Journal of Terrestrial Observation

Volume 2, Issue 2 *Spring 2010* *Article 4*

A Flexible Approach to Help Overcome Limitations of Moderate Resolution Satellite Imagery for Mapping Invasive Saltcedar on the Bighorn River, Montana

Scott L. Powell, Rick L. Lawrence, Christine Sommers Austin, and Shana Wood

 ISSN 1946-1143.

A Flexible Approach to Help Overcome Limitations of Moderate Resolution Satellite Imagery for Mapping Invasive Saltcedar on the Bighorn River, Montana

Scott L. Powell, Rick L. Lawrence, Christine Sommers Austin, and Shana Wood

Montana State University

ABSTRACT

Saltcedar (Tamarix spp.) is an invasive shrub found throughout the northern Great Plains, including along the Bighorn River in southeastern Montana. Extensive saltcedar infestations have the potential to dramatically alter the species composition and hydrology of riparian ecosystems. The need for inexpensive and large-area saltcedar mapping necessitates the use of moderate resolution satellite imagery. Flexible management expectations along with flexible mapping products are important for overcoming data resolution limitations. We demonstrate an approach for mapping saltcedar along 50-km of the Bighorn River in Montana using ASTER (Advanced Spaceborne Thermal Emission and Reflectance Radiometer) imagery and a Random Forests classification tree modeling approach. We modeled the continuous probability of saltcedar presence and evaluated optimal threshold levels in terms of omission and commission error. Reasonable classification accuracy was achieved for some management purposes. The threshold that optimally balanced omission and commission error yielded an overall classification accuracy of 75%. The flexibility in our approach enables land managers to shift the balance between overmapping and undermapping saltcedar occurrence, depending upon management needs and budgets.

ACKNOWLEDGMENTS

This research was funded by a grant from the Upper Midwest Aerospace Consortium. The authors would like to thank Wyatt Cross and Dave McWethy of Montana State University for their fine boatmanship.

INTRODUCTION

Saltcedar (*Tamarix* spp.) is an invasive shrub from Europe and Asia that occurs widely across the western United States and directly competes with cottonwood (*Populus* spp.), willow (*Salix* spp.), and other riparian vegetation for water and nu-

trient resources (Frasier and Johnson, 1991). Extensive saltcedar infestations have the potential to radically alter the habitat structure (Hunter et al., 1988), biodiversity (Hughs, 1993), and hydrology (Brotherson and Winkel, 1986) of riparian ecosystems. In the United States, saltcedar is widely distributed in the southwestern states and has long been established in the northern Great Plains, including eastern Montana and the Bighorn River (Pearce and Smith, 2007). Saltcedar produces small white or pink flowers in the spring and summer and the foliage turns yellow to orange in the late summer to early fall in Montana.

Mapping the distribution of saltcedar is an important yet elusive resource management objective. Traditional methods for mapping invasive species, including field and aerial surveys, tend to be expensive, difficult to repeat, and more suitable for smaller areas (Cooksey and Sheley, 1998). Aerial photo mapping approaches have successfully distinguished saltcedar from surrounding vegetation during certain seasons but are costly for large area coverage (Everitt and Deloach, 1990; Ge et al., 2006). A number of saltcedar mapping applications have emerged more recently that rely on high spatial resolution (Akasheh et al., 2008) or hyperspectral (Hamada et al., 2007) remote sensing methods. These applications, while promising, also tend to be hampered by data availability, cost, and areal coverage. Moderate-resolution satellite remote sensing offers promise, despite known limitations, for some invasive species mapping applications due to the pressing need for inexpensive, repeatable, and large-area monitoring efforts. Published applications of moderate-resolution imagery for saltcedar mapping in particular, however, are few. Saltcedar was mapped in Colorado using multispectral Landsat imagery, and it was found that winter leaf-off imagery aided classification (Groeneveld and Watson, 2008). Landsat TM imagery and ancillary GIS data have also been used to map saltcedar distribution in Nevada (Sengupta et al., 2005). Saltcedar defoliation by beetles has also been mapped in Utah using ASTER (Advanced Spaceborne Thermal Emission and Reflectance Radiometer) imagery (Dennison et al., 2009).

