12-2004

Analysis of Reflected Spectral Signatures and Detection of Geophysical Disturbance Using Hyperspectral Imagery

Forrest McCoy Hoffman
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a thesis written by Forrest McCoy Hoffman entitled "Analysis of Reflected Spectral Signatures and Detection of Geophysical Disturbance Using Hyperspectral Imagery." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Physics.

William E. Blass, Major Professor

We have read this thesis and recommend its acceptance:

Robert N. Compton, Chia C. Shih, Stephen J. Daunt

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a thesis written by Forrest McCoy Hoffman entitled "Analysis of Reflected Spectral Signatures and Detection of Geophysical Disturbance Using Hyperspectral Imagery." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Physics.

William E. Blass

Major Professor

We have read this thesis
and recommend its acceptance:

Robert N. Compton

Chia C. Shih

Stephen J. Daunt

Accepted for the Council:

Anne Mayhew

Vice Chancellor and
Dean of Graduate Studies

(Original signatures are on file with official student records.)

ANALYSIS OF REFLECTED SPECTRAL SIGNATURES AND DETECTION OF GEOPHYSICAL DISTURBANCE USING HYPERSPECTRAL IMAGERY

A Thesis

Presented for the

Master of Science Degree

The University of Tennessee, Knxoville

Forrest McCoy Hoffman

December 2004

Dedication

This thesis is dedicated to Joan, Nathaniel, and Bjørn-Gustaf.

Acknowledgements

This research used data developed with partial support from the U.S. Department of Energy National Petroleum Technology Office (NPTO). Resources for this research were partially provided by the Center for Computational Sciences at Oak Ridge National Laboratory. Oak Ridge National Laboratory is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE–AC05–00OR22725. Site photos and information about the study area were provided by William Hargrove, Rebecca Efroymson, Yetta Jager, and Tom Ashwood (Oak Ridge National Laboratory, Environmental Sciences Division); Peter Earls (Oklahoma State University, Department of Botany); Kerry Sublette (University of Tulsa, Department of Chemical Engineering and Director of the Center for Environmental Research and Technology); Tina Carlsen (Lawrence Livermore National Laboratory, Environmental Restoration Division); Bryan Tapp (University of Tulsa, Department of Geosciences); and Bob Hamilton (The Nature Conservancy, Tallgrass Prairie Preserve). Their assistance is appreciated. I wish to thank my committee members: William E. Blass, Robert N. Compton, Chia C. Shih, and Stephen J. Daunt. In particular, Dr. Blass' sustained support as Major Professor and Dr. Shih's recent encouragement and direction are greatly appreciated.

Abstract

Geophysical disturbances resulting from human activities often have significant consequences for plants and animals, and even for entire ecosystems. Disturbances resulting from petroleum exploration and production activities can have long term impacts on soils, watersheds, rivers and lakes, vegetation, wildlife, and humans. These anthropogenic disturbances are frequently the result of hydrocarbon (oil) or produced water (brine) spills. Brine is usually produced simultaneously with oil or gas. The ability to detect brine spills with remote sensing techniques would be valuable to petroleum companies and industry regulators. The objectives of this research were to 1) determine if brine spills could be detected spectroscopically, 2) determine if spectral analysis could be performed using a statistical method to identify surface features quickly and easily from large imaging spectroscopy data sets without modeling and removing atmospheric effects or performing detailed spectral unmixing, 3) develop a spectral signature for brine spills which could be applied at other locations, and 4) determine if brine spills could be detected using substantially fewer spectral bands so that a smaller and cheaper instrument could be applied to detect these disturbances. Using hyperspectral image cubes acquired by NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over Osage County, Oklahoma, a multivariate statistical clustering technique successfully discerned well-documented brine disturbances on the Tallgrass Prairie Preserve, and the resulting brine spectral signature was applied to locate similar brine disturbances in surrounding image scenes. While validating the prediction results by visiting the site was outside the scope of this project, high resolution aerial photographs were used to assess the success of the predictions and attribute at least 40 of the 87 prediction regions to petroleum activities. While a number of false positives resulted from the analysis, many of these are easily discounted based on objects in the aerial photographs or explained by mineral/salt accumulation. In addition, four bands from the 224-band hyperspectral imagery were used to predict brine disturbances in one of the image cubes. Approximately 90% of the prediction regions detected in the original analysis—which used 187 of the 224 bands—were again detected using only four spectral bands.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Geophysical disturbances resulting from human activities often have significant consequences for plants and animals, and even for entire ecosystems. Disturbances resulting from petroleum exploration and production (E&P) activities, including drilling and pumping, can have long term impacts on soils, watersheds, rivers and lakes, vegetation, wildlife, and humans. These anthropogenic disturbances are frequently the result of hydrocarbon (oil) or produced water (brine) spills. Since hydrocarbons are less toxic to vegetation than brine, areas affected by oil spills recover more quickly than those affected by releases of produced water. Moreover, hydrocarbons may be dispersed or eliminated by volatilization, solubilization, or microbial decomposition. Brine spills, on the other hand, can have a more significant and longer-term impact on the Earth [25]. Most oil and gas E&P operations produce formation water simultaneously with oil or gas. This water is piped via "gather" lines to a central separator where oil is removed. It then travels back to unproductive wells, where it is pumped back down into the Earth to help force up additional oil. This produced water can have salt concentrations exceeding 100,000 mg/kg [8]. Soils affected by brine spills are usually denuded of vegetation and are prone to compaction and erosion [24, 8]. As a result, brine spills can leave a nearly permanent mark on the landscape. Areas which have suffered heavy erosion as a consequence of produced water releases are often referred to as "brine scars." Detecting and remediating such anthropogenic geophysical disturbances can improve ecosystem health and reduce recovery time from decades or centuries to only years.

The ability to detect brine spills and scars with remote sensing techniques would be valuable to both petroleum companies and industry regulators. The objectives of this study were four-fold. First, determine if brine spills could be detected spectroscopically. Second, determine if spectral analysis could be performed using a statistical method to identify surface features quickly and easily from relatively large imaging spectroscopy data sets without the traditional complications of modeling and removing atmospheric effects or performing detailed spectral unmixing. Third, develop a spectral signature for brine spills and scars which could be applied at other locations to detect such geophysical disturbances. Fourth, determine if brine spills or

scars could be detected using substantially fewer spectral bands so that a smaller and cheaper instrument could be applied to detect these disturbances.

The research results presented here suggest a technique for detecting brine disturbances by combining remotely sensed hyperspectral imagery with a multivariate statistical clustering technique. The technique was applied for a region in Oklahoma where oil and gas exploration and production have occurred for more than 80 years. The 224-band hyperspectral imagery was captured at a resolution of $\sim$25 m^2 by AVIRIS, the Airborne Visible/Infrared Imaging Spectrometer operated by NASA's Jet Propulsion Laboratory. Because of the large size of the data set, the relatively low spatial and spectral resolution of the data, and the complexity of unmixing spectra of surface features and removing atmospheric effects from the spectra, traditional spectroscopic analysis techniques cannot be easily applied to detecting surface features over large geographic areas.

Instead, a multivariate statistical clustering technique was applied to separate and group surface features based on their reflected spectral signatures without atmospheric removal. As these results demonstrate, the clustering technique provides a fast, powerful, and quantitative method for spectral analysis of complex hyperspectral imagery. The clustering algorithm was originally developed by the author and collaborators on a parallel computer constructed from surplus personal computers (PCs). Today, the algorithm runs on Linux clusters as well as some of the fastest supercomputers in the world, and it can be applied to very large and complex datasets [22, 19].

Located in the center of the study area is the Tallgrass Prairie Preserve, a 38,000 acre block of land set aside and managed as a natural prairie ecosystem. The mineral rights for the Tallgrass Prairie Preserve belong to the Osage Nation, as is the case for most of the surrounding county. Oil and gas production are the primary economically important activities of Osage County, and production still occurs within the boundaries of the Preserve. Because of its accessibility and the long history of petroleum production within its borders, the Preserve represents an excellent place to validate remote sensing of brine scars. While collecting data on the ground is outside the scope of this project, digital georectified aerial photographs with a resolution of 1 m^2 were used to visually inspect brine spill and scar sites predicted by the cluster analysis.

Chapter 2

Physics Foundation

2.1 Hyperspectral Imaging

In the 1960s and 1970s, sensors aboard artificial satellites orbiting Earth, like the Thematic Mapper (TM) on LandSat, began imaging the planet through a handful of carefully selected color filters. This technique became known as multispectral imaging. It produces a small number of co-registered images from broad non-contiguous spectral bands, seven for the TM. Since this additional spectral information proved useful for understanding more about Earth's environment, such spectroscopic remote sensing was extended to the use of large numbers of contiguous and narrower spectral bands spanning more of the electromagnetic spectrum: from the visible to the near-infrared, and sometimes the ultraviolet and far-infrared regions. Called hyperspectral imaging, this technique produces a large number (sometimes hundreds) of co-registered images which are usually stacked in a three-dimensional image cube with axes of geographic position (x and y) and wavelength (λ) as shown in Figure 2.1.

The hyperspectral imaging technique was first employed in the 1980s by an instrument called the Airborne Imaging Spectrometer at NASA's Jet Propulsion Laboratory (JPL). Today, JPL flies a more advanced instrument called the Airborne Visible/Infrared Imaging Spectrometer or AVIRIS. The main objective of the AVIRIS project is to identify, measure, and monitor constituents of the Earth's surface and atmosphere using molecular absorption and particle scattering signatures.

2.2 Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)

AVIRIS is a unique optical sensor that provides calibrated images of the upwelling spectral radiance in 224 contiguous spectral bands with wavelengths from 400 to 2500 nanometers (nm) and a spectral bandwidth of approximately 10 nm. AVIRIS is usually flown on a NASA ER–2 jet—a modified version of the U2 spy plane—

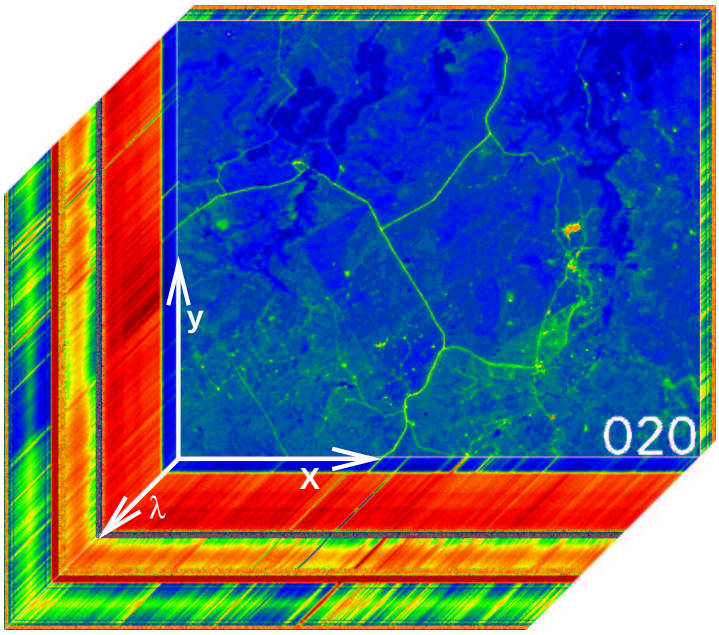


Figure 2.1: A hyperspectral image cube. An image cube has axes of geographic position (x and y) and wavelength (λ). The number 020 superimposed on the image refers to the band number of the image shown on the top of the cube.

approximately 20 km above ground level (AGL) at about 730 km/hr. The instrument, shown in Figure 2.2, has a 30° field of view and uses a scanning mirror to sweep back and forth in "push broom" or "whisk broom" fashion producing 614 pixels for the 224 detectors for each scan. It has a 20.4 Mbps data rate with 12 bit data encoding. Radiation is passed through optical fibers to four separate reflective grating spectrometers, each of which has a double-pass Schmidt optical configuration, and silicon (Si) detectors for the visible range and indium-antimonide (InSb) detectors for the near infrared. Detectors are cooled using liquid nitrogen (LN2). The on-board data recorder writes to 35GB DLT tape IV media. When flown aboard the ER–2, each pixel is approximately 20 m^2 on the ground, yielding a swath about 11 km wide. Ground data are recorded on board the instrument along with navigation and engineering data. Every 512 scan lines, called a "scene," produces about 140 MB of data [2].

2.3 Imaging Spectroscopy

Hyperspectral imaging of the land surface is a passive remote sensing method since the source radiation is not artificially produced. Instead, the Sun is the source of emitted radiation. Solar radiation traverses the Earth's atmosphere, reflects off of the surface, and again traverses the lower atmosphere up to the aircraft altitude.

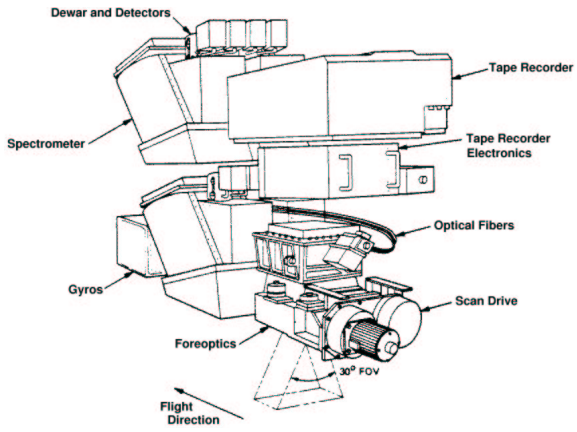


Figure 2.2: The AVIRIS instrument.

Chemical constituents of the atmosphere, differentially distributed at various altitudes (see Table 2.1), absorb this radiation at particular wavelengths while chemical constituents in surface vegetation, water, rocks, and soil absorb radiation at some wavelengths and reflect radiation at other wavelengths. How radiation interacts with matter depend upon the energy of the incident radiation.

Transmitted and reflected radiation from selected portions of the electromagnetic spectrum can be measured to characterize objects of interest. The visible and infrared regions of the spectrum are particularly good for remote sensing of the atmosphere and land surface. In these regions, radiation interacts with valence electrons in molecules (visible) and stimulates molecular vibrations and rotations (infrared) giving rise to the observed absorption features in spectra [13, 32, 14, 6]. As a result, remotely sensed hyperspectral images are snapshots of a complex and dynamic physical system consisting of the Sun, Earth's atmosphere, and Earth's surface as well as the instrument itself.

Each spatial element on the ground has a continuous spectral signature in the resulting image cube, so in addition to viewing data in geographic space, hyperspectral data can also be plotted, one cell or pixel at a time, in spectral space. These spectral reflectance signatures can be used to classify surface features since soil, water, vegetation, and rocks and minerals all exhibit unique reflectance spectra. Figure 2.3 shows three such spectral signatures: one typical of roads, one characteristic of vegetation, and one representing lakes or water bodies. In this figure, the lower x-axis is in units of wavelength and the y-axis is in units of radiance or intensity. The wavelength of emitted radiation depends upon the speed of oscillation of an electric charge, in this case in the photosphere of the Sun. This rate of oscillation, called frequency, is denoted by ν and is related to the wavelength by

$$\nu = \frac{c}{\lambda} \tag{2.1}$$

Table 2.1: The composition of Earth's atmosphere. Adapted and corrected from Goody and Yung [14] and Stephens [32].

Molecule	**Volume Fraction**	**Comments**
N_2	0.7808	Photochemical dissociation high in the ionosphere; mixed at lower levels
O_2	0.2095	Photochemical dissociation above 95 km; mixed at lower levels
H_2O	< 0.04	Highly variable; photodissocates above 80 km
Ar	9.34×10^{-3}	Mixed up to 110 km; diffusive separation above
CO_2	3.45×10^{-4}	Slightly variable; mixed up to 100 km; dissociated above
CH_4	1.6×10^{-6}	Mixed in troposphere; dissociated in mesosphere
N_2O	3.5×10^{-7}	Slightly variable at surface; dissociated in stratosphere and mesosphere
CO	7×10^{-8}	Variable photochemical and combustion product
O_3	$\sim 10^{-8}$	Highly variable; photochemical origin
$CFCl_3$ and CF_2Cl_2	$1\text{–}2 \times 10^{-10}$	Industrial origin; mixed in troposphere, dissociated in stratosphere

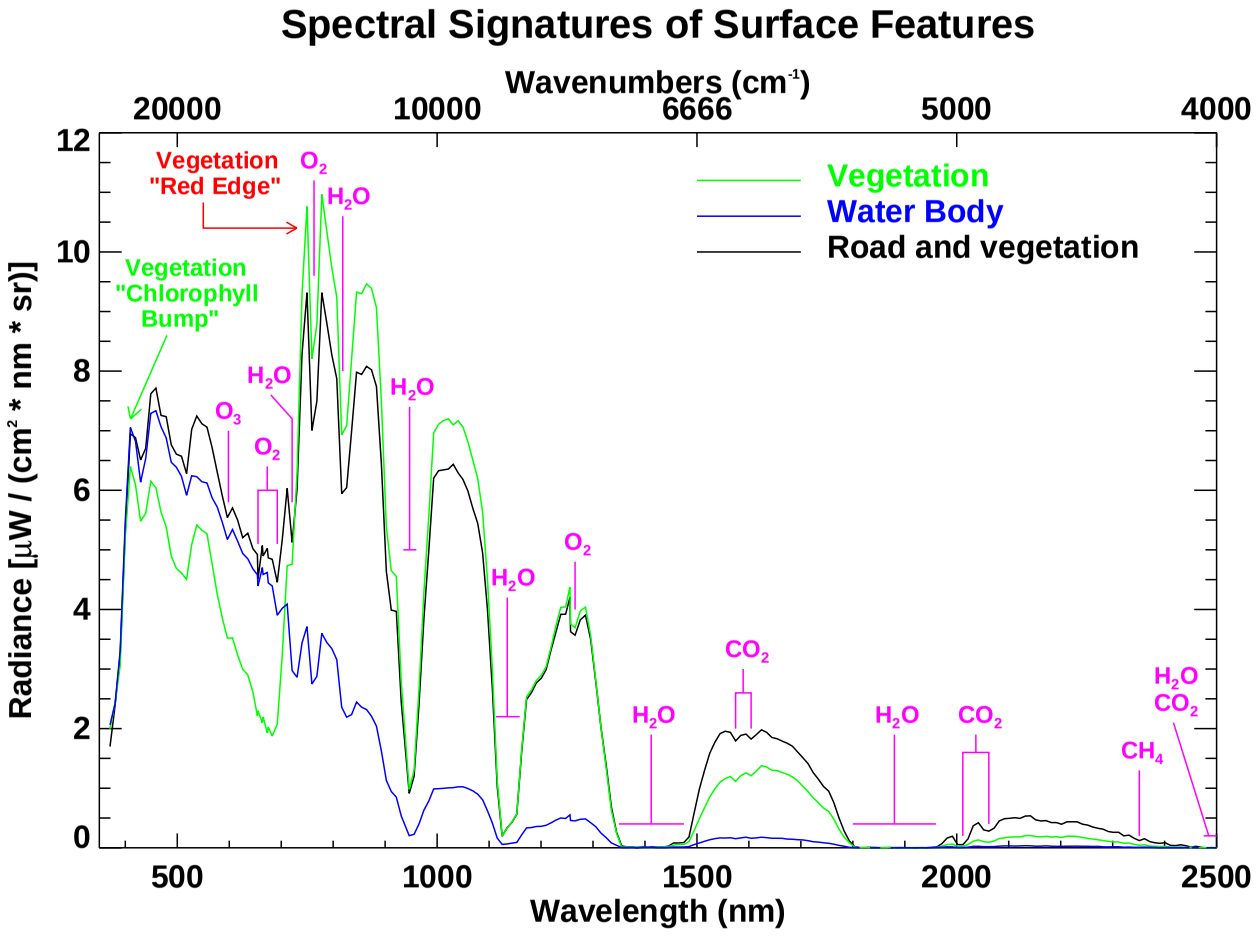


Figure 2.3: Spectral signatures of surface features from hyperspectral imagery.

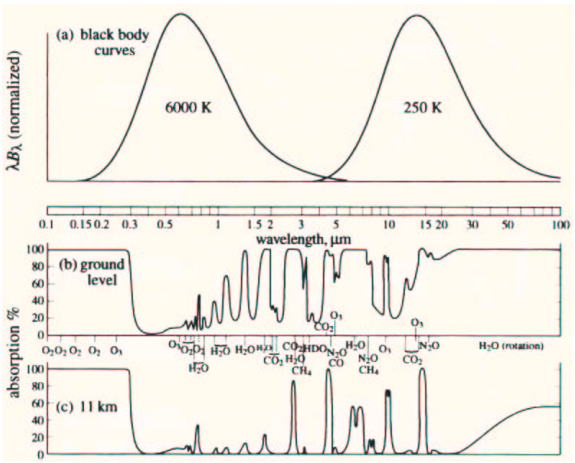


Figure 2.4: Atmospheric absorption spectra. (a) Black body curves for 6000 K and 250 K. (b) Atmospheric absorption spectrum for a solar beam reaching the ground. (c) The same for a beam reaching the temperate tropopause. The areas beneath the curves in (a) are proportional to energy fluxes. Integrated over all angles, and averaged over time and over the globe, solar and terrestrial fluxes must balance; for this reason the two curves in (a) are drawn with equal areas. Conditions are appropriate to middle latitudes, with a solar elevation of about 40°, or for diffuse radiation. From Goody [13].

where c is the speed of light. The units of frequency are oscillations per second or hertz. An alternative measure which does not depend on the speed of light is in terms of wavenumber

$$\tilde{\nu} = \frac{1}{\lambda} \tag{2.2}$$

which is the number of full cycles in a given unit of length, usually 1 cm. The upper x-axis of Figure 2.3 shows the wavenumber in units of cm^{-1} corresponding to the wavelength on the lower x-axis.

2.3.1 Atmospheric Absorption Spectra

The shape of the curves in Figure 2.3 is dominated by the shape of the black body radiation curve of the Sun which peaks in the visible (350–750 nm) and decreases asymptotically with increasing wavelength. Figure 2.4a shows the black body curve for the Sun, which has a temperature near 6000 K, and for the Earth's middle troposphere at approximately 250 K. The temperature distribution of the Earth's atmosphere is shown in Figure 2.5.

The large bands of absorption in all the spectra in Figure 2.3 are due to water vapor (H_2O), carbon dioxide (CO_2), and other chemical constituents in the atmosphere as noted in the figure. Spectra of individual chemical species combine when observed

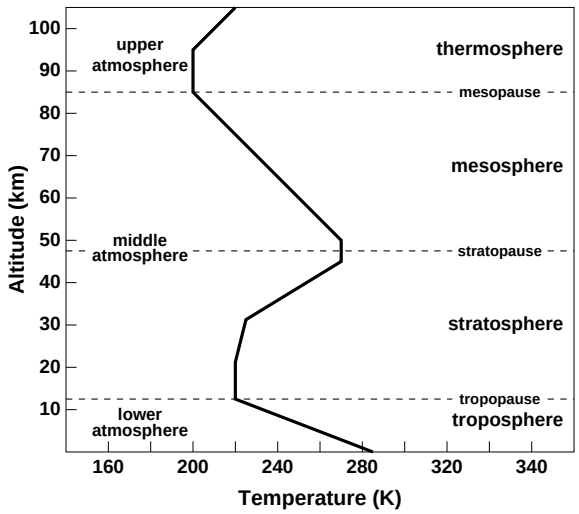


Figure 2.5: Vertical temperature structure of the atmosphere. The curve represents the U.S. Standard Atmosphere [33].

through a column of the atmosphere as shown in Figure 2.6. The combined spectra from important atmospheric species are also shown in Figure 2.4b for an up-looking detector at ground level and in Figure 2.4c for an up-looking detector at 11 km AGL. Absorption by these atmospheric molecules is even stronger in AVIRIS hyperspectral imagery since the detector is down-looking at an altitude of 20 km AGL. In this case, solar radiation traverses the atmosphere above 20 km, then traverses the densest portion of the atmosphere (some 94% of its mass) twice—once before and once after reflecting off the surface—prior reaching the detector.

The pressure and density of the atmosphere decrease exponentially with altitude as follows

$$p(z) \approx p(0) \exp\left(-\frac{z}{H}\right) \tag{2.3}$$

$$\rho(z) \approx \rho(0) \exp\left(-\frac{z}{H}\right) \tag{2.4}$$

where p is the pressure, ρ is the density, z is the altitude, and H is the scale height of the atmosphere. The scale height is taken to be approximately 7 km. This pressure and density profile is shown in a log plot in Figure 2.7. In addition to the natural broadening of atmospheric absorption lines due to the Heisenberg Uncertainty Principle, significant pressure broadening results in wide bands of absorption and line wings, particularly from H_2O and CO_2, over a large part of the infrared spectrum. In fact, the complex vibrational-rotational absorption spectrum of water vapor combined with the significant concentrations of it in the lower atmosphere produce a continuum of absorption throughout much of the infrared region [14, 32].

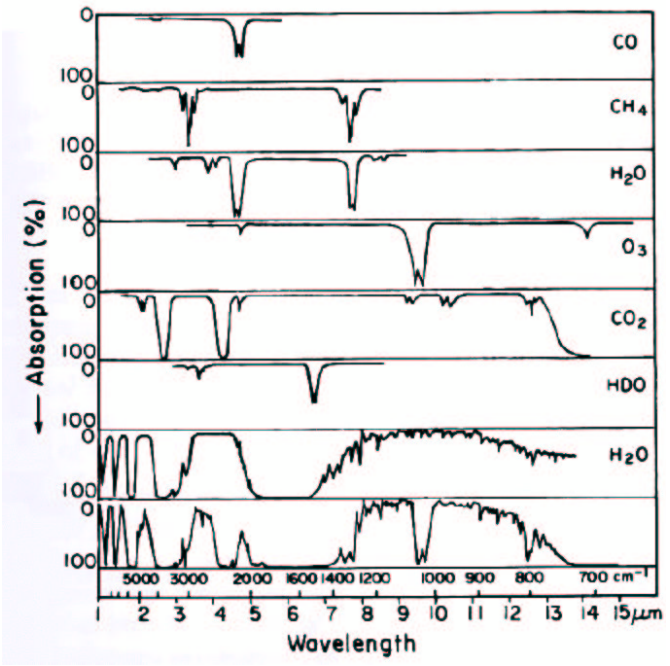


Figure 2.6: Infrared absorption spectra of atmospheric chemical species. The top six panels contain spectra of important chemical species, and the bottom panel is a simulated absorption spectrum of the atmosphere. From Valley [1].

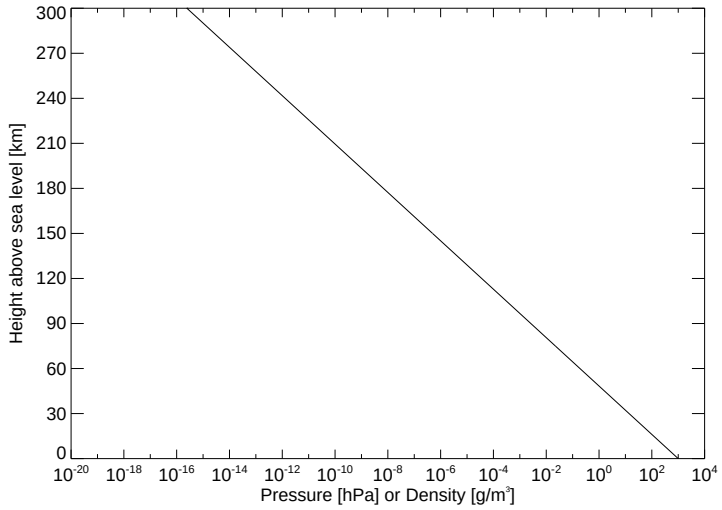


Figure 2.7: Vertical atmospheric pressure and density profile.

2.3.2 Vegetation Reflection Spectrum

Vegetation reflects well at the short end of the visible region, hence its green appearance, but it is very strongly absorbing in most of the remaining visible spectrum because of chlorophyll absorption. The resulting peak in the green is often referred to as the "chlorophyll bump." Plants on Earth are well adapted to the Sun because chlorophyll molecules absorb visible radiation in the process of photosynthesis in which water and carbon dioxide are converted into sugar and oxygen. At the long end of the red and the short end of the near infrared around 700 nm, plant leaves quickly become very strongly reflective. This effect is known as the "red edge" and is a powerful indicator of vegetation for remote sensing. Radiation partially scatters off of leaf surfaces and internal cell walls, and partially refracts through cell walls. Radiation may penetrate leaf surfaces through leaf stomates, and can Mie or Rayleigh scatter off of cell organelles which have sizes on the order of visible light. Reflection off and transmission through leaves is a complex function of cell and stomate shapes and sizes.

Because of the characteristic properties of the reflection spectra of vegetation, an index of biospheric activity is often computed as follows

$$\text{NDVI} = \frac{I_{nir} - I_{vis}}{I_{nir} + I_{vis}} \tag{2.5}$$

where I_{nir} is the intensity at a near-infrared wavelength and I_{vis} is the intensity at a visible wavelength. Referred to as the normalized difference vegetation index (NDVI), this quantity is an indicator of the amount of live vegetation in the field of view of an optical sensor [32]. Since it is normalized, NDVI from various satellite and suborbital platforms can be compared. Moreover, remote sensing can be applied to the determination of plant species and in some cases to the detection of canopy cover, plant health, stress, and senescence [31]. In addition, the "red edge" spectroscopic feature has been studied in Earthshine experiments and is proposed as a surface biomarker in the search for Earth-like extra-solar planets [30].

2.3.3 Spectral Mixing

Included in Figure 2.3 is the reflection spectra for a water body. Surface water tends to be brighter than vegetation in the visible, but darker than any other feature in the infrared since it is a strong absorber of energy. Also included in the figure is a spectral curve for a road surrounded by grassy vegetation. This spectral trace is brighter than vegetation in the visible and is much brighter in the infrared region; however, it also exhibits a somewhat diminished "red edge" feature indicative of vegetation. Both the roadway and the vegetation are contained in the hyperspectral image pixel used to plot this spectral curve. This sort of "spectral mixing" is common in remote sensing, and a number of "spectral unmixing" techniques have been developed in an effort to discriminate between sub-pixel surface features [7]. Such techniques are not

applied here since the objective is to determine if disturbances can be found in the hyperspectral imagery without applying any corrections to the data.

Seven calibrated AVIRIS scenes extracted from two ER–2 flight lines were obtained from the U.S. Geological Survey EROS Data Center for this analysis. The data are in radiance units of microwatts per square centimeter per nanometer per steradian, or μW/(cm^2 $\times$ nm $\times$ sr). All seven scenes are from an over-flight performed in July 1999. The scenes cover the entire Tallgrass Prairie Preserve and some of the surrounding area in Osage County, Oklahoma. This area has both retired and active oil and gas wells and has been used for petroleum production for over 80 years.

Traditional spectral and hyperspectral analyses employ models of atmospheric transmission (e.g., MODTRAN [4]) and databases of molecular absorption (e.g., HITRAN [29]) to simulate the effects of atmospheric composition and structure on remotely sensed spectra. Once modeled, these effects are removed, with varying degrees of success. Software packages like ATREM (ATmospheric REMoval) [11, 12], FLAASH (Fast Line-of-site Atmospheric Analysis of Spectral Hypercubes) [27], and others will automatically perform atmospheric correction on hyperspectral imagery. The resulting spectra may then be compared to catalogs of spectra from surface rocks and minerals like the Splib05a of the U.S. Geological Survey [9]. Radiometric calibration and spectral unmixing are usually required to match cataloged spectra, particularly in highly heterogeneous landscapes.

Figure 2.8 shows reflection spectra for four common minerals from the Splib05a spectral library. These spectra were collected in the laboratory under controlled conditions and are in terms of reflectance. Brine disturbances are expected to have a spectrum very much like the halite (NaCl) spectrum shown in the figure. While kaolinite is almost as reflective from 700–1200 nm, halite is far more reflective than all the other three minerals across the entire spectral range. As a result, brine spills and scars should be strongly reflective in the visible and the near infrared.

Traditional strategies for spectroscopic analysis are of limited utility over large geographic areas because of the large size of the hyperspectral data, the relatively low spatial and spectral resolution of the data, the expensive on-the-ground radiometric calibrations that are required, and the complexity of unmixing surface features and removing atmospheric effects from the spectra. An alternative strategy, like the statistical clustering technique described here, may be more practical for some classes of spectral analysis problems using large, complex hyperspectral imagery over a large geographic area.

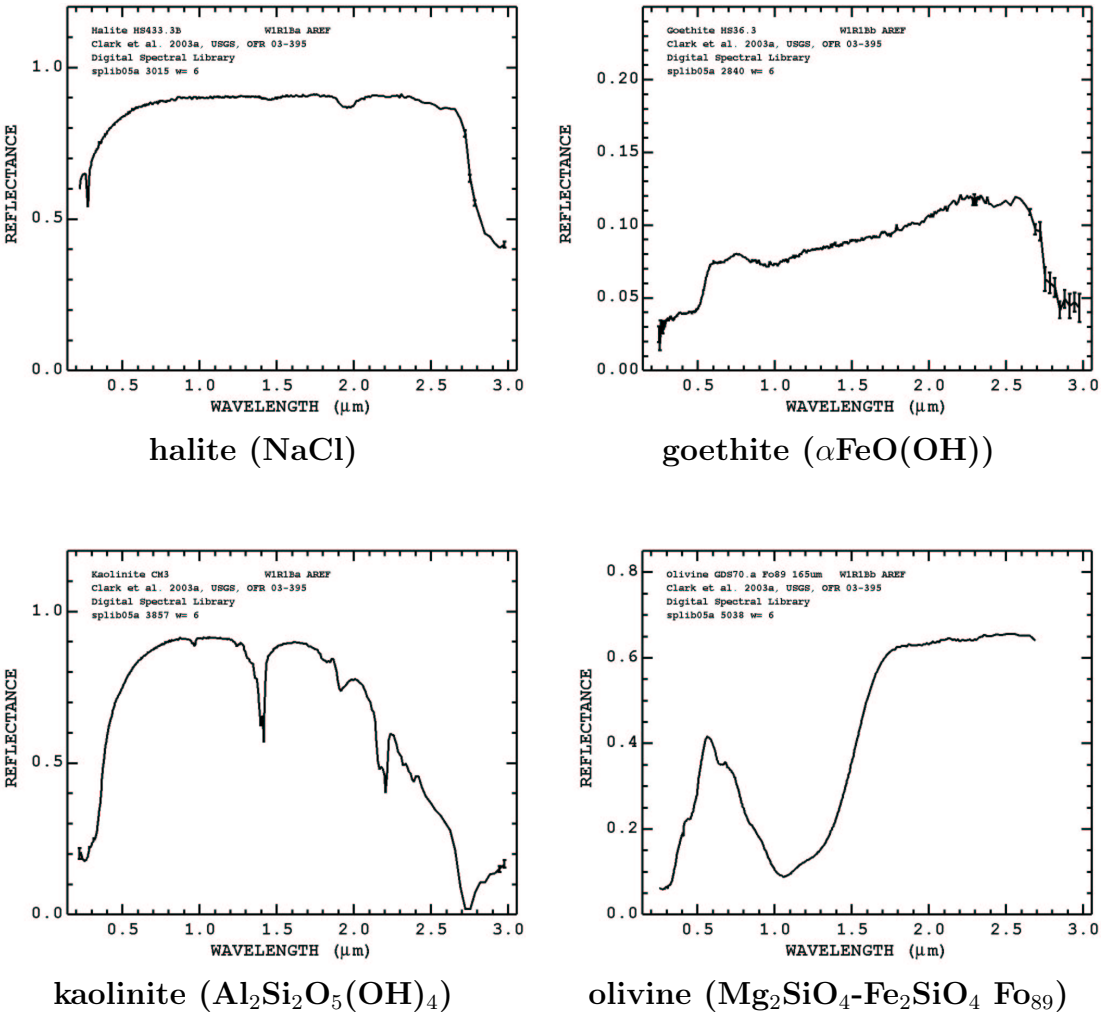


Figure 2.8: Sample mineral reflectance spectra from the Splib05a library [9].

Chapter 3

Multivariate Geographic Clustering (MGC)

Multivariate statistical clustering techniques are frequently used in data mining applications because of their ability to categorize or group a large number of objects based on the similarity of many object properties. The resultant groupings typically define a smaller number of super-objects with properties similar to the objects assigned to each group. Hargrove and Hoffman have pioneered the use of clustering multivariate geophysical properties from maps in a geographic information system (GIS), called Multivariate Geographic Clustering (MGC), as a means of objectively and quantitatively delineating ecoregions from map stacks of 9 and 25 geophysical characteristics or variables for the conterminous U.S. at a 1 km^2 resolution [17, 15]. Ecoregions are land areas having similar characteristics important to plant and animal species. The resulting regionalizations have been used to quantify the representativeness of the AmeriFlux sampling network [18, 16]. More recently, a variant of the technique has been applied to categorize land areas with properties which change through time, notably, modeled global climate conditions evolving over decades to centuries. This methodology, called Multivariate Spatio-Temporal Clustering (MSTC), yields temporally-varying climate regimes which can be used to diagnose model behavior and inherent model climate variability and to understand and interpret model predictions [23].

Originally developed on a custom-built Beowulf-style parallel computer constructed from surplus personal computers (PCs) [19], the clustering algorithm employs a dynamic load balancing scheme with fault tolerance and restart capabilities. The code is written in the C language and uses a master-slave parallel scheme employing MPI (the Message Passing Interface) for interprocess communication on distributed memory parallel supercomputers [22]. The algorithm has been used to successfully analyze large data sets on a wide variety of multi-processor supercomputers from small Linux clusters to some of the largest commercial supercomputers available. The same algorithm has also been used as a metacomputing or grid application to analyze very

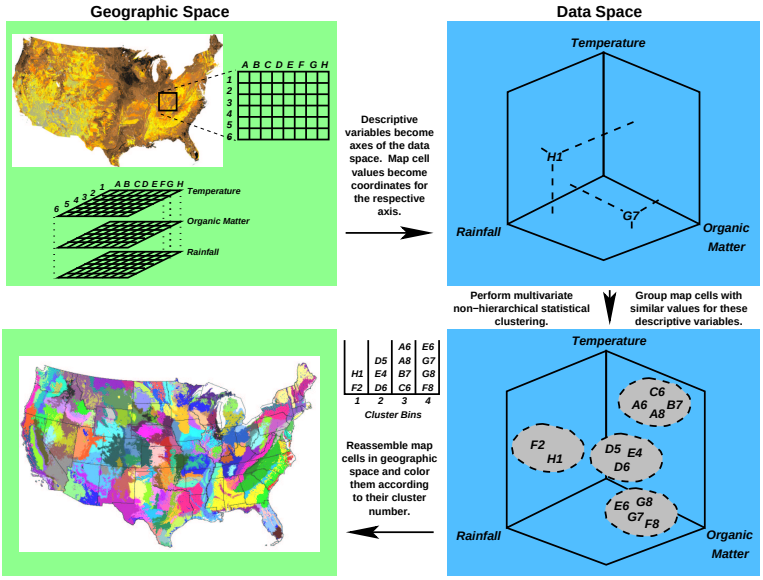


Figure 3.1: The Multivariate Geographic Clustering (MGC) procedure.

large datasets by coupling four supercomputers together across the U.S. via the Internet [26].

The MGC technique, based on the iterative k-means clustering algorithm of Hartigan [21], is outlined in Figure 3.1. The left side of the figure (green) represents geographic space, while the right side (cyan) illustrates the same map cells or observations in a multi-dimensional data space. The N characteristics of each map cell (left top panel) are used as N coordinates for that observation in the N-dimensional data space (right top panel). In Figure 3.1, N is three: temperature, organic matter, and rainfall. Having no information about the geographic coordinates of each observation, the iterative clustering algorithm finds k groups of observations based on their proximity, by simple Euclidean distance, in the N-dimensional data space (right bottom panel). Reassembling the map cells in geographic space and coloring them according to their cluster assignment yields a new map (left bottom panel) showing regions of approximately equal variance with respect to the N characteristics used in the clustering process.

3.1 Data Standardization

Since physical characteristics may be in any units and have any distribution, all input data are standardized prior to the clustering procedure. The standardization process computes the mean and standard deviation of each characteristic or data space axis, then subtracts the mean from each data point and divides by the standard deviation. If x_{ij} represents the value of each characteristic j for each data point i, then the mean

is

$$\bar{\mathbf{x}} = \sum_{j=1}^{N} \bar{x}_j \hat{e}_j \tag{3.1}$$

where

$$\bar{x}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}, \tag{3.2}$$

N is the number of characteristics or data space axes, n is the number of data points or observations, and the $\hat{e}_j$ are orthogonal unit vectors which form the Euclidean data space. The standard deviation is then given by

$$\boldsymbol{\sigma} = \sum_{j=1}^{N} \sigma_j \hat{e}_j \tag{3.3}$$

where

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}{n-1}}. \tag{3.4}$$

New data values are computed for use in clustering as follows

$$\mathbf{y}_i = \frac{\mathbf{x}_i - \bar{\mathbf{x}}}{\boldsymbol{\sigma}} = \sum_{j=1}^{N} y_{ij} \hat{e}_j \tag{3.5}$$

where

$$y_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}. \tag{3.6}$$

Standardization results in a normalized data set where each characteristic has a mean of zero and a standard deviation of one.

The means and standard deviations are saved so that centroid locations resulting from the clustering procedure can be unstandardized to produce mean values in the original units of the observations. If z_{kj} represents the value for each axis j of each resulting centroid k, then the unstandardized centroid values are

$$\mathbf{c}_k = \sum_{j=1}^{N} c_{kj} \hat{e}_j \tag{3.7}$$

where

$$c_{kj} = (z_{kj} \times \sigma_j) + \bar{x}_j. \tag{3.8}$$

The program used to perform data standardization and unstandardization is called *standardize*. The C source code for this program is contained in Appendix A.1.

3.2 The MGC Algorithm

The MGC algorithm starts with a set of k initial seed centroids, one for each of the desired k cluster groups requested by the user. The locations of these initial centroids in data space is determined from the entire collection of input maps using a fast parallel algorithm called the "best of the best." These initial centroids (or seeds) are chosen to be the k most widely distributed observations in data space. On a parallel computer this is accomplished by dividing up the total number of observations among a number of computer processes, p. Each process begins by finding the k best seeds from its portion of the observations. Then half the processes participating in the computation send their k candidate seeds to the other half of the processes. Next, the processes which have $2k$ seeds find the best k seeds from that set. This cycle repeats with the number of active processes reducing by half in each generation until only a single process must find the best k seeds from the best $2k$ candidates, the best of the best.

In the iterative part of the algorithm, each observation is assigned to the centroid to which it is closest by simple Euclidean distance in the N-dimensional data space. In effect, each standardized observation $\mathbf{y}_i$ is assigned to the cluster centroid $\mathbf{c}_k$ which minimizes the function

$$E_{ik} = |\mathbf{c}_k - \mathbf{y}_i| = \sqrt{\sum_{j=1}^{N} (c_{kj} - y_{ij})^2} \tag{3.9}$$

where $\mathbf{c}_k$ and $\mathbf{y}_i$ are N-dimensional vectors. In the cluster code implementation, the square Euclidean distance is used in order to avoid the high computational costs of taking the square root.

At the end of each iteration, once all observations are assigned to a centroid, the N coordinates of all observations assigned to each group are averaged to produce new centroid locations for each cluster

$$\mathbf{c}_k = \frac{\sum_{i=1}^{n_k} \{\mathbf{y}_i \in \mathbf{c}_k\}}{n_k} = \sum_{j=1}^{N} \frac{\sum_{i=1}^{n_k} \{y_{ij} \in \mathbf{c}_k\}}{n_k} \hat{e}_j \tag{3.10}$$

where n_k is the number of observations assigned to the cluster $\mathbf{c}_k$ in the previous iteration and $\mathbf{y}_i \in \mathbf{c}_k$ refers to those n_k observations.

After the centroid locations are computed, another iteration assigning observations to these new centroids begins. The iterative process of classifying observations and adjusting centroid locations continues until fewer than a pre-determined number of observations (or a percentage of the total number of objects) change cluster assignments during an iteration. After the process has converged on a particular classification or grouping, the N coordinates of the k centroids quantitatively define the average characteristics or synoptic conditions for each cluster produced. The nominal stopping rule used is based on the number of observations which change cluster

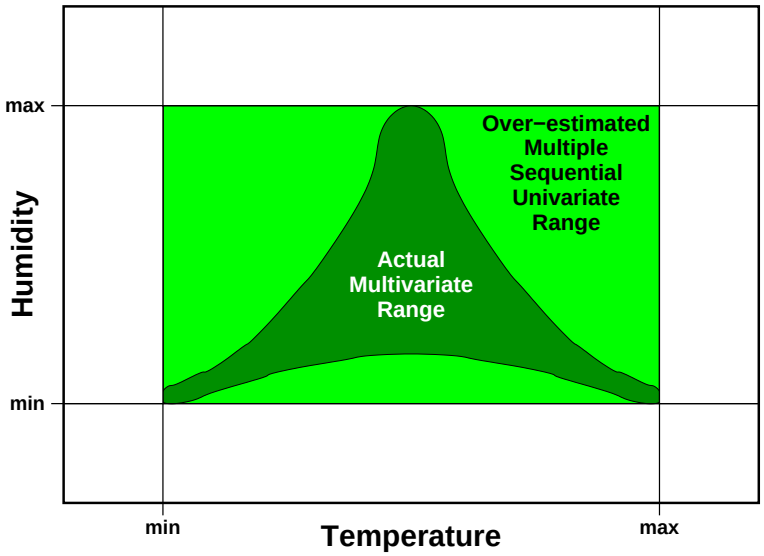


Figure 3.2: Error from sequential univariate analysis. As a true multivariate technique, clustering finds the actual, realized portion of data space.

assignment in a single iteration. When this number drops below 0.5%, clustering iterations stop and the final cluster assignments and centroid locations are saved.

3.3 Examples of Cluster Features

One of the great advantages of clustering techniques is that they are multivariate. That is, they consider all characteristics simultaneously. Some researchers make the mistake of analyzing their data in a sequential univariate way believing they have performed a multivariate analysis. Figure 3.2 illustrates the effect of such an analysis. If some plant or animal can survive within some range of temperatures and some range of humidity values, it is incorrect to assume that all combinations within those ranges are safe. In fact, it could be that this species can withstand temperature extremes only when humidity levels are low. The actual multivariate survival range could be much smaller than that obtained by considering each variable in turn. MGC and other multivariate clustering techniques consider all characteristics simultaneously and would, as a result, yield clusters only within the actual multivariate range.

To better understand how the clustering algorithm partitions data multi-variance, consider the sample data set shown in Figure 3.3. This data set represents some observations projected into a two-dimensional data space, and it has a Gaussian-like appearance similar to many types of geophysical data. Once the data are standardized, as described above, and seed centroid locations are found, it may be clustered at any level of division less than the number of observations. If clustered at a level of division of 15, the data points are grouped into 15 distinct clusters as shown in Fig-

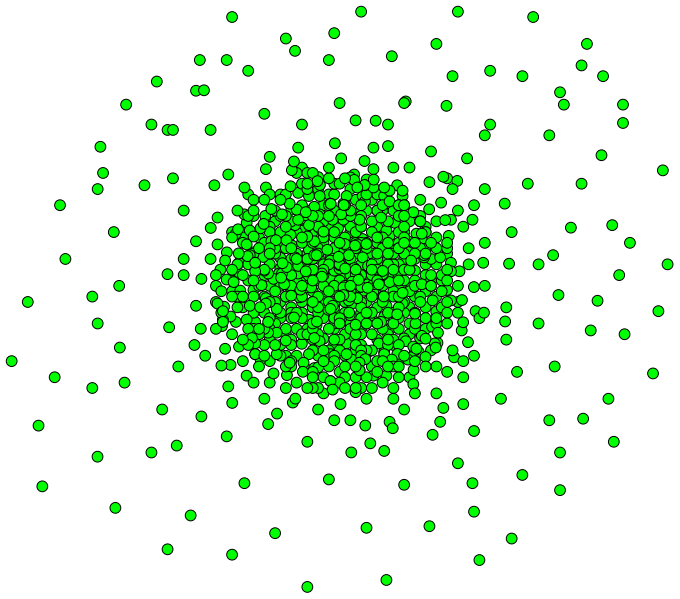


Figure 3.3: An example data set in a two-dimensional data space.

ure 3.4. Here, the data points are colored randomly by their cluster assignment. The centroid locations are represented by the large colored disks which sit at the center of the like-colored observations. Centroids can be considered to form a "skeleton" in the body of observations.

At this level of division, the central data core—the densely populated region of data space—is divided into seven distinct clusters while the sparsely populated regions are divided into eight clusters. In general, centroids are distributed evenly throughout data space; however, their distribution may be rather uneven in more sparsely populated regions. This property is evident in Figure 3.4.

The "radius" of a cluster is often considered to further describe the "fit" of a cluster result. While clusters do not actually have radii, since they would overlap each other in almost any definition, a circular or hyperspherical radius is a natural concept given the Euclidean distance calculations used in k-means clustering techniques. In fact, clusters can be considered to have both inner and outer radii defined to be the distance from the cluster centroid to the nearest and farthest observations assigned to that cluster respectively. Such inner and outer radii are also shown in Figure 3.4.

The outer radii of clusters in the densely populated regions of data space are usually nearly equal; such clusters partition the data variance in an equal fashion. Meanwhile, in sparsely populated regions of data space, outer radii tend to be larger, highly variable, and strongly dependent on the chosen level of division. In the densely populated region of data space, the inner radius of a cluster is usually small or zero, while

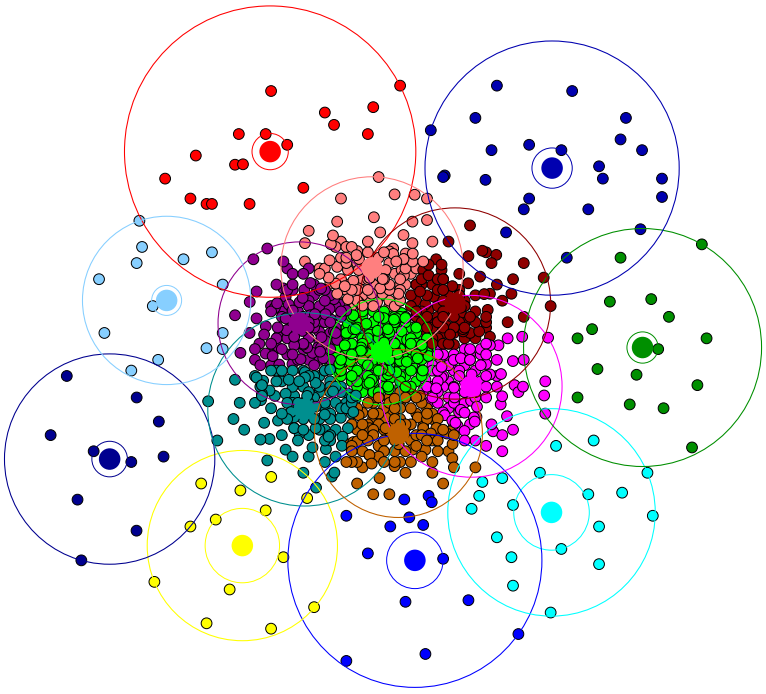


Figure 3.4: Centroid locations and inner and outer radii for 15 clusters. The data points are colored by their cluster assignment.

clusters in the sparsely populated regions of data space are likely to have non-trivial inner radii. Grouping outliers into a few large clusters while equally partitioning the majority of the data makes clustering valuable for scientific analysis.

In a sense, the number of clusters defines the "resolution" at which the multivariance of the data is "sampled." The magnification of a microscope can serve as analogy to this "resolution." When the magnification is low, only the largest features of an object are visible. When the magnification is increased, much more of the detail of an object becomes visible. When a complex object is viewed at a single magnification, that magnification must be sufficiently high to expose the finest level of detail desired, even if the detail of interest is isolated in a particular region of the object. The rest of the object may be homogeneous and not warrant such a high magnification.

Choosing too few clusters will result in clusters with very large outer radii (and very broad regimes when projected back into geographic space) which will likely cause the researcher to miss salient features. Choosing too many clusters will cause very small details and the effects of outliers to be amplified in the analysis. Desired is an adequate sampling of the data multivariance which exposes sought-after features. A suitable number of clusters is usually chosen—just like the magnification of a microscope—by trying a few different orders of magnitude. The exact number, however, is not particularly important.

The results from clustering the example data set at other levels of division are shown in Figure 3.5. Some of the general features of clustering described above do not hold at all levels of division. It is unlikely that choosing 2, 3, 4, or 5 levels of division will produce a useful result. Similarly, choosing 25 or 30 clusters is probably overkill. Instead, requesting between 10 and 20 clusters in this case should produce reasonable results, but the exact number is usually not significant.

MGC has been shown to produce useful clues to plant and animal habitats when certain geophysically-relevant variables are chosen as data space coordinates, and the same technique can be applied to feature detection since almost *any* set of geographically-coincident properties may be used as a data space coordinates to characterize a map cell. In this study, the intensities of reflected radiation at different wavelengths obtained from hyperspectral imagery serve as the characteristics for identifying surface features or patterns. The clusters uniquely classify land surface areas based on their spectral properties.

Hyperspectral imagery can be considered in geographic space (one layer or wavelength at a time), in spectral space (on cell or pixel at a time), and in N-dimensional data space (all cells in the Euclidean space formed by using each wavelength as an orthogonal axis). In the analysis which follows, this imagery will be shown in geographic and spectral space.

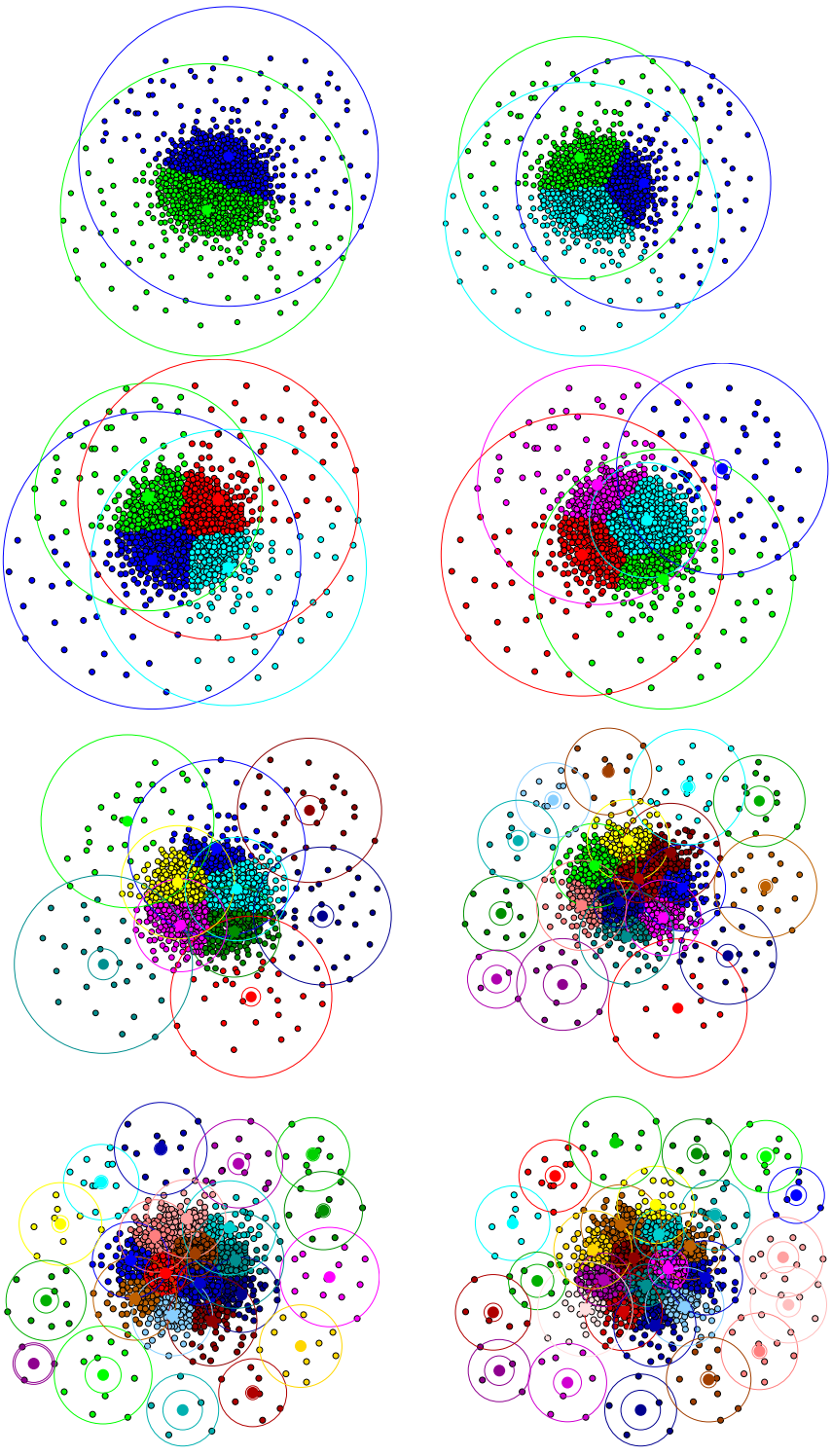


Figure 3.5: Clustering results for example data at various levels of division. The levels of division shown here are 2, 3, 4, 5, 10, 20, 25, and 30.

Chapter 4

Study Area

The Tallgrass Prairie Preserve (TPP) was created in 1989 when The Nature Conservancy (TNC) purchased the 30,000 Barnard Ranch north of Pawhuska in Osage County, Oklahoma. See map in Figure 4.1. Now spanning over 38,000 acres, the TPP partially fulfills TNC's goal to recreate a functioning tallgrass prairie ecosystem. Tallgrass prairie once covered over 142 million acres and was one of North America's major ecosystems supporting bison and many other animal and plant species. Today, large tracts of tallgrass prairie exist only in the Flint Hills of Oklahoma and Kansas. The TPP is managed through a series of "prescribed" burns conducted at different times of the year to mimic the original seasonality of presettlement fires. Fire removes dead vegetation, controls encroaching woody vegetation, and increases the vigor and flowering of many plant species. In 1993, TNC reintroduced a herd of 300 bison to the prairie and established the Adopt-a-Bison program to support and preserve their habitat in the Preserve. The tallgrasses—big bluestem, Indiangrass, and switchgrass—can grow as tall as eight feet in moist, deep soils reaching maximum height in August and September. The TPP and surrounding area consist of large watersheds which extend into cross-timber forests composed primarily of post and blackjack oak. A diversity of grassland and forest habitats support a wide variety of wildlife including the lesser and greater prairie chickens, numerous other bird species, white-tailed deer, bobcats, armadillos, beavers, woodchucks, badgers, and coyotes[1].

The Osage Reservation in Indian Territory was confirmed by Congress in 1872 and later became known as Osage County, Oklahoma. Being too difficult to till, most of the Osage land is used for cattle ranching and oil and gas production. During the "Oil Boom" of the 1920's, the Osage nation became the richest native people in the United States. While oil exploration and production in this area has declined in the intervening years, according to the website of the Osage Nation (http://www.osagetribe.com/), the Osage presently have 13,251 actively producing oil and gas wells yielding 11,738 barrels of oil per day and 15,200 million cubic feet

[1]Description of the Tallgrass Prairie Preserve adapted from The Nature Conservancy website at http://nature.org/wherewework/northamerica/states/oklahoma/preserves/tallgrass.html

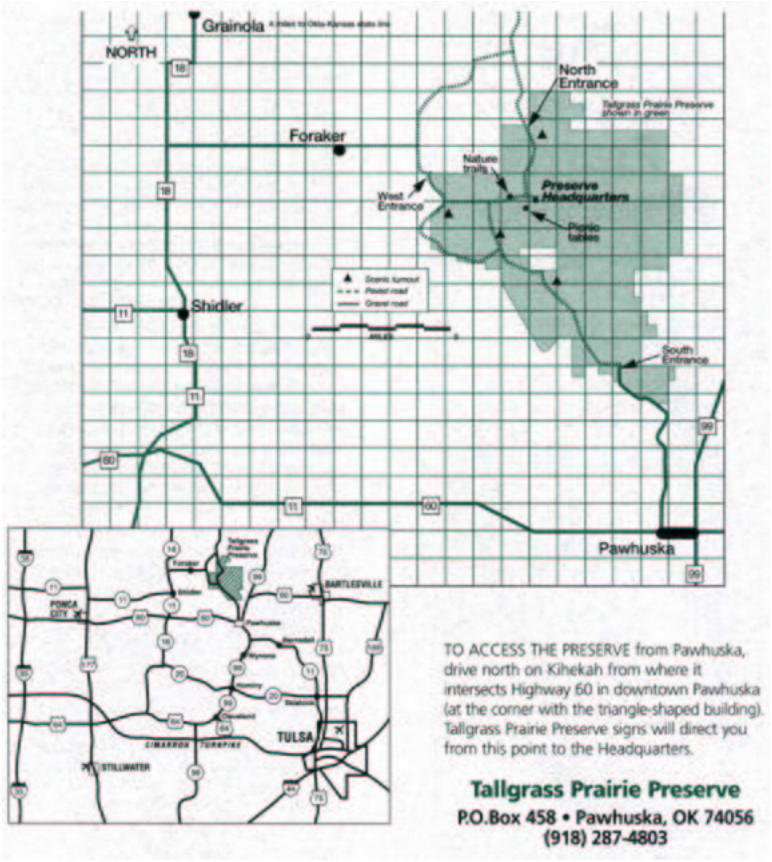


Figure 4.1: Map and driving directions to the Tallgrass Prairie Preserve.

of gas per day. Over 100 active oil wells, out of some 300 wells drilled, are located within the boundaries of the Tallgrass Prairie Preserve. These wells produce 15 to 20 barrels of oil per day and ten times that volume of brine [25], and the brine may have salt concentrations exceeding 100,000 mg/kg [8]. This produced water forms naturally with oil and gas. When it is extracted with the petroleum, it is piped via "gather" lines to a central separator where oil is removed. It then travels back to unproductive wells, where it is pumped back down into the Earth to help force up additional oil. Oil and brine spills resulting from drilling and pumping, referred to as "upstream petroleum activities," present environmental risk to the prairie ecosystem. Brine spills can have a more significant and longer-term impact on the Earth than hydrocarbon spills [25]. Brine spills in this area usually kill all vegetation in the affected area resulting in top soil erosion, soil compaction, and a significant geophysical disturbance referred to as a "brine scar." Recovery from loss of top soil and brine contamination can take decades to centuries if the affected site is not remediated. Remediation usually consists of application of gypsum and hay to the soil; however, significant soil contamination may require leaching and draining strategies while total top soil loss may necessitate complete top soil replacement. Although ecologists have attempted to assess and quantify the risk of spills to wildlife in this region [10], remediation of brine scars remains a topic of current environmental research [24].

Chapter 5

Methods

Since the objective of this study was to detect brine spills and scars from hyperspectral imagery, some independent means of verification was needed. Although ORNL staff visited the TPP for a related project, funding was not available for travel to the site to verify brine scar predictions. Fortunately, 1 m orthophotos for Osage County were found, and loaded into a Geographic Information System (GIS) for storage, display, and manipulation. Since these orthophotos contained geographic coordinate information, they served as the basis for georectification of the AVIRIS hyperspectral imagery as well as a means for visually verifying some brine disturbance predictions.

5.1 Hyperspectral Imagery

Seven AVIRIS scenes from two flight lines over Osage County, Oklahoma, were obtained from the U.S. Geological Survey EROS Data Center for this investigation. Scenes 3, 4, 5, and 6 from flight line Tallgrass Prairie Preserve 4 (labeled f990719t01p03_r05) and scenes 4, 5, and 6 from Tallgrass Prairie Preserve 6 (labeled f990719t01p03_r06) were obtained since they cover almost the entire TPP and some of the surrounding county. These data were acquired by AVIRIS flying on the ER–2 aircraft on 19 July 1999, under clear sky conditions. Their spatial and temporal coordinates are shown in Table 5.1. Quicklooks of the flight lines, which were flown in opposite directions, are shown in Figure 5.1. In the discussion which follows, scenes will be referred to as r0Xsc0Y where X is the flight line number (5 or 6) and Y is the scene number (3, 4, 5, or 6 for flight line 5 or 4, 5, or 6 for flight line 6).

