

Probabilistic classification of forest structures by hierarchical modelling of the remote sensing process*

J. L. Moffett^a, J. E. Besag^b, S. D. Byers^b, and W.-H. Li^c

^aSilviculture Laboratory, College of Forest Resources, Box 352100

^bStatistics Department, Box 354322

^cGeological Remote Sensing Laboratory, Box 351310

University of Washington, Seattle, WA 98195 USA

ABSTRACT

Satellite sensors observe upwelling radiant flux from the Earth's surface. Classification of forest structures from these measurements is a statistical inference problem. A hierarchical model has been developed by linking several sub-models which represent the image acquisition process and the spatial interaction of the classes. The model for blur assumes the underlying, unobserved image is degraded according to the system point spread function. The model for topographic effects assumes the unblurred pixel values are determined by the corresponding bidirectional reflectance distribution function (BRDF) and the mean spectral reflectance of each class. A discrete Markov random field (MRF) model provides information about the spatial contiguity of the classes. Prior distributions are specified for the mean and covariance parameters. Bayes theorem is used to construct a posterior probability distribution for the classification given the data. Due to the high dimensionality of the resulting MRF, estimates of image attributes are obtained using a Markov chain Monte Carlo technique. The marginal posterior modes (MPM) point estimate minimizes the expected number of misclassifications by maximizing the marginal probability with which each pixel is classified. The advantages of this approach include the ability to specify a unique BRDF for each class and to have posterior probability estimates provide spatially explicit information about the certainty of the MPM estimate. Limitations of the model include the assumptions necessary for modelling bidirectional reflectance, the difficulty of defining classes at an appropriate scale, and assessing the accuracy of probabilistic classifications. Specimen results using Landsat TM data are presented.

Keywords: Landsat, hierarchical, bidirectional reflectance, Bayesian, Markov

1. INTRODUCTION

The classification of forest types from image data is an inference process that reduces raw data to useful information. Satellite remote sensing can be used as a means of stratifying land cover in the first stage of a multi-stage sampling scheme for forest inventory. Classified forest maps further provide necessary information for landscape planning such as the distribution of stand structure classes by area across a watershed. Probabilistic contextual classifiers are well suited to modelling the uncertainty inherent in the process.

Section 2 of the paper briefly discusses the remote sensing process. Sections 3, 4 and 5 describe the methodology and components of the probabilistic model. We present some results in Section 6 and conclude with a brief discussion in Section 7.

2. THE REMOTE SENSING PROCESS

2.1. The system point spread function

The Landsat Thematic Mapper (TM) measures upwelling radiance in analogue form as a continuously varying electrical signal. The instantaneous field of view (IFOV) is the area over which the energy measured is integrated at any instant of time. The size of the IFOV depends upon the optical aperture of the sensor and determines the

*To be presented at the SPIE International Symposium on Optical Science, Engineering, and Instrumentation, July 1997, San Diego.
Send correspondence to: jmoffett@silvae.cfr.washington.edu

nominal resolution of the sensor. The continuous signal initially recorded by the sensor is sampled in time to produce discrete measurements across the scan line. The TM sampling interval is 28.5 m compared to the nominal resolution of 30 m.¹ In addition to this oversampling, each pixel detects radiance from beyond the range of the nominal IFOV as determined by the system point spread function (PSF).² The system PSF is the two-dimensional inverse Fourier transform of the system modulation transfer function (MTF), which is the product of the individual MTFs associated with the detectors, satellite motion, electronics, and atmospheric scattering.³ The effect of the overall system PSF is to increase the IFOV with the resulting resolution referred to as the effective instantaneous field of view (EIFOV). The blurring attributable to the system PSF reduces target contrast and diffuses the boundaries between adjacent land cover types.⁴ Path radiance is also incident upon the sensor.

2.2. Bidirectional reflectance

Forest canopies scatter solar irradiance anisotropically. The bidirectional reflectance distribution function (BRDF) is an intrinsic measure of the directional scattering of a surface. The BRDF is defined as the ratio of the radiance scattered into a given direction to the irradiance incident upon a surface from a given direction (in units of sr^{-1}), thus characterizing directional reflectance in terms of all possible sun-surface-sensor geometries. Bidirectional reflectance is often expressed in a unitless form as a bidirectional reflectance factor (BRF), which is the ratio of the radiance reflected into a particular direction to the radiance that would be reflected into the same direction by a perfect Lambertian surface.³

The BRDF is generally considered to be independent of wavelength for the TM bands. However the overall magnitude of reflectance in a given direction depends upon the spectral reflectance of the surface.⁵ For a given forest cover type, the BRDF changes as a function of slope, because trees grow geotropically.⁶ In order to properly account for the effects of topography, it is necessary to have knowledge of the surface BRDF in addition to the illumination geometry.

Several field measurement studies have been conducted to quantify the BRDFs of forest canopies.⁷⁻¹⁰ In general, these studies have shown significant anisotropy in the solar principal plane. Maximum reflectances occur in the backscatter directions, with a hot spot at the zero phase angle. Minimum reflectances occur in the forward scatter directions, attributable to within crown and between crown shadowing effects. Recent research shows that crown structure and height variation have a significant effect on bidirectional reflectance,¹¹ and that BRDFs vary with the type of land cover,^{10,12} suggesting that bidirectional reflectance can provide useful information for discriminating different forest and land cover types, as opposed to complicating topographic correction procedures.

3. BAYESIAN IMAGE ANALYSIS METHODOLOGY

For a given scale, the desired classification can be modelled as a discrete Markov random field (MRF) x with an associated probability distribution $\pi(x)$. This distribution allows prior information about the spatial configuration of the classes to be incorporated into the classification process, along with spectral information. The spatial information pertains to the local characteristics of the classification, so that adjacent pixels can be deemed more likely to belong to the same class, objects can be encouraged to have closed boundaries and noisy configurations can be suppressed.