Early detection of saltcedar invasion, before it attains dominance in an ecosystem, is critical for management purposes. This heightens the challenge of accurate mapping with moderate resolution imagery because of the sporadic and heterogeneous nature of early saltcedar invasion patterns. Resource managers must have flexible expectations to address this challenge and recognize that some forms of classification error might be more tolerable than others. A flexible map product might allow for such shifting management objectives. Reasonable levels of commission error, or overmapping of saltcedar, might be a more acceptable type of error than omission error, or missing saltcedar patches altogether, for some management needs. For example, a land manager might be willing to travel to some locations where saltcedar was mapped but is not present as the cost of reducing the likelihood of missing infestations. A flexible map can tilt the balance between the two forms of error if, conversely, more certainty is required for field visitation or eradication purposes. The resource management expectations, therefore, are extremely important in helping to ameliorate the limitations of mapping with moderate resolution

imagery. Our objective in this study was to demonstrate a straightforward approach for flexible mapping of saltcedar occurrence along the Bighorn River in Montana using moderate-resolution ASTER satellite imagery and Random Forests statistical modeling. We chose to use ASTER imagery because it has better spectral and spatial resolution than Landsat imagery and, therefore, might be better suited for mapping small patches of saltcedar. ASTER imagery has been widely used for ecological applications and has demonstrated similar classification results to Landsat imagery (Powell et al., 2007). Random Forests regression trees were selected as the modeling approach because they are non-parametric, are robust to over-fitting, obviate the need for separate validation data, and are accurate for a wide variety of applications (Cutler et al., 2007).

STUDY AREA

The study area is an approximately 50-km length of the Bighorn River in southeastern Montana near the town of Hardin (Figure 1). This low-gradient, largely agricultural stretch of the Bighorn River begins on the Crow Indian Reservation downstream of the Bighorn Canyon National Recreation Area and ends short of the confluence with the Yellowstone River. We constrained the study area to a 500-m buffer zone on either side of the river to approximate the extent of the field observations and the expected distribution of saltcedar. The riparian vegetation along this length of the Bighorn River, in addition to saltcedar (*Tamarix* spp.), is dominated by scattered stands of cottonwood (*Populus* spp.), Russian olive (*Elaeagnus angustifolia*), willow (*Salix* spp.), and sagebrush (*Artemesia* spp.), interspersed with a variety of grasses, shrubs, agricultural crops, and rangeland.

Methods

Field data were collected via canoe during the summer of 2006 and field verified in the fall of 2008. A total of 72 saltcedar stands were recorded by location, stand diameter, estimated canopy coverage, growth stage (sapling, pole, or mature), and phenology (flowering or non-flowering). Other vegetation types were recorded at 170 locations, including agriculture, cottonwood, grass, Russian olive, sagebrush, rabbitbrush (*Chrysothamnus viscidiflorus)*, willow, and bare soil. We identified an additional 122 observations of water, conifer, fallow-agriculture, and green-agriculture using high-resolution digital aerial photos. We consolidated all reference data into six classes: saltcedar (n=72), forest (n=121), grass (n=50), bare/fallow (n=37), shrub (n=40), and water (n=44).

The satellite imagery that we used for this study consisted of two ASTER images from August 19, 2006. The adjacent ASTER images were acquired with Level 1B processing (registered radiance at sensor) and then were geometrically corrected using Digital Orthophoto Quarter Quads (DOQQs). For the northerly image we used 22 ground control points (GCPs) and for the southerly image we used 18 GCPs. For both images, first order polynomials with RMSE < 0.5 pixels were calculated. The images were then mosaiced together and clipped using a 500-m buffer on either

side of the river. We extracted for each reference plot location the spectral values for each of two 15-m visible bands (green and red), one 15-m near-infrared band (3N – nadir), and six 30-m shortwave-infrared bands (Table 1).

Table 1. ASTER band specifications.

