Tenth International Symposium

Machine Processing of Remotely Sensed Data

with special emphasis on

Thematic Mapper Data and Geographic Information Systems

June 12 - 14, 1984

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

AN ANALYSIS OF THE UTILITY OF LANDSAT THEMATIC MAPPER DATA AND DIGITAL ELEVATION MODEL DATA FOR PREDICTING SOIL EROSION

D.B. GESCH, B.I. NAUGLE

Mid-America Remote Sensing Center
Murray State University
Murray, Kentucky

I. ABSTRACT

The Universal Soil Loss Equation (USLE) is an erosion model which is well-suited for use in a geographic information system format. A georeferenced data base with inputs necessary for the USLE was assembled. Its components included three levels of digitized soil polygons, land cover classifications from Landsat TM and MSS data, and slope information derived from Digital Elevation Model (DEM) data. Six different combinations of the data base components were used in the USLE to calculate erosion estimates over a small watershed in western Kentucky. The six estimates of erosion were tested in pairwise fashion to determine if differences between estimates were statistically significant. Estimates using DEM slope information were significantly different from those using soil survey slope information. Likewise, estimates utilizing the TM land cover classification were significantly different from those using the MSS classification. The level of detail of soil polygons also affects soil loss predictions. Overall, the use of TM data, soil series polygons, and DEM data significantly affected the results of soil loss estimation over the study area.

II. INTRODUCTION

Soil erosion by water has been recognized as a significant resource problem. It reduces the inherent productivity of agricultural soils by removing topsoil, thus decreasing soil depth, nutrient content, and water holding capacity. It also contributes to the degradation of water quality by creating nonpoint source pollution by transportation of absorbed pesticides, and it increases flood danger through sedimentation. Because soil erosion is influenced by many variables it is difficult to quantify directly, but it can be estimated numerically through analysis of data dealing with precipitation, soil type, slope conditions, crop type, and agricultural practice. Beginning about 1940, various methods and procedures led to regionally based soil loss equations. Adaptations of and improvements upon those earlier models resulted in what has become known as the Universal Soil Loss Equation (USLE). The USLE is an erosion model developed to predict the long-term average soil losses from specific field areas in specified cropping and management systems (Wischmeier and Smith, 1978, p. 3). The Universal Soil Loss Equation is:

$$A = R\ K\ L\ S\ C\ P$$

where A = the computed amount of soil loss
R = the rainfall erosivity index
K = the soil erodibility factor
L = the slope length factor
S = the slope gradient factor
C = the cover and management factor
P = the conservation practice factor

The product of R, K, L, S, C, and P gives an estimate of soil loss, which is usually expressed in tons/acre/year.

Because of its design, the USLE is well-suited for use in a geographic information system. The various parameters of the equation form the separate layers of information for the georeferenced data base. With the availability of Landsat 4 Thematic Mapper (TM) data a new source exists for land cover information needed for the USLE. Goward (1983, p. 1) states, "the Thematic Mapper (TM) instrument is designed principally to acquire detailed information about the type, amount, and condition of vegetation, particularly cultivated vegetation." The TM's 30 meter spatial resolution and its seven spectral bands are an improvement over its predecessor, the Landsat Multispectral Scanner (MSS), which has an 80 meter spatial resolution and four spectral bands. Also, a new source exists for the topographic information required by the USLE with the availability of Digital Elevation Model (DEM) data. DEM data, distributed by the National Cartographic Information Center (NCIC), corresponds to USGS 7.5 minute quadrangles. DEM data have a spatial resolution of 30 meters, a vertical resolution of one meter, and a root mean square (RMS) elevation error of seven meters.

The primary purpose of this study was to determine if the use of 30 meter TM data and DEM slope data, in the context of the USLE, would result in significantly different erosion estimates from those obtained by using 80 meter MSS data and general slope data from the soil survey. A secondary objective was to determine if erosion estimates obtained from very detailed digitized soil polygons would differ significantly from estimates obtained by using more general soil association polygons.

