

Decision Boundaries Using Bayes Factors: The Case of Cloud Masks

F. Murtagh, D. Barreto (IEEE Member), J. Marcello

Abstract—We assess the use of an approximation to the Bayes factor for objectively assessing spatial segmentation models. The Bayes factor allows us to automatically determine thresholds, in multidimensional feature space, for such objectives as cloud mask definition. We compare our results with a cloud map currently provided as a data product.

Index Terms—Bayes factor, Bayes information criterion, BIC, model selection, multiband image, principal components analysis, Gaussian mixture model, segmentation, Markov model, Potts prior

I. INTRODUCTION

The problem of classification and related decision-making is essentially the problem of defining multiple thresholds. The optimal definition or selection of such decision boundaries is a difficult problem for a number of reasons: we are usually concerned with irregular regions in decision space; the distribution of signal and/or noise in our data is not amenable to a priori fixed setting of thresholds; and classification or decision-making is usually multivariate.

A Bayesian assessment framework provides an objective and generally-applicable approach to classification and related decision-making. In this article, we apply a Bayes factor approach to the classification of pixels as being cloud or non-cloud. The Bayes factor, developed by Jeffreys in the 1930s [11], is the posterior odds of one model over another when the prior probabilities of the two models are equal. Approximations to the Bayes factor are used in practice: we use the pseudo-likelihood information criterion, BIC_{PL} , in this article, which is the Bayes information criterion in the pseudo-likelihood case.

The contribution of this work lies in the experimental testing of BIC_{PL} as an objective criterion for defining multidimensional decision boundaries. As will be discussed below, cloud mask data products are typically determined from selected spectral bands, engineering data, and what amounts to a trained decision tree classifier. Our approach is unsupervised, fast, objective, and can be applied to any image data.

II. SPATIAL CLUSTERING

A. Model and Algorithm

Following the traditional approach [24], [19], [17] to determining the segment label map, X , we use a maximum a posteriori, or MAP, estimate as follows:

$$\hat{X} = \operatorname{argmax}_X p(X | \theta^X) \prod_k p(Y | X = k, \theta_k^Y) \quad (1)$$

for segment or class k , observed multiband vector Y , and model parameters related to the segment map θ^X and to the observed data and

F. Murtagh is at the School of Computer Science, Queen's University Belfast, Belfast BT7 1NN, Northern Ireland. D. Barreto is at the Departamento de Señales y Comunicaciones, U.L.P.G.C, Campus Univ. de Tafira, 35017 Las Palmas de Gran Canaria, Spain, and was at Queen's University Belfast when working on this paper. J. Marcello is at the Departamento de Señales y Comunicaciones, U.L.P.G.C, Campus Univ. de Tafira, 35017 Las Palmas de Gran Canaria, Spain. Author for correspondence: F. Murtagh, email f.murtagh@qub.ac.uk

segment map, θ_k^Y . To simplify notation, for the present we do not write the pixel index i for each occurrence of the label map X and the observed data Y . The first term, as will be seen below, is defined by a Potts energy function. The second term is taken in this work as a multivariate Gaussian, so that parameters θ_k^Y are segment means, variances and covariances. In equation (1), the prior is $p(X | \theta^X)$ and the likelihood of region k is modeled by $p(Y | X = k, \theta_k^Y)$.

The parameter set, $\Theta_K = \{\theta^X, \theta_k^Y | k = 1, 2, \dots, K\}$ is to be estimated, and we use maximum likelihood for this. There are other, related approaches (e.g. using MAP instead of ML: a lucid discussion can be found in [1]).

What we have described is tractable and has a considerable tradition. If I is some information criterion, then model selection is arrived at by optimizing the additional term, $\operatorname{argmax}_K I(\Theta_K)$. This leads to a fully unsupervised algorithm wherein all parameters, including now K , the number of segments, are estimated. In practice this requires results for various values of K . Model selection will be discussed in section B below.

