

UNIVERSIDAD NACIONAL DE CÓRDOBA
FACULTAD DE MATEMÁTICA, ASTRONOMÍA Y FÍSICA

SERIE “A ”

TRABAJOS DE MATEMÁTICA

Nº 82/07

**The Influence of Training Errors, Context and Number of
Bands in the Accuracy of Image Classification**

Alejandro C. Frery - Susana B. Ferrero - Oscar H. Bustos



Editores: Jorge R. Lauret – Elvio A. Pilotta

CIUDAD UNIVERSITARIA – 5000 CÓRDOBA

REPÚBLICA ARGENTINA

The Influence of Training Errors, Context and Number of Bands in the Accuracy of Image Classification

Alejandro C. Frery
Instituto de Computação
Universidade Federal de Alagoas

Susana B. Ferrero
Departamento de Matemática
Universidad Nacional de Río Cuarto

Oscar H. Bustos
Facultad de Matemática, Astronomía y Física
Universidad Nacional de Córdoba

Abstract

We present the assessment of two classification procedures using both a Monte Carlo experiment and real data. Classification performance is hard to assess with generality due to the huge number of variables involved. We consider the problem of classifying multispectral optical imagery with pointwise Gaussian Maximum Likelihood (ML) and contextual ICM (*Iterated Conditional Modes*), with and without errors in the training stage. Two experimental setups were considered in order to assess the influence of using partial and low quality information and to make a quantitative comparison of ML and ICM in real situations. Using simulation the ground truth is known and, therefore, precise comparisons are possible. The contextual approach proved being superior than the pointwise one, at the expense of requiring more computational resources. Pseudolikelihood parameter estimation for the context is used, and we provide the expression for estimating this parameter when using the multiclass Potts model with eight neighbors. Quantitative and qualitative results are discussed.

1 Introduction

The production of thematic cartography is one of the main goals of remote sensing image processing and analysis. Thematic maps provide the overall inventory of classes in an image; they can be obtained with classification techniques and are essential in many applications as, for instance, crops statistics, mining and hydrological resources studies.

A general setup for image classification was provided by Geman and Geman (1984). They consider separate models for the observed data and for unobserved relevant information. Using the data and the underlying models, one seeks suitable estimators for the desired, but unobserved, information.

The problem of assessing the classification techniques accuracy has been studied in the literature (see, for instance, the works by Belward and Dehoyos, 1987; Emerson et al., 2005; Emrahoglu et al., 2003; Erbek et al., 2004; Frizzelle and Moody, 2001; Gao et al., 2006; Huang and Mausal, 1994; Ince, 1987; Wilson, 1992; Zhuang et al., 1995).

Regarding the explicit use of context by means of statistical models, Di Zenzo et al. (1987) consider the impact of the number of bands, spectral and spatial resolutions and context (using probabilistic relaxation) on the classification of crops. Flygare (1997) assesses the influence

of context, by means of a model of spectral correlation, in classification accuracy. Hubert-Moy et al. (2001) point out that using the ICM algorithm for improved classification under Markovian models, though computationally more expensive than pointwise maximum likelihood, is “worth the pain”. Magnussen et al. (2004) employ first-order autoregressive process on the reflectance model for introducing spatial information. Our contribution consists of making a quantitative assessment of the influence of (i) information content and number of bands, (ii) errors in the training stage, and (iii) context, by means of the Potts model for the (unobserved) classes and the ICM algorithm.

In order to draw the most possible general conclusions, it is desirable to use (a) automatic procedures that do not depend on unknown parameters and (b) data representative of as many as possible situations. We attain the first requisite by estimating all required parameters from available information, and the second by modelling incomplete data and errors in the training stage.

The multivariate Gaussian law is the most widely used distribution for modelling image data acquired with optical remote sensing instruments (Matthews et al., 2001). This is mainly due to the fact that the observed data result from the composition of a large number of non-deterministic independent sources with bounded variability. This distribution will be employed in this work for describing the observed data, since techniques based on this distribution are available in most remote sensing image processing platforms.

The spatial correlation of the classes is an important source of information, and we incorporate it by modelling the classes under estimation as a Markov Random Field.

Training is subjected to errors, since it depends on multiple sources of possibly vague and contradictory information (visual analysis, previous experience, data acquired by other sensors and in a different moment etc.). We assess the impact of training errors in the accuracy of the products by introducing this factor in the Monte Carlo experiments. As explained in section 3, wrong training samples are introduced in order to simulate the worst possible realistic scenario in real applications.

The purpose of this paper is assessing the precision of products obtained by Maximum Likelihood (ML) and by ICM classifications under different situations: regarding training, with and without errors and, regarding spectral, information, with complete and incomplete data. With the exception of the work by Frery et al. (2006), where partial results are presented, a complete assessment is not available in the literature.

The rest of the paper unfolds as follows. Section 2 recalls the basic definitions of statistical classification. Section 3 presents the Monte Carlo experiments and the simulation results. Section 4 shows the results of applying the techniques to real data: Landsat ETM+ (section 4.1) and ASTER (section 4.2) images are analysed, the former under two different setups for assessing the influence of spectral information content and the presence of errors in the training stage. Finally, section 5 comments the results and their consequences.

2 Supervised Statistical Classification

From a mathematical standpoint, a multispectral optical image is a three-dimensional real matrix:

$$\mathbf{z} = [z(i, j, k)]_{0 \leq i \leq M-1, 0 \leq j \leq N-1, 0 \leq k \leq K-1}, \quad z(i, j, k) \in \mathbb{R},$$

and a classification rule is a function that, using the available information, defines a set of $M \times N$ labels, say $\mathbf{c} = [c(i, j)]_{0 \leq i \leq M-1, 0 \leq j \leq N-1}$ on a set of L possible labels $\mathbf{C} = \{c_1, \dots, c_L\}$.

Supervised statistical classification procedures consist of providing such a rule by means of decisions based on the statistical properties of the data, i.e., parameters to be estimated. These procedures are based upon three steps, namely, training, production and testing.

2.1 Multivariate Gaussian Model

This model assumes that the observations related to each of the L classes obey different probability laws characterized by the probability density function

$$f_\ell(\mathbf{z}) = \frac{\exp\left\{-\frac{1}{2} [(\mathbf{z} - \boldsymbol{\mu}_\ell)^t \mathbf{M}_\ell^{-1} (\mathbf{z} - \boldsymbol{\mu}_\ell)]\right\}}{(2\pi)^{K/2} [\det(\mathbf{M}_\ell)]^{1/2}},$$

where K is the number of bands, $\boldsymbol{\mu}_\ell$ is the vector of means and \mathbf{M}_ℓ is the covariance matrix and $1 \leq \ell \leq L$ is the class index. This assumption is usually verified in practice when optical data is used, if a careful choice of classes is made.

The classification rule that stems from this assumption and the hypothesis of independence among different sites is assigning the site (i, j) to class c_{ℓ^*} if

$$f_{\ell^*}(\mathbf{z}(i, j)) \geq f_\ell(\mathbf{z}(i, j)), \quad (1)$$

for every $\ell^* \neq \ell$, i.e., if the likelihood of the observation $\mathbf{z}(i, j)$ is maximized by the model of class ℓ^* . The rule formulated in equation (1) is equivalent to assigning the site (i, j) to class c_{ℓ^*} if

$$\begin{aligned} & \ln(\det(\mathbf{M}_{\ell^*})) - (\mathbf{z}(i, j) - \boldsymbol{\mu}_{\ell^*})^t \mathbf{M}_{\ell^*}^{-1} (\mathbf{z}(i, j) - \boldsymbol{\mu}_{\ell^*}) \\ & \leq \ln(\det(\mathbf{M}_\ell)) - (\mathbf{z}(i, j) - \boldsymbol{\mu}_\ell)^t \mathbf{M}_\ell^{-1} (\mathbf{z}(i, j) - \boldsymbol{\mu}_\ell), \end{aligned}$$

for every $1 \leq \ell \leq L$.

In most practical situations one has to estimate the parameters $\boldsymbol{\mu}_\ell$ and \mathbf{M}_ℓ using training samples. As we will see, the quality of the training data is of paramount importance for the final result.

2.2 Markov Random Fields and the ICM Algorithm

Real data exhibit a great deal of spatial correlation. Markov Random Fields, the spatial generalization of Markov chains, have deserved a great deal of attention in the computer vision literature since they were successfully used in image restoration (Carnevali et al., 1985; Geman and Geman, 1984; Winkler, 2006); see Appendix A for a short review on these models, that will be used here to describe spatial correlation.

