Task Order No. 2
Yukon-Charley Rivers National Preserve
Land Cover Mapping Report

Land Cover Mapping Services
Contract #1443CX991097005

Presented to:

National Park Service
Alaska Regional Office
2525 Gambell Street, Room 107
Anchorage, AK 97508-2892

Presented By:

Ken Stumpf
Director, Resource Management Applications
Geographic Resource Solutions
1125 16th Street, Ste 213
Arcata, CA 95521
(707)-822-8005
(707)-822-2864(fax)
grs@northcoast.com

TABLE OF CONTENTS

INTRODUCTION

The National Park Service, Alaska System Support Office (**AKSO**) is responsible for Land Cover Mapping projects for the Alaska Region. Past efforts have included the development of Geographic Information System (GIS) data bases, as well as image processing of satellite data to develop land cover themes. In 1997, AKSO awarded a Land Cover Mapping Services contract to Geographic Resource Solutions (**GRS**). Services provided by **GRS** under this contract include image processing and classification, as well as the development of detailed land cover database descriptions associated with mapped land cover units.

In September of 1999, the AKSO requested that **GRS** demonstrate the capability to successfully perform an illumination correction and develop detailed land cover database descriptions. The illumination correction would facilitate future image classification efforts and the detailed land cover descriptions would enable more detailed data modeling, mapping, and analysis by AKSO personnel. The AKSO, in conjunction with the Bureau of Land Management, had jointly funded a land cover mapping project during 1997-98 that resulted in the development of a land cover data set in an ARCView grid format that represented the Yukon-Charley Rivers National Preserve (YUCH) and its immediate surrounding area. This grid/map was developed from 1991 Landsat TM data and contained data representing 30 different mapping classes (Table I).

The methodology and results of this project are described in a report titled "Yukon-Charley/Black River/40 Mile Earth Cover Classification User's Guide" written (June 5, 1998) by Ducks Unlimited, Inc. Field descriptions of the different land cover types were based on over 240 training site developed from multiple field data sources that included helicopter reconnaissance, as well as existing AKSO field data descriptions. A portion of the field data was withheld in an effort to assess mapping accuracy. Mapping accuracies estimated during this effort ranged from 72% correct for all 30 classes to 80% correct for 21 grouped classes. Unfortunately, the 21 grouped classes and the 30 more specific classes are still too generalized and do not contain sufficient detail to enable further discrimination or analysis. In addition, data modeling and pixel editing were used in an attempt to resolve confusion of several of the confused land cover types. The insufficient detail, lower than desired accuracy, and reliance on modeling and pixel editing are results that limit the usefulness of such information.

GRS has accomplished similar land cover mapping projects that have resulted in detailed accurate land cover descriptions. These efforts have been based on image processing and classification methodologies that rely on normalized imagery and detailed field data descriptions. **GRS** integrates two significantly different processes into their image classification methodology. One process entails the normalization or illumination correction of the satellite imagery. The other process involves the summarization of complex field data descriptions that preserve cover type descriptive information and yield detailed cover descriptions. Generalized categorical values are developed from the specific estimates and may be easily recalculated if categorical rules or limits are modified.
The AKSO has requested that GRS reprocess the YUCH imagery and field data in an

attempt to develop more detailed land cover descriptions. This effort involves the application of the illumination correction process to the subject area and the generation of detailed land cover information capable of subsequent evaluation, reclassification, modeling, and analysis.

The purpose of this report is to describe this project. This report addresses the illumination correction methodology **GRS** used to normalize the satellite imagery and the apparent results of this effort. This report also addresses the development of detailed land cover descriptions and the ability to summarize the detailed land cover descriptions in different ways, depending on the needs of the data user.

TASK 1: ILLUMINATION CORRECTION

Terrain topography has a direct impact on the quality of data detected by a satellite sensor. This impact is exacerbated in mountainous areas, where the main effect of topography is to introduce greater variability in surface reflectance data (Smith, 1980). In turn, this creates serious difficulties in the analysis of satellite imagery and extraction of meaningful information regarding surface cover (Chiou, 1992). In terms of image classification processes, this translates into greater numbers of cover classes required to describe areas, larger numbers of training areas needed to define the latter, and greater need for post-classification editing. Cost increases and greater difficulty in error detection/tracking during a vegetation mapping project associated with the latter, should make successful illumination correction techniques a highly desirable element of land cover classification methodology.

The successful application of an illumination correction process may potentially reduce the confusion between different land cover classes and decrease the number of training sites required to describe the various land cover types present in the subject area. This decrease in required field data may greatly reduce the cost of image processing projects in remote areas with short data collection seasons. The primary goal of this project task is to apply illumination correction procedures to the Yukon-Charley/Black River imagery; the corrected imagery will be evaluated in an attempt to determine if the correction has a beneficial effect on the corrected imagery.

Basic Methodology

Illumination correction methods devised to address this problem have traditionally taken an empirical or an analytical approach (Smith, 1980). Empirical approaches do not take into account physical elements present in the sensor-surface-light source interaction at the moment of image acquisition. Band ratios are an example of this approach, where a hypothetical “noise” factor (assumed equal and constant for each band) is canceled out by the ratio of two bands and a new band introduced in the analysis. Conversely, analytical approaches are based on the physical and geometric relationships between these elements, perform corrections in each original band without introducing new bands, and have consistently proven to be superior to band ratios (Smith *et al.*, 1980; Colby, 1991; and Chiou *et al.*, 1992).

The image illumination correction implemented in this pilot is based on the backward radiance transformation correction (BRTC) techniques described by Smith (1980), Colby (1991), and Chiou (1992) and further adapted and complemented with custom GRS processing tools for application in Intergraph's modular GIS environment (MGE). Necessary data files provided by AKSO were translated into MGE/GRS compatible formats for processing. Final data files were translated back into Arc/Info GRID, INFO, and Arc/Info-compatible image formats.

The BRTC process takes into account geometric relationships between the sun and sensor's position in the sky, and particular topographic characteristics of the area covered by the sensor, at the time of image acquisition. Digital elevation model (DEM) data is used to calculate slope and aspect; which, in turn, is used to determine the effective incidence angle of solar radiation on the terrain surface. This information is then related to sensor-detected reflectance values by means of Minnaert 's model for satellite radiance, defined as:

$$L(\lambda,e) = Ln(\lambda)\cos^{k(\lambda)} I \cos^{k(\lambda)-1} e$$

The Minnaert constant (k) is descriptive of the type of surface scattering, relates to surface roughness (Smith, 1980), and can be estimated by simple regression applied on the linearized version of the model. Thus, individual "k" values are calculated for each band to be corrected. The correction is then applied by reversing the model equation to calculate the normalized reflectance value (Ln) for each pixel in the band. The assumption is made that terrain surfaces behave as a non-Lambertian or non-perfect reflectors, and that, given equal surface cover characteristics and atmospheric effects, variations in reflectance values detected by the sensor are mainly due to topographic relief.

The correction is applied on a pixel-by-pixel and band-by-band basis. Pixel-defined image reflectance (L), and slope (e) and incidence (I) angle values, make corrections vary among areas within the image; whereas band-defined "k" values, in turn, make correction factors vary from band to band.

The above described methodology was applied to bands 1, 2, 3, 4, 5 and 7 of the AKSO-supplied Landsat TM scene (S6614), which encompasses the Yukon-Charley Rivers National Preserve and surrounding areas. Necessary USGS 15-minute digital elevation models (DEM), which were not the same DEM files used to register the original imagery to ground coordinates, were retrieved from the AGDC web site, projected, re-sampled, and merged to best match corresponding pixels in the area of interest (AOI) in the imagery. The AOI was reduced to cover the extent of pre-defined training set locations (DUFF sites, IMA sites, and ancillary sites). Slope and aspect data sets were developed from the AOI-DEM using Map Algebra and a custom GRS slope-derivation algorithms. Incidence angle data was derived from slope and aspect data sets (incorporating sun elevation and azimuth angles) and subsequently coupled with slope data to calculate a normalization constant (k) for each image band. Relative frequency distributions of calculated incidence and slope angle values within the AOI are shown in Figure 1 and Figure 2, while Table 1 records descriptive statistics for these data sets.

Application of the illumination correction process was performed on all six input (raw) bands of the imagery using the **GRS** process ***batch_illumcor***, which requires the input band name, the incidence angle data set name, and the slope data set name. Input bands were renamed to reflect the area, band number, and band status. Raw band 1 was renamed to yuch1raw.tif. Corrected bands were output with a similar name except the letters 'btm' replaced 'raw' to indicate the band had been corrected. Corrected band 1 was named yuch1btm.tif. Descriptive statistics and calculated Minnaert constant values for all bands are listed in Table 2.

Data Set	Minimum	Maximum	Mean	Standard Deviation	Variance
Slope	0	81	10.67	8.1913	67.0968
Incidence	0	70	35.27	9.2483	85.5305

Table 1: Descriptive Statistics for Slope and Incidence Angle

Initial efforts were completed and the corrected imagery was reviewed relative to the raw imagery using cross tabulation matrices. It was apparent that the algorithms were not normalizing the data in the manner anticipated. Upon further investigation we determined that there was a shift between the imagery and the DEM data. This shift resulted in a terrain shadow that created very dark spots and very bright spots - darker areas that should have been lightened were darkened and lighter areas that should have been darkened were lighted. An evaluation of the data resulted in an estimated shift of 4 pixels in a north-south direction between the imagery and the DEMs, with the belief that the DEMs were in the correct location and the imagery was shifted. As other data were referenced to the imagery, we decided to shift the DEMs to correspond (as closely as possible) to the imagery, rather than attempt to shift the imagery and related training data (we believe there may still be a slight offset in these data sets, but could not shift the data sets by less than a whole pixel value). After performing this shift, the ***batch_illumcor*** process was reapplied, resulting in corrected imagery that showed little in the way of the prior 'black holes' and 'hot spots.'