We now return to the question of choice of prior. We will use a Markov random field, or MRF, prior [4], [24], [19], to express spatial interrelationship. This is an interaction model on a lattice structure. The process is Markov with respect to a neighborhood structure. The Potts prior is given by

$$p(X_i) \propto \exp \left(\phi \sum_{j \in \mathcal{N}_i} \delta(X_i, X_j) \right)$$

where δ is a Dirac delta function with 1 if labels $X_i = X_j$, and 0 otherwise; \mathcal{N}_i is the neighborhood of i defined as the 8 pixel nearest neighbors; and ϕ is a spatial cohesion value, taken as constant over the image pixels. This leads to the following conditional distribution:

$$p(X_i = k | \mathcal{N}_i, \phi) = \exp(\phi U(\mathcal{N}_i, k)) / \sum_j \exp(\phi U(\mathcal{N}_i, j))$$

where the energy term $U(\mathcal{N}_i, k)$ is the number of pixels in \mathcal{N}_i which have label k .

We define the pseudo-likelihood [4], [24] as

$$PL = \prod_k p(X_i = k | \mathcal{N}_i, \phi) \quad (2)$$

i.e., PL is the product of conditional distributions on the MRF, given a current realization of the labeling. We use the conditional distribution, $p(X_i = k | \mathcal{N}_i, \phi)$, as the term $p(X | \theta^X)$ in equation (1), which we estimate using maximum pseudo-likelihood.

The algorithm we use is Besag's [4] Iterative Conditional Modes, ICM. The algorithm is as follows. We use a marginal segmentation to find a first estimate, \hat{X} . Then we update the estimate of parameters $\hat{\theta}_k^Y$ by maximizing the likelihood in equation (1), given the current \hat{X} . Next we estimate ϕ using maximum PL, equation (2), again using \hat{X} as given. Finally, $\hat{\theta}_k^Y$ and $\hat{\phi}$ have now been updated, and we update \hat{X} using the overall a posteriori relation derived from equation (1):

$$\hat{X}_i = \operatorname{argmax}_k p(Y_i | X_i = k) p(X_i = k | \mathcal{N}_i, \phi) \quad (3)$$

The succession of three major stages in the algorithm is iterated until convergence. Convergence is guaranteed, but to a local optimum.

B. Model Selection using BIC and BIC_{PL}

We denote the overall model, which we have fit to our data, as M_K . M_K is defined from the distributions and all parameters, Θ_K , and the spatial interaction model.

We now wish to investigate one such model versus another, i.e. M_K versus $M_{K'}$ for two choices of numbers of classes, K and K' . With reference to cloud classification, we seek optimal K in order to provide a best estimate of cloud labels, in some particular set of segments $k \in K$.

The posterior probability of model M_K , with pixel labeling X , is

$$p(M_K | X) = \frac{p(X | M_K)p(M_K)}{\sum_{j=1}^K p(X | M_j)p(M_j)} \quad (4)$$

Taking each model as equilikely implies that $p(M_K)$ and the M_j are constant.

The Bayes factor is the posterior odds of one hypothesis when the prior probabilities of the two hypotheses are equal: $p(X | M_K)/p(X | M_{K'})$. The term $p(X | M_K)$ is the integrated likelihood rather than the maximized likelihood.

The integrated likelihood, $p(X | M_K)$, is given by

$$p(X | M_K) = \int p(X | \Theta_K, M_K)p(\Theta_K)d\Theta_K \quad (5)$$

We have that $p(X | \Theta_K, M_K)$ is the usual likelihood. Finally $p(\Theta_K)$ is the prior, which we will assume as equilikely for all M_K .

A good approximation to the integrated likelihood is given in terms of BIC, Bayes information criterion [20], [12]:

$$\text{BIC} = 2 \log p(X | \hat{\Theta}_K, M_K) - |\Theta_K| \log N \quad (6)$$

where $\hat{\Theta}_K$ is the maximum likelihood estimator of Θ_K , i.e. the result of the Gaussian mixture fitting. N is the dimensionality of the observation vectors, and $|\Theta_K|$ is the cardinality of the parameter set.

