METHODS FOR ASSESSING THE POSITIONAL AND THEMATIC ACCURACY OF REMOTELY SENSED PRODUCTS

Carlos Antonio Oliveira Vieira¹, Paul M. Mather², Antônio Simões Silva³, Dalto Domingos Rodrigues⁴, Joel Gripp Júnior⁵, Antônio Santana Ferraz⁶ and Julio Cesar de Oliveira⁷.

Universidade Federal de Viçosa, Departamento de Engenharia Civil, Brasil.
School of Geography, University of Nottingham, UK
Recibido: julio 2005                         Aceptado: diciembre 2005

Abstract
The aim of this paper is to review standard methods for assessing the quality control of cartographic products in the context of remote sensing. A further aim is to present a methodology to assess the positional accuracy of spatial databases using generic features (and their spatial distribution) within the image in the validation phase, and also to describe a methodology that specifically takes into account spatial pattern of errors of omission and commission, in order to present the user with an indication of the reliability of the pixel label assignments. Results show that a considerable amount of research and development needs to be accomplished before the spatial characterisation of positional and thematic accuracy associated with remote sensing can be adequately reported in standardised format and legends. Several techniques for the quality control of spatial databases using generic features are adapted to the context of remote sensing. However, one drawback of some of these approaches is the difficulty in obtaining homologous points in both representations. Alternative techniques to overcome such limitations are discussed.

Keywords: Standard Methods, Remote Sensing, Spatial Databases, Image, Positional Accuracy

MÉTODOS PARA EVALUAR LA PRECIOSIÓN POSICIONAL Y TEMÁTICA DE PRODUCTOS GENERADOS POR SENSOMETRÍA REMOTA

Resumen
El objetivo de este artículo es revisar los métodos estándares aplicados para evaluar el control de calidad de productos cartográficos en el contexto de los sensores remotos. Adicionalmente se persigue presentar una metodología que permita evaluar la precisión

¹Ing , Msc en , PhD en Geografía,Universidade Federal de Viçosa, Departamento de Engenharia Civil CCE, Viçosa – MG, 36571-000, Brazil. Email: carlos.vieira@ufv.br
²Professor, School of Geography, University of Nottingham, University Park, Nottingham NG7 2RD, UK. Email: Paul.Mather@nottingham.ac.uk
³Universidade Federal de Viçosa, Departamento de Engenharia Civil CCE, Viçosa – MG, 36571-000, Brazil
⁴Universidade Federal de Viçosa, Departamento de Engenharia Civil CCE, Viçosa – MG, 36571-000, Brazil
⁵Universidade Federal de Viçosa, Departamento de Engenharia Civil CCE, Viçosa – MG, 36571-000, Brazil
⁶Universidade Federal de Viçosa, Departamento de Engenharia Civil CCE, Viçosa – MG, 36571-000, Brazil
⁷Universidade Federal de Viçosa, Departamento de Engenharia Civil CCE, Viçosa – MG, 36571-000, Brazil
posicional de bases de datos espaciales utilizando características genéricas (y su espacial 
distribución) dentro de la imagen durante la fase de validación; así como describir una 
metodología que tome en consideración patrones espaciales por errores de omisión y 
comisión, para darle al usuario un nivel de la confiabilidad del pixel etiquetado. Los 
resultados obtenidos demuestran que hay una necesidad de desarrollar mucho más 
investigación, aún antes que la caracterización espacial y la exactitud temática asociado 
al uso de los sensores remotos pueda ser llevado a formatos estándares y leyendas 
conocidos. Un considerable número de técnicas de control de calidad de base de datos 
espaciales, utilizando características genéricas, son adaptadas al contexto de los sensores 
remotos. Sin embargo, una de las desventajas de estas técnicas radica en la dificultad de 
obtener puntos homogéneos en ambas presentaciones. Las técnicas alternativas para 
solver tales limitaciones son aquí discutidas.

Palabras Claves: Métodos Estándares, Sensores Remotos, Base de Datos Espaciales, 
Imagen, Precisión Posicional

1. Introduction

It is widely acknowledged that classification of remotely sensed imagery has variable, and 
often poor, quality. The cause and nature of these errors have been subjects of extensive 
research in order to improve the accuracy of remotely sensed products. Error in this context 
can be defined as a discrepancy between the situation depicted on the generated image 
(map) and reality (Arbia, Griffith, and Haining, 1998). Performing spatial data analysis 
operations on data of unknown accuracy will result in a product with low reliability and 
limited use in the decision-making process. This can be compounded by errors deriving 
from one source propagating through the database via derived products (Lunetta, Congalton, 
Fenstermaker, Jessen and McGwire, 1991). The quality of data is a function both of the 
inherent properties of those data and the use to which they are to be put. Hence, knowledge 
of error levels is necessary if data quality is to be estimated.