The Potts model will be used in this work as the underlying law for the classes in remote sensing imagery. It states that the log-probability of observing class c at coordinate (i, j) , given the observation of all the other classes, is proportional to $\beta \in \mathbb{R}$ times the number of neighboring sites where class c occurred. Positive values of β assign more probability to maps with clusters of same classes. Using this model as a prior distribution leads to a classification rule that takes context into account. A sample of this model, which is detailed in Appendix A, is shown in Figure 1(a).

The ICM algorithm is an iterative approach to finding better solutions than those provided by a pointwise procedure, such as pixelwise Maximum Likelihood. It starts with an arbitrary solution and improves it replacing the class in every coordinate by the one that maximizes an objective function that, in turn, comprises two terms: the evidence provided by the data (the information on which Gaussian Maximum Likelihood is based upon) and the evidence provided by the context.

In our implementation, ICM starts with the pointwise Gaussian Maximum Likelihood classification. Then, a new classification is obtained using, for every $(i, j) \in S$, $1 \leq \ell \leq L$, $\mathbf{z} \in \mathbb{R}^K$, the following decision rule:

$$g_\ell((i, j), \mathbf{z}, \mathbf{c}, \beta) =$$

$$\frac{1}{2} (-\log(\det(\mathbf{M}_\ell)) - (\mathbf{z} - \boldsymbol{\mu}_\ell)^t \mathbf{M}_\ell^{-1} (\mathbf{z} - \boldsymbol{\mu}_\ell)) + \beta \#\{(i', j') \in V_{(i,j)}^* : c(i', j') = c_\ell\}, \quad (2)$$

where $V_{(i,j)}^* = V_{(i,j)} \cup \{(i, j)\}$. The first term of the second member in equation (2) is the same as the pointwise Maximum Likelihood classification rule under the multivariate Gaussian model. The second term is the contextual component that, provided $\beta > 0$, puts more weight on those classes that surround site (i, j) .

The contextual influence is quantified by the value of the parameter β . When $\beta = 0$ the rule provided by equation (2) reduces to pointwise Maximum Likelihood classification under the multivariate Gaussian model, i.e., context has no effect on the evidence provided by the observed data; when $\beta \rightarrow \infty$ the effect is reversed, i.e., the observed data have no influence on the rule, which is solely the local mode. This parameter is unknown and, therefore, it has to be informed.

Besag (1989) provides a proof of the convergence of the ICM algorithm to a local maximum of the function given in equation (2). Since this equation has the ML rule as a particular case, namely when $\beta = 0$, and that the contextual term is positive, the inclusion of the latter provides solutions with higher likelihood.

Jackson and Landgrebe (2002) use an ICM algorithm with fixed values of the context parameter, and they show that a contextual classification with small samples attains an accuracy comparable with that obtained with pixelwise maximum likelihood. Arbia et al. (1999) also use fixed values of this parameter, chosen by trial-and-error, in a two-class classification setup with simulated data. Descombes et al. (1999) use a Monte Carlo Markov Chain procedure, while Tso and Mather (1999) employ a genetic algorithm approach. Melgani and Serpico (2003) propose the a minimum perturbation approach, but they maintain the estimated parameter along all the iterations. Moser et al. (2005) propose the use of the ICM algorithm under Markov random fields models, estimating its parameters as the solution of a set of linear inequalities, solved by extending the Ho-Kashyap algorithm. Vaccaro et al. (2000) provide a comprehensive account of the use of spatial information in synthetic aperture radar image analysis. All these approaches are either costly or inaccurate.

Our approach consists of iteratively estimating the contextual parameter from the available information by pseudolikelihood (Arnold and Strauss, 1991); see Appendix B for details. This estimation is performed after each iteration, being the first classification the Gaussian Maximum Likelihood rule or, equivalently, the rule provided by equation (2) setting $\beta = 0$. An iteration consists of (i) estimating β from the previous classification and (ii) applying the rule provided in equation (2).

It is observed that the sequence of estimated parameters is non decreasing, i.e, that $\hat{\beta}(0) \leq \hat{\beta}(1) \leq \dots$, so one will always end up with a classification with more homogeneous patches than the first provided as starting solution. The algorithm proceeds until evidence of convergence is achieved. In our implementation at least one of two criteria has to be satisfied in order to stop the procedure: a certain maximum number of iterations (100 in our experiments) or a certain minimum percentage of classes changed (set to 5%).

3 Precision Assessment by Simulation

A Monte Carlo experiment was devised in order to assess classification precision in the aforementioned scenarios. Images are simulated and they are automatically classified. In doing so, one has the ground truth before which the results can be compared. After making this simulation-based

assessment, the two classification techniques are applied to real data: more than fifty samples from a Landsat ETM+ image with many thematic classes. All available bands were employed in all but one situation, being this last one designed to evaluate the impact of using partial and low quality information on the classification procedures; we conclude that contextual information can successfully compensate the lack of quality data.

Three types of class images were used in this work in order to describe typical situations that appear in practice: a hand-painted one (called “Cubism”, see Figure 1(a)) inspired in thematic maps, random blocks, and outcomes of the Potts model.

Maps in the shape of random blocks with L classes are obtained dividing the support S , which is a square of side 64 or 72 in squares of side 4 or 6, respectively, and drawing a class independently from the other for every small square; if the same class is drawn in every small square, the map is discarded and the procedure begins again. A typical outcome for a 64×64 support and $L = 4$ is shown in Figure 1(b).

Figure 1(c) shows a typical outcome of the Potts models, as defined in equation (3) with four classes and $\beta = 1/2$.

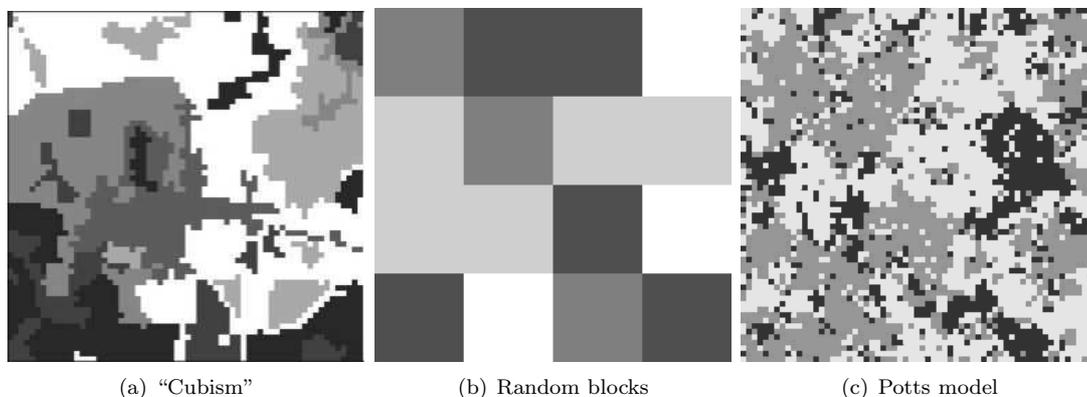


Figure 1: Images used in the assessment

In order to make the assessment in as many as possible representative situations, fourteen situations were considered: the three types of class images of sizes $64 \times 64 \times K$ and $72 \times 72 \times K$, where $K = 3$ or 4 bands and 4 or 6 classes. Besides these models, two training situations were modelled: perfect and imperfect training data. The quality of training data is of paramount importance since, as will be seen, it is critical and this issue has not been fully addressed in the literature.

The parameter values for each situation, i.e., $(\mu_\ell, \mathbf{M}_\ell)$, were chosen with the following rules (see details in Appendix C):

- P1: Values corresponding to the analysis of a real image with six classes, as described by Richards and Jia (1999).
- P2: Three classes with comparable low mean values and three with comparable high mean values; variances for the classes having comparable values are the same.
- P3: The same mean for all the classes; covariances are the same, and variances are increasing.
- P4: All classes with the same mean values, with increasing variances and covariances.

Table 1 presents the fourteen situations.