Table 5.1: AVIRIS flight lines used for analysis

Flight Line	**Latitude**		**Longitude**		**Time**	
	Start	**Stop**	**Start**	**Stop**	**Start**	**Stop**
f990719t01p03_r05	+36°42'54"	+36°54'30"	−96°12'48"	−96°36'36"	17:02:00	17:09:00
f990719t01p03_r06	+37°00'06"	+36°48'48"	−96°36'36"	−96°12'48"	17:13:00	17:21:00

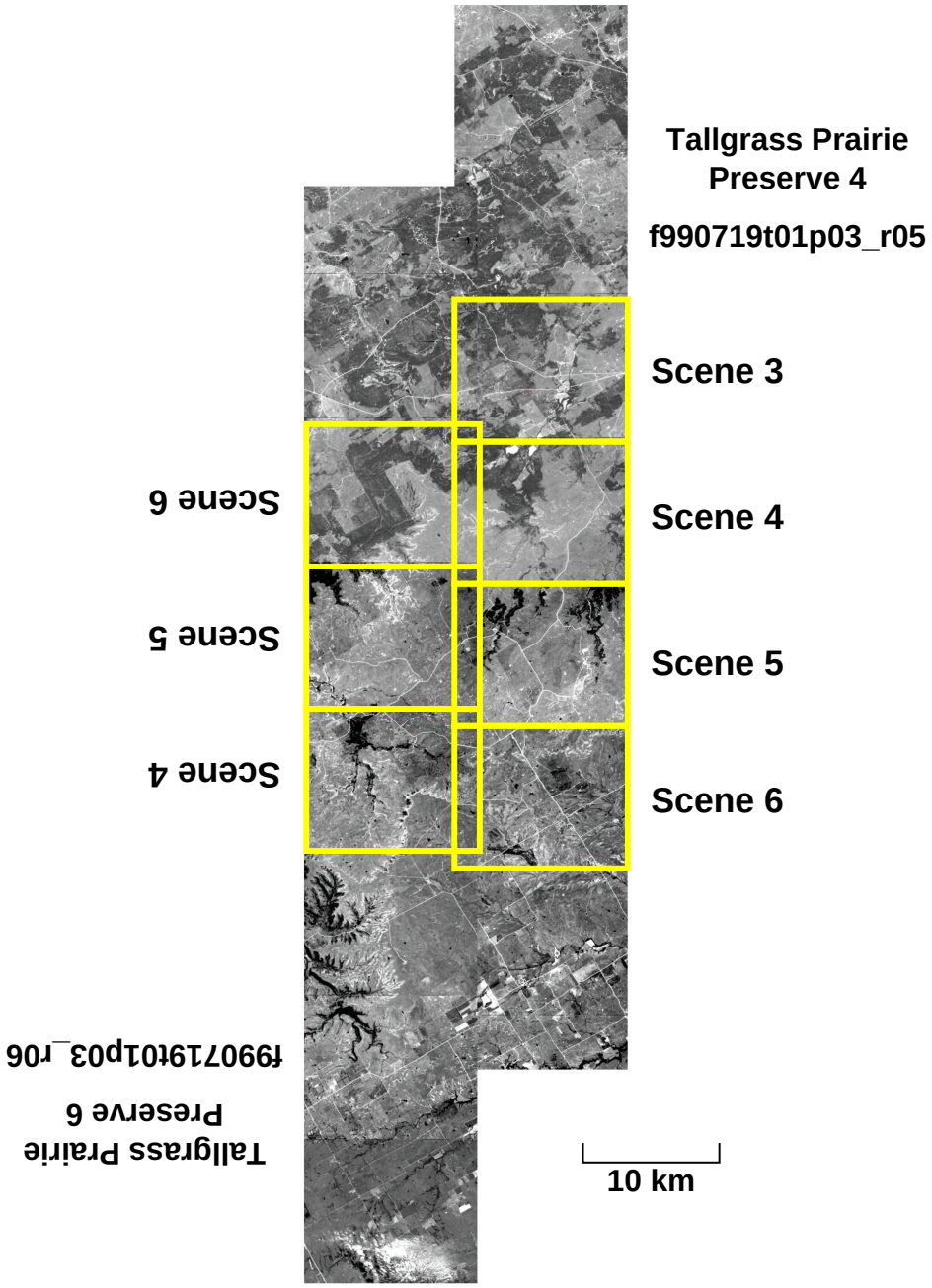


Figure 5.1: The AVIRIS scenes from two flight lines used for analysis.

The AVIRIS data were stored as 16 bit big endian binary integer band-interleaved files, one for each scene. A program called *read_data* was written to extract the data from the scene files into band-interleaved ASCII files for clustering and into separate band files (224 for each scene) for loading into the GIS. The GIS files are written as 16 bit big endian binary 2s complement unsigned integers. Since this conversion was performed on a little endian system, the program performed byte swapping for reading and writing binary files. The source code for *read_data* is listed in Appendix A.2.

5.2 Digital Orthophoto Quadrangles

A digital orthophoto quadrangle (DOQ) is a computer-generated image produced from an aerial photograph. A DOQ has been orthorectified: altered pixel by pixel to produce an image with the geometric features of a map with true geographic positions. The U.S. Geological Survey (USGS) produces standard quads as gray-scale or color-infrared images covering 7.5 minutes longitude by 7.5 minutes latitude. A digital orthophoto quarter quadrangle (DOQQ) covers one quarter of the full 7.5 minute quad. DOQQs have identification numbers, and the last digit is a number from one through four identifying its position in the full DOQ.

While the USGS sells DOQQs via its EarthExplorer system and through its business partners, links to 1-meter resolution DOQQs for Osage County, Oklahoma, were found on the Geo Information Systems website at http://geo.ou.edu/ run by the University of Oklahoma. The files were downloaded from a File Transfer Protocol (FTP) server operated by the Oklahoma State GIS Council. Fifty-two DOQQs were loaded into a GIS where they were used to georectify the AVIRIS hyperspectral scenes and visually identify geophysical features observed in the 25 meter resolution hyperspectral imagery. Scripts used for downloading the DOQQs and loading them into the GIS are listed in Appendix A.4. The source aerial photographs were all taken in February, March, or April 1995 and were produced in August 1998.

5.3 Geographic Information System

The Geographic Resources Analysis Support System (GRASS) GIS was used to store, manage, georectify, and display the hyperspectral imagery. Originally developed by the U.S. Army Corps of Engineers Construction Engineering Research Laboratories (CERL) to support land management and environmental planning by the military, GRASS is now an Open Source[1] software package released under the GNU General Public License (GPL) [28, 3]. Widely used by researchers in academia and government laboratories, it provides raster and vector capabilities, an image processing system, and a graphics production system. GRASS is written in the C language and runs on a variety of computer platforms including Linux and most Unix variants. GRASS

[1]For the Open Source Definition, see http://www.opensource.org/

continues to be developed as a community project. It's development is organized primarily in Europe, and the official GRASS website is in Italy at http://grass.itc.it/. It is also distributed as a part of the FreeGIS Project at http://freegis.org/.

DOQQs were loaded into GRASS using the *r.in.doq* command. A number of the files containing the DOQQs were truncated by a few bytes. When *r.in.doq* encountered the end of the file before it was expected, it printed an error and exited. Since only a few bytes were missing in the last raster line of the affected DOQQs, it would have been preferable for the *r.in.doq* program to fill in the missing data and complete the loading of the imagery into GRASS. To solve this problem, the source code for *r.in.doq* was modified to fill in the value 255 for the missing cells. In *main.c*, the loop which reads the DOQ data, shown in Algorithm 1, was modified to report an error and pad the missing cells as shown in Algorithm 2. If GRASS were not an Open Source package, loading the slightly truncated DOQQs would not have been possible without modifying each affected binary file. The entire set of DOQQs obtained for this study is shown in Figure 5.2.

Hyperspectral scenes were georectified against the DOQQs in GRASS using the *i.group*, *i.points*, and *i.rectify* commands following the procedure described by Harmon and Shapiro [20]. All of the 224 images for each scene, one per band as output by *read_data*, were imported into the GIS. They were grouped together by scene using the *i.group* command. Next, *i.points* was run on each scene to pick points in common between the DOQQs and the AVIRIS imagery. This laborious process results in a set of calculated coefficients for a second order least squares transformation which projects groups of images into the geographic space of a corresponding map in the GIS. Road intersections were the easiest features to match in this rural area. Imagery of AVIRIS band 200 in which roads are strongly visible was used to match points in the 1 m DOQQs. Approximately two scenes were georectified per eight hour day. Since the resolutions were different by a factor of 25, selecting high quality points in the two sets of images was a challenge. It was often necessary to drop previously-selected points to reduce the error in the transformation model. Once a desired set of points was selected, the *i.rectify* command was used to transform all imagery from all bands simultaneously. The rectification model was saved by the GIS so that future rectification of the same AVIRIS images could be performed. Since the cluster analysis was performed on the raw AVIRIS imagery, the cluster results were later rectified using the stored model for each scene.

Additional GIS coverages of roads, oil wells, and rivers and streams were obtained from files originally compiled by Tom Ashwood for a related project [5], but these were not used in the analysis. These coverages contained no metadata, and the roads and streams coverages had a lower resolution than that of the DOQQs. The TPP boundary was also obtained from Ashwood's files and was used to determine which brine scar prediction regions were located within the borders of the TPP. The seven AVIRIS scenes (outlined in yellow) are shown overlain onto the full set of DOQQs (outlined in green) along with roads (gray) and rivers and streams (blue) in Figure 5.3.

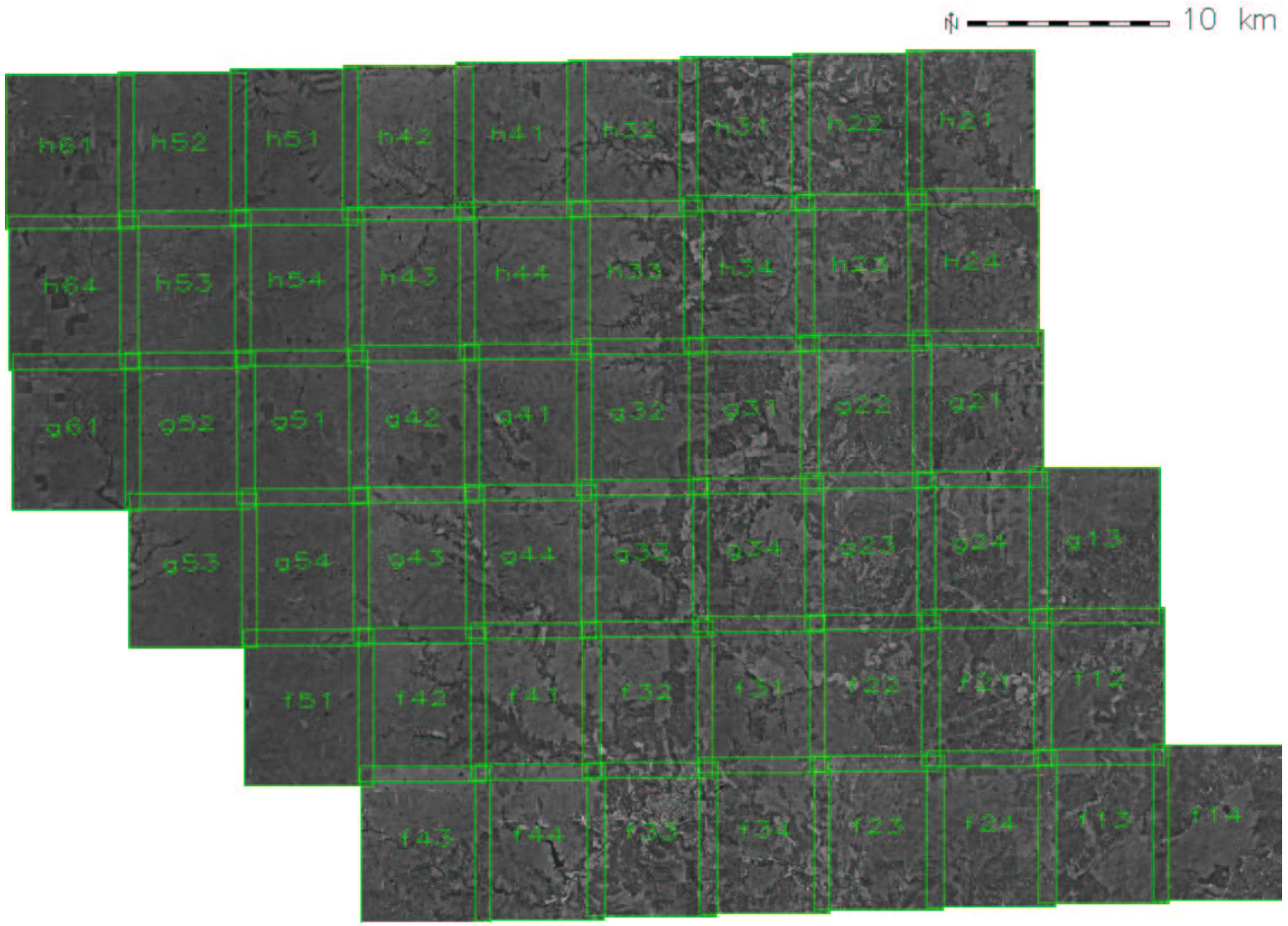


Figure 5.2: All 52 DOQQs obtained for the study area.

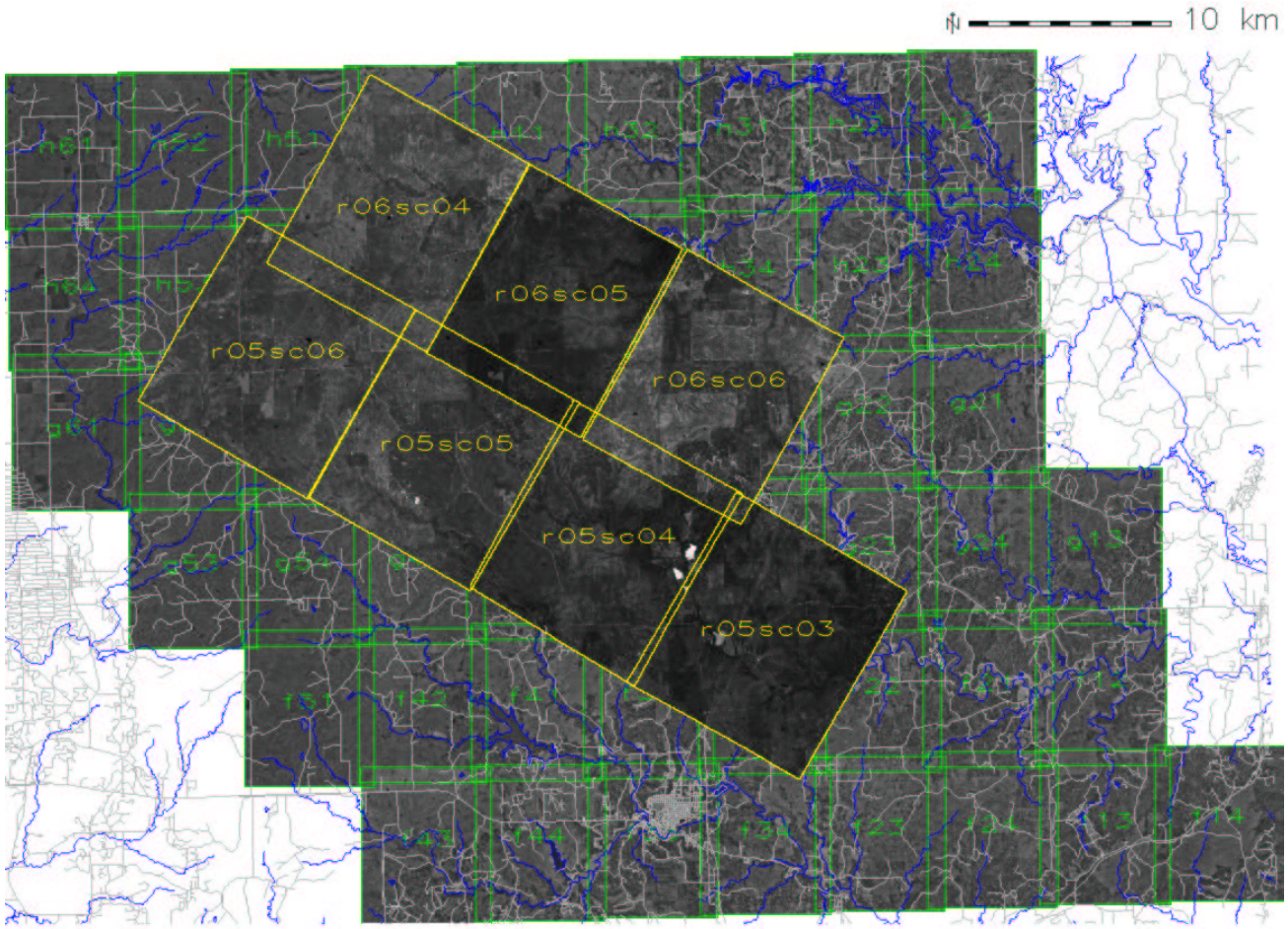


Figure 5.3: The seven georectified AVIRIS scenes. All seven scenes (yellow) are overlain onto the DOQQs (green). Additional coverages of roads (gray) and rivers and streams (blue) are shown.

Since more DOQQs were obtained than were needed, a new region was defined in the GIS which just encapsulated the seven AVIRIS scenes, and the unnecessary DOQQs were eliminated from display. Figure 5.4 shows this new region with the AVIRIS scenes overlain on the relevant DOQQs with the TPP boundary outlined in red.

Algorithm 1 This loop reads DOQ data in *r.in.doq*.

```
   for (row = 0; row < nrows; row++)
   {
      if (fread((char *) bcell,1,ncols,fd) != ncols) {
         char msg[100];
         sprintf(msg,"error in reading row %d",row);
         G_fatal_error (msg);
         exit(1);
         }
      for (i=0;i<ncols;i++) {
         cell[i] = (CELL)bcell[i];
         if (max < cell[i]) max = cell[i];
         }
      G_put_raster_row (cf, cell, CELL_TYPE);
   }
```

Algorithm 2 This modified loop reads DOQ data and pads missing cells in *r.in.doq*.

```
   for (row = 0; row < nrows; row++)
   {
      if ((got_cols = fread((char *) bcell,1,ncols,fd)) != ncols) {
         fprintf(stderr, "Error reading row %d; got %d columns; \
padding remaining %d cells with 255\n", row, got_cols,
            (ncols - got_cols));
         for (i = got_cols; i < ncols; i++)
            bcell[i] = 255;
      }
      for (i=0;i<ncols;i++) {
         cell[i] = (CELL)bcell[i];
         if (max < cell[i]) max = cell[i];
      }
      G_put_raster_row (cf, cell, CELL_TYPE);
   }
```

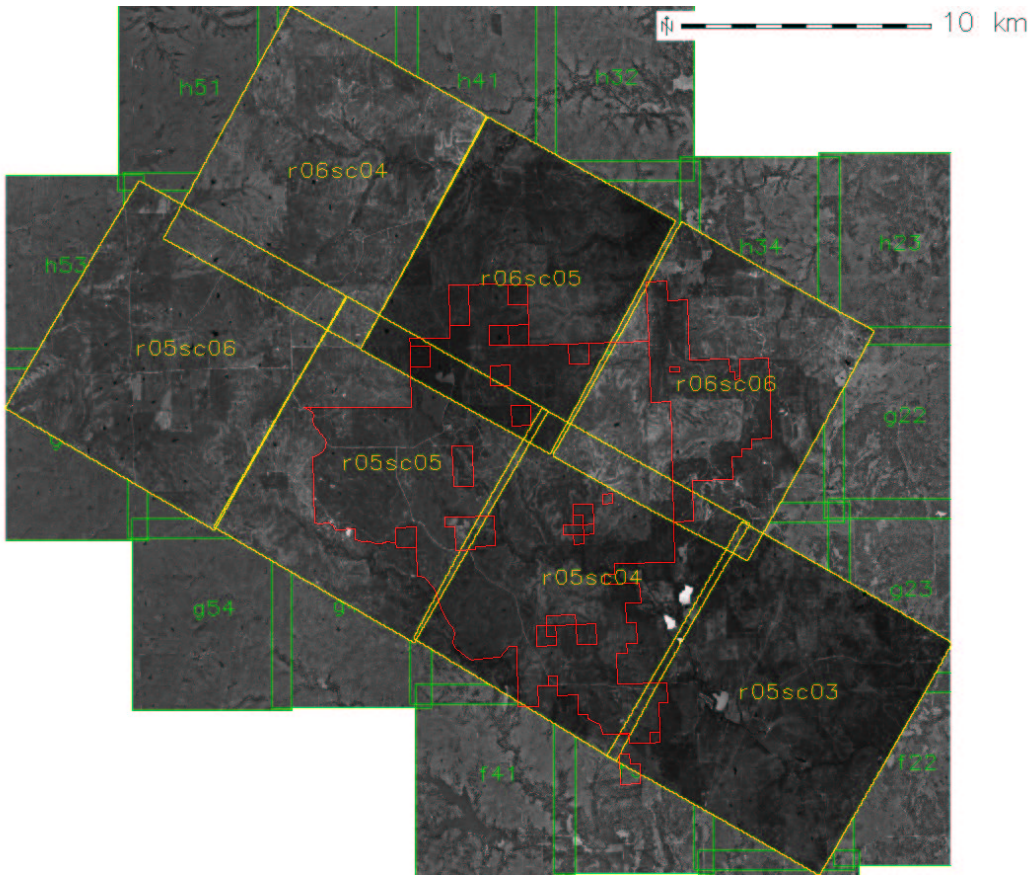


Figure 5.4: The geographic region encapsulating the AVIRIS scenes. The scenes (yellow) and relevant DOQQs (green) are shown with the TPP boundary (red).

5.4 Cluster Analysis

Scene r05sc05 was chosen for an initial cluster analysis because it was known to contain large, well-documented brine scars within the borders of the TPP. This area had been visited by ORNL and LLNL staff, and boundaries of these two brine scars were recorded by Bill Hargrove using a Global Positioning System (GPS) during that site visit.

Scene r05sc05 was clustered at a few different levels of division, and all of the results showed cohesive geographic regions which appeared to be individual surface features. Next, intensity maps of each of the 224 AVIRIS bands for this scene were visually inspected to understand the contribution of each to the clustering results. A stack of such intensity maps is shown in Figure 2.1 with band 20 exposed on the top of the image cube. As expected, the intensity maps for bands at wavelengths of strong water vapor absorption (see Figure 2.3) showed no significant deviation from zero. Most of these maps contained only random or instrumental noise; however, some linear patterns in the noise were visible. Since noise in these bands should not contribute to feature detection, these bands were removed from all the data sets used for subsequent cluster analyses, reducing the number of characteristics or data space axes from 224 to 187. All of the 224 calibrated instrument channels and their uncertainties are shown in Table A.1. Notice that bands 33–35 are out of order with respect to increasing wavelength; as a result, all spectra plots of AVIRIS data were sorted by wavelength. The 187 bands retained for subsequent cluster analysis are denoted in the table by a check (✓) in the sixth column.

Using the 187 remaining bands, all seven AVIRIS scenes were clustered independently at integral levels of division from 13 to 22. The clustering results for a level of division of 16 for all seven scenes after being georectified are shown as randomly colored maps overlain onto the DOQQs in Figure 5.5. As is evident in this figure, interesting surface patterns are extracted by the clustering algorithm; however, because each scene was clustered independently, cluster numbers and definitions (and therefore map colors) are different across all the scenes. Since the extents of two brine scars in scene r05sc05 were known, a method for using the spectral signature information from those disturbances to find additional spills across all the scenes was desired. While all the scenes could have been clustered simultaneously, the fact that each scene was calibrated independently with different absolute intensities compromises the results of this method.

To achieve the desired result, it was first necessary to find a suitable level of division for only the r05sc05 scene. This scene was clustered at every level of division from 13 to 32, and at each level of division, the cluster representing the known brine scars was determined from the resulting cluster maps after georectification. Figure 5.6 shows a histogram of brine scar predictions in terms of the number of cells (after georectification) for the entire scene at all levels of division. In this figure, each group of cells is colored uniquely so that each group can be identified across all the levels of division shown. As expected, as the level of division is increased,

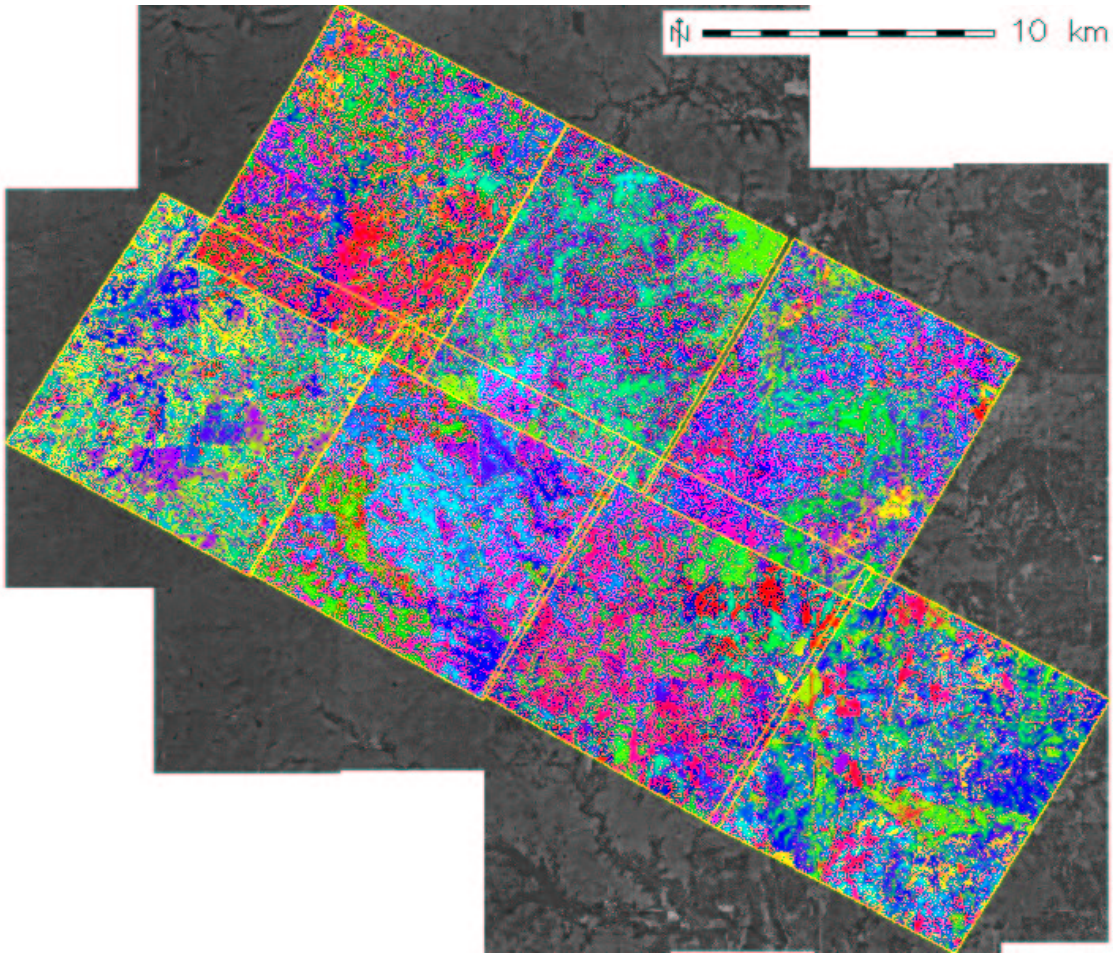


Figure 5.5: All scenes clustered independently into 16 pieces and georectified.

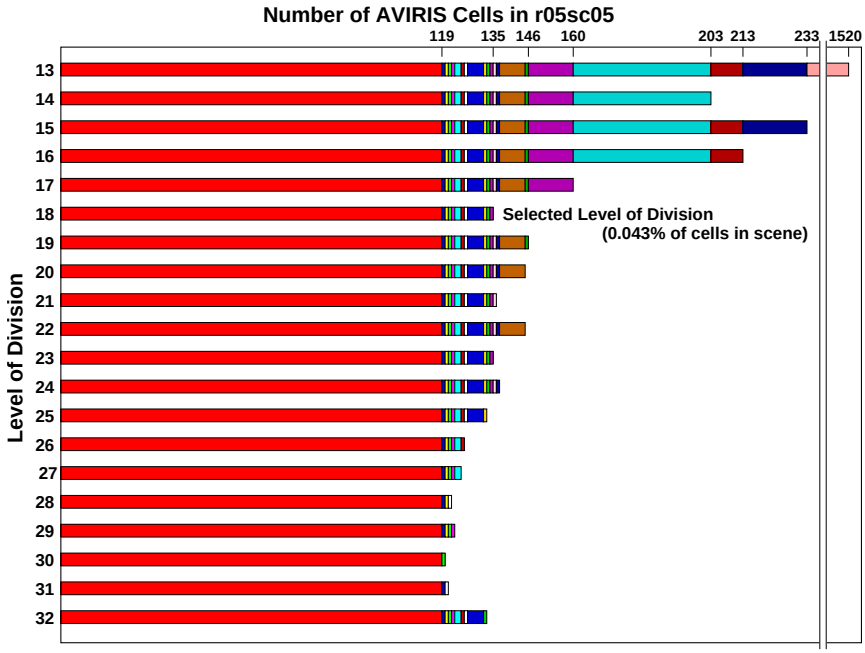


Figure 5.6: Histogram of brine scar prediction cells in r05sc05.

the cluster becomes more unique or tightly defined in data space. Therefore, it's cell membership is correspondingly reduced. However, this is not strictly true as the histogram demonstrates. The same data are presented in tabular form in Table 5.2.

All of the cluster results predict the same 119 cells as brine scars (shown in red on the left in Figure 5.6). At a level of division of 13, far too many cells, 1520 in total, are included as brine scar predictions, but at all other levels of division, the brine scar cell membership is 233 or fewer. At 30 clusters, brine scar cell membership is the lowest at 120. It would appear that this cluster defines a unique and rather stable set of observations in data space across many levels of division. The number of cells defined as brine scars will not depend significantly on the exact level of division chosen between 17 and 32. Based on the stability of the brine scar predictions for this scene and the analysis of brine scar prediction maps which follow, the clustering result for a level of division of 18 was selected. At this level of division, the brine scar cluster consists of 135 cells or 0.043% of the total number of cells in the entire scene. Maps of brine scar predictions for two well-known brine scars were analyzed to verify predictions at all of these levels of division.

Table 5.2: Map statistics across levels of division for scene r05sc05.

13	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**Cells**
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	119
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	1
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	1
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		✓	1
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓			✓	1
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					✓	2
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓						✓	1
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓			✓	✓	1
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓							✓	5
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓								1
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓								✓	1
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓									1
✓	✓	✓	✓	✓		✓	✓	✓	✓		✓									1
✓	✓	✓	✓	✓		✓	✓		✓		✓									1
✓	✓	✓	✓	✓		✓	✓		✓											8
✓	✓	✓	✓	✓		✓														1
✓	✓	✓	✓	✓																14
✓	✓	✓	✓																	43
✓		✓	✓																	10
✓		✓																		20
✓																				1287
																				312848

The two large, well-documented brine scars in r05sc05 are easily visible in the DOQQs in Figure 5.7. The red line is the TPP boundary and the white lines connect the GPS points collected by Hargrove during a site visit in 13 June 2001. The brine scar on the left is a result of a leaking produced water tank north of the roadway. That tank is no longer standing. The path which follows the arc of the TPP boundary on the left then passes south of the brine scar on the right is an old railroad bed which no longer contains tracks. The brine scar on the right is a very old and heavily eroded area which is bisected by a fence which follows the TPP boundary. Hargrove did not cross the fence, so the GPS points follow the fence line. This scar is often referred to as Site 5, and it appears in aerial photographs as far back as 1937[2]. A photograph taken within the interior of this brine scar in 1999 is shown in Figure 5.8. A 2001 photograph of the portion of this scar which extends beyond the TPP boundary is shown in Figure 5.9.

Contained in Figures 5.10 and 5.11 are the brine scar predictions produced from the cluster analysis of r05sc05 at levels of division from 13 to 30. While at 13 clusters too many spots are included in the predictions, all the other figures show very similar results. There are 86 cells common to all the predictions. The same 90 cells are "turned on" in both the 18 and the 23 clusters maps for this area. Aside from the prediction for 13 clusters, all the other levels of division predict the same three discrete brine spills: the two large ones described above and another one alongside the bare railroad bed in between the other two. Consistent with the histogram shown in Figure 5.6, the choice of the level of division between 17 and 30 produced brine scar predictions which were not significantly different from each other.

Brine scar prediction maps for the entire r05sc05 scene were inspected to ensure that roads and other bare ground areas were not substantially included in the predictions. Figure 5.12 shows four such maps for 13, 17, 18, and 30 clusters. The map for 13 clusters shows that many roadways and bare ground areas are included in the cluster representing brine scars. As concluded twice above, this level of division is too low to generate meaningful predictions. However, the predictions from the cluster analyses with levels of division of 17, 18, and 30 are not substantially different.

Figures 5.13 and 5.14 show images of some of the individual clusters which are easily recognizable in the r05sc05 scene at a level of division of 18. Main roads appear in cluster 11; gravel roads and driveways, main road shoulders, and bare ground areas appear in cluster 9; shallow water, including streams and creeks, shallow ponds, and pond edges, appear in cluster 18; deeper water and pond centers appear in cluster 5; shrubs and more woody vegetation, including cross-timbers which tends to lie along the stream and river banks, appear in cluster 3; cluster 17 appears to capture one species or life stage of a grass; and cluster 15 constitutes the brine scar predictions. The dense pattern and hard edge seen in cluster 17 follows the north-south border of the TPP and is due to the managed burns which take place on the Preserve. The last panel in Figure 5.14 shows only these recognizable clusters on the same image.

[2]Kerry Sublette, personal communication.

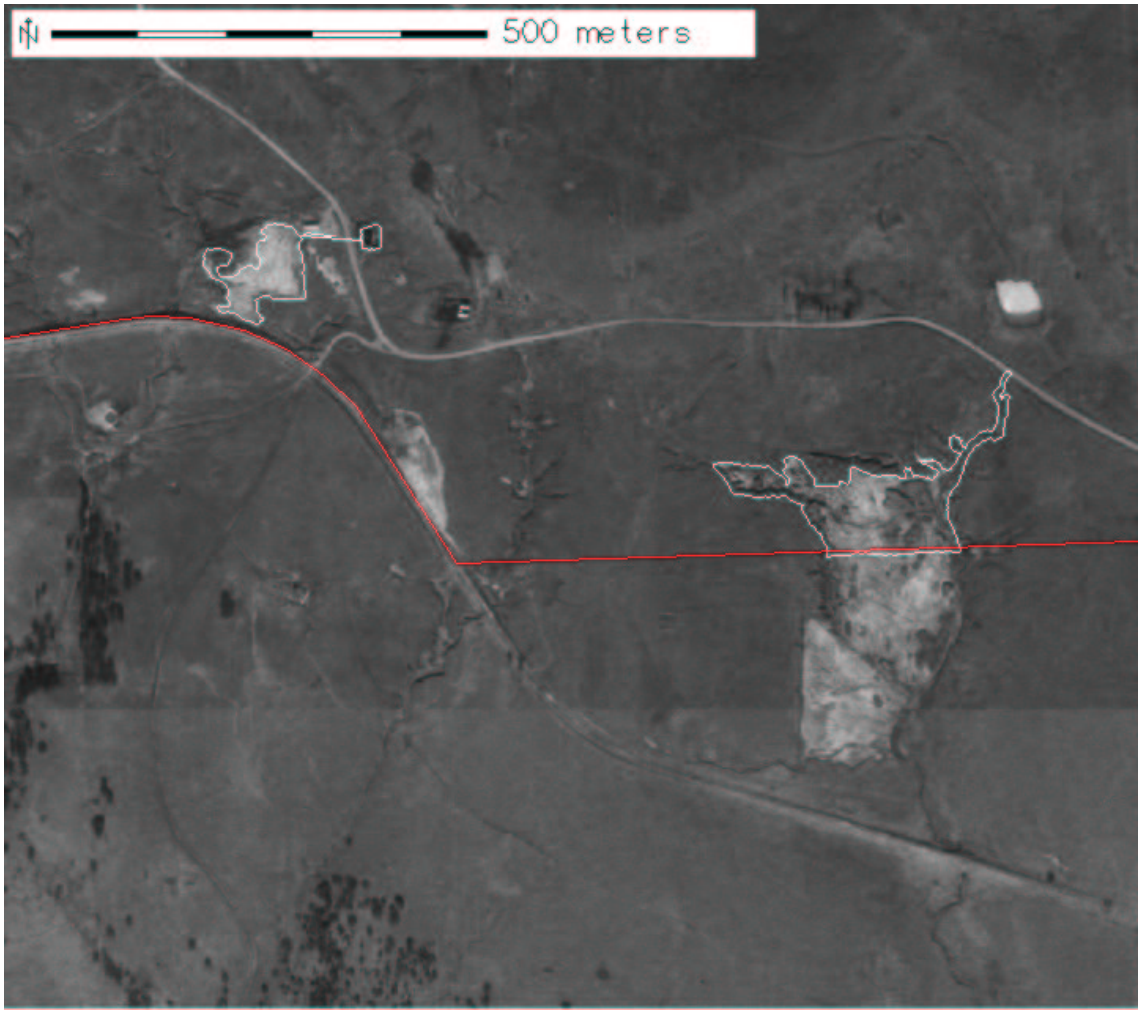


Figure 5.7: Two large, well-documented brine scars near the TPP boundary. The red line is the TPP boundary and the white lines are the brine scar boundaries as determined from Hargrove's GPS points.

Figure 5.8: Part of the interior of the oldest and largest brine scar on the TPP. Photo courtesy of Kerry Sublette (1999).

Figure 5.9: Portion of Site 5 brine scar outside the TPP boundary fence. Photo courtesy of Tina Carlsen (June 13, 2001).

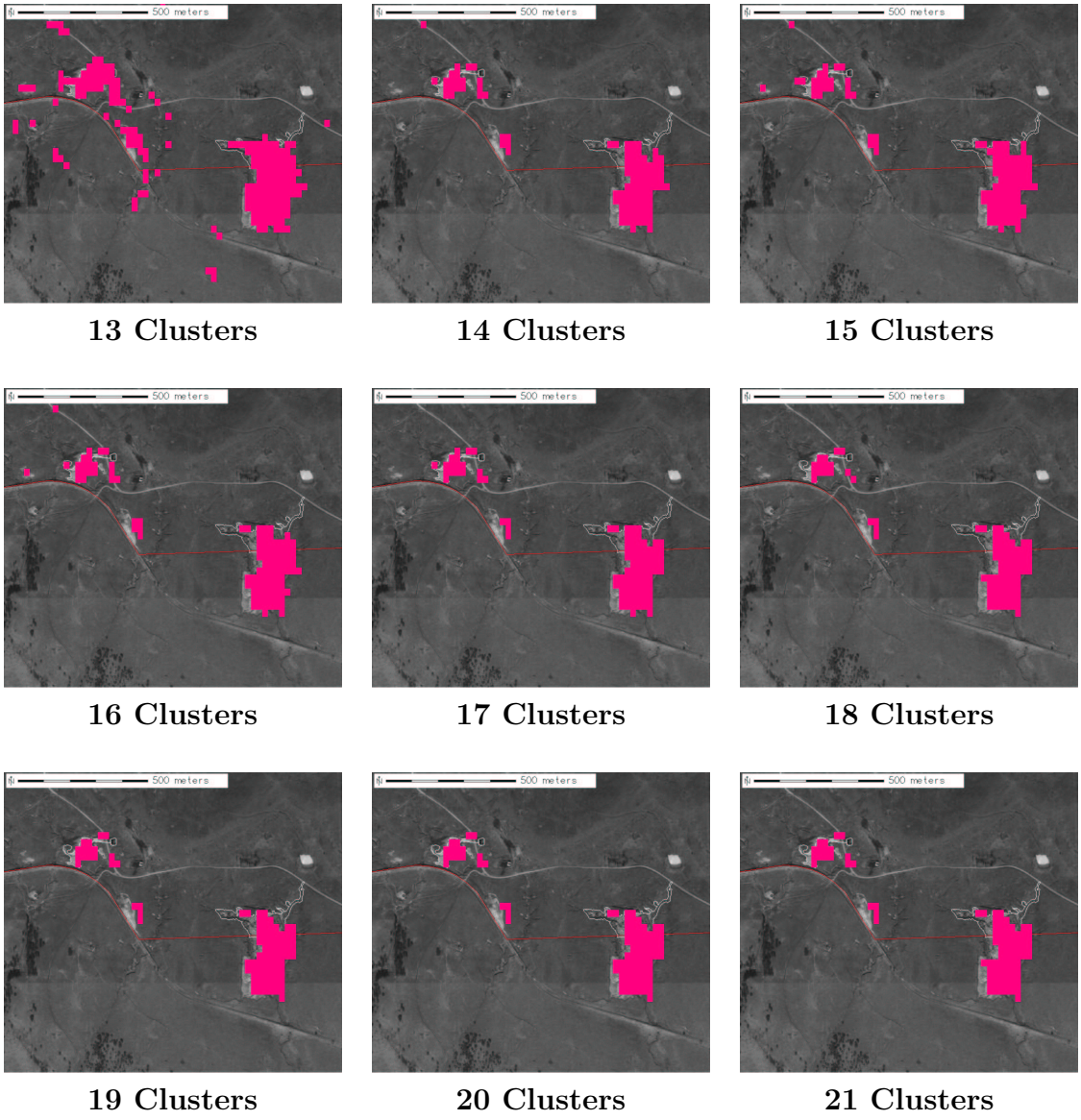


Figure 5.10: Well-known brine scars clustered at levels of division 13–21.

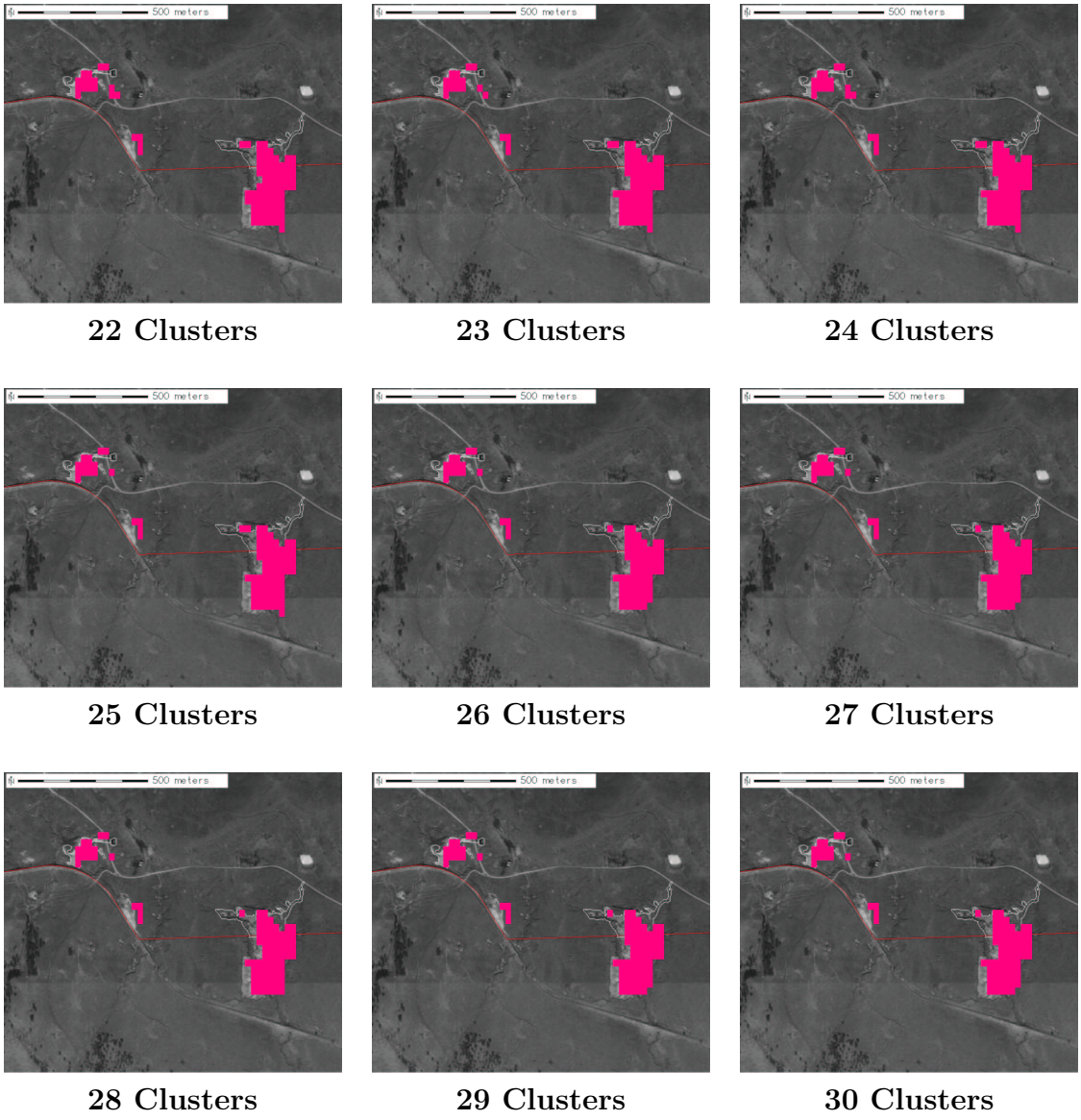
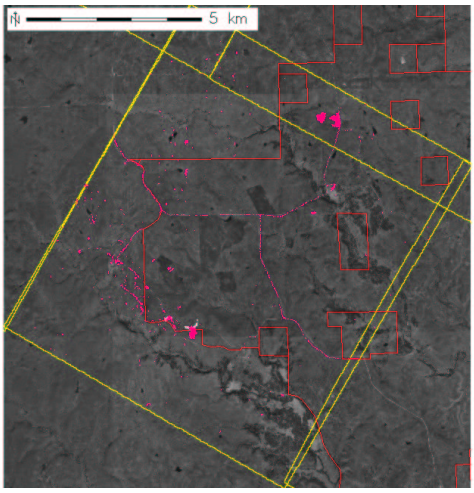
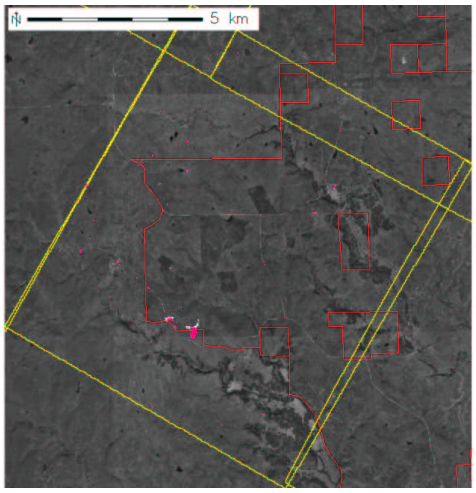


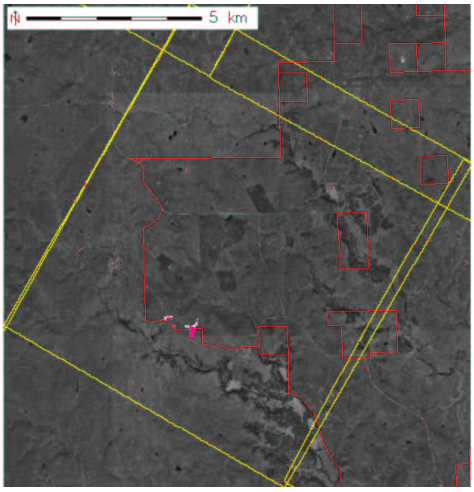
Figure 5.11: Well-known brine scars clustered at levels of division 22–30.



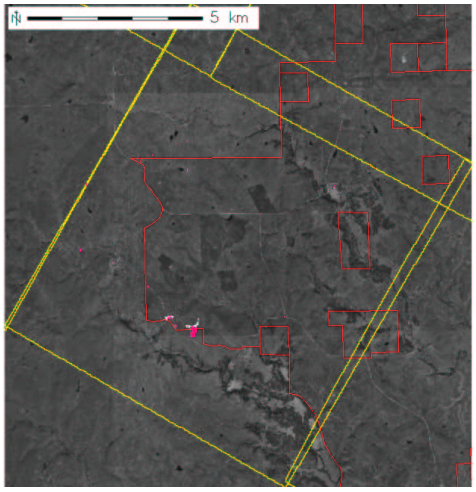
13 Clusters



17 Clusters

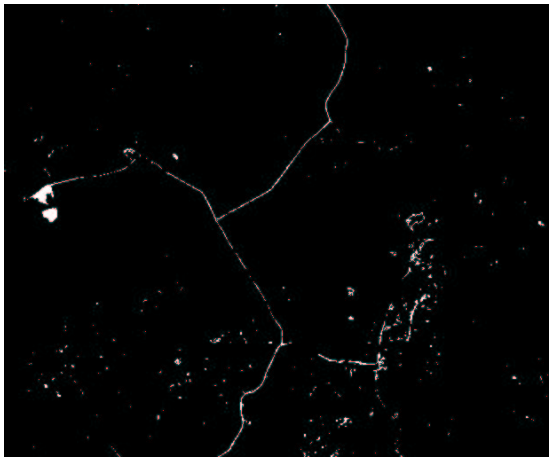


18 Clusters

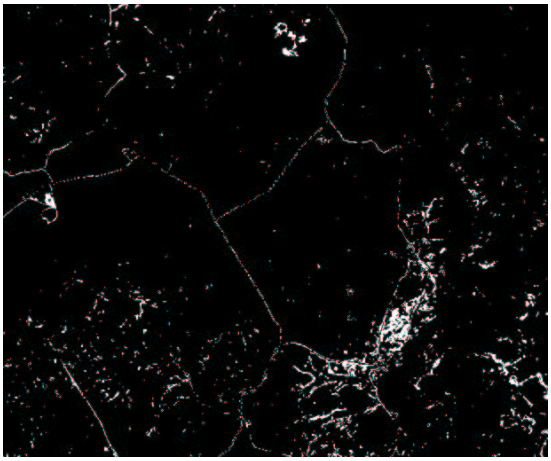


30 Clusters

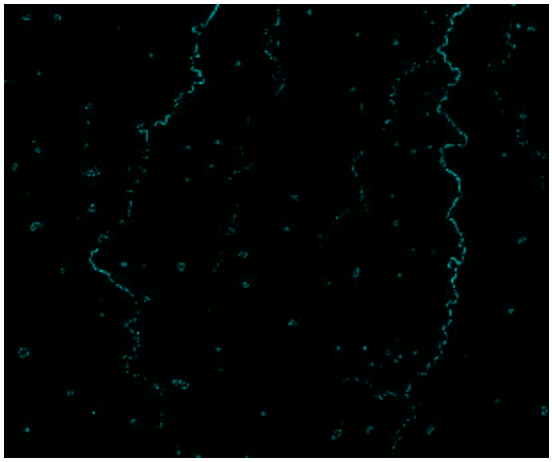
Figure 5.12: Brine scar predictions for 13, 17, 18, and 30 levels of division.



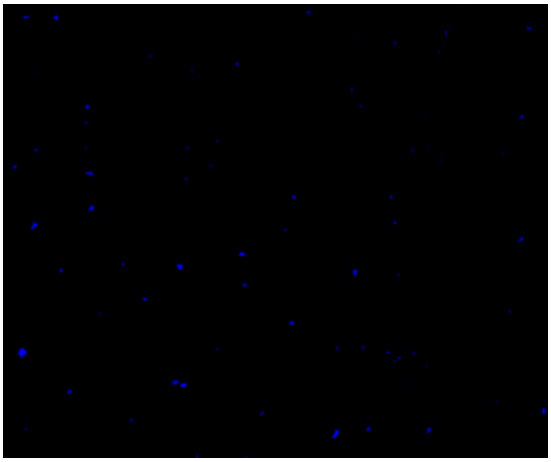
Cluster 11: Main Roads



Cluster 9: Gravel

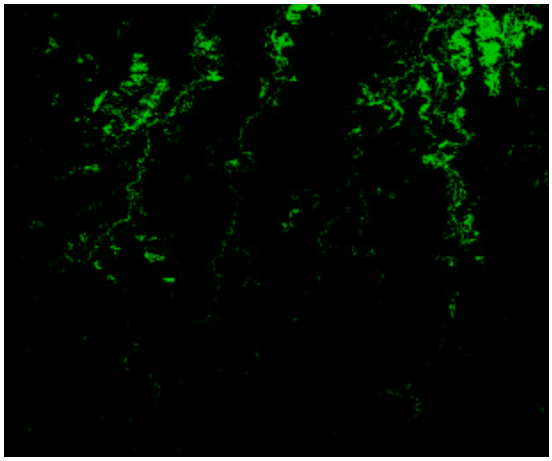


Cluster 18: Shallow Water

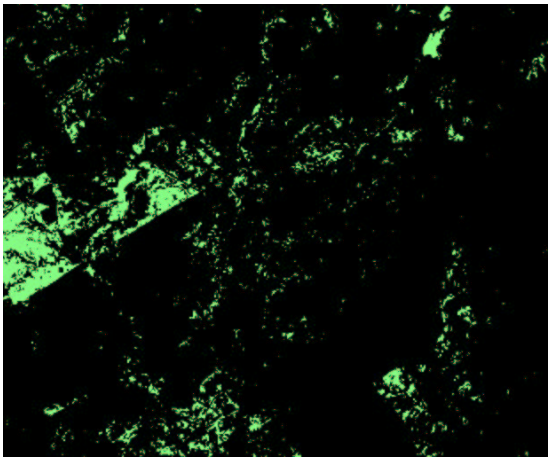


Cluster 5: Deep Water

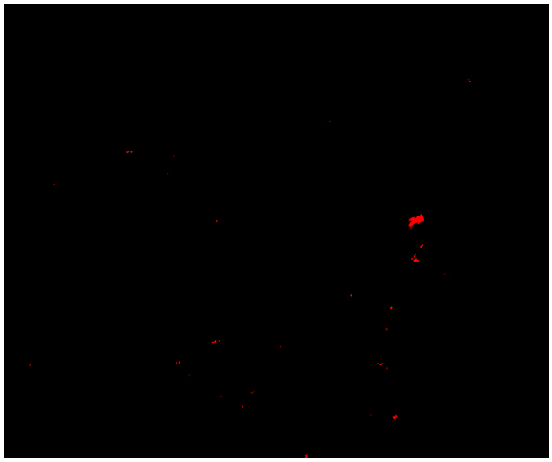
Figure 5.13: Recognizable features in r05sc05 when divided into 18 clusters.



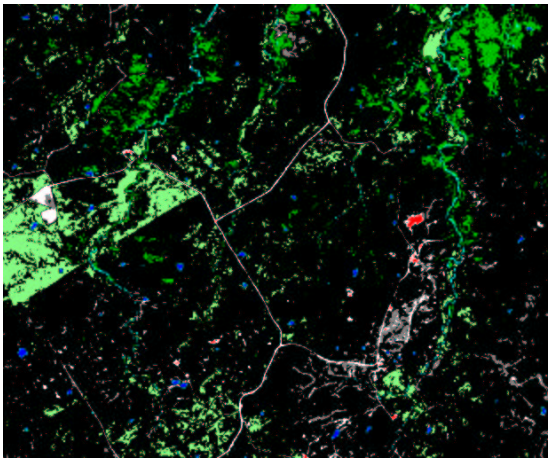
Cluster 3: Woody Vegetation



Cluster 17: Grass A



Cluster 15: Brine Scars



Clusters 11, 9, 18, 5, 3, 17, and 15

Figure 5.14: More recognizable features in r05sc05 when divided into 18 clusters.

Clearly this cluster analysis was successful in extracting surface features from AVIRIS hyperspectral imagery without spectral unmixing or rigorous line-by-line spectral analysis methods.

The second step in extending this statistical model to other scenes was to "classify" the six remaining scenes using the final cluster centroids from the cluster analysis of r05sc05 at the chosen level of division. This classification process amounts to a "single-pass" clustering in which each cell is assigned to the nearest centroid in data space and the process stops immediately after this single iteration of assignment. The centroid locations are not recomputed based on assigned observations. This technique effectively extends knowledge of brine scar spectral signatures from one scene to all others.

Since all the scenes were individually calibrated and had different effective exposures, they have different ranges in data space. As a result, it was not possible to simply use the final centroids from the cluster analysis of r05sc05 as seed centroids for the single-pass classification of the other six scenes. As a part of data standardization, it was necessary to transform each of the six scene data sets to the inference space of the r05sc05 scene or to transform the resulting seed centroids from the r05sc05 cluster analysis to each of the six new inference spaces. Since each scene had to be standardized prior to cluster analysis, it was faster to transform the 18 final unstandardized centroids from the inference space of the r05sc05 scene to the inference space of each of the other six scenes in turn using the means and standard deviations output from each scene standardization process.

The centroids used for the single-pass classification, $\mathbf{c}'_k$, were computed from the unstandardized centroids output by the cluster analysis of r05sc05 with 18 levels of division, $\mathbf{c}_k$, as follows

$$\mathbf{c}'_k = \frac{\mathbf{c}_k - \bar{\mathbf{x}}}{\boldsymbol{\sigma}} \tag{5.1}$$

where $\bar{\mathbf{x}}$ and $\boldsymbol{\sigma}$ are the means and standard deviations calculated for the scene to be classified.

After transforming the centroids and clustering each of the six remaining scenes, new maps were generated showing cohesive geographic clusters across scenes. As shown in Figure 5.15, when all the scenes were classified using transformed centroids and the resulting cluster images were georectified, they fit together seamlessly with a consistent set of random colors across the entire study area. Given the consistent cluster definition, it was now possible to analyze brine scar predictions across all the hyperspectral scenes.

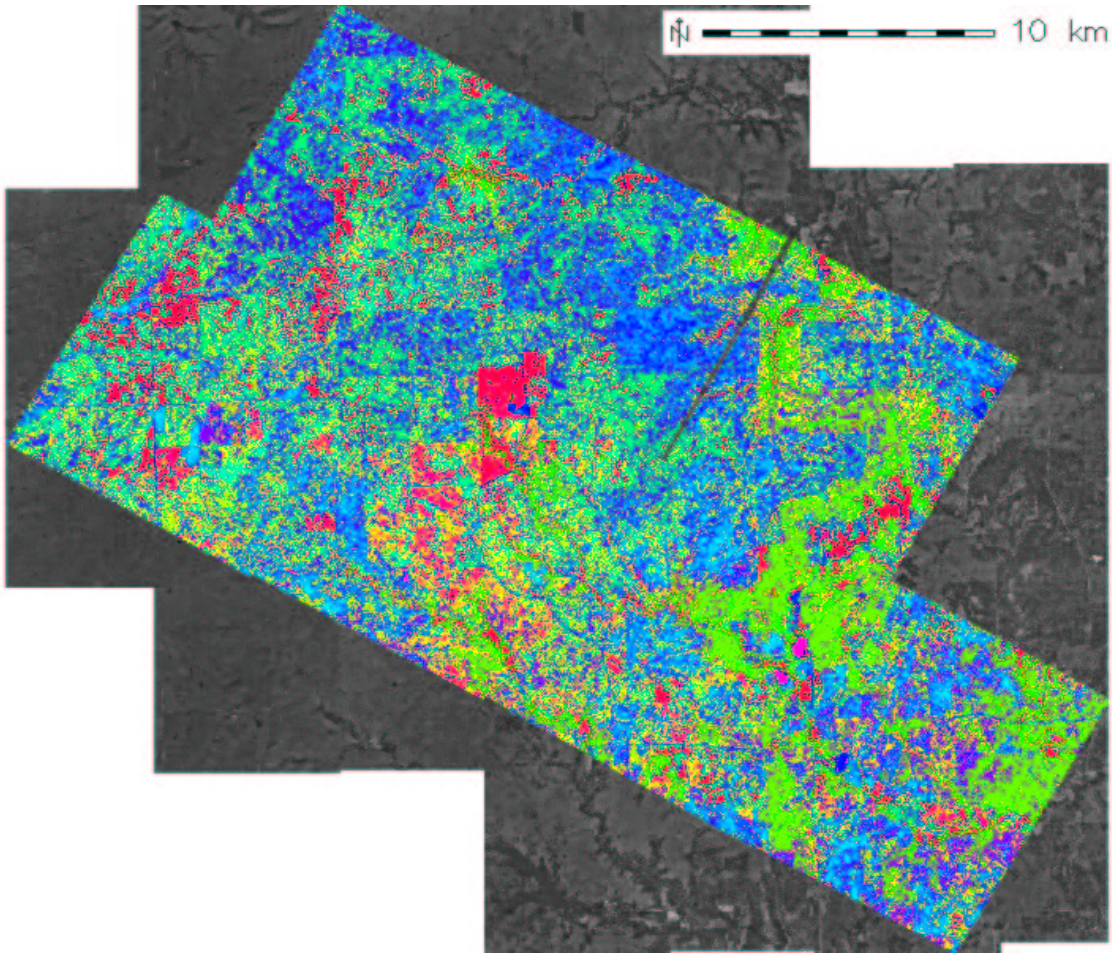


Figure 5.15: All scenes classified using the 18 centroids from r05sc05.

Chapter 6

Results

6.1 Brine Scar Predictions

All of the cells in the seven AVIRIS hyperspectral scenes which were assigned by the single-pass clustering procedure to the cluster which overlies the known brine scars in the cluster analysis of scene r05sc05 were considered predictions of brine scars. Maps of all of the brine scar predictions projected onto the DOQQs from all seven AVIRIS scenes are contained in Appendix A.6. Each of these 87 maps represents a named region in the GIS. The same regions are listed in Table 6.1 along with a short description, the probable cause of the brine scar prediction, the AVIRIS scene in which it is contained, a check (✓) if the region is located within the TPP, the number of brine scar cells (after georectification), and the minimum number of discrete brine spills if the cells are obviously attributable to petroleum activities. The geographic coordinates for the regions are contained in Table A.2.

Of the 87 prediction regions, 40 are attributable to petroleum activities based on objects visible in the DOQQs. Of those 40 petroleum regions, 9 of them are on or within the TPP boundary. That leaves 47 regions which may contain false positives. Due to the differences in time between when the DOQQs were acquired and when the AVIRIS hyperspectral overflights occurred, it is not always possible to see matching features in the DOQQs. Many of these regions may not actually be false positives. The 40 petroleum regions represent at least 65 obviously discrete brine spills; 14 of which are on or within the TPP boundary. The brine scar cluster accounts for approximately 1365 non-overlapping georectified map cells. Of those, 223 are in petroleum regions, and 109 are within the TPP. That leaves some 1142 cells attributable to possible false positives. While that is a large proportion of the total brine scar cells, 845 of those cells are in a single prediction region.

Nearly all of the possible false positives can be explained, and in some cases they should not be considered false positives since they represent similar geophysical disturbances. The region containing the largest contiguous patches of brine scar cells, the 845 cells mentioned above, represents three agricultural fields (see Figure A.30).

Table 6.1: Clustered brine scar predictions.

