

A simulated Linear Mixture Model to Improve Classification Accuracy of Satellite Data Utilizing Degradation of Atmospheric Effect

W.M Elmahboub
Mathematics Department, School of Science, Hampton University
Hampton, Virginia 23668, USA

ABSTRACT

Researchers in remote sensing have attempted to increase the accuracy of land cover information extracted from remotely sensed imagery. Factors that influence the supervised and unsupervised classification accuracy are the presence of atmospheric effect and mixed pixel information. A linear mixture simulated model experiment is generated to simulate real world data with known end member spectral sets and class cover proportions (CCP). The CCP were initially generated by a random number generator and normalized to make the sum of the class proportions equal to 1.0 using MATLAB program. Random noise was intentionally added to pixel values using different combinations of noise levels to simulate a real world data set. The atmospheric scattering error is computed for each pixel value for three generated images with SPOT data. Accuracy can either be classified or misclassified. Results portrayed great improvement in classified accuracy, for example, in image 1, misclassified pixels due to atmospheric noise is 41 %. Subsequent to the degradation of atmospheric effect, the misclassified pixels were reduced to 4 %. We can conclude that accuracy of classification can be improved by degradation of atmospheric noise.

Keywords: Simulation and modeling, remote sensing data.

1. INTRODUCTION

The value of the land cover classification depends on its accuracy. Low classification accuracy of remote sensing data remains the area of growing concern in scientific community. Therefore researchers in remote sensing have attempted to increase the accuracy of the land cover information extracted from remotely sensed imagery. Factors that influence the classification accuracy are the presence of atmospheric effect and mixed pixel information [1]. Natural aerosols such as dust particles and water droplet are the main causes of atmospheric noise in satellite imagery [4]. The atmospheric scattered noise error varies with zenith angle, particle size distribution, and scattering coefficient and increases for zenith angle exceeds 60° [3]. The atmospheric error varies also with land cover type. The error that is distributed within the whole image will lead to misclassification of pixels. A linearly modeled simulated image with atmospheric noise and mixed pixel information (class cover proportion) representing a real world data set using SPOT data was generated, however, the focus will be on correcting the atmospheric scattered noise to improve the image classification accuracy.

The most important requirements for generation of the linear model are estimation of end member spectra (pure pixel information) and class cover proportion (CCP). In general, the end member spectra (EMS) can be obtained by carefully selecting the pure pixel values from a perfectly homogeneous area of the imagery. Multiple member spectra collected from several areas of the imagery should be optimized into a single representative EMS, since collected multiple spectra include spectral variation due to atmospheric effects and imperfect calibration [2]. For the estimation of the EMS and cover class proportions the least squares method is used. Constraining methods, including Lagrangian-Least Square-Cover Class Proportions (L-LS-CCP), Weighted-Least Squares-Cover Class Proportions (W-LS-CCP), and Quadratic programming constraining method-cover class proportions (Q-CCP) are used to make the sum of CCP equal to 1.0 [2]. Random and atmospheric noises are added to simulate real world data sets. The accuracy of classification can be evaluated by comparing classified or misclassified points of the corrected image (CI) and the simulated noisy image (NI). In section 1.1, 1.2, and 1.3, we discuss the principles of the linear mixture model, normalized least squares method, and the linear mixture model noise respectively.

1.1 Principles of the Linear Mixture Model

According to the linear mixture model, we can write the following:

$$X_i = \sum M_{ij} f_j + e_i \quad (1)$$

Where X_i represents the pixel value, M_{ij} is the EMS, f_j is cover class proportions, e_i is a noise term. M_{ij} is independent of f_j . In matrix notation, this is:

$$X = Mf + e = \mu_1 f_1 + \mu_2 f_2 + \dots + \mu_c f_c + e \quad (2)$$

The vector μ_i denotes the columns of M [2]. For least squares estimation of cover class proportion, we assume that all errors are confined to E' , and then equation (2) of the linear mixture model can be modified in the following:

$$X' = \hat{M}F + E' \quad (3)$$

Now the goal of least squares is to optimize the $n \times c$ EMS matrix \hat{M} so that the term $\|X' - MF'\|^2$ is minimized. Using the least squares method, the solution can be decided such that:

$$\hat{M}_{LS} = X'(F')^+ = X'(F')^T [F'(F')]^{-1} \quad (4)$$

$(F')^+$ represents a pseudo inverse matrix of F' [2]. The least squares optimization will be used in the data generation experiment in section 2.

