COUNTDOWN TO THE 21ST CENTURY

ACSM/ASPRS
ANNUAL CONVENTION
MARCH 18-23, 1990
DENVER, COLORADO

POSTCLASSIFICATION CLASS SORTING IN CREATING A GENERAL LAND COVER INVENTORY

Paul R. Baumann
James B. Greenberg
Department of Geography
SUNY College at Oneonta
Oneonta, NY 13820

ABSTRACT

Ancillary information such as topographic data is frequently incorporated into multi-spectral data sets as additional data channels and used in conjunction with spectral data to form spectral classes. These classes are needed in statistically classifying a data set. Each picture element in a data set is assigned to one spectral class and then an investigator tries to relate each spectral class to a single land cover condition. Often spectral classes, developed either with or without ancillary data, relate to more than one land cover making it difficult to determine the appropriate land cover assignment. In this study the spectral classes are created by employing only spectral data and the ancillary information is used to separate pixels associated with a particular spectral class into more than one land cover. Hutchinson refers to this approach as postclassification class sorting built on the logical use of ancillary data. This study examines the application of postclassification sorting on creating a general land cover product. Oneonta, New York, situated in the Catskill Region, is the study area.

INTRODUCTION

Hutchinson (1982) has identified three different approaches to using ancillary data in conjunction with spectral data to improve land cover classifications. These approaches are: 1) preclassification scene stratification, 2) classification modification, and 3) postclassification class sorting. This study centers on the postclassification method, examining some of the issues encountered in employing this approach in attempting to develop a general land cover classification. Postclassification sorting is based on the observation that many spectral classes do not relate to a single land cover class. They generally represent two or more land cover conditions making it difficult for the investigator to assign a spectral class to one particular land cover. In the postclassification method the investigator establishes sorting rules based on the logical use of ancillary data as they pertain to the particular land cover being examined. These rules permit the investigator to sort the pixels associated with a particular spectral class and assign them to different land cover classes. In this study, postclassification sorting was attempted in trying to improve a general land cover classification of a rural area in Upstate New York. The ancillary information was elevation, slope, and aspect data taken from USGS 7 1/2 minutes quadrangles.

STUDY AREA

The study area is Oneonta, New York and its environs. Oneonta is a community of approximately 20,000 people located halfway between Albany, N.Y. and Binghamton, N.Y. It is situated in the Susquehanna River valley and spreads out over the adjacent hills. The region around

Oneonta consists of hilly topography with a local relief ranging from 800 to slightly over 1100 feet. Most of the hill tops and slopes are covered with mixed deciduous-evergreen forests with some pure stand patches of coniferous trees. Brushland and pasture occupy the lower slope areas giving way to cropland in the lower portions of the valleys. The Susquehanna River flows southward on the eastern edge of the study area and then turns westward moving through the center of the area. Otego Creek and its valley link up with the Susquehanna on the west side of the study area. In general, the topography has been heavily modified by glacial activities. This region was selected as the study area due to the availability of the appropriate data sets and the investigators' familiarity with the region. In addition, easy accessibility to the area allowed for quick field checking when needed. The study area was rectangular in shape corresponding to the data sets and covered about 200 square miles.

DATA SETS

The topographic data set of the study area was constructed from elevation data extrapolated from portions of nine USGS 7 1/2 minute quadrangle maps. A grid was superimposed on the maps and orientated to the UTM coordinate system. An elevation reading was taken at every 200 meter mark on the grid resulting in over 13,000 readings being collected and entered into a computer file. From this data base an elevation value was interpreted for every 50 meter grid point and from this new data base an average elevation value as well as slope and aspect values were determined for the center of each grid cell. The USGS maps used to form this ancillary data were developed back in the early 1940s and although the cultural features on these maps have been updated, the contour lines still represent landform conditions of the 1940s. Over the years rivers and streams have shifted and highways and shopping centers have been constructed, resulting in the formation of different land surface conditions. The contour lines no longer represent the true situation in several areas on the maps. However, these maps form the most up-to-date data base for elevation that is available and they contain greater topographic detail than the 1:250,000 USGS map of the area from which the National Cartographic Information Center has prepared a digital terrain data set.