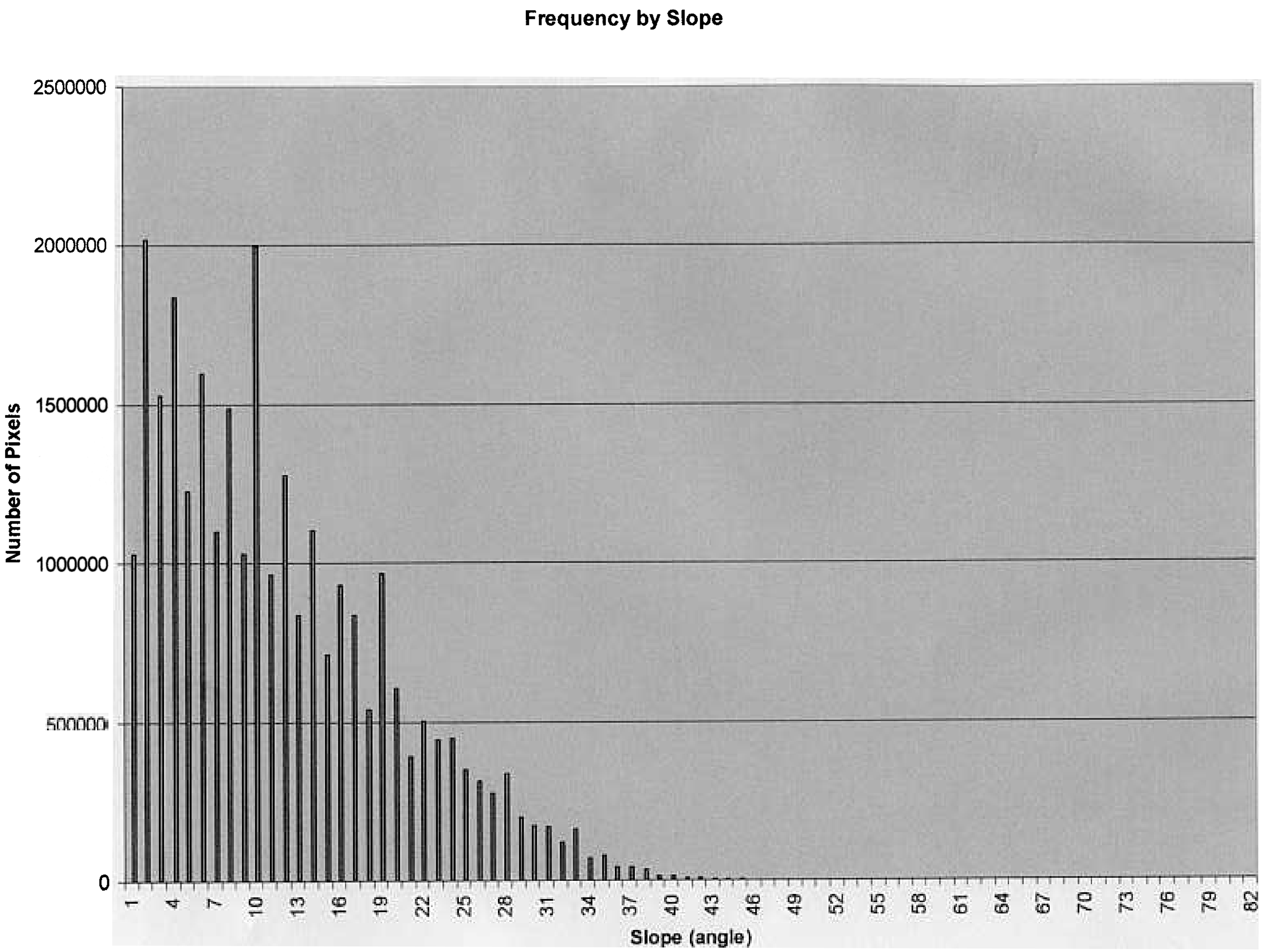

Figure 1: Distribution of Pixels by Slope Angle

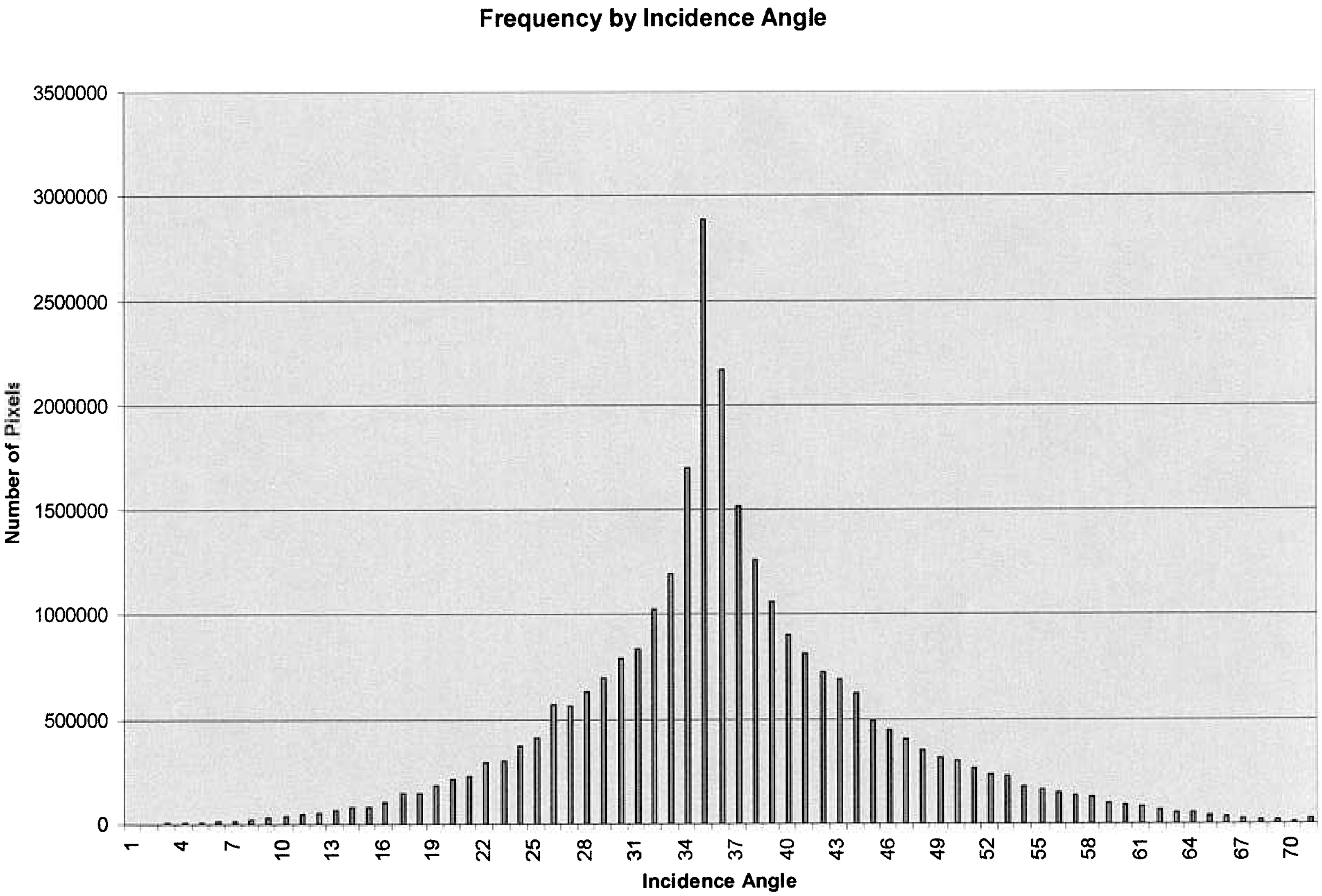

Figure 2: Frequency by Incidence Angle

Training Areas and Classification Methodology

One means of comparing the results of the illumination correction would be to evaluate the raw and corrected imagery within a uniform set of sample areas. A basic set of areas that could be used for this purpose were the training areas that would be defined during an image classification project. These areas may be defined using statistical parameters and training data sets could be developed for the raw and corrected imagery based on the same set of areas. Statistics and other data for these training sites are generated and may be compared.

Training areas were defined using coordinate point locations indicated for the DU classification project. The files were supplied by AKSO and included training and "accuracy assessment" sites. Coordinate points corresponding to 228 training sites included DUFF, IMA, and ancillary site locations and were used to generate training polygons for this project. Rather than using manually delineated training polygons developed by DU, we used region-growing techniques and allowed the software to define training polygons based on user-specified statistical parameters. Namely, training areas were seeded at the DU point coordinate locations and region growth limited to within two standard deviations from the 3 x 3 cell seed size. Several areas that refused to grow to an acceptable extent were grown with 5 x 5 seeds. One site was removed from the data set as it was outside the limits of the DEM data. Two classes were added to the original DU set, for SNOW and WATER (river) classes, completing the 229 areas in the project's training set. Training areas were grown using raw bands 1,2,3,4,5,4/3 ratio band , and 7. Raw statistics were collected. The extent of each raw data-defined training area was maintained in the corrected bands training set; however, class statistics for all sites were re-calculated using corrected bands 1,2,3,4,5, and 7. Corrected band statistics were collected. This scheme allowed us to compare pixel-to-pixel digital values (DNs) for the same groups of pixels in each class, between the raw and corrected data sets.

Another means of comparison is provided by evaluating the classification results from processing the raw and corrected bands of imagery. There will be little difference in the resulting class maps if the illumination correction process has not altered the imagery. The training data sets for the raw and corrected bands were now applied during classification efforts. Classifications were performed applying a Maximum Likelihood classifying algorithm on the above described training sets and bands, ensuring identical classification parameters (95% confidence threshold for the null class, all classes, and all bands). Differences could be identified through a review of the cross tabulation of the different classification maps (the results of the 'corrected' classification would also be used during the second phase of this project as a basis for summarizing pixel data and developing detailed data descriptions) .

Results

Expected results of a successful implementation of the BRTC are the normalization of the raw band data as manifested by an increase in overall variance in the corrected bands; decreased variance between areas of the SAME cover type; retention or increase of variance between areas of DIFFERENT cover types to allow differentiation; and an obvious reduction of impression of relief in mountainous areas of the image, with little or no change in areas of negligible topography. The following is an analysis of the overall results of the application of the BRTC to the YUCH data set.

Reduction of Impression of Relief: Visual Evaluation of RGB Composites

The effectiveness of the BRTC method in reducing topographic effect is readily apparent upon visual examination of the corrected imagery. Images 1 through 4 (See attached BITMAP files) show details of RGB composites for selected areas made from raw and corrected bands. A dramatic reduction /elimination of relief impression can be observed on the corrected imagery.

Image 1 illustrates the marked reduction in the impression of relief in an RGB composite image using bands 4,5 and 3. Areas one through three are examples of the elimination of relief effect in the imagery.

Image 2, illustrates the stronger effect of the BRTC process in areas of marked topographic relief, brightening originally dark areas and darkening originally bright areas (Area 1). Conversely, areas of little topographic relief (Area 2) remain mostly unaffected by the correction.

Image 3 also illustrates the topographically sensitive nature of the BRTC, correcting mostly where topography deems it necessary (Area 1) and leaving areas of little relief, such as the flood plains encircled in Area 2, unchanged. The bright spots observed in the upper part of the corrected image, are due to the slight mis-registration between the DEM model and the imagery. Registration of these data sets is CRUCIAL to implementing true BRTC. The process of ensuring registration between these data sets should begin at the time of the acquisition of the imagery from the supplier. Images should be registered to the entire set of DEM(s) included in the image area, if possible, and these DEM(s) should accompany the imagery when delivered. This would be the required procedure during an actual classification project. In Image 4, a 5-4-3 RGB composite illustrates examples of reduction in relief impression from raw to corrected band sets.

Image 5, also a 5-4-3 RGB composite, shows a detail of how the correction affects specific areas in the imagery. Reflectance patterns previously hidden by topographic shadowing like the pink-colored pixels in Area 1, are brought out and closer to the same type pixels on the illuminated mountainside. It is safe then to assume that these pixels have become available to classifiers in the corrected imagery. Area 2, also illustrates this effect on a smaller, vegetated area.

Increase in Overall Variance in Corrected Bands

A noted effect of the application of BRTC is a significant increase in overall band variance (Colby, 1992). This increased variance can result in greater separability of the spectral data. Changes in overall band variance from raw to corrected data sets are recorded in Table 2. The application of the BRTC method to the YUCH data set yielded significant increases in overall band variance for corrected bands 2,3,4,5, and 7, while slightly decreasing for band 1. This shows that application of the BRTC had a significant effect on the Yukon Charley Rivers data set.

Band	k	Raw MIN	Raw MAX	Raw Mean	Corr MIN	Corr MAX	Corr Mean	Raw VAR	Corr VAR
1	0.231669	29	255	56.11	16	255	57.80	41.80	41.18
2	0.401624	11	245	21.21	8	255	22.81	13.73	14.31
3	0.489009	7	255	21.45	8	255	23.53	39.16	43.54
4	1.150686	1	235	47.51	1	255	61.18	160.53	218.04
5	1.119256	1	250	49.42	1	255	63.03	367.64	522.68
7	1.044643	1	137	16.36	1	255	20.50	77.34	113.21

Table 2: Summary DN Statistics for Raw and Corrected Bands

Decreased Variance Between Areas of EQUAL Cover Type

The effectiveness of BRTC can be best demonstrated by analyzing a set of areas of "known" equal cover type, distinct yet internally homogeneous topography, and differing spectral signatures. Given equal cover types, and other atmospheric effects being equal for all areas, an efficient illumination correction would make these areas more spectrally homogeneous. In other words, the variance between these sites should be reduced in the corrected imagery. This type of demonstration would require utmost control of the data to be analyzed; mainly, absolute certainty regarding the actual cover type, degree of homogeneity, and extent of pre-selected test sites. Lacking this kind of control during this pilot demonstration, it was decided to utilize best available data, in the form of previously defined DU training sites. "Closed Birch" sites were selected to track decreasing variance between areas of equal cover, assuming that the most accurate cover type determinations would have occurred on such distinct and highly homogeneous cover types. Results of the analysis of six "Closed Birch" sites for bands 2,3,4, and 5 are recorded in Table 3. Decreases in variance between Closed Birch sites are expressed by decreasing F ratios between sites for each band (Colby, 1992). The decrease in the F statistic for sites 1018,1047, 253, 293, 380, and 558 was significant for all tested bands at 306 degrees of freedom. Assuming that these areas indeed have equal cover types, we therefore imply that the BRTC brought tested Closed Birch sites spectrally closer to each other, as predicted.

Band	F-Ratio Between Closed Birch Sites		F-Ratio Between Closed Birch Sites	
DF=306	RAW	CORRECTED	RAW	CORRECTED
2	88.87	25.36	150.99	177.89
3	120.73	86.82	1084.48	1045.26
4	143.02	63.77	693.10	162.73
5	197.18	67.69	848.88	440.64

Table 3: Comparison of Variances Between Sites of Equal Cover and Between Cover Types.

Variance Between Areas of Different Cover

Maintaining class separability is a crucial point in the successful application of illumination correction techniques for land cover classification. When “Closed Birch” site 558 data was replaced with data from “Wet Graminoid” site number 372 for the ANOVA, variances between sites increased significantly in all the raw bands. Such increased variances were sufficiently maintained in the normalized bands, indicating that although sites of similar cover became more homogeneous, normalized class sites remained separable from other cover type sites. The marked decrease in variance between cover types for band 4 should be offset by sufficient differentiation in other pertinent bands. Interestingly, when the same analysis was performed only for sites which retained or improved in performance during normalized classification (sites 1047, 253, and 293), the F-ratio for band 4 was 280.15, a 72% increase from the value obtained using good and diminished performers.

Results of the analysis of variances between sites and between cover types should be viewed cautiously, since there was no control over the accuracy of cover descriptions for the selected test sites. However, they point in the right direction in terms of spectral homogenization between sites of similar cover and maintenance of class separability in the corrected imagery, albeit with some deviations. Such deviations may have been caused by inaccuracies regarding cover type determinations and/or actual extent of the said cover type in the test sites. An example of this could be “Closed Birch” site number 1005, which was excluded from the variance analysis after determining that it’s marked decrease in classification performance (from 82% in raw set to 36% in normalized set) was markedly higher than any other site of this class. After revisiting the site on the imagery, it became obvious that site 1005 was noticeably dissimilar to all other sites of the “Closed Birch” type. Moreover, cross-tabulation between the raw and normalized classifications showed that pixels attributed to site-class 1005 in the raw map migrated consistently to “Open Needleleaf” classes in the normalized classification. This, coupled with the fact that pixels from other less drastically performance-diminished “Closed Birch” classes (sites 1018, 380, and 558) migrated to those classes defined by sites with maintained and/or improved classification performance (sites 1047, 253, and 293), led us to believe that site number 1005 did not consist of the same cover type as the other "Closed Birch" sites.

Band Graphs

Linear plots of digital number (DN) frequencies enables a graphic analysis of changes in mean pixel values for individual bands. Data from Bands 2 and 5 provide examples of the changes that occurred during normalization.

Band 2: Linear plots of digital value frequencies for Band 2 in Figure 3 show an increase in mean digital value from raw to corrected data sets. The corrected pixel data distribution exhibits an overall shift to the right and shows a slightly wider spread over the range of possible digital values, resembling a linear stretch. The distribution for the corrected set shows a better approximation to a normal or "bell shaped" distribution, although this difference is difficult to distinguish visually.

Band 5: Linear plots of DN frequencies for Band 5 in Figure 4, show a marked increase in mean DN in the corrected data set. The shift is actually a 27.54% increase in mean DN value The corrected DN value distribution also exhibits an overall shift to the right, with a marked wider spread of DN values over the data range. The distribution for the corrected set shows a noticeably better approximation to a normal or "bell-shaped" distribution.

Cross-tabulation

Cross-tabulation between raw and corrected bands shows the re-distribution of raw pixel values in the corrected data sets. The following evaluations regarding changes in re-distribution of raw pixel DN and origin of corrected pixel DN(s), are based on cross-tabulations between raw bands and their BRTC corrected counterparts. (More on this)

<u>Digital Value Graphs</u>

Disposition of Raw DNs 20 and 50

Figure 5 illustrates the re-distribution of raw digital numbers (DN) 20 after the illumination correction in Band 2. The bulk of the original DN 20 values (approx. 2.75 million) were re-assigned to DN 22, with significant numbers (approx. 1.8 million) re-assigned to DN 21 and DN 23 (approx. 0.6 million) in the corrected band, and were thus "brightened." A smaller fraction (approx. 0.45 million pixels) of raw DN 20 pixels remained unchanged, whereas an even smaller portion (< 50,000) moved to values above DN 23. Small fractions of raw DN 20 (< 50,000) were re-assigned to DN values below 20, and were thus "darkened."