1.2 Normalized Least Squares Method

MATLAB normal random generator is used to estimate the CCP. If an estimated vector \hat{F}_{LS} included negative elements, it will be set to zero. For example, if the vector of the estimated CCP \hat{F}_{LS} is [0.4 -0.05 0.7 -0.06], then the negative proportions -0.05 and -0.06 will first be set to zero and the new \hat{F}_{LS} vector will be represented by [0.4 0 0.7 0]. Secondly, each proportion will be multiplied by [1 / (sum of element)] i.e., [1/(0.4+0+0.7+0)] to yield $\hat{F}_{NLS} = [4/11 \ 0 \ 7/11 \ 0]$ [2].

1.3 Linear Mixture Model Noise

The pixel values of the simulated image can be calculated based on the linear mixture, equation (2). Atmospheric noise is added to the model to simulate the real world data set. The atmospheric noise is derived using the following equations:

$$\xi_n = \sum_{i=0}^n \frac{(-1)^i x^{2i}}{2^{2i} (i!)^2} \quad (5)$$

$$Q_{sca} = \frac{8}{3} x^4 \left[\frac{m^2 - 1}{m^2 + 2} \right]^2 \quad (6)$$

$$C_{sca} = \int_0^{2\pi} \int_0^{\pi} Q_{sca}(\theta) \sin \phi \cos \phi d\phi \quad (7)$$

$$n(r) = C \left(\frac{r}{r_{min}} \right)^{-4.5} \quad (8)$$

$$\mu_{sca} = \frac{\prod_{(-1.5)}^{\pi} * C * \left(\frac{1}{r_{min}} \right)^{-4.5} [r_1^{-1.5} Q_{sca1} + \dots + r_n^{-1.5} Q_{sca n}] * \exp(Z/H) * \sec \theta}{(-1.5)} \quad (9)$$

$$\sigma_{sca} = \frac{(1 + A) \times \mu_{sca} \times \kappa_p \times \gamma \times \eta \times Z}{((1 + A) \times K_{sca} \times Z \times \gamma \times \eta + 1)} \quad (10)$$

$$\kappa_p = L_{min} + ((L_{max} - L_{min}) / Q_{sca1max}) \quad (11)$$

$$\gamma = \exp(K_{sca} \times Z \times \left(\frac{1}{\cos(\theta)} + 1 \right)) \quad (12)$$

$$\eta = \left(\frac{\cos(\theta) + 1}{\cos(\theta)} \right) \quad (13)$$

Where Equation (5) represents Bessel function that is needed to derive the scattering efficiency per single particle; Equation (6)

represents the particle scattering efficiency; Equation (7) represents the scattering coefficient of the medium surrounding the particle; Equation (8) represents the size distribution; Equation (9) represents the scattering coefficient per unit volume; and Equation (10) represents the scattering noise per pixel. $r_{min} \leq r \leq r_{max}$, r is particle size, $r_{min} = 0.1 \mu m$, $r_{max} = 10 \mu m$, and C is the unknown constant representing the highest concentration of particles in a unit volume. Z is altitude, x is the parameter size H is the scale height, m is the complex index of refraction and $H = 0.8$ in a clear sky conditions. $Q_{sca1}, Q_{sca2}, \dots, Q_{sca n}$ are the scattering efficiencies for different particle sizes (r_1, r_2, \dots, r_n) at each image bandwidth assuming that the particle sizes are arbitrary (0.1- 0.5 μm for band 1; 0.1- 0.6 μm for band 2; 0.1 - 0.7 μm for band 3). In equation (6), σ_{sca} represents an error assuming that a fraction of noise of the scattered radiation is assumed to be transmitted to the sensor and can be converted back from radiances to digital numbers using Equation (11) [4]. A is the pixel area, κ_p is the pixel value converted to radiances, θ is the zenith angle, L_{min} is the minimum pixel value, L_{max} is the maximum pixel value, and $Q_{sca1max}$ is 255 within the image [5]. The noise error for each pixel varies with the zenith angle, size distribution, scattering coefficient, and pixel value. Using a sequence of iterations in addition to employing randomly collected training sets, we can determine the unknown constant C in Equation (9). To remove the noise, subtraction of the error is essential. Subtraction technique relies on the following expression:

