SPATIAL CLASSIFICATION ERROR RATES
RELATED TO PIXEL SIZE

by
Paul Switzer
and
Achilles Venetoulias
Stanford University

TECHNICAL REPORT NO. 11
AUGUST 1985

PREPARED UNDER THE AUSPICIES
OF
NATIONAL SCIENCE FOUNDATION
GRANT MCS 81-09584

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
SPATIAL CLASSIFICATION ERROR RATES
RELATED TO PIXEL SIZE

by
Paul Switzer
and
Achilles Venetoulis
Stanford University

TECHNICAL REPORT NO. 11
AUGUST 1985

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION
GRANT MCS 81-09584

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Spatial Classification Error Rates  
Related to Pixel Size  

Paul Switzer  
and  
Achilles Venetoulias  
Stanford University  

1. Introduction.

Suppose that a spatial domain is partitioned, in an unknown way, into classes I and II with areal proportions $Q$ and $1-Q$ respectively. The same domain is partitioned into an array of observation pixels. For the pixel centered at $t$ we observe $X(t)$, the value of a function integrated over that pixel; the observations $X(t)$ may be vector-valued. A simplified model for $X(t)$ is:

$$X(t) = a + b \cdot p(t) + e(t)$$  \hspace{1cm} (1)

where

$p(t)$ is the unknown areal proportion of the pixel $t$ which belongs to class I  
$e(t)$ is a mean zero residual  
$a, b$ are constants.

In one common situation "training" data are available for pixels where it is known that $p(t)$ is zero or one, i.e. for pixels which are homogeneous and known to belong to either class I or class II. From such data one may estimate in a standard fashion the constants $a$ and $b$ and the stationary covariance of $e(t)$. For the illustrative purposes of this paper we assume these estimates are obtained without error and furthermore that $e(t)$ is a gaussian spatial process.

The goal is to classify the central point of the pixel at $t$. If we do not exploit any spatial structure and pretend that pixels are class homogeneous then the optimal classifier for $t$ based on $X(t)$ is the Fisher linear discriminant function appropriately modified by the prior probabilities $Q$ and $1-Q$. Then there is no further loss of generality by taking $X(t)$ to be univariate, $a = 0$, $b > 0$. The parameter $b$ may then be interpreted as the distance between class means. Using the simplifications available for this data model, our main task will be to
examine the probability for misclassification as a function of pixel size.

Larger pixels will have the property that the residual or noise component of the observation, \( e(t) \), will have smaller variance, but at the same time larger pixels are less likely to be class-homogeneous. These properties work oppositely with respect to classification error probabilities even with computational questions aside. For modest increases in pixel size it will turn out, for the most part, that larger pixels are better than smaller pixels. Furthermore, it will be interesting to note how error probabilities depend on the parameters of the simplified descriptions of the underlying spatial class pattern and the observation model.

The pixel sizes we consider are of the form \( N \times N \) where \( N \) is an integer greater than or equal to unity. Thus if the data are available at a resolution of \( 1 \times 1 \) then we may form larger pixels by averaging the available data over \( N \times N \) neighborhoods. Overall, the qualitative effect of local averaging prior to application of the discriminant function is to remove isolated misclassifications in the interior of contiguous homogeneous regions at the expense of an increased classification error rate near the boundary between the two classes. To quantify this error rate trade-off a simple probabilistic model is used for the local behavior of the class boundary.

Other approaches to incorporating spatial information into classification algorithms may attempt to first identify the location of any boundary segment which may be present within a neighborhood of unit pixels as was done by Dehnad (1984) for example. Related approaches attempt to estimate the probabilities of various class-type configurations within pixel neighborhoods as was done by Owen (1985), Hjort (1985), and Haslett (1985) for example.