Table 1: Parameters for the observations

Situation	Image Type	Size	Classes	$(\mu_\ell, \mathbf{M}_\ell)$	Error
1	Blocks	$64 \times 64 \times 4$	4	P1	N
2	Blocks	$72 \times 72 \times 3$	6	P2	N
3	Blocks	$64 \times 64 \times 4$	4	P3	N
4	Blocks	$64 \times 64 \times 4$	4	P3	Y
5	Potts	$64 \times 64 \times 4$	4	P1	N
6	Potts	$64 \times 64 \times 4$	4	P1	Y
7	Potts	$72 \times 72 \times 3$	6	P2	N
8	Potts	$72 \times 72 \times 3$	6	P2	Y
9	Potts	$64 \times 64 \times 4$	4	P3	Y
10	Potts	$64 \times 64 \times 4$	4	P3	Y
11	Cubism	$64 \times 64 \times 3$	6	P2	N
12	Cubism	$64 \times 64 \times 3$	6	P2	Y
13	Cubism	$64 \times 64 \times 3$	6	P4	N
14	Cubism	$64 \times 64 \times 3$	6	P4	Y

Two hundred replications were made in every situation in order to assess the performance of the pointwise and contextual procedures. This number was found by exploratory analysis of the results obtained in selected situations, and provides enough precision for the desired comparison. Each replication consists of assuming a certain image class, sampling from its distribution if it is of type random blocks or Potts model, transforming classes into observations following the assumed models, obtaining samples for each class (with or without errors), producing the two classifications and validating them.

The training stage consists of randomly sampling, for each class, 10% of the sites and using the corresponding observations for parameter estimation. In the presence of training errors (situations 4, 6, 8, 9, 10, 12 and 14), 1/10 of those sampled observations is replaced by data from another class uniformly chosen among the others. This percentage of errors in the training samples was specified by experienced users as a higher limit observed in real applications, so we assess here the worst possible situation.

Since the true class image is known beforehand, it is possible to compute the actual error matrix and the coefficients of overall accuracy and Kappa with their respective confidence intervals (Agresti, 1990; Congalton, 1991; Fitzgerald and Lees, 1994).

The two hundred replications for each situation allow us to draw the following conclusions:

- Most situations produce values of Kappa higher than 0.70, so most classifications can be considered “good”.
- The lowest coefficients (overall accuracy and Kappa) were achieved in situations 13 and 14, where there was a high level of confusion: same mean values for every class and increasing variances and covariances.
- Situations 3 and 4 also produced low coefficients, but in this case ICM doubled the quality of pixelwise classification.
- Coefficients computed on ICM classifications are higher than the others in those situations where training was subjected to error.

- All coefficients are significantly different, and in most cases the evidence provided is that ICM is better than pixelwise classification.

Figure 2 summarizes some of these results, showing the 95% confidence intervals of the Kappa coefficient in some of the simulated situations. Light lines correspond to the Maximum Likelihood algorithm, while thick ones show the results obtained with ICM. It is clear that ICM significantly and consistently outperforms ML.

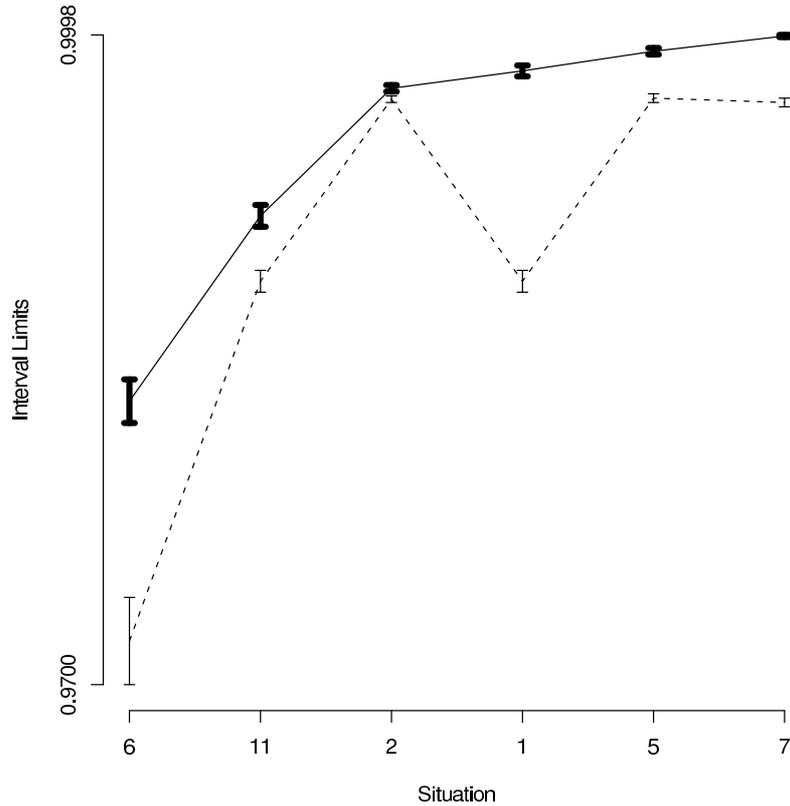


Figure 2: Confidence intervals for Kappa 95%

4 Analysis of real data

4.1 Landsat ETM+

Two experimental setups were considered: one where a 400×233 pixels image was analyzed and other where the whole dataset (6920×5960 pixels) was treated. The first experience aims at assessing the influence of using partial and low quality information; ML and ICM classifications obtained with the three bands that provide the least separation are compared. In the second setup, 50 subimages were generated from the complete data set in order to make a quantitative comparison of ML and ICM in real situations.

4.1.1 Setup 1

An area of 400×233 pixels from the 229083 Landsat 7 ETM+ image (30 m resolution) acquired in 2000 over the city of Río Cuarto, Argentina, was analyzed. The 453 RGB composition of the image is shown in Figure 3(a). Six thematic classes were identified using prior knowledge, exploratory data analysis and photointerpretation: River (predominantly Black in the RGB composition, type # 1, Red in the classification), Urban (Light Blue, # 2, Green), Bare Soil (Light Green, # 3, Blue), Natural Pasture (Dark Green, # 4, Yellow), Managed Pasture (Orange, # 6, Cyan) and Trees (Red, # 6, Magenta).

In order to estimate the vectors of means and the covariance matrices, 5672 training samples were chosen (about 6% of all the pixels). These observations were subjected to a careful exploratory analysis, since the quality of these samples is paramount for obtaining good results. Test samples were also identified in order to assess classification accuracy; in this study 4041 pixels were labelled as test samples.

The reference classification was obtained with the seven available bands by ML; it is shown in Figure 3(b) and its estimated accuracy is 0.86 (see Table 2 for details).

The three bands that provided the weakest separation between the classes of interest are 1, 3 and 5; the parameters were estimated using this information, and ML classification was obtained (see Figure 3(c)). It was then used as the starting point of the ICM algorithm, that ended with $\hat{\beta}(2) = 0.81$ and the classification shown in Figure 3(d). Quantitative results are shown in Table 2; all estimated accuracy values are high, showing that the classification procedure is excellent (see Landis and Koch, 1977).

As a general conclusion, one sees that when there is complete and reliable information, the accuracy improvement provided by ICM with respect to ML is small (from 0.84 to 0.86 when seven bands are used), though statistically significant. The biggest contribution the contextual classification provides arises when only incomplete data are available: if the worst three bands are employed, the accuracy improves from 0.79 to 0.84. Even in the former situation, i.e., when seven bands are used, the classification improvement is visually noticeable (see in figures 3(b) and 3(e) how the trees spots in magenta are better resolved).

Table 2: Influence of partial information

Technique, data	Accuracy	Kappa	95% Confidence Interval
ML, 7 bands	0.86	0.8188	[0.8049, 0.8328]
ML, 3 bands	0.79	0.7141	[0.6976, 0.7306]
ICM, 3 bands	0.84	0.7872	[0.7722, 0.8021]
ICM, 7 bands	0.88	0.8447	[0.8317, 0.8579]

ML with all the available information provides an accuracy of 0.86 but using the three worst bands this figures drops to 0.79. Using contextual information on the three worst bands improves the result to 0.84, which is close to the accuracy obtained with seven bands. Confidence intervals for the accuracy show that these values are significantly different. Incidentally, the accuracy achieved by ICM and seven bands is of 0.88; this classification is shown in Figure 3(e). Figure 3 shows that classifications obtained by ICM are less grainy than those that employed only spectral information.

