George R. Harker
Western Illinois University
Macomb, IL 61455
J. W. Rouse, Jr.
Texas A&M University
College Station, TX 77843

Flood-Plain Delineation Using Multispectral Data Analysis

Flood-plain boundaries determined from multispectral scanner data correlated well with those produced from established techniques.

Abstract: *The delineation of flood plains with the use of current techniques is both costly and time-consuming. This paper explores the application of a remote sensing technique that may permit the determination of flood-plain areas without the extensive work associated with existing techniques.*

In this study, multispectral scanner data were simulated by utilizing the density differences in a color-infrared transparency for a section of the Navasota River in Texas. The transparency was taken from a low-flying aircraft and covered an area approximating a square mile. The simulated data were processed by an automatic classification technique previously developed in the remote sensing field.

The technique used involves the application of the maximum likelihood rule in order to categorize the data being processed. An attempt was made to distinguish between areas known to be in the flood plain and those without. A reasonable correlation was found between boundaries based on computer processed multispectral data and those produced by techniques currently in use.

Introduction

The current need for improved methods of delineation and mapping of flood plains over large areas has been clearly identified by Sollers *et al.*[1], Harker[2], and others. The federal government spends approximately $500 million each year on federal protection projects, yet the losses due to flooding exceed $1 billion annually.[3] The continuing effort of the U.S. Corps of Engineers to map the flood plains of the nation's rivers utilizes proven methods, often including the use of aerial photography, but the process is slow and costly. One such activity by the State of Connecticut, using an approach similar to that of the Corps of Engineers, produced a comprehensive flood plain map of 46.6 miles of river at a cost of approximately $5,000 per mile.[4] Although this figure is thought to be above the average cost, the fact remains that the cost is sufficiently high that a more cost-effective approach is needed if the flood plains of the nation's rivers are to be mapped within a reasonable number of years.

One possible method of increasing the efficiency of flood-plain mapping is to employ the well-established computer techniques for analysis of multispectral reflectance data. Sollers *et al.*,[1] in a preliminary study of

Photogrammetric Engineering and Remote Sensing,
Vol. 43, No. 1, January 1977, pp. 81-87.

applicable methodology, suggested that such an approach is feasible. This paper presents results of an attempt to delineate the flood plain for a test site along a Texas river by using the maximum likelihood classification scheme.[2] The computer-generated flood plain is compared with (1) The Corps of Engineers 100-year flood boundary, (2) the boundary defined by the flood-plain soils as mapped by the Soil Conservation Service, (3) the U.S. Geological Survey (USGS) topographic lines in the region, and (4) a flood-plain boundary based on interpretation of aerial photography.

The Experimental Approach

The experimental approach consists of the following phases:

(1) Documentation of the test site and the collection of ground observations;

(2) Acquisition of color-infrared aerial photography;

(3) Conversion of aerial photography to density values which simulate multispectral scanner data and are in the appropriate format for computer analysis;

(4) Data processing, which includes a variety of programs used to manipulate the format of the data as well as the application of automatic classification techniques previously developed in the remote sensing field; and

(5) Analysis of computer-generated flood-plain boundaries to examine their relationship to other flood-plain boundaries.

The site chosen for the experiment is located north and east of College Station, Texas. It consists of a portion of the Navasota River Valley which exists in a relatively natural state. Site selection was made with the knowledge that a great deal of ground truth information helpful to the study was already available. The Corps of Engineers had been studying the area for a proposed dam and had compiled a comprehensive study of the area which included the identification of flood zones associated with different discharge rates. The Soil Conservation Service had also been active in the area, having prepared a comprehensive soil survey for the county which contained the test area.

The flood-plain areas of the Navasota River are composed of typical bottom land and terrace fluvial sediments consisting of various mixtures of sand, silt, gravel, and clay. The portion within the study area falls in the Blackland Prairie belt of the West Coastal Plain physiographic section. In the study area, the flood plain is reasonably well defined and the river bed follows a meandering course.

The study area was photographed by a RC-8 Wild camera with a 500 mμ av 22 filter. The exposure was made at an altitude of 3,500 feet on aerial color infrared film #8443. Atmospheric conditions during the flight ranged from a light haze to clear with scattered clouds.