There are two different components of accuracy in the context of remote sensing: 
position and thematic accuracy (Janssen and Van der Wel, 1994). Positional accuracy 
determines how closely the position of discrete objects shown on a rectified image (map) 
or in a spatial database agree with the true position on the ground, while thematic accuracy 
refers to the non-positional characteristic of a spatial data entity, the so-called attributes 
(which are derived from radiometric information).

Quality control of cartographic products is usually accomplished by computing the 
discrepancy between each member of a set of well defined points present in one cartographic 
document with the corresponding points observed in the field, using a technique that 
guarantees sufficient accuracy for the analysis. In spatial databases generated from remotely 
sensed data, it is equally necessary to have knowledge of the discrepancies (errors). However, in some imagery where the number of control points is not sufficient, or where 
their spatial distribution is not adequate, the use of generic features (such as roads, edges, 
polygons, etc) to provide a means of relating two spatial data sets is an important alternative.

The aim of this paper is to review standard methods for assessing the quality of 
cartographic products in the context of remote sensing. A further aim is to present a
methodology to assess the positional accuracy of spatial databases using generic features (and their spatial distribution) within the image in the validation phase, and also to present the user with an indication of the thematic reliability of the remote sensing products.

2. Study Area and Data
Two different datasets were used in order to assess the positional and thematic accuracy of remote sensing products.

2.1 Positional Accuracy
The study area is located near the town of Uberaba-MG, in south-eastern Brazil. This area is located at approximately 700m above sea level, and possesses an undulating topography. The economic activities of the region are based on the milk and derived products as most of the area is covered by grassland.

Two remotely sensed multi-band images were used in this study, one acquired by the Landsat Thematic Mapper (at 30m of resolution) and the other by the High Resolution CCD Camera (HRCCC) carried by the China-Brazil Earth Resources Satellite (CBERS). This camera has a spatial resolution of 20 m. For the purposes of this study, a single waveband of each of the two multi-band images (TM and HRCCC), were used. A 1:25,000 scale map of the study region was used to provide ground reference data.

The two single-band images were geometrically registered to the UTM reference system (zone -23 S) using the Córrego Alegre horizontal datum (Brazil). Image to map registration used 14 and 12 ground control points respectively for the Landsat TM and CBERS HRCCC images, with nearest neighbour resampling, since this technique maintains the original pixel values (Jensen, 1986). In each case, the root-mean-square (RMS) error associated with registration was less than 0.5 pixels (i.e., the RMSTM = 0.4721 and the RMSGBERS = 0.479). Atmospheric correction was not performed since comparisons are not being made directly between images.

2.2. Thematic Accuracy
A SPOT High Resolution Visible (HRV) multispectral (XS) image (14 June 1994) of a region of flat agricultural land located near the village of Littleport (E. England) was used in the second case study, together with Field Data Printouts for the summer 1994. These printouts were derived from survey data supplied by individual farms, and provide details of the crop or crops growing in each field in the study area. On the basis of examination of the areas covered by each crop, the geographical scale of the study, and the spectral separability of the crops, seven crop categories were selected: potatoes, sugar beet, wheat, fallow, onions, peas and bulbs.

Image processing operations were performed using ERDAS Imagine (version 8.0) and the IDRISI GIS. Neural network application used the SNNS software. Some in-house programs were written to carry out specific procedures. Registration of the image to the Ordnance Survey (GB) 1:25,000 map was performed using 17 ground control points and nearest neighbour interpolation. The RMS error was 0.462 pixels.
3. Techniques for Estimating Positional and Thematic Accuracy

3.1 Positional Accuracy

A standard method of assessing the positional accuracy of cartographic products is based on comparison of deviations between homologous control points that can be accurately located on both the reference map and the geometrically corrected image. Statistics computed from the deviations at these homologous points are used to perform specific tests to evaluate the accuracy of the geometrically corrected image.

An alternative approach to assess the positional accuracy of cartographic products is based on the use of geometric features (Galo, Poz, and Ferreira, 2001). Examples of geometric features are roads, edges, and other boundaries. They should be easily located and represented as a set of sequential coordinates in both documents (i.e., the image to be corrected and the reference map). Three feature-based methods have been used by Vieira, Mather, and Borges (2002) to measure the correspondence between features shown on a reference map and a remotely sensed image. These are the Generated Point Method, the Areal Method and the Equivalent Rectangle Method. As these alternative methods work with the relative distances between homologous points, there is no need to apply trend analysis to check for the presence of systematic errors in the E and N directions (E and N are the directions X and Y respectively on the Universal Transverse Mercator coordinate system). One of these methods is presented in the following sections.