$$Y_{ij} = X_{ij} - \sigma_{sca} \quad (10)$$

Where X_{ij} represents the input pixel value at line i (row) and sample j (column). The error σ_{sca} represents the noise per pixel as in Equation (10). If deriving the constant C and removing the noise led to reduction in the error matrix (confusion matrix of the classification result) elements that are located off the major diagonal (misclassified pixels), then the overall accuracy represented by the major diagonal elements would exhibit improvement. The final step is the sequence of classifications and evaluations using independent training sets of known ground truth locations, error matrices, and KHAT statistics [4]. In section 2, we discuss the experiment of the linear mixture model data generation.

2. DATA GENERATION EXPERIMENTS

The generation of the experimental data sets requires creating end member's spectra and the CCP values according to equation (2). Each known EMS data set represents real world pixel values, for example, combinations of grass, roads, buildings, deep water, and shallow water were manually extracted from actual SPOT (band 1, band 2, band 3) as shown in figures 3, 4, and 5. The EMS was optimized as shown in equation (4). MATLAB random number generator initially generated the simulated CCP values for each class. The different combinations of CCP, L-LS-CCP, W-LS-CCP, and Q-CCP are used as constraining methods to normalize the CCP and make the sum of three proportions equal to 1. The Q-CCP proved to be the best method. Figures 1 and 2 indicate that this method

produced better results because of a lower RMSE that does not change with sample size. Random noise was intentionally added to pixel values and CCP by different combinations of noise levels to simulate a real world data set using MATLAB program.

The atmospheric scattering error that is derived in Equation (10) has been computed and added for each pixel value. The output image can be displayed and classified using IRDAS IMAGINE (commercial remote sensing software). The simulated image is also classified following the correction of atmospheric noise.

3. RESULTS

The Linear Mixture Model is tested using statistical tests to determine the reliability of the model. The first test indicates that the model is appropriate at 99 % confidence level. The second test depicts that the regression model is significant at 90 % confidence level. The post-processing was performed as the final step to test the results by comparing misclassified points of the corrected image (CI) and the noisy image (NI). Higher misclassified points indicate low accuracy. The accuracy assessments and results of the linear mixture model are presented in Tables 1, 2, and 3. In Table 1, the total reference points are 157200, total misclassification percent error of CI is 4 %, and total misclassification percent error of NI is 41 %. Total error due to atmospheric noise is 37 % for deep and shallow water. In Table 2, the total points are 36000, total misclassification percent error of CI is 21 %, and total misclassification percent error of NI is 51 %. Total error due to atmospheric noise for grass and trees is 30 %. In Table 3, the total points are 36000, total misclassification percent error of CI is 7 %, and total misclassification percent error of NI is 47 %. Total error due to atmospheric noise for road and buildings is 40 %. Therefore the atmospheric error is significant, which resulted in low classification accuracy that led to higher misclassified pixels. Image 1, Image 2, and Image 3 are integrated into single image. The classification for the integrated image was performed and the result is shown in Table 4.

4. DISCUSSIONS and CONCLUSION

The main purpose for adopting atmospheric correction in the linear mixture model is to investigate its impact on the accuracy of classification. After performing all the necessary experiments, Table 3 shows great improvement in classification accuracy. For image 1, the CI shows an error of 1.26% of misclassified pixels (deep water) at the top of the image compared to an error of 17 % of the NI. There is an error of 0.21 % of the CI, compared to an error of 13 % of the NI for misclassified pixels of shallow water. There is also an error of 2 % of CI compared to an error of 11 % of NI misclassified pixels of deep water.

The classification results performed with higher accuracy. For image 2 there is 11% error of CI misclassified pixels compared to 21% error of the NI misclassified pixels of grass at the upper part of the image. There is also 6 % error of CI compared to 17 % error of NI for misclassified pixels of trees, and 4 % error of CI compared to 13 % error of NI for misclassified pixels of grass on the lower part of the image. For image 3. There is 4 % error of CI compared to 15 % error of NI misclassified pixels of roads on the upper part of the image. There is 1.25% error of CI compared to 19 % error of NI misclassified pixels of buildings, and 1.34 % error of CI compared to a larger 13 % error of NI

for roads on the lower part of the image. The total error due to atmospheric noise of image 3 is 40 % and there are 7 % total random error, due to computations, truncations, and mixed pixel information.