Different processes, such as atmospheric scattering and topographic distortion, degrade the image data. The Bayesian approach allows a succession of sub-models for the several sources of variation within a hierarchical framework. Rather than preprocessing the image data to correct for, or “remove” unwanted variation, this approach links the raw image data with the eventual classification through a series of physically based models of remote sensing processes. In addition, Bayesian methods incorporate spatial information *without* losing fidelity to the data, in contrast to post-classification smoothing algorithms, such as majority filters.

Previously a simple MRF model for forest classification in mountainous terrain demonstrated the benefits of incorporating spatial information and the need to account for the influence of topography.¹³ Examples of applied hierarchical models are presented in Refs. 14,15. Green *et al.* used multivariate Landsat data to classify the rural-urban interface in New Jersey and Wilson *et al.* developed a univariate model incorporating various sensor effects for classifying land cover on flat terrain.

4. A BAYESIAN HIERARCHICAL MODEL

4.1. Observed data

Consider a two-dimensional rectangular array of pixels over a finite region S , with each pixel identified by an index $i = 1, 2, \dots, n$. For each pixel in S , and for each spectral band, a remote sensor measures the upwelling radiance degraded by a stochastic process. The imperfect observations form the raw data $y = \{y_i : i \in S\}$. For the Landsat Thematic Mapper, each y_i has six components, corresponding to bands 1-5 and 7. More generally, we index the components of y_i by $p, q = 1, \dots, r$.

4.2. Markov random fields

Let $x = \{x_i : i \in S\}$ denote a random vector that classifies each pixel in S into one of m unordered classes, labeled $1, \dots, m$. We do not differentiate between random quantities and particular realizations in our notation. The primary objective of classification is to estimate the “true” class image x^* from the data y . In Bayesian image analysis, x^* is represented as a realization of a *locally dependent Markov random field* (MRF) having a probability distribution $\pi(x)$. We discuss the choice of $\pi(x)$ further below but the basic idea here is that, in forest classification, stands are typically much larger than individual pixels and that this type of contextual information should be incorporated in estimating x^* from y .

The MRF approach allows one to focus on a conditional distribution for each x_i . However, unlike time-dependent variables, there is no concept of past, present and future and the only logical conditioning set for x_i is x_{-i} , the set of *all* other pixel values in the configuration x . An MRF is formulated by first defining a Markov property, namely a set of *neighbors* $x_{\partial i}$ for each pixel i , such that $\pi(x_i|x_{-i}) \equiv \pi(x_i|x_{\partial i})$. Then the identification of these n *conditional* distributions provides an explicit formula for the joint distribution $\pi(x)$. Indeed, it can be shown that, even without this final identification, $\pi(x)$ must satisfy

$$\pi(x) \propto \exp \left\{ - \sum_C V_C(x_C) \right\}, \quad (1)$$

where the V_C s are arbitrary finite *potential functions* defined on the configurations x_C of the cliques C of the neighborhood system; here a clique is a maximal set of pixels, all of which are mutual neighbors. For further details, see Ref. 16. More obviously, any choice of potential functions determines the conditional distributions $\pi(x_i|x_{\partial i})$. In practice, the specification of an MRF often involves an interplay between these two perspectives of conditional probability and potential functions.

As an example, consider the *second-order pairwise interaction* MRF,¹⁷ for which the neighbors of each (interior) pixel are defined to be the eight pixels surrounding it, and $\pi(x)$ is given by

$$\pi(x) \propto \exp \left\{ \sum_{1 \leq k \leq m} \alpha_k n_k - \sum_{1 \leq k < l \leq m} \beta_{kl} n_{kl} \right\}, \quad (2)$$

where n_k is the number of pixels in class k and n_{kl} is the number of neighboring pairs of pixels having class (k, l) . The α_k 's and the β_{kl} 's govern the percentage of pixels in each class and the strength of interaction between different classes; for $\beta_{kl} > 0$ classes k and l are discouraged from being neighbors. Often the model is reduced to the Potts form in which $\alpha_k = 0$ for all k and $\beta_{kl} = \beta$ for all $k < l$, so that classes are treated as exchangeable. This MRF, with a particular choice or estimate of β , has proved popular in imaging applications but the restriction to pairwise interactions does not allow recognition of region shape.

As an alternative, consider the class of MRFs introduced by Ref. 18 and elaborated by Ref. 19. Such models allow greater control over the spatial configuration of the class areas but at the expense of additional complexity. Thus, the neighborhood of each pixel i consists of 24 other pixels, in a 5×5 block centered on i but of course excluding i itself. The resulting cliques C are 3×3 blocks of pixels and the corresponding functions V_C are specified in terms of three constants that together regulate the suppression of noise, the contiguity properties of different regions and the configuration of edges. Another potentially useful class of MRFs is described in Ref. 20.

4.3. Bidirectional reflectance model

The uncertainty of the BRDF and the variability in class reflectances can be modelled by specifying an unobserved, unblurred image $z = \{z_i : i \in S\}$. Assuming the z_i to be Gaussian and conditionally independent, given x , leads to the following distribution

$$\pi(z|x) \propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (z_i - T_{x_i} \mu_{x_i})' \Theta^{-1} (z_i - T_{x_i} \mu_{x_i}) \right\}, \quad (3)$$

where T_{x_i} is a diagonal matrix representing the BRFs for class x , pixel i . The mean spectral reflectances, μ_{x_i} , also for class x , pixel i , are expressed as digital numbers. The common covariance matrix is Θ . Note that here and below, we use π generically to denote a probability density.