We will act as though a pixel is either class-homogeneous or else it is class-heterogeneous and is cut by a class boundary which is a line segment. The probability measure for this boundary segment is taken to be the affine-invariant measure which is a uniform measure, with intensity parameter \( v \), on the space of polar coordinates of a line. See Solomon (1978) for a very good discussion of this random line measure. For pixels which are not too large relative to the scale of class alternation this probability model should be reasonable; it has been used in the classification context by Switzer (1983) and others.
With this model for simplified class boundaries within pixels we may now regard \( p(t) \), the class I proportion for the pixel at \( t \), as a random variable independent of the noise random variable \( e(t) \). The distribution of \( p(t) \), given that the central point of the pixel at \( t \) belongs to class I and that the pixel is class-heterogeneous \( [1/2 \leq p(t) < 1] \), may be written as [see Appendix]

\[
G(p) = \text{Prob}\{p(t) \leq p\} = \left\{ \frac{(1-u^2)}{\sqrt{1+u^2}} - \sqrt{\frac{u^2}{2}} \int_0^\theta(u) \sqrt{\cos t} dt \right\}
\]

where \( u = 2(1-p) \) and \( \theta(u) = \arctan \left( \frac{1-u^2}{u} \right) \). \hspace{1cm} (2)

This conditional distribution of \( p(t) \) does not depend on any of the three parameters \( N, \nu, \) and \( Q \), respectively pixel size, boundary intensity, and prior class probabilities.

The distribution of \( e(t) \), the noise component of an observation, will be gaussian with mean zero and standard deviation depending on the pixel size \( N \) and the noise autocorrelation structure. In the example calculations of this paper the pointwise values of the noise are taken for convenience to have isotropic spatial correlation

\[
r(d) = \rho^d \quad \text{where} \quad \rho > 0 \quad \text{is a persistence parameter and} \quad d \quad \text{is distance}.
\]

The normalization we use is that pixels of unit size always have unit standard deviation for \( e(t) \). Now let

\[
V(N, \rho) = \int_U \int_U \rho^{N|t-s|} d\mu(s) d\mu(t)
\]

where \( U \) is a unit square and \( \mu \) is Lebesgue measure in the plane. Then the standard deviation of the noise \( e(t) \) for pixels of size \( N \times N \) may be written as

\[
S(N, \rho) = [V(N, \rho)/V(1, \rho)]^{1/2}.
\]

(4)

The classification rule we consider here for classifying the pixel at \( t \) is the common Bayes rule which uses only the data \( X(t) \) and which ignores the possibility of a boundary crossing, i.e. the rule assumes naively that \( p(t) = 0 \) or \( p(t) = 1 \). The rule is the usual modified linear discriminant, viz., assign the central point of the pixel at \( t \) to class I if \( X(t) > L(b, N, \rho, Q) \) where

\[
L(b, N, \rho, Q) = \frac{1}{2} b + \frac{S^2(N, \rho)}{b} \ln \left( \frac{1-Q}{Q} \right).
\]

(5)
When the class prior probabilities are equal, \( Q = 1/2 \), then the rule depends only on the class mean separation \( b \), but not on \( N \) or \( \rho \). However, the classification error rate will depend on all parameters.

For the above classification rule it is now possible to calculate the probabilities of misclassifying a class I point, both in the case of a class-homogeneous pixel and in the case of a class-heterogeneous boundary pixel. These two error rates are given respectively below and may be derived simply by first conditioning on \( p(t) \).

\[
E_H = \Phi[(L(b, N, \rho, Q) - b)/S(N, \rho)]
\]

\[
E_B = \int_{1/2}^{1} \Phi[(L(b, N, \rho, Q) - u \cdot b)/S(N, \rho)]dG(u)
\]

where \( G(u) \) is given by (2).

Overall classification error rates are then calculated using the relative frequencies of homogeneous and heterogeneous boundary pixels. These relative frequencies depend on an intensity parameter \( u \) of the boundary line measure, the pixel size \( N \), and the prior class probability \( Q \). If the pixels are not too large relative to the alternation of the class pattern then it may be safe to ignore the probability of more than one boundary crossing in a pixel. Then the boundary model gives the approximate relative frequency of heterogeneous pixels, centered at class I, as

\[
f(u, N, Q) = [1 - Q][1 - \exp(-uN)].
\]

The next section contains calculations of classification error rates for various combinations of the parameters \( b, N, \rho, Q, u \) in a possibly too parsimonious model. However, these calculations do show the extent to which error rates might be sensitive to distance between class means, pixel size, spatial noise autocorrelation, prior probabilities, and complexity of the class boundary.