III. LITERATURE REVIEW

The utility of remote sensing for predicting soil erosion has been realized by numerous authors. Morris-Jones and Kiefer (1978, pp. 159-186) successfully used manual photo-interpretation techniques on high altitude infrared photos and medium altitude color and color infrared photos to gather all types of land use/land cover information needed to predict soil erosion. Morgan et al (1978, pp. 291-293) also used high altitude color and color infrared photos for C-factor determination for the USLE. Their soil loss estimates compared favorably with conventional field estimates. They also suggested the application of satellite imagery to areawide erosion studies because of its up-to-date, repetitive coverage. Fenton (1980, pp. 217-231) noted that soil information and land cover information from remotely sensed data provide a valuable resource data base. He added that computerization of this type of information in a data base is an efficient means of storing, retrieving, and analyzing data.

Several studies have illustrated the utility of using a geobased information system format for erosion predictions. Mooneyhan (1980, pp. 30-50) cites a study in which Landsat MSS data was used to produce land cover classifications specific to those required as input to the USLE. Also in that study, a digitized soil survey, soil erodibility information, and slope data were entered into the geographic data base.

Digital Elevation Model data has been previously used for soil loss prediction with the USLE in a geographic information system. Spanner (1983, pp. 314-321) compared DEM data and Defense Mapping Agency digital terrain tapes. He concluded that the DEM data was very useful and that it was better than the DMA topographic data for determining the slope factors for the USLE.

IV. STUDY AREA

The study area is a small watershed, approximately 15 square miles (39 sq. km), located in McCracken County in western Kentucky. The watershed of the East Fork Massac Creek lies on two USGS 7.5 minute quadrangles, Paducah West, Ky. and Melber, Ky. The area has nearly level to sloping topography, and the soils have formed in thick loess. The major land use of the area is rural agriculture. Common practices on the cropland include wheat and soybeans in a double crop system and corn or soybeans in a single crop system. Both conventional and minimum tillage practices are used in the area. This specific area was chosen for study for several reasons. TM imagery and MSS imagery from the same date are available. Also available are the DEM data for the Paducah West and Melber quadrangles, and a detailed soil survey of McCracken County. The East Fork Massac Creek watershed was chosen because of its manageable size and because of its variable topography, soil types, and land cover.

V. DATA PROCESSING AND ANALYSIS

Processing of all data was accomplished using the 16 bit Univac V-77 computer, COMTAL image processing system, and TALOS digitizer housed at the Mid-America Remote Sensing Center (MARC). The software used to process the Landsat and DEM data is the Earth Resources Laboratory Applications Software (ELAS) developed at NASA-Earth Resources Laboratory. Processing of digitized soil data was done with programs from ERL and other software developed at MARC.

The data collection procedure included these steps:

1. The watershed boundary and all soil boundaries from the soil survey were transferred to a geographically correct map base by use of the zoom transfer scope.

2. All boundaries were digitized from the map base, which corresponded to the quadrangles containing the study area. The digitized data was then edited, and appropriate corrections were made. The editing and organization of the digitized soil boundaries turned out to be the most time-consuming task of all the data processing, due mainly to the great number of segments needed for the detailed soil polygons. However, the task was greatly simplified by a series of programs developed at MARC to organize the digitized segments, enabling them to be easily put together to form closed area polygons. The first program in the series corrects the nodes, those places where two or more digitized segments meet. The endpoint of each segment at a node is set equal to a common point so that none of the segments extend past or fall short of where they should meet. A correct node is crucial later on when the individual segments are put together to form closed area polygons. When all nodes have been corrected, a second program is used to identify which individual segments are needed to form the specific area polygons. This algorithm was invaluable to this study, again due to the great number of segments which had to be manipulated. Once the listing of area polygons is found to be correct, a third program is employed to build the input for ELAS. An ELAS module actually puts together the segments to form the area polygons and saves them with a unique identifying number. At this point, the soil polygons were ready to be loaded into the data base.

