Mapping Willow Distribution Across the Northern Range of Yellowstone National Park

---

2004 Final Report

Jeremy P. Shive and Robert L. Crabtree
Yellowstone Ecological Research Center

April 1, 2004

Table of Contents

1.1. Introduction

Based on evidence, such as repeat photographic surveys, willow (*Salix spp.*) communities have shown declines across Yellowstone National Park's Northern Range (NR) over the past 80 years. Although the observed declines are fairly widespread, there have been local increases in willow biomass since 1996 (Crabtree 2003, unpublished data). This time period is marked by climatic warming and drought conditions, significant flooding events, and possible shifts in ungulate herbivory. A number of studies are currently focused on identifying which of these factors or interactions between factors are responsible for the observed willow declines.

In order to understand further changes in willow and riparian shrub distribution, a comprehensive baseline inventory database must be produced to completely understand the current willow distribution and status. Over the past two decades numerous investigators have repeatedly identified the need for an accurate baseline willow database to facilitate future research and monitoring efforts.

1.2. Project Goals

This was the initial Phase 1 investigation of a multi-year project to create a potential and current willow distribution map across the NR incorporating multiple remotely sensed datasets. The goal was to test some preliminary willow classification methods using current remote sensing datasets and field training and validation data. The results of this initial investigation will form the basis for proposing the most appropriate methods of producing accurate willow classifications in Phase 2 of this project. The goals of the Phase 2 investigation are to 1) co-register all field plot data from previously conducted research throughout the NR; 2) create a current base map for future change

detection analysis; 3) provide a basis for developing a NR-wide sampling and long-term monitoring strategy.

2.1. Methods

2.1.1. Field Data Collection

A field crew spent 6 weeks during September and October of 2003 exploring floodplains and stream drainages throughout the NR visually searching for willow and deciduous riparian vegetation (Figure 1). The goal of the ground reconnaissance survey was to identify large (> 15 m x 15 m) homogenous patches of willow (i.e., larger than the spatial resolution of the imagery), and other riparian vegetation species. Willow tends to have a patchy and linear distribution along drainages, which makes finding homogenous locations larger in spatial extent than the resolution of the imagery very difficult. The largest near contiguous patches were selected and Global Positioning System (GPS) point coordinates were collected at the patch center and polygons were collected around the patch boundary (Table 1).

All field data coordinates were collected using Trimble GeoExplorer XT GPS units. The GPS data were collected in the WGS-84 Universal Transverse Mercator (UTM) Zone 12. All GPS data were post processed using Trimble Pathfinder Office software and differentially corrected using local base station correction files.

We recorded some common descriptive information describing each sampling location. The minimum and maximum vegetation height was measured and the average patch height was visually estimated. Dominant species density was visually estimated followed by a description of the fractional percent of each additional vegetation cover

type present. Digital photographs were also taken in the four cardinal directions at each sampling location to provide a visual interpretation of each site.

The intent of the ground surveys was to collect field data that can be used as training and validation data. Training data are used to build a statistical description of what surface features (e.g., willow) "look" like across the spectral bands collected. Numerous training pixels are chosen to create a Region of Interest (ROI) file for each feature class. ROI's are used as the input for various classification algorithms and are responsible for providing the statistical variation in each feature class that determines the final classification map. A subset of the ground survey data were selected and used solely for developing training ROI's, while the remaining data were used as a validation dataset. The validation dataset was used to determine the accuracy of the classification map by comparing predicted pixels against ground truth data that was not used to create the training ROI's.

