INT. J. REMOTE SENSING, 1989, VOL. 10, NO. 11, 1805–1812

A technique for extrapolating and validating forest cover across large regions

Calibrating AVHRR data with TM data

L. R. IVERSON

Center for Biodiversity, Illinois Natural History Survey and Department of Forestry, University of Illinois, Champaign, Illinois 61820, U.S.A.

E. A. COOK

Center for Biogeographic Information, Illinois Natural History Survey, Champaign, Illinois 61820, U.S.A.

and R. L. GRAHAM

Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, U.S.A.

(Received 31 January 1989; in final form 28 April 1989)

Abstract. An approach to extending high-resolution forest cover information across large regions is presented and validated. Landsat Thematic Mapper (TM) data were classified into forest and nonforest for a portion of Jackson County, Illinois. The classified TM image was then used to determine the relationship between forest cover and the spectral signature of Advanced Very High Resolution Radiometer (AVHRR) pixels covering the same location. Regression analysis was used to develop an empirical relationship between AVHRR spectral signatures and forest cover. The regression equation developed from data from the single county calibration area in southern Illinois was then applied to the entire AVHRR scene, which covered all or parts of ten states, to produce a regional map of forest cover. This map was used to derive estimates of forest cover, within a geographical information system (GIS), for each of the 428 counties located within the boundaries of the original AVHRR scene. The validity of the overall regional map was tested by comparing the AVHRR/TM-derived estimates of county forest cover with independent estimates of county forest cover developed by the U.S. Forest Service (USFS). The overall correlation coefficient of the AVHRR/TM and USFS county forest cover estimates was $r=0{\cdot}89$ ($n=428$ counties). Not surprisingly, some individual states and the areas nearer to the southern Illinois calibration centre had higher correlation coefficients. Absolute estimates of forest cover percentages were also significantly well predicted. With the future inclusion of multiple calibration centres representing a number of physiographic regions, the method shows promise for predicting continental and global estimates of forest cover.

1. Introduction

Advanced Very High Resolution Radiometer (AVHRR) data have been used to monitor gross correlates to continental primary productivity (Goward *et al.* 1985, 1987, Tucker *et al.* 1985, Townshend and Justice 1986), rangeland condition (Sadowski and Westover 1986) and tropical deforestation (Nelson and Holben 1986). The approach used in many of these studies can be referred to as 'top-down' because the AVHRR data were classified first and then related to ground or other ancillary information to identify relationships between AVHRR data and the vegetation topic of interest. This approach is valuable, but it is sometimes difficult to assign ecologically and/or statistically meaningful information to the AVHRR classes because of the mixed nature of the pixels and the difficulty in acquiring ground information or even aerial photographs, over areas encompassing thousands of square kilometres.

Our approach was to use the Landsat Thematic Mapper (TM) data to measure precisely the forest cover and then calibrate the less-precise AVHRR data with the TM data for forest cover estimates over the vast areas monitored by the AVHRR. This approach can be thought of as a 'bottom-up' method. In this letter we describe our three-stage method for extending TM-based values of forest cover to regional estimates (figure 1).

First, a partial TM scene, one for which adequate aerial photographs or ground information is available, is classified into forest and non-forest classes. In the second stage, AVHRR data are overlaid on the classified TM data and statistical relationships are derived between AVHRR spectral data and the amount of forested landscape, as classified in the TM data. In the third stage, the statistical model previously developed is applied to a larger AVHRR scene to generate regional maps of forest cover. For purposes of validation, the regional forest cover estimates are then compared to independent estimates.

In summary, our approach is to use nested scales of imagery to generate quantifiably accurate regional estimates of forest cover which cannot necessarily be