One color transparency (Plate 1) from the air photo series was chosen for conversion to density values. The test area reported here covered approximately a square mile. Since the 9 by 9-inch format of the transparency was too large for the available densitometer equipment, it has to be measured in six sections. In all, the total transparency was scanned four times using (1) no filter, (2) red filter Wratten No. 92, (3) green filter Wratten No. 93, and (4) blue filter Wratten No. 94. The scan cell size was such that the portion of the transparency measured corresponded to an area on the ground, rectangular in shape, 2 metres by 9 metres.

Density values were computed and transferred to magnetic tapes for computer analysis. Computer gray scale printouts of the entire frame were generated. These maps were used for selection of training areas. Areas known to be within the flood plain were selected. Some of these training areas were tree-covered and some were covered by low herbaceous vegetation. Similar selections were made in the areas known to be beyond the flood plain.

By using the statistics developed from the training sets chosen to characterize areas within and without the flood plain, the data were processed according to the maximum likelihood classification rule[5]. By using this approach, the data representative of each scan cell was processed and categorized as belonging to one of four categories. The categories (Plate 1) were identified as:

Class 1—low vegetation not in the flood plain
Class 2—low vegetation in the flood plain
Class 3—trees characterizing the flood plain
Class 4—trees characterizing the uplands

On the resultant computer printouts, each class is characterized by the same numeric symbol as its category. An additional category identified by "0" is used for all other entries that could not be categorized in one of the four defined classes.

The basic data consisted of density values essentially corresponding to three different wavelength bands and a simulation of reflected white light. The subsequent data analysis examined all possible wavelength

PLATE 1. Portion of Navasota River Valley showing training set locations.

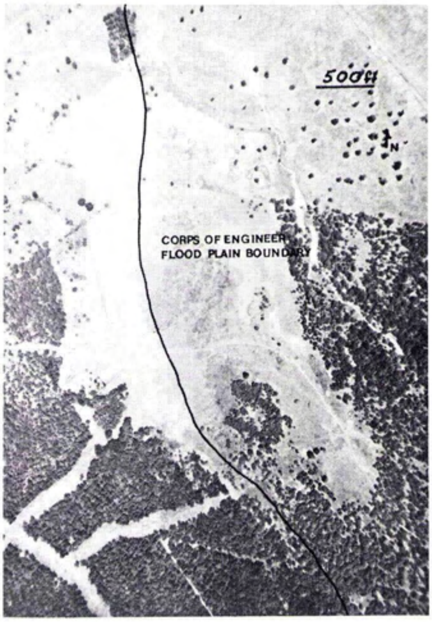

FIG. 1. Corps of Engineer—100-year flood boundary.

band combinations, as well as the use of the simulated white light band. The data analysis consisted in generating a computer printout of Class 2 (which was determined to provide the best correlation with the flood plain boundaries) for each combination of wavelength bands. On the resulting computer printout, a boundary was drawn in such a manner that most of the points categorized as Class 2 would fall within the boundary. A matter of personal judgment dictated the exclusion or inclusion of certain points within the boundary depending on the relative density of the points being classified. This boundary was compared against other boundaries generally considered to depict the flood plain under different circumstances. The boundaries used in the comparison consisted of (1) the Corps of Engineers boundary associated with a 100-year flood and a flood discharge of 50,000 cfs (Figure 1); (2) a boundary based on a Soil Conservation Service soil map[6]; (3) the U.S. Geological Survey contour line; and (4) a boundary based on photo-interpretation of a set of stereo pairs.

RESULTS

DATA ANALYSIS—CORPS OF ENGINEER BOUNDARY

All three wavelength bands, (Table 1; band 2-infrared, band 3-red, and band 4-green), the computer-based boundary, and the Corps 100-year flood boundary exhibited the relationship summarized in Table 2. The two boundaries come to within 30 feet (9.2 m) at one point and diverge to 640 feet (194 m) at another. Half of the values are closer than 200 feet (61 m), with the greatest divergence representing 10 percent of the overall distance from the Corps' boundary to the river. The coefficient of correlation between the boundaries was 0.8708, using a sample composed of 12 measurements from the river channel to the respective boundaries.