Sub-System	Band Number	Spectral Range (μm)	Spatial Resolution
visible and near-infrared (VNIR)	1 2 3N (nadir) 3B (backward)	0.52-0.60 0.63-0.69 0.78-0.86 0.78-0.86	15 m
shortwave-infrared (SWIR)	4 5 6 7 8 9	1.60-1.70 2.15-2.19 2.19-2.23 2.24-2.29 2.30-2.37 2.36-2.43	30 m
thermal infrared (TIR)	10 11 12 13 14	8.13-8.48 8.48-8.83 8.93-9.28 10.25-10.95 10.95-11.65	90 m

We opted not to use the 90-m thermal bands due to spatial resolution constraints with the reference data. We then plotted the reference data in two-dimensional spectral space to visually examine the spectral separability of saltcedar with respect to the other vegetation types.

We modeled the distribution of saltcedar as a binary presence/absence response variable (saltcedar presence = 1 [n=72] versus saltcedar absence = 0 [n=292]), using Random Forests (Breiman, 2001) classification trees and "out-of-bag" (OOB) error estimation, implemented through the R package ModelMap (Freeman, 2009; Freeman et al., in review). Random Forests is an increasingly popular statistical modeling approach for ecological applications (Cutler et al., 2007), including invasive species classification and mapping (Lawrence et al., 2006). Random Forests is a non-parametric ensemble modeling approach that constructs numerous small classification trees that vote on predictions, and is considered to be robust to overfitting (Breiman, 2001). Each tree is constructed with a random subset of the data, and the remaining data are used for OOB error estimation. OOB estimation has been shown to be unbiased and practical for eliminating the need for independent validation data (Lawrence et al., 2006). Classification error was evaluated with the aid of the PresenceAbsence package in R (Freeman, 2007). The relative importance of each of the potential predictor variables was assessed in terms of the mean decrease in accuracy with and without a given variable.

Random Forests and PresenceAbsence provide information with respect to the probability that an observation will be a member of a target class. In the case of a binary classification the output is similar to that of a logistic regression. In this

case, we estimated the probability of a pixel being saltcedar. This enabled us to examine the continuum of probability thresholds between 0 and 100% for saltcedar presence to evaluate the error tradeoffs. We evaluated a suite of four threshold optimization criteria, including several premised on the notions of model sensitivity and specificity. In this context, sensitivity is akin to the model's ability to correctly predict observed presence, or put another way, it is equivalent to the inverse of saltcedar omission error. Specificity is akin to the model's ability to correctly predict observed absence, or equivalent to the inverse of absence omission error. The four optimization criteria identified the threshold: (1) that maximized the sum of sensitivity and specificity (MaxSens+Spec), (2) where sensitivity equaled specificity (Sens=Spec), (3) that maximized overall accuracy (Max Overall), and (4) that maximized K_{hat} (Max K_{hat}).

Based upon these four criteria, we selected an optimum threshold for further error evaluation. We reclassified the continuous predictions into a binary saltcedar presence/absence map using the 25% probability threshold, and calculated omission and commission error rates, as well as overall classification accuracy and K_{hat}.

RESULTS AND DISCUSSION

An examination of the reference data in two-dimensional spectral space plots revealed significant spectral overlap between saltcedar and other vegetation types, especially shrub and to a lesser extent forest (Figure 2). There was considerably less spectral overlap between saltcedar and water, grass, and bare soil/fallow. The variable importance plot derived from the Random Forests model indicated that the two 15-m visible bands (green and red) were the most influential predictor variables, according to the relative change in accuracy from variable permutations (Figure 3). This was likely attributable to better spectral separability in the visible bands, perhaps as a result of higher spatial resolution in these wavelengths.

Analysis of the continuum of probability thresholds between 0 and 100% for saltcedar presence revealed significant tradeoffs in error evaluation (Figure 4). Based on the four threshold optimization criteria that we examined, the optimal threshold was between 15 and 45% (Table 2).

Table 2. Optimization criteria and probability thresholds, with associated errors.