3. Calculations.

The standard deviation of the noise component of an observation, \( S(N, \rho) \), was computed using (4) for values of the unit-distance autocorrelation \( \rho = 0.00, 0.14, 0.37, 0.55, \) and \( 0.82 \) and pixel sizes \( N \times N \) for \( N = 3, 5, 7, 9, 11 \). Recall that the normalization convention makes the standard deviation for pixels of unit size equal to unity for all values of \( \rho \). The table of standard deviations is given below.
Table 1. Standard Deviations of Observations

<table>
<thead>
<tr>
<th>$N$ = 1</th>
<th>1.0</th>
<th>1.0</th>
<th>1.0</th>
<th>1.0</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.33</td>
<td>55</td>
<td>.67</td>
<td>.76</td>
<td>.91</td>
</tr>
<tr>
<td>5</td>
<td>.20</td>
<td>.34</td>
<td>.48</td>
<td>.60</td>
<td>.82</td>
</tr>
<tr>
<td>7</td>
<td>.14</td>
<td>.26</td>
<td>.38</td>
<td>.50</td>
<td>.75</td>
</tr>
<tr>
<td>9</td>
<td>.11</td>
<td>.21</td>
<td>.31</td>
<td>.42</td>
<td>.69</td>
</tr>
<tr>
<td>11</td>
<td>.09</td>
<td>.17</td>
<td>.26</td>
<td>.36</td>
<td>.64</td>
</tr>
</tbody>
</table>

$\rho$ is the unit distance autocorrelation

$N \times N$ are the pixel dimensions

The calculations for Table 1 were done using Monte Carlo integration. 10,000 pairs of uniformly distributed points in the unit square were used to evaluate (3) for all values of $\rho$ and $N$.

A representation for the distribution of $p(t)$, the class I area fraction within a heterogeneous pixel of size $N \times N$ centered at $t$, is given by (2) for the simple class boundary model described above. The expression is conditional on the point $t$ belonging to class I, i.e. $p(t) > 1/2$. However, the expression (2) was not computed directly. Instead 10,000 random straight line boundary segments were simulated. These were constructed by uniform sampling in the polar coordinate parameter space $[0, 2\pi] \times [0, \sqrt{2}/2]$. Parameter values which resulted in lines not crossing the unit square centered at the origin were rejected. This distribution for the class I area fraction is summarized below.
Table 2. Class I Area Fraction

<table>
<thead>
<tr>
<th>Class I Fraction</th>
<th>Cumulative Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>.50</td>
<td>.00</td>
</tr>
<tr>
<td>.60</td>
<td>.14</td>
</tr>
<tr>
<td>.70</td>
<td>.29</td>
</tr>
<tr>
<td>.80</td>
<td>.44</td>
</tr>
<tr>
<td>.90</td>
<td>.61</td>
</tr>
<tr>
<td>.95</td>
<td>.73</td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The simulated class I area fractions were used to compute the integral (7). This integral gives the error probability \( E_B \) of misclassifying a class I point using pixel-averaged data – for the simplest normal theory case of the two-class discriminant of (5) – given that the pixel is cut by a linear class boundary. Roughly speaking \( E_B \) is the average probability of misclassifying points near a class boundary.

The \( E_B \) error probability is plotted in Figures 2a, 2b, 2c, 2d as a function of \( S \), the standard deviation of the noise component for the case of equal class prior probabilities. These four figures use, respectively, four different values of the mean separation between classes, viz., \( b = 0.5, 1.0, 1.5, 2.0 \) in units such that \( S = 1.0 \) for pixels of unit size. Figures 1a, 1b, 1c, 1d and 3a, 3b, 3c, 3d show the same error probability plots for the case of unequal class prior probabilities. The lower curve on each of these figures shows the homogeneous-pixel error probability \( E_H \) as a function of \( S \). This \( E_H \) is the normal tail probability (6) and corresponds to the naive error rate which assumes all pixels are class-homogeneous. The discrepancy between \( E_B \) and \( E_H \) is quite striking, especially when the noise variability is small.

The overall error rate for misclassifying class I points will be some weighting of the two error rates \( E_B \) and \( E_H \), for a value of the standard deviation \( S \) determined by the pixel size \( N \) and the autocorrelation parameter \( \rho \). The weights for the two component error rates are the relative frequencies of boundary pixels versus homogeneous pixels. These frequencies in turn depend on the pixel size \( N \), the boundary intensity parameter \( v \), and the class I prior probability \( Q \). An approximation to the frequency of boundary pixels for class I points is
given by (8). This approximation is used to construct the plots of Figures 4a–6d, which show overall classification error rates for class I points as a function of pixel size $N$. Each figure corresponds to a pair of different values of the autocorrelation persistence parameter, $\rho$, and the class I prior probability, $Q$. Figures 4a, 4b, 5a, 5b, 6a, 6b correspond to a value of the class mean difference, $b$, equal to 0.5 while Figures 4c, 4d, 5c, 5d, 6c, 6d correspond to a value of the class mean difference equal to 1.5. Finally, separate curves on these plots correspond to the boundary intensity parameter values $v = 0, 0.10, 0.25, 0.50$, respectively from bottom to top of each plot. However, the heterogeneous pixel frequency approximation (8) may not be reasonable for values of $v \cdot N$ greater than 1.5.

