Vegetation Classification using Seasonal Variations of σ^o

D. G. Long, G. Watt, and P. J. Hardin
Microwave Earth Remote Sensing Laboratory, Brigham Young University
Provo, Utah 84602 USA

ABSTRACT

Scatterometers were originally designed and deployed based on their proven ability to measure near-surface winds over the ocean. However, they are also proving useful in global and regional studies of vegetation and soil moisture. In this paper we examine the C-band and Ku-band radar signatures of vegetation over South America using data from the ERS-1/2 and NSCAT scatterometers. We compare the seasonal responses of various types of vegetation based on both Matthew's classification and the University of Maryland AVHRR-based classification and use the seasonal response in a simple vegetation classification experiment. A time-series of enhanced resolution σ^o images are generated for both sensors. The classifier is used to classify the vegetation coverage of each pixel in the image frame into broad classes of vegetation type. Considering the accuracy and resolution limitations of the reference vegetation maps, the classifications results exhibit a high degree of accuracy and consistency with the primary confusion observed between related vegetation classes of similar vegetation canopy density.

Keywords: vegetation classification, scatterometry, radar scattering, tropical vegetation

1. INTRODUCTION

Both the ERS-1/2[1] and the NSCAT[2] scatterometers were originally designed for measuring near-surface winds. However, these instruments have also been found useful in studies over land,[3,4] particularly forest areas.[5–7] It has been shown that during the summer, the variation in the backscatter coefficient over these land areas is dependent on soil and vegetation moisture, while during the winter, soil freezing and snow cover drastically changes the response of σ^o.[3,4,6] Scatterometers have been previously used in vegetation studies over the Amazon at Ku-band.[7]

One reason scatterometers are useful in studying vegetation is because they are insensitive to cloud cover. They also obtain global coverage over a reasonably short time and therefore are useful in global studies of seasonal changes over vegetated areas. This paper considers a simple large-scale vegetation classification experiment in South America based on the seasonal scatterometer response of vegetation at multiple frequencies. First, the seasonal response of key vegetation types are examined. Then, the classification approach is described. Finally, the classification results are presented.

2. THE SIR ALGORITHM

Scatterometers make measurements of the normalized radar backscatter coefficient (σ^o) at 25 km (NSCAT) or 50 km (ERS-1/2) resolution. Over the ocean these measurements are used to estimate the near-surface wind. σ^o measurements are also obtained over land and ice regions. Over vegetated regions σ^o exhibits a near-linear relationship with incidence angle over the incidence angle range of $22° - 55°$.

$$10 \log_{10} \sigma^o = \mathcal{A} + \mathcal{B}(\theta - 40°) \tag{1}$$

where $\mathcal{A}$ is the incidence angle normalized σ^o at 40° incidence angle and $\mathcal{B}$ is the incidence angle dependence of σ^o (i.e., the slope of σ^o versus incidence angle). In this paper the $\mathcal{A}$ and $\mathcal{B}$ values are estimated using the Scatterometer Image Reconstruction (SIR) resolution enhancement algorithm.[8]

Images of $\mathcal{A}$ and $\mathcal{B}$ are generated from ERS-1/2 scatterometer mode and NSCAT data with the aid of SIR. Fourteen days of data are processed to produce each image. For NSCAT, the SIR algorithm with filtering (SIRF) is used, resulting in an estimated effective resolution of 10 km. The effective resolution of the ERS-1/2 data is estimated to be 25-30 km. Sample NSCAT V-polarization and ERS-2 $\mathcal{A}$ images are presented in Fig. 1.

Correspondence: Email: long@ee.byu.edu; Telephone: 801-378-4383; Fax: 801-378-6586

Part of the SPIE Conference on Earth Observing Systems IV • Denver, Colorado • July 1999
SPIE Vol. 3750 • 0277-786X/99/$10.00