Region Name	Description	Probable Cause	AVIRIS Scene	In TPP	Cells	Spills
arm ef12 site	TPP bison paddocks and the U.S. Department of Energy's Atmospheric Radiation Measurement (ARM) Extended Facility #12 located near the Tallgrass Prairie Preserve Headquarters. Figure A.1.	Mineral gravel or salt contamination from ice control	r05sc05	✓	1	
tpp hq	Tallgrass Prairie Preserve (TPP) Headquarters. Figure A.2.	Mineral gravel or salt contamination from ice control	r05sc05	✓	5	
bigscar	Two of the largest known brine scars in the entire study area the boundaries of which were recorded using a GPS. Figure A.3.	Multiple brine spills	r05sc05	✓	90	4
lower disturbance	Single brine scar associated with a well near the two large brine scars. Figure A.4.	Brine spill	r05sc05		1	1
bottom tanks	Brine scar near collection of tanks south of two large brine scars. Figure A.5.	Brine spill	r05sc05		1	1
upper well N scar	Medium-sized brine scar just north of the TPP boundary. Figure A.6.	Brine spill	r05sc05		2	1
upper well tank farm	Tanks just north of the TPP boundary. Figure A.7.	Brine spill	r05sc05		1	1
upper well roadside pump	Road-side jackpump just north of TPP boundary. Figure A.8.	Brine spill	r05sc05		1	1
upper well main road	Area near main road but just northwest of an active well just within the TPP boundary. Figure A.9.	Brine spill	r05sc05	✓	1	1
upper well big scar	Old, large brine scar between a number of wells. Figure A.10.	Brine spill	r05sc05	✓	4	1
left upper well scar	Multiple brine scars among a collection of wells and tanks. Figure A.11.	Brine spill	r05sc05 and r05sc06		11	4
tanks	Collection of tanks. Figure A.12.	Brine spill	r05sc05		1	1

Table 6.1: Continued

Region Name	Description	Probable Cause	AVIRIS Scene	In TPP	Cells	Spills
roadside tanks	Road-side tanks just within TPP western boundary. Figure A.13.	Brine spill	r05sc05	✓	4	1
middle disturbance	Old, large brine scar. Figure A.14.	Brine spill	r05sc05	✓	1	1
erosion7	Eroded area with no obvious wells or tanks nearby. Figure A.15.	Unknown	r05sc05		1	
ranch	A ranch with two large buildings and a parking area. Figure A.16.	Metal roofs, mineral gravel, or salt contamination from ice control	r05sc05		8	
ranch2	Big ranch with large driveways. Figure A.17.	Metal roof, mineral gravel, or salt contamination from ice control	r05sc05		5	
intersection5	Highway intersection. Figure A.18.	Road debris or salt contamination from ice control	r05sc05	✓	1	
intersection6	Highway intersection. Figure A.19.	Road debris or salt contamination from ice control	r05sc05	✓	1	
intersection7	Highway intersection (same pixel in both scenes). Figure A.20.	Road debris or salt contamination from ice control	r05sc05 and r06sc05	✓	1	
erosion2	Eroded area on a back road. Figure A.21.	Unknown	r05sc06		7	
intersection2	Intersection of roads. Figure A.22.	Road debris or salt contamination from ice control	r05sc06		5	
intersection8	Highway intersection and eroded area nearby. Figure A.23.	Road debris or salt contamination from ice control	r05sc06		4	
junction2	Highway interchange. Figure A.24.	Road debris or salt contamination from ice control	r05sc06		2	
farm3	Ranch area with large turnaround. Figure A.25.	Mineral gravel or salt contamination from ice control	r05sc06		1	
smallfacility	Road-side facility with parking area. Figure A.26.	Metal roof, mineral gravel, or salt contamination from ice control	r05sc06		3	

Table 6.1: Continued

Region Name	Description	Probable Cause	AVIRIS Scene	In TPP	Cells	Spills
disturb7	Brine scar around a well (same pixel in two scenes). Figure A.27.	Brine spill	r05sc04 and r06sc06	✓	1	1
oilarea	Brine scars at two wells and a tank area. Figure A.28.	Brine spill	r05sc04	✓	6	3
hilltop	Object on riverbank beside roadway near agricultural area. Figure A.29.	Unknown	r05sc04		2	
bigfields	Three very large agricultural fields and farm area. Figure A.30.	Application of mineral fertilizer or other soil amendment	r05sc04 and r05sc03		845	
strip	Gravel road or strip of agricultural field in area with tanks. Figure A.31.	Unkown	r05sc04		5	
farm2	Farm or ranch with large driveway. Figure A.32.	Mineral gravel or salt contamination from ice control	r05sc04	✓	9	
compound3	Building and large field lot near highway and bright shoulders on the highway nearby. Figure A.33.	Unknown	r05sc04	✓	5	
roadside2	Bright areas on highway. Figure A.34.	Unknown	r05sc04	✓	2	
intersection9	Highway intersection. Figure A.35.	Road debris or salt contamination from ice control	r05sc04	✓	1	
road	Bridge and highway turnaround. Figure A.36.	Road debris or salt contamination from ice control	r05sc04	✓	3	
tank cluster	Large tank farm. Figure A.37.	Brine spill	r05sc03		6	1
tank group	A collection of tanks near wells. Figure A.38.	Metal roof	r05sc03		0	
waterfront	Ag field near waterfront. Figure A.39.	Mineral fertilizer or other soil amendment	r05sc03		1	
facility intersection	Some kind of facility at a large intersection. Figure A.40.	Road debris, mineral gravel, salt contamination from ice control, or brine spill	r05sc03		3	

Table 6.1: Continued

Region Name	Description	Probable Cause	AVIRIS Scene	In TPP	Cells	Spills
compound4	Facility situated among various wells. Figure A.41.	Brine spill	r05sc03		2	1
ag intersection	Large roadway intersection. Figure A.42.	Road debris or salt contamination from ice control	r05sc03		3	
farm4	Farm house and barn area. Figure A.43.	Metal roof, mineral gravel, or salt contamination from ice control	r05sc03		1	
disturb1 new	Brine scar around well. Figure A.44.	Brine spill	r05sc03		2	1
disturb2 new	Two obvious brine scars; one near holding tanks and the other at a well site. Figure A.45.	Brine spill	r05sc03		2	2
disturb3	Brine scar on road surrounded by wells. Figure A.46.	Brine spill	r05sc03		1	1
tank station	Tank depot in well area. Figure A.47.	Brine spill	r05sc03		3	1
agriculture	Agricultural field. Figure A.48.	Mineral fertilizer or other soil amendment	r05sc03		1	
agriculture2	Multiple spots on agricultural fields. Figure A.49.	Mineral fertilizer or other soil amendment	r05sc03		57	
erosion1	Bright eroded area near road. Figure A.50.	Unknown	r05sc03		1	
roadshoulder1	Bright road shoulders. Figure A.51.	Mineral gravel, road debris, or salt contamination from ice control	r05sc03		3	
junction	Highway junction and a large building and fenced compound near wells. Figure A.52.	Metal roof, road debris, or salt contamination from ice control	r05sc03		19	
intersection1	Major road intersection. Figure A.53.	Road debris or salt contamination from ice control	r05sc03		1	
farm1	Large ranch. Figure A.54.	Metal roof, mineral gravel, or salt contamination from ice control	r05sc03		11	

Table 6.1: Continued

Region Name	Description	Probable Cause	AVIRIS Scene	In TPP	Cells	Spills
compound2	Fenced-off compound with parking area. Figure A.55.	Mineral gravel or salt contamination from ice control	r05sc03		2	
disturb4	Well site. Figure A.56.	Brine spill	r06sc06	✓	1	1
disturb5	Large area of erosion in wooded area with wells and tanks nearby. Figure A.57.	Brine spill	r06sc06		11	1
disturb6	Well site, tank farm, or new feature. Figure A.58.	Brine spill	r06sc06		1	1
disturb10	Tank or well site near road. Figure A.59.	Brine spill	r06sc06		2	1
wells in 66	Brine scars at many well and tank sites. Figure A.60.	Brine spill	r06sc06		21	14
group of spills	Brine scars at three well/tank sites. Figure A.61.	Brine spill	r06sc06		7	1
ranch3	Ranch with large driveway. Figure A.62.	Mineral gravel or salt contamination from ice control	r06sc06		6	
erosion8	Runoff from pond and nearby well site. Figure A.63.	Brine spill	r06sc06		4	2
runoff	Highly eroded area and pond downstream of well site. Figure A.64.	Brine spill	r06sc05		11	1
intersection10	Road intersection. Figure A.65.	Road debris or salt contamination from ice control	r06sc05		1	
disturb11	Tanks and well site. Figure A.66.	Brine spill	r06sc05 and r06sc04		2	2
disturb12	Spot near well and tank areas. Figure A.67.	Brine spill	r06sc05		3	1
creek to creek spill	Spill from one well drained to and eroded the bank of a secondary stream. Figure A.68.	Brine spill	r06sc05	✓	1	1
tank wash	Runoff from tank which may have fed larger area in region possible runoff. Figure A.69.	Brine spill	r06sc05		2	1
puka	Runoff from possible well (also contained in possible runoff). Figure A.70.	Brine spill	r06sc05		1	1
lower stream	Sandbars, two bridges, and erosion/outcrop spots (also contained in possible runoff). Figure A.71.	Unknown	r06sc05		41	

Table 6.1: Continued

Region Name	Description	Probable Cause	AVIRIS Scene	In TPP	Cells	Spills
disturb13	Large tank or well facility. Figure A.72.	Brine spill	r06sc05		5	1
erosion6	Two bright spots near river. Figure A.73.	Unknown	r06sc05		2	
spill drainage	Drainage from wells or tanks. Figure A.74.	Brine spill	r06sc05		3	1
by river spot	Spot near river. Figure A.75.	Unknown	r06sc05		1	
compound6	Compound with multiple driveways. Figure A.76.	Metal roofs, mineral gravel, or salt contamination from ice control	r06sc05		4	
jackpump1 new	Well area. Figure A.77.	Brine spill	r06sc05		2	1
sandbars new	A bridge, many sandbars, and erosion/outcrops with possible old well site north of river. Figure A.78.	Unknown	r06sc05		59	
erosion3	Drainage area into pond. Figure A.79.	Unknown	r06sc04		3	
erosion4	Erosion or drainage near back road. Figure A.80.	Unknown	r06sc04		1	
erosion5	Erosion area near back road intersection. Figure A.81.	Unknown	r06sc04		1	1
erosion9	Area near small shack or tank. Figure A.82.	Brine spill	r06sc04		1	1
intersection3	Long building at multi-way intersection. Figure A.83.	Metal roof, road debris, mineral gravel, or salt contamination from ice control	r06sc04		1	
intersection4	Major roadway intersection. Figure A.84.	Road debris or salt contamination from ice control	r06sc04		1	
bridge and farm river	Bridge over river with sandbars and ranch with large driveways. Figure A.85.	Road debris, metal roof, mineral gravel, or salt contamination from ice control	r06sc04		2	
two spills	Two obvious brine scars around wells. Figure A.86.	Brine spill	r06sc04		2	2
set of tanks	Cluster of holding tanks. Figure A.87.	Brine spill	r06sc04		1	1

Kerry Sublette, professor of Chemical Engineering at the University of Tulsa, speculated that these fields may have been fertilized just prior to acquisition of the AVIRIS hyperspectral imagery[1]. It is reasonable to assume that mineral fertilizer might have a similar spectral signature to mineral produced water on bare soil. Additional possible explanations include saline-seep, which can occur when vegetation is removed and groundwater in the area subsequently rises, or irrigation with high salinity water. In areas with a near-surface water table, like these fields which are near a river, high salinity conditions can build up each time vegetation is suddenly removed [8].

In addition to features which are very likely brine scars, other surface features which occasionally appear as brine scar predictions include portions of some gravel driveways, major road intersections where dirt and gravel form piles, a few road shoulders, some metal rooftops, large sand bars in rivers, and heavily eroded areas which may or may not be attributable to petroleum activities. Examples of many such predictions are contained in Appendix A.6. Metal rooftops are real false positives; however, many of the other surface features may be included as brine scar predictions because they represent similar, often anthropogenic, geophysical disturbances. Batteries of tanks used for petroleum activities are not considered false positives, but in cases where large white tanks are clustered together, brine scar predictions may actually be the result of their high spectral reflectance. It would be important to perform chemical analyses around these tanks to disprove brine disturbance predictions.

It is speculated that some gravels may have a mineral composition consistent with that of the produced water extracted in the region. Moreover, some driveways and parking lots are known to be salted to control ice in the winter, and residual salt contamination is likely to have a spectral signature very similar to actual brine scars. Some road intersections probably appear as brine scars because of the large piles of rocks, dirt, dust, and minerals which accumulate in such urban intersections. These areas, along with road shoulders, may also be contaminated by chemicals applied to control salt or from produced water spilled from trucks during transport. Sand bars on river banks probably collect mineral crusts, including those resulting from spilled produced water, salt used for ice control on bridges, and fertilizers which run off into creeks and streams in the area.

Given these reasonable explanations for many of the brine scar predictions, one may not consider many of them to be false positives. It is easy to use the DOQQs to identify the probable cause for most brine scar predictions. Without visiting these locations and performing chemical analyses of rocks and soils, it is not possible to absolutely discount clustered brine scar predictions. While permission could likely be obtained from the Nature Conservancy to sample locations on the TPP, the surrounding land is private property and getting such permission in those areas could be difficult. Collecting data on the ground is outside the scope of this study.

Nevertheless, based on visual verification using the DOQQs, 65 or more brine scars associated with wells or tanks were successfully detected using cluster analysis. Some

[1]Kerry Sublette, personal communication.

of these spill areas were considerably smaller than the size of a single AVIRIS pixel, yet the hyperspectral data captured enough reflected radiation for the analysis to detect these disturbances. In many cases, only a single well or tank location would be associated with a brine scar prediction even though it was surrounded by many other wells or tanks which appear in the DOQQs to be the same. This is not a surprising result since produced water is often piped from neighboring wells to a single collection of tanks for separation and storage or to a single well for re-injection into the ground.

6.2 Cluster Spectral Signatures

Since the centroids resulting from cluster analysis represent the average combination of conditions (or spectra) of their membership, they can be viewed as representative spectra. Figure 6.1 shows spectral traces or signatures for each of the 18 clusters derived from the r05sc05 scene which were used in classifying the other six hyperspectral scenes. The curves are discontinuous because the hyperspectral data were subset, as described above, reducing the number of bands from 224 to 187. The curves for clusters 3, 5, 9, 11, 15, 17, and 18 match the images presented in Figures 5.13 and 5.14.

As one would expect, cluster 5 (deep water) has the lowest intensity throughout the near infrared. Cluster 18 (shallow water) is the next darkest in the near infrared, and cluster 3 (woody or shrubby vegetation) is the third darkest in the near infrared. This shrubby vegetation also exhibits the strongest chlorophyll bump, *i.e.*, it was very green in mid-July. Since this form of vegetation tends to grow primarily along streams, creeks, and around other water sources, as seen in Figure 5.14, it is more likely to remain green and well-hydrated throughout the summer months. It may also be spectrally mixed with shallow water.

The brine scar cluster (cluster 15) has the strongest intensity in the visible and the long near infrared. This unique spectral signature explains why the brine scar cluster is so stable across many levels of division. Moreover, only the brine scar cluster exhibits no significant red edge effect from vegetation. The next brightest surface features in the visible and the long near infrared are main roads (cluster 11) and gravel (cluster 9). This explains why some of the road intersections and bare ground areas appear in what would otherwise be a brine scar cluster at low levels of division; they have similar spectral signatures.

Interestingly, cluster 17 (grass A) has the strongest intensity in the short near infrared. This is a very strong red edge effect which uniquely identifies an individual grass species or life stage. It may be possible to distinguish vegetation types using hyperspectral imagery, even at the relatively low spatial resolution of AVIRIS.

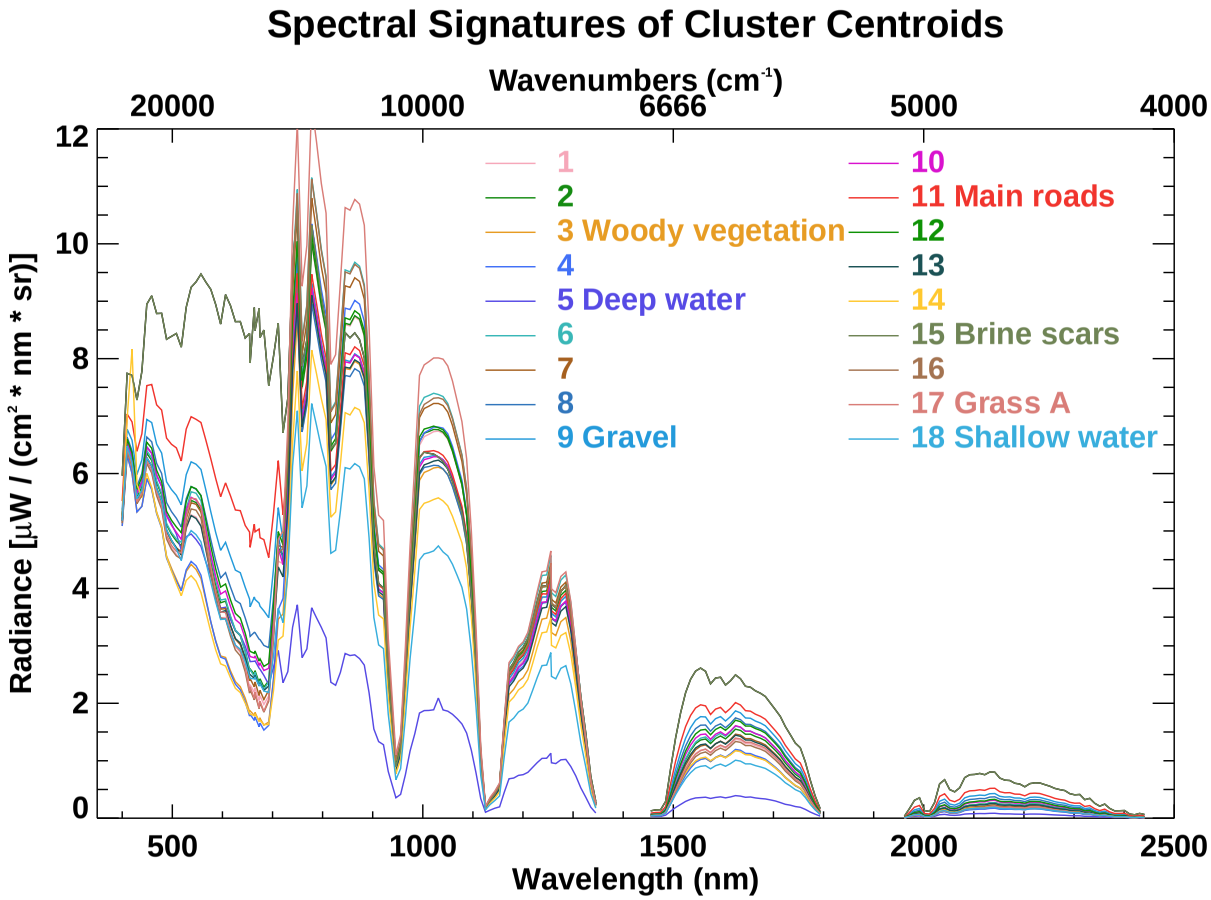


Figure 6.1: Spectra for the 18 cluster centroids. The curves for clusters 3, 5, 9, 11, 15, 17, and 18 match the images presented in Figures 5.13 and 5.14.

6.3 Brine Scar Representativeness

After a spectral signature for brine scars has been quantitatively determined using cluster analysis, it is useful to investigate the degree to which predictions are represented by this signature. This representativeness measure provides an indication of the similarity between a brine scar prediction and the mean of the collection of all brine scar predictions. It is calculated as the Euclidean distance from each observation or brine scar cell to the brine scar centroid in data space. Based on the radius definitions in Chapter 3, the brine scar prediction cell most closely represented by the brine scar centroid will be nearest to the centroid in data space, and that distance will be the "inner radius" of the brine scar cluster. The prediction cell least-well represented by the brine scar centroid will be the farthest from the centroid in data space, and its position will determine the "outer radius" of the brine scar cluster.

Maps were produced showing the same brine scar predictions as in Appendix A.6, but with each georectified prediction cell colored according to how well it was represented by the brine scar cluster centroid. These representativeness colors vary from green (highly representative and short data space distance) to yellow to red to white (poorly representative and large data space distance). Inset on the maps included below are randomly colored spectral traces of each of the original AVIRIS image pixels from the same brine scar predictions. The number of map cells is often different from the number of AVIRIS pixels (and therefore spectral curves) because georectification warps the image onto the DOQQs. Nevertheless, these spectral curves show the scatter characteristic of patches of prediction map cells which are mapped in representativeness colors. Each of the spectral plots includes the spectral signature of the brine scar cluster centroid in black for comparison. For maps of the DOQQs in each region without the overlying brine predictions, see the corresponding figures in Appendix A.6.

As expected for the two well-known brine scars, their spectra match that of the brine scar cluster centroid very well. Figure 6.2 shows both of these brine scars, along with the third scar in that region bordering the old railroad bed, as green to green-yellow map cells. For all three scars, the largest variance is in the visible and short near infrared. The scar on the right (Site 5) has a few rather bright pixels in the visible. In the long near infrared, all three scars have a similar spread around the centroid curve. Most pixels are slightly stronger than the centroid in the long near infrared.

Figure 6.3 is a photograph of another well-known brine scar on the TPP. The map of this scar is shown in representativeness colors in Figure 6.4. In this region, all the brine scar map cells are green, reflecting the strong similarity between the spectral signature of the pixels and the cluster centroid representing brine scars. The spectral plot details this similarity for each pixel. While there is some spread in the visible and short near infrared, there is nearly complete overlap in the long near infrared. Similarly, the scar shown in Figure 6.5 is well represented by the brine scar cluster centroid; however, its intensity in the visible is mostly below that of the centroid.

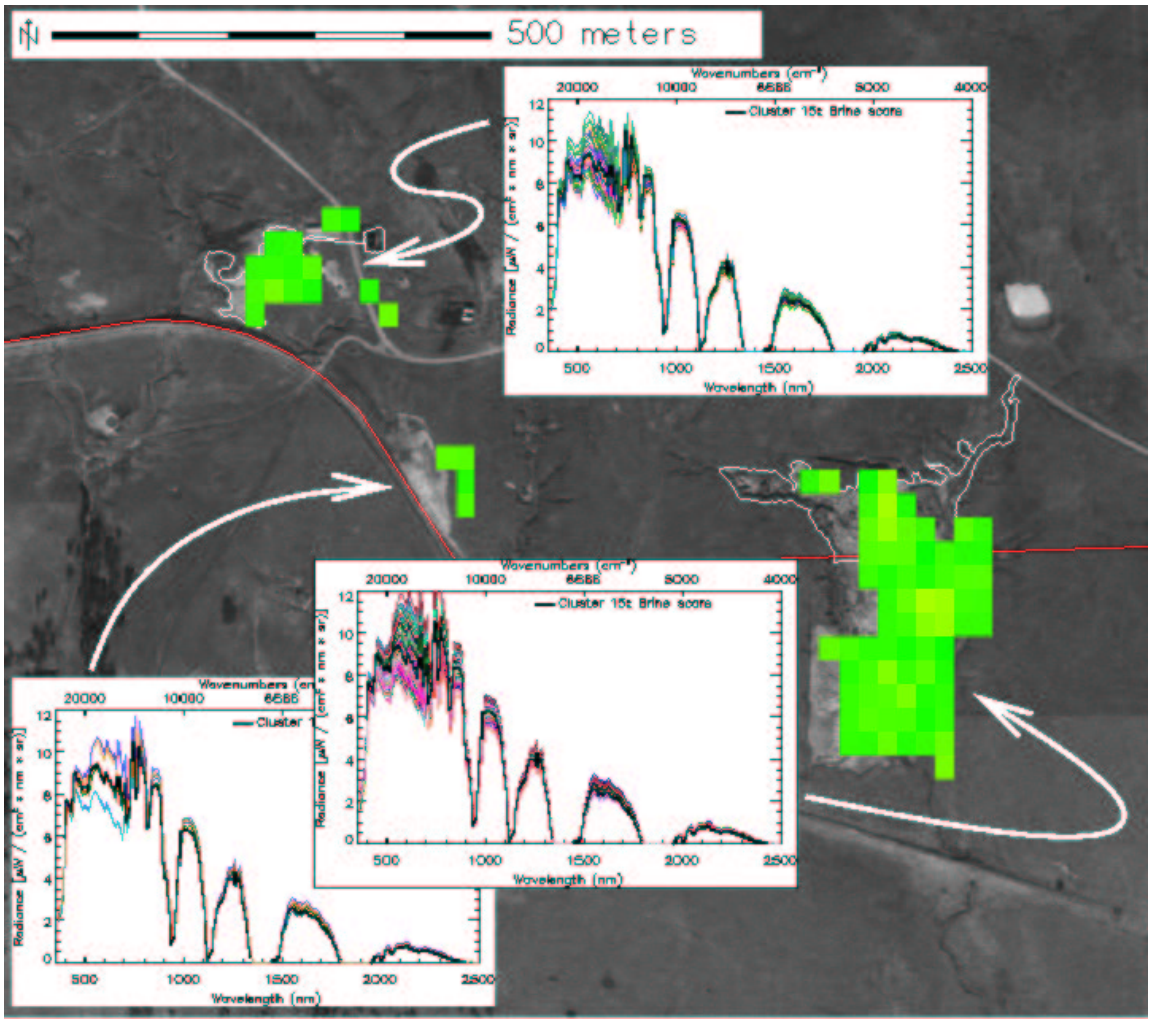


Figure 6.2: The two large, well-documented brine scars shown in representativeness colors. All three brine scars include spectra comparing each original AVIRIS pixel against the brine scar cluster centroid. Region: bigscar, Scene: r05sc05. Corresponds to Figure A.3.

Figure 6.3: Photo of another well-known brine scar on the TPP. Photo courtesy of Tina Carlsen (June 13, 2001).

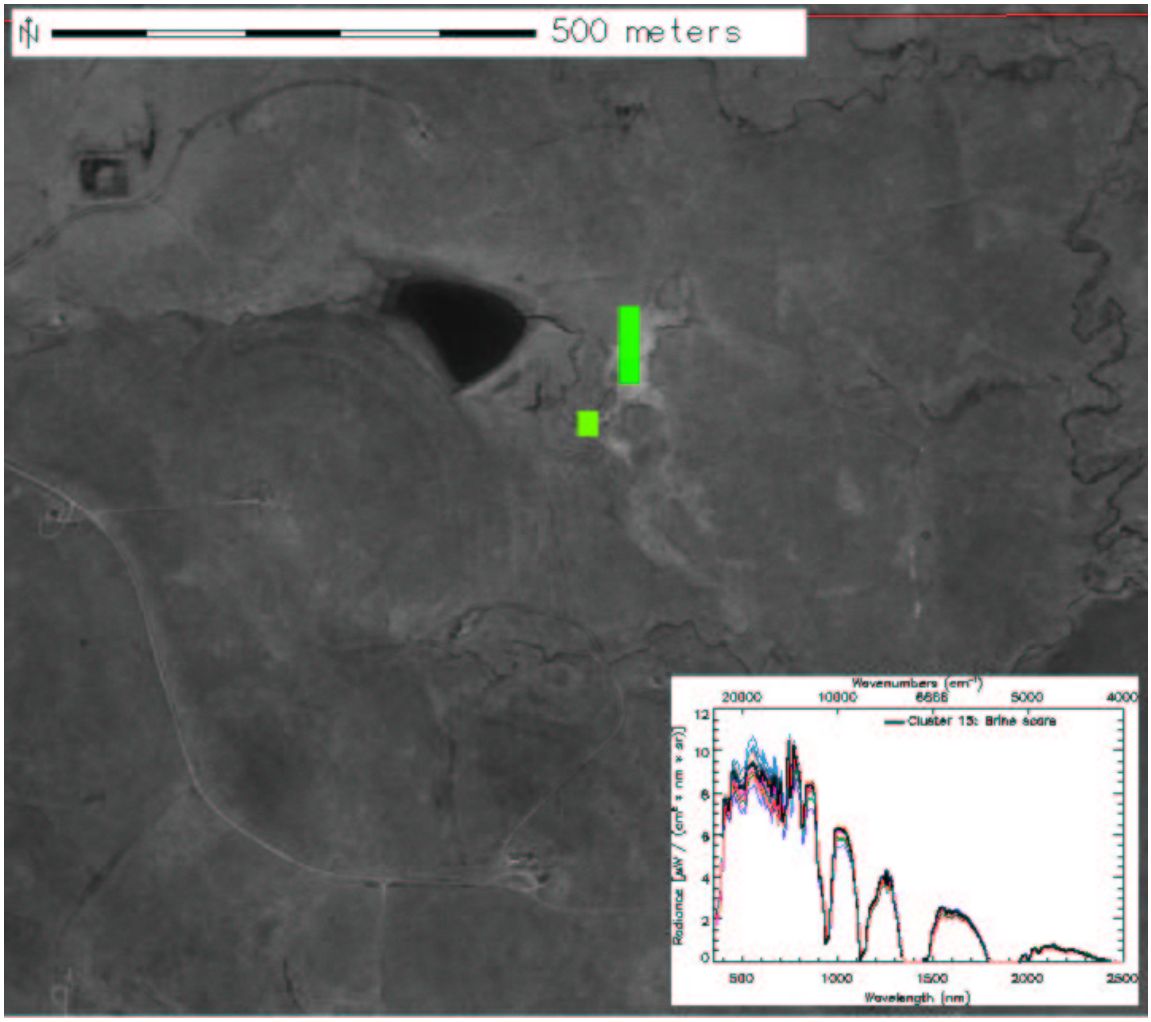


Figure 6.4: Another well-known brine scar on the TPP shown in representativeness colors. Region: upper_well_big_scar, Scene: r05sc05. Corresponds to Figure A.10.

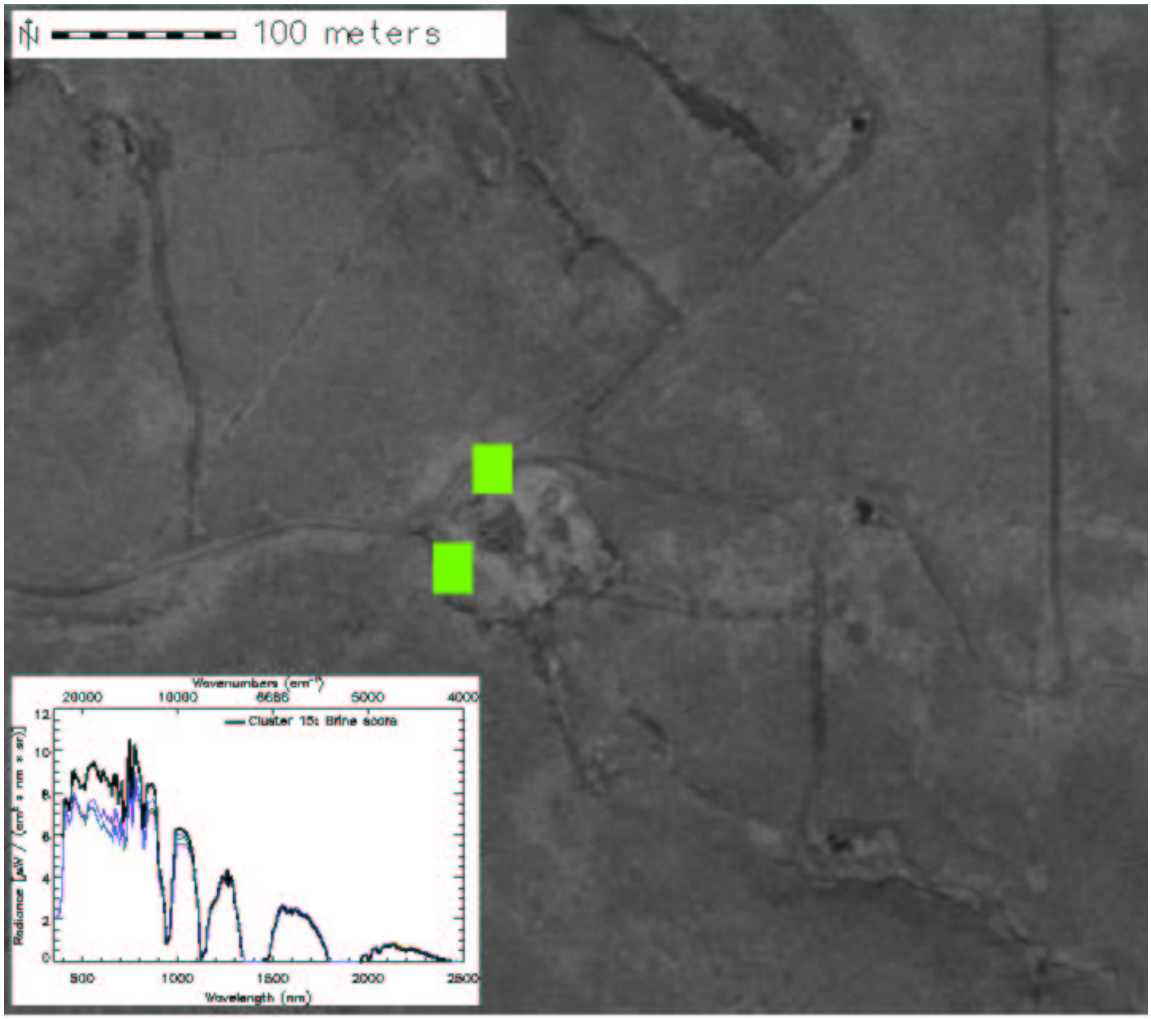


Figure 6.5: A brine scar outside the TPP shown in representativeness colors. Region: upper_well_N_scar, Scene: r05sc05. Corresponds to Figure A.6.

Figure 6.6 shows a representativeness map for the large agricultural fields thought to have been freshly tilled or fertilized prior to the AVIRIS overflight. While some cells are a reasonably close match to the brine scar cluster centroid (green), many of them deviate enough to appear yellow to orange. In the spectral plot, the original pixels appear to have a large spread around the brine scar centroid in the visible, and the intensity in the near infrared for many pixels is significantly stronger. The pixel or pixels corresponding to the cell over the farm buildings on the right of the figure is not included in the spectral plot. The intensity in the long near infrared and the short near infrared around 1 μm is uniquely strong. This may be an indicator of the condition of these fields or the soil amendments which may have been recently applied.

The TPP Headquarters are shown in Figure 6.7 using representativeness colors. The cells are very close to green, so they are well represented by the brine scar cluster centroid even though they probably are not actually the result of spilled produced water. This suggests that gravel in the parking area has a brine-like spectral signature or that residual contamination from repeated application of salt or other chemicals used to control ice on parking and walking surfaces in winter shows up strongly even in the summer.

Figure 6.8 shows a ranch with two parallel barns and a parking area in between. Large animal pens can be seen just south of the barns. While the parking area appears as two green cells (closely matching the brine scar cluster centroid), the south barn contains a green cell, a yellow cell, and an orange cell indicating a significant deviation from the brine scar cluster centroid. The north barn contains a pink cell which, in this color scheme, indicates an even larger deviation from the centroid. It is likely that these rooftops are painted a bright color and are reflecting significant amounts of visible radiation and reasonably significant amounts of near infrared radiation which appear similar to the brine scar spectral signature. While these structures are understandably generating false positive brine scar predictions, they are easy to discount based on visual inspection of the DOQQs. The spectral signatures and representativeness colors also provide clues that some of these cells are poor matches with the brine scar cluster centroid to which they were assigned in the MGC procedure. In any case, the parking area shows up probably because of the gravel mineral content or because of salt contamination.

Figure 6.9 shows a group of buildings and driveways in representativeness colors. The white cell is the result of direct reflection of solar radiation from a metal roof. This cell deviates the farthest of any cell assigned to the brine scar cluster, and it defines the "outer radius" of that cluster. The corresponding pixels in the AVIRIS imagery has the highest intensity of any other pixel across the entire spectrum. This roof is probably unpainted, and it acts nearly as a perfect reflector. The relative positions of the sun and the aircraft were such that significant solar radiation was directed toward the AVIRIS instrument. The spectral trace of these pixels, shown as a cyan and green curves, are off the scale in the visible and very strong throughout

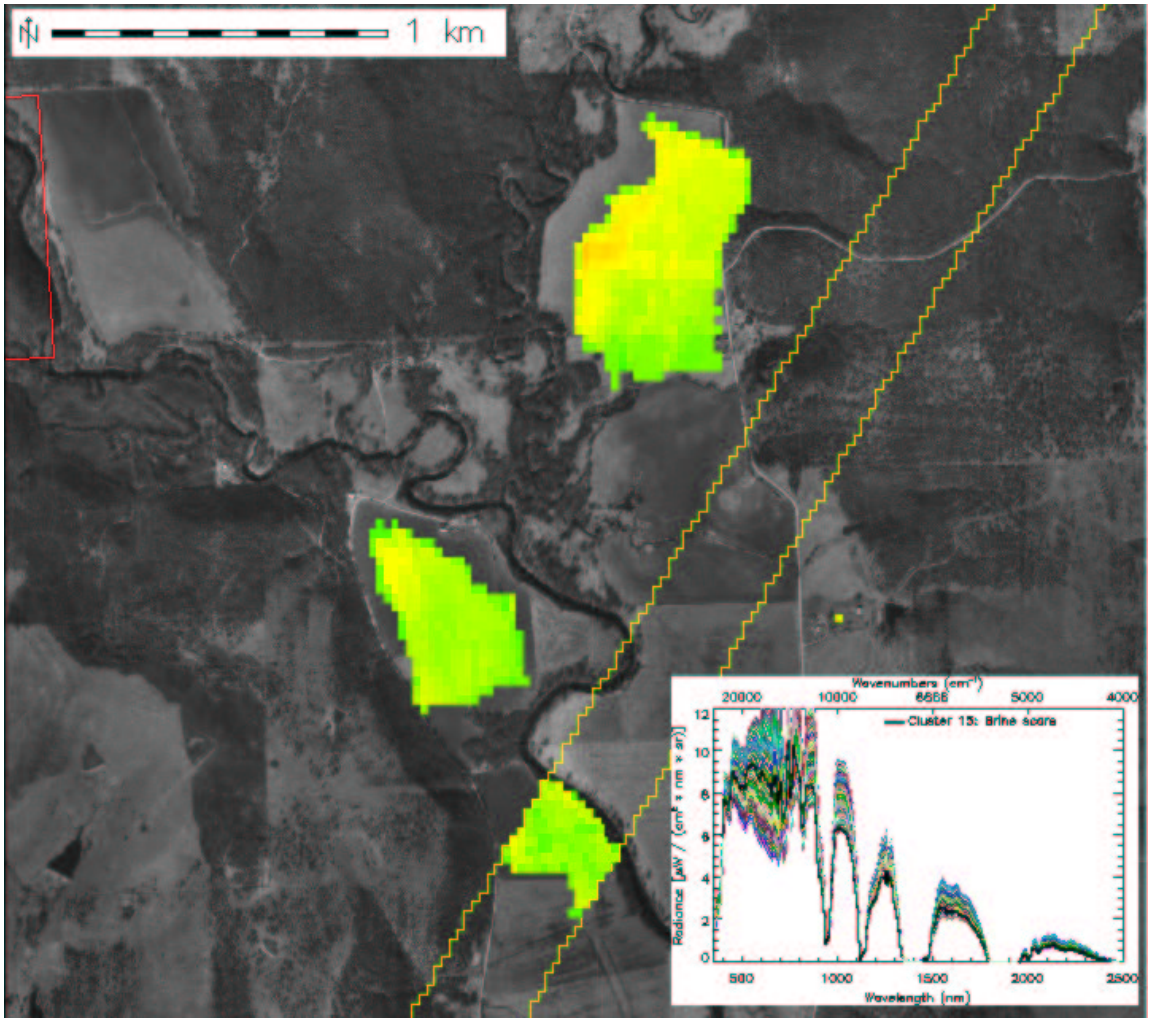


Figure 6.6: The large agricultural fields shown in representativeness colors. Region: bigfields, Scene: r05sc04 and r05sc03. Corresponds to Figure A.30.

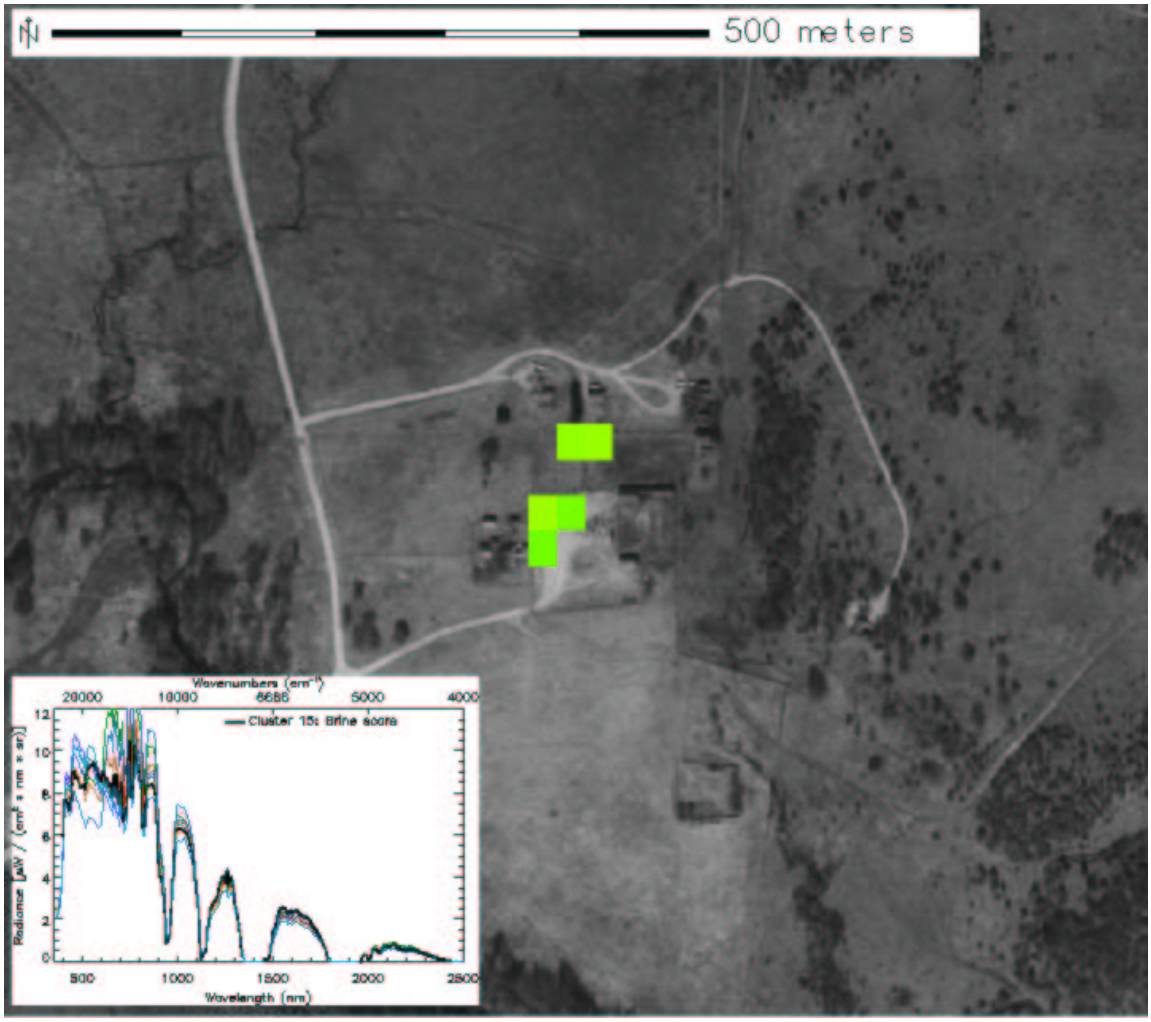


Figure 6.7: The TPP Headquarters shown in representativeness colors. Region: tpp_hq, Scene: r05sc05. Corresponds to Figure A.2.

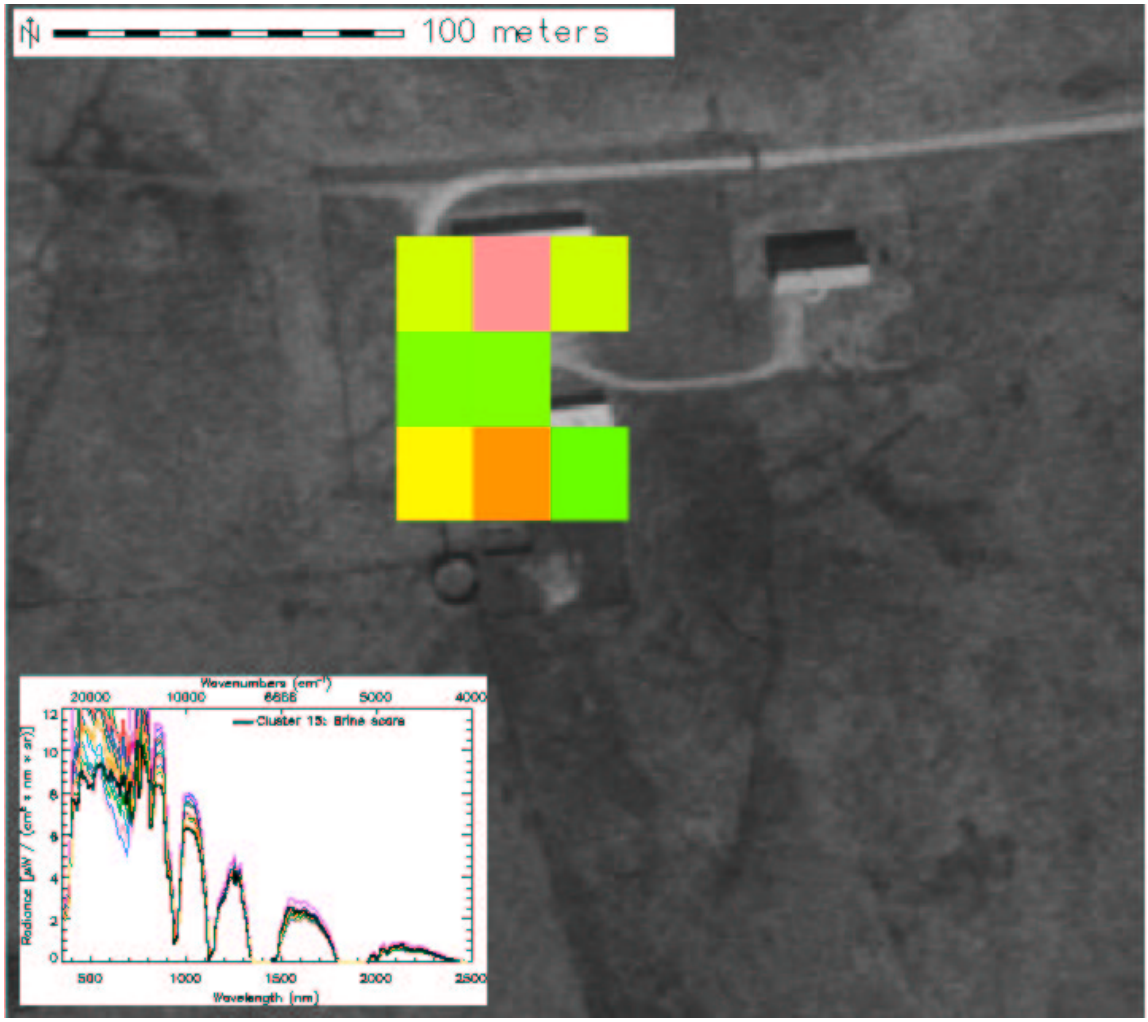


Figure 6.8: Ranch buildings and parking area shown in representativeness colors. Region: ranch, Scene: r05sc05. Corresponds to Figure A.16.

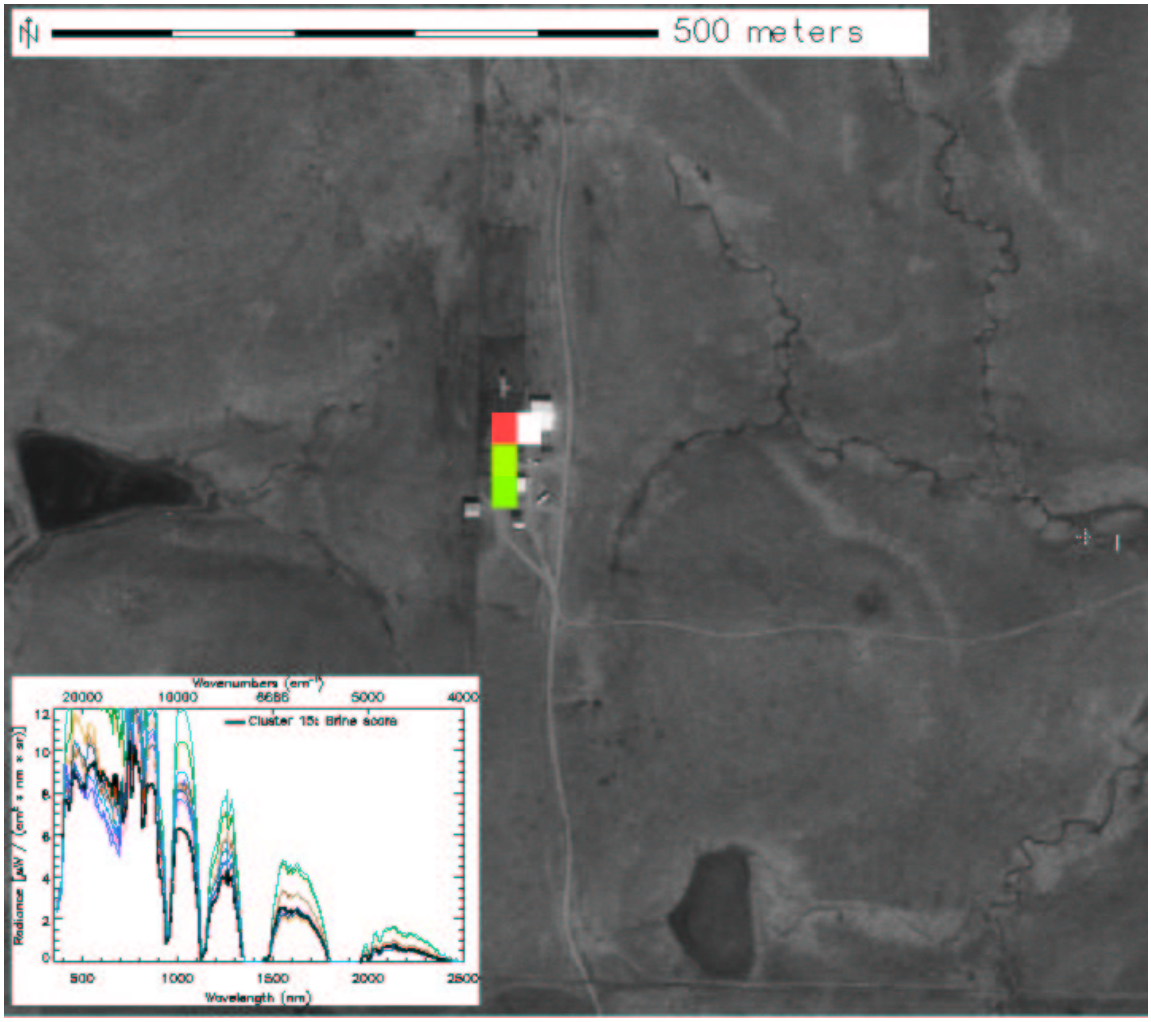


Figure 6.9: A group of buildings and driveways shown in representativeness colors. Region: compound6, Scene: r06sc05. Corresponds to Figure A.76.

the near infrared. The cyan spectral trace approximates the exact solar spectrum without the atmosphere removed. The red cell probably represents a mixture of the edge of the roof and nearby vegetation since it exhibits very strong visible intensity and a strong red edge effect. The green cells are much more closely represented by the brine scar cluster centroid spectrum. Their spectral curves indicate they are likely the driveway in front of the middle building and that vegetation is mixed in the same pixel because of the strong red edge effect.

Figure 6.10 shows one of the brine scar prediction regions for which no corresponding surface feature appears in the DOQQs. Shown in representativeness colors, it appears the prediction cells have a measurable deviation from the brine scar cluster centroid. From the spectral plot, it appears that the object, which must have appeared at this location between the time the DOQQs were taken and the time of the AVIRIS overflight, reflects strongly in the visible, but corresponds well to the brine scar cluster centroid in the long near infrared. Based on the similarity of the cells and spectra with that of the buildings contained in Figure 6.8, it is hypothesized that a building with a painted metal roof was constructed at this location between 1995 and 1999.

6.4 Complete Convergence

Since the MGC algorithm is frequently applied to the analysis of extremely large data sets, it uses a convergence criterion of 0.5%. That is, the procedure stops once less than 0.5% of the observations change cluster assignment between iterations. This avoids time-consuming and expensive computation spent fine tuning the statistical fit for an insignificant number of observations. This "close is close enough" philosophy is not necessary for smaller data sets like the hyperspectral scenes. To test the effect of this stopping rule on the brine scar predictions, a second cluster analysis of scene r05sc05 was started using the 18 centroids from the original analysis, which had converged to 0.5% in 34 iterations, as seed centroids. This second analysis was run until no observations changed cluster assignment from one iteration to the next. This complete convergence of the fit, which ran for 23 iterations, resulted in two cells being added to the brine scar cluster in scene r05sc05. Those two cells are shown in yellow in the two regions in Figure 6.11. The cells are merely edge cells on brine scars already predicted by the original analysis. In this case, close was close enough.

6.5 Four Band Subset Analysis

Another objective of this study was to determine if a significantly smaller subset of bands could be used, instead of all 224 or 187 bands, to find brine scars with similar skill. Based on inspection of the spectra of the 18 cluster centroids, four bands were chosen to serve as characteristics for new cluster analyses of the r05sc05 scene.

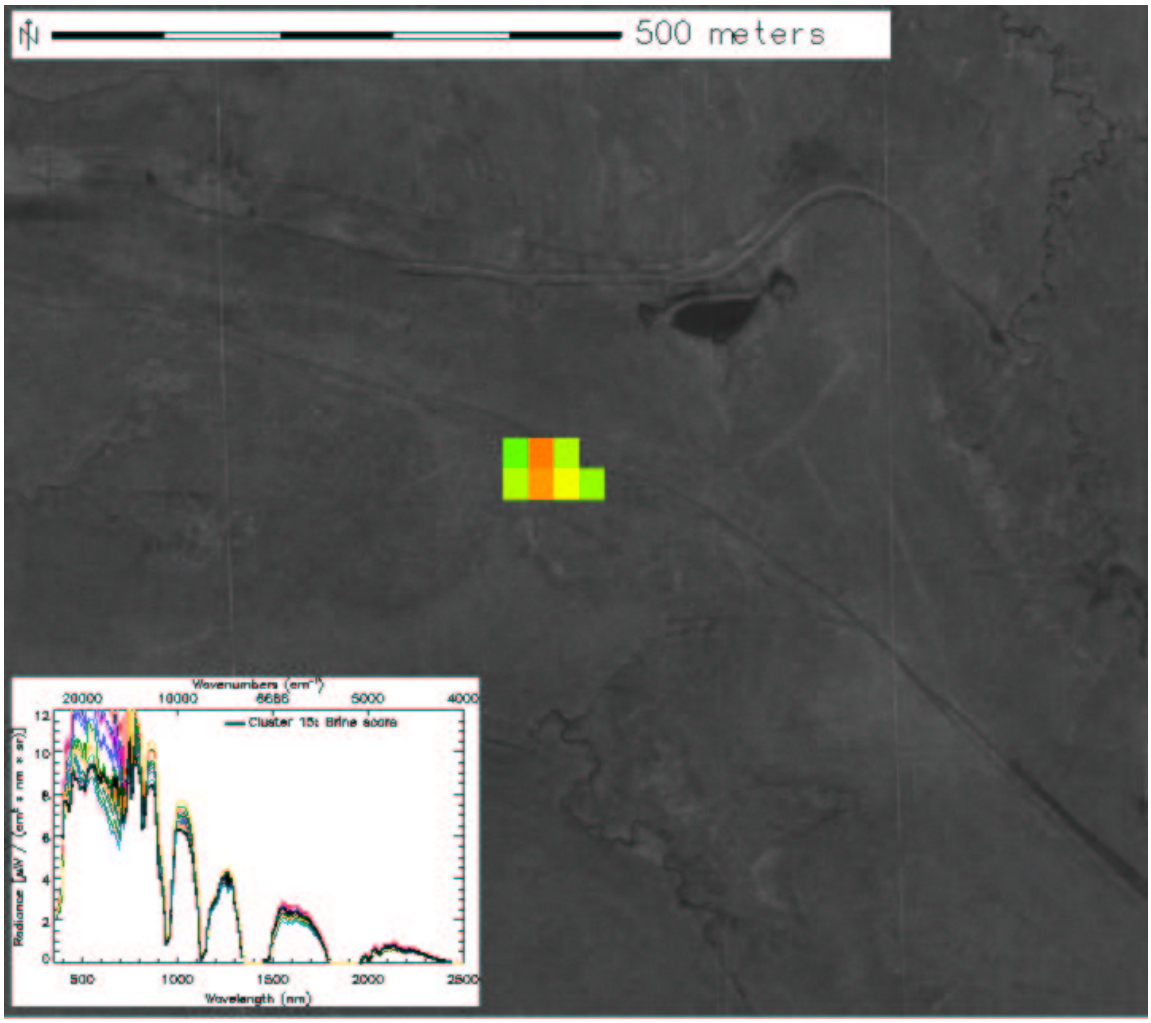
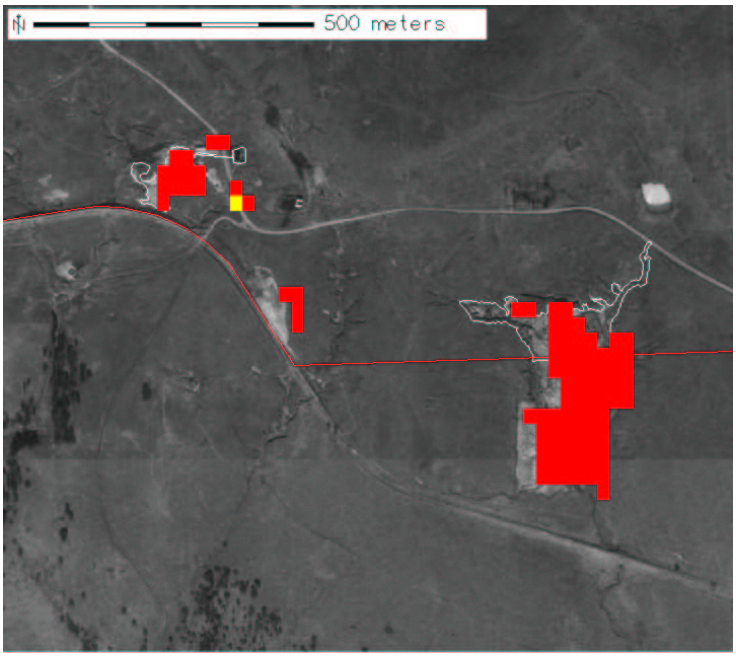
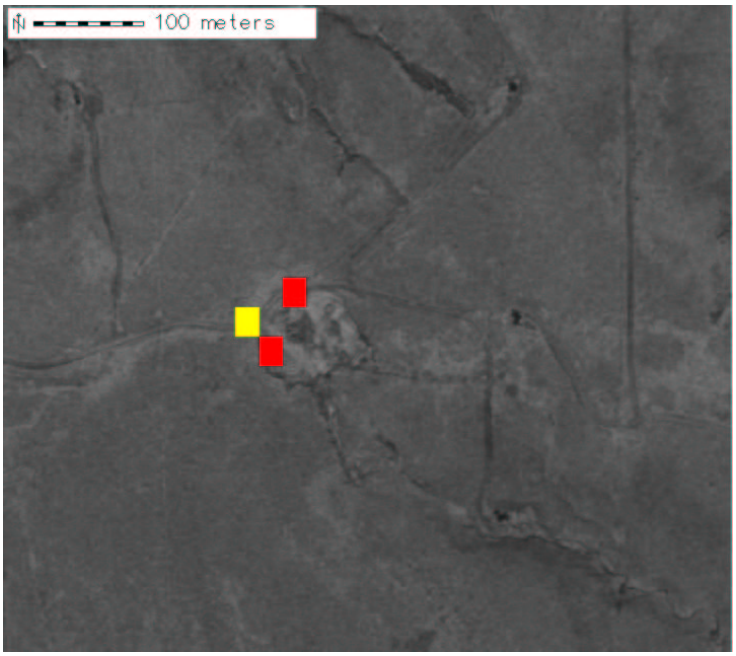


Figure 6.10: A brine scar prediction with no corresponding surface feature in the DOQQs shown in representativeness colors. Region: erosion2, Scene: r05sc06. Corresponds to Figure A.21.



Region: bigscar, Scene: r05sc05.
Corresponds to Figure A.3.



Region: upper_well_N_scar, Scene: r05sc05.
Corresponds to Figure A.6.

Figure 6.11: Two cells added to the brine scar cluster after complete convergence. The cells added are shown in yellow.

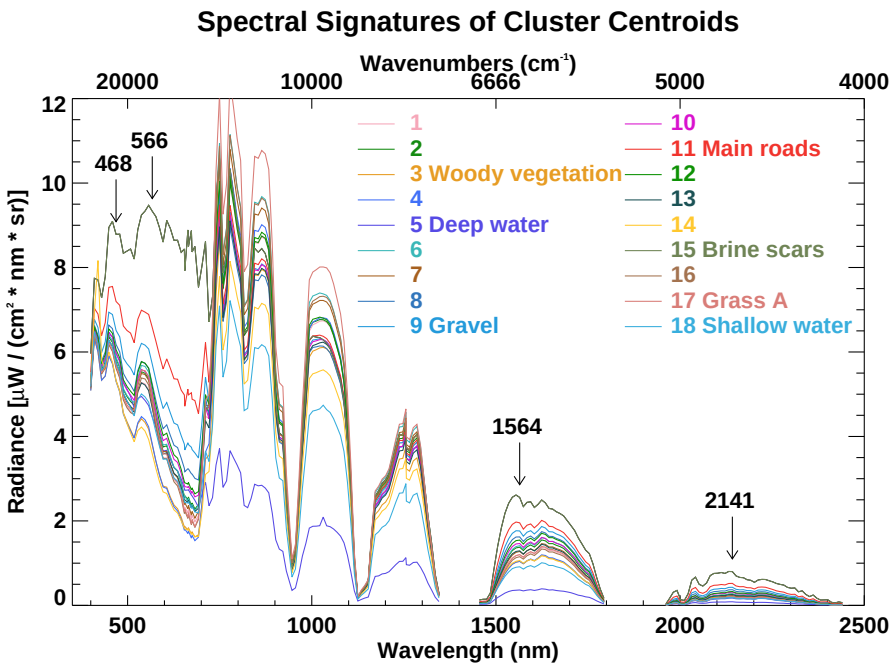


Figure 6.12: The four wavelengths chosen for a second subset superimposed on the 18 original cluster centroids.

The four selected bands—at 468, 566, 1564, and 2141 nm—were chosen because the intensity of the brine scar cluster centroid at those wavelengths was significantly different from that of all other centroids at 18 levels of division. Figure 6.12 shows the locations and intensities of these bands for the brine scar cluster. The same four bands are denoted in Table A.1 by a check (✓) in the seventh column. Many other combinations of bands are possible, and it is recognized that this choice may not be optimal. Nevertheless, these seem to be reasonable choices for discriminating brine scars from other surface features. This much smaller data subset was clustered at every level of division between 10 and 32. These analyses were not seeded with a four band subset of the resulting centroids from the original 187 band analysis. Table 6.2 shows the map statistics for levels of division 10–28 as compared with the original cluster analysis result at a level of division of 18 (left-most column).

The four band subset at a level of division of 18 has 111 cells in common with the original cluster analysis which used 187 bands. The original analysis contained 24 additional cells in its brine scar cluster; however, the 18 cluster four band subset did not include in its brine scar cluster any cells which were not included in the predictions of the original analysis. A better match to the original results can be found at a level of division of 12. At this level of division, the four band subset has 120 cells in common with the original analysis which used 187 bands. The original

Table 6.2: Map statistics for 10–28 clusters in the four band subset. Comparison can be made with the original 187 band analysis with 18 clusters (first column).

18	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**Cells**
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	76
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	6
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓		12
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓			✓				✓		1
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓							✓		2
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓									✓		8
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓											6
✓	✓	✓	✓	✓	✓	✓	✓													6
✓	✓	✓	✓	✓	✓		✓													1
✓	✓	✓	✓	✓	✓															1
✓	✓	✓	✓																	1
✓	✓	✓																		7
✓																				8
	✓	✓	✓	✓	✓	✓	✓													1
	✓	✓	✓	✓	✓		✓													2
	✓	✓																		16
																				314214

analysis has 15 cells which were not included in the brine scar predictions of the four band analysis, and the four band analysis has three cells which were not included in the original predictions.

Selected maps comparing brine scar predictions from the two different hyperspectral subsets (187 bands versus 4 bands) are included in Figures 6.13 and 6.14. In these figures, cells predicted as brine scars by both analyses are colored green, cells predicted as brine scars only by the four band analysis are colored yellow, and cells predicted as brine scars only by the original analysis are colored red. While some cells are missed in the four band analysis as compared with the 187 band analysis, 88% of the cell predictions are the same and 90% of the same prediction regions contain brine scar predictions using the much smaller data set. Two prediction regions, arm_ef12_site (Figure A.1) and bottom_tanks (Figure A.5) were not included in the predictions from the four band analysis. Two of the three cells which were added as brine scar predictions in the four band analysis are shown in Figure 6.14. These cells are connected to objects found by both analyses. The third cell added in the four band analysis is shown in Figure 6.15. This cell in r05sc05 was not predicted as a brine scar in the original 187 band analysis. However, it overlaps with other scenes, and it was predicted as a brine scar in scene r06sc04 in the 187 band analysis. It corresponds to the brine scar prediction on the left of Figure A.66.