3.1.1 Map Accuracy Standards (MAS) Determination

An important method for cartographic evaluation of generated maps is the use of Map Accuracy Standards – MAS (Galo et al., 2001), which is based on the comparison of deviations between homologous control points easily located on the reference map \((X_m, Y_m)\) and the image \((X_i, Y_i)\). Statistics computed from the deviations at these homologous points \((\Delta E = X_m - X_i, \Delta N = Y_m - Y_i)\) are used to perform specific tests to evaluate the trend and the accuracy of the geometrically corrected image. In this analysis it was considered that the reference map was sufficiently accurate for our purposes, i.e. to assess the geometric quality of the images.

According to Merchant (1982), the aim of trend analysis is to check for the presence of systematic errors. This check uses the sample mean of the deviations \((\bar{\Delta E}, \bar{\Delta N})\). A statistical test is applied, using the null hypothesis that the sample mean is estimating a true (population) mean of zero. If the null hypothesis can be accepted at some significance level then the conclusion to be drawn is that there is no trend or systematic error in the directions X and Y respectively.

The Student’s t Test is normally used to carry out this statistical test. The critical value \(t_{n-1,\alpha/2}\), is obtained from statistical tables (where \(n\) is the total number of control points and \(\alpha\) is the confidence level, for example 0.1). If the calculated value of \(t\) for the deviations along the north-south dimension \(|t_N|\) is less than \(t_{n-1,\alpha/2}\) and the calculated value of \(t\) for deviations in the east–west dimension \(|t_E|\) is less than the tabled value \(t_{n-1,\alpha/2}\) then the generated product (e.g., image) is free from systematic errors on the directions N and E respectively. The estimated values of \(t_N\) and \(t_E\) can be estimated using the following equations:
\[ t_N = \frac{1}{\sigma_N} \Delta_N n^{1/2} \]
\[ t_E = \frac{1}{\sigma_E} \Delta_E n^{1/2} \]
where \( \sigma_E \) and \( \sigma_N \) are the standard deviations of the discrepancies \( \Delta_E \) and \( \Delta_N \) in the directions E and N respectively.

Accuracy analysis uses comparison of the variance of sample deviations (\( \text{var}(\Delta x, \text{var} \Delta y) \)) to their respective pre-defined (tabled) values. The test is performed using a hypothesis about the mean and standard deviation of the sample for each of the geometric coordinates. The statistical procedure employed is the Chi-square (\( \chi^2 \)) test. The accuracy of the geometrically corrected image can be estimated separately for the directions N and E, using standard statistical methodology involving the comparison of a sample value of \( \chi^2 (\chi^2_{N,n-1}) \) and the tabled value \( (\chi^2_{N,n-1}, \alpha) \), or \( (\chi^2_{E,n-1}) \) and \( (\chi^2_{E,n-1}, \alpha) \) respectively. The values of \( \chi^2_{N,n-1} \) and \( \chi^2_{E,n-1} \) are estimated using the following equations:

\[ \chi^2_{N,n-1} = (n-1) \cdot \left( \frac{\sigma_N^2}{\theta_N^2} \right) \]
\[ \chi^2_{E,n-1} = (n-1) \cdot \left( \frac{\sigma_E^2}{\theta_E^2} \right) \]

where \( \theta_N \) and \( \theta_E \) are obtained from Table 1 and vary as a function of map scale using the formulae: \( \theta_N = \theta_E = \sigma_E / \sqrt{2} \). The values of the Standard Error (SE) for the Brazilian Map Accuracy Standards, for instance, is defined by the decree n° 89.817 of 1984, which classifies cartographic products in relation to their geometric quality (see Table 1).

<table>
<thead>
<tr>
<th>CLASSES</th>
<th>PLANIMETRY</th>
<th>SCALE 1:100.000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CAS</td>
<td>SE</td>
</tr>
<tr>
<td>A</td>
<td>0.5 mm</td>
<td>0.3 mm</td>
</tr>
<tr>
<td>B</td>
<td>0.8 mm</td>
<td>0.5 mm</td>
</tr>
<tr>
<td>C</td>
<td>1.0 mm</td>
<td>0.6 mm</td>
</tr>
</tbody>
</table>

3.1.2. Generated Point Method

One drawback of some of the standard approaches to quality control of cartographic products is the difficulty of obtaining homologous points in both representations. It is therefore necessary to consider the use of generic features to complement the control points.