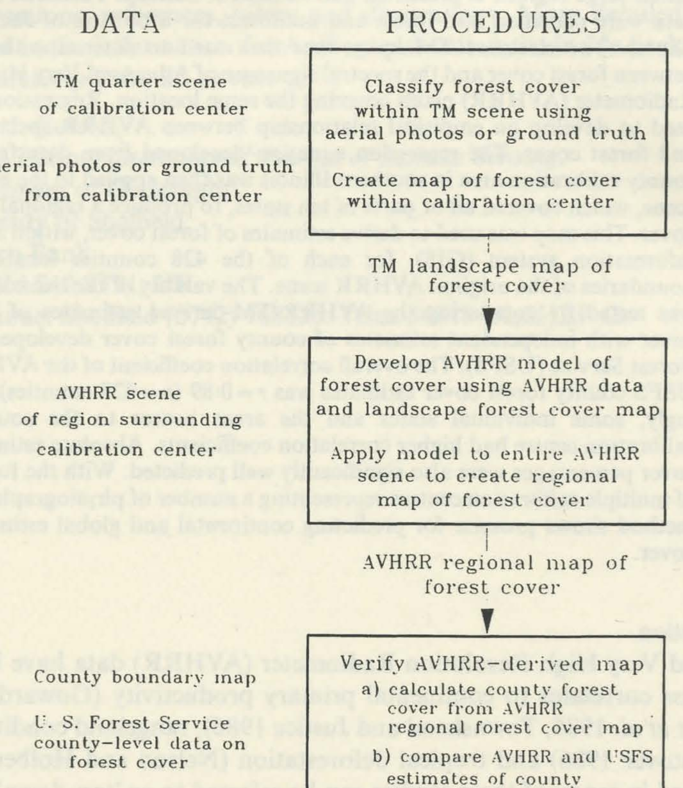

Figure 1. Schematic flow chart of the methodology.

directly measured by a sensor but is functionally related to surface reflectance characteristics of a forest canopy that TM and AVHRR sensors can detect.

2. Methods

The methods described here were developed using AVHRR and TM data on an area centred in southern Illinois, U.S.A. This area was selected because of the availability of satellite sensor and ancillary data on the forests, the fragmented nature and range of forest cover and its familiarity to the authors.

2.1. *Development of TM forest cover map*

TM data were extracted from an 18 July 1984 scene (path 24, row 34) over the south-west quarter of Jackson County, in southern Illinois, and grouped into forest and non-forest classes using unsupervised classification techniques. Verification of class definitions was accomplished using National High Altitude Photography (NHAP) and topographic maps. The binary image that resulted was then available for combining with and calibrating the AVHRR data.

2.2. *Development of AVHRR forest-cover model*

A local-area coverage (LAC) AVHRR data set from the NOAA-10 satellite and which covered 564 175 km^2 centred on Illinois (latitude 34°–44° N, longitude 86°–94° W) for the afternoon of 4 June, 1987, was acquired from the EROS Data Center in Universal Transverse Mercator (UTM) format with a pixel size of 1110 $\times$ 1110 m. The data covered 428 counties from all or parts of ten states. The 4 June date of acquisition corresponded to a time when all forests in the study areas were in full-leaf and the reflectances from row-crop agriculture (the primary non-forest feature of the region) were dominated by bare soil. Therefore, the vast majority of highly-reflecting, chlorophyll-rich land area was forestland on the date of AVHRR data acquisition. The assumption that forests will have the majority of chlorophyll reflectance at the time of AVHRR data acquisition needs to be met for this technique to be successful. One hundred and fifty-four AVHRR pixels corresponding to the Jackson County TM study area were extracted and overlaid in a geographical information system (GIS) such that 1369 TM pixels (37 $\times$ 37, 30 m pixels) defined each AVHRR pixel. The AVHRR pixels were each subset into 1369 subpixel units to allow direct correspondence to the TM data. An observation in the final data file of 30 m subpixels was then composed of (1) a unique AVHRR pixel identifier, (2) the AVHRR spectral band values associated with the subpixel unit (all will be identical within an AVHRR pixel) and (3) the TM forest class (forest or non-forest) associated with the subpixel unit.

A sampling program extracted data from every fourth line and fourth column ($\frac{1}{16}$th sample) to reduce data density; approximately 100 30 m pixels were used from each of 154 AVHRR pixels to derive estimates of forest cover for each AVHRR pixel. The TM-derived forest class information of the subpixel units was then aggregated to estimate the percentage forest cover within each AVHRR pixel. The relationship between AVHRR spectral values and forest cover within the AVHRR pixel was empirically determined using multiple regression analysis in which the TM-derived forest cover estimate was used as the independent variable. Various AVHRR spectral characteristics, including AVHRR bands 1 to 4, the normalized difference vegetation index (NDVI = (NIR − VIS)/(NIR + VIS) where NIR = near-infrared wavelength measurement and VIS = visible wavelength measurement), band ratios and other of interest. This approach is valuable, but it is sometimes difficult to assign ecologically and/or statistically meaningful information to the AVHRR classes because of the mixed nature of the pixels and the difficulty in acquiring ground information or even aerial photographs, over areas encompassing thousands of square kilometres.