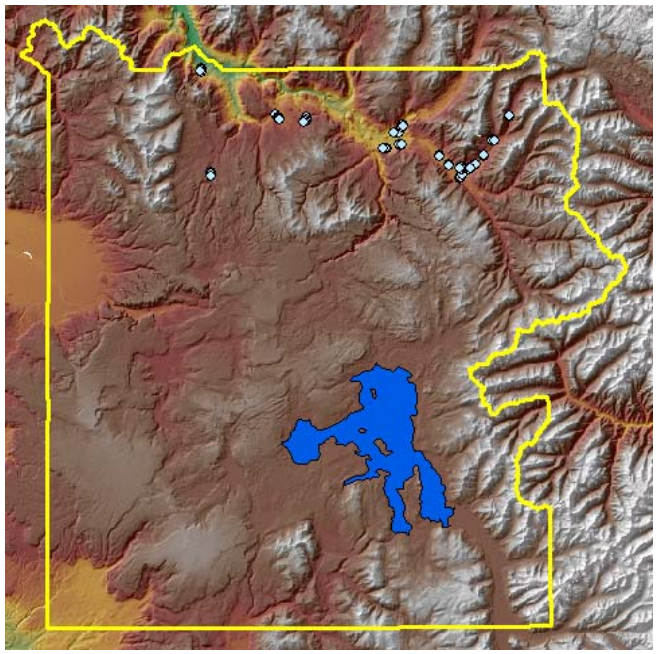


Figure 1. A display of all the riparian vegetation sampling locations (shown in light blue) across Yellowstone National Park's Northern Range.

Table 1. A listing of riparian vegetation classes and the total number of differentially corrected GPS points and polygons collected at each location.

Vegetation Feature	Total GPS Points	Total GPS Polygons
Aspen	30	30
Cottonwood	12	12
Willow	53	44

2.2. Remote Sensing Data

We incorporated numerous remotely sensed datasets to evaluate the feasibility of mapping willow habitat distributions. Both passive optical and active RADAR imagery were analyzed to assess the most appropriate method for developing willow or riparian habitat classification maps.

We used data collected by NASA's Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) sensor in June of 2001. ASTER collects 9 spectral bands dispersed between the visible and short-wave infrared (i.e., 520-2430 nm) wavelengths (Table 2). The spatial resolution, or pixel size, of ASTER data is 15 m for the first three bands and increases to 30 m for the remaining six spectral bands. For this project we resampled the 30 m bands to 15 m resolution to maintain consistent spatial resolution across all spectral bands, which is needed to run classification algorithms.

We also evaluated data collected by Intermap's Star3i Interferometric Synthetic Aperture Radar (IFSAR) across the entire NR of Yellowstone in 1999. The Star3i data produce a Digital Elevation Model (DEM) with a spatial resolution of 10 m and a vertical accuracy of about 1-2 m.

Table 2. Summary table of the remotely sensed datasets used in the preliminary analysis.

Sensor	Platform	Data Type	Spatial Resolution	# of Bands	Band Description
ASTER	Spaceborne	Multispectral	15 m	9	Band 1 = 520 - 600 nm
			15 m		Band 2 = 630 - 690 nm
			15 m		Band 3 = 760 - 860 nm
			30 m		Band 4 = 1600 - 1700 nm
			30 m		Band 5 = 2145 - 2185 nm
			30 m		Band 6 = 2185 - 2225 nm
			30 m		Band 7 = 2235 - 2285 nm
			30 m		Band 8 = 2295 - 2365 nm
			30 m		Band 9 = 2360 - 2430 nm
Star 3i	Airborne	DEM	10 m	1	X-band = 3 cm

2.3. Image Processing

Initially, we attempted to separate and distinguish a willow class explicitly. Willow is commonly associated with sedges and moist soils, and we found that the willow class actually classified "Mesic Meadow" and not willow explicitly due to a mixing of spectral reflectance within a single pixel. Other riparian vegetation, such as Cottonwood (*Populus spp.*) and Aspen (*Populus tremuloides*) were grouped into a "Deciduous" class. Following the initial classification attempts, we combined all riparian vegetation into a broad "Riparian" class. Previous field data collected by Yellowstone Ecological Research Center (YERC) was also used to create training and validation datasets of other common cover types found across the NR.

All image processing work was conducted using Research Systems Inc. ENVI 4.0 software package. A Minimum Noise Fraction (MNF) transformation (Green et al. 1988) was applied to the nine original ASTER spectral bands. The MNF transform can be thought of as a modified Principal Components transformation that attempts to minimize the noise and maximize the variance in spectral data. We selected the first six MNF transformed bands and omitted the last three bands, which were dominated by noise. The

first six MNF transformed bands served as the input data for all following image processing steps.