4.4. System point spread function model

In order to model the effect of the system PSF, each y_i is assumed to be Gaussian with a mean that is a weighted average of the z_j for pixels j in the proximity of pixel i . Assuming that the y_i are conditionally independent, given z , we obtain

$$\pi(y|z) \propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (y_i - \sum_j h_{ij} z_j)' \Phi^{-1} (y_i - \sum_j h_{ij} z_j) \right\}, \quad (4)$$

where h is the system PSF detecting radiance from pixels j in the proximity of pixel i , and including i itself, and Φ is the common covariance matrix.

4.5. Prior distributions for the mean and precision parameters

Traditional classification methods, such as maximum likelihood, estimate the class means and covariances from training data. The Bayesian framework provides a method for modelling the uncertainty of such estimates by specifying prior distributions for the parameters. The class means are assigned the same Gaussian distribution with each mean weighted by training data, achieved by fixing a certain number of x_i in each class corresponding to training sites for each class. The two precision matrices are each assigned conjugate Wishart distributions. These priors are written as follows

$$\mu_c \sim N_r(\eta, \Gamma), \quad c = 1, 2, \dots, m, \quad (5)$$

$$\Phi^{-1} \sim W_r([\omega R]^{-1}, \omega), \quad (6)$$

$$\Theta^{-1} \sim W_r([\omega R]^{-1}, \omega). \quad (7)$$

5. IMAGE ESTIMATION

5.1. The joint posterior distribution

As mentioned above, the primary objective is to estimate x^* , together with an assessment of uncertainty. Inferences about x^* are based on the posterior distribution for the classification, given the data, $\pi(x|y)$. This distribution arises from integrating out all the other variables from the full joint posterior distribution for the random field, which is constructed using Bayes theorem and written as

$$\pi(x, z, \mu_x, \Phi, \Theta|y) \propto \pi(y|z, \Phi) \pi(z|x, \mu_x, \Theta) \pi(x) \pi(\mu_x) \pi(\Phi) \pi(\Theta). \quad (8)$$

Note that further marginalization over x_{-i} provides the posterior probability $\pi(x_i = c|y)$ that pixel i falls in any particular class c . In practice all such marginalizations are carried out automatically; see Section 5.3.

5.2. The marginal posterior modes estimate

The marginal posterior modes (MPM) estimate minimizes the expected number of misclassifications under the model by choosing, for each pixel, the class that has maximum posterior probability given the data, and is written as follows

$$\hat{x}_i = \arg \max \pi(x_i|y). \quad (9)$$

The MPM estimate has been adopted quite widely for classification problems; see, for example, Refs. 21,15,14.

5.3. Markov chain Monte Carlo

The MPM estimate cannot be calculated by analytical or deterministic numerical methods because of the high dimensionality of the problem and so we resort to Markov chain Monte Carlo (MCMC) methods, which have become the standard tool in Bayesian estimation in recent years; see, for example, Ref. 22. A sequence of stochastically dependent classifications is generated by simulating a very long discrete-time Markov chain having (8) as its stationary distribution. On each iteration, all the components of x , z , μ_x , Φ^{-1} and Θ^{-1} are updated in accordance with their full conditional distributions arising from the posterior (8); see below. Information about any subset of the components is extracted from the output, simply by ignoring the remaining components. For example, $\pi(x_i = c|y)$ is estimated by the relative frequency of class c on pixel i and hence the MPM estimate \hat{x} is obtained by noting the most frequently occurring class for each pixel.

5.4. Full conditional distributions

When the full conditional distributions are sufficiently simple, MCMC updates can be carried out via Gibbs sampling²³ but otherwise Metropolis²⁴ or Hastings²⁵ steps are recommended; see Ref. 22. In the present case, we use Gibbs steps for all the components, except for those of x itself, which we update via the Metropolis method.

The full conditional for z_i^p , the component of z at pixel i , band p , is Gaussian with mean $\mu_{z_i^p}$ and variance $\sigma_{z_i^p}^2$ defined as follows:

$$\begin{aligned} \frac{\mu_{z_i^p}}{\sigma_{z_i^p}^2} &= \Theta_{p,p}^{-1} t_{x_i}^p \mu_{x_i}^p + \sum_{q \neq p} \Theta_{p,q}^{-1} (t_{x_i}^q \mu_{x_i}^q - z_i^q) \\ &+ \Phi_{p,p}^{-1} \left\{ \sum_{k \in (\nu_i \cup i)} h_{ki} y_k^p - h_{ii} \sum_{j \in \nu_i} h_{ij} z_j^p - \sum_{k \in \nu_i} h_{ki} \sum_{j \in \xi_k^i} h_{kj} z_j^p \right\} \\ &+ \sum_{q \neq p} \Phi_{p,q}^{-1} \left\{ \sum_{k \in (\nu_i \cup i)} h_{ki} y_k^q - z_i^q \sum_{k \in (\nu_i \cup i)} h_{ki}^2 - h_{ii} \sum_{j \in \nu_i} h_{ij} z_j^q - \sum_{k \in \nu_i} h_{ki} \sum_{j \in \xi_k^i} h_{kj} z_j^q \right\}, \end{aligned} \quad (10)$$

where ν_i represents the pixels in the proximity of i as given by the PSF, excluding pixel i , and $\xi_k^i = (\nu_k \cup k) \setminus i$ indicates all of the pixels in the neighborhood of k , including pixel k but excluding pixel i ;

$$\sigma_{z_i^p}^2 = \frac{1}{\Theta_{p,p}^{-1} + \Phi_{p,p}^{-1} \sum_{k \in (\nu_i \cup i)} h_{ki}^2}. \quad (11)$$

The full conditional for μ_c^p , the mean spectral reflectance of class c , band p , is also Gaussian with analogous formulae for its mean and variance.