These simulation-testing results from images 1, 2, and 3 were investigated for the same hypothesis; we can conclude that the linear mixture model experiments show significant reduction in the misclassified pixels when comparing CI with NI; therefore there is significant improvement in classification accuracy when the atmospheric error is removed including all the simulated images.

5. REFERENCES

- [1] G. M., Foody, D.P., Cox, "Subpixel Land Cover Composition Estimation Using a Linear Mixture Model and Fuzzy Membership Functions", **International Journal of Remote Sensing**, Vol. 15, No. 3, 1994, pp 240-271.
- [2] J.J., Settle, N. A., Drake, "Linear Mixing and the Estimation of Ground Cover Proportions", **International Journal of Remote Sensing**, Vol. 14, No. 6, 1993, pp. 230-265.
- [3] S. M., Singh, "Lowest Order Correction for Solar Zenith Angle to Global Vegetation Index (GVI) data", **International Journal of Remote Sensing**, Vol. P, No. 10 and 11, 1988, pp 100-125.
- [4] W.M. Elmahboub, "An Integrated Atmospheric Correction and Classification Accuracy", Ph.D. Dissertation, University of Wisconsin – Madison, Madison WI, 2000, pp. 250.
- [5] W.M. Elmahboub, F. Scarpace, B. Smith, "An Integrated Methodology to Improve Classification Accuracy of Remote Sensing Data", **IEEE International Geosciences and Remote Sensing Symposium Proceedings**, Vol. IV, 2003, pp 2161-2163.
- [6] W.M. Elmahboub and E. Yankey, "Spectral Analysis for Mars Surface Minerals using Hubble Telescope Digital", **SIP 2005**, pp. 420-422.

6. FUTURE RECOMMENDATIONS

The simulated linear mixture model shows promising results for improving classification accuracy through degradation of atmospheric scattering noise of satellite data. The accuracy can be significant for a large scene with zenith angle exceeding 60 ° where the noise caused by aerosol scattering varies significantly across the scene.

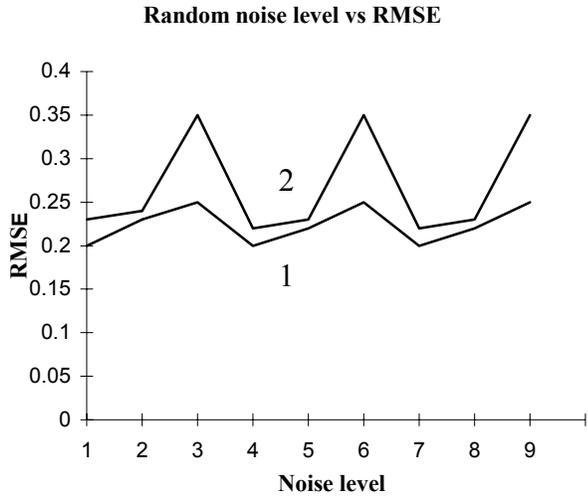


Figure 1 RMSE of CCP estimation and constraining methods. Series 1 represents QCCP; series 2 represents L, W-LS-CCP.

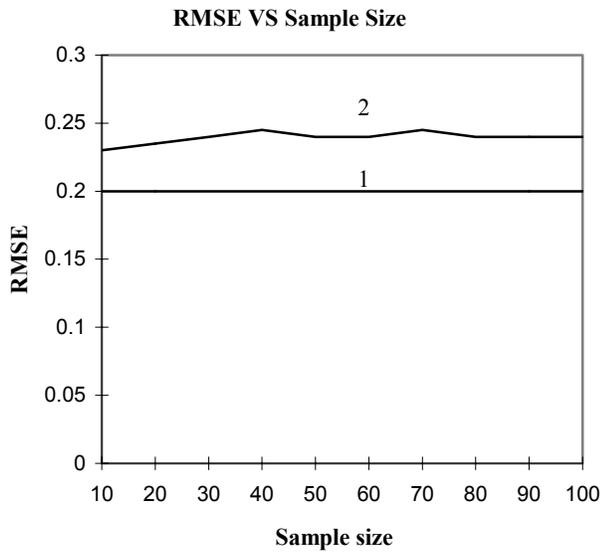


Figure 2 RMSE of CCP estimation and constraining methods by changing sample sizes.