4.1.2 Setup 2

From the complete data set (seven bands 6920×5960 image), 50 non-overlapping sub images of size 200×160 with seven bands each were generated. Each was subjected to visual and

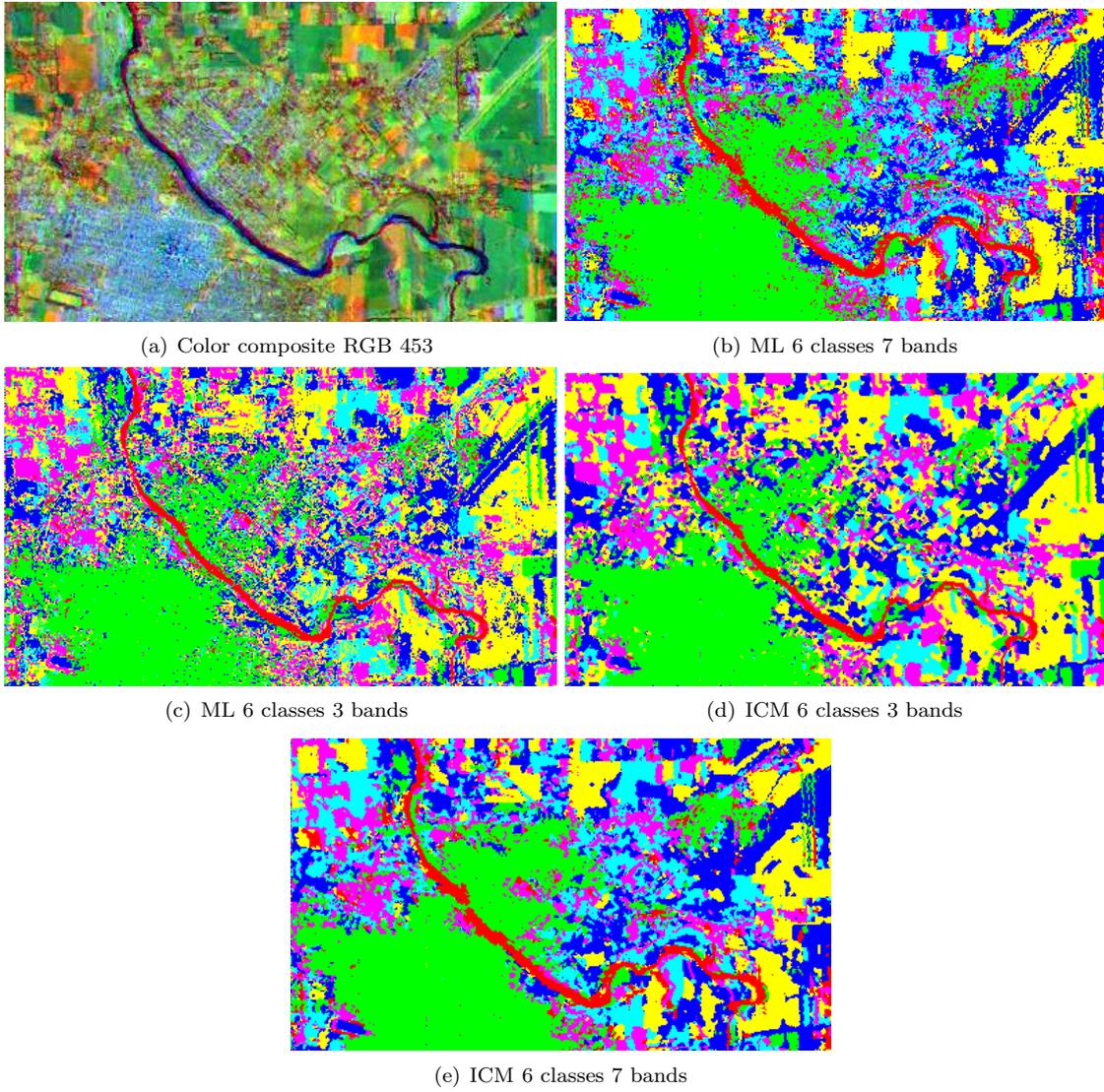


Figure 3: Color composition and maps

descriptive analysis for the identification of classes, having found 4, 5, 6 and 7 land covers.

Training and test areas were then selected, with at least 100 sites for each class, and each image was classified by ML and ICM. ICM required at most two iterations, ending with $\hat{\beta}(\cdot) \in [0.60; 0.85]$ in all situations and within the upper half of the interval in 23 out of 50 situations. After classification, the Kappa coefficient of agreement (along with its 90% confidence interval) and accuracy were estimated.

In most situations the coefficient of agreement is close to 1 regardless the classification procedure, so we can conclude that both techniques are in good agreement with the ground truth.

Regarding Kappa, ICM produce equal or better classifications than ML, and in eight out of fifty situations the improvement is statistically significant at the 90% level. Figure 4 shows the estimated values of Kappa for both classification techniques (ML squares and dashed lines, ICM circles and solid lines) in a few situations.

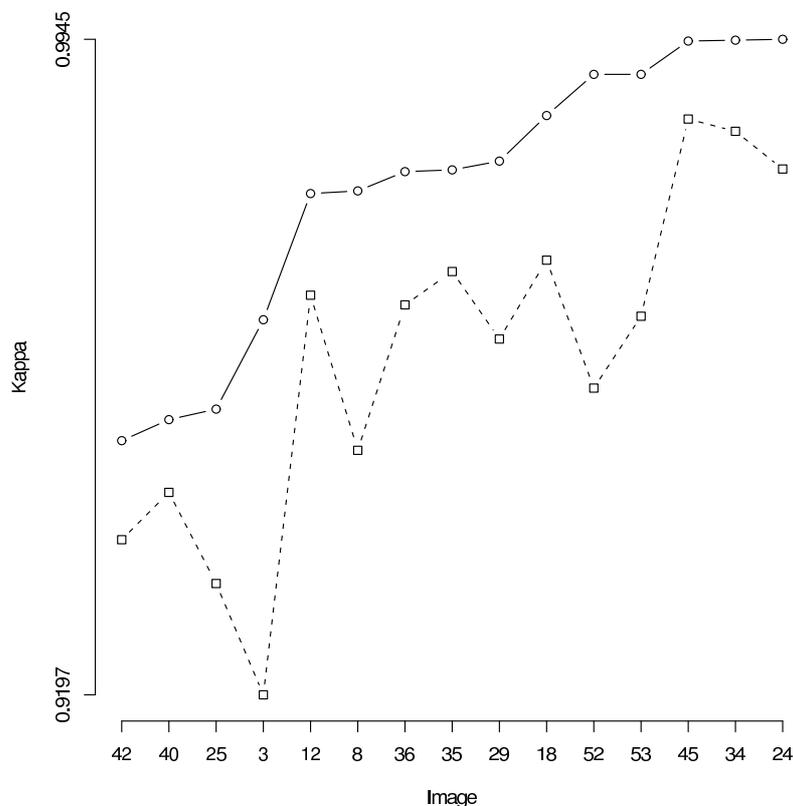


Figure 4: Kappa from ICM and ML maps

4.2 Analysis of ASTER Data

Figure 5 presents the original data and two classifications using a 560×360 pixels ASTER dataset obtained in 2001 over the Comechingones area, Córdoba, Argentina. The nominal resolution is of 15 m, and the area consists mainly of agricultural fields. The classifications were obtained using the bands 1, 2 and 3N presented in Figure 5(a).

In a first approach, three classes were defined and no significant difference between ML and

ICM classifications was found. Then, by visual inspection and using field data, six classes were defined and the two algorithms were applied obtaining the results presented in figures 5(b) and 5(c).

This result is in agreement with Flygare (1997), i.e., that when well separated classes are sought, simple classification procedures provide good results. As the number of classes grows, the potential confusion also increases and, in such cases, more sophisticated techniques are better suited.

ML classification (Figure 5(b)) has a Kappa coefficient of 0.959684, while ICM (Figure 5(c)) attains 0.978861. These results are statistically significant, so the latter classification is better than the former beyond random fluctuations.

5 Conclusions

We presented the comparison of two classification procedures for remote sensing multispectral imagery: pointwise maximum likelihood and contextual ICM. In both cases, observations were described by the multivariate Gaussian distribution. The factors under assessment were (i) number of bands (ii) number of classes and (iii) errors in the training stage (iv) spectral situations often encountered in real applications.