We experimented using traditional supervised classification algorithms, such as maximum likelihood and parallelpiped (Lillesand and Kiefer 1994). After numerous classification attempts, we decided the most promising classification algorithm was the Spectral Angle Mapper (SAM). The SAM classification algorithm determines the similarity between image spectra and training ROI spectra based on the angle between them calculated as a vector in n-dimensional space, where "n" equals the number of input bands (Kruse et al. 1993). Smaller angles represent better matches to ROI reference spectra. We experimented by altering the angular tolerance threshold until known riparian sites were being correctly classified while simultaneously minimizing the amount of observed overpredictions.

The Star3i DEM data were used to create a Bare-Earth Elevation Model (BEM) and secondarily to derive vegetation heights. A BEM is a surface elevation model that is calculated by identifying and separating the ground elevation points from the top of vegetation. After the ground elevation points are predicted a surface is interpolated between these points to create a new bare-earth surface elevation model. A BEM can then be subtracted from the original DEM to derive a residual vegetation heights layer. We used ESRI's ArcView 3.2 and the Rapid Terrain Visualization LiDAR Toolkit extension, developed through the Department of Defense, to calculate the BEM.

Vegetation heights can be used to estimate a finer discrimination of a broad classification class. For example, the multispectral ASTER data may only be able to differentiate a broad riparian class encompassing numerous species and community

types. We can divide the broad riparian class into smaller sub-classes based on estimated vegetation height, which provides the ability to map discrete feature classes not possible with a single remotely sensed dataset. We used this approach to identify and separate mature Cottonwood and Aspen trees from other riparian vegetation.

2.4. Accuracy Assessment

Error matrices serve as the basis for descriptive statistical techniques used to evaluate classification accuracy (Congalton and Green 1999) (Table 3). Producer's accuracy is calculated by dividing the total number of correct pixels in a category by the total number of pixels actually identified from ground truth reference data (Congalton and Green 1999). Producer's accuracy represents the probability a true positive location on the ground is correctly classified. User's accuracy is calculated by dividing the total number of correctly classified pixels by the total number of pixels classified in that category (Congalton and Green 1999). User's accuracy represents the probability that a classified image pixel is actually that category on the ground. Omission and commission errors are calculated by subtracting producer's and user's accuracy from 100% respectively. Overall accuracy is calculated by summing each class's true positive pixels, and dividing by the sum of all cells in the error matrix.

An additional accuracy statistic that is commonly reported is the Kappa coefficient, or KHAT. KHAT is a measure of the difference between the agreement of ground truth referenced data and an automated classification, and the chance agreement between the ground truth reference and a random classification (Lillesand and Kiefer1994). KHAT values usually range from 0 to 1, and this number can be interpreted

as a measure of how much better the observed classification is than random chance. KHAT can conceptually be described as (Lillesand and Kiefer 1994):

$$KHAT = \frac{\text{observed accuracy- chance agreement}}{\text{1- chance agreement}}$$

Table 3. A sample generic error matrix with commonly reported accuracy statistics described.

	Ground Truth	
	Present	**Not Present**
Predicted	A (True Positive)	B (False Positive)
Not Predicted	C (False Negative)	D (True Negative)
A/(A+C) = Producer's Accuracy		
A/(A+B) = User's Accuracy		
A+D/(A+B)+(C+D) = Overall Accuracy		
100% - Producer's Accuracy = Omission Error		
100% - User's Accuracy = Commission Error		

3.1. Results and Discussion

3.1.1. ASTER Image Classification

The first ASTER image classification using separate riparian classes exhibited poor accuracies for the "Deciduous" and "Mesic Meadow" classes (Figure 2). The overall accuracy of the ASTER classification with the riparian class split into "Deciduous" and "Mesic Meadow" classes was 60.9%, and the corresponding Kappa coefficient was 0.54. The "Deciduous" class had a producer's and user's accuracy of 6.14% and 41.67% respectively (Table 4). A producer's accuracy of 6.14% means that nearly 94% of known ground truth locations (i.e., omission error) of deciduous vegetation was not classified. A user's accuracy of 41.6% means that about 60% of the predicted deciduous locations are overpredicted misclassifications. The "Mesic Meadow" class had a calculated producer's and user's accuracy of 37.93% and 33.33% respectively.