Finally the Bayes factor is approximated by the difference of BIC terms, which in turn are the maximized likelihood results of model fits for different numbers of classes, K and K' :

$$2 \log \frac{p(X | M_K)}{p(X | M_{K'})} \approx \text{BIC}(K) - \text{BIC}(K') \quad (7)$$

In operation, a plot of BIC for increasing numbers of classes, K , generally shows increase to an approximate plateau. We can usually increase the model fit indefinitely by increasing K . It is usual to consider the first peak in this plot, or the effective reaching of the plateau, to provide the optimal value of K . This is motivated by the Ockham's razor parsimony principle [10]. We can also derive the BIC term as a minimum information measure [9], [16].

In the MRF spatial interaction situation, a combinatorial explosion comes about with use of the likelihood. BIC involves integrating the likelihood of the observed data given in equation (3). There are K^N possible configurations for N pixels.

Therefore we use the more tractable pseudo-likelihood instead of the likelihood. Rather than summing over all possible configurations of X , we consider configurations close to the ICM estimate of X , \hat{X} . We consider each pixel vector Y_i in turn and condition on \hat{X}_{N_i} , i.e. the neighborhood of \hat{X}_i . Then, instead of the likelihood,

$$\sum_{k=1}^K p(Y_i | X_i = k)p(X_i = k)$$

we use the pseudo-likelihood,

$$\sum_{k=1}^K p(Y_i | X_i = k)p(X_i = k | \mathcal{N}_i, \phi)$$

which is the sum of the same expression used in equation (3). As already noted, the first term of this expression is a multivariate Gaussian distribution, and the second term is the conditional distribution used with the Potts prior density.

The pseudo-likelihood of the image (the Y_i terms are independent, conditional on the underlying hidden states) is then:

$$\begin{aligned} & \prod_i p(Y_i | X_{N_i}, \phi) \\ &= \prod_i \sum_{k=1}^K p(Y_i | X_i = k)p(X_i = k | \mathcal{N}_i, \phi) \end{aligned} \quad (8)$$

This is the likelihood integrated over the approximate posterior distribution of a set of models near the MAP estimate of X . Denoting equation (8) as $\text{PL}(Y | K)$, BIC which addresses the computational implications of the spatial model becomes [21], [22]:

$$\text{BIC}_{PL}(K) = 2 \log \text{PL}(Y | K) - |\Theta_K| \log N \quad (9)$$

BIC_{PL} , spatially-friendly BIC, is used in the same way as BIC, i.e. we vary $K = 2, 3, \dots$ and find a first relative maximum. We may overrule this principle (cf. experimentation below) if a less parsimonious result is more meaningful. In this situation the parsimony principle is the first relative maximum of BIC_{PL} above a smallest acceptable number of segments: i.e. for $K > K_0$. The curves produced do not usually increase to a plateau as in the case of BIC. Alternative spatial-respecting approximations of BIC are studied in [7].

The model selection procedure described provides a principled framework, which we show in the following sections to be scalable to real problems involving multiband (and potentially the multi-temporal, multisource data, with ancillary GIS data, used in [19]). The ICM iterative procedure used is also fast: other optimization techniques are far costlier in computational requirements (among which are those discussed in [15], [8]). It remains an open problem to carry out a comprehensive comparative assessment of trade-offs between computation time and quality of solution; or indeed the choice of priors [3].

III. CLOUD MASK ALGORITHMS

Clouds are generally characterized by higher reflectance and lower temperature than the underlying earth surface. The MODIS (Moderate Resolution Imaging Spectroradiometer) instrument provides high radiometric sensitivity in 36 spectral bands ranging in wavelength from $0.4 \mu\text{m}$ to $14.4 \mu\text{m}$. The MODIS Flight Instrument used, ProtoFlight Model or PFM, is integrated on the Terra (EOS AM-1) spacecraft commissioned as part of NASA's Earth Observing System. Terra successfully launched in December 1999.

