USGS
science for a changing world

Mapping Surface Disturbance of Energy-Related Infrastructure in Southwest Wyoming—An Assessment of Methods

Scientific Investigations Report 2012–5025

U.S. Department of the Interior
U.S. Geological Survey

Mapping Surface Disturbance of Energy-Related Infrastructure in Southwest Wyoming—An Assessment of Methods

By Stephen S. Germaine, Michael O'Donnell, Cameron L. Aldridge, Lori Baer, Tammy Fancher, Jamie McBeth, Robert R. McDougal, Robert Waltermire, Zachary H. Bowen, James Diffendorfer, Steven Garman, and Leanne Hanson

Scientific Investigations Report 2012–5025

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia: 2012

Suggested citation:
Germaine, S.S., O'Donnell, M.S., Aldridge, C.L., Baer, L.A., Fancher, T.S., McBeth, J.L., McDougal, R.R., Waltermire, R.G., Bowen, Z.H., Diffendorfer, J.E., Garman, S.L., and Hanson, Leanne, 2012, Mapping surface disturbance of energy-related infrastructure in southwest Wyoming—An assessment of methods: U.S. Geological Survey Scientific Investigations Report 2012–5025, 42 p.

Contents

Figures

Tables

Mapping Surface Disturbance of Energy-Related Infrastructure in Southwest Wyoming—An Assessment of Methods

By Stephen S. Germaine, Michael O'Donnell, Cameron L. Aldridge, Lori Baer, Tammy Fancher, Jamie McBeth, Robert R. McDougal, Robert Waltermire, Zachary H. Bowen, James Diffendorfer, Steven Garman, and Leanne Hanson

Executive Summary

We compared the ability of three leading information-extraction software programs and manual hand digitization to interpret information from remotely sensed imagery of a visually complex gas field in Wyoming. Using ground control points and manual hand digitization of QuickBird (Digital Globe, Inc.) satellite imagery as benchmarks, we evaluated how accurately VLS Feature Analyst®, ENVI Feature Extraction®, Definiens eCognition®, and hand digitization mapped the area of, and classified the type of infrastructural disturbance features present on each of three remotely sensed images. Imagery included 30 meter resolution Landsat, 10 meter resolution SPOT (Satellite Pour l'Observation de la Terre), and 0.6 meter resolution pan-sharpened QuickBird scenes of a 64 square-kilometer area on the Jonah Energy Field in Sublette County. After evaluating mapping and classification performance of each method-image combination (MIC), we calculated cost-benefit ratios for each by comparing the time required to complete maps with each MIC (cost) against two descriptors of map quality: (1) degree of geometric overlap shared between each MIC and the map derived using hand digitization of the Quickbird image, and (2) kappa index scores describing feature classification accuracy for each MIC based on the ground-control-point reference data.

Our first objective was to compare how accurately each MIC mapped the area, or footprint, of infrastructural disturbance present on each image. To assess this, we first summed and compared the total area of disturbance mapped by each MIC, then compared the dissimilarity among footprints of features mapped by each MIC using a spatial overlap statistic. For both comparisons, we considered the map produced by hand digitizing QuickBird imagery our benchmark. We found that: (1) well pads, disturbed areas, and maintained roads accounted for the majority of area mapped by each of the MICs; (2) Feature Extraction mapped the spatial area of disturbance features most accurately on the Landsat and QuickBird imagery, while hand digitization was most accurate on the SPOT imagery; (3) most of the footprint dissimilarity error present on the QuickBird and SPOT imagery was contributed by the disturbed area and well pad feature classes, while these plus maintained roads and two-tracks contributed the majority of footprint dissimilarity error on the LS image; (4) footprint dissimilarity error was smallest on the Feature Analyst map of the Landsat imagery, the hand digitization map of the SPOT imagery, and the Feature Extraction map of the QuickBird imagery; and (5) the Feature Extraction map of the QuickBird scene was the only map that had less than 20 percent footprint dissimilarity error relative to our benchmark map.

Our second objective was to evaluate how accurately each MIC classified disturbance features present on the ground. To assess this, we defined eight feature classes such that each infrastructural disturbance feature present in our study area fit into only one class, then drove all roads in the study area while identifying and mapping precise coordinate locations of every feature present. From this pool, we randomly selected 176 feature locations against which we compared the classification made by each of the 12 MICs. We then geometrically corrected and mapped all 12 MIC maps to a common cartographic projection, compared the feature class assigned at each of the 176 locations with the ground control reference values, and evaluated the level of feature classification agreement among each MIC and the reference data using kappa statistics. We found that: (1) Feature Analyst, Feature Extraction, and hand digitization classified features with similar success on the QuickBird and SPOT imagery; (2) eCognition classified features poorly relative to the other methods on the QuickBird and SPOT imagery; and (3) all maps derived from Landsat imagery classified disturbance features poorly.

In addition, we digitally tagged the 176 disturbance feature locations onto the map derived from hand digitized QuickBird imagery to serve as a reference for the other 11 MICs, rasterized all of the maps, and assessed classification agreement on a pixel by pixel basis. We found that: (4) Feature Extraction classified features best overall on the QuickBird imagery, but eCognition, Feature Analyst, and Feature Extraction all classified disturbance features on the QuickBird imagery with a high level of success; (5) Feature Analyst

classified features best overall on the SPOT imagery, but all four MICs classified features with a fair–good accuracy level; and (6) Feature Analyst classified features best on the Landsat imagery, but all four MICs again classified features with a fair–good accuracy level.

Our third objective was to conduct a cost-benefit assessment of each MIC. To assess this, we compared the time (hours) required to extract and classify all of the disturbance information using each MIC against two estimators of product quality: footprint dissimilarity values and kappa scores associated with the ground control point (GCP) classification accuracy assessment. When we plotted footprint dissimilarity against production time we observed that: (1) The MICs formed different patterns within each image resolution; (2) on the Landsat image, the eCognition extraction resulted in high footprint dissimilarity error rates; (3) the Feature Analyst extraction took approximately eight times longer to complete than the other MICs; (4) on the SPOT image, eCognition again resulted in a high level of footprint dissimilarity error; (5) Feature Analyst, Feature Extraction, and hand digitization all performed similarly in terms of time and error-rate efficiency; (6) extractions performed on the QuickBird imagery took from four to seven times longer than all other MICs except the Feature Analyst extraction of the Landsat image; (7) Feature Extraction and hand digitization took the least amount of time to extract and had lower error rates than the other two MICs.

When we plotted overall GCP kappa index scores against extraction time, we learned that: (8) on the Landsat imagery all four MICs had poor scores, while Feature Analyst took approximately eight times longer to extract and eCognition took approximately two times longer than Feature Extraction or hand digitization; (9) on the SPOT image, Feature Analyst, Feature Extraction, and hand digitization took half as long to extract as eCognition did, and they each had higher classification scores than did eCognition; (10) on the QuickBird image, Feature Analyst, Feature Extraction, and hand digitization all had high classification scores, but Feature Analyst took far longer to extract, and eCognition had worse classification scores than the other three MICs.

On the Landsat and QuickBird imagery, Feature Extraction performed best overall based on the suite of tasks we evaluated, while hand digitization performed best overall on the SPOT imagery. eCognition performed worst overall on all three images. Consideration of degree of map accuracy required, cost associated with software, operator, computation time, and tradeoffs in the form of spatial extent versus resolution should all be considered when evaluating which combination of imagery and information-extraction method might best serve any given land use mapping project. With any automated information extraction product, time and training to develop operational expertise may need to be considered. Error rates for both area measurements and feature classification were prohibitively high on Landsat imagery, while QuickBird was time and cost prohibitive for mapping large spatial extents. The SPOT imagery produced map products that were far more accurate than Landsat and did so relatively inexpensively. When resources permit, attaining imagery that supports the highest classification and measurement accuracy possible will result in dramatically improved map accuracy.

Introduction

Land-use and land-cover changes are occurring with increasing rapidity (Antrop, 2005; Drummond and Loveland, 2010) and are impacting the physical and biological environment on every continent (Vitousek and others, 1994; Lambin and others, 2001; Foody, 2002). These changes alter function, habitat quality, and biodiversity in both terrestrial and aquatic environments (Fahrig, 2003; Driscoll, 2004; Ripa and others, 2006). In the United States, fragmented landscapes are now far more common than intact ones. For instance, more than 80 percent of all of the land in the conterminous United States lies within 1.06 kilometers (km) of a road (Riitters and Wickham, 2003). Public transportation infrastructure in the U.S. includes 6.3 million km (3.9 million miles (mi)) of roadways, 3.2 million km (2 million mi) of oil and gas pipelines, and 192,000 km (120,000 mi) of railroads (National Atlas, 2010), while the length of the nation's river and stream system is 5.6 million km (3.5 million mi; U.S. Environmental Protection Agency, 2004).

Accurate measurement of land use and land cover change is critical for understanding their effects on natural systems. However, errors still exist in many commonly used map products derived from remotely-sensed imagery (Foody, 2002; Hawbaker and Radeloff, 2004). These errors continue to limit our ability to accurately quantify land use and land cover changes, mainly because existing land-cover maps contain feature classification and measurement errors to the degree that they may be …"judged to be of insufficient quality for operational applications." (Foody, 2002, p. 186).

Reducing classification and measurement error rates is not straightforward. Newer satellites such as SPOT 5 and QuickBird produce images that have dramatically increased spatial resolution and clarity, and it is well understood that resolution affects both feature classification and spatial measurement accuracy (Lillesand and Kiefer, 2000). However, high-resolution images are expensive, have small footprints, and are relatively new, making comparisons with past land use change difficult with these formats. Clearly, the most appropriate spatial resolution for any given mapping exercise depends on the spatial structure of the landscape under consideration, the level of detail desired, and (if land cover change is being investigated) the time period during which change is being evaluated (Lillesand and Kiefer, 2000). For these and other reasons, Landsat and other medium-resolution imagery are still frequently used in land use and land cover change evaluations.

Automated image-extraction software has the potential to improve image-interpretation efficiency and reduce mapping errors, but few evaluations of this potential have been

made. While visual interpretation followed by manual editing has been the traditional method of classifying remotely sensed imagery, manual interpretation is labor intensive, the full range of spectral information present may not be discernible by the human eye, and patterns among objects may present a challenging level of complexity (Lillesand and Kiefer, 2000). Furthermore, high-resolution imagery often contains multiple pixels within single features, rendering traditional pixel-based interpretation methods less effective than when interpreting medium- and low-resolution imagery where a single pixel represents multiple features (Mladinich, 2010). Unlike previous extraction methods, automated image-extraction programs use both spatial context and spectral information to perform object-oriented classification, and incorporate machine learning capabilities to iteratively improve upon initial classifications.

We compared three leading information-extraction software programs and manual hand digitization while interpreting and classifying information from remotely-sensed imagery of a visually complex gas energy field. We also tested three types of imagery: 30 meter (m) resolution Landsat, 10 m resolution SPOT (Satellite Pour l'Observation de la Terre), and 0.6 m resolution pan-sharpened QuickBird (Digital Globe, Inc.) imagery. We anticipated that information interpretation would be positively related to image resolution, and therefore focused primarily on within-image comparisons among the four extraction methods. For each of the 12 method-image combinations (MIC), we had three objectives. They were to compare: (1) ability to accurately measure areas of disturbance, (2) classification accuracy among types of disturbance, and (3) cost-benefit efficiency while performing the first two objectives.

Study Area

All imagery was collected over a 64 km^2 area on the Jonah Energy Field in southwest Wyoming, an area undergoing intensive gas energy development that encompasses a gradient of development intensity and a variety of infrastructure categories. Our evaluation site was centered on 606116 Easting, 4702679 Northing, (UTM Zone 12, datum WGS84), and was west of State Highway 191 and south of the town of Boulder, Wyoming (fig. 1). Infrastructural disturbance at the site included approximately 320 km (198.6 mi) of improved roads and 298 gas well pads at the time of scene acquisition. Other types of disturbance included borrow areas, bare earth from which all vegetation and topsoil had been removed, pipeline rights-of-way, holding ponds, buildings, equipment parking and storage lots, and topsoil storage mounds. An unquantified amount of unimproved 2-track trails occurred throughout the study area, as well as occasional stock pens and bermed stock ponds.

Topography was gentle to steep rolling hills interspersed with washes and small drainages that were dry most of each year. Surface materials included rock outcrops and mixed loamy soils of Typic Torriorthent, Lithic Torriorthent, and Typic Natrargid types (University of Idaho, 2009). Vegetative cover was typical of intermountain sagebrush steppe (Johnson and O'Niel, 2001; Davies and others, 2006). Dominant shrub species included Wyoming big sagebrush (*Artemisa tridentata wyomingensis*) and Basin big sagebrush (*A. t. tridentata)*, with rabbitbrush (*Chrysothamnus* spp.) present in some areas. Sparse understory vegetation included a mixture of herbs and forbs, with exotic species well established in disturbed areas. Elevation ranged from 2,144 m (7,035 feet (ft)) to 2,241 m (7,353 ft).

Materials and Methods

Imagery Pre-processing

We evaluated VLS Feature Analyst (FA) (Visual Learning Systems, 2008), ENVI's Feature Extraction (FX) (ITT Visual Information Solutions, 2008), and Definiens' eCogntion (EC) (Definiens, 2006) automated information-extraction software. We also evaluated hand digitization (HD). We acquired study area imagery from three remote sensors: pan-sharpened QuickBird (QB) (August, 2006) at 0.6-m cell size; multispectral SPOT imagery (SP) (August, 2006) at 10-m cell size; and multispectral Landsat 7 imagery (LS) (June, 2006) at 30-m cell size (table 1). All imagery was acquired in standard format and was geometrically corrected and mapped to a common cartographic projection. We coregistered the SPOT to the QuickBird image using Leica Photogrammetry Suite (Leica Geosystems, 2008). The overall residual mean squared error for the SPOT coregistration was 2.8 pixels in the x-direction and 1.6 pixels in the y-direction. We attempted to georeference the Landsat imagery to the Quickbrid imagery, but did not do so because the Landsat registration could not be improved over the commercially applied georeferencing. Because orthorectification would have negligible effects on our comparisons, we did not orthorectify any of the imagery.

Image Interpretation

To minimize any operator-inherent error in the post-processing phase of each MIC interpretation, we applied the following common ground rules.

- Operators were not allowed to consult supplementary digital data sets such as roads or gas-well locations to assist in interpreting the sample imagery, but were allowed to use software extensions when they aided information interpretation or extraction.
- Some disturbance class features did not have distinct common boundaries (for example, gravel roads and bare pads). In these cases, each operator was given

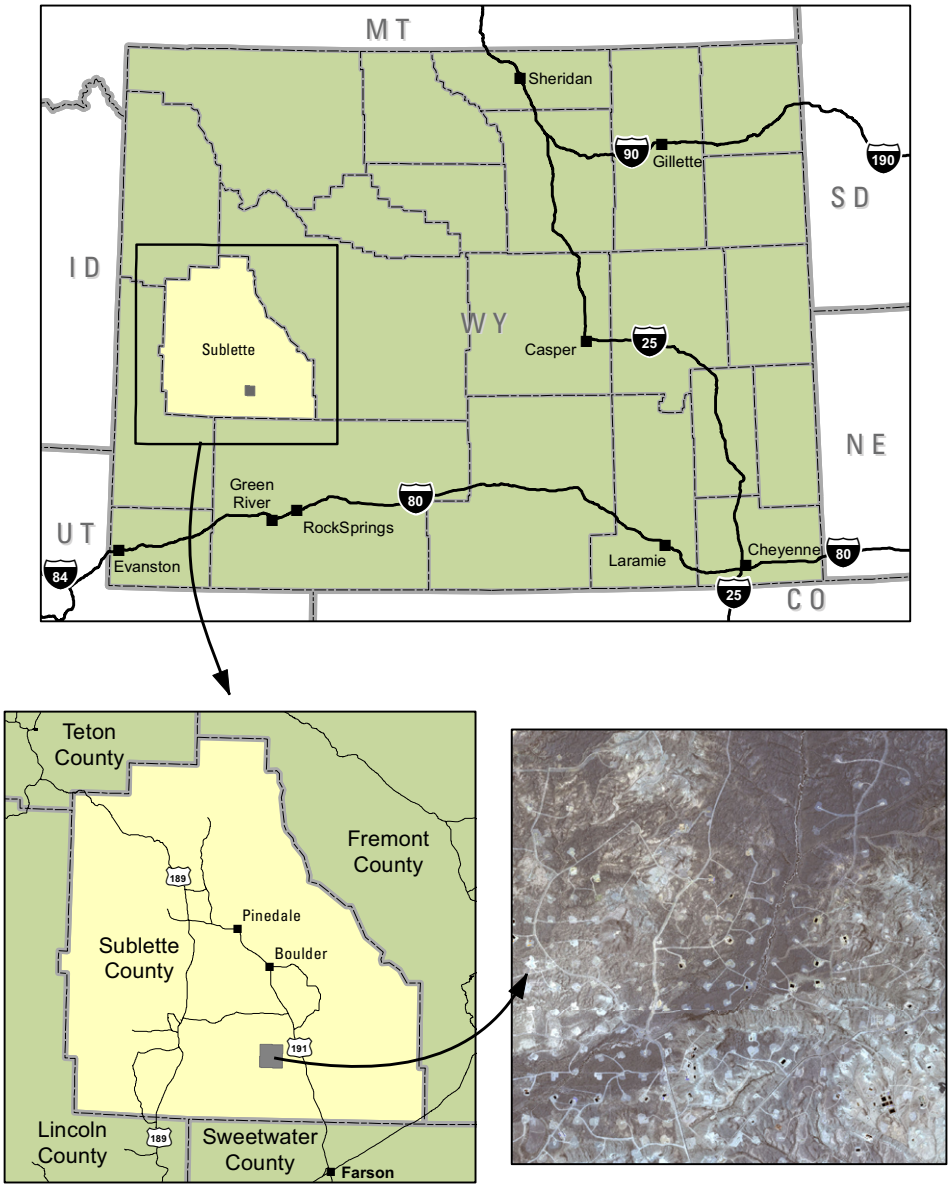

Figure 1. Jonah gas energy field in Sublette County, Wyo. Detail shows the 64 km² area where we compared information-extraction methods using QuickBird, SPOT, and Landsat imagery. Shown is the QuickBird scene, collected during August, 2006. (km², square kilometers)

Table 1. Relevant parameters for three sensors (Landsat, SPOT, QuickBird) used to evaluate extraction method–image combination performance from the Jonah gas energy field, Sublette County, Wyo. (µm, micrometers; m, meters; km, kilometers; km^2, square kilometers)

Remotely sensed imagery	Acquisition date	Band	Wavelength (µm)	Resolution (m)	Swath width (km)	Revisit (days)	Cost / scene size[a]
Quick Bird	August 14, 2006	1	0.45–0.52	2.44	16.5	1–3.5 (depends on latitude (30° off-nadir))	$6,528 / 272 km^2
		2	0.52–0.6	2.44			
		3	0.63–0.69	2.44			
		4	0.76–0.9	2.44			
		Pan	0.45–0.9	0.61			
SPOT 5	August 9, 2006	1	0.50–0.59	10	60	26 (off-nadir 1–3)	$6,350 / 3,600 km^2
		2	0.61–0.68	10			
		3	0.78–0.89	10			
		4	1.58–1.75	20			
		5	Panchromatic (not purchased)	2.5/5			
Landsat 7 ETM	June 23, 2006	1	0.45–0.52	30	185	16	Free as of April 2008 / 31,450 km^2
		2	0.52–0.60	30			
		3	0.63–0.69	30			
		4	0.76–0.90	30			
		5	1.55–1.75	30			
		6	10.4–12.5	60			
		7	2.08–2.35	30			
		8	Panchromatic (not used)	15			

[a] Prices shown represent acquisition costs for imagery used in this study as of 21 June 2011. Multiple online vendors may exist for QuickBird and SPOT imagery, and prices may vary over time and among vendors.

example figures demonstrating how boundaries should be delineated.

- We allowed operators to change viewing scales of images as needed to identify and measure features optimally, and did not set a common minimum mapping unit for disturbance features.
- No smoothing of final edited features was allowed.
- Operators were required to process the imagery from lowest resolution (LS) to highest resolution (QB) so that memory of specific details would not bias interpretation of lower-resolution image features.
- Finally, we defined eight unique classes of disturbance (table 2) such that any feature encountered on the imagery could be correctly assigned to only one class.

Hand Digitization (LSHD, SPHD, QBHD)

Before beginning actual hand digitization, operators familiarized themselves with the images, visually recognized feature patterns, and referred to the class definitions to mentally sort objects into appropriate classes. During hand digitization, we used six aids to help delineate and classify features: shape (form; on landscape scale imagery, regular uniform shapes are often human-caused), pattern (spatial arrangement of objects), size (measure of an object's area or width), tone/color (spectral qualities of an object relative to other objects), texture (physical characteristics of an object such as grain size and coarseness), and association/site (presence of one object is often linked with another; for example, all well pads have access roads).

