Using Remote Sensing to Map Vegetation Density on a Reclaimed Surface Mine[1]

Michael Shank[2]

Abstract. The West Virginia Department of Environmental Protection, in a cooperative agreement with the Office of Surface Mining's Charleston Field Office, is evaluating the utility of high resolution satellite images for characterizing vegetation patterns on reclaimed surface mines. This paper details the results of the first phase of this project, which sought to determine whether satellite images could be used to estimate percentage vegetation cover.

Depending on the post-mining land use, percent vegetation cover may be employed as a standard for evaluating re-vegetation success prior to bond release. However, the increasing size of surface operations, combined with labor intensive field techniques required to estimate percent cover, make implementing the standard difficult.

This paper details a simple technique for estimating percent vegetation cover based on the widely-used Normalized Difference Vegetation Index (NDVI). NDVI exhibited a 0.96 correlation with percent vegetation cover for 34 reference samples collected on a 94 acre study area in southern West Virginia. Based on this relationship, a technique was developed that produced a mean error of 6.41% (+/- 2.68% at the 90% confidence level) when estimating percent cover for the 34 field sites.

[1]Paper was presented at the 2008 "Incorporating Geospatial Technologies into SMCRA Business Processes", March 25 – 27, 2008, Atlanta, GA.

[2]Michael Shank is a geographer at the West Virginia Department of Environmental Protection, in Charleston, West Virginia.

Introduction

The re-establishment of vegetation on mined lands is an integral part of the reclamation process in Appalachia. Regulatory authorities monitor and evaluate vegetation re-growth with reference to standards which must be met prior to bond release. Two of these standards specify percentages of vegetative ground cover and stem counts per acre. This paper addresses the first problem—estimating percent ground cover. For our purposes, p*ercent ground cover* is defined as the percentage of an area on the earth's surface occupied by vegetation, including grasses, herbaceous plants, shrubs, trees, and litter (Farmer *et. al*. 1981). In other words, it is the total portion of an area where vegetative material obscures the underlying substrate when viewed from above at an ordinary human scale of observation.

The field techniques used by regulatory authorities in West Virginia are laid out by Farmer *et. al*. (1981), who addresses the problem of sampling a contiguous tract of land in a statistically valid way to arrive at an overall estimate of percent ground cover. In contrast, the application of remote sensing to this problem does not involve sampling from a population in order to make an inference about the population as a whole. Rather, it calculates a percent cover estimate for each image pixel, where each pixel represents a minute section of the area being evaluated. This produces a continuous map of vegetation cover across an entire area. An overall percent cover statistic then can be estimated from the set of pixels located within the area of interest. The success of this approach depends on the ability to develop a suitable technique that produces reliable results with minimal error. This problem forms the basis for this paper.

Characterizing vegetation is a persistent theme in remote sensing research. Techniques for monitoring stress and disease, estimating crop yield, estimating biomass and leaf area index, and mapping vegetation classes abound in the literature. These techniques were developed largely for analysis of medium resolution sensor products, suitable for characterizing relatively large areas. However, the advent of commercial high-resolution imaging satellites presents new opportunities for characterizing vegetation in greater detail than previously was practical. As an example, Digital Globe's Quickbird satellite effectively collects more than 1,600 samples per acre of multi-spectral information over a potential mine site, which may be useful for mapping disturbed areas, vegetation types, coal stockpiles, ponds (and perhaps seeps) and other features.

This paper represents the first deliverable in a cooperative agreement between the West Virginia Department of Environmental Protection and the Office of Surface Mining, Charleston Field Office, to examine the potential of analyzing high-resolution satellite images to characterize vegetation on reclaimed mines. In addition to estimating vegetation cover and mapping vegetation types, initial investigations suggest high-resolution images can be useful for monitoring vegetation change over time, including overall biomass increase, expansion of particular species, and die-off events.

Study Area

The analysis utilized a quickbird multispectral scene acquired on June 14th, 2007. The image encompassed 17 surface mining permits issued to Hobet Mining between 1977 and 2005. The total permitted area for this complex is approximately 18 square miles.

The image was ortho-rectified using the USGS 1/9 arc-second National Elevation Dataset for West Virginia. The elevation model was derived from stereo photography acquired in April 2003, which post-dated backfill and grading operations within the study area. The image initially was georeferenced using coincident points identified on high-resolution digital ortho-photography, and was later supplemented by ground control points obtained by GPS in the field.

The study area occupied an irregular area, approximately 94 acres in size, in the southern section of the West Ridge Surface Mine (figure 1). The area was selected because it included a range of revegetation conditions, from fully vegetated grass to mostly barren areas. An initial field reconnaissance was conducted on May 1st, 2007, which revealed that some of the mostly 'barren' areas actually appeared to be commercial forestry reclamation, in which tree plantings are accompanied by relatively sparse herbaceous species. Numerous examples of oaks, tulip poplar, pine and others were noted in these areas. The substrate for the entire area was predominately gray sandstone ranging from pebbles to fist size rocks, with occasional large boulders. Scattered, darker shale also was common. Moisture was not a factor affecting substrate reflectance, as the area was subject to a moderate drought at the time of acquisition.