Of the 36 MODIS spectral channels available, 17 visible and infrared radiances are used to develop the MODIS cloud mask [13]. The approach is the same as used by NOAA, based on the Saunders and Kriebel algorithm [18] (see also [5]). This scheme consists of five tests applied to each individual pixel to determine if that pixel is cloud free. The pixel is only identified as cloud if it passes every test to certain conditions. The basic idea of this cloud detection is sustained in the different contrast for cloud/cloud free pixels. MODIS has 250 m resolution in two of the visible channels, 500 m resolution in five visible and near-infrared channels, and 1000 m resolution in the remaining channels.

IV. MODIS DATA ANALYSIS

A. Description of Data

Unlike in the case of a trainable classifier (e.g., multilayer perceptron) the unsupervised classification techniques used here require careful choice of input data. If this were not so, the multivariate Gaussian $p(Y | X)$, for example, would try to cater equally for appropriate spectral bands and for irrelevant spectral bands. From the 36-band MODIS radiance dataset of the Atlantic region we chose bands 1, 2, 18, 20, 27,

29, 31, 32 and 35. These bands give information about clouds (high clouds and low clouds). The rest of the bands used by MODIS to produce their cloud mask, which are bands 4, 5, 6, 7, 19, 22, 23 and 26, give information about snow, shadow contamination, aerosol and thin cirrus. We did not include bands 5, 6 and 19 since they had holes and were therefore incomplete; bands 4, 6 and 7, associated with snow and aerosol, since they are used to set thresholds in the cloud algorithm, and thresholds are not used by us; bands 5, 19, 22 and 23, particularly useful for shadow, and again not of benefit to us; and band 26 which is a particular cirrus detector, and adequately catered for in other bands. All bands discussed here are shown in [2].

A first principal component of the selected bands is shown in Figure 1. Each band was of dimensions 2030×1354 and original pixel values were 64 bit double precision floating. The data are from 2000 September 24 at 12:10 UT. The cloud mask image provided by MODIS product MOD035.L2 is used to assess our results. The MODIS cloud mask product was designed to meet the needs of a wide variety of users, and therefore there are several ways to define a cloud mask depending on the tolerance for cloud or clear. In our case we are only interested in the best estimate as to whether there is cloud or not.



Fig. 1. MODIS first principal component image.

B. BIC_{PL} on Multiband Data

For the MODIS data, we will use the MRF-based classification, with BIC_{PL} as the parsimony model fit criterion.

Figure 2 shows the BIC_{PL} values found for a succession of numbers of segments, based on the processing of the 9-band dataset. There is

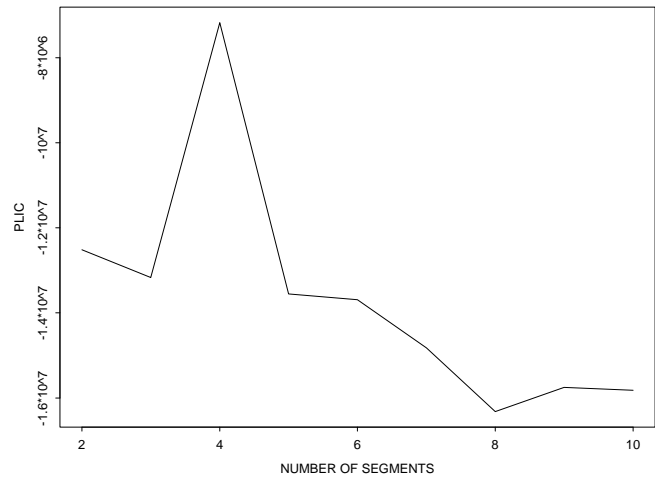


Fig. 2. BIC_{PL} values for varying number of segments, using the MODIS 9-dimensional multiband image.

TABLE I

RESULTS OF COMPARING THE SPATIAL CLUSTERING RESULT OF FIGURE 3 WITH THE GROUND TRUTH IMAGE OF FIGURE 4.