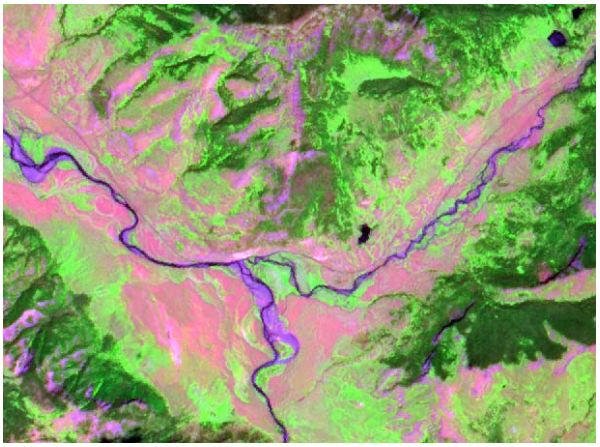
This feature class failed to identify nearly 62% off all known "Mesic Meadow" sites and falsely overpredicted about 64% of the mapped "Mesic Meadow" locations.

There are a number of potential reasons to explain the inaccuracies found with these classes. A "Deciduous" class will create classification problems due to the patchy distribution of this type of vegetation across the landscape. In many cases Aspen stands and Cottonwood trees do not occur in large homogenous patches, but rather in local patches commonly following linear drainages. Understory vegetation and exposed soil will commonly contribute to mixed pixels of deciduous vegetation. A mixed pixel is found when the spectral response (i.e., the measured reflectance in each spectral band) of any given image pixel is a mixed combination of surface feature reflectances (e.g., tree, understory grasses, and bare soil). The spectral response of a mixed pixel of deciduous vegetation can be very different than a "pure" pixel of 100% deciduous vegetation. The spectral variability introduced by a mixed pixel can commonly cause misclassifications similar to what we experienced in this study.

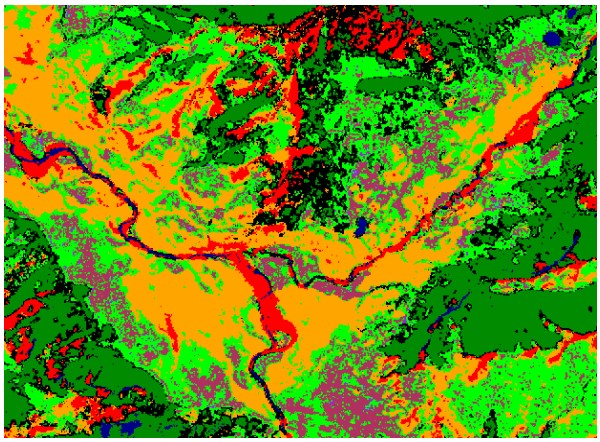
The "Mesic Meadow" class actually encompasses numerous vegetation types such as grasses, sedges, and willow. These different vegetation types are typically found distributed as mixtures and our results suggest that isolating and differentiating these vegetation types with ASTER data is not possible in this study area. The low accuracy statistics reported for the "Mesic Meadow" class are caused by a mismatch between image classification categories and ground truth categories. The accuracy statistics were calculated based on ground truth data that identified species and not species associations or communities. So even though the "Mesic Meadow" may have correctly identified a