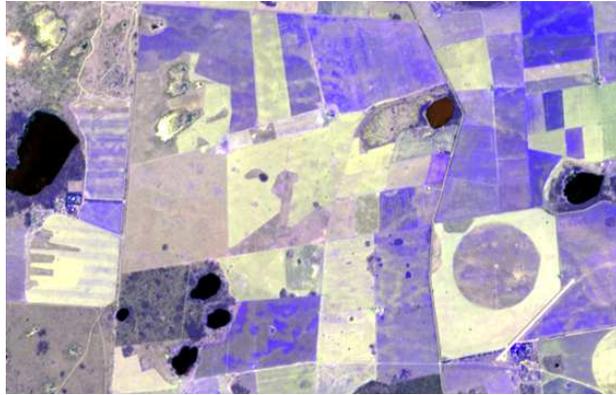
Using real data we conclude that (i) training and test samples were carefully chosen, leading to good classification results; the influence of errors in the training stage was quantified, and (ii) ICM is always better than ML, but performs the best when there is less than optimal available quality information compensating the lack of dependable spectral information with contextual evidence.

The evidence collected allows us to say that the ICM contextual classification technique is the most adequate in every situation, even if the training data are collected in a non-dependable manner, so if the computational effort required is not an issue it is always recommended.

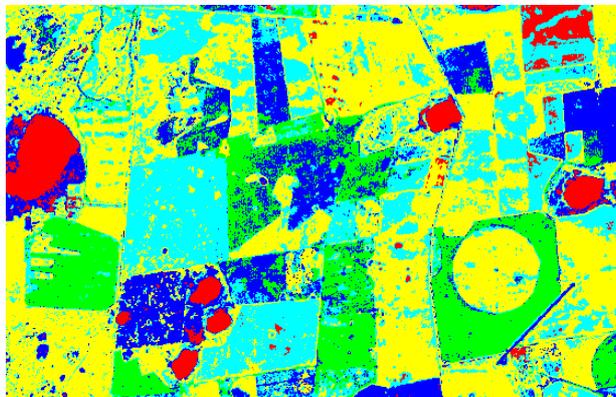
Other results are expected with models for other sensors as, for instance, speckled data from synthetic aperture radar; in this case, distributions from the Multiplicative Model (see, for instance, Frery et al., 1997) a new assessment for the worst case training errors should be used. The methodology, though, can be easily used to cover any situation that can be described in terms of statistical modelling.

References

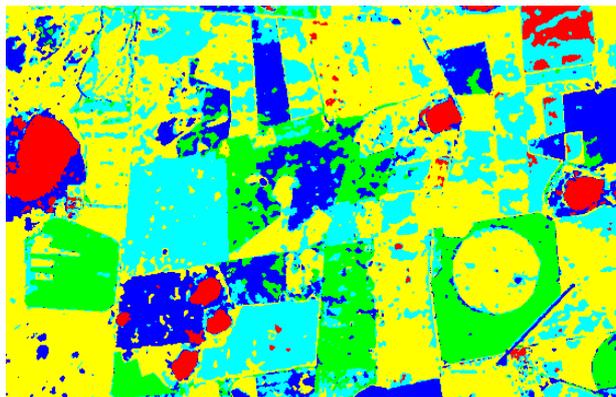
- AGRESTI, A. 1990. *Categorical Data Analysis*. John Wiley & Sons, Inc.
- ARBIA, G., BENEDETTI, R., AND ESPA, G. 1999. Contextual classification in image analysis: An assessment of accuracy of ICM. *Computational Statistics & Data Analysis* 30:443 – 455.
- ARNOLD, B. C. AND STRAUSS, D. 1991. Pseudolikelihood estimation: some examples. *Sankhya: The Indian Journal of Statistics Series B* 53:233–243.
- BELWARD, A. S. AND DEHOYOS, A. 1987. A comparison of supervised maximum-likelihood and decision tree classification for crop cover estimation from multitemporal Landsat MSS data. *International Journal Of Remote Sensing* 8:229–235.
- BESAG, J. 1989. Towards Bayesian image analysis. *Journal of Applied Statistics* 16:395–407.
- BUSTOS, O. H., FRERY, A. C., AND OJEDA, S. 1998. Strong markov processes in image modelling. *Brazilian Journal of Probability and Statistics* 12:149–194.



(a) Three bands color composition



(b) Maximum likelihood classification



(c) ICM classification

Figure 5: Classification of ASTER data

- CARNEVALLI, P., COLETTI, L., AND PATARNELLO, S. 1985. Image processing by simulated annealing. *IBM Journal of Research and Development* 29:569–579.
- CONGALTON, R. G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment* 37:35–46.
- DESCOMBES, X., MORRIS, R. D., ZERUBIA, J., AND BERTHOD, M. 1999. Estimation of Markov random field prior parameters using Markov Chain Monte Carlo maximum likelihood. *IEEE Transactions on Image Processing* 8:954–963.
- DI ZENZO, S., DEGLORIA, S. D., BERNSTEIN, R., AND KOLSKY, H. G. 1987. Gaussian maximum-likelihood and contextual classification algorithms for multicrop classification experiments using Thematic Mapper and Multispectral Scanner Sensor data. *IEEE Transactions on Geoscience and Remote Sensing* 25:815–824.
- EMERSON, C. W., LAM, N. S. N., AND QUATTROCHI, D. A. 2005. A comparison of local variance, fractal dimension, and Moran's i as aids to multispectral image classification. *International Journal Of Remote Sensing* 26:1575–1588.
- EMRAHOGLU, N., YEGINGIL, I., PESTEMALCI, V., SENKAL, O., AND KANDIRMAZ, H. M. 2003. Comparison of a new algorithm with the supervised classifications. *International Journal Of Remote Sensing* 24:649–655.
- ERBEK, F. S., OZKAN, C., AND TABERNER, M. 2004. Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities. *International Journal Of Remote Sensing* 25:1733–1748.
- FITZGERALD, R. W. AND LEES, B. G. 1994. Assessing the classification accuracy of multisource remote sensing data. *Remote Sensing of Environment* 47:368–368.
- FLYGARE, A.-M. 1997. A comparison of contextual classification methods using Landsat TM. *International Journal of Remote Sensing* 18:3835–3842.
- FRERY, A. C., FERRERO, S., AND BUSTOS, O. H. 2006. Accuracy of statistical classification strategies in remote sensing imagery. In M. M. Oliveira and R. L. Carceroni (eds.), *Proceedings XIX Brazilian Symposium on Computer Graphics and Image Processing*, pp. 255–262, Manaus, AM. IEEE Computer Press.
- FRERY, A. C., MÜLLER, H.-J., YANASSE, C. C. F., AND SANT'ANNA, S. J. S. 1997. A model for extremely heterogeneous clutter. *IEEE Transactions on Geoscience and Remote Sensing* 35:648–659.
- FRIZZELLE, B. G. AND MOODY, A. 2001. Mapping continuous distributions of land cover: A comparison of maximum-likelihood estimation and artificial neural networks. *Photogrammetric Engineering And Remote Sensing* 67:693–705.
- GAO, Y., MAS, J. F., MAATHUIS, B. H. P., ZHANG, X. M., AND VAN DIJK, P. M. 2006. Comparison of pixel-based and object-oriented image classification approaches - a case study in a coal fire area, Wuda, inner Mongolia, China. *International Journal Of Remote Sensing* 27:4039–4055.
- GEMAN, D. AND GEMAN, S. 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6:721–741.