Entire image, segments 3, 4 taken as cloud	Spatial
Cloud pixels recovered (% of ground truth)	86.18%
Cloud pixels lost (% of ground truth)	13.82%
False alarms (as % of result)	27.53%
Entire image, segment 4 taken as cloud	Spatial
Cloud pixels recovered (% of ground truth)	48.09%
Cloud pixels lost (% of ground truth)	51.91%
False alarms (as % of result)	1.22%

very clear evidence in favor of a 4-segment fit. Figure 3 shows the segmentation associated with this result.

The cloud mask image is shown in Figure 4. Our results, seeking to determine cloud pixels, may also include aerosols, snow or shadow contamination, which are not likely to be found as cloud in the MODIS cloud data product.

Table I compares our result, shown in Figure 3, with the ground truth image. Taking segments 3 and 4 as cloud resulted in land areas being included too. The top part of Table I therefore yields the good result of 86% recovery of cloud pixels, but at the high cost of nearly 28% false alarms. By taking a more restrictive definition of cloud, in the lower part of Table I, we find that the false alarm rate drops to nearly 1%. But here the cost is high in that recovery rate is now less than half.

As stated, segmentation was carried out in 9-dimensional space. Such segmentation is quite unstable, for the following reasons: (i) *curse of dimensionality*: the number of pixels in a segment may become relatively small, and therefore allow singularities to come about; (ii) *determinism as opposed to stochasticity*: whether in areas related to land, cloud or sea, pixel vectors in a segment may have very tight variance, again leading to singularity; and (iii) *correlated bands*: the input data is quite highly correlated across bands, to begin with, and this is likely to be accentuated within clusters. Singularity manifests itself in problems in generating Gaussian probabilities, due to the use of the inverse of the determinant. These problems are known, of course: see [25].

We therefore sought a more stable segmentation procedure, based on the same segmentation model, and using as input a more directly interpretable data set. For this, we used a number of principal com-



Fig. 3. MODIS segmented image with 4 segments. Input: 9-dimensional multiband image. Motivating choice of number of segments is the relative maximum value in the plot of Fig. 2.



Fig. 4. MODIS cloud mask.

ponent bands, leading to the same problem as before, and finally a single principal component band. An additional justification for using a principal components analysis can be added: quality of results will always depend on quality of input data, and the denoising and variance-maximizing properties of this preprocessing are of potential value.

C. BIC_{PL} on Multiband MODIS Eigen-Image

Cumulative percentage variances explained were found for the first three components to be: 71, 87, 95.7. The eigen-band or principal component band thus accounts for 71% of the variation in this 9-band data set. Results of segmentation analysis on principal component 1 will now be discussed.

Figure 5 shows the BIC_{PL} values found for the principal component image. A 3-segment model is favored as the first relative maximum. Figure 6 shows the segmentation associated with this result. We will examine the highest valued segment only. This is shown in Figure 7. Results found were as follows: cloud pixels recovered, as a percentage of ground truth: 95.91%; cloud pixels lost, as a percentage of ground truth: 4.09%; and false alarms, as a percentage of the result: 25.88%.

From the BIC_{PL} values in Figures 5, after the 3-segment solution the next best solution is a 9-segment one. Consider the case where we rule the 3-segment solution out of order on the grounds of physical interpretation, because it confuses cloud cover with land mass. In such a

situation, we may require a larger number of segments. The 9-segment solution is shown in Figure 8, and in this 9-segment result we have a means of separating land from cloud.

In Figure 8, we have included land areas with cloud cover. It is not difficult to remove land areas explicitly, through a land map. However in the case of our images, an easy way to avoid land areas was simply to take the first 1000 pixels in the horizontal dimension, out of a total of 1354. Repeating the comparison on the non-land region gave considerably better results, as can be seen in Table II. While the accuracy of detection remained at a high level, the percentage of false alarms, i.e. pixels found by us to be cloud pixels when they were not so defined in the ground truth cloud map, fell drastically. Obtaining a good cloud mask over sea is important in its own right, for example in order to get a good Sea Surface Temperature product.

