*WETLANDS,* Vol. 26, No. 2, June 2006, pp. 465–474


# MAPPING WETLANDS AND RIPARIAN AREAS USING LANDSAT ETM+ IMAGERY AND DECISION-TREE-BASED MODELS

Corey Baker, Rick Lawrence, Clifford Montagne, and Duncan Patten
*Department of Land Resources and Environmental Sciences*
*Montana State University*
*Bozeman, Montana, USA 59717-3490*

*Abstract:* The location and distribution of wetlands and riparian zones influence the ecological functions present on a landscape. Accurate and easily reproducible land-cover maps enable monitoring of land-management decisions and ultimately a greater understanding of landscape ecology. Multi-season Landsat ETM+ imagery from 2001 combined with ancillary topographic and soils data were used to map wetland and riparian systems in the Gallatin Valley of Southwest Montana, USA. Classification Tree Analysis (CTA) and Stochastic Gradient Boosting (SGB) decision-tree-based classification algorithms were used to distinguish wetlands and riparian areas from the rest of the landscape. CTA creates a single classification tree using a one-step-look-ahead procedure to reduce variance. SGB uses classification errors to refine tree development and incorporates multiple tree results into a single best classification. The SGB classification (86.0% overall accuracy) was more effective than CTA (73.1% overall accuracy) at detecting a variety of wetlands and riparian zones present on this landscape.

*Key Words:* wetland mapping, riparian zones, Landsat, decision tree classification, stochastic gradient boosting, classification tree analysis

## INTRODUCTION

Wetland and riparian zones provide a variety of ecological services that contribute to ecosystem functions at local, watershed, and regional scales (Semilitsch and Bodie 1998, Tabacchi et al. 1998, Ehrenfeld 2000, Mitsch and Gosselink 2000). Wetlands can effectively minimize sediment loss, control runoff volume, purify surface water, and enhance aquifer recharge (Ehrenfeld 2000, Tiner 2003). The shape, size, and distribution of wetland and riparian zones are largely determined by geologic, topographic, and hydrologic conditions (Peck and Lovvorn 2001, Toyra et al. 2002). The ecological contributions of wetlands and riparian zones, if factored into land values, suggest that these ecosystems are more economically and ecologically valuable than most other land cover types (Mitsch and Gosselink 2000).

Wetlands are ''[areas] that under normal circumstances do support. . . a prevalence of vegetation typically adapted for life in saturated soil conditions'' (U.S. EPA 2003. p.1) while riparian areas are ''ecosystems [that] occupy the transitional areas between the terrestrial and aquatic ecosystems'' (Montgomery 1996. p.2). Several fundamental ecological differences exist between wetlands and riparian zones; however, the ecological importance and human interaction between these ecosystems are very similar. These common characteristics enable synonymous discussion for purposes of landscape resource mapping. The term wetland, therefore, will be used to describe both wetland and riparian areas unless specified.

Accurate wetland mapping is an important tool for understanding wetland function and monitoring wetland response to natural and anthropogenic actions. Wetlands are often damaged or overwhelmed by increased surface flows in urban or suburban areas with high densities of impervious surfaces (i.e., buildings and paved surfaces) (Ehrenfeld 2000, Mitsch and Gosselink 2000, Wang et al. 2001). Wetland mapping is used to evaluate land-use decisions and monitor the effectiveness of mitigation efforts (Muller et al. 1993). Landscape scale mapping of these scarce habitats facilitates understanding of floral and faunal population dynamics (Semilitsch and Bodie 1998).

The susceptibility of wetlands to human activities and human dependence on the ecological contributions of wetlands illustrate the importance of mapping wetland resources. Establishing the role of wetlands in increasingly urban landscapes requires an understanding of wetland density and distribution (Tiner 2003). The three primary inventory techniques currently used to map wetland ecosystems are on-site evaluations, aerial photo interpretation, and digital image processing. Wetland mapping projects using on-site measurements of environmental conditions provide highly detailed data including lists of floral and faunal species, water chemistry, and soil characterization information (Tiner

1993). The added expense of personnel, equipment, and time rarely justifies the more detailed level of data collected through on-site evaluations when mapping wetlands at a landscape or watershed scale (Harvey and Hill 2001).

Aerial photographs provide synoptic views of study areas, allowing ''big picture'' understanding of hydrology and vegetation patterns (Harvey and Hill 2001). Additionally, aerial photograph archives are available for many regions of the United States, providing a valuable historical record of past landscape conditions. Many concerns are still associated with the use of aerial photos for wetland mapping, despite improvements in the quality of aerial photos. A primary concern with landscape-scale wetland maps derived from aerial photos is the extensive time lapse between imagery acquisition and production of the final wetland map (Ramsey and Laine 1997). Repeatability is another concern with human-derived photo-interpretation products. As concern over global wetland resources continues to escalate, so does the need for automated and reproducible wetland maps (Finlayson and van der Valk 1995). Using quantitatively derived wetland inventory maps in change detection analyses reduces inconsistencies associated with human interpretation and thus improves the power to identify actual wetland changes.

Multispectral sensors provide data with increased spectral and radiometric resolutions and decreased spatial resolutions compared to conventional aerial photography. Systeme Pour l'observation de la Terre (SPOT) and Landsat are two satellites with sensors that have been used to produce accurate maps of a variety of wetland types in Australia, Canada, and the United States (Sader et al. 1995, Narumalani et al. 1997, Kindscher et al. 1998, Harvey and Hill 2001, Townsend and Walsh 2001, Toyra et al. 2002). Data from the Indian Remote Sensing Satellite–Linear Imaging Self Scanning II (IRS–LISS-II) multispectral sensor were used to map wetland meadows in Grand Teton National Park, Wyoming, USA. The lack of middle infrared (MIR) detection on the IRS instrument inhibited the detection of vegetation and soil moisture, which are distinctive features of wetland areas (Johnston and Barson 1993, Mahlke 1996).