We digitized disturbance features by manually tracing them on-screen over the georeferenced imagery using ArcGIS™ 9.3 (ESRI, 2008). We used IMAGINE Easytrace™ (ERDAS Imagine, 2009) to expedite the feature-extraction process by using interactively placed seed points and applying an adaptive line-fitting algorithm to accurately trace the feature between the seed locations. We used Easytrace only to capture distinct linear disturbances (such as maintained roads) because point-to-point hand digitization produced more reliable and consistent results for more complex features.

While digitizing, we attributed each disturbance feature in a manner consistent with the predefined classification scheme, and ensured that the following attribution and topological criteria were met: each feature was assigned to exactly one disturbance class, each feature was correctly classified, each polygon had accurate boundaries, and all polygons were closed (that is, began and ended at the same point) and contained no gaps or overlapping features. We regularly compared our digitized interpretation against the source imagery to ensure positional and classification accuracy and to confirm that no features had been missed (fig. 2). Topological corrections were made using standard ArcInfo Workstation tools (for example, createlabels, clean, build, labelerrors, and nodeerrors) or geodatabase topology rules. Manual feature classification and topological corrections followed each of the three automated software extractions (described below) and were performed in similar fashion (table 3).

eCognition (LSEC, SPEC, QBEC)

To extract data using eCognition, we first performed a multiresolution segmentation of the scene. This was an automated process in which the rasterized image data were converted to object-oriented polygons. We adjusted the following program parameters to optimize within-object homogeneity for each feature disturbance class on each type of imagery:

- Color homogeneity (color): color is the spectral value assigned each pixel. Homogeneity was measured as the sum of the standard deviations of spectral values for all pixels in a defined object. In most cases, color was the most important criterion used in defining objects.
- Shape: included two geometric parameters:
  - Compactness: an object-area criterion defined as the ratio between object length (l) and the square root of the object size (n) in pixels.
  - Smoothness (h): an object-edge criterion with a ratio defined by an object's edge length (l) and the edge length of a square bounding box $(b)^2$, where:

$$h=l/b.$$

- Scale parameter: an operator-defined value that determines the maximum heterogeneity allowed in a defined object. Modifying this value results in a change in the size of a defined image object. In heterogeneous imagery, the size of defined objects for a given scale parameter will be smaller than in more homogeneous data.

Optimizing within-object homogeneity was an iterative process that required up to four iterations. After each run, we visually inspected results to assess the accuracy of the segmentation, and when necessary, adjusted program parameters and re-ran the segmentation. We manually ensured classification and topological accuracy as described above (fig. 3).

Feature Analyst® For ArcGIS (LSFA, SPFA, QBFA)

Feature Analyst (FA) used a machine algorithm that combined object- and pixel-based classification techniques. To train the FA software to each image, we created training shapefiles for each class and hand digitized samples of features. The hand-digitized training sample for each feature class covered the range of spectral variation present and spanned the image extent. We extracted one feature class at a time and performed 3–5 processing iterations for each feature class. Between each

Table 2. Disturbance-feature classification scheme used for attributing types of infrastructural disturbance present on the Jonah gas energy field, Sublette County, Wyo. in 2007. (m^2, square meters)

Numeric class value	Feature class	Definition
1	Building	(1) All **built/structural features** (polygonal data format) ≥ 2 m^2 in area including all types of trailers, barns, sheds, enclosed tanks, houses, etc; (2) Excluded motorized vehicles (cars, trucks, Caterpillar equipment).
2	Disturbed	(1) any area where vegetation or soil appeared to have been removed or visibly reduced by direct human activity (for example, excludes cattle grazing); (2) included road (including railroad) borrow areas, railroads, utilities (pipelines, powerlines, and communication lines); (3) included areas that were devegetated during gas or oil well installation; (4) included any other areas that satisfy A, B, or C, above but that do not meet the criteria to be any other feature class; **AND** (5) excluded road loops, buildings, and other areas on well pads that show evidence of frequent, repeated disturbance by vehicular traffic; (6) excluded naturally bare areas.
3	Paved	All paved impervious surface transportation routes and vehicle parking areas.
4	Road_maintained	All non-paved transportation routes generally having a borrow-pit and a non-vegetated center strip. Occasional vegetation may occur on road surface in rare instances.
5	Road_track	All non-paved transportation routes generally not having a borrow-pit along the feature and having a (sometimes discontinuous) vegetated center strip.
6	Water_lotic	All flowing anthropogenic water features including canals and ditches. Excludes all natural water features. Must contain surface water.
7	Water_lentic	(1) All non-flowing anthropogenic water features including energy-related holding ponds, stock tanks, stock ponds, and wildlife water catchments (for example guzzlers); (2) All non-holding pond features must have water present to qualify. If else, label "disturbed"; (3) Excludes natural water features.
8	Bare_pad	(1) All sites from which vegetation and/or soil had been removed or visibly reduced during installation of gas/oil wells or related infrastructure such as pumping stations or compressors; (2) Included loop roads, areas enclosed by loop roads, and any other areas showing visible evidence that suggests frequent, repeated disturbance by vehicular or other human traffic.

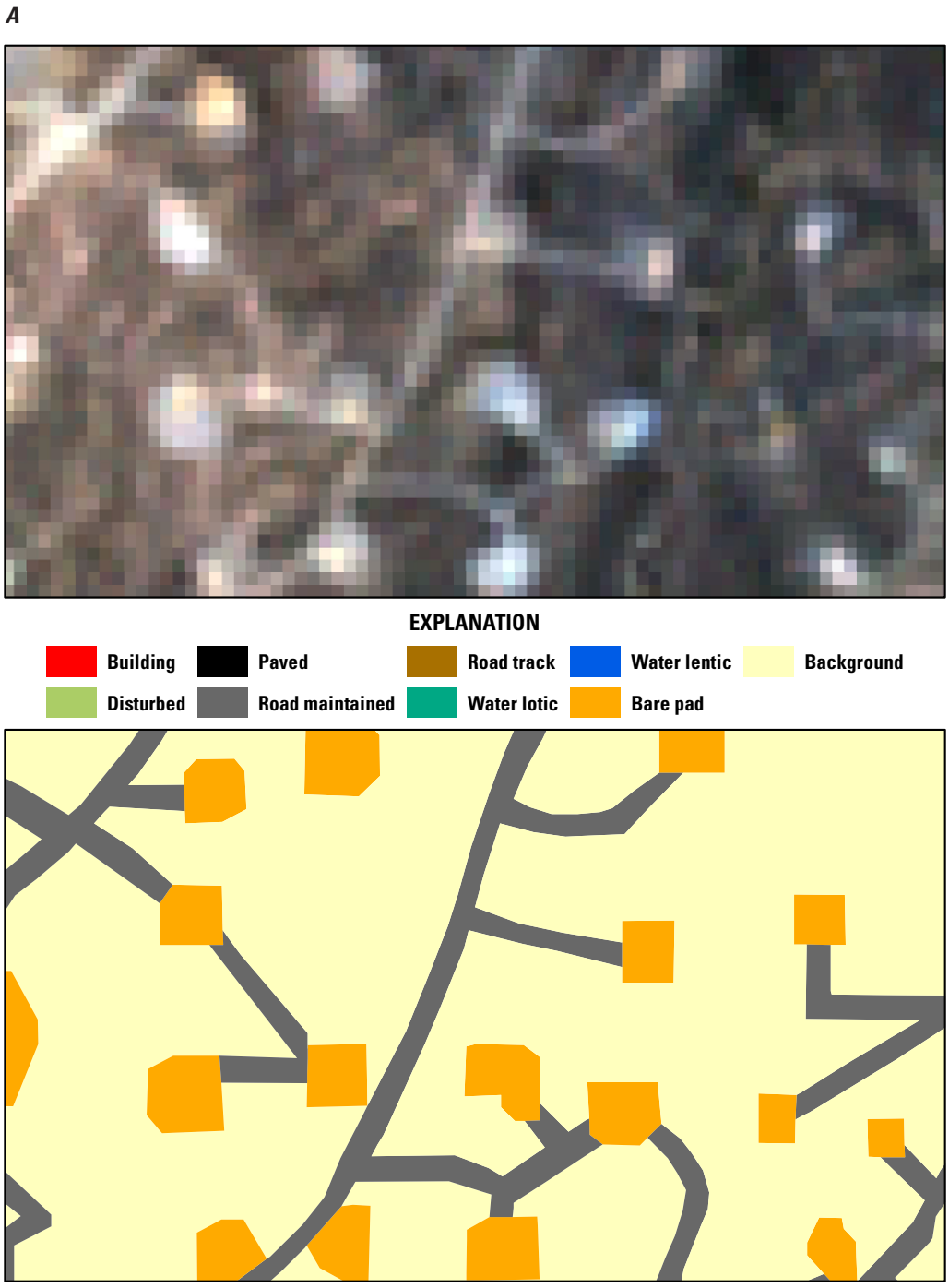

Figure 2. Computer screen captures of each of three image types (*A*, Landsat; *B*, SPOT; *C*, QuickBird) prior to hand digitization (top panels in each figure pair) and after digitization and classification (bottom panels). Explanations of bottom-panel symbology accompany each figure.

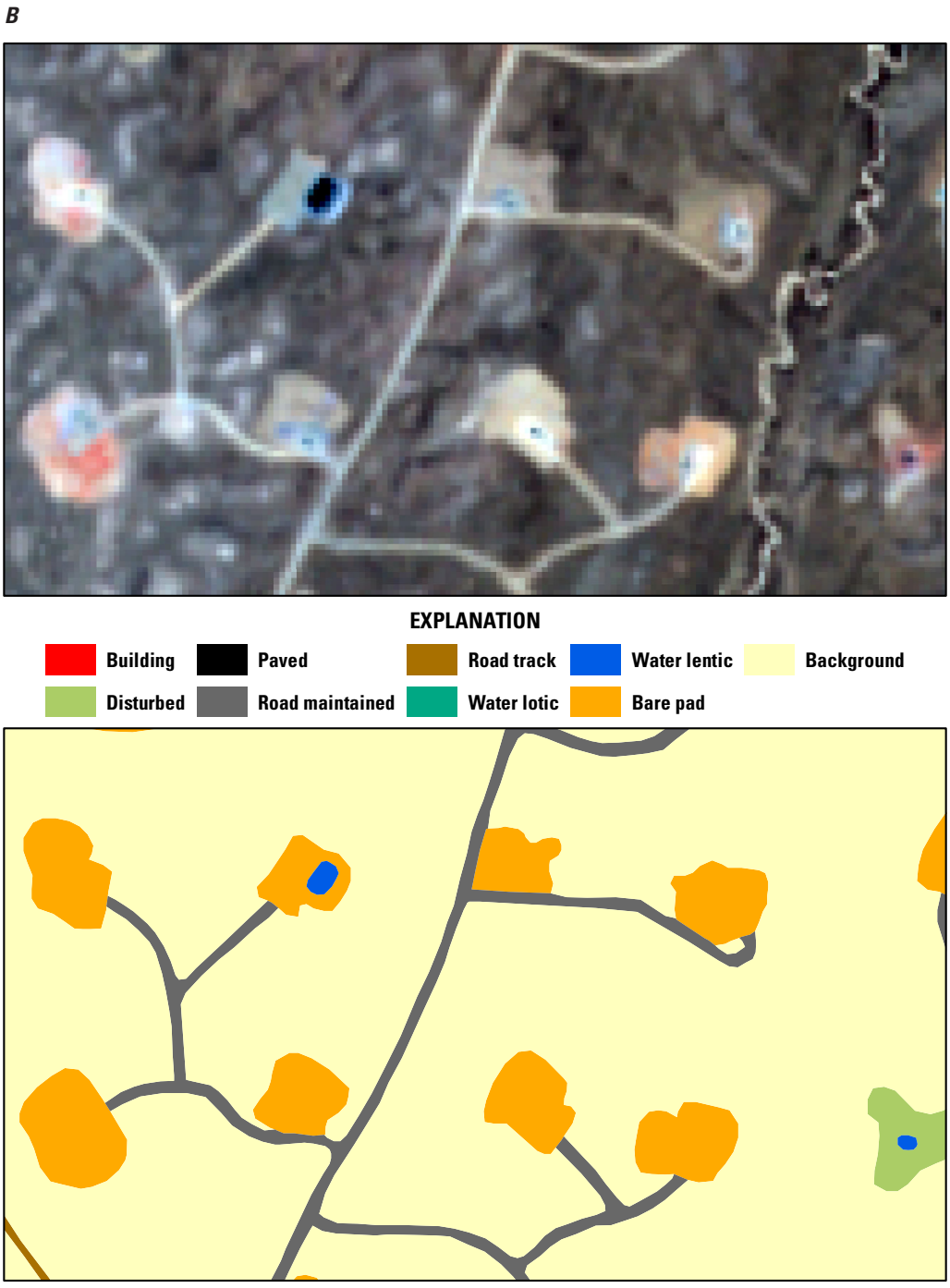

Figure 2. Computer screen captures of each of three image types (*A*, Landsat; *B*, SPOT; *C*, QuickBird) prior to hand digitization (top panels in each figure pair) and after digitization and classification (bottom panels). Explanations of bottom-panel symbology accompany each figure.—Continued

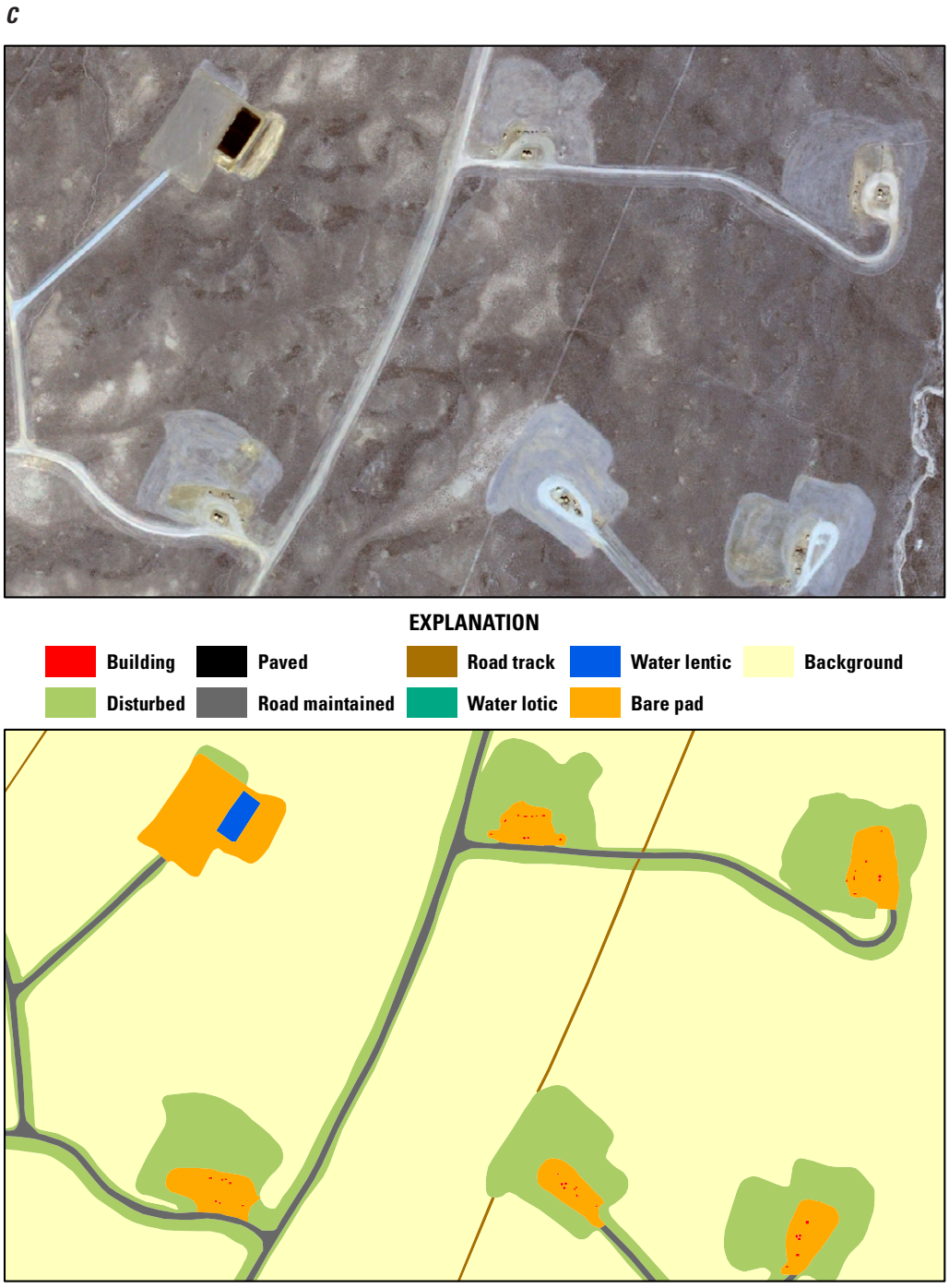

Figure 2. Computer screen captures of each of three image types (*A*, Landsat; *B*, SPOT; *C*, QuickBird) prior to hand digitization (top panels in each figure pair) and after digitization and classification (bottom panels). Explanations of bottom-panel symbology accompany each figure.—Continued

Table 3. Image processing steps for each of four image extraction methods used to map a 64 km^2 section of the Jonah gas energy field, Sublette County, Wyo., in 2007. (km^2, square kilometers)

	Information extraction method			
Workflow	Hand digitization	eCognition	Feature analyst	Feature extraction
	Visual assessment / familiarization	Image rasterization	Create training shapefiles	Convert to ENVI file format
	Manual digitizing	Processing	Hand digitize sample features	Processing
	Feature classification	Visual inspection of output	Processing	Convert to Arc coverage
	Topological error checking and correction	Adjust parameters and reprocess if necessary	Visual inspection of output	Manual feature classification and topological correction
		Manual feature classification and topological correction	Adjust parameters and reprocess if necessary	
			Manual feature classification and positional correction of each shapefile	
			Union feature class shapefiles and convert to Arc coverage	
			Error check feature classification and topological accuracy	
			Dissolve to produce final coverage, convert back to shapefile	

iteration we adjusted our training sample set and/or program parameters to achieve an optimal extraction. We manipulated the following program parameters during processing:

- Image bands: we used spectral bands 1–4 to process the QuickBird and SPOT images, and converted band 4 to a 'texture' band-interpretation type. We created a four-band stacked image from the original Landsat image using Landsat bands 2, 3, 4, and 6.
- Input representation: this option determines the shape and size of the "window" of pixels surrounding each focal pixel while it is being classified. Information from neighboring pixels are incorporated into the extraction algorithm, helping to distinguish groups of pixels into feature classes. Input representation types used during processing included Manhattan, BullsEye 1 and 2, square, and circle options (Visual Learning System, 2008), and were up to 15 pixels in width.
- Masks: we masked out areas of pixels that had already been assigned to particular feature classes, enabling more accurate and efficient feature classification during ensuing extraction iterations.
- Resample factor: this option sets a multiplier used to reduce image resolution. For example, if the resample factor is set to 10, imagery with 1-m resolution is resampled and processed with 10-m resolution (that is, one pixel equivalent to 100 m^2). We resampled the QuickBird imagery from 0.6 m to 1–2-m resolution because we achieved equivalent results in less processing time when working at this resolution in past work.
- Aggregation area: this option allows the operator to set a minimum pixel area below which polygons are not identified or extracted.
- Remove Clutter: we used this tool to derive an optimal extraction for bare_pad on the Landsat image. This is another of FA's Hierarchical Learning tools that allows the operator to create a clutter-training file by selecting correct and incorrect features (polygons) from a previous extraction result.

We left all other parameter adjustment options at default settings. Our decision to accept a final extraction was made after reviewing each iterative result and subjectively choosing the one requiring the least manual editing.