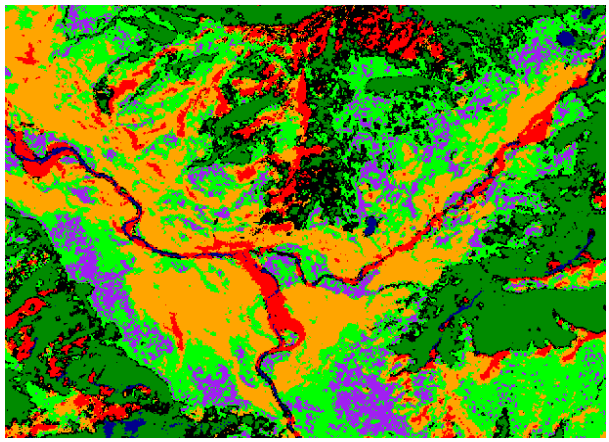
moist region where willow and grasses are present, the ground truth datasets were not collected to validate species mixtures and this location would be reported as an error.

a)



b)





c)

Figure 2. a) A false-color (RGB = Bands 7, 3, 2) ASTER image subset around the Soda Butte Creek and Lamar River confluence. b) The first image classification result of the ASTER subset (black = unclassified, red = rock/exposed soil, blue = water/shadow, dark green = conifer forest, purple = deciduous, orange = sagebrush, light green = grasslands, maroon = mesic meadow). c) The second image classification result for a single "Riparian" class (black = unclassified, red = rock/exposed soil, blue = water/shadow, dark green = conifer forest, orange = sagebrush, light green = grasslands, purple = riparian)

The second ASTER image classification with a single broad "Riparian" class performed slightly better, but accuracies still fell below an acceptable level to be considered a successful application. The overall accuracy was consistent with the first classification approach at 60.9%, and the corresponding Kappa coefficient was 0.53. The "Riparian" class had a producer's and user's accuracy of 32.11% and 66.3% respectively

(Table 5). Even with a single broad encompassing "Riparian" class, roughly 68% of the known riparian sites were missed. The combined class exhibited a lower commission error rate overpredicting about 34% of the classified "Riparian" pixels.

Similar to the first classification, the reported classification errors can be attributed to the mixed pixel problem. There is such a large amount of spectral variation among riparian vegetation it is difficult to produce a classification map that accurately spans the variation present in the imagery without misclassifying other surface features. Willow distributions typically follow streams or moisture influenced soils creating a linear pattern. Remote sensing imagery samples the environment with square pixels that are not ideal for linearly distributed features such as willow (Figure 3). A dataset with an increased spatial resolution may be able to isolate individual "pure" willow pixels and produce a more accurate willow classification map

Table 4. The first ASTER classification accuracy assessment summary with riparian vegetation divided into "Deciduous" and "Mesic Meadow" classes.

	Producer's Accuracy (%)	User's Accuracy (%)	Omission Error (%)	Commission Error (%)
Rock/Exposed Soil	89.23	93.55	10.77	6.45
Water/Shadow	83.78	98.41	16.22	1.59
Conifer	38.33	95.83	61.67	4.17
Deciduous	6.14	41.67	93.86	58.33
Sagebrush	96.18	82.35	3.82	17.65
Grassland	43.84	25.81	56.16	74.19
Mesic Meadow	37.93	33.33	62.07	66.67

Table 5. The second ASTER classification accuracy assessment summary with a single broad "Riparian" class.

	Producer's Accuracy (%)	User's Accuracy (%)	Omission Error (%)	Commission Error (%)
Rock/Exposed Soil	89.23	93.55	10.77	6.45
Water/Shadow	83.78	98.41	16.22	1.59
Conifer	38.33	95.83	61.67	4.17
Sagebrush	96.18	76.83	3.82	23.17
Grassland	42.47	21.09	57.53	78.91
Riparian	32.11	66.3	67.89	33.7

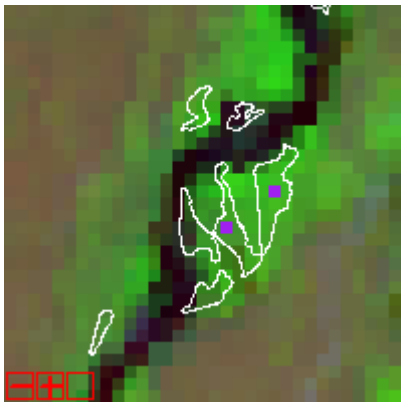


Figure 3. The white polygons were collected with GPS in the field and depict the boundaries of some willow patches overlaid onto a false-color ASTER image. The two purple pixel selected were used as willow training locations. The outline of the polygons depict the difficulty in identifying "pure" willow pixels.