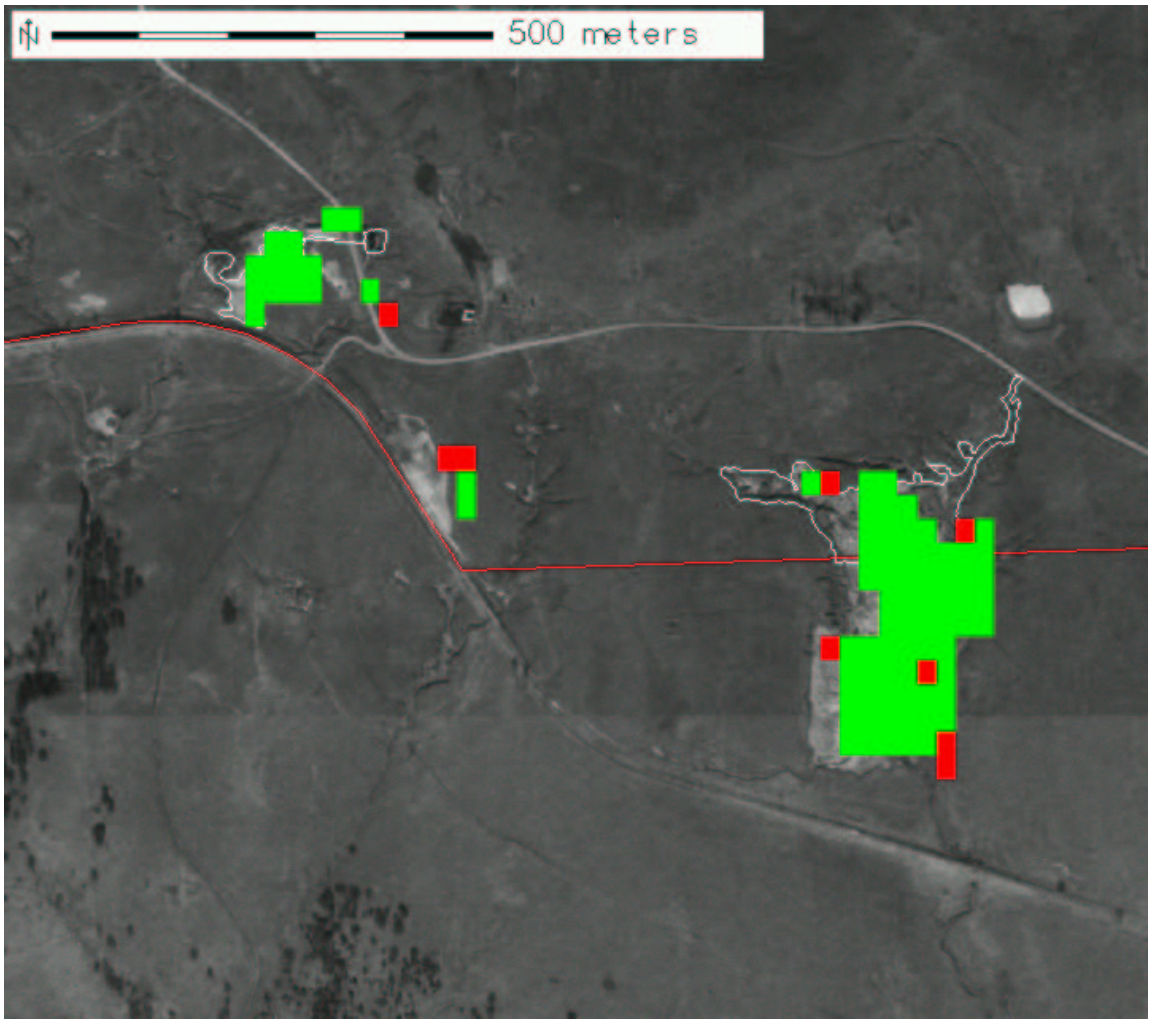
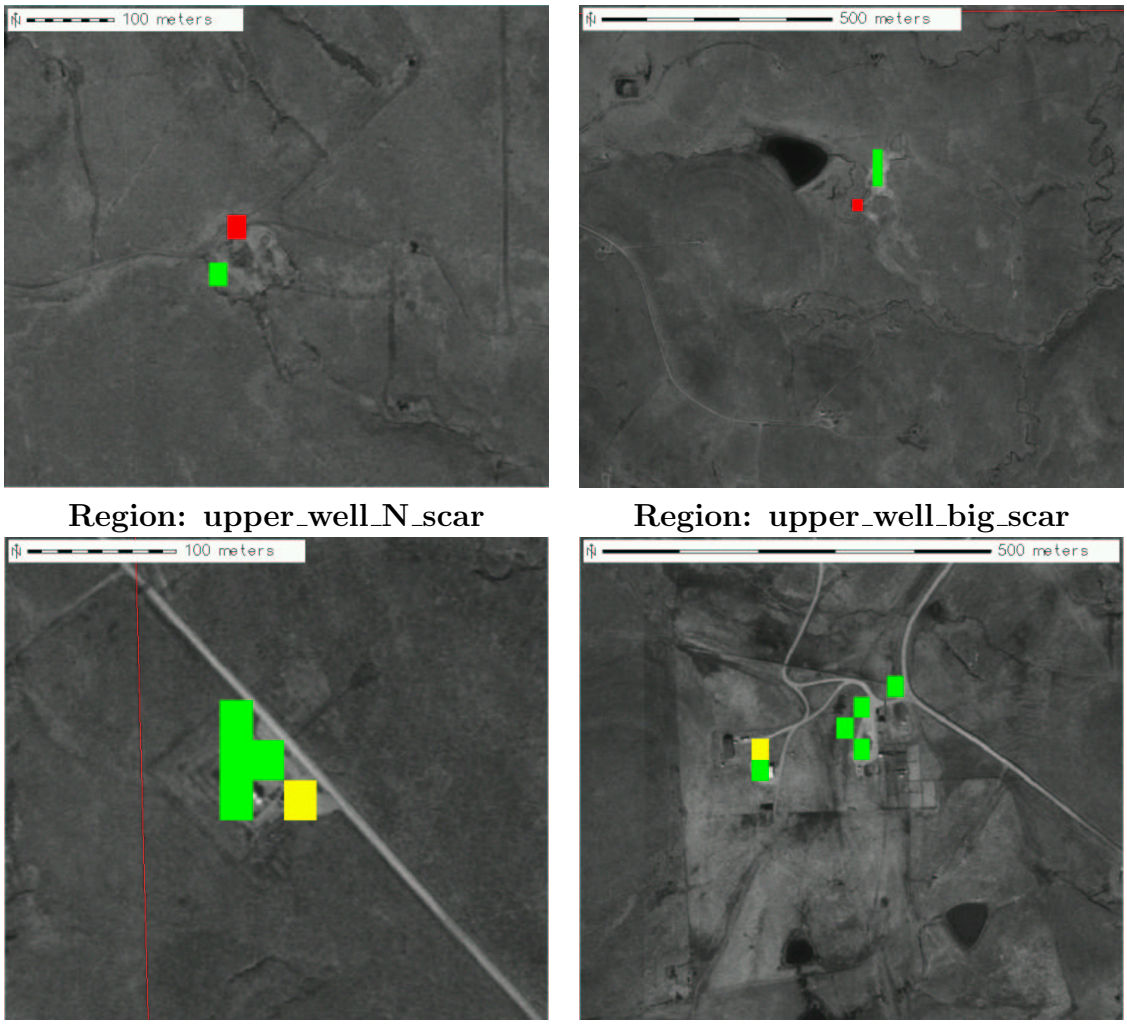


Figure 6.13: Brine scar predictions for the bigscar region using four bands. The green cells appear in both brine scar clusters while red cells are included only in predictions from the original analysis using 187 bands. Corresponds to Figure A.3.



Region: upper_well_N_scar **Region: upper_well_big_scar**

Region: roadside_tanks **Region: ranch2**

Figure 6.14: Other brine scar predictions using four bands. The green cells appear in both brine scar clusters, the yellow cells appear in predictions from the four band analysis, and red cells are included only in predictions from the original analysis using 187 bands. The maps correspond to Figure A.6 (top left), Figure A.10 (top right), Figure A.13 (bottom left), and Figure A.17 (bottom right).

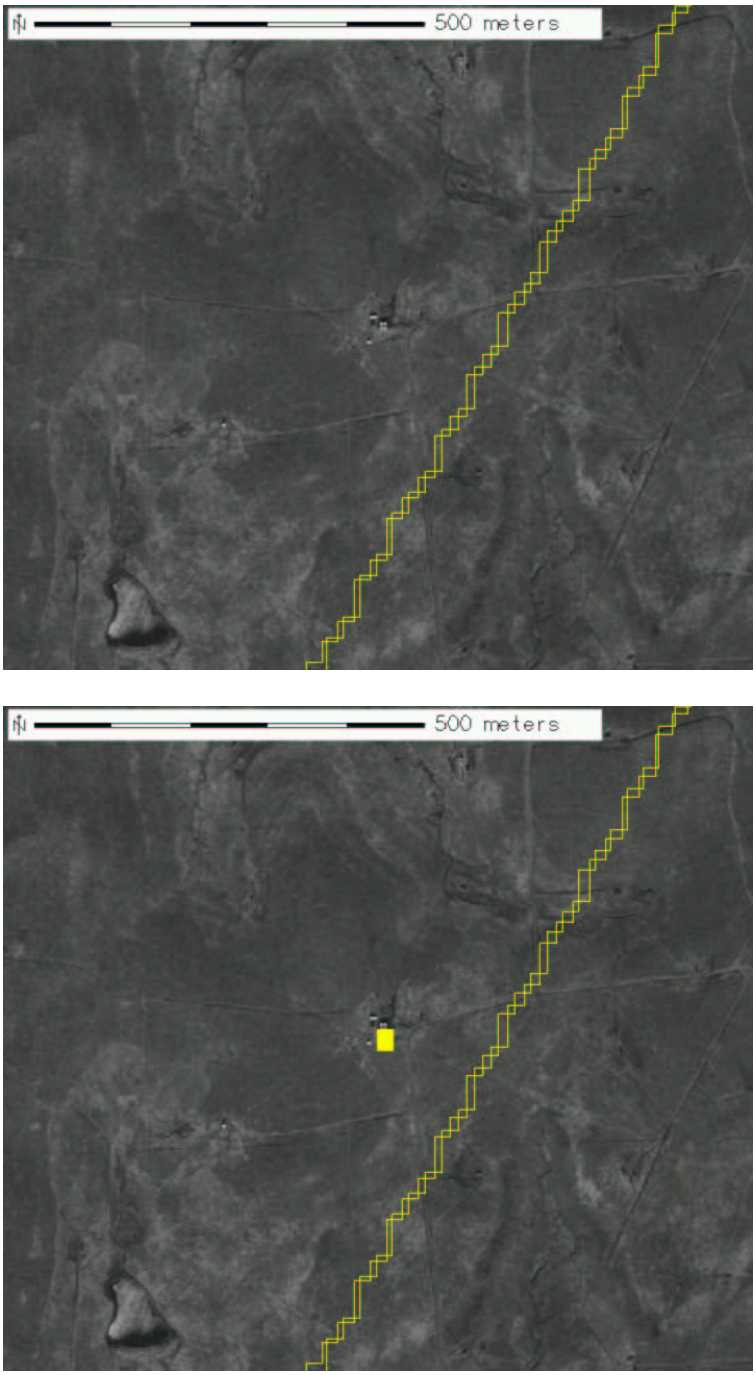


Figure 6.15: New brine scar prediction from four band analysis. This cell in r05sc05 was not predicted as a brine scar in the original 187 band analysis. However, it overlaps with other scenes, and it was predicted as a brine scar in scene r06sc04 in the 187 band analysis. Corresponds to Figure A.66.

Chapter 7

Conclusions

The four objectives of this research were to 1) determine if brine spills could be detected spectroscopically, 2) determine if spectral analysis could be performed using a statistical method to identify surface features quickly and easily from hyperspectral imagery without removing atmospheric effects from the spectra or performing spectral unmixing, 3) develop a spectral signature for brine spills and scars which could be applied at other locations, and 4) determine if brine spills or scars could be detected using substantially fewer spectral bands so that a smaller and cheaper instrument could be applied to detect such disturbances.

An area in Osage County, Oklahoma, centered over the Tallgrass Prairie Preserve was studied using available hyperspectral imagery from AVIRIS. Oil and gas exploration and production have occurred in this area for more than 80 years. While site visits for verifying brine scar predictions were not possible, 1 m orthophotos were used to check prediction results.

As these research results indicate, brine disturbances have a fairly unique spectral signature which is particularly bright in the visible and the long near infrared. Using hyperspectral imagery, it was possible to locate probable brine disturbances. Cluster analysis proved to be a powerful method for extracting surface features from the AVIRIS imaging spectroscopy data sets over a wide geographic area. MGC's ability to find brine scars and other similar geophysical disturbances without removing atmospheric effects or performing complicated spectral unmixing could prove useful in quickly locating areas for remediation or further study by more traditional means. Identification of brine disturbances using remote sensing, as successfully demonstrated here, should help petroleum companies and industry regulators in managing resources for cleanup and remediation.

One of the significant benefits of cluster analysis is its ability to sample and group data sets in a true multivariate fashion and define centroids representing mean conditions in the resulting data space region occupied by relevant observations. For spectral data, this means signatures can be quickly and easily defined and applied to similar data sets for object identification or classification. In this study, a spectral signature for brine spills and scars was developed using MGC by tuning the level of

division for a single scene of AVIRIS data containing known brine scars. This spectral signature was then applied to other AVIRIS scenes to identify other candidate brine disturbances.

As explained herein, some areas classified as brine scar predictions were attributable to other features using the DOQQs. Some of these predications—bright roof reflections in particular—are rightly considered to be false positives. These strong reflections appear in the brine scar cluster because its centroid has the highest radiance across the entire spectral range. However, many other predictions constitute geophysical disturbances not unlike brine scars. Freshly plowed or fertilized fields, parking lots, gravel driveways, road shoulders, road intersections, and large sandbars are occasionally identified as possible brine scars because of their similar spectral signatures. Many of these areas are likely to collect mineral crusts from human activities. Representativeness analysis using maps with cells colored by their similarity to the brine scar centroid (spectral) definition along with spectral plots help to further refine the analysis, and high resolution DOQQs make eliminating false positive straightforward in most cases. Applying a similarity constraint on data space distances could also significantly reduce the number of actual false positives.

A total of 87 regions were predicted as containing brine spills or scars, and 40 were attributable to petroleum activities based on objects visible in the DOQQs. Nine of these petroleum regions are contained within the TPP boundary. Petroleum prediction regions represent at least 65 discrete brine spills, 14 of which occurred on the TPP. As many as 47 prediction regions may represent false positives or anthropogenic disturbances which may not be the result of produced water releases. Since the DOQQs were acquired in 1995 and the AVIRIS overflight occurred in 1999, it was not always possible to identify features in the DOQQs which corresponded to predictions derived from the hyperspectral imagery.

Since the MGC algorithm is frequently applied to the analysis of extremely large data sets, it uses a convergence criterion of 0.5%. That is, the procedure stops once less than 0.5% of the observations change cluster assignment between iterations. To test the sensitivity of this criterion, data from an AVIRIS scene were clustered until no observations changed cluster assignment between iterations. This complete convergence of the clustering fit is probably not necessary in most cases, and since it resulted in only two cells being added to the brine scar cluster, it was not necessary here. Those two cells were part of brine scars which had already been identified with 0.5% convergence.

In order to determine if brine disturbances could be detected with fewer bands, four bands were chosen, based on the spectra of cluster centroids from the original analysis, to serve as characteristics for new cluster analyses of an AVIRIS scene. While 15 cells were missed in the four band analysis with 12 clusters as compared with the 187 band analysis with 18 clusters, 90% of the same regions contained brine scar predictions using the much smaller data set. Three cells were added to the predictions in the four band analysis. Two of these cells are merely extensions of

objects found by both analyses, and the third corresponds to a prediction cell from another scene. A multispectral remote sensing instrument which looked only at these four bands of the spectrum could be constructed and operated much more cheaply than the AVIRIS instrument for detecting these geophysical disturbances at petroleum areas all over the Earth. Moreover, other nearby bands may be even better indicators for brine scars. Since they were chosen by inspecting the spectra of centroids obtained in the original analysis, the four bands chosen for this study may not be the optimal wavelengths for brine scar detection. Coefficients of a discriminant analysis trained using the MGC results could be used to identify the four best bands for brine scar discrimination. Nevertheless, a large proportion of the original predictions were found again with the four bands used here while only two objects (one of which may be a false positive) were missed.

Bibliography

[1] Handbook of geophysics and space environment. Technical report, U.S. Air Force Geophysics Directorate, Washington, D.C., 1965.

[2] Airborne visible/infrared imaging spectrometer (AVIRIS): A description of the sensor, ground data processing facility, laboratory calibration, and first results. Technical Report 87–38, NASA Jet Propulsion Laboratory, November 1987.

[3] GNU General Public License. On-line publication, Free Software Foundation, June 1991.

[4] G. P. Anderson, A. Berk, P. K. Acharya, M. W. Matthew, L. S. Bernstein, J. H. Chetwynd, H. Dothe, S. M. Adler-Golden, A. J. Ratkowski, G. W. Felde, J. A. Gardner, M. L. Hoke, S. C. Richtsmeier, B. Pukall, J. Mello, and L. S. Jeong. *MODTRAN4: Radiative Transfer Modeling for Remote Sensing*, volume 4049, pages 176–183. Proceedings of SPIE, 2000.

[5] Tom L. Ashwood. Site statistics for the nature conservancy's tallgrass prairie preserve. Project status report report, Oak Ridge National Laboratory, October 2001.

[6] Colin Neville Banwell. *Fundamentals of Molecular Spectroscopy*. McGraw-Hill, London, second edition, 1972.

[7] J. D. Bayliss, J. A. Gualtieri, and R. F. Cromp. Analyzing hyperspectral data with independent component analysis. In *Proceedings of the SPIE Applied Image and Pattern Recognition Workshop*, October 1997.

[8] David J. Carty, Stephen M. Swetish, William F. Priebe, and Wayne Crawley. Remediation of salt-affected soils at oil and gas production facilities. API Publication 4663, American Petroleum Institute, October 1997.

[9] R. N. Clark, G. A. Swayze, R. Wise, K. E. Livo, T. M. Hoefen, R. F. Kokaly, and S. J. Sutley. USGS digital spectral library splib05a. USGS Open File Report 03–395, U.S. Geological Survey, 2003.

[10] Rebecca A. Efroymson, Tina M. Carlsen, Henriette I. Jager, Tanya Kostova, Eric A. Carr, William W. Hargrove, James Kercher, and Tom L. Ashwood. *Toward a Framework for Assessing Risk to Vertebrate Populations from Brine and Petroleum Spills at Exploration and Production Sites*, pages 261–285. ASTM STP 1458, West Conshohocken, PA, 2004.

[11] B.-C. Gao, K. B. Heidebrecht, and A. F. H. Goetz. Derivation of scaled surface reflectances from aviris data. *Remote Sensing of the Environment*, 44:145–163, 1993.

[12] A. F. H. Goetz, J. W. Boardman, B. Kindel, and K. B. Heidebrecht. Atmsopheric corrections: On deriving surface reflectance from hyperspectral imagers. In *Proceedings of the SPIE Annual Meeting*, volume 3118, pages 14–22, 1997.

[13] Richard M. Goody. *Principles of Atmospheric Physics and Chemistry*. Oxford University Press, Inc., New York, 1995.

[14] Richard M. Goody and Yuk Ling Yung. *Atmospheric Radiation*. Oxford University Press, Inc., New York, second edition, 1989.

[15] William W. Hargrove and Forrest M. Hoffman. Using Multivariate Clustering to Characterize Ecoregion Borders. *Computing in Science & Engineering*, 1(4):18–25, July 1999.

[16] William W. Hargrove and Forrest M. Hoffman. A Flux Atlas for Representativeness and Statistical Extrapolation of the AmeriFlux Network. Technical Memorandum ORNL/TM–2004/112, Oak Ridge National Laboratory, April 2004.

[17] William W. Hargrove and Forrest M. Hoffman. The potential of multivariate quantitative methods for delineation and visualization of ecoregions. *Environmental Management*, 2004. In Press.

[18] William W. Hargrove, Forrest M. Hoffman, and Beverly E. Law. New analysis reveals representativeness of the AmeriFlux network. *Eos Trans. AGU*, 84(48):529, 535, December 2003.

[19] William W. Hargrove, Forrest M. Hoffman, and Thomas Sterling. The Do-It-Yourself Supercomputer. *Scientific American*, 265(2):72–79, August 2001.

[20] Victoria Harmon and Michael Shapiro. GRASS Tutorial: Image Processing. Technical report, U.S. Army Construction Engineering Research Laboratory, 1991.

[21] J. A. Hartigan. *Clustering Algorithms*. John Wiley & Sons, New York, 1975.

[22] Forrest M. Hoffman and William W. Hargrove. Multivariate Geographic Clustering Using a Beowulf-style Parallel Computer. In Hamid R. Arabnia, editor, *Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA '99)*, volume III, pages 1292–1298, Las Vegas, Nevada, June 1999. CSREA Press. ISBN 1–892512–11–4.

[23] Forrest M. Hoffman, William W. Hargrove, David J. Erickson, and Robert J. Oglesby. Using clustered climate regimes to analyze and compare predictions from fully coupled general circulation models. *Earth Interactions*, 2004. In Press.

[24] W. Colgan III, M. C. Vavrek, and J. Bolton. Re-vegetation of an oil/brine spill: Interaction between plants and mycorrhizal fungi. In *Proceedings of the 9th Annual International Petroleum Enviornmental Conference*, Albuquerque, NM, October 2002.

[25] Henriette I. Jager, Rebecca A. Efroymson, Kerry Sublette, and Tom L. Ashwood. Unnatural landscapes in ecology: Generating the spatial distribution of brine spills. *Environmetrics*, 2004. Submitted.

[26] Gnanamanika Mahinthakumar, Forrest M. Hoffman, William W. Hargrove, and Nicolas T. Karonis. Multivariate Geographic Clustering in a Metacomputing Environment Using Globus. In *Proceedings of the ACM/IEEE SuperComputing 99 Conference*, Portland, Oregon, November 1999.

[27] M. W. Matthew, S. M. Adler-Golden, A. Berk, S. C. Richtsmeier, R. Y. Levin, L. S. Bernstein, P. K. Acharya, G. P. Anderson, G. W. Felde, M. P. Hoke, A. Ratkowski, H.-H. Burke, R. D. Kaiser, and D. P. Miller. Status of atmospheric correction using a MODTRAN4-based algorithm. In *Proceeding of SPIE: Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery IV*, volume 4049, pages 199–207, 2000.

[28] M. Neteler and H. Mitasova. *Open Source GIS: A GRASS GIS Approach*. Kluwer Academic Publishers, Boston, 2002.

[29] L. S. Rothman, A. Barbe, D. Chris Benner, L. R. Brown, C. Camy-Peyret, M. R. Carleer, K. Chance, C. Clerbaux, V. Dana, V. M. Deri, A. Fayt, J.-M. Flaud, R. R. Gamache, A. Goldman, D. Jacquemart, K. W. Jucks, W. J. Lafferty, J.-Y. Mandin, S. T. Massie, V. Nemtchinov, D. A. Newnham, A. Perrin, C. P. Rinsland, J. Schroeder, K. M. Smith, M. A. H. Smith, K. Tang, R. A. Toth, J. Vander Auwera, P. Varanasi, and K. Yoshino. The HITRAN molecular spectroscopic database: Edition of 2000 including updates though 2001. *Journal of Quantitative Spectroscopy and Radiative Transfer*, February 2003. In press.

[30] S. Seager and E. B. Ford. The vegetation red edge spectroscopic feature as a surface biomarker. In *Proceedings of the Astrophysics of Life*, Baltimore, Maryland, December 2002. e-Print Archive astro-ph/0212550.

[31] N. Short. The Landsat tutorial workbook: Basics of satellite remote sensing. NASA Reference Publication 1078, National Aeronautics and Space Administration, U.S. Government Printing Office, Washington, DC, 1982.

[32] Graeme L. Stephens. *Remote Sensing of the Lower Atmosphere: An Introduction*. Oxford University Press, Inc., New York, 1994.

[33] U.S. standard atmosphere. Technical Report NOAA–S/T76–1562, National Oceanic and Atmospheric Administration, National Aeronautics and Space Administration, U.S. Air Force, U.S. Government Printing Office, Washington, D.C., 1976.

Appendix

A.1 The *standardize* Program

The following is the C source code, *standardize.c*, which was used to standardize data sets prior to cluster analysis, unstandardize cluster centroids following cluster analysis, and transform data to the inference space of other data sets using stored means and standard deviations. The program reads ASCII data files and writes out either ASCII or native binary files suitable for use by the parallel clustering algorithm. It will save means and standard deviations to files, and can read them for unstandardizing or transforming data sets.

```
#ifdef linux
   #define _LARGEFILE_SOURCE
   #define _LARGEFILE64_SOURCE
#endif /* linux */

#include <stdio.h>
#include <unistd.h>
#include <stdlib.h>
#include <string.h>
#include <math.h>
#include <error.h>

void usage(char *name)
{
    fprintf(stderr, "Usage: %s [-v] [-b] [-h] -r num_obs \
-c num_coord [-o output.dat] [-m mean.out] [-s stddev.out] [-t] \
[-u] [-M mean.in -S stddev.in] input.dat\n", name);
    fprintf(stderr, "where options are:\n");
    fprintf(stderr, "  -b\t\twrite binary output file (defaults to \
ASCII)\n");
    fprintf(stderr, "  -c num_coord\tnumber of coordinates (or \
fields)\n");
    fprintf(stderr, "  -h\t\tprints this help information\n");
    fprintf(stderr, "  -M mean.in\tthe name of a file from which \
to read the means (in ASCII)\n");
    fprintf(stderr, "  -m mean.out\tthe name of a file in which \
to write the means (in ASCII)\n");
    fprintf(stderr, "  -o output.dat\tthe name of the output file \
(defaults to stdout)\n");
    fprintf(stderr, "  -r num_obs\tnumber of observations \
(or records)\n");
    fprintf(stderr, "  -S stddev.in\tthe name of a file from \
which to read the standard deviations\n  \t\t(in ASCII)\n");
    fprintf(stderr, "  -s stddev.out\tthe name of a file in \
which to write the standard deviations\n  \t\t(in ASCII)\n");
    fprintf(stderr, "  -t\t\tcauses the program to transform \
the input data using the mean\n  \t\tand standard deviations \
provided (requires use of -M and -S\n  \t\targuments)\n");
    fprintf(stderr, "  -u\t\tcauses the program to \
unstandardize the input data (requires\n  \t\tuse of -M \
and -S arguments)\n");
    fprintf(stderr, "  -v\t\tprints means, variances, and \
standard deviations to stdout when\n  \t\tstandardizing\n");
```

```
    fprintf(stderr, "  input.dat\tthe name of the ASCII input \
data file (cannot be stdin)\n");
    fprintf(stderr, "Setting any output file name to - causes \
the program to use stdout.\n");
    fprintf(stderr, "Because two passes are made through the \
input file when standardizing, the\ninput file cannot be stdin.  \
For consistency, the input data file can never be\nstdin.\n");
    fprintf(stderr, "Using -v in combination with -m - would be \
redundant.\n");
    fprintf(stderr, "Using -v in combination with -s - would be \
redundant.\n");
    fprintf(stderr, "Using -v and/or -m - and/or -s - \
(without specifying an output file or with\n-o -) could be \
confusing.\n");
    fprintf(stderr, "Using -v and/or -m - and/or -s - \
in combination with -b (without specifying an\noutput file \
or with -o -) is usually a bad idea.\n");
    fprintf(stderr, "The -m and -s arguments cannot be \
used when unstandardizing (i.e., in\nconjunction with -u).\n");
    fprintf(stderr, "The -m and -s arguments cannot be \
used when transforming (i.e., in\nconjunction with -t).\n");
    fprintf(stderr, "When unstandardizing or transforming, \
the input means and standard\ndeviations may be in the same \
file.  The means are read first.\n");
    fprintf(stderr, "Using -u in combination with -t is \
impossible.\n");
    fprintf(stderr, "Using -M requires the use of -S, and \
vice versa.\n");
    fprintf(stderr, "Using -v in combination with -u \
affects nothing.\n");
    fprintf(stderr, "Using -v in combination with -t \
affects nothing.\n");
    exit(0);
}

int main(int argc, char **argv)
{
    FILE *fobs, *fout, *fsout, *fmout, *fsin, *fmin;
    register int i, j, c;
    char *ifile, *ofile, *sfile, *mfile, *sifile, *mifile;
    char bin_flag, verbose, unstd, tform;
    int num_obs, num_coord, warn;
    float std_obs;
    double *data, *sq_sum, *sum, *var, *std_dev, *mean, *s, *m;

    bin_flag = verbose = unstd = tform = 0;
    num_obs = num_coord = 0;
    fmout = fsout = fmin = fsin = NULL;
    ifile = ofile = mfile = sfile = mifile = sifile = NULL;

    /* Check command-line flags and switches */
    while ((c = getopt(argc, argv, "bc:hM:m:o:r:S:s:tuv")) != EOF) {
        switch (c) {
            case 'b':
                bin_flag = 1;
                break;
```

```
            case 'c':
                num_coord = atoi(optarg);
                break;
            case 'h':
                usage(argv[0]);
                break;
            case 'M':
                mifile = strdup(optarg);
                break;
            case 'm':
                mfile = strdup(optarg);
                break;
            case 'o':
                ofile = strdup(optarg);
                break;
            case 'r':
                num_obs = atoi(optarg);
                break;
            case 'S':
                sifile = strdup(optarg);
                break;
            case 's':
                sfile = strdup(optarg);
                break;
            case 't':
                tform = 1;
                break;
            case 'u':
                unstd = 1;
                break;
            case 'v':
                verbose = 1;
                break;
        }
    }

    if (!ofile)
        ofile = "-";

    if (optind < argc)
        ifile = strdup(argv[optind]);
    else
        usage(argv[0]);

    for (optind++; optind < argc; optind++)
        fprintf(stderr, "Argument %s ignored.\n", argv[optind]);

    /* Error checking */
    if (num_obs == 0) {
        fprintf(stderr, "Number of observations must be \
greater than zero.\n");
        usage(argv[0]);
    }
    if (num_coord == 0) {
        fprintf(stderr, "Number of coordinates must be \
greater than zero.\n");
```

```
        usage(argv[0]);
    }
    if (unstd && tform) {
        fprintf(stderr, "Using -u in combination with -t is \
impossible.\n");
        usage(argv[0]);
    }
    if ((mifile && !sifile) || (!mifile && sifile)) {
        fprintf(stderr, "Using of the -M argument requires \
use of the -S argument, and vice versa.\n");
        usage(argv[0]);
    }
    if (unstd || tform) {
        if (!mifile) {
            fprintf(stderr, "A file of means must be provided \
using the -M argument when unstandardizing or transforming.\n");
            usage(argv[0]);
        }
        if (!sifile) {
            fprintf(stderr, "A file of standard deviations must \
be provided using the -S argument when unstandardizing or \
transforming.\n");
            usage(argv[0]);
        }
        if (mfile) {
            fprintf(stderr, "Means cannot be output when \
unstandardizing or transforming.\n");
            usage(argv[0]);
        }
        if (sfile) {
            fprintf(stderr, "Standard deviations cannot be output \
when unstandardizing or transforming.\n");
            usage(argv[0]);
        }
    }

    /* Open input file and output files.  We open them all up front
       so that the user does not get half-way through before
       discovering there may be a problem with the output files. */
    if (!(fobs = fopen64(ifile, "r"))) {
        perror(ifile);
        exit(2);
    }
    if (strcmp("-", ofile)) {
        if (!(fout = fopen64(ofile, "w"))) {
            perror(ofile);
            exit(3);
        }
    }
    else {
        fout = stdout;
    }
    if (mfile) {
        if (strcmp("-", mfile)) {
            if (!(fmout = fopen64(mfile, "w"))) {
                perror(mfile);
```

```
            exit(3);
        }
    }
    else {
        fmout = stdout;
    }
}
if (sfile) {
    if (strcmp("-", sfile)) {
        if (!(fsout = fopen64(sfile, "w"))) {
            perror(sfile);
            exit(3);
        }
    }
    else {
        fsout = stdout;
    }
}
if (mifile) {
    if (strcmp("-", mifile)) {
        if (!(fmin = fopen64(mifile, "r"))) {
            perror(mifile);
            exit(3);
        }
    }
    else {
        fmin = stdin;
    }
}
if (sifile) {
    if (strcmp("-", sifile)) {
        if (strcmp(sifile, mifile)) {
            if (!(fsin = fopen64(sifile, "r"))) {
                perror(sifile);
                exit(3);
            }
        }
        else {
            fsin = fmin;
        }
    }
    else {
        fsin = stdin;
    }
}

/* Allocate memory for data, sq_sum, and sum variables */
if (!(data = (double *)malloc(num_coord * sizeof(double)))) {
    perror("data");
    exit(4);
}
if (!unstd && !tform) {
    if (!(sq_sum = (double *)malloc(num_coord *
        sizeof(double)))) {
        perror("sq_sum");
        exit(4);
```

```
    }
    if (!(sum = (double *)malloc(num_coord *
        sizeof(double)))) {
        perror("sum");
        exit(4);
    }
    if (!(var = (double *)malloc(num_coord *
        sizeof(double)))) {
        perror("var");
        exit(4);
    }
    if (!(mean = (double *)malloc(num_coord *
        sizeof(double)))) {
        perror("mean");
        exit(4);
    }
    if (!(std_dev = (double *)malloc(num_coord *
        sizeof(double)))) {
        perror("std_dev");
        exit(4);
    }
}
else {
    sq_sum = sum = var = mean = std_dev = NULL;
}
if (!(s = (double *)malloc(num_coord * sizeof(double)))) {
    perror("s");
    exit(4);
}
if (!(m = (double *)malloc(num_coord * sizeof(double)))) {
    perror("m");
    exit(4);
}

/* Read in means and standard deviations if desired;
   otherwise just initialize */
if (mifile && sifile) {
    for (j = 0; j < num_coord; j++)
        fscanf(fmin, " %lg ", &m[j]);
    for (j = 0; j < num_coord; j++)
        fscanf(fsin, " %lg ", &s[j]);
    if (fsin == fmin) {
        if (fmin != stdin)
            fclose(fmin);
    }
    else {
        if (fmin != stdin)
            fclose(fmin);
        if (fsin != stdin)
            fclose(fsin);
    }
}
else {
    for (j = 0; j < num_coord; j++) {
        m[j] = (double)0.0;
        s[j] = (double)1.0;
```

```
        }
    }

    /* UNSTANDARDIZING */
    if (unstd) {
        /* Read observations, unstandardize data, and write
           it to the output file.  Only a couple lines are
           different, but I duplicate the entire look here
           because I'm after SPEED. */
        if (bin_flag) {
            fprintf(stderr, "Reading %d observations with %d \
coordinates from ASCII input file and\nwriting %d unstandardized \
observations to binary output file\n", num_obs, num_coord, num_obs);
            for (i = 0; i < num_obs; i++) {
                for (j = 0; j < num_coord; j++) {
                    fscanf(fobs, " %lg ", &data[j]);
                    std_obs = (float)((data[j]*s[j])+m[j]);
                    fwrite(&std_obs, sizeof(float),1,fout);
                }
            }
        }
        else {
            fprintf(stderr, "Reading %d observations with %d \
coordinates from ASCII input file and\nwriting %d unstandardized \
observations to ASCII output file\n", num_obs, num_coord, num_obs);
            for (i = 0; i < num_obs; i++) {
                for (j = 0; j < num_coord; j++) {
                    fscanf(fobs, " %lg ", &data[j]);
                    std_obs = (float)((data[j]*s[j])+m[j]);
                    fprintf(fout, " %f", std_obs);
                }
                fprintf(fout, "\n");
            }
        }
    }
    /* TRANSFORMING */
    else if (tform) {
        /* Read observations, transform data, and write
           it to the output file.  Only a couple lines are
           different, but I duplicate the entire look here
           because I'm after SPEED. */
        if (bin_flag) {
            fprintf(stderr, "Reading %d observations with %d \
coordinates from ASCII input file and\nwriting %d transformed \
observations to binary output file\n", num_obs, num_coord, num_obs);
            for (i = 0; i < num_obs; i++) {
                for (j = 0; j < num_coord; j++) {
                    fscanf(fobs, " %lg ", &data[j]);
                    std_obs = (float)((data[j]-m[j])/s[j]);
                    fwrite(&std_obs, sizeof(float),1,fout);
                }
            }
        }
        else {
            fprintf(stderr, "Reading %d observations with %d \
coordinates from ASCII input file and\nwriting %d transformed \
```

```
observations to ASCII output file\n", num_obs, num_coord, num_obs);
            for (i = 0; i < num_obs; i++) {
                for (j = 0; j < num_coord; j++) {
                    fscanf(fobs, " %lg ", &data[j]);
                    std_obs = (float)((data[j]-m[j])/s[j]);
                    fprintf(fout, " %f", std_obs);
                }
                fprintf(fout, "\n");
            }
        }
    }
    /* STANDARDIZING */
    else {
        /* Initialize sq_sum and sum to zero */
        for (j = 0; j < num_coord; j++)
            sq_sum[j] = sum[j] = (double)0.0;

        fprintf(stderr, "Pass 1: Reading %d observations with %d \
coordinates from ASCII input file\n", num_obs, num_coord);
        /* Pass 1: Read observations and compute squared sum
           and sum in order to compute variance */
        for (i = 0; i < num_obs; i++) {
            for (j = 0; j < num_coord; j++) {
                fscanf(fobs, " %lg ", &data[j]);
                sq_sum[j] += data[j] * data[j];
                sum[j] += data[j];
            }
        }

        warn = 0;
        /* Compute mean, variance, and standard deviation
           for each coordinate */
        for (j = 0; j < num_coord; j++) {
            mean[j] = sum[j] / (double)num_obs;
            var[j] = (sq_sum[j] - (sum[j] * sum[j] /
                (double)num_obs)) / ((double)num_obs-(double)1);
            std_dev[j] = sqrt((double)var[j]);
            if (std_dev[j] == (double)0.0)
                ++warn;
        }

        if (verbose) {
            printf("Means:\n");
            for (j = 0; j < num_coord; j++) {
                printf(" %f", mean[j]);
            }
            printf("\n");
            printf("Variances:\n");
            for (j = 0; j < num_coord; j++) {
                printf(" %f", var[j]);
            }
            printf("\n");
            printf("Standard Deviations:\n");
            for (j = 0; j < num_coord; j++) {
                printf(" %f", std_dev[j]);
            }
```

```
            printf("\n");
        }

        /* Output means and standard deviations to files,
           if desired */
        if (mfile) {
            for (j = 0; j < num_coord; j++)
                fprintf(fmout, " %f", mean[j]);
            fprintf(fmout, "\n");
            if (fmout != stdout)
                fclose(fmout);
        }
        if (sfile) {
            for (j = 0; j < num_coord; j++)
                fprintf(fsout, " %f", std_dev[j]);
            fprintf(fsout, "\n");
            if (fsout != stdout)
                fclose(fsout);
        }

        if (warn)
            fprintf(stderr, "WARNING: %d standard deviations are \
zero!\n", warn);

        /* Move back to the beginning of the input file for pass 2 */
        rewind(fobs);

        /* Write binary or ASCII output depending on
           command-line flag.  Only a couple of lines are
           different, but I duplicate the entire loop here
           because I'm after SPEED. */
        if (bin_flag) {
            fprintf(stderr, "Pass 2: Writing %d standardized \
observations with %d coordinates to binary output \
file\n", num_obs, num_coord);
            /* Pass 2: Read observations, standardize data,
               and write it to the output file */
            for (i = 0; i < num_obs; i++) {
                for (j = 0; j < num_coord; j++) {
                    fscanf(fobs, " %lg ", &data[j]);
                    std_obs = (float)((s[j] * (data[j] -
                        mean[j]) / std_dev[j]) + m[j]);
                    fwrite(&std_obs, sizeof(float),1,fout);
                }
            }
        }
        else {
            fprintf(stderr, "Pass 2: Writing %d standardized \
observations with %d coordinates to ASCII output \
file\n", num_obs, num_coord);

            /* Pass 2: Read observations, standardize data,
               and write it to the output file */
            for (i = 0; i < num_obs; i++) {
                for (j = 0; j < num_coord; j++) {
                    fscanf(fobs, " %lg ", &data[j]);
```

```
                    std_obs = (float)((s[j] * (data[j] -
                        mean[j]) / std_dev[j]) + m[j]);
                    fprintf(fout, " %f", std_obs);
                }
                fprintf(fout, "\n");
            }
        }
    }

    /* Close data files */
    if (fobs != stdin)
        fclose(fobs);
    if (fout != stdout)
        fclose(fout);

    /* Free memory */
    free(data);
    if (!unstd && !tform) {
        free(sq_sum);
        free(sum);
        free(var);
        free(mean);
        free(std_dev);
    }
    free(s);
    free(m);

    exit(0);
}
```

A.2 The *read_data* Program

The following is the C source code, *read_data.c*, written to extract the AVIRIS hyperspectral data from the big endian binary band-interleaved files provided by JPL. This program was run against all seven scene files to create band-interleaved ASCII files for clustering and separate band files for loading into the GIS. When compiled with MAKE_GIS_FILES defined, the program creates 224 individual files (one for each wavelength) as big endian binary 2s complement unsigned integers which can be loaded directly into the GRASS GIS.

```
#include <stdio.h>
#include <limits.h>
#include <endian.h>
#include <sys/param.h>

/*
 * Read the band-interleaved AVIRIS hyperspectral image files,
 * and write out layers for each wavelength as 2s complement
 * unsigned integers for loading directly into the GRASS GIS
 * and send band-interleaved raw ASCII file to stdout for
 * use in clustering.  AVIRIS data are in BIG ENDIAN so byte
 * swapping is required on LITTLE ENDIAN machines.
 */

#undef MAKE_GIS_FILES   /* If defined, the GIS files are made;
               otherwise, only the ASCII data is output */

#define NCHANNELS   224
#define NSAMPLES    614
#define NLINES      512

#define BSWAP16(x) \
   ( (((x) & 0xff00) >>  8) | (((x) & 0x00ff) << 8) )

#define BSWAP32(x) \
   ( (((x) & 0xff000000) >> 24) | (((x) & 0x00ff0000) >>  8) | \
     (((x) & 0x0000ff00) <<  8) | (((x) & 0x000000ff) << 24) )


int main(int argc, char **argv)
{
    FILE *fp;
    int i, j;
    short int ival, oval;
#ifdef MAKE_GIS_FILES
    FILE *gfp[NCHANNELS];
    unsigned short int gval;
    char fname[MAXPATHLEN];
#endif /* MAKE_GIS_FILES */

    if (!(fp = fopen(argv[1], "r"))) {
        perror(argv[1]);
        exit(1);
```

```
    }

#ifdef MAKE_GIS_FILES
    for (i = 0; i < NCHANNELS; i++) {
        sprintf(fname, "%s.%03d.gis", argv[1], i);
        if (!(gfp[i] = fopen(fname, "w"))) {
            perror(fname);
            exit(2);
        }
    }
#endif /* MAKE_GIS_FILES */

    for (j = 0; j < NSAMPLES*NLINES; j++) {
        for (i = 0; i < NCHANNELS && fread(&ival, sizeof(short int),
            1, fp); i++) {
#if __BYTE_ORDER == __LITTLE_ENDIAN
            oval = BSWAP16(ival);
#endif /* __BYTE_ORDER == __LITTLE_ENDIAN */
            printf(" %hd", oval);
#ifdef MAKE_GIS_FILES
#if __BYTE_ORDER == __LITTLE_ENDIAN
            gval = BSWAP16((unsigned short)(oval < 0 ?
                (SHRT_MAX - oval + 1) : oval));
#else /* __BYTE_ORDER == __LITTLE_ENDIAN */
            gval = (unsigned short)(oval < 0 ?
                (SHRT_MAX - oval + 1) : oval);
#endif /* __BYTE_ORDER == __LITTLE_ENDIAN */
#endif /* MAKE_GIS_FILES */
        }
        printf("\n");
    }
    fprintf(stderr, "Read %d values from %s\n", (i*j), argv[1]);

#ifdef MAKE_GIS_FILES
    for (i = 0; i < NCHANNELS; i++)
        fclose(gfp[i]);
#endif /* MAKE_GIS_FILES */
    fclose(fp);
    return 0;
}
```

A.3 AVIRIS Spectral Calibration

Table A.1: AVIRIS spectral calibration.

Band	Wavelength Center [nm]	Uncertainty in Wavelength Center	FWHM for Equivalent Gaussian [nm]	Uncertainty in FWHM	Included in Clustering	Included in Small Subset
1	370.43	0.45	9.17	0.30		
2	380.26	0.30	9.16	0.30		
3	390.09	0.22	9.15	0.30		
4	399.92	0.09	9.13	0.19	✓	
5	409.75	0.07	9.12	0.11	✓	
6	419.58	0.06	9.12	0.08	✓	
7	429.41	0.06	9.11	0.06	✓	
8	439.23	0.06	9.10	0.05	✓	
9	449.06	0.05	9.10	0.05	✓	
10	458.89	0.05	9.10	0.05	✓	✓
11	468.71	0.05	9.09	0.05	✓	
12	478.54	0.05	9.09	0.05	✓	
13	488.37	0.05	9.09	0.04	✓	
14	498.19	0.05	9.09	0.04	✓	
15	508.02	0.05	9.10	0.04	✓	
16	517.84	0.05	9.10	0.04	✓	
17	527.67	0.05	9.11	0.04	✓	
18	537.49	0.05	9.12	0.04	✓	
19	547.32	0.05	9.12	0.04	✓	
20	557.14	0.05	9.13	0.04	✓	✓
21	566.96	0.05	9.14	0.04	✓	
22	576.79	0.05	9.16	0.04	✓	
23	586.61	0.05	9.17	0.04	✓	
24	596.43	0.05	9.18	0.04	✓	
25	606.25	0.05	9.20	0.04	✓	
26	616.08	0.06	9.22	0.04	✓	
27	625.90	0.05	9.24	0.04	✓	
28	635.72	0.05	9.26	0.04	✓	
29	645.54	0.05	9.28	0.04	✓	
30	655.36	0.05	9.30	0.04	✓	
31	665.18	0.05	9.33	0.04	✓	
32	675.00	0.07	9.35	0.06	✓	
33	654.17	0.06	9.27	0.05	✓	
34	663.71	0.06	9.26	0.05	✓	
35	673.25	0.06	9.25	0.05	✓	
36	682.79	0.06	9.25	0.05	✓	
37	692.33	0.06	9.26	0.05	✓	
38	701.87	0.06	9.27	0.05	✓	
39	711.41	0.06	9.30	0.05	✓	
40	720.95	0.06	9.33	0.05	✓	
41	730.49	0.06	9.36	0.05	✓	
42	740.03	0.07	9.41	0.05	✓	

Table A.1: Continued

Band	Wavelength Center [nm]	Uncertainty in Wavelength Center	FWHM for Equivalent Gaussian [nm]	Uncertainty in FWHM	Included in Clustering	Included in Small Subset
43	749.57	0.06	9.46	0.05	✓	
44	759.12	0.06	9.52	0.05	✓	
45	768.66	0.07	9.58	0.04	✓	
46	778.20	0.06	9.65	0.05	✓	
47	787.75	0.07	9.73	0.04	✓	
48	797.29	0.06	9.82	0.05	✓	
49	806.84	0.06	9.92	0.05	✓	
50	816.39	0.07	10.02	0.05	✓	
51	825.93	0.06	10.13	0.05	✓	
52	835.48	0.06	10.24	0.05	✓	
53	845.03	0.06	10.37	0.05	✓	
54	854.58	0.06	10.50	0.05	✓	
55	864.12	0.06	10.64	0.05	✓	
56	873.67	0.06	10.78	0.05	✓	
57	883.22	0.06	10.94	0.05	✓	
58	892.77	0.06	11.10	0.05	✓	
59	902.33	0.06	11.26	0.05	✓	
60	911.88	0.06	11.44	0.05	✓	
61	921.43	0.06	11.62	0.05	✓	
62	930.98	0.06	11.81	0.05	✓	
63	946.35	0.06	9.72	0.06	✓	
64	955.76	0.06	9.72	0.06	✓	
65	965.17	0.06	9.72	0.06	✓	
66	974.58	0.06	9.72	0.06	✓	
67	983.99	0.06	9.72	0.06	✓	
68	993.39	0.06	9.72	0.06	✓	
69	1002.80	0.06	9.72	0.06	✓	
70	1012.21	0.09	9.72	0.06	✓	
71	1021.62	0.06	9.72	0.06	✓	
72	1031.03	0.06	9.72	0.06	✓	
73	1040.44	0.06	9.72	0.06	✓	
74	1049.84	0.06	9.72	0.06	✓	
75	1059.25	0.08	9.72	0.06	✓	
76	1068.66	0.06	9.72	0.06	✓	
77	1078.06	0.06	9.72	0.06	✓	
78	1087.47	0.06	9.73	0.06	✓	
79	1096.88	0.06	9.73	0.06	✓	
80	1106.28	0.06	9.73	0.06	✓	
81	1115.69	0.06	9.73	0.06	✓	
82	1125.10	0.06	9.73	0.06	✓	
83	1134.50	0.06	9.74	0.06	✓	
84	1143.91	0.06	9.74	0.06	✓	
85	1153.31	0.06	9.74	0.06	✓	
86	1162.72	0.10	9.74	0.07	✓	
87	1172.12	0.06	9.74	0.06	✓	
88	1181.52	0.06	9.75	0.06	✓	

Table A.1: Continued

Band	Wavelength Center [nm]	Uncertainty in Wavelength Center	FWHM for Equivalent Gaussian [nm]	Uncertainty in FWHM	Included in Clustering	Included in Small Subset
89	1190.93	0.06	9.75	0.06	✓	
90	1200.33	0.06	9.75	0.06	✓	
91	1209.73	0.06	9.75	0.06	✓	
92	1219.14	0.06	9.76	0.06	✓	
93	1228.54	0.06	9.76	0.06	✓	
94	1237.94	0.06	9.76	0.06	✓	
95	1247.35	0.06	9.77	0.06	✓	
96	1256.75	0.06	9.77	0.06	✓	
97	1255.57	0.09	10.50	0.07	✓	
98	1265.54	0.09	10.52	0.07	✓	
99	1275.51	0.09	10.54	0.07	✓	
100	1285.48	0.09	10.57	0.07	✓	
101	1295.46	0.09	10.59	0.07	✓	
102	1305.43	0.09	10.61	0.07	✓	
103	1315.40	0.09	10.63	0.07	✓	
104	1325.37	0.09	10.64	0.07	✓	
105	1335.34	0.09	10.66	0.07	✓	
106	1345.30	0.09	10.67	0.07	✓	
107	1355.27	0.09	10.69	0.07		
108	1365.24	0.09	10.70	0.07		
109	1375.21	0.09	10.71	0.07		
110	1385.17	0.09	10.72	0.07		
111	1395.14	0.09	10.73	0.07		
112	1405.10	0.09	10.74	0.07		
113	1415.07	0.09	10.75	0.07		
114	1425.03	0.09	10.76	0.07		
115	1434.99	0.09	10.76	0.07		
116	1444.96	0.09	10.77	0.07		
117	1454.92	0.09	10.77	0.07	✓	
118	1464.88	0.09	10.77	0.07	✓	
119	1474.84	0.09	10.77	0.07	✓	
120	1484.80	0.09	10.77	0.07	✓	
121	1494.76	0.09	10.77	0.07	✓	
122	1504.72	0.09	10.77	0.07	✓	
123	1514.68	0.09	10.76	0.07	✓	
124	1524.64	0.10	10.76	0.08	✓	
125	1534.59	0.09	10.75	0.07	✓	
126	1544.55	0.11	10.74	0.09	✓	
127	1554.51	0.10	10.74	0.08	✓	✓
128	1564.46	0.10	10.73	0.08	✓	
129	1574.42	0.10	10.72	0.08	✓	
130	1584.37	0.10	10.70	0.08	✓	
131	1594.32	0.10	10.69	0.08	✓	
132	1604.28	0.10	10.68	0.08	✓	
133	1614.23	0.09	10.66	0.07	✓	
134	1624.18	0.10	10.65	0.08	✓	

Table A.1: Continued

Band	Wavelength Center [nm]	Uncertainty in Wavelength Center	FWHM for Equivalent Gaussian [nm]	Uncertainty in FWHM	Included in Clustering	Included in Small Subset
135	1634.13	0.09	10.63	0.07	✓	
136	1644.09	0.10	10.61	0.08	✓	
137	1654.04	0.10	10.59	0.08	✓	
138	1663.99	0.09	10.57	0.07	✓	
139	1673.94	0.10	10.55	0.08	✓	
140	1683.88	0.09	10.53	0.07	✓	
141	1693.83	0.11	10.50	0.09	✓	
142	1703.78	0.12	10.48	0.10	✓	
143	1713.73	0.12	10.45	0.11	✓	
144	1723.67	0.12	10.42	0.12	✓	
145	1733.62	0.11	10.40	0.08	✓	
146	1743.56	0.09	10.37	0.07	✓	
147	1753.51	0.10	10.34	0.08	✓	
148	1763.45	0.11	10.30	0.09	✓	
149	1773.40	0.12	10.27	0.10	✓	
150	1783.34	0.09	10.24	0.07	✓	
151	1793.28	0.09	10.20	0.07	✓	
152	1803.22	0.09	10.17	0.07		
153	1813.16	0.09	10.13	0.07		
154	1823.11	0.09	10.09	0.07		
155	1833.05	0.09	10.05	0.07		
156	1842.98	0.09	10.01	0.07		
157	1852.92	0.09	9.97	0.07		
158	1862.86	0.10	9.92	0.08		
159	1872.80	0.09	9.88	0.09		
160	1882.74	0.09	9.84	0.07		
161	1880.96	0.06	10.85	0.06		
162	1891.03	0.06	10.86	0.07		
163	1901.09	0.06	10.86	0.06		
164	1911.15	0.07	10.87	0.06		
165	1921.21	0.06	10.87	0.06		
166	1931.27	0.06	10.87	0.06		
167	1941.32	0.06	10.87	0.06		
168	1951.37	0.07	10.87	0.06		
169	1961.42	0.06	10.87	0.06	✓	
170	1971.47	0.07	10.87	0.06	✓	
171	1981.51	0.07	10.86	0.07	✓	
172	1991.55	0.07	10.86	0.06	✓	
173	2001.59	0.07	10.86	0.07	✓	
174	2011.63	0.07	10.85	0.07	✓	
175	2021.66	0.07	10.85	0.06	✓	
176	2031.69	0.07	10.84	0.07	✓	
177	2041.72	0.07	10.84	0.07	✓	
178	2051.75	0.07	10.83	0.07	✓	
179	2061.77	0.07	10.82	0.07	✓	
180	2071.79	0.06	10.81	0.06	✓	

Table A.1: Continued

Band	Wavelength Center [nm]	Uncertainty in Wavelength Center	FWHM for Equivalent Gaussian [nm]	Uncertainty in FWHM	Included in Clustering	Included in Small Subset
181	2081.81	0.07	10.80	0.06	✓	
182	2091.82	0.07	10.79	0.06	✓	
183	2101.83	0.07	10.78	0.06	✓	
184	2111.84	0.07	10.77	0.06	✓	
185	2121.85	0.06	10.76	0.06	✓	
186	2131.86	0.06	10.75	0.06	✓	✓
187	2141.86	0.07	10.73	0.07	✓	
188	2151.86	0.07	10.72	0.06	✓	
189	2161.85	0.07	10.70	0.07	✓	
190	2171.85	0.06	10.69	0.06	✓	
191	2181.84	0.07	10.67	0.07	✓	
192	2191.83	0.07	10.65	0.07	✓	
193	2201.81	0.07	10.63	0.07	✓	
194	2211.80	0.07	10.62	0.07	✓	
195	2221.78	0.07	10.60	0.07	✓	
196	2231.76	0.07	10.58	0.07	✓	
197	2241.73	0.07	10.55	0.07	✓	
198	2251.71	0.07	10.53	0.07	✓	
199	2261.68	0.07	10.51	0.07	✓	
200	2271.65	0.07	10.49	0.07	✓	
201	2281.61	0.07	10.46	0.07	✓	
202	2291.57	0.07	10.44	0.07	✓	
203	2301.53	0.08	10.41	0.07	✓	
204	2311.49	0.07	10.39	0.07	✓	
205	2321.45	0.07	10.36	0.07	✓	
206	2331.40	0.08	10.33	0.07	✓	
207	2341.35	0.08	10.31	0.06	✓	
208	2351.30	0.07	10.28	0.07	✓	
209	2361.24	0.07	10.25	0.07	✓	
210	2371.18	0.07	10.22	0.07	✓	
211	2381.12	0.07	10.19	0.07	✓	
212	2391.06	0.07	10.16	0.07	✓	
213	2400.99	0.07	10.12	0.07	✓	
214	2410.93	0.07	10.09	0.07	✓	
215	2420.85	0.07	10.06	0.07	✓	
216	2430.78	0.07	10.02	0.07	✓	
217	2440.71	0.06	9.99	0.07	✓	
218	2450.63	0.07	9.95	0.07		
219	2460.55	0.07	9.91	0.07		
220	2470.46	0.07	9.88	0.07		
221	2480.37	0.06	9.84	0.07		
222	2490.29	0.06	9.80	0.07		
223	2500.19	0.06	9.76	0.07		
224	2510.10	0.06	9.72	0.06		

A.4 Scripts for Manipulating GIS Data

Script 1 This script was used to download the DOQQs, contained in the band interleaved (.bil) files, from the Oklahoma State GIS Council's FTP site.

```
#!/bin/bash
NUM=36096
URL=ftp://okmaps.onenet.net/doq/${NUM}
for i in f43 f44 f33 f34 f23 f24 f13 f14 \
   f51 f42 f41 f32 f31 f22 f21 f12 \
   g53 g54 g43 g44 g33 g34 g23 g24 g13 \
   g61 g52 g51 g42 g41 g32 g31 g22 g21 \
   h64 h53 h54 h43 h44 h33 h34 h23 h24 \
   h61 h52 h51 h42 h41 h32 h31 h22 h21; do
   wget -nv ${URL}/${NUM}${i}.bil
   wget -nv ${URL}/${NUM}${i}.hdr.NAD27
   wget -nv ${URL}/${NUM}${i}.hdr.NAD83
   wget -nv ${URL}/${NUM}${i}.hdr
   wget -nv ${URL}/${NUM}${i}.RawHeader
done
```

Script 2 After the DOQQs were downloaded and the *r.in.doq* program was modified, this script was used to load the DOQQs into the GRASS GIS.

```
#!/bin/bash
for i in *.bil; do
   MAPNAME=‘echo ${i} | awk -F. ’{print substr($1,6,3);}’‘
   echo "-----------------------------------------------------------"
   echo "Now loading ${i} into ${MAPNAME}"
   r.in.doq -s input=${i} output=${MAPNAME}
done
```

Script 3 Once the DOQQs were loaded into the GIS, this script was used to generate vector outlines of each DOQQ.

```
for i in *.bil; do
   MAPNAME=‘echo ${i} | awk -F. ’{print substr($1,6,3);}’‘
   echo "------------------------------------------------------------"
   echo "Creating vector outline for ${MAPNAME}"
   g.region rast=${MAPNAME}
   r.mapcalc << EOF
junk = if(${MAPNAME})
exit
EOF
   r.poly input=junk output=${MAPNAME}
done
g.remove rast=junk
```

A.5 Brine Scar Prediction Locations

Table A.2: Brine scar prediction locations. Coordinates are in UTM meters.

Region Name	North	South	West	East	In TPP
arm ef12 site	4080457	4079802	728826	729581	✓
tpp hq	4081037	4080412	729394	730266	✓
bigscar	4077178	4076245	724771	726073	✓
lower disturbance	4077076	4076249	723980	725094	
bottom tanks	4074784	4073758	727270	728272	
upper well N scar	4082139	4081823	725474	725877	
upper well tank farm	4082143	4081750	725069	725491	
upper well roadside pump	4081784	4081468	724425	724794	
upper well main road	4081220	4080991	724261	724505	✓
upper well big scar	4081502	4080575	724974	726160	✓
left upper well scar	4081279	4080174	721999	723282	
tanks	4078299	4078035	724017	724328	
roadside tanks	4078034	4077708	724351	724681	✓
middle disturbance	4078811	4078152	724671	725445	✓
erosion7	4079561	4079018	722511	723132	
ranch	4079069	4078741	722399	722774	
ranch2	4078861	4078236	723283	723927	
intersection5	4077417	4076653	727945	728841	✓
intersection6	4080594	4079353	727017	728331	✓
intersection7	4082655	4082081	729618	730352	✓
erosion2	4086016	4085139	715620	716627	
intersection2	4085907	4085069	718215	719077	
intersection8	4083304	4082813	723189	723748	
junction2	4083343	4082894	721540	722284	
farm3	4080593	4079967	714892	715610	
smallfacility	4082053	4081185	716794	717501	
disturb7	4080862	4080325	734466	735116	✓
oilarea	4079616	4078333	732806	734381	✓
hilltop	4075730	4075269	739008	739543	
bigfields	4074787	4071758	737456	740676	
strip	4070699	4069691	731613	732654	
farm2	4069399	4068814	736327	736978	✓
compound3	4072503	4071888	732629	733349	✓
roadside2	4073240	4072392	732177	733168	✓
intersection9	4074316	4073869	732077	732623	✓

Table A.2: Continued

Region Name	North	South	West	East	In TPP
road	4075748	4075375	730539	730936	✓
tank cluster	4072708	4071899	748351	749371	
tank group	4070577	4069723	747114	748326	
waterfront	4074425	4073798	746295	747053	
facility intersection	4071885	4070849	747785	748975	
compound4	4071777	4071135	744236	745051	
ag intersection	4069534	4068835	740721	741504	
farm4	4073316	4072561	739299	740336	
disturb1 new	4065613	4065021	741102	741816	
disturb2 new	4064436	4063781	742817	743516	
disturb3	4073055	4071282	742438	744379	
tank station	4066375	4065580	740559	741590	
agriculture	4069429	4069207	741464	741712	
agriculture2	4070348	4069398	739793	741021	
erosion1	4069268	4069006	742301	742564	
roadshoulder1	4070614	4070180	742831	743423	
junction	4066463	4065306	739579	740700	
intersection1	4069248	4069060	741879	742036	
farm1	4068800	4068374	742374	742867	
compound2	4068239	4067729	743202	743819	
disturb4	4084211	4083469	737439	738304	✓
disturb5	4087212	4086393	741585	742389	
disturb6	4088447	4087805	738258	739070	
disturb10	4084719	4084093	741978	742628	
wells in 66	4085117	4082454	744878	747323	
group of spills	4081658	4080384	742515	744164	
ranch3	4079776	4078707	741925	743071	
erosion8	4079232	4077756	740196	741904	
runoff	4088170	4087189	731306	732435	
intersection10	4086573	4085695	728685	729705	
disturb11	4085081	4084347	726422	727443	
disturb12	4088535	4087626	732628	733819	
creek to creek spill	4085859	4085322	731627	732325	✓
tank wash	4086364	4085766	736553	737217	
puka	4085179	4084211	735864	736815	
lower stream	4086072	4084576	734327	735627	
disturb13	4086282	4085819	733103	733629	
erosion6	4089796	4089514	730543	730812	

Table A.2: Continued

Region Name	North	South	West	East	In TPP
spill drainage	4090704	4090318	735048	735468	
by river spot	4091284	4090800	734691	735403	
compound6	4091344	4090610	732890	733835	
jackpump1 new	4090896	4090046	732331	733626	
sandbars new	4092861	4090687	729913	732717	
erosion3	4090892	4090233	724197	724871	
erosion4	4093986	4093452	724901	725472	
erosion5	4094721	4094370	725849	726163	
erosion9	4090965	4090100	724923	725901	
intersection3	4088521	4087960	721295	721913	
intersection4	4089427	4088675	728491	729273	
bridge and farm river	4091481	4090606	729073	730159	
two spills	4087237	4086373	726797	727843	
set of tanks	4085036	4084064	725756	727031	

A.6 Regions Containing Brine Scar Predictions

Figure A.1: Region: arm_ef12_site, Scene: r05sc05. TPP bison paddocks and the U.S. Department of Energy's Atmospheric Radiation Measurement (ARM) Extended Facility #12 located near the Tallgrass Prairie Preserve Headquarters.

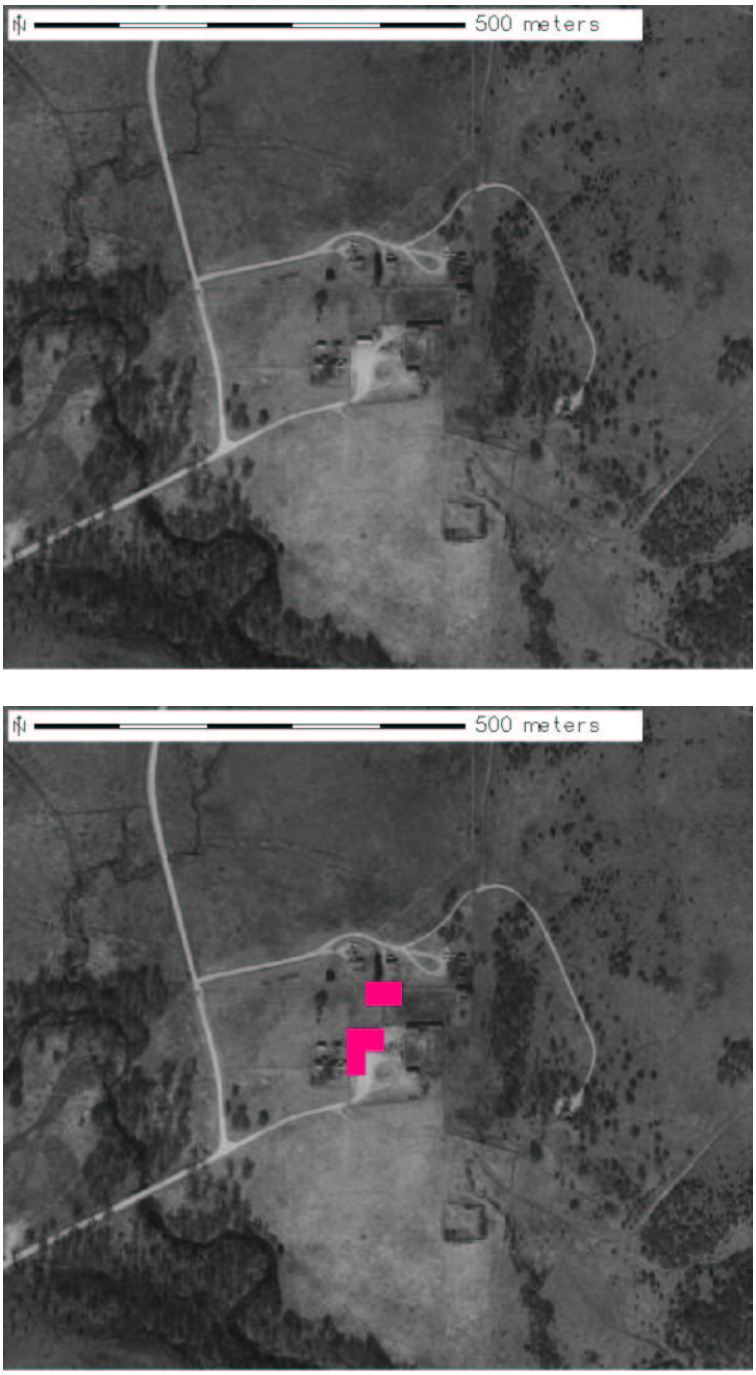


Figure A.2: Region: tpp_hq, Scene: r05sc05. Tallgrass Prairie Preserve (TPP) Headquarters.