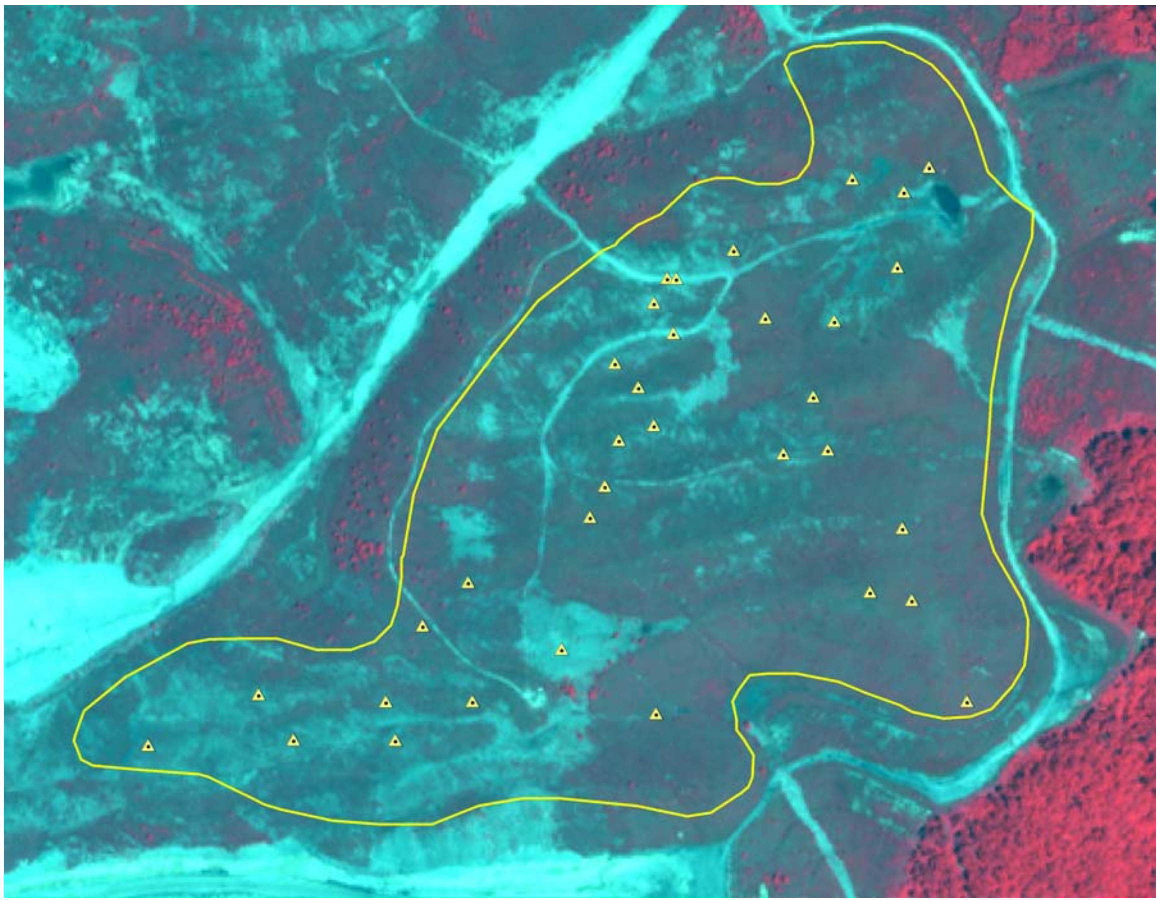

Figure 1. Percent cover study area. Yellow icons identify locations of reference samples acquired in the field.

Field Sample Design

Selection of field sample locations was guided by several compromising factors, including time constraints, accuracy issues, and the desire to sample a range of vegetation densities. Sample locations were identified from the satellite image by examining Normalized Difference Vegetation Index (NDVI) statistics calculated for a 5x5 moving window. NDVI is a common tool for identifying and characterizing vegetation. In this instance, NDVI served as a surrogate measure of vegetation density and homogeneity in the neighborhood surrounding an image pixel. The average NDVI value for the 5x5 window was used to stratify image pixels into ten groups representing conditions ranging from bare earth to fully revegetated. For each group, an equal number of image pixels were selected randomly from the set of pixels with a low neighborhood variance. Variance was used as a way to constrain the selection to pixels associated with relatively homogeneous patches of vegetation, and thereby limit the impact of positional errors. In other words, the area sampled on the ground should represent an area of relatively homogeneous vegetation cover that is larger than the expected shift in pixel location due to positional errors in the image.

The selected pixel coordinates were extracted from the image and uploaded to a GPS unit for navigation in the field. Field data collection consisted of navigating to a sample site, making a visual confirmation of site suitability, and taking a set of vertically-oriented digital photographs. For each site, the camera was positioned vertically using a bubble level at approximately chest height. Typically, a total of five photographs were collected at a location, representing the center and four corners (figure 2). The corners were located by counting paces on a compass bearing from the center point. Absolute precision was not required in capturing the four corner photos. However, the center of the site was fixed using a Trimble geo-XH receiver in open sky conditions. The requirements for collecting 5 photographs at each site were relaxed for sites what were completely barren or unquestionably revegetated, in which case the GPS position was acquired and the vegetation condition was simply noted.

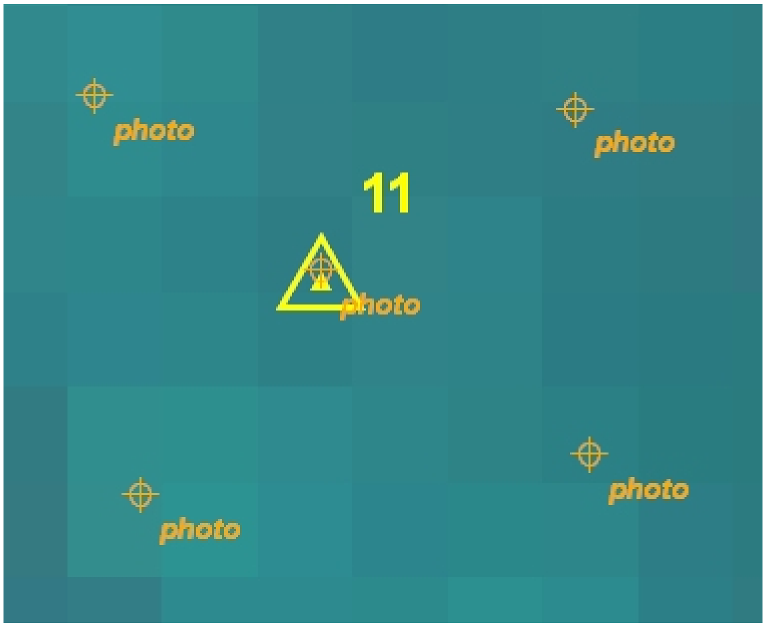

Figure 2. Pattern of field photos taken for sample 11.

The camera used for field photographs was a Casio 500SE, which produced an image covering approximately 1.6m^2 at chest height. A 5x5 pixel block represents 144m^2, so the five images represent about 5.5% of the total sample area.