Several wetland-mapping studies suggest that Landsat-based classifications provide greater overall accuracies than other space-borne sensors (Civco 1989, Hewitt 1990, Bolstad and Lillesand 1992a). A test of this theory found that Landsat Thematic Mapper (TM) based classifications provided wetland maps with 82% accuracy for forested wetlands in Maine, USA (Sader et al. 1995). A similar overall accuracy (80%) was achieved when mapping riparian zones in xeric ecosystems of Eastern Washington, USA with Landsat-TM data (Hewitt 1990). Wetland classifications using aerial photos (1-m resolution), SPOT (20-m resolution), and Landsat (30-m resolution) image data were compared to determine the accuracy and applicability of each data source (Harvey and Hill 2001) and found that the sensitivity of Landsat band-2 (green), band-3 (red), band-4 (near infrared, NIR), and band-5 (MIR) provided a more accurate classification than SPOT, and overall accuracy comparable to that of aerial photos. These results demonstrate that accuracy is not sacrificed with automated wetland identification methods or with coarser spatial data for landscape-scale analyses.

The combination of readily interpretable classification results and accurate class separations has contributed to the increasing popularity of rule-based and decision tree methods for classification of multispectral data (Bolstad and Lillesand 1992b, Sader et al. 1995, Lawrence and Wright 2001). Interpretation using classification rules enables the image analyst to identify inconsistencies in the data and validate true ecological variation existing on the landscape. A supervised rule-based classification method produced an overall accuracy of 80% in wetland specific classifications of forested wetlands in Maine, an 8% improvement over the statistical clustering functions of unsupervised classifications (Sader et al. 1995). The classification rules used by Sader et al. were developed using ancillary topography, geology, and hydrology Geographic Information System (GIS) data sources to model forested wetland characteristics.

Classification tree analysis (CTA) is a rule-based technique that has produced highly accurate classifications based on a variety of spectral and ancillary data sources (Lawrence et al. 2004). Similar to neural networks, CTA is a non-parametric technique that does not assume normal distributions in the available datasets. CTA forms dichotomous decision trees using continuous or categorical data (Lawrence et al. 2004). The CTA algorithm works to reduce both intra-class and inter-class variability through recursive binary splitting of training data values (Venables and Ripley 1997). The results of such binary splits are displayed as branching dichotomous trees that serve as readily interpretable illustrations of variability within the data. Splits are applied to the classification of an image through classification rules (Lawrence and Wright 2001). Combinations of multispectral and ancillary data have been used in decision trees to produce highly accurate land-cover classifications. Decision trees are easily interpreted and can provide valuable insight into ecological conditions.

Recent refinements of CTA approaches can result in more accurate classifications, albeit easily interpretable classification rules are often sacrificed when using more complicated refinements. Since CTA trees are formed using a one-step-look-ahead, initial splits to reduce the greatest variability largely determine the effectiveness of the tree to distinguish more detailed separations further down the tree (Venables and Ripley 1997, Lawrence et al. 2004). Less effective splitting occurs when outliers are present in the data or when attempting to classify land cover containing high within-class variability. Additionally, if the class of interest represents a small portion of the landscape and the training data are collected in similar proportions, the less dominant land-cover types can be under-classified with CTA (Lawrence et al. 2004). These issues are applicable to wetland classification within a large landscape and thus encouraged a closer examination as part of our analysis.

Bagging, which uses random subsets of the data to develop decision trees, and boosting, which uses errors in trees to refine new trees, both use iterative tree development to address some of the previously mentioned shortcomings inherent in the one-step at a time CTA algorithm (Lawrence et al. 2004). Stochastic gradient boosting (SGB) has the potential to provide improved classification accuracies over CTA by combining the beneficial aspects of bagging and boosting techniques (for comprehensive discussion, see Lawrence et al. 2004). Using a steepest gradient boosting algorithm, the most readily corrected classification problems are emphasized in iterations of tree development and the resulting collection of trees (a grove) vote on the correct classification using a plurality rule (Lawrence et al. 2004). Bagging and boosting procedures develop large numbers of trees with minimal user interaction to provide accurate and reproducible results. Broad applicability of SGB for purposes of land-cover classification has yet to be tested due to the recent development of this technique and limited software distribution, although lately this and related techniques have become more readily available, notably through contributions to the free R statistical program. This technique has the potential to identify distinctive characteristics of small and highly diverse ecosystems, such as wetlands, from spectral and topographic data.

Our objective was to develop an accurate and easily reproducible procedure for mapping wetlands across natural and human dominated landscapes. Ancillary environmental data were incorporated into spectrally based classifications to improve the detection of isolated or ecologically unique wetlands (Sader et al. 1995). The applicability and accuracy of two decision tree algorithms, CTA and SGB, were compared to determine the efficacy of both techniques for wetlands mapping. Additionally, CTA and SGB were compared on urban and rural subsets of the study area to determine specific strengths and weaknesses of each classification on different landscapes. The ultimate goal of these analyses was to help identify a rapid, accurate, and reproducible technique for mapping wetland and riparian zones in landscape-scale analyses. The recent introduction of bagging and boosting software for decision tree classifications (e.g., TreeNet and See5) and highly favorable results in studies using these methods encourages land-cover classifications based on these statistical algorithms. High diversity and inter-class variability makes wetlands a difficult land-cover type to classify accurately, therefore making wetlands excellent testing sites for these classifications.

## METHODS

### Study Area

The 135,570-ha study site was the lower basin of the Gallatin River watershed, located in the Gallatin Valley of Southwestern Montana, USA (Figure 1). The project area boundary generally follows the boundary of the Gallatin Local Water Quality District. The foothills and mountainous terrain of the Bridger, Gallatin, and Madison ranges surround the plains of the Gallatin Valley. The Gallatin and East Gallatin rivers have formed the majority of landscape features on the valley floor (Willard 1935). A semi-arid climate and fertile soils support the prevalence of irrigated and dryland agriculture in the valley. Primary crops of the region are alfalfa, barley, wheat, and hay for livestock. Population growth over the past 50 years has resulted in localized conversions of agricultural land to residential and commercial development (Kendy 2001).