Optimization Criteria	**Threshold**	**Overall Accuracy**	**Saltcedar Omission Error**	**Absence Omission Error**	**K_{hat}**
MaxSens+Spec	15%	71%	7%	34%	39%
Sens=Spec	25%	75%	24%	26%	39%
Max K_{hat}	45%	82%	53%	9%	40%
Max Overall	45%	82%	53%	9%	40%

Overall accuracy increased as the threshold level increased. Higher thresholds, however, equated to higher saltcedar omission error, meaning that saltcedar would likely be undermapped at these thresholds. The lower thresholds, with lower saltcedar omission error, conversely, had higher absence omission error, implying

that saltcedar would likely be overmapped at these thresholds. K_{hat} changed little by threshold level (between 15 and 45%), reflecting the shortcomings of this statistic for binary classifications (since predicted chance agreement can be high in these classifications).

The 25% probability threshold was identified based upon the optimization criterion where sensitivity equaled specificity (Figure 4). This criterion balances the competing demands of presence and absence omission errors. The saltcedar omission error at this threshold (24%) was nearly equivalent to the absence omission error (26%), and the overall classification accuracy was 75% (Table 3).

Table 3. Classification accuracies for the 25% probability threshold for saltcedar presence.

25% Probability Threshold		**Predicted**		
		absence	saltcedar	total
Observed	absence	217	75	292
	saltcedar	17	55	72
	total	234	130	364
saltcedar omission error = 24%; absence omission error = 26% saltcedar commission error = 58%; absence commission error = 7% overall accuracy = 75% K_{hat}= 39%				

We produced a saltcedar continuous probability map, with values ranging between 0 and 1, and saltcedar presence maps derived from the 15%, 25%, and 45% probability thresholds (Figure 5), which we overlaid on ASTER false-color composite images.

Our results demonstrate that despite the many limitations to mapping saltcedar with moderate resolution satellite imagery, reasonable accuracy can be achieved depending on management needs. The flexibility in the thresholding approach from an invasive species management perspective is critical for adapting to changing budgets and management needs. A lower probability threshold will optimize for lower saltcedar omission error if, for example, the management goal is to produce an accurate map of saltcedar presence given a dataset of field observations. At the 15% probability threshold level, we mapped 4,377 ha of saltcedar infestation (Figure 5). A higher probability threshold will optimize for lower saltcedar commission error if, conversely, the management goal is to use a map in the field for saltcedar eradication purposes, and the budget does not allow for numerous false positive field visits. At the 45% probability threshold level, we mapped 1,247 ha of saltcedar infestation (Figure 5). At the 25% probability threshold level, we mapped 3,015 ha of saltcedar infestation. At this level, saltcedar omission error was relatively low (24%), but the saltcedar commission error was quite high (58%), suggesting that over half of our mapped predictions were inaccurate. Higher specificity, or absence omission error, could be obtained with a higher probability threshold, but this would be at the expense of sensitivity, or saltcedar omission error. The permissibility of these types of errors depends largely upon the management strategy.

The accuracy of our results underscores the challenge of mapping saltcedar with moderate-resolution remote sensing data. Saltcedar is a relatively minor component of the overall Bighorn River riparian vegetation; therefore the stands that were observed in the field were generally small and interspersed with other vegetation types. The majority of field observed saltcedar stands were not only mixed stands, but spatially smaller than single 15-m VNIR or 30-m SWIR ASTER pixels, compounding the challenge of accurate co-location of the reference plots with the image data. Moreover, the saltcedar observations were more likely than not mixed pixels, for which we had no percent composition information. Finally, our saltcedar observations were spectrally indistinguishable from some of the other riparian vegetation found along the Bighorn River, namely, willow, further compounding our classification challenge.

Reasonable accuracy was achieved for certain management purposes using a simple presence/absence Random Forests statistical model despite these known limitations. There are several potential improvements that could be made to likely yield higher prediction accuracy. A logical improvement, based on the above discussion, would be to use a sub-pixel mapping approach to reconcile the patchy distribution of saltcedar with moderate spatial resolution imagery. Another potential improvement would be to use temporally suitable fall satellite imagery when saltcedar foliage is more spectrally distinct from surrounding vegetation (Everitt and Deloach, 1990). Finally, the inclusion of ancillary GIS predictor variables such as soil type and distance from the river might yield additional predictive power.