Using digital photography maximized productivity in the field, allowing more sites to be visited during the site visit. The photographs provided a mechanism for estimating percent cover that was more rigorous than visual estimation. Once collected, field photos were processed to create a binary vegetation mask that matched patterns of vegetation in each photo (figure 3). The masks then were used to calculate the proportional area covered by vegetation in the five photos and assigned as the percent cover value for the site. Time in the field allowed for the collection of 34 sample sites within the study area.

Delineating the vegetation mask manually was a very labor intensive process. Early in the study, experiments were performed with infrared images, taken with an infrared filter, which could be composited with a corresponding color image. These composites were used with an unsupervised classification algorithm to identify the extent of vegetation in the photo. While the results were encouraging, the practice was somewhat cumbersome as it required the use of a camera tripod to capture two photos of each scene.

This reduced the number of field sites that could be sampled within a given time period, and for this reason, the manual process of delineating vegetation was used instead.

Figure 3. Field photo with manually interpreted vegetation mask overlay in red.

Relationship Between NDVI and Percent Cover

Vegetation indices, particularly the Normalized Difference Vegetation Index (NDVI), have been applied to a wide variety of remote sensing vegetation studies. Vegetation indices exploit the characteristic of vegetation to reflect significantly more light in the near-infrared portion of the electromagnetic spectrum than adjacent red frequencies (figure 4). In contrast, substrate materials tend to reflect similarly in both segments.

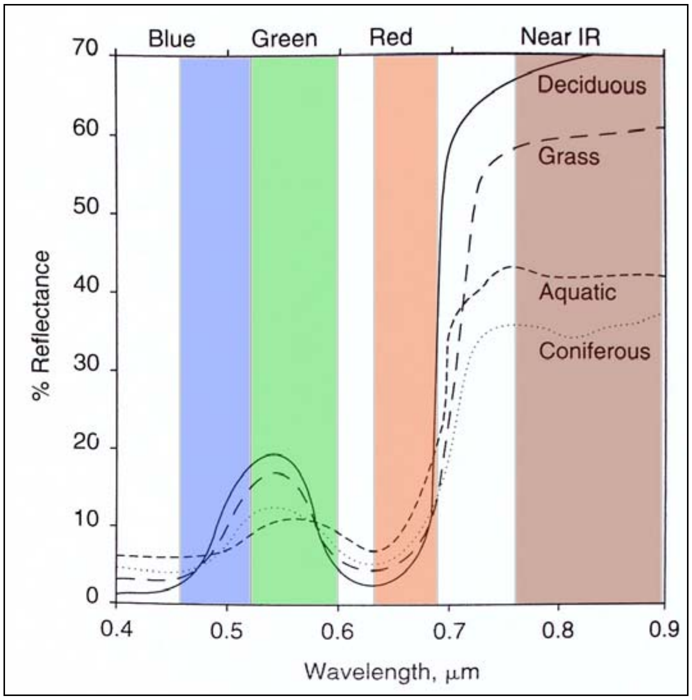

Figure 4. Typical vegetation reflectance patterns, superimposed with sensor bands associated with the quickbird multi-spectral sensor

NDVI has been used to estimate leaf area index, to reveal stressed vegetation, to identify deforestation, and to monitor desertification (Avery and Berlin, 1992). It also has been used to identify trace quantities of vegetation (Elvidge *et. al.*, 1993) and estimate percent vegetation cover (Purevdorj *et al.*, 1998) using AVHRR data. NDVI is specified as:

$$NDVI = (NIR - RED) / (NIR + RED) \qquad \textbf{(1)}$$

Where *NIR* is the recorded radiance in the near infrared, and *RED* is the recorded value in the red portion of the spectrum for a particular image pixel. NDVI values for non-vegetation typically produce small or slightly negative values, while vegetated areas produce values starting around 0.4 and approaching 1.0.

Figure 5 depicts the relationship between NDVI and percent cover for the 34 field sites. The solid line traces the best fit equation calculated using simple linear regression:

$$PCT\_COVER = -0.140224 + 2.5886 * NDVI \qquad \textbf{(2)}$$

The equation produces an R^2 value of 0.9195. Residual errors for this model ranged from 0 to 20.27%, averaging 6.78% for the entire sample set (RMSE was 8.66%).

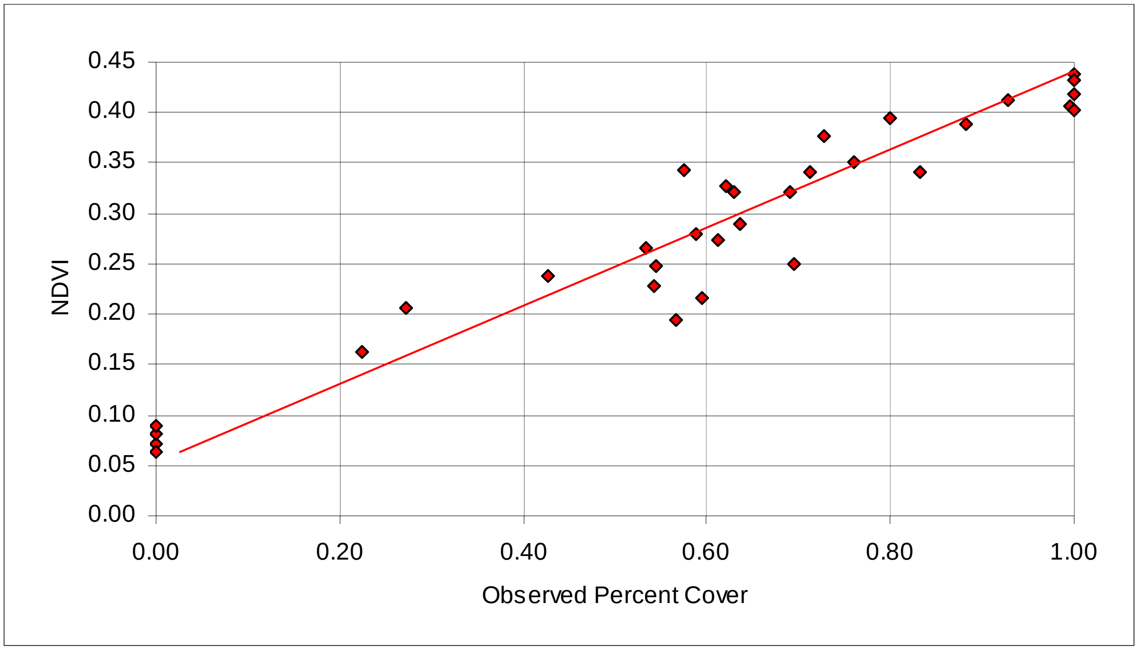

Figure 5. Relationship between NDVI and percent vegetation cover for field samples.