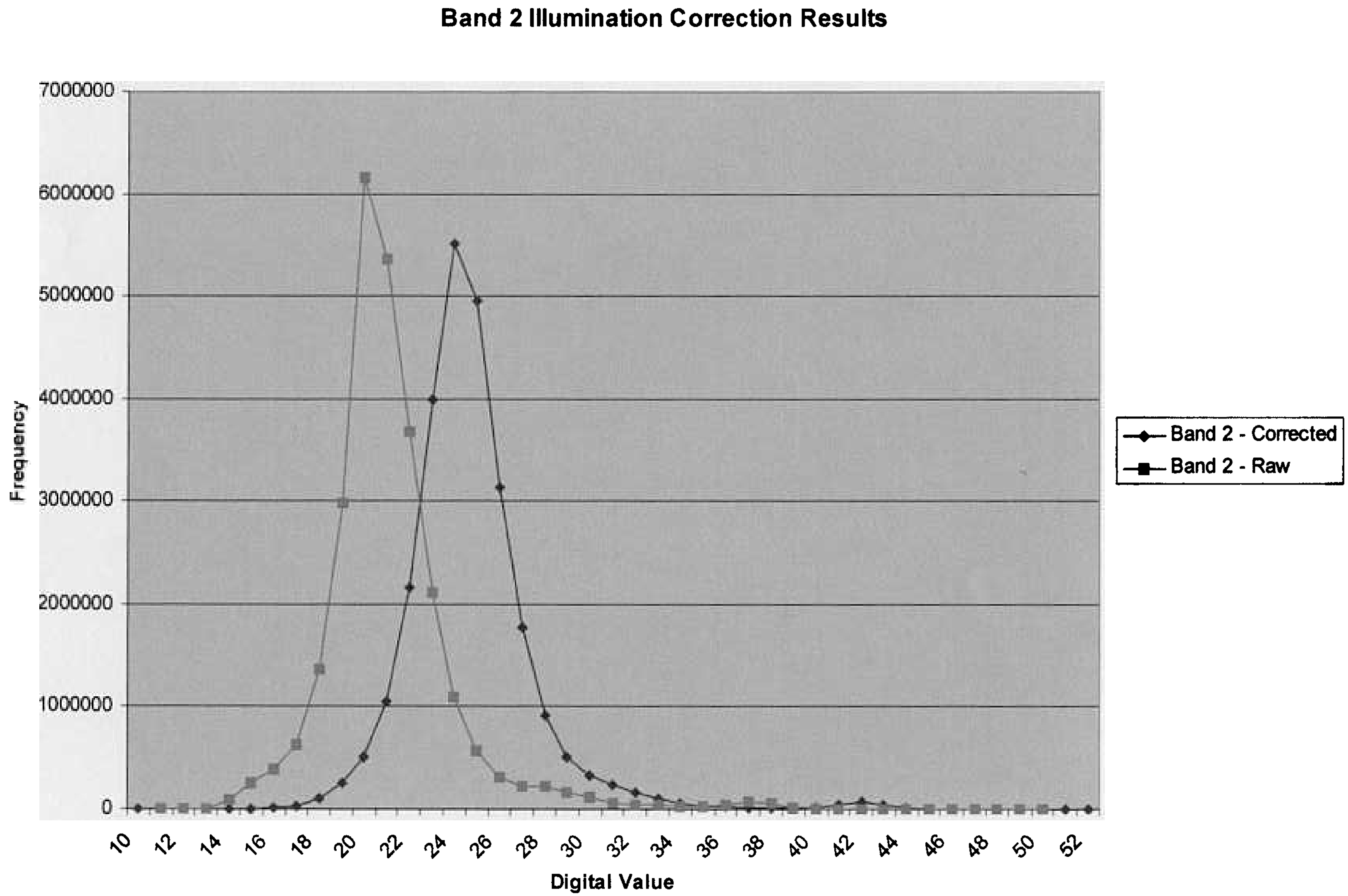

Figure 3: Relative DN Distribution for Raw and Corrected Band 2

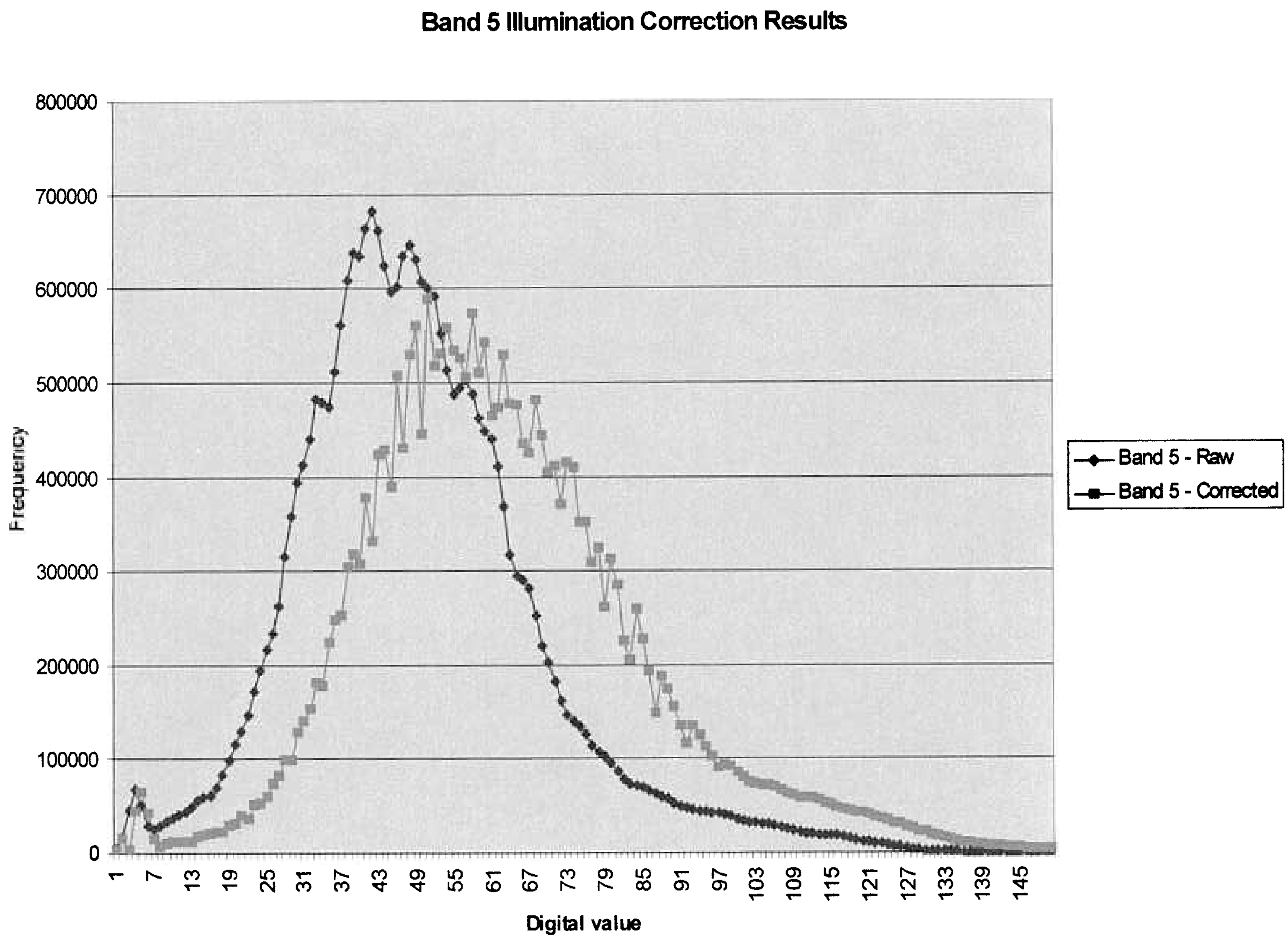

Figure 4: Relative DN Distributions for Raw and Corrected Band 5

A similar re-distribution of raw DN 50 can be observed for the same band in Figure 6. However, in this case, the original raw DN pixels were distributed more evenly over a wider range of corrected values, within an overall shift to the right.

Almost all raw DN 20 pixels were brightened in the corrected Band 5, as shown in Figure 7. The bulk of raw DN 20 pixels were re-assigned to corrected DN values ranging from 23 to 32; however, the overall number of DN 20 pixels in the raw band was relatively low (<< 10,000) and would make this effect less noticeable in the image.

A look at raw DN 50 in Band 5, also shows that roughly all raw DN 50 value pixels were assigned higher DN values in the corrected band, as shown in Figure 8. The bulk of raw DN 50 pixels were assigned corrected values ranging from 59 and 65, with some corrected values ranging above DN 68.

Origin of Corrected DNs 20 and 50

Figure 9 shows relative contributions from raw DN value classes to the DN 20 class in corrected Band 2. In other words, raw digital number values that were modified to yield DN 20 pixels in the corrected data set. The graph shows that the major contributions came from raw DN 19, 18, 20, 17, 21, and 16, in order of magnitude. Relative raw DN value class contributions to the corrected DN 50, shown in Figure 10, indicate a wider range of contributing value classes (DN 36 to DN 56) for the same band.

Following correction, patterns for Band 5 illustrate patterns already discerned from Figure 7 and Figure 8; contributions to corrected DN 20 and DN 50 came, unvariably, from lower raw DN values, as described in Figure 11 and Figure 12.

Training Data

Statistical comparison of means for all training sites, considered as a group of observations, resulted in acceptance of the equality between raw and corrected means, as shown in Table 4. Normalized means were adjusted for average (non-zero) difference in mean values (on an individual band basis) between the normalized and raw bands (means were based on non-zero values, as zero value pixels were not adjusted during normalization). This demonstrates that the data has been consistently stretched across the range of band DN values in a fairly linear fashion as the data have been normalized.