An October 12, 1972 Landsat MSS data set was used for the study. This fall data set is cloud free and each of its four spectral channels possesses a high quality rating. It is very difficult to obtain a cloud free period over the Oneonta region. Winter is the best time to find such a condition but the snow which frequently covers the ground during this period generally does not provide a desirable study surface. Summer is either too hazy or cumulus clouds exist over the area. More recent cloud free MSS data sets are available for the region but due to the various problems associated with the operation of the Landsat multi-spectral scanners over the years none of these data sets offer good quality data in all four spectral channels. In addition to being cloud free and having high data quality, this data set shows land cover in Oneonta before several major changes occurred. These changes modified the topography, and thereby, outdated some of the contours shown on the USGS maps. This data set is the earliest good data set that can relate to topography which was mapped in the early 1940s.

METHODOLOGY

The spectral and topographic data sets were merged to form a seven channel set. Classifications based on various combinations of spectral and topographic channels were attempted with varying degrees of success (Baumann, 1985). Due to the nature of the data, aspect was not used in any of these initial tests. Elevation and slope tended to dominate the classification results at the expense of losing some of the spatial detail provided by the spectral channels. For this study only the four spectral channels were used to create a classified data set. An unsupervised technique known as SEARCH was employed to produce the spectral classes for this classified data set. SEARCH, developed by NASA's Earth Resources Laboratory (1980) and modified by the investigators, sends a 3 pixel by 3 pixel window across the raw data set seeking training fields which are spectrally homogeneous. These training fields are used to form spectral classes. Homogeneity is defined by certain parameter limits specified by the investigator. Eighty-two spectral classes were generated by SEARCH from the four spectral channels and the parameter limits established. The statistics generated for the spectral classes as well as the raw data set with the four spectral channels formed the input into the maximum likelihood classifier which was used to develop the classified data set.

In previewing the classified data set many of the spectral classes demonstrated characteristics associated with more than one land cover condition. Working through the Cornell University Theory Center, the investigators were able to use the University's supercomputer and special graphics facilities to create an interactive computer graphics environment for testing in a postclassification approach the use of the topographic data to improve spectral class/land cover class associations. Since no software package was readily available based on the postclassification approach, the investigators prepared a program which allowed them to display the classified data set on a high resolution IBM 5080 graphics workstation. This program permitted them to highlight the pixels associated with a particular spectral class and establish rules to subdivide these pixels into different groups based on the topographic data. Since the program was interactive, the investigators were able to test quickly a variety of rule conditions with respect to a particular spectral class. However, to develop the appropriate rule conditions for each of the spectral classes that related to more than one land cover would require a considerable amount of time. In addition, the postclassification approach introduced a new set of issues that the investigators had to appraise. Due to this situation the investigators centered their work on certain spectral classes and did not attempt to deal with each spectral class possessing a problem. This study concentrates on four spectral classes and tries to produce a water category for a general land cover classification from these four classes.

Of the eighty-two spectral classes generated four showed in varying degrees an association with water conditions. Spectral class No. 70 illustrated a strong association with the open water areas of lakes and ponds but did not define well the edge and inlet areas of these water bodies. This class offered little assistance in identifying the Susquehanna River or its tributaries. Class No. 74 outlined certain sections of the river and defined some of the edge areas around the lakes. However, this class was also associated with steep, north and northwest facing slopes containing large stands of evergreen trees, namely hemlock, and river cuts. Like class No. 74 both classes No. 77 and No. 78 identified certain portions of the river and areas around the lakes. In addition, No. 77 related to old residential areas situated on glacial terraces as well as the evergreen covered north and northwest facing slopes. No. 78 possessed a strong association with Oneonta's large railroad yard located on the river's floodplain. Collectively, these four classes had information needed to develop a reasonable water category for a general land cover classification, but except for No. 70, these classes related also to other land cover conditions. Figure 1 shows the mean spectral signatures of these four classes. Class 70 demonstrates a normal water signature with low means in both channels 6 and 7. The other three classes show mixed signature patterns and little indication toward water conditions. However, the relatively low mean values in channel 7 of these three classes suggests the existence of wetlands but not large, open water conditions.