Linear NDVI Model

The reasonable correlation between NDVI and percent cover observations suggests percent cover can be estimated in a scene, provided a reasonable model of the relationship between NDVI and percent cover can be constructed. One possibility is to fix the endpoints of the model, then interpolate linearly between the two extremes. In practice, the endpoints can be estimated by averaging NDVI values from an unvegetated area, and those from a fully revegetated area. Both sample areas can be interpreted from the satellite image itself, and optionally confirmed in the field.

To implement this idea, two sample areas were selected within the study area, one barren of vegetation and one completely revegetated. The sample areas yielded 80 unvegetated and 62 fully vegetated samples for which NDVI values were calculated and averaged for both conditions. NDVI values averaged 0.089 for the unvegetated samples, and 0.44 for the revegetated samples. Intermediate values were calculated as a simple linear scaling between these two points. Vegetation percentages calculated from this model then were compared to the 34 field sites. Figure 6 shows the relationship between percent cover estimated from the linear model (on the y axis) and the observed percent cover (x axis). The red line in figure 6 is the linear model calculated from the two sample areas. Figure 6 presents a similar pattern to what was observed in figure 5, which appears to indicate that errors in this approach are associated with limitations in correlating NDVI and percent cover, or with field observations, rather than with the model itself.

Residual errors generated by the model ranged from 0.12% to 26.47% for individual samples, with an average of 7.27% and an RMSE of 9.81%. Figure 7 is a histogram of the error distribution of the sample set, showing that 27 of the 34 samples exhibited errors of less than 10 percent.

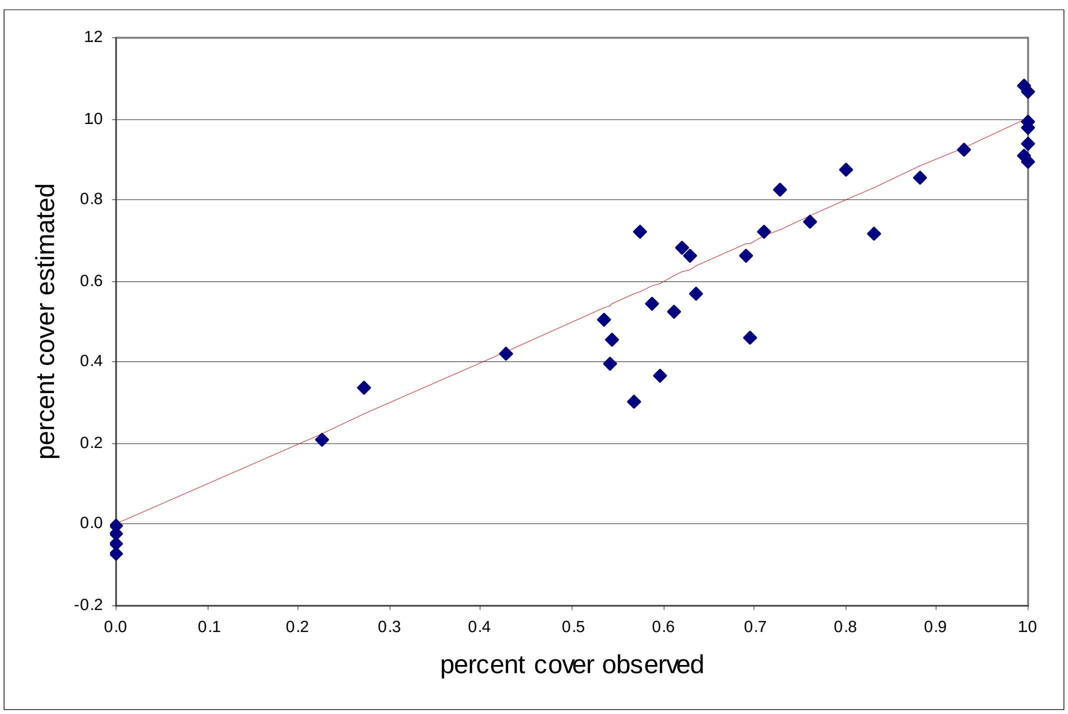

Figure 6. Relationship between percent cover estimates using the linear model and percent cover observed for field samples.

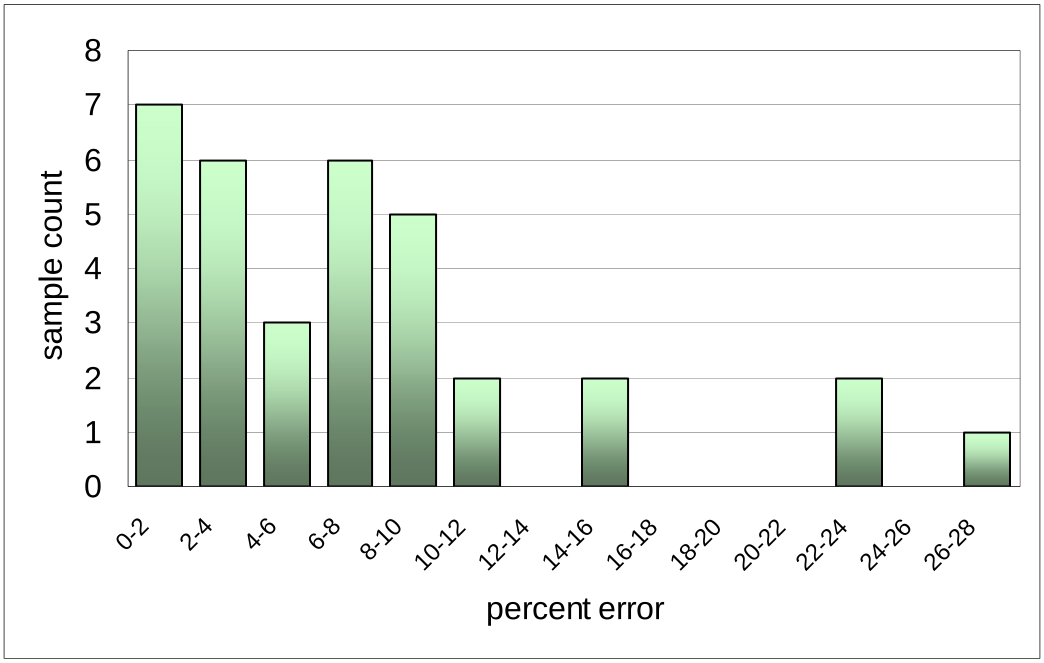

Figure 7. Error histogram associated with the linear model.