- HUANG, K. Y. AND MAUSAL, P. M. 1994. Comparing a piecewise-linear classifier with Gaussian maximum-likelihood and parallelepiped classifiers in terms of accuracy and speed. *Photogrammetric Engineering And Remote Sensing* 60:1333–1338.
- HUBERT-MOY, L., COTONNEC, A., LE DU, L., CHARDIN, A., AND PEREZ, P. 2001. A comparison of parametric classification procedures of remotely sensed data applied on different landscape units. *Remote Sensing of Environment* 75:174–187.
- INCE, F. 1987. Maximum-likelihood classification, optimal or problematic - a comparison with the nearest neighbor classification. *International Journal Of Remote Sensing* 8:1829–1838.
- JACKSON, Q. AND LANDGREBE, D. A. 2002. Adaptive Bayesian contextual classification based on Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing* 40:2454–2463.
- LANDIS, J. R. AND KOCH, G. C. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33:159–174.
- MAGNUSSEN, S., BOUDEWYN, P., AND WULDER, M. 2004. Contextual classification of Landsat TM images to forest inventory cover types. *International Journal Of Remote Sensing* 25:2421–2440.
- J. A. Matthews, E. M. Bridges, C. J. Caseldine, A. J. Luckman, G. Owen, A. H. Perry, R. A. Shakesby, R. P. D. Walsh, R. J. Whittaker, and K. J. Willis (eds.) 2001. *The Encyclopaedic Dictionary of Environmental Change*. Arnold, London.
- MELGANI, F. AND SERPICO, S. B. 2003. A Markov random field approach to spatio-temporal contextual image classification. *IEEE Transactions on Geoscience and Remote Sensing* 41:2478–2487.
- MOSER, G., SERPICO, S. B., AND CAUSA, F. 2005. MRF model parameter estimation for contextual supervised classification of remote-sensing images. *In Geoscience and Remote Sensing Symposium Proceedings*. IEEE Computer Press.
- PICKARD, D. K. 1987. Inference for discrete Markov fields: The simplest nontrivial case. *Journal of the American Statistical Association* 82:90–96.
- RICHARDS, J. A. AND JIA, X. 1999. *Remote Sensing Digital Image Analysis: An Introduction*. Springer-Verlag, New York, 3 edition.
- TSO, B. C. K. AND MATHER, P. M. 1999. Classification of multisource remote sensing imagery using a genetic algorithm and Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing* 37:1255–1260.
- VACCARO, R., SMITS, P. C., AND DELLEPIANE, S. G. 2000. Exploiting spatial correlation features for SAR image analysis. *IEEE Transactions on Geoscience and Remote Sensing* 38:1212–1223.
- WILSON, J. D. 1992. A comparison of procedures for classifying remotely-sensed data using simulated data sets. *International Journal Of Remote Sensing* 13:365–386.
- WINKLER, G. 2006. *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction*. Stochastic Modelling and Applied Probability. Springer, 2 edition.

ZHUANG, X., ENGEL, B. A., XIONG, X. P., AND JOHANNSEN, C. J. 1995. Analysis of classification results of remotely-sensed data and evaluation of classification algorithms. *Photogrammetric Engineering and Remote Sensing* 61:427–433.

A Markov Random Fields

Let $C = \{c_1, \dots, c_L\}$ be the set of classes that describes the ground truth. Each element of the support $S = \{(i, j) : 0 \leq i \leq M - 1, 0 \leq j \leq N - 1\}$ is assigned one of these classes, C^S is the set of functions on S with value in C .

A family $\mathbf{V} = \{V_{(i,j)} : (i, j) \in S\}$ of subsets of S is a *neighborhood* of S if

1. $(i, j) \notin V_{(i,j)}$ for every (i, j) ,
2. $(i, j) \in V_{(i',j')} \Leftrightarrow (i', j') \in V_{(i,j)}$, and
3. $S = \cup_{(i,j) \in S} V_{(i,j)}$.

The pair $\mathcal{G} = (S, \mathbf{V})$ is a graph.

In this work we will consider the so-called “eight-neighbors structure”: $V_{(i,j)}^2 = \{(i', j') \in S : (i' - i)^2 + (j' - j)^2 \leq 2, (i', j') \neq (i, j)\}$, and $\mathbb{V}^2 = \{V_{(i,j)}^2 : (i, j) \in S\}$ for every $(i, j) \in S$.

A Random Field with space state C^S is a random matrix

$$\mathcal{C} = [C(i, j)], \text{ with } 0 \leq i \leq M - 1, 0 \leq j \leq N - 1,$$

such that $C(i, j)$ is a random variable with values in C . An outcome of \mathcal{C} is an element of C^S . The probability distribution of \mathcal{C} on C^S will be denoted by $\Pr_{\mathcal{C}}$, i.e., for each $\mathbf{c} \in C^S$ one has that $\Pr_{\mathcal{C}}(\mathbf{c})$ is the probability that $C(i, j) = c(i, j)$ for every $(i, j) \in S$.

Given a graph $\mathcal{G} = (S, \mathbf{V})$ and a Random Field \mathcal{C} with space state C^S , we say that \mathcal{C} is a \mathcal{G} -Markov Random Field if for every $\mathbf{c} \in C^S$ and every $(i, j) \in S$ holds that

$$\Pr_{\mathcal{C}}(A | B) = \Pr_{\mathcal{C}}(A | V),$$

where $A = \{\mathbf{c}' : \mathbf{c}'(i, j) = c(i, j)\}$, $B = \{\mathbf{c}' : \mathbf{c}'(i', j') = c(i', j'), \forall (i', j') \neq (i, j)\}$, and $V = \{\mathbf{c}' : \mathbf{c}'(i', j') = c(i', j'), \forall (i', j') \in V_{(i,j)}\}$. In words, the conditional distribution of the random variable at site (i, j) given the observation of all other sites depends only on the observed outcomes at the neighboring sites $V_{(i,j)}$.

A particular Markov Random Field model, namely the Potts model, has been widely used to describe the spatial distribution of classes in thematic maps (Bustos et al., 1998; Carnevalli et al., 1985; Pickard, 1987). In order to define it, consider $\beta \in \mathbb{R}$ a real number. A Random Field \mathcal{C} with space state C^S is a Potts model with parameter β with respect to the graph $\mathcal{G} = (S, \mathbf{V}^2)$ if

$$\Pr_{\mathcal{C}}(\mathbf{c}) = \frac{\exp(\beta \sum_{(i,j) \in S} N_{(i,j)}(\mathbf{c}))}{Z_{\beta}}, \quad (3)$$

where

$$N_{(i,j)}(\mathbf{c}) = \#\{(i', j') \in V_{(i,j)}^2 : c(i', j') = c(i, j)\},$$

and the so-called partition function is given by

$$Z_{\beta} = \sum_{\mathbf{c}' \in C^S} \exp\left(\beta \sum_{(i,j) \in S} N_{(i,j)}(\mathbf{c}')\right).$$

B Pseudolikelihood estimation

Given \mathbf{c} an outcome of the Potts model space state C^S and eight-neighbors structure, any point $\hat{\beta} \in \mathbb{R}$ that maximizes the product of all the conditional distributions $\prod_{(i,j) \in S} \Pr(C(i,j) | V_{(i,j)})$ is called “maximum pseudolikelihood estimator of β ”. Finding such points is easier than finding the maximum likelihood estimator, and it amounts to solving a nonlinear equation consisting of sixty-seven terms. Each term involves the number of sites in S for which a certain configuration has been observed. Next equation provides the expression that is used to compute $\hat{\beta}$, where k_8 denotes the number of sites where $C(i,j)$ coincides with all its eight neighbors, k_{71} denotes the number of sites where $C(i,j)$ coincides with only seven of its eight neighbors, k_{233} denotes the number of sites where $C(i,j)$ coincides with only two of its eight neighbors and the remaining six are divided in two sets of equal classes and so on; finally, $k_{011111111}$ denotes the number of sites where $C(i,j)$ is different from all its eight neighbors, and they are different among them. Note that the order in which the neighbors are arranged is irrelevant for this estimator.