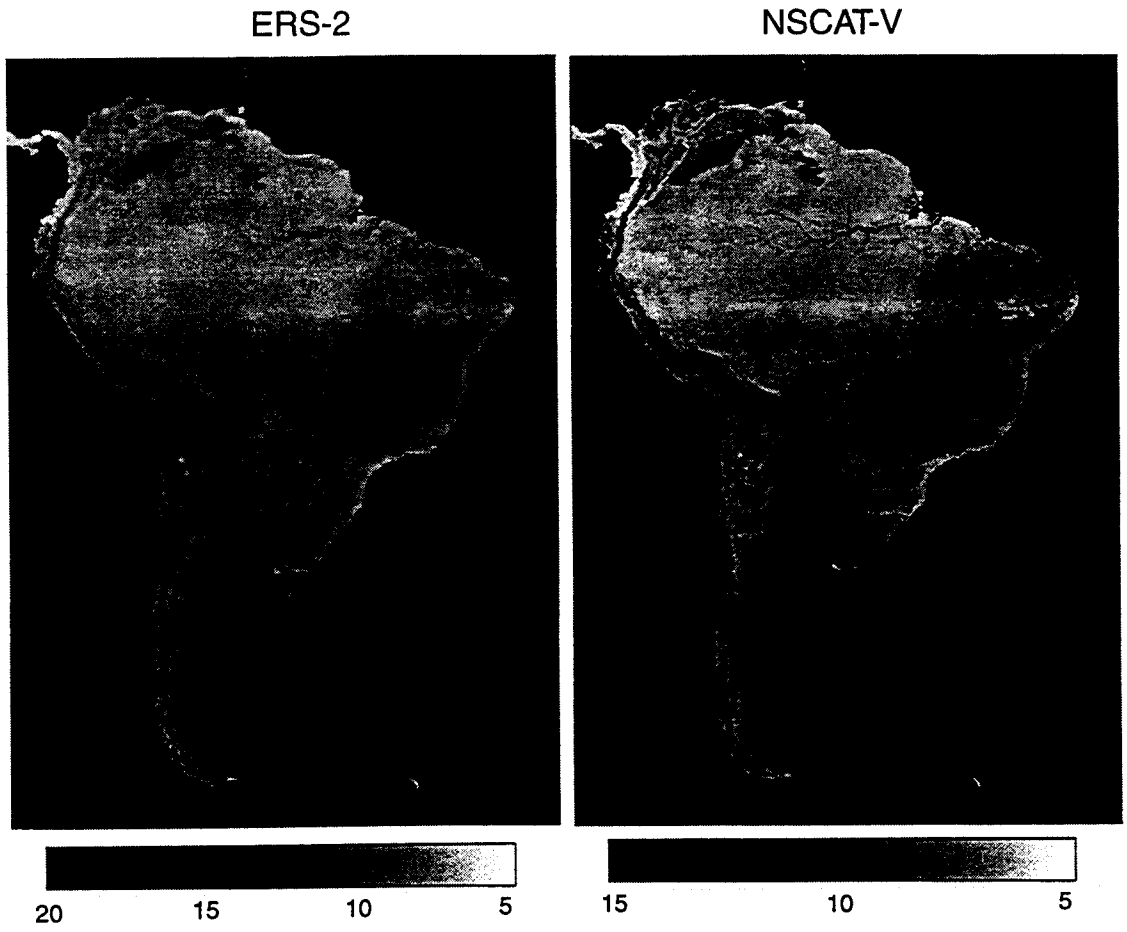

Figure 1. Sample ERS-2 and NSCAT $\mathcal{A}$ images over South America.

3. SEASONAL VARIATION IN σ^O FOR KEY VEGETATION TYPES

To study the normalized radar backscatter coefficient, σ^o, over different vegetation regions, six vegetation types from South America were chosen. These vegetation types, or classes, are chosen from two main vegetation maps. One is the Matthews Global Vegetation, Land Use, and Seasonal Albedo[9] data set from the National Oceanic and Atmospheric Administration (NOAA) and the other is a vegetation map derived from AVHRR data created by the Geography department at the University of Maryland.[10,11] Each map defines a set of vegetation types and provides a spatial map of each type. Complicating matters, the maps are at different resolutions.

The Matthews vegetation map has 32 separate categories for vegetation throughout the world, while the AVHRR-derived map has only thirteen vegetation categories. To simplify intercomparison, a simple set of compatible vegetation categories was developed. In combing the vegetation descriptions care was taken to preserve similar vegetation characteristics over the vegetation types found in South America. Six major vegetation categories were defined. The resulting reduced category maps are shown in Fig. 2 with the corresponding vegetation classes listed in Table 1. Note that some vegetation types are not included in the combined classes. Pixels identified as these vegetation types are not considered in the following analysis.

Using a sequence of SIR images of South America, average values of $\mathcal{A}$ and $\mathcal{B}$ at a given time are computed for each vegetation class using all the pixels identified as that pixel class based on the vegetation map. These average values are plotted versus time for at both C and Ku bands and $\mathcal{B}$ values at Ku-band in Figs. 3 and 4. Due to noise in the C-band $\mathcal{B}$ value images, this data is not shown. The response of ERS-1/2 is plotted over the same time frame as NSCAT so that they can more easily be compared. Because the Matthews and AVHRR vegetation maps

Table 1. Vegetation classes over South America

Veg Class	Description	% Pixels Matthews	% Pixels AVHRR
1	Tropical Rainforests/Evergreen Broadleaf forests	40.3	42.7
2	Evergreen Rainforests	10.2	5.8
3	Deciduous broadleaf forests/mixed forests	9.6	3.2
4	Tropical/subtropical drought deciduous forests	2.6	7.0
5	Xeromorphic shrubland/shrublands/closed bushlands	27.7	25.4
6	Open shrublands/grassland with shrub cover	9.6	16.0
	Xeromorphic forest/ woodlands/wooded grasslands		
	grassland with 10-40% woody tree cover		
	Meadows/short, medium, tall grasslands		

