LOCAL-SCALE EVALUATION OF A TECHNIQUE FOR LAND-COVER CLASSIFICATION BASED ON COMPOSITED NDVI DATA

JORDAN BORAK, PAUL FISHER, AND ALAN STRAHLER
DEPARTMENT OF GEOGRAPHY
BOSTON UNIVERSITY
675 COMMONWEALTH AVE.
BOSTON, MA 02215

AARON MOODY
DEPARTMENT OF GEOGRAPHY
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NC 27599

ABSTRACT

Thresholding based on biophysical variables derived from time trajectories of satellite data is a new approach to classifying land cover via remote sensing at coarse resolutions. This approach is attractive because it is much simpler than conventional alternatives. Further, it operates on biophysical variables and thus should be more robust than more data-dependent techniques. The present research evaluates one such technique for performance at the local scale of a Sierra Nevada test site. The input data are composited values of the Normalized Difference Vegetation Index (NDVI), derived from the red and infrared channels of the Advanced Very High Resolution Radiometer (AVHRR) instrument as flown on NOAA polar-orbiting space platforms. They are presented at a nominal 1-km spatial resolution. Associated with these values are radiances in three thermal bands that are used to estimate surface temperature. The classification algorithm, proposed by Running and Nemani (1994), accepts mean growing-season NDVI, mean growing-season near-infrared radiance, NDVI amplitude and surface temperature as input parameters from the composited NDVI and surface temperature data. The units recognized are broad life-form vegetation classes, such as evergreen needleleaf forest, evergreen broadleaf forest, shrubs, etc. They were compared to a ground truth map of the Plumas National Forest, California, made for timber inventory using Thematic Mapper (TM) imagery with extensive ground validation. Classification accuracies were variable, depending on the class and the comparison method. Although some units were recognized with good accuracy, others were not. Our analysis indicates a potential for successful application of thresholding techniques to land-cover classification, but clearly further research into the utilizations of such methods is necessary.

INTRODUCTION

A primary goal of the United States Global Change Research Program (USGCRP) is global-scale modeling of earth-system changes and processes. NASA plays a fundamental role in the USGCRP through its Mission to Planet Earth, a broad initiative in earth observation and modeling of earth-system phenomena that includes the Earth Observing System (EOS). As part of EOS, satellite platforms carrying suites of instruments for observing land, oceans and atmosphere are being constructed and launched in a comprehensive program stretching well beyond the turn of the century. A keystone instrument on the EOS-AM platform, scheduled for launch in 1998, is MODIS, the Moderate Resolution Imaging Spectroradiometer (Salomonson et al., 1989). An instrument in the heritage of NOAA's Advanced Very High Resolution Radiometer (AVHRR), the MODIS will acquire global data on a near-daily basis in 36 spectral bands at nadir spectral resolutions ranging from 250 m to 1 km.

A key dataset that will be of use in the generation of many MODIS products is a land-cover map produced from MODIS data that recognizes broad life-form classes for input

to ecosystem process models (Running et al., 1995). In the MODIS pre-launch era, an important objective is identifying and refining the proper classification technique for these land covers. Running et al. (1993) recently proposed a temporal thresholding classification algorithm applied to time trajectories of composited NDVI data and associated values of surface temperature that has potential application to MODIS data. To explore this method further, we applied it to a forested region in the Sierra Nevada of California. This paper describes our application of this technique and its results, emphasizing per-pixel classification accuracy at the 1-km scale.

STUDY AREA AND DATA SET

The study area for the NDVI thresholding classification was the Plumas National Forest in northeastern California. The Plumas, an area encompassing approximately 7320 square km, lies at the transition between the northern margin of the Sierra Nevada and the Southern edge of the Cascade Range. It is an area of high relief (2247 meters in total elevation range) with vegetation covers characterized by shrub, pine, and oak formations at low elevations, while mixed conifer forests and brush vegetation are encountered at high elevations. Generally, areas of brush and grasslands are distributed throughout the area, along with barren rock outcrops which are predominantly found at the highest elevations.