$$\begin{aligned}
& k_8 \frac{8L-1}{e^{8\beta}+L-1} + k_{71} \frac{6e^{\hat{\beta}}+7L-14}{e^{7\hat{\beta}}+e^{\hat{\beta}}+L-2} + k_{62} \frac{4e^{2\hat{\beta}}+6L-12}{e^{6\hat{\beta}}+e^{2\hat{\beta}}+L-2} + k_{611} \frac{10e^{\hat{\beta}}+6L-18}{e^{6\hat{\beta}}+2e^{\hat{\beta}}+L-3} + \\
& + k_{53} \frac{2e^{3\hat{\beta}}+5L-10}{e^{5\hat{\beta}}+e^{3\hat{\beta}}+L-2} + k_{521} \frac{3e^{2\hat{\beta}}+4e^{\hat{\beta}}+5L-15}{e^{5\hat{\beta}}+e^{2\hat{\beta}}+e^{\hat{\beta}}+L-3} + k_{5111} \frac{12e^{\hat{\beta}}+5L-20}{e^{5\hat{\beta}}+3e^{\hat{\beta}}+L-4} + \\
& k_{44} \frac{4L-8}{2e^{4\hat{\beta}}+L-2} + k_{431} \frac{e^{3\hat{\beta}}+3e^{\hat{\beta}}+4L-12}{e^{4\hat{\beta}}+e^{3\hat{\beta}}+e^{\hat{\beta}}+L-3} + k_{422} \frac{4e^{2\hat{\beta}}+4L-12}{e^{4\hat{\beta}}+2e^{2\hat{\beta}}+L-3} + \\
& + k_{4211} \frac{2e^{2\hat{\beta}}+6e^{\hat{\beta}}+4L-16}{e^{4\hat{\beta}}+e^{2\hat{\beta}}+2e^{\hat{\beta}}+L-4} + k_{41111} \frac{10e^{\hat{\beta}}+4L-20}{e^{4\hat{\beta}}+4e^{\hat{\beta}}+L-5} + k_{35} \frac{-2e^{5\hat{\beta}}+3L-6}{e^{3\hat{\beta}}+e^{5\hat{\beta}}+L-2} + \\
& + k_{341} \frac{-e^{4\hat{\beta}}+2e^{\hat{\beta}}+3L-9}{e^{3\hat{\beta}}+e^{4\hat{\beta}}+e^{\hat{\beta}}+L-3} + k_{332} \frac{e^{2\hat{\beta}}+3L-9}{2e^{3\hat{\beta}}+e^{2\hat{\beta}}+L-3} + k_{3311} \frac{4e^{\hat{\beta}}+3L-12}{2e^{3\hat{\beta}}+2e^{\hat{\beta}}+L-4} + \\
& + k_{3221} \frac{2e^{2\hat{\beta}}+2e^{\hat{\beta}}+3L-12}{e^{3\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}+L-4} + k_{32111} \frac{e^{2\hat{\beta}}+6e^{\hat{\beta}}+3L-15}{e^{3\hat{\beta}}+e^{2\hat{\beta}}+3e^{\hat{\beta}}+L-5} + k_{311111} \frac{10e^{\hat{\beta}}+3L-18}{e^{3\hat{\beta}}+5e^{\hat{\beta}}+L-6} + \\
& k_{26} \frac{-4e^{6\hat{\beta}}+2L-4}{e^{2\hat{\beta}}+e^{6\hat{\beta}}+L-2} + k_{251} \frac{-3e^{5\hat{\beta}}+e^{\hat{\beta}}+2L-6}{e^{2\hat{\beta}}+e^{5\hat{\beta}}+e^{\hat{\beta}}+L-3} + k_{242} \frac{-2e^{4\hat{\beta}}+2L-6}{2e^{2\hat{\beta}}+e^{4\hat{\beta}}+L-3} + \\
& + k_{2411} \frac{-2e^{4\hat{\beta}}+2e^{\hat{\beta}}+2L-8}{e^{2\hat{\beta}}+e^{4\hat{\beta}}+2e^{\hat{\beta}}+L-4} + k_{233} \frac{-2e^{3\hat{\beta}}+2L-6}{e^{2\hat{\beta}}+2e^{3\hat{\beta}}+L-3} + k_{2321} \frac{-e^{3\hat{\beta}}+e^{\hat{\beta}}+2L-8}{2e^{2\hat{\beta}}+e^{3\hat{\beta}}+e^{\hat{\beta}}+L-4} + \\
& k_{23111} \frac{-e^{3\hat{\beta}}+3e^{\hat{\beta}}+2L-10}{e^{2\hat{\beta}}+e^{3\hat{\beta}}+3e^{\hat{\beta}}+L-5} + k_{2222} \frac{2L-8}{4e^{2\hat{\beta}}+L-4} + k_{22211} \frac{2e^{\hat{\beta}}+2L-10}{3e^{2\hat{\beta}}+2e^{\hat{\beta}}+L-5} + \\
& k_{221111} \frac{4e^{\hat{\beta}}+2L-12}{2e^{2\hat{\beta}}+4e^{\hat{\beta}}+L-6} + k_{2111111} \frac{6e^{\hat{\beta}}+2L-14}{e^{2\hat{\beta}}+6e^{\hat{\beta}}+L-7} + k_{17} \frac{-6e^{7\hat{\beta}}+L-2}{e^{\hat{\beta}}+e^{7\hat{\beta}}+L-2} + k_{161} \frac{-5e^{6\hat{\beta}}+L-3}{3e^{\hat{\beta}}+e^{6\hat{\beta}}+L-3} + \\
& + k_{152} \frac{-4e^{5\hat{\beta}}-e^{2\hat{\beta}}+L-3}{e^{\hat{\beta}}+e^{5\hat{\beta}}+e^{2\hat{\beta}}+L-3} + k_{1511} \frac{-4e^{5\hat{\beta}}+L-4}{3e^{\hat{\beta}}+e^{5\hat{\beta}}+L-4} + k_{143} \frac{-3e^{4\hat{\beta}}-2e^{3\hat{\beta}}+L-3}{e^{\hat{\beta}}+e^{4\hat{\beta}}+e^{3\hat{\beta}}+L-3} + \\
& + k_{1421} \frac{-3e^{4\hat{\beta}}-e^{2\hat{\beta}}+L-4}{2e^{\hat{\beta}}+e^{4\hat{\beta}}+e^{2\hat{\beta}}+L-4} + k_{14111} \frac{-3e^{4\hat{\beta}}+L-5}{4e^{\hat{\beta}}+e^{4\hat{\beta}}+L-5} + k_{1331} \frac{-4e^{3\hat{\beta}}-L-4}{2e^{\hat{\beta}}+2e^{3\hat{\beta}}+L-4} + \\
& + k_{1322} \frac{-2e^{3\hat{\beta}}-2e^{2\hat{\beta}}+L-4}{e^{\hat{\beta}}+e^{3\hat{\beta}}+e^{2\hat{\beta}}+e^{2\hat{\beta}}+L-4} + k_{13211} \frac{-2e^{3\hat{\beta}}-e^{2\hat{\beta}}+L-5}{3e^{\hat{\beta}}+e^{3\hat{\beta}}+e^{2\hat{\beta}}+L-5} + k_{131111} \frac{-2e^{3\hat{\beta}}+L-6}{5e^{\hat{\beta}}+e^{3\hat{\beta}}+L-6} + \\
& + k_{12221} \frac{-3e^{2\hat{\beta}}+L-5}{2e^{\hat{\beta}}+3e^{2\hat{\beta}}+L-5} + k_{122111} \frac{-2e^{2\hat{\beta}}+L-6}{4e^{\hat{\beta}}+2e^{2\hat{\beta}}+L-6} + k_{1211111} \frac{-e^{2\hat{\beta}}+L-7}{6e^{\hat{\beta}}+e^{2\hat{\beta}}+L-7} + \\
& k_{11111111} \frac{L-8}{8e^{\hat{\beta}}+L-8} + k_{08} \frac{-8e^{8\hat{\beta}}}{1+e^{8\hat{\beta}}+L-2} + k_{071} \frac{-7e^{7\hat{\beta}}-e^{\hat{\beta}}}{1+e^{7\hat{\beta}}+e^{\hat{\beta}}+L-3} + k_{062} \frac{-6e^{6\hat{\beta}}-2e^{2\hat{\beta}}}{1+e^{6\hat{\beta}}+e^{2\hat{\beta}}+L-3} + \\
& + k_{0611} \frac{-6e^{6\hat{\beta}}-2e^{\hat{\beta}}}{e^{6\hat{\beta}}+e^{\hat{\beta}}+e^{\hat{\beta}}+L-3} + k_{053} \frac{-5e^{5\hat{\beta}}-3e^{3\hat{\beta}}}{e^{5\hat{\beta}}+e^{3\hat{\beta}}+L-2} + k_{0521} \frac{-5e^{5\hat{\beta}}-2e^{2\hat{\beta}}-e^{\hat{\beta}}}{e^{5\hat{\beta}}+e^{2\hat{\beta}}+e^{\hat{\beta}}+L-3} + \\
& + k_{05111} \frac{-5e^{5\hat{\beta}}-3e^{\hat{\beta}}}{e^{5\hat{\beta}}+3e^{\hat{\beta}}+L-4} + k_{044} \frac{-8e^{4\hat{\beta}}}{2e^{4\hat{\beta}}+L-2} + k_{0431} \frac{-4e^{4\hat{\beta}}-3e^{3\hat{\beta}}-e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{3\hat{\beta}}+e^{\hat{\beta}}+L-3} + \\
& + k_{0422} \frac{-4e^{4\hat{\beta}}-4e^{2\hat{\beta}}}{e^{4\hat{\beta}}+2e^{2\hat{\beta}}+L-3} + k_{04211} \frac{-4e^{4\hat{\beta}}-2e^{2\hat{\beta}}-2e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{2\hat{\beta}}+2e^{\hat{\beta}}+L-4} + k_{041111} \frac{-4e^{4\hat{\beta}}-4e^{\hat{\beta}}}{e^{4\hat{\beta}}+4e^{\hat{\beta}}+L-5} + \\
& + k_{0332} \frac{-6e^{3\hat{\beta}}-2e^{2\hat{\beta}}}{2e^{3\hat{\beta}}+e^{2\hat{\beta}}+L-3} + k_{03311} \frac{-6e^{3\hat{\beta}}-2e^{\hat{\beta}}}{2e^{3\hat{\beta}}+2e^{\hat{\beta}}+L-4} + k_{03221} \frac{-3e^{3\hat{\beta}}-4e^{2\hat{\beta}}-e^{\hat{\beta}}}{e^{3\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}+L-4} + \\
& + k_{032111} \frac{-3e^{3\hat{\beta}}-2e^{2\hat{\beta}}-3e^{\hat{\beta}}}{e^{3\hat{\beta}}+e^{2\hat{\beta}}+3e^{\hat{\beta}}+L-5} + k_{0311111} \frac{-3e^{3\hat{\beta}}-5e^{\hat{\beta}}}{e^{3\hat{\beta}}+5e^{\hat{\beta}}+L-6} + k_{02222} \frac{-8e^{2\hat{\beta}}}{4e^{2\hat{\beta}}+L-4} + \\
& + k_{022211} \frac{-6e^{2\hat{\beta}}-2e^{\hat{\beta}}}{3e^{2\hat{\beta}}+2e^{\hat{\beta}}+L-5} + k_{0221111} \frac{-4e^{2\hat{\beta}}-4e^{\hat{\beta}}}{2e^{2\hat{\beta}}+4e^{\hat{\beta}}+L-6} + \\
& k_{02111111} \frac{-2e^{2\hat{\beta}}-6e^{\hat{\beta}}}{e^{2\hat{\beta}}+6e^{\hat{\beta}}+L-7} + k_{011111111} \frac{-8e^{\hat{\beta}}}{8e^{\hat{\beta}}+L-8} = 0
\end{aligned}$$