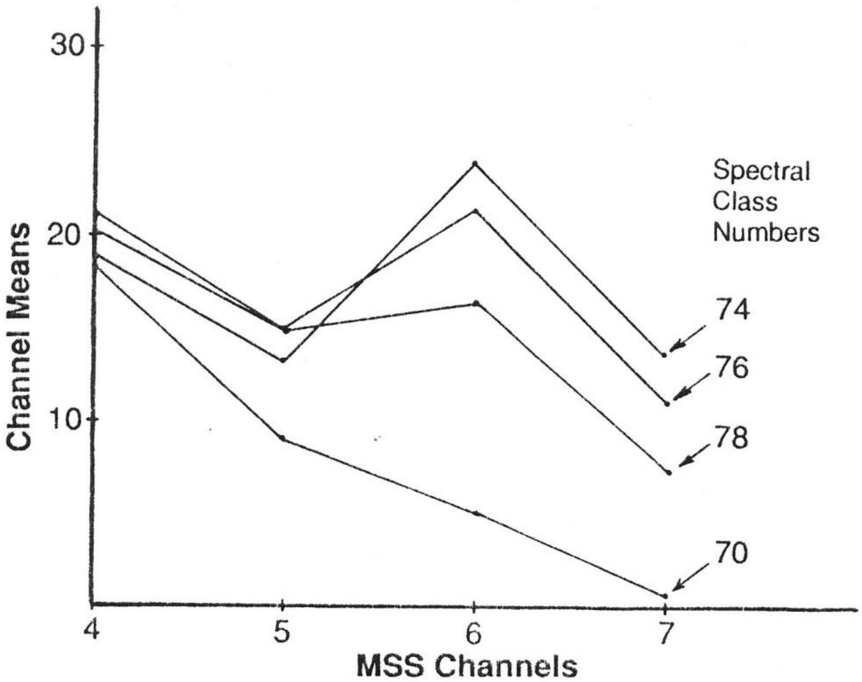

Fig. 1: Spectral Signatures of Water Classes

The Susquehanna River originates in Otsego Lake which is about twenty miles north of the study area; thus, the River is not extremely wide as it cuts through this region. The River also experiences low level conditions generally during the fall season when its rocky and muddy banks are exposed and create high reflectances. The spectral data set represents a dry fall period. The River meanders in several sections and possesses a number of back water areas. A dam has been placed on the river to form a small reservoir known as Goodyear Lake.

SORTING RULES

Since Class No. 70 was clearly associated with water, it was not necessary to establish any sorting condition to separate it from other land cover. In order to sort out the water areas from the steep, evergreen covered, north and northwest facing slopes related with Class No. 74, several different sorting rules were attempted. Elevation was initially used to form a separation between the valleys and the higher locations. Any pixel associated with Class No. 74 and below 1150 feet in elevation was assigned to water. This rule successfully removed the slope areas but it also removed the water areas around the edge of lakes and ponds located at higher elevations. Since the lakes and ponds are found at various elevations, it was not possible to find an elevation range which would only eliminate the slopes. A rule based on aspect was also developed which centered on separating north and northwest facing slopes from other angles. Although most of the evergreen forests assigned to Class No. 74 were located on these types of slopes, some appeared on other aspects, and thus, were not separated. Finally, slope was investigated, and after trying several variations, all pixels between 0 and 3 percent slope were classified as water. This rule eliminated many of the problem conditions found with elevation and aspect but did identify another problem area. Sections along the river where normally a low slope condition prevailed were not being classified as water. These areas related to the outer banks of meanders in the river and possessed slopes greater than 3 percent. More specifically, these areas represented sections where the river was cutting into the glacial terraces located along the valley edge. The spectral-based pixels associated with the river were superimposed on these steep slope cuts, creating a non-logical relationship. This problem was the result of either poor registration between the spectral and topographic data sets when they were merged or the difference between the date of the two data sets. Since a very low root mean square occurred between the control points of the two data sets and no apparent registration problems materialized in other sections of the study area, the latter explanation appeared more feasible. Between 1942 and 1972, the two dates associated with the data sets, the river cut more into the terraces resulting in the 1972 river being superimposed on the 1942 cut slope areas as shown on the 1942 maps.