By utitilizing the different combinations of two wavelength bands, the best results were obtained with the infrared and red bands. The linear distance between the computer-based boundary and that of the Corps' for this band combination ranged from one foot to a maximum of 270 feet (83 m). The coefficient of correlation was 0.9394.

Working with a single wavelength band, the best results were obtained with the reflected infrared portion (band 2) of the spectrum. The resulting distances ranged from 30 feet (9.2 m) to 230 feet (70 m). The coefficient of correlation between the two boundaries was 0.9215.

DATA ANALYSIS—BOUNDARY SUGGESTED BY FLOOD-PLAIN SOILS

The boundary suggested by the flood-plain soil of the Navasota River was compared also with that suggested by the computer. Results from three wavelength bands yielded a correlation of 0.1532. The linear relationship between the two boundaries ranged from 190 feet (57 m) to 770 feet (231 m).

Using two wavelength bands, a correlation of 0.1595 was found; with the reflected infrared as a single band, the result was 0.1423. As Table 2 indicates, the correlations in all categories were similarly low.

TABLE 1. MEANING OF BAND DESIGNATION

Wavelength Band #	Color	Approx. Wavelength
1	White	500 to 800 $m\mu$
2	Infrared	700 to 800 $m\mu$
3	Red	600 to 700 $m\mu$
4	Green	500 to 600 $m\mu$

TABLE 2. COMPUTER-GENERATED BOUNDARY VERSUS CONVENTIONAL TECHNIQUES

Wavelength Band	Corps		S.C.S.		Air Photography		U.S.G.S.	
	Correlation	Linear Difference	C.	L.D.	C.	L.D.	C.	L.D.
2 Infrared 3 Red 4 Green	0.8708	30 to 640ft	0.1532	190 to 770	0.6759	130 to 940	0.8108	0 to 630
2 3	0.9394	1 to 270	0.1595	190 to 830	0.7883	100 to 510	0.8629	0 to 300
3 4	0.8621	1 to 500	0.1950	140 to 800	0.6718	130 to 780	0.8005	0 to 600
2 4	0.8585	1 to 500	0.2159	100 to 830	0.6789	130 to 700	0.8010	0 to 600
1 White	0.7524	1 to 910	0.3014	0 to 810	0.4186	100 to 1220	0.7530	10 to 900
2	0.9215	30 to 230	0.1423	140 to 850	0.7629	20 to 440	0.8421	0 to 330
3	0.5213	0 to 710	−0.1413	240 to 1180	0.2632	100 to 1010	0.4441	10 to 710
4	0.8427	30 to 540	0.2246	140 to 770	0.6322	130 to 810	0.7934	20 to 640

DATA ANALYSIS—U.S. GEOLOGICAL SURVEY CONTOUR LINES AS BOUNDARY

An analysis of the computer-developed boundary was made against a 235-foot contour line on a U.S. Geological Survey contour map. This particular contour was chosen because it fell in the vicinity of the boundary suggested by the computer.

By utilizing three wavelength bands, the coefficient of correlation was 0.8108. The linear distance between the two boundaries ranged from zero to 630 feet (192 m). With the use of two wavelength bands, infrared and red, for example, the best result obtained had a correlation of 0.8629. The linear distances between boundaries ranged from zero to 370 feet (114 m). Comparable results were found with other combinations of two bands.

By employing a single wavelength band, the best result again was obtained with reflected infrared which yielded a coefficient of correlation of 0.8421.

DATA ANALYSIS—BOUNDARY BASED ON PHOTO INTERPRETATION

Working with a mirror stereoscope and a pair of color-infrared transparencies, a boundary was developed by using apparent differences in the vegetative cover, color of soil and elevation. This boundary then was compared with the computer-developed boundary. With three wavelength bands the coefficient of correlation was 0.6759. The linear distances ranged from 130 to 940 feet. With two bands, the best correlation (0.7883) resulted when infrared and red bands were used. As a single wavelength band, band 2 again yielded the best correlation with a value of 0.7629. A comparison of the computer-developed boundary and the several other flood-plain boundaries is shown in Figure 2.

SUMMARY OF FINDINGS

(1) In areas where ground cover was of a low herbaceous variety, a boundary approximating the flood-plain boundary could be discerned on computer printouts derived from multivariate data analysis.

