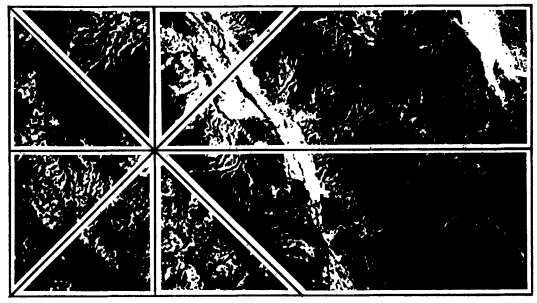
For geographic applications, satellite image data must be combined with aerial photos, census data, and existing maps. Integrating the various data forms is a complex, multistep operation.

Pictorial Information Processing of Landsat Data for Geographic Analysis



Albert L. Zobrist
Jet Propulsion Laboratory

Georgy Nagy
University of Nebraska

Geographic analysis can be traced to ancient roots: the establishment of the Roman road nework, the Egyptian census (used for levying taxes, of course), and the layout of the Sumerian canal system. However, not until the invention of the balloon in 1789 and of photography in 1839 could large areas of the earth be pictured. The Civil War saw the first tactical application of airborne photography, which blossomed during the first half of the twentieth century into the systematic coverage of entire countries for cartographic purposes. The Apollo, Gemini, and Skylab flights dramatically increased the area that could be covered in a single picture, while the introduction in the late fifties of the multispectral scanner and of sideways-looking radar extended pictorial data collection to the invisible portions of the electromagnetic spectrum.

Satellite image acquisition

The systematic application of remote sensing and digital image processing to geographic analysis began with the launching of the ERTS-A earth resources satellite, the first of the Landsat series currently in orbit. (The resolution of the Nimbus, Tyros, and ITOS meteorological satellites was too low for most geographic purposes.) ERTS-A, subsequently renamed Landsat 1, was preceded by five years of intense planning led by NASA. Responsibility for satellite operation and image acquisition and distribution is gradually being transferred to the National Oceanic and Atmospheric Agency of the Department of Commerce, with NASA retaining major responsibility for advanced development and research. Currently, Landsat products may be obtained (for a fee) from the Earth Resources Observation System—EROS—Data Center of the US Geological Survey. Private ownership and operation of the US program, under federal government regulation, is the ultimate objective.

Landsat 3, the more recently launched of the two currently operational earth resources satellites, is in a near-polar, sun-synchronous orbit—that is, it revisits each location when the sun is at about the same angle—at a mean altitude of 920 kilometers. It circles the earth every 103 minutes, and returns to the same orbit every 18 days. With two satellites, therefore, repetitive coverage every nine days is obtainable, in principle. The position and attitude of the satellite are tracked continuously; this tracking data is the input for the registration procedure used to find the approximate location of the satellite images. The location of selected scenes can then be determined more precisely by matching portions of the images to stored images whose locations are known.

The current Landsats have two independent image acquisition systems. One is provided by a four- or five-channel multispectral radiometer—MSS—and two or three panchromatic return-beam vidicon—RBV—cameras. The multispectral scanners use an oscillating mirror to obtain six rows of 3240 samples, at sampling intervals of 9.94 milliseconds, on each mirror sweep. The motion of the satellite produces a frame consisting of 2400 scan lines. A single MSS frame represents a 185 × 170-kilometer area on the surface of the earth. The intensity of radiation is quantized in 64 intervals, corresponding to one 6-bit byte. Each four-band picture thus requires about 30,000,000 bytes, with each one-byte pixel corresponding to a 57 × 80-meter area. The spectral windows are in the 0.5 to 1.1-micrometer range, with two bands in the visible region and two in the near-infrared. The fifth channel on Landsat 3 is in the thermal-infrared region of 10.4 to 12.6 micrometers. The spectral channels were selected to correspond to transparent windows of the atmosphere, which are also useful for characterizing

0018-9162/81/1100-0034$00.75 © 1981 IEEE

land cover types for various applications. The MSS on Landsat 3 ceased operation in 1981.

The format of the Landsat 3 return-beam vidicons is 5375 × 4125 pixels. The two cameras each view a 99 × 99-kilometer segment of the earth, called a subscene. Four subscences are combined to form a scene approximately as big as a single MSS frame. One RBV picture element corresponds to a 19 × 19-meter area on earth.

The image information is either transmitted directly to earth (many countries, most recently the People's Republic of China, are considering or already operating ground receiving stations) or, when the satellite is out of range of the ground receiver, stored on board on magnetic tape for subsequent transmission. On-board storage has proved unreliable and will be obviated when the Tracking and Data Relay Satellite System is installed. Each geosynchronous TDRSS satellite will have very high bandwidth (up to 300 megabits/second) transponders and will relay the Landsat data to a single station at White Sands, N.M. From White Sands, the data will be retransmitted to Goddard via Domsat and, after preprocessing at Goddard, transmitted via Domsat to the EROS Data Center for product generation and archiving.

Landsat image processing, archiving, and distribution

The combined Goddard and EROS facilities for processing Landsat data must surely be the largest—or at least the highest-volume—civilian image processing facility in existence. During the 1980 fiscal year, for example, about 26,000 Landsat scenes were received and processed at the EROS Data Center. The sales volume in fiscal 1980 was $2,400,000—128,000 scenes were sold in photographic form (down from a 1976 peak of 297,000 frames) and about 4100 scenes in digital form. The federal government accounted for 26 percent of the purchases, 3 percent were sold to state/local government, 25 percent to industrial users, 9 percent to academic users, 4 percent to individuals, and a startling 42 percent to non-US customers.