The full conditional for the precision matrix Θ^{-1} is a Wishart distribution, written as

$$\pi(\Theta^{-1} | \dots) \sim W_r \left(\left[(\omega R) + \sum_{i=1}^n (z_i - T_{x_i} \mu_{x_i})(z_i - T_{x_i} \mu_{x_i})' \right]^{-1}, n + \omega \right), \quad (12)$$

where “...” denotes “all other variables.” A similar expression is found for Φ^{-1} . These matrices are sampled using a method outlined in Ref. 26.

Finally, with the Potts MRF, the full conditional for $\pi(x_i = c | \dots)$ is given by

$$\begin{aligned} \pi(x_i = c | \dots) &\propto \exp \left\{ \sum_{p=1}^r \Theta_{p,p}^{-1} (z_i^p t_{c_i}^{p,p} \mu_c^p - \frac{1}{2} (t_{c_i}^{p,p})^2 (\mu_c^p)^2) \right. \\ &\left. + \sum_{p < q} \Theta_{p,q}^{-1} (z_i^p t_{c_i}^{p,q} \mu_c^q + z_i^q t_{c_i}^{q,p} \mu_c^p - t_{c_i}^{p,p} t_{c_i}^{q,q} \mu_c^p \mu_c^q) + \sum_{j \in \partial i} \beta_{ij} 1_{[x_i = x_j]} \right\}, \end{aligned} \quad (13)$$

with an adjustment to the final term for the alternative Descombes *et al.* model.

Given the current class x_i , the Metropolis method requires a proposal x'_i , chosen uniformly from the remaining classes, which is then accepted with probability

$$\min \left\{ 1, \frac{\pi(x'_i | \dots)}{\pi(x_i | \dots)} \right\}, \quad (14)$$

else x_i is retained. Alternatively, Gibbs steps could easily be implemented here as well.

6. RESULTS

6.1. Landsat image data and location

As an illustration of the above methodology, a TM image was classified using both Potts and Descombes *et al.* models for spatial contiguity. A 281×235 pixel subset was taken from an image acquired July 7, 1991, of the Cascade Mountains in the southern part of Washington State. The rugged terrain in this region complicates forest classification. Six forest types common to the Pacific Northwest were defined in terms of structure²⁷: stand initiation (0), shrub/sapling (1), stem exclusion (2), understory reinitiation (3), old growth (4), and hardwood (5). Figure 1 shows a “ground truth” classification of field data developed using a GIS. The numbers in Figure 1 identify the forest classes mentioned above. Rock outcrops, denoted by diagonal shading, have not been classified.

The image was registered to a USGS digital elevation model (DEM) using the UTM coordinates provided with the image. The DEM was resampled to 25 m. Path radiance offsets were subtracted from each band, by identifying the minimum values in each band over a 3400×4000 region of the full scene.²⁸

6.2. Fixed parameters

6.2.1. BRFs

The BRFs were calculated from a canopy surface application of a radiosity model²⁹; this accounts for the effects of multiple scattering among tree crowns, a significant component of forest canopy reflectance.^{29,11} We first developed canopy functions for three different forest structures: hardwood (relatively high density), high density conifer and low density conifer. For each type, we computed BRFs as a function of sun elevation, local slope and aspect, tree shape and size, and stem density. Slope and aspect were calculated from the DEM. In addition, we added an atmospheric scattering term to account for diffuse sky irradiance in each wavelength. Modelling the contribution of sky irradiance yielded improved results in similar applications; see, for example, Refs. 30,31.

A unique set of BRFs for each class were then specified from the canopy functions in accordance with an assumed spectral reflectance, or reflectivity, of the scene elements in each wavelength. For example, the BRFs for TM band 4 of the understory reinitiation class were computed from the low density conifer canopy function, assuming 45 percent reflectivity, while the BRFs for the same band of the old growth class were computed from the same canopy function, but with a 35 percent reflectivity.

6.2.2. PSF

Accurate measurements of the system PSF are difficult to obtain. Several attempts suggested the TM EIFOV ranges from 35 to 50 m with a mean between 40 and 45 m and probable sidelobes,^{32,1,33} perhaps from oversampling along the scan line. Given a resampled pixel size of 25 m and assuming an EIFOV of 45 m gives the following estimate of h in terms of a nine pixel region

$$\begin{pmatrix} .05 & .10 & .05 \\ .15 & .30 & .15 \\ .05 & .10 & .05 \end{pmatrix}.$$

Although, in reality, h is wavelength dependent, we assume that it is constant here, with plans to model a wavelength dependent PSF in future applications.