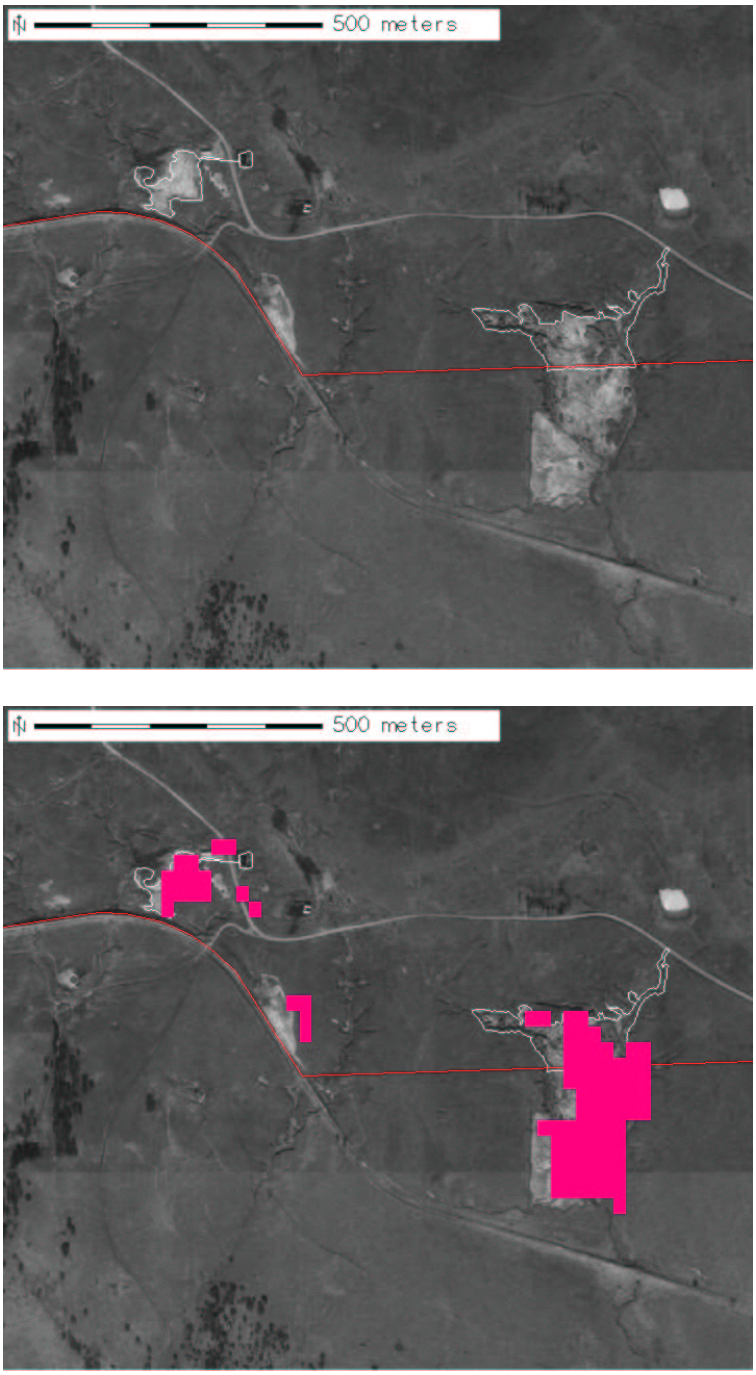


Figure A.3: Region: bigscar, Scene: r05sc05. Two of the largest known brine scars in the entire study area the boundaries of which were recorded using a GPS.

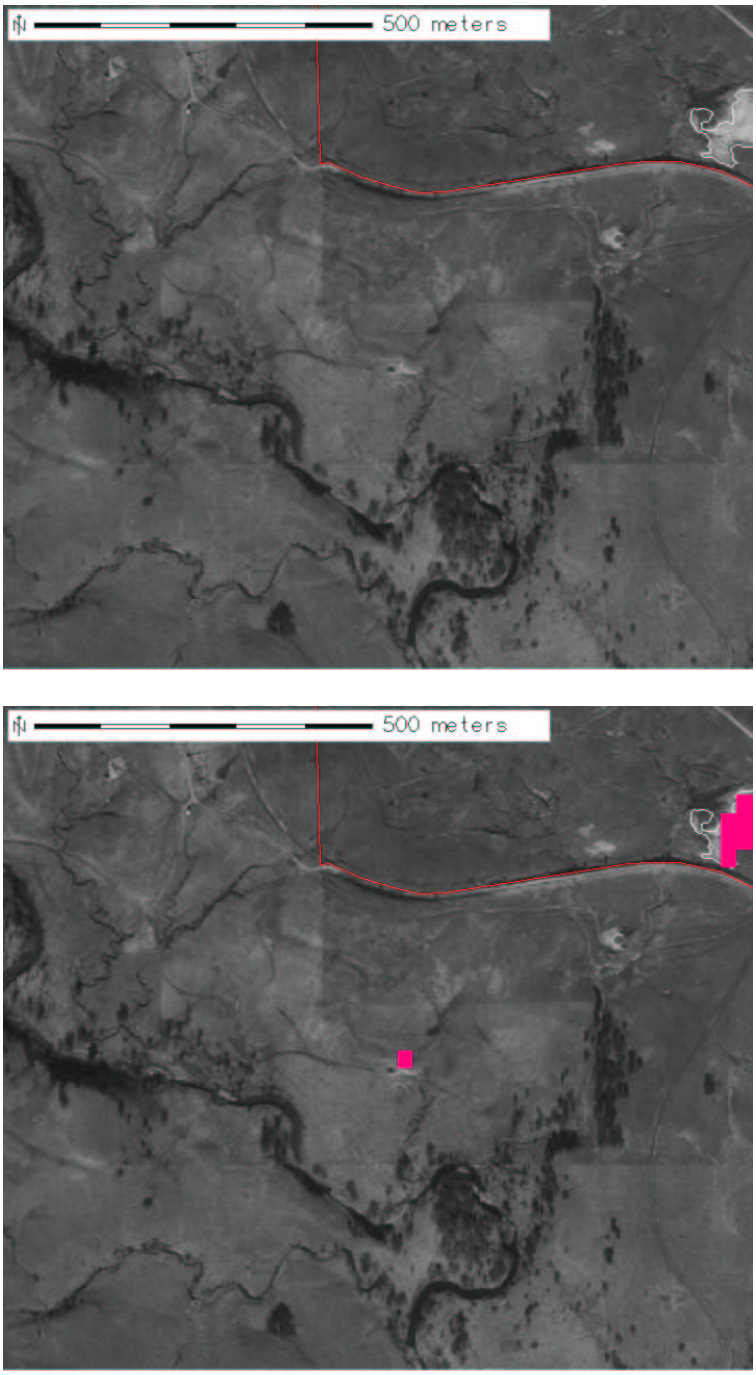


Figure A.4: Region: lower_disturbance, Scene: r05sc05. Single brine scar associated with a well near the two large brine scars.

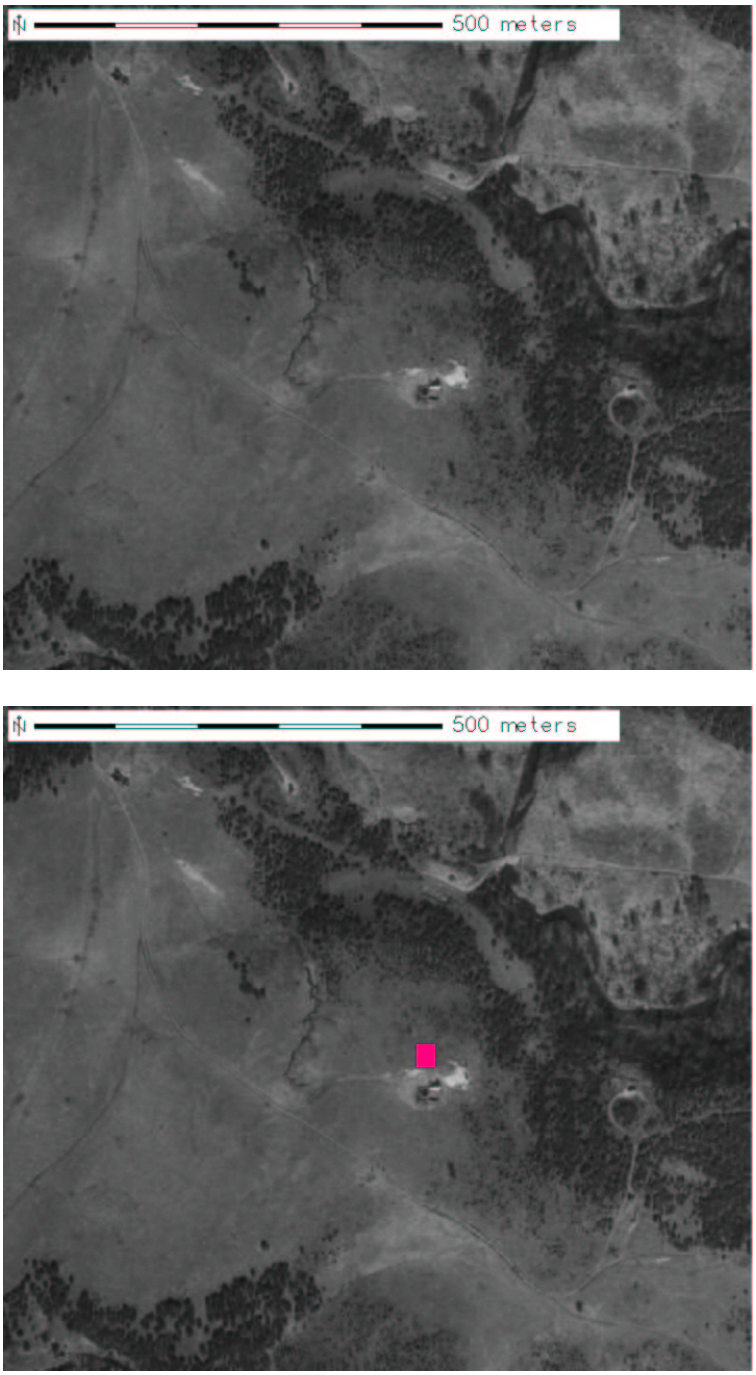


Figure A.5: Region: bottom_tanks, Scene: r05sc05. Brine scar near collection of tanks south of two large brine scars.

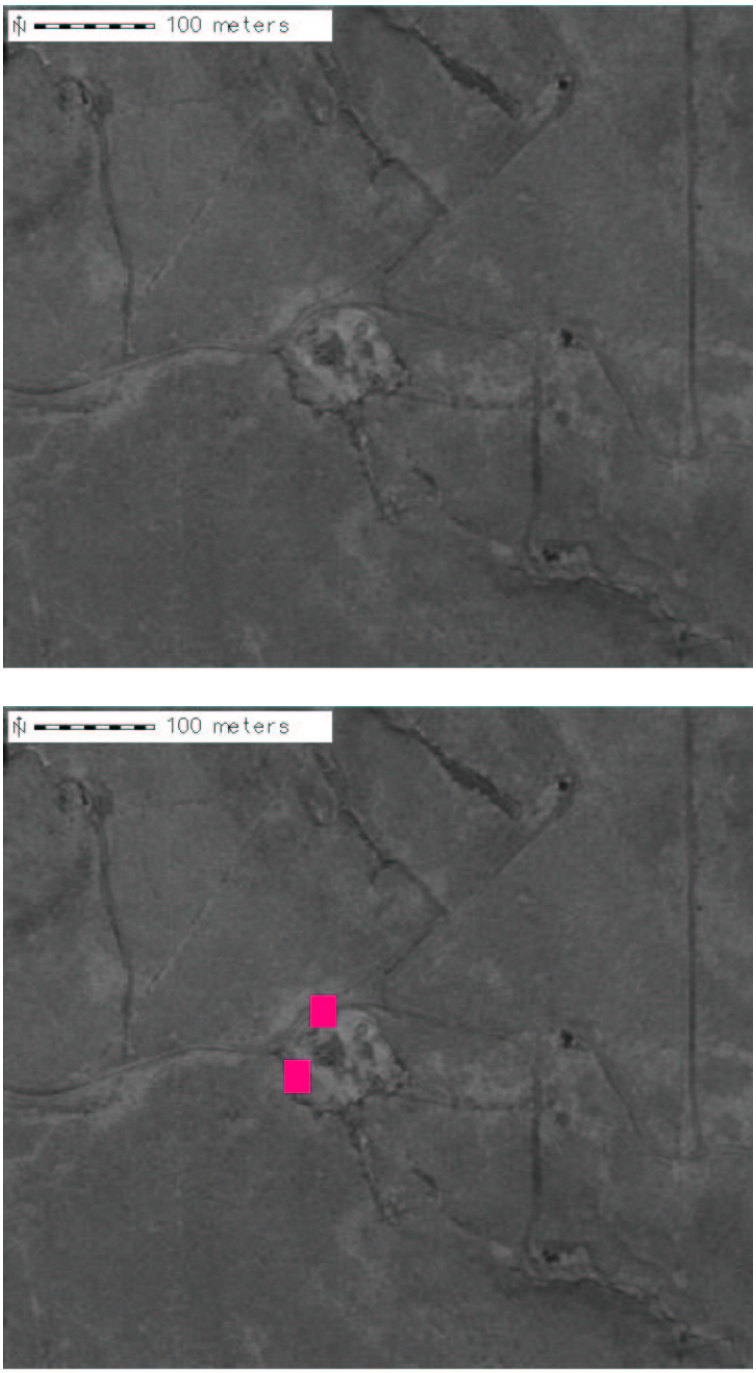


Figure A.6: Region: upper_well_N_scar, Scene: r05sc05. Medium-sized brine scar just north of the TPP boundary.

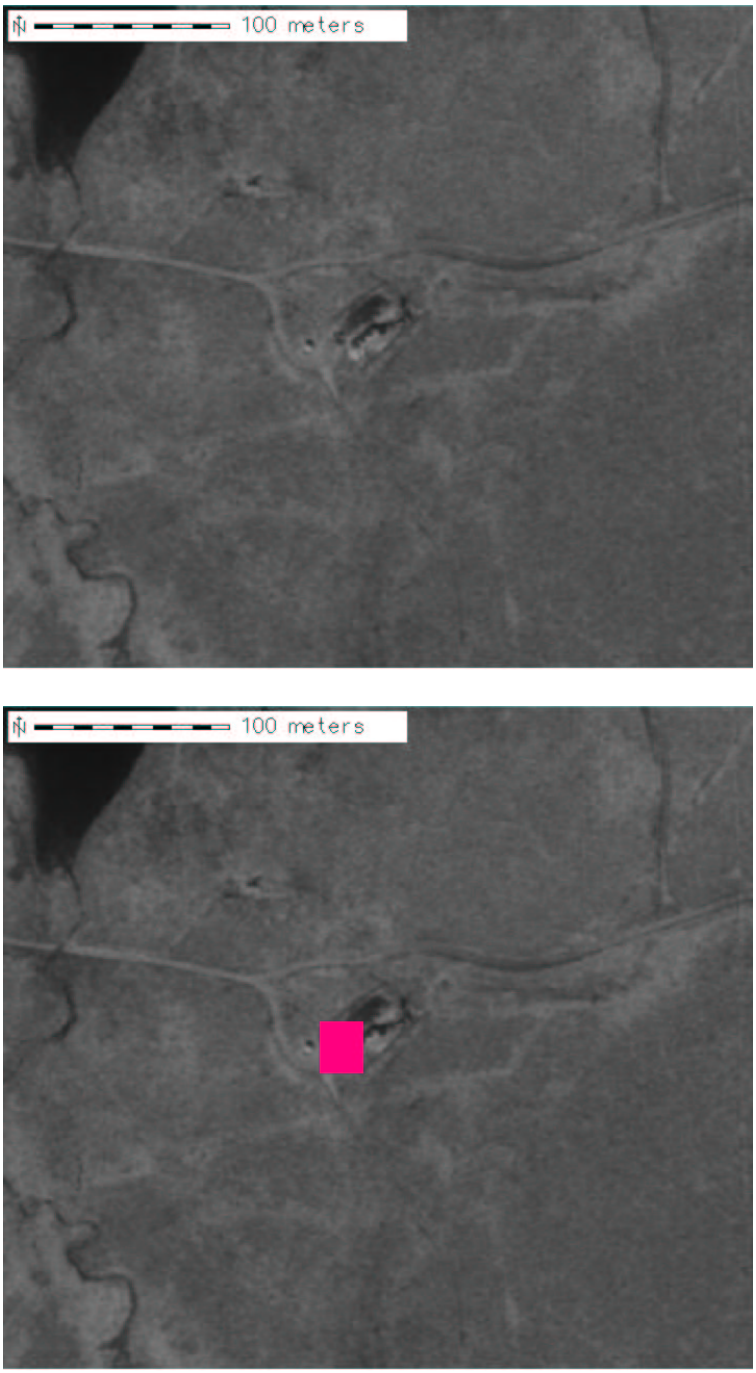


Figure A.7: Region: upper_well_tank_farm, Scene: r05sc05. Tanks just north of the TPP boundary.

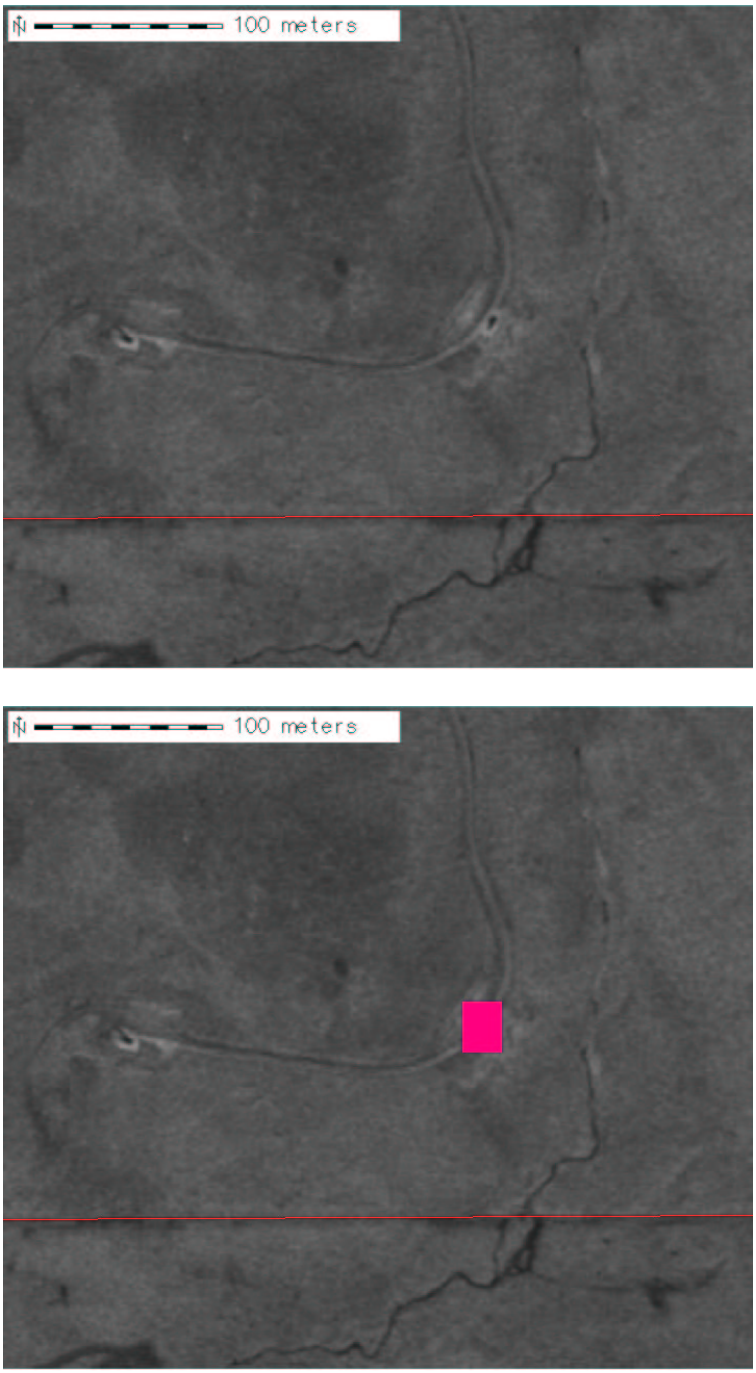


Figure A.8: Region: upper_well_roadside_pump, Scene: r05sc05. Road-side jackpump just north of TPP boundary.

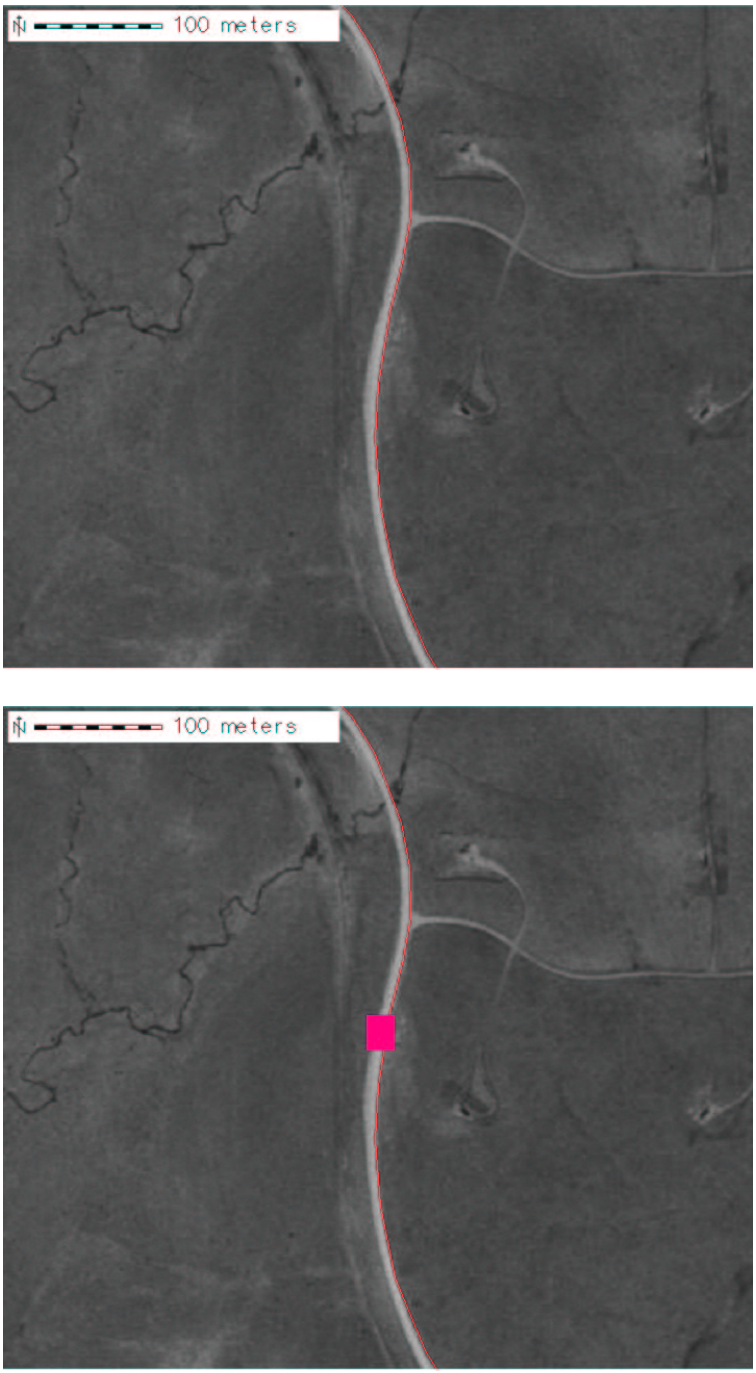


Figure A.9: Region: upper_well_main_road, Scene: r05sc05. Area near main road but just northwest of an active well just within the TPP boundary.

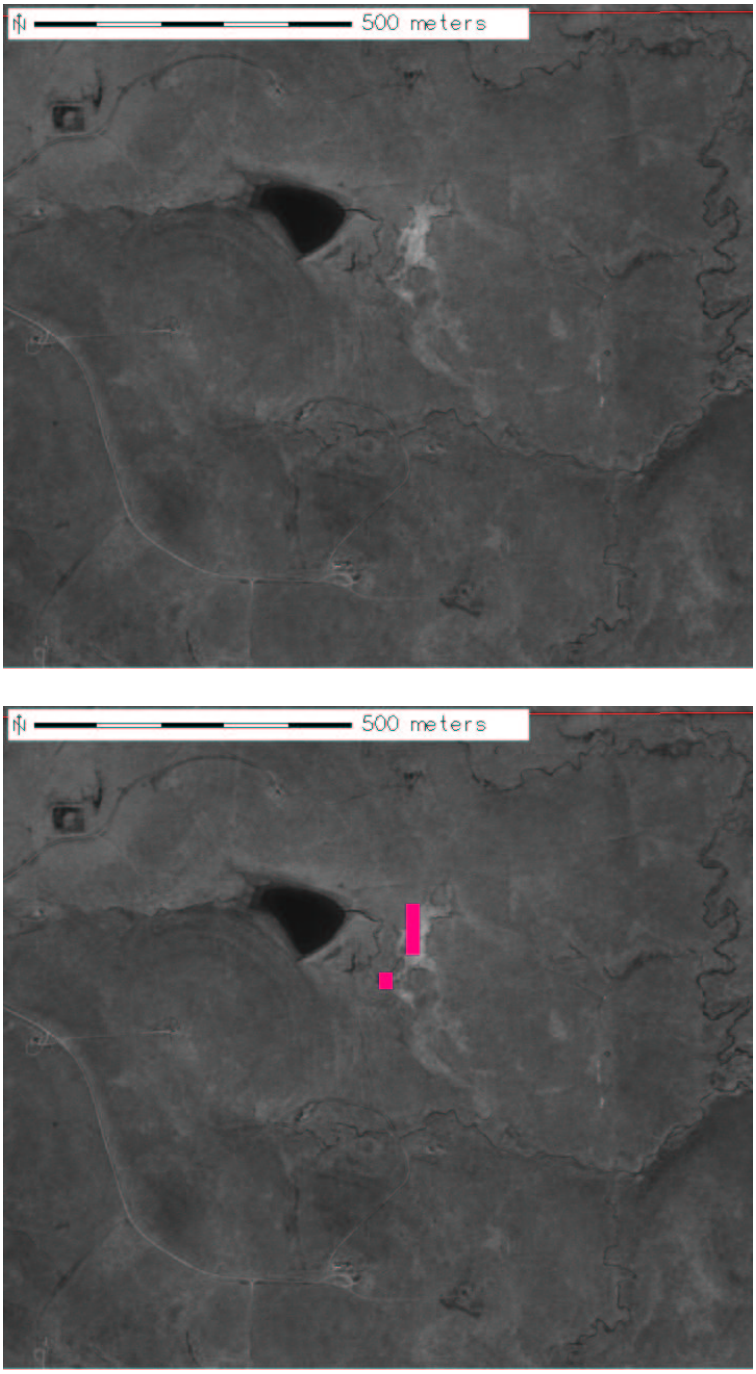


Figure A.10: Region: upper_well_big_scar, Scene: r05sc05. Old, large brine scar between a number of wells.

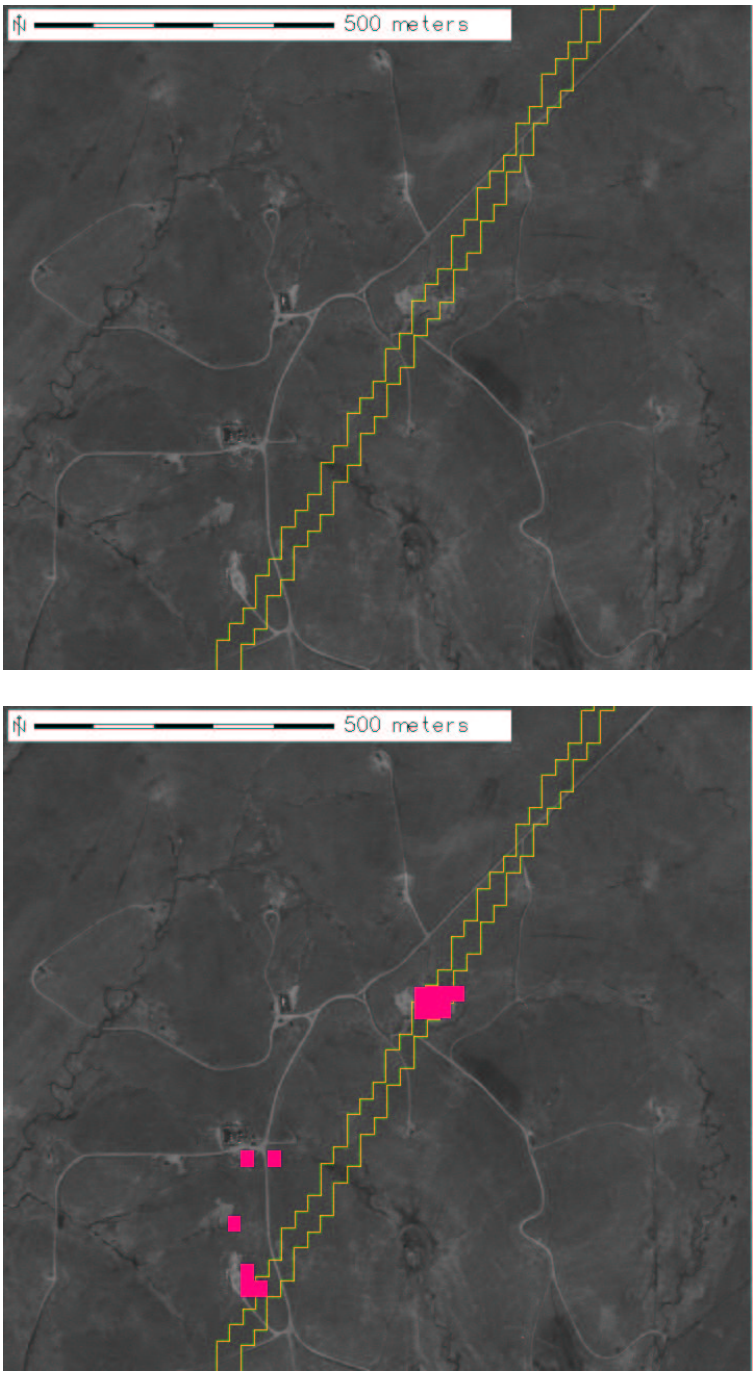


Figure A.11: Region: left_upper_well_scar, Scene: r05sc05 and r05sc06. Multiple brine scars among a collection of wells and tanks.

Figure A.12: Region: tanks, Scene: r05sc05. Collection of tanks.

Figure A.13: Region: roadside_tanks, Scene: r05sc05. Road-side tanks just within TPP western boundary.

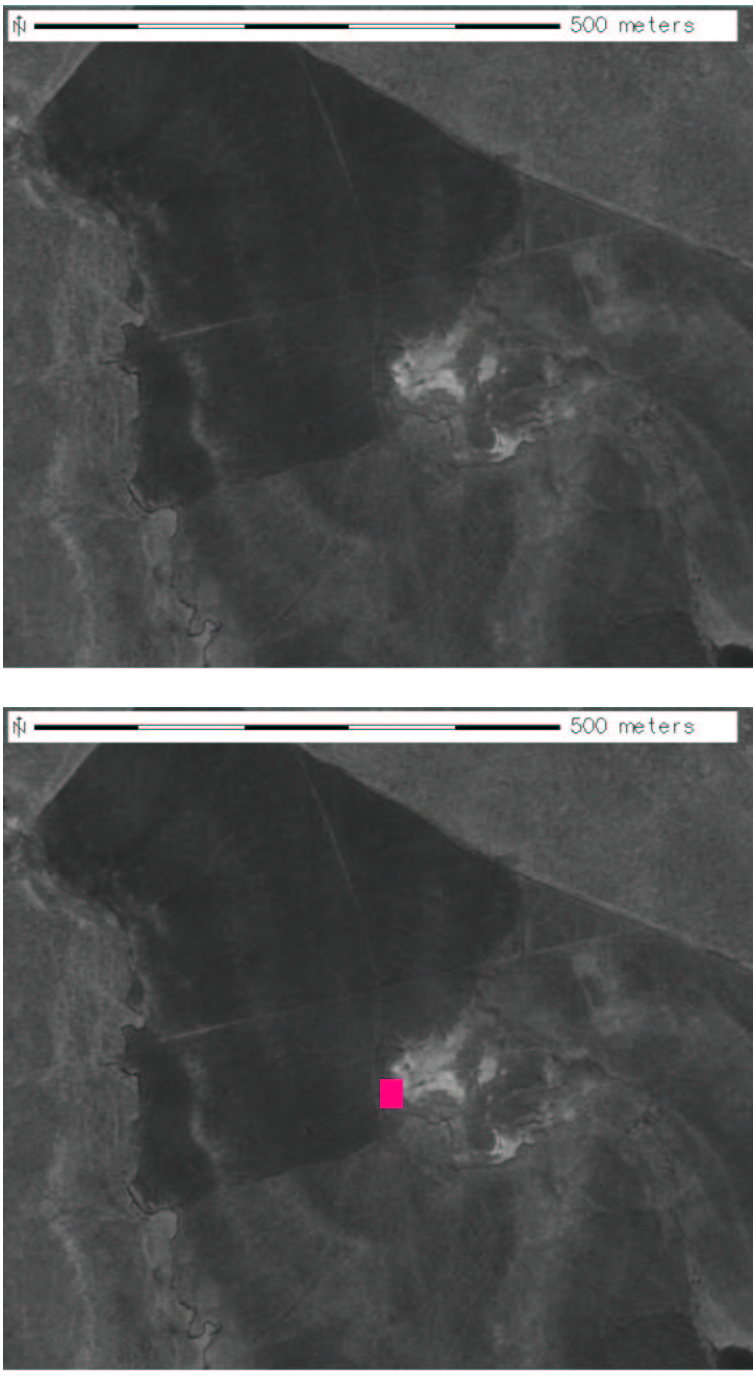


Figure A.14: Region: middle_disturbance, Scene: r05sc05. Old, large brine scar.

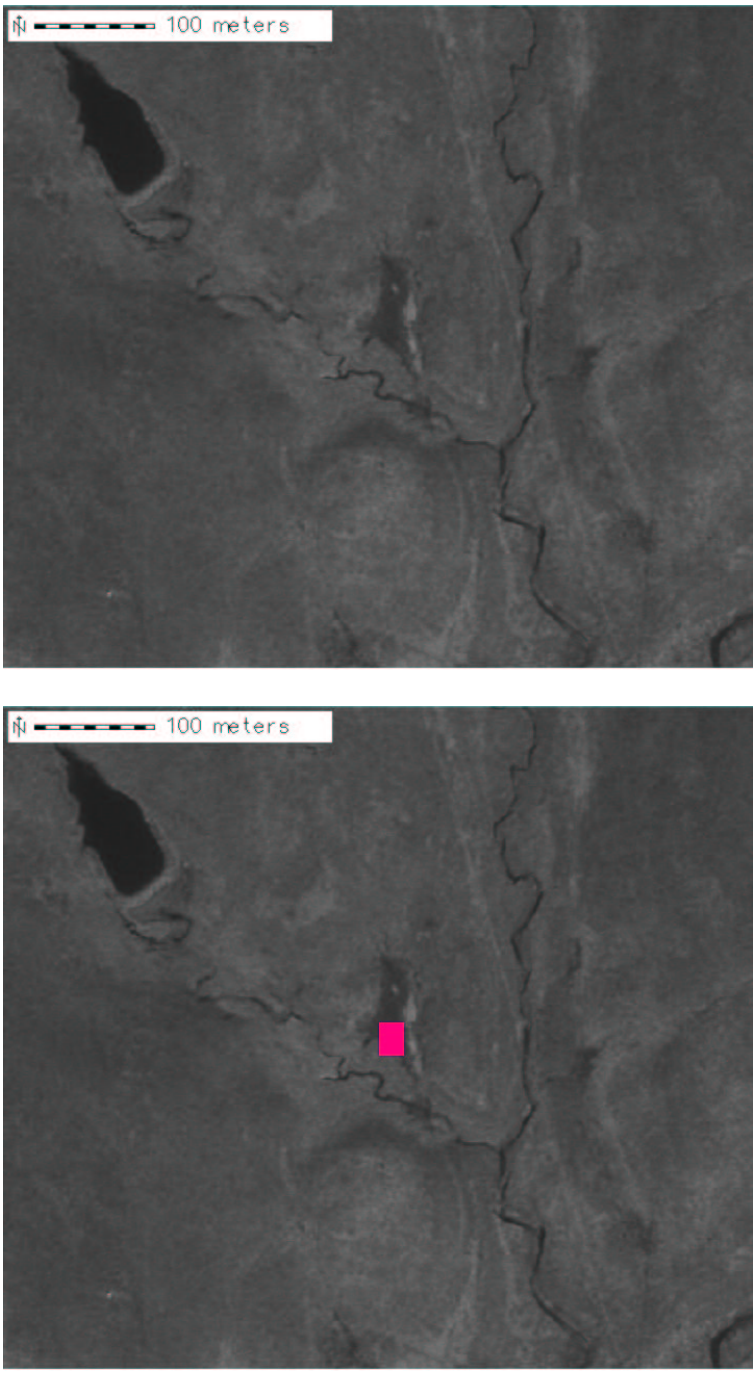


Figure A.15: Region: erosion7, Scene: r05sc05. Eroded area with no obvious wells or tanks nearby.

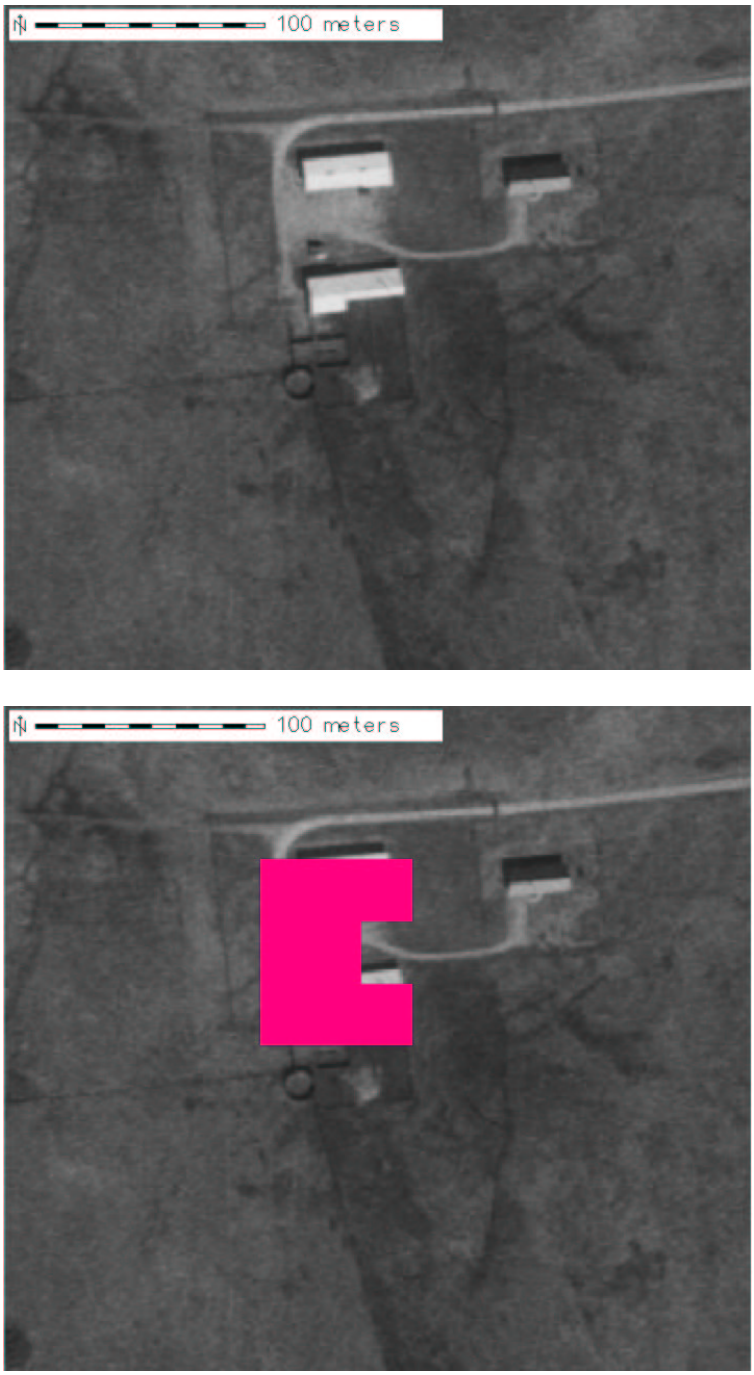


Figure A.16: Region: ranch, Scene: r05sc05. A ranch with two large buildings and a parking area.

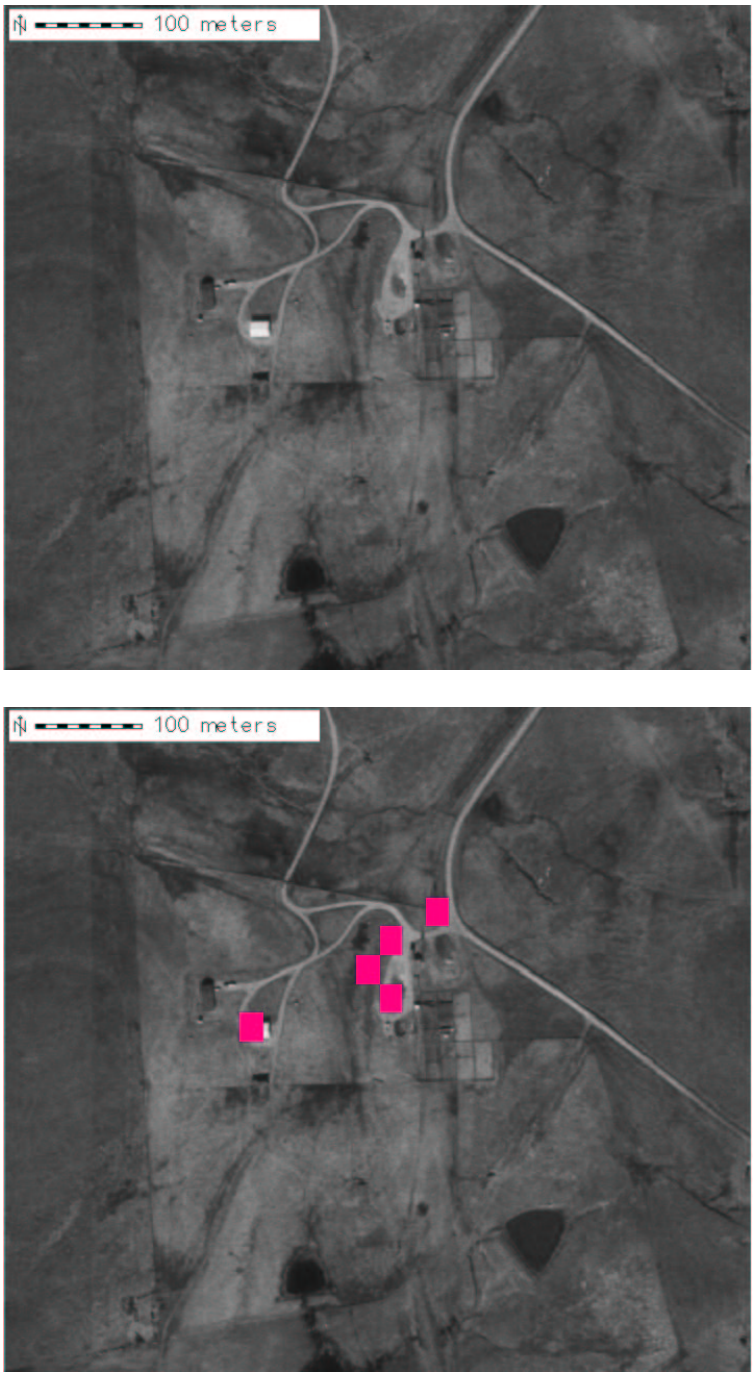


Figure A.17: Region: ranch2, Scene: r05sc05. Big ranch with large driveways.

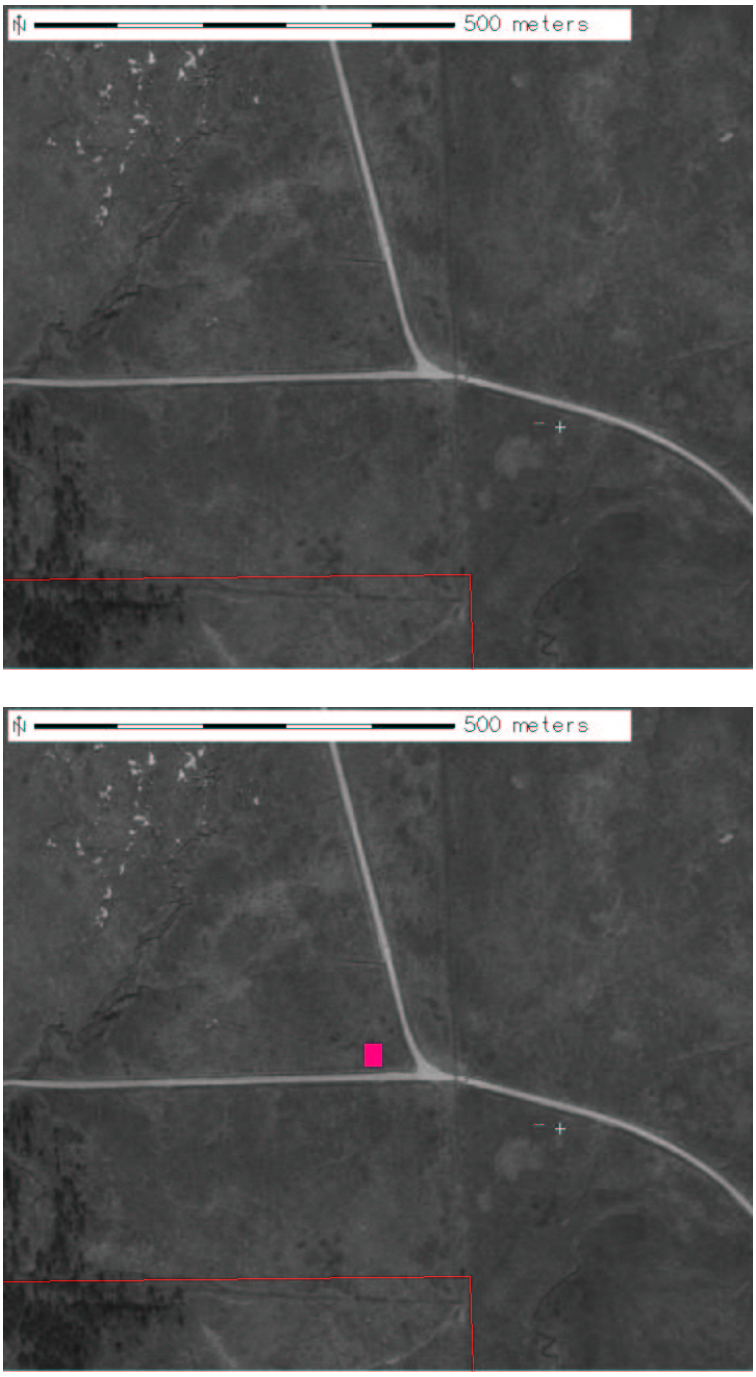


Figure A.18: Region: intersection5, Scene: r05sc05. Highway intersection.

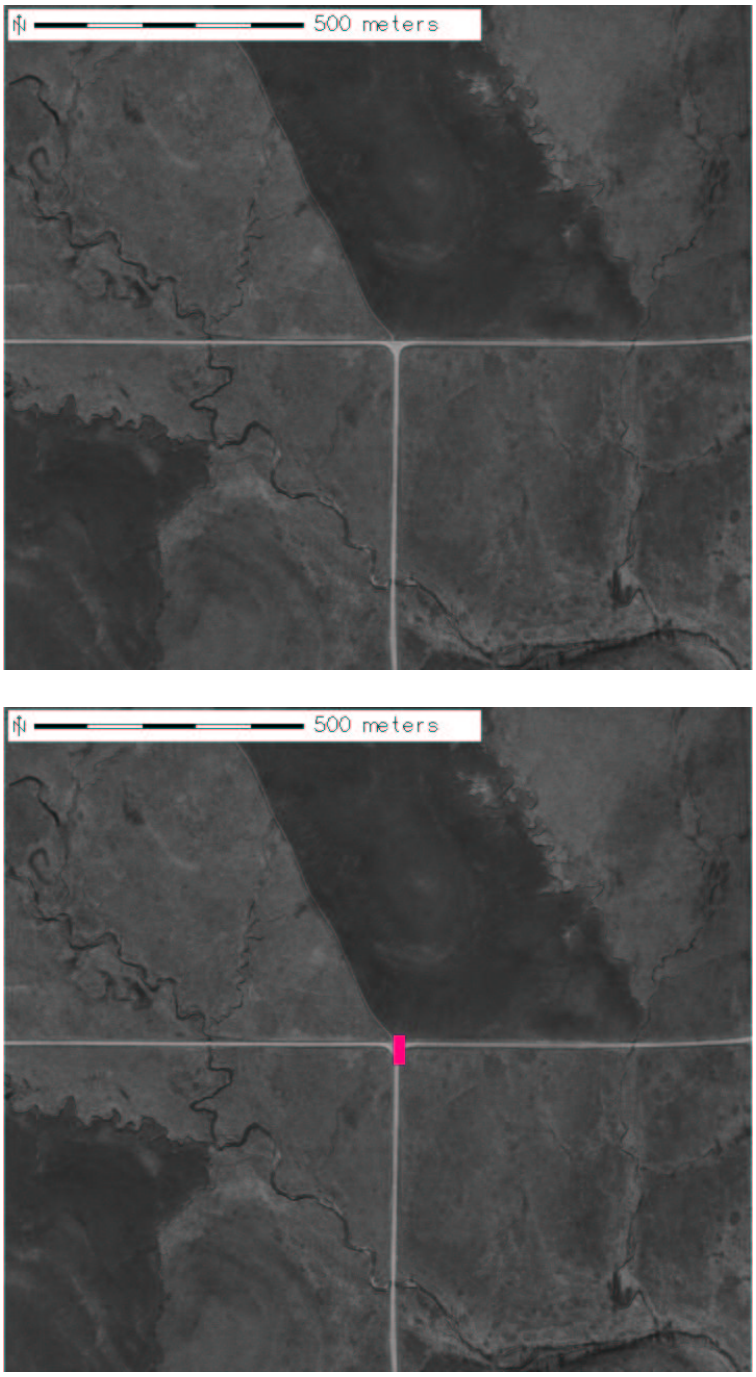


Figure A.19: Region: intersection6, Scene: r05sc05. Highway intersection.

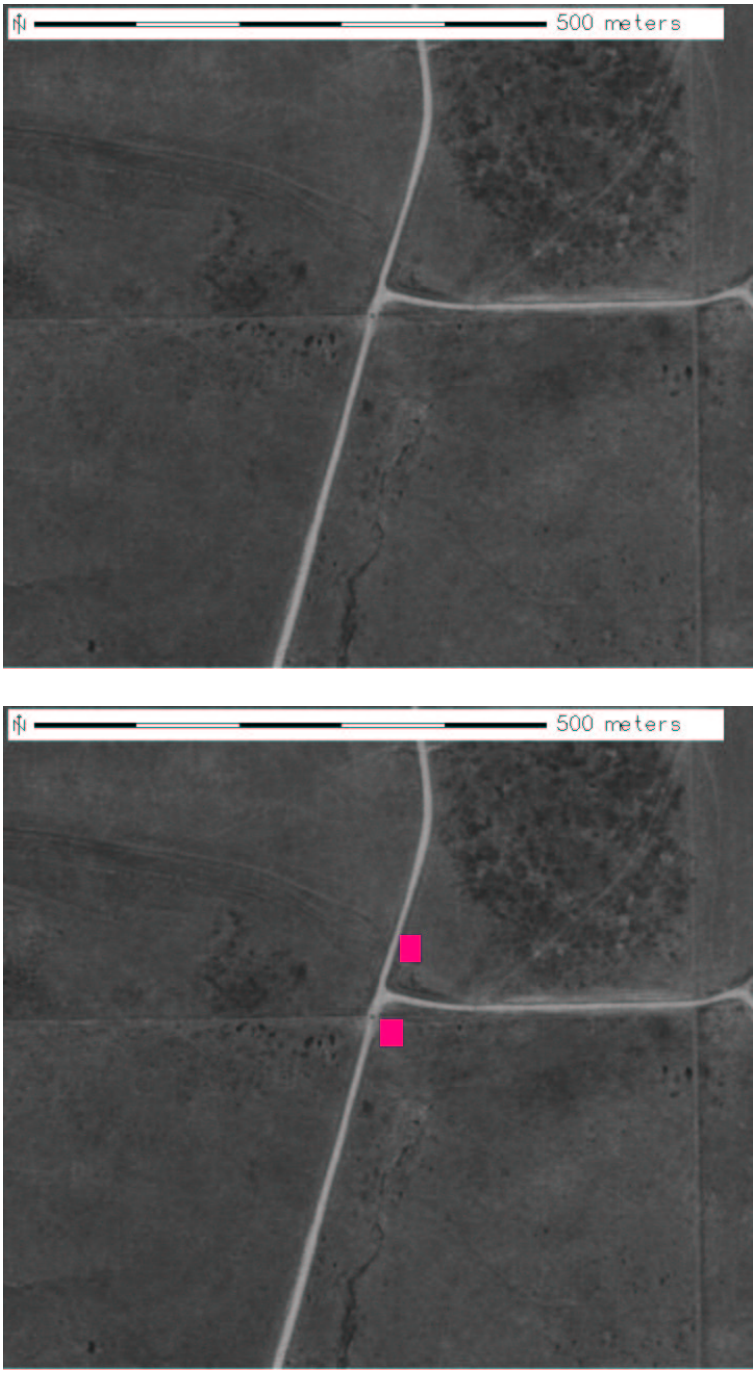


Figure A.20: Region: intersection7, Scene: r05sc05 and r06sc05. Highway intersection (same pixel in both scenes).

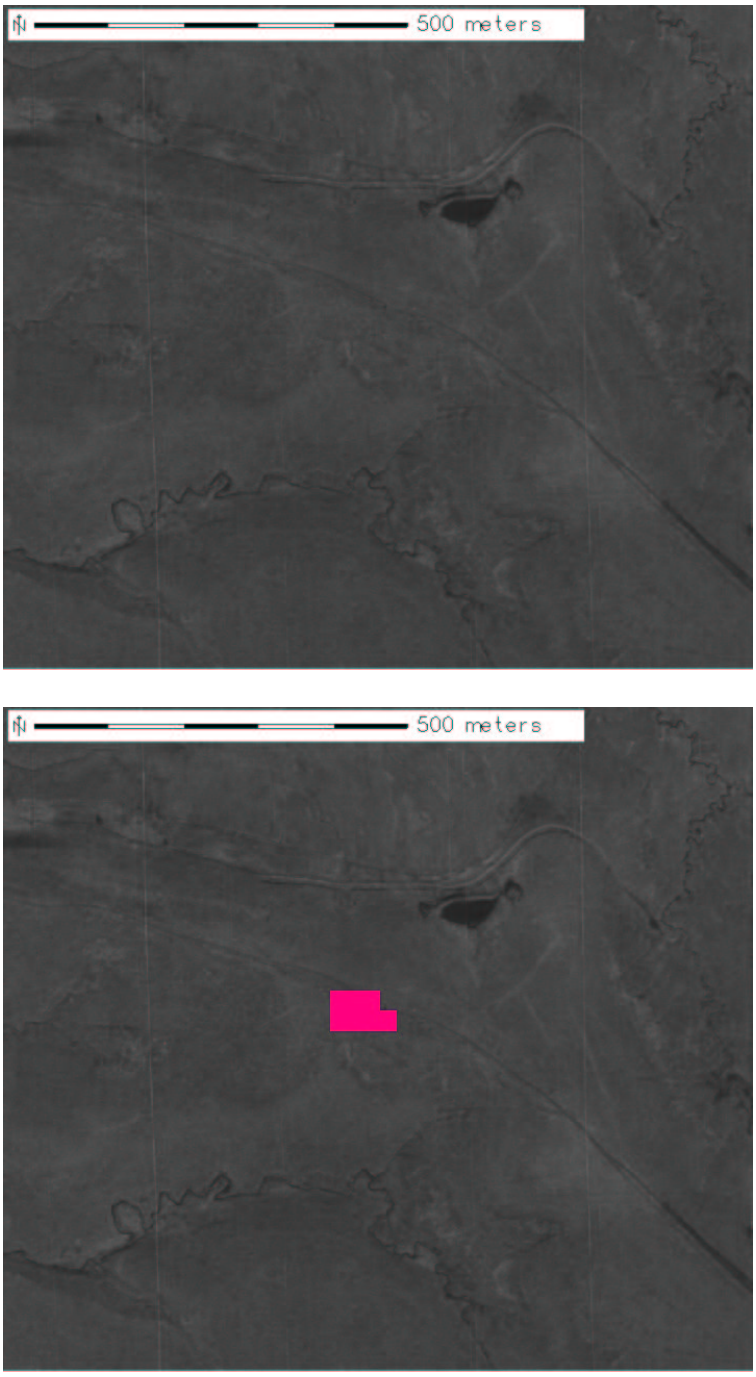


Figure A.21: Region: erosion2, Scene: r05sc06. Eroded area on a back road.

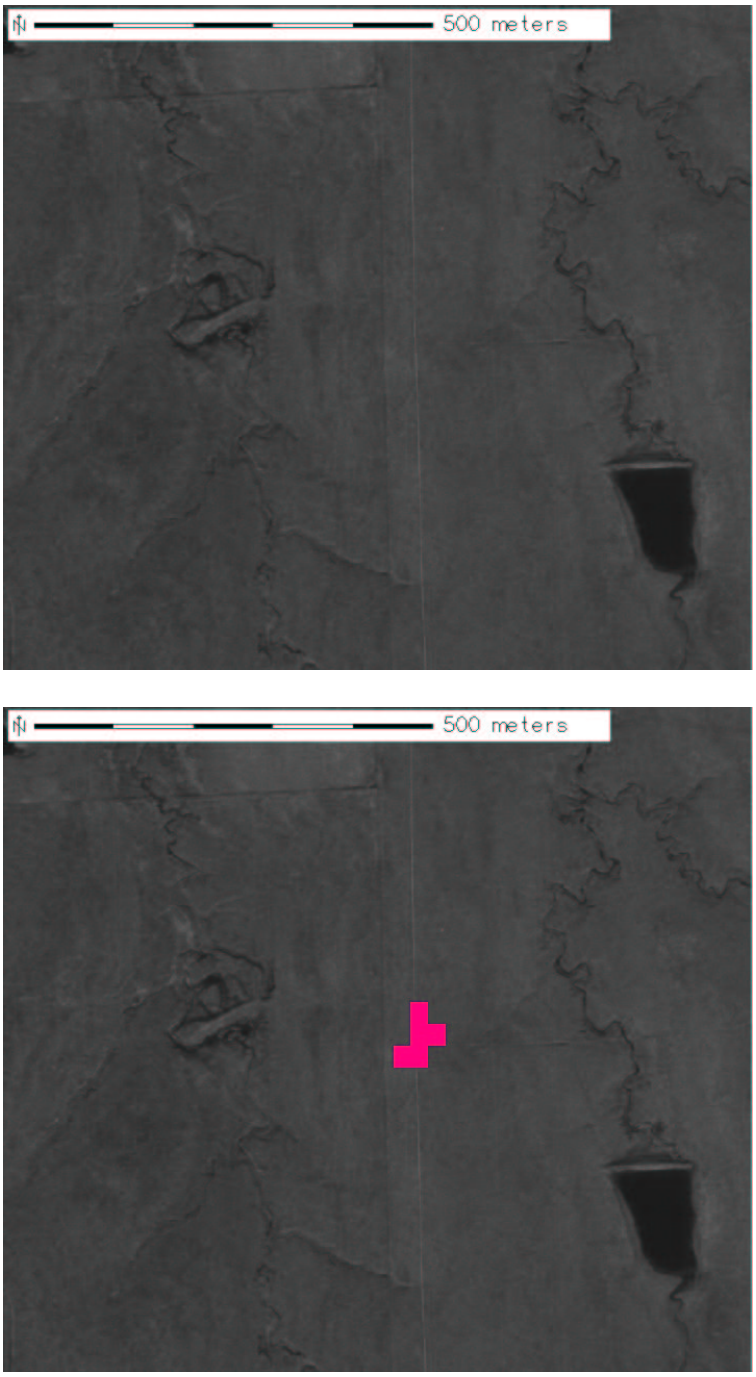


Figure A.22: Region: intersection2, Scene: r05sc06. Intersection of roads.

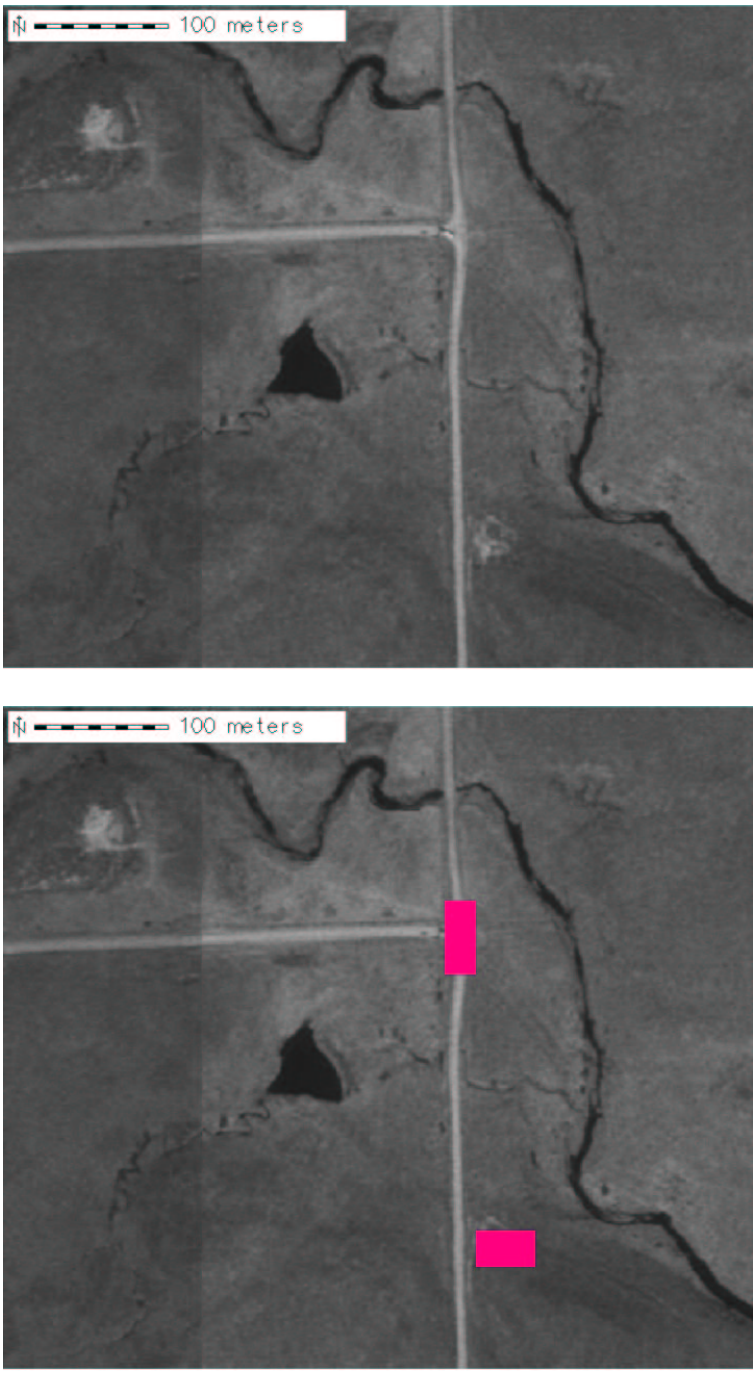


Figure A.23: Region: intersection8, Scene: r05sc06. Highway intersection and eroded area nearby.