The classifications were performed on AVHRR-LAC maximum value NDVI biweekly composites compiled for 1990 and 1991 by EROS Data Center. There are 19 composites from March 1990 to December 1990 and 21 composites from January 1991 to December 1991, providing a near-continuous time series for 1990-1991. These data were resampled to 250 m resolution, as this format is similar to what will eventually be produced by the MODIS instrument (the 250 m MODIS spectral bands will be red; 620 - 670 nm and NIR; 841 - 876 nm).

Digital data from all five AVHRR bands and the associated NDVI information were geometrically coregistered to the Lambert Azimuthal Equal Area Projection, calibrated to reflectance, and scaled to one-byte data by EROS Data Center. In order to draw justifiable comparisons to our land-cover reference map of the Plumas, all AVHRR data were transformed to UTM coordinates and registered to the reference map using a nearest-neighbor resampling algorithm. Control points for this resampling were selected on the basis of a lakes and streams data layer, which is registered to the same projection as the original AVHRR data. This data-to-reference registration process contained a root mean square error of 0.902 pixels along the horizontal axis and 0.652 pixels along the vertical axis. No edge effects indicative of registration problems were apparent, and after considering the magnitude of the difference in spatial resolution between the image data and the reference map, registration errors were regarded as negligible.

The reference vegetation map of the Plumas was originally produced from 1990 Landsat TM data (Woodcock, et. al., 1994), and contained the following classes:

1. barren lands/grasslands
2. brush/shrubs
3. hardwood trees (mostly evergreen broadleaf)
4. water
5. conifer trees (evergreen)
6. meadow
7. conifer plantation

Since there were no growth-stage data available for the conifer plantations, the actual class compositions of these features were impossible to designate in the reference map, and were therefore eliminated from our analyses.

1995

ACSM/ASPRS

Annual Convention & Exposition Technical Papers

ACSM 55th Annual Convention
ASPRS 61st Annual Convention

Volume Three: ASPRS

Charlotte, North Carolina
February 27 - March 2, 1995

METHODOLOGY

The method which was used to separate the Plumas into vegetation classes was developed by Running and Nemani (1994). This procedure used an NDVI and surface temperature-based thresholding algorithm (hereafter referred to as the NDVI-Ts algorithm for purposes of brevity) in a six biome classification scheme. Originally, six vegetation classes were defined as follows:

1. Evergreen Needleleaf Forests
2. Evergreen Broadleaf Forests
3. Deciduous Needleleaf Forests (Larix)
4. Deciduous Broadleaf Forests (including shrubs)
5. Broadleaf Annual Vegetation (Crops)
6. Grasses

These were later modified to include water and barren areas.

The NDVI-Ts algorithm used NDVI, NIR radiance, and surface temperature time-series data as input parameters on a per-pixel basis. The NDVI was calculated from calibrated digital radiance values of AVHRR bands 1 and 2. (Eq. 1)

$$\text{NDVI} = (\text{Band 2} - \text{Band 1})/(\text{Band 2} + \text{Band 1}) \quad (1)$$

The NIR radiance was equal to the band 2 calibrated digital radiance value. A split window temperature algorithm was used to calculate surface temperature using data from the brightness temperatures in AVHRR bands 4 and 5 (Price, 1984). (Eq. 2)

$$Ts = [\text{Tb(Band 4) brightness temperature} + 3.33\ (\text{Tb(Band 4)} - \text{Tb(Band 5)})] \quad (2)$$

Initially, the surface temperature was calculated for every pixel in each NDVI composite scene. Each pixel was then classified into one of four broad-based land-cover categories based on both an NDVI threshold and a surface temperature threshold. As proposed by Nemani and Running (1994), the thresholds used were 0.4 and 35^{o} C, respectively:

1. Barren Land, Shrubs and Grasses:
   Observed NDVI < NDVI Threshold of 0.4
   Observed Surface Temp. >= Surface Temp. Threshold of 35^{o} C
2. Broadleaf Crops:
   Observed NDVI >= NDVI Threshold of 0.4
   Observed Surface Temp. >= Surface Temp. Threshold of 35^{o} C
3. Water:
   Observed NDVI < NDVI Threshold of 0.4
   Observed Surface Temp. < Surface Temp. Threshold of 35^{o} C
4. Forests:
   Observed NDVI >= NDVI Threshold of 0.4
   Observed Surface Temp. < Surface Temp. Threshold of 35^{o} C