Our approach was to use the Landsat Thematic Mapper (TM) data to measure precisely the forest cover and then calibrate the less-precise AVHRR data with the TM data for forest cover estimates over the vast areas monitored by the AVHRR. This approach can be thought of as a 'bottom-up' method. In this letter we describe our three-stage method for extending TM-based values of forest cover to regional estimates (figure 1).

First, a partial TM scene, one for which adequate aerial photographs or ground information is available, is classified into forest and non-forest classes. In the second stage, AVHRR data are overlaid on the classified TM data and statistical relationships are derived between AVHRR spectral data and the amount of forested landscape, as classified in the TM data. In the third stage, the statistical model previously developed is applied to a larger AVHRR scene to generate regional maps of forest cover. For purposes of validation, the regional forest cover estimates are then compared to independent estimates.

In summary, our approach is to use nested scales of imagery to generate quantifiably accurate regional estimates of forest cover which cannot necessarily be

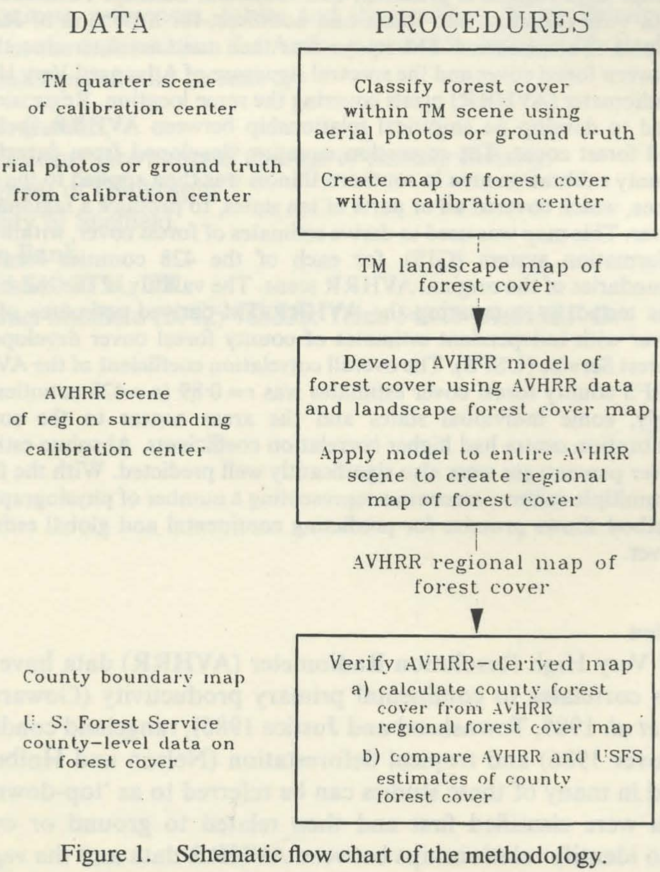

Figure 1. Schematic flow chart of the methodology.

directly measured by a sensor but is functionally related to surface reflectance characteristics of a forest canopy that TM and AVHRR sensors can detect.

2. Methods

The methods described here were developed using AVHRR and TM data on an area centred in southern Illinois, U.S.A. This area was selected because of the availability of satellite sensor and ancillary data on the forests, the fragmented nature and range of forest cover and its familiarity to the authors.

2.1. *Development of TM forest cover map*

TM data were extracted from an 18 July 1984 scene (path 24, row 34) over the south-west quarter of Jackson County, in southern Illinois, and grouped into forest and non-forest classes using unsupervised classification techniques. Verification of class definitions was accomplished using National High Altitude Photography (NHAP) and topographic maps. The binary image that resulted was then available for combining with and calibrating the AVHRR data.