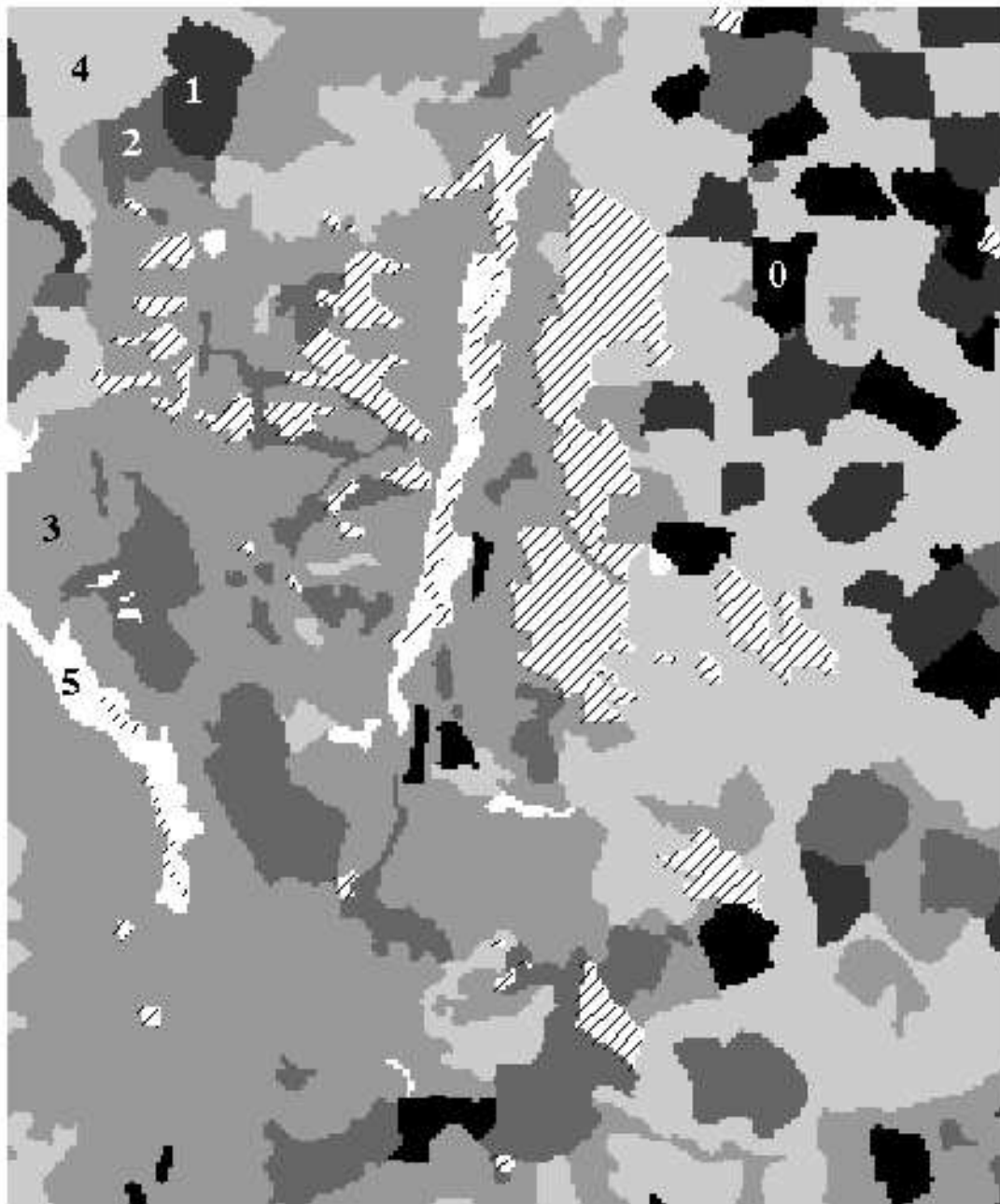


Figure 1. “Ground truth” classification of six forest structure classes for a portion of the Gifford Pinchot National Forest. This was created by categorizing field observations of canopy cover, stem density, size class and vegetation type with a GIS. Rock outcrops were left unclassified and denoted by diagonal shading.

6.2.3. Parameters for the prior distributions

In classifying the image using the Potts prior, the value of β was set to 0.5 with the diagonal pixel interaction downweighted by a factor of $\sqrt{2}$. The three parameters required for the Descombes *et al.* model were assigned the values 1.33 for noise suppression, 0.67 for region boundaries and 0.67 for boundary interaction.

In the absence of strong prior information, we assigned vague priors to the class means and the precisions.

6.3. Simulation

To begin simulation, we assigned arbitrary initial classes to x . Initial values are required for the auxiliary image z and parameters μ_x , Θ^{-1} and Φ^{-1} . Initial values of the class means and precisions were estimated from the training data, and refined by taking the convergence values from a Potts model simulation with a low value of β . The initial z values were specified as the product of the class means and corresponding BRFs.

In order to eliminate the influence of the initialization and to ensure that the Markov chain had converged to its equilibrium distribution before collecting samples, the simulation was run for a predetermined number of iterations prior to sampling, referred to as the burn-in period. We assessed convergence by monitoring the stability of the number of pixels in each class and their mean spectral reflectances, as well as the number of pixels that changed on each iteration. We also conducted several runs with very different initial classifications, such as random and single class, to be confident that the correct limiting distribution had been reached.

6.4. MPM and posterior probability estimates

The MPM estimate obtained using the Descombes *et al.* prior, see Figure 2, is characterized by large contiguous areas corresponding to the homogeneous regions of understory reinitiation and old growth forests classified with high probability. In the several areas corresponding to the rock outcrops and younger forest classes the MPM classification is characterized by noisy configurations. The MPM estimate obtained using the Potts model, not shown due to page limitations, exhibits a greater degree of noisy patterns.

The marginal posterior probabilities of the presence of stem exclusion forest in each pixel are shown in Figure 3. Brighter pixels indicate a higher probability of this class. These probabilities were estimated using the Descombes *et al.* prior. While the MPM estimate does not correspond to the “ground truth” as well for this class, Figure 3 shows that many of the regions classified as stem exclusion using the field data have a moderate probability of belonging to this class. This is especially true for slopes facing away from the sun. In areas classified with moderate probability, the most likely classes always appear to be spectrally similar.