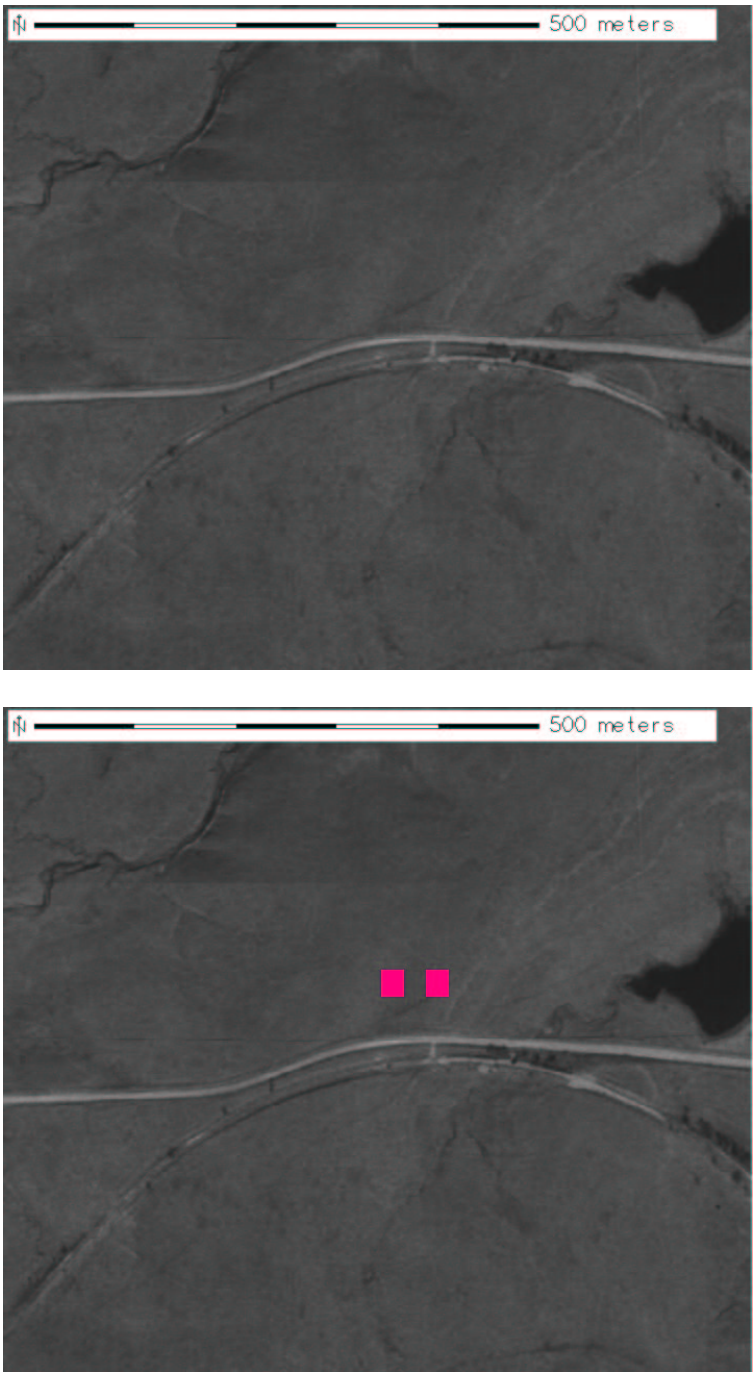


Figure A.24: Region: junction2, Scene: r05sc06. Highway interchange.

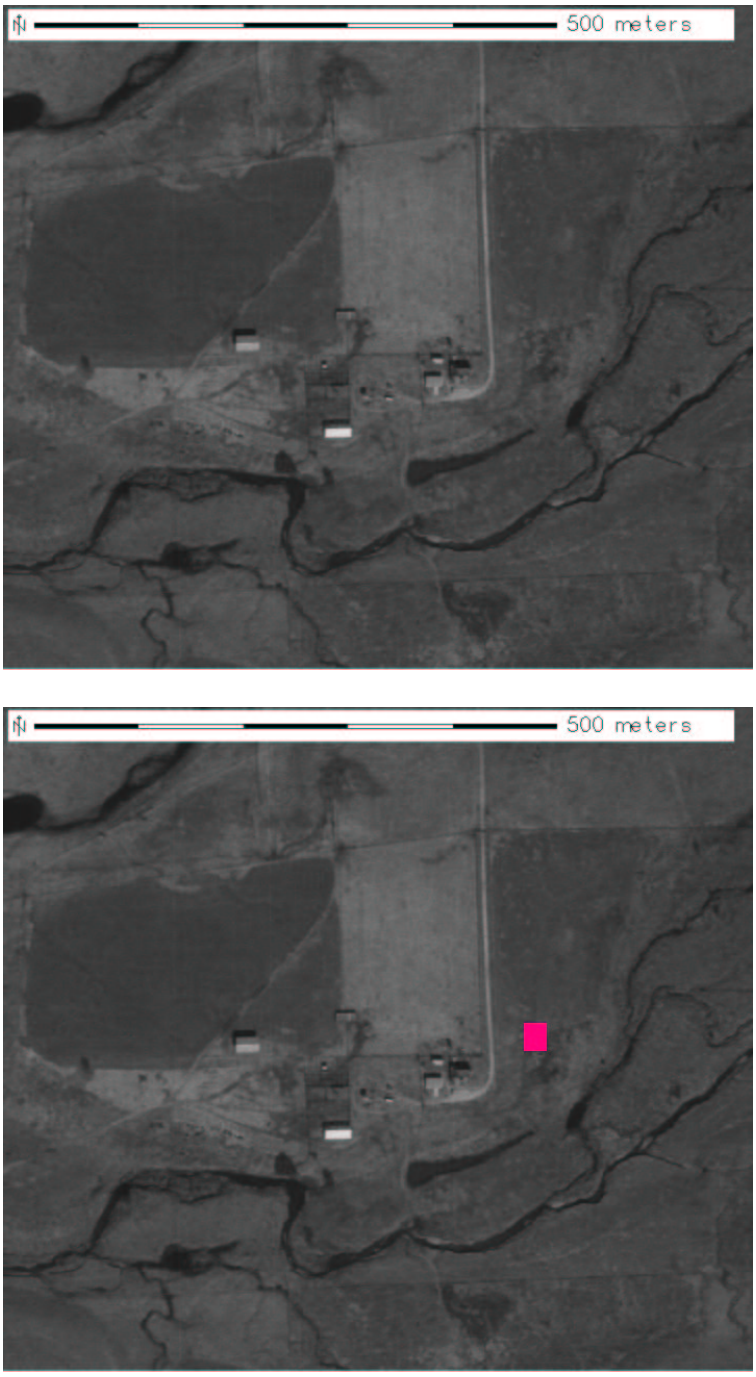


Figure A.25: Region: farm3, Scene: r05sc06. Ranch area with large turnaround.

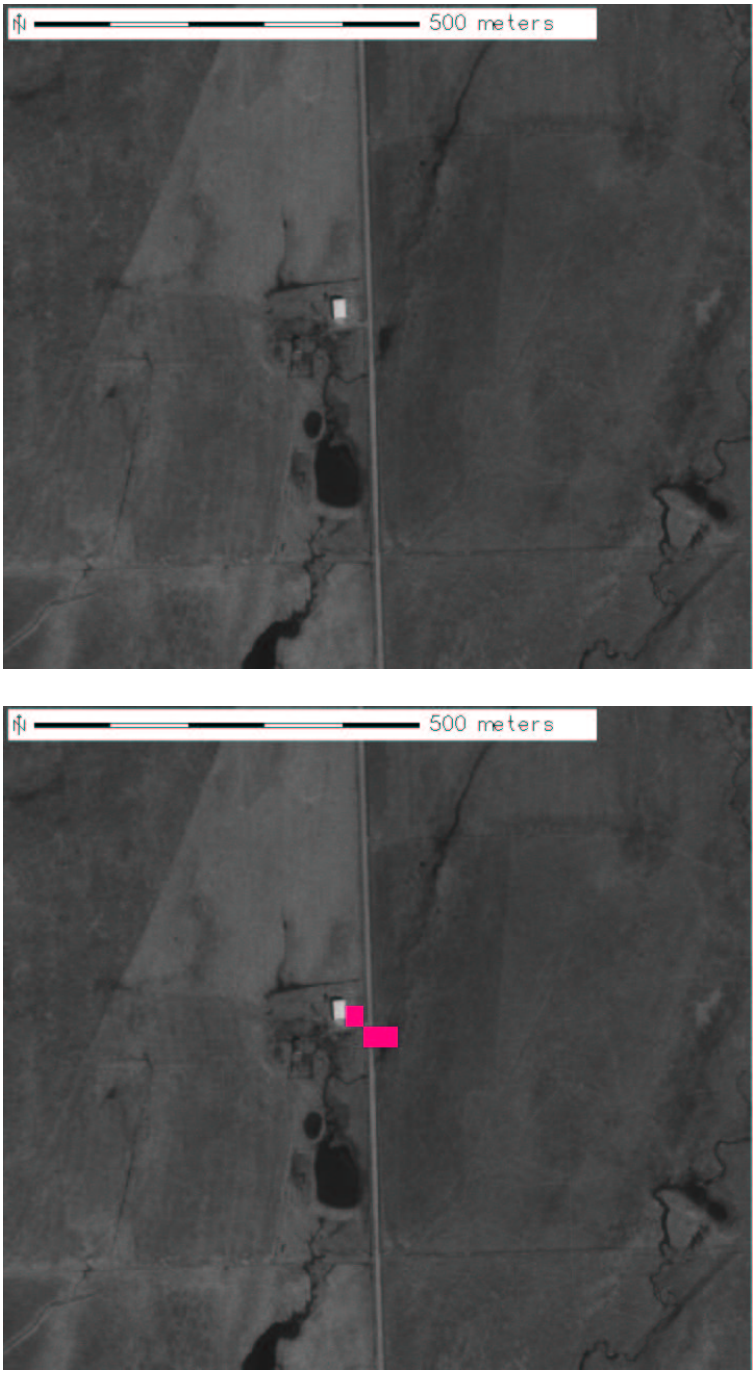


Figure A.26: Region: smallfacility, Scene: r05sc06. Road-side facility with parking area.

Figure A.27: Region: disturb7, Scene: r05sc04 and r06sc06. Brine scar around a well (same pixel in two scenes).

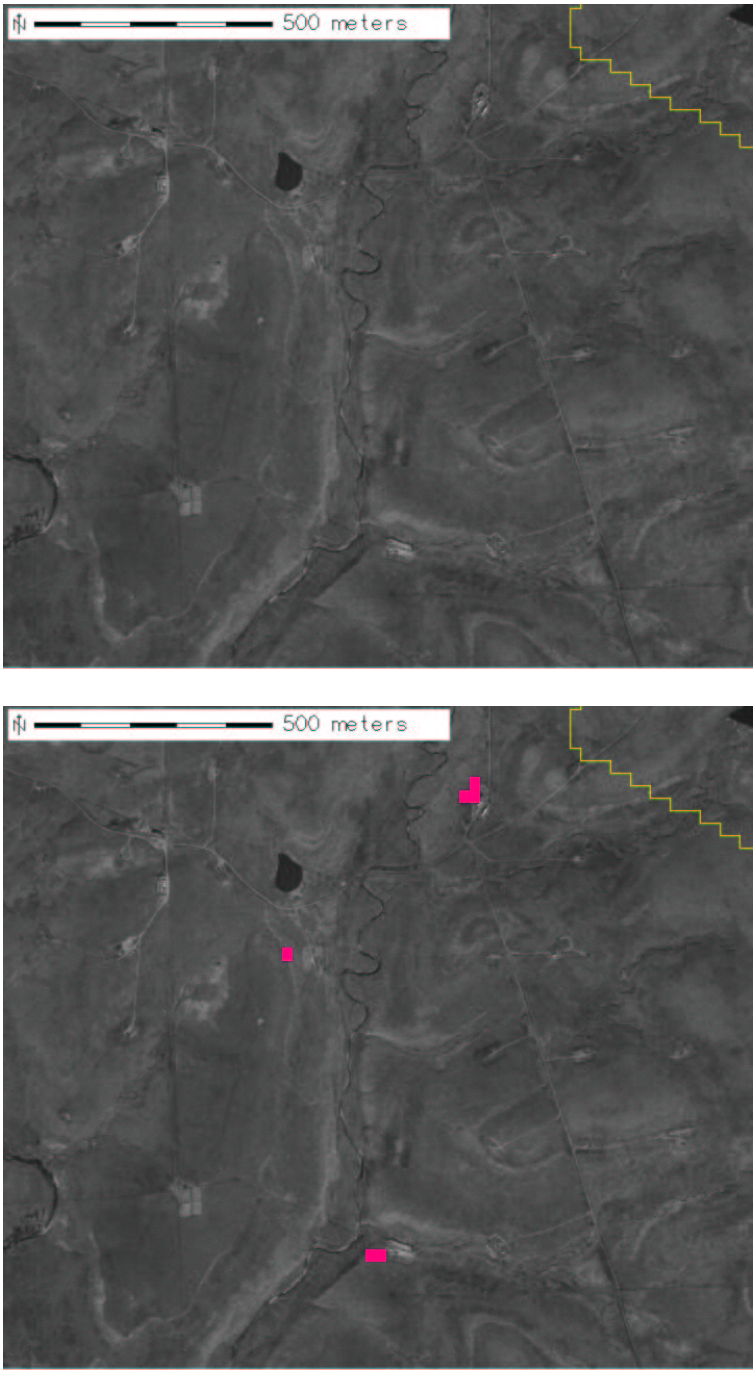


Figure A.28: Region: oilarea, Scene: r05sc04. Brine scars at two wells and a tank area.

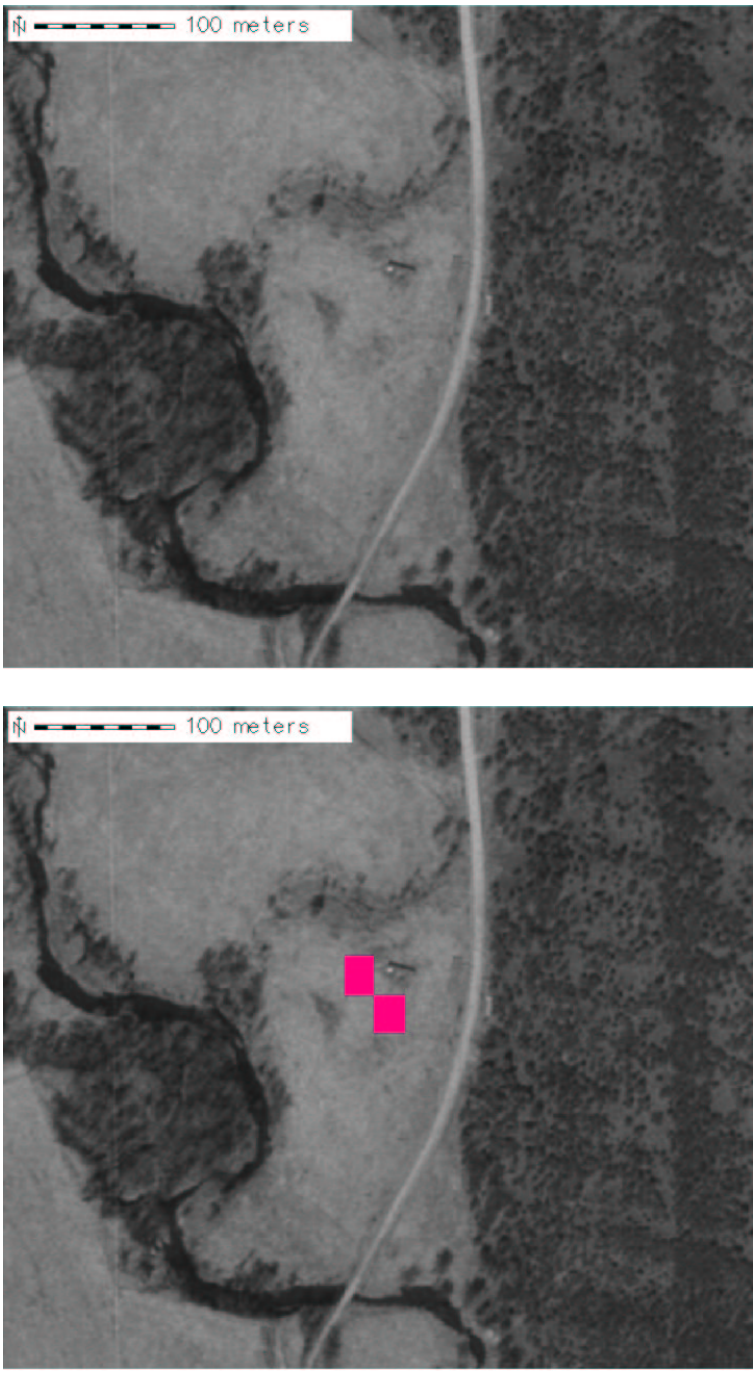


Figure A.29: Region: hilltop, Scene: r05sc04. Object on riverbank beside roadway near agricultural area.

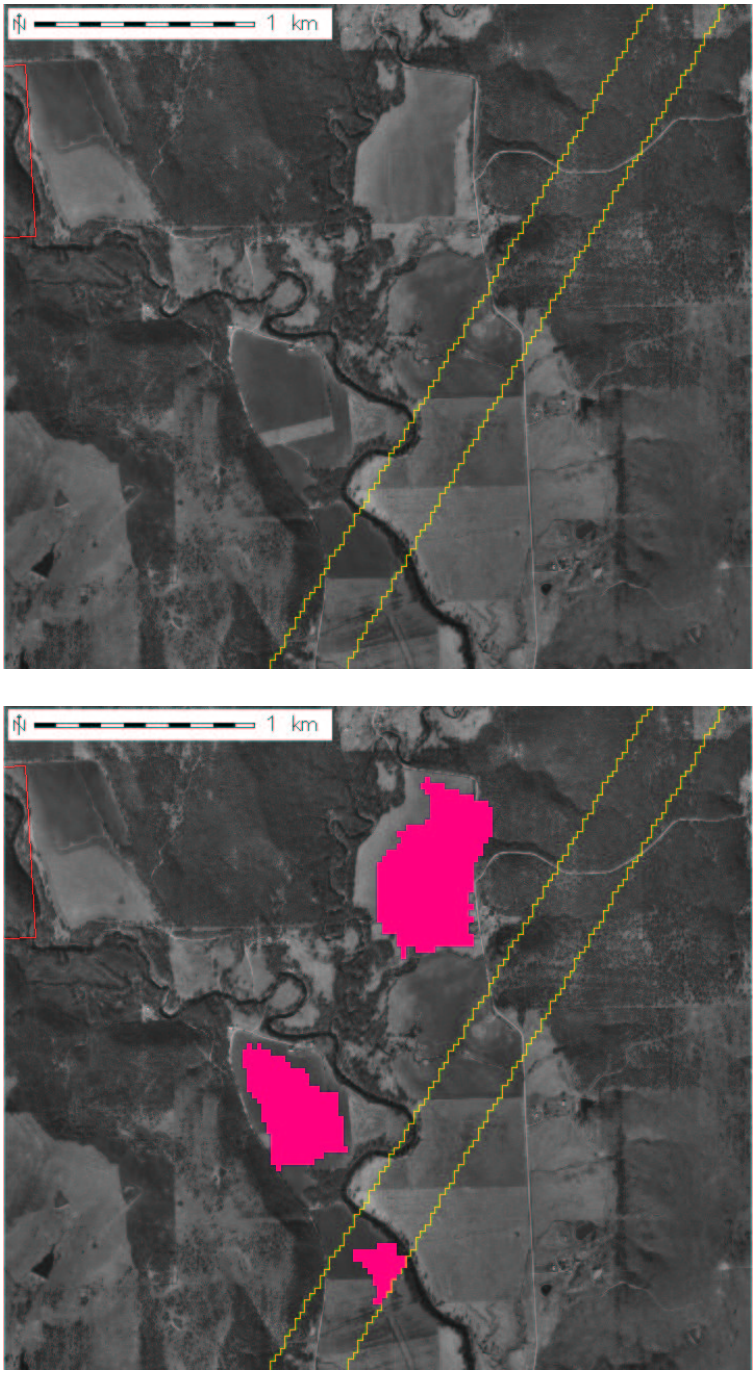


Figure A.30: Region: bigfields, Scene: r05sc04 and r05sc03. Three very large agricultural fields and farm area.

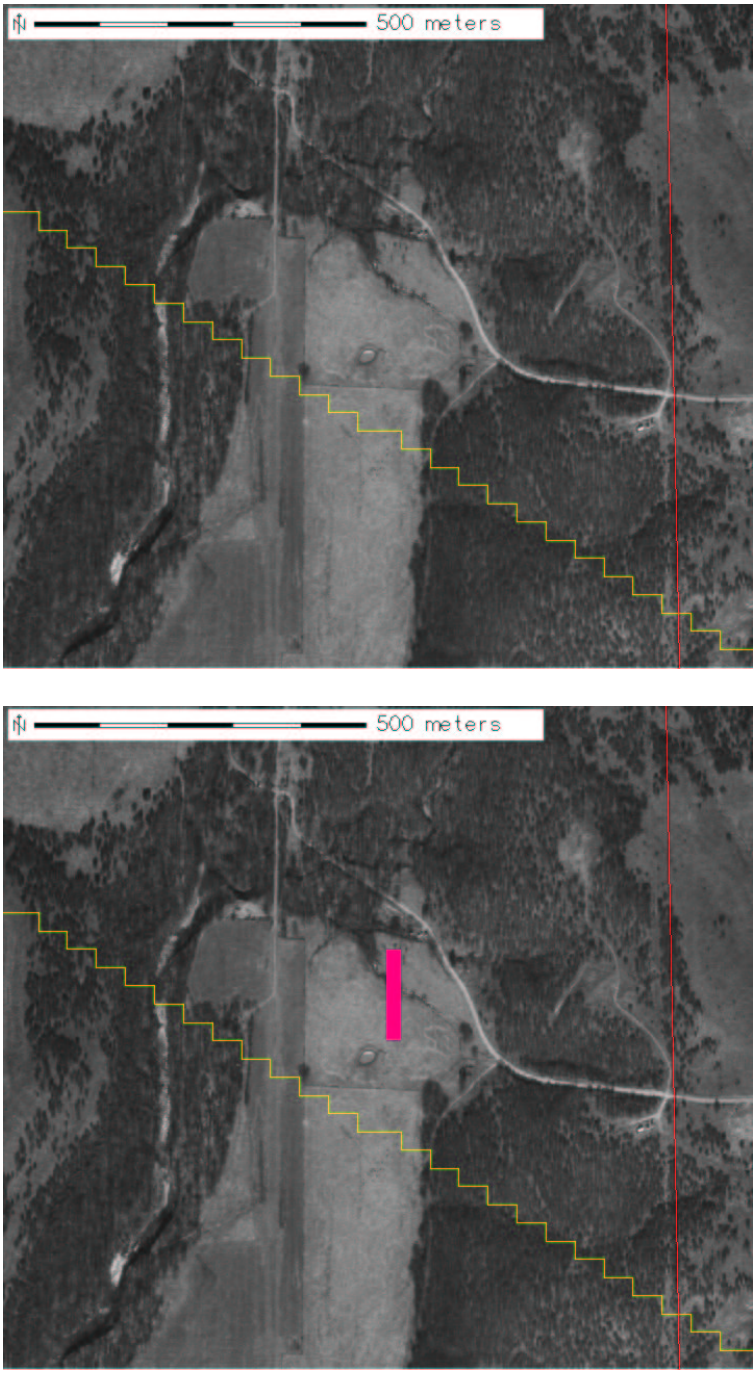


Figure A.31: Region: strip, Scene: r05sc04. Gravel road or strip of agricultural field in area with tanks.

Figure A.32: Region: farm2, Scene: r05sc04. Farm or ranch with large driveway.

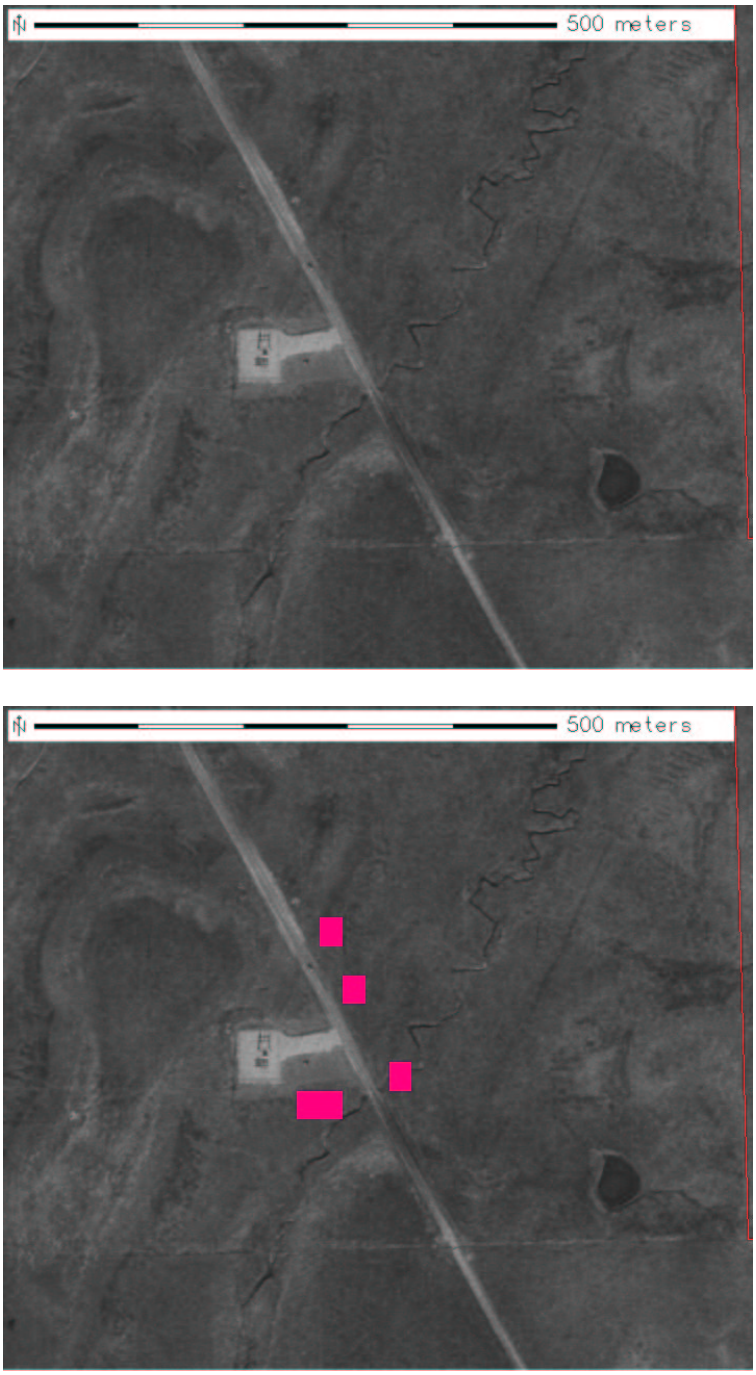


Figure A.33: Region: compound3, Scene: r05sc04. Building and large field lot near highway and bright shoulders on the highway nearby.

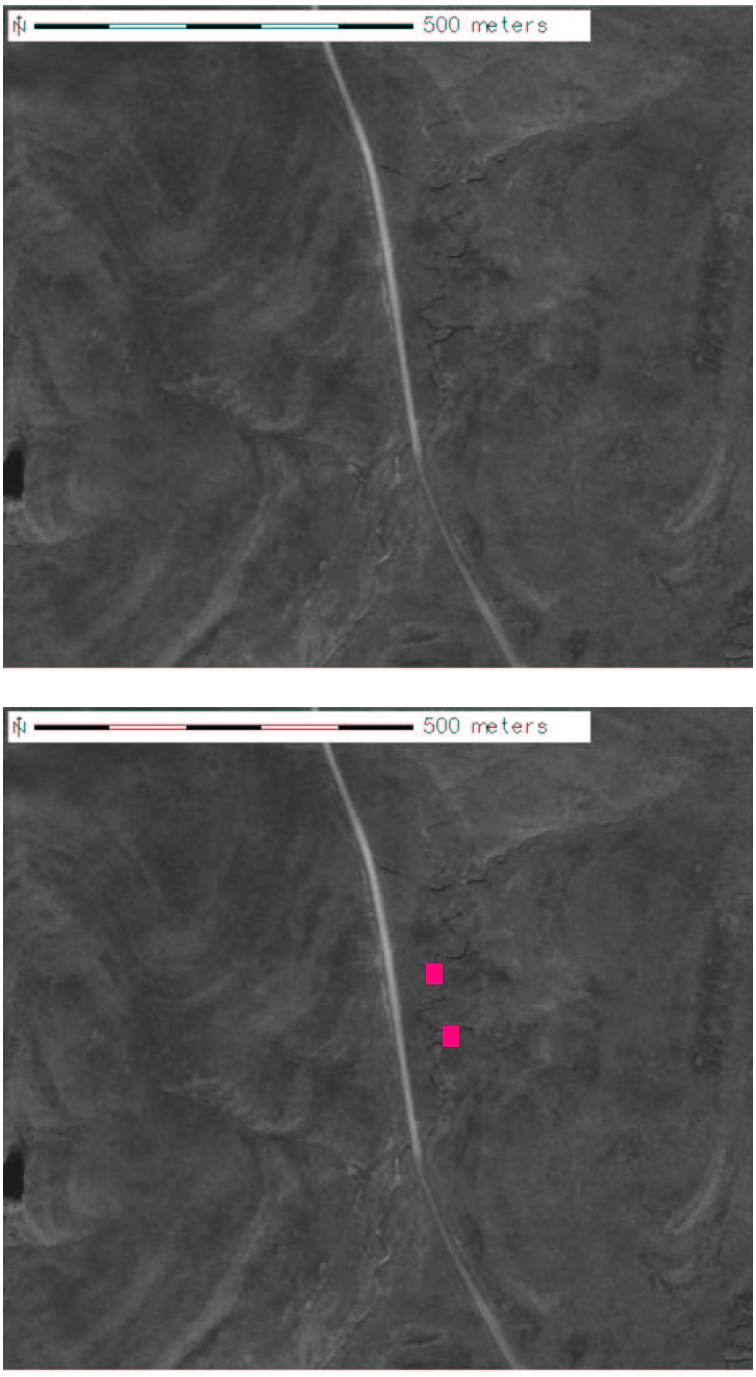


Figure A.34: Region: roadside2, Scene: r05sc04. Bright areas on highway.

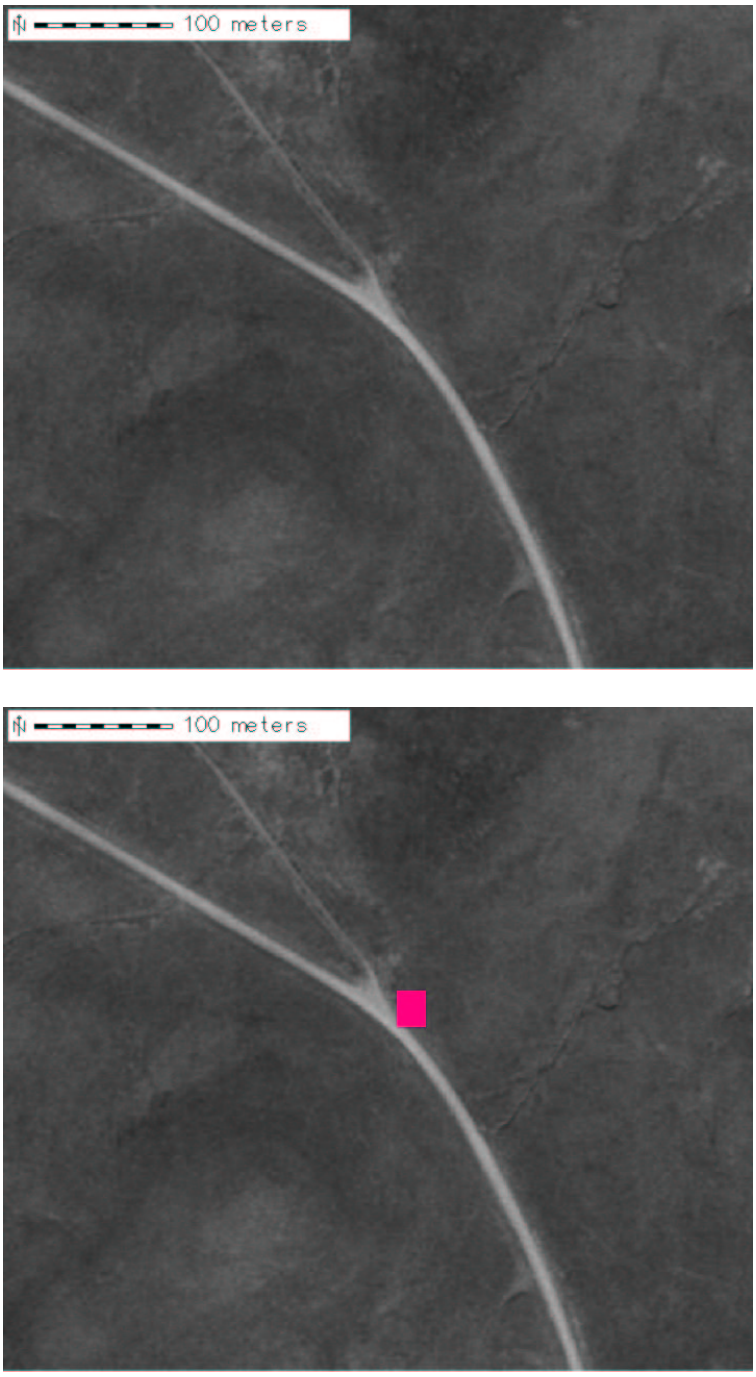


Figure A.35: Region: intersection9, Scene: r05sc04. Highway intersection.

Figure A.36: Region: road, Scene: r05sc04. Bridge and highway turnaround.

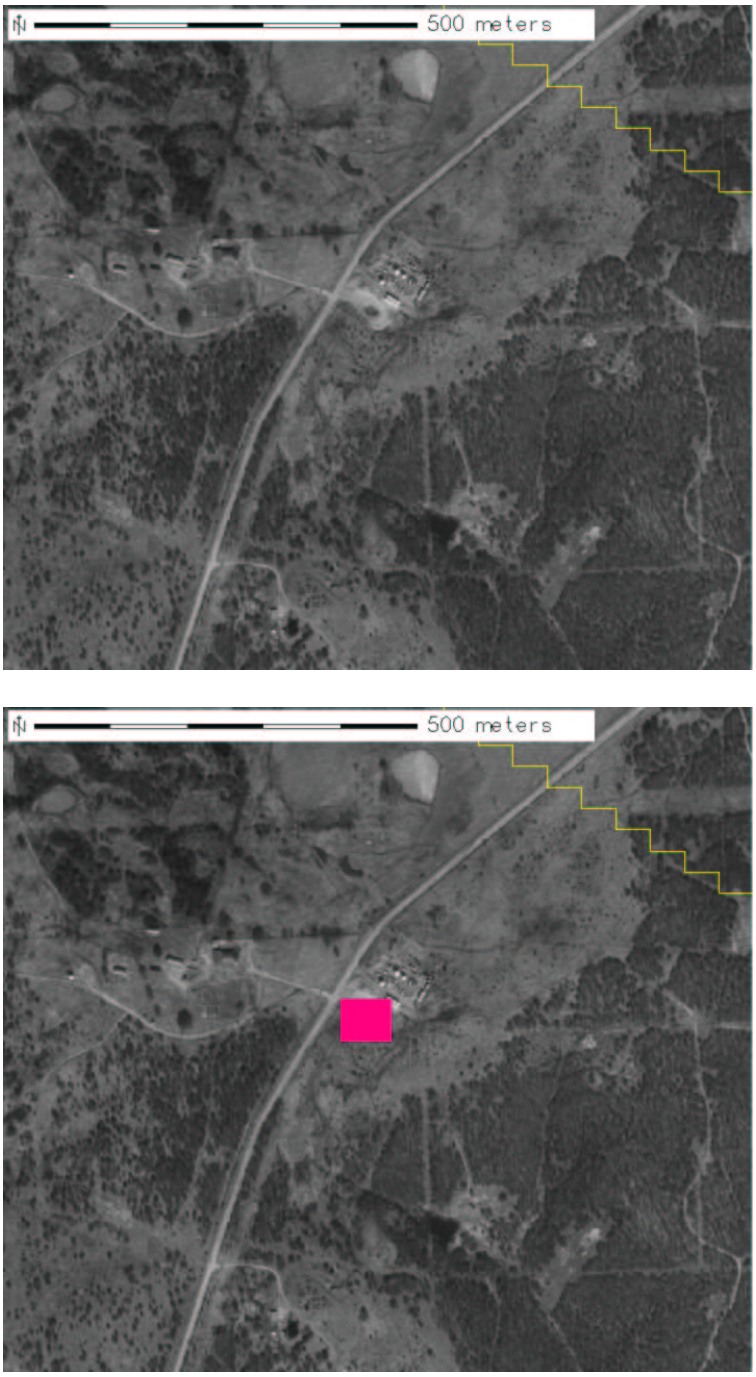


Figure A.37: Region: tank_cluster, Scene: r05sc03. Large tank farm.

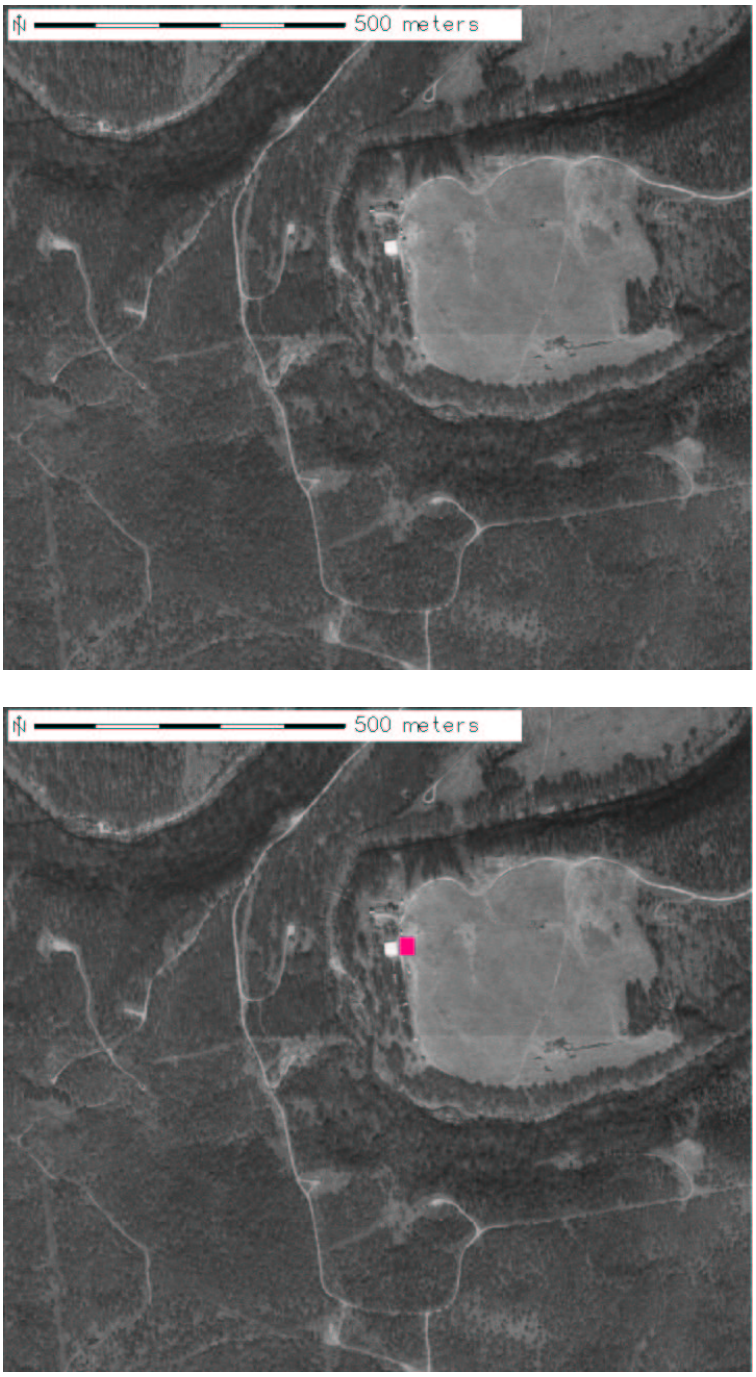


Figure A.38: Region: tank_group, Scene: r05sc03. A collection of tanks near wells.

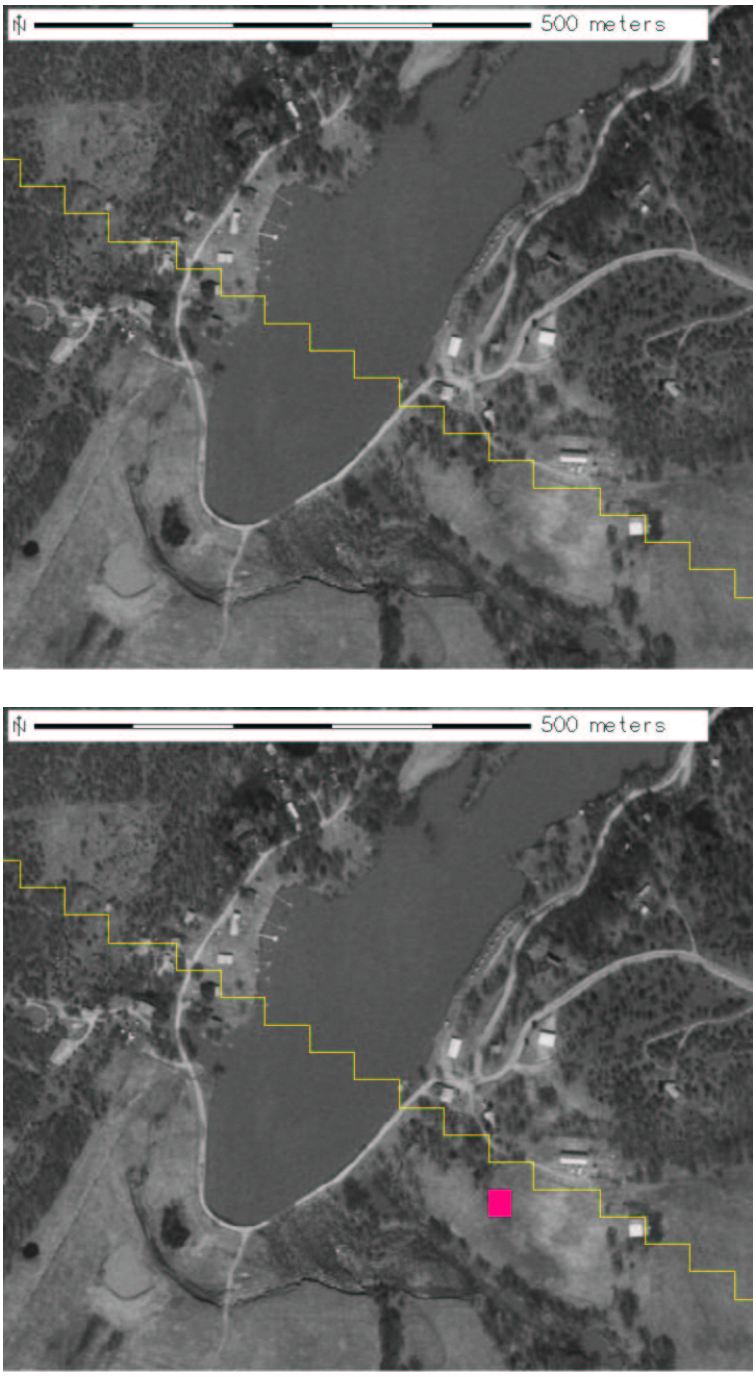


Figure A.39: Region: waterfront, Scene: r05sc03. Ag field near waterfront.

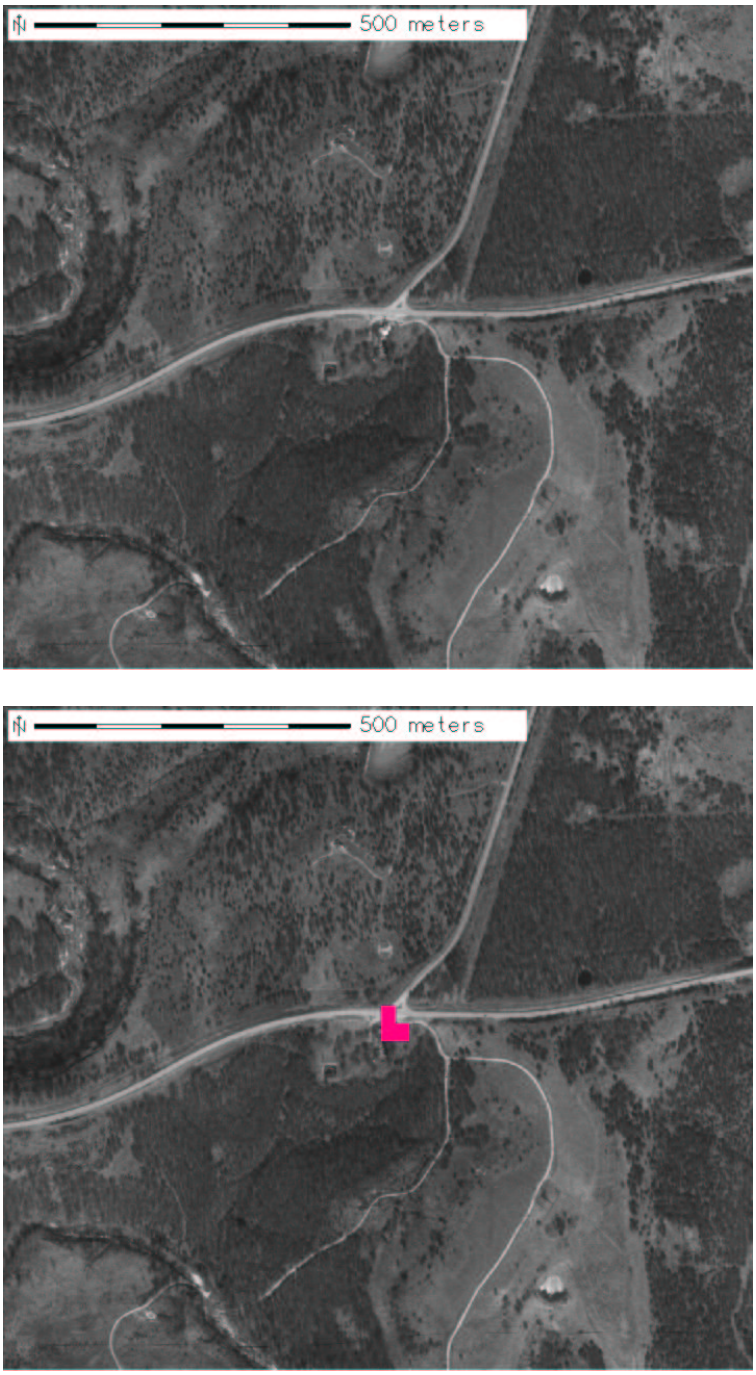


Figure A.40: Region: facility_intersection, Scene: r05sc03. Some kind of facility at a large intersection.

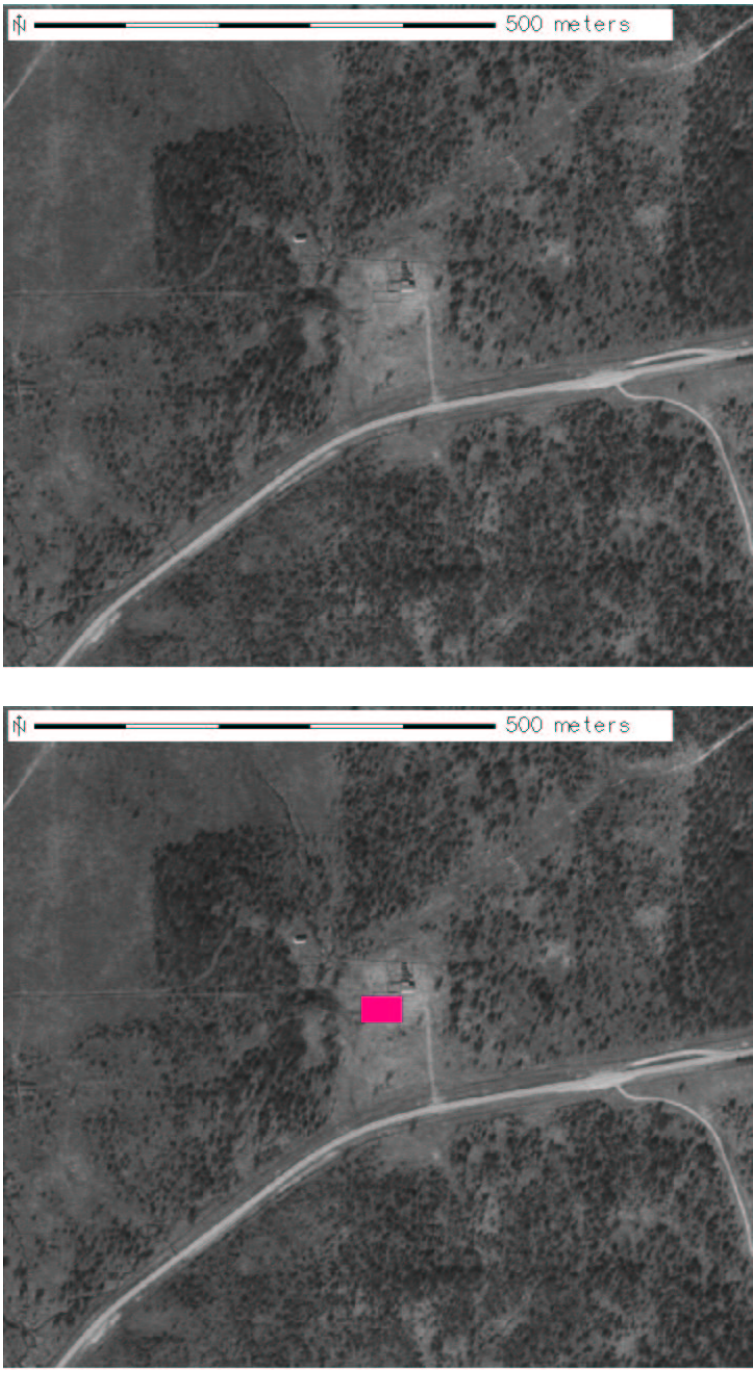


Figure A.41: Region: compound4, Scene: r05sc03. Facility situated among various wells.

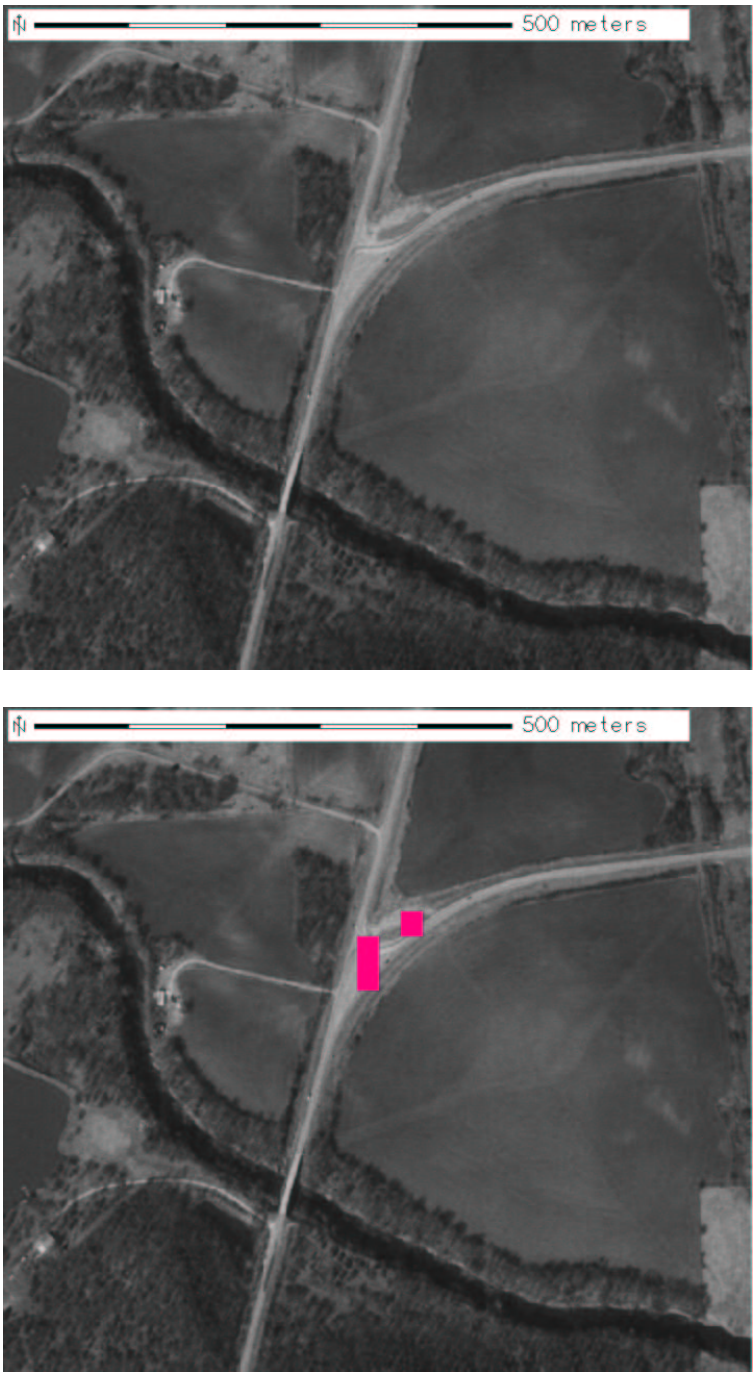


Figure A.42: Region: ag_intersection, Scene: r05sc03. Large roadway intersection.

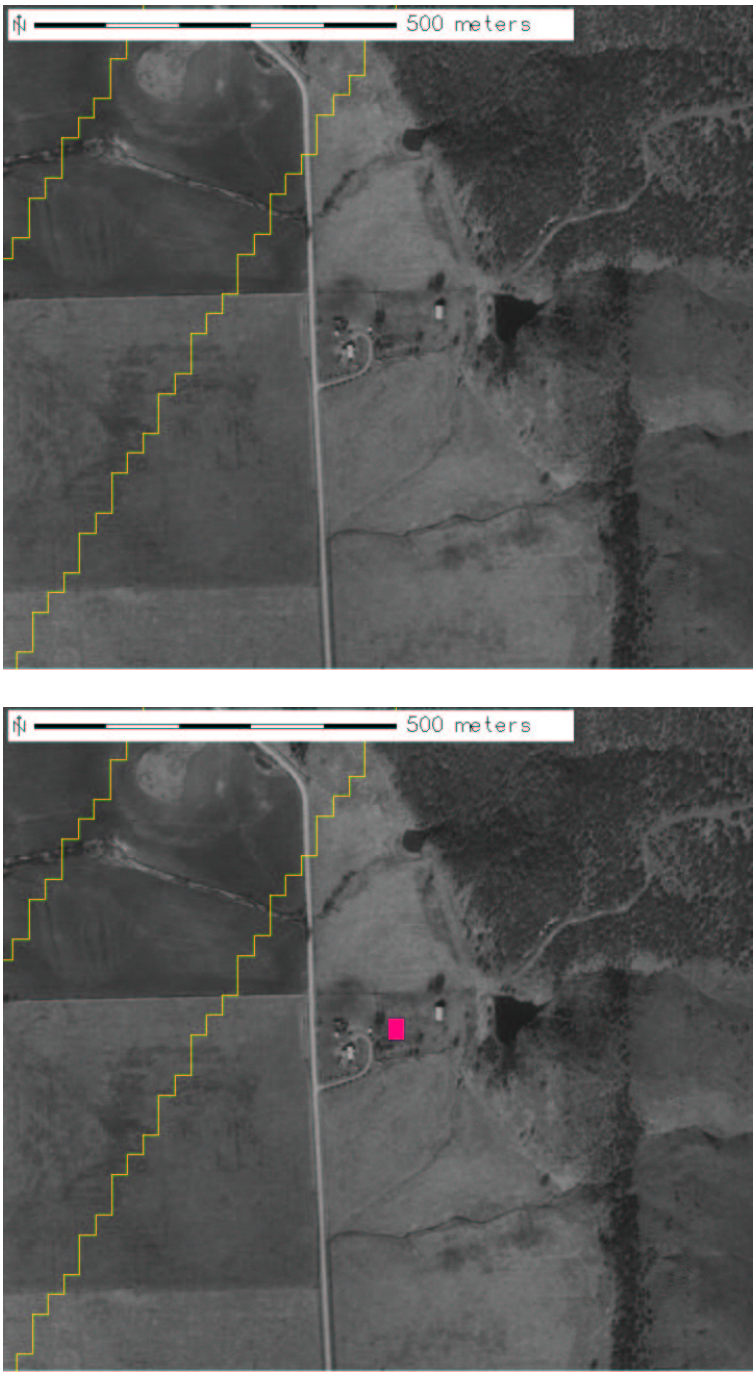


Figure A.43: Region: farm4, Scene: r05sc03. Farm house and barn area.

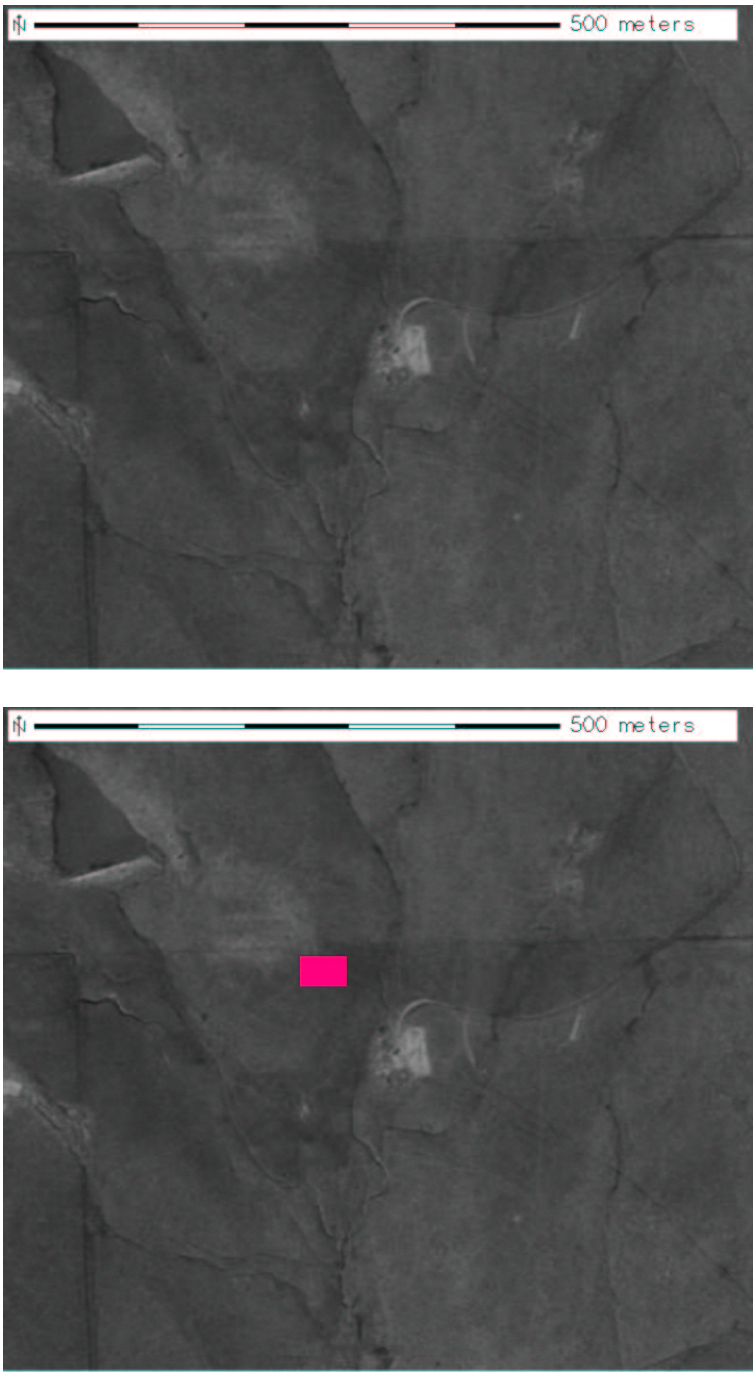


Figure A.44: Region: disturb1_new, Scene: r05sc03. Brine scar around well.

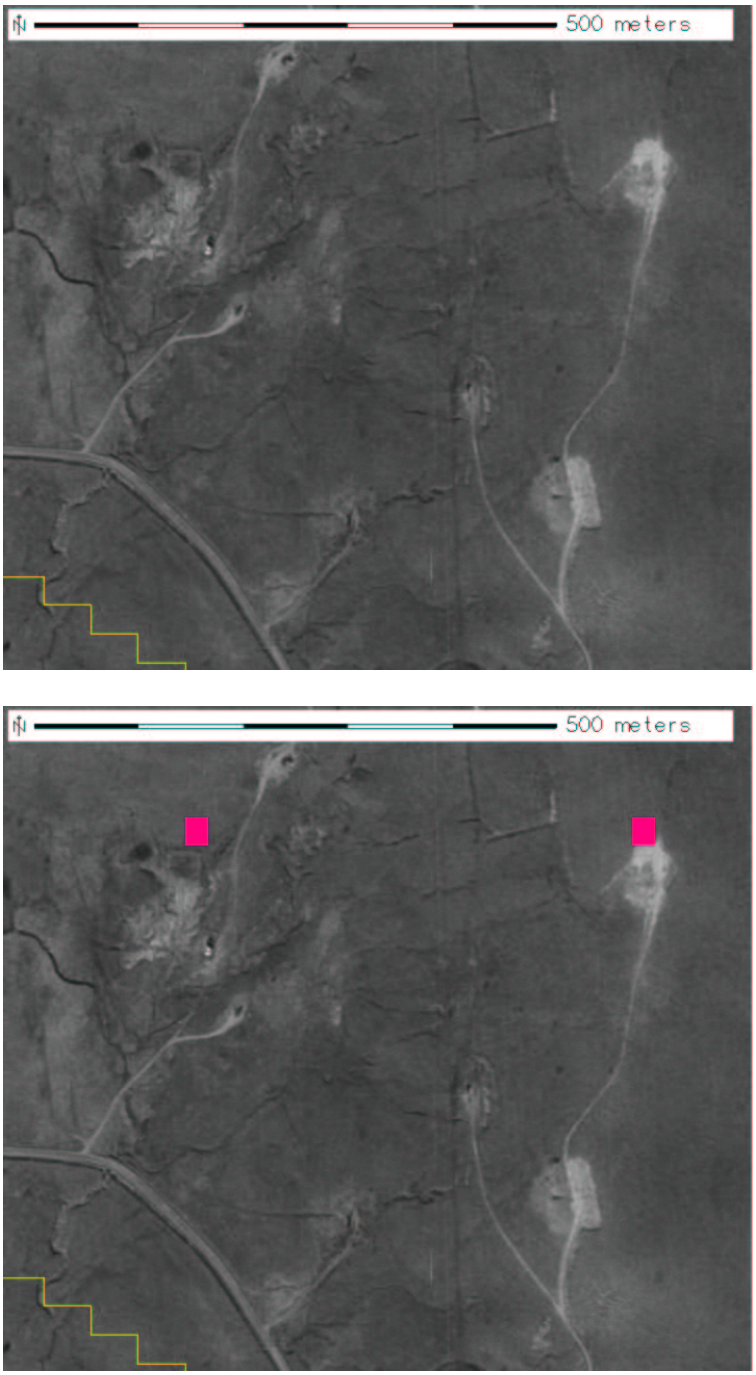


Figure A.45: Region: disturb2_new, Scene: r05sc03. Two obvious brine scars; one near holding tanks and the other at a well site.