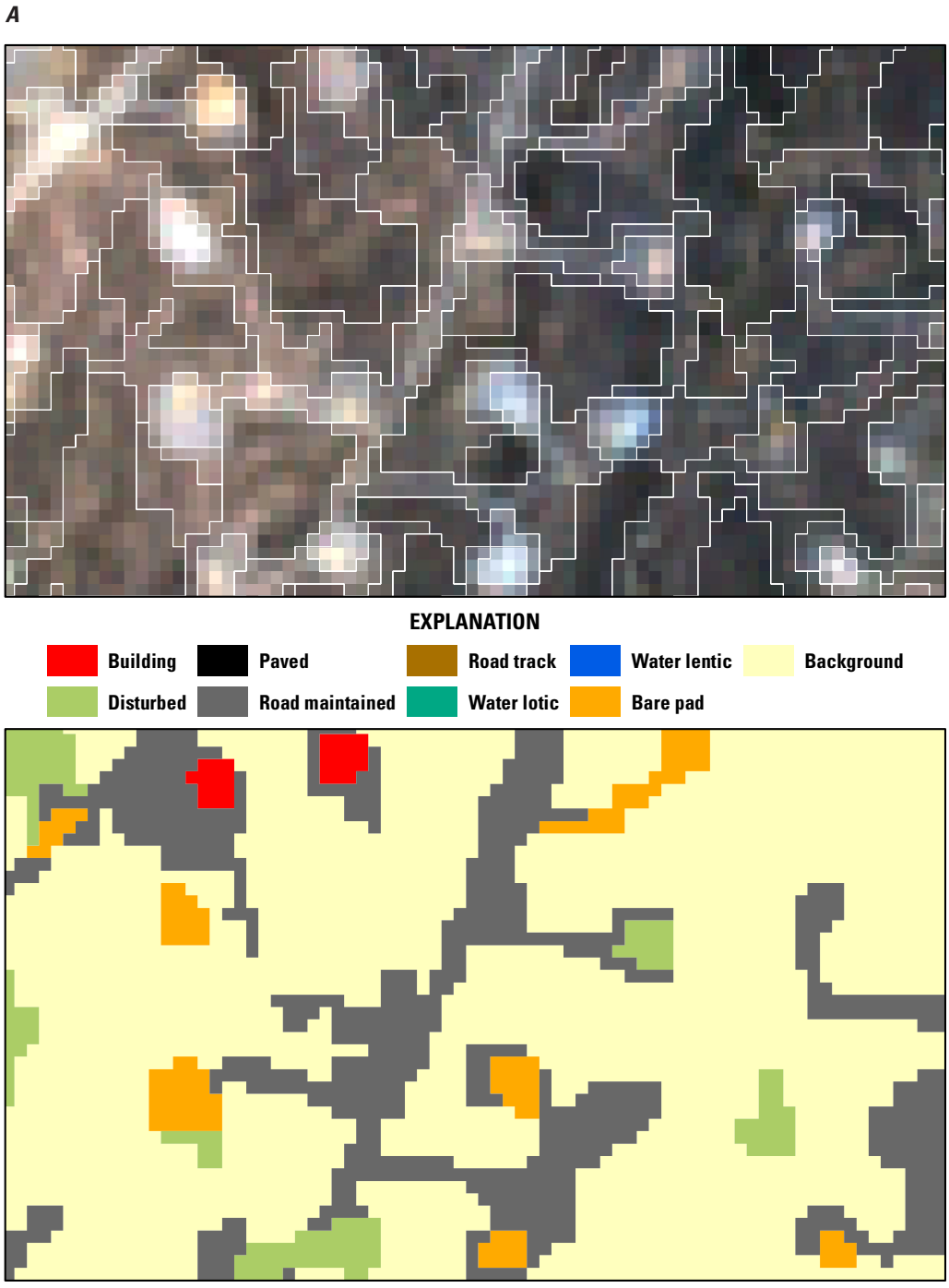

Figure 3. Computer screen captures of each of three image types (*A*, Landsat; *B*, SPOT; *C*, QuickBird) after automated extraction by eCognition software (top panels) and after post-processing (bottom panels). In each figure, top panels demonstrate the level of automated segmentation achieved, bottom panels demonstrate the post-processing classification achieved. Explanations of bottom panel symbology accompany each figure.

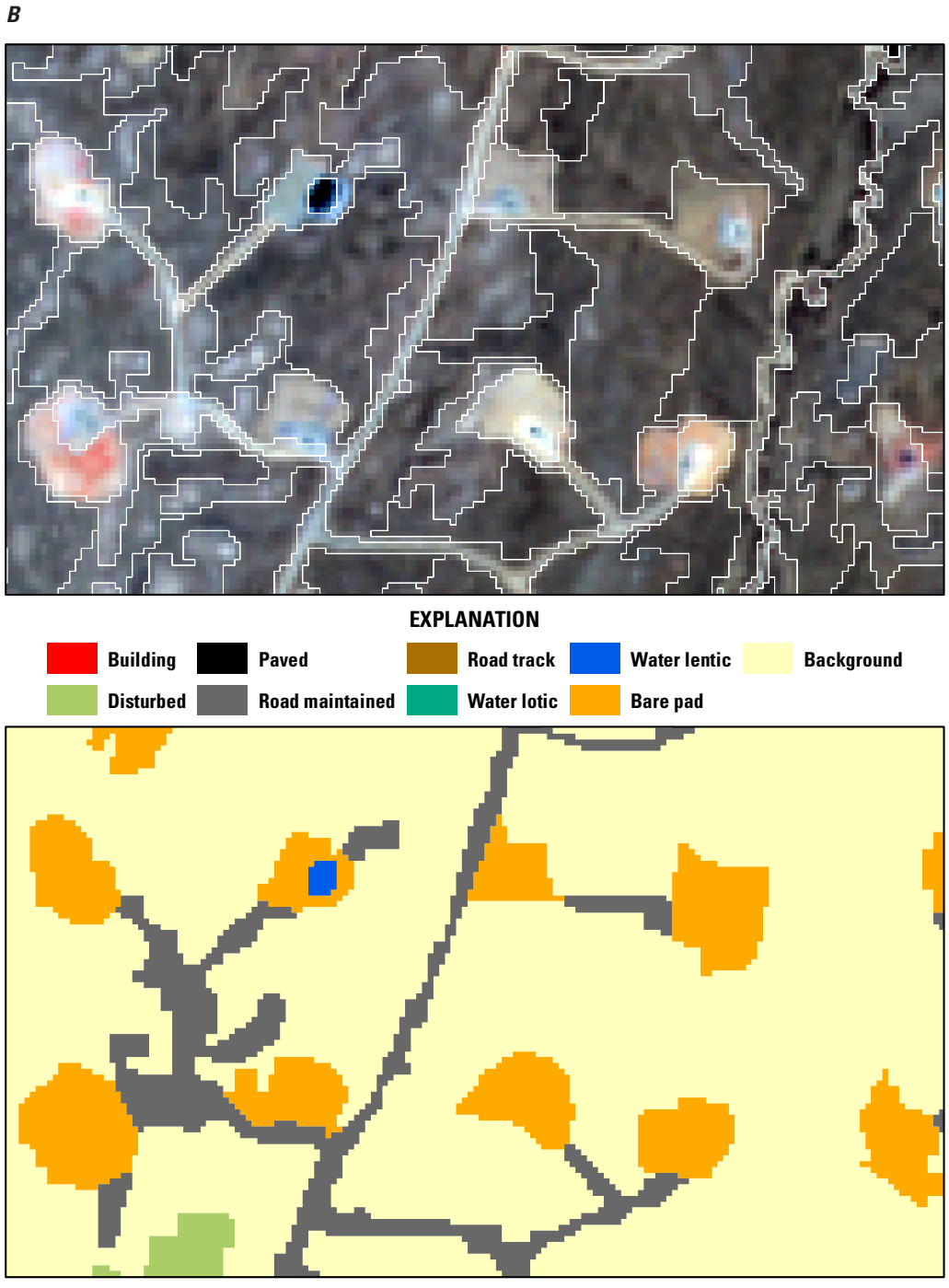

Figure 3. Computer screen captures of each of three image types (*A*, Landsat; *B*, SPOT; *C*, QuickBird) after automated extraction by eCognition software (top panels) and after post-processing (bottom panels). In each figure, top panels demonstrate the level of automated segmentation achieved, bottom panels demonstrate the post-processing classification achieved. Explanations of bottom panel symbology accompany each figure.—Continued

C

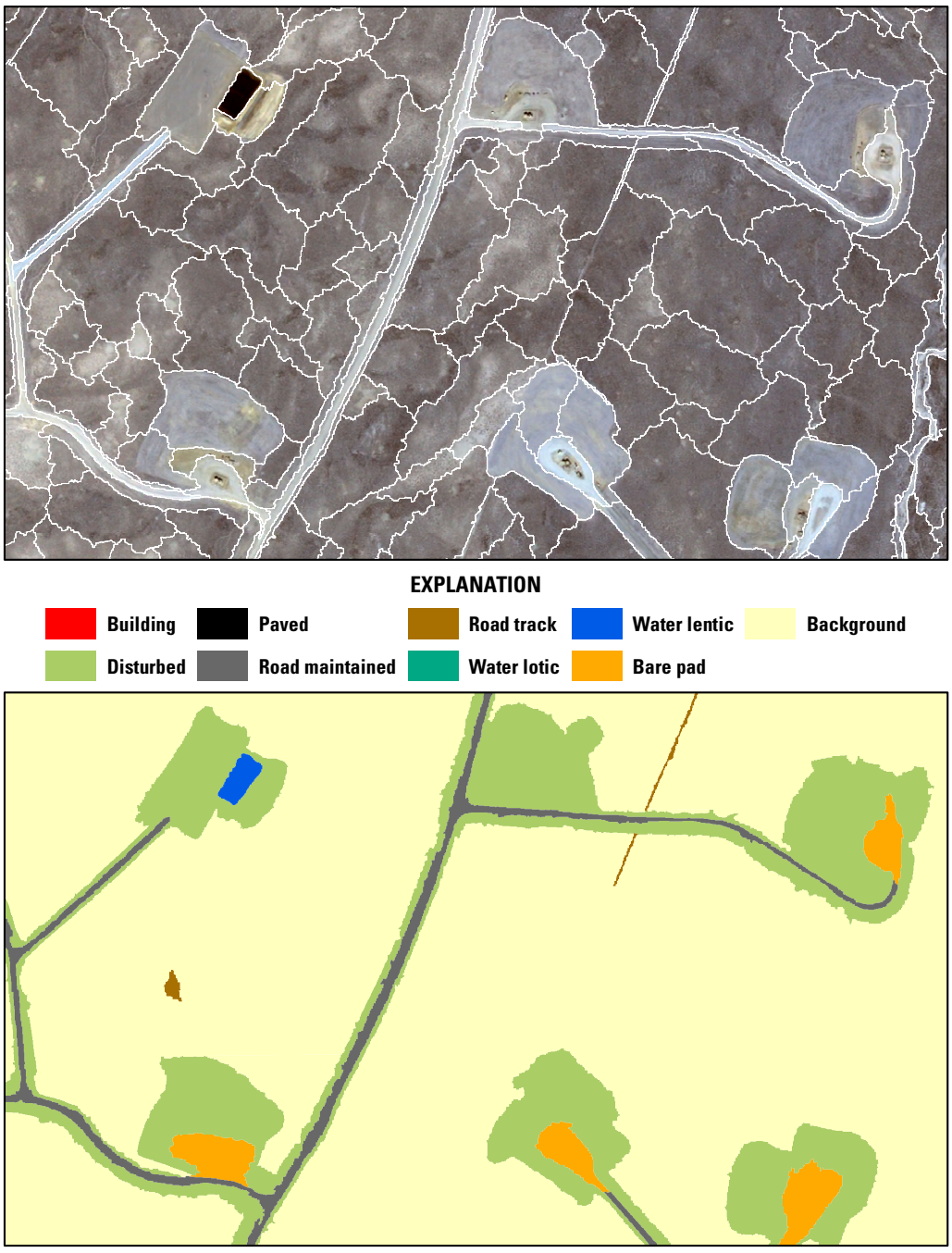

Figure 3. Computer screen captures of each of three image types (*A*, Landsat; *B*, SPOT; *C*, QuickBird) after automated extraction by eCognition software (top panels) and after post-processing (bottom panels). In each figure, top panels demonstrate the level of automated segmentation achieved, bottom panels demonstrate the post-processing classification achieved. Explanations of bottom panel symbology accompany each figure..—Continued

We manually ensured classification and topological accuracy for each extraction, unioned all feature class layers into an Arc coverage, then re-checked the coverage for topological and attributional accuracy. Once any errors were corrected, we performed a dissolve on the coverage using the final classification field as the dissolve item (fig. 4). The overlay and quality control tasks were completed using both ArcGIS Desktop 9.2 and ArcInfo Workstation 9.2.

ENVI Feature Extraction® (LSFX, SPFX, QBFX)

To extract Landsat image data using ENVI's Feature Extraction module, we first converted the image from ERDAS Imagine (ERDAS Imagine, 2009) format to ENVI format. To enhance image interpretation, we manipulated the following processing parameter values:

- Segmentation: segmentation groups pixels with neighboring pixels that have similar values for color, brightness, and related characteristics, and was set at 20 (scale range 0–100). The optimal value maximizes the contrast between the segmentation output and the image. Lower values result in a higher number of arcs and polygons, and higher values result in fewer arcs and polygons defining features in the image. The objective is to use the highest possible segmentation value that identifies features of interest.

- Merging: merging joins segmented pixel groups with neighboring groups to further optimize within-class feature delineation. We set the merge scale value to 61 (range 0–100). The value is adjusted until the operator's ability to visually distinguish merged segments from the image background is optimized. Low values result in fewer merged segments and higher polygonal complexity. Higher values result in simplified output. The objective is to use the highest possible merge value that adequately outlines features of interest.

We then manually classified all polygons into one of the eight feature classes and ensured classification and topological accuracy by converting the file to a coverage and by editing on-screen. We processed the SPOT and QuickBird data in a manner similar to that used for the Landsat imagery. For SPOT, we used a segmentation scale value of 40 and a merge scale value of 90 to reduce the time spent selecting and reclassifying polygons, despite increased time required for final editing. To speed up the manual classification for QuickBird, we divided the vector file derived from the entire image into six segments (fig. 5). We processed the QuickBird image using a segmentation scale level of 30 and a merge scale value of 90 (except for the southeast one-sixth of the study area where we used a segmentation scale level of 10 and a merge scale level of 80 to optimize feature delineation). We then appended the six segments of the QuickBird scene and manually edited all polygons on both the SPOT and QuickBird images in the same manner as described above (fig. 6).

Ground-Truthing

We collected ground control-point data (GCP) for comparison with feature classifications made during each information extraction process (fig. 7). During ground truthing, two observers drove all navigable roads during June–July 2007 and identified every disturbance feature encountered. At each feature, we took a digital photograph using Ricoh Caplio Pro G3® cameras (Ricoh Company Ltd., Tokyo, Japan), noted the feature class(es) present in the camera data dictionary, and logged the corresponding global positioning system (GPS) coordinate and photo orientation (in degrees azimuth) in a GPS receiver (GlobalSat SiRF STAR III model BC-337) embedded in each camera. To help organize driving routes, each vehicle was equipped with a Dell Inspiron laptop computer loaded with ArcGIS 9.2 (ESRI, 2006) software and the QuickBird image of the study site. Garmin hand-held GPS units were located on the dashboard of each vehicle and linked to the computers to enable us to manage travel routes through the study site in real time.

Post processing involved two main tasks: (1) organizing the inventory of disturbance features into classes and (2) corresponding the photograph point coordinates with the polygon-classification data derived from each method. Organizing the digital photograph data required creating a spatial reference layer, then applying stratified random selection criteria to select photos for use in the classification accuracy assessments. We used the GPS-Photo Link extension (GeoSpatial Experts, 2006) in ArcGIS 9.2 (ESRI, 2006) to create a point shapefile of the coordinate locations that linked to each of the 3,260 digital photographs. However, the cameras experienced a high rate of failure in assigning coordinates to photographs. Two problems occurred: (1) many photographs had no coordinates assigned and (2) serial sequences of photos were frequently assigned the same set of coordinates despite being at different locations. We removed all photos that were lacking coordinates and considered only the first photo in a series of repeated coordinates to have the correct assignment. We retained only those photos assigned unique and complete coordinate information.

We then randomly selected up to 30 photos per class (table 4), visually confirmed that each photo was correctly classified (on the QB imagery), and manually relocated each point to the exact location of the target feature. This resulted in 156 useable photo points for accuracy assessments. Manual relocation was necessary because the photographer was occasionally up to 200 m away from the object being photographed.

Prior to analysis, we topologically corrected each MIC layer as described above to achieve correct spatial alignment, then generated summary data and accuracy assessments using R (R Core Development Team, 2008) and Microsoft Excel software.

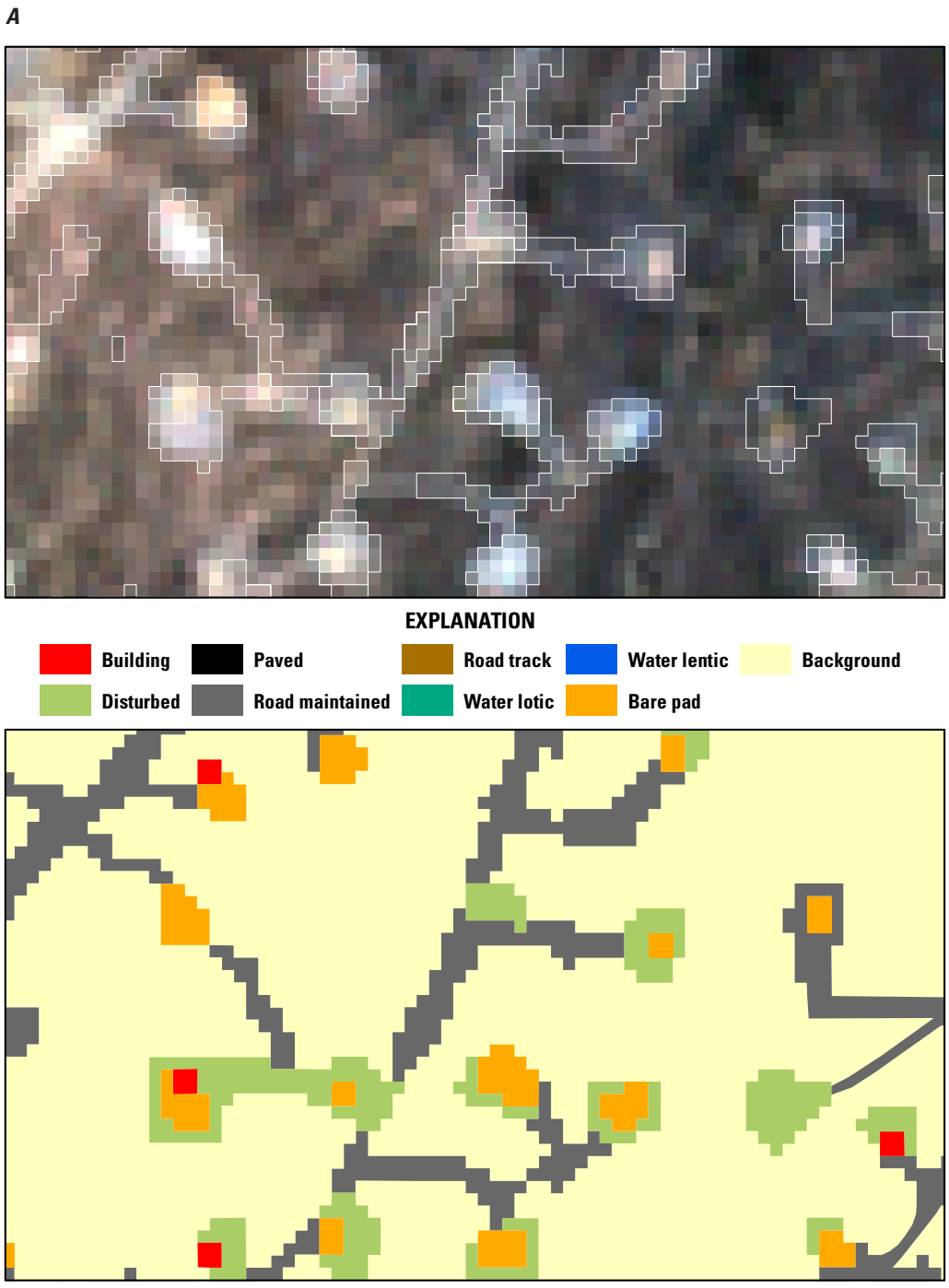

Figure 4. Computer screen captures of each of three image types (*A*, Landsat; *B*, SPOT; *C*, QuickBird) after automated extraction by Feature Analyst software (top panels) and after post-processing (bottom panels). In each figure, top panels demonstrate the level of automated segmentation achieved, bottom panels demonstrate the post-processing classification achieved. Explanations of bottom panel symbology accompany each figure.