3. Land cover classifications were run on the Landsat MSS and TM data. The date of the scenes, Path 22/Row 34, is October 18, 1982. With each data set the first step was an unsupervised clustering procedure. A 3 x 3 pixel window was passed through the data to acquire homogeneous training field statistics. This unsupervised search of the MSS data resulted in 27 classes. The search over the TM data resulted in 32 classes. Both data sets were then classified according to the appropriate statistics with a maximum likelihood classifier. Each classification was then interpreted with color infrared photos over the study area. In each case, there were a number of classes which represented mixed cover types. For both the MSS and the TM there were classes which included pixels of cropland and grassland. This is due to the date of the imagery. In mid-October most double crop soybeans are still in the field, and these appear spectrally similar to grassland. For the MSS, 11 of the original 27 classes were a mixed cover type. Some of these 11 classes were the soybean/grass mix, while others were a grass/forest mix, most likely "edge" pixels. For the TM, 9 of the original 32 classes showed a mix, and all of these were of the soybean/grass type. A procedure called within class clustering was used to better separate the cover types in these mixed classes. It is a point clustering procedure which effectively breaks down the original classes. The 11 mixed MSS classes were broken down into 26 classes, while the 9 mixed TM classes resulted in 48 new classes. A maximum likelihood classifier then classified only those pixels on which clustering was performed. These classifications were then interpreted with the CIR photos. The individual TM classes were mostly restricted to one cover type, and where there was still a mix, one cover type was a high percentage of the class. Thus, the class was grouped with the cover type with which it was judged to be most closely associated. The MSS classes showed more of a mix than the TM classes, probably due to the resolution of the MSS. Again, though, the classes were grouped with the most representative cover type. For the purposes of this study, three final classes were decided upon: cropland, grassland/pasture, and forest. It was initially thought that cropland could be divided into double crop and single crop cropland, but as classification results were reviewed this was seen to be impossible without a great amount of error. Acreages of each cover type from the MSS and TM classifications are given in Table 1.

Table 1. Acreages of Cover Types

Data Source	Acres and % of Watershed Cropland	Grassland	Forest
MSS	4078(41.4)	2284(23.2)	3484(35.4)
TM	5206(52.9)	1760(17.9)	2879(29.2)

4. The Digital Elevation Model data required a great deal of processing before meaningful slope data could be derived. When the DEM data was viewed in its raw form it showed severe striping across the data set. When it was processed for slope information in this form it showed many "false" slopes, even some in the middle of a large water body. An algo-

rithm called histogram matching was employed to rid the data of the stripes. In this routine, a histogram of data values is calculated for a window of a certain number of lines. The histogram for the line at the center of the window is then compared to the overall histogram. If the difference is greater than a predetermined tolerance the values of the center line are changed so that its histogram matches the overall histogram of the window. This procedure effectively removed most of the regular striped pattern seen in the raw data. The data was then smoothed with a simple averaging filter. At that point the DEM was processed for slope information. The algorithm to output the percent slope passes a 3 x 3 pixel window through the data. The largest gradient across the window is identified and the center pixel is assigned that slope. Nineteen slope classes were used, 0-15% slope in increments of 1% and 15-30% slope in increments of 5% with the last class being slopes greater than 30%.

5. All soil, slope, and land cover information was loaded into the geographically referenced data base with a cell size of 30 meters. The MSS and TM classifications were georeferenced to the UTM grid with ground control points by a method using nearest neighbor resampling. Appropriate USLE factor values were determined and assigned. Weighted averages were used for K-factor and slope information from the soil survey.
6. Erosion estimates were calculated with the Universal Soil Loss Equation for various combinations of the data base components. The individual components were soil series polygons, soil association polygons, DEM slope categories, MSS land cover, and TM land cover.
7. Erosion estimates from the various calculations were tested in pairs to see if differences were statistically significant. The method of analysis was the paired difference t-test. Corresponding cells in different estimates formed the pairs. A sample was created by taking the difference in each pair, and then the mean of the differences was tested against zero.

The six different combinations (Method 1-Method 6) of the data base components used for erosion calculations are shown in Table 2. As can be seen, constants were used for the R and P factors.