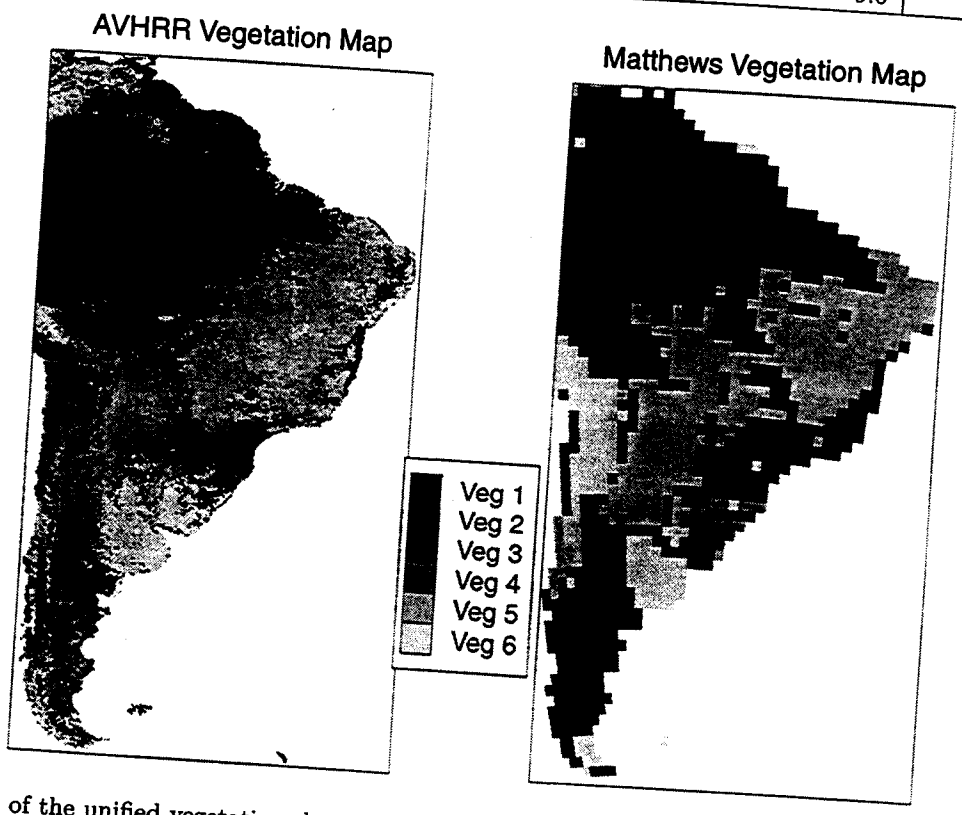

Figure 2. Maps of the unified vegetation classes from the Matthews and the AVHRR vegetation maps over South America.

give different vegetation types for some areas, separate analyses are made based on each map. Figure 5 shows the year-to-year seasonal changes in the vegetation response for ERS-1/2. Note that in this simple analysis the data has not been stratified by latitude or topography, though this can be expected to improve the classification attempted later.

We note that different σ^o responses are seen for each vegetation class and that the signatures vary from C-band to Ku-band. The dominant signature is the difference in the mean σ^o for each vegetation class; however, some seasonal variations are apparent. While tropical rainforests exhibit very little seasonal variation, other vegetation classes exhibit up to two dB of variation over an annual. The seasonal differences are much larger for the $\mathcal{A}$-values than for the $\mathcal{B}$-values. Based on Fig. 5 the average σ^o response over each vegetation type in South America exhibits relatively little interannual variability.

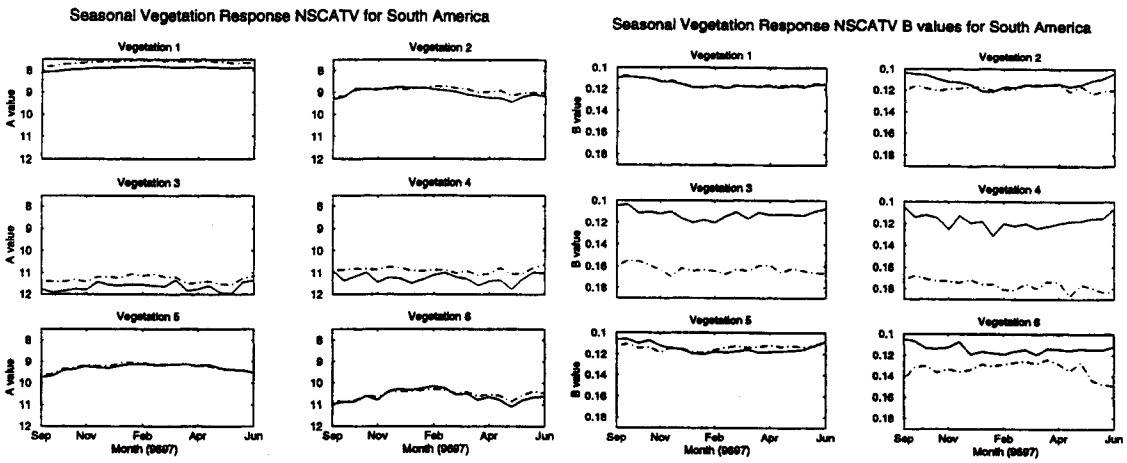

Figure 3. Seasonal response of (left) $\mathcal{A}$-values and (right) $\mathcal{B}$-values over each vegetation region of South America for NSCAT V-pol. The solid line is the σ^o response computed based on the AVHRR map while the dashed line is based on the Matthews map.

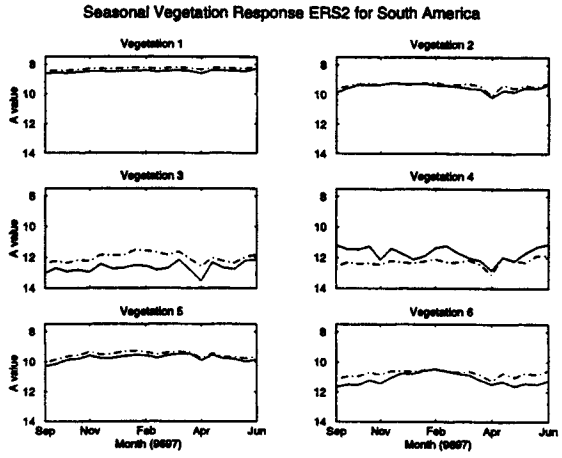

Figure 4. Seasonal response of $\mathcal{A}$-values over each vegetation region of South America for ERS-2. The solid line is the σ^o response computed based on the AVHRR map while the dashed line is based on the Matthews map.

We again note that this analysis assumes a similar seasonal response over the entire area covered by a given vegetation class and neglects spatial variations in the σ^o response of a given vegetation class. This is seen as a limitation of this simple study.