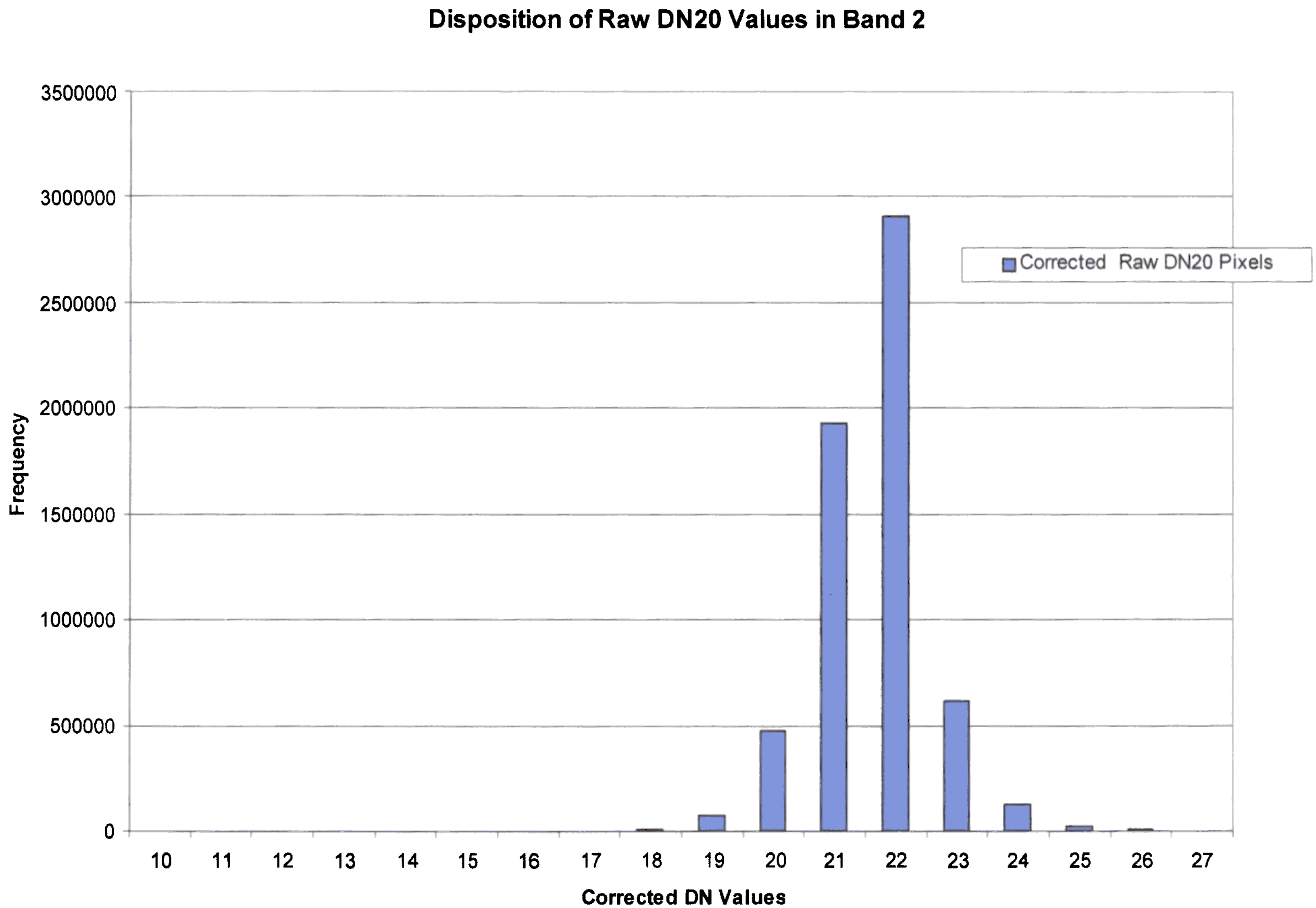

Figure 5: Disposition of Raw DN 20 Values in Band 2

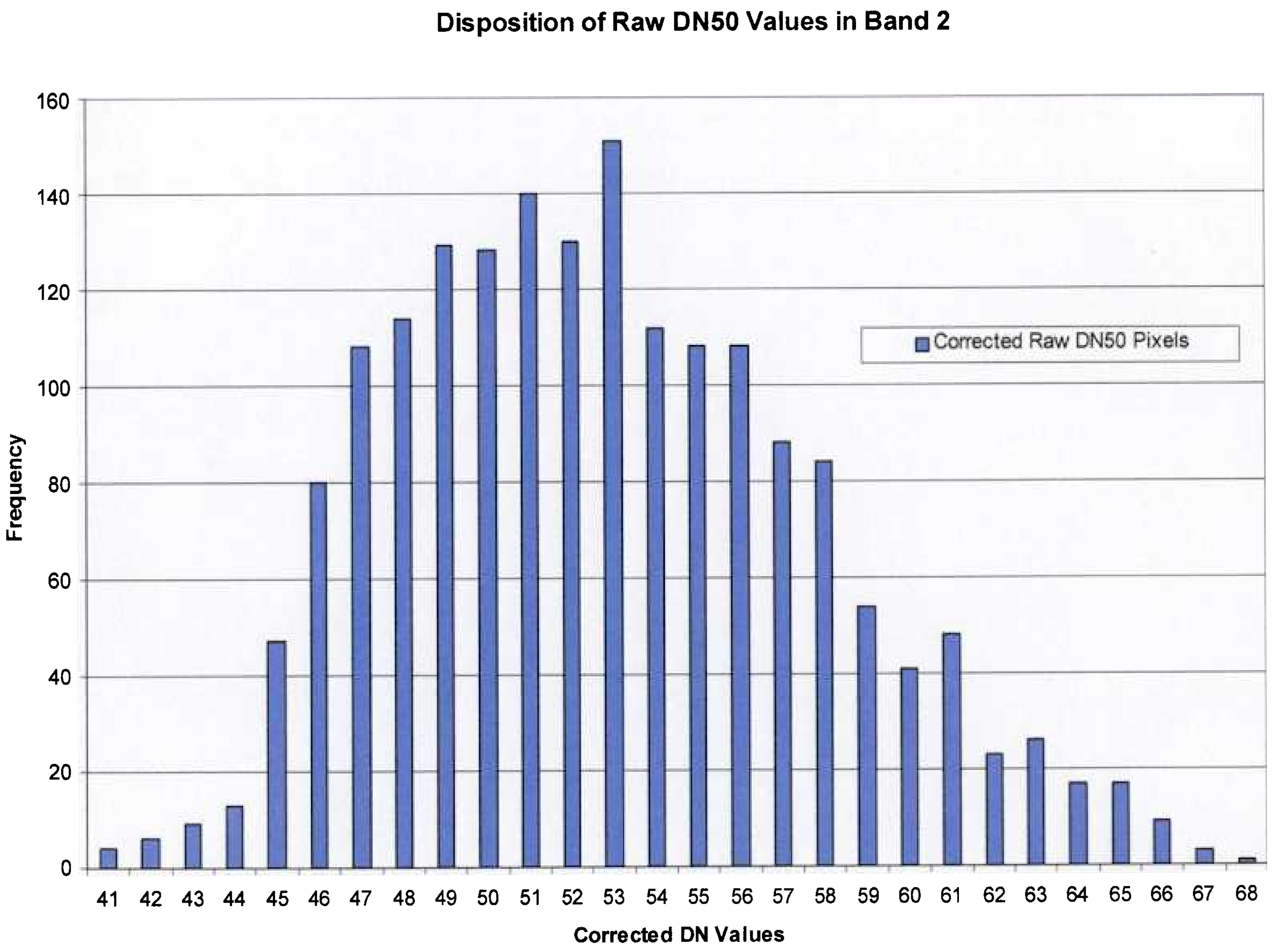

Figure 6: Disposition of Raw DN 50 Values in Band 2

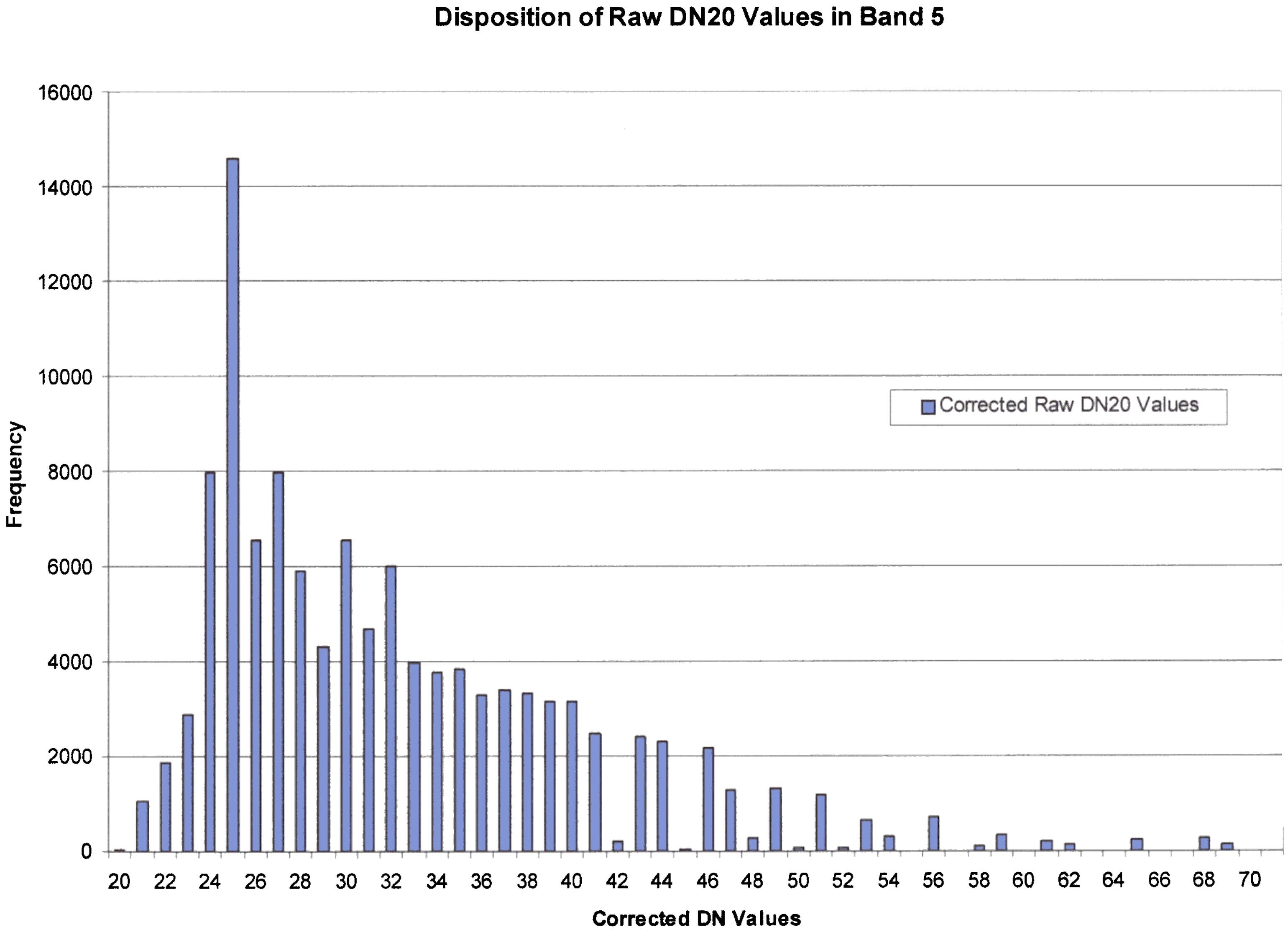

Figure 7: Disposition of Raw DN 20 Values in Band 5

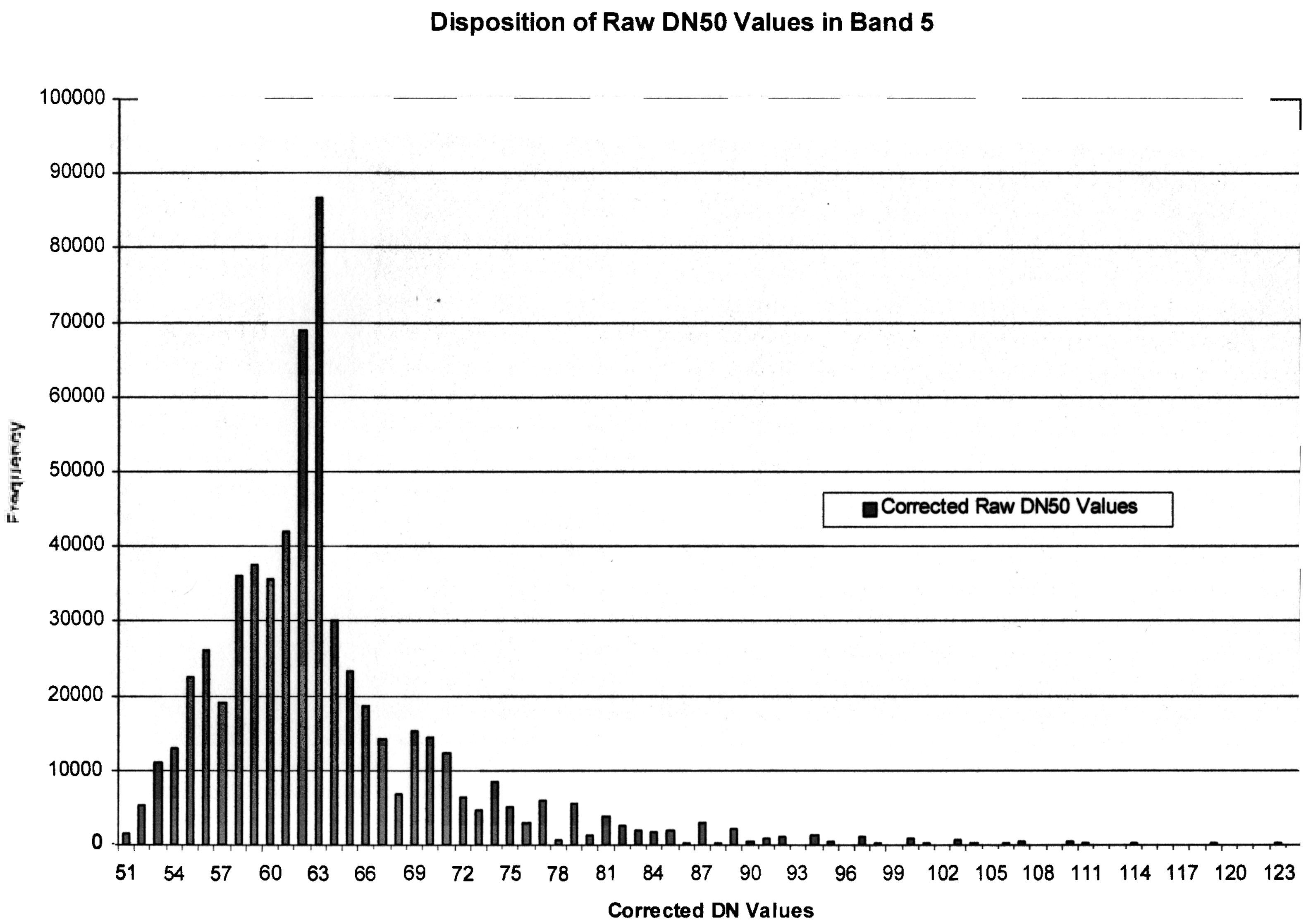

Figure 8: Disposition of Raw DN 50 Values in Band 5

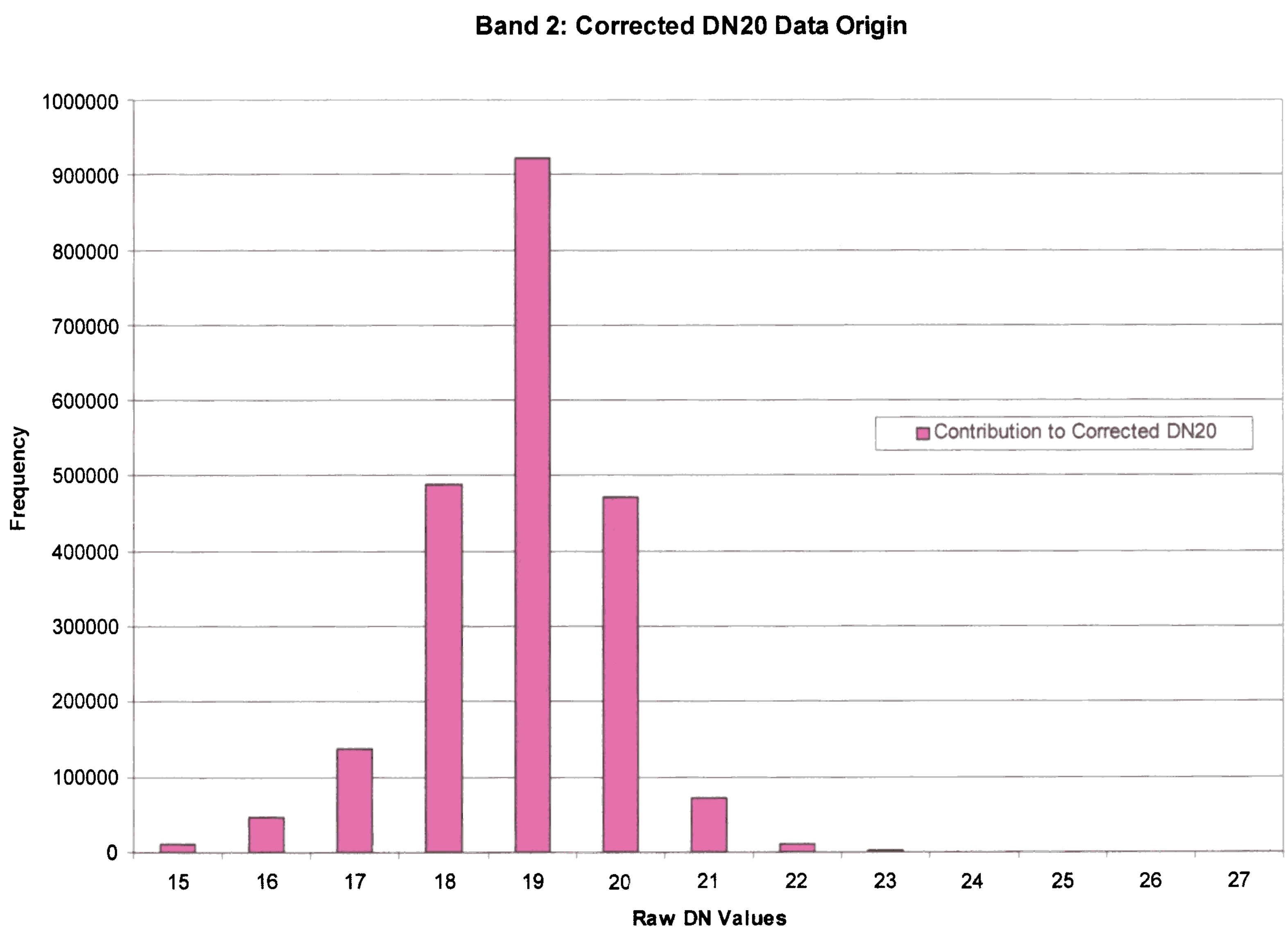

Figure 9: Origin of Corrected DN 20 Values in Band 2

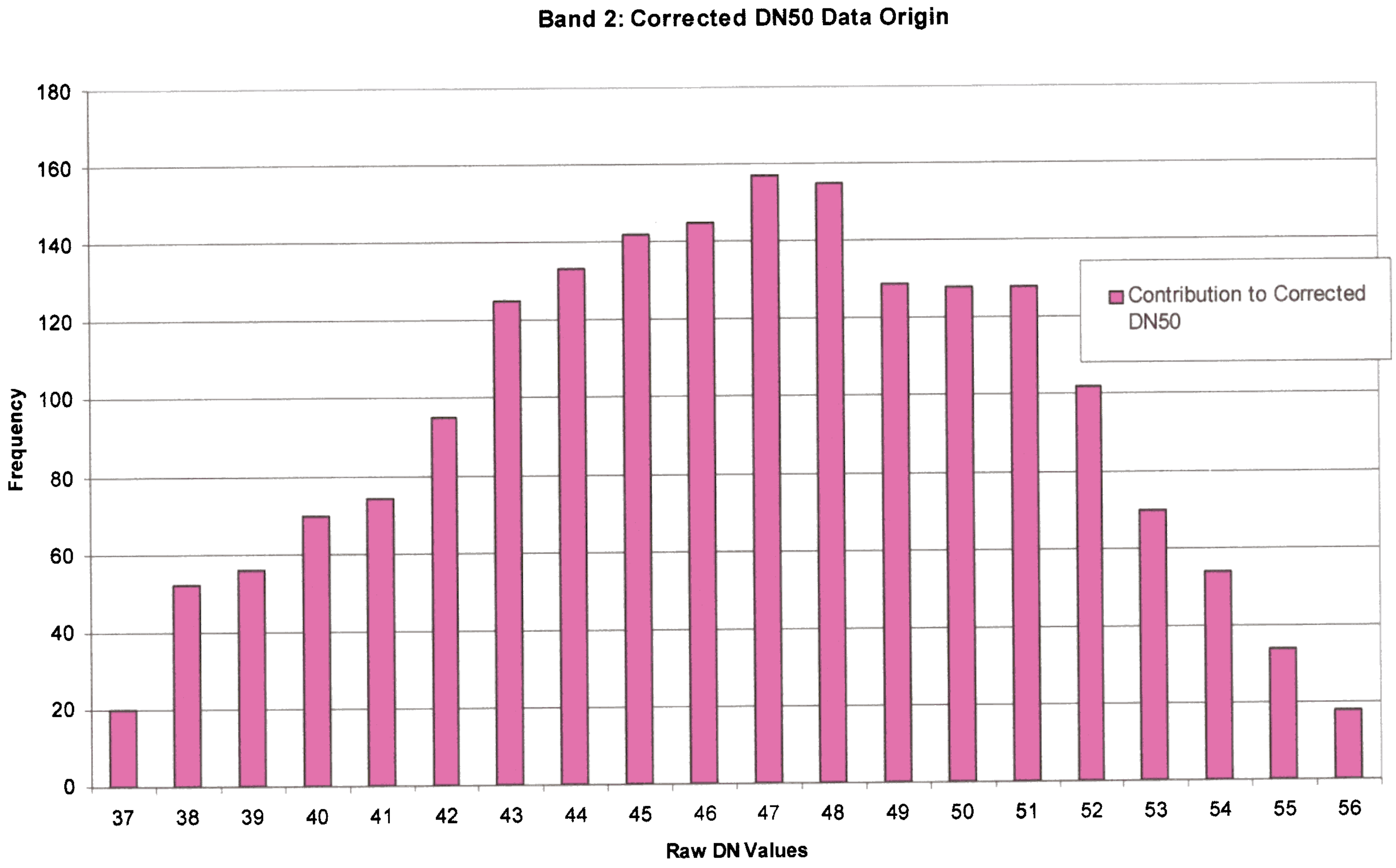

Figure 10: Origin of Corrected DN 50 Values in Band 2

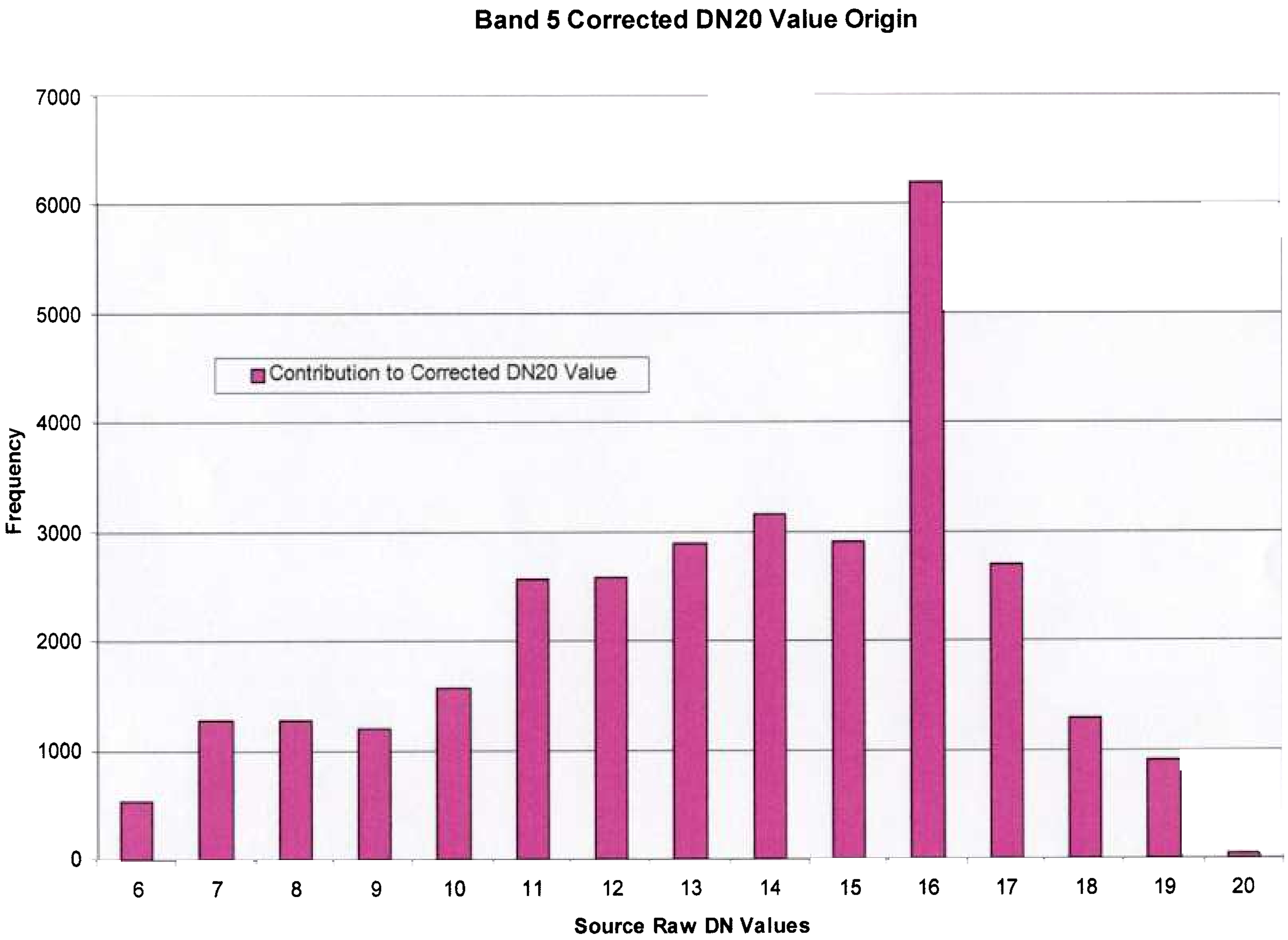

Figure 11: Origin of Corrected DN 20 Values in Band 5

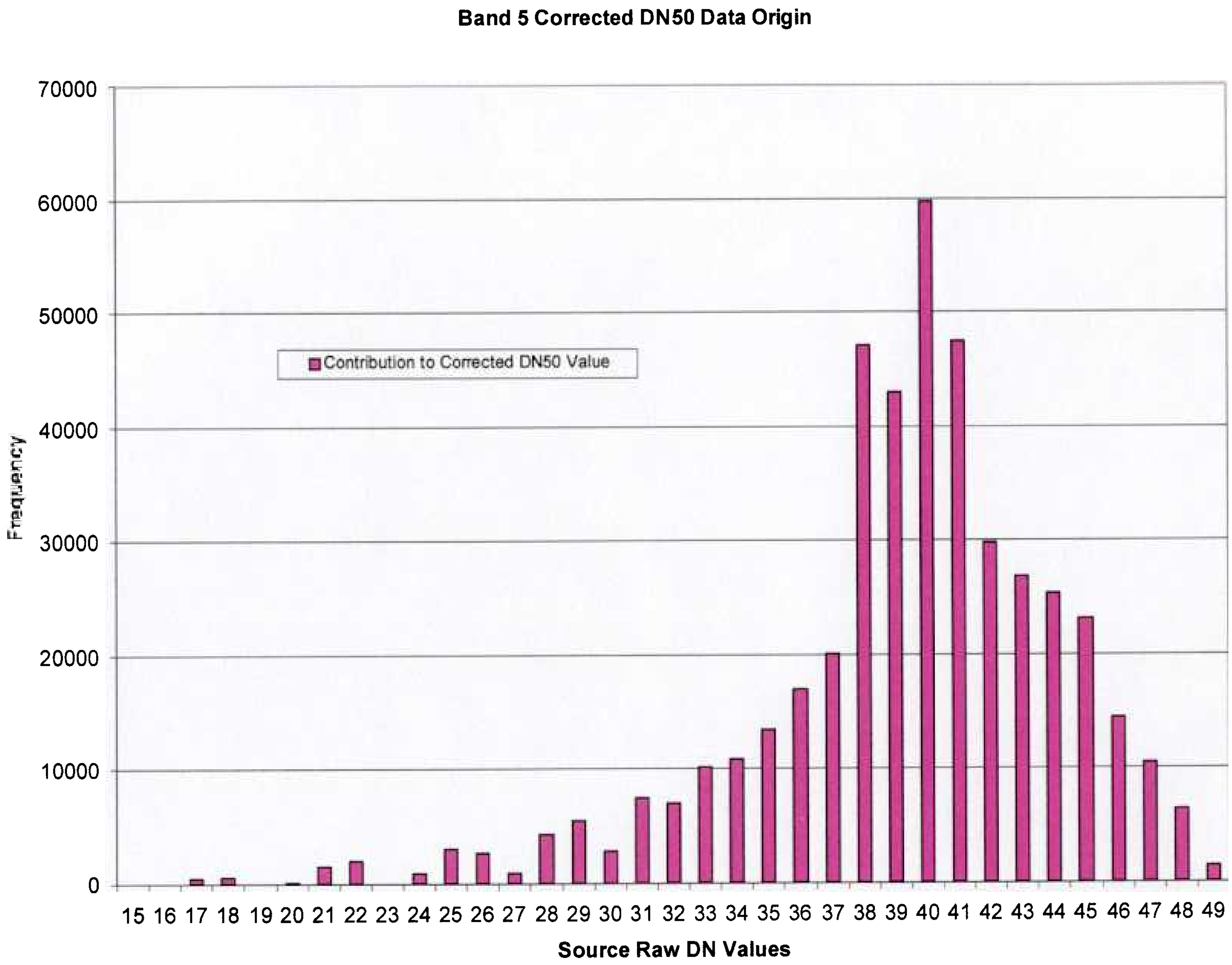

Figure 12: Origin of Corrected DN 50 Values in Band 5

Comparison of variances between raw and corrected training sites as a group of observations yielded acceptance of equal variances for bands one, two, and three and rejected equality for bands four, five, and seven, as shown in Table 5. Raw training data variances were used as "parametric" (expected) variance values, against which corrected or "observed" variances were tested.

Band	t(sample)	p(0.05)	5% Decision
1	0.0398	0.9681	Accept
2	0.1450	0.8848	Accept
3	0.0485	0.9614	Accept
4	0.4336	0.6648	Accept
5	-0.0949	0.9244	Accept
7	-0.1268	0.8991	Accept

Table 4: Test for Equal Means Between Raw and Corrected Training Sets.

We would expect a balance between darkened and brightened sites, assuming the training sites are randomly distributed with respect to aspect and slope throughout the project area. The expected balance between darkened and brightened sites is readily discernible in bands 1, 2, and 3; however, the large k values obtained for bands 4, 5, and 7, preclude the model from implementing any reductions in brightness values in these bands. However, it can still bring out shaded areas throughout these bands, therein allowing the normalization of such areas. The end result is that there is more than sufficient data to generate separable statistics.

Cross-tabulation between raw and corrected class maps

Training areas were "grown" based on statistical properties of the raw data sets around training point locations (seed pixels). Pixels were incorporated into the growing training areas if they were within two standard deviations from the mean value of the seed (a 3x3 pixel window around the seed pixel). Therefore, training areas are fairly homogeneous within the raw data set.

Having grown training areas as described, their areal extent was maintained as training areas for the corrected data set, with their statistics re-calculated using corrected bands. Because the extent of training areas was defined with respect to statistical properties of the raw data and artificially kept constant for the normalized classification, it would be expected that the same training areas would perform differently on each data set depending on the actual characteristics of each training area.

BAND	OBS VAR	EXP VAR	DF	CHI-SQR	X^2(0.05)[227]	H0: OVAR=EVAR
1	155.49	171.41	227	205.91	262.865	Accept
2	53.20	54.82	227	220.30	262.865	Accept
3	134.90	131.59	227	232.70	262.865	Accept
4	289.06	221.37	227	296.40	262.865	Reject
5	552.67	414.98	227	302.32	262.865	Reject
7	124.30	95.78	227	294.58	262.865	Reject

Table 5: Test for Equal Variances Between Raw and Corrected Training Sets.

Training area growth was based on existing homogeneity in raw data sets. Therefore, by definition, a certain degree of homogeneity existed with respect to the raw data. In turn, homogeneity observed in a raw training area could be due to different causes. Three main scenarios may be presumed:

- The site(s) is a truly homogeneous cover type on fairly homogeneous topography (Type 1).

- The site(s) includes heterogeneous cover types (maybe closely related types?) on heterogenous topography. The apparent DN homogeneity of the site is then really the result of topographic effect on reflectance (i.e., different cover types whose reflectance values have become similar due to different combinations of slope and effective incidence angles) - (Type 2).