Six of the reference samples produced estimates that were outside the allowable range of 0..1. This result is not particularly surprising—since an average was used to fix the end points of the model, about half of the reference samples themselves could be expected to produce values in excess of 1, or less than 0. Additionally, relatively large values would be expected to result from tall dense grasses, or deciduous tree canopy. Though these areas contain a higher biomass, and more reflective leaf area, they are not 'more revegetated' than a short grass area that completely covers the substrate material. Reclassifying negative values to 0, and values greater than 1 to a value of 1, should not affect the validity of the result, and is essential for estimating area statistics. After reclassifying the out-of-bounds results, the mean error for the reference samples dropped to 6.41%, with an RMSE of 9.51%.

Selection of an appropriate area to calculate the revegetated mean is a subject of importance. What the model requires is essentially the *lowest* NDVI value that is representative of a fully revegetated area. Therefore, an area of short grass is better than and area of dense high grass, because the dense high grass will initially scale to over 100%, which as we have seen is not a problem. However, selecting an area of dense high grass could cause a significant under estimate throughout a scene. For example, mean error for the 34 field sites increased to 12.5% (+/- 4.43%) when a sample area was selected from an area of dense vegetation within the study area, producing underestimates for 32 of the 34 field sites. An experienced remote sensing analyst may be able to make an informed decision about where to sample in order to find an optimal NDVI value, though a short field reconnaissance of the area would be extremely valuable.

The model was used to produce a percent cover map of the entire study area, shown in figure 8, which clearly shows areas of sparse vegetation, including roads, a pond (northwest section, in purple), and the intentionally sparse vegetation associated with commercial forest reclamation in the westernmost part of the study area.

Percent cover estimates for the entire study area initially ranged from –0.47 to 1.83 (figure 9) before the out-of-bounds values were adjusted. The vast majority of the out-of-bounds estimations occurred at the high end of the scale, and many of the highest values were associated with early invader trees, such as ailanthus, and the black locust shown in figure 10. The photo in figure 10 is identified by the yellow marker in figure 11, which illustrates the relationship between very high percent cover estimates (over 1.25, shown as a semi-transparent green overlay) and individual tree canopies. The lowest negative values were associated with a small pond.

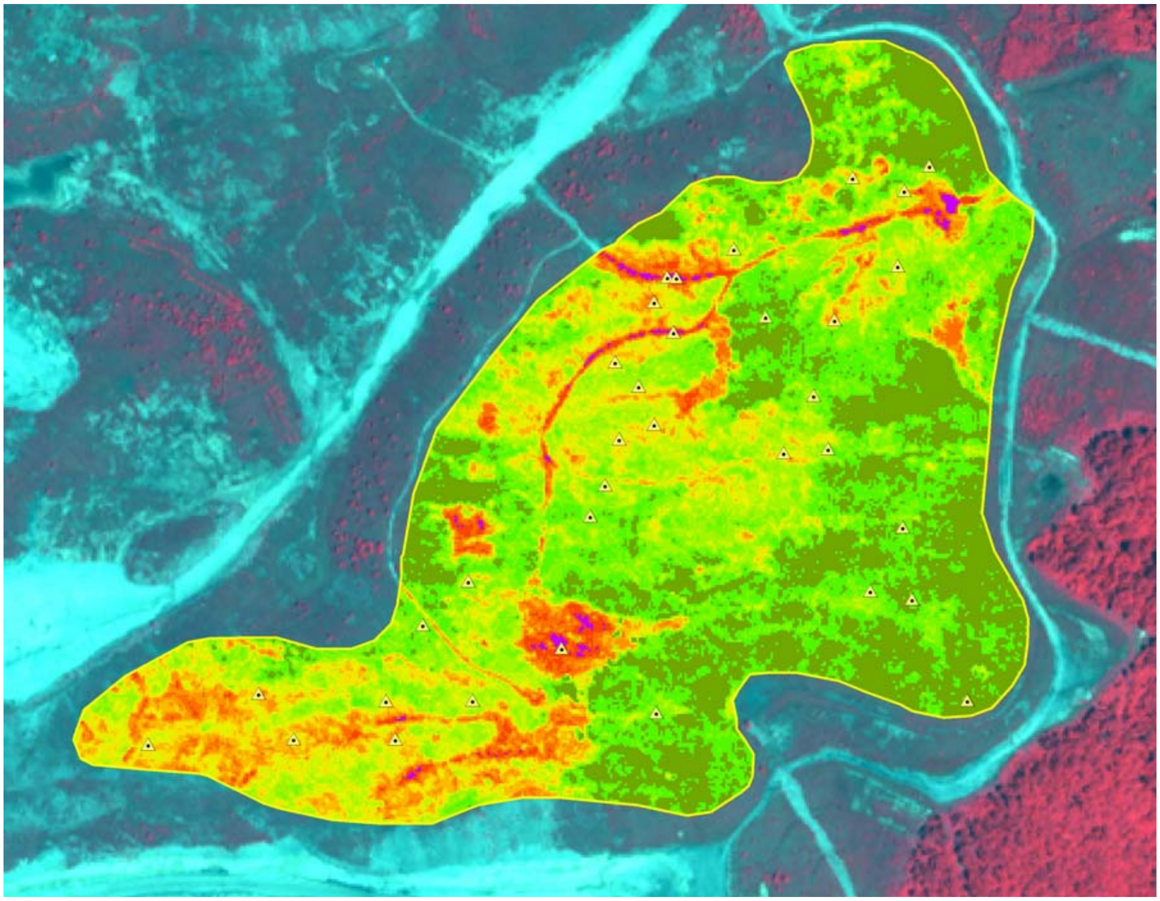

Figure 8. Percent cover map produced from NDVI measurements. Percent cover for the entire area was estimated to be 73.7%.