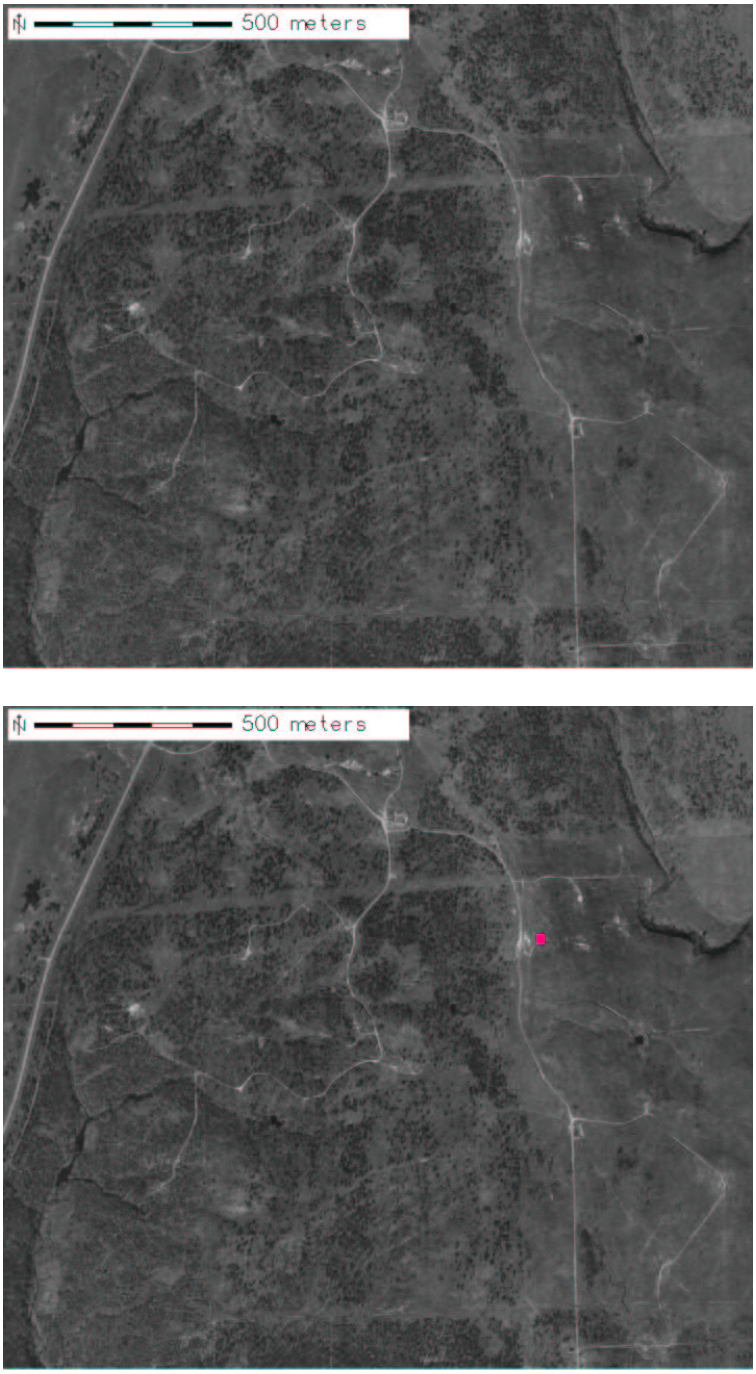


Figure A.46: Region: disturb3, Scene: r05sc03. Brine scar on road surrounded by wells.

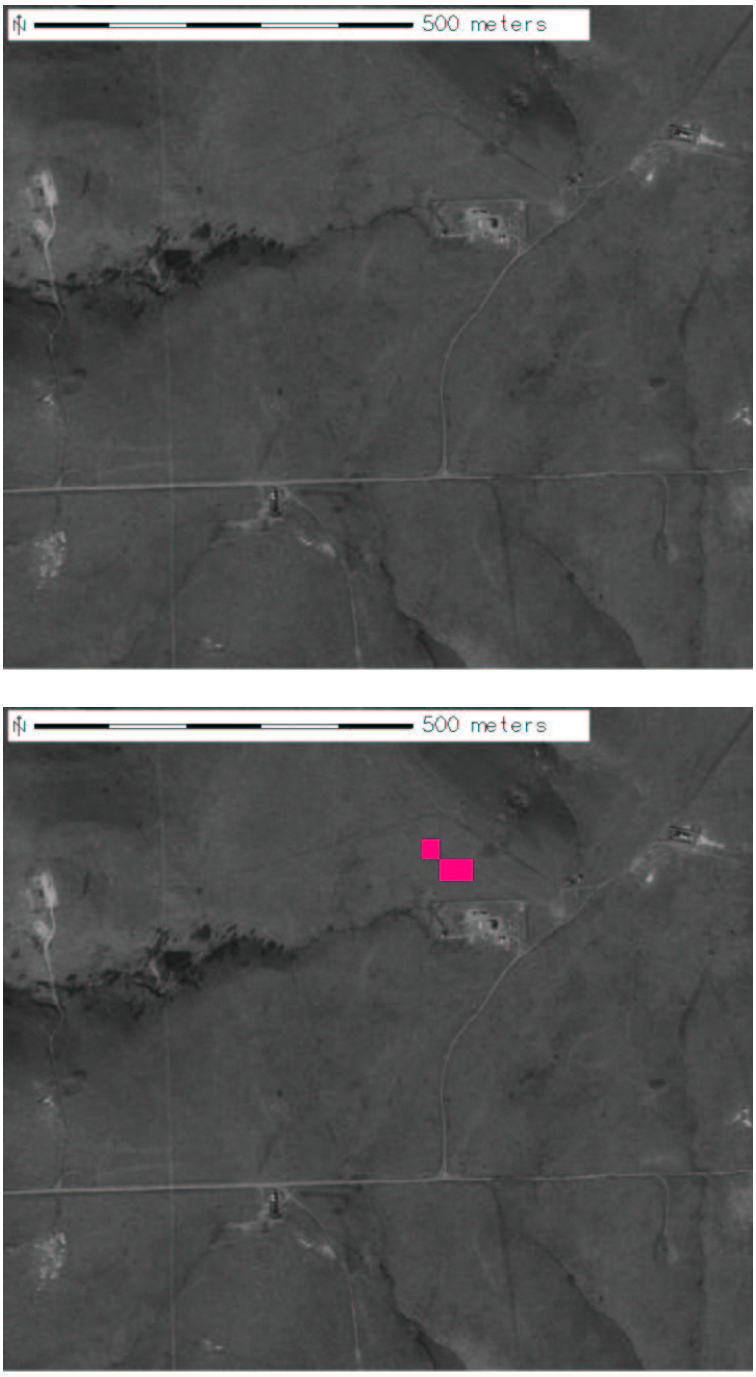


Figure A.47: Region: tank_station, Scene: r05sc03. Tank depot in well area.

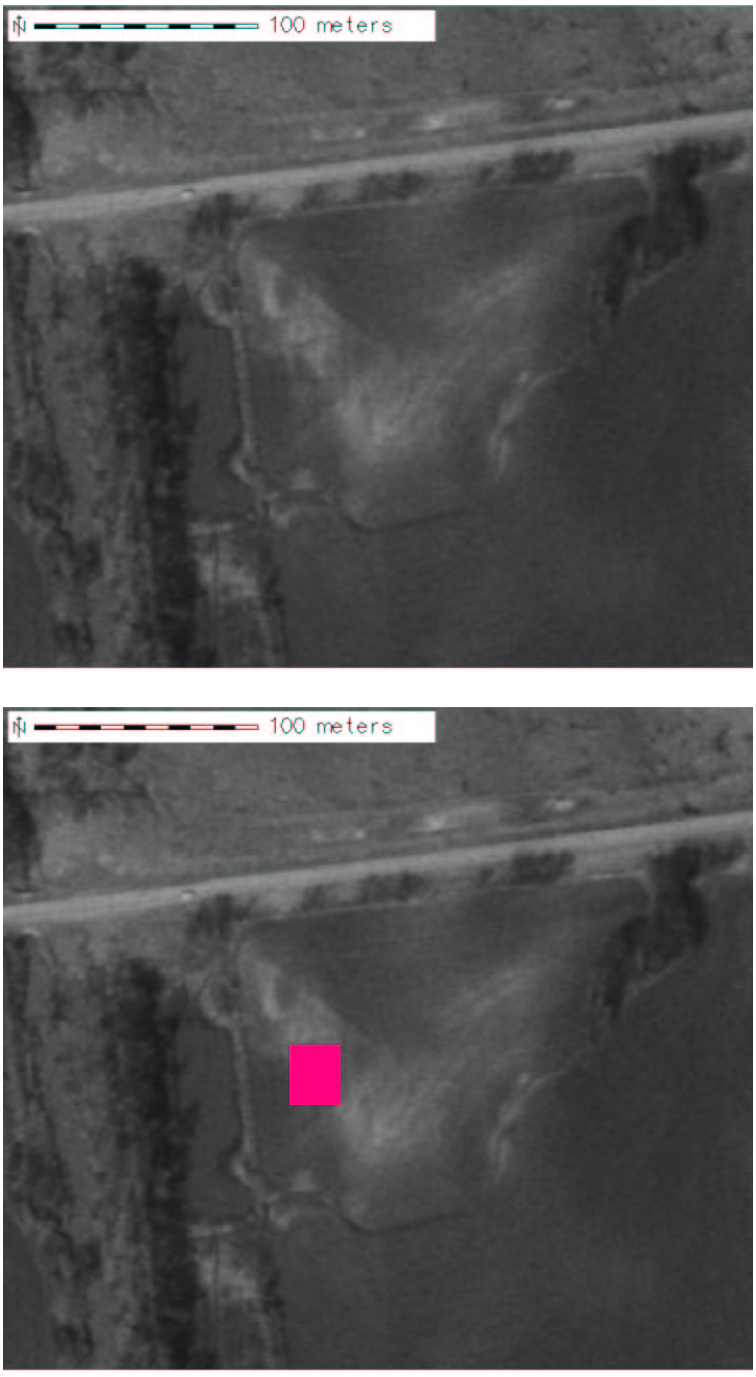


Figure A.48: Region: agriculture, Scene: r05sc03. Agricultural field.

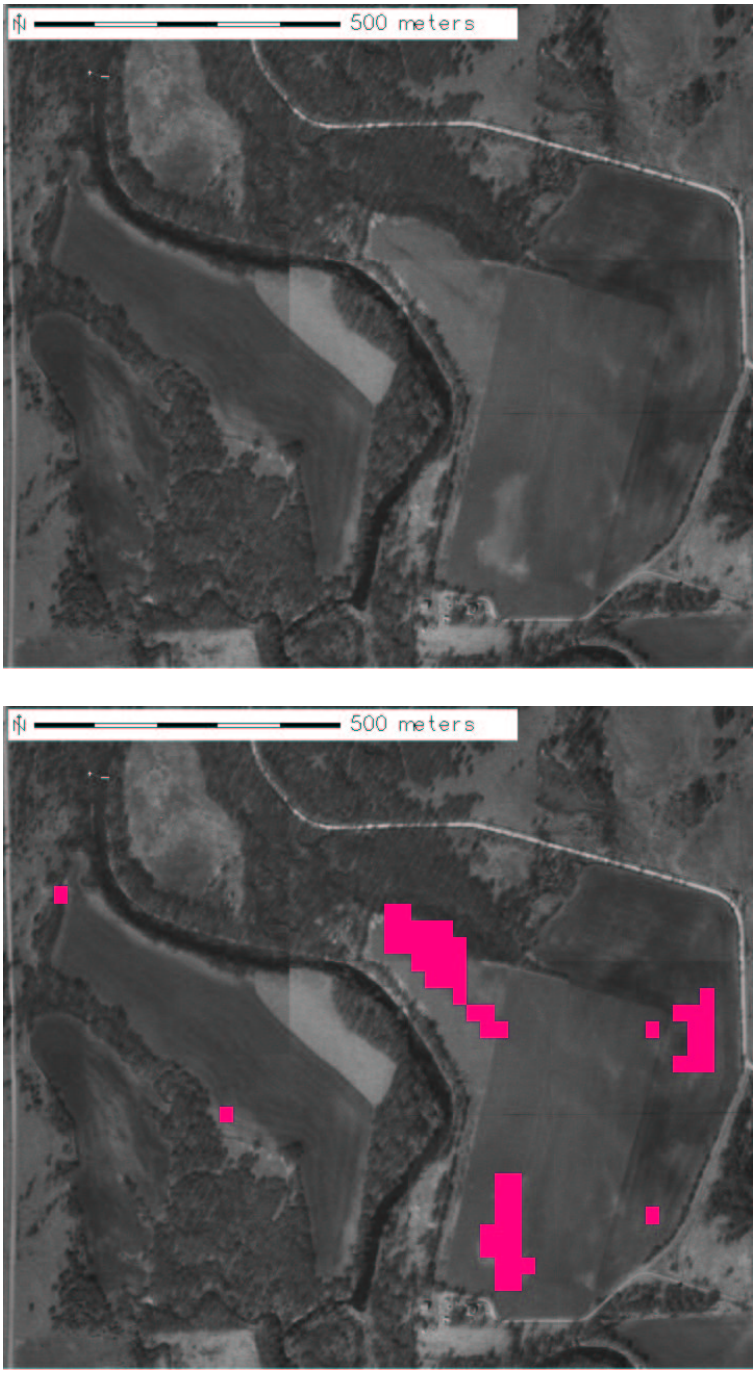


Figure A.49: Region: agriculture2, Scene: r05sc03. Multiple spots on agricultural fields.

Figure A.50: Region: erosion1, Scene: r05sc03. Bright eroded area near road.

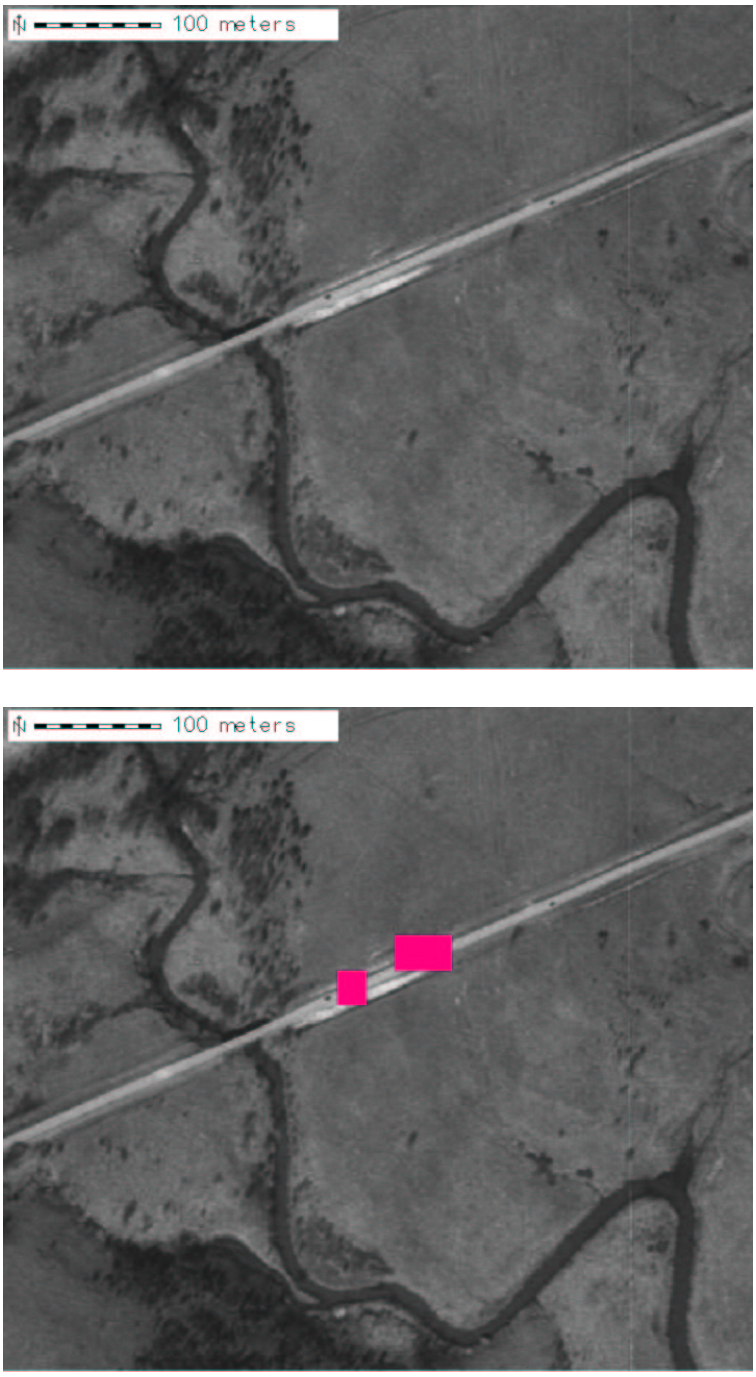


Figure A.51: Region: roadshoulder1, Scene: r05sc03. Bright road shoulders.

Figure A.52: Region: junction, Scene: r05sc03. Highway junction and a large building and fenced compound near wells.

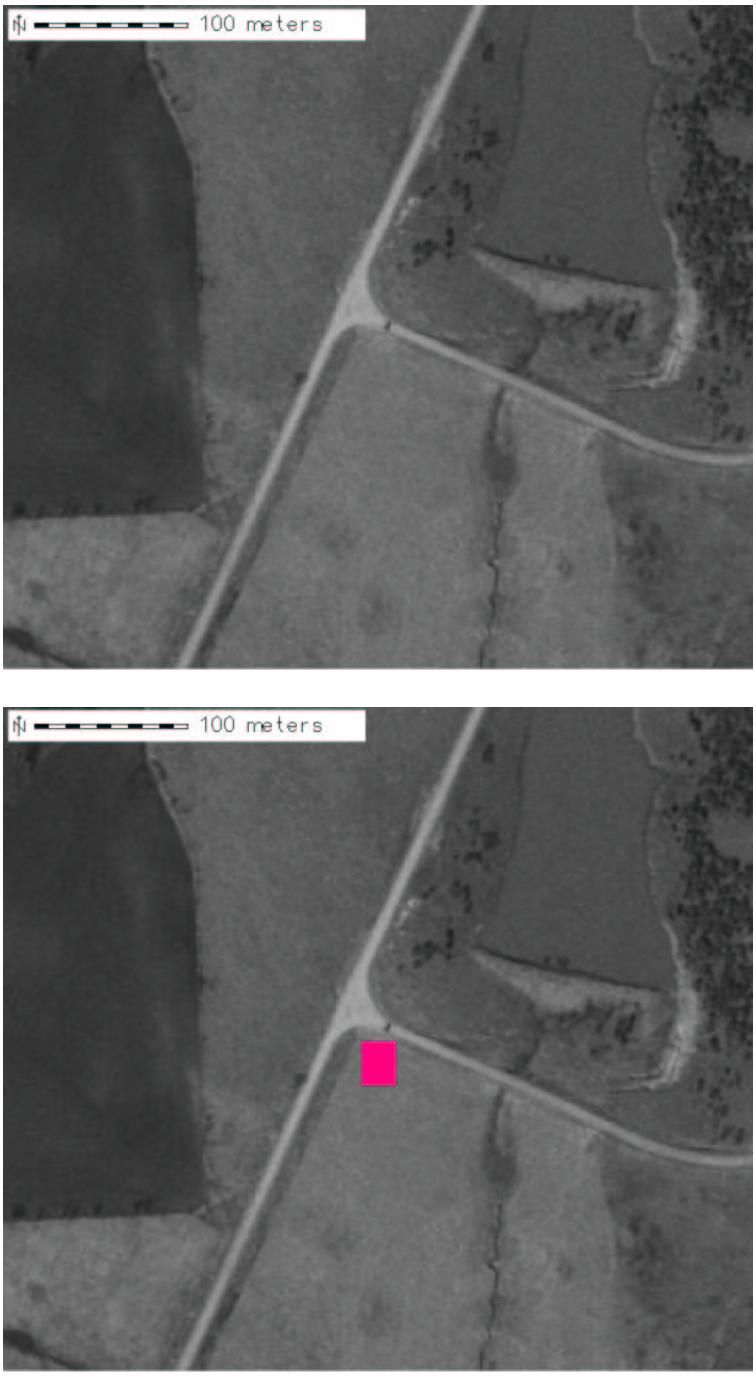


Figure A.53: Region: intersection1, Scene: r05sc03. Major road intersection.

Figure A.54: Region: farm1, Scene: r05sc03. Large ranch.

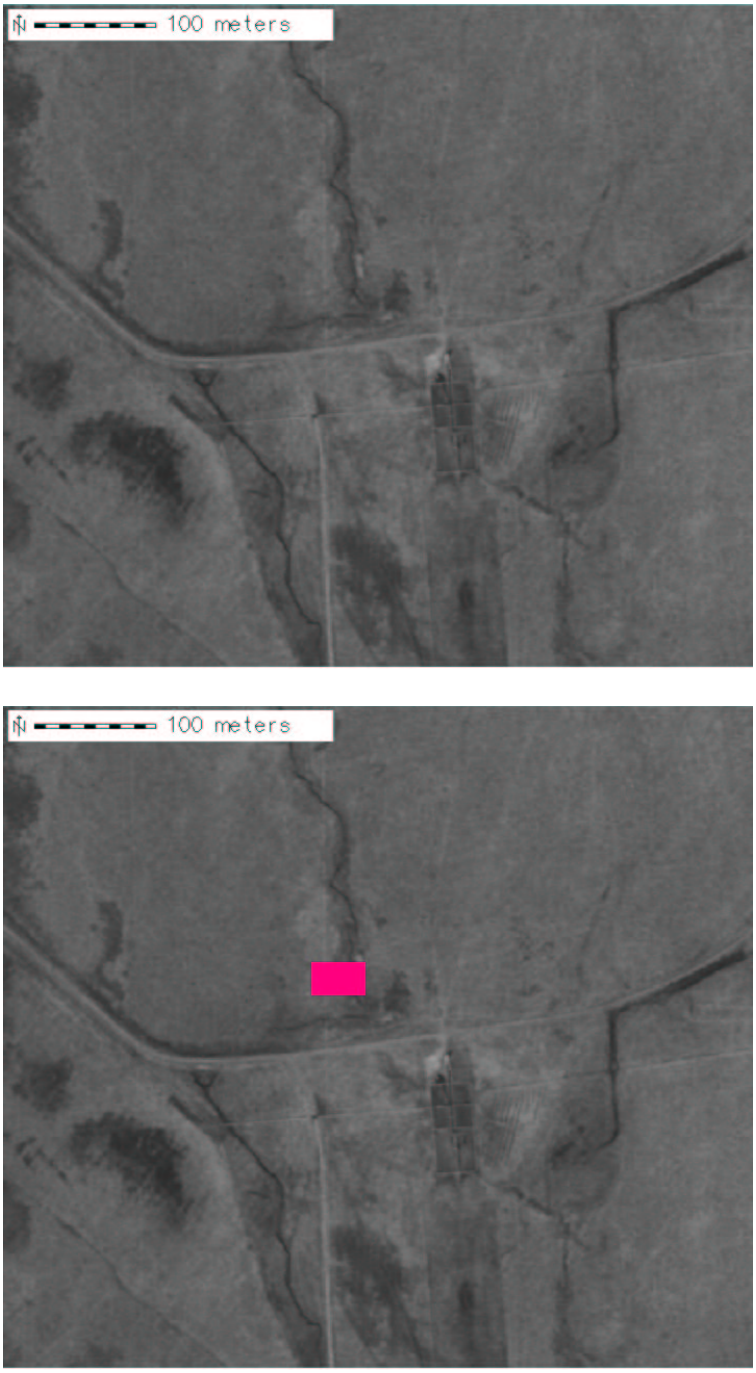


Figure A.55: Region: compound2, Scene: r05sc03. Fenced-off compound with parking area.

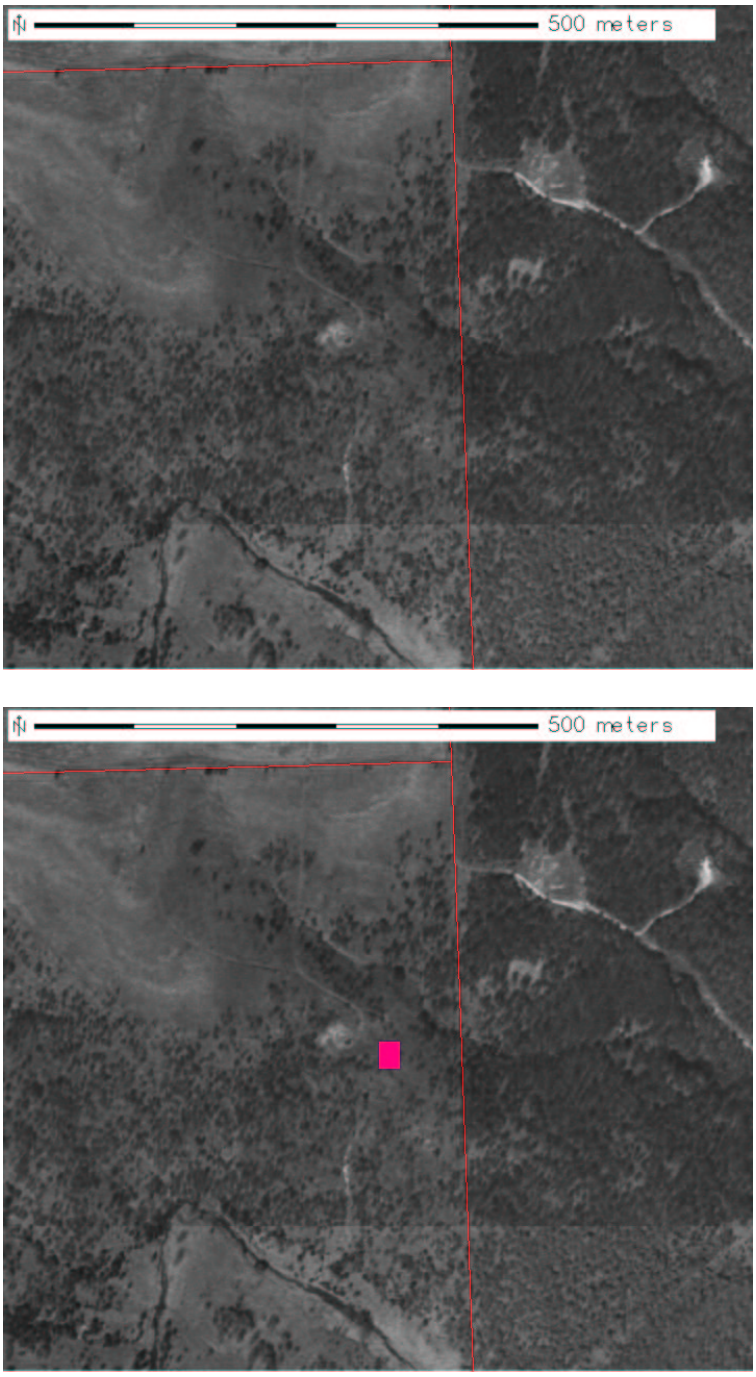


Figure A.56: Region: disturb4, Scene: r06sc06. Well site.

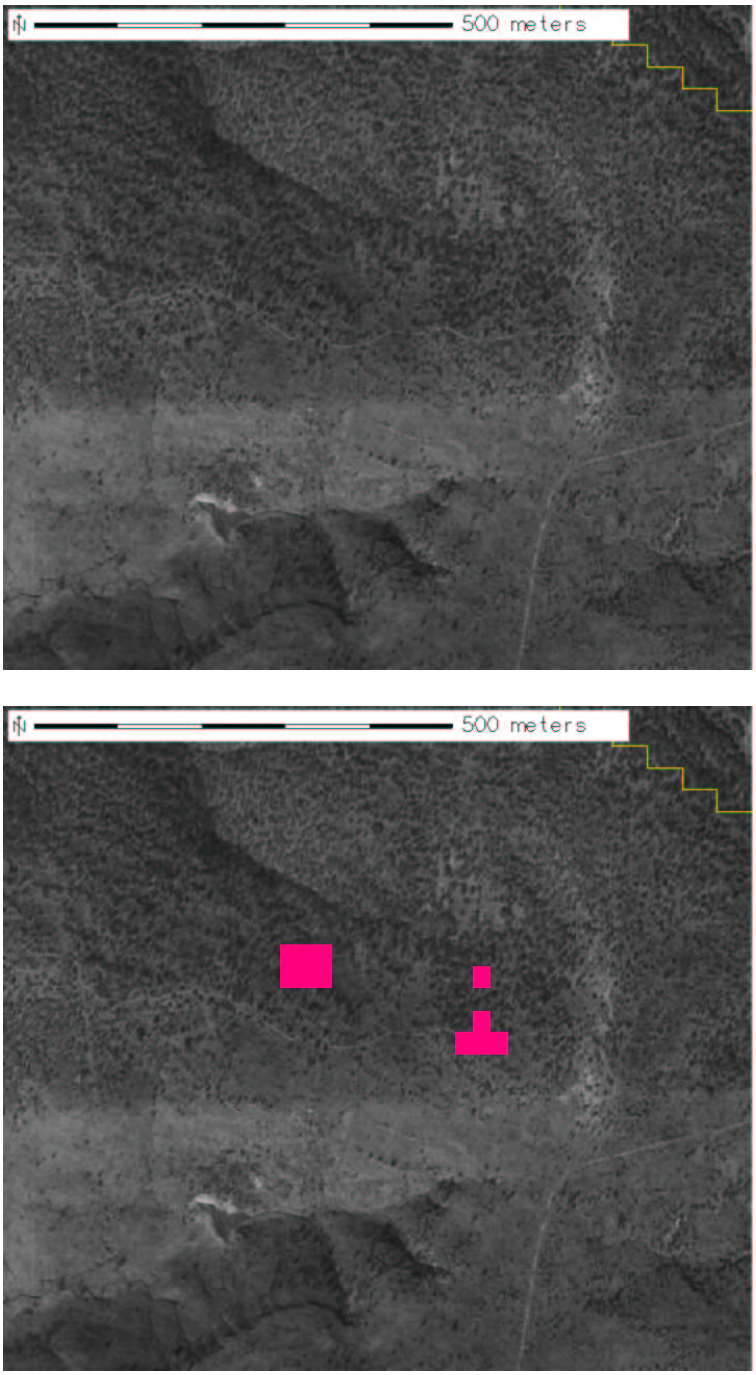


Figure A.57: Region: disturb5, Scene: r06sc06. Large area of erosion in wooded area with wells and tanks nearby.

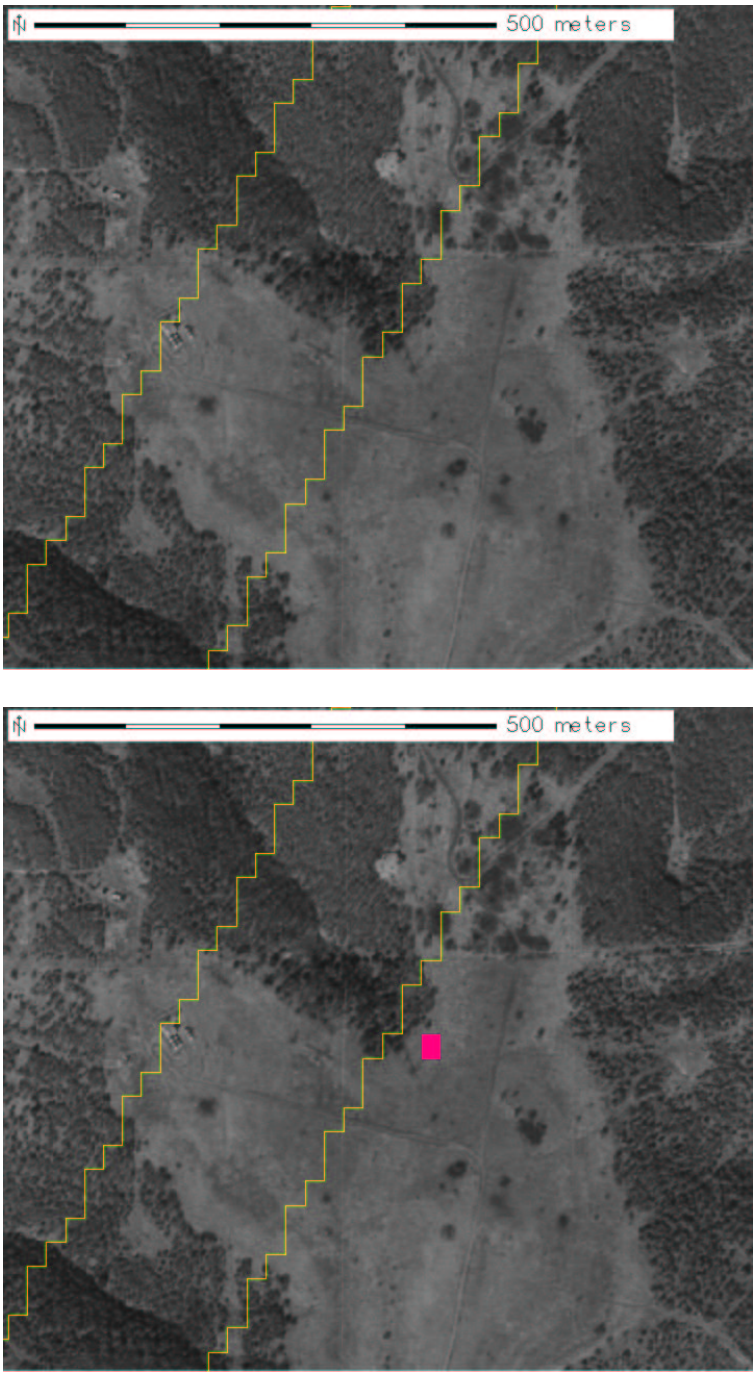


Figure A.58: Region: disturb6, Scene: r06sc06. Well site, tank farm, or new feature.

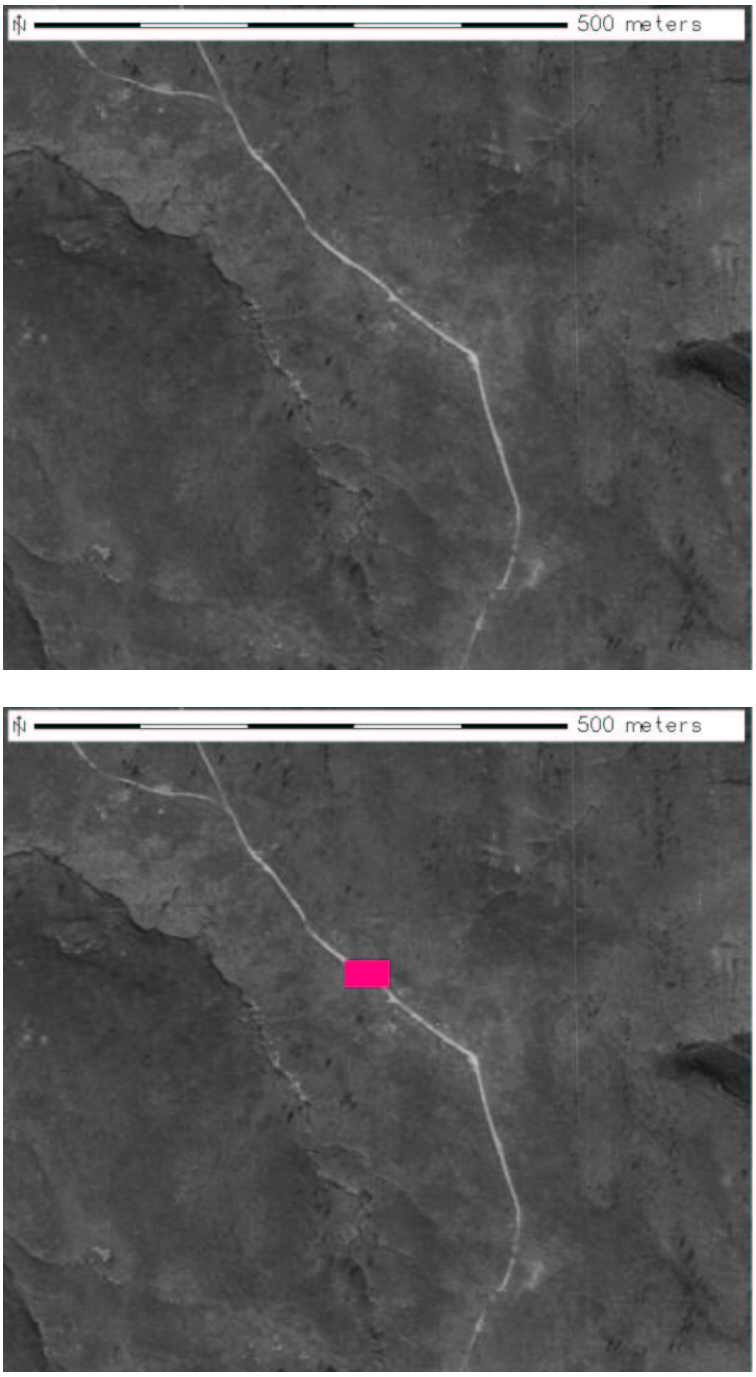


Figure A.59: Region: disturb10, Scene: r06sc06. Tank or well site near road.

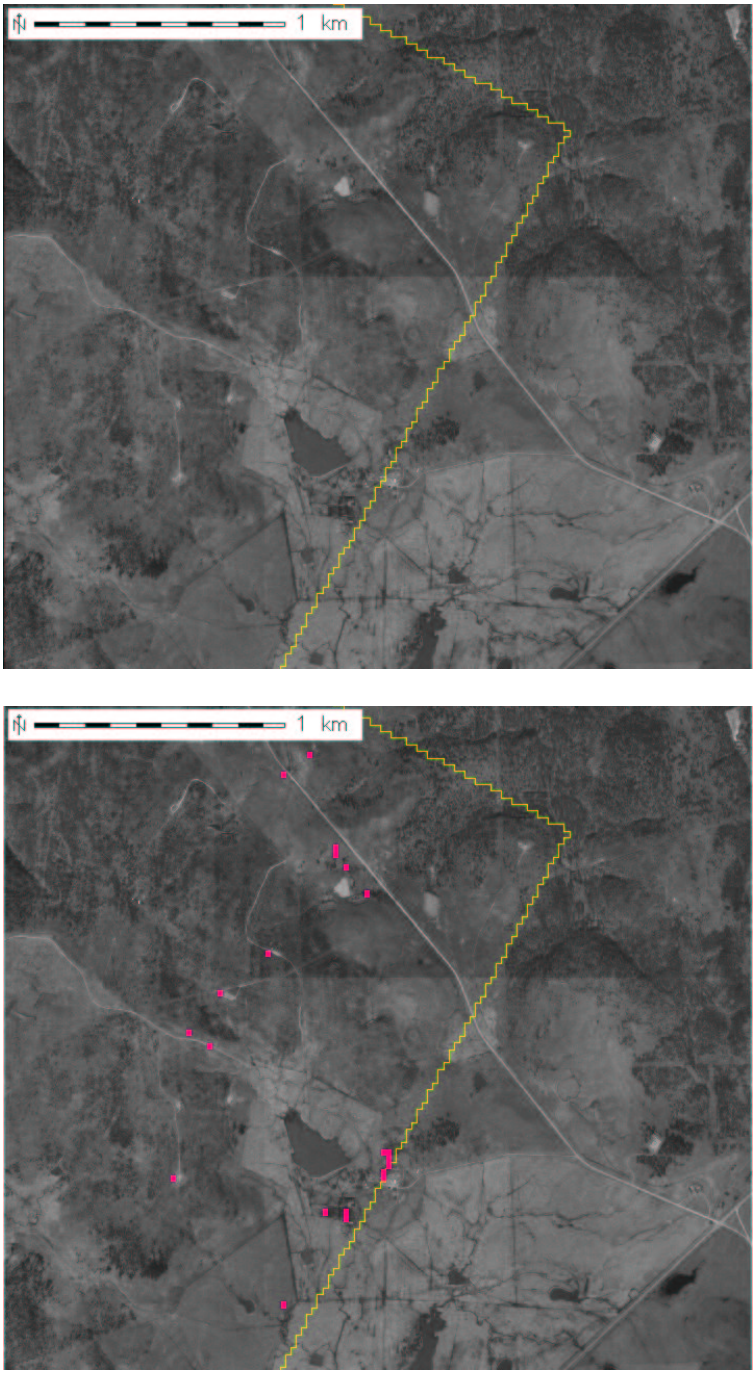


Figure A.60: Region: wells_in_66, Scene: r06sc06. Brine scars at many well and tank sites.

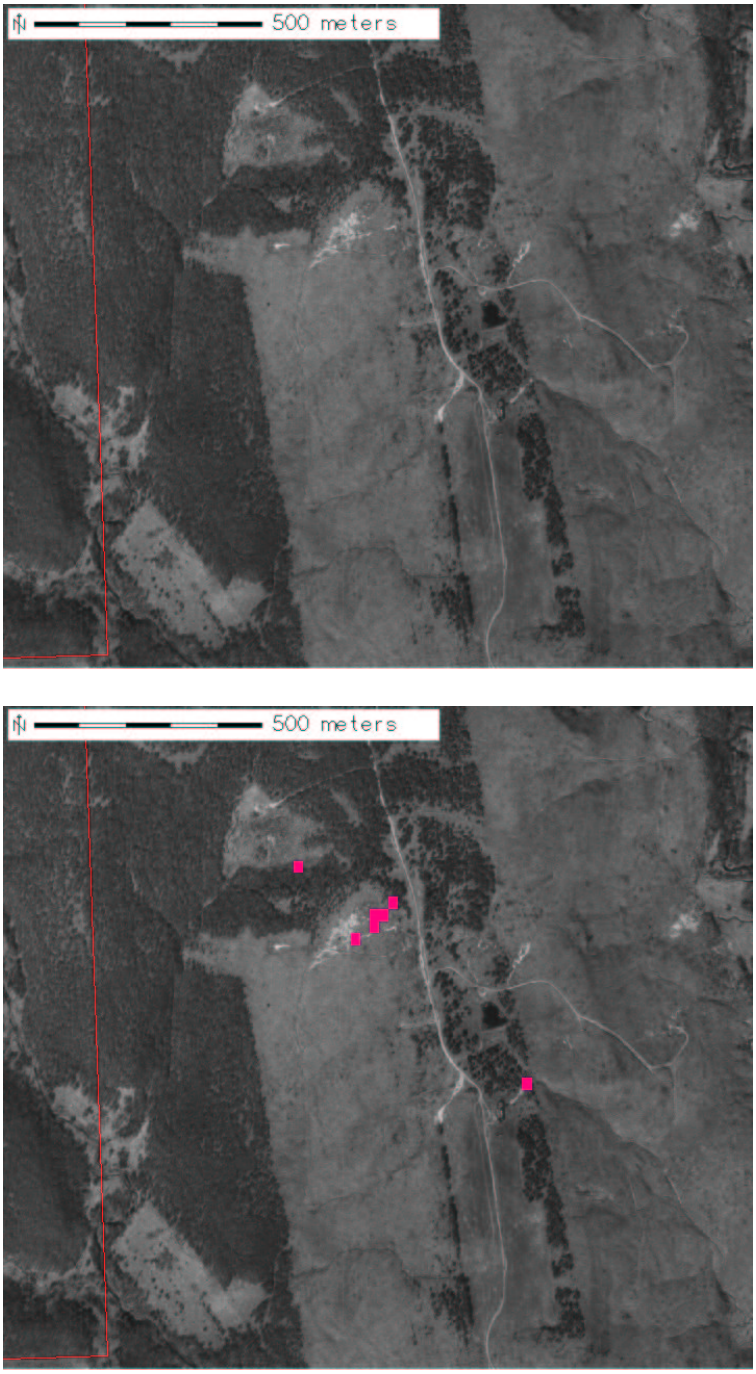


Figure A.61: Region: group_of_spills, Scene: r06sc06. Brine scars at three well/tank sites.

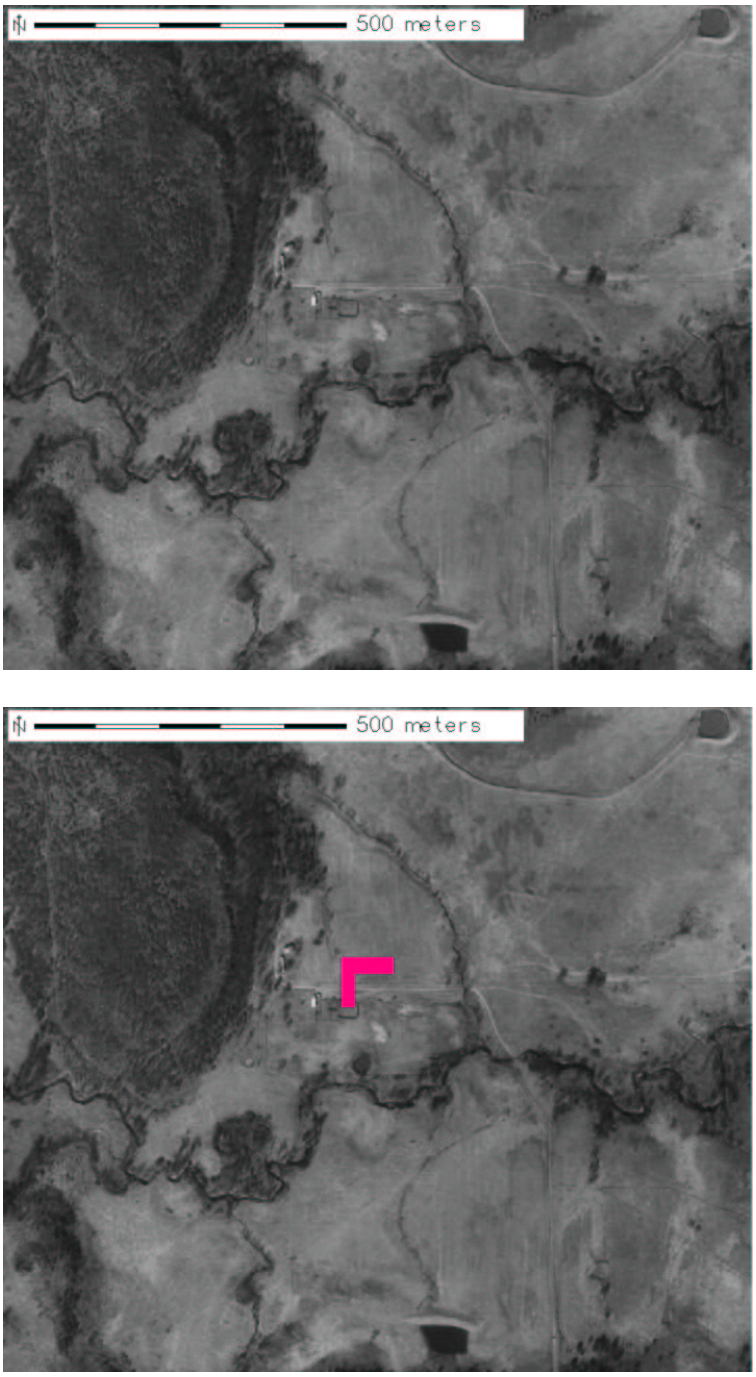


Figure A.62: Region: ranch3, Scene: r06sc06. Ranch with large driveway.

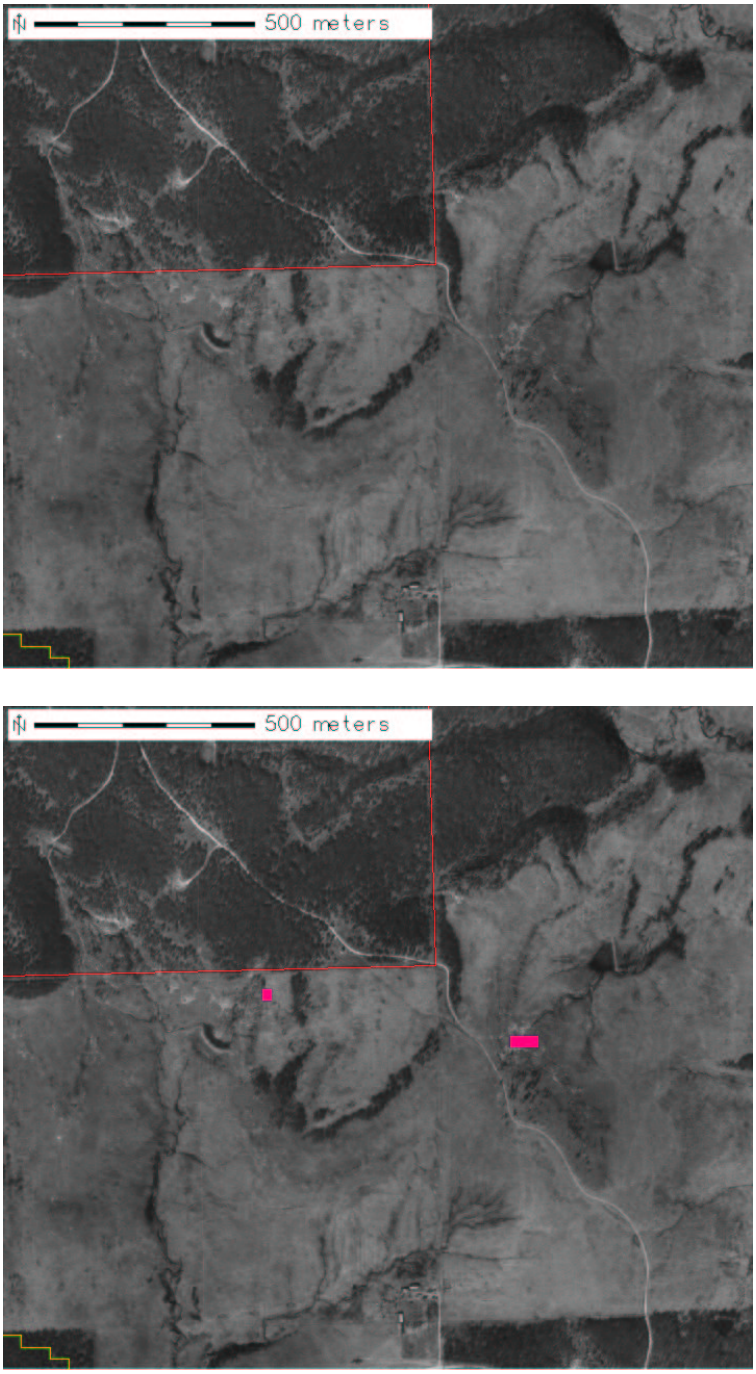


Figure A.63: Region: erosion8, Scene: r06sc06. Runoff from pond and nearby well site.

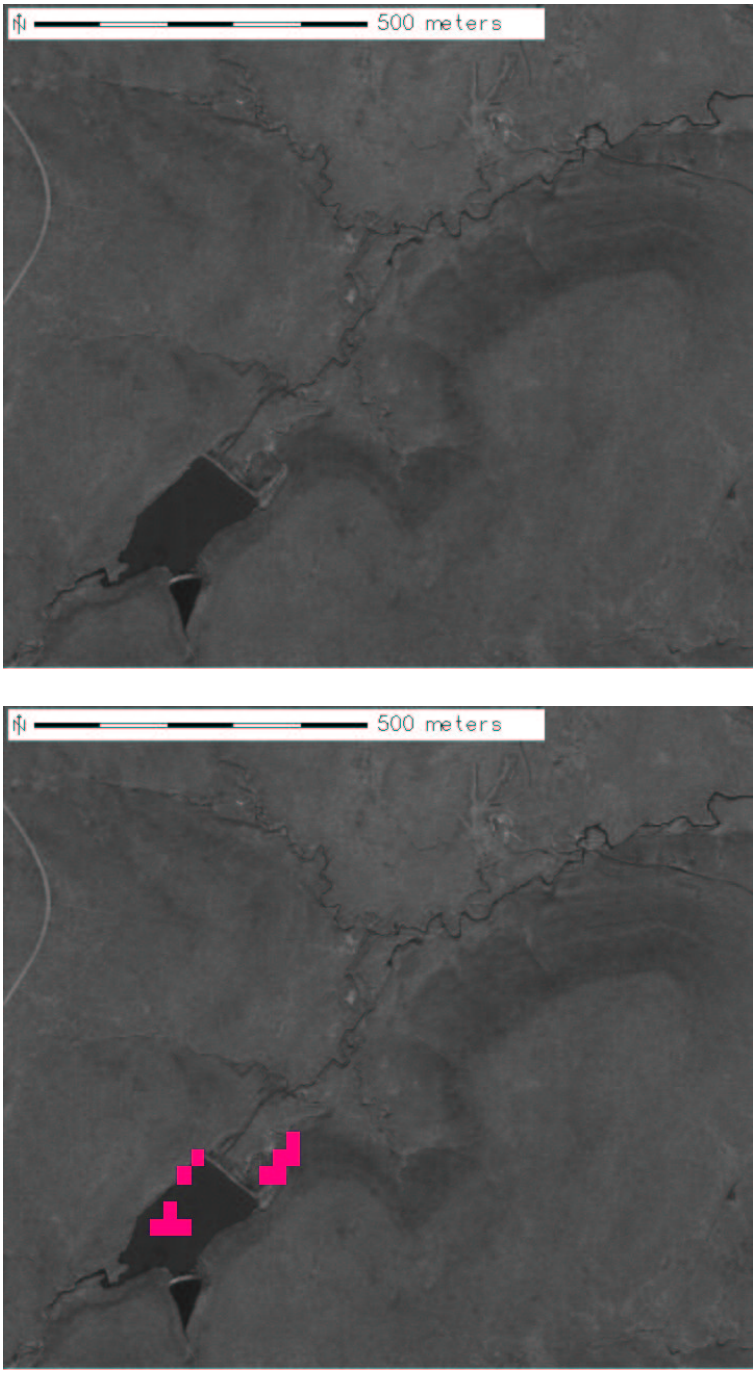


Figure A.64: Region: runoff, Scene: r06sc05. Highly eroded area and pond downstream of well site.

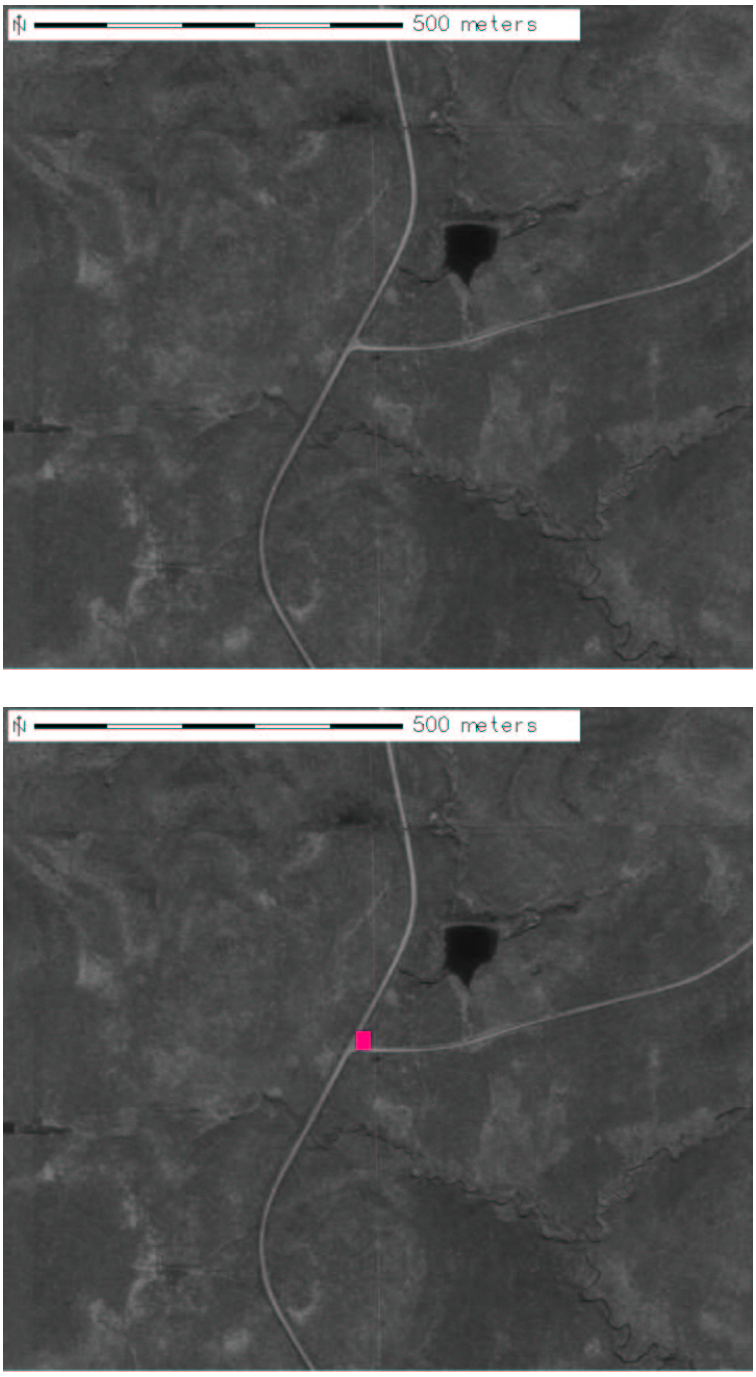


Figure A.65: Region: intersection10, Scene: r06sc05. Road intersection.

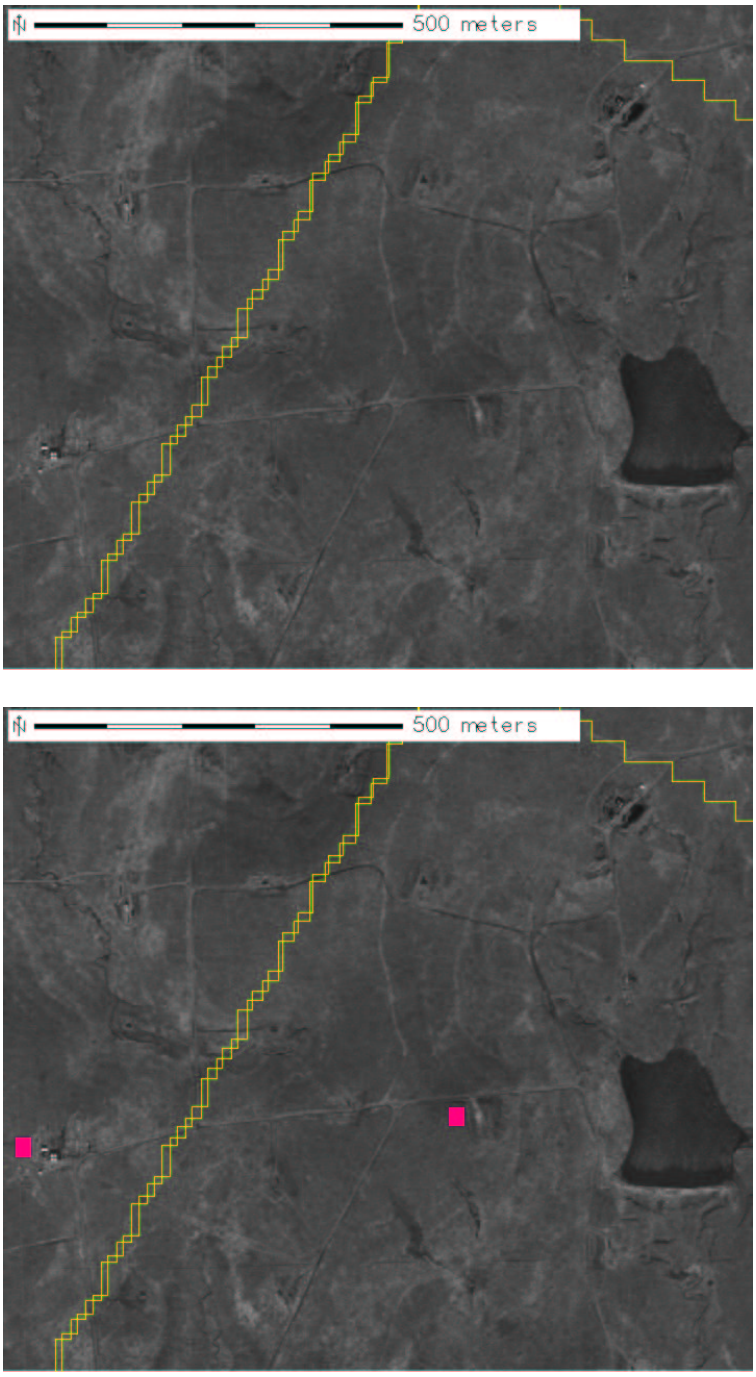


Figure A.66: Region: disturb11, Scene: r06sc05 and r06sc04. Tanks and well site.

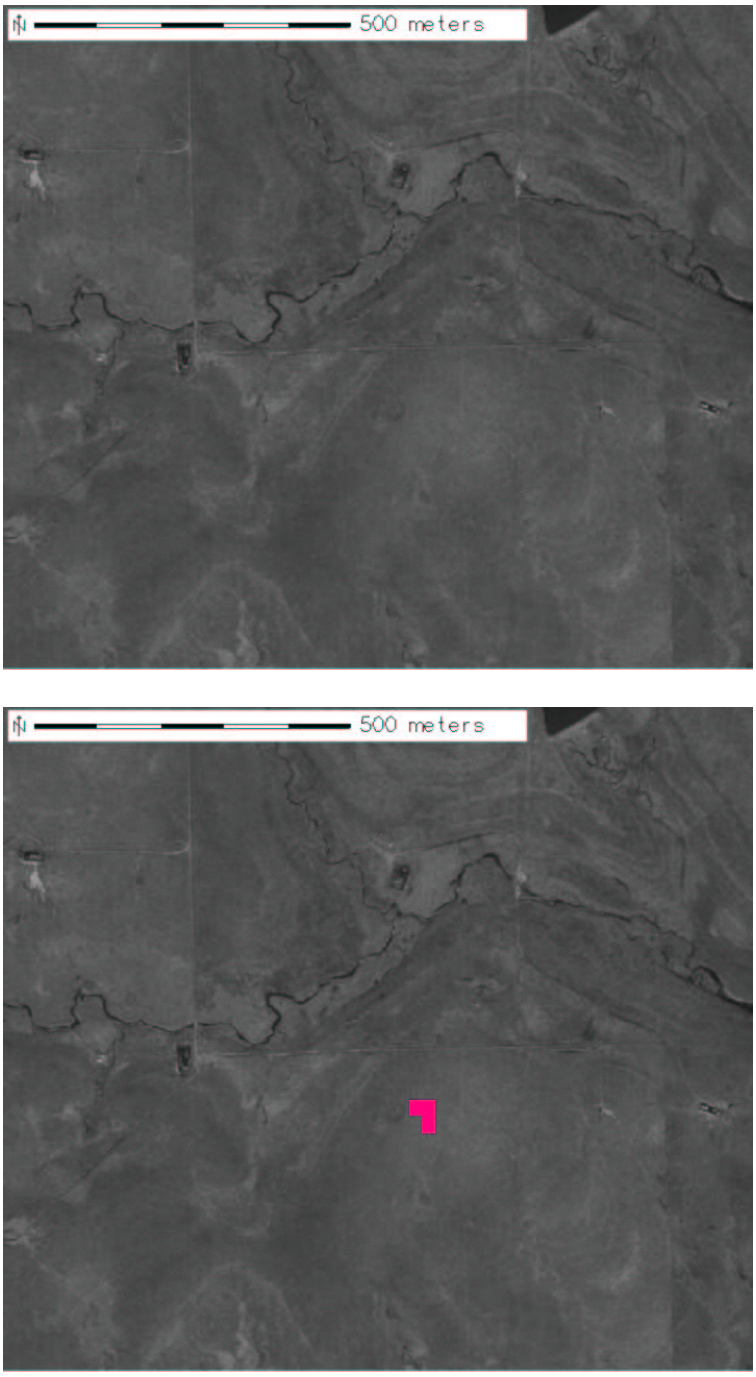


Figure A.67: Region: disturb12, Scene: r06sc05. Spot near well and tank areas.

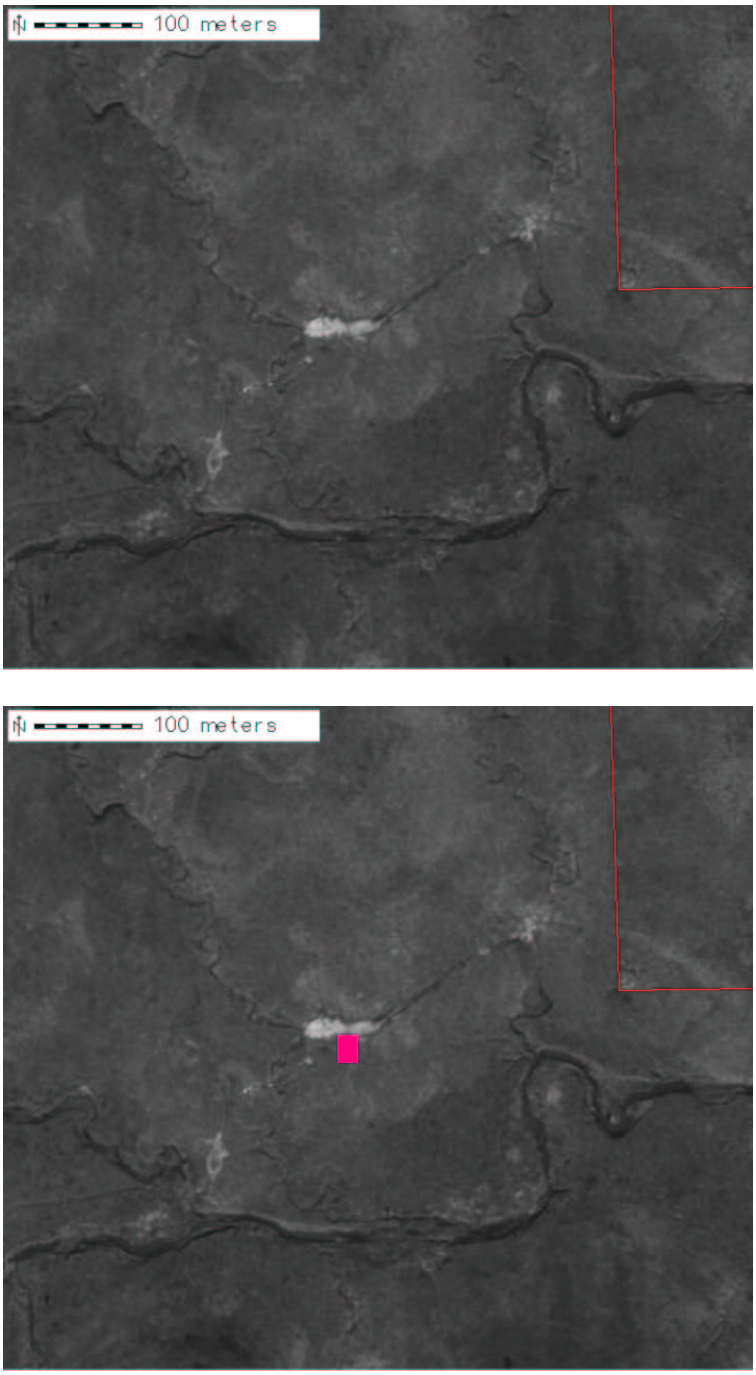


Figure A.68: Region: creek_to_creek_spill, Scene: r06sc05. Spill from one well drained to and eroded the bank of a secondary stream.