The MSS and RBV images received from Landsat are system-corrected at the NASA/Goddard Image Processing Facility. Radiometric correction removes sensor and digitizer anomalies. Geometric correction is necessary to compensate for satellite attitude changes, slewing introduced by satellite motion during each scan, and the effects of the angle subtended by the satellite's field of view. The characteristics of the corrected images are given in the Landsat Data User's Handbook.[1] The images are resampled to a Hotine Oblique Mercator projection. (Resampling means that the gray value assigned to each point in the target image is a function—usually a weighted average—of the points corresponding to the target point in the source image.) The resampled MSS frames contain 3548 × 2983 pixels, while the resampled RBV subscenes contain 5322 × 5322 pixels.

The standard error in the system-corrected data is about 160 meters in each direction. Linear least-square analysis indicates that using linear correction factors to remove certain errors in scale, earth rotation effect, and aspect ratio could reduce the standard error to 50 meters (less than one pixel), which would meet the National Map Standards for 1:250,000 scale. The current products are within the required accuracy for 1:1,000,000 scale.

Haze removal, an optional process, compensates for atmospheric scatter. Another optional process, contrast stretching, allows the use of the full dynamic range of the system for low-contrast images. Edge enhancement, which exaggerates the difference between a given pixel and a user-specified neighborhood (kernel), is also available. Users who prefer to avoid the loss of discrimination due to resampling may request only radiometrically corrected data; this results in some loss in positional accuracy.

In summary, the systematic "bulk" correction of all Landsat image data takes place at the NASA/Goddard Image Processing Facility. IPF receives ephemeris data from the tracking stations, including postion, pitch, yaw, and altitude as a function of time. The data from the satellite itself contains radiometric and calibration data: readings from the MSS calibration wedges, RBV calibration lamp readings, and reseau marks from the RBV vidicons. This information is compiled with prelaunch calibration measurements, characteristics of the transmission channels, and cartographic constants describing the geoid to allow precise reconstruction of each scene. Both fully corrected images and partially corrected images accompanied by the necessary correction factors are recorded on high-density (20,000-bpi) magnetic tape and transmitted to the EROS data center in Sioux Falls.

At the EROS Image Data Processing Facility the data is archived on high-density tapes. Archival data is generated at a rate roughly corresponding to the output of 200,000 keypunch operators working day and night. The image data is distributed on request in the form of either high-resolution (laser-recorded) photographic products or as standard digital computer tapes. A digital browse file of selected catalog information is also available through dial-up terminals via Telenet. Special processing of both film and digital data is undertaken at user request on one of the three major image-processing computer systems. User training is available through special workshops designed to meet particular needs.

Landsat and geographical analysis

The availability of Landsat image data has brought about major changes in geographical analysis. While pilot studies in the mid-sixties attempted to demonstrate what can be done with image data alone, it has become increasingly clear that most important applications require the combination of image data with many other sources of geographical information. Some pictorial information systems consider pictures as data base elements. Yet picture content is inaccessible (although preprocessing may have extracted descriptors from the content). Pictorial information processing that deals with the *content* of images seems more appropriate for most Landsat applications.

A primary advantage of Landsat coverage is the ability to quickly obtain time-varying information about large areas. Examples include sea ice, snow cover distribution for predicting water availability, changes in vegetative cover, urban spread, water quality, and shoreline erosion. Equally important is cartographic input for detailed mapping of the large areas of the globe hitherto mapped only at larger scales. Secondary indicators permit drawing conclusions about demographic factors, wildlife population, agricultural practices, and land management. For all of these applications a crucial ingredient is the combination of satellite image data with aerial photography, census data, existing maps, small-sample data, and so on. This generally requires transformation of the various forms of data into a compatible format and extraction of the significant items.

The resolution of the satellite image is constrained not by any inherent limitation in the instrumentation but by the immense volume of data in even relatively small areas of the image. Because of the large amounts of data, the efficiency of the transformation and data processing algorithms is a paramount consideration. Another requirement introduced by the sheer mass of the data is built-in checking and verification procedures to reduce the adverse effects of noise, artifacts, and inconsistencies of various sorts. It is simply not feasible to visually inspect and manually correct the images on a pixel-by-pixel basis. Geographic analysis may, in fact, be viewed primarily as a data-reduction process; a vast amount of data is reduced to a few elements that can be assimilated by human decision-makers. At the same time, there is an increasingly wide range of applications of image processing technology to geographic analysis. Systems originally designed to do pattern recognition and image manipulation are being expanded to perform the full gamut of operations needed for data integration and spatial analysis. As with any new technology, there are technical, economic, and conceptual barriers to wider use.

Two particular problems are worthy of a concentrated examination here: first, the need for algorithms to enable or simplify the basic operations of geographic analysis, and second, the complexity of applications in terms of the sequence of operations performed. In the projects described below, we attempt to illustrate this complexity by simply listing the major steps—without much explanation or detail. The letter codes next to each step indicate that one or more of the following aspects are significant in that step:

- A—algorithmic problems,
- C—high computational expense,
- G—vector or polygon data structure,
- M—manual operations,
- R—raster data structure, and
- T—tabular data structure.

Editing and validation steps—generally ignored below—are, of course, significant to real data handling.