3.2. Vegetation Heights

The preliminary results of predicting vegetation heights suggest that this approach may be a useful for dividing a general "Riparian" class into finer sub-classes differentiated by vegetation height. We were not able to confidently predict any vegetation heights less than 1.5 m because the spatial resolution of the imagery precludes estimates of subtle height differences. Taller vegetation height predictions followed general patterns of known vegetation in the field. There was a lot of error, or noise, observed in the 1.5 - 4 m height category likely introduced from minor topographic

variations within the floodplain (Figure 4). The > 4 m height class produced results that corresponded well with known ground vegetation distributions (Figure 4). The > 4m height class is useful for distinguishing mature Aspen and Cottonwood trees from other riparian vegetation. Since the ASTER data alone are not capable of identifying deciduous trees accurately, the combined information from two datasets increases the likelihood of accurately mapping and identifying riparian vegetation types.

Deriving relative vegetation heights in floodplain environments seems feasible, and the lack of topographic relief across floodplains provides ideal conditions for evaluating this process. In order to derive vegetation heights, an accurate BEM must be interpolated. The BEM processing algorithm assumes the lowest elevation within a moving window to be the ground elevation. Given the 10 m spatial resolution of the imagery, there can be substantial topographic and vegetation height variation within the extent of a single pixel. This variation makes it difficult to identify what is "true" ground elevations and what is a mix of topography and vegetation. The results show that predicting vegetation heights in uplands, where topography is complex, with 10 m resolution is not effective. Upland habitats commonly have more drastic changes in topography within small distances. This makes the process of identifying ground elevations difficult, and consequently the tops of ridges and peaks are "smoothed" creating error in height predictions caused by confusion with topographic variation.

4.1. Different Sensor Potential

An additional goal of this project was to create a road map of other potential remotely sensed data sources capable of improving willow habitat classification. We

evaluated a number of different datasets that covered the Soda Butte Creek and Lamar River confluence.

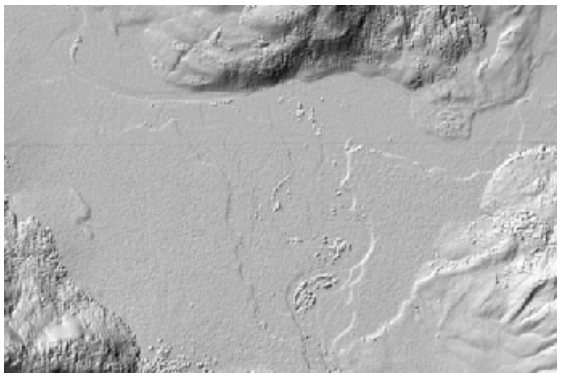
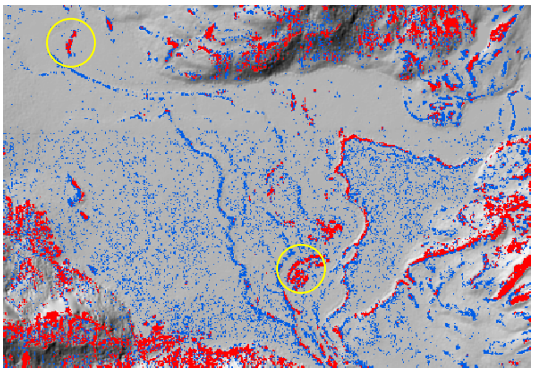
a)  b) 

Figure 4. a) Star3i IFSAR hillshade DEM at the Soda Butte Creek and Lamar River confluence. b) The derived vegetation heights classification (blue = 1.5-4 m, red = >4 m) overlaid onto the original DEM. The yellow circles identify example regions of known Cottonwood trees greater than 4 m in height.