The analysis on the first principal component image scores over the analysis in the 9-dimensional multiband spaces in the following respects: results shown in Table II are better than those in Table I; the algorithm is stable and robust; and the single band input image, defined in terms of its information content (percentage variance explained) is more easily interpreted.

V. CONCLUSIONS

This article develops a cloud mask using statistical pattern recognition based on spatial clustering. The MODIS cloud mask product

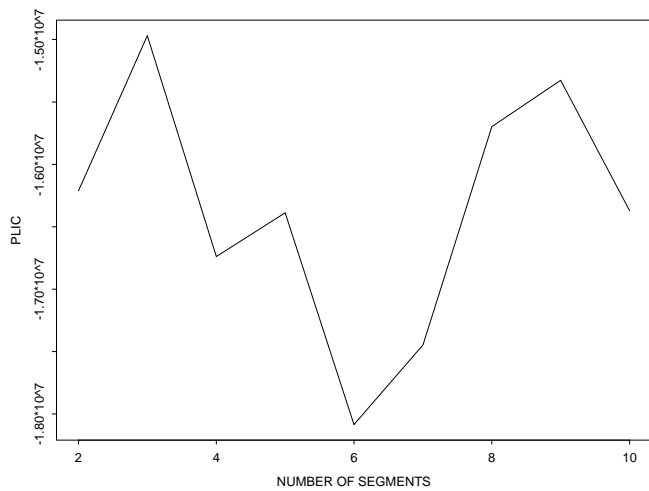


Fig. 5. BIC_{PL} values for varying number of segments, using the MODIS principal component image.

TABLE II

RESULTS OF COMPARING THE SPATIAL CLUSTERING RESULTS OF FIGURE 8 WITH THE GROUND TRUTH IMAGE OF FIGURE 4.

Entire image, segments 7,8,9 taken as cloud	Spatial
Cloud pixels recovered (% of ground truth)	99.73%
Cloud pixels lost (% of ground truth)	0.27%
False alarms (as % of result)	35.75%
Sea area only, segments 7,8,9 taken as cloud	Spatial
Cloud pixels recovered (% of ground truth)	99.68%
Cloud pixels lost (% of ground truth)	0.32%
False alarms (as % of result)	13.50%
Entire image, segment 8,9 taken as cloud	Spatial
Cloud pixels recovered (% of ground truth)	74.64%
Cloud pixels lost (% of ground truth)	25.36%
False alarms (as % of result)	18.95%
Sea area only, segments 8,9 taken as cloud	Spatial
Cloud pixels recovered (% of ground truth)	69.14%
Cloud pixels lost (% of ground truth)	38.86%
False alarms (as % of result)	1.36%

MOD035_L2 was used for assessment. It is important to take into account that validation of satellite products is a very difficult issue. Therefore the ground truth cloud masks have some errors although they are good enough to make comparisons with.

The procedure followed in this article is quite general. In our initial work we used AVHRR (Advanced Very High Resolution Radiometer) data. This sensor is carried on NOAA's Polar Orbiting Environmental Satellites (POES). Sensor AVHRR/2 was used, which is a 1100 m resolution five channel instrument on board NOAA 14 (launched December 1994). We used bands 1, 2, 4 and 5. Band 3 was not used because it is affected by sun glint and therefore is mainly useful for night images. We used the first principal component of these images. The procedure we followed, and the results obtained, were very similar to those discussed for MODIS data in this article.

A number of cloud detection algorithms are in use, but all either need set thresholds or training sets for cloud detection, and operate in localized geographic areas only [6]. The clustering technique with the objective Bayes factor model selection approach introduced in this article is useful for fast, robust and approximate cloud detection, prior to



Fig. 6. MODIS segmentation with 3 segments. Cf. first relative peak in Fig. 5.

use of, or as a cross-check on, data product masks. While the results of Tables I (based on 9-band data) and Table II (based on principal component data) are not dissimilar, the latter approach has been found to give better quality results, and it provides for a much more stable analysis method. Given how knowledge of land areas yielded a better result (in section IV), it is clear that one future direction of investigation will be tree-based methods [17], based on Bayes factors [14].