(2) A successful discrimination between trees within or without the flood plain could not be made in this study of single-date data.

(3) A strong correlation exists between the Corps of Engineers' 100-year frequency flood boundary and the computer-developed boundary.

(4) A poor correlation between the computer-developed boundaries and soil survey information was observed. However, ground truth information and surface soil color and texture suggested that the Soil Conservation Service boundary is in error for this site.

(5) The use of two wavelength bands yielded results essentially the same as if three wavelength bands were used.

(6) The most successful combination of

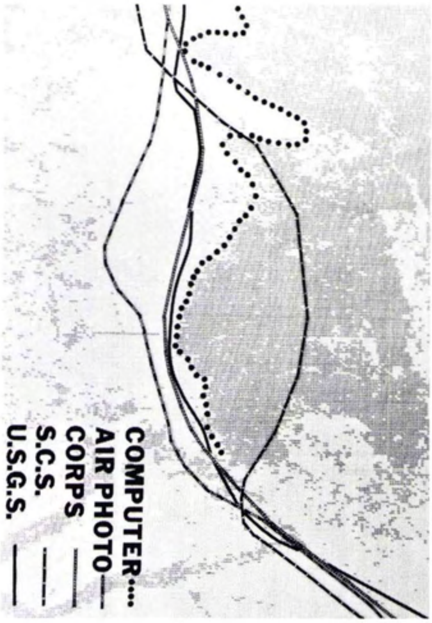

FIG. 2. Computer-Developed Boundary Using Green, Red, and Infrared Bands.

wavelength bands employed the near infrared and red portions of the electromagnetic spectrum.

(7) A single wavelength band utilizing reflected infrared radiation yielded a boundary which also exhibited a strong correlation to the Corps of Engineers' boundary.

This experiment appears to show that a meaningful flood-plain boundary can be identified by using automatic classification techniques on multispectral data in areas where the vegetational cover of the flood plain does not consist of dense trees and the surface soil is visibly different from upland soil.

COMMENT

Flood plains by nature are a very intricate and variable phenomena not readily amenable to precise calibration. With this in mind, it has never been reasonable to view the remote sensing technique explored here as an end-all, be-all method of flood-plain delineation. Multispectral analysis cannot take the place of detailed ground observation and study. Its role must be in supplementing ground observations.

While the responsible government agencies must continue producing flood information in much the same way as they have in the past, there are useful ways in which multispectral analysis could be used. The Corps of Engineers, for example, could produce preliminary maps suggesting areas possibly subject to flooding based on results of the multispectral analysis technique. A possible advantage of the multispectral map is that it is based on observable natural phenomena which indicate areas flooded often enough to produce detectable changes in the environment. In many cases, it may be more useful to recognize these areas than to know the 100-year flood boundary.

ACKNOWLEDGMENT

This study was supported by National Aeronautics and Space Administration NASA Grant NGL 44-001-001.

REFERENCES

1. Sollers, S. C., G. W. Peterson, D. L. Henninger, and F. Y. Borden. "The Use of Remote Sensing and Natural Indicators to Delineate Flood Plains—Preliminary Findings," *Proceedings of the Ninth Symposium on Remote Sensing of Environment,* Ann Arbor, The University of Michigan, p. 667, 1974.
2. Harker, George R., *The Delineation of Flood Plains Using Automatically Processed Multispectral Data,* Technical Report RSC-60, College Station, Texas, Remote Sensing Center, Texas A&M University, August 1974.
3. The Task Force on Federal Flood Control Policy, *A Unified National Program for Managing Flood Losses,* Washington, United States Government Printing Office, 1966.
4. Pelletier, Charles J., *Connecticut's Program for Establishing Stream Encroachment Limits,* Connecticut, September 13, 1960.
5. White, D. A., J. W. Rouse, Jr. and J. A. Schell, *Analysis of Simulated Multispectral Data from Earth Resources Satellites,* Technical Report RSC-35, College Station, Texas, Remote Sensing Center, Texas A&M University, 1971.
6. United States Department of Agriculture, Soil Conservation Service, *Soil Survey, Brazos County Texas,* Series 1951, No. 1, Washington, D.C., United States Government Printing Office, 1958.