2.2. *Development of AVHRR forest-cover model*

A local-area coverage (LAC) AVHRR data set from the NOAA-10 satellite and which covered 564 175 km^2 centred on Illinois (latitude 34°–44° N, longitude 86°–94° W) for the afternoon of 4 June, 1987, was acquired from the EROS Data Center in Universal Transverse Mercator (UTM) format with a pixel size of 1110 × 1110 m. The data covered 428 counties from all or parts of ten states. The 4 June date of acquisition corresponded to a time when all forests in the study areas were in full-leaf and the reflectances from row-crop agriculture (the primary non-forest feature of the region) were dominated by bare soil. Therefore, the vast majority of highly-reflecting, chlorophyll-rich land area was forestland on the date of AVHRR data acquisition. The assumption that forests will have the majority of chlorophyll reflectance at the time of AVHRR data acquisition needs to be met for this technique to be successful. One hundred and fifty-four AVHRR pixels corresponding to the Jackson County TM study area were extracted and overlaid in a geographical information system (GIS) such that 1369 TM pixels (37 × 37, 30 m pixels) defined each AVHRR pixel. The AVHRR pixels were each subset into 1369 subpixel units to allow direct correspondence to the TM data. An observation in the final data file of 30 m subpixels was then composed of (1) a unique AVHRR pixel identifier, (2) the AVHRR spectral band values associated with the subpixel unit (all will be identical within an AVHRR pixel) and (3) the TM forest class (forest or non-forest) associated with the subpixel unit.

A sampling program extracted data from every fourth line and fourth column ($\frac{1}{16}$th sample) to reduce data density; approximately 100 30 m pixels were used from each of 154 AVHRR pixels to derive estimates of forest cover for each AVHRR pixel. The TM-derived forest class information of the subpixel units was then aggregated to estimate the percentage forest cover within each AVHRR pixel. The relationship between AVHRR spectral values and forest cover within the AVHRR pixel was empirically determined using multiple regression analysis in which the TM-derived forest cover estimate was used as the independent variable. Various AVHRR spectral characteristics, including AVHRR bands 1 to 4, the normalized difference vegetation index (NDVI = (NIR − VIS)/(NIR + VIS) where NIR = near-infrared wavelength measurement and VIS = visible wavelength measurement), band ratios and other possible vegetation indices, were used as dependent variables. A standard error around the mean and 95 per cent confidence intervals were calculated to estimate the variance of the regression predictions. Correlation analysis was also used to identify individual relationships between spectral characteristics and the percentage forest cover.

2.3. *Generation of regional map of forest cover*

The AVHRR regression model that best (highest adjusted r^2) predicted percentage forest was applied to the overall AVHRR data set in the following manner. An unsupervised classification was performed on the four-band AVHRR data to mask pixels completely dominated by water, bare ground, or other non-forest data. These pixels were assigned to the 0 per cent forest class. A very conservative approach was taken in assigning pixels to the 0 per cent forest class. Topographic maps and aerial photographs were used in such a way that if a pixel was interpreted as having any forest, it was classed as a forested pixel, so that the regression equation could be applied to each pixel having at least some forest cover. An estimate of percentage forest could then be generated for each pixel over the entire region. The percentage forest estimates were aggregated into five cover classes: 0–10, 11–30, 31–50, 51–70 and 71–100 per cent.

2.4. *Validation of regional forest cover estimates*

The TM/AVHRR-derived estimates of forest cover were compared with the U.S. Forest Service (USFS) inventory data, which are acquired by county nationwide. These data were available through the Oak Ridge National Laboratory (ORNL) geo-ecology database (Olson *et al.* 1980) and were a compilation of data published by USFS from 1965 to 1980. For several states more recent USFS inventories were available and were used in place of the geo-ecology database; these included Wisconsin (Raile 1985), Illinois (Hahn 1987), Indiana (Hansen 1987) and Arkansas (Hines 1988). USFS forest cover estimates typically have about a 4 per cent sample error (at a 68 per cent level of confidence) for counties in this region (Hahn 1987).

To compare the USFS county information with AVHRR estimates, the forest cover values of the AVHRR pixels were aggregated to county level and summarized by overlaying pixel data with U.S. county boundaries and averaging the forest class values of all pixels within a county. There were a total of 428 county-level observations in the data set.

Correlation analyses were performed to compare the AVHRR estimates to the USFS estimates of percentage forest. This comparison was made in three ways: all counties grouped together, counties stratified by state and counties stratified by distance from the calibration centre. For the latter comparison, circular buffers away from the centre point of the calibration area of 0–100, 101–200, 201–300, 301–400 and >400 km were used.