4. USING THE SEASONAL RESPONSE FOR VEGETATION CLASSIFICATION

In this section the observed σ^o seasonal response from the SIR $\mathcal{A}$ images of NSCAT-V and ERS-1/2 scatterometer data are used as the training set for an algorithm to classify vegetation based on the scatterometer observations. For each data set and vegetation map we form a different training matrix, X, where the columns of X correspond to the seasonal response of each vegetation class discussed above. For example, the vegetation response over each class for the Matthews vegetation map using NSCAT-V data forms the columns of a training matrix, while the NSCAT-V vegetation response for the AVHRR-derived map forms the columns of another training matrix. For each training matrix, principle component analysis is used.

To study the minimum number of eigenvectors required in the classification, the log of the eigenvalues from highest

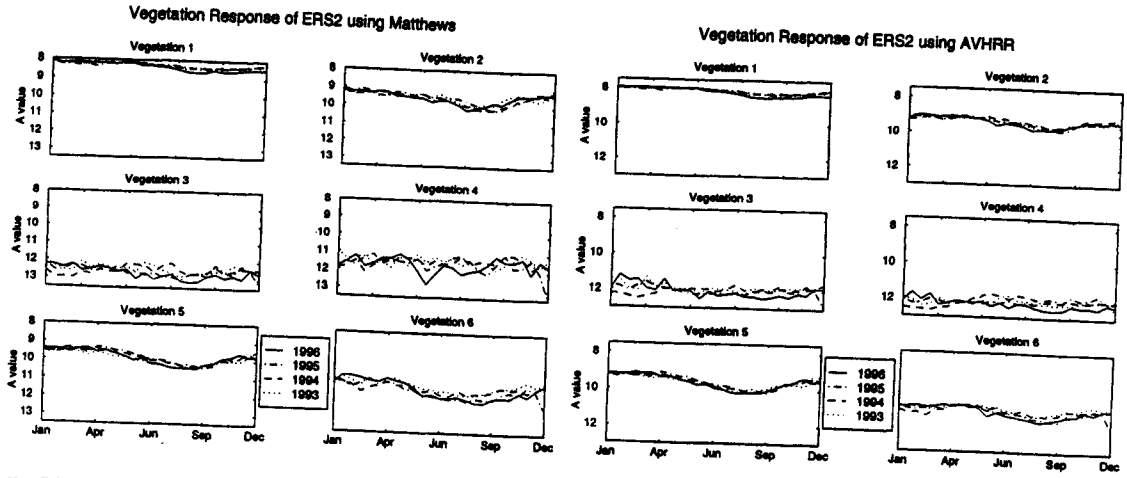

Figure 5. Yearly seasonal response of $\mathcal{A}$-values over each vegetation region of South America for ERS-1/2 over the years 1993-1996.

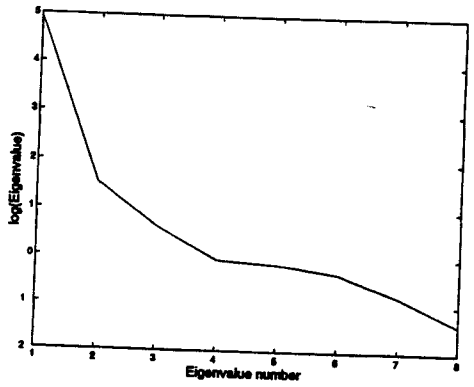

Figure 6. The log of the principal component values of X. Both maps produce similar eigenvalue plots.

to lowest are plotted and are shown in Figure 6 for NSCAT-V using the Matthews vegetation map. Figure 7 shows the corresponding eigenvectors for the eigenvalues. Notice that the first eigenvector, although it has some variations, is mainly a DC response. The other vectors represent different seasonal variations in the vegetation responses. While the top two eigenvalues account for the most variation in the vegetation responses all of the eigenvectors are retained in this study.

The eigenvectors produced from PCA are seen as a basis set that contains the most prominent seasonal responses (i.e., the principal component responses) of each vegetation region. The eigenvalues can be viewed as the variance along each basis set. In other words, the axes in the new vector space lie along the most prominent vegetation features of the various vegetation responses. Our motivation in doing this is to orthogonalize the most prominent vegetation characteristics between vegetation classes so as to make it easier to distinguish between one vegetation class and another. From here X can be expressed as a linear combination of eigenvectors or, $X = \sum_i a_i Q$. Where Q contains the eigenvectors. In matrix notation the equation becomes $X = AQ_1$ where the rows of A are a_i. The a_i's are calculated for each of the chosen vegetation regions above. These a_i's are used as the centroids in the next step.

As part of the classification, a decision metric or distance measure is required. Two different distance criteria are used for comparison. One is the usual L^2-norm distance and the other is a maximum likelihood metric. In both cases the seasonal response of each pixel in the image is projected onto the given basis set, $r_i = Q^{\dagger} x_i$, where x_i is the seasonal response of the i_{th} pixel. r_i is the response of the pixel in the new basis set.

The first distance criteria, the minimum L^2-Norm distance, is defined in the usual way, i.e. $d = ||x - y||$, where