Though this is a lengthy function, it is easily implemented and, since it is smooth on $\widehat{\beta}$, most available routines that compute the real roots of a univariate nonlinear function will provide a result.

C Parameters for each simulation situation

C.1 Situation P1

The parameters are those reported in Richards and Jia (1999, p. 188), obtained from an image with water, fire burn, vegetation and urban areas:

$$\begin{aligned} \boldsymbol{\mu}_1 &= \begin{pmatrix} 44.27 \\ 28.82 \\ 22.77 \\ 13.89 \end{pmatrix}, \quad \boldsymbol{\mu}_2 = \begin{pmatrix} 42.85 \\ 35.02 \\ 35.96 \\ 29.04 \end{pmatrix}, \quad \boldsymbol{\mu}_3 = \begin{pmatrix} 40.46 \\ 30.92 \\ 57.50 \\ 57.68 \end{pmatrix}, \quad \boldsymbol{\mu}_4 = \begin{pmatrix} 63.14 \\ 60.44 \\ 81.84 \\ 72.25 \end{pmatrix}, \\ \mathbf{M}_1 &= \begin{pmatrix} 14.36 & 9.55 & 4.49 & 1.19 \\ 9.55 & 10.51 & 3.71 & 1.11 \\ 4.49 & 3.71 & 6.95 & 4.05 \\ 1.19 & 1.11 & 4.05 & 7.65 \end{pmatrix}, \quad \mathbf{M}_2 = \begin{pmatrix} 9.38 & 10.51 & 12.30 & 11.00 \\ 10.51 & 20.29 & 22.10 & 20.62 \\ 12.30 & 22.10 & 32.68 & 27.78 \\ 11.00 & 20.62 & 27.78 & 30.23 \end{pmatrix}, \\ \mathbf{M}_3 &= \begin{pmatrix} 5.56 & 3.91 & 2.04 & 1.43 \\ 3.91 & 7.46 & 1.96 & 0.56 \\ 2.04 & 1.96 & 19.75 & 19.71 \\ 1.43 & 0.56 & 19.71 & 29.27 \end{pmatrix}, \quad \mathbf{M}_4 = \begin{pmatrix} 43.58 & 46.42 & 7.99 & -14.86 \\ 46.42 & 60.57 & 17.38 & -9.09 \\ 7.99 & 17.38 & 67.41 & 67.57 \\ -14.86 & -9.09 & 67.57 & 94.27 \end{pmatrix}. \end{aligned}$$

C.2 Situation P2

Three bands and six classes; three of them with low mean values and the remaining three with high mean values, same covariance matrices for classes with close mean values:

$$\begin{aligned} \boldsymbol{\mu}_1 &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{\mu}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \boldsymbol{\mu}_3 = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}, \quad \boldsymbol{\mu}_4 = \begin{pmatrix} 125 \\ 125 \\ 125 \end{pmatrix}, \quad \boldsymbol{\mu}_5 = \begin{pmatrix} 142 \\ 142 \\ 142 \end{pmatrix}, \quad \boldsymbol{\mu}_6 = \begin{pmatrix} 234 \\ 234 \\ 234 \end{pmatrix}, \\ \mathbf{M}_1 = \mathbf{M}_2 = \mathbf{M}_3 &= \begin{pmatrix} 0.0100 & 0.0030 & 0.0009 \\ 0.0030 & 0.0100 & 0.0030 \\ 0.0009 & 0.0030 & 0.0100 \end{pmatrix}, \quad \mathbf{M}_4 = \mathbf{M}_5 = \mathbf{M}_6 = \begin{pmatrix} 25.00 & 7.50 & 2.25 \\ 7.50 & 25.00 & 7.50 \\ 2.25 & 7.50 & 25.00 \end{pmatrix}. \end{aligned}$$

C.3 Situation P3

Four classes and four bands; the classes have equal mean vectors and covariances, being differentiated by the variances only:

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \boldsymbol{\mu}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

$$\mathbf{M}_1 = \begin{pmatrix} 1.0000 & 0.3000 & 0.0900 & 0.0081 \\ 0.3000 & 1.0000 & 0.3000 & 0.0900 \\ 0.0900 & 0.3000 & 1.0000 & 0.3000 \\ 0.0081 & 0.0900 & 0.3000 & 1.0000 \end{pmatrix}, \quad \mathbf{M}_2 = \begin{pmatrix} 2.0000 & 0.3000 & 0.0900 & 0.0081 \\ 0.3000 & 2.0000 & 0.3000 & 0.0900 \\ 0.0900 & 0.3000 & 2.0000 & 0.3000 \\ 0.0081 & 0.0900 & 0.3000 & 2.0000 \end{pmatrix},$$

$$\mathbf{M}_3 = \begin{pmatrix} 4.0000 & 0.3000 & 0.0900 & 0.0081 \\ 0.3000 & 4.0000 & 0.3000 & 0.0900 \\ 0.0900 & 0.3000 & 4.0000 & 0.3000 \\ 0.0081 & 0.0900 & 0.3000 & 4.0000 \end{pmatrix}, \quad \mathbf{M}_4 = \begin{pmatrix} 8.0000 & 0.3000 & 0.0900 & 0.0081 \\ 0.3000 & 8.0000 & 0.3000 & 0.0900 \\ 0.0900 & 0.3000 & 8.0000 & 0.3000 \\ 0.0081 & 0.0900 & 0.3000 & 8.0000 \end{pmatrix}.$$

C.4 Situation P4

Six classes and three bands; the classes have equal zero mean vectors and proportional covariance matrices:

$$\mathbf{M}_1 = \begin{pmatrix} 1.00 & 0.30 & 0.09 \\ 0.30 & 1.00 & 0.30 \\ 0.09 & 0.30 & 1.00 \end{pmatrix}, \quad \mathbf{M}_j = j^2 \mathbf{M}_1, 2 \leq j \leq 6.$$

D Computational information

Developments were made in the IDL platform (www.rsinc.com) and incorporated into ENVI, an image processing platform developed in IDL. Plots were produced in R (www.r-project.org).