3. Results

3.1. *AVHRR forest cover model*

Within the Jackson County calibration centre, several spectral characteristics were significantly correlated with percentage forest as ascertained by TM classification. The NDVI, calculated from AVHRR data, was correlated to percentage forest cover ($r=0{\cdot}585$, $n=154$, $p<0{\cdot}0001$), as were individual AVHRR bands 1 ($r=0{\cdot}599$, $n=154$, $p=<0{\cdot}0001$) and 2 ($r=0{\cdot}334$, $n=154$, $p<0{\cdot}0001$). The best two-variable regression model, which predicted forest cover over an AVHRR pixel and which violated no assumptions related to multi-collinearity, used a linear combination of bands 1 and 2 and had an adjusted r^2 of 0·407 ($p<0·0001$): percentage forest cover $=232-3·056$ (AVHRR band 1) $+0·615$ (AVHRR band 2).

This equation, when applied to all forested pixels (63 per cent of all pixels) of the AVHRR scene, and included with the remaining non-forested pixels, produced a regional estimate of percentage forest cover of 19·6 per cent, with 95 per cent confidence limits of 13·5 to 25·7 per cent. This finding compares well to the USFS calculated mean for the area of 20·8 per cent forest.

3.2. *Regional map of forest cover*

A map depicting county forest cover over all or parts of ten states was generated by applying the regression equation to each forested AVHRR pixel in the study region and aggregating the forest cover values within each county (figure 2). The AVHRR-based forest cover map realistically shows large regions of Illinois and Iowa with very low forest cover, with increased forest cover in the Ozarks and Mark Twain Forest of

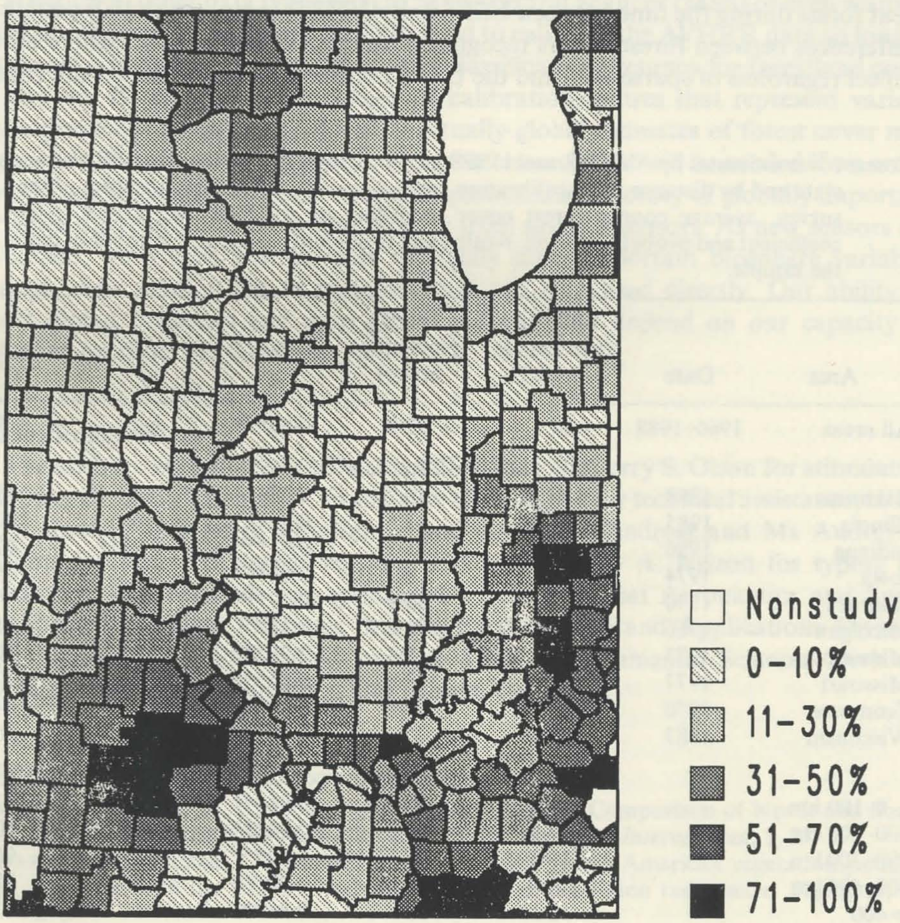

Figure 2. Estimated percentage forest cover using AVHRR data.