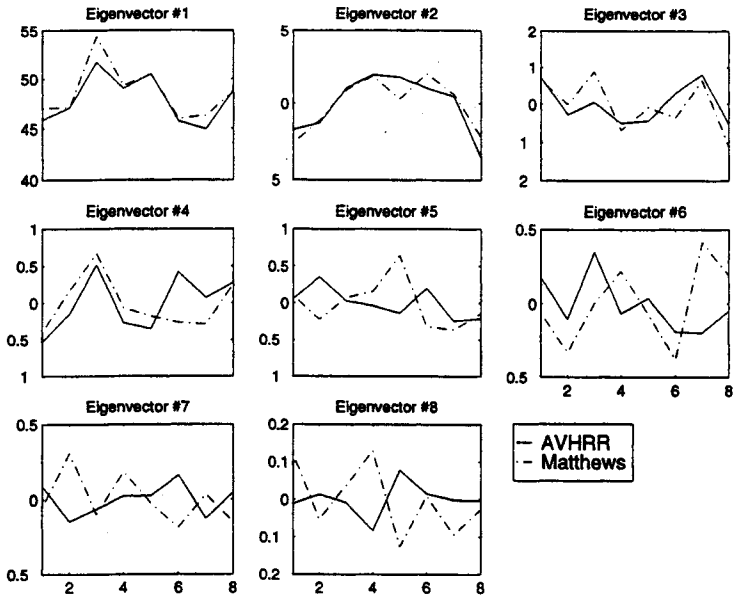

Figure 7. The eigenvectors of X. The solid line is computed based on the AVHRR map while the dashed line is based on the Matthews map.

x and y are two vectors in the space. In our case, for each pixel in the algorithm calculates the distance from r_i, the vegetation response of the pixel, to each vegetation centroid in the new orthogonal space. The centroid that is the closest to the pixel response is chosen as the vegetation class of that pixel, i.e.

$$Im(i) = \min_j ||r_i - a_j|| \tag{2}$$

where j is the number of vegetation regions, i is the pixel number in the image, and $Im(i)$ is the final classification map.

Although the L^2 norm classifier is easy to implement, it does not take into account the shape or statistics of the data. An alternate distance classifier which considers the prior statistics is also used. The pixel value is chosen according to

$$Im(i) = \min_j [(r_i - a_j)' R_j^{-1} (r_i - a_j) + \ln |R_j|] \tag{3}$$

where R_j is the sample covariance of the projection of the seasonal response of each vegetation region, r_i is the pixel response, and a_j is the centroid of each vegetation region.

This classifier is derived from maximizing the multivariate Gaussian distribution. The first term of Eq. (3) is the Mahalanobis distance and is often used as a distance metric. In our analysis, the entire maximum likelihood expression is used.

5. RESULTS

The classification algorithm results are illustrated in Figures 8, 9, 10, and 11. The values in the image correspond to the region numbers and vegetation types of the above study regions. Tables 2-9 show the percentage of pixels properly classified over each class.

To quantitatively evaluate the classification schemes, accuracy and inaccuracy assessments are done on the classifications. In Ref. 12 a new accuracy and inaccuracy assessment is proposed that better assess the data than the usual kappa statistic. The new accuracy assessment is derived from the Kullback-Leibler information. The usual class-averaged and overall accuracies are also given.

Table 2. Vegetation classification percentages using NSCAT V pol data over South America based on using the Matthew's training set and the L^2 norm.

Vegetation Class	L^2 Norm Confusion Matrix					
	Veg 1	Veg 2	Veg 3	Veg 4	Veg 5	Veg 6
1	79.2	6.2	1.7	1.7	5.7	3.4
2	24.8	33.3	0.8	1.6	26.0	13.2
3	3.9	4.3	59.1	13.5	4.2	12.3
4	1.0	2.5	35.7	27.2	8.9	22.7
5	16.1	25.6	2.3	2.6	41.5	11.2
6	13.9	11.6	28.2	7.9	11.7	24.0

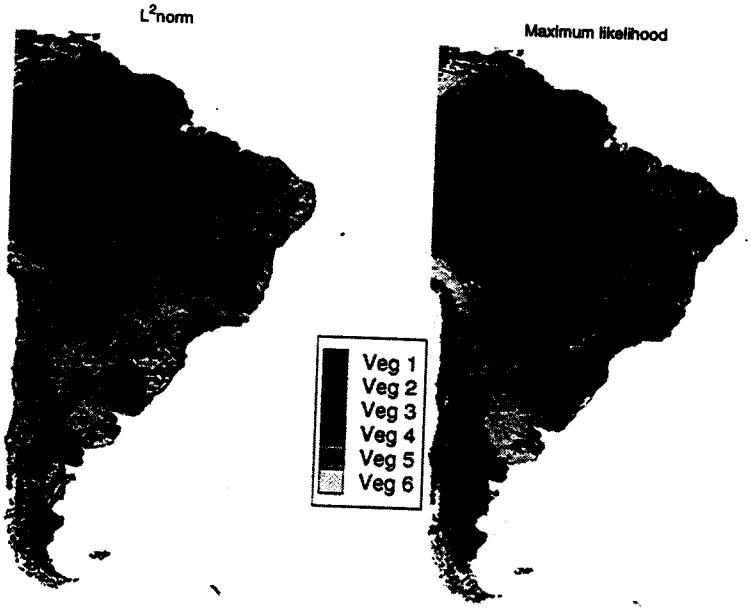

Figure 8. NSCAT vegetation classification results over South America based on the Matthews vegetation map training set. On the left is the classification using the L^2-norm. On the right is the classification using the maximum likelihood (ML) classifier.

Given an error matrix X, the class-averaged coefficient, C, and overall accuracy coefficient, A, are computed using:

$$C(X) = \frac{1}{r}\sum_{i=1}^{r}\frac{x_{ii}}{n_i} \tag{4}$$

$$A(X) = \frac{1}{N}\sum_{i=1}^{r}x_{ii} \tag{5}$$

where r is the number of classification categories, n_i is the number of samples for each category, and N is the total number of samples. These two accuracy assessments are shown for South America in Table 10.

From looking at Table 10 it is apparent that using C-band data gives somewhat better accuracy than Ku-band data. The accuracy is also slightly higher using the training set generated using the Matthews vegetation map than the AVHRR-derived vegetation map. Classification using the maximum likelihood classifier is more accurate than the regular L^2 Norm distance.