Precipitation averages range from 40 cm in the valley (1,250 m) to over 100 cm in the higher elevations (3,350 m) (Custer et al. 1996, Western Regional Climate Center 2002). Snow and rain from March through June provide the majority of precipitation. Surface and subsurface flow regimes have been altered through the widespread construction of irrigation canals. Canals reduce in-stream flows and distribute water throughout the interior and periphery of the valley. The perennial streams contain much herbaceous and woody vegetation, including chokecherry (*Prunus virginiana* (Nutt) Torr.), willow (*Salix* spp.), black cottonwood (*Populus trichocarpa* Torr. and Gray), narrowleaf cottonwood (*P. augustifolia* James), quaking aspen (*Populus tremuloides* Michx.), and several other native and non-native species. Vegetation strips along the ephemeral natural streams and artificial canals are narrower, with less vegetation density and species diversity than perennial systems.

### Image Processing

Landsat Enhanced Thematic Mapper Plus (ETM+) images from May 22, 2001 and September 11, 2001

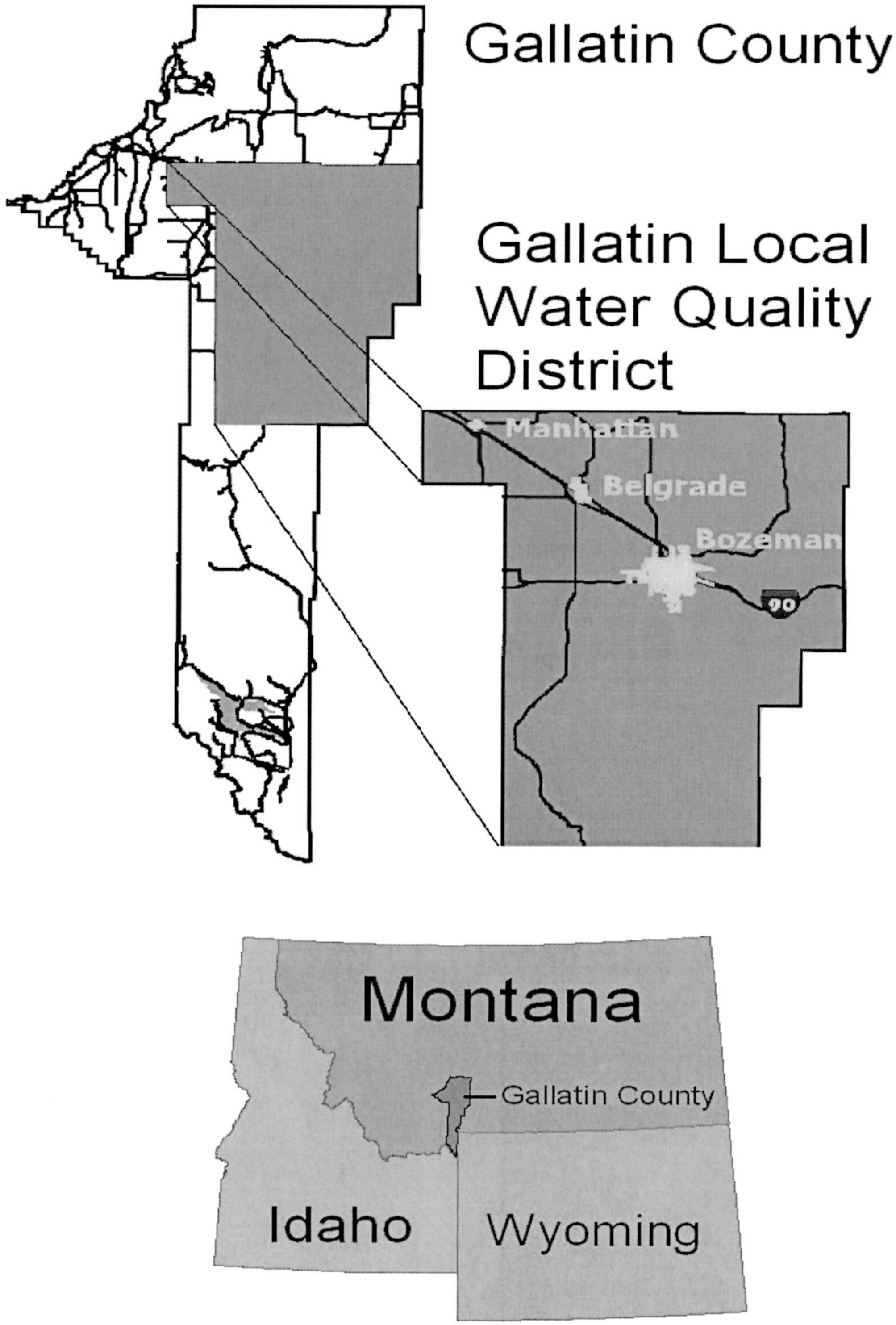


Figure 1. Location map for the Gallatin Local Water Quality District.

were the spectral data sources used in the classification procedure. The Landsat ETM+ sensor records 7 bands of spectral data in the visible, infrared, and thermal portions of the electromagnetic spectrum. The spatial resolution of this sensor is 30 m (the 60-m thermal band-6 was resampled to 30 m using nearest neighbor interpolation), resulting in a 900 $m^2$ (0.09 ha) minimum mapping unit. Multi-date imagery was used to capture the extent of seasonal variation between wet (May) and dry (September) conditions. To help identify seasonal wetlands, the wet and dry images were merged into a single classification using known upland, riparian, and wetland areas as training sites. A total of 65,467 training pixels were used to classify the 1,507,429 pixels contained in the study area.