The generated point method can be applied to digitised homologous features (i.e., the same features represented in both map and image) in order to generate a set of equally spaced homologous coordinates following the path of both features.

The initial control points for each feature should be homologous, so the relative...
distances (or deviations D) between the generated homologous points are used to perform the comparison. It is evident that if there is no deviation between the generated homologous points, the image is well corrected with reference to the map; otherwise, it is necessary to apply statistical analysis to in order to check the positional accuracy.

A comparison between the median and standard deviation of these deviations and published Map Accuracy Standards (MAS) are carried out considering a specific scale. The accuracy test uses a Chi-square ($\chi^2$) test based on the specified Standard Error (SE). The geometrically corrected image will be accepted as accurate if $\chi^2_{D,n-1} < \chi^2_{n-1, \alpha}$. The sample value of $\chi^2_{D,n-1}$ can be estimated using equation (5):

$$\chi^2_{D,n-1} = (n-1) \cdot (\sigma^2_D / \theta^2_D)$$

where, as before, $\theta_D$ is obtained from the Table 1. However, it is not directional, and separate tests are not carried out for the N and E dimensions separately. The value $\theta_D$ is computed from the formula $\theta_D = SE/\sqrt{2}$.

3.2. Thematic Accuracy

Current accuracy assessment methods are based on non-spatial statistics derived from the confusion or error matrix, which compares the output of a classifier and known test data (Table 2). These statistics include overall accuracy, individual class accuracy, user's and producer's accuracy, and several other statistics. Although these measures are in widespread use, none of them considers the spatial distribution of erroneously classified pixels, either implicitly or explicitly.

Table 2. Confusion or error matrix (source Vieira and Mather, 2001).

<table>
<thead>
<tr>
<th>Classes</th>
<th>Reference Data</th>
<th>Users (%)</th>
<th>Z</th>
<th>K (cond)</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>344</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>33</td>
<td>349</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>2</td>
<td>386</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>20</td>
<td>5</td>
<td>390</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>394</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TOTAL</td>
<td>400</td>
<td>400</td>
<td>400</td>
<td>400</td>
<td>400</td>
</tr>
</tbody>
</table>

3.2.1 Characterising the Spatial Distribution of the Errors

One possible way to characterise the spatial distribution of the errors in a thematic classification is by generating a "distance image" (see Figure 1(a)) showing the distance from individual pixels to the multivariate means of the classes to which they have been assigned. Either the Euclidean distance or the Mahalanobis distance measure can be used. The former, however, implies spherical clusters in feature space, while the latter takes into account the covariance between the features on which the classification is based. The individual distances are scaled onto a 0-255 range, and displayed as a grey scale image. Darker pixels are spectrally "nearer" to their class centroid (in the sense of statistical distance), and are thus more likely to have been classified correctly. On the other hand, pixels with higher distance values are spectrally further from the centroid of the class to
which they were assigned, and are thus more likely to have been misclassified. A threshold can be applied to the distance image to identify those pixels that are most likely to have been misclassified.

![Distance Image](image1.png) ![Error Image](image2.png)

**Figure 1.** Spatial characterisation of classification errors using thematic image generated by Maximum Likelihood Classifier (385 by 385 pixels).

An alternative way of looking at the spatial distribution of the errors present in a classified image is by directly comparing thematic images with their respective ground truth maps. One of the products of this comparison should be a binary error image (Figure 1(b)) in which each point takes the value 0 (correctly labelled) or 1 (erroneously labelled). By examining the spatial distribution of such pixels in Figure 1 we can make a number of observations. It is apparent that misclassified pixels are spatially correlated. Possible reasons for these correlation effects include the presence of mixed pixels at field boundaries, variation in the reflectance spectrum (most probably caused by variations in soil type within a field) and the effects of crop management practices such as the use of fertilisers. Spatial analysis measures (e.g., Join Count Statistics) could be used in order to determine whether these correlation effects are random or clustered in their spatial distribution. Looking at the spatial distribution of the remaining errors can help to refine the classification process.

### 3.2.2 Visualising the Reliability

Any distance measure between the pixel and the mean pixel values (or prototypes) of each class can be used to compute a measure of reliability of a pixel's label. These measures of reliability could be then combined to the already assigned class label in order to generate a new thematic value for the pixel, which not only indicates the class to which the pixels was assigned but also the level of reliability achieved. A separate colour is assigned to each class. Within-class levels are also assigned separate shades of that colour, so that
each class is represented by five shades of the given colour (see Figure 2(a)). This kind of representation allows the visual appreciation of the level of reliability of the classified crop. A contour representation of the reliabilities can also be used (Figure 2(b)). These types of representation help the user to identify portions of the thematic map that have reduced reliability. Although the final map may look uniform in its accuracy, it is actually a representative assemblage from several image processing procedures and refinements. It is important for the user to known how these accuracies are spatially distributed in the image through a thematic reliability map.