It is a moot point as to how unsupervised our approach is, given that particular image bands were selected, followed on occasion by our overruling of the Bayes factor outcome in favor of the next most parsimonious result. It is true that some supervision has been used just as we have overruled the Occam razor parsimony principle. However we have also demonstrated quite clearly that these principles can inform the algorithms used for approximate or confirmatory cloud mask finding. Even the most unsupervised of classification or other algorithms requires careful choice of input data and algorithm parameters.

REFERENCES

- [1] S.A. Barker, *Image Segmentation using Markov Random Field Models*, PhD thesis, Department of Engineering, University of Cambridge, 1998.
- [2] D. Barreto, "Selection of MODIS bands for cloud map", technical report, <http://main.cs.qub.ac.uk/~fmurtagh/papers/MODISbands.pdf>, 20 pp., 2003.



Fig. 7. From MODIS segmentation with 3 segments, shown here is the segment with largest mean.

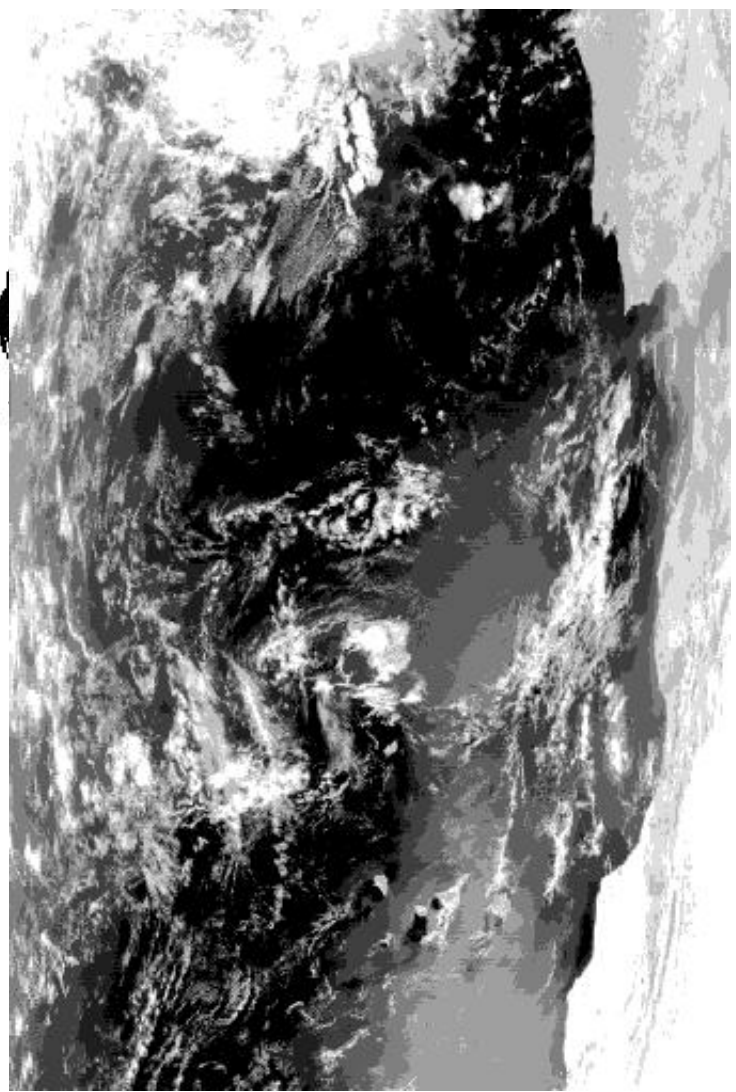


Fig. 8. MODIS segmentation with 9 segments. Cf. relative peak in Fig. 5 corresponding to the 9-segment result.