REFERENCES

Akasheh, O.Z., C.M.U. Neale, and H. Jayanthi. 2008. Detailed mapping of riparian vegetation in the middle Rio Grande River using high resolution mulit-spectral airborne remote sensing. *Journal of Arid Environments* 72:1734-1744.

Breiman, L. 2001. Random forests. *Machine Learning* 45:5-32.

Brotherson, J.D., and V. Winkel. 1986. Habitat relationships of saltcedar (*Tamarix ramosissima*) in Central Utah. *Great Basin Naturalist* 46:535-541.

Cooksey, D., and R. L. Sheley. 1998. Montana noxious weed survey and mapping system. MontGuide 9613, Montana State University Extension Service, Bozeman, MT.

Cutler, D.R., T.C. Edwards, K.H. Beard, A. Cutler, and K.T. Hess. 2007. Random forests for classification in ecology. *Ecology* 88:2783-2792.

Dennison, P.E., P.L. Nagler, K.R. Hultine, E.P. Glenn, and J.R. Ehleringer. 2009. Remote monitoring of tamarisk defoliation and evapotranspiration following saltcedar leaf beetle attack. *Remote Sensing of Environment* 113:1462-1472.

Everitt, J.H., and C.J. Deloach. 1990. Remote Sensing of Chinese Tamarisk (*Tamarix chinensis*) and associated vegetation. *Weed Science* 38:273-278.

Frasier, G. W. and T. N. Johnson, Jr. 1991. Saltcedar (*Tamarix*): Classification, distribution, ecology, and control. *In* L. F. James, J. O. Evans, M. H. Ralphs,

and R. D. Child, eds. *Noxious range weeds,* pp. 377-386. Westview Press, Boulder, Colorado.

Freeman, E.A. 2007. PresenceAbsence: An R package for presence-absence model evaluation. USDA Forest Service, Rocky Mountain Research Station, 507 25th St., Ogden, Utah. Eafreeman@fs.fed.us, URL:http://cran.rproject.org/web/packages/PresenceAbsence/index.html.

Freeman, E.A. 2009. ModelMap: An R package for modeling and map production using Random Forest and Stochastic Gradient Boosting. USDA Forest Service, Rocky Mountain Research Station, 507 25th St., Ogden, Utah. Eafreeman@fs.fed.us, URL:http://cran.rproject.org/web/packages/ModelMap/index.html

Freeman, E.A., T.S. Frescino, and G.G. Moisen. In review. ModelMap: An R package for model creation and map production. *Journal of Statistical Software*.

Ge, S., R. Carruthers, P. Gong, and A. Herrera. 2006. Texture analysis for mapping *Tamarix Parviflora* using aerial photographs along the Cache Creek, California. *Environmental Monitoring and Assessment* 114:65-83.

Groeneveld, D.P., and R.P. Watson. 2008. Near-infrared discrimination of leafless saltcedar in wintertime Landsat TM. *International Journal of Remote Sensing* 29:3577-3588.

Hamada, Y., D.A. Stow, L.L. Coulter, J.C. Jafolla, and L.W. Hendricks. 2007. Detecting tamarisk species (*Tamarix* spp.) in riparian habitats of Southern California using high spatial resolution hyperspectral imagery. *Remote Sensing of Environment* 109:237-248.

Hughs, L.E. 1993. The devils own: Tamarisk. *Rangelands* 15:151-155.

Hunter, W.C., R.D. Ohmart, and B.W. Anderson. 1988. Use of exotic saltcedar (*Tamarix chinensis*) by birds in arid riparian systems. *The Condor* 90:113-123.

Lawrence, R.L., S.D. Wood, and R.L. Sheley. 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). *Remote Sensing of Environment* 100:356-362.

Pearce, C.M., and D.G. Smith. 2007. Invasive saltcedar (*Tamarix*): Its spread from the American southwest to the northern Great Plains. *Physical Geography* 28:507-530.

Powell, S.L., D. Pflugmacher, A.A. Kirschbaum, Y. Kim, and W.B. Cohen. 2007. Moderate resolution remote sensing alternatives: A review of Landsat-like sensors and their applications. *Journal of Applied Remote Sensing*, Volume 1, doi: 10.1117/1.2819342.

