classification accuracy generally reduces with an increase in the standard deviations. The sum of standard deviations for the class industrial in the four spectral channels were wuch higher as compared to the sum of standard deviations for any of the other classes. The intervals of its spectral response in each of the four spectral channels were very large and that is why it has low errors of omission. The errors of onission as well as commission for the class institutional are very large. An institution can have office buildings,parking areas, recreation areas, housing for employees, workshops, trees, etc. Thus, it would require a large numbe: of training areas for them to be truly representative of the class institutional. The authors believe that the training areas of the class institutional were perhaps not truly representative of the class institutional and hence, the errors of omission as well as commission were found to be rather large.

The first element of the last row of Table 6 shows the total number of pizels that have been incorrectly classified as class residential divided by the total number of training and test fields of all the classes except the class residential. In this calculation, it should be noted that many pixels may have been classified incorrectly more than once as a class residential. Similarly, the other clements of the last row are self explanatory.

Table 7 shows as expected, that the multicell option increases the errors of omission and decreases the errors of comission. Considering the errors of omission as well as : errors of comaission, multicell option for $m=1$ considerably decreases the percentage if correct classification for each of the classes. This is because the number of pixels used for training of each ciass were relatively small for statistieal purpose. Thus, the unit cells in the four dinensional spectral purpose were sparsely populated. Thus, there may be many cells

Which are actually representative of the class but do not have any pixel because the total number of pixels for training for each of the elastes was rather small. Similarly, for the multicell option, the errors of onission increase and the errors of comission decrease 3 lawly as we go fromm a 1 to $m=2$ to $m \times 3$. Considering the ertors of omission as well as the errors of conmission, the percentage of cortect classification decreases slowly as we go frommal to $m=2$ to $m$ a 3 .

Table 8 shows the results of classification using a sample chassifier. As pointed out earlier, the same training and test fields were used in this case as in the "single cell" or "multicell" option of the Image-100. Unfortunately, one field of the class institutional could not be classified using sample classifier due to temporary malfunctioning of the Image-100 at that time. Comparing these results to the single-cell option, we find that the errorg of commission are reduced significantly, whereas the errors of omiasion are. increased for some classes and decreased for the obhers. The errors of omission for the class industrial are very large because the standard deviation of the class industrial was extremely large in each of the four spectral channels. For most classes, the larger were the standard deviations, the $10-$ wer was the percentage of correct classification using,sample classifiu. Comparing sample classifier to the bulticell option ${ }^{i}$ Find that it gives much swaller errors of omission, wherens it gives greater errors of commission for sone ciasses and smaller for the others. On the whole, the sample classifier gives percentage of correct classification much better than the single-cell or multi-cell options of the Image100. Similar analysis will be done using the 'pixel by pi*el maximun likelihood gaussian classifier'. In addi:ion, computer compatible tapes of S.J. dos campos of other times will be analysed to investigate the effect of tine on these results

The authors gratefully acknowledge the assistance of Lr. Nelson de Jesus Parada, Director of the Instituto de Pesgutsas Espaciais (INPE); Dr. Celso de Renna e Scuza, Mr. J.C. Moreira and Mrs. M.S.S. Barros of INPE for their assistance with this work.

Which are actually representative of the class but do not have any pixel because the cotal number of pixels for training for each of the classes was rather small. Similarly, for the multicell option, the errors of omission increase and the errors of comission decrease slowly as we go fromm $=1$ to $m=2$ to $m=3$. Considering the errors of omission as well as the errors of commission, the percentage of correct elassification decreases slowly as we go from m = 1 to $m=2$ to $m=3$.

Table 8 shows the results of classification using a sample classifier. As pointed out earlier, the same training and test fields were used in this case as in the "single cell" or "multicel1" option of the Image-100. Unfortunately, one field of the class institutional could not be classified using sample classifier due to temporary malfunctioning of the Image-100 at that time. Comparing these results to the single-cell option, we find that the errors of commission are reduced significantly, whereas the errors of omission are increased for some classes and decreased for the whers. The errors of omission for the class industrial are very larece because the standard deviation of the class industrial was extremely large in each of the four spectral channels. For most classes, the larger were the standard deviations, the 10wer was the percentage of correct classification using sample classif. Comparing sample classifier to the 'multicell option' Find that it gives much sualler errors of omission, whersas it gives greater errors of commission for some classes and smaller for the others. On the whole, the sample classifier gives percentage of correct classification much better than the single-celi or multi-celloptions of the Image100. Similar analysis will be done using the 'pisel by pixel maximm likelihood gaussian classifier'. In addition, computer compatible tapes of S.J. dos Campos of other times will be analysed to investigate the effect of tine on these results

The authors gratefully acknowledge the assistance of Dr. Nelson de Jesus Parada, Director of the Instituto de Pesquísas Espaciais (INPE); Dr. Celso de Renna e Scuza, Mr. J.C. Moreira and Mrs. M.S.S. Barros of INPE for their assistance with this work.

> ORLGLNA PABL IS
> or Puols wrther'
.......


ORIGINAL PAGE IS
$\square$