The May image was geo-registered to the September scene (registration error less than 6.0 m). Both scenes were corrected to at-sensor reflectance using the United States Geological Survey (USGS) equation (Huang et al. 2001) and ETM+ gain/bias header file data. Tasseled Cap (TC) transformations, which produce components representing brightness, greenness, and wetness, were performed using the at-sensor reflectance values and USGS TC coefficients (Huang et al. 2002). Ancillary data used in this project included a 30-m USGS digital elevation model (DEM), slope map (calculated from the 30-m DEM), and digital hydric soils data from the 1985 Natural Resource Conservation Service (NRCS) soil survey for Gallatin County. Classification training sites were developed for wetland, riparian, and other land cover using recently digitized wetland and riparian data acquired from 1:24,000 color infrared (CIR) aerial photography of the study area and on-site surveys.

### Image Classification

Seven land-cover types were identified in the primary classification procedure, including open water, forest, urban, agriculture, grass/shrub, riparian, and wetland. The first five cover classes were combined into a ''non-wet'' class that was used for the remainder of the analysis. The ''wetland'' class was primarily composed of marshes, wet meadows, and slope wetlands. The ''riparian'' class included riparian wetlands, ephemeral drainages, and woody riparian vegetation (i.e., cottonwood and willow).

CTA decision trees were created using a combination of S-Plus™ statistical software and ERDAS Imagine™ image processing software (ERDAS 2001, Insightful 2001). Overfitting of CTA decision trees was avoided through cross validation of the training data (Lawrence and Wright 2001). The SGB decision tree grove was created using the same training data sets as CTA and was developed with TreeNet™ software (Salford Systems 2001). The decision trees provided in the TreeNet™ grove file were then used to produce a classified map of the study area.

### Accuracy Assessment

Accuracy assessment points were randomly generated in a stratified random format to define approximately 100 points each for the wetland and riparian classes and 150 points for the more predominant non-wet class. On-site evaluations, CIR photographs taken September 9, 2001, and a 5-m digital image derived from the 2001 CIR photos were used as reference data for classification accuracy assessments. Land-cover class assignments for accuracy assessment pixels were determined using a modification the 50% vegetation rule (Tiner 1993). In this project at least 20% of a 30-m pixel had to contain hydrophytic vegetation in order to be classified as wetland or riparian.

A spatial analysis of classification sensitivity was performed to determine the accuracy of the two classification techniques on different landscapes. In this analysis, we examined mis-classified pixels to ascertain if errors of omission or commission prevailed with either classification technique on specific landscapes. The first subset was located in a primarily rural setting with abundant agricultural land, and the second subset included the urban/sub-urban regions surrounding the town of Bozeman. The rural landscape contained larger wetlands and riparian sites with greater diversity, while the urban subset comprised smaller and more distinct wetland types. Accuracy assessment of this sensitivity analysis also used a stratified random design to identify reference points for each of the three land-cover classes. A focused accuracy assessment of these distinct subsets exposed the strengths and weaknesses of each technique in regards to wetland detection in both heavily diversified and homogenous landscapes.

## RESULTS AND DISCUSSION

### Overall Classification Accuracies

Overall classification accuracy was 73.1% for CTA and 86.0% for SGB, a 12.9% improvement over CTA results (Table 1). Producer's accuracies for wetland and riparian classes in the SGB classification (93.2% and 88.3%, respectively) were markedly higher than CTA (58.3% and 57.5%, respectively). The producer's accuracy is a measurement of omission error and is calculated by determining the probability that a reference pixel for each class is correctly classified. The majority of the error in the CTA classification resulted from wetland and riparian areas that were mis-classified as non-wet. Conversely, the majority of error in the SGB classification resulted from non-wet areas mistakenly classified as wetland. Simply stated, the CTA tended to miss marginal wetland and riparian sites, while SGB errantly classified moist upland sites as wetland or riparian.

User's accuracy is used to measure commission errors and represents the mapping accuracy for each class. User's accuracy of SGB (94.5%) was 28.1% higher than CTA (66.4%) for the non-wet class. The tendency of CTA to underestimate wetland and riparian areas was the primary cause of the large difference. The user's accuracy values for the wetland and ripar-

Table 1. Error matrices using classified and reference data pixels for CTA and SGB classifications.

| | Reference Data | | | | |
|---|---|---|---|---|---|
| Classified Data | Non-wet | Wetland | Riparian | | Users Accuracy |
| **CTA classification** | | | | | |
| Non-wet | 142 | 38 | 34 | 142/214 | 66.40% |
| Wetland | 10 | 60 | 6 | 60/76 | 79.00% |
| Riparian | 1 | 5 | 54 | 54/60 | 90.00% |
| | 142/153 | 60/103 | 54/94 | | |
| Producers Accuracy | 92.80% | 58.30% | 57.50% | | |
| ***Overall Accuracy*** | ***73.10%*** | | Kappa = 0.569 | | |
| **SGB Classification** | | | | | |
| Non-wet | 122 | 3 | 4 | 122/129 | 94.50% |
| Wetland | 23 | 96 | 7 | 96/126 | 76.20% |
| Riparian | 8 | 4 | 83 | 83/95 | 87.40% |
| | 122/153 | 96/103 | 83/94 | | |
| Producers Accuracy | 79.70% | 93.20% | 88.30% | | |
| ***Overall Accuracy*** | ***86.00%*** | | Kappa = 0.788 | | |

ian classes were similar for the two classifications. The primary source of error in the wetland class for both classifications was the inclusion of non-wet sites into the wetland class. Commission errors in the riparian class were more evenly distributed, with approximately equal numbers of non-wet and wetland sites erroneously placed in this class.

A notably smaller percentage of classification errors resulted from confusion between riparian and wetland pixels. The presence of woody vegetation in riparian zones appeared to minimize confusion, despite the hydrologic similarities of these sites. The over-inclusion of wetlands in the non-wet class was primarily attributed to the prevalence of flood-irrigated fields with elevation, soils, and spectral values similar to those of wetlands. Differences in the vegetation patterns between these two land covers were visible in the CIR photographs, although this variability was not visually discernable in the coarser resolution Landsat images.