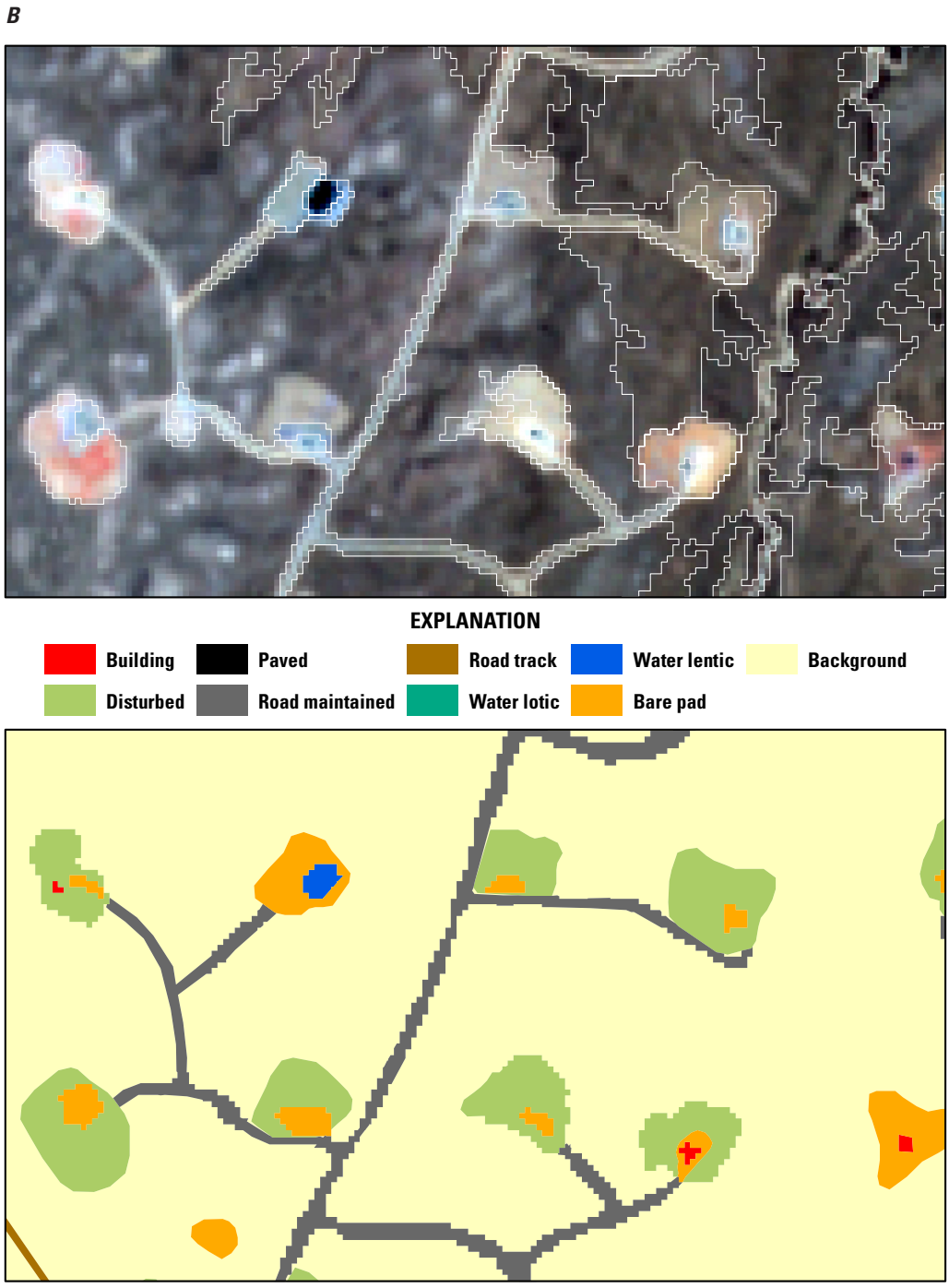

Figure 4. Computer screen captures of each of three image types (*A*, Landsat; *B*, SPOT; *C*, QuickBird) after automated extraction by Feature Analyst software (top panels) and after post-processing (bottom panels). In each figure, top panels demonstrate the level of automated segmentation achieved, bottom panels demonstrate the post-processing classification achieved. Explanations of bottom panel symbology accompany each figure.

C

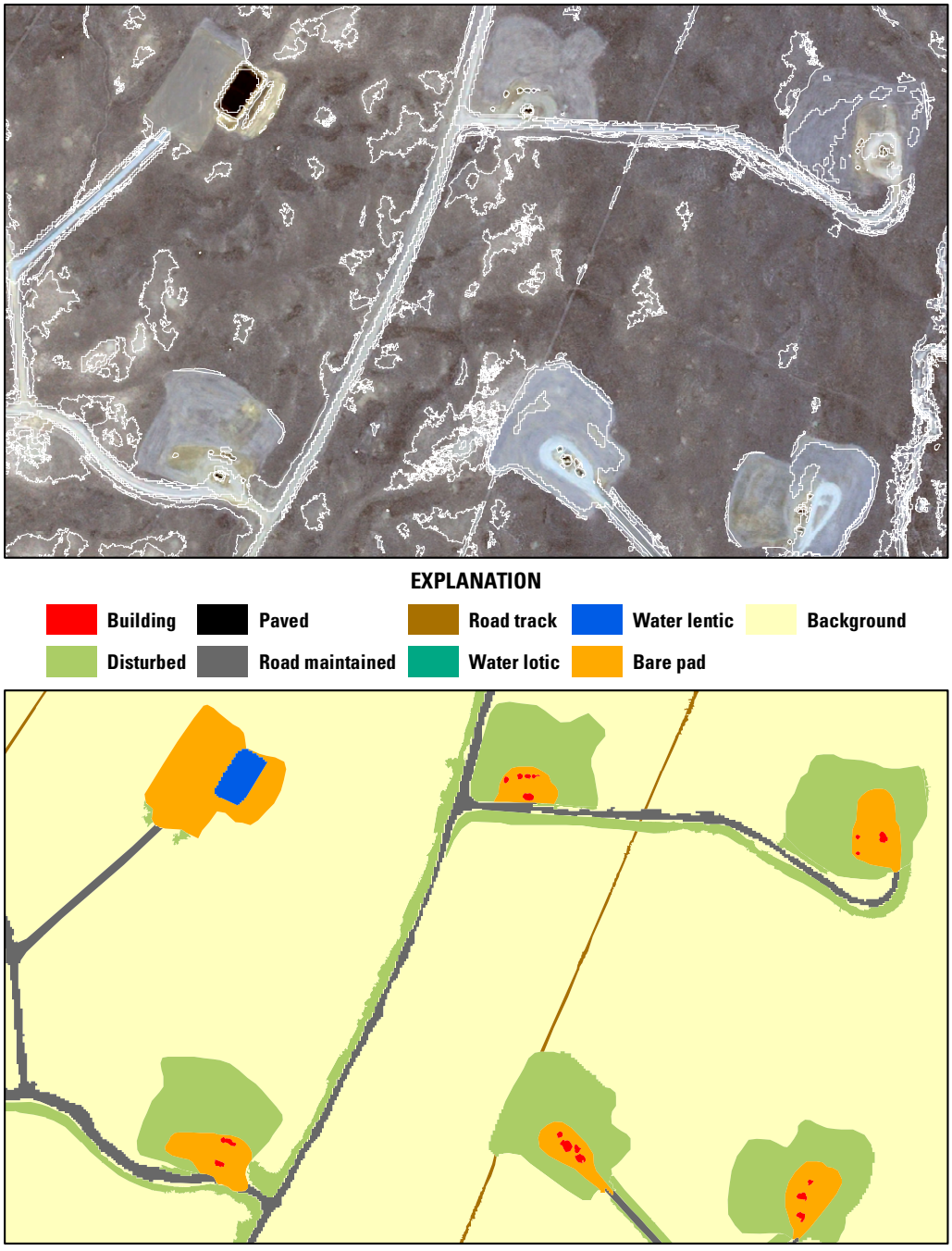

Figure 4. Computer screen captures of each of three image types (*A*, Landsat; *B*, SPOT; *C*, QuickBird) after automated extraction by Feature Analyst software (top panels) and after post-processing (bottom panels). In each figure, top panels demonstrate the level of automated segmentation achieved, bottom panels demonstrate the post-processing classification achieved. Explanations of bottom panel symbology accompany each figure.—Continued

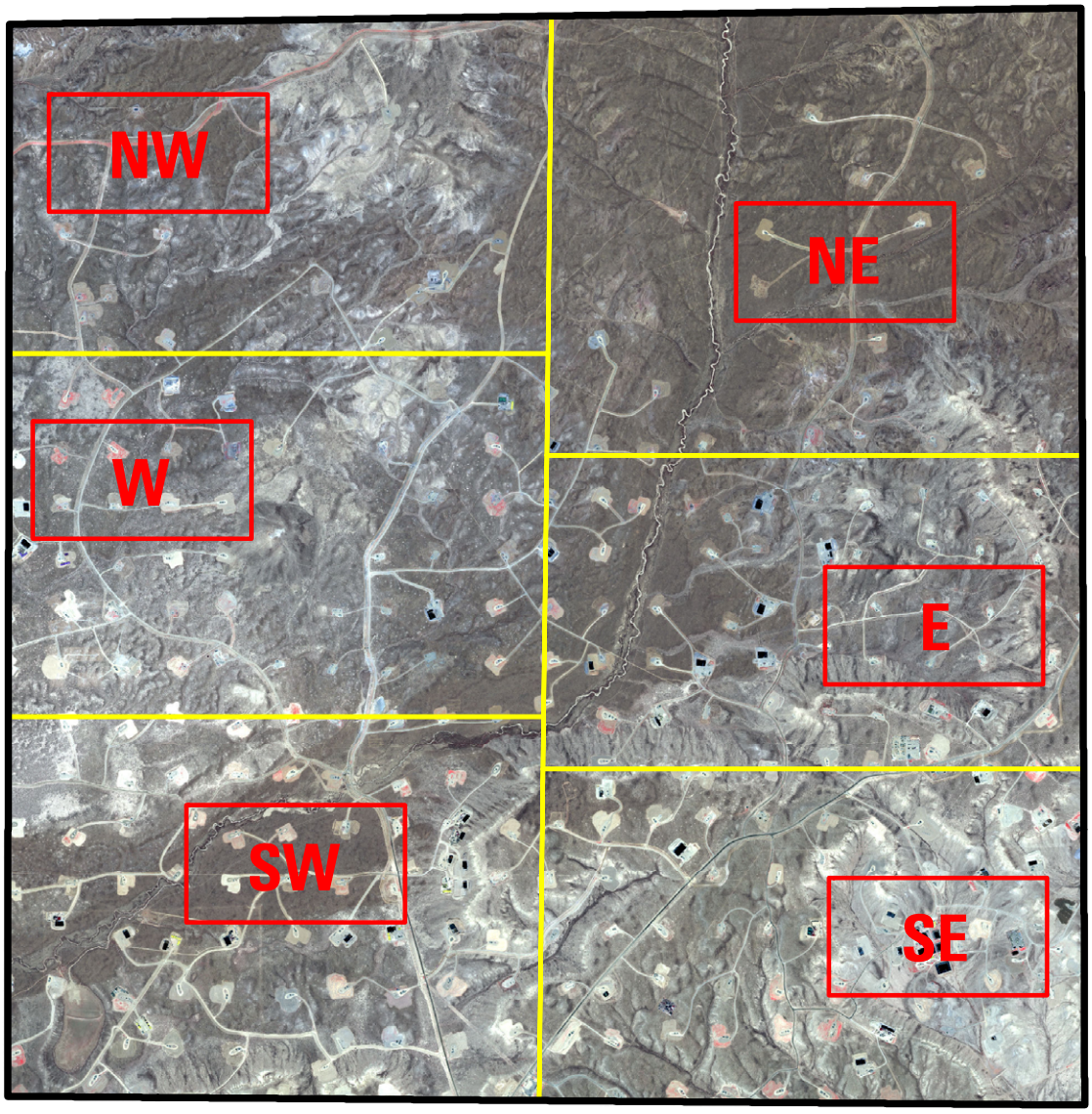

Figure 5. Segmentation of the QuickBird scene prior to automated information extraction process by ENVI Feature Extraction software. Shown are the six areas (yellow outline) within which the vector polygons were manually classified after inserting the classified vector data for each of the six approximately 148.5-hectare subset areas (red outline).

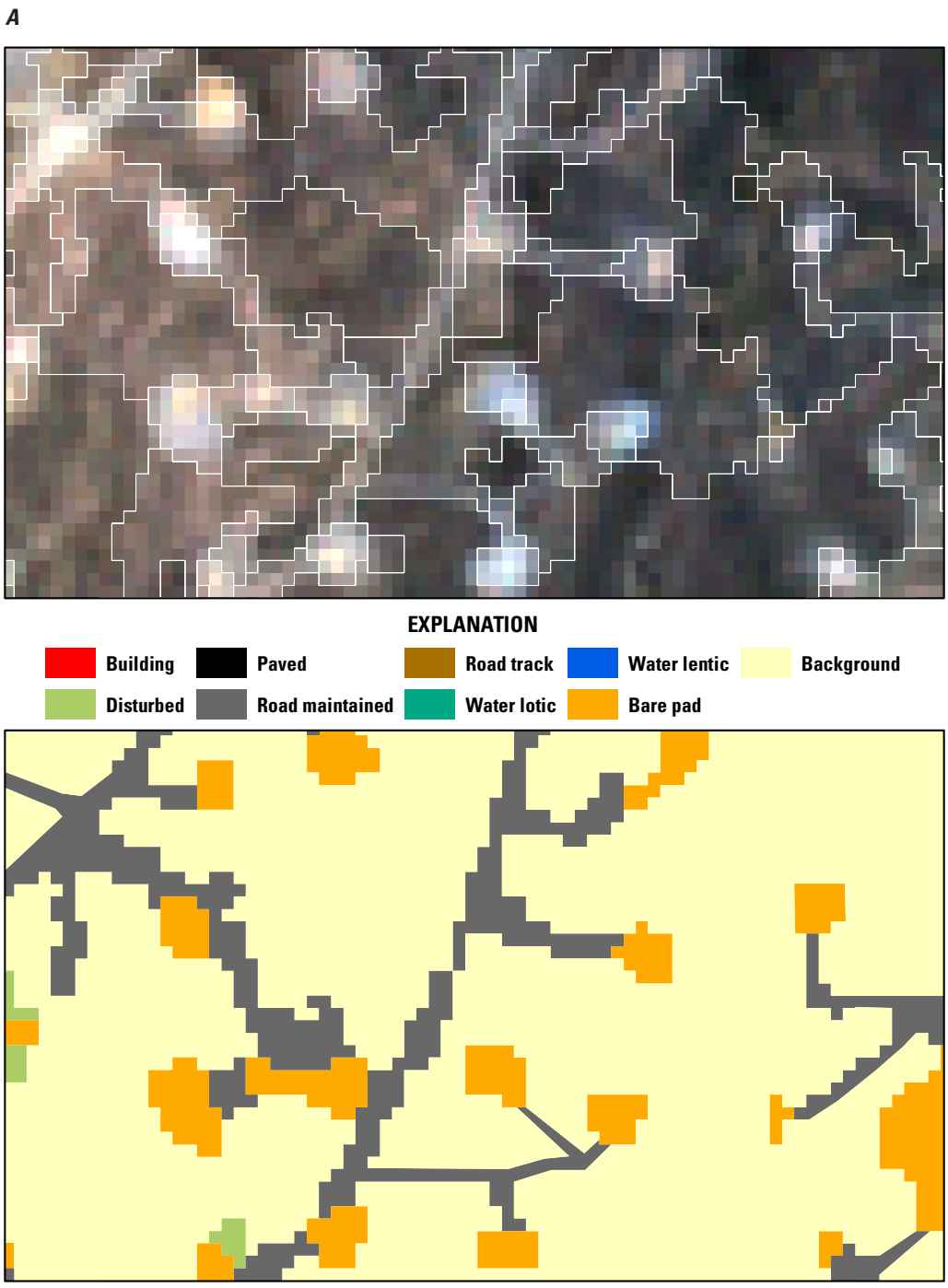

Figure 6. Computer screen captures of each of three image types (*A*, Landsat; *B*, SPOT; *C*, QuickBird) after automated extraction by ENVI Feature Extraction software (top panels) and after post-processing (bottom panels). In each figure, top panels demonstrate the level of automated segmentation achieved, bottom panels demonstrate the post-processing classification achieved. Explanations of bottom panel symbology accompany each figure.

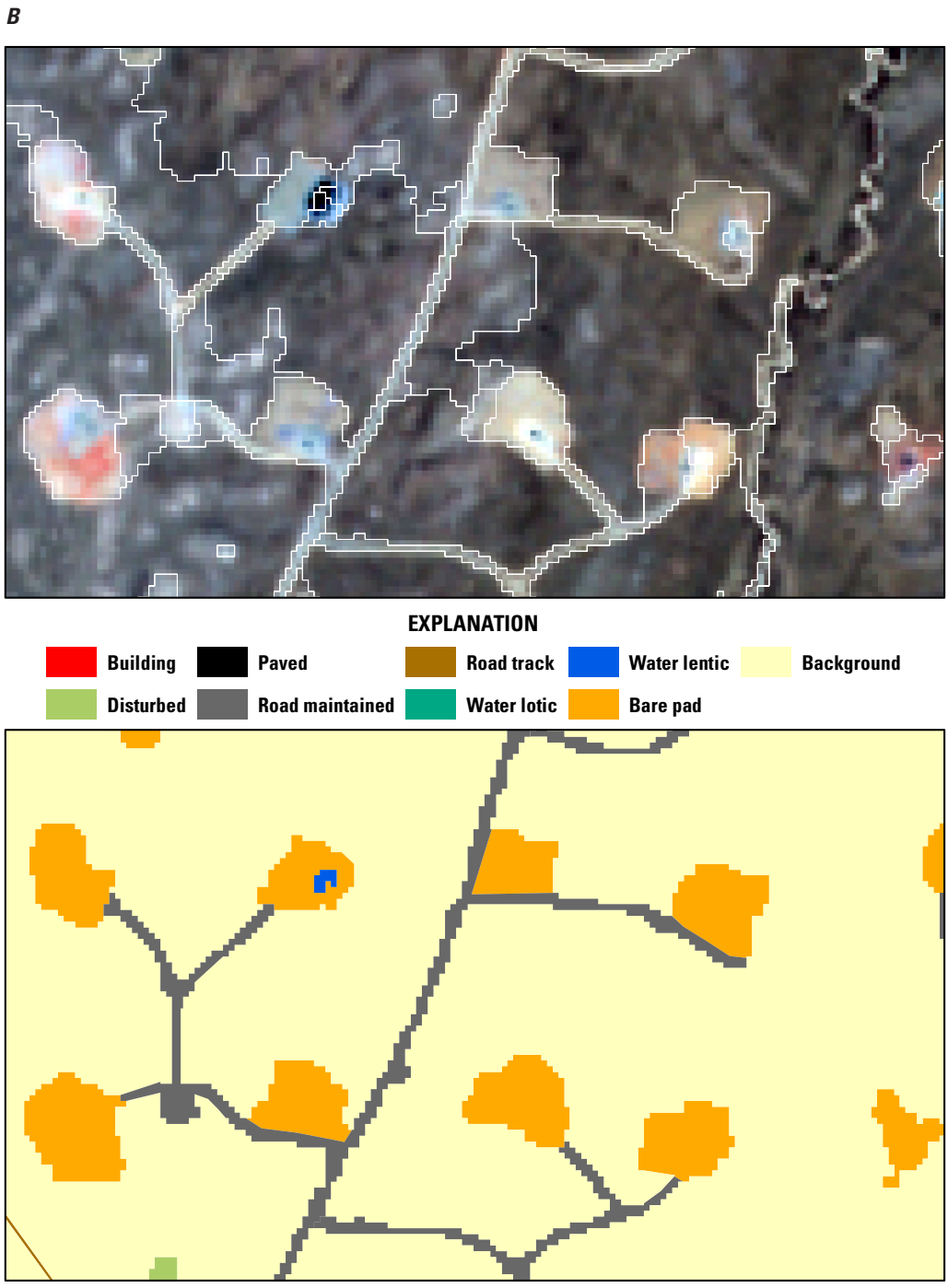

Figure 6. Computer screen captures of each of three image types *(A*, Landsat; *B*, SPOT; *C*, QuickBird) after automated extraction by ENVI Feature Extraction software (top panels) and after post-processing (bottom panels). In each figure, top panels demonstrate the level of automated segmentation achieved, bottom panels demonstrate the post-processing classification achieved. Explanations of bottom panel symbology accompany each figure.—Continued