7. DISCUSSION AND CONCLUSIONS

Overall, the classification in Figure 2 compares favorably with the classified field data. We did not collect the field data ourselves and cannot be certain of its lineage. We believe it provides a good comparison, but should not be considered the “true” classification. In places where the MPM estimate differs from the field data classification, the inconsistencies can be explained by examining the original image data and considering the terrain context. For example, hardwood forests are typically found in riparian areas where large areas of the surface are covered by gravel bars and water, both of which are spectrally very different from tree canopies. Such variation is not accounted for in the field data classification but is apparent in the MPM estimates as noisy configurations with a greater degree of uncertainty. This implies that the results can be improved by increasing the number of classes to account for other types of surface cover, but also shows that the scale of classification appropriate for the spectral resolution of the remote sensor may not correspond to the scale at which the classes are defined and/or desired. In addition, the results may be limited by only having developed three canopy functions.

Forests on aspects facing away from the sun have been classified with more uncertainty, suggesting that the lower radiance from those slopes reduces the ability to discriminate classes in those regions. Space limitations prevent us from presenting posterior probability images for each class. The ability to model the differences between the spectral reflectances of the classes on the basis of BRFs implies that it may be possible to supervise classification without traditional training sites based on field data. Supervising a classification in this manner may require distinct canopy functions for each class. However, this approach to supervision would be limited by the need for a corresponding DEM.

Other areas requiring further research include the need to assess the sensitivity of the posterior probability estimates to various model specifications and parameter values. Specifically, we expect that more detailed MRFs, such as Refs. 18,20 would yield more realistic estimates than Potts, but this has yet to be confirmed for this application. In addition, our experience in applying these models indicates that strong noise suppression using the Potts model leads to excessive homogeneity, while the Descombes *et al.* model reduces noise without producing undesirable homogeneity, however further work will be required to confirm this result.

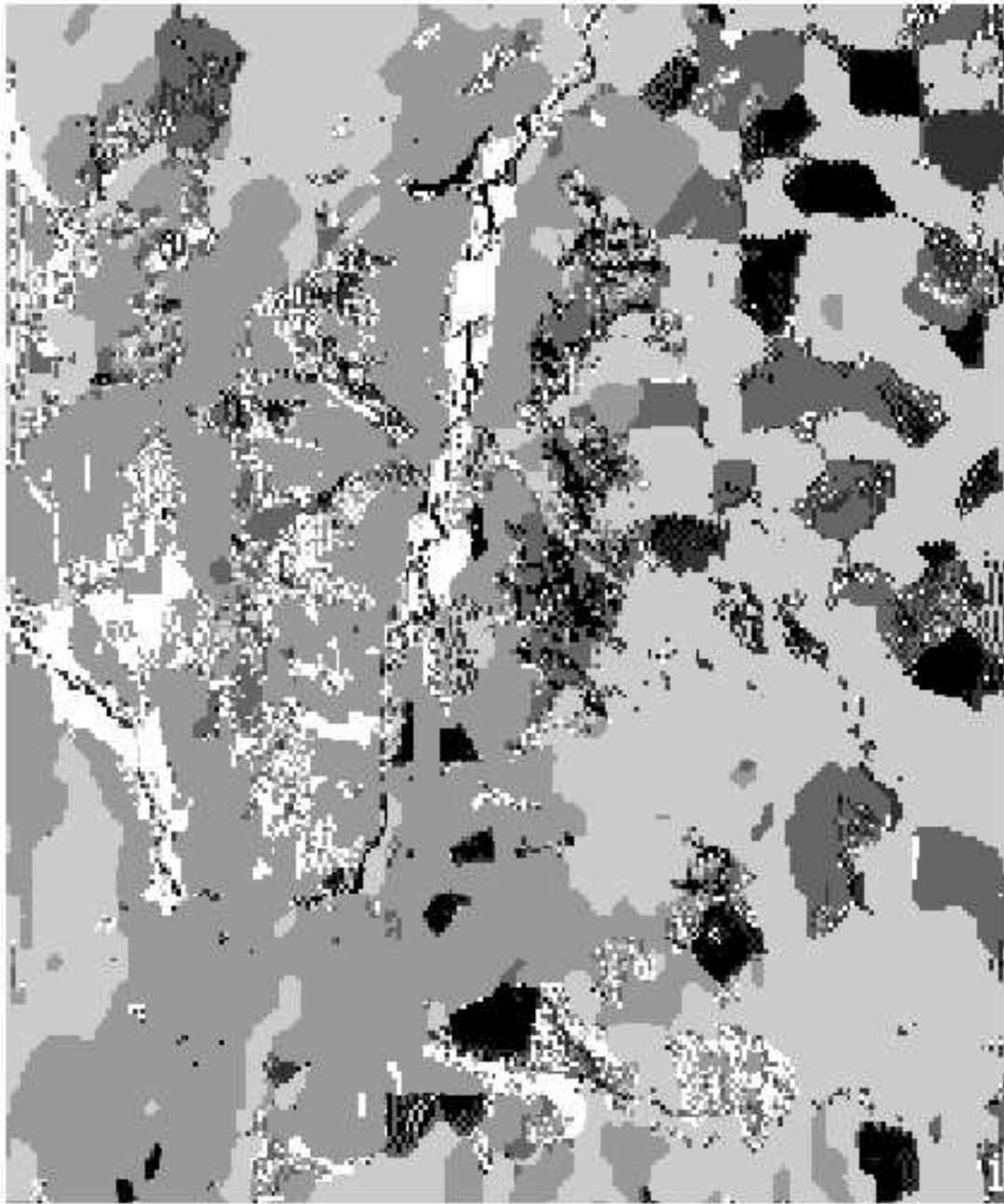


Figure 2. The MPM estimate of the “true” classification using the Descombes *et al.* prior. Many of the larger areas of homogeneous forest are correctly estimated, while many areas of noisy patterns correspond to rock outcrops.