Both techniques classified some wet and/or heavily vegetated upland areas as wetlands, although the inclusion of marginal and severely impaired wetlands was intentional. Detection of wetland and riparian sites was a source of error in both classifications; however, the overall and class accuracies were lower with CTA. Recent investigations of CTA classifications indicate that high within-class variability might positively influence the performance of SGB classifications compared to CTA (Lawrence et al. 2004). This theory would apply to the diversity of wetland and riparian systems in the Gallatin Valley and might explain the markedly improved producer's accuracies of these classes with SGB. The SGB tree development method concentrates on correcting classification errors on the most similar data and separating more distinctive classes on subsequent iterations of tree development. In this manner, SGB can be more adept at separating spectrally similar classes (Lawrence et al. 2004).

The classified images created through CTA and SGB contain substantially different proportions of wetland and riparian areas (Figure 2). CTA classified 6.8% of the pixels as wetland and 2.3% as riparian. The SGB classification placed 13.1% of the pixels in the wetland class and 5.3% in the riparian. These percentages, however, cannot be used to estimate the total area occupied by wetlands and riparian areas because each pixel classified as wetland or riparian can be comprised of as little as 20% or as much as 100% wetland or riparian vegetation. The buffers surrounding most wetland and riparian zones were therefore notably larger than aerial photo based inventories. Our objective was to determine the accuracy of classification procedures designed to distinguish wetland and riparian areas from other land-cover types. It was advantageous, therefore, to locate all areas potentially containing wetlands or riparian areas rather than to neglect marginal or smaller hydrologic ecosystems. In this respect, isolated pixels classified as wetland can be interpreted as a 900m$^2$ site where 20% or more of the area had wetland characteristics. These classification parameters could be refined to detect specific wetland types by selecting training sites that have the wetland characteristics desired in a classification or change detection analysis.

## Classification Accuracy for Urban and Rural Subsets

Results of the sensitivity analysis for the rural subset had an overall accuracy of 90.0% for SGB and 66.0% for CTA (Table 2). The SGB method was more apt to

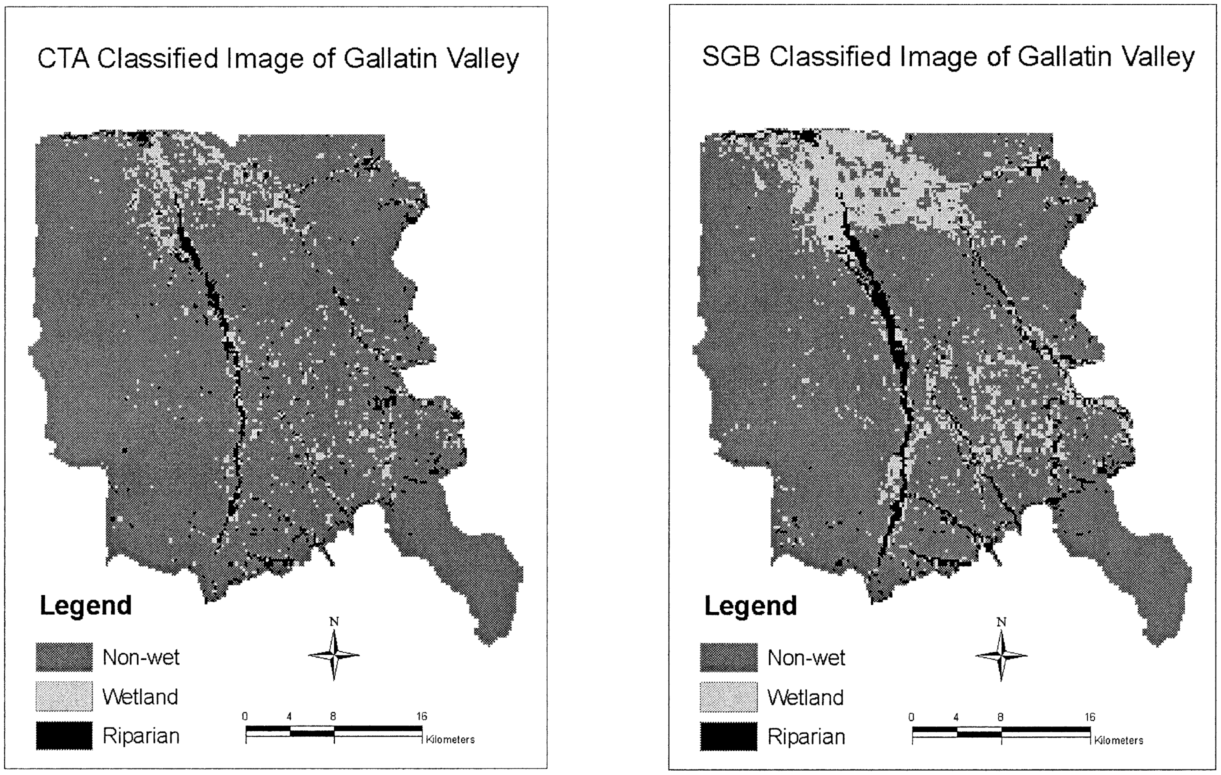


Figure 2. Classified images from CTA and SGB procedures.

include marginal wetlands and moist ecotones in the wetland class. Inclusion of marginal and degraded wetlands is advantageous when performing comprehensive wetland inventories that identify all possible wetland sites. SGB more successfully classified altered or impaired wetlands, such as cropped wetland sites that were partially converted to agriculture or heavily grazed.