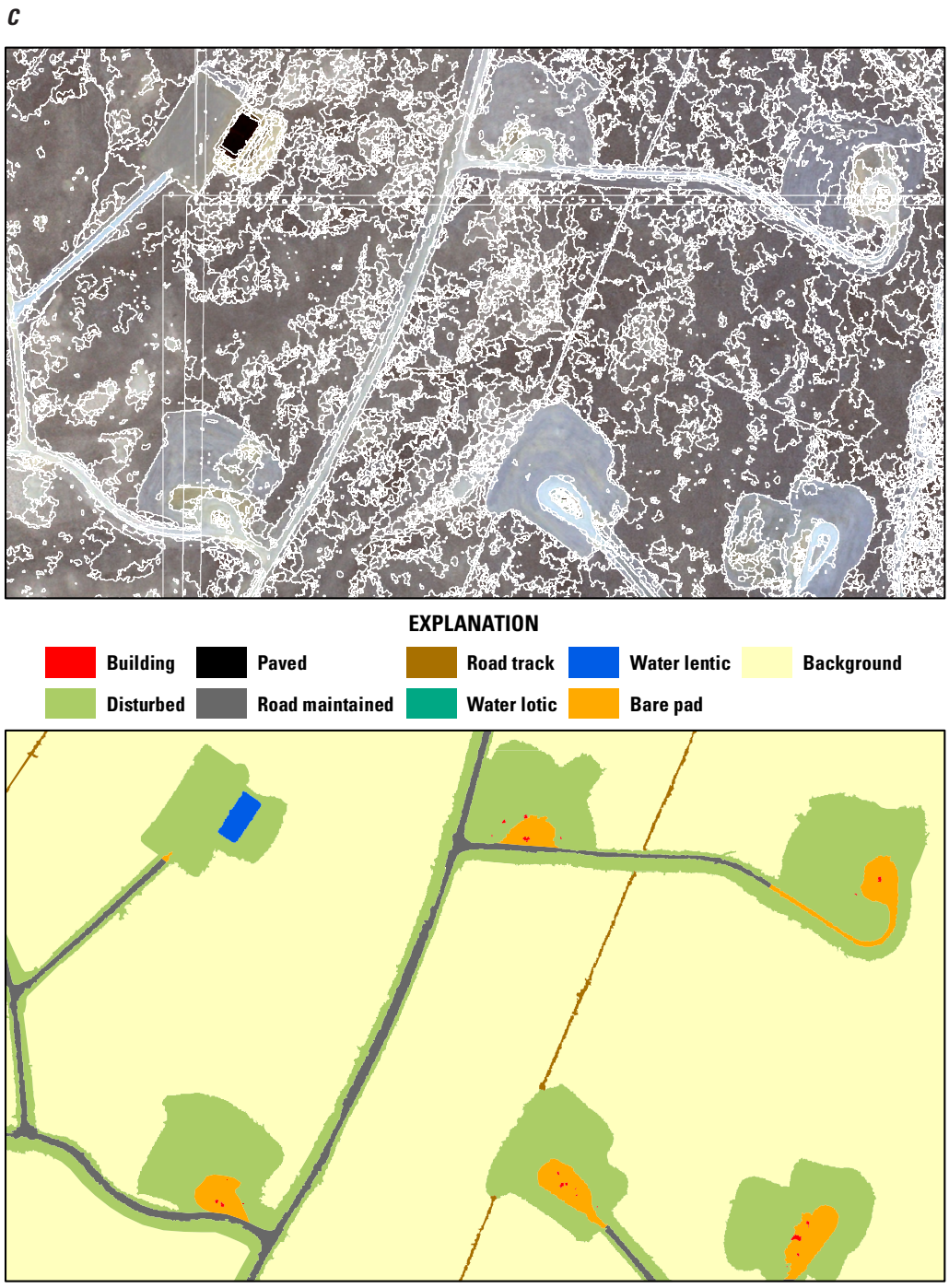

Figure 6. Computer screen captures of each of three image types (*A*, Landsat; *B*, SPOT; *C*, QuickBird) after automated extraction by ENVI Feature Extraction software (top panels) and after post-processing (bottom panels). In each figure, top panels demonstrate the level of automated segmentation achieved, bottom panels demonstrate the post-processing classification achieved. Explanations of bottom panel symbology accompany each figure.—Continued

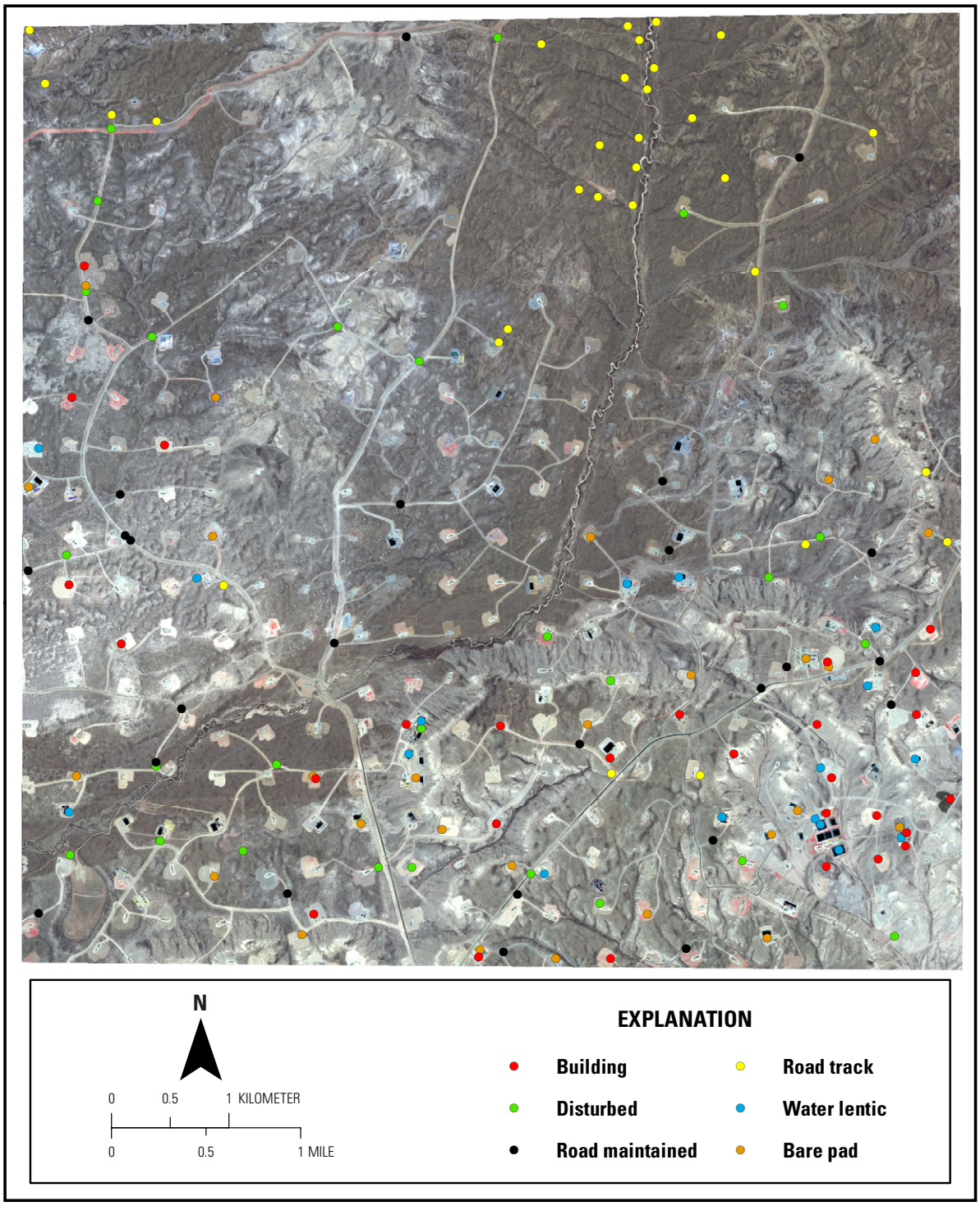

Figure 7. Distribution of ground control photograph reference points on our Jonah gas field study area, Sublette County, Wyo.

Table 4. Ground control point photographs used for classification-accuracy assessment among extraction method–image combinations.

Feature class	n	Other
Building	29	
Disturbed	27	
Paved	0	Not observed during ground truthing
Road_maintained	25	
Road_track	30	
Water_lotic	0	Not observed during ground truthing
Water_lentic	17	
Bare_pad	28	
Background	0	

Analysis

We compared 12 MIC combinations (table 5), focusing primarily on within-image resolution comparisons, because the relationship between image resolution and specific detail is fundamentally well understood. Prior to the comparisons, we assessed the degree of spatial autocorrelation present in our ground control photo data using a Moran's I test (Congalton, 1991). Our first comparison (Objective 1) was of two measures of the area of mapped features among competing MICs, because neither measure gave the "complete picture" about the spatial geometry of disturbance features. The first measure was a comparison of the physical area mapped in each feature class by each MIC. To provide information on spatial agreement of mapped features among MICs, we also assessed the geometric congruence of polygons among MICs using a spatial overlap statistic (Jacquez, 1995). This statistic measured the area of polygon non-overlap generated by each MIC when compared against a reference map and summed the area of footprint dissimilarity (that is, over-mapping + under-mapping error) within each feature class (fig. 8).

For both of these comparisons, we evaluated the degree of variation that existed among MICs for each feature class and noted the MICs that performed best at mapping features within each image resolution. We also clustered MICs into groups based on similarity of area-mapping performance to gain a multivariate assessment of MIC similarity. To cluster MICs, we standardized each data set (mean = 0, SD = 1), calculated Pearson correlation coefficients, and used the complete linkage algorithm to generate tree diagrams (Wilkinson and others, 2007). We determined the optimal number of clusters generated in each diagram by evaluating root mean square standard deviation, pseudo-F, and pseudo-T indices (Wilkinson and others, 2007). We considered QBHD our reference dataset against which we evaluated other within-resolution MICs, because hand digitization has traditionally been the industry-standard method of delineating image features and because of the high level of detail visible on the QB scene.

To account for edge incongruities resulting from data-format conversions during feature extraction and from image registration errors, we applied fuzzy tolerance buffers (Chrisman, 2001) to each MIC map while evaluating area of footprint dissimilarity values. Fuzzy tolerances reduce feature edge-location dissimilarities as long as they fall within a specified tolerance distance. Based on calculated residual mean square errors for image misregistration, we set tolerances at 2.5 m for QB, 15 m for SP, and 30 m for LS. We then graphed the deviations between the QBHD measure of total area of disturbance and that of each other MIC to visually assess the

Table 5. Information extraction method–image resolution combinations and in-text acronyms used in this report.

Remotely sensed imagery	eCognition	Feature analyst	Feature extraction	Hand-digitized
QuickBird	QBEC	QBFA	QBFX	QBHD
SPOT (Satellite Pour l'Observation de la Terre)	SPEC	SPFA	SPFX	SPHD
Landsat TM	LSEC	LSFA	LSFX	LSHD

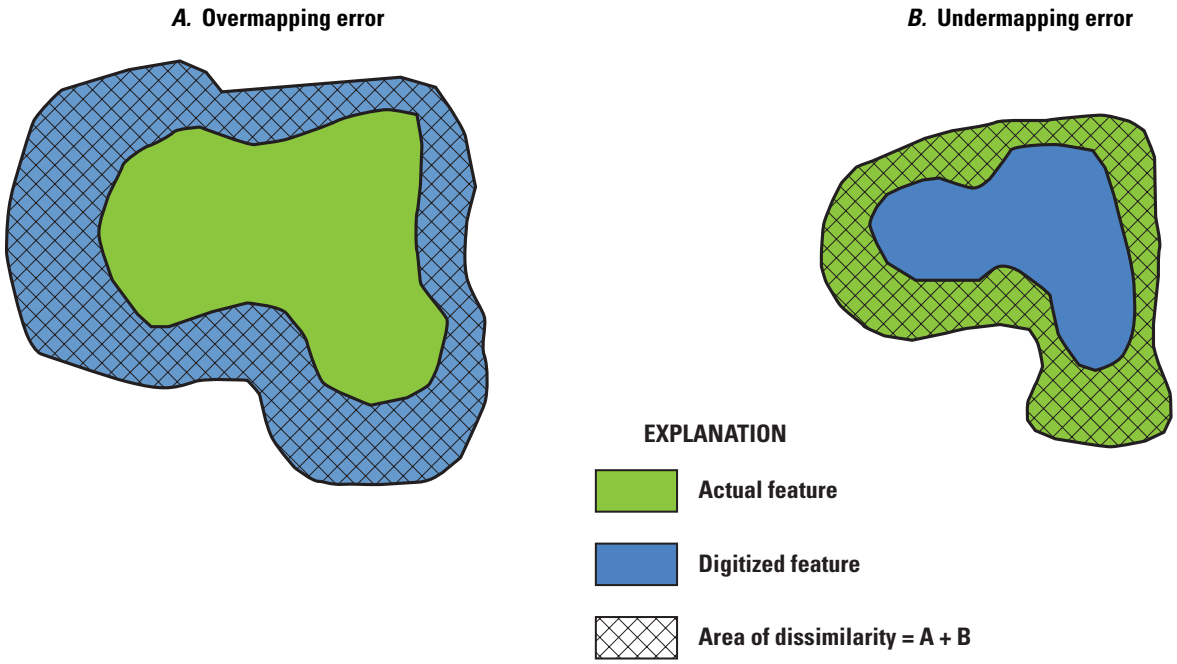

Figure 8. Illustration of spatial mapping errors associated with feature delineation from remotely sensed imagery. *A*, Overmapping error occurs when the mapped footprint of a feature is larger than the actual size of the feature; *B*, undermapping error occurs when the mapped footprint is smaller than reality.

relative accuracy of each at identifying amount of disturbance present.

We next evaluated the classification accuracy of each MIC for each type of disturbance (Objective 2). First, for each disturbance feature selected in our random GCP sample, we compared the class assignment against the assignment made to the corresponding polygon (or pixel) of each MIC. Because of geometric registration inaccuracies and because MIC products were developed using different image resolutions, we used fuzzy tolerances to account for spatial offset between the GCPs and each MIC, and set tolerance values at 10-, 45-, and 60-m radii for QuickBird, SPOT, and Landsat, respectively. Second, we rasterized all of the MIC products and compared the feature class assigned to each pixel using HD as our within-resolution benchmark. To minimize error during rasterization, we resampled the raster output data at a 2-m resolution and snapped all data sets to the QB scene so that all had identical column and row pixel dimensions and co-registration. We again applied a fuzzy tolerance adjustment this time setting tolerance values to 3, 7, and 15 m respectively for QuickBird, SPOT, and Landsat.

For both GCP and pixel-by-pixel comparisons, we summed classification success values across feature classes within each MIC, then calculated overall Cohen's Kappa index scores and plus or minus 95 percent confidence interval (CI) estimates, and considered differences between MICs significant when CIs did not overlap. We also calculated conditional Kappa values to assess the level of classification agreement between each MIC and the reference data within each feature class. For each misclassification, we tabulated omission and commission error rates, where any feature identified in the reference data but not by an MIC was considered an omission for that MIC, and any feature identified by an MIC but not by the reference data was considered a commission error.

Kappa index values range from 0 (no relationship) to 1 (perfect agreement), indicating the level of inter-analyst (that is, MIC values versus reference values) agreement while adjusting for agreement that might occur by chance (Cohen, 1960; Fleiss, 1971). Because no single interpretation of kappa values is universally accepted (all are arbitrarily derived), we used a conservative interpretation, considering Kappa values greater than 0.75 to indicate strong agreement, values between 0.40–0.75 to indicate fair agreement, and values less than 0.40 to indicate poor agreement, after an interpretation suggested by Bowman and Shetty (2007).

Next, we conducted a cost-benefit assessment of each MIC (Objective 3). Our goal was to provide a comparison of cost (in terms of time) and benefits (in terms of map classification quality) to guide others who may wish to map landscapes in similar ways. Therefore, each GIS analyst recorded the

amount of time required to extract and classify all of the information from each map image. We then plotted two estimators of product quality (area of footprint dissimilarity values and kappa scores associated with the overall GCP classification accuracy assessment) against image-processing time, and visually assessed the scatter of values across plots. Spatial analyses and tabulations were performed in program R (R Development Core Team, 2008), while cluster analyses were conducted in Systat (Systat Software Inc., 2007).

To discriminate among MICs, we ranked them based on the five metrics we evaluated: accuracy while mapping the area of disturbance, footprint dissimilarity error values, GCP-based feature classification success, pixel-based classification success, and the amount of time required to complete each extraction. For each MIC, we scored area of disturbance and footprint dissimilarity as: - number of cases of most-accurate within-resolution measurement / (number of cases of least-accurate within-resolution measurement + 1); we scored each feature-classification metric as:- number of cases of strong Kappa agreement / (number of cases of poor agreement + 1); and scored time as the number of hours required to complete each MIC extraction. We then ranked each MIC based on scores for each metric (1 = highest rank, 4 = lowest rank), summed overall scores across the five metrics, and assigned final rankings within each image resolution.

Results

Based on the QBHD reference values, 10.62 km^2 (16.6 percent) of land in our study area had been altered by some form of disturbance, and approximately 67 percent of this land fell in the Disturbed land class (see table 2 to review feature class designations). Bare_pad (18 percent), Road_maintained (9 percent) and Road_track (5 percent) were the next most prevalent disturbance types. Road_paved and Water_lotic were identified by one MIC each but were not observed in our study area during ground truthing (table 6).

Mapped Area of Disturbance Features

There was a great deal of variation in accuracy among MICs while mapping the area of individual features, although within-resolution similarity was greatest on the QB image and lower on the other images (table 6). Bare_pad, Disturbed, and Road_maintained accounted for the majority of area mapped for all 12 of the MICs, but the relative proportion of these features changed among image resolutions (fig. 9).

On the LS imagery, EC and FX each mapped two types of features (Disturbed and Water_Lentic; Building and Road_maintained, respectively) most accurately relative to the QBHD reference, while FA and HD each mapped one feature (Bare_pad and Building, respectively) most accurately (table 6). EC and HD were each least accurate at mapping two of the features, while FA was least accurate at mapping one, and FX was not least accurate while mapping any of the features at this image resolution.

On the SP imagery, HD mapped three features most accurately (Building, Road_maintained, and Road_track), while EC, FA, and FX each mapped one feature most accurately (Disturbed, Bare_pad, and Water_lentic, respectively). Least accurate were EC (four features), FX (two features), then HD (one feature), while FA was not least accurate at mapping any features on the SP image. On the QB imagery, FX mapped four features most accurately (Building, Road_maintained, Road_track, and Water_lentic, while EC and FA each mapped one type of feature most accurately (Disturbed and Bare_pad, respectively). Least accurate were EC and FA (three features each; table 6).

Pooling across individual feature classes, EC mapped disproportionately higher total measures of disturbance than the other methods on the LS and SP imagery (fig. 10). In contrast, FA, FX, and HD measured relatively similar total levels of disturbance on the SP and LS imagery, and all methods measured similar total levels of disturbance on the QB imagery. However, using the QBHD measure of all disturbed land area (10.62 km^2) as reference, five of the 11 other MICs (QBEC, QBFA, QBFX, SPFA, SPHD) mapped area of disturbance with an accuracy within plus or minus 10 percent (fig. 10).

Four clusters were generated in the multivariate analysis of the amount of area measured in each feature class by each MIC (fig. 11). Cluster one contained the four QB MICs, all of which measured area of disturbance among disturbance features in a similar manner. The nearest outgroup cluster contained SPFA and SPEC, indicating that these MICs mapped the area of features in a manner more similar to the QB cluster than did the remaining MICs. Cluster three contained LSFX, LSHD, SPFX, and SPHD, while LSEC and LSFA comprised cluster four.

Dissimilarity of Mapped Disturbance Features

The majority of footprint dissimilarity error was contributed by the Disturbed and Bare_pad feature classes on the QB and SP imagery, while in addition to these, Road_maintained and Road_track contributed large amounts of footprint dissimilarity error to the MICs on the LS image (fig. 12). On the LS image, FA and HD each most-accurately mapped the spatial location of three features (for FA: Building, Water_lentic, Bare_pad, for HD: Disturbed, Road_maintained, Water_lentic; table 7). While all four methods mapped Road_track with equal accuracy, neither EC nor FX outperformed the other MICs at mapping any LS image features. On the SP image, HD most accurately mapped the spatial location of four features (Building, Disturbed, Road_maintained, Water_lentic, while the other three MICs optimally mapped one feature each. On the QB image, FX most accurately mapped the spatial location of four features (Building, Disturbed, Water_lentic, and Bare_pad), while EC and FA each mapped

Table 6. Measurements of area of disturbance (km²) by disturbance-feature class for 12 information extraction method–image resolution combinations. (km², square kilometers; - , no data)

	Landsat				SPOT				QuickBird				CV[a]
Class	LSEC	LSFA	LSFX	LSHD	SPEC	SPFA	SPFX	SPHD	QBEC	QBFA	QBFX	QBHD*	
Building	0.69	0.14	**0.00**	**0.00**	0.23	0.30	0.00	**0.01**	0.02	0.15	**0.03**	0.04	1.51
Disturbed	**6.20**	3.30	0.68	0.04	**8.49**	4.53	0.88	1.23	**7.39**	6.48	7.45	7.08	0.69
Road_paved	0.00	0.00	0.00	0.00	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.46
Road_maintained	14.57	7.08	**6.17**	6.40	4.85	2.88	2.86	**2.68**	1.00	1.10	**0.97**	0.93	0.92
Road_track	0.00	0.00	0.00	0.00	0.15	0.27	0.14	**0.45**	0.23	0.45	**0.48**	0.56	0.93
Water_lotic	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00[b]	0.00	-
Water_lentic	**0.16**	0.06	0.19	0.19	0.00	0.22	**0.13**	0.18	0.11	0.14	**0.13**	0.13	0.45
Bare_pad	2.62	**1.95**	5.39	7.37	3.79	**3.34**	5.10	5.86	1.02	**1.79**	1.50	1.88	0.59
Total	24.24	12.53	**12.44**	13.99	17.77	11.55	9.11	**10.4**	9.78	10.11	**10.58**	10.62	
Average	3.03	1.57	1.56	1.75	2.22	1.44	1.14	1.30	1.22	1.26	1.32	1.33	
Std. deviation	5.14	2.54	2.63	3.18	3.18	1.83	1.88	2.06	2.53	2.20	2.54	2.41	

* Reference values. [a] Coefficient of variation. [b] QBFX reported 0.00004 km² of Water_Lotic as present on the study area.