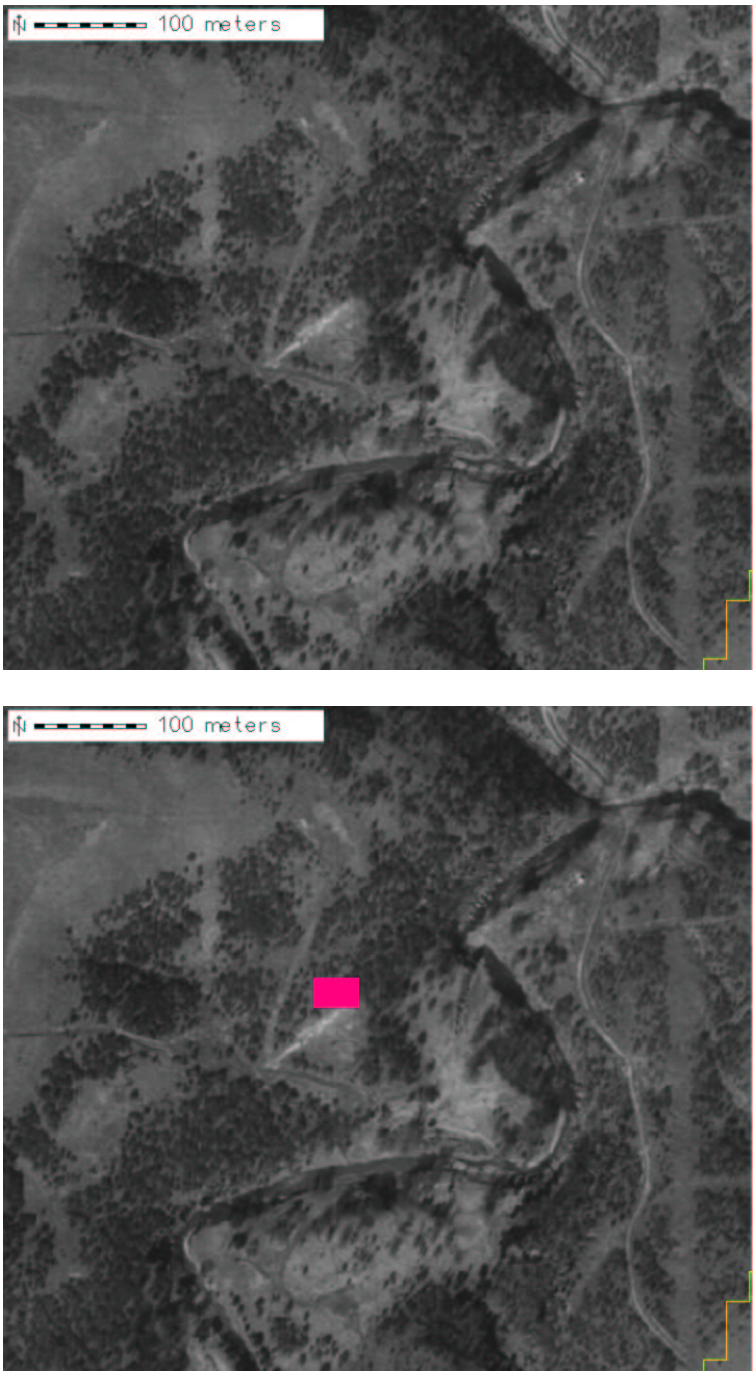


Figure A.69: Region: tank_wash, Scene: r06sc05. Runoff from tank which may have fed larger area in region possible runoff.

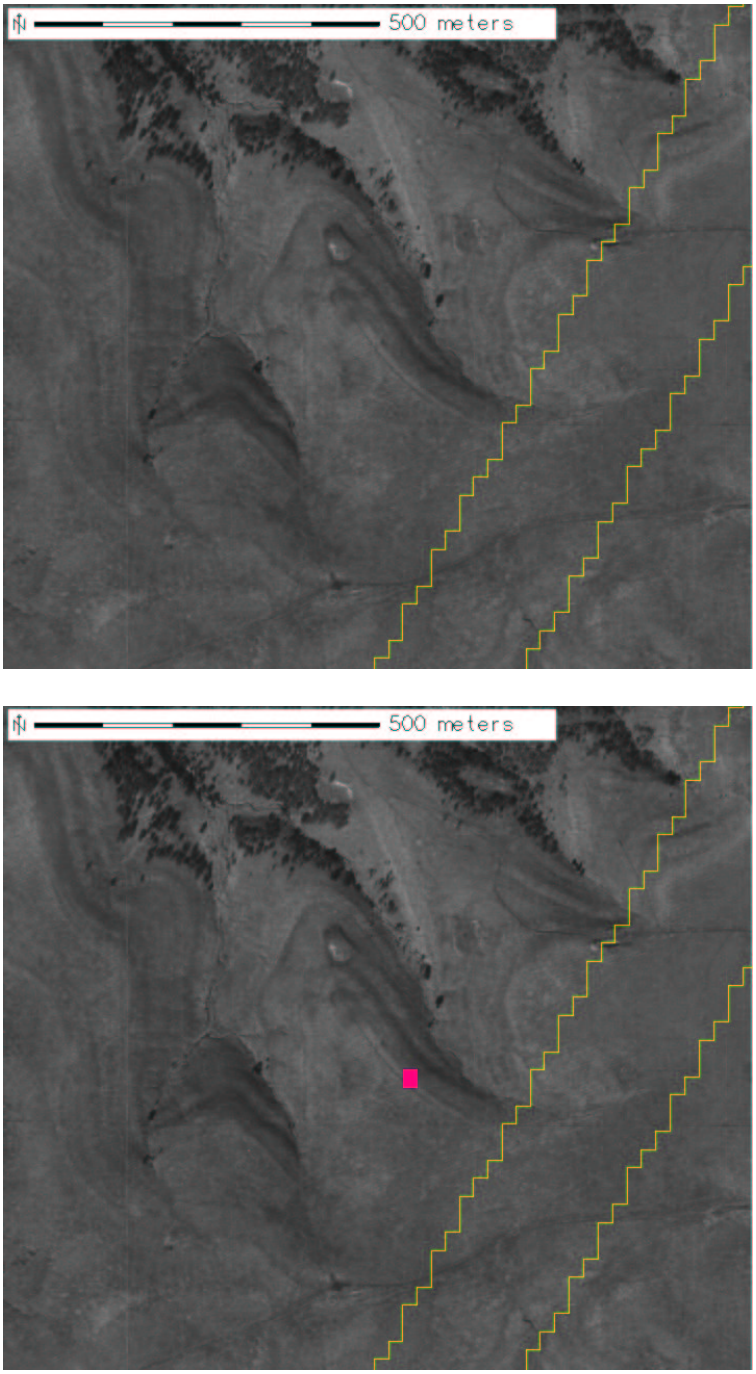


Figure A.70: Region: puka, Scene: r06sc05. Runoff from possible well (also contained in possible runoff).

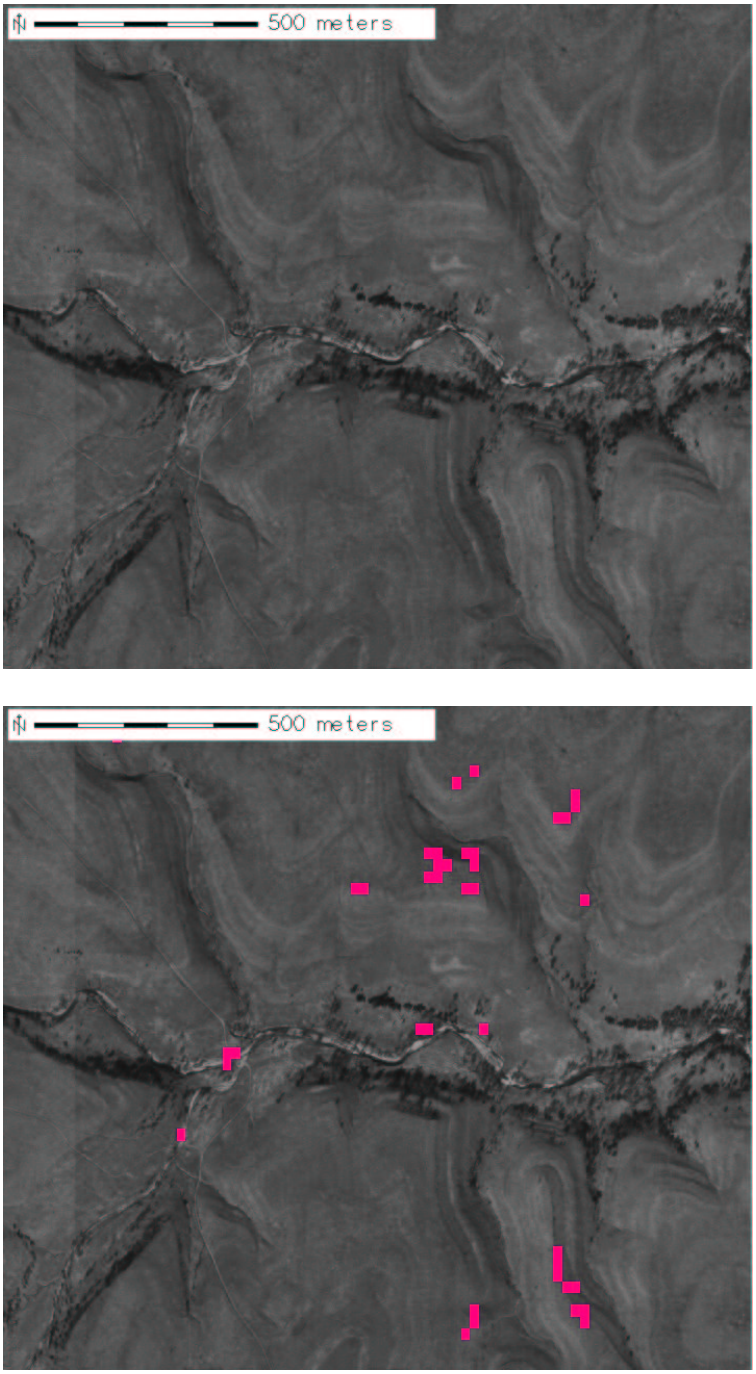


Figure A.71: Region: lower_stream, Scene: r06sc05. Sandbars, two bridges, and erosion/outcrop spots (also contained in possible runoff).

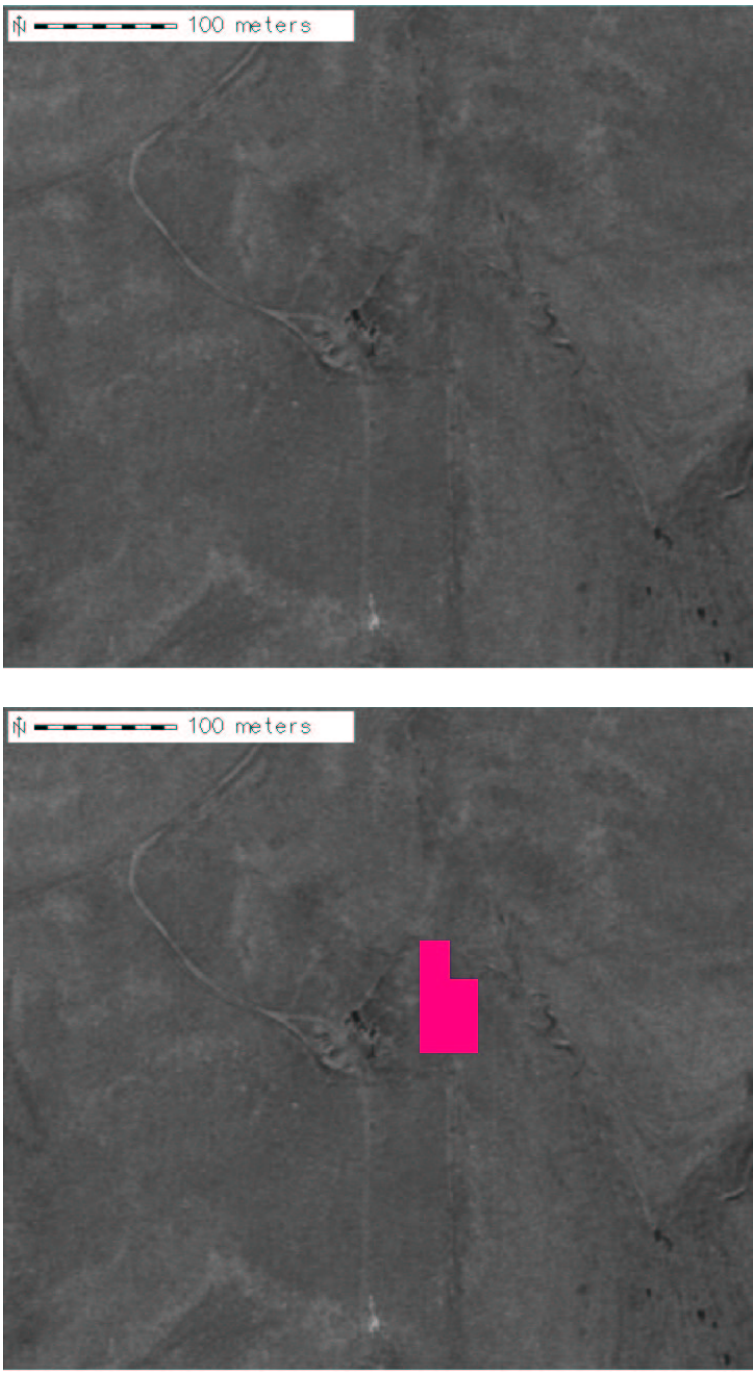


Figure A.72: Region: disturb13, Scene: r06sc05. Large tank or well facility.

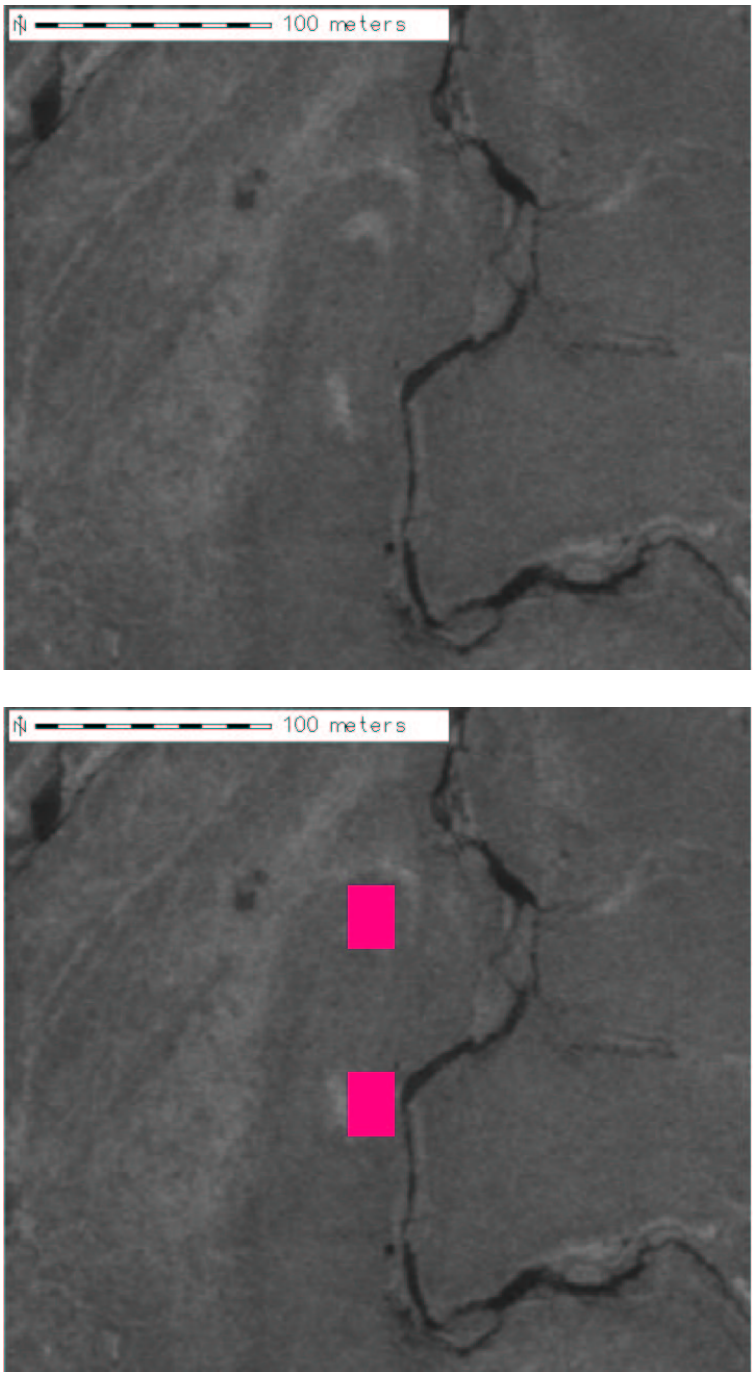


Figure A.73: Region: erosion6, Scene: r06sc05. Two bright spots near river.

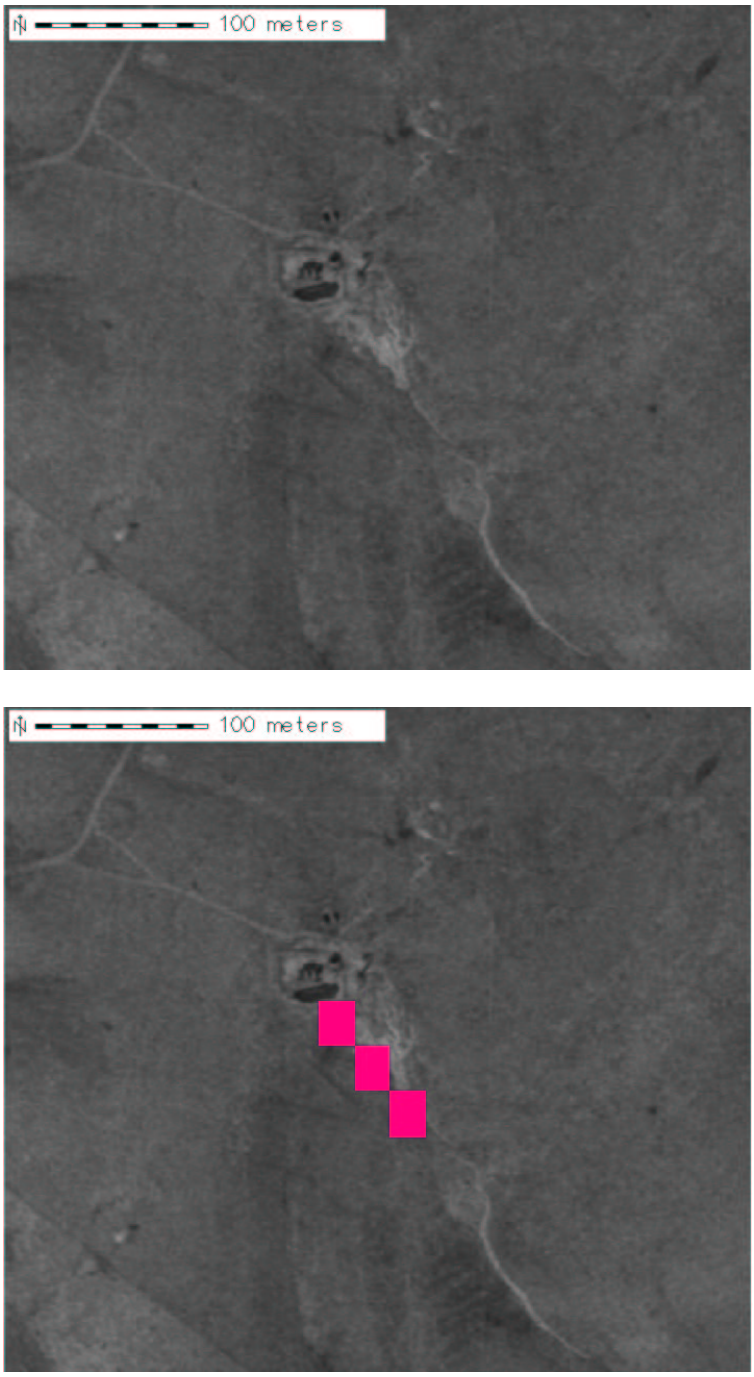


Figure A.74: Region: spill_drainage, Scene: r06sc05. Drainage from wells or tanks.

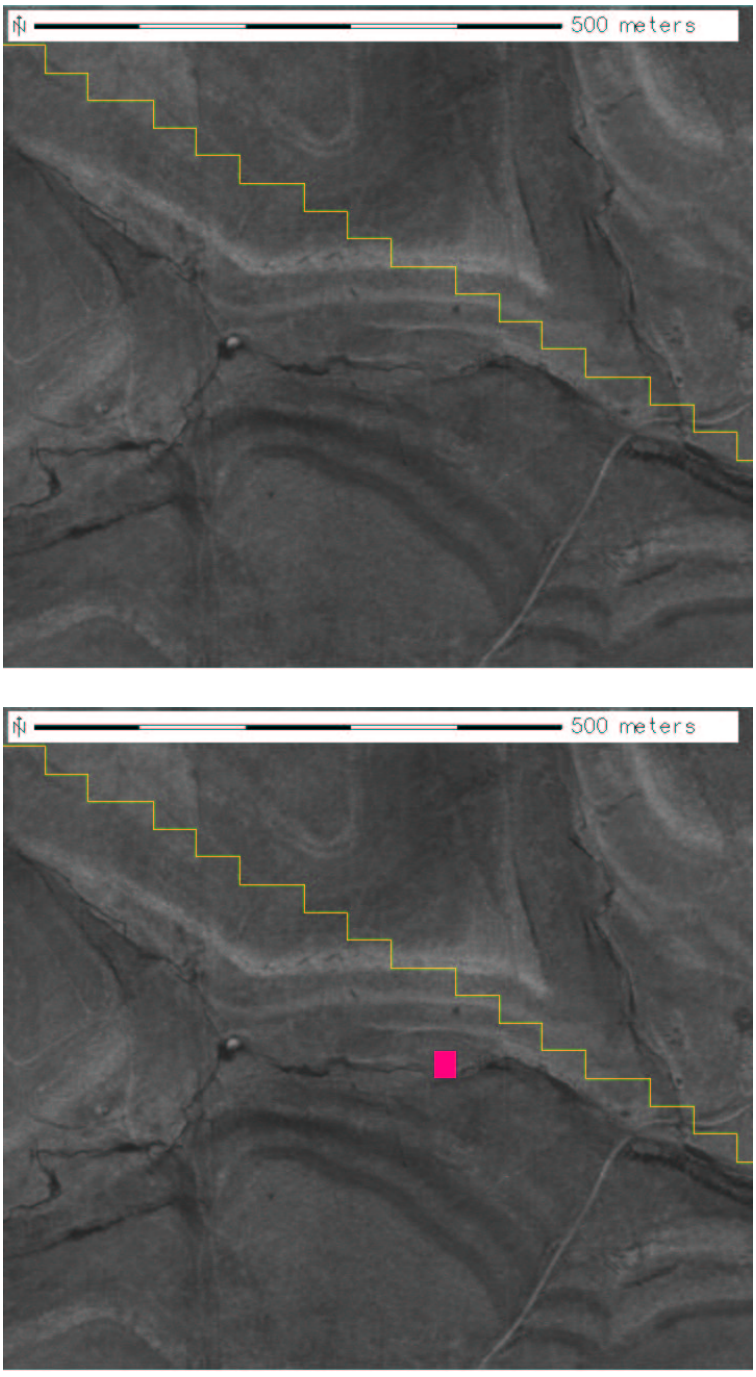


Figure A.75: Region: by_river_spot, Scene: r06sc05. Spot near river.

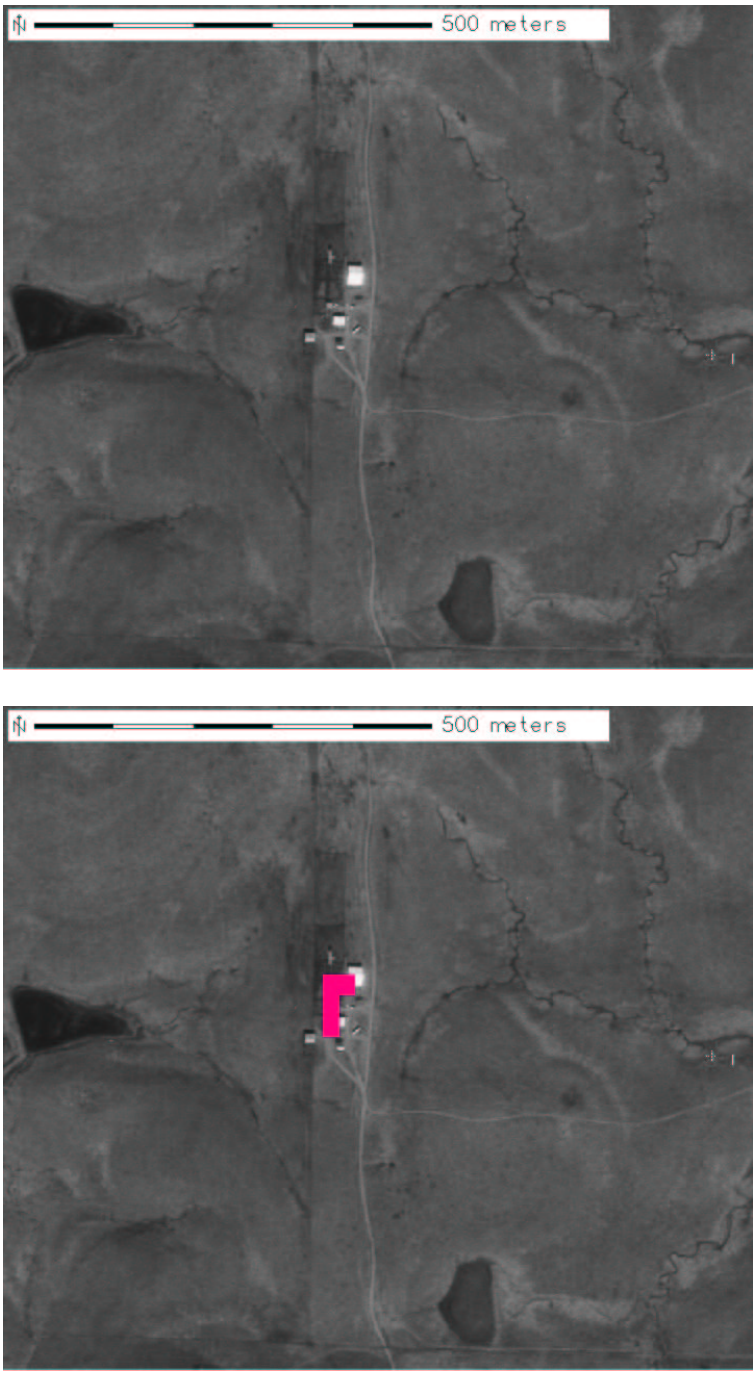


Figure A.76: Region: compound6, Scene: r06sc05. Compound with multiple driveways.

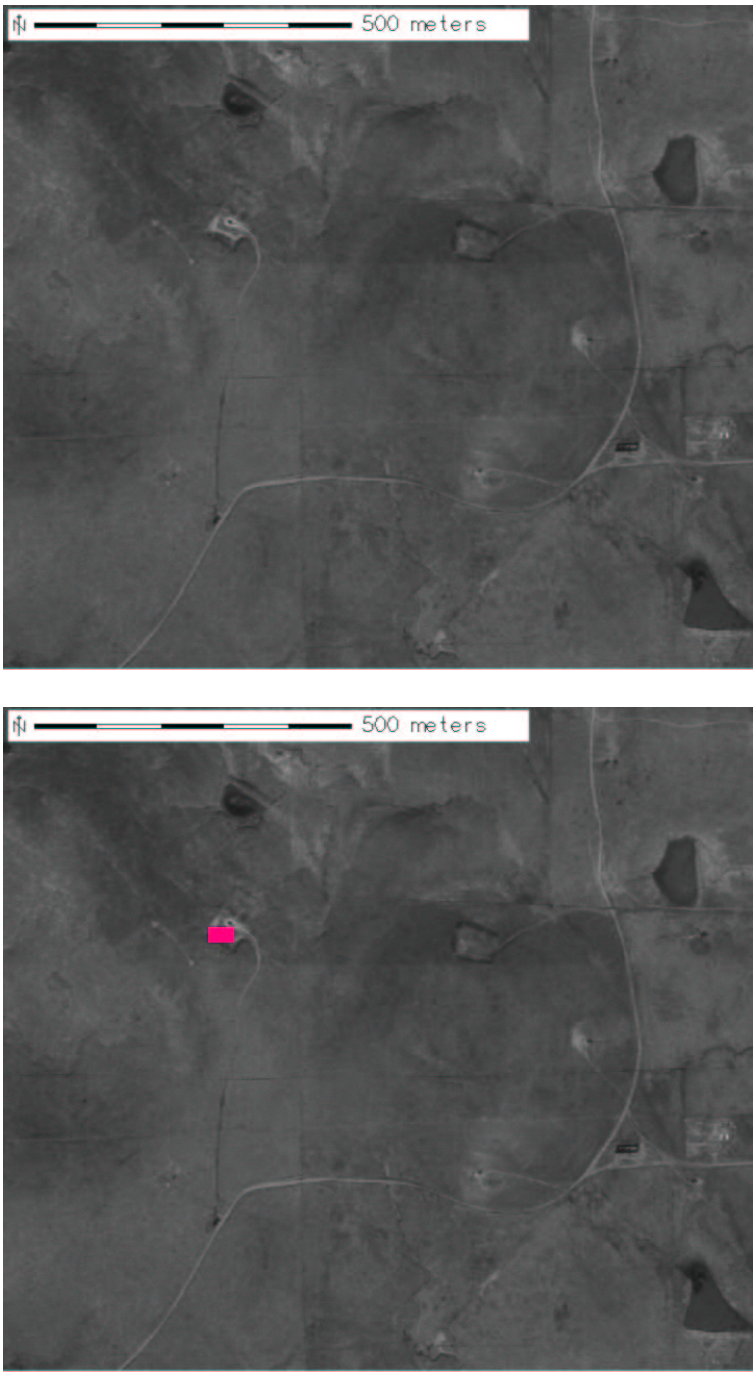


Figure A.77: Region: jackpump1_new, Scene: r06sc05. Well area.

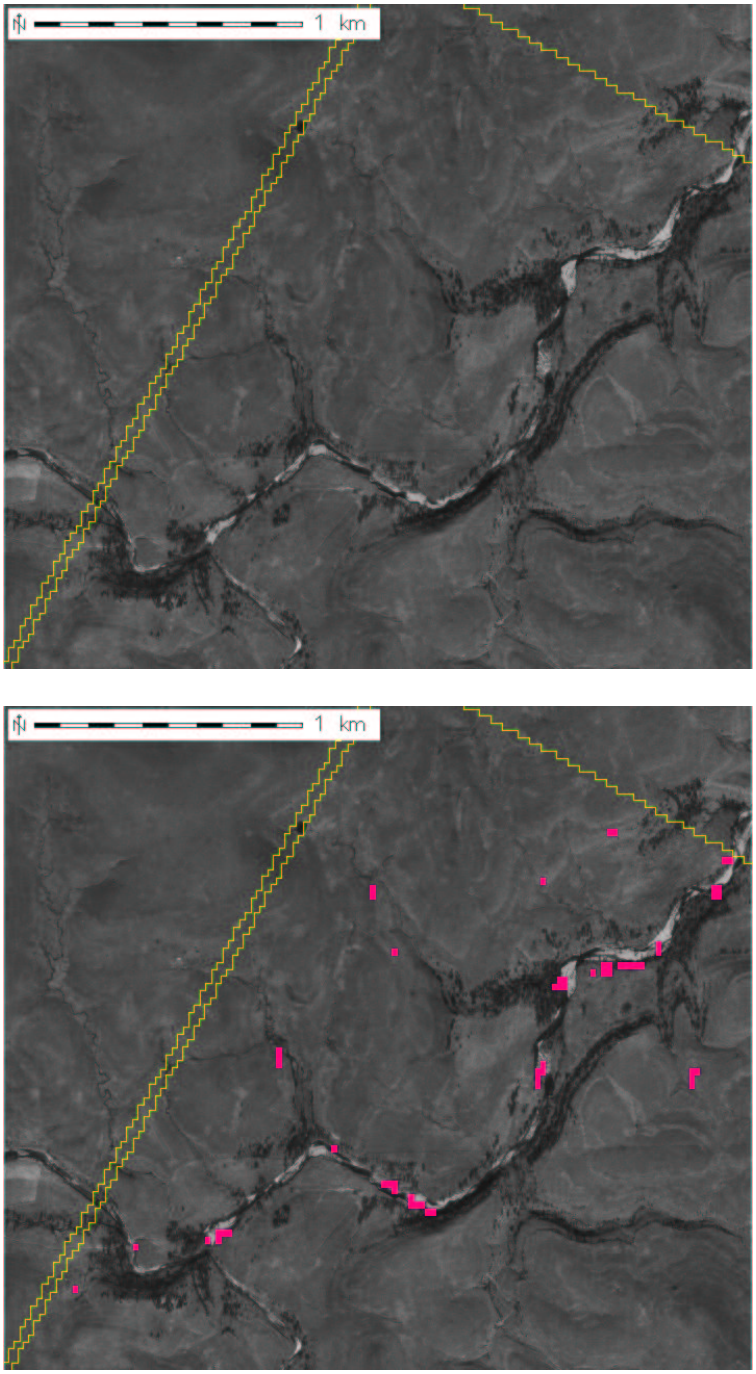


Figure A.78: Region: sandbars_new, Scene: r06sc05. A bridge, many sandbars, and erosion/outcrops with possible old well site north of river.

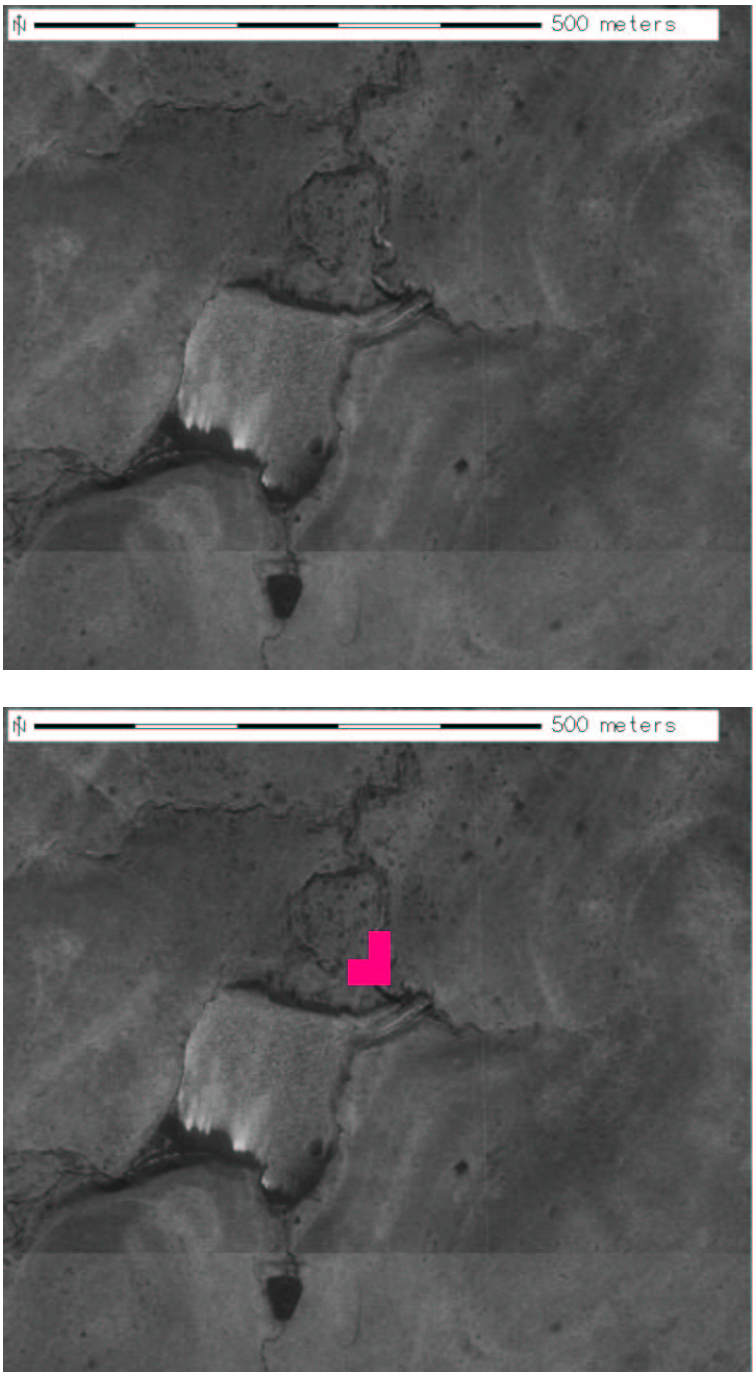


Figure A.79: Region: erosion3, Scene: r06sc04. Drainage area into pond.

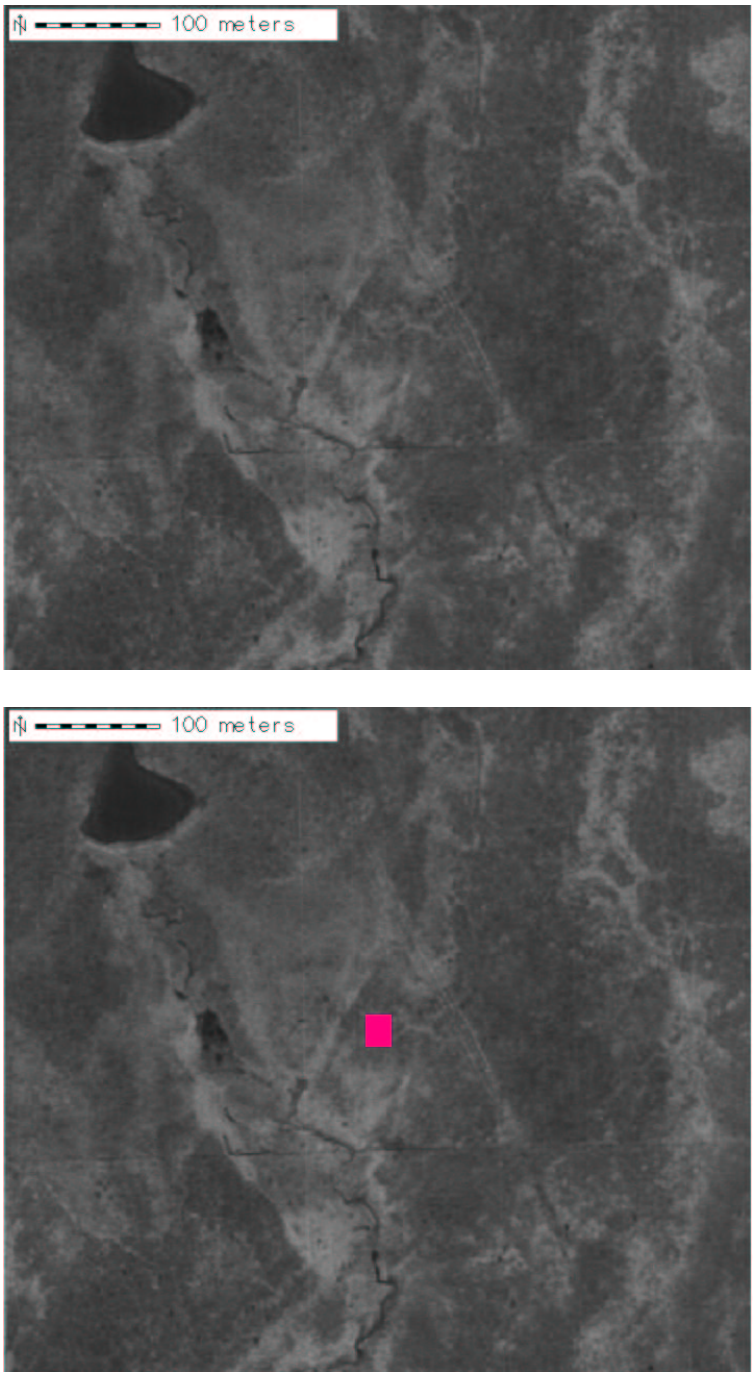


Figure A.80: Region: erosion4, Scene: r06sc04. Erosion or drainage near back road.

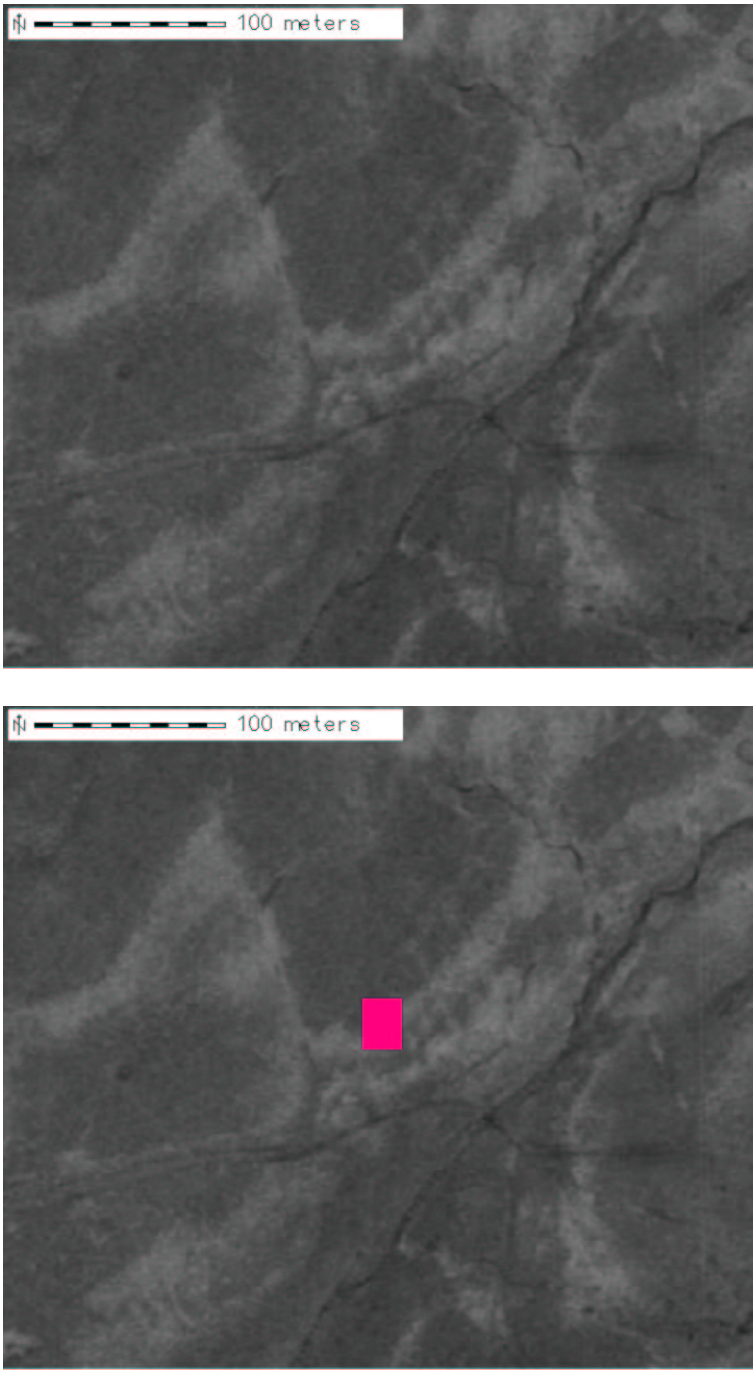


Figure A.81: Region: erosion5, Scene: r06sc04. Erosion area near back road intersection.

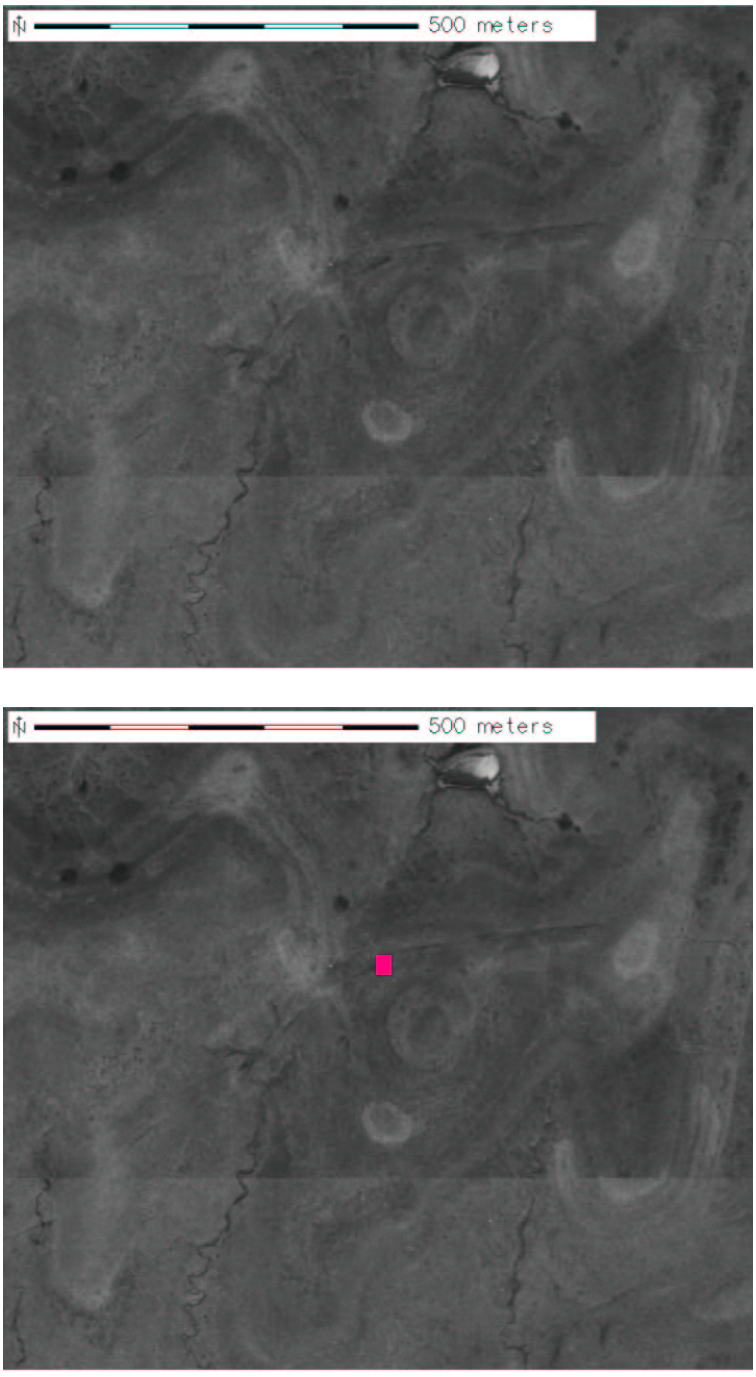


Figure A.82: Region: erosion9, Scene: r06sc04. Area near small shack or tank.

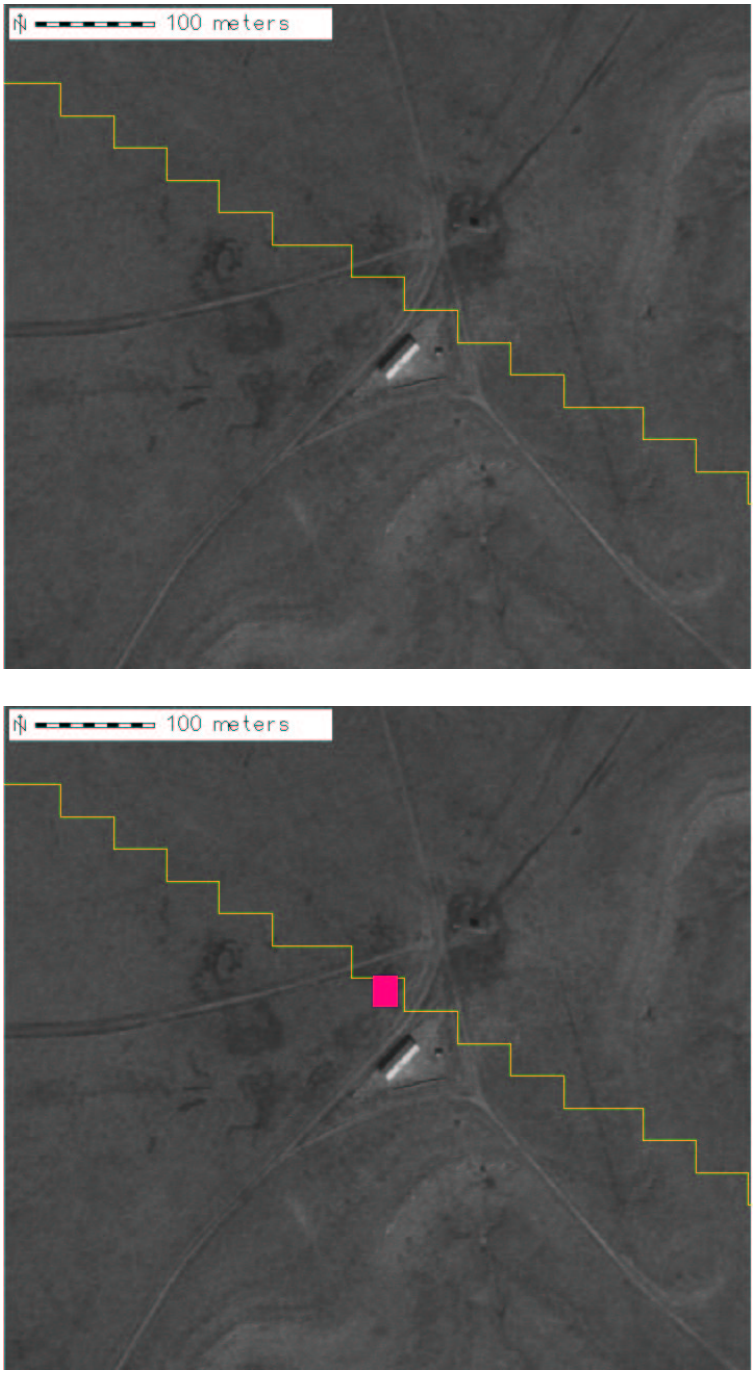


Figure A.83: Region: intersection3, Scene: r06sc04. Long building at multi-way intersection.

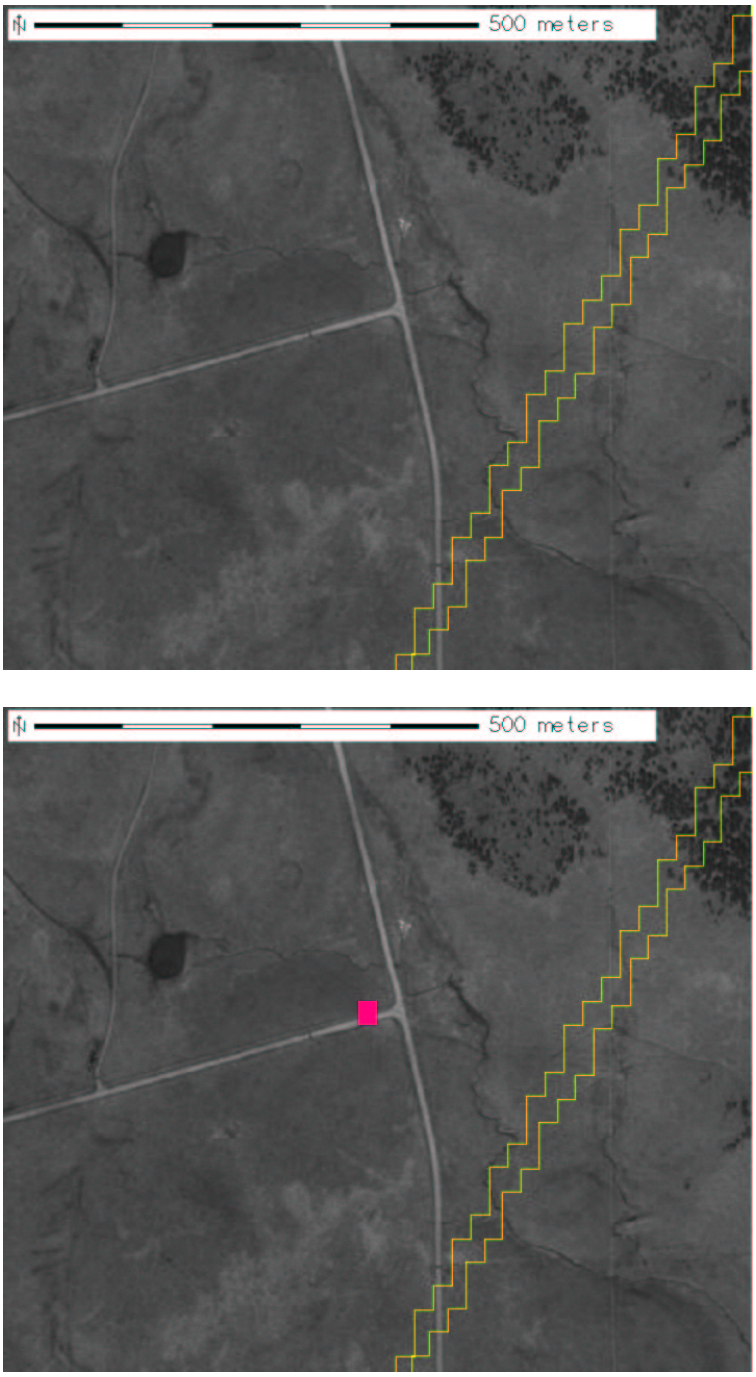


Figure A.84: Region: intersection4, Scene: r06sc04. Major roadway intersection.

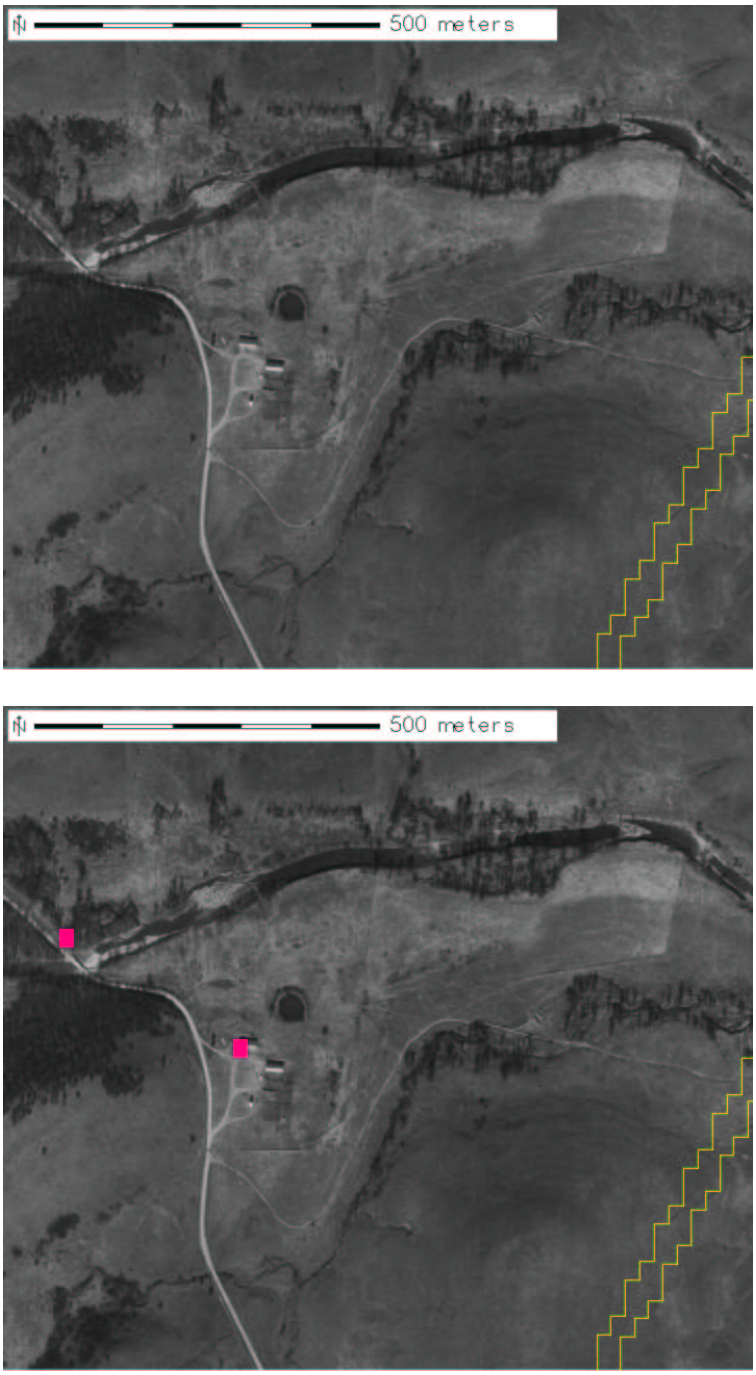


Figure A.85: Region: bridge_and_farm_river, Scene: r06sc04. Bridge over river with sandbars and ranch with large driveways.

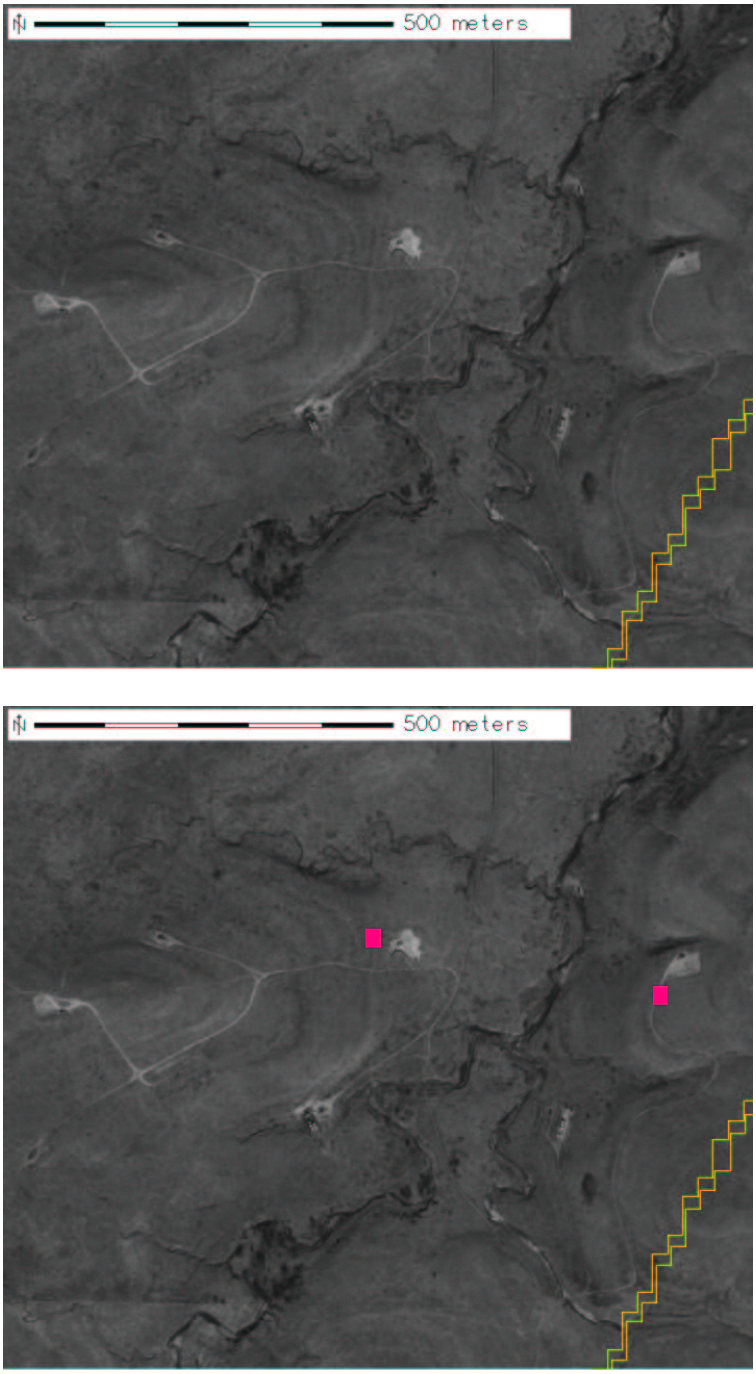


Figure A.86: Region: two_spills, Scene: r06sc04. Two obvious brine scars around wells.

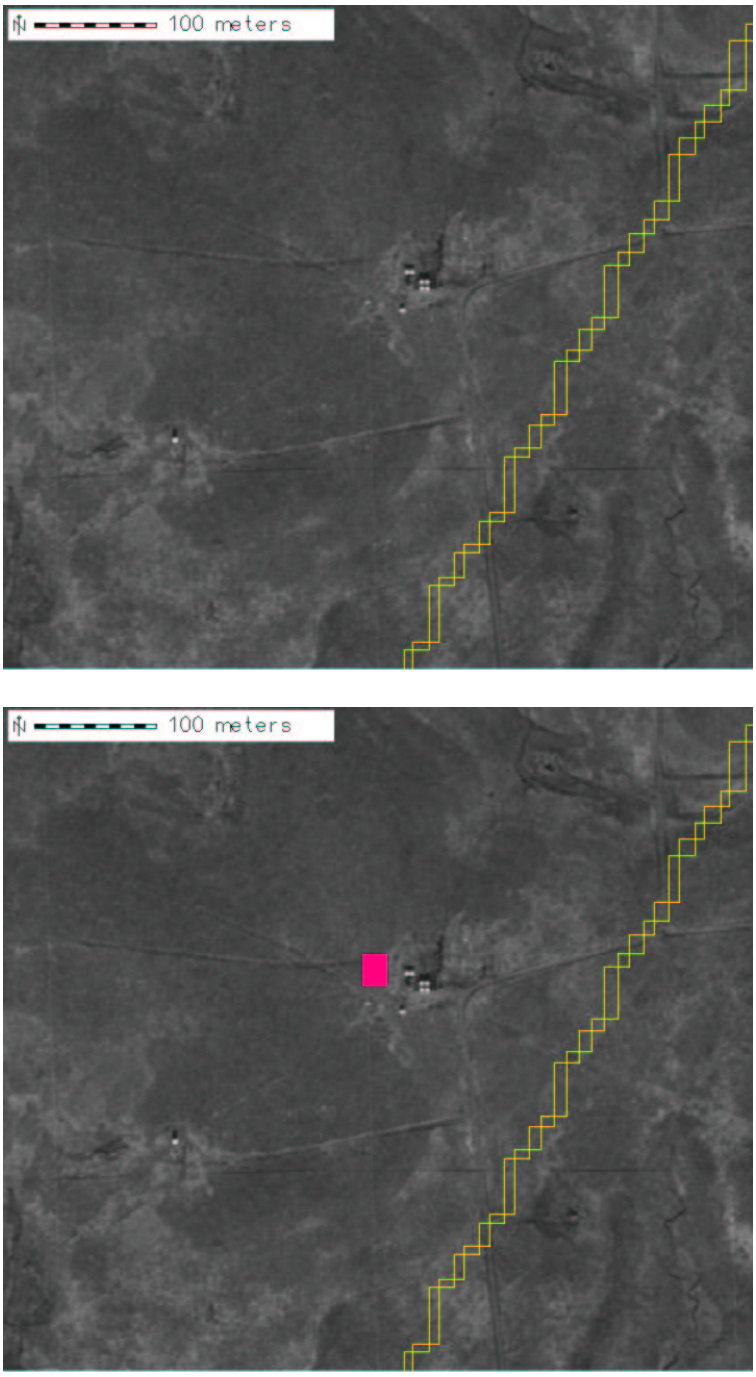


Figure A.87: Region: set_of_tanks, Scene: r06sc04. Cluster of holding tanks.

Vita

Forrest Hoffman is a research scientist at Oak Ridge National Laboratory where he holds joint appointments in the Computer Science & Mathematics and the Environmental Sciences Divisions. At ORNL Forrest established the first World Wide Web site at the Laboratory in 1995 and built ORNL's first Beowulf-style parallel computer, called The Stone SouperComputer, in 1997. A member of Sigma Pi Sigma ($\Sigma\Pi\Sigma$), the American Geophysical Union, the Institute of Electrical and Electronics Engineers (IEEE) Computer Society, and the Smoky Mountain Chapter of the American Meteorological Society, Forrest performs computational science research in global climate, landscape ecology, and terrestrial biogeochemistry on Linux clusters as well as some of the world's largest supercomputers in ORNL's Climate & Carbon Research Institute (CCRI). Forrest writes a monthly column for *Linux Magazine* called "Extreme Linux" and presently serves on the Advisory Committee for Advanced Research Computing (ARC) at Georgetown University.