The research presented in this paper demonstrates the flexibility of the Bayesian approach to image classification. We have integrated physically based models of the remote sensing process and the spatial interaction of the classes to classify Landsat TM data of forested landscapes. Not only have we “corrected for topography”, by modelling the anisotropic reflectance at the pixel scale for various forest canopy types, but also we use the information provided by the directional reflectance to, in part, discriminate between the classes. By adopting this approach, we are able to take advantage of the flexibility of a radiosity model to calculate BRFs for a variety of forest canopy structures.

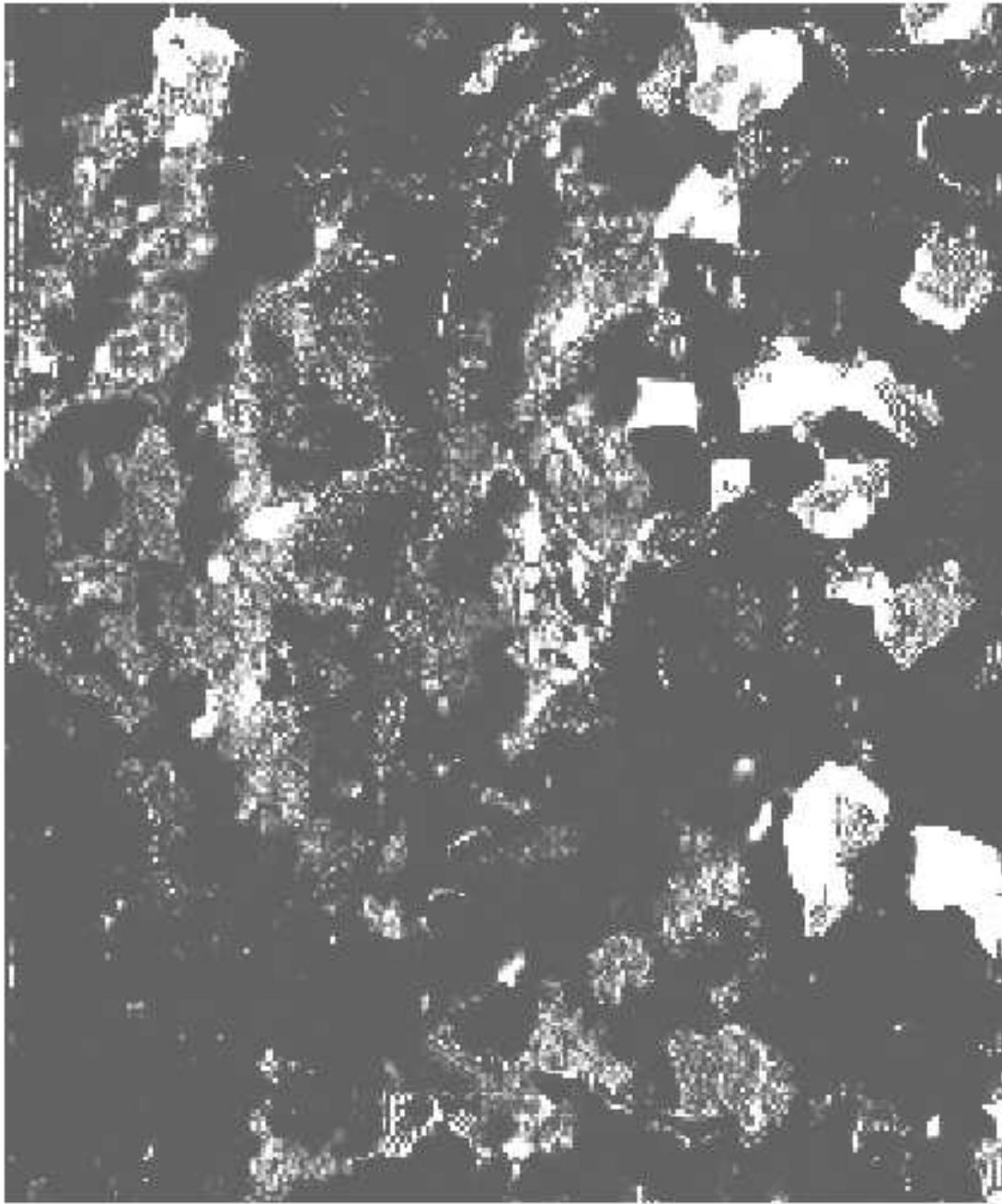


Figure 3. The marginal posterior probabilities that each pixel belongs to the stem exclusion class estimated using the Descombes *et al.* spatial prior. Brighter pixels indicate higher probabilities. Many slopes facing away from the sun have moderate probabilities, as do areas of rock outcrop.

8. ACKNOWLEDGMENTS

We are very grateful for the variety of individual and institutional support we have received from Chad Oliver, Jim McCarter, Milton Smith, Robin Weeks, Alan Gillespie, Jim McCaa, Steve Cothorn and the USDA Forest Service, Pacific Northwest Research Station. Julian Besag acknowledges support from the National Science Foundation and from the National Research Center for Statistics and the Environment.