The ability of SGB to detect isolated and drier-end wetlands also served as a source of error for irrigated pastures and cropland. CTA was less susceptible to the inclusion of wetlands in the non-wetland class but more likely to exclude drier wetland and riparian areas. Evidence of such predictable differences might allow analysts to select a classification technique based on the level of hydrologic sensitivity desired in the classification. It is possible that classification of broad and spectrally distinctive land-cover types might be more accurately performed with CTA, while detection of under-represented or highly variable land cover will re-

Table 2. Summary accuracy data for classification sensitivity analysis of urban and rural data subsets.

| **Rural Subset** | Users Accuracy | Producers Accuracy | **Urban Subset** | Users Accuracy | Producers Accuracy |
|---|---|---|---|---|---|
| **SGB** | | | **SGB** | | |
| Non-wet | 100.0% | 89.3% | Non-wet | 96.0% | 53.3% |
| Wetland | 86.0% | 86.0% | Wetland | 36.0% | 69.2% |
| Riparian | 84.0% | 95.5% | Riparian | 56.0% | 82.4% |
| Overall Accuracy | 90.0% | | Overall Accuracy | 62.7% | |
| Kappa | 0.850 | | Kappa | 0.440 | |
| **CTA** | | | **CTA** | | |
| Non-wet | 57.8% | 100.0% | Non-wet | 78.5% | 93.3% |
| Wetland | 81.8% | 36.0% | Wetland | 27.6% | 30.8% |
| Riparian | 80.7% | 56.8% | Riparian | 71.4% | 29.4% |
| Overall Accuracy | 66.0% | | Overall Accuracy | 68.0% | |
| Kappa | 0.476 | | Kappa | 0.381 | |

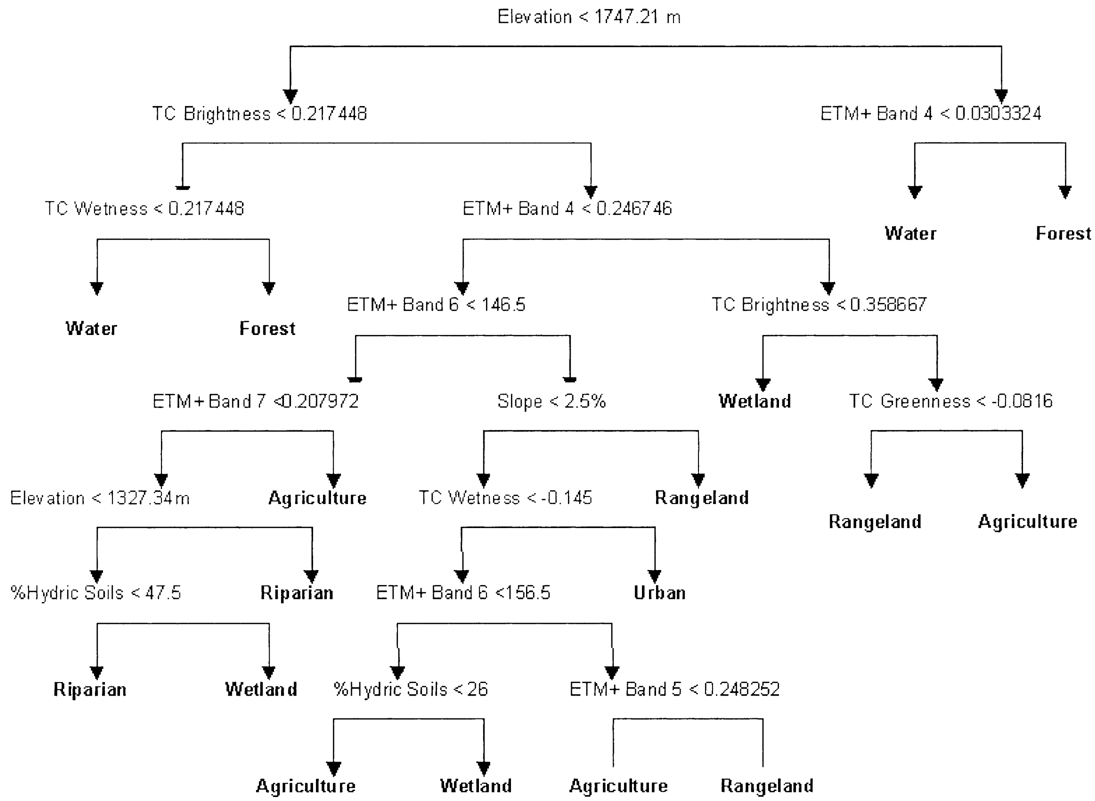


Figure 3. CTA decision tree for wetland, riparian, and non-wet classes (urban, agriculture, rangeland, forest, and water). Rules at each tree split indicate the conditions for the left branch at that split.

quire the increased sensitivity of SGB. Choosing between classification methods (such as CTA or SGB) or data sources (moderate spatial resolution or high spatial resolution) could enable stakeholders to select the level of classification detail.

Both classification techniques produced lower accuracies in the urban dominated landscape subset. While the increased sensitivity of SGB to wet conditions was advantageous for rural landscapes, this served as a source of error in the urbanized areas. Classification errors for SGB in the urban subset partially resulted from irrigated forests (e.g., city parks and cemeteries) erroneously classified as riparian areas and heavily irrigated pastures that were mistakenly classified as wetlands.

The accuracy of decision-tree-based classifications was potentially dependent on the inherent variability within the landscape, as demonstrated by the sensitivity analysis. The modest performance of CTA and SGB on the urban landscape subset was not necessarily indicative of limitations with either technique but, rather, a result of the inherent similarity of certain urban land uses to wetlands and, potentially, inadequate training for complicated urbanized wetland and riparian areas. Furthermore, the 30-m spatial resolution of ETM+ limited the detection of small, yet ecologically healthy, wetland and riparian systems present in the highly fragmented framework of urban and suburban areas. Higher spatial resolution data and a concerted effort to sample the variability of urban wetland and riparian sites could potentially improve identification of these areas in spectrally diverse landscapes.

## Evaluation of Variables Used

SGB developed 80 total decision trees, which was later reduced to 29 trees to avoid overfitting. Overfitting of the single CTA decision tree was avoided using cross validation to reduce the number of terminal nodes from 39 to 17 (Figure 3). SGB produces a large number of trees that can neither be displayed practically nor interpreted individually. SGB does, however, indicate the relative importance of variables within the model. Despite the distinctive statistical approaches of CTA and SGB, both algorithms relied on several common spectral and ancillary variables. These similarities are evident in the decision splits of the CTA tree and the variable importance table from the SGB output (Table 3). SGB used data from 19 of the 23 available variables while CTA used 18 out of the same 23.