Desert conservation. As part of the planning activity for the 25-million-acre California Desert Conservation Area, the Bureau of Land Management needed a survey of vegetation and animal forage in a form that could be integrated with other elements of the plan. Multispectral classification of Landsat imagery was chosen as the best means of obtaining this survey, considering time and money constraints. The goal was to integrate digital terrain data, such as elevation, slope, and aspect, that have a strong effect on vegetation, with ground information from maps, surveys, and photointerpretation to aid in classification. The output was required in 1° × 1° quadrangles in Lambert Conformal Conic projection and in tabular form, categorized by grazing allotments, land ownership, and other types of boundaries.

The procedure for vegetation and forage classification consisted of 24 major steps:

1. Obtain Landsat raw data (10 frames).
2. Obtain map control points from conventional maps (M).
3. Convert the map coordinates to Lambert Conformal Conic grid (G).
4. Find the coordinates in the Landsat frames (G).
5. Find edge-match control points for all overlapping Landsat frames (A, C, R).
6. Adjust edge-match control points according to a distortion model determined by the map control points (A, G).
7. Perform geometric correction and map projection of the Landsat frames according to the control points (A, C, R).
8. Perform brightness adjustment of the Landsat frames according to differences noted in the edge-matching control points (C, R).
9. Trim edges of the Landsat frames to remove bad data (R).
10. Mosaic the 10 frames together to obtain a single, large 7400 × 7500 image (C, R).
11. Repeat steps 7-10 for three more spectral bands of Landsat.
12. Section Landsat mosaic into 1° quadranges (A, R, G).
13. Perform geometric correction on digital terrain file to obtain 1° quadrangles in the same grid map projection (A, C, R).
14. Classify land cover by unsupervised initial clustering (1993 clusters) with supervised pooling (25 final classes) by a discipline scientist (A, C, T).
15. Test the statistics on 12 areas of 400,000 acres each (R, G, M, T). Refine the statistics (M, T).
16. Extend the classification to the entire California Desert Conservation Area (A, C, R).
17. Coordinate digitize various boundary files such as land ownership (G).
18. Change the digitizer coordinates to Lambert Conformal Conic grid (G).
19. Change the boundary file from vector to raster format (A, C, R, G).
20. Identify the polygons determined by the boundaries (A, C, R).
21. Perform polygon overlay of classification files with boundary files (A, C, R, T). This yields a table of acreages of classes per district or parcel (T).
22. Perform data management operations such as

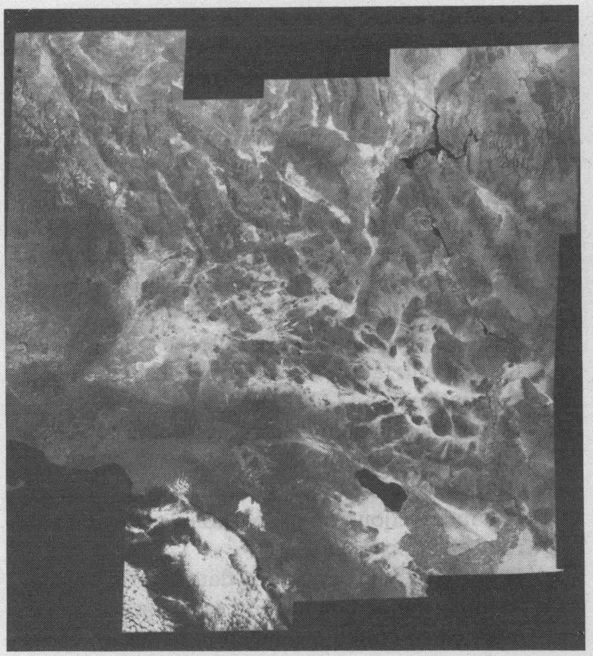


Figure 1. California desert mosaic of 12 Landsat scenes. The image size is 7400 × 7500 for each spectral band. The near-infrared band is shown here.

Figure 2. One degree quadrangle extracted from the desert mosaic with grazing allotments superimposed.

sort, merge, merge correlate, aggregate, and cross-tabulate on the results of polygon overlays (T).

23. Report the tables by printer listing or magnetic tape (T).
24. Report the image products by photographic image playback or magnetic tape (R).

The results of the California Desert Task were used as a primary input to the Desert Master Plan for range forage allocation for livestock and wild burros. The work was performed within a two-year deadline, primarily in digital format to facilitate 10-year updates mandated by Congress. Details of the work and utilization of the results can be found in McLeod and Johnson[2] and in the Bureau of Land Management's report,[3] respectively. Figure 1 shows the California desert mosaic with two additional Landsat frames added. Figure 2 shows a 1° quadrangle with grazing allotments superimposed digitally. Figure 3 illustrates the tabulation of biomass by grazing allotment.

Air pollution study. Several agencies in the Portland, Oregon, area have been working under the general coordination of the Pacific Northwest Regional Commission to determine the utility of Landsat digital images for urban applications. A map of land use from Landsat was prepared at the USGS EROS Data Center, with assistance from the NASA Ames Research Center. The application described here used the map to allocate population from the 1973 census update to traffic zones and two-kilometer grid cells. The grid-cell population was combined with air pollution data from the Oregon Department of Environmental Quality to obtain a health impact map.

The procedure for the health impact interpretation consisted of 14 major steps:

1. Find control points in boundary files (M).
2. Find the same control points in the Landsat image (G).
3. Calculate a distortion surface from the control points (A, G).