- The site(s) includes heterogenous cover types, but individual reflectance values are overwhelmed (dampened) by strong overall topographic effect on illumination (i.e., heterogenous cover types on steep shaded mountain sides) which makes the site appears homogeneous (Type 3).

Upon proper application of a BRTC, pixel values should migrate closer to the true reflectance values of their respective cover types within and without training areas thus:

. Training areas of Type 1 would remain homogeneous after normalization, either by retaining raw DN(s), as in the case of Type 1 areas whose homogeneous topography would not solicit significant corrections) or by changing homogeneously toward the true spectral value of the cover type they represent, as in the case of Type 1 areas whose homogenous topography solicits the same overall correction for the entire area. The variance of this sites would remain similar to its raw counterparts or even decrease upon normalization, allowing these areas to perform as well or better than their raw versions, during normalized data classification.

. In training areas with topo-induced homogeneity (Types 2 and 3), the latter would be reduced by normalization, as pixel values migrate toward their corrected values. Differences in DN should arise. The actual number of migrating DN(s) and the magnitude of DN changes would, in turn, determine the magnitude of the subsequent increase in variance within training areas. In fact, some of the migrating pixels may not have been incorporated into the training area given the applied region-growth parameters; however, they remained because training areas were forced to remain constant for the corrected set. The inclusion of these, now deviant, pixel values in the normalized training areas could diminish their performance during classification, as observed in Closed Birch Training sites 1005, 1018, 380, and, to a lesser extent, in site 558.

Throughout the scene bands, pixel values would also migrate closer to the true reflectance value of the cover type they represent. Pixel value arrays defined in the raw training set would not necessarily exist in the same patterns in the normalized bands. Thus, some previously classified pixels would no longer fit within area-defined corrected training class statistics and become un-assigned during classification of the normalized data set. Given the applied methodology for the YUCH data set, we should expect a number of pixels that were assigned to training site types 2 and 3 in the raw data, to become un-classified in the normalized classification.

Conversely, previously unclassified pixels would find class assignments in the corrected training data set, depending on how many training areas remained good classification performers upon normalization. For training sites of the same cover type, it would be expected to see pixels previously assigned to raw site classes whose performance was diminished by normalization, migrate toward sites of the same class whose performance was enhanced upon normalization.

Results of the cross-tabulation between raw and normalized YUCH classifications show that while 1,238,512 previously classified pixels in the raw data set became un-classified in the normalized data set, another 2,321,553 previously unclassified pixels in the raw data set were assigned to classes in the normalized data set. Since identical training data sets (including number and actual extent of training areas) and classification algorithm (ML Classifier, 95% threshold, all bands, all classes) were applied to both raw and normalized data sets, the net gain of 1,083,041 (3.4%) newly classified pixels should be attributed to the normalization process.

The same cross-tabulation results show that, were we to maintain training site number 1005 assignment to be Closed Birch, that class would have experienced an overall 25% loss during normalized classification; however, when suspect site number 1005 is pulled from the group, the Closed Birch class experiences an increase of over 17,000 pixels or a 7% gain during classification of the normalized set (Table 6).