Boldface identifies the value(s) within each image resolution for each feature class (across rows) that is closest to the QBHD reference value, except in cases where greater than two ties exist for any within-class value. Underline identifies the value(s) within each image resolution for each feature class (across rows), that is farthest from the QBHD reference value.

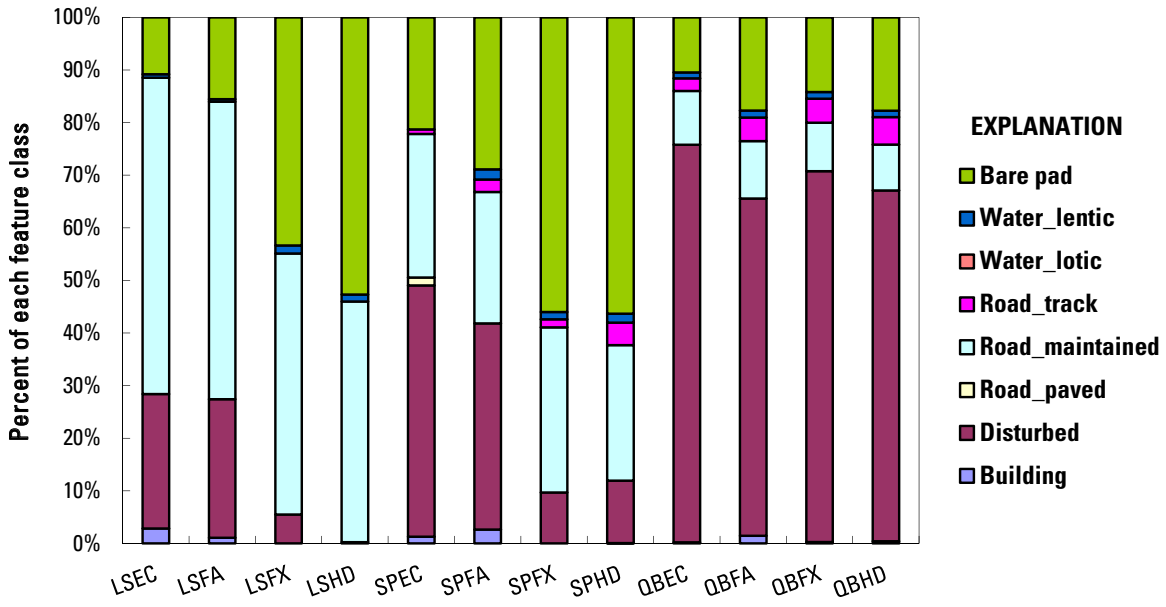

Figure 9. Proportional amount of area classified into each of eight disturbance classes by each of 12 information extraction method–image resolution (MIC) combinations. Differing colors within bars represent percent of total area contributed by each feature class (see explanation, right side of figure). (For each four-letter code (X axis labels): LS, Landsat; SP, SPOT; QB, QuickBird; EC, eCognition; FA, Feature Analyst; FX, Feature Extraction; and HD, hand digitization.)

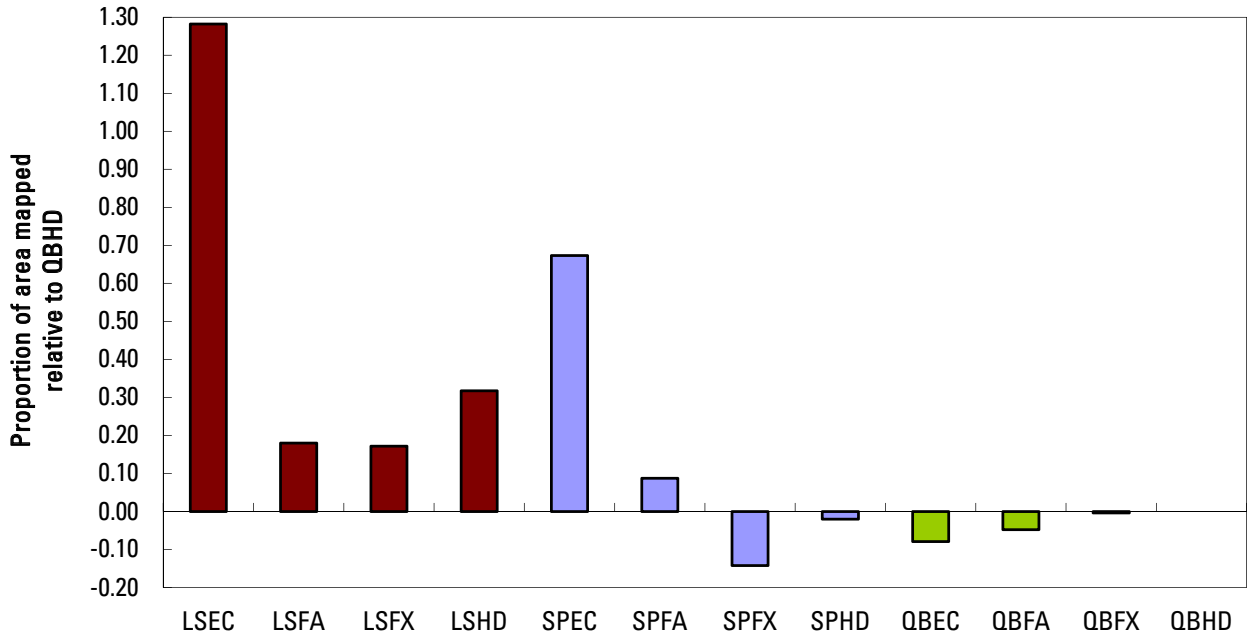

Figure 10. Amount of disturbance (all disturbance-feature classes pooled) mapped by each of 11 information extraction method–image resolution combinations (MICs) relative to the QBHD reference data. (For each four-letter code (X axis labels): LS, Landsat; SP, SPOT; QB, QuickBird; EC, eCognition; FA, Feature Analyst; FX, Feature Extraction; and HD, hand digitization.)

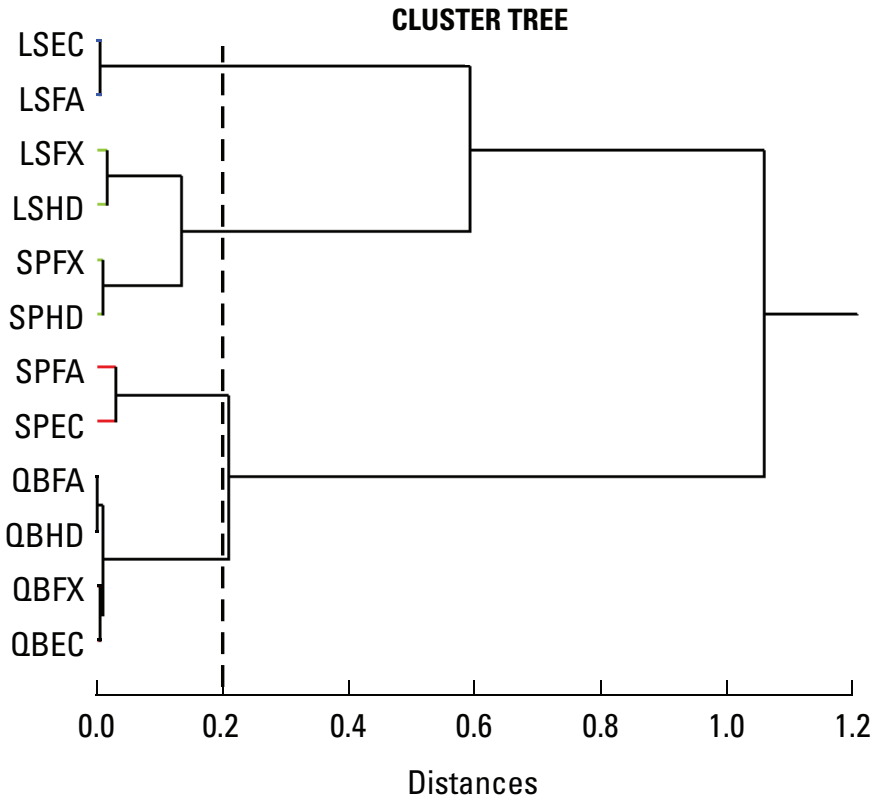

Figure 11. Tree diagram depicting extraction method–image resolution (MIC) groupings based on patterns of area mapped in each disturbance-feature class. Root mean squared standard deviation, pseudo F, and pseudo T indices (not shown) indicated four distinct clusters, identifiable by assuming a vertical line (shown) at distance = 0.2. Branch tips (left side of tree) representing each cluster are colored differently. (For each four-letter code (X axis labels): LS, Landsat; SP, SPOT; QB, QuickBird; EC, eCognition; FA, Feature Analyst; FX, Feature Extraction; and HD, hand digitization.)

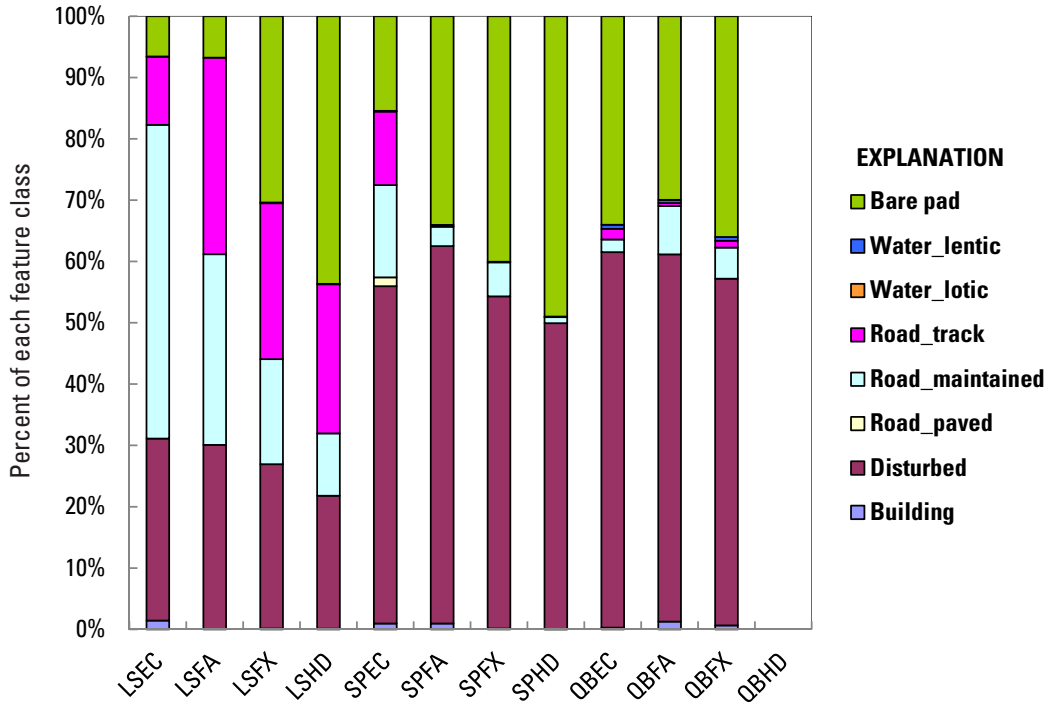

Figure 12. Degree of footprint dissimilarity present in each of 11 information extraction–image resolution combinations (MICs) when superimposed on the QBHD reference data set. Differing colors within bars represent percent of error contributed by each feature class in each MIC (see explanation, right side of figure). (For each four-letter code (X axis labels): LS, Landsat; SP, SPOT; QB, QuickBird; EC, eCognition; FA, Feature Analyst; FX, Feature Extraction; and HD, hand digitization.)

Table 7. Measures of footprint dissimilarity error (km²) among 11 information extraction method–image resolution combinations and the QBHD reference map. (km², square kilometers; - , no data)

	Landsat				SPOT				QuickBird				CV[a]
Class	LSEC	LSFA	LSFX	LSHD	SPEC	SPFA	SPFX	SPHD	QBEC	QBFA	QBFX	QBHD*	
Building	0.26	**0.00**	0.01	0.01	0.18	0.06	0.01	**0.00**	**0.01**	0.03	**0.01**	0.00	1.71
Disturbed	5.42	1.90	2.14	**1.81**	10.21	3.63	3.75	**3.36**	2.02	1.60	**0.97**	0.00	0.87
Road_paved	0.00	0.00	0.00	0.00	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-
Road_maintained	9.35	1.97	1.37	**0.85**	2.80	0.19	0.38	**0.07**	**0.07**	0.21	0.09	0.00	1.83
Road_track	2.03	2.03	2.03	2.03	2.23	0.00	0.00	0.00	0.06	**0.01**	0.02	0.00	1.22
Water_lotic	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Water_lentic	0.01	**0.00**	0.01	**0.00**	0.02	0.02	**0.00**	**0.00**	0.02	**0.01**	**0.01**	0.00	0.77
Bare_pad	1.20	**0.43**	2.42	3.64	2.87	**2.01**	2.77	3.30	1.12	0.80	**0.62**	0.00	0.69
Total	18.27	**6.33**	7.97	8.34	18.57	**5.90**	6.92	6.74	3.29	2.67	1.71	0.00	
Average	3.04	1.05	1.33	1.39	2.65	0.98	1.15	1.12	0.55	0.44	0.29	0.00	
Std. Deviation	3.65	1.01	1.08	1.40	3.56	1.51	1.67	1.71	0.84	0.64	0.41	0.00	

* Reference values. [a] Coefficient of variation.

Boldface identifies the value(s) within each image resolution for each feature class (across rows), that has the lowest footprint dissimilarity error, except in cases where greater than two ties exist for any within-class value. Underline identifies the value(s) within each image resolution for each feature class (across rows), that has the largest footprint dissimilarity error.

two features best ((1) Building, Road_maintained and (2) Road_track, Water_lentic, respectively; table 7).

Pooling footprint dissimilarity errors across all eight disturbance-feature classes revealed that EC had higher error rates than the other MICs on all three image resolutions (fig. 13). Although we did not assess the significance of these differences, FA had the lowest total footprint dissimilarity error rates for the LS and SP imagery, and FX had the lowest footprint dissimilarity error rate for the QB imagery.

Four clusters were identified in the analysis of footprint dissimilarity error rates generated for each MIC while mapping each feature class (fig. 14). Cluster one was the QBHD reference group. Cluster two contained the remaining three QB MICs and all four SP MICs, with SPEC least-similar among them. SPFX and SPHD were slightly less similar, and SPEC was by itself in a separate, still less-similar sub-cluster. Cluster three contained LSFX and LSHD, which joined with clusters one and two lower on the tree than did cluster four, which contained LSEC and LSFA.

Feature Classification using Ground Control Points

The GCPs were not randomly distributed throughout the study area (global Moran's I = -0.1752; $P = 0.0004$). This was at least partly because most of the infrastructural disturbance was concentrated in the southern and western parts of the study area (fig. 7). However, the Moran's I index value was small, suggesting that the departure from a random point distribution was low. For this reason, we analyzed our GCP data without further subsetting.

Based on the overall Kappa assessment of feature classification accuracy on the QB image, FA, FX, and HD all achieved strong levels of agreement with the GCPs, while EC achieved only a fair—good index score (fig. 15). On the SP imagery, FA, FX, and HD were again all similar and achieved a fair–good index score, but EC achieved only a poor index score. None of the products mapped from the LS imagery differed, and all had poor index scores.

On the LS imagery, commission error rates were low enough to result in strong levels of agreement between MICs and the GCP references in only five cases; all four MICs accurately classified Road_maintained features, and HD accurately classified Bare_pad features (table 8). Conversely, EC had poor index scores for four features (Building, Road_track, Water_lentic, and Bare_pad), FX and HD both had poor scores for three (Building, Disturbed, and Road_track for both), and FA had poor index scores for two (Building and Road_track). Omission error rates on the LS imagery followed a similar trend. Each of the four MICs had strong index values for the Water_lentic feature. For the remainder, FX received three poor scores (Disturbed, Road_maintained, and Bare_pad), EC had two (Road_maintained and Bare_pad), and FA (Building) and HD (Bare_pad) had one each.

Commission error rates associated with classification of features on the SP image varied among MICs (table 8). EC achieved no strong index scores, while FA, FX, and HD had

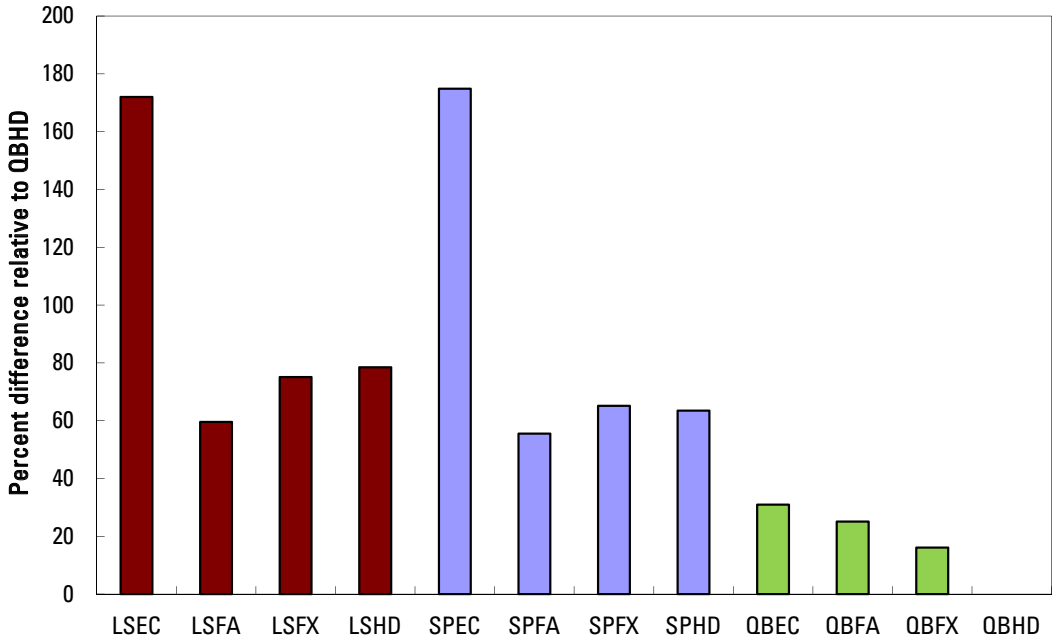

Figure 13. Footprint dissimilarity error rates after pooling disturbance-feature classes within each of 11 information extraction method–image resolution combinations (MICs) relative to the QBHD reference data. (For each four-letter code (X axis labels): LS, Landsat; SP, SPOT; QB, QuickBird; EC, eCognition; FA, Feature Analyst; FX, Feature Extraction; and HD, hand digitization.)

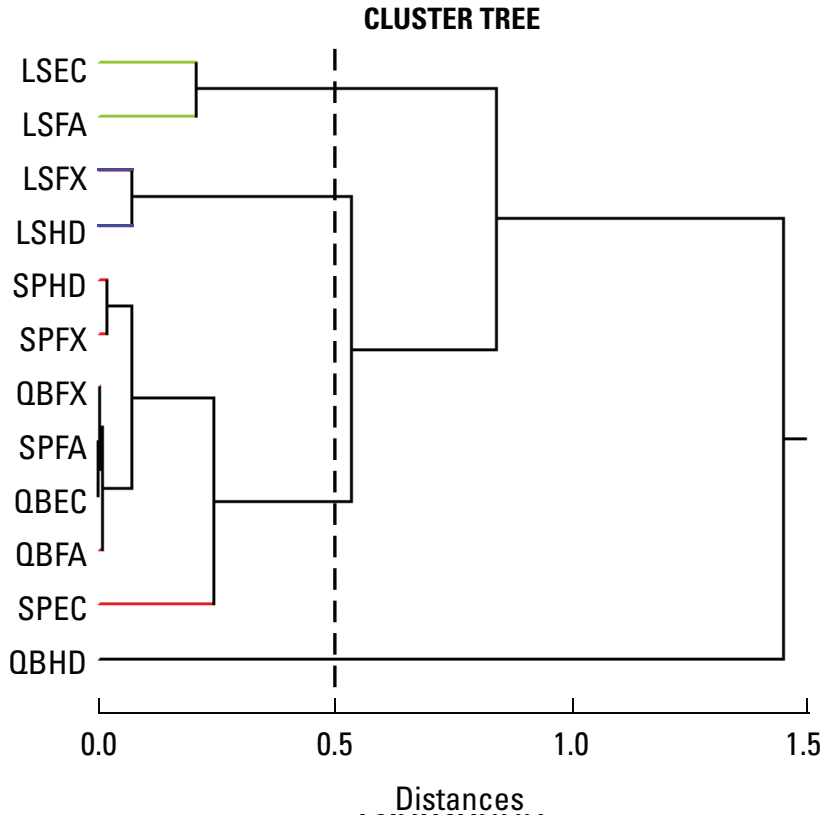

Figure 14. Tree diagram depicting method-image combination groupings based on patterns of footprint dissimilarity error in each disturbance-feature class. Root mean squared standard deviation, pseudo F, and pseudo T indices (not shown) indicated four distinct clusters, identifiable by assuming a vertical line (shown) at distance = 0.5. Branch tips (left side of tree) representing each cluster are colored differently. (For each four-letter code (X axis labels): LS, Landsat; SP, SPOT; QB, QuickBird; EC, eCognition; FA, Feature Analyst; FX, Feature Extraction; and HD, hand digitization.)