Of the 23 total variables, elevation (DEM), hydric soils, NIR-Band 4 (September), TC-Brightness (Sep-

Table 3. Variables used for classification listed in order of importance from SGB output. The number of CTA decision nodes utilizing the same classification variables.

| Variable | SGB Rank | Variable | # of CTA Decision Nodes |
|---|---|---|---|
| Soils | 1 | Soils | 2 |
| Elevation (DEM) | 2 | Elevation (DEM) | 2 |
| TC Greenness | 3 | TC Greenness | 1 |
| TC Brightness | 4 | TC Brightness | 2 |
| ETM+ Band 4 | 5 | ETM+ Band 4 | 2 |
| ETM+ Band 3 | 6 | ETM+ Band 3 | 0 |
| ETM+ Band 6 | 7 | ETM+ Band 6 | 2 |
| ETM+ Band 1 | 8 | ETM+ Band 1 | 0 |
| ETM+ Band 7 | 9 | ETM+ Band 7 | 1 |
| ETM+ Band 2 | 10 | ETM+ Band 2 | 0 |

tember), TC-Wetness (September), and thermal-Band 6 (September) were used in the primary splits of the CTA tree and were among the top 10 most important variables listed for SGB. Topographic position and moisture-sensitive middle infrared response provided the greatest reductions in deviance on the CTA output. These responses can be interpreted as the most distinguishable characteristics between the riparian or wetland sites and the rest of the landscape. DEM data was most useful in separating the forests and lakes in the surrounding mountains from features on the valley bottom. Similarly, slope data were most evident in splits between sloping rangelands and the flatter agricultural or wetland features. Hydric soils data proved helpful in separating wetlands from irrigated agricultural land and riparian zones. These sites often contained similar vegetation types and surface moisture conditions, which enabled non-spectral variables, such as soils, greater power of separability.

Spectral data from the September image were more frequently used by both classification algorithms to separate landcover types than the May image. Moisture and vegetation vigor was sharply contrasting in the September image between moderately-to-extremely moist wetlands and the senescent upland vegetation. Such contrasts were not visible in the May image, where the majority of the landscape was irrigated by spring rains and snowmelt.

## CONCLUSIONS

The results of this study supported previous findings that applying SGB techniques to decision trees can improve classification accuracy (Lawrence et al. 2004). Using a combination of Landsat imagery and ancillary environmental data with an SGB classification algorithm was a highly effective technique for distinguishing a variety of wetland conditions from the surrounding landscape. Wetland and riparian areas were classified with minimal omission errors and an aptitude for detecting isolated and marginal wetland areas. Mapping this landscape with 86% accuracy provides a valuable resource inventory map of hydrologically dependent ecosystems. These results also demonstrate that boosted decision trees provide improved sensitivity to characteristics of marginal and damaged wetlands that are often missed in other wetland mapping procedures. Further investigation is necessary to determine the ability of SGB classifications for mapping specific wetland types, with the potential to use higher resolution sensors such as IKONOS or QuickBird. Wetland maps of this spatial resolution would enable calculations of wetland area in addition to rapid change-detection methods.

Some recently introduced boosting procedures are somewhat of a hybrid between the CTA and SGB algorithms and therefore might result in more balanced classifications. Investigating such balance might enable the development of one classification procedure that is equally accurate on rural and urban landscapes. See5 (which provides CTA with or without boosting) and R (which has packages available for CTA, a regression version of SGB, and some related techniques) are two such software packages that are much more affordable (R is available for free) than either S-Plus or TreeNet and therefore might warrant a thorough investigation for purposes of wetland detection. Future research in this area would include the use of higher resolution sensors, such as IKONOS or QuickBird, along with SGB algorithms to improve detection of small wetland sites and narrow riparian zones.

Wetlands and riparian areas are highly diverse ecosystems that have significant variability of physical properties. Our results provide further evidence that highly accurate detection of such diverse land-cover is feasible using automated classification procedures. Repeat temporal coverage, unbiased data collection, and effective sampling of landscape variability are advantages provided by remotely sensed data that enable systematic inventories of these ecosystems (Lakshmi et al. 1997). Combining automated classifications with recently acquired remote sensing data can quickly and accurately determine the location of small, isolated, and highly variable ecosystems, thus enabling the systematic monitoring of these important ecological resources.

## LITERATURE CITED

Bolstad, P. V. and T. M. Lillesand. 1992a. Improved classification of forest vegetation in northern Wisconsin through a rule-based