Table 3. Vegetation classification percentages using NSCAT V pol data over South America based on using the Matthew's training set and the maximum likelihood (ML) norm.

Vegetation Class	ML Confusion Matrix					
	Veg 1	Veg 2	Veg 3	Veg 4	Veg 5	Veg 6
1	78.2	6.3	2.0	2.2	6.3	3.0
2	18.6	63.9	0.8	5.4	5.4	5.4
3	4.0	8.3	66.4	5.9	3.5	9.4
4	1.8	11.7	3.9	76.5	1.0	3.1
5	17.5	30.2	3.2	4.6	40.5	3.3
6	10.2	17.2	9.6	14.0	7.3	38.9

Table 4. Vegetation classification percentages using NSCAT V pol data over South America based on using the AVHRR training set and the L^2 norm.

Vegetation Class	L^2 Norm Confusion Matrix					
	Veg 1	Veg 2	Veg 3	Veg 4	Veg 5	Veg 6
1	82.0	9.0	1.0	0.5	4.8	1.7
2	25.6	37.0	3.1	3.1	21.7	9.1
3	2.3	8.1	59.7	6.7	10.6	11.7
4	8.0	11.1	50.8	10.1	9.8	9.9
5	8.6	33.4	3.5	2.1	40.7	11.1
6	3.1	9.2	36.3	7.1	23.7	19.1

The third accuracy assessment is computed using:

$$J_p(X) = \prod_{i=1}^{r} \left(\frac{x_{ii} + 1/2}{n_i + 1/2} \right)^{n_i/N} \tag{6}$$

where r is the number of classification categories, n_i is the number of samples from each category, and N is the total number of samples. This accuracy assessment has advantages over the kappa statistic. For an in-depth discussion of the comparisons between the two see Ref. 12. The J_p assessment gives a coefficient between 0 to 1 only, where the coefficient is 0 if and only if all samples are misclassified and 1 indicates total classification. In Table 10 the J_p accuracy assessment coefficients for all the classifications over South America are given. We conclude that the ML distance metric provides the best performance for all cases. ERS-2 data and the Matthew's map results in the best performance. This may be due, in part, to the reduced resolutions of the data and map.

Table 5. Vegetation classification percentages using NSCAT V pol data over South America based on using the AVHRR training set and the maximum likelihood norm.

Vegetation Class	ML Confusion Matrix					
	Veg 1	Veg 2	Veg 3	Veg 4	Veg 5	Veg 6
1	82.9	7.8	1.1	1.3	4.7	1.4
2	27.8	40.6	2.5	4.4	18.9	5.4
3	3.0	10.9	28.8	44.7	4.8	6.9
4	3.7	10.9	11.8	61.7	6.6	4.9
5	14.1	16.3	2.7	5.7	52.9	7.8
6	2.3	10.3	10.7	16.6	20.9	37.6

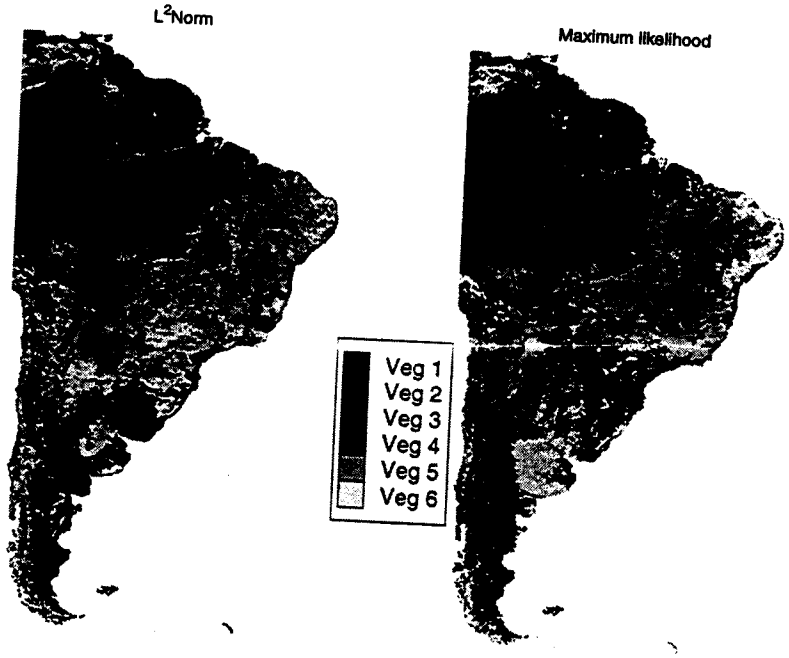

Figure 9. NSCAT Vegetation Classification over South America based on the AVHRR vegetation map training set. On the left is the classification using the L^2-norm. On the right is the classification using the maximum likelihood (ML) classifier.

Table 6. Vegetation classification percentages using ERS-2 data over South America based on using the Matthew's training set and the L^2 norm.

Vegetation Class	L^2 Norm Confusion Matrix					
	Veg 1	Veg 2	Veg 3	Veg 4	Veg 5	Veg 6
1	82.2	6.4	0.7	1.4	4.9	2.1
2	29.6	30.9	0.5	2.4	24.6	11.1
3	4.0	3.4	65.9	8.0	4.8	10.6
4	1.0	3.6	32.8	31.2	11.4	18.1
5	22.0	20.8	1.9	2.7	41.6	9.7
6	13.7	10.5	25.0	9.5	13.7	25.0

6. CONCLUSION

This paper has presented the results of a simple large scale vegetation classification experiment based on the seasonal response of scatterometer σ^o observations. The classification algorithm was applied to σ^o data obtained at two different frequencies using two different vegetation maps to train it. The results were compared with the vegetation maps. It was shown that the algorithm performed well over large areas containing homogeneous vegetation types. Misclassification most commonly occurs with classes of similar vegetation types. Overall, the results were consistent with previous work. The 1999 launch of SeaWinds will extend the Ku-band data set begun with Seasat in 1978 and continued with NSCAT in 1996-1997. ERS-1/ C-band data continues from 1993 to the present.