These broad categories were further divided based on growing season NDVI and NIR statistics. For each pixel, the growing season was defined as the subset of the full time-series where surface temperature exceeded a specified growing season temperature minimum in the corresponding ground resolution cell. The mean NDVI across this growing season ($NDVI_{gs}$) was calculated for each pixel and used to divide the first land cover group into barren, shrub, and grass vegetation classes. Pixels exhibiting high $NDVI_{gs}$ were classified as grass, while pixels that displayed low $NDVI_{gs}$ were

classified as barren land. Mid-range $NDVI_{gs}$ pixels were classified as shrub vegetation.

$NDVI_{amp}$, defined as the maximum NDVI over the time-series minus the minimum NDVI occurring during the growing season, was used to separate the forest class into deciduous and evergreen categories. Each pixel with a low $NDVI_{amp}$ was classified as evergreen, while each with a high $NDVI_{amp}$ was classified as deciduous. Additionally, average NIR radiance during the growing season was employed to further distinguish forest vegetation pixels based on leaf shape. A forested pixel with a high NIR_{gs} radiance was classified as broadleaf, while any other was classified as needleleaf.

As a final step, the output classifications from each of the data sets were compared using a plurality rule. Whichever class was assigned to a particular pixel most often throughout the full time-series was selected as the final classification for that pixel. The only exception to this rule was the water class, where pixels were required to be classified as such in each data set of the time series in order to be labeled as water on the class map. Theoretically, all true water pixels should have been identified as such over each time-series scene. Otherwise, the next-to-most common assignment was accepted as the correct label.

The original parameter values suggested by Nemani and Running did not produce satisfactory classification accuracies, so the values were modified in an attempt to increase the classification accuracies. A comparison of the original and modified parameter settings is presented in table 1.

Table 1: Comparison of Parameter Settings, NDVI-T_s Method

Parameter	Parameter Setting: Original NDVI-T_s Parameters	Parameter Setting: Plumas-derived
NDVI	0.4	0.4
Surface temperature	35^{o} C	37^{o} C
Growing Season Temperature	5^{o} C	5^{o} C
Grass $NDVI_{gs}$	0.25	0.29
Barren $NDVI_{gs}$	0.15	0.15
$NDVI_{amp}$	0.3	0.5
NIR_{gs}	15% (DN=39)	28% (DN=72)

RESULTS

Each thresholding algorithm produced a class map as output. After registering the thresholded class maps of the Plumas to the TM-derived 30 m reference map, we created pixel-by-pixel confusion matrices in order to measure the accuracies of the thresholding methods. The matrices illustrated percentages of correctly classified pixels for each class. The finer resolution of the 30m product also facilitated investigation into sub-pixel attributes of the thresholded class maps.

Two approaches to accuracy assessment were examined in this research. The user's accuracy, a measure of commission error, referred to the percentage of classified pixels that were correctly labeled by the NDVI-Ts algorithm. The producer's accuracy, a measure of omission error, was the percentage of reference map pixels that were correctly classified by the classification procedure (Congalton, 1986).

Additionally, several vegetation classes were not directly comparable to categories on the reference map. Therefore, for purposes of clarity, we collapsed or renamed classes with similar attributes wherever possible as illustrated in tables 2 and 3. We combined the barren and grasses categories in the thresholded maps to correspond to the barren/grasses class of the reference map. We treated the conifer class and evergreen needleleaf forest as equivalent categories, as well as the hardwood and evergreen broadleaf forest classes. Also, because the brush class could contain trees, wherever the thresholded maps categorized pixels as shrubs, evergreen broadleaf or deciduous broadleaf forest, but the reference map classified them as brush, we determined these to be accurate classifications. This decision was based on both the knowledge communicated to us by the developers of the reference map and a degree of ambiguity associated with the definition of the shrub class. The meadow class, having no clear analogue among the thresholded map classes, and constituting a miniscule percentage of the land cover of the study site, was consequently eliminated from the study. Finally, the broadleaf annual and deciduous needleleaf categories were not applicable in this analysis because neither class was found within the study site.