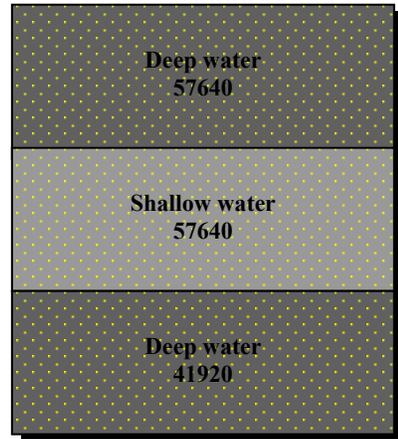


Figure 3 Simulation of Image 1

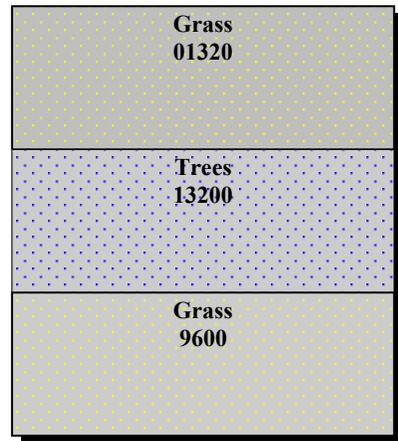


Figure 4 Simulation of Image 2

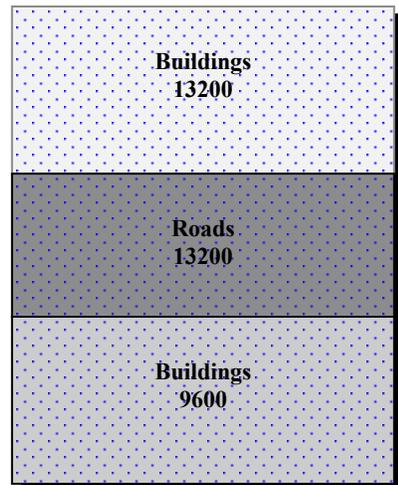


Figure 5 Simulation of Image 3

Deep water 57640	Grass 13200	Roads 13200
Shallow water 57640	Trees 13200	Buildings 13200
Deep water 41920	Grass 9600	Roads 9600

Figure 6 The integrated simulated image

Class	Ref. points	Mis. Class. Pixels Of CI	Mis. Class. Pixels Of NI	% Error Of CI	% Error Of NI
Roads	13200	1572	5363	4	15
Buildings	13200	450	6960	1.25	19
Roads	9600	484	4560	1.34	13
Total	36000	-	-	7	47

re 6.6

Table 1. Comparison between misclassified points of CI and NI for image 1.

Class	Ref. points	Mis- Class. Pixels of CI	Mis- Class. Pixels of NI	% Error of CI	% Error of NI
Deep water	57640	1977	26365	1.26	17
Shallow water	57640	330	20400	0.21	13
Deep water	41920	2898	19934	2	11
Total	157200	-	-	4	41

Table 4 Accuracy assessment for the integrated image

Class type	Total points	Total points of CI	Total points of NI	Acc Of CI	Acc of NI
Unknown	0	8466	104379	-	-
Deep water	99560	98639	61854	43	27
Shallow water	57640	57310	37240	25	16
Grass	22800	20240	5231	89	2
Trees	13200	11051	7427	5	3
Roads	22800	20744	5039	9	2
Buildings	13200	12750	6240	6	3
Total %	229200	220734	124821	96	53

Table 2. Comparison between misclassified points of CI and NI for image 2.

Class	Ref. points	Mis. Class. Pixels of CI	Mis. Class. Pixels of NI	% Error Of CI	% Error of NI
Grass	13200	3860	7193	11	21
Trees	13200	2149	5773	6	17
Grass	9600	1300	4369	4	13
Total	36000	-	-	21	51

Table 3. Comparison between misclassified points of CI and NI for image 2.