Previous YERC research has focused on the application of hyperspectral data to map riparian vegetation. Given the increased spectral resolution and number of spectral bands hyperspectral data provide, subtle spectral differences can be differentiated. Since many riparian species are spectrally similar, remotely sensed data with a larger number of spectral bands would provide the best opportunity to distinguish riparian vegetation types (Figure 5). In addition too more spectral bands, an increased spatial resolution will reduce the probability of mixed pixels. With smaller pixel sizes, it is easier to identify training and validation sites that are "pure" willow and not mixtures of willow, grass, and soil for example. The combination of high spectral and spatial resolution data provides the greatest opportunity to exploit the advantages of these data types to potentially produce willow classification maps approaching the accuracies accomplished through ground surveys.

Another dataset that we will continue to evaluate is NASA's AirSAR sensor. The AirSAR sensor is an airborne RADAR that records three different wavelengths (i.e., C-band, L-band, and P-band) and all wave polarizations (i.e., HH, HV, VV). RADAR data respond to different physical properties of surface features than optical data, which provides useful complementary information. RADAR wavelengths respond to surface roughness, structure, and the dielectric constant of surface features. Water exhibits a high dielectric constant, which means regions of the environment that hold a lot of moisture (e.g., inundated soils in riparian habitat) will show a high backscatter response. This enables us to produce classification maps predicting areas where soil moisture levels are higher than background moisture levels. Willow and other riparian vegetation distribution patterns are directly correlated with soil moisture levels, and this type of classification can be used to predict potential or current willow habitat distributions (Figure 6).

Figure 5. A Probe 1 hyperspectral image (128 spectral bands and 1 m spatial resolution) classification showing the potential for riparian species discrimination.

We also evaluated additional high resolution DEM datasets to understand how increased spatial resolution contributes to predicting vegetation heights. The two

additional DEM datasets were collected in July of 2003 by the Department of the Army's Joint Precision Strike Demonstration Project's Rapid Terrain Visualization (RTV) program. The first dataset was collected by RTV's airborne Ku-band IFSAR sensor and produced a 3 m resolution DEM. The second dataset was collected by RTV's airborne LiDAR sensor and produced a first and last return 1 m resolution DEM.

Figure 6. An AirSAR image draped over a 30 m USGS DEM near the Soda Butte Creek and Lamar River confluence. The red-orange colors show the distribution of moisture influenced soils and represent a potential willow distribution map.

We used the same methodology described for the Star3i DEM processing. Higher spatial resolution reduced the amount of observed error in the vegetation height classifications (Figure 7). With smaller pixel sizes, we are able to create more accurate BEM's where there is less confusion between topographic and vegetation heights variations. A more accurate BEM allows for a more accurate estimate of relative

vegetation height categories, and at the 1m resolution provided by the LiDAR data we can actually distinguish individual trees (i.e., as opposed to stands or mixed patches) and corresponding heights across the floodplain.

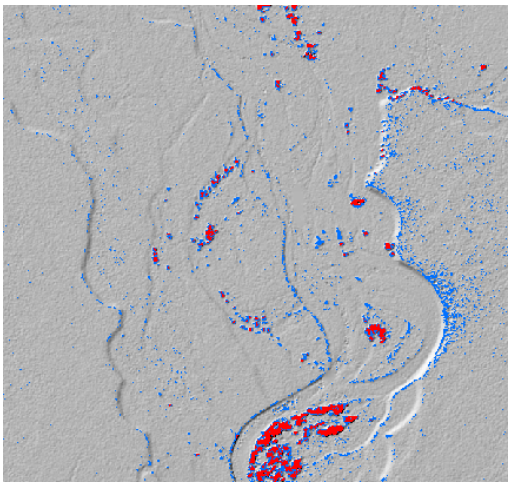
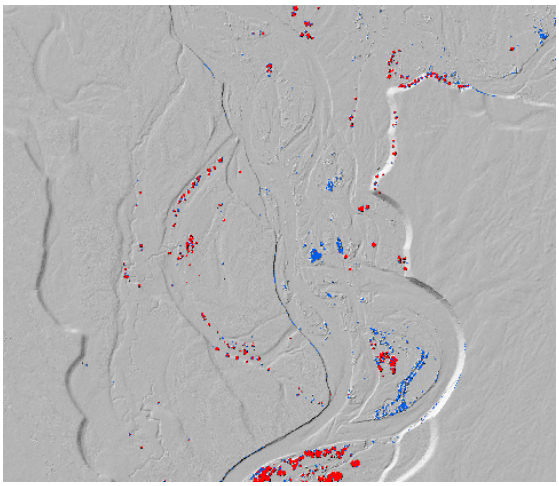
a)  b) 