<u>Table 2: Unit Comparison - User's Accuracy</u>
<u>NDVI-Ts Method</u>

Thresholded Cover Class (250 m)	Correct Ground Truth Class (30 m)
Barren, grasses	Barren/grasses
Shrubs	Brush
Evergreen broadleaf	Hardwood, brush
Deciduous broadleaf	Brush
Evergreen needleleaf	Conifer
Water	Water
Deciduous needleleaf	(NA)
Broadleaf annual	(NA)

<u>Table 3: Unit Comparison - Producer's Accuracy</u>
<u>NDVI-Ts Method</u>

Thresholded Cover Class (250 m)	Correct Ground Truth Class (30 m)
Barren, grasses	Barren/grasses
Evergreen broadleaf	Hardwood
Evergreen broadleaf, Deciduous broadleaf, Shrubs	Brush
Evergreen needleleaf	Conifer
Water	Water
Deciduous needleleaf	(NA)
Broadleaf annual	(NA)

The results of the accuracy assessments were summarized in tables 4 and 5. In each table, the first row of results referred to the classifications produced employing the original (straight) set of parameters as proposed by Nemani and Running (1994). With respect to user's accuracies, the classifier performed best on the evergreen broadleaf class. After both the brush and hardwood categories of the reference map were taken into consideration, 39.97% of the pixels classified as evergreen broadleaf forest by the thresholding algorithm were correctly labeled. For producer's accuracies, the brush class performed best with 80.78% of the brush pixels on the reference map correctly classified by the thresholding algorithm. It was clear that the classification algorithm did not produce satisfactory results for the stated purposes. Consequently, some of the threshold parameters were adjusted in an attempt to raise classification accuracies. In addition to these threshold adjustments, the data set was also varied to determine whether the use of protracted or shifted time-series data had significant effects.

For the analysis performed on the 1990 composite data with the modified threshold values, results were as stated in the second row of tables 5 and 6. Evergreen needleleaf represented the NDVI-T_S thresholding class with the highest user's accuracy with 51.75% correct. Examination of producer's accuracies yielded conifer as the most accurate category with 80.73% correctly classified.

As a data set variation, an entire year of composites spanning June, 1990 to June, 1991 was selected, again with the modified parameters. The results of this analysis were as shown in the third row of tables 5 and 6. The class with the highest user's accuracy was evergreen needleleaf with 51.96% correctly classified. Again, the conifer class exhibited the best producer's accuracy with 82.64% correctly classified.

As another variation in the data set, two complete years of biweekly composites spanning the years 1990 to 1991 were selected with results as summarized in the fourth row of tables 5 and 6. The class with the highest user's accuracy was evergreen needleleaf with 51.90% correctly classified. As before, the reference map class with the greatest accuracy was conifer with 79.69% classified correctly.

Table 4: NDVI-Ts Classification User's Accuracies (%)

Sub-Method	Barren, grasses	Shrubs	Evergr. broadleaf	Decid. broadleaf	Evergr. n'leaf	Water
Straight	24.72	15.39	39.97	22.00	0.00	0.01
Modified	22.61	23.27	45.23	20.42	51.75	0.04
Modified--June-June	22.96	25.81	46.26	0.00	51.96	0.91
Modified--Two Years	21.32	25.43	45.87	18.44	51.90	0.02

Table 5: NDVI-Ts Classification Producer's Accuracies (%)

Sub-Method	Barren/grasses	Brush	Hardwood	Conifer	Water
Straight	24.95	80.78	68.65	0.00	0.00
Modified	14.01	16.98	28.25	80.73	0.05
Modified--June-June	10.11	16.79	26.32	82.64	2.64
Modified--Two Years	10.75	19.26	34.36	79.69	0.02

DISCUSSION

Overall, the results of the NDVI-T_S classification method were not satisfactory. Employing the original parameters of the thresholding algorithm, all of the user's accuracies were under 40%. In terms of the producer's accuracies, however, over 80% of the brush pixels, comprising 24% of the reference map, were classified correctly. However, none of the conifer pixels, which account for 46% of the reference map, received the correct classification by using the original set of parameters.