TABLE 1: SORTING RESULTS

Class No.	Sorting Rule	Water Pixels	Non-Water Pixels	Total Pixels	Percent Water
70	None	495	0	495	100.0
74	Slope 0-3.0%	763	2035	2798	27.2
77	Slope 0-2.5%	569	815	1384	41.1
78	Elevation 1100-2180 ft.	193	165	358	53.9
	TOTAL	2020	3015	5035	40.1

Class No. 77 also demonstrated confusion between steep slopes with evergreen coverage and water along lake edges and river paths. In addition, this class identified the older sections of the Oneonta urban area which consisted of the central business district and the tree covered residential neighborhoods. The newer sections of the city as well as those residential sections which had expanded to the adjacent higher slope areas did not fall into this class. Elevation and aspect rules were tested with this class but the results were similar to those encountered with Class No. 74. A large portion of the city is situated on a terrace with the 1100 foot contour representing the break between the terrace and the floodplain. Thus, an elevation rule based on this contour was formulated. Everything above 1100 feet was not classified as water. This rule worked well for the city but farther up the valley, where the floodplain was above 1100 feet, river pixels were being classified as non-water areas. From this experience the investigators noted that more sophisticated rules were needed which might allow sorting to be limited to certain defined sections of a study area or permit the parameters assigned to a rule to change at certain increments as the data set was being sorted. Since elevation did not produce acceptable results, slope was tested. Those pixels with slopes between 0 and 2.5 percent were classified as water. The flatter sections of the terrace associated with Class No. 77 could not be separated from the conditions along the floodplain. Consequently, part of the city, mainly the central business district, was classified as water. The investigators elected to use this rule rather than the elevation rule since empirically this rule appeared to present better results. The investigators also noted in working with this class that although they were able to separate with reasonable success water from non-water areas, they were not able to handle separations within the non-water category. In other words, they could not distinguish between the urban areas and the steep, evergreen covered slopes without impacting the water situation. Postclassification sorting can become very complex when a spectral class relates to more than two land covers.

Class No. 78 related to only 358 pixels, many of which identified the large railroad yard in Oneonta. This yard was developed during the period of the steam engine train which was generally confined to areas with grades of less than 2 percent. Consequently, the yard was built on the floodplain where a very low slope condition existed and few people wanted to live. This class also related to edge areas around lakes and sections of the river. It was apparent that neither slope nor aspect, which on the floodplain and around water bodies would be level, would separate the railroad yard from water areas. A rule based on elevation was developed that grouped all pixels associated with the class as being water if they were above 1100 feet. The railroad yard was clearly situated outside this group but so were water pixels along the river which were located below the yard at a lower elevation. Again, if a rule could be designed that would focus on a certain portion of the data set, it might be possible to handle conditions like the railroad yard.

SUMMARY

Table 1 summarizes the sorting rules for the four classes and shows how many pixels were classified as water versus non-water conditions. If all the pixels associated with the four classes were identified as water prior to postclassification sorting, approximately 59.9 percent of them would be inaccurately classified. Most likely one or two of the spectral classes would have been identified as another land cover but then some obvious water areas would be misclassified. The sorting process greatly improved classification accuracy; however, this approach consumed a considerable amount of time in working out the individual rules for each spectral class. Also, better sorting procedures are needed which would allow: 1) rules to apply under certain conditions to specific areas within the data set and not the entire study area, 2) rules to be established which can separate spectral classes into more than two land covers, and 3) rules to change their parameter limits as the sorting process is occurring.

ACKNOWLEDGEMENTS

This research was conducted using the Cornell National Supercomputer Facility, a resource of the Cornell Theory Center, which is funded in part by the National Science Foundation, New York State, the IBM Corporation and members of the Center's Corporate Research Institute.

REFERENCES

Baumann, P. R. 1985, "Employing Topographic Data To Rectify Certain False Water Readings In Landsat Multi-Spectral Scanner Land Cover Classifications," Proceedings of the Middle States Division of the Association of American Geographers and the New England-St. Lawrence Valley Geographical Society, pp. 2-10.

Hutchinson, C. F. 1982, "Techniques for Combining Landsat and Ancillary Data for Digital Classification Improvement," Photogrammetric Engineering and Remote Sensing, Vol. XLVIII, No. 1, pp. 123-130.

NASA/Earth Resources Laboratory, 1980, ELAS: Earth Resources Laboratory Applications Software, National Space Technology Laboratories, Mississippi.