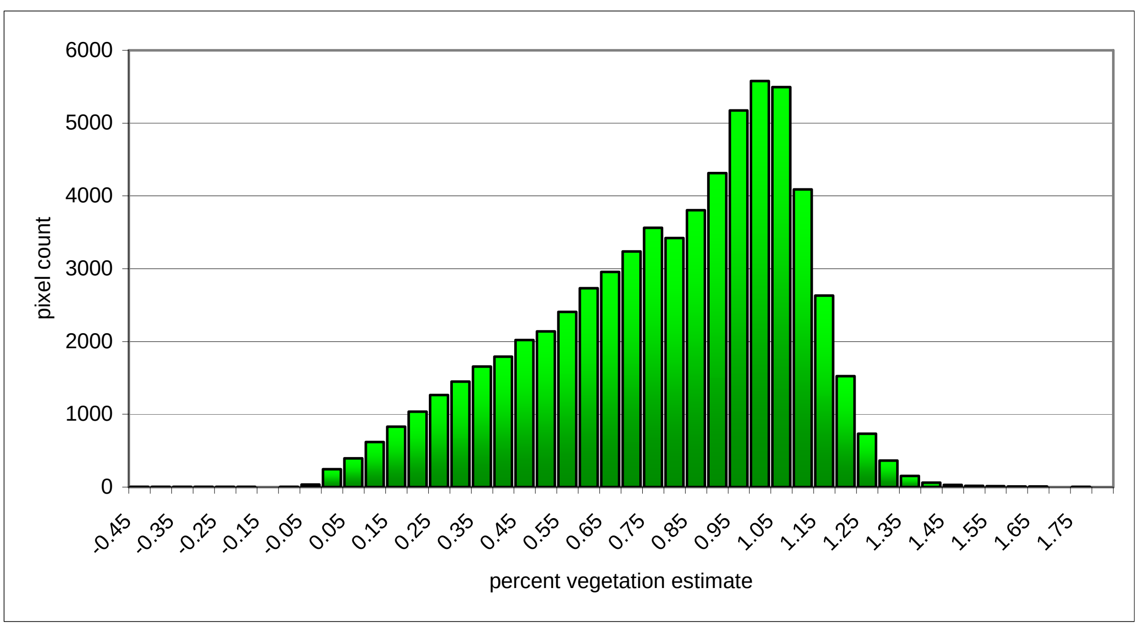

Figure 9. Histogram of percent cover for pixels within the study area. Categories over 1.0 typically are associated with dense vegetation and tree canopy.

Figure 10. Field photo of black locust.

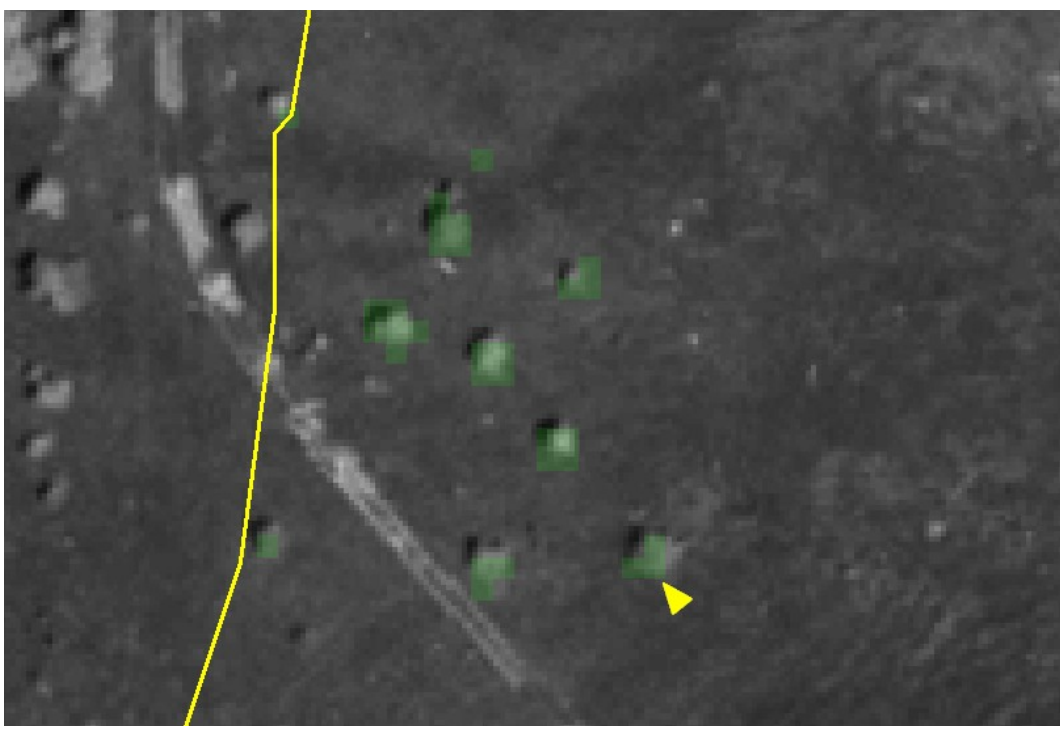

Figure 11. Location and direction of field photo (yellow) shown in figure 10. The green overlay depicts percent cover values over 1.25, which correspond with individual tree canopies. The yellow line depicts the boundary of the study area.

Other Vegetation Indexes

Remote Sensing literature is replete with variations, modifications, and transformations of basic vegetation indexes, all purporting to improve some aspect vegetation analysis. Two of the more commonly cited variations include the Soil Adjusted Vegetation Index (SAVI) and the Perpendicular Vegetation Index (PVI). SAVI (Huete, 1988) was developed to minimize the problem of variability in soil reflectance, and is specified as:

$$[(NIR - RED / (NIR + RED + L)] * (1 + L) \qquad \textbf{(3)}$$

where *NIR* is the near-infrared radiance, *RED* is the visible red radiance value, and *L* is an adjustment factor, usually set between 0 and 1. SAVI produced comparable correlation results to NDVI for cases where *L* was set at 0.5 and 0.9, so no significant advantage was observed over the more common NDVI index.