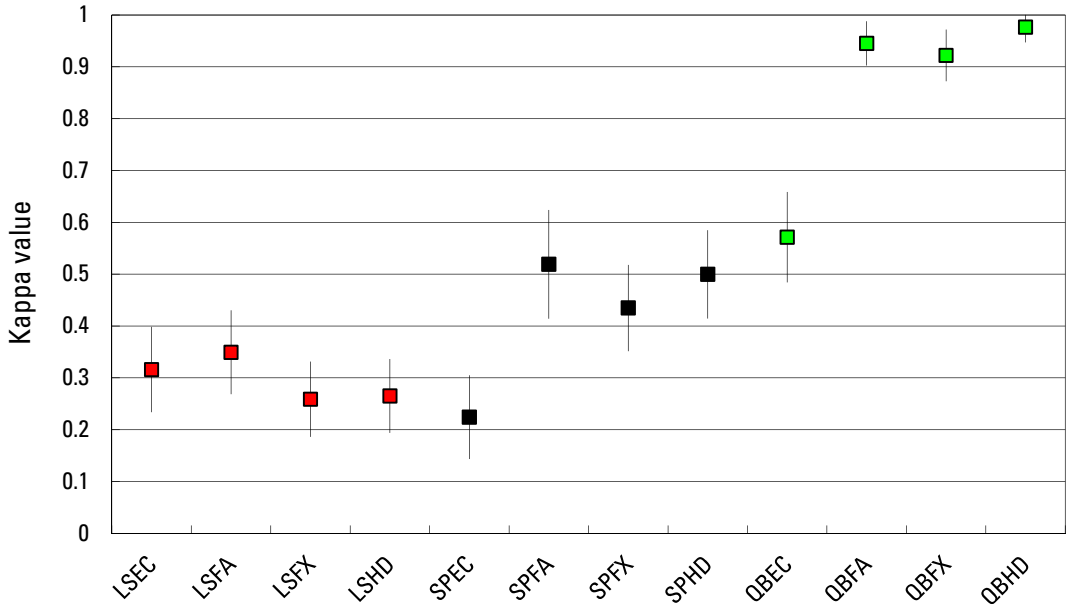

Figure 15. Kappa index scores (plus or minus 95 percent CIs) indicating strength of feature classification agreement between each of 12 information extraction method–image resolution combinations (MICs) and ground control point (GCP) reference data. We interpreted kappa scores greater than 0.75 as indicative of strong agreement, 0.40–0.75 indicative of fair–good agreement, and scores less than 0.40 indicative of poor agreement. (For each four-letter code (X axis labels): LS, Landsat; SP, SPOT; QB, QuickBird; EC, eCognition; FA, Feature Analyst; FX, Feature Extraction; and HD, hand digitization.)

Table 8. Classification accuracy index scores among 12 information extraction method–image resolution combinations and the ground control point (GCP) feature class data. (--, identifies instances where a class was not assigned either to the GCP or to a map; NaN, the conditional Kappa could not be calculated)

	LSEC	LSFA	LSFX	LSHD	SPEC	SPFA	SPFX	SPHD	QBEC	QBFA	QBFX	QBHD
Class	Conditional Kappa (based on Commission error rates)											
Building	0.154	0.040	0.000	0.000	0.021	0.596	0.000	0.029	0.000	**0.914**	**0.957**	**0.957**
Disturbed	0.457	0.474	0.024	0.000	0.095	0.447	0.125	0.258	**1.000**	**0.867**	**0.954**	**1.000**
Road_paved	--	--	--	--	NaN	--	--	--	--	--	--	--
Road_maintained	**0.884**	**0.846**	**0.774**	**0.791**	0.652	**0.948**	**0.946**	**0.947**	**1.000**	**1.000**	**0.952**	**1.000**
Road_track	0.000	0.000	0.000	0.000	0.000	0.227	0.282	0.381	0.475	**1.000**	**0.919**	**1.000**
Water_lentic	0.384	0.500	0.617	0.557	0.560	**0.870**	**0.870**	**0.870**	**0.870**	**0.934**	**0.934**	**0.934**
Bare_pad	0.303	0.533	0.616	**0.752**	0.275	**0.908**	**0.941**	**1.000**	0.430	**0.954**	**0.823**	**0.955**
	Conditional Kappa (based on Omission error rates)											
Building	0.444	0.267	NaN	NaN	0.083	**1.000**	NaN	**1.000**	NaN	**1.000**	**1.000**	**1.000**
Disturbed	0.437	0.545	0.394	NaN	0.091	0.636	**1.000**	**1.000**	0.499	**0.952**	**0.838**	**0.957**
Road_paved	--	--	--	--	0.000	--	--	--	--	--	--	--
Road_maintained	0.378	0.579	0.363	0.503	0.523	0.602	0.505	0.560	**0.912**	**0.954**	**0.952**	**0.954**
Road_track	NaN	NaN	NaN	NaN	NaN	**1.000**	**0.887**	**1.000**	**0.927**	**1.000**	**1.000**	**1.000**
Water_lentic	**1.000**	**1.000**	**0.906**	**0.898**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.934**
Bare_pad	0.367	0.432	0.272	0.237	0.192	**0.798**	0.331	0.352	0.322	**0.838**	**0.903**	**1.000**

Boldface identifies Kappa values indicative of strong agreement between MICs and ground control point reference values. Underline identifies Kappa values having poor agreement between MICs and ground control point reference values.

three cases each (Road_maintained, Water_lentic, and Bare_pad for each). Conversely, EC had four poor scores (Building, Disturbed, Road_track, and Bare_pad), FX and HD had three each (Building, Disturbed, and Road_track for both), and FA had one (Road_track). Omission error rates resulted in EC receiving a strong index score for only one feature (Water_lentic) and poor scores for four features (Building, Disturbed, Road_paved, and Bare_pad). FA and HD had strong scores for four features each (Building, Road_track, and Water_lentic for both, Bare_pad for FA, Disturbed for HD), and FX had strong scores for three features (Disturbed, Road_track, and Water_lentic. Both FX and HD had one poor score each (Bare_pad) while FA had none.

The conditional Kappa assessment of omission error rates resulted in strong index scores for most of the QB-derived products: only three scores were poor or fair–good, and these all occurred in the EC extraction (table 8). Similarly, low commission error rates resulted in FA, FX, and HD each having strong index scores for six feature classes, while EC had strong scores for three, one fair–good score, and one poor score (Bare_pad).

Feature Classification using Pixel-Based Comparisons

Based on the overall Kappa assessment of pixel by pixel classification accuracy on the QB image, EC, FA, and FX all achieved strong scores, indicating high levels of agreement with the HD reference data (fig. 16). On both the SP and LS imagery, all four MICs classified features with an error rate that resulted in a fair–good index score. EC consistently had the lowest scores on the LS and SP imagery, while FA consistently had the highest scores.

Kappa scores based on commission error rates on the LS imagery resulted in two strong scores: both EC and FA identified Bare_pads well. HD had three poor scores (Building, Disturbed, and Water_lentic, EC had two (Disturbed and Water_lentic, FA had one (Disturbed), and FX had no poor scores (table 9). The omission error rate assessment of the LS image resulted in two strong index scores: FX (Buildings) and HD (Water_lentic. EC and HD each had two poor scores (Road_track for EC, Building for HD, and Bare_pad for both), while FA and FX had none.

The Kappa assessment of commission error rates on the SP imagery resulted in all four MICs having strong scores for classifying Bare_pad, and FX and HD also had strong scores for Buildings (table 9). EC had two poor scores (Disturbed and Water_lentic, while FX had one (Road_maintained), and FA and HD had none. Based on omission error rates, FA and HD each had four strong scores (Building, Road_track, and Water_lentic for both, Bare_pad for FA, Disturbed for HD), FX had three (Disturbed, Road_track, Water_lentic, and EC had one (Water_lentic).

Based on the Kappa evaluation of commission error rates on the QB imagery and using QBHD as the reference, FA

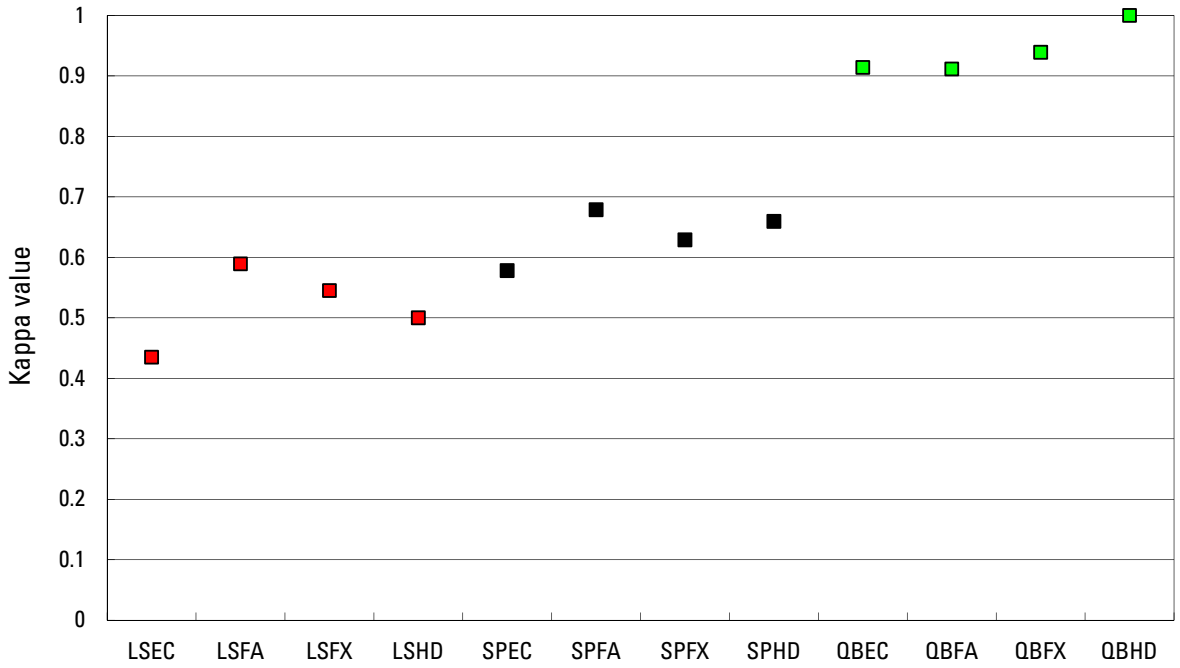

Figure 16. Kappa index scores (plus or minus 95 percent CIs, which are too narrow to see in this figure) indicating strength of feature classification agreement between each of 11 information extraction method—image resolution combinations (MICs) and QBHD reference data in a pixel by pixel comparison. We interpreted kappa scores greater than 0.75 as indicative of strong agreement, 0.40–0.75 indicative of fair–good agreement, and scores less than 0.40 indicative of poor agreement. (For each four-letter code (X axis labels): LS, Landsat; SP, SPOT; QB, QuickBird; EC, eCognition; FA, Feature Analyst; FX, Feature Extraction; and HD, hand digitization.)

Table 9. Pixel by pixel classification accuracy index scores among 11 information extraction–image resolution combinations and the QBHD reference data. (--, identifies instances where a class was not assigned either to the GCP or to a map; NaN, the conditional Kappa could not be calculated)

	LSEC	LSFA	LSFX	LSHD	SPEC	SPFA	SPFX	SPHD	QBEC	QBFA	QBFX	QBHD
Class	Conditional Kappa (based on Commission error rates)											
Building	0.404	0.646	0.610	0.000	0.508	0.715	**0.814**	**0.851**	**0.872**	**0.903**	**0.916**	1.000
Disturbed	0.236	0.355	0.438	0.392	0.330	0.510	0.481	0.579	**0.898**	**0.831**	**0.893**	1.000
Road_paved	--	--	--	--	--	--	--	--	--	--	--	--
Road_maintained	NaN	NaN	NaN	NaN	NaN	0.534	0.387	0.584	**0.777**	**0.883**	**0.835**	1.000
Road_track	0.565	0.620	0.604	0.568	0.648	0.583	0.687	0.641	**0.938**	**0.958**	**0.927**	1.000
Water_lentic	0.329	0.537	0.401	0.373	0.378	0.457	0.463	0.457	**0.897**	**0.823**	**0.923**	1.000
Bare pad	**0.793**	**0.756**	0.739	0.726	**0.893**	**0.845**	**0.812**	**0.824**	**0.957**	**0.951**	**0.971**	1.000
	Conditional Kappa (based on Omission error rates)											
Building	0.546	0.727	**0.791**	0.000	**0.852**	**0.850**	0.745	**0.820**	**0.943**	**0.942**	**0.952**	1.000
Disturbed	0.631	0.605	0.631	0.561	0.585	0.621	0.604	0.588	**0.780**	0.735	**0.865**	1.000
Road_paved	--	--	--	--	--	--	--	--	--	--	--	--
Road_maintained	NaN	NaN	NaN	NaN	NaN	0.469	0.524	0.418	**0.853**	**0.862**	**0.873**	1.000
Road_track	0.282	0.459	0.655	0.710	0.620	0.691	**0.842**	**0.821**	0.539	**0.829**	**0.785**	1.000
Water_lentic	0.561	0.711	0.723	**0.768**	0.729	**0.803**	**0.833**	**0.868**	**0.867**	**0.906**	**0.903**	1.000
Bare pad	0.356	0.526	0.452	0.399	0.468	0.606	0.549	0.587	**0.929**	**0.924**	**0.951**	1.000

Boldface identifies Kappa values indicative of strong agreement between MICs and QBHD reference values. Underline identifies Kappa values having poor agreement between MICs and reference values.

and FX received strong index scores for all six of the features assessed. EC had five strong scores and one poor score (Building; table 9). The Kappa assessment of omission error rates resulted in FX having strong scores for each of the feature classes. FA had strong scores for five classes and EC had strong scores for four classes.

Cost-Benefit Assessment

The MICs formed different patterns within each image resolution when we plotted footprint dissimilarity against production time (fig. 17). On the LS image, the EC extraction resulted in very high footprint dissimilarity error rates and the FA extraction took a very long amount of time to complete. FX and HD were similar in terms of time efficiency and mapping accuracy. On the SP image, EC again resulted in an unacceptably high level of footprint dissimilarity error, while FA, FX, and HD all performed similarly in terms of time and error-rate efficiency. The QB MICs contained less footprint dissimilarity error than extractions performed on the other imagery, but all four MICs took from two to seven times longer to produce than the other MICs except LSFA (table 10). Among the QB MICs, FX and HD took the least amount of time to extract and they had lower error rates than either EC or FA.

When we plotted overall GCP kappa index scores against extraction time on the LS imagery, all four MICs had poor scores, but FA took more than eight times longer to extract than the other MICs and EC took approximately twice as long as FX and HD (fig. 18). On the SP image, FA, FX, and HD took only half the time to extract as did EC, and FA, FX, and HD each had fair–good kappa scores, while EC had a poor score. On the QB image, FA, FX, and HD all achieved strong kappa scores, but FA took approximately 40 more hours to complete than did FX and HD. EC took the longest time to complete and received only a fair–good kappa score.

MIC Rankings

On the LS imagery, FX ranked highest overall, while FA and HD tied for second highest, and EC ranked lowest. FX ranked highest at mapping the spatial area of disturbance and

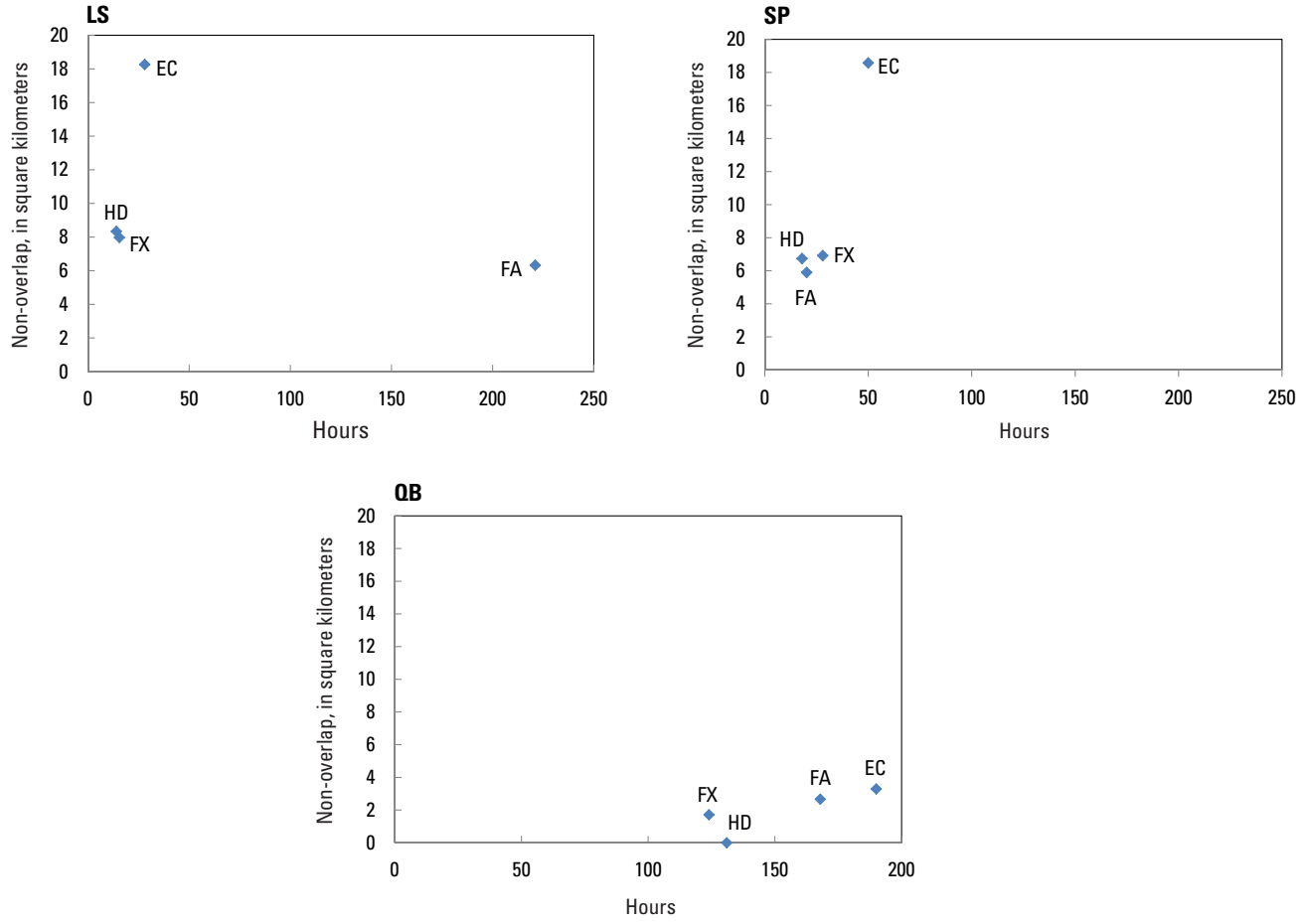

Figure 17. Extraction time in hours versus footprint dissimilarity error in square kilometers for each of 11 information extraction method–image resolution combinations when compared against the QBHD reference data (shown on the QB figure). (LS, Landsat; SP, SPOT; QB, Quickbird; EC, eCognition; FA, Feature Analyst; FX, Feature Extraction; HD, hand digitization.)

Table 10. Time (hours) required to complete automated extraction and post-extraction processing for each of three software products and hand digitization.

	Landsat	SPOT	QuickBird
eCognition			
Extraction	15	18	27
Editing	13	32	163
Total	28	50	190
Feature Analyst			
Extraction	16.6	6.3	15.1
Editing	204.5	14	152.8
Total	221.1	20.3	167.9
Feature Extraction			
Extraction	0.5	0.5	3.5
Editing	15	27.5	120.5
Total	15.5	28	124
Hand Digitization			
Extraction	0	0	0
Editing	14	18	131
Total	14	18	131

pixel-based feature classification. HD ranked highest at GCP-based feature classification and extraction time, FA ranked highest for mapping features with minimal footprint dissimilarity error, and EC did not rank highest at performing any of the operations we evaluated (table 11).

On the SP imagery, HD ranked highest overall, followed by FA, FX, and EC in decreasing rank order. HD ranked highest at mapping spatial area of disturbance, footprint dissimilarity, pixel-based feature classification, and extraction time, while FA ranked highest at GCP based feature classification. EC and FX did not rank highest for any metric.