SITE NUMBER	GRID VALUE	TRAINING SITE PERFORMANCE		TOTAL PIXELS IN MAP	
		RAW	CORR	RAW	CORR
253	14	90.6%	96.9%	11,791	23,601
380	78	75.9%	62.0%	113,833	66,266
1018	84	64.2%	48.1%	27,056	27,069
293	96	81.5%	81.5%	48,780	99,830
558	104	81.3%	79.2%	19,342	15,076
1047	113	43.5%	58.0%	32,136	38,235
1005	182	81.8%	36.4%	436,789	246,767
SUBTOTAL (less 1005)				252,925	270,077
TOTAL				689,714	516,844

Table 6: Cross-tabulation Between Raw and Corrected Closed Birch Classes.

Cross-tabulation of raw and corrected training polygon data

A review of the cross tabulation of the raw class map and the corrected class map data within training areas is quite useful in providing understanding of the effects of the normalization process on the subsequent ability of the classifier to differentiate the digital data better and produce a more comprehensive and less confused map. In a perfect application with all training classes being separable with no overlap we would expect that the pixels within a training polygon would be classified as that particular training class. As the data are normalized we can expect two different results. The separability is maintained and the training area is still classified as itself or there is a shift in the classification and some of the pixels are now classified as another type. This latter result is due to normalization creating training class data that now overlaps another training class. In theory, this class should be of a similar or same cover type that was different in the raw bands due to differential slope and aspect conditions, but is now the same in the normalized bands. In this case, normalization results in a possible replication of training data, a situation that is undesirable as an unnecessary training set has been included in the training data set. We would thereforeanticipate, in some cases, a decrease in the percent of pixels in a training class that are classified as itself. This is in actuality what we see. However, we also see the opposite - we see the number of pixels classified as itself increase. This happens because in actuality, we find that every pixel within a training polygon does not get classified as it’s own value. There is typically some sort of mixture of values depending on how pure and homogeneous the site is and how many similar overlapping sites exist in the training set. As the data are normalized we do see a decrease in the number of pixels or percent of the area that is classified as itself as pixels are assigned a value of a similar class. But we can also find situations where the numbers increase. In this case, normalization has reduced confusion or overlap with other classes that were very similar in the raw data, but are now separable in the corrected data. The result of the improved separability in this case is an increase in the percent of the area that is classified as itself. This type of class now stands on its own statistically and is not confused with other classes. This is a highly desirable class, as it is unique. Thus we find that some sites stay the same, some sites now overlap other training sites, and some sites are now separable from other sites. A sample of a portion of the pixel fidelity spreadsheet is shown in Table 7. This table shows for each training class the number of pixels that were classified as the same value as the training class number for the raw and corrected imagery.

CLASS NUMBER	AREA COUNT	SAME RAW	RAW PCT SAME	SAME BTM	BTM PCT SAME	NUMBER DIFF	PCT DIFF
159	79	42	53.16%	18	22.78%	(24.00)	-30.38%
160	54	45	83.33%	20	37.04%	(25.00)	-46.30%
162	162	130	80.25%	138	85.19%	8.00	4.94%
163	72	20	27.78%	44	61.11%	24.00	33.33%
164	45	32	71.11%	26	57.78%	(6.00)	-13.33%
165	72	56	77.78%	54	75.00%	(2.00)	-2.78%
166	295	151	51.19%	168	56.95%	17.00	5.76%
167	82	60	73.17%	39	47.56%	(21.00)	-25.61%

Table 7: Pixel Fidelity Data

Accuracy Assessment

Nearly all of the aforementioned means of evaluating the effect(s) of the illumination correction as based on the use of inference and attempting to show that the imagery has been altered and that the alteration results in a beneficial results. The most valid approach to determining the validity of the normalization process is through an evaluation of the classification results. Did the normalized classification result in a map with a higher level of accuracy than the classification based on the raw data ? Only through a comparison of this nature can the difference between the raw and ‘corrected’ classifications be demonstrated. If confusion has been resolved, a classes are more separable, then the accuracy assessment should reflect these differences. While this was not a component of this study, it is a highly desirable means of evaluation. A new accuracy assessment should be performed on the basis of the prior accuracy assessment site and the ‘corrected’ class map. These results should be compared with the prior results based on the raw classification. Differences should be reflected through a comparison of this nature.

Summary

Normalization can have a positive effect on image classification efforts. What becomes apparent in this evaluation of raw versus normalized data is that potential training data sites should be reviewed in the imagery after the imagery is normalized, but before the sites are visited to perform field data collection activities. The magnitude of training data collection efforts and subsequent analysis will be minimized, if the land cover can be accurately classified using the minimum number of training sites. The overall cost of data collection efforts should be lower if the minimum number of sites are visited and evaluated. In addition, mapping accuracy may be increased if confusion is minimized and the chance of mislabeling a type is reduced.

If possible, training sites should be reviewed in the corrected imagery prior to initiating field data collection efforts. An evaluation of potential training sites should be performed in the imagery to determine the uniqueness of the spectral data, the applicability of the class to the project as a whole, and the cost of acquiring data from that particular site. The use of similar and overlapping training sites should be minimized and the best training sites, in terms of the lowest cost and difficulty in acquiring data, should be scheduled for data collection activities. The reduction in the number of training sites required for classification of an image is a major benefit of the illumination correction process.

TASK 2: LEVELS OF SUMMARIZATION

The second major effort of this project concerned the development of detailed cover attributes based on the image classification results. Demonstration of the capability to summarize pixel data and develop detailed estimates was originally planned to involve the processing of the original Ducks Unlimited (DU) raw pixel class map within the 5-acre minimum size polygons present in the delivered DU YUCH ARCView data set. The raw pixel class data would also be summarized relative to other pertinent data, such as the generalized DU land cover classes, to yield more detailed cover descriptions of the subject area. This effort would also demonstrate GRS' ability to use and integrate existing AKSO and DU data sources in the summarization process. This effort would also demonstrate how detailed information could be developed using the same basic data sources as were available during the initial mapping effort.

As work progressed on this task two discoveries were made: the original DU pixel class map was not available and the 5-acre YUCH map was not processed such that land cover units met a 5-acre minimum. The 5-acre YUCH map included numerous individual pixels as distinct land cover units. A total of over 3,160,495 cover units were present in this data set that included over 26,000,000 pixels, meaning that the average cover unit was close to 9 pixels or just less than 2 acres, rather than the anticipated minimum size of 5 acres. As the data necessary to demonstrate summarization were not available, the nature of this effort was altered to include the summarization of the GRS 'corrected' classification results with respect to the more generalized land cover units present in the YUCH map. To further demonstrate the capability to summarize data at other levels of interest, the corrected classification results would also be summarized with respect to the YUCH Major Type areas and the training site polygons used to develop training statistics.

The Development of Detailed Land Cover Information

The development of detailed land cover descriptions requires that the pixel data classes are each represented by accurate and sufficiently descriptive estimates of land cover characteristics and that these detailed data descriptions may be processed and summarized relative to another data set, such as distinct cover unit areas (polygons) or generalized lifeforms. In a typical image processing project, these specific land cover descriptions are referred to as training data; the data would be developed through field data collection efforts that result in quantitative estimates of cover by the various land cover components (percent cover by species or land cover type). These field data estimates are related to the different spectral classes that are identified during classification efforts and are used to 'describe' those different pixel classes. These data must provide representation of the entire project area as well as each individual training site. If the entire image is to be classified and mapped, then the training data must represent each of the different cover classes thought to exist within the project area. Additionally, all land cover components at each training site must be included to account for all 100 percent of the land cover that is represented by each pixel class. Missing cover types or land cover components will result in holes or gaps in the land cover data set. A sample cover description is shown in Table 8. These detailed data descriptions provide comprehensive information regarding the various cover components, even if those components are not part of the predominant cover type (e.g. the duff/litter component in this closed mixed Needleleaf/Broadleaf type). The retention and use of the complete cover description enables the subsequent development of accurate and detailed cover information (note: any references to size class in these distributions are purely arbitrary in this project and have no significance - they are a byproduct of the software which is capable of generating information describing tree size). The intent of this task is to demonstrate the ability to develop such detailed information from an image classification effort.

Training Site Cover Description
Training Area: 2
Training Site Number: 62

Training Area Density Summary:

Size Class:	0-4"	5-8"	9-12"	>=13"	%Tree Cover	%Other Cover	Total Cover
Black Spruce	0.0%	0.0%	40.0%	0.0%	40.0%		40.0%
Hardwood	0.0%	0.0%	10.0%	0.0%	10.0%		10.0%
Tall Shrub						15.0%	15.0%
Low Shrub						10.0%	10.0%
Moss						20.0%	20.0%
Duff						5.0%	5.0%
Total Cover	0.0%	0.0%	50.0%	0.0%	50.0%	10.0%	100.0%
Total Tree Cover					50.0%		

Tree Density Summary:

Size Class:	0-4"	5- 8"	9-12"	>=13"	%Tree Cover
Black Spruce	0.0%	0.0%	80.0%	0.0%	80.0%
Hardwood	0.0%	0.0%	20.0%	0.0%	20.0%
Total Tree Cover	0.0%	0.0%	100.0%	0.0%	100.0%

Table 8: Training Site Cover Description

Land Cover Class (Training) Cover Descriptions

As this project did not include image processing efforts by **GRS**, the training data descriptions that **GRS** would typically use were not developed using transect-based field sampling techniques. However, other potential data sources were available for use in this effort. The original DU mapping project used training data representing 236 land cover data collection sites during the classification of the YUCH Land Cover Mapping Project. These training data descriptions were developed from both DUFF and AKSO IMA data collection efforts. These descriptions were primarily developed during on-site data collection efforts based on ocular estimates (nearest 5%) of cover made from a helicopter hovering over the different sites. As with any data collection effort based on less than a 100% sample, some errors will be associated with the field data. In an image classification effort these errors may be significant, as they will be imbedded throughout the resulting data descriptions wherever that class is applied. All data must be reviewed to determine their validity and appropriateness for use in this project.

After review and evaluation of the prospective data, a total of 228 existing training site descriptions were used to describe the land cover at these 228 field data locations. Data descriptions were reviewed and sites having more or less than 100% cover were adjusted to reflect the calculated cover label. In a few cases, cover equaled or exceeded 200%. These cases involved sites that lacked unique site numbers or areas for which too much cover was recorded. Duplicate site numbers were resolved by means of segregating the data by the area in which it was collected - either Yukon-Charley or Blacks River. Areas with excessive cover were adjusted downward to a total of 100%, typically by reducing cover values that were apparently understory cover components, rather than overstory components.

After data cleaning was completed, these data were processed and loaded into the database to provide cover descriptions for each of the training sites that would be used in the GRS classification and summarization efforts. Data specific to genus and species, or the lowest level of specification (e.g. gravel, duff, and so forth) were loaded into the database. Data were processed in a manner consistent with prior efforts GRS accomplished for the AKSO, when data were developed for the Wrangell-St Elias Natural Preserve based on the definitions of The Alaska Vegetation Classification (Viereck). Although this classification system was not identical to the YUCH mapping effort, it was sufficiently similar to facilitate the development of the detailed cover information without need for additional programming and modification of existing processes. While this classification system's rules and definitions were not exactly the same, they were similar enough for use in this task.

The cover description for one of the existing training sites (number 9), a closed mixed-Needleleaf/Broadleaf type, is shown in Table 9. Data like these data were processed and loaded into the relational database into the *tr_site_info* table, which stores training site field estimation information. A listing of this record from this database table is shown in Table 10.

Training Site Cover Description
Training Area: 9
Training Site Number: 190

Training Area Density Summary:

Size Class:	0-4"	5-8"	9-12"	>=13"	%Tree Cover	%Other Cover	Total Cover
Black Spruce	0.0%	0.0%	50.0%	0.0%	50.0%		50.0%
Hardwood	0.0%	0.0%	40.0%	0.0%	40.0%		40.0%
Duff						10.0%	10.0%
Total Cover	0.0%	0.0%	90.0%	0.0%	90.0%	10.0%	100.0%
Total Tree Cover					90.0%		

Tree Density Summary:

Size Class:	0-4"	5- 8"	9-12"	>=13"	%Tree Cover
Black Spruce	0.0%	0.0%	55.6%	0.0%	55.6%
Hardwood	0.0%	0.0%	44.4%	0.0%	44.4%
Total Tree Cover	0.0%	0.0%	100.0%	0.0%	100.0%

Table 9: Training Site Cover Description

site_num	[190]
pix_ct	[45]
acreage	[10.0]
lform	[T]
ltype	[PHw]
pr_species	[Black Spruce]
pred_sp_pct	[55.6]
closure_class	[3]
density	[90.0]
pct_conifer	[55.6]
pct_hdwood	[44.4]
other_cover	[0.0]
cv_shr	[0.0]
cv_hrb	[0.0]
cv_aqu	[0.0]
cv_bar	[10.0]
cv_nonveg	[10.0]

Table 10: Training Site Database Record Listing

Detailed Data Development

The detailed land cover characteristics of any particular area of interest are developed by summarizing the different pixel class data that overlap or correspond with the subject area. The overlap is determined by cross tabulating the pixel classes and the subject area ids (or values). It is represented by listing the subject area id (value), the pixel class number, and the frequency of that pixel class in that subject area. For n classes of pixels this information is represented symbolically as follows:

area_id_1 , pixelClass_1 , frequency_1
area_id_1 , pixelClass_2 , frequency_2
area_id_1 , pixelClass_3 , frequency_3
...
...
...
area_id_1 , pixelClass_n , frequency_n

The final estimate of any land cover unit's vegetation characteristics is based on the weighted tabulation and summarization of the different classes of pixels found within each area. This task is accomplished for each cover area in the project area using the process ***GRS_polysum***. ***GRS_polysum*** computes the weighted average cover description for each subject area on the basis of the frequencies of the different pixel classes present in each subject area. The process generates reports and formatted data files that may be imported into relational database tables. Output is based upon user defined criteria and guidelines. Each cover description yields an estimated distribution of land cover by cover characteristic (see Table 12 for an example).

Data Definition

A listing of the land cover database definition is provided in Table 11. This listing reflects the different land cover attributes that were loaded into the relational database table during this particular effort. More data items may be loaded if the data are present in the cover data descriptions and the user(s) wishes to include the data items. These data items reflect only those items loaded in this particular case.

An example of the land cover unit characteristics that may be estimated for a specific area is listed in Table 13. This listing is based on the weighted average cover distribution developed from summarization of the specific number of pixels of each pixel class present in the subject area. These characteristics include generalized categorical characteristics such as the cover type class, the canopy closure class, and the primary vegetation type, as well as discrete estimates of other characteristics, such as the specific percent total tree cover, percent conifer, percent hardwood, shrub cover, and herbaceous cover.