The PVI (Richardson and Wiegand, 1977) builds on the idea of a ‘soil line’ which can be defined as a least squares line calculated from bare earth pixels present in an image. Figure 12 shows a soil line calculated from a set of vegetation-free sample pixels (red) extracted from the image used in this study, along with a set of samples taken from revegetated areas (green). The PVI equation, given as:

$$(NIR - a * RED + b) / SQRT(a^2 + 1) \qquad \textbf{(4)}$$

PVI essentially calculates a perpendicular distance from the soil line, where *a* is the slope, and *b* is the intercept of the line. Applied to the percent cover problem, PVI produced average errors that were larger than NDVI when using simple regression (10.61% vs. 6.78%).

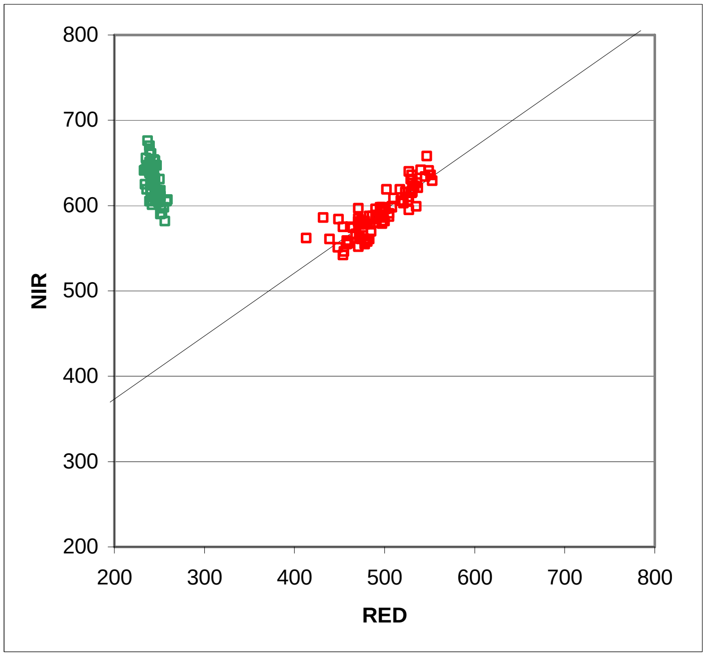

Figure 12. The idea of a soil line, calculated from barren image pixels, shown in red. Vegetation pixels are shown in green.

Neither of the alternative soil indexes—SAVI and PVI—improved on the results obtained using NDVI to correlate percent cover estimates obtained in the field. At first this appears to be a curious result, as many researchers have identified problems arising from variations in soil properties, and proposed numerous solutions for mitigating the problem, *e.g.* Rondeaux, Steven and Baret (1996), Clevers (1989), Baret and Guyot (1991) and Todd and Hoffer (1998). The relative success in this study may be attributed to more homogeneity in the substrate material, and perhaps the fact that dry conditions removed any impacts that moisture might have played in the analysis.

While NDVI values happen to correlate fairly well with the field samples, the relationship may not be optimal. Figure 13 shows the field samples plotted on a 2-dimensional graph, along with representative pixels taken from either barren or fully revegetated areas of the image. Best fit lines are calculated to find the 'soil line' from the barren pixels, and an intersecting 'vegetation line' from the revegetated pixels. Visually, the line that best describes the transition from barren to revegetated does not run parallel to the soil line, or linearly between the means of the two populations. Rather the transition from fully vegetated to completely barren traces a shape similar to a parabola, or ellipse. As the percentage vegetation falls there is an initial steep falloff in near infrared response, while red response increases only slightly. At the point where percent vegetation reaches about 50%, radiance begins to increase rapidly in the red, but also the near infrared, as the reflectance properties of the underlying substrate begin to dominate the contribution to pixel radiance. Given this pattern, an improved metric, designed to model this distribution explicitly, might involve measuring the angle formed by a line between a pixel in question and the focal point of a properly fitted parabola or ellipse. This is contrast to NDVI, where values are directly related to the angle formed by a line from the origin of the graph (0,0) and a particular point on the graph. This result also seems to run counter to assumptions, such as linear mixing, that underlie several approaches to estimating sub-pixel landcover percentages.