4. References.


Appendix.

Let \((r, \theta)\) be the polar coordinates of the line segment crossing a square pixel where \(r\) is perpendicular distance measured from the center of the square and \(\theta\) is the angle formed with the diagonal of the square, as shown below. This line segment divides the square into a class I part and a class II part, i.e., the class boundary is taken to be a locally linear line segment. By convention the center of the square belongs to class I. Let \(A\) be the area proportion of the square pixel which belongs to class I; then \(1/2 \leq A \leq 1\).

\[
\begin{align*}
\text{class I} & \quad \text{class II} \\
\text{Figure 1}
\end{align*}
\]

To calculate the area of \(A\) note that the class II region will be either triangular or trapezoidal depending on \((r, \theta)\). By symmetry we need only consider the case \(0 \leq \theta \leq \pi/4\). Then the event \(\{A \leq t\}\) occurs if and only if

\[
0 \leq r \leq \frac{\sqrt{2}}{2} (t - 1/2) (\sin \theta + \cos \theta) \quad \text{for} \quad 0 \leq \theta \leq \theta^*(t) \quad [\text{trapezoid case}],
\]

or

\[
0 \leq r \leq \frac{\sqrt{2}}{2} \cos \theta - \sqrt{1 - t \cos 2\theta} \quad \text{for} \quad \theta^*(t) \leq \theta \leq \pi/4 \quad [\text{triangle case}]
\]

where

\[
\theta^*(t) = \arctan \left( \frac{2t - 1}{3 - 2t} \right).
\]

For each \(t\) the inequalities above describe a finite region in the \((r, \theta)\) parameter space whose area we denote by \(\mu(t)\). Since the invariant random line measure assigns a uniform distribution over any subset of \((r, \theta)\) space, the probability of the event \(\{A \leq t\}\) will be \(\mu(t)/\mu(1)\) where \(\mu(1) = 2\). The calculation of these areas in parameter space gives the result (2) which may be
written as

\[ P \left\{ A \leq 1 - \frac{u}{2} \right\} = \frac{1 - u^2}{\sqrt{1 + u^2}} - \sqrt{\frac{u}{2}} \int_0^{\theta(u)} \cos \omega d\omega \]

where \( \theta(u) = \arctan \left( \frac{1 - u^2}{2u} \right) \).
Conditional Error Rates versus Pixel Standard Deviation

Fig. 1a
b = 0.5

Fig. 1b
b = 1.0

Fig. 1c
b = 1.5

Fig. 1d
b = 2.0

Class I Prior Probability Q = 0.25

FIGURE 1
Conditional Error Rates versus Pixel Standard Deviation

Fig. 2a
b = 0.5

Fig. 2b
b = 1.0

Fig. 2c
b = 1.5

Fig. 2d
b = 2.0

Class I Prior Probability Q = 0.5

FIGURE 2
Conditional Error Rates versus Pixel Standard Deviation

FIGURE 3

Class I Prior Probability Q = 0.75
Overall Classification Error Rate versus Pixel Size

Fig. 4a
b = 0.5

Fig. 4b
b = 0.5

Fig. 4c
b = 1.5

Fig. 4d
b = 1.5

ρ = 0  Class I Prior Probability Q = 0.25  ρ = 0.82

FIGURE 4
Overall Classification Error Rate versus Pixel Size

\[ \rho = 0 \quad \text{Class I Prior Probability} \quad Q = 0.5 \quad \rho = 0.82 \]

FIGURE 5
Overall Classification Error Rate versus Pixel Size

Fig. 6a
b = 0.5

Fig. 6c
b = 0.5

Fig. 6b
b = 1.5

Fig. 6d
b = 1.5

$\rho = 0$  Class I Prior Probability $Q = 0.75$  $\rho = 0.82$

FIGURE 6