BUREAU OF LAND MANAGEMENT
CALIFORNIA DESERT CONSERVATION AREA

STANDING BIOMASS AND RENEWABLE FORAGE
PRODUCTION BY LAND OWNERSHIP
METRIC TONS PER GRAZING ALLOTMENT

GRAZING ALLOTMENT	PUBLIC STANDING BIOMASS	PRIVATE STANDING BIOMASS	TOTAL STANDING BIOMASS	PUBLIC RENEWABLE FORAGE PRODUCTION	PRIVATE RENEWABLE FORAGE PRODUCTION	COMBINED RENEWABLE FORAGE PRODUCTION	TOTAL PIXELS
1	573.43	0.0	573.43	13.49	0.0	13.49	208
2	21842.43	550.86	22393.28	599.51	12.72	612.22	8608
3	68366.00	7194.39	75560.37	1955.37	184.23	2139.60	27095
4	141637.75	4849.91	146487.62	3490.77	146.66	3637.42	56767
5	33470.54	4695.17	38165.71	809.38	112.47	921.85	15311
6	22196.81	1505.12	23701.92	593.48	40.64	634.13	8745
7	11296.34	1665.53	12961.87	344.33	52.25	396.58	5334
8	139427.94	0.0	139427.94	3323.06	0.0	3323.06	38192
9	314165.19	25231.98	339397.06	8476.55	573.11	9049.66	135714
10	19509.17	3438.52	22947.70	520.31	87.37	607.68	9023
11	56630.12	2442.11	59072.22	1319.12	54.01	1373.13	23156
12	7041.93	465.47	7507.39	79.66	4.50	84.16	2002
13	7231.32	1413.39	8644.71	144.09	27.45	171.55	4131
14	500826.56	387987.69	888813.56	7816.42	6845.37	14661.80	387870
15	22378.24	32479.74	54857.96	463.68	669.81	1133.49	27028
16	3010.14	656.75	3666.89	8.44	2.60	11.04	241
17	652944.62	465476.56	1118423.00	4447.50	1819.20	6266.70	136732
18	901.88	99.76	1001.64	13.33	1.43	14.75	269
19	432873.44	22079.96	454953.44	2348.64	207.47	2556.11	61197
20	58174.70	2036.44	60211.12	1002.19	27.52	1029.70	32161
21	311833.56	373846.00	685680.12	952.24	974.11	1926.35	53176
22	5839.24	4806.54	10645.79	121.75	99.02	220.77	5287
23	30181.32	10677.65	40858.97	622.04	221.52	843.55	20439
24	12276.98	1360.91	13637.89	189.13	22.40	211.52	7060
25	16345.71	450.46	16796.18	160.05	3.42	163.47	3537
26	80726.37	79305.50	160031.87	1576.94	1561.15	3138.08	78288
27	6678.58	1713.64	8392.22	138.62	35.30	173.92	7057
28	30041.36	0.0	30041.36	349.34	0.0	349.34	10762
29	62567.74	1477.66	64045.37	1087.17	26.08	1113.25	27943
30	32510.35	9718.66	42228.99	586.66	169.86	756.52	19704
31	77149.69	1275.94	78425.56	1391.06	24.19	1415.24	40658
32	211109.50	69559.69	280669.12	3675.54	1242.76	4918.30	136994
33	205180.75	88035.44	293215.69	3659.83	1563.19	5223.02	144672
34	449.92	5.95	455.88	8.11	0.12	8.22	225
35	43849.18	9625.40	53474.59	539.71	122.50	662.21	20087
36	79676.75	15420.87	95097.62	495.46	52.30	547.76	14661
37	258320.75	338278.37	596598.50	201.51	258.65	460.16	22731
38	86641.12	31552.29	118193.50	1289.35	470.86	1760.21	71582
39	95094.87	5760.78	100855.62	1316.41	80.12	1396.53	43844
40	214672.31	67114.81	281787.12	3378.36	1099.03	4477.39	125985
41	65707.19	2502.88	68210.06	1076.81	41.92	1118.73	21053
42	51492.22	14042.72	65534.94	686.14	200.17	886.31	44491
43	326987.75	63653.74	390641.37	5248.66	1002.89	6251.55	195874
44	51885.25	5821.68	57706.96	728.26	95.89	824.15	11702
45	211070.50	11355.71	222426.25	2735.20	133.79	2868.99	74886
46	367052.50	25894.94	392947.69	5952.80	512.28	6465.07	158421
47	582347.25	194143.69	776491.00	10028.17	3738.23	13766.41	286216
48	47954.75	5104.79	53059.52	675.98	77.37	753.36	23048
49	1888.57	0.0	1888.57	31.29	0.0	31.29	422
50	377574.37	22397.44	399971.56	5709.79	350.49	6060.28	192224
51	338445.81	17272.24	355718.12	4386.19	228.20	4614.39	159221
52	13587.26	1791.88	15379.15	160.25	19.10	179.35	12253
53	14448.06	0.0	14448.06	181.34	0.0	181.34	7523
109	424994.00	15300.44	440294.12	11313.63	457.66	11771.29	134064
111	22050.65	720.26	22770.91	632.64	14.95	647.58	8860
129	82658.62	0.0	82658.62	1288.87	0.0	1288.87	34597
139	14798.37	1436.55	16234.92	188.98	17.98	206.96	6769
146	32937.52	2136.58	35074.06	660.68	45.95	706.63	13124
200	336593.37	682.56	337275.87	8324.46	8.91	8333.37	214519
201	2325040.00	572.30	2325612.00	65177.15	15.08	65192.24	1202333
202	1459711.00	348.84	1460060.00	22038.81	6.67	22045.48	651984
204	261293.37	83147.69	344440.81	5249.59	1612.54	6862.13	182515
205	9223.51	702.88	9926.39	117.80	13.49	131.29	3208
206	1133.05	50.47	1183.52	8.18	0.24	8.42	1036
207	0.0	1056.43	1056.43	0.0	4.31	4.31	1383
210	1569.93	298.43	1868.37	30.33	5.37	35.70	873
220	906498.50	4264.30	910762.75	12805.93	58.10	12864.03	348069
221	588128.25	3886.41	592014.31	8042.64	50.12	8092.75	368124
222	217.99	4303.91	4521.90	2.27	57.23	59.50	1325
223	130391.25	304520.44	434910.94	976.81	1851.89	2828.70	354402
224	397890.25	7935.29	405825.50	1760.04	32.12	1792.16	287700
225	312.80	0.0	312.80	1.58	0.0	1.58	277
254	11264749.0	8282591.00	19547392.0	159066.12	83480.12	242546.81	9095850