Sengupta, D., C. Geraci, S. Kolkowitz, Y. Komandyan, and K. Cheng. 2005. Assessing tamarisk in Nevada by combining field and remote sensing techniques. ASPRS 2005 Annual Conference. March 7-11, 2005. Baltimore, Maryland.

Figure 1. Location of the 50-km length study area of the Bighorn River in relation to the state of Montana.

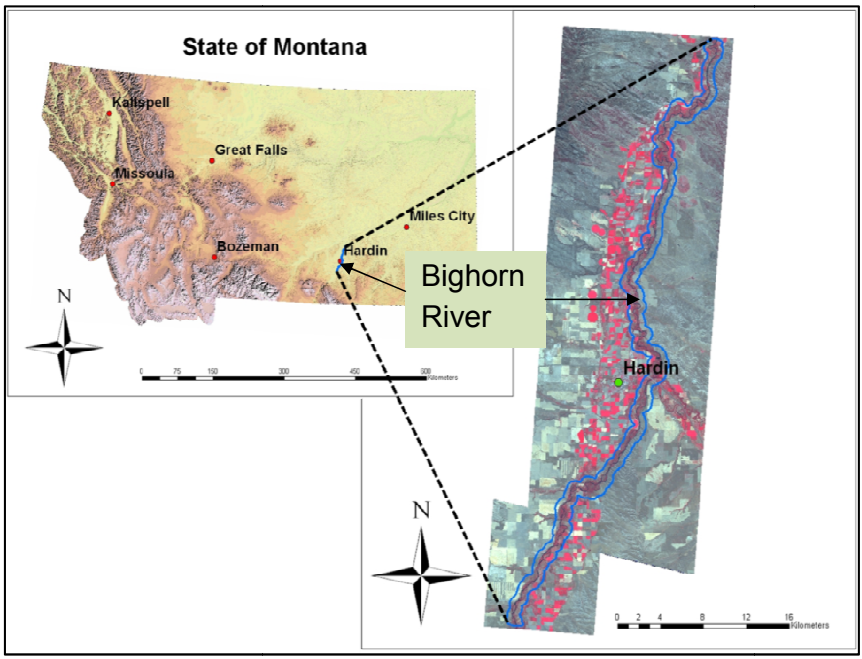

Figure 2. Reference data plotted in two-dimensional spectral space. The top graph plots the six reference classes in SWIR v. Red spectral space, and the bottom graph plots the six reference classes in N-IR v. Red spectral space.