Figure 7. a) RTV IFSAR (3 m) predicted vegetation heights near the Soda Butte Creek and Lamar River confluence (blue = 1-4 m, red = >4 m). b) RTV LiDAR (1 m) predicted vegetation heights near the Soda Butte Creek and Lamar River confluence (blue = 1-4 m, red = >4 m).

4.2. Recommendations for Future Studies

Topographic relief introduces shadows and large changes in surface illumination across large landscapes. Changes in brightness cause the spectral response of surface features to vary. A successful image classification requires the input training ROI's to encompass the range of class spectral variability across the image extent (Lillesand and Kiefer 1994). DEM's (e.g., Star3i or Shuttle Radar Topography Mission) allow a researcher to create a stratified sampling scheme by calculating aspect and slope raster images, and using these layers to select sampling sites across representative topographic categories (e.g., south-facing aspects with 0-20° slope or northeast-facing aspects with 30-40 slope°, etc.). Collecting data across different topographic categories will yield greater classification accuracies and the focus of future remote sensing studies should

emphasize the importance of how training and validation data are collected in the field. It is also very important that field-based ground truth data collections focus on collecting training and validation data that match the final classes used to produce the classification map (i.e., if the ground truth classes and classification classes are different, it is difficult to confidently assess classification class accuracies).

ASTER multispectral data alone are not capable of producing accurate and widespread willow distribution maps. The limited number of spectral bands makes it extremely difficult to distinguish spectrally similar features such as riparian vegetation. The large spatial resolution of ASTER data creates mixed pixels of riparian vegetation, and it is also very difficult to identify homogenous "pure" pixels required for effective classification applications. Future willow mapping research would benefit from higher spatial resolution data and an increased number of spectral bands.

The process of calculating vegetation heights holds much promise for future studies of riparian vegetation. Given the limitations of multispectral data mentioned above, any additional information that can be derived from different datasets will provide a more complete and accurate classification. There are plans to launch satellite IFSAR sensors in the near future capable of providing 1-2.5 m DEM's. At this resolution, there is a greater probability of predicting accurate vegetation heights that can contribute to the production of widespread willow habitat maps, and also serve as means to investigate changes through a monitoring program.

There are a number of additional datasets, such as AirSAR, that have a lot of potential for increasing the accuracies of willow habitat mapping. Our conclusions suggest that a single sensor will have problems creating accurate willow distribution

maps, and multi-sensor data fusion will ultimately produce the most accurate and detailed willow habitat maps.

5.1. Literature Cited

Congalton, R.G., and K. Green. 1999. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. CRC Press Inc., Boca Raton, Florida.

Green, A.A., M. Berman, P. Switzer and M.D. Craig. 1988. A transform for ordering multispectral data in terms of image quality with implications for noise removal. IEEE Transactions on Geoscience and Remote Sensing 26(1): 65-74.

Kruse, F.A., A.B. Lefkoff, J.B. Boardman, K.B. Heidebrecht, A.T. Shapiro, P.J. Barloon, and A.F.H. Goetz. 1993. The spectral image processing system (SIPS)-Interactive visualization and analysis of imaging spectrometer data. Remote Sensing of Environment 44: 145-163.

Lillesand, T.M., and R.W. Kiefer. 1994. *Remote Sensing and Image Interpretation*, 3rd Edition. Wiley, New York.