Missouri, the Hoosier Forest of Indiana, some southwestern Michigan forests, much of Wisconsin and the Shawnee National Forest of southern Illinois.

3.3. *Validation of estimates with USFS data*

To validate the accuracy of the AVHRR-derived estimates of county forest cover, the data were correlated to USFS-derived cover estimates over 428 counties. There was a high correlation between the AVHRR and USFS county estimates, with $r=0{\cdot}89$ overall (see the table). When evaluated by buffer distance from the original TM/AVHRR calibration site, the highest r values occurred within the 0 to 100 km radius ($r=0{\cdot}96$), with the correlation decreasing only slightly beyond 200 km (see the table). Analysis of states with over 30 county samples showed highly significant correlation coefficients ranging from $r=0{\cdot}72$ in 39 Kentucky counties to $r=0{\cdot}96$ in 77 Missouri counties (see the table).

Comparisons between mean county forest cover using pair-wise t-tests revealed a statistically similar estimate for the AVHRR data (24·1 per cent forest) compared to the USFS data (23·2 per cent forest) over all counties (see the table). Counties within the 100 km buffer zone matched exceptionally well (see the table). Some of the differences between estimates could be expected because of (1) actual changes in per cent forest during the time between USFS and AVHRR sampling and (2) definitional differences between forestland as recognized by the AVHRR sensor (where all trees reflect regardless of sparseness) and the USFS (where minimum densities are required

Forest cover estimates by AVHRR and USFS methods. Overall relationship and summaries by state and by distance from calibration centres are given. Data include date of the USFS survey, average county forest cover according to AVHRR and USFS, correlation coefficient and probability level, t-value and probability level, and number of counties in the sample.

Area	Date	AVHRR (percentage cover)	USFS (percentage cover)	r	t	n
All areas	1966–1988	24·1	23·2	0·89***	1·7	428
		Analysis by state				
Arkansas	1988	39·7	40·7	0·89***	−0·3	15
Illinois	1985	12·7	13·6	0·89***	−1·6	100
Indiana	1986	30·3	21·2	0·91***	6·9***	62
Iowa	1974	4·5	4·9	0·80***	−0·8	55
Kentucky	1978	42·1	33·4	0·72***	4·0**	39
Michigan	1966	35·6	41·8	0·78*	−1·6	11
Minnesota	1977	6·0	8·1	0·99***	−4·3*	10
Missouri	1977	32·8	32·6	0·96***	0·3	77
Tennessee	1970	34·1	36·6	0·80***	−0·7	24
Wisconsin	1983	24·6	25·6	0·79***	−0·4	36
		Distance from calibration centre				
0–100 km		28·5	28·9	0·96***	−0·3	27
100–200 km		27·4	29·9	0·94***	−2·7*	70
200–300 km		36·1	33·0	0·89***	3·1*	96
300–400 km		28·1	24·4	0·78***	2·2	82
>400		12·1	12·4	0·86***	−0·5	153

$*P \leqslant 0{\cdot}01$, $**P \leqslant 0{\cdot}001$, $***P \leqslant 0{\cdot}0001$.

before being classified as forestland). For example, in Illinois, USFS categories which are classified as 'non-forest with trees' (e.g., wooded strips, wooded pasture, windbreaks and urban forest) accounted for 364 000 ha statewide, or 2·5 per cent of the State's land area (Iverson *et al.* 1989). Given the limitations of direct correspondence between the two estimates and the inaccuracies associated with each method, the AVHRR-derived estimates compare very favourably to those derived from the USFS.