Table 2. Data Base Components Used for USLE

Method	R	K	L	S	C	P
1	250	Assoc.	Assoc.	Assoc.	TM	1.0
2	250	Series	Series	Series	TM	1.0
3	250	Series	Series	Soil Slope	TM	1.0
4	250	Assoc.	Assoc.	DEM	TM	1.0
5	250	Series	Series	DEM	TM	1.0
6	250	Series	Series	DEM	MSS	1.0

The K-, L-, and S-factors for the associations are weighted averages of the values for the series that form the associations. The soil slope classes used for S in Method 3 are the slope categories within the series as given in the soil survey. Acreages of erosion estimates (in t/ac/yr) in five erosion categories are shown for the six methods in Table 3.

Table 3. Acreages of Erosion Classes.

		Method				
Erosion Classes	1	2	3	4	5	6
Acceptable (<5)	8257	5800	5801	4828	4785	5898
Slight (5-10)	0	128	304	744	1170	964
Moderate (10-15)	0	140	0	1315	776	642
Severe (15-25)	0	4	10	1132	1550	1232
Extreme (>25)	1588	3773	3730	1826	1564	1109

VI. RESULTS

Differences in erosion estimates between the various methods were tested for statistical significance. A stratified random sample of 100 cells was picked within the watershed. Method 5 was used for the stratification because all five of the erosion categories were represented in its results. Thus, 20 random points were chosen in each of the five erosion classes. The six erosion estimates were then recorded for each chosen cell. A logarithmic transformation was done on all data values to stabilize the variance. All statistical processing was accomplished with the Statistical Analysis System (SAS).

The null hypothesis in each of the pairwise tests was that there was no significant difference between the erosion estimates from the two methods. At the beginning of the study an alpha level of 0.05 was set to determine significance. However, all the tests which showed significant differences were significant at levels far below 0.05. The first set of tests were conducted to see what effect DEM slope data had on erosion predictions. Method 1, which used average slope from soil associations, was tested against Method 4, which used DEM slope data. All other factors besides slope were held constant. This test was significant with a p-value at 0.033. A second test compared Method 2, which used average slope from soil series, with Method 5, which used DEM slope data. Again, all other factors were held constant. This test was highly significant with a p-value of 0.0007. The third test compared Method 3, which used soil slope classes, with the DEM slope data of Method 5. The p-value of 0.01 of this test shows the significant difference between these two methods. Thus, in all three cases where DEM slope data was used in the USLE, erosion estimates were significantly different than those obtained by using soil survey slope data.

A second set of tests were conducted to determine what effect the detail of digitized soil polygons had on erosion predictions. Method 1 using associations was tested against Method 2 using series. DEM slope data was not used in either method, so the test was really a comparison of average slope from the associations and average slope from the series. The differences were highly significant with a p-value at 0.0004. A second test looked at Method 2 and Method 3. Again, DEM slopes were not used in either method so it was a comparison of soil slope classes with average slope from the series. With a p-value of 0.146 the differences were not significant and the null hypothesis was accepted. The third test compared Methods 4 and 5. DEM slope classes were used in both methods, so it was really a test of K- and L-factors between the associations and the series. This test was significant at the 0.0019 level. The tests in this set show that the detail of soil polygons, and their associated K- and L-factors, does make a significant difference in erosion estimates. This is true between the association and series level. When soil slope classes from the survey were used for estimates and then compared to estimates using average slopes for series, no significant difference was detected. If average slopes within soil polygons are to be used for calculations, digitizing the slope categories within the series does not significantly change results.

The last test conducted looked at Method 5 and Method 6. DEM slopes were used in both, and all other factors were held constant, except the C-factor. Method 5 used C-factors derived from the TM land cover classification, while Method 6 used C-factors from the MSS classification. The differences in estimates were highly significant at a p-value of 0.0001. This result was expected because, as it was seen earlier, the TM classification resulted in many more acres of cropland.