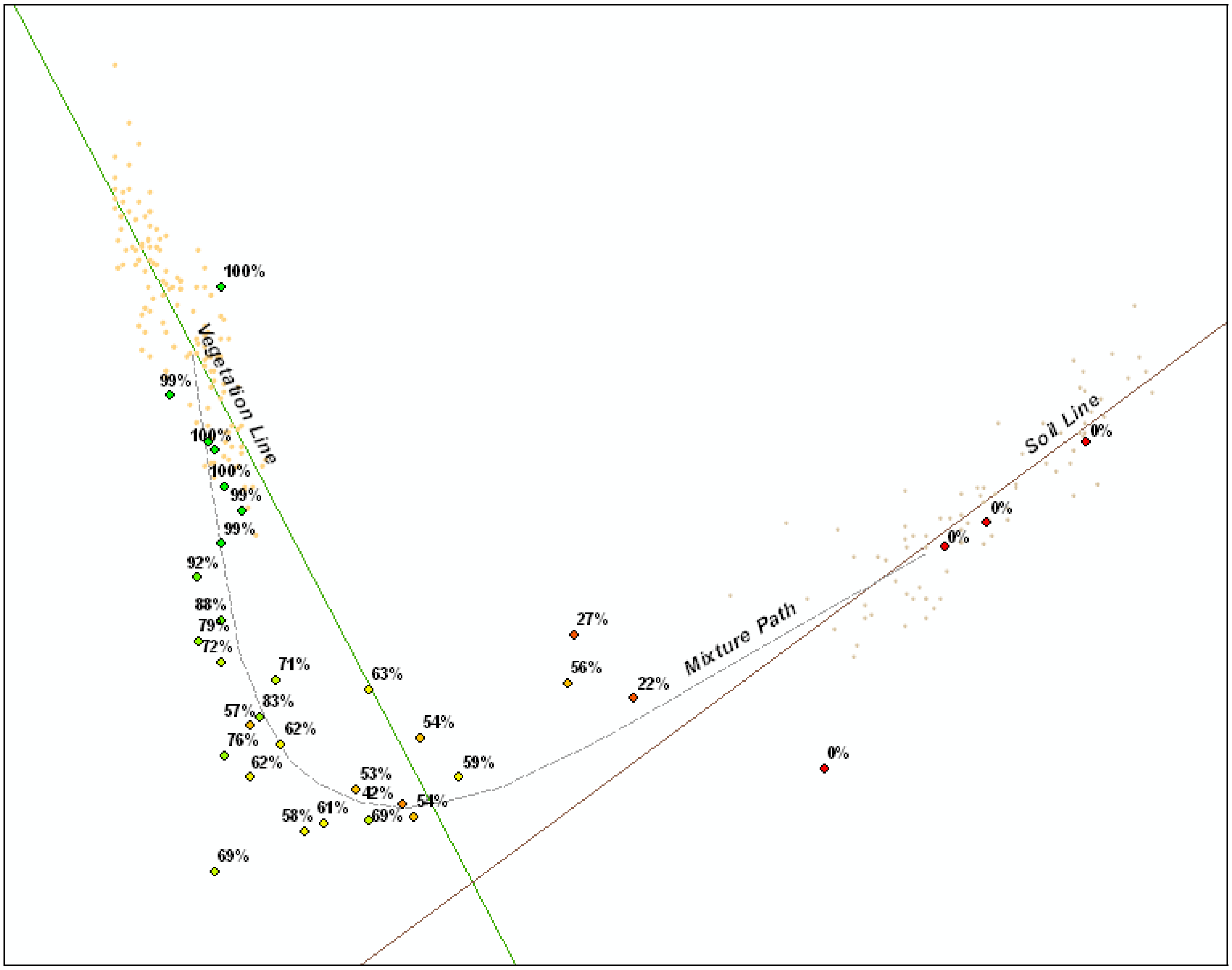

Figure 13. Plot of field samples, with labels indicating percent vegetation. Also shown are samples of barren pixels and vegetation pixels, along with associated regression lines. The figure illustrates that the transition from fully vegetated to unvegetated describes a parabolic, or elliptical, shape in the feature space. The Y axis represents near infrared radiance; the X axis is visible red.

Conclusions

This research demonstrated that NDVI can be used to estimate proportional vegetation cover on a reclaimed mine with reasonable accuracy. Two conditions may have contributed to the success of this analysis. First, the substrate material was fairly uniform throughout the study area. Second, soil moisture was not a factor because the substrate material was rock, and conditions were dry. These environmental conditions probably ameliorated difficulties faced by many researchers due to variations—both spatially and temporally—in soil moisture.

On the other hand, techniques developed to estimate percent cover in the field may have contributed error to the analysis. For example, the field sample that produced the single largest error (26.47%) was associated with an image pixel that occupied an area of relatively sparse vegetation. However, the corner photos appear to encroach into areas of much denser vegetation, which probably resulted in a percent cover average for the entire area that was significantly higher than the image pixel itself. This case represents a failure in the site selection process to only consider areas of relatively homogeneous vegetation.

Control points collected in the field, which were used to produce a final georeferenced image, all showed similar error vectors of a few meters. This indicated that the image could be georeferenced fairly precisely. If conducting the study again, the vertical photo pattern probably could be reduced in size, perhaps approximating a 3x3 pixel block. This would cause the photo set to cover a larger percentage of the sample area, and could potentially reduce problems with heterogeneous vegetation patterns.

Future investigations of this technique should focus on comparing the technique with results obtained from established field methods, demonstrating repeatability, and investigating the parabolic vegetation curve—and related assumptions about mixed-pixel composition—depicted in figure 13.

References

Avery, Thomas and Gradon Berlin. 1992. Fundamentals of Remote Sensing and Airphoto Interpretation, Fifth Edition. Macmillan Publishing Company, New York.

Baret, F and G. Guyot 1991. Potentials and limits of vegetation indices for LAI and APAR assessment (absorbed photosynthetically active radiation). *Remote Sensing of Environment*, Vol. 35, pp. 161-173.

Clevers, J. 1989. The application of a weighted infrared-red vegetation index for estimating leaf area index by correcting for soil moisture *Remote Sensing of Environment*, Vol. 29, pp. 25-37.

Elvidge, Christopher D, Zhikang Chen and David P Groeneveld. 1993. Detection of trace quantities of green vegetation in 1990 AVIRIS data. *Remote Sensing of Environment,* Vol. 44, pp. 271-279.

Farmer, Robert E., John C. Rennie, David H. Scanlon III, and Thomas G. Zarger. 1981. Technical Guides on the use of Reference Areas and Technical Standards for Evaluating Surface Mine Vegetation in OSM Regions I and II. Office of Surface Mining Region II, Knoxville, Tennessee.

Huete, A. R. 1988. A soil-adjusted vegetation index (SAVI), *Remote Sensing of Environment*, Vol. 25, pp. 53-70.

Purevdorj, I., R. Tateishi, T. Ishiyama, and Y. Honda. 1998. Relationships between percent vegetation cover and vegetation indices. *International Journal of Remote Sensing*, Vol. 19, pp. 3519-3535.

Richardson, A. J. and C. L. Wiegand. 1977. Distinguishing vegetation from soil background information, *Photogrammetric Engineering and Remote Sensing*, Vol. 43, pp. 1541-1552.

Rondeaux, Geneviève, Michael Steven and Frédéric Baret. 1996. Optimization of soil-adjusted vegetation indices**.** *Remote Sensing of Environment*, Vol. 55, pp. 95-107.

Todd, Stella W and Roger Hoffer. 1998. Responses of spectral indices to variations in vegetation cover and soil background. *Photogrammetric Engineering & Remote Sensing*, Vol. 64, pp. 915-921.