COLUMN	ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC
1	VALUE	4	8	B	-
5	DU_CLASS	6	6	I	-
11	LTYPE	3	3	C	-
14	PR_SPECIES	14	14	C	-
28	CLOSURE_CLASS	1	1	C	-
29	DENSITY	4	14	F	4
33	OTHER_COVER	4	14	F	4
37	PCT_CONIFER	4	14	F	4
41	PCT_HDWOOD	4	14	F	4
45	SIZE_CLASS	6	6	I	-
51	QMD_ALL	4	14	F	4
55	QMD_CON	4	14	F	4
59	QMD_HWD	4	14	F	4
63	CV_SHR	4	14	F	4
67	CV_HRB	4	14	F	4
71	CV_AQU	4	14	F	4
75	CV_BAR	4	14	F	4
79	CV_OTH	4	14	F	4
83	PRED_SP_PCT	4	14	F	4
87	PIX_CT	11	11	I	-
98	COUNT	4	10	B	-
102	LFORM	2	2	C	-
104	RMP_VAL	3	3	I	-

Table 11: Database Record Definition (ITEMS)

All of these different estimates are significant with respect to the possible use of the data. The estimation of these specific values enables future reclassification of these subject areas using modified or alternative class definitions. The data provide flexibility to the data user as they may be re-evaluated with respect to modified class boundaries and definitions to develop new class maps or themes of data. For example, the subsequent development of lifeform (column name lform) estimates demonstrates the flexibility of the data. Generalized class data alone cannot be re-evaluated in this manner. These additional data also enable cross-walking to different vegetation classification systems.

There are other benefits of using these detailed data. One major benefit is that the detailed data provide a much more meaningful basis for comparison compared to categorical values. Accuracy assessment data may be compared statistically with these data to test confidence limits and the matching of accuracy assessment data with the mapped data, rather than using fuzzy logic to estimate differences of categorical estimates. Field estimates of cover or density may be tested relative to the mapped estimates and differences in these values may be tested for significance even when they span the thresholds or limits of the class boundaries. Species composition is another value that is described with detailed data much better than general categorical values that only indicate a single type, or a mixture of that type with another type. Thresholds and limits may change, but these detailed data are not effected by those changes; they enable the recalculation of the categorical values when the limits are modified. Differences between a field estimate and a mapped estimate are much easier to test as quantitative values rather than general types. A mixed type is very similar to a single species type when the field data and the mapped data are very close in value, but span the threshold of the definition (e.g. field data indicate composition of 73% Black Spruce and 27% Broadleaf - a Mixed Needleleaf/Broadleaf type - as opposed to mapped data that indicate composition of 78% Black Spruce and 22% Broadleaf - a Needleleaf type).

Summary Example #1: Summary by YUCH Land Cover Unit

The most useful level of summarization of the pixel class data is with respect to the distinct land cover units that have been mapped using image classification methods.
A total of 3,160,495 distinctly separable land cover units were mapped in the DU YUCH data set. Each of these areas was assigned a unique id and processed relative to the GRS 'corrected' class map to yield a data set that contained the frequency of pixel class by cover unit for each cover unit in the YUCH data set. An example of a listing of pixel frequency by pixel class for cover unit 437543 follows:

437543 , 8 , 1
437543 , 22 , 1
437543 , 35 , 1
437543 , 48 , 1
437543 , 52 , 2
437543 , 54 , 2
437543 , 99 , 1
437543 , 116 , 5
437543 , 122 , 2
437543 , 160 , 1
437543 , 166 , 1
437543 , 182 , 2
437543 , 221 , 6

In this case, cover unit 437543 is represented by a total of 26 pixels in 13 different classes. These data were then processed by ***GRS_polysum*** to yield cover unit land cover summaries. An example of such a summary is shown in Table 12. The process also generates categorical estimates, based on the rules and limits that have been defined to represent the different categorical values. Each land cover area's characteristics were estimated by evaluating the distribution of cover using the user-defined classification rules and definitions. Density class values, lifeform, and land cover type values are three examples of categorical values that can be generated from the cover descriptions. The rules that define these categorical estimates are programmable and can be modified to reflect different information needs. These data were then loaded into the relational database table associated with the unique values representative of the individual land cover units. A database record listing for this cover unit is shown in Table 13.

Land Cover Density Summary:
Land Cover Area #: 437543
Total Number of Pixels: 26
Contributing Pixels: 26

Size Class:	0-4"	5-8"	9-12"	>=13"	%Tree Cover	%Other Cover	Total Cover
White Spruce	0.0%	0.0%	4.2%	0.0%	4.2%		4.2%
Black Spruce	0.0%	0.0%	7.5%	0.0%	7.5%		7.5%
Hardwood	0.0%	0.0%	16.9%	0.0%	16.9%		16.9%
Tall shrub						24.0%	24.0%
Dwarf shrub						13.8%	13.8%
Misc shrub						5.2%	5.2%
Graminoid						6.3%	6.3%
Forb						11.0%	11.0%
Wet moss						2.9%	2.9%
Barren						8.1%	8.1%
Total Cover	0.0%	0.0%	28.7%	0.0%	28.7%	71.3%	100.0%
Total Tree Cover					28.7%		

Tree Density Summary:

Size Class:	0-4"	5- 8"	9-12"	>=13"	%Tree Cover
White Spruce	0.0%	0.0%	14.8%	0.0%	14.8%
Black Spruce	0.0%	0.0%	26.2%	0.0%	26.2%
Hardwood	0.0%	0.0%	59.1%	0.0%	59.1%
Total Tree Cover	0.0%	0.0%	100.0%	0.0%	100.0%

Table 12: Land Cover Unit Cover Summarization

The storage of the detailed information in relational database tables greatly facilitates the use of this information and the development of related data. The development of categorical values is a major benefit of generating detailed information. Some values may be reassigned if definitions are modified. For example, in this effort there were three tree density classes defined - woodland (10-24.9%), open (25-59.9%), and dense (>=60%). If more classes were desired, for example, five classes of width equal to 20%, the assignment of this new density class may be made using SQL commands in the relational database management system (e.g. update tablename set closure_class = 5 where density >= 80.0 or update tablename set closure_class = 4 where density between 60.0 and 80.0).

In this particular project the detailed data made possible the estimation of a new database item - a lifeform estimate. This value was not developed during the prior mapping/image processing project. Such a value was desired to represent the predominant component of the cover, even if that component was not represented in the label/name. For example, a Woodland Needleleaf cover unit might be 20% spruce cover and 80% shrub cover. This cover unit's lifeform would be shrub(**S**) rather than tree(**t**), as the shrub component was the dominant component of the unit's cover. The estimation of lifeform was accomplished by applying SQL statements to the relational database information after all processing and database loading were finished. This column was populated using the logic shown in Appendix A. These rules may be modified to result in the assignment of different lifeform values depending on the user's needs, thereby resulting in many possible (different) maps. Other database values, such as an index of vegetation diversity or fuel load, may be generated, provided the detailed cover data support the development of this additional information. The limitations on the use of these data are now defined by the accuracy of the data and the validity of the usage and relationships, rather than the generalized nature of the information.

```
value                [437543     ]
du_class             [17    ]
lform                [s]
ltype                [PHw]
pr_species           [Hardwood      ]
pred_sp_pct          [59.1      ]
closure_class        [2]
density              [28.7      ]
other_cover          [63.3      ]
pct_conifer          [40.9      ]
pct_hdwood           [59.1      ]
cv_shr               [43.1      ]
cv_hrb               [20.2      ]
cv_aqu               [0.0       ]
cv_bar               [0.0       ]
cv_nonveg            [8.1       ]
pix_ct               [26          ]
count                [26          ]
rmp_val              [437543   ]
```

Table 13: Land Cover Unit Database Record Listing

Summary Example #2: Summary by Major DU Class

The preceding example of summarization involved the development of detailed information for very specific cover units that averaged approximately 2 acres in size. A second potential level of summarization is with respect to the major land cover classes present in the DU YUCH data set. There were 30 major classes present in the DU YUCH class map that ranged from various tree, shrub, and herbaceous types to non-vegetated types representing water and barren areas. The development of detailed information descriptive of these general classes may be useful in evaluating the estimated cover characteristic differences represented by these different classes. This information may also be useful in determining the effectiveness of the field data collection and classification efforts, and the potential mislabeling or confusion/overlap of data amongst these Major Classes.

In this example, the GRS 'corrected' pixel class map was processed with respect to the original pixel grid containing the DU Major Class values to generate a listing of pixel class frequency by Major Class value. An example of a listing of pixel frequency by pixel class for DU Major Class 1 follows:

1,0,9816	1,48,27	1,92,607	1,132,6	1,187,614
1,1,5	1,52,137	**1,93,2043**	1,135,49	**1,188,16255**
1,2,1872	1,54,54	**1,94,1791**	1,137,3	1,189,24
1,3,247	1,58,1	1,95,132	**1,138,15758**	1,190,100
1,4,1	1,59,5	1,96,12	1,139,1	1,191,5
1,5,2	1,60,1	1,97,5	1,140,18	1,193,234
1,6,4	1,61,5	**1,98,1117**	1,141,10	1,194,10
1,7,4	1,62,122	1,99,103	1,142,335	1,196,86
1,8,141	1,63,11	1,100,3	1,143,18	1,198,66
1,9,86	**1,64,1206**	1,102,3	**1,144,22595**	1,199,2
1,10,62	1,65,3	1,103,4	1,149,1	1,200,126
1,11,57	1,66,4	1,104,3	**1,151,1272**	1,201,10
1,12,1	1,67,6	**1,105,1461**	1,152,1	**1,203,9229**
1,14,1	1,68,943	**1,106,1371**	**1,153,1120**	1,204,8
1,16,1	1,69,2	1,107,12	**1,155,1783**	1,206,4
1,17,6	1,70,167	1,108,507	1,156,7	1,208,2
1,18,1	**1,71,1465**	1,109,203	1,157,137	1,211,23
1,20,267	**1,73,2149**	**1,110,1561**	1,158,1	**1,212,2806**
1,21,1	1,74,146	1,111,16	1,159,11	1,213,363
1,22,9	**1,75,12651**	1,112,120	1,160,56	**1,214,8572**
1,24,11	1,76,497	1,113,2	1,166,19	1,215,4
1,25,2	1,78,29	1,114,48	1,168,519	1,216,2
1,27,7604	1,79,844	1,115,122	1,169,1	1,219,881
1,28,569	1,80,689	1,116,60	1,170,29	1,221,27
1,29,1698	1,81,1	**1,119,1818**	1,172,3	1,222,5
1,30,9	1,82,226	1,120,3	1,173,13	1,225,2
1,31,2404	1,83,1	1,121,12	**1,174,2783**	1,226,3
1,32,12746	1,84,8	1,122,178	1,176,1	1,227,78
1,34,11	1,85,983	1,123,237	1,177,134	1,230,1
1,35,58	**1,86,1735**	1,124,5	1,180,46	
1,36,895	1,87,1	1,125,1	1,182,152	
1,37,3	1,88,21	1,126,1	1,183,56	
1,41,3	1,89,23	1,127,9	1,184,29	
1,44,3	1,90,13	1,130,164	1,185,83	
1,45,1	1,91,9	1,131,1	1,186,17	

In this case, Major Class 1 is represented by a total of 163,130 pixels in 169 different classes. Some of these classes are represented by tens of thousands of pixels, while others are represented by as few as 1 pixel. This situation is illustrative of the variety of different pixel classes that would be found in a very generalized class, such as the Closed Needleleaf class. The pixel classes with large numbers of pixels are most likely individual training classes with characteristics similar to those of the Major Class characteristics. The isolated types are likely small isolated types that may have been ‘cleaned’ or removed from the original pixel map prior to development of the final DU YUCH cover map, which were not filtered out of the GRS ‘corrected’ pixel map (note: the large number of types present in this Major Class tends to indicate that some cleaning of the DU YUCH data set did take place and that many small isolated types were removed from the final data set).

These data were then processed by ***GRS_polysum*** to yield Major Class land cover summaries. An example of such a summary is shown in Table 14, which is a summary of Major Class 1, the Closed Needleleaf type. The process also generates categorical estimates, based on the rules and limits that have been defined to represent the different categorical values. These detailed data are similar to the detailed data developed for the specific cover units, except each set of values describes a DU Major Class. These data were loaded into a relational database table (*duclassbtm.dat*) and used to describe the general Major Classes.

A listing of some of the primary data items by DU Major Class is shown in Table 15. These data were developed during this summarization effort and may be helpful in the further use of this information. It is interesting to note that the average tree density of 50.4% that is shown in the summary is outside the bounds of the ‘closed’ definition, a class that represents cover units with a tree density of >= 60% cover. When comparing these detailed estimated to the generalized values the ’closed’ class is under represented in the Needleleaf and Broadleaf types, but barely above the lower limit in the Mixed Needleleaf/Broadleaf class. This is indicative of a general trend that GRS has noted in other mapping projects - ocular estimates of cover often overestimate the actual cover present at the site.

Area Cover Density Summary:
DU Class: 1
Total Number of Pixels:163130
Contributing Pixels:153314

Size Class:	0-4"	5-8"	9-12"	>=13"	%Tree Cover	%Other Cover	Total Cover
White Spruce	0.0%	0.0%	26.0%	0.0%	26.0%		26.0%
Black Spruce	0.0%	0.0%	16.7%	0.0%	16.7%		16.7%
Hardwood	0.0%	0.0%	7.7%	0.0%	7.7%		7.7%
Tall shrub						9.2%	9.2%
Low shrub						1.6%	1.6%
Dwarf shrub						1.3%	1.3%
Misc shrub						5.9%	5.9%
Graminoid						2.2%	2.2%
Forb						6.1%	6.1%
Wet moss						5.0%	5.0%
Lichen						3.7%	3.7%
Aquatic						1.9%	1.9%
Water						2.7%	2.7%
Barren						10.0%	10.0%
Total Cover	0.0%	0.0%	50.4%	0.0%	50.4%	49.6%	100.0%
Total Tree Cover					50.4%		

Tree Density Summary:

Size Class:	0-4"	5- 8"	9-12"	>=13"	%Tree Cover
White Spruce	0.0%	0.0%	51.6%	0.0%	51.6%
Black Spruce	0.0%	0.0%	33.2%	0.0%	33.2%
Hardwood	0.0%	0.0%	15.2%	0.0%	15.2%
Total Tree Cover	0.0%	0.0%	100.0%	0.0%	100.0%

Table 14: DU Class Cover Summarization

DU Cl#	ltype	tree cover	percent conifer	percent hardwood	shrub cover	forb cover	pixel count
1	UnP	50.40	84.80	15.20	17.90	17.10	163130
2	PHw	39.40	74.70	25.30	26.00	27.90	10518338
3	UnP	36.40	89.60	10.40	21.80	36.30	7603
4	PHw	22.60	59.60	40.40	36.10	37.00	4094321
5	PHw	19.60	68.10	31.90	35.20	41.90	162686
6	PHw	15.40	74.50	25.50	36.20	45.30	30331
10	Hwd	63.40	19.50	80.50	19.20	11.30	1184296
13	Hwd	35.00	18.30	81.70	31.00	26.00	124614
16	PHw	57.40	45.80	54.20	17.50	16.00	1289556
17	PHw	42.10	38.70	61.30	29.90	21.60	1118249
20	PHw	25.40	28.70	71.30	40.70	28.20	246577
21	PHw	15.40	40.00	60.00	41.30	36.70	2370239
22	DSh	1.80	0.00	0.00	37.20	57.30	686
23	PHw	12.90	51.90	48.10	37.90	43.80	1118477
24	DSh	8.60	26.70	73.30	36.20	22.90	820999
32	DSh	6.70	56.90	43.10	28.50	61.40	15363
34	UnP	18.30	78.90	21.10	23.40	46.10	2848
40	DSh	6.50	24.70	75.30	40.80	32.20	133984
50	DSh	8.00	54.40	45.60	35.50	52.50	235323
51	DSh	2.10	0.00	0.00	28.90	64.70	14278
60	UnP	12.10	89.10	10.90	10.40	21.10	367
70	UnP	17.80	81.20	18.80	14.50	27.90	80350
71	???	0.00	0.00	0.00	0.00	0.00	193462
72	???	0.00	0.00	0.00	0.00	0.00	33839
80	PHw	11.20	35.00	65.00	29.60	23.00	486104
81	DSh	5.60	28.80	71.20	27.40	14.30	589284
92	PHw	11.00	47.00	53.00	50.10	13.20	4315
93	UnP	29.30	85.20	14.80	11.50	14.40	2571
94	UnP	37.20	78.20	21.80	15.90	19.00	849238
96	PHw	14.40	61.90	38.10	22.40	44.40	329077

Table 15: Estimated DU Class Cover Attributes

Summary Example #3: Summary by Training Area

In the same way that a land cover unit may be summarized with respect to the pixel classification map, so may other area features. An example of such a feature is the training site polygon used to collect training statistics. The development of land cover descriptions for training polygons is a particularly useful operation, as it provides a cross-check on the data collection and classification processes - the land cover description generated by processing the classified pixels should be quite similar to the original land cover description developed during the field examination of the site. Discrepancies between the field data descriptions and the generated descriptions indicate a problem and raise questions regarding the appropriateness of the use of this training site in the image classification process.