Employing modified parameters, it was possible to increase user's accuracies for four of the six classes, with a decrease in accuracy of less than 10% for the remaining two. Specifically, over 50% of the thresholded evergreen needleleaf class corresponded correctly with the reference map, while over 80% of the conifer reference pixels were accurately classified. Yet, the absolute classification accuracy of all classes remained under 52%, with only 0.04% of the known bodies of water correctly identified by the algorithm.

Sub-canopy and sub-pixel signals were likely responsible for some misclassifications in the NDVI-T_S thresholding algorithm. Certain understory vegetation types exhibit strong seasonal signals that may have contributed confounding information to the overall signals of the dominant canopy types. Similarly, mixed pixels may contain small amounts of vegetation with strong signals that may add confusion to the classification procedure. Although it was possible to analyze the sub-pixel land-cover components of the output classifications as compared to the reference map, we were

unable to determine if there were any ordered confusions produced by the thresholding algorithm. However, it was recognized that the low classification accuracies were at least partially due to characteristics of the reference data.

At the finer resolution of the basemap, landscape heterogeneity was generally much greater than at the coarser resolution of the AVHRR-LAC data. Also, since the reference map was produced as part of a United States Forest Service timber inventory project, it was heavily biased with respect to the presence of conifers. Thus, some pixels labeled as conifer forest in the reference map actually contained very small numbers of conifer trees (Woodcock, et. al., 1994). Both of these reference data attributes lead to the conclusion that employing an unbiased, 1 km reference map would markedly increase classification accuracies.

It was likely that additional class-specific reasons existed for the poor performance of the NDVI-T_S thresholding procedure as proposed by Nemani and Running. The accuracies associated with the evergreen needleleaf/conifer class employing modified parameters were predictably high owing to the relatively large percentage (46%) of this class in the reference data. However, for the water class the algorithm was extremely stringent in that no pixels could be categorized as water unless classified as such in every scene of the time series. Since the signal received at the satellite is affected by spectrally-dependent multiple scattering and spectral mixing within the atmosphere, and by adjacency and directional effects at the surface, the misclassification of a water pixel in at least one of the scenes in a particular data series was a strong possibility. There was also the issue of ephemerality, which became important when time series included periods of drought or other uncharacteristic precipitation regimes. These effects would be magnified when the bodies of water were split between adjacent pixels. Therefore, it is probable that the algorithm can never reliably identify water at a pixel or sub-pixel scale. The mapping of large lakes and coastlines may also be subject to these effects at water/land boundaries.

To further examine the sensitivity of the NDVI-T_S thresholding algorithm, the structure of time-series in the data set were altered. In addition to the original calendar year time-series for 1990, the algorithm was applied to an entire year of composites collected between the summers of 1990 and 1991 and to two full years of composites for 1990 and 1991. The outcome of thresholding the Plumas between the summers of 1990 and 1991 was similar to that of calendar year 1990, the exception being a decrease in the producer's accuracy for classification of the barren/grass category. This may have been due to natural succession of vegetation over the Plumas. For example, a pixel classified as barren/grass in 1990 might have contained a new growth of brush in 1991 which would have produced a different classification.

There appeared to be few strong contrasts between the results obtained using the single calendar year data and the results produced by thresholding the two-year time-series data. Although the deciduous broadleaf class experienced a slight increase in user's accuracy, the water and barren/grasses categories experienced slight decreases and the shrubs class showed a major decrease. With respect to producer's accuracy, the hardwood and brush classes experienced moderate increases, but the performance of the barren/grasses category decreased by a similar percentage. The water class decreased by a large percentage. It is clear that inter-annual pixel variation was sufficient to produce increased classification accuracies for hardwood/deciduous broadleaf vegetation by highlighting differences across the years and yielding larger values for $NDVI_{amp}$. However, changes in surface features such as water level and vegetative succession caused decreases in classification accuracies for some classes over the same time period.

Although the six-biome classification scheme of the NDVI-T_S algorithm was designed to separate differing types of vegetation into distinct classes, this was not possible with respect to the shrubs category. Most of the brush found in the Plumas consists of evergreen broadleaf species that grow amidst the taller, scattered conifer and hardwood