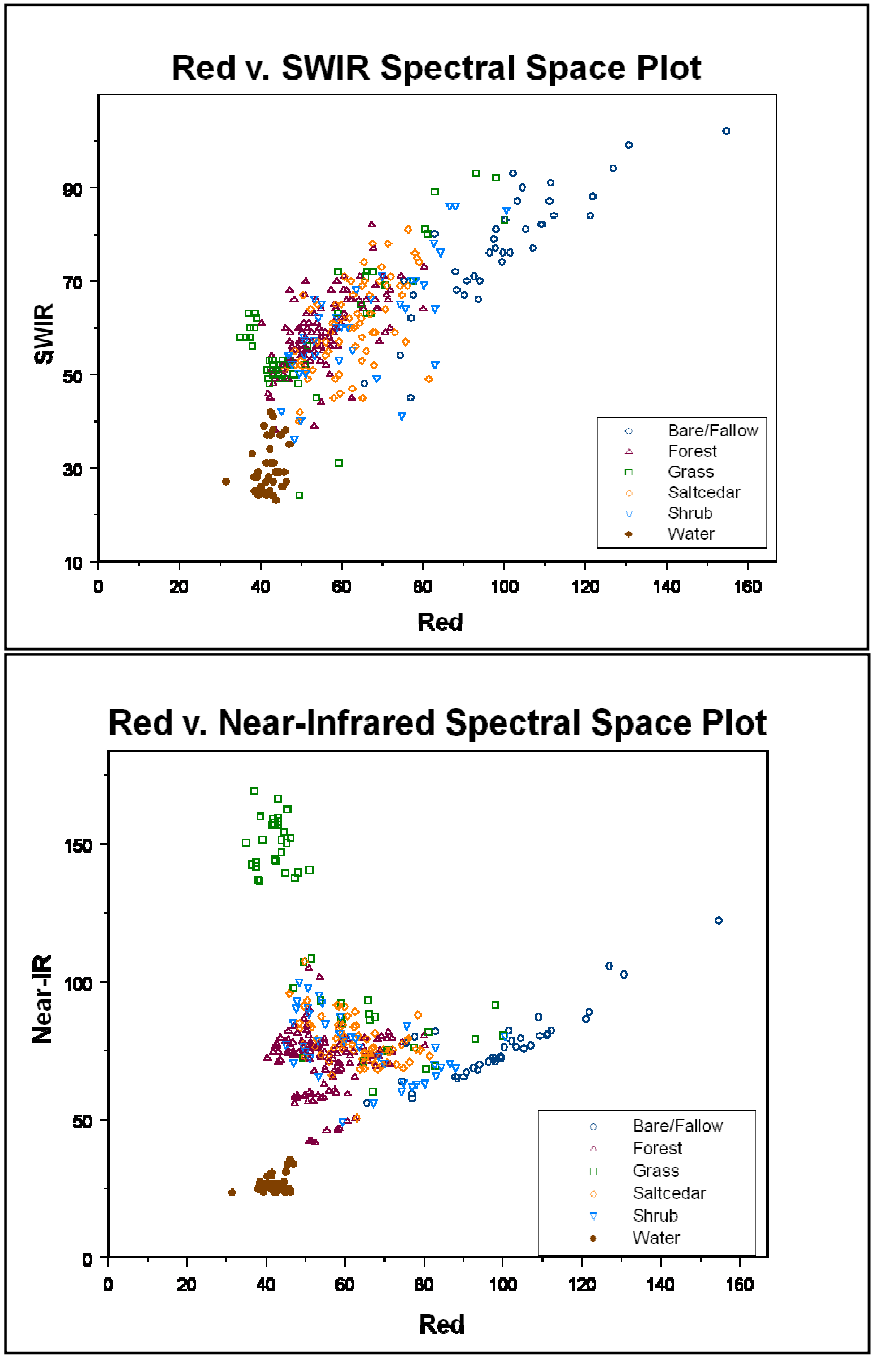

Figure 3. Variable importance plot from the Random Forests model.