combination of soils, terrain, and Landsat Thematic Mapper data. Forest Science 38:5–20.
Bolstad, P. V. and T. M. Lillesand. 1992b. Rule-based classification models: flexible integration of satellite imagery and thematic spatial data. Photogrammetric Engineering and Remote Sensing 58: 965–971.
Civco, D. L. 1989. Knowledge-based land use and land cover mapping. p. 276–291. *In* Technical Papers, 1989 Annual Meeting of the American Society for Photogrammetry and Remote Sensing, Baltimore, MD, USA.
Custer, S. G., P. Farnes, J. P. Wilson, and R. D. Snyder. 1996. A Comparison of Hand- and Spline-Drawn Precipitation Maps for Mountainous Montana. Journal of the American Water Resources Association 32:393–405.
Ehrenfeld, J. G. 2000. Evaluating wetlands within an urban context. Ecological Engineering 15:253–265.
ERDAS. 2001. ERDAS Imagine® Configuration Guide. ERDAS Incorporated, Atlanta, GA, USA.
Finlayson, C. M. and A. G. van der Valk. 1995. Wetland classification and inventory: A summary. Vegetatio 118:185–192.
Harvey, K. R. and G. J. E. Hill. 2001. Vegetation mapping of a tropical freshwater swamp in the Northern Territory, Australia: a comparison of aerial photography, Landsat TM and SPOT satellite imagery. International Journal of Remote Sensing 22:2911–2925.
Hewitt, M. J. 1990. Synoptic inventory of riparian ecoystems: The utility of Landsat Thematic Mapper data. Forest Ecology and Management 33/34:605–620.
Huang, C., B. Wylie, L. Yang, C. Homer, and G. Zylstra. 2002. Derivation of a Tassled Cap transformation based on Landsat and at-satellite reflectance. International Journal of Remote Sensing 23:1741–1748.
Huang, C., L. Yang, C. Homer, B. Wylie, J. Vogelman, and T. DeFelice. 2001. At-satellite reflectance: a first order normalization of Landsat and ETM+ images. USGS White Papers, http://landcover.usgs.gov/pdf/huang2.pdf, last accessed February 13, 2006.
Insightful. 2001. S-Plus 6 User's Guide. Insightful Corporation, Seattle, WA, USA.
Jensen, J. R. 1996. Introductory Digital Image Processing, second edition. Prentice Hall, Upper Saddle River, NJ, USA.
Johnston, R. M. and M. M. Barson. 1993. Remote sensing of Australian wetlands: An evaluation of Landsat TM data for inventory and classification. Australian Journal of Marine and Freshwater Resources 44:235–252.
Kendy, E. 2001. Ground-water resources of the Gallatin Local Water Quality District, southwestern Montana. U.S. Geological Survey Fact Sheet 007–01.
Kindscher, K., A. Fraser, M. E. Jakubauskas, and D. M. Debinski. 1998. Identifying wetland meadows in Grand Teton National Park using remote sensing and average wetland values. Wetlands Ecology and Management 5:265–273.
Lakshmi, V., E. F. Wood, and B. J. Choudhury. 1997. Evaluation of Special Sensor Microwave/Imager satellite data for regional soil moisture estimation over the Red River Basin. Journal of Applied Meteorology 36:1309–1328.
Lawrence, R. L. and A. Wright. 2001. Rule-based classification systems using classification and regression tree (CART) analysis. Photogrammetric Engineering and Remote Sensing 67:1137–1142.
Lawrence, R., A. Bunn, S. Powell, and M. Zambon. 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. Remote Sensing of Environment 90:331–336.
Mahlke, J. 1996. Characterization of Oklahoma Reservoir wetlands for preliminary change detection mapping using IRS-1B Satellite imagery. IGARSS 1996: 1996 International Geoscience and Remote Sensing Symposium, 1769–1771.
Mitsch, W. J. and J. G. Gosselink. 2000. The value of wetlands: importance of scale and landscape setting. Ecological Economics 35:25–33.
Montgomery, G. R. 1996. RCA III Riparian areas: reservoirs of diversity. Working paper No. 13, http://www.nrcs.usda.gov/technical/land/pubs/wp13text.html, last accessed February 13, 2006.
Muller, E., H. Decamps, and K. D. Michael. 1993. Contribution of space remote sensing to river studies. Freshwater Biology 29:301–312.
Narumalani, S., Y. Zhou, and J. R. Jensen. 1997. Application of remote sensing and geographic information systems to the delineation and analysis of buffer zones. Aquatic Botany 58:393–409.
Peck, D. E. and J. R. Lovvorn. 2001. The importance of flood irrigation in water supply to wetlands in the Laramie Basin, Wyoming, USA. Wetlands 21:370–378.
Ramsey, E. W. and S. C. Laine. 1997. Comparison of Landsat Thematic Mapper and high resolution aerial photography to identify change in complex coastal wetlands. Journal of Coastal Research 13:281–292.
Sader, S. A., D. Ahl, and W. S. Liou. 1995. Accuracy of Landsat-TM and GIS rule-based methods for forest wetland classification in Maine. Remote Sensing of Environment 53:133–144.
Salford Systems. 2001. TreeNet stochastic gradient boosting: An implementation of the MART methodology. Salford Systems, San Diego, CA, USA.
Semilitsch, R. D. and R. Bodie. 1998. Are small, isolated wetlands expendable? Conservation Biology 12:1129–1133.
Tabacchi, E., D. L. Correll, R. Hauer, G. Pinay, A. Planty-Tabacchi, and R. C. Wissmar. 1998. Development, maintenance and role of riparian vegetation in the river landscape. Freshwater Biology 40: 497–516.
Tiner, R. W. 1993. Using plants as indicators of wetlands. Proceedings of the Academy of Natural Sciences of Philadelphia 144:240–253.
Tiner, R. W. 2003. Geographically isolated wetlands of the United States. Wetlands 23:494–516.
Townsend, P. A. and S. J. Walsh. 2001. Remote sensing of forested wetlands: application of multitemporal and multispectral satellite imagery to determine plant community composition and structure in southeastern USA. Plant Ecology 157:129–149.
Toyra, J., A. Pietroniro, L. W. Martz, and T. D. Prowse. 2002. A multisensor approach to wetland flood monitoring. Hydrological Processes 16:1569–1581.
U.S. EPA 2003. Section 404 of the Clean Water Act: how wetlands are defined and identified. http://www.epa.gov/OWOW/wetlands/facts/fact11.html (last updated September 26, 2003).
Venables, W. N. and B. D. Ripley. 1997. Modern Applied Statistics with S-PLUS, second edition. Springer, New York, NY, USA.
Wang, L., J. Lyons, and P. Kanehl. (2001). Impacts of urbanization on stream habitat and fish across multiple spatial scales. Environmental Management 28:255–266.
Western Regional Climate Center. 2002. Historical Climate Information. http://www.wrcc.dri.edu/index.html (last accessed 20 January 2003).
Willard, D. E. 1935. Montana: the Geological Story. The Science Press Printing Company, Lancaster, PA, USA.

Manuscript received 15 October 2004; revisions received 3 November 2005; accepted 6 February 2006.