On the QB imagery, we could not rank HD because it was used as a reference for evaluating all other MICs. FX ranked highest overall, followed by FA, then EC. FX ranked highest at mapping spatial area of disturbance, footprint dissimilarity, pixel-based feature classification, and extraction time, and tied with FA for first at GCP based feature classification. EC did not rank highest at any metric.

Discussion

Landscapes impacted by human disturbance may have an array of infrastructural disturbance types present, and mapping them accurately is an important first step in evaluating land use/land cover changes and their effects on environmental phenomena. Image resolution, contrast, color, grain, pattern, shape, and the spatial distribution of features present in remotely sensed imagery all influence how accurately this information can be identified and measured (Wade and others, 1999; Hawbaker and Radeloff, 2004; Mladinich, 2010). We found that within image resolutions—our primary comparative foci—the MIC combinations that we examined did not perform similarly at identifying the amount of disturbance, the type of disturbance, or in extraction time in our southwest Wyoming study area.

Based on our MIC rankings, the Feature Extraction software performed best over the widest range of image resolutions and extraction tasks. Feature Extraction ranked first at two tasks performed on the Landsat imagery and all five tasks performed on the QuickBird imagery. Feature analysts also performed well over a wide range of resolutions and tasks, and ranked first in processing tasks on the SPOT image. Hand digitization was a highly efficient data extraction method, although we could not rank its performance in extracting the QuickBird imagery because we used that output as benchmark data. eCognition performed poorly on all three image resolutions.

It is not clear why eCognition performed poorly relative to the other information-extraction methods. Mladinich (2010) compared eCognition, Feature Analyst, and ENVI Feature Extraction while identifying features in a disturbed landscape containing irregular and poorly defined features caused by extensive off-road vehicle use. She found that while overall Kappa agreement rates did not differ significantly among the three software packages, eCognition performed slightly better than the other two programs at classifying features into disturbance categories. Emphasizing analytical objectivity, Mladinich (2010) processed her study area imagery only to the point at which manual editing became necessary, then

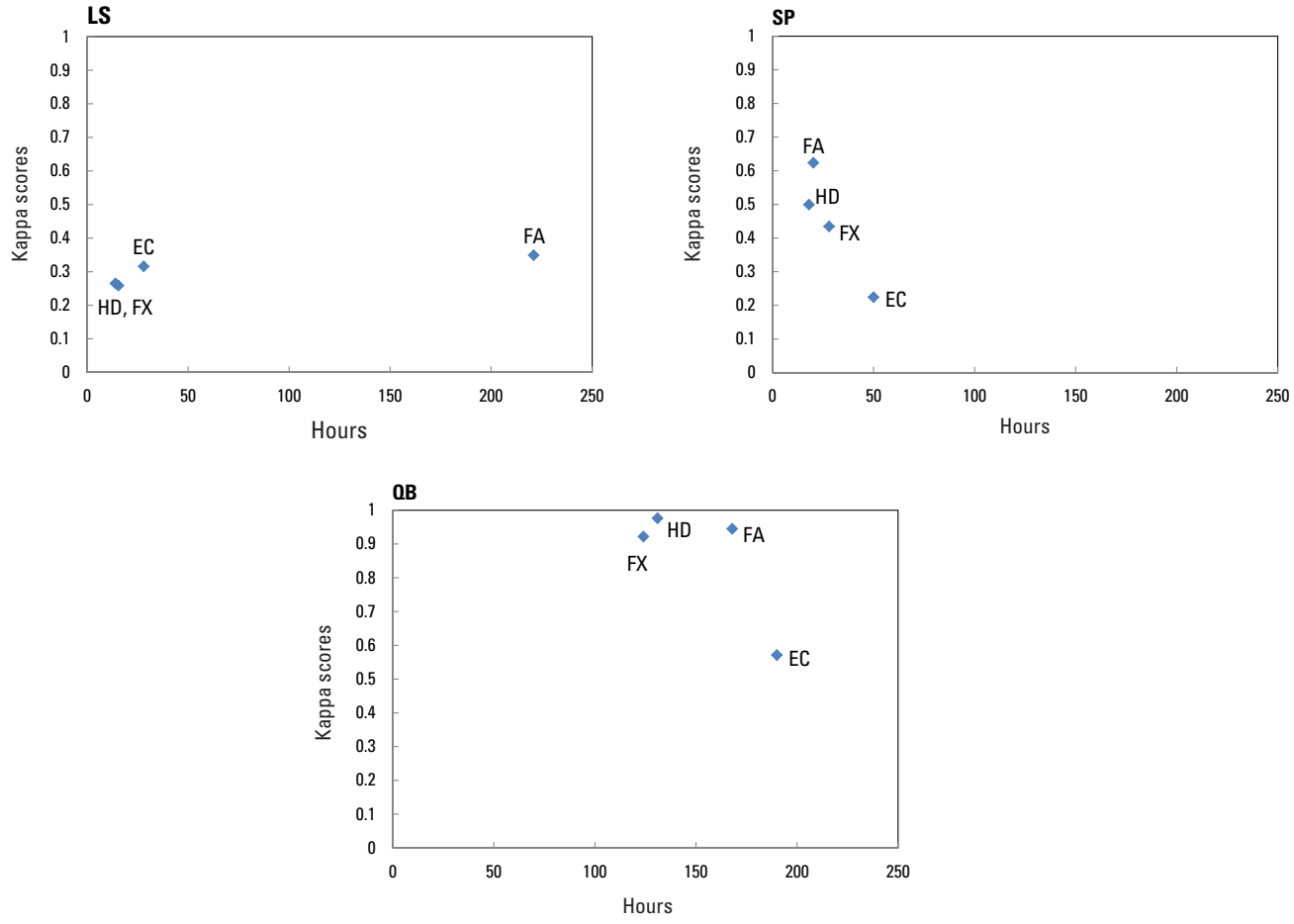

Figure 18. Extraction time in hours versus ground control point (GCP) kappa index scores for each of 12 information extraction method–image resolution combinations. (LS, Landsat; SP, SPOT; QB, Quickbird; EC, eCognition; FA, Feature Analyst; FX, Feature Extraction; HD, hand digitization.)

Table 11. Within-image resolution rankings among information extraction method–image resolution combinations based on performance while performing four common image mapping tasks and the time required to generate maps from remotely sensed images of a 64 km² section of the Jonah gas energy field in Sublette County, Wyo. (km², square kilometers)

	Landsat				SPOT				QuickBird		
Metric[a]	EC	FA	FX	HD	EC	FA	FX	HD	EC	FA	FX
Spatial Area	2.0	3.0	1.0	4.0	4.0	2.0	3.0	1.0	2.5	2.5	1.0
Dissimilarity	3.5	1.0	3.5	2.0	4.0	3.0	2.0	1.0	3.0	2.0	1.0
GCP class	3.5	2.0	3.5	1.0	4.0	1.0	3.0	2.0	3.0	1.5	1.5
Pixel class	3.0	2.0	1.0	4.0	4.0	2.0	3.0	1.0	3.0	2.0	1.0
Time	3.0	4.0	2.0	1.0	4.0	2.0	3.0	1.0	4.0	3.0	1.0
Score	15.0	12.0	11.0	12.0	20.0	10.0	14.0	6.0	15.5	11.0	5.5
Final rank	**4.0**	**2.5**	**1.0**	**2.5**	**4.0**	**2.0**	**3.0**	**1.0**	**3.0**	**2.0**	**1.0**

[a] Algorithms used to determine ranks for each information extraction method–image resolution combination (MIC) were as follows. For spatial area and spatial non-overlap, # instances of most accurate within-resolution measurement / (# instances of least accurate within-resolution measurement + 1). For classification accuracy, # instances of strong agreement / (# instances of poor agreement + 1). Time was the number of hours required to complete each extraction process. For all metrics, actual values were transformed to ranked scores, which are presented here. Final rank scores are sums for each MIC (columns), with lowest scores ranking highest overall within each image resolution.

analytically compared the "raw" extraction results. She also limited her classification scheme to two categories: disturbed and not disturbed.

We chose to incorporate post-processed interpretations of the automated output and a more complex classification scheme to more closely represent real-world mapping efforts. We attributed disturbance features into eight classes instead of two, resulting in more complex computations required by each software package. We did so because in many applications (for example, landscape ecology) investigators are interested in modeling the relative effects of specific disturbance features against one another. Also, some of our features were well defined on the SPOT and QuickBird images, while others were poorly defined. While we did not formally assess the influence of well-defined (for example, water_lentic, road_maintained) versus poorly-defined (for example, disturbed, road_track) features, tables 6–9 suggest that feature definition did not affect eCognition's extraction performance. Our eCogntion analyst used the simplest rule sets available for all of the classifications, because in his experience, customization of rule sets in eCognition can be very time consuming. It may be that the feature complexity present in our study area imagery required more complex rule sets, and that eCognition generalized its process at a lower level of discrimination than we designed into our study. Alternatively, post-processing misinterpretation of bare or sparsely vegetated areas as disturbed or roaded areas could have contributed to both area and classification errors. However, evaluation of these possibilities was beyond the scope of this study.

Leydsman and others (2008) performed a qualitative comparison of Feature Analyst and ENVI Feature Extraction in an investigation of the feasibility of using automated extraction software to map buildings in Sevier County, Utah. They found that both products failed to extract some (amount unquantified) of the built structures present in 1-m resolution NAIP imagery, even after training the algorithms with data from the image, and concluded that Feature Analyst provided "improved results" compared to ENVI Feature Extraction. Vanderzanden and Morrison (2002) compared the ability of Feature Analyst to classify forest stand types against direct photo interpretation and a texture band interpretation (Ryherd and Woodcock, 1996) of a QuickBird panchromatic 0.6-m image. They found FA to be more accurate than direct photo interpretation in each of three comparisons made (size/structure, crown closure, and cover type) and found FA to be more accurate than the texture band interpretation in two of the three comparisons (Vanderzanden and Morrison 2002).

The difference in overall performance that we observed among hand digitization, Feature Extraction, and Feature Analyst was small in many cases. When evaluated against our ground control photograph points (the only assessment that did not use hand-digitized data from the QuickBird scene as reference data), all three MICs classified features with a high degree of success on the QuickBird imagery, which was of similar resolution to the NAIP image evaluated in Leydsman and others' (2008) study. We did not find either Feature Analyst or Feature Extraction to be more accurate than direct image interpretation (hand digitization) as did Vanderzanden and Morrison (2002).

The accurate feature measurements and classification success associated with the QuickBird MICs is due to the high information content and detail present in this imagery. However, the processing times associated with extracting information from the QuickBird MICs rendered them less cost effective than the Landsat or SPOT imagery. The SPOT Feature Analyst and hand-digitized extractions were most cost effective in our GCP comparison, and along with hand digitization and Feature Extraction on the Landsat imagery, were most cost effective in our non-overlap comparison as well. Because of the comparatively low information content in the LS imagery, we expected that information extraction would proceed quickly. This occurred for three of the MICs, but Feature Analyst was a clear outlier in terms of extraction time. We attribute this to an analyst effect; post-analysis discussions revealed that the analyst who performed this extraction operated in a more meticulous and detail-oriented manner than did the other analysts.

Large-area mapping efforts are becoming increasingly common. To map an area the size of one Landsat scene (31,450 km^2) would require 116 QuickBird scenes or nine SPOT scenes (see table 1). Based on image costs for our study, purchasing QuickBird imagery for an area of this size would cost $757,248, while SPOT imagery for the same area would cost $57,150. Landsat imagery is free. Clearly, cost per unit area must also be seriously considered during the planning stages of landscape mapping projects.

Our comparison of cost-benefit ratios among MICs objectively treated extraction time and the quality of extracted information equally. In practice, the effect(s) of incorporating increased spatial and/or classification error (that is, long term costs) into mapping efforts will render maps permanently less accurate. This loss of map accuracy must be carefully weighed against extraction time, imagery acquisition, and other (short term) costs associated with higher-resolution imagery.

Cluster analyses of the area occupied by disturbance features and the footprint dissimilarity data were the only multivariate analyses we performed, and they corroborated our univariate assessments for these mapping tasks. When mapping the area occupied by the footprint of the disturbance features on the SPOT imagery, eCognition and Feature Analyst were most similar to the reference cluster because they mapped the area of disturbed land and bare well pads in a manner that was proportionally similar to the mapping of the reference group. All of the extraction methods performed well at minimizing footprint dissimilarity error on the QuickBird and SPOT imagery, but eCognition extraction of the SPOT image was least similar of all MICs included in the first outgroup cluster, and this was due to the high level of footprint dissimilarity error relative to the other methods.

In conclusion, consideration of the degree of map accuracy required, cost associated with software, operator, and computation time, and tradeoffs in the form of spatial extent

versus resolution should all be considered when evaluating which combination of imagery and information extraction method might best serve any given land-use-mapping project. With any automated information extraction product, time and training to develop operational expertise may need to be considered. Error rates for both area measurements and feature classification were prohibitively high on Landsat imagery, while QuickBird was time and cost prohibitive for mapping large spatial extents. The SPOT imagery produced map products that were far more accurate than Landsat and did so relatively inexpensively. On the SPOT imagery, hand digitization and Feature Analyst outperformed the other two methods. However, when resources permit, attaining the highest possible classification and measurement accuracy through optimizing image resolution will result in dramatically improved map accuracy.

Acknowledgments

Charlene Nielsen and Scott Nielsen helped develop the study design and performed initial data summarization. Alisa Coffin, Mark Drummond, and Scott Nielsen provided thoughtful reviews of the final report. Hannah Moyer assisted with hand digitization mapping.

References Cited

Antrop, M., 2005, Why landscapes of the past are important for the future: Landscape and Urban Planning, v. 70, p. 21–34.

Bowman, C., and Shetty, H.N., 2007, Crosstabulation (one-way, two-way, and multiway), *in* Systat ver. 12, volume 1: San Jose, Calif., Systat Software Inc., p. I-219–296.

Chrisman, N., 2001, Exploring geographic information systems, (2d ed.): Hoboken, N.J., John Wiley and Sons, 320 p.

Cohen, J., 1960, A coefficient of agreement for nominal scales: Educational and Psychological Measurement, v. 20, p. 37–46.

Congalton, R., 1991, A review of assessing the accuracy of classifications of remotely sensed data: Remote Sensing of Environment, v. 37, p. 11.

Davies, K.W., Bates, J.D., and Miller, R.F., 2006, Vegetation characteristics across part of the Wyoming big sagebrush alliance: Rangeland Ecology and Management, v. 59, p. 567–575.

Definiens, 2006, eCognition 5.0 professional: München, Germany, Definiens AG, Trappentreustr. 1, D-80339.

Driscoll, D.A., 2004, Extinction and outbreaks accompany fragmentation of a reptile community: Ecological Applications, v.4, p. 220–240.

Drummond, M.A., and Loveland, T.R., 2010, Land-use pressure and a transition to forest-cover loss in the eastern United States: Bioscience, v. 60, p. 286–298.

ERDAS Imagine, 2009, ERDAS IMAGINE Professional and IMAGINE Easytrace, Version 9.3: Norcross, Ga., Leica Geosystems, Inc. [software]

ESRI, 2006, ArcGIS Release 9.2: Redlands, Calif., Environmental Systems Research Institute, 2006. [software]

ESRI, 2008, ArcGIS Release 9.3; Redlands, Calif., Environmental Systems Research Institute, 1999–2008. [software]

Fahrig, L., 2003, Effects of habitat fragmentation on biodiversity: Annual Review of Ecology and Evolutionary Systematics, v. 34, p. 487–515.

Fleiss, J.L., 1971, Measuring nominal scale agreement among many raters: Psychological Bulletin, v. 76, p. 378–382.

Foody, G.,M., 2002, Status of land cover classification accuracy assessment: Remote Sensing of Environment, v. 80, p. 185–201.

GeoSpatial Experts, 2006, GPS-Photo Link ArcView 9.x extension—Release 4.1 [software]: Thornton, Colo., GeoSpatial Experts LLC, 2001–2006.

Hawbaker, T.J., and Radeloff, V.C., 2004, Roads and landscape pattern in northern Wisconsin based on a comparison of four road data sources: Conservation Biology, v. 18, p. 1233–1244.

ITT Visual Information Solutions, 2008, ENVI Zoom version 4.5: Boulder, Colo., ITT Visual Information Solutions. [software]

Jacquez, G.M., 1995, The map comparison problem—Tests for the overlap of geographic boundaries: Statistics in Medicine, v. 14, p. 2343–2361.

Johnson, D.H., and O'Niel, T.A., eds., 2001, Wildlife-habitat relationships in Oregon and Washington: Corvallis, Oreg., Oregon State University Press, 736 p.

Lambin, E.F., Turner, B.L., Geist, H.J., Agbola, S.B., Angelsen, A., Bruce, J.W., Coomes, O.T., Dirzo, R., Günther, F., Folke, C., George, P.S., Homewood, K., Imbernon, J., Leemans, R., Li, X., Moran, E.F., Mortimore, M., Ramakrishnan, P.S., Richards, J.F., Skånes, H., Steffen, W., Stone, G.D., Svedin, U., Veldkamp, T.A., Vogel, C., and Xu, J., 2001, The causes of land-use and land-cover change—Moving beyond the myths: Global Environmental Change, v. 11, p. 261–269.

Leica Geosystems, 2008, Leica photogrammetry suite version 9.2: St. Gallen, Switzerland, Leica Geosystems Geospatial Imaging.

Leydsman, E.I., Lowry, J.H., McGinty, C.M., and Ramsey, R.D, 2008, Feature extraction of built structures. Investigation into the utility and feasibility of object-based image segmentation tools for identifying built structures using high resolution digital imagery—A case study in Sevier County, Utah. [Unpublished report prepared by the Remote sensing/GIS Laboratory, Utah State University, submitted to the Utah Automated Geographic Reference Center and the U.S. Geological Survey.]

Lillesand, T.M., and R.W. Kiefer, 2000, Remote sensing and image interpretation: New York, Wiley and Sons, 736 p.

Mladinich, C.S., 2010, An evaluation of object-oriented image analysis techniques to identify motorized vehicle effects in semi-arid ecosystems of the American west: GIScience & Remote Sensing, v. 47, p. 53–77.

National Atlas, 2010, Transportation of the United States—Scope of the American transportation system: Washington, D.C., The National Map, accessed February 17, 2010, at *http://www.nationalatlas.gov/transportation.html*.

R Core Development Team, 2008, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing, ISBN 3-900051-07-0, *http://www.R-project.org*.

Riitters, K.H., and Wickham, J.D., 2003, How far to the nearest road?: Frontiers in Ecology and the Environment, v. 1, p. 125–129.

Ripa, M.N., Leone, A., Garnier, M., and Lo Porto, A., 2006, Agricultural land use and best management practices to control nonpoint pollution: Environmental Management, v. 38, p. 253–266.

Ryherd, S., and Woodcock, C., 1996, Combining spectral and texture data in the segmentation of remotely sensed images: Photogrammetric Engineering and Remote Sensing, v. 62, p. 181–194.

SYSTAT, 2007, SYSTAT Software, Inc., v. 12, San Jose, Calif.

U.S. Environmental Protection Agency, 2004, National Water Quality Inventory—Report to Congress for the 2004 reporting cycle: Washington, D.C., U.S. Environmental Protection Agency Report, EPA 841-R-08-001, accessed on February 17, 2010, at *http://www.epa.gov/owow/305b/2004report/report2004pt1.pdf*.

University of Idaho, 2009, The twelve soil orders—Soil taxonomy: Moscow, Idaho, University of Idaho, accessed on April 17, 2010, at *http://soils.cals.uidaho.edu/soilorders/index.htm*.

Vanderzanden, D., and Morrison, M, 2002, High resolution image classification—A Forest Service test of Visual Learning System's Feature Analyst: Salt Lake City, Utah, USDA Forest Service Remote Sensing Applications Center, Report RSAC-3999-RPT1, 18 p.

Visual Learning Systems, 2008, Feature Analyst Version 4.2: Missoula, Mont., Visual Learning Systems, Inc. [software and reference manual]

Vitousek, P.M., Mooney, H.A., Lubchenco, J., and Melillo, J.M., 1994, Human domination of Earth's ecosystems: Science, v. 277, p. 494–499.

Wade, T.G., Wickham, J.D., and Bradford, D.F., 1999, Accuracy of road density estimates derived from USGS DLG data for use in environmental applications: Photogrammetric Engineering and Remote Sensing, v. 65, p. 1419–1425.

Wilkinson, L., Engelman, L. Carter, J., Coward, M., Athreya, S., Dutta, M., and Peri, G., 2007, Cluster analysis: San Jose, Calif., Systat Software, Inc., Systat v. 12.

Publishing support provided by:
Denver Publishing Service Center

For more information concerning this publication, contact:
Center Director, USGS Fort Collins Science Center
2150 Centre Ave., Bldg. C
Fort Collins, CO 80526-8118
(970)226-9398

Or visit the Fort Collins Science Center Web site at:
http://www.fort.usgs.gov/