Figure 3. Partial tabulation of biomass for analysis of grazing capacity.

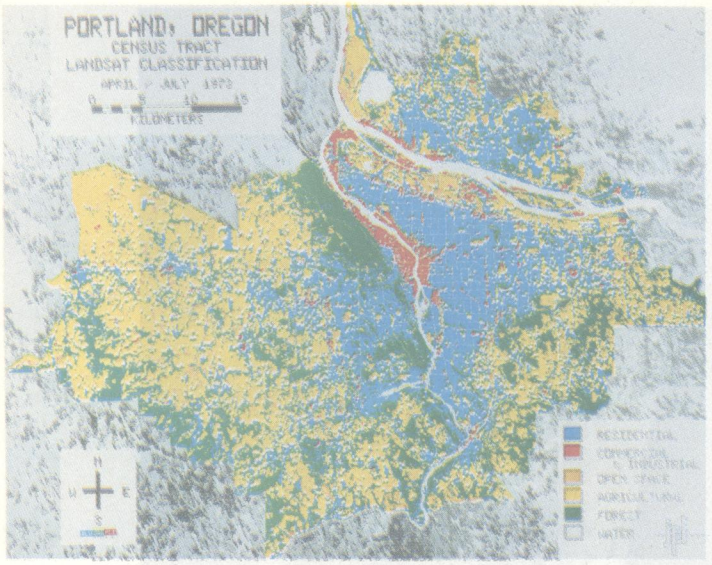


Figure 4. Land use map of Portland, Oregon, derived from Landsat data by digital classification with census-tract boundaries superimposed.

PORTLAND, OREGON
POLLUTION BY GRID CELL

GRID CELL	POPULATION	POLL VALUE	POLL INDEX	COVERAGE
2,10	116	28	3236	92
3,10	147	29	4266	100
3,11	112	30	3363	100
3,13	1140	28	31911	91
4, 9	100	29	2912	100
4,10	135	29	3929	100
4,11	152	29	4397	100
4,12	294	29	8513	100
4,13	1155	29	33490	100
4,14	1279	29	37098	96
4,17	79	28	2203	
5, 9	132	29	3832	
5,10	199	28	5565	
5,11	163	28		
5,12	118	28		
5,13	766			
5,14	827			
5,15	479			100
5,16				98
5,17				100
			80775	100
	1160	31	34708	100
35,14	974	30	29227	93
36,11	825	30	24753	100
36,12	599	30	17982	100
	915092		36101440	

Figure 5. Partial tabulation of health effect of air pollution of two-kilometer grid cell. This is a function of population and air pollution.

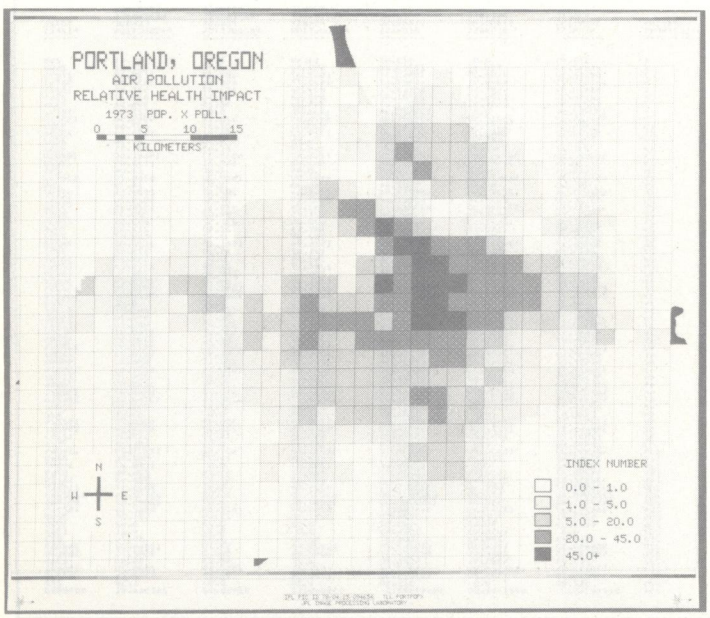


Figure 6. Choroplethic map of health effect by grid cell generated from the tabulation.