VII. CONCLUSIONS

The use of C-factors derived from classified Landsat TM and MSS data in the USLE resulted in significantly different erosion estimates. The TM data was easier to classify, and its classification resulted in fewer classes representing mixed cover types. Use of slope information from the DEM data resulted in significantly different erosion estimates from those obtained by using soil survey slope data. The detail of digitized soil polygons also had a significant effect on erosion estimates. However, the use of digitized slope categories within the soil series does not significantly affect the erosion calculations. Although it cannot be concluded definitely that the erosion estimates from the more detailed data sources are more accurate than those from the less detailed sources, it is presumed to be the case.

Even though the combination of soil association polygons and DEM slope data (Method 4) gave significantly different erosion estimates from soil series polygons used with DEM slopes (Method 5), Method 4 should be investigated more completely because of the ease of its application. If Method 4 can be shown to be addquate from an erosion management standpoint, the effort of digitizing and manipulating soil series polygons could be avoided.

Finally, the quality of the DEM data used in this study should be discussed. As mentioned earlier, a great deal of processing was necessary before meaningful slope data could be derived from the DEM. The authors found no evidence of this problem in other studies which used DEM data. Some of these studies were in areas of more rugged terrain where the error caused by striping would have probably been less significant. The striping in the DEM falls within the stated seven meter RMS error, but in an area of low relief the slopes from line to line where striping occurs will predominate. These slopes are not to be found on the actual land surface.

VIII. ACKNOWLEDGEMENTS

The authors would like to thank Dr. Thomas C. Kind for his input and assistance on this project, and also all the faculty and staff at the Mid-America Remote Sensing Center for their support.

IX. BIBLIOGRAPHY

Goward, Samuel N., 1983, "Utilizing Thematic Mapper Data in Vegetation Analysis," presented at the Applied Remote Sensing Workshop, 1983 AAG National Meeting, Denver, Colorado.

Fenton, T. E., 1980, "Estimating Soil Erosion by Remote Sensing Techniques," in Remote Sensing for Resource Management, edited by C. J. Johannsen and J. L. Sanders, Soil Conservation Society of America, Ankeny, Iowa, 1982, pp. 217-231.

Mooneyhan, D. Wayne, 1980, "Organizing Information for Effective Resource Management," in Remote Sensing for Resource Management, edited by C. J. Johannsen and J. L. Sanders, Soil Conservation Society of America, Ankeny, Iowa, 1982, pp. 30-50.

Morgan, Ken M., Gerhard B. Lee, Ralph W. Kiefer, T. C. Daniel, Gary D. Bubenzer, and John T. Murdock, 1978, "Prediction of Soil Loss on Cropland with Remote Sensing," Journal of Soil and Water Conservation, Vol. 33, pp. 291-293.

Morris-Jones, Donald R. and Ralph W. Kiefer, 1978, "Application of Remote Sensing to Estimating Soil Erosion Potential," in Remote Sensing of Earth Resources, Volume VII, edited by F. Shahrokhi, 1980, pp. 159-186.

Spanner, Michael A., 1983, "The Use of Digital Elevation Model Topographic Data for Soil Erosion Modeling within a Geographic Information System," in Technical Papers of the 49th Annual Meeting of the American Society of Photogrammetry, 1983, pp. 314-321.

Wischmeier, W. H. and D. D. Smith, 1978, "Predicting Rainfall Erosion Losses - A Guide to Conservation Planning," U.S. Department of Agriculture, Agriculture Handbook No. 537, 58 p.

X. BIOGRAPHICAL DATA

Dean Gesch is a graduate student in the Department of Geosciences at Murray State University. He received his B.S. degree in Physical Geography at Carroll College in Waukesha, Wisconsin. He has worked for two years as a research assistant at the Mid-America Remote Sensing Center and has been involved in several projects, including a pine forest inventory, an investigation of the utility of remote sensing data for archeological applications, and a regional land cover and soil loss study.

Burl Naugle, a Professional Civil Engineer, is an Assistant Professor in the Department of Geosciences at Murray State University, with a joint appointment at the Mid-America Remote Sensing Center. He acquired his B.S. and M.S. degrees in Agricultural Engineering at Purdue University and his Ph.D in the same field at the University of Kentucky. His research interests are in digital processing of remotely sensed data and digital cartography.