Acknowledgements

ERS-1/2 and NSCAT data for this study was obtained from the PO.DAAC at the Jet Propulsion Laboratory.

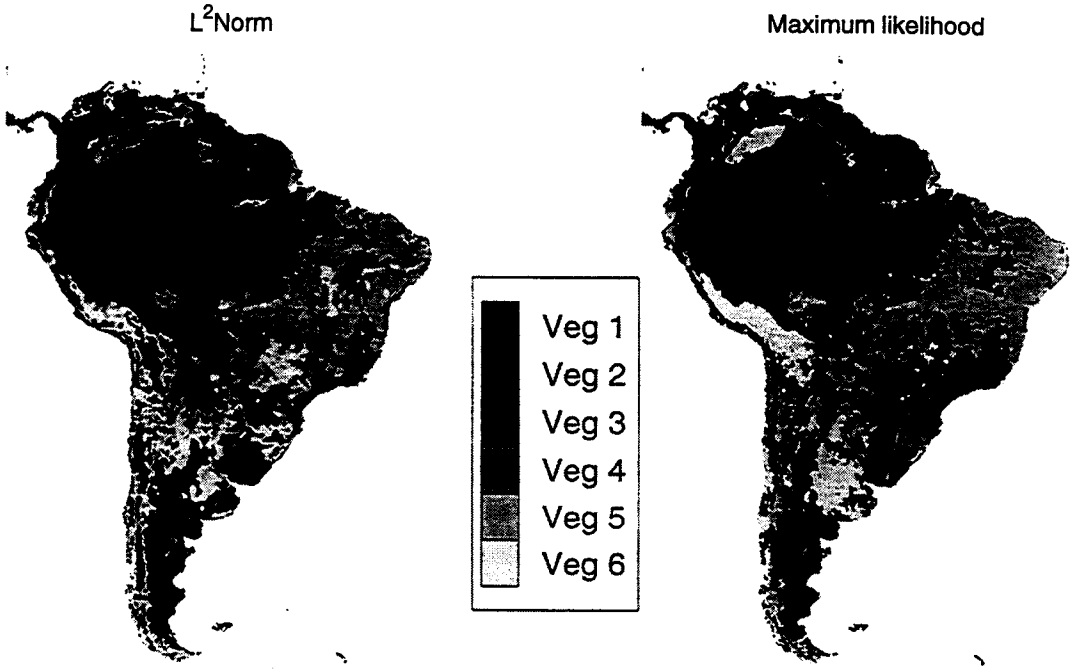

Figure 10. ERS-2 Vegetation Classification over South America based on the Matthews vegetation map training set. On the left is the classification using the L^2-norm. On the right is the classification using the maximum likelihood (ML) classifier.

Table 7. Vegetation classification percentages using ERS-2 data over South America based on using the Matthew's training set and the maximum likelihood norm.

Vegetation Class	ML Confusion Matrix					
	Veg 1	Veg 2	Veg 3	Veg 4	Veg 5	Veg 6
1	82.1	4.6	1.5	1.8	5.8	2.0
2	24.6	58.4	0.8	4.8	7.2	3.3
3	9.7	4.5	64.4	2.2	3.8	12.0
4	1.5	11.4	0.8	81.1	0.2	3.2
5	23.8	18.2	2.8	2.6	47.0	4.3
6	16.7	6.5	12.0	10.6	8.5	43.1

REFERENCES

1. E. Attema, "The Active Microwave Instrument On-board the ERS-1 Satellite," *Proc. IEEE*, vol. 79, no. 6, pp. 791-799, 1991.
2. F.M. Naderi, M.H. Freilich, and D.G. Long, "Spaceborne Radar Measurement of Wind Velocity Over the Ocean – An Overview of the NSCAT Scatterometer System," *Proc. IEEE*, vol. 79, no. 6, pp. 850–866, 1991.
3. K. Boehnke and V. Wismann, "Thawing Processes During Siberian Spring Observed by ERS Scatterometer and SAR," *Proc. Int. Geosci. Rem. Sens. Sym.*, pp. 1826-1828, 4-8 August 1997.
4. J.T. Pulliainen, T. Manninen, and M.T. Hallikainen, "Application of ERS-1 Wind Scatterometer Data to Soil Frost and Soil Moisture Monitoring in the Boreal Forest Zone," *IEEE Trans. Geosci. Rem. Sens.*, vol. 36, no. 3, pp. 849-863, 1998.

Table 8. Vegetation classification percentages using ERS-2 data over South America based on using the AVHRR training set and the L^2 norm.