- [3] J.O. Berger, "Bayesian analysis: a look at today and thoughts of tomorrow", *Journal of the American Statistical Association*, 95, 1269–1276, 2000.
- [4] J. Besag, "Statistical analysis of dirty pictures", *Journal of the Royal Statistical Society, Series B*, vol. 48, pp. 259–302, 1986.
- [5] A.V. Di Vittorio and W.J. Emery, "An automated, dynamic threshold cloud-masking algorithm for daytime AVHRR images over land", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, pp. 1682–1694, 2002.
- [6] F. Eugenio, *Processing, Modeling and Geo-Referencing Remote Sensing Images: Automatic Methodology to Obtain Oceanographic Parameters*. PhD thesis, Departamento de Señales y Comunicaciones de la Universidad de las Palmas de G.C., 2000.
- [7] F. Forbes and N. Peyrard, "Hidden Markov model selection criteria based on mean field-like approximations", technical report 4371, INRIA Rhône-Alpes, 2001.
- [8] P.J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination", *Biometrika*, 82, 711–732, 1995.
- [9] M.H. Hansen and Bin Yu, "Model selection and the principle of minimum description length", *Journal of the American Statistical Association*, vol. 96, pp. 746–774, 2001.
- [10] W.H. Jefferys and J.O. Berger, "Ockham's razor and Bayesian analysis", *American Scientist*, vol. 80, pp. 64–72, 1992.
- [11] H. Jeffreys, *Theory of Probability*, 3rd ed., Oxford University Press, 1961.
- [12] R.E. Kass and A.E. Raftery, "Bayes factors", *Journal of the American Statistical Association*, vol. 90, pp. 773–795, 1995.
- [13] MODIS Cloud Mask Team, Discriminating clear-sky from cloud with MODIS algorithm theoretical basis document (MOD35), 1997.
- [14] F. Murtagh, J.L. Starck and A.E. Raftery, "Bayesian inference for multi-band image segmentation via model-based cluster trees", submitted.
- [15] J.M. Pérez and J. Berger, "Expected posterior prior distributions for model selection", *Biometrika*, 89, 491–512, 2002.
- [16] J. Rissanen, "Stochastic complexity and modeling", *Annals of Statistics*, 14, 1080–1100, 1986.
- [17] A. Sarkar, M.K. Biswas, B. Kartikeyan, V. Kumar, K.L. Majumder and D.K. Pal, "A MRF model-based segmentation approach to classification of multispectral imagery", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, pp. 1102–1113, 2002.
- [18] R.W. Saunders and K.T. Kriebel, "An improved method for detecting clear sky and cloudy radiances from AVHRR data", *International Journal of Remote Sensing*, vol. 9, pp. 123–150, 1988.
- [19] A.H. Schistad Solberg, T. Taxt and A.K. Jain, "A Markov random field model for classification of multisource satellite imagery", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, pp. 100–113, 1996.
- [20] G. Schwarz, "Estimating the dimension of a model", *The Annals of Statistics*, vol. 6, pp. 461–464, 1978.
- [21] D.C. Stanford, *Fast Automatic Unsupervised Image Segmentation and Curve Detection in Spatial Point Patterns*, PhD thesis, Department of Statistics, University of Washington, 1999.
- [22] D.C. Stanford and A.E. Raftery, "Approximate Bayes factors for image segmentation: the pseudolikelihood information criterion (PLIC)", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1517–1520, 2002.

- [23] K. Strabala, *MODIS Cloud Mask User's Guide*, 1999.
 - [24] Chee Sun Won and H. Derin, "Unsupervised segmentation of noisy and textured images using Markov random fields", *CVGIP: Graphical Models and Image Processing*, vol. 54, pp. 308–328, 1992.
 - [25] K. Yamazaki and S. Watanabe, "Singularities in mixture models and upper bounds of stochastic complexity", *International Journal of Neural Networks*, forthcoming, 2003.
-