4. Geometrically correct the boundary files according to the distortion surface (G).
5. Change the boundary file from vector to raster format (A, C, R, G).
6. Identify the polygons determined by the boundaries (A, C, R).
7. Perform polygon overlay of classification files with boundary files (A, C, R, T). This yields a table of acreage of classes per zone or cell (T).
8. Rearrange the table of acreages and produce a printed report (T).
9. Perform polygon overlay of the census tracts with the grid cells taking the sum of residential pixels in each intersection (A, C, R, T).
10. Merge the population data according to census tract keys (T).
11. Disaggregate the population data over the tract-cell intersections in proportion to the sum of residential pixels (T).
12. Reaggregate the population data by grid cell (T).
13. Multiply population per grid cell by pollution per grid cell to obtain a tabulation of health effect and produce a printed report (T).
14. Inject the health-effect tabular values into the grid-cell image to give a choroplethic map of health effect (R, T).

The Portland agencies can now combine these results with new data sets produced from the grid pollution model to calculate health effect on 1973 population. It is also easy to use census updates or projections to obtain health effects at a desired date. Parenthetically, census updates and projections are rather hard to do, but are undertaken by the Bureau of the Census and other agencies. Details are given in Bryant et al.[4] Figure 4 shows the land use map derived from Landsat with political boundaries overlaid. Figure 5 is a partial tabulation of health effect by grid cell. Figure 6 is the map of health effect by grid cell.

Urban expansion delineation. The Geography Division of the US Bureau of the Census was interested in the urban growth of the Orlando, Florida, region during 1970-1975. Since Landsat 1 was not placed into orbit until 1972, this study had to compare a 1975 land cover map derived from Landsat to a 1970 base constructed from the decennial census.

The procedure for assessing urban expansion consisted of 16 major steps:

1-6. Same as Portland case (above).
7. Count the pixels in each polygon to produce a table of areas (R, T).
8. Merge the census population data into the table of areas according to census track keys (T).
9. Divide population by area and use 1000 people per square mile as a threshold for urban/non-urban (T).
10. Inject the urban/non-urban values into the census track image to give a choroplethic map (R,T).
11. Use classification to extract basic land cover from the 1975 Landsat (A, C, R, G, T, M).
12. Use an image cell-value translation routine to

convert the classes to urban/non-urban (R).

13. Use polygon overlay to produce tables of urban/non-urban land use by census tract (A, C, R, T).
14. Merge the tables according to census-tract key to produce a single table comparing urban growth by census tract for 1970-1975 (T).
15. Digitally overlay the census tract boundaries on the two urban area images to produce maps (R).
16. Use an image arithmetic combination routine to produce a change map (A, R).

It is interesting to note that Landsat data forms an historical archive over large areas of the earth dating back to 1972. Here, the status of an area in 1975 was retrieved. For a report, see Friedman and Angelici.[5] Figure 7 shows the map of urban change, 1970-1975.

Solar power study. Our final example, a rooftop inventory for photovoltaic potential in California's West San Fernando Valley, consisted of 14 major steps:

1-6. Same as steps 1-6 of the Portland case using power station district boundaries.
7. Use classification to extract land cover from Landsat (A, C, G, R, T, M).
8. Correlate sample data from aerial photographs to estimate percentage of available rooftop in each land use class (G, M).
9. Perform polygon overlay to determine areas of land use per power station district (A, C, R, T).
10. Merge correlate power-station-district electricity use data into the land use table (T).
11. Convert land use area to rooftop area using rooftop percentages from step 8 (T).
12. Convert rooftop area to electric generation capacity using known characteristics of photovoltaic arrays and assuming 50 percent utilization of rooftop (T).
13. Compare with power-station-district electricity use and produce printed report (T).
14. Inject the potential photovoltaic generating capacity (as percentages) into the power district image to produce a choroplethic map (R, T).

Engineers may tell us the characteristics of a single rooftop silicon array, but it is the geographer's task to provide data on widespread use. This study has shown that a surprising amount of the power needed by a suburban area can be provided by rooftop generation. Details are described by Angilici and Bryant.[6] Figure 8 shows part of the tabulation of rooftop photovoltaic generating capacity.

Structuring the data

All of these projects used Landsat data to derive a land use base for analyses of various kinds. Although a land use base could be in vector mode (as are the USGS LUDA files, for example), the use of remotely sensed digital data seems to foster a raster mode of operation. The 50 to 80-meter cell size is also necessitated by Landsat. In some ways, particularly in the remote sensing field, Landsat is a driving force behind technology developments. In other ways, Landsat processing must adapt to conventions developed in preexisting applications areas. One trend, exemplified by the California desert study, is to convert Landsat data to a standard map projection. Another trend is data integration, especially the linking of Landsat-derived data with the other major data bases at the regional, state, and national levels. The kinds of analyses that result from data integration are demonstrated in the projects described

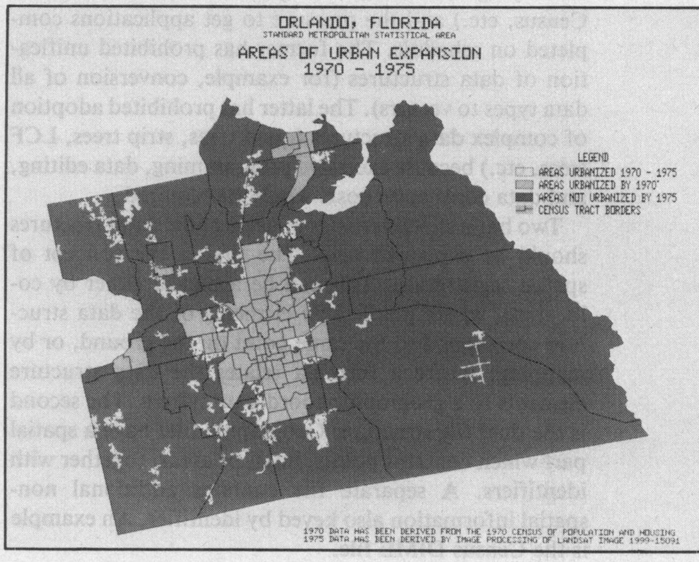


Figure 7. Final map of urban change for Orlando, Florida, derived from census data and Landsat data.