Vegetation Class	L^2 Norm Confusion Matrix					
	Veg 1	Veg 2	Veg 3	Veg 4	Veg 5	Veg 6
1	83.9	7.3	0.4	0.4	4.9	2.3
2	31.4	28.5	2.2	1.4	21.9	14.2
3	3.0	5.9	15.2	50.7	9.9	14.3
4	2.8	4.5	12.3	56.2	6.3	17.4
5	17.8	21.9	1.9	1.5	40.3	16.0
6	9.6	9.0	15.4	19.3	20.3	24.9

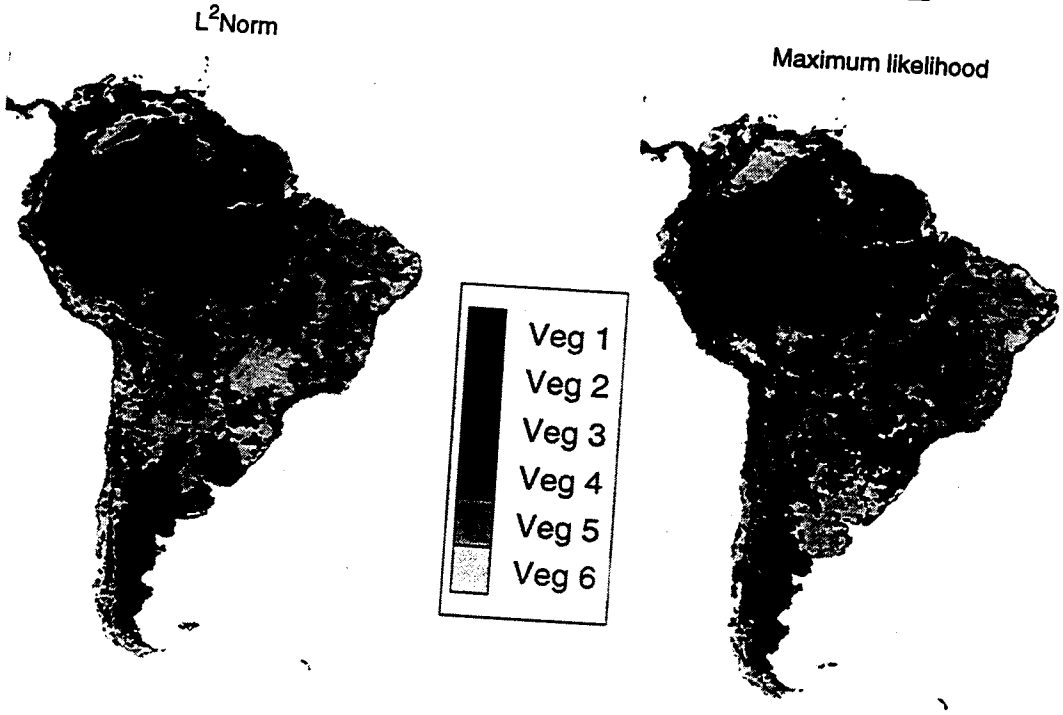

Figure 11. ERS-2 Vegetation Classification over South America based on the AVHRR vegetation map training set. On the left is the classification using the L^2-norm. On the right is the classification using the maximum likelihood (ML) classifier.

5. J. Pulliainen, N. Walkder, T. Manninen, M. Hallikainen, and J. Grandel, "Land Applications of ERS-1 Wind Scatterometer in Boreal Forest Zone," *Proc. Int. Geosci. Rem. Sens. Sym.*, pp. 1826-1828, 4-8 August 1997.
6. C.J. Wilson and D.G. Long, "Analysis of the Canadian Boreal Forest using Enhanced Resolution ERS-1 Scatterometer Imagery," *Proc. Int. Geosci. Rem. Sens. Sym.*, pp. 218-220, 27-31 May 1996.
7. D.G. Long and P.J. Hardin, "Vegetation Studies of the Amazon Basin Using Enhanced Resolution Seasat Scatterometer Data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 2, pp. 449-460, 1994.
8. D. Long, P. Hardin, and P. Whiting, "Resolution Enhancement of Spaceborne Scatterometer Data," *IEEE Trans. Geosci. Remote Sens.*, vol. 31, pp. 700-715, 1993.
9. E. Matthews, "Global Vegetation and Land Use: New High Resolution Data Bases for Climate Studies," *J. Climatology Appl. Meteor.*, vol. 22, pp. 474-487, 1983.

Table 9. Vegetation classification percentages using ERS-2 data over South America based on using the AVH... training set and the maximum likelihood norm.

Vegetation Class	ML Confusion Matrix					
	Veg 1	Veg 2	Veg 3	Veg 4	Veg 5	Veg 6
1	84.9	5.8	0.6	1.2	5.3	1.5
2	36.9	29.9	1.4	2.6	21.4	7.3
3	7.9	5.1	38.9	37.3	3.4	6.7
4	7.6	6.2	10.0	66.9	6.1	2.6
5	21.8	12.8	3.2	3.1	50.8	7.7
6	10.0	9.4	11.8	8.7	21.9	36.7

Table 10. Class-average and overall accuracy coefficients for each training set and distance metric.

		L^2 Norm			ML		
		C(X)	A(X)	$J_p(X)$	C(X)	A(X)	$J_p(X)$
NSCAT	Matthews	.4488	.5645	.5193	.6183	.6235	.5962
	AVHRR	.4175	.5356	.4477	.5115	.6294	.5979
ERS-2	Matthews	.4718	.6050	.5524	.6399	.6721	.6495
	AVHRR	.4180	.5745	.5078	.5174	.6389	.6005

10. T. Purevdorj and R. Tateishi, "The Assessment of Green Vegetation Coverage of Grasslands Using AVHRR Data," *Proc. Int. Geosci. Rem. Sens. Sym.*, pp. 1283-1285, 6-10 July 1998.
11. U. of Maryland web page, *Global Land Cover Mapping Using AVHRR Data*, `http://www.geog.umd.edu/landcover/lc/projects.html`, 1999.
12. R. Nishii and S. Tanaka, "Accuracy and Inaccuracy Assessments in Land-Cover Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 491-497, 1999.