REFERENCES

1. E. Malaret, L. A. Bartolucci, D. F. Lozano, P. E. Anuta, and C. D. McGillem, "Landsat-4 and Landsat-5 Thematic Mapper data quality analysis," *Photogrammetric Engineering and Remote Sensing* **51**, pp. 1407–1416, 1985.
2. M. J. Duggin, "Factors limiting the discrimination and quantification of terrestrial features using remotely sensed radiance," *International Journal of Remote Sensing* **6**, pp. 3–27, 1985.
3. J. R. Schott, *Remote Sensing, the Image Chain Approach.*, Oxford University Press, 1997.
4. Y. J. Kaufman, "The atmospheric effect on the separability of field classes measured from satellites," *Remote Sensing of Environment* **18**, pp. 21–34, 1985.
5. P. N. Slater, *Remote Sensing, Optics and Optical Systems*, Addison-Wesley Publishing Company, 1980.
6. C. B. Schaaf, X. Li, and A. H. Strahler, "Topographic effects on bidirectional and hemispherical reflectances calculated with a geometric-optical canopy model," *IEEE Transactions on Geoscience and Remote Sensing* **32**, pp. 1186–1193, November 1994.
7. D. W. Deering, E. M. Middleton, and T. F. Eck, "Reflectance anisotropy for a spruce-hemlock forest canopy," *Remote Sensing of Environment* **47**, pp. 242–260, 1994.
8. K. J. Ranson, J. R. Irons, and D. L. Williams, "Multispectral bidirectional reflectance of northern forest canopies with the advanced solid-state array spectroradiometer (ASAS)," *Remote Sensing of Environment* **47**, pp. 276–289, 1994.
9. A. A. Abuelgasim and A. H. Strahler, "Modeling bidirectional radiance measurements collected by the advanced solid-state array spectroradiometer (ASAS) over Oregon transect conifer forests," *Remote Sensing of Environment* **41**, pp. 261–275, 1994.
10. K. T. Kriebel, "Measured spectral bidirectional reflectance properties of four vegetated surfaces," *Applied Optics* **17**, pp. 253–259, 1978.
11. P. R. J. North, "Three dimensional forest light interaction model using a Monte Carlo method," *IEEE Transactions on Geoscience and Remote Sensing* **34**, pp. 946–956, July 1996.
12. E. C. B. de Colstoun, C. L. Walthall, A. T. Ciallla, E. R. Vermote, and R. N. Halthore, "Variability of BRDF with land cover type for the west central HAPEX-Sahel super site," International Geoscience and Remote Sensing Symposium, 1996.
13. J. L. Moffett and J. Besag, "Spatial and probabilistic classification of forest structures using Landsat TM data," Second International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, (Fort Collins, Colorado), 1996.
14. E. J. Green, A. F. M. Smith, and W. E. Strawderman, "Construction of thematic maps from satellite imagery," pp. 75–104, Bayesian Statistics 5: Proceedings of the 4th Valencia International Meeting, 1994.
15. J. D. Wilson, P. J. Green, and I. S. Weir, "A Bayesian analysis of remotely sensed data using a hierarchical model," Tech. Rep., Statistics group, Department of Mathematics, University of Bristol, 1994.
16. J. E. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society, Series B* **36**(2), pp. 192–236, 1974.
17. J. E. Besag, "Statistical analysis of dirty pictures," *Journal of the Royal Statistical Society, Series B* **48**, pp. 259–302, 1986.
18. X. Descombes, J.-F. Mangin, E. Pechersky, and M. Sigelle, "Fine structures preserving Markov model for image processing," pp. 349–356, Proceedings of the 9th Scandinavian Conference on Image Analysis, 1995.
19. S. Byers, "Extensions and applications of a fine structures preserving Markov model," Tech. Rep., Department of Statistics, University of Washington, 1997.
20. H. Tjelmeland and J. Besag, "Markov random fields with higher order interactions," *Scandinavian Journal of Statistics*, 1995. To appear.
21. J. E. Besag, J. York, and A. Mollié, "Bayesian image restoration with two applications in spatial statistics," *Annals of the Institute of Statistical Mathematics*. **43**(1), pp. 1–59, 1991.
22. J. E. Besag, P. J. Green, D. M. Higdon, and K. Mengersen, "Bayesian computation and stochastic systems (with discussion)," *Statistical Science* **10**, pp. 3–66, 1995.
23. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images.," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6**, 1984.

24. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *Journal of Chemistry and Physics* **21**, pp. 1087–91, 1953.
25. W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika* **57**, pp. 97–109, 1970.
26. P. L. Odell and A. H. Feiveson, "A numerical procedure to generate a sample covariance matrix," *Journal of the American Statistical Association* **61**, pp. 199–203, 1966.
27. C. D. Oliver and B. C. Larson, *Forest Stand Dynamics*, McGraw-Hill, 1990.
28. J. A. Richards, *Remote Sensing Digital Image Analysis, an Introduction*, Springer-Verlag, 2nd ed., 1993.
29. W.-H. Li, R. Weeks, and A. R. Gillespie, "Multiple scattering in the remote sensing of natural surfaces," pp. I 205–215, Eleventh Thematic Conference and Workshops on Applied Geologic Remote Sensing, 1996.
30. K. I. Itten and P. Meyer, "Geometric and radiometric correction of TM data of mountainous forested areas," *IEEE Transactions on Geoscience and Remote Sensing* **31**, pp. 764–770, July 1993.
31. D. S. Kimes and J. A. Kirchner, "Modeling the effects of various radiant transfers in mountainous terrain on sensor response," *IEEE Transactions on Geoscience and Remote Sensing* **GE-19**, pp. 100–108, April 1981.
32. R. A. Schowengerdt, C. Archwamety, and R. C. Wrigley, "Landsat Thematic Mapper image-derived MTF," *Photogrammetric Engineering and Remote Sensing* **51**, pp. 1395–1406, September 1985.
33. C. D. McGillem, P. E. Anuta, and E. Malaret, "Spatial resolution estimation of Landsat-4 Thematic Mapper data," *Proc. Landsat-4 Science Characterization Early Results Symposium NASA Conference Pub. 2355*, pp. 527–535, Landsat-4 Science Characterization Early Results, 1985.