SAN FERNANDO VALLEY
POTENTIAL OF SOLAR ENERGY TO DISPLACE ENERGY CONSUMPTION
BY POWER SUB-DISTRICT (WITH INDUSTRIAL STATIONS INCLUDED)
FOR THE MONTHS OF SEPTEMBER AND OCTOBER (1978)
PRELIMINARY RESULTS

	SEPTEMBER			OCTOBER		
POWER SUB-DISTRICT	SOLAR POTENTIAL (KWH)	ACTUAL DEMAND (KWH)	(%)	SOLAR POTENTIAL (KWH)	ACTUAL DEMAND (KWH)	(%)
130- 4	274995	1080449	25.5	236947	1064046	22.3
130- 5	162856	455334	40.2	157553	428173	36.8
130- 6	189722	102373	185.3	163472	96256	169.8
130- 7	469107	441264	106.3	404201	414896	97.4
130- 8	120862	600118	20.1	104140	564258	18.5
130- 9	748125	699577	106.9	644615	660787	97.6
130-10	48025	353011	13.6	41380	331917	12.5
	3460820	5077351		2981936	4835465	
131- 1	582298	1410755	41.3	501732	1390317	36.1
131- 2	566393	718497	78.8	488028	721513	67.6
131- 3	588075	781470	75.3	506710	784751	64.6
131- 4	321756	675251	47.6	277238	678086	40.9
131- 5	462449	622142	74.3	398465	624753	63.8
131- 6	497640	1093428	45.5	428787	1059036	40.5
131- 7	741980	720774	102.9	639321	723799	88.3
131- 8	455211	735948	61.9	392228	739037	53.1
131- 9	322108	687390	46.9	277542	690276	40.2
131-10	431642	723490	59.7	371921	830519	44.8
131-11	793297	792341	88.8	615984	859768	70.5
131-12	567131	1028515	55.1	488707	1116314	43.8
131-13	760256	1305590	58.2	655068	1233132	53.1
131-14	416363	905899	46.0	358755	909702	39.4
131-15	166681	397564	41.9	143610	399233	36.0
	7583320	12599047		6554098	12760231	
133- 1	437102	535679	81.6	376625	421294	89.4
133- 2	570237	877404	65.0	491340	690051	71.2
133- 3	324186	618601	52.4	279334	486668	57.4
133- 4	644873	1174825	54.9	555647	931187	59.7
133- 5	592452	983234	60.0	510481	777215	65.7
133- 6	166256	1979518	8.4	143253	1707687	8.4
133- 8	424775	837690	50.7	366004	658817	55.6
133- 9	350410	821989	42.6	301927	646469	46.7
133-10	622283	1212994	51.7	536188	992959	54.0
133-12	448412	833496	53.8	386370	655912	58.9
133-13	428515	720395	59.5	369226	566568	65.2
133-14	389595	723166	53.9	335617	568747	59.0
133-15	475145	685299	69.3	409405	538966	76.0
133-16	630266	676063	93.2	543063	531703	102.1
133-17	597842	1086959	55.0	515125	887646	58.0

Figure 8. Partial tabulation of rooftop photovoltaic generating capacity. Particular months and particular power substation districts are analyzed.

above. All four of the projects use the polygon overlay operation to yield tabulations of spatial data. Perhaps the most interesting of these cases was the use of Landsat data to allocate population in Portland when everyone knows a satellite cannot count people.

The previous section has given a taste of the richness of data structures (R, G, and T codes) and prevalence of algorithmic problems (A and C codes). What can be said about the impact of research in mathematics, image processing, and computational geometry on the development of remote sensing information systems? With regard to data structures, the constraints have been the underlying structure of the great data bases (Landsat, Census, etc.) and the pressure to get applications completed on schedule. The former has prohibited unification of data structures (for example, conversion of all data types to vectors). The latter has prohibited adoption of complex data structures (quad trees, strip trees, LCF trees, etc.) because excessive programming, data editing, and data conversion costs would be required.

Two basic characteristics of geographic data structures should be explained here. The first is the concept of spatial registration. This can be achieved either by colocation, where predictable elements of the data structure correspond to the same point on the ground, or by mapping, where a formula relates the data structure elements to a geographic coordinate system. The second is the dual file structure. Geographic files have a spatial part which contains points, lines, or areas, together with identifiers. A separate file contains additional nonspatial information also keyed by identifier. An example is the Census DIME file.

Research on geometric algorithms has been a source of inspiration for geographic information systems, but a major limiting factor has been the need to retain conventional data structures, as outlined in the previous paragraph. Three related problems are of major concern.

First, most of the theoretical work has aimed at acceptable worst-case performance, whereas good expected performance or even the development of unpredictable heuristic methods would seem appropriate for economic reasons. McLemore and Zobrist[7] describe a heuristic for coloring a map in four colors with no two adjacent regions the same color, with a worst case of $0(n^2)$ (if it fails to work, then it also reports that in $0(n^2)$ time). Manacher and Zobrist[8] describe a method for triangulating a point set in shortest-edge-first fashion which has a worst case of $0(n^3)$ but which uses a technique that obtains $0(n^2)$ average case behavior (proof not published).