(a) Thematic Reliability           (b) Reliability Contour Representation

Figure 2. Representation of the reliabilities using a Maximum Likelihood classifier (385 by 385 pixels).

4. Results and Discussion
A summary of the data used in this study is presented in Table 3.

4.1 Trend Analysis
Trend analysis was carried out using the Student’s t distribution and a confidence level of 90% ($\alpha = 0.10$). The critical value obtained from statistical table for 28 control points is: $T_{27, 0.10} = 1.703$. Equations (1) and (2) were used to compute estimated values of $t_N$ and $t_E$. 
Table 3. Summary of the mean deviation, standard deviations used in the Trend and Accuracy analysis; Generated Point method; and Areal and equivalent Rectangles approach.

**LANDSAT:**
\[ t_N = 0.1926 < 1.703 \text{ and } t_E = 0.3977 < 1.703 \] (there is no trend in these directions)

**CBERS:**
\[ t_N = 0.2828 < 1.703 \text{ and } t_E = 1.2172 < 1.703 \] (there is no trend in these directions)

### 4.2 Accuracy Analysis

Accuracy analysis was carried out for both Landsat and CBERS images. The results of the estimates of \( \chi^2_{N,n-1} \) and \( \chi^2_{E,n-1} \), computed using Equations (3) and (4), are presented in Table 4.

**Table 4.** Estimated values of \( \chi^2_{N,n-1} \) and \( \chi^2_{E,n-1} \) used in the accuracy analysis
Adopting the critical value of $\chi^2 = 36,7412$ and considering the limit values for the classes A, B, and C as shown in Table 1 for the Brazilian Map Accuracy Standards (scale 1:100.000), these two images have accuracy equivalent to the class C.

### 4.3 Generated Point Method

The values of the Standard Error (SE) for the generated point method are the same as the Brazilian Map Accuracy Standards, for instance, (see Table 1). As this method works with relative distances between homologous points, there is no need to apply trend analysis to check for the presence of systematic errors in the E and N directions. The accuracy analysis was also performed for both Landsat and CBERS images using a set of equally spaced homologous coordinates in order to check the positional feature accuracy for a given confidence level. The results are presented in Table 5.

#### Table 5. Estimated values of $\chi^2_{D,n-1}$ used in the accuracy analysis for the Landsat and CBERS images.

<table>
<thead>
<tr>
<th>CLASSES</th>
<th>LANDSAT</th>
<th>CBERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>128,0141</td>
<td>441,4366</td>
</tr>
<tr>
<td>B</td>
<td>46,0852</td>
<td>158,9175</td>
</tr>
<tr>
<td>C</td>
<td>32,0035</td>
<td>110,3592</td>
</tr>
</tbody>
</table>

The critical value of $\chi^2$ is 48,2329 and the limit values are given in Table 1 for the Brazilian Map Accuracy Standards (scale 1:100.000). The Landsat image has an accuracy equivalent to class B, while the CBERS product does not fit in any specify class.

### 5. Conclusions

Trend analysis indicates that there is no systematic error in any direction for the Landsat and CBERS geometrically corrected images at the 90% confidence level. On the other hand, both corrected image have an accuracy equivalent to class C using the Brazilian Map Accuracy Standards for a map scale of 1:100.000. Only the Landsat image fits the class B specification, using the generated point method.

These results show that a considerable amount of research needs to be undertaken before the spatial characterization of positional and thematic accuracy associated with remote sensing data can be adequately reported in standardized format and legends. Several techniques for the quality control of spatial databases using generic features are adapted to the context of remote sensing. However, one drawback for some of these approaches is the difficulty in obtaining homologous points in both representations. Alternative techniques for overcoming such limitations could be the use of a spline Fitted method, in which the shape of curves fitted using splines are compared instead of isolated homologous control points (Galo et al., 2001).

### 6. Acknowledgements

Preliminary research by Mr. Vieira was supported by the Brazilian Research Council (CAPES). The later stages of this study were conducted as part of the CNPq/COAIE/Kit
Enxoval (Project Number 68.0045/01-5) performed in collaboration with The University of Nottingham. We are grateful to Logica PLC, INPE and SPOT Image for permission to use their images.

Bibliography references