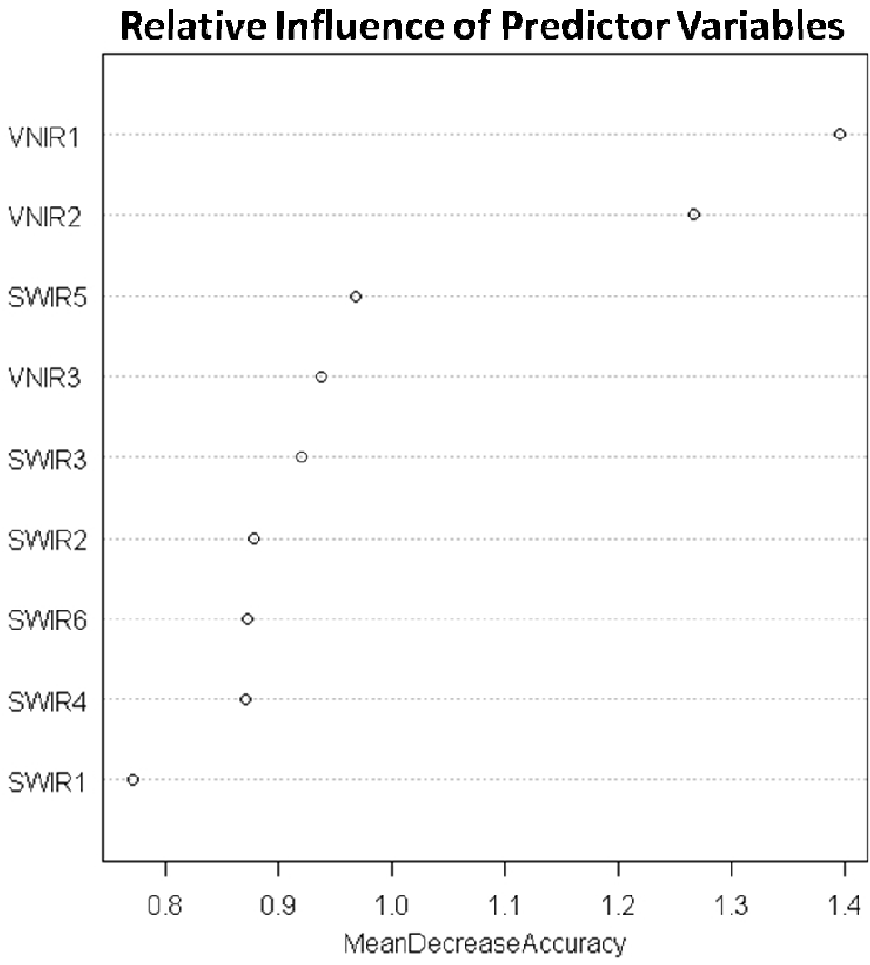

Figure 4. Graph of the tradeoffs in sensitivity, specificity, and K_{hat} along the continuum of probability threshold levels.

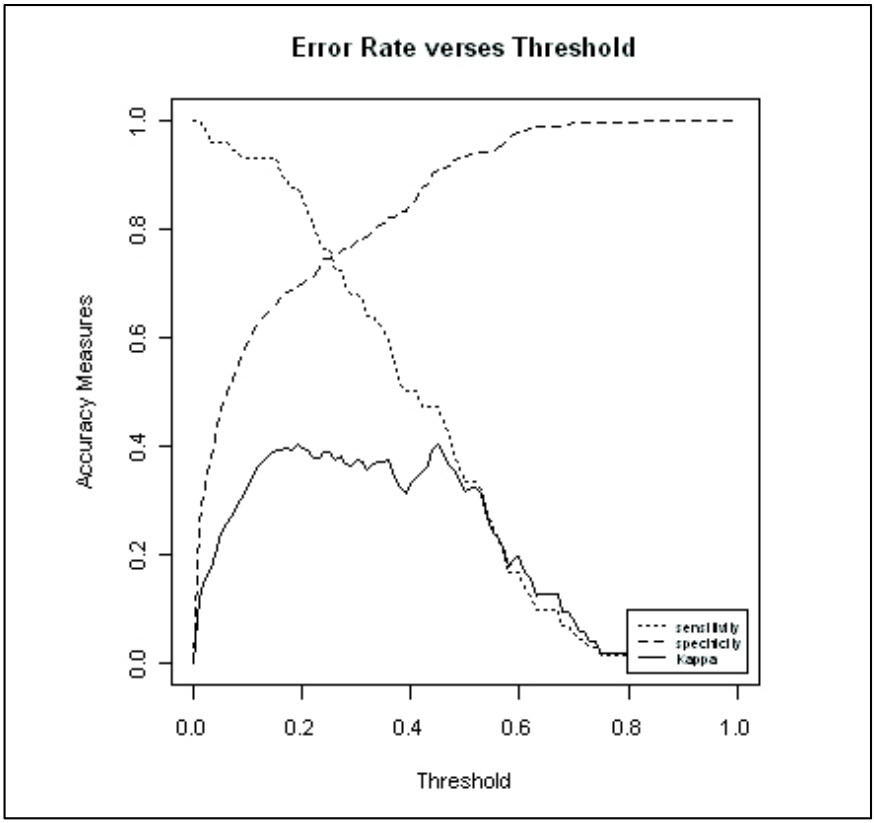

Figure 5. Final maps of (A) the continuous probability of saltcedar presence along the Bighorn River, scaled between 0 (Low) to 1 (High); B) the saltcedar presence map derived from the 25% probability threshold; C) the saltcedar presence map derived from the 15% probability threshold; and D) the saltcedar presence map derived from the 45% probability threshold.