A total of 230 training areas were processed to generate the cross tabulation of the training area number and the GRS 'corrected' pixel class. A sample distribution for training unit #9 (site_num 190) , an area 45 pixels in size, follows:

9 , 2 , 12
9 , 9 , 28
9 , 28 , 5

These data were then processed by ***GRS_polysum*** to yield cover unit land cover summaries. An example of such a summary is shown in Table 16. The process also generates categorical estimates, based on the rules and limits that have been defined to represent the different categorical values. Each training area's characteristics were estimated by evaluating the distribution of cover using the user-defined classification rules and definitions. Density class values, lifeform, and land cover type values are three examples of categorical values that can be generated from the cover descriptions. These data were then loaded into the relational database table (*trsiteinfo.dat*) associated with the unique values representative of the individual training units. These data may be easily related to the training field data estimates and compared with the training site field cover description. An example of the corresponding database listing is shown in Table 17. A side-b-side listing of the original field training site estimates relative to the cover attributes based on the pixel class summarization is shown in Table 18. The data are fairly similar, except that the tree cover is estimated at only 75% on the basis of the pixel classes as opposed to the field estimate of 90% tree cover. Tree specie composition is nearly the same, representing a mixed Needleleaf/Broadleaf type, except there are also minor shrub and forb components in the pixel based estimate that were not present in the field description. It is interesting to note that the two other pixel classes assigned to this training area were supposedly an 'open' 50% cover Needleleaf type (2) and an 'open' 55% cover Needleleaf type (28), as described in the field descriptions. These two training areas are described as an 'open' 55.3% cover Needleleaf/Broadleaf type and an 'open' 54.3% cover Needleleaf type on the basis of the pixel summarization data. Both these area descriptions correspond much better with their field descriptions than does the 'closed' 90% cover field description with the 75.4% cover pixel summary based value.

Training Site Area Description
Training Area: 9

Training Area Density Summary:
Total Number of Pixels: 45
Contributing Pixels: 45

Size Class:	0-4"	5-8"	9-12"	>=13"	%Tree Cover	%Other Cover	Total Cover
Black Spruce	0.0%	0.0%	47.3%	0.0%	47.3%		47.3%
Hardwood	0.0%	0.0%	28.1%	0.0%	28.1%		28.1%
Tall shrub						5.7%	5.7%
Misc shrub						2.7%	2.7%
Forb						1.1%	1.1%
Wet moss						7.0%	7.0%
Barren						8.1%	8.1%
Total Cover	0.0%	0.0%	75.4%	0.0%	75.4%	24.6%	100.0%
Total Tree Cover					75.4%		

Tree Density Summary:

Size Class:	0-4"	5- 8"	9-12"	>=13"	%Tree Cover
Black Spruce	0.0%	0.0%	62.7%	0.0%	62.7%
Hardwood	0.0%	0.0%	37.3%	0.0%	37.3%
Total Tree Cover	0.0%	0.0%	100.0%	0.0%	100.0%

Table 9A: Training Site Area Cover Summary

tr_site_id	[9]
pix_ct	[45]
acreage	[10.0]
lform	[T]
ltype	[PHw]
pr_species	[Black Spruce]
pred_sp_pct	[62.7]
closure_class	[3]
density	[75.4]
pct_conifer	[62.7]
pct_hdwood	[37.3]
other_cover	[16.2]
cv_shr	[8.4]
cv_hrb	[8.1]
cv_aqu	[0.0]
cv_bar	[8.1]
cv_nonveg	[8.1]
grid_val	[9]

Table 17: Training Site Area Database Record Listing

	Training Area Pixel Summary	Training Site Field Description
tr_site_id	[9]	
pix_ct	[45]	
acreage	[10.0]	
lform	[T]	[T]
ltype	[PHW]	[PHW]
pr_species	[Black Spruce]	[Black Spruce]
pred_sp_pct	[62.7]	[55.6]
closure_class	[3]	[3]
density	[75.4]	[90.0]
pct_conifer	[62.7]	[55.6]
pct_hdwood	[37.3]	[44.4]
other_cover	[16.2]	[0.0]
cv_shr	[8.4]	[0.0]
cv_hrb	[8.1]	[0.0]
cv_aqu	[0.0]	[0.0]
cv_bar	[8.1]	[10.0]
cv_nonveg	[8.1]	[10.0]

Table 18: Training Site Data Comparison

Whether or not this is a significant difference and if so, which estimate is more accurate, cannot be known from these data, but the evaluation of the training unit field data estimates relative to the pixel classification summary data does provide a useful means of identifying potential inconsistencies and problems that may effect the data set. Serious data problems can be identified and resolved prior to the production of a final map thereby improving the quality of the mapping effort.

Data Modeling and Grid Reclassification

The development of detailed data enables the modeling of these data in a grid format. As a result, many different grids or data sets may be developed to represent a variety of different land cover characteristics. In this particular project, a grid representative of the individual land cover units was created in an ARCGrid format. The detailed land cover data were associated with each grid value (cover unit). Through the application of SQL statements the cover data may be processed to develop different data sets. An example of this capability is the estimation of lifeform that was developed through the application of the SQL logic shown in Appendix A. The statements generated a lifeform value for every cover unit. This value may then be mapped in ARCGrid by application of the reclass function. A resulting grid data set may be generated to represent lifeform values across the project area. Other maps may be developed in a similar manner by modeling the detailed information to develop other data values and then mapping those values. The site specific detailed data make this process possible.

Reclassification of a grid is a useful means of developing new grids representative of other land cover information. This reclassification is available through the use of the ArcInfo/GRID reclass function, which outputs a new grid on the basis of the value of one of the exisiting grid's .vat table column's values.

During this project two data sets were developed in this manner. One data set represented the 30 DU major land cover classes represented by the 230 raw training classes in the raw class map. The other data set represented the same 30 DU major land cover classes, but was based on the 'corrected' or normalized class map. In each case, the individual pixel training class values (230 values) were mapped to the more general DU major classes (30 values) using SQL statements applied to the .vat table data. Each of the 230 pixel training classes was remapped to a DU major classes, by storing the DU Class value associated with each pixel training class value in the rmp_val column. New grids, each indicative of the DU Class values with respect to the raw and 'corrected' image class maps are easily generated by applying the the reclass function to the .vat tables. An example of this ArcInfo/GRID command statement is:

duclass = BTM_CLSMP.rmp_val

A new grid called *raw_clsmp* is created with the same projection, orientation, and pixel size as the original grid *grs_yuch*. As long as the detailed map information in the .vat enables the development of 'additional' data items, new grids representative of these new data may be generated. This function may be used repeatedly with different column values to generate new and different grid data sets. The detailed information generated through the different levels of pixel summarization greatly facilitates the development of the new and varied data sets.

DELIVERABLES

Data and information is delivered to the **AKSO** on cdrom. Data will be delivered in a format compatible with the **AKSO** hardware and the ArcInfo GIS. All data is delivered in a native ArcInfo format, as an excel spreadsheet, or as an ascii text file.

1. Imagery

   A. *yuch_norm.img* - the 6 bands of illumination corrected imagery (ERDAS .img format).

2. ArcInfo/GRID Data

   A. *grs_yuch* - the **GRS** 'corrected' class map and associated ArcInfo/GRID database table containing over 3-million land cover unit summary records. Land cover units were mapped on the basis of unique areas of DU classes in the DU YUCH data set.

   B. *du_class* - the grid map based on the original Ducks Unlimited land cover map containing 30 pixel classes representing major land cover classes.

   C. *raw_clsmp* - the classifcation map based on the raw imagery containing pixel classes 1-230 representing different training sites.

   D. *btm_clsmp* - the classifcation map based on the corrected imagery containing pixel classes 1-230 representing different training sites.

   E. *rawnobtm* - the grid data set representing classified raw pixel locations populated with DU Class values for which corrected locations were NOT classified.

   F. *btmnoraw* - the grid data set representing classified corrected pixel locations populated with DU Class values for which raw locations were NOT classified.

   G. *yuch_dem* -the DEM data set associated with the imagery.

   H. *yuch_slpe* - the slope data set associated with the imagery.

   I. *yuch_asp* - the aspect data set associated with the imagery.

3. ArcInfo Coverages

   A. The *trainsites* point coverage that contains field-based land cover locations and associated descriptive information developed from the DUFF and IMA data sets.

4. ArcInfo/INFO Data

A. The *duclassbtm.dat* table that contains detailed land cover summary information for the 30 major land cover classes in the DU YUCH map.

B. The *trsiteinfo.dat* table that contains detailed land cover field summary information for the 230 training class polygon areas used in classification of the DU YUCH map.

C. The *traindubtm.dat* table that contains detailed land cover summary information based on the correct/normalized image classification.

5. Other Data Files

A. *metadata.txt* - a text file containing metadata concerning data developed during this project.

B. *comp_train.rpt* - ASCII report listing the IMA/DUFF field land cover estimates and the raw and corrected training area detailed estimates based on summarization of the raw and corrected pixel classification data within training site polygons.

C. *trn_pixel_fidelity.xls* - an Excel spreadsheet representing the pixel fidelity (percent classified as same pixel) for the 230 training site polygons based on classification of the raw and corrected imagery.

D. *rawxbtmducl.xls* - an Excel spreadsheet representing DU Class cross-tabulation values with respect to the evaluation of the *raw_clsmp* and *btm_clsmp* grid data sets that have been reclassified to represent DU class values.

E. *duclass.xls* - an Excel spreadsheet that contains training class numbers (1-230) and their corresponding DU major land cover class. This file was the basis for the development of the btm and raw DU major class data sets using the reclass function with the raw and ‘corrected’ class map data.

F. *reclass.aml* - an ArcInfo aml script that generates a *lifeform* grid from the *grs_yuch* grid on the basis of the rmp_val column values.

G. *npsa02prj.txt* - an ArcInfo projection file containing information describing the projection of the project data.

RECOMMENDATIONS

GRS has several recommendations regarding mapping efforts of this nature. These recommendations involve efforts that concern the performance of an illumination correction and the collection and application of training data.

With respect to the illumination correction of the satelite imagery, the collection of image processing training data, and the development of detailed site-specific land cover estimates **GRS** recommends:

1. Perform BRTC to normalize imagery (prior to determining location of TS(s)). This may reduce the number of training sites necessary to classify the image.

2. Do not collect TS data in areas of high slope (> 45 degree slope is OUT), even if snow or barren land cover types.

3. Do not collect TS data in areas of variable aspect and/or slope. Attempt to find highly homogeneous sites with respect to the topography.

4. Ensure a high correlation/registration between imagery and DEM data sets, by using same set of DEM(s) used in image rectification process.

5. Perform an Accuracy Assessment to test the difference between classifications developed based on raw and 'corrected' imagery. This assessment should include sample areas across all land cover types that provide a suitable sample of different slopes and aspects. Review the sample data with respect to slope and aspect to determine whether or not there is any correlation os success or failure with respect to topography.

6. Describe all field data collection sites using a methodology that accounts for all of the various cover components. Care must be taken to be certain that the cover estimates total to 100% cover.

7. Develop estimates of pixel fidelity to estimate the number of pixels within any given training polygon that are classified as that pixel value. These data provide very useful information that may be used to identify training classes that are being classified as the same training class as opposed to classes that are being classified as another class. This information identifies training areas that are seperable from other classes from those that are confused with other training classes. Review of land cover estimates by training class enable the determination of confusion between similar land cover classes as opposed to different land cover classes. Confusion between different land cover classes can be identified and further investigated.

BIBLIOGRAPHY

Chiou, C.R., L. T. Vernon, and R. M. Hoffer. 1992. Comparison of four techniques for topographic normalization of LANDSAT TM data. American Society of Photogrammetry and Remote Sensing. Vol(?), No. (?), pp. (?).

Colby, J. D. 1991. Topographic normalization in rugged terrain. Photogrammetric Engineering and Remote Sensing. Vol. 57, No. 5, pp. 531-537.

Smith, J. A., T. L. Lin, and K. J. Ranson. 1980. The Lambertian assumption and LANDSAT data. Photogrammetric Engineering and Remote Sensing. Vol. 46, No. 9, pp. 1183-1189.

The Lambertian assumption and LANDSAT data. Photogrammetric Engineering and Remote Sensing. Vol. Vol(?), No. (?), pp. (?).

APPENDIX A: Pixel Class Lifeform Determination Logic

Lifeform Determination:	Lform Class
if tree cover >= 50%	**T**
else if shrub cover >= 50%	**S**
else if herbaceous cover >= 50%	**H**
else if aquatic cover >= 50%	**A**
else if barren cover >= 50%	**B**
else if water cover >= 50%	**W**
else if unknown/cloud/shadow cover >= 50%	**U**
else if tree cover > (shrub, herbaceous, barren, aquatic, and other cover)	**t**
else if shrub cover > (herbaceous, barren, aquatic, and other cover)	**s**
else if herbaceous cover > (barren, aquatic, and other cover)	**h**
else if aquatic cover > (barren and other cover)	**a**
else if barren cover > other cover	**b**
else	**u**