4. Discussion and conclusions

Our work demonstrates that nested scales of imagery in conjunction with ground-based data and a GIS can successfully generate landscape and regional estimates of forest cover. Furthermore, the approach permits the error associated with such estimates to be documented and is extremely thrifty in its use of imagery. Of course, certain assumptions have been made with this approach, the major ones being that trees were providing the predominant source of chlorophytic reflectance at the time of the AVHRR overflight, similar atmospheric conditions existed across the image, minimal scan angle effects were apparent and the calibration centre was representative of the region. Although TM data were used exclusively in this study to calibrate AVHRR data, other data types such as Multispectral Scanner (MSS) or even scanned aerial photography could potentially be used to calibrate the AVHRR data so long as positional accuracy of the data layers and classification accuracy for forestland could be assured. By relying on a network of calibration centres that represent various physiographic regions, regional and eventually global estimates of forest cover may be possible. Our research is a prototype of the research that will be needed if we are to develop spatially-extensive estimates with quantifiable accuracy of globally important variables that cannot be measured directly from satellite sensors. As new sensors are developed, we will be able to sense indirectly other important biosphere variables through their relationship to variables that can be sensed directly. Our ability to detect global processes and map global patterns will depend on our capacity to capitalize on these relationships.

Acknowledgments

The authors are indebted to Drs Paul G. Risser and Jerry S. Olson for stimulating this research, to Dr Thomas D. Frank and Mr Ying Ke for technical assistance, to Mr Mark Joselyn, Drs Mark MacKenzie and Anton G. Endress and Ms Audrey S. Hodgins for review of earlier versions and to Ms Betty A. Nelson for typing the manuscript. The research was sponsored by the National Aeronautics and Space Administration's Land Processes Branch, Earth Science and Applications Division grant NASA-28781. Publication number 3298, Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

References

Goward, S. N., Dye, D., Kerber, A., and Kalb, V., 1987, Comparison of North and South American biomes from AVHRR observations. *Geocarto International*, **1**, 27–39.

Goward, S. N., Tucker, C. J., and Dye, D. G., 1985, North American vegetation patterns observed with the NOAA-7 advanced very high resolution radiometer. *Vegetatio*, **64**, 3–14.

Hahn, J. T., 1987, Illinois forest statistics, 1985. USDA Forest Service, *North Central Forest Experiment Station Research Bulletin* NC-103 (St Paul, Minnesota: USDA).

HANSEN, M. H., 1987, Forest area in Indiana, 1986, USDA Forest Service, *North Central Forest Experiment Station Research Note* RN-NC-341 (St Paul, Minnesota: USDA).

HINES, F. D., 1988, Forest statistics for Arkansas' Ozark counties, 1988. USDA Forest Service, *Southern Forest Experiment Station Research Bulletin* SO-131 (New Orleans, Louisiana: USDA).

IVERSON, L. R., OLIVER, R. L., TUCKER, D., RISSER, P. G., RAYBURN, R., and BURNETT, C., 1989, The forest resources of Illinois: an atlas and analysis of spatial and temporal trends. *Illinois Natural History Survey Special Publication,* 11, Champaign, Illinois.

IVERSON, L. R., and RISSER, P. G., 1988, Analyzing long-term vegetation change utilizing geographic information system and remotely sensed data. *Advances in Space Research*, **7**, 183–194.

NELSON, R., and HOLBEN, B., 1986, Identifying deforestation in Brazil using multiresolution satellite data. *International Journal of Remote Sensing*, **7**, 429–448.

OLSON, R. J., EMERSON, C. J., and NUNGESSER, M. K., 1980, *Geoecology: a county-level environmental data base for the conterminous United States.* Oak Ridge National Laboratory Environmental Sciences Division Publication No. 1537 (Oak Ridge, Tennessee: Oak Ridge National Laboratory Environmental Sciences Division).

RAILE, G. K., 1985, Wisconsin forest statistics, 1983. USDA Forest Service, *North Central Forest Experiment Station Research Bulletin* NC-94 (St Paul, Minnesota: USDA).

SADOWSKI, F. G., and WESTOVER, D. E., 1986, Monitoring the fire-danger hazard of Nebraska rangelands with AVHRR data. *Proceedings of the Tenth Canadian Symposium on Remote Sensing held in Edmonton, Alberta, Canada*, pp. 355–363.

TOWNSHEND, J. R. G., and JUSTICE, C. O., 1986, Analysis of the dynamics of African vegetation using the normalized difference vegetation index. *International Journal of Remote Sensing*, **11**, 1435–1445.

TUCKER, C. J., TOWNSHEND, J. R. G., and GOFF, T., 1985, African land-cover classification using satellite data. *Science, New York*, **227**, 369–375.