A second problem arises when algorithms are based on complex data structures that impose unrealistic constraints on data capture or editing. For example, many data bases have files of polygons which are intended to partition an area but actually overlap or underlap each other. This topological imperfection rules out many algorithms for polygon processing.

The third problem is numerical accuracy (quantization effects) in performance of elementary steps of an algorithm. An example is the oft-advocated method for determining whether a point is inside a polygon: drop a perpendicular from the point and count crossings of the polygon boundary. An odd count implies that the point is inside. The problem is that the perpendicular may be tangent or end-intersecting to an edge in the polygon, thereby producing an error. Furthermore, this problem is spatially ill-conditioned because the regions of error can be arbitrarily far from the polygon boundary.

Based on present successes and expected advances in remote sensing from space and in communication, data storage, and data processing technologies, the next 30 years hold an exciting promise: no less than a complete inventory of the world's major resources. Assessment of food and fiber crops, forage biomass, forests, wetlands, coastline, urban development, growth of deserts, water resources, minerals, soils, oceans, weather, and climate will all be affected by this vast enterprise. The inventory process will not be a simple operation of pixel counting from remotely sensed data. Instead, the methods and models in present use by discipline scientists will be enhanced by the incorporation of remotely sensed data sets and rendered more efficient by the computerized techniques of geographic analysis now being developed. ■

Acknowledgments

This article presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract No. NAS 7-100, sponsored by the National Aeronautics and Space Administration, and at the University of Nebraska with the support of the Conservation and Survey Division through NASA grant 28-004-020.

References

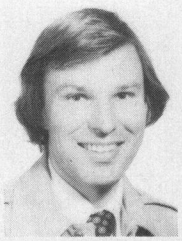
1. *Landsat Data User's Handbook,* revised edition, US Geological Survey, Arlington, Va., 1979.
2. R. G. McLeod and H. B. Johnson, "Resource Inventory Techniques Used in the California Desert Conservation Area," presented at the Arid Lands Resource Inventory Workshop, La Paz, Mexico, Nov. 30-Dec. 6, 1980.
3. *The California Desert Conservation Area Final Environmental Impact Statement and Plan,* US Department of Interior, Bureau of Land Management, Oct. 1, 1980.
4. N. A. Bryant, A. J. George, Jr., and R. Hegdahl, "Tabular Data Base Construction and Analysis from Thematic Classified Landsat Imagery of Portland, Oregon," *Proc., Fourth Annual Symposium on Machine Processing of Remotely Sensed Data,* Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, June 21-23, 1977, pp. 313-318.

5. S. Z. Friedman and G. L. Angelici, "The Detection of Urban Expansion from Landsat Imagery," *Remote Sensing Quarterly,* Vol. 1, No. 1, Jan. 1979. Originally presented at spring meeting of Association of American Geographers, New Orleans, Apr. 1978.

6. G. L. Angelici and N. A. Bryant, "Solar Potential Inventory and Modelling," *Data Resources and Requirements: Federal and Local Perspectives,* contributed papers from the 16th Annual Conference of the Urban and Regional Information Systems Association, Washington, DC, Aug. 6-10, 1978, pp. 442-453.

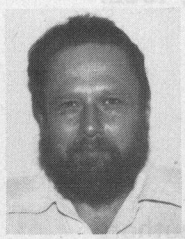
7. B. D. McLemore and A. L. Zobrist, "A Five Color Algorithm for Planar Maps," presented at the Second Caribbean Conference on Combinatorics and Computing, University of the West Indies, Cave Hill, Barbados, Jan. 1977.

8. G. K. Manacher and A. L. Zobrist, "Fast Average Case Greedy Triangulation of a Planar Point Set," *Proc. 16th Annual Allerton Conference on Communication, Control, and Computing,* Monticello, Ill., Oct. 4-6, 1978.



Albert L. Zobrist is presently a member of the technical staff at the Jet Propulsion Laboratory. He was an associate professor of computer sciences at the University of Arizona, Tucson, 1974-75, and from 1970-74 was a professor of electrical engineering and computer science at the University of Southern California, Los Angeles, where he was also a member of the research staff of the Information Sciences Institute from 1973-74. Before that, he was a member of the technical staff at the Aerospace Corporation, El Segundo, California.

Zobrist received the BS degree in mathematics from Massachusetts Institute of Technology, Cambridge, in 1964, and the MS degree in mathematics and the PhD degree in computer science, both from the University of Wisconsin, Madison, in 1966 and 1970.



George Nagy was appointed chairman of the Computer Science Department of the University of Nebraska in 1972 and teaches courses on computer organization, pattern recognition, and remote sensing. Prior to joining the University of Nebraska, Nagy spent 10 years on the staff of the IBM T. J. Watson Research Center, developing pattern classification techniques for optical character recognition, speech processing, data compression, and remote sensing. His current research areas are geographic data processing, digital image registration, and the quantitative evaluation of the human-computer interface.

Nagy has served as a research consultant for IBM, Bell Telephone Laboratories, Tektronix, Compression Laboratories, and NASA. He has given lectures at over 100 universities and professional conferences in the United States and abroad and is the author of numerous research and survey articles. A member of the IEEE and ACM, Nagy received the PhD degree in electrical engineering from Cornell University in 1962.